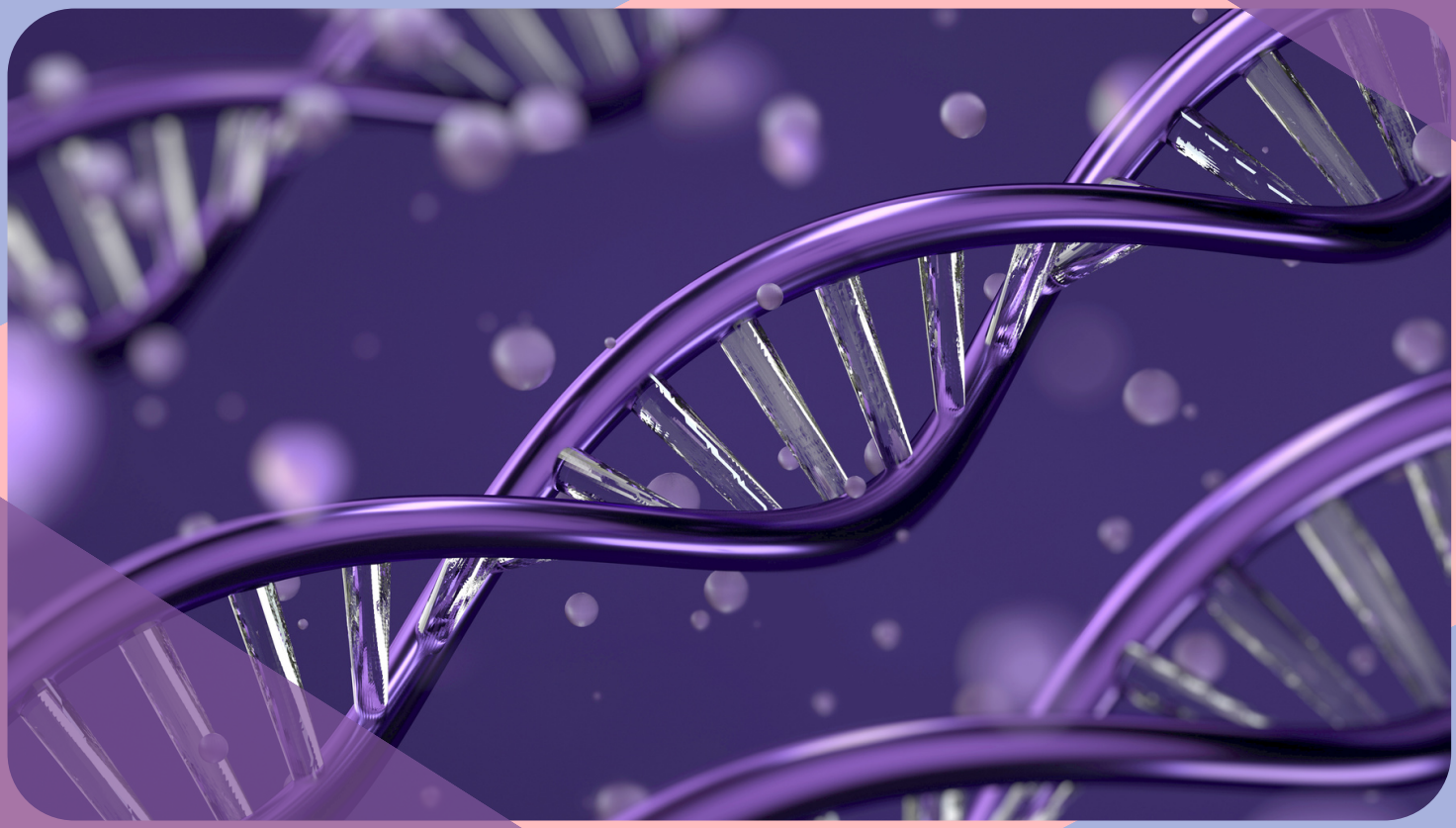
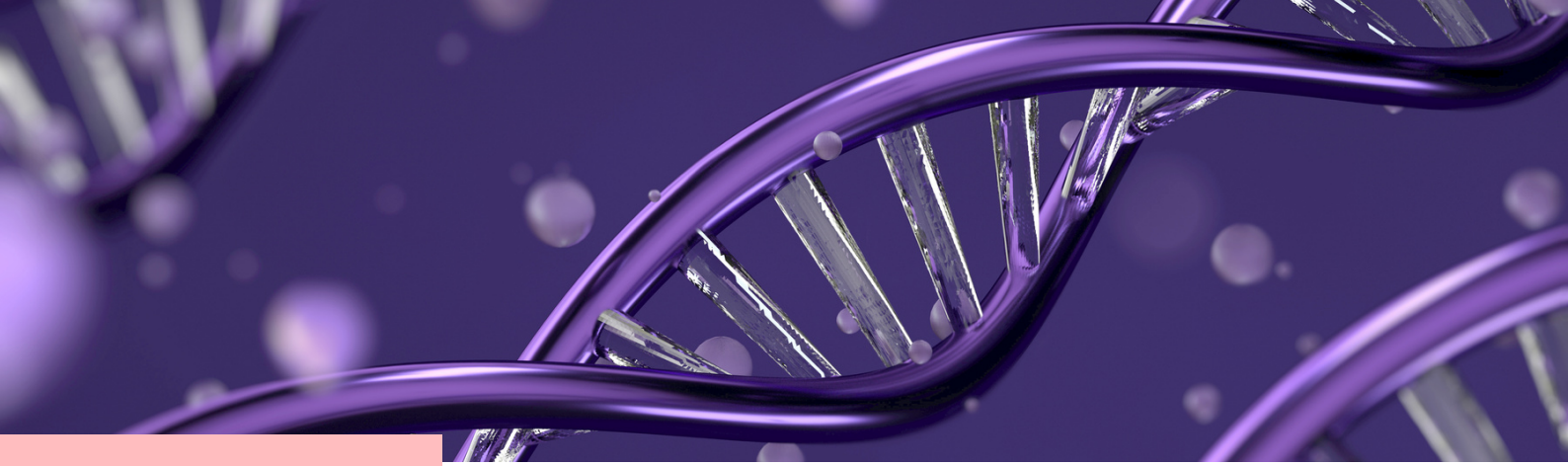
BioTechniques LEARN



Lab essentials:
DNA extraction methods

fsg

www.biotechniques.com



ABOUT

This eBook explores DNA extraction methods, with key content from *BioTechniques*. *BioTechniques* is the international journal of life sciences, publishing cutting-edge lab-based methodologies and techniques. The partnering site is an online platform uniting a global community of lab researchers, industry experts and budding scientists, with the latest news, features and exclusive interviews.

CONTENTS

Ten tips for high-throughput DNA sequencing

BY EMMA WATERS, QUADRAM INSTITUTE (NORWICH, UK)

DNA sequencing is a revolutionary technology that enables scientists to decipher the life-giving instructions contained within the genetic code of all organisms. At its core, DNA sequencing is the process of determining the precise order of nucleotide bases (adenine, cytosine, guanine and thymine) within a DNA molecule. This information provides a blueprint of the genetic instructions encoded in the DNA, allowing scientists to explore the structure, function and variability of genes and genomes.

To assist researchers in harnessing the full potential of high-throughput DNA sequencing, here's a top ten tip guide.

Introduction

DNA sequencing has transformed the field of biology. In medical microbiology, it has significantly advanced our understanding of genetic diseases, enabling the identification of disease-causing mutations and the development of personalized treatments. In evolutionary biology, it has shed light on the history and relationships between species, providing insights into the mechanisms of adaptation and speciation. Moreover, DNA sequencing has opened new frontiers in fields like forensic science, agriculture, conservation biology and microbial ecology, to name just a few.

The history of DNA sequencing is a remarkable journey that spans the last few decades. It all began in the 1970s when Frederick Sanger pioneered the first short-read DNA sequencing technique. This technique was used to sequence the first full genome of a virus, which was around 5 thousand bases long. In the 1980s, efforts were made to automate Sanger sequencing, which greatly increased the speed and efficiency, enabling the sequencing of larger genomes. These advancements allowed the launch of the Human Genome Project in 1990, where a group of scientists wanted to determine the 3 billion bases of the human genome, which was only achieved 13 years later! During that time, other scientists sequenced the genomes of different organisms, celebrating each time they sequenced a bigger genome; from 1995 when the first bacterial genome was sequenced, containing a couple of million bases [1], to 5 years later when a fly genome was sequenced, which contained a couple of hundred million bases [2].

Over this period, sequencing and computer power continued to improve, allowing for massive parallel sequencing of DNA fragments and thus the introduction of next-generation sequencing (NGS) technologies. These advancements – made only 7 years after the original Human Genome Project ended – facilitated the sequencing of thousands of human genomes and marked the start of understanding genetic diseases. In 2014, Public Health England started to use short-read sequencing at a national scale to

Ten tips for high-throughput DNA sequencing

manage infectious diseases, particularly for tuberculosis and bacterial foodborne diseases, as this type of sequencing is highly accurate and can rapidly identify single nucleotide polymorphisms (SNPs) [3]. However, short-read sequencing, which sequences DNA fragments of hundreds of bases in length, is unable to resolve long-repeat sequences found in genomes that can span over several thousand bases.

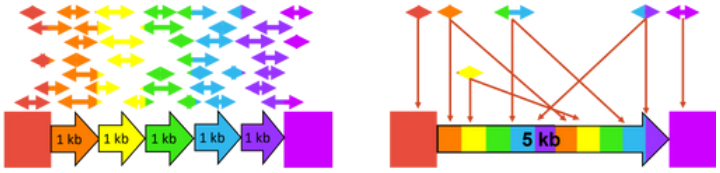


Figure 1: (Left) Short-reads of 100-1000 bases in length, mapped onto a section of genome containing easy-to-resolve genes. (Right) The mapping of short-reads is confused by sections of genomes that contain long-repeat sequences, like those of ribosomal operons found in all living organisms.

In 2009, the introduction of third-generation sequencing technologies brought about a further breakthrough in DNA sequencing [4]. These long-read sequencing techniques can sequence much longer DNA fragments (thousands to millions of bases long) and assemble genomes with increased accuracy whilst identifying large structural mutations where large genomic fragments rearrange within genomes and can shuffle the order of genetic instructions [5].

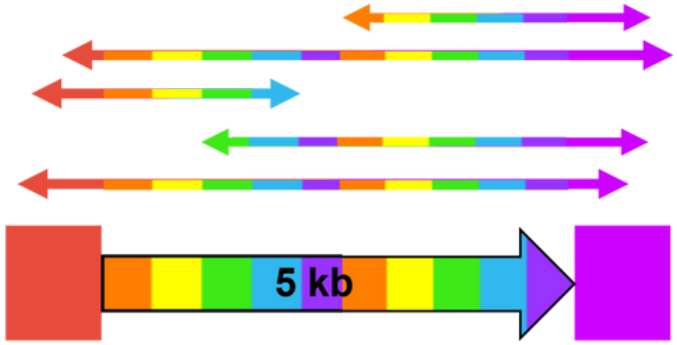


Figure 2: Long-read sequences mapped onto and resolving a section of a genome, which contains long-repeat sequences.

In this era of genomics, DNA sequencing continues to evolve rapidly. Sequencing technologies and their associated library preparation have become more accurate, faster, cost-effective and portable, leading to the potential for DNA sequencing to be performed on anything, by anyone and anywhere.

Tip 1: Choose the appropriate sequencing platform

Consider desired requirements like read length, error rate and data output to match your needs. Generally, if you are interested in SNPs and not worried about complete assemblies, short-read platforms like Illumina's MiSeq, NextSeq and NovaSeq offer high-throughput capabilities. However, if you are interested in structural variation, you will require long-read sequencing capabilities such as PacBio's Sequel II or Revio systems, or Oxford Nanopore's MinION or PromethION. You could also perform both short- and long-read sequencing to generate the gold standard, hybrid assemblies.

Ten tips for high-throughput DNA sequencing

Tip 2: Optimize DNA extraction

Optimize sample preparation and DNA extraction for your desired sequencing method. If you are performing long-read sequencing, use appropriate high-molecular-weight DNA extraction protocols to minimize DNA shearing and degradation, as fragmented DNA can impact long-read sequencing performance. Quality control measures, such as checking DNA integrity via gel electrophoresis or fluorometric quantification, should be implemented to ensure the samples meet the sequencing requirements.

Tip 3: Choose the appropriate library preparation

Follow optimized library preparation protocols specific to your chosen sequencing platform. These protocols generally follow the same outlines: DNA fragmentation and size selection (which might be omitted in long-read methods), end repair, barcode ligation for multiplexing, adapter ligation, amplification and library clean-up steps. There may also be size selection steps to enrich long fragments and remove shorter ones that may affect sequencing accuracy.

Tip 4: Multiplexing for simultaneous sequencing of multiple samples

Some library preparation kits come with a selection of unique barcodes, which allow multiple samples to be sequenced simultaneously, increasing throughput and cost effectiveness. The barcode ensures that you can identify and separate individual samples during downstream data analysis. If you do use multiplexing in your sequencing, use sample plate maps and multichannel pipettes to ensure samples are allocated the correct barcodes and throughput is increased. Also ensure samples are pooled to provide equal representation of each one, whilst remembering to consider desired data output and genome size if these are different between samples.

Tip 5: Library quality control

Before loading prepared libraries on the sequencing platform do a final library quality control check. You do not want to waste an expensive sequencing flowcell on a poor, failed library! Assess library quality using gel electrophoresis and fluorometric quantification, or similar methods, to confirm that library size distribution, quality and quantity are roughly correct. Monitor the performance of the sequencer using control samples or PhiX spike-ins to ensure optimal sequencing conditions.

Tip 6: Determine desired sequencing depth

Determine the appropriate sequencing depth for your sample. Consider factors such as genome size, complexity and the level of sensitivity required for variant detection, and consult published literature to help you determine an appropriate sequencing depth. Once you have reached the desired sequencing depth for your library, any data past this point might not be useful. Instead, you may be able to wash the current library off and then reuse the flowcell

Ten tips for high-throughput DNA sequencing

for a different sample, to further increase cost effectiveness.

Tip 7: Bioinformatic analysis pipelines
Develop and adopt an appropriate, automated bioinformatics pipeline for data analysis. This could include tools to perform read quality assessment and filtering, assembly and polishing, assembly quality checks, variant calling and other downstream analyses to extract meaningful insights from the raw data. Remember to stay up to date with the latest versions of bioinformatics tools and methodologies relevant to your sequencing platform and analysis.

Tip 8: Data storage and backup
DNA sequencing generates vast amounts of data. Ensure you have adequate storage capacity and backup systems to handle and protect the generated data. Implement a reliable data management strategy to organize and store the sequencing data efficiently, facilitating data access and analysis.

Tip 9: Stay up to date
DNA sequencing and subsequent analysis are continuously evolving. Try to keep up to date with advancements in sequencing technologies, protocols and data analysis tools. Regularly review scientific literature, attend conferences and workshops, and engage with the scientific community to stay informed about the latest techniques and best practices.

Tip 10: Automate where possible
Many sequencing technologies have the capability to simultaneously sequence hundreds of samples at once. Now, the rate-limiting steps are generally in the preparation of these samples for sequencing and the subsequent data analysis. There are automated systems for DNA extraction, quality control and library preparation that can speed up the wet lab side of this. There are also ways to automate data backup and analysis via the command line or open, web-based platforms like Galaxy. You may even be able to outsource the entire process to external DNA sequencing companies who will take your extracted DNA (and money) and in return will give you processed sequencing data. This may be beneficial for some projects; however, you may generally have less control and will still have to consider the points covered above.

P.S. Final tip: never load a library at the end of the working day… Something always goes wrong, and it is always better to have friendly support around when this happens.

Ten tips for high-throughput DNA sequencing

References

1. Fleischmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223) 496–512 (1995).
2. Adams MD, Celniker SE, Holt RA *et al.*The genome sequence of *Drosophila melanogaster. Science* 287(5461), 2185–2195 (2000).
3. UK Health Security Agency. A revolution in genomic sequencing. [Date accessed: 11th July 2023].
4. Hayden EC. Genome sequencing: the third generation. *Nature* doi: 10.1038/news.2009.86 (2009) (Epub ahead of print).
5. Waters EV, Tucker LA, Ahmed JK, Wain J and Langridge GC. Impact of *Salmonella* genome rearrangement on gene expression. *Evol. Lett.* 6(6) 426–437 (2022).

About the Author

Emma Waters (right) is a Postdoctoral Research Scientist at the Quadram Institute (Norwich, UK). Emma originally trained as a Chemist at the University of East Anglia (Norwich, UK), before receiving her PhD in Biochemistry where she investigated how to attach tiny chemical solar panels to bacteria so they could harness solar energy to produce useful chemicals.

Since then, Emma changed her research field to medical microbiology where she uses different sequencing techniques to investigate bacterial niche adaptation. Currently, she uses long-read and RNA sequencing to identify and study the effects of large structural rearrangements in the genomic DNA of bacteria. This type of mutation shuffles genetic instructions to survive new stresses and environments. In the case of pathogenic bacteria, this can aid adaption to asymptomatic lifestyles, silent transmission and detection aviation.

Emma Waters is funded by Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Microbes and Food Safety and has no identified conflicts of interest with the contents of this article.

Interview

BioTechniques

Releasing secrets bound to ancient remains with modern DNA extraction techniques: an interview with Elena Essel

Elena Essel[*,1]

[1]*Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, 04103 Leipzig, Germany; *Author for correspondence: elena_essel@eva.mpg.de*

BioTechniques 75: 43–46 (August 2023) 10.2144/btn-2023-0067

First draft submitted: 5 June 2023; Accepted for publication: 26 July 2023; Published online: 17 August 2023

ABSTRACT

Elena Essel (Msc) spoke to Ebony Torrington, Managing Editor of *BioTechniques*. Essel is a molecular biologist in Matthias Meyer's Advanced DNA Sequencing Techniques group at the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany). Essel studied biology at University of Erlangen-Nuremberg (Erlangen, Germany) for her bachelor's and in Martin-Luther-University Halle-Wittenberg (Halle an der Saale, Germany) for her master's. Essel worked in Meyer's group on DNA extraction of very degraded material for her master's thesis. Meyer is an expert in developing new cutting-edge methods for researching ancient DNA, with a focus on skeletal remains, and more recently on sediment remains. Essel now focusses on DNA sampling and extraction aspects of the pipeline at Meyer's lab for the ancient DNA workflow.

KEYWORDS:

ancient DNA • Denisova • DNA extraction • molecular biology • Paleolithic pendant • temperature-controlled DNA extraction

You recently published an article on ancient human DNA recovered from a Paleolithic pendant [1]. Could you provide a brief overview of this work?

One of the big questions in many archeological sites is, who made a tool or artifact? This is a question that many of our collaborators from the archeological field ask us. When people think about ancient DNA, they often assume skeletal remains, but at many sites there are no skeletal remains of ancient humans, so no clues about who made or used these artifacts or tools.

We took a big step forward with sediment work, we can now go to a site and analyze the sediment for DNA, even if there are no human remains [2–4]. Even in the absence of skeletal remains, at least in some sites, we can detect human DNA in the sediments and can give more insights into who made the artifacts. However, it is indirect proof, so we were really excited about the challenge of directly associating an artifact with the DNA of the person who handled, made or used an artifact. Such questions were driven by our collaborators in the field of archeology.

We started our sediment work by focusing on stone tools because stone is an easy substrate to deal with, being more dense, more solid, more robust, and less likely to break during the extraction process compared to artifacts made of osseous material. The results were disappointing; it was much more challenging than we expected. We realized that we had to switch focus, so we went back to the material that we know best, skeletal remains. In the field of ancient DNA, we have been working with skeletal remains for many years. We know a lot about how DNA is preserved in bone and the chemical interactions between bones or teeth and DNA.

The challenge with bones and teeth is that they are very fragile, and much more porous and brittle than a stone tool. It is crucial to ensure that the integrity of the objects is maintained during the extraction of any DNA- this was where I joined this project. We had to think completely differently because the standard procedure in the ancient DNA workflow involves drilling a little hole into the bone. This was not an option for the bone and tooth artifacts, because you would lose crucial information that is stored on the surface. The first challenge was to develop a protocol that preserved the integrity of the artifacts without altering the surface.

We first used material that was similar to the artifacts, or material that could have been the raw material from those artifacts, to develop the protocol. We applied different DNA extraction techniques and analyzed the surface before and after the extraction to evaluate their effects. We found that one of the methods could extract DNA from objects made from bones and teeth, and also did not alter the surface, so we started to delve into this method.

The method was reported in a previous publication in *BioTechniques* as a pretreatment for decontaminating bone powder [5]. In this method, we drilled a small hole before collecting powder, so we had to tweak the method a bit for these objects, which can range from just a few centimeters up to 10 or 20 cm. After optimizing the method, we arrived at a protocol that allowed us to extract DNA from the objects while preserving their structure and integrity.

How do the extraction techniques differ for different specimens, for example bone vs. artifact?

For the bone material that we often work with, skeletal remains, we are interested in the DNA of the animal or the human being that the remains came from. We usually take a dentist's drill, and we collect a bit of bone powder while drilling, then we dissolve the powder in ethylenediaminetetraacetic acid (EDTA), which is a strong decalcifying agent. EDTA is used in laundry or dishwasher detergents to decalcify the water and protect your washing machine; it actually dissolves calcium.

This is great for bone powder because we want to dissolve the bone matrix to release the DNA into solution, but it is a nightmare for the artifacts because if we use a decalcifying agent, we actually dissolve the informative surface of the bone or tooth artifact, so we had to use a different approach.

From previous experiments and publications, we know that one can use a trick to release the DNA from the bone matrix without actually dissolving the bone matrix [5,8]. The bone matrix consists of about 70% of a mineral called hydroxyapatite, which is mostly calcium, while the DNA backbone contains phosphate. The calcium is positively charged, and the phosphate is negatively charged so DNA binds very tightly to the bone matrix via electrostatic interactions. We can add an excess of free phosphate ions using a phosphate buffer, then these free phosphate ions start to compete for the calcium-binding sites in the hydroxyapatite, pushing out the DNA and replacing the DNA that was formerly bound to the bone matrix. The DNA is then released into solution, so it is accessible for us to purify it, enrich it and work with it. Thus, we can get the DNA out without dissolving the bone matrix, keeping the bone intact.

How important was temperature-controlled extraction of DNA from ancient bones to your research?

Carrying out the phosphate extraction at room temperature did not yield as much DNA as with the EDTA-based approaches, where we dissolved the entire bone matrix. The temperature-controlled DNA extraction that we published in *BioTechniques* was a game changer for this project [2]. We are making use of a very simple physical principle: applying heat increases the movement of the molecules; more heat means more movement, and with more movement, we can push more DNA molecules from the bone matrix into solution. This also allows us to release DNA that is bound tightly or deeply in the bone matrix into solution without dissolving the bone.

Once you have extracted the DNA, how do you analyze it?

We use two different approaches. In the first approach, we sequence everything. For the pendant and similar artifacts, we expect to find at least two sources of DNA: the DNA of the animal that the pendant or the artifact was made from – in this case a cervid (deer) – and contaminant DNA from all the microbes that inhabited the bone after it was deposited. If we are lucky, we also find a third component, the ancient human DNA of the person who handled the object.

In the second approach, we take this DNA 'soup', which contains the cervid DNA, the human DNA, and the microbe DNA, and we specifically enrich the DNA of interest. If we are looking for ancient human DNA, we use so-called probes, which are artificial DNA fragments bound to magnetic beads. Since DNA is complementary, with its double-stranded structure, these single-stranded probes attract the complementary strand in our DNA. As the probes are bound to magnetic beads, we can use a magnet to collect them and wash off all the DNA that we are not interested in. In the case of the pendant, we obviously wanted to look for ancient human DNA, so we used probes containing artificial pieces of human DNA to enrich for the human fraction. This allows us to sequence the human DNA fraction in greater depth compared to what we could do with shotgun sequencing, enabling us to look at population genetic aspects or determine which biological sex the DNA comes from. It would be too costly to sequence the sequences of interest when sequencing all the DNA because the human fraction would be less than a percent.

What temperatures are animal, human, & microbe DNA extracted at?

Microbial DNA is predominantly released in the low-temperature fractions. We think this is because the microbe DNA is mostly surface bound, and it hasn't penetrated too deeply into the bone matrix. We find the animal's DNA in all the temperature fractions, which makes sense because during the life of the organism, there was DNA surrounding the bone in bodily fluids.

Interestingly, human DNA was only extracted in the very high-temperature fractions, above 90°C. As an artifact is handled, skin cells get left behind. For example, sometimes I might chew on my pen, so saliva can be deposited, or you sweat or cut yourself, leaving blood. The bone matrix is porous, it acts like a sponge, it absorbs liquids and carries the DNA with the bodily fluids into the inner parts of the bone matrix of the artifacts. When an object is handled, used a lot or worn in close contact with skin and sweat and saliva, the DNA of that person can penetrate deeply into the artifact. The DNA is then bound, sitting in the matrix, so we need high temperatures to be able to extract the DNA.

That's incredible & it's remarkable how this DNA is still there, even though it's 20,000 years old

It really is incredible! As a methods person I was very excited about the method, but I think the most exciting finding of the paper is that DNA of the person handling an osseous object can preserve over such great time periods. When we started this whole project, we knew that it could be a helpful method for non-destructively extracting DNA from objects, but the chances of finding ancient human DNA were

very slim. We thought that it might not work out and we kept our expectations low to avoid disappointment, so it was really mind-blowing for us to find ancient human DNA.

Forensics people might find it less surprising as they have common techniques for extracting contact DNA. For example, they can actually extract DNA from a fingerprint. For me, it was amazing that this is still possible after 20,000 years. The DNA is still there and there's still enough for in-depth population and genetic analysis.

How long would the DNA take to break down? Could it be there for hundreds or even thousands more years & is there any way of predicting that?

This is a tricky question. I think the preservation mechanism of ancient DNA is likely to be similar for all bone objects. Independent of being an artifact of an unmodified bone or tooth DNA is bound to the bone matrix. Since the DNA preservation mechanism is the same, I think the rules are the same. Whether the DNA comes from the animal itself or from the maker or user, the preservation conditions are important. The oldest DNA that we know from bones was recovered from a 1,000,000-year-old mammoth coming from permafrost. If we found an object that was a million years old under permafrost, I would love to try our DNA extraction method on it.

You mentioned that you work with a lot of archeologists. Did they uncover any other objects that they were interested in you sampling?

We are working closely with our collaborators excavating archeological sites and we do have some samples that we are currently working on. The objects are still super rare, so if you are on a dig for, let's say four weeks, finding five or six objects would be a huge success. And finding the one that was handled intensely enough and preserved the DNA of the person handling feels like winning in the lottery.

You've already mentioned quite a few challenges when it comes to extracting DNA. Were there any other challenges that you faced?

The biggest challenge that we faced was contamination with modern human DNA. We started this project with artifacts from collections that were excavated decades ago, in the 1970s, 80s, and 90s. Back then nobody knew about ancient DNA yet and the problem of leaving modern human DNA on them. So, the objects were excavated without gloves, without any precautions to limit the introduction of modern DNA. They have been heavily studied, so people were handling them intensely.

Even worse, it used to be common to lick a sample to test whether it was bone or stone. If it's stone, the liquid saliva will stay on the surface, but if it's bone, it will penetrate. I was a bit shocked when I first heard this since this is exactly how we think the ancient DNA of the person handing this thing got into the sample, so then the modern human DNA is competing with the DNA that was absorbed 20,000 years ago. The modern DNA is perfectly preserved and there are millions of copies of it: we were just drowning in modern human contamination, so even if there was ancient human DNA, we would have a very hard time detecting it.

How did you separate the ancient human DNA from the modern human DNA?

We can use some characteristics that are typical for ancient DNA. Over time, DNA degrades and, especially on the ends, we have an accumulation of what we call C to T substitutions, whereby cytosine is converted to uracil over time, so that when we sequence the DNA, cytosine shows as a thymine.

We look at the C to T substitution patterns, we actually call it a smiley plot because if you look at your DNA fragment in the interior part, you see the normal frequency of C to T substitutions, while at the outer ends of the plot, you see these elevations, so it really looks like a smiling face. Fragments that carry these C to T substitutions at the end can be considered authentic ancient molecules. DNA fragments without these signals at the ends can be put into the modern category and disregarded.

How long does it take for that degradation to happen?

This is not easy to answer: It really depends on the preservation conditions. You can already see these patterns quite strongly in forensic samples that are only 50 years old, but then also you have samples that are several thousand years old, but come from favorable preservation conditions, such as permafrost, so the degradation rate is a lot slower. You can't put a timestamp on it, but if you see the C to T substitutions then you know you have something that underwent degradation – whether that be 50 or 20,000 years – then you have to take a careful look.

How did you confirm the DNA you found was in fact ancient?

We conducted several tests to determine the age of the DNA. First, we looked for C to T substitutions, which are a common marker of ancient DNA. We found these substitutions, which gave us confidence that we were indeed looking at ancient DNA. We then used a genetic dating approach to estimate the age of the DNA. This approach involves counting the number of mutations in the mitochondrial DNA and comparing it to the number of mutations in known DNA samples. We found that the DNA we were studying was approximately 20,000 years old. Finally, we compared the DNA to known ancient populations. We found that it closely matched two samples from the same time period and geographical area. These samples were from a region that is slightly further east of Denisova Cave, but they are still within the same geographical area.

Can you tell me about how your research interlinks with the archeologists. How are the discoveries you made with your team important in understanding the past & how ancient humans lived at that time?

As I mentioned earlier this is the first time that we have been able to directly link a historical object from Paleolithic times to a genetic profile directly. When you think about medieval times, you have a burial and you have grave goods, and you can directly associate these grave goods with the person who was buried in this grave.

It is more difficult when studying Paleolithic times. I'm not aware of a case where we have this direct burial context in Paleolithic times, so it is really tricky to make any assumptions about connections between artifacts and humans at a site. This is really the first time that we have been able to start looking into those connections.

In the case of the Denisova pendant, we were able to identify that the DNA that was left on this pendant came from a woman. Now we cannot say much from that one sample about society structure or division of labor. We would need many more samples to start seeing patterns, but maybe we could start looking into such aspects of behaviors and sociality. We might see that one type of tool contains predominantly male DNA while predominantly female DNA is on another tool.

We might be able to see how people were traveling, or how ideas were traveling. If we find a specific type of tool at one site, and then we see it 5000 years later at a different site, then we can start linking them genetically and see how these objects or ideas moved geographically and timewise.

By extracting DNA from more samples, we hope to find more ancient human DNA that allows us to open the window to the past and get a better idea about these ancient societies. It would be awesome to help shed light on aspects like the social structures. Until now, this has been tricky, if not impossible, for the Paleolithic period.

What other aspects of your research are you excited by & what are your hopes for the future of your research?

The first goal is to reproduce the method on other artifacts. We want to confirm that the method doesn't just work for one pendant from the Denisova cave but that it works systematically for samples that have been freshly excavated with precautions to limit modern human contamination.

Once we have confirmed that the Denisova cave case was not a single lucky shot, then, as a methods person, I'm really interested in digging into the technical aspects. Is DNA preserved better in one type of artifact than another? We have the DNA from a pendant, which is a personal object. Are personal objects more promising sources of ancient DNA than tools because they were worn in close body contact over many years?

Something that I would really like to look into more closely is, can we overcome the problem of modern DNA contamination? There are so many exciting artifacts sitting in collections that have been contaminated with modern DNA but I'm sure they have exciting stories to tell. I would love to improve the methods so that we can also use contaminated samples for this type of study. I'm looking forward to getting back into the lab to find some parameters that can help to overcome the problem of modern human contamination.

If you could sample any artifact in the world to extract DNA, what would you chose?

Rather than a specific object, I am more interested in objects from a particular time frame: the transitional phase where we know that Neanderthals and modern humans coexisted. There are some technological complexes where it is unclear if they were made by modern humans or Neanderthals, or maybe they lived there together. If we could extract the DNA of the people handling, making and using such objects, we can maybe contribute to answering the question of who made them.

I would love to read about that if you managed to find out more! I can see why you get so excited about all of the work that you do

I'm a methods person, so when I started working here, I wasn't terribly interested in Neanderthals or early modern humans. When I came here, it was the technical challenges that excited me and the possibilities to recover DNA that is 400,000 years old [6,7]. I find it incredible to see that with these rather technical, very chemical methods, we can then tell something about humankind. I think it's a unique place here in Leipzig, where we can really combine these two things: science and history. And as a bonus, celebrate the methods and the biochemistry behind them.

It is interesting how this sort of research has been used to overcome those challenges. Without the techniques that you spend hours developing in the lab, we might never be able to understand the history of ancient humans

I never thought that I would be excited about Neanderthals, but the team here in Leipzig changed my mind! Sometimes, we're interested in one sample so we specifically develop a method for this one sample, but then it also works on other samples, making the techniques more sensitive, helping to squeeze out the teeny tiny bits of DNA that are there.

A new method might be developed to answer one specific question for one or a handful of samples, but then it can be applied to a broader set of samples, helping us gain so much more insight into the history of humans. There is a lot of crossover between the

forensics field and the ancient DNA field, so some of these methods for studying the past might help to do something good today as well.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Disclaimer

The opinions expressed in this interview are those of the interviewee and do not necessarily reflect the views of Future Science Ltd.

References

1. Essel E, Zavala EI, Schulz-Kornas E *et al.* Ancient human DNA recovered from a Palaeolithic pendant. *Nature* 618, 328–332 (2023).
2. Slon V, Hopfe C, Weiss CL *et al.* Neandertal and Denisovan DNA from Pleistocene sediments. *Science* 356(6338), 605–608 (2017).
3. Vernot B, Zavala EI, Gómez-Olivencia A *et al.* Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science* 372(6542), (2021).
4. Zavala EI, Jacobs Z, Vernot B *et al.* Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nat.* 595, 399–403 (2021).
5. Essel E, Petra Korlević P, Meyer M. A method for the temperature-controlled extraction of DNA from ancient bones. *BioTechniques* 71(1), 382–386 (2021).
6. Meyer M, Fu Q, Aximu-Petri A *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nat.* 505, 403–406 (2014).
7. Meyer M, Arsuaga JL, de Filippo C *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nat.* 531, 504–507 (2016).
8. Korleciv P, Gerber T, Gansauge MT *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques.* 59(2), (2018).

Benchmark

BioTechniques

Simple, fast and inexpensive hot sodium hydroxide and tris DNA extraction method for genotyping tomato and melon seeds

Yolanda García-Abolafio[1], Francisco Villanueva[1] & María Urrutia*,[2]

*[1]Huerta Valle Hibri2 S.L. Camino Remanente, Huerta Valle, Vélez-Málaga, Málaga, 29700, Spain; [2]Instituto de Hortofruticultura Subtropical y Mediterránea La Mayora (IHSM. UMA-CSIC), Plant Breeding & Biotechnology, Consejo Superior de Investigaciones Científicas, Algarrobo-Costa, Málaga, 29750, Spain; *Author for correspondence: m.urrutia@csic.es*

BioTechniques 76: 245–249 (December 2023) 10.2144/btn-2023-0054

First draft submitted: 30 June 2023; Accepted for publication: 12 October 2023; Published online: 2 November 2023

ABSTRACT

Seed commerce is a highly profitable global market. Most commercialized seeds are hybrid seeds originating from a controlled cross between two selected parental lines. The market value of hybrid seeds depends on their hybrid genetic purity. DNA molecular markers are a reliable and widespread tool to genotype plant materials; however, DNA extraction from seeds is challenging, often laborious and expensive. With the ultimate goal of creating a tomato and melon hybrid seeds purity test, various challenges arise. To overcome these problems and with the purpose of crude DNA extraction, a simple, fast, inexpensive and easily scalable adaptation of the hot sodium hydroxide and tris method coupled to a competitive allele-specific PCR genotyping method is proposed.

TWEETABLE ABSTRACT

Adaptation of hot sodium hydroxide and tris method for easy, fast and cheap DNA extraction for seed genotyping.

METHOD SUMMARY

A simple adaptation of the hot sodium hydroxide and tris method was established for molecular marker assessment in purity check analysis for tomato and melon seed lots.

KEYWORDS:

crude DNA extraction • *Cucumis melo* • hybrid seed • KASP • molecular markers • seed purity test • *Solanum lycopersicum*

In modern agriculture, the sowing of improved hybrid seeds (F1) is widespread. In the case of horticultural crops such as tomato and melon, F1 seeds are commercialized in a highly profitable and competitive global market. These F1 seeds are obtained from controlled crosses between two selected parental lines, and their market value is dependent on their hybrid genetic purity. In this context, reliable seed purity tests are of crucial importance.

DNA analysis via molecular markers is a reliable, convenient and effective approach for fast and cost-effective varietal identification and seed purity tests. These tests are routinely implemented in regular seed production and commerce. Plants are typically genotyped from leaf tissue. For this sampling process to happen, the plant needs to grow to the cotyledonary stage or, ideally, to its first true leaf. However, this sampling process is costly in time, space and human resources when performed at an industrial level. Therefore, performing genetic screenings directly on seeds or early emerging roots (radicles) would be ideal for the characterization of seed lots. However, the main pitfall resides in obtaining DNA extracted from seed tissue. Seeds are rich in reserve components such as lipids, oils, proteins, polysaccharides and polyphenols, which can hamper DNA extraction [1]. Several commercial seed DNA extraction kits are available (e.g., Sbeadex™ plant kit for seed extractions from LGC Biosearch Technologies (Hoddesdon, UK), Quick-DNA Plant/Seed 96 kit from Zymo Research (CA, USA), Extract-N-Amp™ Seed PCR Kit from Sigma-Aldrich Corporation (MO, USA), DNeasy Plant Pro and Plant Kits from Qiagen (Hilden, Germany)); however, they are expensive for large screenings. Alternative laboratory methods tend to be long, laborious, use toxic reagents or result in all these problems combined [2–6].

The hot sodium hydroxide and tris (HotSHOT) method [7] is commonly implemented for DNA extraction from various vertebrate and insect tissues [8–10]. However, it is not routinely used in plants. The HotSHOT technique presents several advantages. First, it is simple, inexpensive, fast, nontoxic and easily scalable. Second, it requires only basic laboratory equipment. Third, it does not require liquid nitrogen. Finally, the product recovered is directly usable in downstream genotyping protocols such as competitive allele-specific

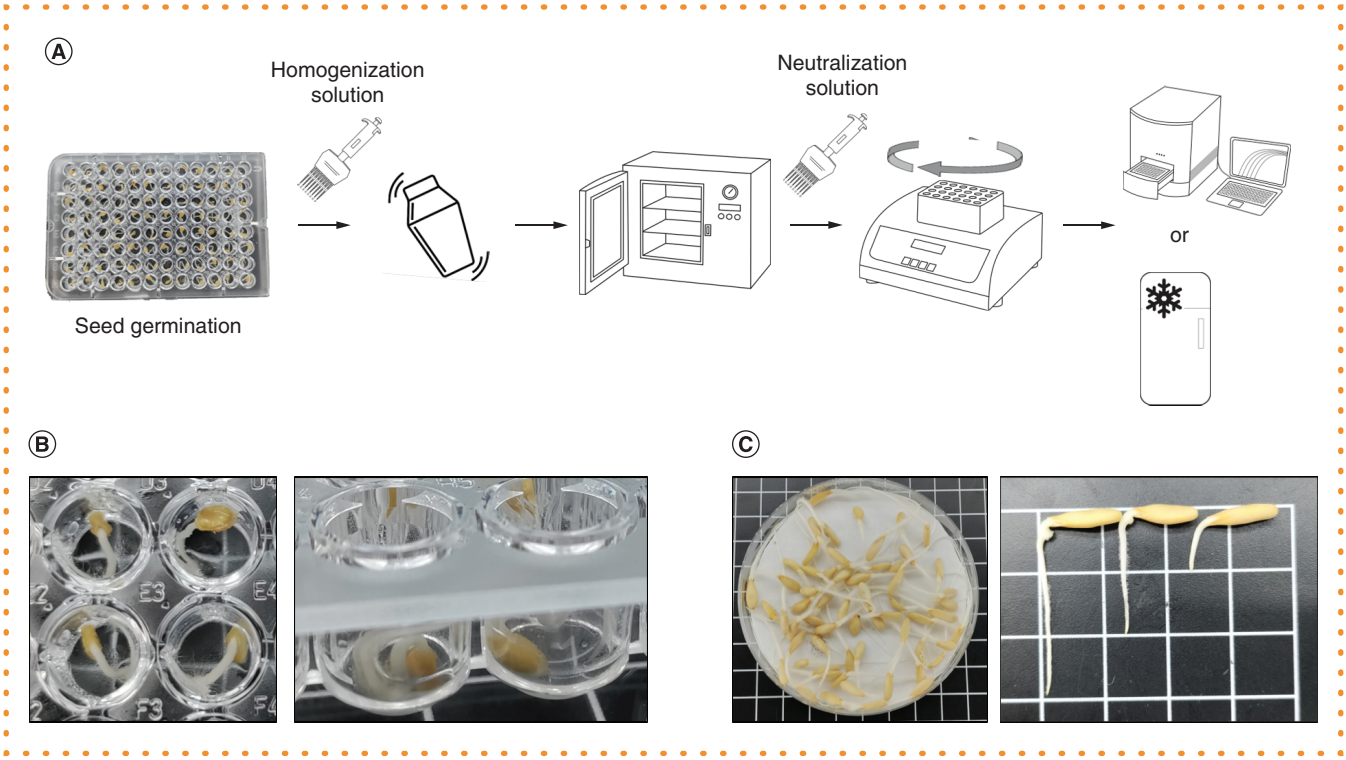


Figure 1. Hot sodium hydroxide and tris seed extraction workflow overview. (A) HotSHOT DNA extraction protocol for tomato and melon seeds workflow scheme: First, addition of homogenization solution; second, grind; third, incubate; fourth, addition of neutralization solution; fifth, mix and, finally, directly use on PCR test or store at -20°C. **(B)** Detail of properly germinated tomato seeds on microtiter plate. **(C)** Detail of properly germinated melon seeds on Petri plate. Example of large-, medium- and small-sized melon radicles. Square side = 1 cm.
HotSHOT: Hot sodium hydroxide and tris.

Table 1. DNA quality.

Sample	Extraction method	Concentration (ng/μl)	A260/A280	A260/A230
1	HotSHOT	46.5	1.58	0.50
2	HotSHOT	35.0	1.50	0.41
3	HotSHOT	45.0	1.58	0.48
4	HotSHOT	50.4	1.62	0.48
5	HotSHOT	28.2	0.99	0.21
6	CTAB	1500.0	2.09	2.02

DNA concentration (ng/μl) and quality parameters (A260/A280; A260/A230) of typical HotSHOT DNA extraction from radicles (samples 1–5) and an example of DNA quality values extracted following traditional CTAB-based DNA extraction from leaf tissue (sample 6).
CTAB: Cetyl trimethylammonium bromide; HotSHOT: Hot sodium hydroxide and tris.

PCR (KASP). In this article, a modification of the HotSHOT protocol linked to KASP genotyping for inexpensive, reproducible and high-throughput seed purity tests from tomato and melon seed lots is presented (Figure 1).

F1 seed lots were obtained from specific crosses of parental lines in tomato (*Solanum lycopersicum*) and melon (*Cucumis melo*). Tomato seeds were germinated directly in a 96-well U-bottom microtiter plate (1 seed per well) with 30 μl of distilled or tap water and allowed to germinate fully (Figure 1B). Melon seeds did not fit on microtiter plates for germination and, therefore, they were germinated on a Petri plate with moist filter paper and later sealed with parafilm. After full germination, small-, medium- and large-size radicle pieces were transferred to a 96-well U-bottom microtiter plate to evaluate the impact of input tissue quantity on the final PCR genotyping results (Figure 1C). Full germination is typically achieved in dark conditions after 5–10 days at 28°C for tomato and 4–6 days at 26°C for melon. For tomato, wells containing ungerminated seeds were annotated and excluded from further genotyping analysis as they tend to produce unreliable results. In the case of melon, only properly germinated radicles were transferred from the Petri plate to the microtiter plate for extraction.

Samples were homogenized in 100 μl of lysis solution (25 mM NaOH; pH 12) with one 4-mm stainless steel bead, covered with a cap mat for the microtiter plate and placed in a vertical homogenizer at 1500 r.p.m. for 2 min or until root tissue was homogenized. Stainless

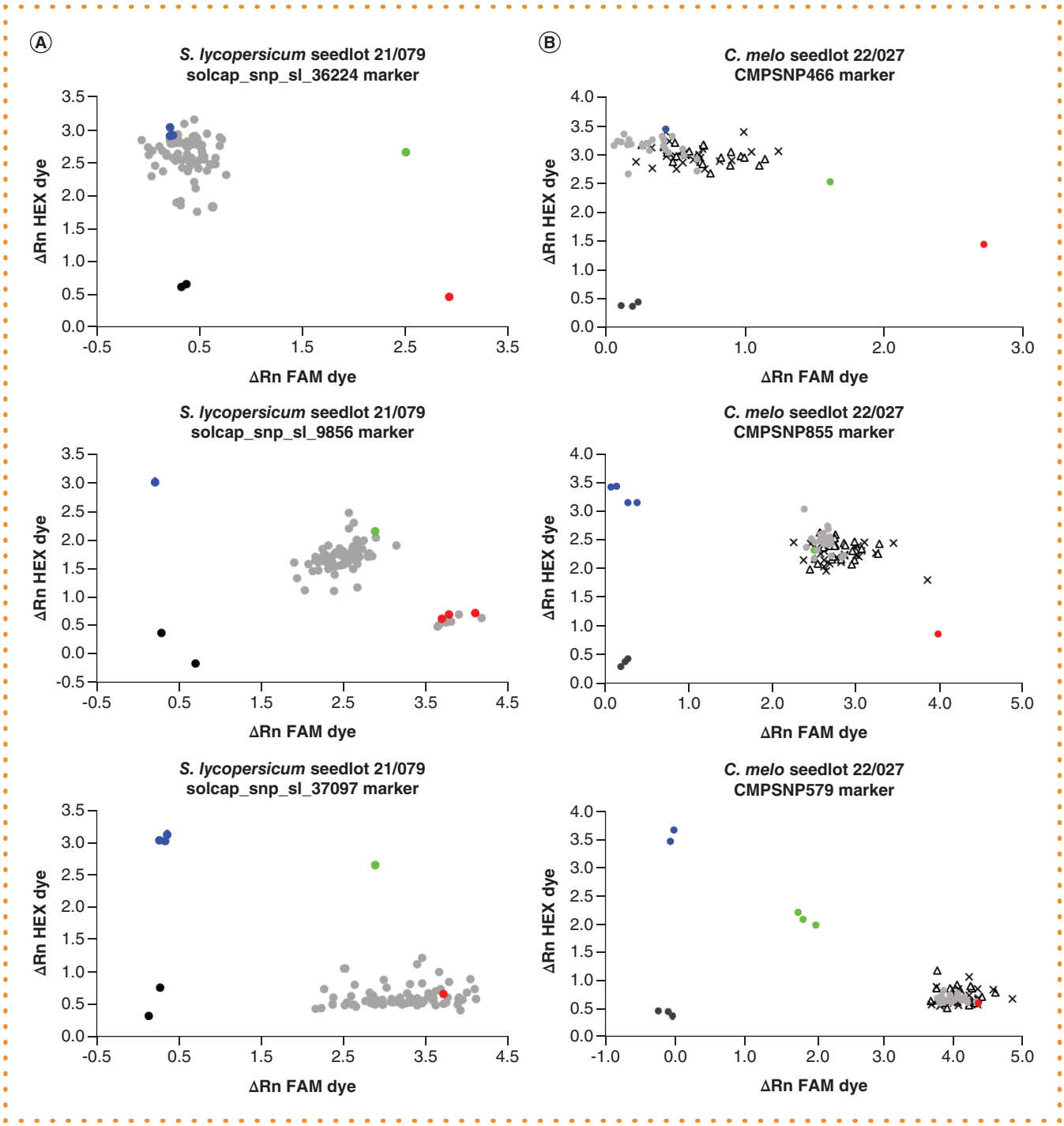


Figure 2. KASP assay genotyping results for hybrid tomato and melon seeds. Discrimination plots show florescence at PCR end point expressed as △Rn for HEX and FAM dyes. Control DNA is indicated as: red dot (●) for homozygous allele 1; blue dot (●) for homozygous allele 2 and green dot (●) for heterozygous 1/2. Nontemplate negative control is shown as black dot (●). **(A)** Tomato hybrid seed lot 21/079 genotyped with molecular markers solcap_snp_sl_36224, solcap_snp_sl_9856 and solcap_snp_sl_37097. Grey dots (●) indicate unknown tested samples. **(B)** Melon hybrid seed lot 22/027 genotyped with molecular markers CMPSNP466, CMPSNP855 and CMPSNP579 considering DNA samples extracted from large (cross), medium (triangle) and small (grey dot) radicle pieces (size according to Figure 1C). △Rn = experimental Rn signal minus baseline Rn signal generated by the instrument. Rn = fluorescence reporter value (FAM or HEX) normalized with passive reference dye (ROX).
FAM: Fluorescein; HEX: Hexachlorofluorescein; ROX: Carboxy-X-rhodamine.

steel beads are cheap and easily available as ball bearings in hardware stores. Stainless steel (420-grade) is strongly recommended for its corrosion resistance (beads remain undamaged in the lysis solution for long periods of time) and magnetic properties, which allow easy bead recovery. Stainless steel beads can also be reused in several DNA extractions after proper cleaning [11]. The process is as follows. First, dirty balls were incubated for 15 min at room temperature in a decontamination solution made of 10% household bleach, 1% NaOH, 1% Fairy® dish soap or a similar product and 90 mM sodium bicarbonate. They are then washed with tap water and later rinsed with distilled water and autoclaved. After homogenization, cap mats were carefully removed to avoid splashing and contamination between wells. Then, plates were sealed with ELISA plate stickers and incubated at 70°C for 30–50 min. Although the original HotSHOT protocol requires near-boiling incubation temperatures (95°C) [7], it is crucial to keep the incubation temperature at 70°C in this step, as polystyrene microtiter plates may not resist higher incubation temperatures. It is equally important to substitute the cap mats, which tend to deform at high temperatures, with ELISA plate stickers during incubation. Finally, samples were neutralized by adding 100 μl of neutralization solution (10 mM TrisHCl; 0.5 mM EDTA; pH 8) and mixing at 450 r.p.m. for 30 s either manually or in a horizontal mixer. At this point, crude DNA extraction is completed. This crude DNA solution can be used downstream immediately or stored at -20°C until further molecular analysis. The quantity and quality of recovered DNA was low. Average concentration values estimated by absorbance at 280 nm ranged between 20 and 50 ng/μl. A280/A260 ratios were between 1.5 and 1.6 while A260/A230 ratios were below 0.5 (see Table 1 for an example of typical DNA quality values following this procedure). These results were expected as small quantities of tissue are required as input material and no cleaning or precipitation steps are performed. Nevertheless, crude DNA extracts are suitable for KASP amplification.

To test the efficacy of DNA extraction for seed genotyping purposes, three molecular markers per seed lot were analyzed. For tomato seed lot 21/079, markers solcap_snp_sl_37097, solcap_snp_sl_9856 and solcap_snp_sl_36224 were selected from the KASP™ assay library (Figure 2A [12]). For melon seed lot 22/027, markers CMPSNP466, CMPSNP855 and CMPSNP579 were selected from Esteras *et al.* (Figure 2B) [13]. In both cases, the KASP assays were designed by LGC Biosearch Technologies (Hoddesdon, UK). Crude DNA obtained from HotSHOT extraction was used as a template. The DNA plate was centrifuged at 4°C and 3500 r.p.m. for 4 min to pellet debris (when stored at -20°C, it was thawed first). Then, 1 μl or 2.5 μl of DNA solution was pipetted into a 384-well or 96-well PCR plate, respectively. The DNA solution may be viscous, and that is why it is important to pipette carefully to avoid the tip clogging. The PCR plate was centrifuged briefly to collect the extract at the bottom of the plate and dried in a stove at 50°C for 15–30 min. Typically, multiple plates with a dispensing function multichannel pipettor were filled at a time, then dried and stored long term at -20°C for further analysis. Finally, the PCR mix was directly loaded into the PCR plate containing the dried DNA and a PCR test was conducted. Each PCR assay plate included three positive controls for allelic discrimination, three no-template negative controls and 90 target/unknown samples to be genotyped. Amplification and final reads were run in a QuantStudio 5 thermal cycler (Thermo Fisher Scientific, MA, USA) under a standard KASP protocol for a total of 42 cycles (LGC Biosearch Technologies). Allelic discrimination plots were done in QuantStudio Design and Analysis Software 1.5.2 (Thermo Fisher Scientific) with default settings for fluorescein (FAM), hexachlorofluorescein (HEX) and carboxy-X-rhodamine (ROX; used for normalization) fluorophores.

Tomato seed lot 21/079 parentals were monomorphic for markers solcap_snp_sl_36224 and solcap_snp_sl_37097 and polymorphic for solcap_snp_sl_9856, the latter being informative for seed purity test. As expected, F1 seed genotyping results were homozygous for allele 2 for markers solcap_snp_sl_36224 and homozygous for allele 1 for solcap_snp_sl_37097 (Figure 2A). However, solcap_snp_sl_9856 showed 92.3% of heterozygous and 7.7% of allele1 homozygous samples. This suggests a possible self-pollination event or the presence of 15.4% of male or female parentals heterozygous for this locus. Another polymorphic marker not linked to solcap_snp_sl_9856 should be evaluated to elucidate these possibilities.

Regarding melon seed lot 22/027, molecular markers CMPSNP466 and CMPSNP579 were monomorphic between parentals, whereas CMPSNP855 was polymorphic. Consequently, F1 genotyping results for markers CMPSNP466 and CMPSNP579 were homozygous for alleles 2 and 1, respectively, and heterozygous for CMPSNP855. In addition, different radicle sizes (small, medium and large) were evaluated as starting material for DNA extraction (Figure 1C). Allelic discrimination assays revealed that a small amount of input material in the HotSHOT DNA extraction tends to give more accurate and tighter clusters in the genotyping assay (Figure 2B).

This process is very simple since DNA extraction only requires basic laboratory equipment and two homemade buffers. In addition, all consumables and reagents are inexpensive, common laboratory materials. Furthermore, it is easily scalable as several plates can be processed in parallel allowing a high number of extractions to be performed in less than 1 h. The crude DNA extract is readily usable for genotyping by a KASP method; however, other PCR-based genotyping methods could be explored. The implementation of this method could be advantageous for screening and purity test analysis of large seed batches.

Author contributions

Y García-Abolafio, F Villanueva and M Urrutia conceived and designed the experiments; Y García-Abolafio and F Villanueva performed data acquisition; F Villanueva and M Urrutia interpreted the results and wrote the manuscript. All authors revised and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Acknowledgments

The authors thank Enrique Moriones, Eva M. Domínguez and Rafael Fernández-Muñoz for their critical reading of the manuscript and Sara Echarri-Recarte for grammar and style revision in English.

Financial disclosure

M Urrutia was funded by Proyecto QUAL21 012 IHSM, Consejería de Universidad, Investigación e Innovación, Junta de Andalucía, Spain. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Júnior CDS, Teles NMM, Luiz DP, Isabel TF. DNA extraction from seeds. In: *Sample Preparation Techniques for Soil, Plant, and Animal Samples.* Micic M (Ed.). Springer, NY, USA, 265–276 (2016).
2. Sudan J, Raina M, Singh R, Mustafiz A, Kumari S. A modified protocol for high-quality DNA extraction from seeds rich in secondary compounds. *J. Crop Improv.* 31(5), 637–647 (2017).
3. Gao S, Martinez C, Skinner DJ, Krivanek AF, Crouch JH, Xu Y. Development of a seed DNA-based genotyping system for marker-assisted selection in maize. *Mol. Breed.* 22(3), 477–494 (2008).
4. Mutou C, Tanaka K, Ishikawa R. DNA extraction from rice endosperm (including a protocol for extraction of DNA from ancient seed samples). In: *Cereal Genomics: Methods and Protocols.* Henry RJ, Furtado A (Eds). Humana Press, NJ, USA, 7–15 (2014).
5. Kamiya M, Kiguchi T. Rapid DNA extraction method from soybean seeds. *Breed. Sci.* 53(3), 277–279 (2003).
6. Liang H, Deng Y, Wang C, Xu X. A high-throughput DNA extraction method from rice seeds. *Biotechnol. Biotechnol. Equip.* 30(1), 32–35 (2016).
7. Truett GE, Heeger P, Mynatt RL, Truett AA, Walker JA, Warman ML. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *BioTechniques* 29(1), 52–54 (2000).

•• **Fist description of HoTSHOT DNA extraction method.**

8. Martincová I, Aghová T. Comparison of 12 DNA extraction kits for vertebrate samples. *Anim. Biodivers. Conserv.* 43.1, 67–77 (2020).
9. Dahle G, Johansen T, Westgaard J-I, Aglen A, Glover KA. Genetic management of mixed-stock fisheries "real-time": the case of the largest remaining cod fishery operating in the Atlantic in 2007–2017. *Fish. Res.* 205, 77–85 (2018).
10. Gutiérrez-López R, Martínez-de la Puente J, Gangoso L, Soriguer RC, Figuerola J. Comparison of manual and semi-automatic DNA extraction protocols for the barcoding characterization of hematophagous louse flies (Diptera: hippoboscidae). *J. Vector Ecol.* 40(1), 11–15 (2015).
11. Pipette Jockey. Just bleach it (2020). https://pipettejockey.com/2016/11/01/just-bleach-it/
12. LGC Biosearch Technology. KASP™ assay library for tomato genotyping. www.biosearchtech.com/products/pcr-reagents-kits-and-instruments/pcr-probes-and-assays/kasp-genotyping-panels/tomato-genotyping-panel

• **Database for tomato molecular markers using KASP technology.**

13. Esteras C, Formisano G, Roig C *et al.* SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. *Theor. Appl. Genet.* 126, 1285–1303 (2013).

• **Reference providing a comprehensive set of molecular markers for melon genotyping.**

Reports

BioTechniques

DNA extraction from clotted blood in genotyping quality

Kira J Stanzick[1], Josef Simon[1], Martina E Zimmermann[1], Michael Schachtner[2], David Peterhoff[2,3], Hans-Helmut Niller[2], Klaus Überla[4], Ralf Wagner[2,3], Iris M Heid[1] & Klaus J Stark*[,1]

*[1]Department of Genetic Epidemiology, University of Regensburg, Franz-Josef-Strauß-Allee 11, Regensburg, 93053, Germany; [2]Institute of Medical Microbiology & Hygiene, Molecular Microbiology (Virology), University of Regensburg, Franz-Josef-Strauß-Allee 11, Regensburg, 93053, Germany; [3]Institute of Clinical Microbiology & Hygiene, University Hospital Regensburg, Franz-Josef-Strauß-Allee 11, Regensburg, 93053, Germany; [4]Institute of Clinical & Molecular Virology, University Hospital Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, Schlossgarten 4, Erlangen, 91054, Germany; *Author for correspondence: Tel.: +49 941 944 5214; klaus.stark@klinik.uni-regensburg.de*

BioTechniques 74: 23–29 (January 2023) 10.2144/btn-2022-0061

First draft submitted: 1 June 2022; Accepted for publication: 5 December 2022; Published online: 4 January 2023

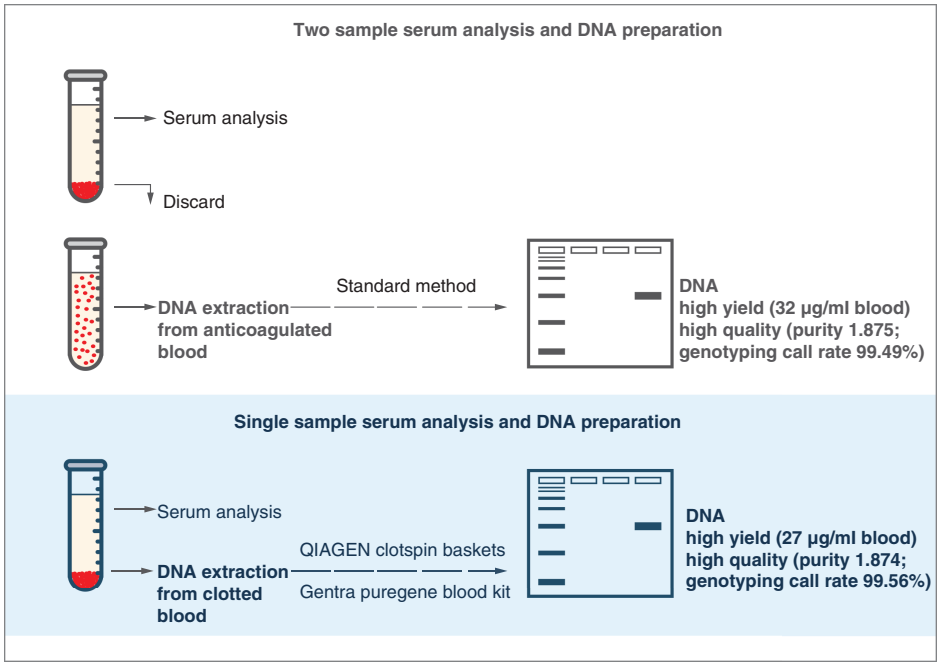
ABSTRACT

DNA extraction from frozen blood clots is challenging. Here, the authors applied QIAGEN Clotspin Baskets and the Gentra Puregene Blood Kit for DNA extraction to cellular fraction of 5.5 ml whole blood without anticoagulating additives. The amount and quality of extracted DNA were assessed via spectrophotometer and gel electrophoresis. Results from array-based genotyping were analyzed. All steps were compared with DNA isolated from anticoagulated blood samples from a separate study. The quality and concentration of DNA extracted from clotted blood were comparable to those of DNA extracted from anticoagulated blood. DNA yield was on average 27 µg per ml clotted blood, with an average purity of 1.87 (A260/A280). Genotyping quality was similar for both DNA sources (call rate: 99.56% from clotted vs 99.49% from anticoagulated blood).

METHOD SUMMARY

This study describes DNA extraction from frozen clotted blood using QIAGEN Clotspin Baskets and the Gentra Puregene Blood Kit. The concentration, yield per milliliter of blood, purity and integrity of the DNA were further confirmed via spectrophotometer and gel electrophoresis. In addition, performance for genotyping on the Illumina Global Screening Array was analyzed. Comparisons with DNA extracted from anticoagulated blood were performed.

GRAPHICAL ABSTRACT



KEYWORDS:
anticoagulated blood • AugUR • clotted blood • DNA concentration • DNA extraction • DNA purity • DNA yield • frozen samples • genotyping • TiKoCo

Epidemiological studies at field centers are important for effective recruitment in short time. For practical constraints "in field," but also ethical concerns, the participants' burden must be reduced to a minimum, especially regarding blood collection.

For most biomarker analyses, whole blood is collected in tubes without additives to gain serum, and cellular fractions are usually discarded after centrifugation. Gold-standard DNA extraction requires buffy coat or whole anticoagulated blood [1,2], and thus blood draw with a separate tube with anticoagulant additives. The additional blood sampling can be avoided when extracting DNA from serum tubes (i.e., clotted blood), for which several methods have been described [3–11]. However, the applicability and effectiveness of these methods to obtain high-quality DNA in epidemiological-scale studies for whole-genome genotyping are unclear.

Here the authors describe a protocol for DNA extraction from blood clots after storage at -20°C for more than 1 year using commercial products. The authors applied this protocol for DNA extraction for an epidemiological-scale field study, the Tirschenreuth SARS-CoV-2 antibody study (TiKoCo, n = 4204) [12–14]. DNA yield, quality and utility for array-based genotyping were compared with DNA extracted from anticoagulated blood from the AugUR study (n = 1282) [15]. Blood samples from this study were chosen for comparison because DNA extraction from AugUR was conducted in the same laboratory and by the same personnel. In addition, AugUR was genotyped with the very same array as TiKoCo.

Materials & methods

Sample collection & DNA extraction materials

Whole blood was collected for the baseline survey of the TiKoCo cohort, a population-based study to determine the SARS-CoV-2 serostatus in the German county of Tirschenreuth [12-14]. The TiKoCo study was approved by the Ethics Committee of the University of Regensburg, Germany (vote 12-101-0258). A single blood tube to obtain a serum sample was taken from each of the 4204 study participants in sitting position (S-Monovette® 5.5 ml, Sarstedt, Nümbrecht, Germany). The samples were processed on the same day. After centrifugation (2000× *g*, 10 min at room temperature), serum was used for anti-SARS-CoV-2 tests and clotted blood was stored at -20°C up to 415 days prior to DNA extraction.

The authors excluded 70 participants without informed consent for genetic analyses and an additional three participants without available clotted blood. For the remaining 4131 participants, blood clot preparation and DNA extraction were performed using Clotspin® Baskets and the Gentra® Puregene® Blood Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions [16] with slight modifications. The Gentra Puregene Blood Kit contains red blood cell (RBC) lysis solution, cell lysis solution, protein precipitation solution (salting-out method) and DNA hydration solution. Proteinase K, isopropanol (2-propanol) and ethanol at molecular biology grade were purchased from Thermo Fisher Scientific (MA, USA).

Protocol for DNA extraction from blood clots

Blood clot preparation & RBC lysis

In detail, frozen blood clots in 15 ml tubes were transferred from -20°C to a warming cabinet (55°C) for 10 min and thereafter immediately placed on ice. The tube was inverted to loosen the clot. The blood clot was completely poured with 5 ml RBC lysis solution into the Clotspin Basket placed on a 50 ml tube (Sarstedt). To disperse the clot, the sample was centrifuged at 2000× *g* for 5 min. The remaining clot material from the Clotspin Basket was transferred through the basket to the filtrate with 10 ml RBC lysis solution and the basket was discarded. To completely disperse the clotted material, the filtrate was vortexed vigorously for 3 s and placed for 5 min at room temperature on a circulating shaker (250 1/min). The tubes were again vortexed vigorously for 3 s and centrifuged at 2000× *g* for 5 min. The supernatant was carefully discarded, taking care that the pellet remained in the tube. If no pellet was visible, about 0.5 ml of the supernatant was kept in the tube. The tube was vortexed rigorously for 10 s and an additional 5 ml RBC lysis solution was added to the pellet, followed by vortexing for 3 s and incubation on a circulating shaker (250 1/min) for 5 min at room temperature.

White blood cell lysis

After centrifugation at 2000× *g* for 5 min to pellet the DNA-containing white blood cells, the supernatant was carefully discarded, leaving about 0.2 ml of residual liquid. The tube was vortexed rigorously for 10 s. Addition of 5 ml cell lysis solution and 25 μl of proteinase K (20 mg/ml) was followed by rigorous vortexing for 10 s. For complete lysis of DNA-containing cells, the samples were incubated overnight at 55°C.

Protein precipitation

On the next day, the samples were cooled on ice for 5 min, and 1.7 ml protein precipitation solution was added, followed by rigorous vortexing for 20 s. After centrifugation at 2000× *g* for 10 min, the samples were incubated for 2 min on ice.

DNA precipitation & washing

For the precipitation of DNA, the supernatant was carefully transferred in a 50 ml tube, containing 5 ml isopropanol. The samples were mixed by gently inverting the tube 50-times and centrifuged at 2000× *g* for 3 min. The supernatant was carefully discarded, and the tube was drained on a clean piece of absorbent paper without losing the pellet. In preliminary experiments, the purity of DNA from samples that showed red- to brown-colored pellets after DNA precipitation was below 1.6 (A260 nm/A280 nm). Therefore, the DNA pellet was inspected visually and with the two following options. (1) If the pellet was not small and white but large and red to brown, 2 ml of cell lysis solution and 10 μl proteinase K were added, followed by incubation at 55°C for 2 h in a warming cabinet or at room temperature overnight on a circulating shaker (250 1/min). For removal of residual proteins and other cellular contaminants, the samples were cooled on ice for 5 min and 1 ml protein precipitation solution was added, followed by rigorous vortexing for 20 s. After centrifugation at 2000× *g* for 10 min, the samples were incubated for 2 min on ice. DNA precipitation was done with 3 ml isopropanol. The samples were mixed by gently inverting the tube 50-times and centrifuged at 2000× *g* for 3 min. The supernatant was carefully discarded. The tube was drained on a clean piece of absorbent paper, taking care that the pellet remained in the tube. (2) For small and white pellets – or after option 1 – 5 ml of 70% ethanol was added immediately and the tubes were inverted until the pellet was detached. After centrifugation at 2000× *g* for 3 min, the supernatant was carefully discarded, and the DNA pellet was air-dried at room temperature for 10 min or until the pellet got glassy.

DNA hydration

The addition of 0.5 ml DNA hydration solution for a large pellet or 0.3 ml for a smaller pellet was followed by incubation at 65°C in a warming cabinet for an hour. To fully dissolve the DNA, the samples were put on a shaker overnight at room temperature. On the next day, the samples were centrifuged briefly and the solved DNA was transferred to a 2 ml cup and stored at -20°C. To avoid cross-contamination of samples, all pipetting steps were carried out using filter tips (different manufacturers).

The protocol is available at www.protocols.io/view/dna-extraction-5qpvobjw7l4o/v1.

DNA extraction from anticoagulated blood

For 1282 participants from the second baseline survey of the German AugUR study [15], DNA samples extracted with a very similar protocol, but from EDTA-anticoagulated blood, were available (AugUR2). Both protocols use a salting-out method to remove proteins and other cellular contaminants. In brief, the differences in the AugUR2 DNA extraction protocol from the one used for clotted blood are the following: whole blood from AugUR2 was stored at -20°C in 2.7 ml EDTA-containing monovettes (Sarstedt). No Clotspin Baskets were used. After thawing, RBC lysis was done in 9 ml lysis buffer (155 mM NH_4Cl, 20 mM $KHCO_3$, 0.1 mM Na_2EDTA, pH 7.4). Lysis of DNA-containing cells was performed with 1.6 ml SE buffer (75 mM NaCl, 25 mM Na_2EDTA, pH 8.0), 10 μl proteinase K (20 mg/ml) and 100 μl 20% sodium dodecyl sulfate overnight. For salting-out, 1.7 ml SE buffer and 1 ml saturated NaCl were used, followed by centrifugation at 40°C. Extracted DNA was solved in 0.3 ml DNA hydration solution (10 mM Tris/HCl pH 8.0, 1 mM Na_2EDTA) overnight.

Evaluation of DNA concentration & quality

Two μl of DNA was used to determine quantity (A260 nm) and quality (A260 nm/A280 nm) on a Tecan Infinite® 200 PRO plate reader with i-control 1.10.4.0 software and a NanoQuant plate (Tecan Life Sciences, Männedorf, Switzerland). To assess potential DNA degradation or shearing, electrophoresis with 1 μl undiluted DNA was performed on 0.8% agarose gels.

Evaluation of DNA by genotyping

Genotyping of single nucleotide polymorphisms for both studies, TiKoCo and AugUR2, was conducted by Life & Brain GmbH, Bonn, Germany, using the Infinium Global Screening Array-24, GSA-MD, version 3.0 (Illumina, Inc., CA, USA) following the manufacturer's instructions, applying 200 ng of DNA. A minimum DNA concentration of 60 ng/μl for genotyping is recommended. All samples meeting this criterion were genotyped. A total of 730,059 variants were available on the genotyping array ("GSAMD-24v3-0-EA_20034606_A1" manifest file).

Data management, statistical analysis & literature search

Data management and statistics were done with SPSS 26.0.0.1 for Windows (IBM Corp., NY, USA). Box and whisker plots were created with GraphPad Prism, version 8.4.3, for Windows (GraphPad Software, CA, USA). PLINK 1.9 was used to analyze the genotype data [17]. For comparison of continuous parameter between two groups, the *t*-test was applied. The level of significance was set at $p < 0.05$.

A comprehensive PubMed search was conducted using the search terms "DNA extraction method human clotted blood" (access date: 1 April 2022). In addition, Labome was searched for available DNA extraction protocols (last access date: 1 April 2022) [18].

Results & discussion

DNA from clotted blood was extracted from 4131 samples, one each per participant of the TiKoCo study. Purity of the extracted DNA was high, with a mean of 1.874 (A260 nm/A280 nm; Table 1); n = 4111 (99.5%) samples yielded a purity >1.80. No difference in DNA purity between extraction from clotted blood and anticoagulated blood (n = 1282) could be observed (Table 1). Total yield of DNA per ml whole

Table 1. Comparison of concentration and quality for DNA extracted from serum clotted blood and EDTA-anticoagulated blood.

	Clotted blood	Anticoagulated blood
Study	TiKoCo	AugUR2
Blood available, n	4131	1282
Blood volume, ml	5.5	2.7
DNA purity (A260 nm/A280 nm), mean ± SD	1.874 ± 0.042	1.875 ± 0.023
DNA per 1 ml blood (μg), mean ± SD (maximum)	26.92 ± 12.56 (111.71)	32.15 ± 12.45 (103.64)
DNA concentration <60 ng/μl, n (%)	71 (1.72%)	2 (0.16%)
DNA samples used for genotyping	4060	1280
Per person genotyping call rate on Illumina Global Screening Array-24, GSA-MD, version 3.0 (%), mean ± SD (n <95%)	99.560 ± 0.003 (3)	99.493 ± 0.001 (0)
Per single nucleotide polymorphism genotyping call rate on Illumina Global Screening Array-24, GSA-MD, version 3.0 (%), mean ± SD	99.502 ± 0.050	99.427 ± 0.052

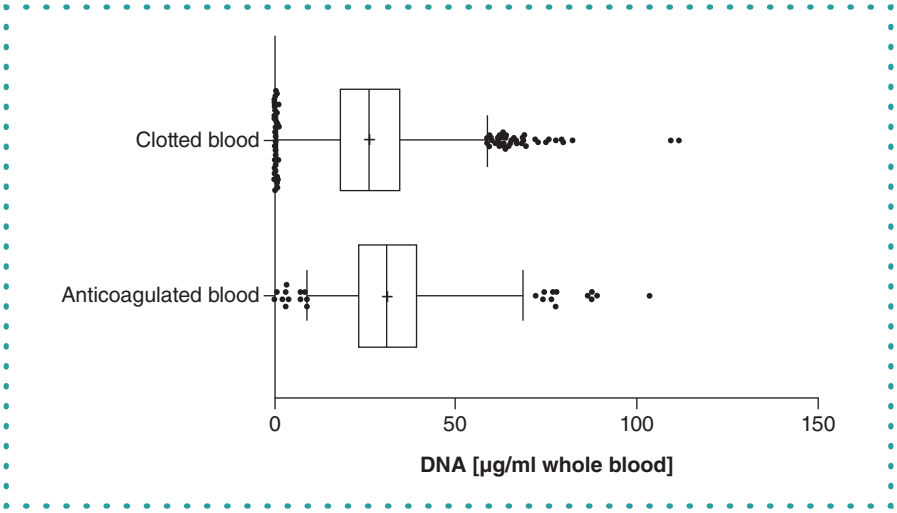


Figure 1. Distribution of total amount of DNA normalized to 1 ml of whole blood from the two DNA sources: DNA extraction from clotted blood (n=4131, TiKoCo), DNA extraction from anticoagulated blood (n=1282, AugUR2). Boxes show 25th, 50th and 75th percentiles. Whiskers range from 1st to 99th percentile with values beyond these percentiles as single points. The mean of DNA amount is shown as "+" (26.9 μg and 32.2 μg respectively; $p=2.4*10^{-38}$).

blood was 26.9 ± 12.6 μg from clotted and 32.2 ± 12.5 μg from anticoagulated samples (Table 1). The clotted blood samples that were washed twice after DNA precipitation (n = 140) showed comparable good purity (1.876 in comparison with 1.874 of samples from clotted blood without second washing step, n = 3991) and 31% lower DNA yield (18.68 μg/ml blood vs 27.21 μg/ml blood), but with sufficient mean DNA concentration (226.6 ng/μl vs 313.5 ng/μl). The minimum DNA concentration of 60 ng/μl for genotyping was achieved for 4060 (98.3%) clotted and for 1280 (99.8%) anticoagulated blood samples (Table 1); all these samples were used for genotyping. For 15 (0.4%) clotted blood samples, low amounts of DNA with concentration <5 ng/μl (too little for most high-throughput analyses) were achieved. The minimum yield for the remaining samples with a concentration between 5 and 59 ng/μl was 1.5 μg DNA in total, which is sufficient for most downstream applications after concentrating the DNA.

DNA yield per ml blood was significantly lower after extraction from clotted blood in comparison with anticoagulated blood (Figure 1). Especially, DNA concentration in samples from clotted blood was below the genotyping threshold of 60 ng/μl in 1.72% of TiKoCo samples in comparison with 0.16% from AugUR2 (Table 1). However, this is not only due to the different extraction protocols but also due to the different blood sampling processes. While AugUR2 blood collection was performed under standardized circumstances in the authors' study center, TiKoCo blood collection was performed "in the field" at three locations, including home visits for immobile individuals [12]. In AugUR2, for all samples, 2.7 ml anticoagulated blood was available. The authors observed that some TiKoCo samples did not contain the requested 5.5 ml blood. From the 71 samples resulting in DNA concentration below 60 ng/μl in TiKoCo, 63 were marked to be not fully filled blood samples. Therefore, very low DNA concentration was rather caused by the low amount of blood in the tube than in dependence of the DNA extraction method. Taking only fully filled tubes into account, the rate of samples with DNA concentration below 60 ng/μl was comparable between extraction from clotted blood (0.19%) and extraction from anticoagulated blood (0.16%).

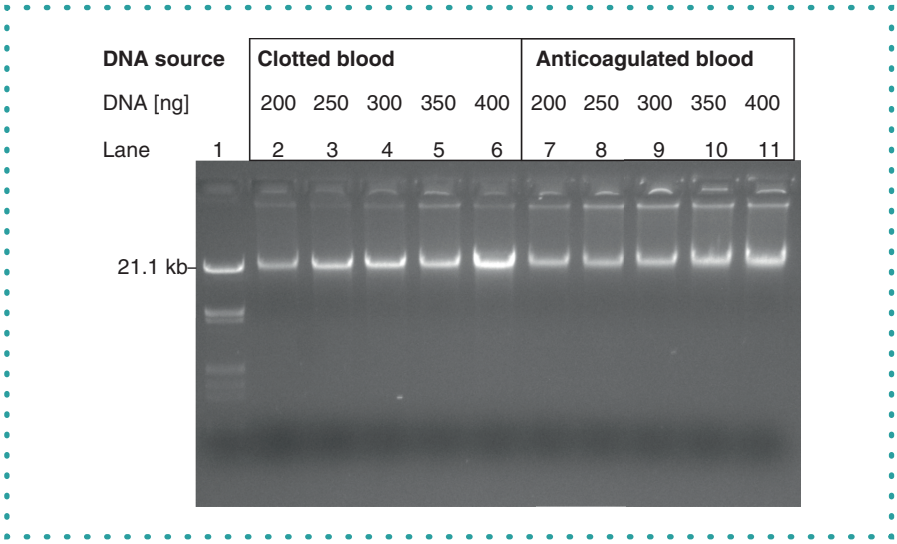


Figure 2. Gel electrophoresis of 1 μl extracted DNA on 0.8% agarose gel. Lane 1 represents λ-EcoRI-HindIII DNA marker (500 ng) with band at 21,226 bp. Lane 2 to 6: DNA samples extracted from clotted blood (TiKoCo) with 200, 250, 300, 350 and 400 ng/μl genomic DNA. Lane 7 to 11: DNA samples extracted from anticoagulated blood (AugUR2) with 200, 250, 300, 350 and 400 ng/μl genomic DNA.

To determine degradation and the molecular weight of the genomic DNA, gel electrophoresis was performed. With the DNA extraction method for clotted blood described here, no degradation of DNA could be observed. The integrity of the DNA was comparable to extraction from anticoagulated blood, showing a high molecular, compressed band (Figure 2).

Genotyping for both TiKoCo and AugUR2 was performed identically at Life & Brain GmbH, Bonn, Germany, applying the Illumina Global Screening Array, version 3. The per person call rate was very high for both DNA sources, with 99.6% for DNA from clotted and 99.5% from anticoagulated blood. The per single nucleotide polymorphism call rate for all 730,059 variants was at the same range, with 99.5% for DNA from clotted and 99.4% from anticoagulated blood (Table 1).

The authors evaluated the influence of storage time on DNA yield for TiKoCo clotted blood samples. While all 4131 samples were stored at -20°C at the same day of blood draw, DNA extraction processing stretched over a period, resulting in a range of storage time from 190 to 415 days. The authors found a small but significant decrease in DNA yield over blood storage time: the difference between the 25% shortest (190–247 days) and the 25% longest storage times (386 to 415 days) was 2.2 μg DNA per ml blood ($p = 3.6*10^{-5}$). However, the observed decrease of DNA yield after 1-year -20°C storage time was low (8%) in comparison with previous reports of 1-month storage at 4°C, with a drop from 37.1 μg to 0.439 μg per 4 ml clotted sample [5]. Interestingly, significant differences in genotyping call rate over the storage time of blood clots at -20°C was observed, but with higher call rate in samples that were stored longer (25% shortest vs 25% longest storage time = 99.50% vs 99.63%; $p = 1.6*10^{-30}$). This may have been by chance, but it shows that genotyping performance did not decrease with blood samples stored at -20°C for a longer period. Storage of blood clots at -20°C for more than 1 year or at even lower temperature warrants further investigations. Since serum tubes do not include DNA stabilization agents (e.g., EDTA), frozen storage is highly recommended in general.

Several protocols and reports on DNA extraction from human clotted blood and after different storage times exist in the literature. The authors performed a PubMed search with the terms "DNA extraction method human clotted blood," resulting in 151 hits. Manual curation of these PubMed hits pointed out several fitting protocols and studies, summarized in Table 2. Methods with manual homogenization or slicing of blood clots were not considered further, since they are impractical for high-throughput DNA extraction.

Adkins *et al.*, in their 2002 published analysis, also used Clotspin tubes and the PureGene DNA purification kit (formerly distributed by Gentra) and tested the extracted DNA for performance in single nucleotide polymorphism genotyping [10]. This was done in only 15 samples. The authors of the present study used the protocol presented in this paper on more than 4000 clotted blood samples, added an optional washing step to improve DNA purity, analyzed the effect of blood clot storage at -20°C over time and tested performance on whole-genome genotyping.

Time needed for DNA extraction is critical for planning personnel resources and feasibility, particularly for epidemiological-scale studies. Hands-on time for DNA extraction from clotted blood was moderate and comparable to our standard DNA extraction method from anticoagulated blood, both with two overnight steps that increase DNA yield. An additional 20 min should be planed for the Clotspin step needed for clotted blood, with effective time depending on the number of parallelly processed samples. The throughput for both methods is limited by the availability of equipment, such as centrifugation capacity and place on circulating shakers. Under ideal equipment conditions, a single technician could process more than 200 samples per day.

Table 2. Selected publications from PubMed search "DNA extraction method human clotted blood."

Publication	Key message	Ref.
Simple and rapid method for extraction of DNA from fresh and cryopreserved clotted human blood. Garg UC *et al. Clin. Chem.* 1996.	Nylon mesh and Qiagen (Gentra) PureGene. Reduced DNA yield compared with extraction from anticoagulated blood. No reduction of DNA yield from frozen clots (at least 1 week).	[3]
Utilizing genomic DNA purified from clotted blood samples for single nucleotide polymorphism genotyping. Adkins KK *et al. Arch. Pathol. Lab. Med.* 2002.	Comparable to the protocol presented in this paper. Extracted DNA tested for single nucleotide polymorphism genotyping. No report on the influence of long-term frozen storage of blood clots.	[4]
A simple method for DNA isolation from clotted blood extricated rapidly from serum separator tubes. Se Fum Wong S *et al. Clin. Chem.* 2007.	Good yield with fresh samples, massive reduction for stored blood clots.	[5]
An alternate method for DNA and RNA extraction from clotted blood. Zakaria Z *et al. Genet. Mol. Res.* 2013.	Extraction of DNA and RNA from clotted blood, sonication for 3 h.	[6]
High-quality and -quantity DNA extraction from frozen archival blood clots for genotyping of single-nucleotide polymorphisms. Bank S *et al. Genet. Test. Mol. Biomarkers* 2013.	Testing different commercial kits. Promega Maxwell 16 Blood Purification Kit and Qiagen Puregene performed best.	[7]
Higher DNA yield for epidemiological studies: a better method for DNA extraction from blood clot. Zhou G *et al. Genet. Test. Mol. Biomarkers* 2019.	High performance of extracted DNA in next-generation sequencing; needs homogenizer.	[8]
A method for improving the efficiency of DNA extraction from clotted blood samples. Mardan-Nik M *et al. J. Clin. Lab. Anal.* 2019.	Ball bearing metal shots in blood clot-containing tubes to break down the blood clot in combination with modified salting-out DNA extraction method. "The main challenge in this method is pipetting in each step to dissolve the pellet."	[9]

Conclusion

This protocol for DNA extraction from clotted blood resulted in a high yield and high quality of DNA, comparable to a similar protocol using anticoagulated blood. We also established the utility of clotted blood-derived DNA for genotyping arrays and documented comparability across the two DNA sources indicated by similar genotype call rates.

DNA extraction from frozen clotted blood using QIAGEN Clotspin Baskets and the Gentra Puregene Blood Kit can be recommended as an efficient way to get high-quality material for genetic analyses from the same blood tube used for serum biomarker analytics, which can be very interesting for epidemiological field studies. Importantly, the DNA yield from blood clots with a mean of 27 μg/ml blood is sufficient for genotyping, whole-genome sequencing and methylation analysis.

Future perspective

Field work in epidemiological studies is a fast way to recruit participants and to obtain health-related data. Genetic analyses are becoming increasingly important to understand molecular mechanisms of diseases. An effective workflow at the point of recruitment is required, especially in collecting biomaterials. Many human biomarkers and antibodies can be measured in serum. To avoid separate blood draws, DNA extraction from remaining blood cells is recommended and feasible with the protocol reported here, even after storage at -20°C for 1 year.

Executive summary

- Our report "DNA extraction from clotted blood in genotyping quality" aims to establish a feasible method to isolate high quality DNA from frozen blood clots and to compare performance for large scale genotyping with DNA extracted from anticoagulated blood.
- The main objective is to reuse stored blood clots from serum samples to avoid additional blood draw with separate tubes. Quantity and quality of DNA isolated from clotted blood should be the same as for gold-standard DNA extraction.
- Our extraction protocol results in DNA with no degradation and high purity. Measures for DNA quantity and quality are presented. Genome-wide genotyping for DNA extracted with our protocol performs at the same high-quality level as with gold-standard extracted DNA.
- In conclusion, DNA extraction from clotted blood applying our protocol is an effective way to avoid additional blood draw in epidemiological studies without any limitations in quality for downstream applications.

Author contributions

KJ Stanzick: methodology and experimentation, validation, data analysis, manuscript writing; J Simon: methodology and experimentation, validation, reviewing and editing of the manuscript; ME Zimmermann: methodology, validation, reviewing and editing of the manuscript; R Wagner, K Überla: reviewing and editing of the manuscript, study PIs of TiKoCo; D Peterhoff, M Schachtner, H-H Niller: sample preparation, reviewing and editing of the manuscript; IM Heid: reviewing and editing of the manuscript, study PI of AugUR; KJ Stark: supervision, methodology, validation, data analysis, manuscript writing

Acknowledgments

The authors thank all study participants for contributing to the TiKoCo and AugUR studies. They would like to thank Melanie Waldenberger, Eva Reischl and Nadine Lindemann from the Research Unit Molecular Epidemiology at the Helmholtz Zentrum Munich for their help in establishing the AugUR2-DNA extraction protocol. The authors also thank Magdalena Scharl for the excellent support during the work.

Financial & competing interests disclosure

The TiKoCo study was supported by the Bavarian Ministry of Science and Arts as well as by the National Research Network of the University Medicine. The AugUR study was supported by grants from the German Federal Ministry of Education and Research (BMBF 01ER1206 and BMBF 01ER1507), by the German Research Foundation (DFG HE 3690/7-1 and BR 6028/2-1) and by institutional budget (University of Regensburg). Author I Heid has received support from Roche Diagnostics for a biomarker project in the AugUR study but it is unrelated to the work presented here. No conflicting relationship exists for the other authors. The authors have not received any funding for this project from QIAGEN. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The TiKoCo study protocol was approved by the Ethics Committee of the University of Regensburg (vote 20-1867-101) and adopted by the Ethics Committee of the University of Erlangen (vote 248_20 Bc). The AugUR study protocol, study procedures and data protection strategy were all approved by the Ethics Committee of the University of Regensburg, Germany (vote 12-101-0258). The studies comply with the 1964 Declaration of Helsinki and its later amendments. All participants provided written informed consent.

Data sharing statement

The datasets generated and analyzed during the current study are not publicly available due to data privacy of study participants. Data on DNA quality are available on request.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Austin MA, Ordovas JM, Eckfeldt JH *et al.* Guidelines of the National Heart, Lung, and Blood Institute Working Group on Blood Drawing, Processing, and Storage for Genetic Studies. *Am. J. Epidemiol.* 144(5), 437–441 (1996).
2. Gail MH, Sheehy T, Cosentino M *et al.* Maximizing DNA yield for epidemiologic studies: no more buffy coats? *Am. J. Epidemiol.* 178(7), 1170–1176 (2013).
3. Garg UC, Hanson NQ, Tsai MY, Eckfeldt JH. Simple and rapid method for extraction of DNA from fresh and cryopreserved clotted human blood. *Clin. Chem.* 42(4), 647–648 (1996).
4. Adkins KK, Strom DA, Jacobson TE, Seemann CR, O'Brien DP, Heath EM. Utilizing genomic DNA purified from clotted blood samples for single nucleotide polymorphism genotyping. *Arch. Pathol. Lab. Med.* 126(3), 266–270 (2002).
   - **Describes the Clotspin tubes.**
5. Se Fum Wong S, Kuei JJ, Prasad N *et al.* A simple method for DNA isolation from clotted blood extricated rapidly from serum separator tubes. *Clin. Chem.* 53(3), 522–524 (2007).
   - **Reports the dramatic decrease of DNA yield in clotted blood samples stored at 4°C.**
6. Zakaria Z, Umi SH, Mokhtar SS *et al.* An alternate method for DNA and RNA extraction from clotted blood. *Genet. Mol. Res.* 12(1), 302–311 (2013).
   - **Describes parallel extraction of DNA and RNA from clotted blood.**
7. Bank S, Nexø BA, Andersen V, Vogel U, Andersen PS. High-quality and -quantity DNA extraction from frozen archival blood clots for genotyping of single-nucleotide polymorphisms. *Genet. Test Mol. Biomarkers* 17(6), 501–503 (2013).
8. Zhou G, Du Y, Wu Y, Wang WC. Higher DNA yield for epidemiological studies: a better method for DNA extraction from blood clot. *Genet. Test Mol. Biomarkers* 23(1), 66–72 (2019).
   - **Compares four methods and describes the success rate in next-generation sequencing.**
9. Mardan-Nik M, Saffar Soflaei S, Biabangard-Zak A *et al.* A method for improving the efficiency of DNA extraction from clotted blood samples. *J. Clin. Lab. Anal.* 33(6), e22892 (2019).
10. Xu R, Ye P, Luo L, Wu H, Dong J, Deng X. A simple and efficient method for DNA purification from samples of highly clotted blood. *Mol. Biotechnol.* 46(3), 258–264 (2010).
11. Zeillinger R, Schneeberger C, Speiser P, Kury F. A simple method for isolation of DNA from blood clots suited for use in PCR. *BioTechniques* 14(2), 202–203 (1993).
    - **Describes an early but pioneering method for DNA extraction from blood clots.**
12. Wagner R, Peterhoff D, Beileke S *et al.* Estimates and determinants of SARS-CoV-2 seroprevalence and infection fatality ratio using latent class analysis: the population-based Tirschenreuth study in the hardest-hit German county in spring 2020. *Viruses* 13(6), 1118 (2021).
13. Einhauser S, Peterhoff D, Niller HH *et al.* Spectrum bias and individual strengths of SARS-CoV-2 serological tests – a population-based evaluation. *Diagnostics (Basel)* 11(10), 1843 (2021).
14. Peterhoff D, Einhauser S, Beileke S *et al.* Comparative immunogenicity of COVID-19 vaccines in a population-based cohort study with SARS-CoV-2-infected and uninfected participants. *Vaccines (Basel)* 10(2), 324 (2022).
15. Stark K, Olden M, Brandl C *et al.* The German AugUR study: study protocol of a prospective study to investigate chronic diseases in the elderly. *BMC Geriatr.* 15, 130 (2015).
16. Qiagen. Gentra Puregene handbook (2022). https://www.qiagen.com/de/resources/resourcedetail?id=a9e6a609-4600-4b03-afbd-974318590ce5&lang=en
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).
18. Dhaliwal A. A comprehensive review of DNA extraction and purification kits cited in the literature. *Mater. Methods* 3, 191 (2013).

Reports

BioTechniques

A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities

Justin P Shaffer[1], Carolina S Carpenter[2,3], Cameron Martino[1,3,4], Rodolfo A Salido[1,5], Jeremiah J Minich[6], MacKenzie Bryant[1], Karenina Sanders[1], Tara Schwartz[1], Gregory Humphrey[1], Austin D Swafford[3,7] & Rob Knight*[,1,3,8,9]

*[1]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA; [2]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA; [3]Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA 92093 ,USA; [4]Bioinformatics & Systems Biology Program, University of California San Diego, La Jolla, CA 92093, USA; [5]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA; [6]Marine Biology Research Division, University of California, San Diego, La Jolla, CA 92093, USA; [7]InterOme, Inc. Carlsbad, CA 92008, USA; [8]Micronoma Inc. San Diego, CA 92121, USA; [9]Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093, USA; *Author for correspondence: Tel.: +1 858 822 2379; robknight@ucsd.edu*

BioTechniques 73: 35–46 (July 2022) 10.2144/btn-2022-0032

First draft submitted: 4 March 2022; Accepted for publication: 26 May 2022; Published online: 17 June 2022

ABSTRACT

Microbial communities contain a broad phylogenetic diversity of organisms; however, the majority of methods center on describing bacteria and archaea. Fungi are important symbionts in many ecosystems and are potentially important members of the human microbiome, beyond those that can cause disease. To expand our analysis of microbial communities to include data from the fungal internal transcribed spacer (ITS) region, five candidate DNA extraction kits were compared against our standardized protocol for describing bacteria and archaea using 16S rRNA gene amplicon- and shotgun metagenomics sequencing. The results are presented considering a diverse panel of host-associated and environmental sample types and comparing the cost, processing time, well-to-well contamination, DNA yield, limit of detection and microbial community composition among protocols. Across all criteria, the MagMAX Microbiome kit was found to perform best. The PowerSoil Pro kit performed comparably but with increased cost per sample and overall processing time. The Zymo MagBead, NucleoMag Food and Norgen Stool kits were included.

METHOD SUMMARY

To allow for downstream applications involving fungi in addition to bacteria and archaea, five DNA extraction kits were compared with a previously established, standardized protocol for extracting DNA for microbial community analysis. Across 10 diverse sample types, one extraction kit was found to perform comparably to or better than the standardized protocol. This conclusion is based on per-sample comparisons of DNA yield, the number of quality-filtered sequences generated, the limit of detection of microbial cells, microbial community alpha-diversity, beta-diversity, taxonomic composition and the extent of well-to-well contamination.

KEYWORDS:

Earth Microbiome Project (EMP) • high-throughput sequencing • Katharoseq • Macherey-Nagel • MagAttract PowerSoil • mock community • mycobiome • rRNA • whole genome sequencing

Research into microbial communities continues to reveal links that help support both human and environmental sustainability [1–4]. To identify important connections between microbiota and human health and well-being [5–7], parallel work must foster the innovation of methods that refine our view of microbial communities [4,8–10]. However, widespread adoption of standardized methods for performing microbiome studies continues to be hindered by a lack of approaches that capture information from all organisms or from across diverse sample types [11–13].

Whereas methods for capturing information from bacteria and archaea have been well developed and widely adopted, those that additionally consider microbial eukaryotes such as fungi have received less attention [14–16]. Similar to bacteria and viruses for mammals, fungi are the most important plant pathogens worldwide [17,18] and represent an increasing threat to certain groups of animals, including amphibians [19–21]. Fungi are also invaluable components of soils and forests [22–24], and are currently emerging as important members of the human microbiome (i.e., mycobiome) [25–27]. Currently, our protocol for DNA extraction for high-throughput microbiome sequencing focuses on describing bacterial/archaeal taxa and has not yet been tested to additionally describe fungi.

Here, we aimed to identify an extraction kit that extracts DNA from fungal communities while producing DNA output and community composition for bacteria/archaea similar to our previously established, standardized protocol [28]. We compared DNA yield, the number

of quality sequences, microbial community alpha- and beta-diversity and taxonomic composition, as well as technical differences in the limits of detection of bacterial and fungal cells [29] and the extent of sample-to-sample (i.e., well-to-well) contamination [30–32] among extraction protocols.

Materials & methods

Sample collection

To compare each candidate extraction kit against our standardized protocol, we collected a wide selection of samples from human body sites and the environment, centered on types common in studies of microbial communities, following Marotz *et al.* and Shaffer *et al.* [28,33]. This set of sample types and protocols for collecting each, the 'Earth Microbiome Project (EMP) in a box,' was drafted for widespread use in benchmarking and similar studies [33]. For this study and following Shaffer *et al.*, we included a total of six human skin samples, six human oral samples, four built environment samples, 10 fecal samples, six human urine samples, two human breastmilk samples, six soil samples, four water samples, four fermented food samples and two tissue samples. Except where described otherwise, we collected samples using Puritan wood-handled, cotton swabs following the EMP standard protocol [14,33].

We collected samples in a way that allowed technical replication across extraction protocols (i.e., three technical replicates per protocol), and aliquoted each unique sample across all extraction kits for comparison of extraction efficiency, following Marotz *et al.* and Shaffer *et al.* [28,33]. Human skin samples included those from the foot and armpit, which were collected from three individuals by rubbing five cotton swabs simultaneously on the sole of each foot or armpit, respectively, for 30 s. Human oral samples included saliva, which was collected from 12 individuals by active spitting into a 50-ml centrifuge tube. Built environment samples included floor tiles (0.3 m^2) from each of two separate laboratory bays, which were sampled separately with nine cotton swabs rubbed simultaneously across one tile surface for 30 s, and computer keyboards from each of two individuals also sampled with nine cotton swabs for 30 s. Fecal samples included those from cats, mice and humans. Cat feces were collected from two individuals and stored in plastic zip-top bags. Mouse feces were collected from two individuals by hand using sterile technique and stored in 1.5-ml microcentrifuge tubes. Human feces were collected from five individuals using the Commode Specimen Collection System (cat. no. 02-544-208; Thermo, CA, USA). Human urine was collected from three male individuals separately into 50-ml centrifuge tubes and three female individuals separately first into the Commode Specimen Collection System and then transferred to 50-ml centrifuge tubes using sterile technique. Soil samples included soil from the rhizosphere of trees and bare soil. For each type, soil from two sites at the Scripps Coastal Reserve was collected down to a depth of 20 cm using a sterile trowel and stored in plastic zip-top bags. Water samples included freshwater from two sites at the San Diego River, and seawater from two sites at the Scripps Institution of Oceanography. All water samples were collected and stored in 50-ml centrifuge tubes. Fermented food samples included yogurt and kimchi. For each, two varieties of a single brand were purchased at a local grocery store and transferred to 50-ml centrifuge tubes under sterile conditions. Tissue samples included jejunum tissue from six male mice and six female mice. For each individual, 3.8 cm of the middle small intestine was removed and any particles squeezed out; each tissue section was added to a 2-ml microcentrifuge tube containing 1-ml sterile 1x phosphate-buffered saline and 40 mg sterile 1-mm silicone beads and homogenized at 6000 rpm for 1 min with a MagNA Lyser (Roche Diagnostics, CA, USA); the liquid homogenate from intestinal sections from six mice of one gender was pooled to create a single sample. We stored all samples at -80°C within 3 h of collection. To compare limits of detection of microbial cells across kits [29], we included serial dilutions of a mock community containing both bacterial and fungal taxa (i.e., ZymoBIOMICS Mock Community Standard I, cat. no. D6300; Zymo Research, CA, USA). Input cell densities ranged from 140.00–1.40E+09 cells for bacteria, and 2.66–2.66E+07 cells for fungi. Finally, to compare well-to-well contamination [32], we included plasmid-borne, synthetic 16S rRNA gene spike-ins [34] (i.e., 4 ng of unique spike-in to one well of columns 1–11 in each plate), and at least five extraction blanks per plate.

DNA extraction

We compared our standardized extraction protocol that uses a 96-sample, magnetic bead cleanup format, the Qiagen MagAttract PowerSoil DNA Isolation Kit (cat. no. 27000-4-KF; Qiagen, CA, USA), against five other extraction kits: the Qiagen MagAttract PowerSoil Pro DNA Isolation Kit (cat. no. 47109; Qiagen), the Norgen Stool DNA Isolation Kit (cat. no. 65600; Norgen Biotek, Ontario, Canada), the Applied Biosystems MagMAX Microbiome Ultra Nucleic Acid Isolation Kit (cat. no. A42357; Applied Biosystems, CA, USA), the Macherey-Nagel NucleoMag Food kit (cat. no. 744945.1, Macherey-Nagel, Düren, Germany) and the ZymoBIOMICS 96 MagBead DNA Kit (cat. no. D4302, Zymo Research). We previously showed that the MagMAX Microbiome kit performs comparably or better than our standardized protocol, considering a majority of the criteria included in this benchmark [33]. However, that experiment was focused on establishing the MagMAX kit as an alternative that also allows for downstream RNA-based applications, and did not examine fungi [33]. Importantly, whereas the PowerSoil, PowerSoil Pro, Norgen and MagMAX extraction kits employ a 96-deepwell plate format for sample lysis, the NucleoMag Food and Zymo MagBead extraction kits employ a lysis rack, which instead has 12 eight-tube strips arranged in a 96-well rack. This latter format can potentially reduce well-to-well contamination, which is known to occur primarily during the lysis step [32].

For logistical purposes, extractions were performed in two iterations (hereafter referred to as Round 1 and Round 2), with a fresh set of samples collected in each instance. Both Round 1 and Round 2 included our standardized protocol as a baseline for comparison. Round 1 centered on the Powersoil Pro and the Norgen kits and Round 2 the MagMAX, NucleoMag Food and Zymo MagBead kits. For extraction, and following Marotz *et al.* and Shaffer *et al.* [28,33], aliquots of each sample were transferred to unique wells of a 96-

Table 1. Limits of detection of microbial cells across extraction kits.

Round	Extraction kit	Threshold (%)	16S				ITS			
			LOD	Read depth	Samples retained	Samples retained (%)	LOD	Read depth	Samples retained	Samples retained (%)
1	PowerSoil	50	1.40E+03	73	81	86	2.66E+00	3	94	100
		80	1.40E+05	449	55	59	2.66E+02	223	45	48
		90	1.40E+05	1812	43	46	>2.66E+07	199,755	0	0
		95	1.40E+08	8676	39	41	>2.66E+07	165,759,100,584	0	0
	PowerSoil Pro	50	1.40E+02	46	79	84	2.66E+01	49	70	80
		80	1.40E+02	692	61	65	2.66E+04	1094	39	45
		90	1.40E+06	7520	33	35	2.66E+05	18,669	14	16
		95	>1.40E+09	145,206	0	0	>2.66E+07	698,158	0	0
	Norgen Stool	50	1.40E+07	1960	18	19	2.66E+00	2	88	100
		80	1.40E+07	4705	12	13	2.66E+05	850	7	8
		90	1.40E+08	8224	10	10	>2.66E+07	314,049,798	0	0
		95	1.40E+08	14,215	9	10	>2.66E+07	>1.00E+12	0	0
2	PowerSoil	50	1.40E+02	88	93	97	2.66E+00	3	91	100
		80	1.40E+02	3224	80	83	2.66E+00	6	91	100
		90	>1.40E+09	88,894	0	0	2.66E+00	12	86	95
		95	>1.40E+09	6,228,705	0	0	2.66E+00	24	75	82
	MagMAX Microbiome	50	1.40E+03	1286	79	82	2.66E+00	5	92	100
		80	1.40E+04	4147	66	69	2.66E+00	479	64	70
		90	1.40E+04	8970	63	66	>2.66E+07	451,442	0	0
		95	1.40E+04	19,454	39	41	>2.66E+07	176,467,642,943	0	0
	NucleoMag Food	50	1.40E+05	8379	61	64	2.66E+00	1	88	100
		80	1.40E+05	14,121	45	47	2.66E+00	2	88	100
		90	1.40E+06	19,430	27	28	2.66E+02	74	69	78
		95	1.40E+07	26,321	18	19	>2.66E+07	26,901,301	0	0
	Zymo MagBead	50	1.40E+02	489	42	44	2.66E+00	1	90	100
		80	1.40E+02	1838	32	33	2.66E+00	4	90	100
		90	1.40E+06	4531	22	23	2.66E+01	466	38	42
		95	1.40E+06	11,445	10	10	>2.66E+07	5,179,906,841	0	0

Titrations of a mock community containing known numbers of cells of bacterial and fungal species (see Materials & Methods) were used to identify the number of per-sample reads needed to meet certain LOD thresholds (i.e., the percentage of reads mapped to expected taxa vs contaminants). For each dataset, the read depth corresponding to a threshold of 50% was used for filtering samples before community analyses, as recommended [29,33]. The number and percentage of samples retained after filtering based on the read depth for each threshold and LOD estimates for bacterial and fungal cells are shown for 16S- and fungal ITS data, respectively. Rounds 1 and 2 indicate different sequencing runs; because sampling effort was not normalized such to compare absolute read counts, comparisons should not be made across sequencing runs.
ITS: Internal transcribed spacer; LOD: Limits of detection.

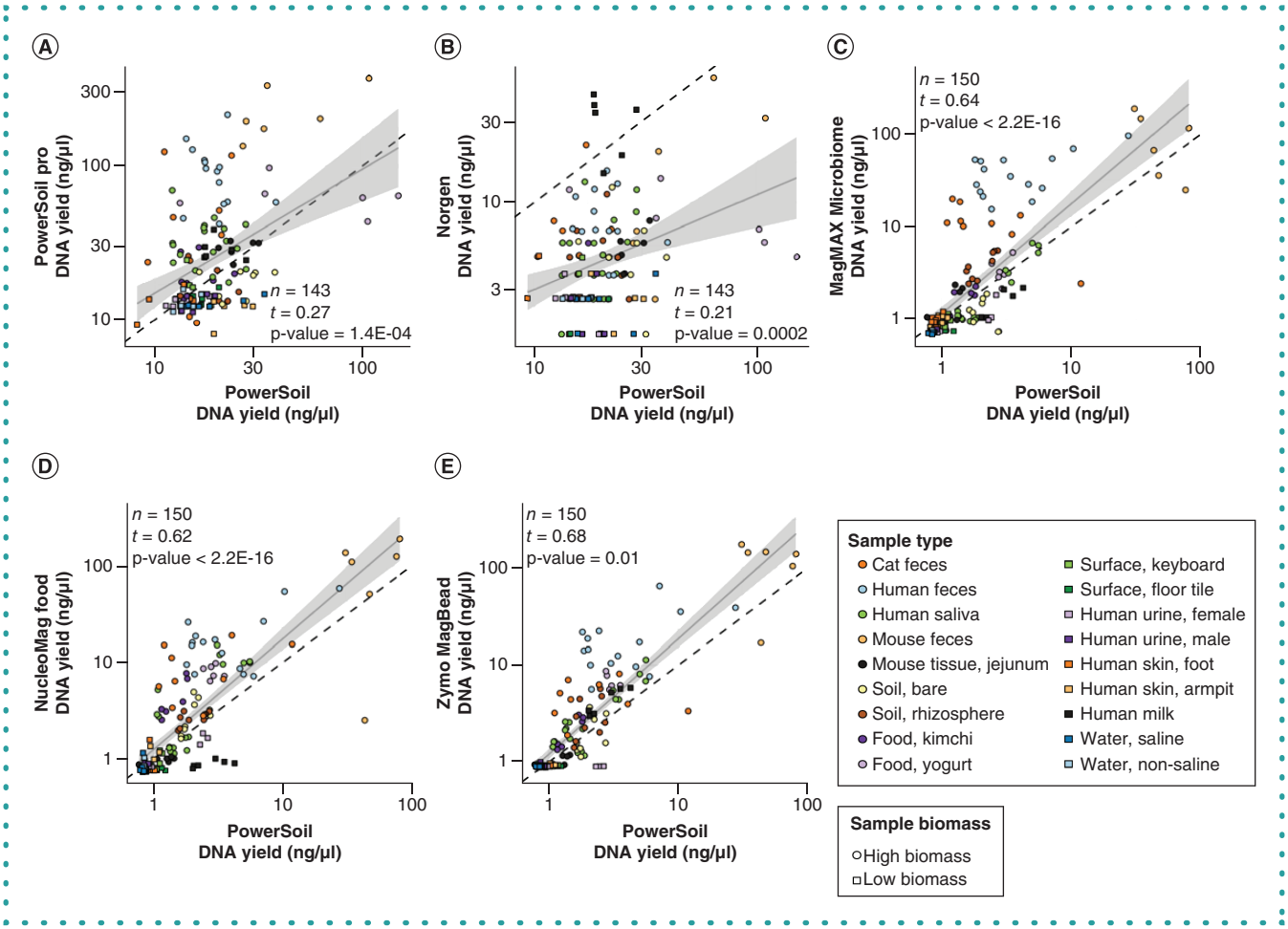


Figure 1. DNA yield (ng/µl) per sample for each candidate extraction kit (y-axis) compared to a standardized protocol (x-axis). (A) PowerSoil Pro. **(B)** Norgen Stool. **(C)** MagMAX Microbiome. **(D)** NucleoMag Food. **(E)** Zymo MagBead. For all panels, colors indicate sample type and shapes sample biomass, and dotted gray lines indicate 1:1 relationships between methods. Results from tests for correlation between per-sample DNA yield for each respective candidate kit versus our standardized protocol are shown (*t* = Kendall's tau). For significant correlations, results from a linear model including a 95% CI for predictions are shown. Both axes are presented in a $\log_{10}$ scale. A miniaturized, high-throughput Quant-iT PicoGreen dsDNA assay was used for quantification, with a lower limit of 0.1 ng/µl. Yields below this value were estimated by extrapolating from a standard curve.

deepwell extraction plate (or lysis rack). For samples collected with swabs, we broke the entire swab head off into the well (or tube). For liquid samples, we transferred 200 µl. For bulk samples, we used cotton swabs to collect roughly 100 mg of homogenized material, and broke the entire swab head off into the well (or tube). For each extraction protocol, all samples, including mock community dilutions, were plated in triplicate. Extractions were performed following the manufacturer's instructions, with lysis performed using a TissueLyser II (Qiagen), and bead clean-ups performed using the KingFisher Flex Purification System (ThermoFisher Scientific, MA, USA). Extracted nucleic acids were stored at -80°C before quantification of DNA yield and subsequent sequencing.

16S rRNA gene, fungal internal transcribed spacer (ITS), and shotgun metagenomics sequencing and data analysis

DNA was prepared for 16S rRNA gene amplicon (16S), fungal internal transcribed spacer (ITS) amplicon, and shallow shotgun metagenomics sequencing as described previously [10,33,35–38]. Extracts from Rounds 1 and 2 were sequenced separately on distinct runs. For bacterial/archaeal 16S data, raw sequence files were demultiplexed using Qiita [39], suboperational taxonomic units (sOTUs) generated using Deblur with the default positive filter for 16S data [40], taxonomy assigned using QIIME2's *feature-classifier* plugin's *classify-sklearn* method with the prefitted classifier trained on GreenGenes V4 (13_8) data [41–44], and phylogeny inferred using QIIME2's *fragment-insertion* plugin's *sepp* method with the GreenGenes (13_8) SEPP reference [45–47]. For fungal ITS data, raw sequence files were demultiplexed using Qiita [39], sOTUs generated using Deblur with a positive filter representing the UNITE8 reference database (dynamic OTUs, including global 97% singletons) [48], and taxonomy assigned using QIIME2's *feature-classifier* plugin's *classify-sklearn* method with a classifier trained on the UNITE8 reference database (described earlier) [41,48]. For shallow shotgun metagenomics data, raw sequence files were demultiplexed using BaseSpace (Illumina, CA, USA) and uploaded to Qiita [39] for additional pre-processing. Demultiplexed

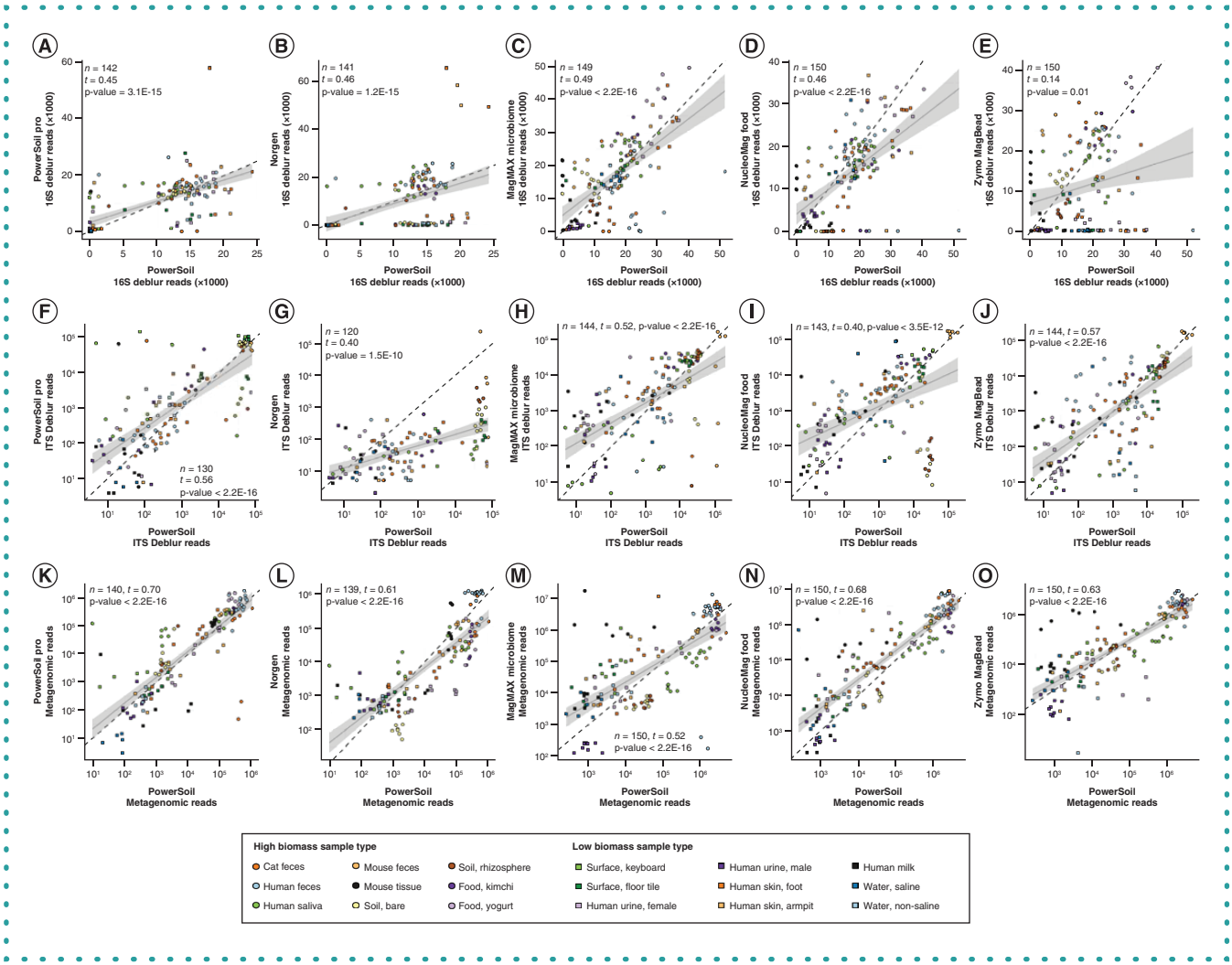


Figure 2. Sequences per sample for each candidate extraction kit (y-axis) compared to a standardized protocol (x-axis). (A–E) 16S data. **(F–J)** Fungal ITS data. **(K–P)** Shotgun metagenomic data. For each panel, colors indicate sample type and shapes sample biomass, and dotted gray lines indicate a 1:1 relationship between methods. For each dataset, results from tests for correlation between read counts from each respective candidate kit versus our standardized protocol are shown (*t* = Kendall's tau). For significant correlations, results from a linear model including a 95% CI for predictions are shown. Both axes are presented in a $\log_{10}$ scale.

sequence data were quality-filtered using fastp [49] and human read depleted by alignment to human reference genome GRCh38 using minimap2 [50]. Filtered reads were aligned to the Web of Life database [51] using bowtie2 [52], and alignment profiles translated to feature-tables using Woltka [53]. Raw sequence data were deposited at the European Nucleotide Archive (accession no. ERP124610), and raw and processed data are available via Qiita (study ID 12201). For all three datasets, subsequent normalization of sampling effort and estimation of alpha- and beta-diversity were performed using QIIME2 [9]. Analyses of taxonomic composition and beta-diversity were performed using custom Python scripts. Correlation tests and Kruskal–Wallis tests were performed in R [54]. All processing and analysis code is available on GitHub (github.com/justinshaffer/Extraction_kit_testing).

Results & discussion

For each of the five candidate extraction kits tested, we observed similar DNA extraction efficiency to our standardized protocol (hereafter referred to as PowerSoil), with the exception of the Norgen kit, which had lower yields across all sample types except human milk (Figure 1 & Supplementary Figure 1A). Across the majority of sample types, the PowerSoil Pro, NucleoMag Food, MagMAX Microbiome and Zymo MagBead kits performed comparably or better than PowerSoil (Figure 1 & Supplementary Figure 1A). We also observed similar trends in the number of quality-filtered reads generated from sequencing for each of the five candidate extraction kits compared with PowerSoil, for 16S, fungal ITS, and shotgun metagenomics data (Figure 2 & Supplementary Figures 1B & 2). Exceptions include the Zymo MagBead kit, which generated fewer high-quality 16S reads across samples from the built environment, water, human urine and human

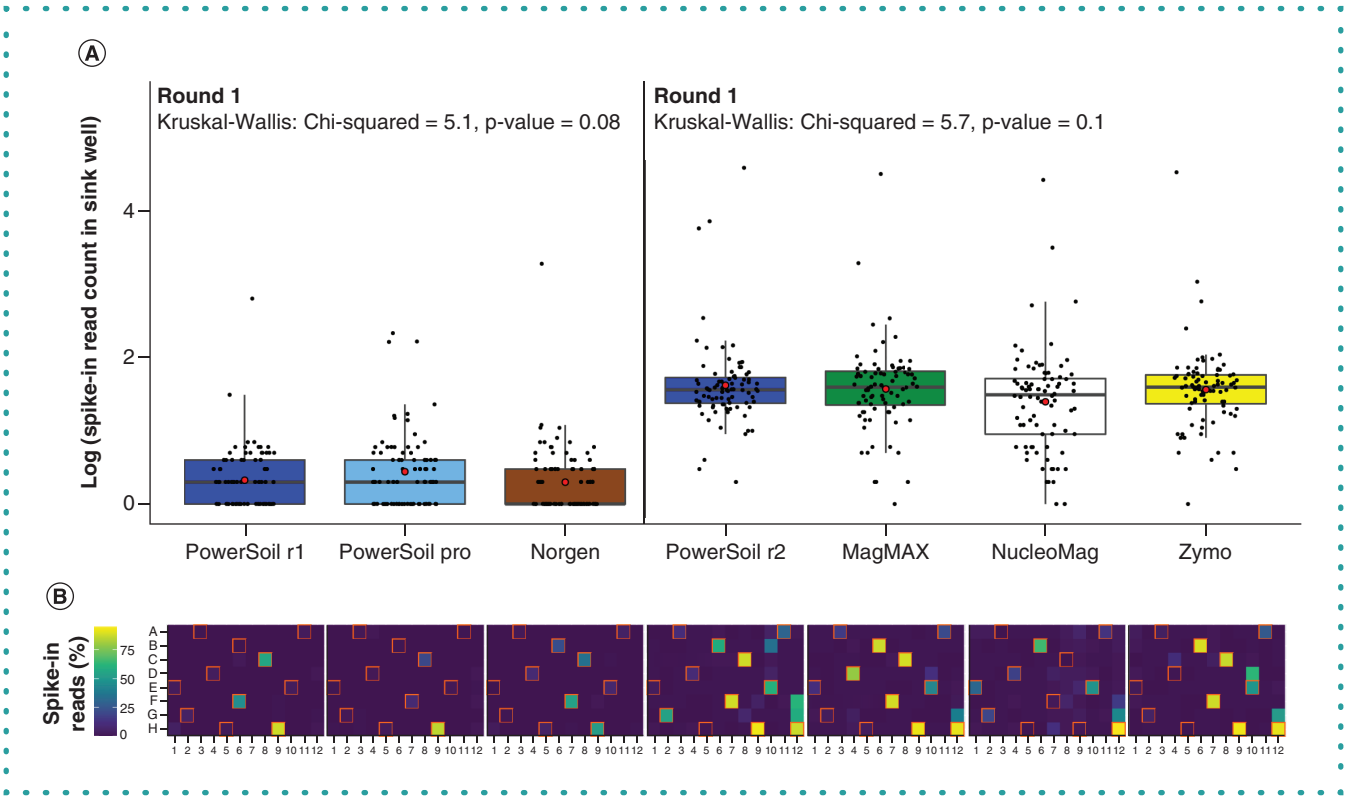


Figure 3. Well-to-well contamination across candidate extraction kits compared to a standardized protocol. Plasmids harboring synthetic 16S sequences were spiked into a single well per plate column (i.e., 1–11) of each high-biomass sample plate before extraction. **(A)** The number of reads matching synthetic 16S sequences was quantified for all wells that did not receive a spike-in (i.e., sink wells). Rounds 1 and 2 indicate different sequencing runs; because sampling effort was not normalized such to compare absolute read counts, comparisons should not be made across sequencing runs. For each sequencing run, results from a Kruskal–Wallis test are shown. **(B)** The percentage of spike-in reads among all reads per well shown as a heatmap. Wells into which plasmids were spiked (i.e., source wells) are outlined in orange.

skin (Figure 2E & Supplementary Figure 1B), and the Norgen kit, which generated fewer high-quality fungal ITS reads across samples from the built environment, food, water and soil (Figure 2G & Supplementary Figure 2A). Interestingly, the reduced performance of the Norgen kit in extracting DNA (Figure 1B) did not influence the number of high-quality shotgun metagenomics reads generated, which was in-line with PowerSoil (Figure 2L & Supplementary Figure 2B).

Considering the limit of detection (LOD) of microbial cells, we observed differences in the ability to detect bacteria versus fungi across the five candidate extraction kits compared with PowerSoil (Table 1). Compared with PowerSoil, the LOD for bacteria was one order of magnitude lower for the PowerSoil Pro kit, the same for the Zymo MagBead kit and one order of magnitude higher for MagMAX Microbiome kit (Table 1). However, when considering sample retention following filtering based on LOD thresholds – an important consideration due to the costs of obtaining/processing samples, and for maintaining reasonable sample sizes for analysis – only the PowerSoil Pro and MagMAX Microbiome kits retain ≥80% of samples, similar to PowerSoil (Table 1). For bacteria, the LOD for the Norgen and NucleoMag Food kits were much higher, implying that they may not be optimal for profiling of rare taxa (Table 1). For fungi, the LOD and frequency of sample retention after filtering were similar for all kits compared with PowerSoil, except for the PowerSoil Pro kit, which had an LOD that was one order of magnitude higher and also retained only 80% of samples compared with 100% across all other protocols (Table 1). Surprisingly, the frequency of well-to-well contamination was similar among protocols (Figure 3). This is especially informative considering the unique lysis rack provided by both the NucleoMag Food and Zymo MagBead kits (see Materials & Methods), which we expected to greatly reduce the frequency of well-to-well contamination compared with lysis in a traditional 96-deepwell plate. We suspect that well-to-well contamination can still occur when using a lysis rack in part due to movement of aerosols during uncapping tubes. We emphasize that without a reduction in well-to-well contamination provided by the lysis rack versus a traditional plate, the roughly 20-fold increase in processing time to open 96 tubes versus to unseal a plate (i.e., ~100 vs 5 s, respectively) argues against adoption of the lysis rack (Figure 3). Future experiments should consider single-tube lysis, which is available for the MagMAX kit (cat. no. A42351) and has been shown to reduce well-to-well contamination [32], although at the cost of increased processing time. Similarly, automated opening and closing of individual, racked tubes, such as that offered by the Matrix Barcoded Storage Tube system (Thermo Scientific), should reduce processing time and potential aerosol transfer.

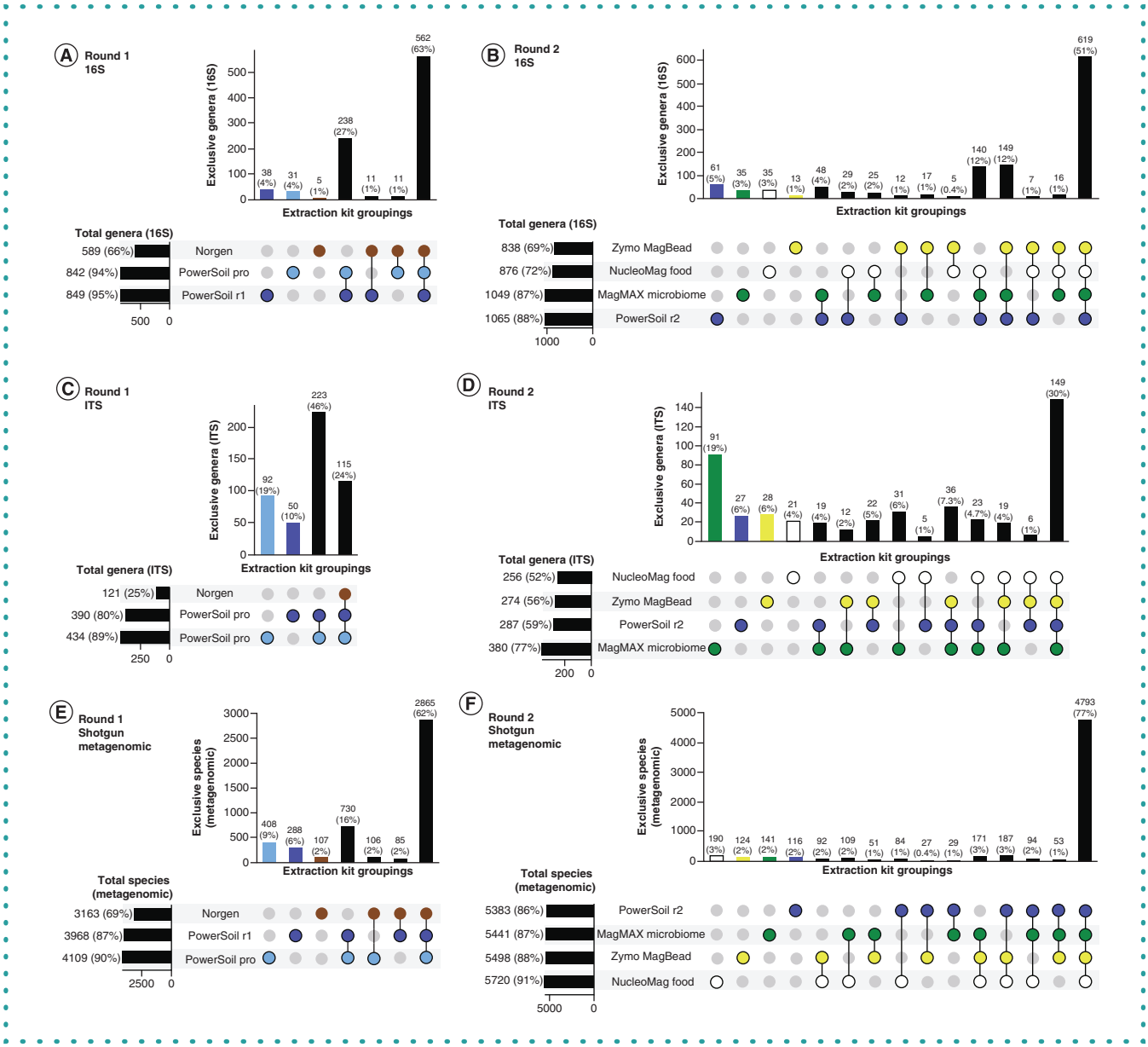


Figure 4. Taxonomic bias among extraction protocols. Upset plots showing **(A & B)** genera for bacterial/archaeal 16S data, **(C & D)** genera for fungal ITS data and **(E & F)** species for bacterial/archaeal metagenomics data, highlighting taxa shared among extraction protocols. Values indicate counts and percentages are respective to all taxa across all protocols. Associations representing fewer than five taxa were excluded for clarity. Rounds 1 and 2 indicate different sequencing runs; because sampling effort was not normalized such to compare absolute taxon counts, comparisons of counts (i.e., vs percentages) should not be made across sequencing runs.

Considering the taxonomic composition of microbial communities across samples, we observed a greater degree of taxon bias among extraction kits compared with PowerSoil for fungal taxa (i.e., ITS data) versus bacterial/archaeal taxa (i.e., 16S and shotgun metagenomics data) (Figure 4C & D). This is likely due in part to the relatively diverse morphologies among fungal spores and propagules compared with those of bacteria/archaea, which may be variably compromised among distinct lysis approaches. Both the PowerSoil Pro and MagMAX Microbiome kits recovered the greatest number of exclusive fungal genera (i.e., those not recovered by other protocols), with each taxon set representing roughly 19% of all fungal genera recovered in a given round of extractions (Figure 4C & D). Similarly, both the PowerSoil Pro and MagMAX Microbiome kits shared a greater number of exclusive fungal genera with PowerSoil, compared with the other candidate extraction kits (Figure 4C & D). We observed a similar trend in our 16S data, except that for both rounds of extraction, PowerSoil recovered the greatest number of exclusive bacterial/archaeal genera compared with any candidate kit (Figure 4A & B). For our shotgun metagenomics data, all candidate extraction kits except for the Norgen kit recovered a greater number of exclusive bacterial/archaeal species than PowerSoil (Figure 4E & F). The PowerSoil Pro and NucleoMag Food kits recovered the

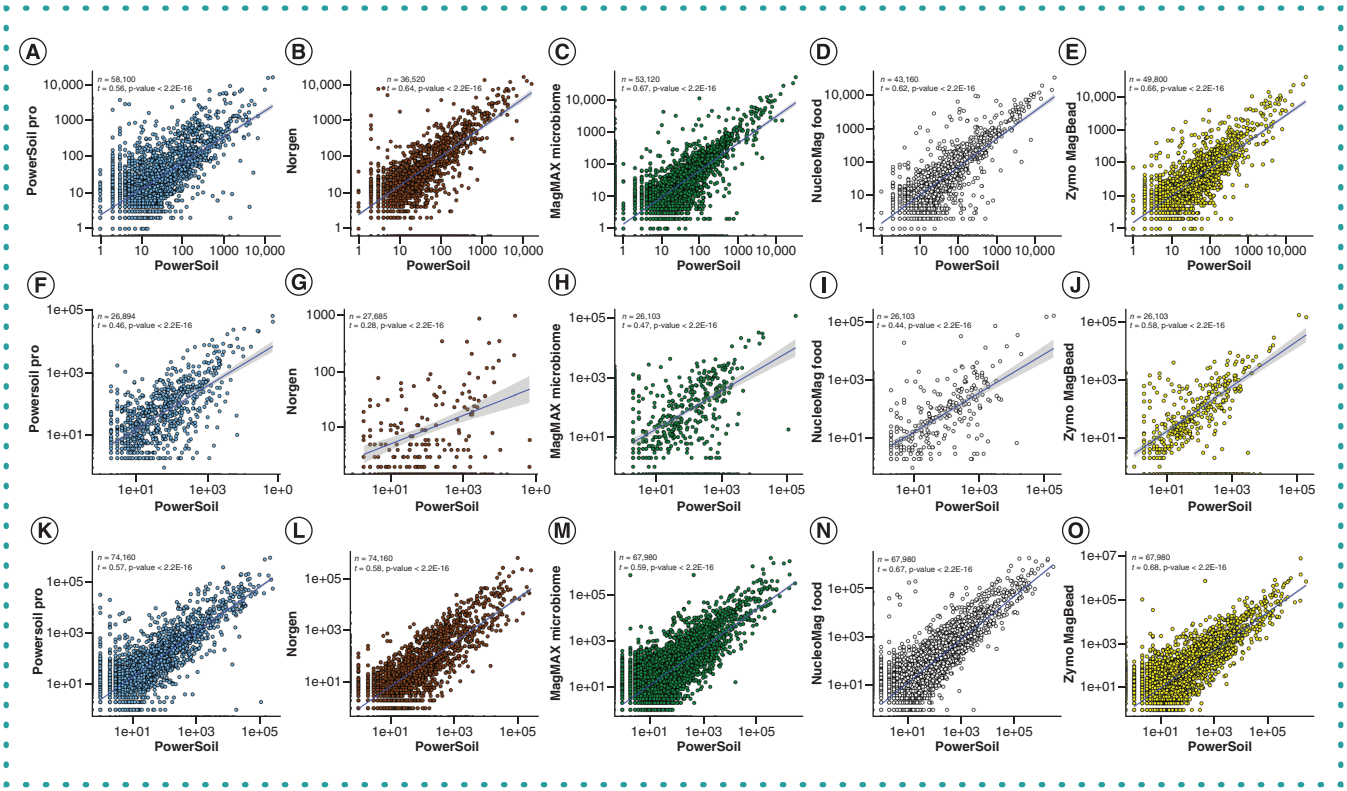


Figure 5. Taxon relative abundances across all samples for each candidate extraction kit (y-axis) compared to a standardized protocol (x-axis). Colors indicate candidate kits. **(A–E)** 16S data, where each point represents a bacterial/archaeal genus. **(F–J)** Fungal ITS data, where each point represents a fungal genus. **(K–O)** Shotgun metagenomic data, where each point represents a bacterial/archaeal genus. For each dataset, results from tests for correlation between taxon abundances from each respective candidate kit versus the standardized protocol are shown (*t* = Kendall's tau). For significant correlations, results from a linear model including a 95% CI for predictions are shown. Both axes are presented in a $\log_{10}$ scale.

greatest percentage of exclusive species, with the MagMAX Microbiome and Zymo MagBead kits recovering only slightly less (Figure 4E & F). The PowerSoil Pro kit shared the greatest percentage of exclusive species with PowerSoil (16%), whereas the NucleoMag Food and MagMAX Microbiome kits shared much less (~1%) (Figure 4F).

For nonexclusive taxa, we observed strong correlations in the relative abundance estimates from each candidate extraction kit compared with PowerSoil (Figure 5). For 16S data, the strongest correlation was observed between PowerSoil and the MagMAX Microbiome kit (Kendall's tau = 0.67) (Figure 5C), followed by the Zymo MagBead kit (tau = 0.66) (Figure 5E) and the Norgen kit (tau = 0.64) (Figure 5B). For fungal ITS data, the strongest correlation was observed with the Zymo MagBead kit (tau = 0.58) (Figure 5J), followed by the MagMAX Microbiome kit (tau = 0.47) (Figure 4H) and the PowerSoil Pro kit (tau = 0.46) (Figure 5F). For shotgun metagenomics data, the strongest correlation was observed with the Zymo MagBead kit (tau = 0.68) (Figure 5O), followed by the NucleoMag Food kit (Kendall's tau = 0.67) (Figure 5N) and the MagMAX Microbiome kit (tau = 0.59) (Figure 5M).

We also observed strong correlations in estimates of microbial community alpha-diversity from each candidate extraction kit compared with PowerSoil (Figure 6A–E), and in general, correlations were stronger for 16S and shotgun metagenomics data compared with ITS data (Figure 6F–J). Specifically, correlations between candidate kits and PowerSoil for 16S alpha-diversity (i.e., Faith's Phylogenetic Diversity [PD]) were all strong (tau > 0.75), except for the Norgen kit, which had the weakest correlation and also the greatest sample dropout from normalization (Figure 6B). Similarly, correlations between candidate kits and PowerSoil for fungal ITS alpha-diversity (i.e., Fisher's alpha) were also strong (tau > 0.60), except for the PowerSoil Pro kit, which had a relatively weak relationship (Figure 6F), and the Norgen kit, which had no relationship and also significant sample dropout (Figure 6G). Correlations between candidate kits and PowerSoil for shotgun metagenomics alpha-diversity (i.e., Faith's PD) were all strong (tau > 0.65), and sample dropout was minimal for all protocols (Figure 6K–O).

With respect to microbial community composition, we found variation explained by bias among extraction protocols to be negligible compared with that explained by host subject identity (i.e., one or two orders of magnitude weaker in explaining beta-diversity) (Table 2). For 16S and shotgun metagenomics data, the variation explained by extraction protocol is one order of magnitude weaker for presence/absence metrics versus abundance-based metrics (Table 2). This supports the analysis of variation among technical replicates from the same sample, which we observed to be small for all extraction protocols across 16S (Supplementary Figure 3), fungal

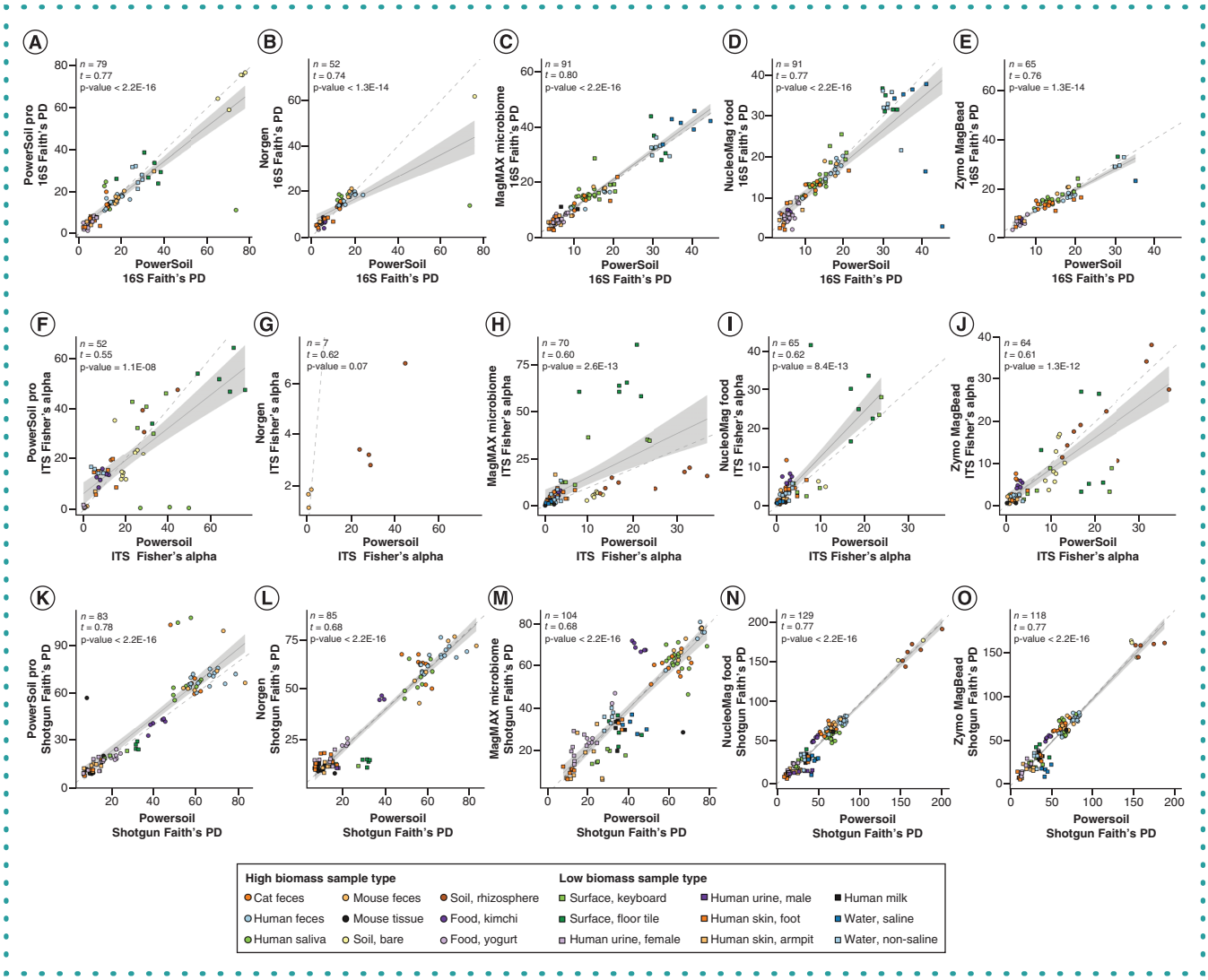


Figure 6. Alpha-diversity per sample across sample types for each candidate extraction kit (y-axis) compared to a standardized protocol (x-axis). (A–E) 16S data. (F–J) Fungal ITS data. (K–P) Shotgun metagenomic data. For each panel, colors indicate sample type and shapes sample biomass, and dotted gray lines indicate a 1:1 relationship between methods. Results from tests for correlation between alpha-diversity values from each respective candidate kit versus our standardized protocol are shown (*t* = Kendall's tau). For significant correlations, results from a linear model including a 95% CI for predictions are shown. Sample types absent from any panel lacked representation by the respective candidate extraction kit and the standardized protocol.

ITS (Supplementary Figure 4) and shotgun metagenomics data (Supplementary Figure 5). We also observed strong correlations in microbial community beta-diversity (i.e., sample–sample distances) from each candidate extraction kit compared with PowerSoil; as for alpha-diversity, in general correlations were stronger for 16S and shotgun metagenomics data compared with ITS data (Supplementary Tables 1–3). Specifically, correlations in sample–sample distances between each of the five candidate kits and PowerSoil for 16S data were strong (rho > 0.75), except for in low-biomass samples processed with the Norgen kit, which exhibited no relationship with PowerSoil for two-of-four distance metrics examined (Supplementary Table 1). For ITS data, correlations in sample–sample distances were consistently weaker for low- versus high-biomass samples, for all candidate extraction kits and distance metrics examined. The PowerSoil Pro kit alone had correlation coefficients >0.50 for low biomass samples and >0.75 for high biomass samples (Supplementary Table 2). For shotgun metagenomics data, correlations in sample–sample distances were also consistently weaker for low- versus high-biomass samples, whereas the magnitude of the difference was smaller compared with ITS data, and correlations for high-biomass samples were strong (rho > 0.85) for all candidate extraction kits and distance metrics examined (Supplementary Table 3).

Importantly, whereas agreement with PowerSoil regarding the results of analyzing 16S or shotgun metagenomics data is desirable, deviation from PowerSoil among the five candidate extraction kits based on fungal ITS data was expected. In that regard, the MagMAX Microbiome kit alone consistently maintains a high degree of correlation with PowerSoil for both 16S and shotgun metagenomics

Table 2. Assessment of factors influencing microbial community beta-diversity in this study.

Data type	Distance metric	Factor	Adjusted R^2	df	AIC	F	p-value
16S	Unweighted UniFrac	Host identity	0.93	60	-979.83	152.03	0.0002
	Unweighted UniFrac	Extraction kit	0.001	6	-988.95	3.20	0.0004
	Weighted UniFrac	Host identity	0.84	60	-428.10	58.62	0.0002
	Weighted UniFrac	Extraction kit	0.02	6	-533.75	19.27	0.0002
	Jaccard	Host identity	0.96	60	-1234.51	231.05	0.0002
	Jaccard	Extraction kit	0.002	6	-1254.66	4.92	0.0002
	RPCA	Host identity	0.89	60	-644.89	86.15	0.0002
	RPCA	Extraction kit	0.004	6	-663.81	4.73	0.0002
ITS	Jaccard	Host identity	0.80	59	-191.58	33.90	0.0002
	Jaccard	Extraction kit	0.02	6	-235.28	8.51	0.0002
	RPCA	Host identity	0.72	59	-30.05	22.32	0.0002
	RPCA	Extraction kit	0.02	6	-62.07	6.64	0.0002
Metagenomic	Unweighted UniFrac	Host identity	0.92	65	-1063.58	141.03	0.0002
	Unweighted UniFrac	Extraction kit	0.005	6	-1107.59	8.79	0.0002
	Weighted UniFrac	Host identity	0.87	65	-672.15	81.26	0.0002
	Weighted UniFrac	Extraction kit	0.02	6	-807.08	24.46	0.0002
	Jaccard	Host identity	0.93	65	-1195.00	168.79	0.0002
	Jaccard	Extraction kit	0.003	6	-1225.27	6.57	0.0002
	RPCA	Host identity	0.85	65	-579.25	70.95	0.0002
	RPCA	Extraction kit	0.01	6	-633.10	10.40	0.0002

Results are from forward, stepwise model selection, following Shaffer et al. (2021) [33]. Values are based on permutation tests of variation explained by redundancy analysis (n = 5000 runs), done separately for unique distance metrics for 16S, the fungal ITS, and shotgun metagenomic data. The full model included extraction round (i.e., Round 1 vs 2), sample biomass (i.e., high vs low biomass), sample type, host subject identity and extraction kit as model variables. 16S data were rarefied to 10,000 quality-filtered reads per sample or had samples with fewer than 10,000 reads excluded when using RPCA distances (n = 640 samples). Fungal ITS data were rarefied to 630 quality-filtered reads per sample or had samples with fewer than 630 reads excluded when using RPCA distances (n = 978 samples). Shotgun metagenomic data were rarefied to 2100 host- and quality-filtered reads per sample or had samples with fewer than 2100 reads excluded when using RPCA distances (n = 1044 samples). Rarefaction depths were selected to maintain at least 75% samples (50% for fungal ITS data) from both high- and low-biomass datasets.
AIC: Akaike information criterion; df: degrees of freedom; ITS: Internal transcribed spacer; RPCA: Robust principal components analysis (i.e., Robust Aitchison distance).

data (Figure 2C & M, Figure 5C & M, Figure 6C & M) while also maintaining a relatively high number of samples from both high- and low-biomass subsets following normalization across all three data layers (Supplementary Tables 1–3). The NucleoMag Food and Zymo MagBead kits have slightly stronger correlations with PowerSoil compared with MagMAX for shotgun metagenomics data for certain analyses of microbial community diversity (Figure 6M–O & Supplementary Tables 1 & 3). However, the MagMAX Microbiome kit recovered the greatest number of exclusive fungal genera while also sharing the greatest number of exclusive fungal genera with PowerSoil (Figure 4D). Although the PowerSoil Pro kit also recovered a similar frequency of exclusive fungal genera (Figure 4C), the increased sample dropout following normalization for that kit versus the MagMAX Microbiome kit (i.e., particularly for low-biomass samples; Supplementary Tables 1–3), increased processing time (i.e., ~3.5 h for PowerSoil and PowerSoil Pro vs 1.0 h for MagMAX Microbiome) as well as the increased cost of consumables (i.e., 3.5x reagent reservoirs and 7x tips for PowerSoil and PowerSoil Pro vs MagMAX Microbiome) combined with previous work showing that the MagMAX kit can also extract high-quality RNA from similar samples [33], argue strongly for the use of the MagMAX Microbiome kit. However, the PowerSoil Pro kit is a good alternative if there are no downstream RNA applications and if time and cost are not important factors.

Conclusion

We conclude that the MagMAX Microbiome extraction kit is comparable to our standardized PowerSoil protocol with respect to characterizing microbial community composition using both 16S and shotgun metagenomic data, and is optimal compared with other candidate kits and our standardized protocol for doing so using fungal ITS data, as it recovers the greatest number of unique fungal genera. The PowerSoil Pro kit is a good alternative, as it excels in the same regards, but it does not extract RNA and is a more time- and cost-intensive protocol compared with the MagMAX Microbiome kit. Regardless, data from the PowerSoil, PowerSoil Pro and MagMAX Microbiome extraction kits should allow for comparisons such as meta-analysis across 16S, ITS and shotgun metagenomics data produced using those protocols and downstream processing and analytical methods similar to those used here. In addition to recovering a greater number of fungal taxa, the more rapid processing time, and use of fewer consumables highlight the MagMAX Microbiome kit as a comparable and efficient alternative to the PowerSoil protocol that also allows for downstream applications including fungi.

Future perspective

Future efforts should continue to focus on optimizing microbiological and molecular methods that capture all organisms in a sample regardless of their evolutionary history and from a diversity of sample types such as those examined here. Such methods provide invaluable resources and should serve as gold standards to be adopted widely by the community [55]. In parallel, further advances in computational methods should focus on reducing technical effects in meta-analyses across studies using distinct methods [6,39]. In concert, such advances will allow us to maximize our understanding of microbial communities and to harness that knowledge to foster human and environmental sustainability.

Executive summary

- Our previously established, standardized protocol for DNA extraction was compared against five alternative DNA extraction kits.
- A diverse panel of sample types was included, ranging from host-associated to environmental.
- Controls for detecting well-to-well contamination and the limit of detection of microbial cells were also included.
- Sample-type-specific differences in DNA extraction efficiency was observed among all extraction protocols.
- Sample type and host identity were stronger drivers of microbial community beta-diversity compared with the extraction protocol used.
- One protocol that generates high-quality data for the fungal internal transcribed spacer (ITS) region and produces 16S and shotgun metagenomics data with high similarity to our established protocol with respect to microbial community alpha-diversity, beta-diversity and taxonomic composition was identified.
- The similarity between the optimal protocol and our existing one will allow for meta-analyses across both with negligible technical bias.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2022-0032

Author contributions

JP Shaffer, CS Carpenter, RA Salido, JJ Minich, G Humphrey, AD Swafford and R Knight designed the study; JP Shaffer, CS Carpenter, RA Salido, M Bryant and G Humphrey provided samples; JP Shaffer, CS Carpenter and T Schwartz performed extractions; RA Salido, M Bryant, K Sanders and G Humphrey performed quality control and sequencing; JP Shaffer, C Martino and RA Salido performed data analyses; JP Shaffer wrote the manuscript, with contributions from all authors.

Acknowledgments

The authors thank S Miller-Montgomery for helping to conceptualize the study; C Marotz, P Belda-Ferre, L Bode, C Brennan, M Estaki, T Kosciolek, SJ Song, J Toronczak and L Zaramela for providing samples; the Scripps Coastal Reserve for permitting collection of soil samples; L Zaramela for providing synthetic plasmids for well-to-well contamination assay; J DeReus for assistance with data processing; G Ackermann for curating metadata and making the data publicly available; and SJ Song and KJ Cavagnero for assistance with visualizing taxon abundance correlations.

Financial & competing interests disclosure

JP Shaffer was supported by NIH San Diego Institutional Research and Academic Career Development Award (5K12GM068524-17) and the US Department of Agriculture – National Institute of Food and Agriculture (USDA-NIFA) (2019-67013-29137). C Martino was supported by the NIH (1RF1-AG058942-01) and Semiconductor Research Corporation and Defense Advanced Research Projects Agency (SRC/DARPA) (GI18518). JJ Minich was supported by the NSF (2011004). G Humphrey was supported by the NIH (U19AG063744,U01 AI124316), the Office of Naval Research (ONR) (N00014-15-1-2809) and the Emerald Foundation (3022). R Knight was supported by the NIH (1RF1-AG058942-01, 1DP1AT010885, R01HL140976, R01DK102932, R01HL134887), USDA-NIFA (2019-67013-29137), SRC/DARPA(GI18518), CCFA (675191), ONR (N00014-15-1-2809) and the Emerald Foundation (3022). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The human subjects work conducted here is approved through UCSD IRB#150275.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

Accession numbers

Raw sequence data were deposited at the European Nucleotide Archive (accession no.: ERP124610) and raw and processed data are available at Qiita (study ID: 12201). All processing and analysis code is available on GitHub (github.com/justinshaffer/Extraction_kit_testing).

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Kelly CR, Ihunnah C, Fischer M *et al.* Fecal microbiota transplant for treatment of *Clostridium difficile* infection in immunocompromised patients. *Am. J. Gastroenterol.* 109(7), 1065–1071 (2014).
2. Panke-Buisse K, Poole AC, Goodrich JK *et al*. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* 9, 980–989 (2015).
3. Bell TH, Stefani FOP, Abram K *et al.* A diverse soil microbiome degrades more crude oil than specialized bacterial assemblages obtained in culture. *Appl. Environ. Microbiol.* 82(18), 5530–5541 (2016).
4. Mills RH, Dulai PS, Vásquez-Baeza Y *et al.* Multi-omics analysis of the ulcerative colitis gut microbiome links *Bacteroides vulgatus* proteases with disease severity. *Nat. Microbiol.* 7, 262–276 (2022).
5. Dominguez-Bello MG, Godoy-Vitorino F, Knight R, Blaser MJ. Role of the microbiome in human development. *Gut* 68, 1108–1114 (2019).
6. Poore GD, Kopylova E, Zhu Q *et al.* Microbiome analysis of blood and tissues suggest cancer diagnostic approach. *Nature* 579, 567–574 (2020).
7. Erdman S. Oxytocin and the microbiome. *Curr. Opin. Endocrin. Metabol. Res.* 19, 8–14 (2021).
8. Knight R, Vrbanac A, Taylor BC *et al.* Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422 (2018).
9. Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible, interactive, scalable, and extensible microbiome data science using QIIME2. *Nat. Biotechnol.* 37, 852–857 (2019).
10. Sanders JG, Nurk S, Salido RA *et al.* Optimizing sequencing protocols for leaderboard metagenomics by combining long and short reads. *Genome Biol.* 20, 1–14; 226 (2019).

•• **This work is important for enabling cost-effective, high-throughput shotgun metagenomics for describing microbial taxa.**

11. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18(5), 1403–1414 (2016).
12. Bjerre RD, Hugerth LW, Boulund F *et al*. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. *Sci. Rep.* 9, 1–11; 17287 (2019).

• **This work is important for understand the influence of technical factors on analysis of microbial communities.**

13. Marotz C, Cavagnero KJ, Song SJ *et al.* Evaluation of the effect of storage methods on fecal, saliva, and skin microbiome composition. *mSystems* 6(2), e01329-20 (2021).

•• **This study highlights the importance of considering how samples are maintained before processing for analysis of microbial communities.**

14. Thompson LR, Sanders JS, McDonald D *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463 (2017).
15. Lloyd-Price J, Arze C, Ananthakrishnan AN *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel disease. *Nature* 569, 655–662 (2019).
16. Nayfach S, Roux S, Seshadri R *et al.* A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* 39, 499–509 (2020).
17. Dean R, Van Kan JAL, Pretorius ZA *et al.* The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* 13(4), 414–430 (2012).
18. Bebber DP, Gurr SJ. Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet. Biol.* 74, 62–64 (2015).
19. Seyedmousavi S, Bosco SMG, Hoog S *et al.* Fungal infections in animals: a patchwork of different situations. *Med. Mycol.* 56, S165–S187 (2018).
20. Scheele BC, Pasmans F, Skerratt LF *et al.* Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science* 363(6434), 1459–1463 (2019).
21. Cheng TL, Reichard JD, Coleman JTH *et al.* The scope and severity of white-nose syndrome on hibernating bats in North America. *Conserv. Biol.* 35(5), 1586–1597 (2021).
22. Ma Z, Guo D, Xu X *et al.* Evolutionary history resolves global organization of root functional traits. *Nature* 555, 94–97 (2018).
23. Simard SW. Mycorrhizal networks facilitate tree communication, learning, and memory. In: *Memory and Learning in Plants.* Baluska F, Gagliano M, Witzany G (Eds). Signaling and Communication in Plants, Springer, Cham (2018).
24. Tedersoo L, Bahram M, Zobel M. How mycorrhizal associations drive plant population and community biology. *Science* 367(6480), 867 (2020).
25. Aykut B, Pushalkar S, Chen R *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature* 574, 264–267 (2019).
26. Coker OO, Nakatsu G, Dai RZ *et al.* Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut* 68, 654–662 (2019).
27. Liu NN, Jiao N, Tan JC *et al.* Multi-kingdom microbiota analysis identify bacterial-fungal interactions and biomarkers of colorectal cancer across cohorts. *Nat. Microbiol.* 7, 238–250 (2022).
28. Marotz C, Amir A, Humphrey G *et al.* DNA extraction for streamlined metagenomics of diverse environmental samples. *Biotechniques* 62(6), 290–293 (2017).

•• **This work is important for understanding differences in DNA extraction methods, particularly different bead clean-up methods, for downstream analysis of microbial communities.**

29. Minich JJ, Zhu Q, Janssen S, Hendrickson SR *et al.* KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems* 3(3), 1–16; e00218.17 (2018).

•• **This work is important for understanding differences in DNA extraction methods for regular or high- versus low-biomass sample types for analysis of microbial communities.**

30. Salter SJ, Cox MJ, Turek EM *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analysis. *BMC Biol.* 12, 1–12; 87 (2014).
31. Eisenhofer R, Minich JJ, Marotz C *et al*. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol.* 27(2), 105–117 (2019).
32. Minich JJ, Sanders JG, Amir A *et al*. Quantifying and understanding well-to-well contamination in microbiome research. *mSystems* 4(4), 1–8; e00186-19 (2019).

• **This work is important for understanding the influence of technical factors such as well-to-well contamination on results of high-throughput sequencing for microbiome studies.**

33. Shaffer JP, Marotz C, Belda-Ferre P *et al.* A comparison of DNA/RNA extraction protocols for high-throughput sequencing of microbial communities. *Biotechniques* 70, 149–159 (2021).
34. Tourlousse DM, Yoshiike S, Ohashi A *et al*. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res.* 45(4), 1–14; e23 (2016).
35. Caporaso JG, Lauber CL, Walters WA *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624 (2012).
36. Minich JJ, Humphrey G, Benitez RAS *et al.* High-throughput miniaturized 16S rRNA amplicon library preparation reduces costs while preserving microbiome integrity. *mSystems* 3(6), 1–8; e00166-18 (2018).
37. Hillmann B, Al-Ghalith GA, Shields-Cutler RR *et al.* Evaluating the information content of shallow shotgun metagenomics. *mSystems* 3(6), 1–12; e00069-18 (2018).
38. Shaffer JP, Nothias LF, Thompson LR *et al.* Multi-omics profiling of Earth's biomes reveals that microbial and metabolite composition are shaped by the environment. *bioRxiv* doi: 10.1101/2021.06.04.446988 (2022).
39. Gonzalez A, Navas-Molina JA, Kosciolek T *et al.* Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* 15, 796–798 (2018).
40. Amir A, McDonald D, Navas-Molina JA *et al.* Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2(2), 1–7; e00191-16 (2017).
41. Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in python. *J. Machine Learn. Res.* 12, 2825–2830 (2011).
42. Robeson MS, O'Rourke DR, Kaehler BD *et al.* RESCRIPt: Reproducible sequence taxonomy reference database management for the masses. *bioRxiv* doi: 10.1101/2020.10.05.326504 (2020).
43. Bokulich NA, Kaehler BD, Rideout JR *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME2's q2-feature-classifier plugin. *Microbiome* 6, 90 (2018).
44. McDonald D, Price MN, Goodrich J *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analysis of bacteria and archaea. *ISME J.* 6, 610–618 (2012).
45. Eddy SR. Accelerated profile hmm searches. *PLoS Computation. Biol.* 7(10), 1–16 (2011).
46. Janssen S, McDonald D, Gonzalez A *et al.* Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems* 3(3), e00021-18 (2018).
47. Matsen FA, Kodner RB, Armbrust EV. Pplacer: linear time maximum likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformat.* 11(1), 538 (2010).
48. Abarenkov K, Zirk A, Piirmann T *et al.* UNITE QIIME release for fungi. Version 10.05.2021. UNITE Community doi: 10.15156/BIO/1264708 (2021).

49. Shifu C, Yanqing Z, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformat.* 34(17), i884-i890 (2018).
50. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformat.* 34(18), 3094–3100 (2018).
51. Zhu Q, Mai U, Pfeiffer W *et al.* Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. *Nat. Commun.* 10, 1–14; 5477 (2019).
52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4), 357–359 (2012).
53. Zhu Q, Huang S, Gonzalez A *et al.* Phylogeny-aware analysis of metagenome community ecology based on matched reference genomes while bypassing taxonomy. *mSystems* 7(2), e0167-22 (2022).
54. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2022). www.R-project.org
55. Greathouse KL, Sinha R, Vogtmann E. DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol.* 20, 2–4; 212 (2019).

Reports

BioTechniques

An optimal genomic DNA extraction method for shoots of four *Dendrocalamus* species based on membership function analysis

Yingchun Ma[‡,1], Bowen Li[‡,1], Dejia Yang[1], Shuguang Wang[1,2,3], Lixia Yu[1,2], Hui Zhan[1,2] & Juan Li[*,1,2]

[1]*College of Life Sciences, Southwest Forestry University, Kunming, 650224, China;* [2]*Yunnan Sympodial Bamboo Key Laboratory, Southwest Forestry University, Kunming, 650224, China;* [3]*Institute of Bamboo & Rattan Science, Southwest Forestry University, Kunming, 650224, China; *Author for correspondence: lijuan@swfu.edu.cn;* [‡]*Authors contributed equally*

BioTechniques 76: 00–00 (March 2023) 10.2144/btn-2023-0087

First draft submitted: 1 September 2023; Accepted for publication: 28 November 2023; Published online: 22 December 2023

ABSTRACT

High-quality genomic DNA extraction is fundamental for the study of gene cloning and expression in plants. Therefore, this study evaluated several methods for extracting genomic DNA from shoots of four *Dendrocalamus* species to determine the optimal technique. Genomic DNA was extracted using three different methods: a commercial DNA extraction kit method, a modified cetyltrimethylammonium bromide method and a sodium dodecyl sulfate method. A membership function analysis was employed to compare these methods. The results demonstrated that the commercial DNA extraction kit method was the most effective and comprehensive approach for extracting genomic DNA from shoots of four *Dendrocalamus* species. Furthermore, this study provided valuable insights into optimizing techniques for extracting genomic DNA in other bamboo species.

TWEETABLE ABSTRACT

The efficiency of three DNA extraction methods was assessed for the shoots of four *Dendrocalamus* species, with the commercial DNA extraction kit method demonstrating optimal performance.

METHODS SUMMARY

Genomic DNA was extracted from the shoots of four *Dendrocalamus* species (sample mass: 0.1 g; n = 3/species/method) using three different methods: a commercial DNA extraction kit method, a modified cetyltrimethylammonium bromide method and a sodium dodecyl sulfate method. The DNA yield, integrity and purity were evaluated for each method, and membership function analysis was employed to comprehensively assess the advantages and disadvantages of each approach.

KEYWORDS:

bamboo • *Dendrocalamus* • extraction • genomic DNA • membership function method

According to collected data, a total of 116 genera and 1439 species of bamboo plants have been found worldwide, with 62% being native to Asia. China possesses the most abundant and diverse bamboo species. In China, there have been reports of a total of 861 species and infraspecific taxa in 43 genera, including 707 species, 52 varieties, 98 forma and four hybrids. These bamboos are naturally distributed across 21 provinces, municipalities and autonomous regions [1]. *Dendrocalamus* is a large bamboo genus in the Gramineae family that holds significant economic and environmental value [2]. Notably, *Dendrocalamus brandisii*, *Dendrocalamus giganteus*, *Dendrocalamus asper* and *Dendrocalamus hamiltonii* are widely distributed in Yunnan Province in China and produce high-quality bamboo shoots. Furthermore, the bamboo shoots of *D. brandisii* and *D. hamiltonii* exhibit distinct sweetness when consumed raw; however, the precise molecular mechanisms underlying this phenomenon remain unexplored but hold great research potential. Previous studies on bamboo plants have mainly concentrated on physiological structure [3] and physical and chemical properties [4]. Specifically, studies of *Dendrocalamus* have mostly focused on anatomical structure [5] and cultivation techniques [6]. However, discussions on genomic DNA extraction methods for bamboo are scarce, which hinders progress in bamboo molecular biology research.

Nucleic acid extraction represents a fundamental technique in molecular biology, with the quality of the extracted material playing a pivotal role in determining the success or failure of subsequent experiments. The classical methods for DNA extraction from plant tissue primarily include the cetyltrimethylammonium bromide (CTAB) and sodium dodecyl sulfate (SDS) methods, which have been demonstrated to be effective in extracting high-quality DNA [7,8]. Additionally, DNA extraction kits are commonly used to isolate plant DNA. Currently, solid-phase extraction techniques utilizing nanomaterials and magnetic solvents are extensively utilized to facilitate the efficient extraction of high-quality nucleic acids from samples containing metal ions and proteins [9–11].

The majority of plant tissues contain cellulose and are abundant in metabolites such as phenolic compounds, proteins and polysaccharides [12]. As gramineous monocots, bamboo shoots also possess a substantial amount of cellulose fiber, polysaccharides and proteins [13], which pose challenges for the extraction and purification of genomic DNA. Hence, it is imperative to identify the optimal method for extracting high-quality genomic DNA from bamboo shoots to ensure the success of subsequent experiments.

A membership function analysis is a comprehensive approach that evaluates various indicators. For example, it is commonly used to comprehensively evaluate the drought resistance of plants [14–16]. This approach can also be employed for comprehensive assessment of DNA extraction methodologies. Therefore, in this study, we utilized four species of *Dendrocalamus* shoots and conducted a membership function analysis to evaluate three genomic DNA extraction methods with the aim of identifying their respective advantages and disadvantages and ultimately determining the most cost-effective and convenient method for obtaining high-quality DNA for subsequent molecular experiments.

Materials & methods

Sample collection & processing

Shoots from four species of bamboo – namely, *D. giganteus*, *D. asper*, *D. hamiltonii* and *D. brandisii* – with an average length of approximately 20 cm were collected from the Pu'er City bamboo garden in Yunnan Province (longitude and latitude: 101.143309 and 22.752514). Immediately after excavation, the bamboo shoots were sectioned into small pieces, rapidly frozen using liquid nitrogen for transportation to the laboratory and subsequently stored in an ultra-low temperature freezer. Each nucleic acid extraction method group consisted of three replicates per species, each weighing 0.1 g. The extracted DNA was dissolved in 40 μl of diethyl pyrocarbonate water and refrigerated at -20°C.

Instruments & reagents

The following instruments were utilized: SS-325 high-pressure steam sterilizer and DYY-6C electrophoresis instrument (Beijing Liuyi Instrument Factory, Beijing, China); CS601 Super Thermostatic Water Bath (Shanghai Buxun Industrial Co., Ltd., Shanghai, China); SQP electronic balance (Sartorius Scientific Instruments [Beijing] Co., Beijing, China); LGIOOB oven and BCD-215TD GA refrigerator (Qingdao Haier Co., Ltd., Qingdao, China); YDS-20 liquid nitrogen tank and centrifuge tubes (Shanghai Rongpeng Information Technology Co., Ltd., Shanghai, China); gel imager (Bio-Rad Laboratories, Inc., CA, USA); and CF16RX II floor-standing high-speed centrifuge (Hitachi Koki Co., Ltd., Tokyo, Japan). The following reagents were employed: chloroform, isopropanol and anhydrous ethanol (procured from a unified school source) and 75% ethanol, 2% CTAB buffer (spermidine), chloroform/isoamyl alcohol (24:1), 3 mol/l sodium acetate, TE buffer, deionized water and a laboratory-prepared reagent consisting of 20% SDS.

Genomic DNA extraction methods

Modified CTAB extraction method

Approximately 0.1 g of pre-ground frozen bamboo was utilized for DNA extraction following the modified protocol of Zhang *et al.* [17]. The procedure involved the following steps: 1) placing 1 ml of 2% CTAB solution in a preheated water bath at 65°C until it became transparent and then adding 20 μl β-thioethanol (β-thioethanol is an improved reagent in the original test scheme and added to increase the extraction efficiency. NaCl was added to SDS extraction method for the same purpose.); 2) adding the pre-ground sample to a new 2-ml centrifuge tube containing 750 μl prewarmed 2% CTAB, thoroughly mixing and incubating for 30 min in a water bath set at 65°C; 3) incorporating 750 μl chloroform/isoamyl alcohol (24:1), ensuring proper mixing, and subsequently centrifuging at 12,000 rpm for 10 min; 4) aspirating and transferring the supernatant into another autoclaved, pipette-equipped centrifuge tube with the addition of RNase (10 mg/ml) at a volume of 0.5 μl and incubating this mixture for 20 min at 37°C; 5) introducing 750 μl chloroform/isoamyl alcohol (24:1), thoroughly mixing and then centrifuging at 12,000 rpm for 10 min; 6) transferring 500 μl of supernatant to a new 1.5-ml centrifuge tube, adding 700 μl of precooled isopropanol and vortexing to mix thoroughly; 7) incubating on ice for 10 min; 8) centrifuging at 12,000 rpm and 4°C for 15 min; 9) washing twice with 1 ml of precooled 75% ethanol; and 10) allowing the ethanol to fully evaporate until the pellet became translucent and then dissolving the dried pellet in deionized water. The final step was conducted on an ultra-clean benchtop.

DNA kit extraction method

A plant genomic DNA extraction kit (catalog no. CW0531) developed by Jiangsu Kangwei Century Biotechnology Co., Ltd. (Beijing, China) was employed for the isolation of DNA from approximately 0.1 g of pre-ground bamboo samples. The manufacturer's protocol was strictly followed during the DNA extraction procedure.

SDS extraction method

The SDS extraction method consisted of the following steps. First, 0.1 g of pre-ground frozen sample was added to a 2-ml centrifuge tube followed by the addition of 600 μl SDS extraction solution composed of 10 ml of 1 mol/l Tris HCl, 10 ml of 5 mol/l NaCl, 10 ml of 0.5 mol/l EDTA, 6.2 ml of 20% SDS, 30.38 g of $NaHSO_3$ and deionized water up to a final volume of 100 ml. Second, the centrifuge tube was incubated in a water bath at 65°C for 1 h with intermittent manual shaking to enhance extraction efficiency. Third, after cooling to room temperature, emulsification was achieved by adding and mixing thoroughly with 500 μl of chloroform/isoamyl alcohol (24:1). Fourth,

centrifugation was performed at a speed of 10,000 rpm and maintained at 4°C for 10 min. Fifth, equivalent volumes of supernatant and chloroform/isoamyl alcohol (24:1) were pipetted into a new tube and mixed by rotating the tube around its axis. This mixture was then incubated at room temperature for 30 min followed by centrifugation at 10,000 rpm and 4°C for 10 min. This step was then repeated. Sixth, the supernatant was transferred into a new centrifuge tube, and 2× the volume of anhydrous ethanol was slowly added to the supernatant to precipitate DNA. Seventh, centrifugation was performed at 12,000 rpm and 4°C for 5 min. Eighth, 1 ml of precooled 70% alcohol was added to the centrifuge tube. Ninth, centrifugation was performed at 10,000 rpm and 4°C for 5 min, and then the supernatant was discarded and the tube was placed on an ultra-clean bench to dry. Tenth, after drying, the DNA pellet was dissolved in a solution containing TE buffer (60 μl) and RNase (1 μl of a concentration of 10 mg/ml), incubating in a water bath at 37°C for 30 min. Eleventh, 180 μl of TE butter was mixed with 240 μl of chloroform/isoamyl alcohol (24:1), and the mixture was then added to the aforementioned solution, mixed well and centrifuged at 13,500 rpm and 4°C for 5 min. Twelfth, the supernatant was transferred to another centrifuge tube, to which one-tenth the volume of 3 mol/l sodium acetate (pH 5.5) and twice the volume of prechilled 95% ethanol were added. The mixture was incubated at 4°C for 30–40 min. Thirteenth, after incubation, the mixture was subjected to centrifugation at 12,000 rpm and 4°C for another 5 min. Following this step, the supernatant was discarded. Fourteenth, the DNA pellet obtained from step three was washed with 1 ml of 75% ethanol, and step 13 was repeated. Fifteenth, the centrifuge tube containing the DNA pellet was placed on an ultra-clean benchtop and allowed to dry until it became translucent. Sixteenth, the pellet was dissolved in TE buffer (50 μl) and stored at -20°C.

Genomic DNA integrity

The integrity of DNA was evaluated by analyzing the molecular weight and mass of DNA bands following electrophoresis on 1% agarose gel. Electrophoresis samples were prepared by combining 5 μl of DNA sample with 1 μl of 6× loading buffer along with a marker measuring 100–10,000 kb. Gel electrophoresis was conducted for a duration of 30 min at 120 V followed by visualization using a gel imager.

Genomic DNA purity & concentration

The purity and concentration of the DNA sample were assessed using a NanoDrop ultraviolet spectrophotometer (Thermo Fisher Scientific, MA, USA) by measuring the optical density (OD) values at 260/280 nm with 1 μl of the sample.

Data processing & analysis

The data were sorted using Excel 2016 (Microsoft Corporation, WA, USA), and statistical analysis and calculation of membership function values for each index were performed using SPSS Statistics 21 (IBM Corporation, NY, USA). The membership function value was evaluated in accordance with the computational formula

$$\mu(X_i) = \frac{(X_i - X_{\min})}{(X_{\max} - X_{\min})} \quad i = 1,\ 2,\ 3,\ \cdots,\ n$$

where Xi represents the i-th comprehensive index, whereas X_{max} denotes the maximum value within this index and X_{min} indicates the minimum value. The magnitude of individual comprehensive indices

$$W_i = P_i / \sum_{i=1}^{n} P_i \quad i = 1,\ 2,\ 3,\ \cdots,\ n$$

where W_i represents the weight assigned to the i-th comprehensive index within the entire set of comprehensive indices, whereas P_i denotes the contribution rate of the i-th comprehensive index attributed to various extraction methods. The Pi value was calculated by SPSS. Evaluation of comprehensive metrics for various extraction methodologies

$$D = \sum_{i=1}^{n} [\mu(x_i) \times w_i] \quad i = 1,\ 2,\ 3,\ \cdots,\ n$$

where the D value is the comprehensive evaluation score of different DNA extraction methods obtained by calculating the comprehensive index.

Results

Genomic DNA quality

An electrophoresis method was used to evaluate the integrity of the genomic DNA. The DNA bands obtained from all four methods appeared as single, intact and compact entities, thereby demonstrating the efficacy of each method. Notably, DNA extracted from the shoots of *D. brandisii* and *D. giganteus* using the DNA extraction kit method exhibited superior clarity and brightness without any signs of degradation. Conversely, the modified CTAB and SDS methods resulted in narrower and darker bands, indicating relatively lower DNA

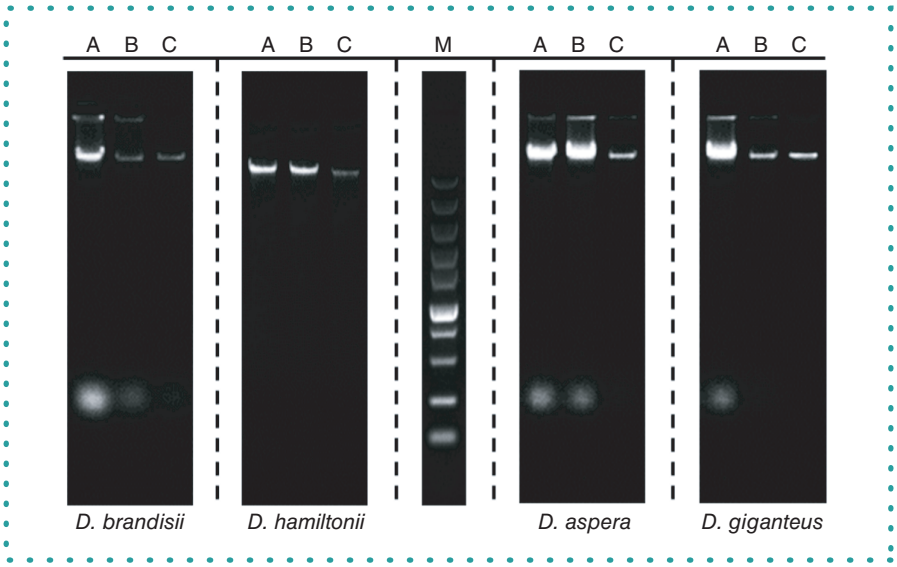


Figure 1. Electrophoresis gel map of genomic DNA extracted from shoots of four *Dendrocalamus* species. (A) DNA extraction kit. **(B)** Modified cetyltrimethylammonium bromide method. **(C)** Sodium dodecyl sulfate method. **(D)** Marker (100–10,000 kb).

Table 1. Purity and concentration of genomic DNA extracted from bamboo shoots.

Experimental material	Extraction method	OD 260/280	Concentration (ng/μl)
Dendrocalamus giganteus	DNA kit	1.84 ± 0.01	186.53 ± 16.82
	CTAB	1.88 ± 0.05	121.93 ± 25.41
	SDS	1.84 ± 0.02	69.49 ± 4.77
Dendrocalamus brandisii	DNA kit	1.96 ± 0.11	57.07 ± 9.05
	CTAB	1.97 ± 0.21	34.47 ± 6.08
	SDS	2.23 ± 0.08	38.90 ± 18.57
Dendrocalamus asper	DNA kit	1.9 ± 0.05	198.80 ± 28.23
	CTAB	1.85 ± 0.06	138.5 ± 20.68
	SDS	1.87 ± 0.07	77.74 ± 6.89
Dendrocalamus hamiltonii	DNA kit	1.93 ± 0.07	50.00 ± 12.31
	CTAB	1.96 ± 0.12	47.39 ± 17.07
	SDS	2.16 ± 0.10	27.37 ± 5.62

Data are presented as mean ± standard error.
CTAB: Cetyltrimethylammonium bromide; OD: Optical density; SDS: Sodium dodecyl sulfate.

concentrations. By contrast, for shoots of *D. hamiltonii* and *D. asper*, the DNA extraction kit method and modified CTAB method yielded similar clear and bright DNA bands; however, inferior extraction efficiency was evident with the SDS method, resulting in overall reduced brightness (Figure 1).

The concentrations of four bamboo shoots obtained through DNA extraction using three distinct methods were analyzed, revealing a strong correlation between concentration and strip brightness. Notably, the DNA extraction kit method yielded the highest DNA concentration. For the shoot of *D. brandisii*, no significant difference was observed between the modified CTAB and SDS methods. However, for the other three bamboo shoots, the modified CTAB method yielded significantly higher DNA concentration compared with the SDS method (Figure 2).

Sample tissues contain varying amounts and types of secondary metabolites, including polyphenols, polysaccharides and proteins, which can impact the quality of DNA. Generally, OD 260/280 ratios ranging from 1.8 to 2.1 indicate optimal purity and high-quality DNA [17]. Ratios exceeding 2.1 suggest potential RNA contamination, whereas ratios below 1.8 indicate possible protein and phenol contamination [7]. The DNA extraction kit yielded pure products with a high concentration. The OD 260/280 ratios (indicating DNA purity) remained consistent across species and methods overall. However, the SDS method resulted in lower DNA concentrations unsuitable for subsequent molecular experiments (Figure 3 & Table 1).

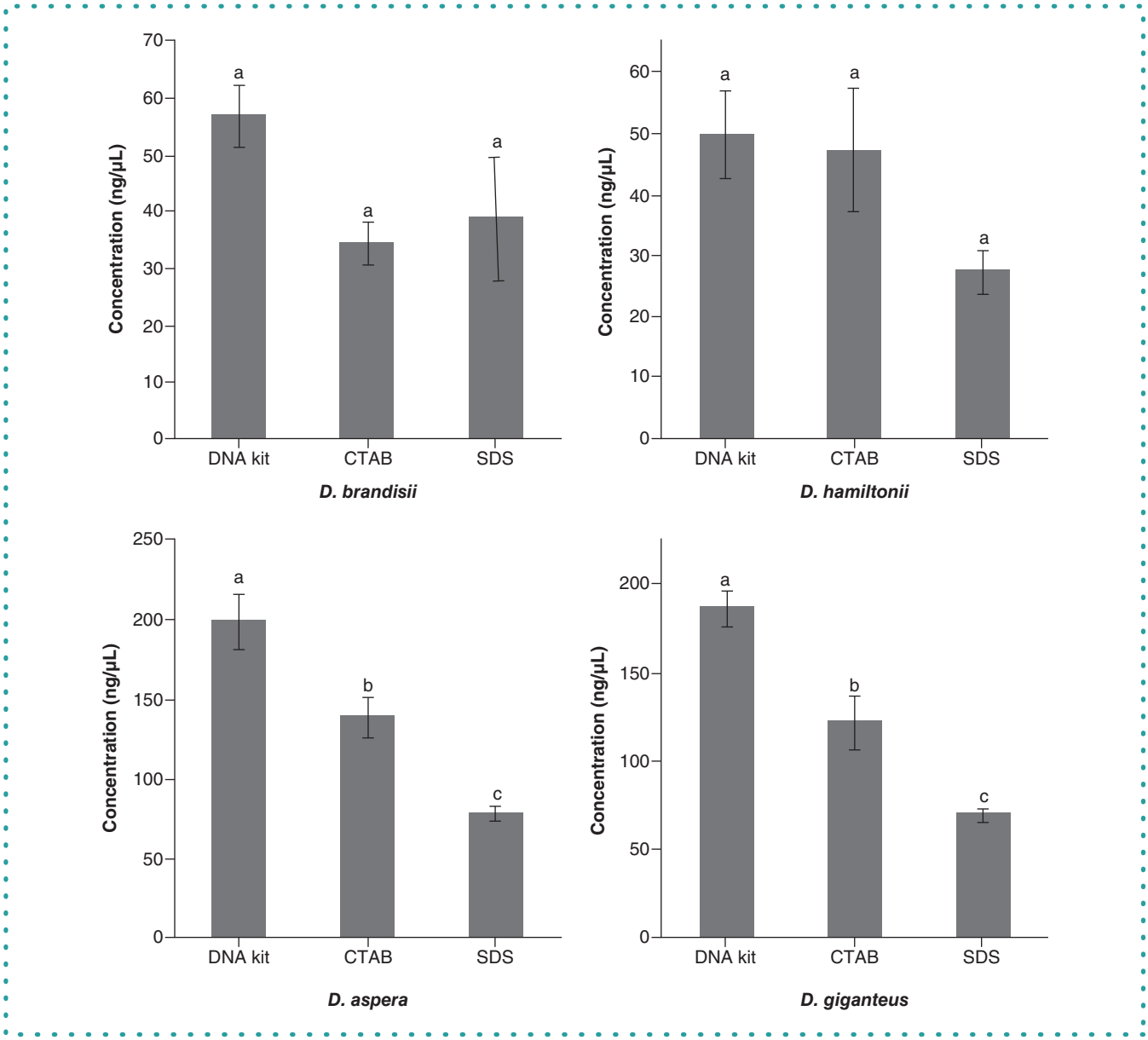


Figure 2. Comparison of DNA concentrations in bamboo shoots of four *Dendrocalamus* species by three extraction methods. Data are presented as mean ± standard error. Statistical differences ($p < 0.05$) exist between letters (i.e., a, b and c).

Polymerase chain reaction of DNA product

The qualities of the extracted DNA were analyzed using polymerase chain reaction (PCR) (Figure 4). Clear target bands without evident tailing were observed in the DNA extracted from all species. Hence, the quality of DNA fulfilled the requirements for conventional PCR experiments, indicating it was suitable for subsequent molecular biology assays.

Cost analysis

The time required, cost, protocol complexity and DNA integrity and concentration were compared (Table 2). Although the complete DNA extraction kit could extract only 50 plant samples, making it relatively expensive (¥7.96/sample), it offered the shortest protocol time (1.5 h) and demonstrated superior efficiency among the three methods. By contrast, both the CTAB and SDS methods had a lower cost per sample, at ¥0.37 and ¥0.42, respectively; however, their protocols were intricate and time-consuming, significantly impacting DNA quality. Despite a slightly lower concentration compared with the kit method, DNA extracted using the CTAB method exhibited an OD value within the range of 1.8–2.1, meeting molecular experiment requirements. Additionally, PCR experiments confirmed that DNA extracted using the CTAB method adequately fulfilled the prerequisites for subsequent molecular experiments. Therefore, to obtain the

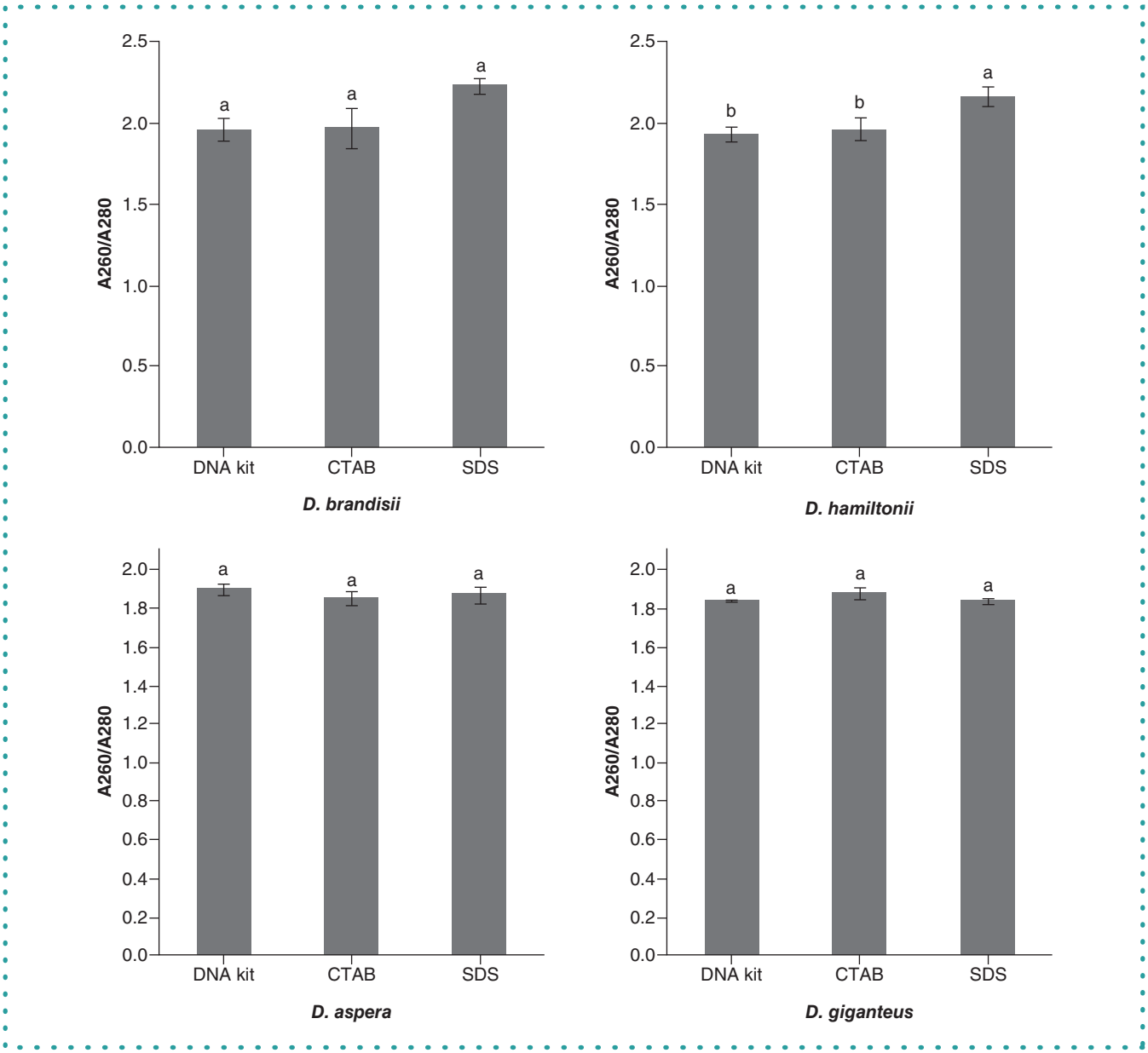


Figure 3. Comparison the optical density values of DNA in bamboo shoots of four *Dendrocalamus* species by three extraction methods. Data are presented as mean ± standard error. Statistical differences ($p < 0.05$) exist between letters (i.e., a and b).

Table 2. Cost analysis of extraction methods.

Extraction method	Operation	Time (h)	Cost per sample (¥)	*Dendrocalamus giganteus*		*Dendrocalamus asper*		*Dendrocalamus hamiltonii*		*Dendrocalamus brandisii*	
				A†	B‡	A†	B‡	A†	B‡	A†	B‡
DNA kit	Simple	1.5	7.96	GR	GR	GR	GR	GR	GR	GR	GR
CTAB	Complex	6	0.37	GR	MR	GR	MR	GR	MR	GR	MR
SDS	Complex	8	0.42	MR	PR	MR	PR	MR	PR	MR	PR

†DNA integrity.
‡DNA concentration.
CTAB: Cetyltrimethylammonium bromide; GR: Good results; MR: Moderate results; PR: Poor results; SDS: Sodium dodecyl sulfate.

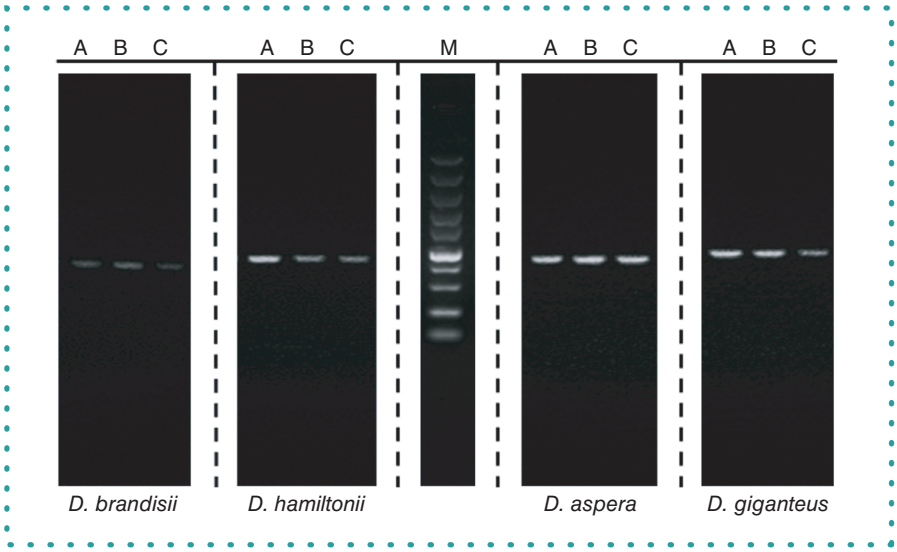


Figure 4. Electrophoresis gel map of polymerase chain reaction products amplified from DNA extracted from shoots of four *Dendrocalamus* species. (A) DNA extraction kit. (B) Modified cetyltrimethylammonium bromide method. (C) Sodium dodecyl sulfate method. (D) Marker (100–10,000 kb).

Table 3. Membership function analysis of extraction methods.

Extraction method	OD 260/280	Extraction efficiency†	Cost per sample (¥)	DNA concentration (ng/μl)	Membership function value	Sort‡
DNA kit	1.908 ± 0.05	16	7.96	123.1 ± 80.535	0.974	1
CTAB	1.915 ± 0.059	4	0.37	85.572 ± 52.258	0.302	2
SDS	2.025 ± 0.199	3	0.42	53.375 ± 24.077	0.143	3

Data are presented as mean ± standard error unless otherwise noted.
† Number of samples extracted at the same time (h).
‡ Sort according to the D values computed in the membership function.
CTAB: Cetyltrimethylammonium bromide; OD: Optical density; SDS: Sodium dodecyl sulfate.

high-purity DNA necessary for advanced experiments, we recommend utilizing a DNA extraction kit. However, for experiments with less stringent requirements, the CTAB method could serve as an alternative for DNA extraction.

Extraction method comparisons by membership function analysis

According to the formula for calculating D-values of membership functions mentioned in the experimental method, a comprehensive analysis was conducted on three DNA extraction methods and their respective D-value scores were calculated. The closer the score is to 1, the higher the comprehensive score of this method is, and the more suitable it is for extracting DNA from bamboo shoots of *Dendrocalamus*. Among the assessed methods, the membership function value of the DNA extraction kit was closest to 1 (0.974); this was followed by the modified CTAB method (0.302) and the SDS method (0.143) (Table 3). Consequently, the kit method demonstrated optimal performance followed by the CTAB method, whereas the SDS method displayed inferior results.

Discussion

DNA extraction is a fundamental technique in molecular biology and an essential prerequisite for gene-based studies. The acquisition of high-purity and high-quality DNA serves as the cornerstone for various molecular applications, including restriction enzyme digestion, PCR amplification, molecular hybridization, genetic polymorphism analysis and genomics research. However, the presence of interfering substances such as polyphenols, polysaccharides and proteins in plant tissues poses a significant challenge to isolating and purifying DNA from plants compared with animals and bacteria [18–20]. Furthermore, because of their high sugar and protein concentrations, extracting bamboo shoots presents even greater difficulties. These secondary metabolic components coexist with DNA [20], resulting in a sticky gel that hampers dissolution while causing browning and degradation of the extracted DNA fragments. Additionally, laborious extraction processes can lead to loss of DNA fragments, which adversely affects the quality and purity of the obtained DNA.

Chemical components and other characteristics vary among plants; thus, the optimal method for genomic DNA extraction also varies depending on the plant species [21]. Currently, CTAB and SDS are widely employed as DNA extraction methods. CTAB is a cationic detergent that effectively precipitates DNA and polysaccharides at low ion concentrations. However, at high ion concentrations (>0.7 mol/l NaCl), it binds to proteins and polysaccharides but cannot precipitate nucleic acids. During DNA extraction, organic solvents are used

to eliminate impurities such as proteins, polysaccharides and phenols followed by alcohol precipitation of the DNA. Generally, the CTAB extraction method demonstrates high efficiency for isolating citrus plant DNA [22] as well as successfully extracting DNA from geranium plants [20] as well as various other botanical species [21]. We used a modified CTAB method to extract genomic DNA from bamboo shoots, resulting in pure DNA at high concentrations, albeit slightly lower than the kit method. However, the cost of the CTAB method was lower than that of the DNA extraction kit method, making it suitable for bulk genomic DNA extractions, whereas the kit method was better suited for small-scale extractions. SDS is an ionic detergent that precipitates nucleic acids and acidic polysaccharides from weak ion solutions. EDTA in SDS chelates divalent metal ions, inhibits RNA enzyme activity and prevents DNA degradation [23]. High-quality genomic DNA for subsequent PCR assays has been obtained from *Populus tomentosa* leaves using the SDS method [24]. This method is also optimal for extracting genomic DNA from *Turpinia arguta* leaves [25]. However, when extracting DNA from *Rhododendron*, impurities and low concentrations have been observed [26], consistent with our results. The SDS method performed worst among all three methods in terms of extracting DNA from bamboo shoots.

The CTAB and SDS methods possessed additional drawbacks: the protocols were complicated, with many steps, and time-consuming and exhibited high reagent toxicity, rendering them unsuitable for large-scale DNA extraction. Hence, it is of the utmost importance to enhance these extraction techniques in order to achieve an efficient, rapid and low-toxicity method for obtaining high-quality DNA.

Based on the traditional CTAB extraction method, the modified CTAB extraction method has been applied in other people's research. For instance, β-thioethanol and moderate ammonium acetate were added to improve DNA extraction from dry *Carum carvi* leaves [27]. Furthermore, ascorbic acid and Triton X-100 were added for rinsing before cell lysis for a genomic DNA extraction protocol from the leaves of Fagaceae plants, which removed polyphenols and polysaccharides, resulting in good quality and integrity [28].This modified CTAB extraction method can also be tried in other plant DNA extraction. DNA was also extracted from *Polygonatum odoratum* using a modified CTAB method that added polyvinylpyrrolidone reagent and β-thioethanol [29], resulting in a product suitable for PCR amplification. In this study, we initially added β-thioethanol to modify the CTAB method for DNA extraction from bamboo shoots. However, the integrity of the total DNA bands extracted was low. Therefore, only adding β-thioethanol to improve the traditional CTAB extraction method can not significantly improve the extraction efficiency of DNA, and it needs to be improved in many aspects. Bamboo shoots were found to be abundant in carbohydrates, and this modified version of the CTAB method proved insufficient for completely eliminating all polysaccharides. Moreover, it was time-consuming and prone to impurities. The CTAB method was employed in one study for DNA extraction from *Bambusa vulgaris* and *Guadua angustifolia*, yielding satisfactory extraction results [8]. In that study, NaCl was incorporated to enhance the final salt concentration, effectively removing polysaccharides and improving extraction outcomes. The DNA extracted using the CTAB method fulfilled the requirements for subsequent PCR experiments and complied with molecular biology experiment standards regarding DNA quality. Therefore, future experiments should consider utilizing the modified CTAB method for extracting DNA from *P. odoratum* and *Passiflora edulis*, as it may facilitate extraction from sweet bamboo shoots.

Genomic DNA was extracted from the shoots of four *Dendrocalamus* species using three different techniques, and the integrity, purity and concentration of each were assessed. Among these methods, the commercial DNA extraction kit yielded the highest quality DNA in the shortest amount of time. Furthermore, this kit eliminated the need for reagents such as phenol and chloroform, simplifying and enhancing the safety of the protocol. However, it should be noted that this method demonstrated a significantly higher cost compared with the other two techniques. The DNA extraction kit proved suitable for modular DNA extraction with small sample sizes and high-precision experiments because of its ability to provide high-quality results. Nevertheless, such as the need to extract a large number of samples of DNA at the same time, or the experiment doesn't require high-quality DNA, alternative approaches such as the modified CTAB extraction method. Although slightly lower in terms of achieved DNA concentration compared with the kit method, the DNA extracted by the modified CTAB extraction method can meet the DNA quality required by the general experiment, and its price is more economical. Conversely, despite its affordability and ability to perform parallel extractions on multiple samples simultaneously, the SDS method resulted in poor-quality DNA unsuitable for subsequent experimentation.

In summary, the DNA extraction kit method produced high-quality DNA at a higher cost and was well suited for high-precision DNA extraction. The modified CTAB method could meet daily experimental requirements and serve as an alternative to the DNA extraction kit method for extracting DNA. However, utilization of the SDS method for bamboo DNA extraction yielded suboptimal results and should be avoided in bamboo-specific extractions.

The quality of DNA extraction was significantly influenced by various factors. In this study, to avoid the need for repeated freezing and thawing of samples, they were promptly ground and stored in an ultra-low temperature refrigerator. To minimize DNA degradation, the entire experiment was conducted on ice and maintained at a low temperature. To minimize the impact of human factors, the entire process was carried out by a single individual, and numerous pre-experiments were performed to ensure the stability and accuracy of the test results. The entire process strictly complied with relevant guidelines for DNA extraction. Sterile gloves and masks were worn during the sample collection, processing and DNA extraction steps to ensure rigor and applicability of the obtained results.

Conclusion

The results demonstrated that, in comparison with the CTAB and SDS methods, the DNA extraction kit method exhibited superior efficacy in extracting DNA from bamboo shoots, especially for experiments involving small sample sizes and high DNA requirements. However,

for experiments with lower DNA requirements, CTAB could serve as an optimal alternative to DNA extraction kit extraction. Furthermore, these methodologies provide valuable insights for optimizing genomic DNA extraction from other plant species.

Future perspective

In future experiments, the DNA extraction kit method is still the most suitable method to extract genomic DNA from bamboo plants and other plants. Because it can perform efficient DNA extraction and can provide high-quality DNA. Therefore, it continues to be a favorable choice for subsequent investigations.

Executive summary

Background

- Isolation of high-quality DNA from bamboo shoots is a challenging process because of their high cellulose fiber, polysaccharide and protein content.
- Identifying the optimal method for extracting high-quality genomic DNA from bamboo shoots is crucial for subsequent experiments.

Materials & methods

- We compared three methods for extracting genomic DNA from the shoots of four *Dendrocalamus* species: a commercial DNA extraction kit, the modified cetyltrimethylammonium bromide (CTAB) method and the sodium dodecyl sulfate method.
- We used membership function analysis to assess the advantages and disadvantages of each method based on DNA yield, integrity and purity.

Results & discussion

- The commercial DNA extraction kit method exhibited superior efficacy in extracting DNA from the shoots of *Dendrocalamus* species compared with the CTAB and sodium dodecyl sulfate methods. However, for experiments with a lower requirement for DNA purity, the modified CTAB method could be chosen as a result of its affordability.

Conclusion

- The commercial DNA extraction kit was the most suitable for extracting DNA from the shoots of *Dendrocalamus* species in this experiment, whereas the improved CTAB method could serve as an alternative.

Author contributions

Y Ma and B Li acted as co-first authors. Y Ma mainly conducted experiments, performed data analysis, created illustrations and wrote and revised the manuscript. B Li mainly conducted experiments and wrote the manuscript. Y Dejia mainly conducted experiments. S Wang, L Yu and H Zhan provided ideas for revising the manuscript. J Li provided experimental ideas and directions, manuscript writing directions and opinions regarding revision.

Financial disclosure

This research was supported by the Xingdian Talent Support Plan Project (XDYC-QNRC-2022-0229), Natural Science Foundation of Yunnan Province (202001AT070133, 202001AT070108) and National Key Research and Development Program of China (2021YFD2200503). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with an interest in or conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

References

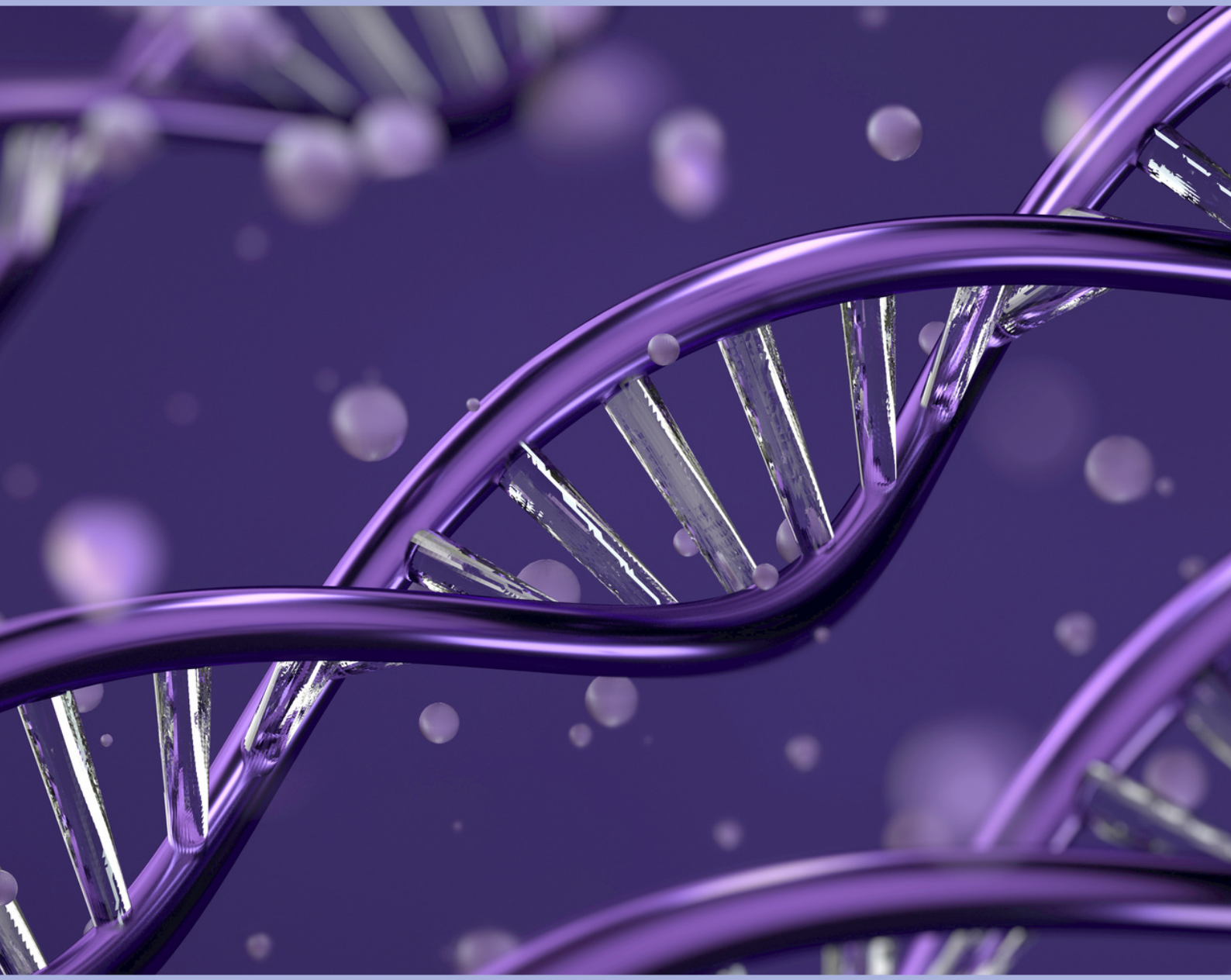
Papers of special note have been highlighted as: • of interest

1. **Khali HA. *Bamboo: Current and Future Prospects.* Books on Demand, IntechOpen, UK, 133–142 (2018).**
2. **Shi Z-J, Xiao L-P, Deng J, Xu F, Sun R-C. Isolation and characterization of soluble polysaccharides of *Dendrocalamus brandisii*. *Bioresources* 6(4), 5151–5166 (2011).**

• **Analyzed and determined the chemical composition of bamboo, providing valuable insights for avoiding potential issues during DNA extraction.**

3. Jing X, Cai C, Fan S, Luo H. Physiological response characteristics of moso bamboo under drought stress based on calcium signal. *Forests* 12(12), 1699 (2021).
4. Wang Y, Zhan H, Ding Y, Wang S, Lin S. Variability of anatomical and chemical properties with age and height in *Dendrocalamus brandisii*. *Bioresources* 11(1), 1202–1213 (2016).
5. Chu CH, Huang L, Wang SG. Comparison on the anatomical structures of the roots generated from different parts of *Dendrocalamus brandisii*. *Acta Bot. Boreal. Occident. Sin.* 40(1), 43–52 (2020).
6. Ramanayake S, Yakandawala K. Micropropagation of the giant bamboo (*Dendrocalamus giganteus* Munro) from nodal explants of field grown culms. *Plant Sci.* 129(2), 213–223 (1997).
7. Li J, Wang S, Yu J, Wang L, Zhou S. A modified CTAB protocol for plant DNA extraction. *Chin. Bull. Bot.* 48(1), 72 (2013).
8. Silva RM, Ribeiro NP. Establishment of a genomic DNA extraction protocol for bamboo species. *Rev. Cienc. Agrar.* 41(3), 825–831 (2018).
- **Yielded favorable outcomes in the extraction of DNA from various bamboo species, thereby serving as a valuable reference for this investigation.**
9. Peng F, Liu M, Wang X, Ding X. Synthesis of low-viscosity hydrophobic magnetic deep eutectic solvent: selective extraction of DNA. *Anal. Chim. Acta* 1181, 338899 (2021).
10. Wang X, Liu M, Peng F, Ding X. Hydrophobic magnetic deep eutectic solvent: synthesis, properties, and application in DNA separation. *J. Chromatogr. A* 1659, 462626 (2021).
11. Liu M, Ding X, Wang X, Li J, Yang H, Yin Y. Extraction of DNA from complex biological sample matrices using guanidinium ionic liquid modified magnetic nanocomposites. *RSC Adv.* 9(40), 23119–23128 (2019).
12. Amani J, Kazemi R, Abbasi AR, Salmanian AH. A simple and rapid leaf genomic DNA extraction method for polymerase chain reaction analysis. *Iran. J. Biotechnol.* 9(1), 69–71 (2011).
- **Presents a novel approach for DNA extraction that holds significant implications for future investigations in the field of DNA extraction.**
13. Choudhury D, Sahu JK, Sharma G. Bamboo shoot: microbiology, biochemistry and technology of fermentation – a review. *Indian J. Tradit. Knowl.* 11(2), 242–249 (2012).
14. Xiu WY, Zhu Y, Chen B, Hu Y, Dawuda MM. Effects of paclobutrazol on the physiological characteristics of *Malus halliana* Koehne seedlings under drought stress via principal component analysis and membership function analysis. *Arid Land Res. Manag.* 33(1), 97–113 (2019).
- **Presents a method for conducting a comprehensive analysis that holds significant reference value for the comprehensive analysis of various extraction methods.**
15. Yan M, Huang W, Hu J, Lv Z, Lei S, Huang C. Evaluation of rice drought resistance by subordinate function. *Hybrid Rice* 24(5), 76–79 (2009).
16. Peng Y, Shi G, Cui H. Evaluation of drought resistant of different processing tomato at seed germination stage under PEG-6000 stress. *Seed* 32(7), 44–49 (2013).
17. Zhang X, Wang L, Shou L. Modified CTAB method for extracting genomic DNA from wheat leaf. *Agric. Sci. Technol.* 14(7), 946 (2013).
- **Employed the cetyltrimethylammonium bromide (CTAB) method to achieve favorable outcomes in wheat, which holds significant reference value for DNA extraction of bamboo plants.**
18. Mafra I, Silva SA, Moreira EJ, da Silva CSF, Beatriz M, Oliveira P. Comparative study of DNA extraction methods for soybean derived food products. *Food Control* 19(12), 1183–1190 (2008).
19. Terry CF, Harris N, Parkes HC. Detection of genetically modified crops and their derivatives: critical steps in sample preparation and extraction. *J. AOAC Int.* 85(3), 768–774 (2002).
20. Esfandani-Bozchaloyi S, Sheidai M, Kalalegh MH. Comparison of DNA extraction methods from *Geranium* (Geraniaceae). *Acta Bot. Hung.* 61(3–4), 251–266 (2019).
21. Aboul-Maaty Na-F, Oraby Ha-S. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bull. Natl Res. Cent.* 43(1), 1–10 (2019).
- **Demonstrates applicability of the CTAB method for DNA extraction in various plant species, providing a valuable reference for this investigation.**
22. Cheng Y-J, Guo W-W, Yi H-L, Pang X-M, Deng X. An efficient protocol for genomic DNA extraction from *Citrus* species. *Plant Mol. Biol. Rep.* 21, 177–178 (2003).
23. Chen D, Fu XY, Yin XM, Zhang RY. Assessment of several extraction methods of total RNA from fruits. *Guangdong Agric. Sci.* (11), 30–33 (2007).
24. Li JD, Liang QW, Wu WJ, LU WJ, Chen HQ. Compare on the different genome DNA extracting methods for mature *Populus tomentosa* leaves. *Henan Sci.* 27(03), 285–288 (2009).
25. Liu XP, Guo CY, Zhou R *et al.* Comparison of different extraction methods for *Turpinia arguta* genomic DNA. *South China For. Sci.* 50(3), 5–9 (2022).
26. Zhao XH, Zhang LH, Wang MY. Extraction of genomic DNA from *Rhododendron* and its RAPD examation. *Biotechnology* 15(05), 43–45 (2005).
27. Zhou XL, Tian X, Shen ND, Zhu HQ. Study on extraction method of total DNA from *Carum carvi*. *Mol. Plant Breed.* 17(07), 2217–2222 (2019).
28. Zheng SS, Li Y, Huang QJ, Deng M. Optimization of high purity genomic DNA extraction protocol for Fagaceae plant. *Mol. Plant Breed.* 18(20), 6725–6733 (2020).
29. Chen YX, Zhou RB, Pan QP, Tong QZ, Chen Y, Liu XD. Extraction and analysis of DNA in *Polygonatum odoratum*. *Northwest Pharm. J.* 22(04), 171–173 (2007).

BioTechniques LEARN



Editorial Department

Senior Editor
Tristan Free
tfree@biotechniques.com

Business Development and Support

Commercial Director
Evelina Rubio Hakansson
e.rubiohakansson@future-science-group.com

@MyBioTechniques @biotechniques BioTechniques

www.biotechniques.com