Received: 5 August 2019 | Revised: 17 March 2020 | Accepted: 26 March 2020

DOI: 10.1111/1755-0998.13162

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

An optimized method for the extraction of ancient eukaryote DNA from marine sediments

Linda Armbrecht[1] | Salvador Herrando-Pérez[1] | Raphael Eisenhofer[1] | Gustaaf M. Hallegraeff[2] | Christopher J. S. Bolch[3] | Alan Cooper[4]

[1]School of Biological Sciences, Faculty of Sciences, Australian Centre for Ancient DNA, The University of Adelaide, Adelaide, SA, Australia

[2]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tas., Australia

[3]Institute for Marine and Antarctic Studies, University of Tasmania, Launceston, Tas., Australia

[4]South Australian Museum, Adelaide, SA, Australia

Correspondence
Linda Armbrecht, School of Biological Sciences, Faculty of Sciences, Australian Centre for Ancient DNA, The University of Adelaide, Adelaide, SA, Australia.
Email: linda.armbrecht@adelaide.edu.au

Funding information
Australian Research Council, Grant/Award Number: CE170100015, DP170102261, DP170104665 and FL140100260

Abstract

Marine sedimentary ancient DNA (*sed*aDNA) provides a powerful means to reconstruct marine palaeo-communities across the food web. However, currently there are few optimized *sed*aDNA extraction protocols available to maximize the yield of small DNA fragments typical of ancient DNA (aDNA) across a broad diversity of eukaryotes. We compared seven combinations of *sed*aDNA extraction treatments and sequencing library preparations using marine sediments collected at a water depth of 104 m off Maria Island, Tasmania, in 2018. These seven methods contrasted frozen versus refrigerated sediment, bead-beating induced cell lysis versus ethylenediaminetetraacetic acid (EDTA) incubation, DNA binding in silica spin columns versus in silica-solution, diluted versus undiluted DNA in shotgun library preparations to test potential inhibition issues during amplification steps, and size-selection of low molecular-weight (LMW) DNA to increase the extraction efficiency of *sed*aDNA. Maximum efficiency was obtained from frozen sediments subjected to a combination of EDTA incubation and bead-beating, DNA binding in silica-solution, and undiluted DNA in shotgun libraries, across 45 marine eukaryotic taxa. We present an optimized extraction protocol integrating these steps, with an optional post-library LMW size-selection step to retain DNA fragments of ≤500 base pairs. We also describe a stringent bioinformatic filtering approach for metagenomic data and provide a comprehensive list of contaminants as a reference for future *sed*aDNA studies. The new extraction and data-processing protocol should improve quantitative paleo-monitoring of eukaryotes from marine sediments, as well as other studies relying on the detection of highly fragmented and degraded eukaryote DNA in sediments.

KEYWORDS

ancient DNA, diatoms, dinoflagellates, haptophytes, Maria Island, metagenomics, plankton, seafloor, Tasmania

1 | INTRODUCTION

Ancient DNA (aDNA) encompasses nonmodern genetic material (i.e., from dead organisms), which is typically highly fragmented and damaged (Shapiro & Hofreiter, 2012). Analyses of aDNA in marine sediments have become increasingly common (e.g., Coolen et al., 2006; Hou et al., 2014; More et al., 2018; Shaw, Weyrich, Hallegraeff, & Cooper, 2019). The increased interest in marine sedimentary ancient DNA (*sed*aDNA) has been stimulated by improved capacity for contamination control during shipboard

sediment core sampling, the possibility of processing samples at ultraclean aDNA facilities, and advances in high-throughput sequencing technologies and bioinformatic tools (Armbrecht et al., 2019). *Sed*aDNA applications are appealing because they require (a) less than a gram of sediment to reconstruct marine palaeo-communities and food webs facilitating efficient sampling, transport and storage; (b) short processing times for multiple samples; and (c) have broad taxonomic coverage including nonfossilizing organisms such as soft-bodied species within Foraminifera and Radiolaria (Lejzerowicz et al., 2013; Morard et al., 2017). Overall, *sed*aDNA approaches hold great potential for improving knowledge of past marine communities and their responses to climate over much longer time scales than possible through ongoing long-term biological monitoring programmes.

To date, a variety of methods have been used to isolate *sed*aDNA from marine sediments (Armbrecht et al., 2019). While a standardized method for direct interstudy comparisons would be ideal, protocols often need to be adjusted depending on sediment types or target organisms. Most extraction protocols utilize commercial kits with slight modifications tailored to specific study objectives and taxa (Armbrecht et al., 2019). The DNeasy Soil Kits (Qiagen) have been shown to yield abundant and high-quality *sed*aDNA from both prokaryotes and eukaryotes in marine sediments (e.g., Coolen et al., 2013; De Schepper et al., 2019; Epp et al., 2019; More et al., 2018; Orsi et al., 2017; Shaw et al., 2019). Such commercial kits enable rapid sample processing of the multiple marine *sed*aDNA samples needed to assemble high-resolution palaeo-community data. However, the kits commonly rely on DNA-binding steps using silica spin columns, which have been reported to result in DNA losses, probably due to competitive column-binding of organic matter (Lloyd, MacGregor, & Teske, 2010), and also selectively retain high molecular-weight (HMW) DNA fragments (Rohland, Glocke, Aximu-Petri, & Meyer, 2018). Such biases can dramatically impact the resolution of marine *sed*aDNA studies, where the amount of aDNA per sample is invariably low. To overcome this issue, aDNA studies commonly use silica-solutions rather than spin columns in the DNA-binding step (Brotherton et al., 2013; Frisia et al., 2017; Weyrich et al., 2017). Most protocols typically involve mechanical force (bead-beating) to isolate DNA as the first step in the extraction process (Armbrecht et al., 2019), whereas gentler chemical protocols (such as sample incubation in ethylenediaminetetraacetic acid, EDTA) have also been successfully applied to isolate eukaryote *sed*aDNA (Slon et al., 2017). However, so far neither the silica-solution nor the EDTA-based approaches have been assessed for marine sediments, nor have they been tested for taxonomic eukaryote coverage from *sed*aDNA preserved in the seafloor.

Due to their sensitivity to changes in environmental conditions and high turnover rates (when alive), planktonic organisms are commonly used as environmental indicators in marine *sed*aDNA research (e.g., Coolen et al., 2013; More et al., 2018, De Schepper et al., 2019). Nonetheless, phyto- and zooplankton, and their predators, exhibit highly diverse morphologies and cell-wall compositions, resulting in low taxonomic coverage from standard DNA extraction protocols. For example, three of the major phytoplanktonic groups (diatoms, dinoflagellates, coccolithophores) differ considerably in their outer protective layer; diatoms are armoured by robust silica shells (frustules), dinoflagellates are encapsulated in either cellulose plates (thecate dinoflagellates), enveloped in membranaceous pellicles (athecate dinoflagellates), or enclosed in tough sporopollenin-like resting stages (cysts), whereas coccolithophore cell walls are covered by calcium carbonate plates (coccoliths). Common phytoplankton predators (crustaceans, cnidarians, molluscs) also feature an assortment of exoskeletons and surface materials including chitin, gelatine and calcium carbonate (Hamm et al., 2003; Krampitz, Drolshagen, Häusle, & Hof-Irmscher, 1983; Netzel & Duerr, 1984; Neugebauer, Bykowski, Neugebauer, & Zjednoczenia, 1986; Sarras et al., 1991). Thus, efficient cell-lysis, the initial step of *sed*aDNA extractions, is challenging for such variable materials, and the effectiveness of protocols remains untested.

Advances in next-generation sequencing techniques (massively parallel DNA amplification and sequencing) and metagenomics (analysis of total DNA from samples) allow assessments of DNA proportional to its presence in any given sample, with highly quantitative potential (Nayfach & Pollard, 2016; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). Due to the low abundance of eukaryote *sed*aDNA in marine sediments, many previous studies have applied metabarcoding approaches popular in modern marine genomics, i.e., where specific taxonomically informative gene regions are amplified using the polymerase chain reaction (PCR; Taberlet et al., 2012). Metabarcoding and PCR have several caveats in aDNA research, e.g., even trace amounts of modern DNA (from reagents, laboratories or living cells) will preferentially amplify over short degraded aDNA, thus contaminating the ancient signal (Salter et al., 2014; Ziesemer et al., 2015); length variations in metabarcoding regions can skew taxonomic profiles in ancient DNA (Ziesemer et al., 2015); coextracted impurities (such as humic substances, pigments, heavy metals) frequently inhibit PCR (Webster, Newberry, Fry, & Weightman, 2003); PCRs with low amounts of template DNA are prone to bias due to stochastic amplification effects occurring in the first few PCR cycles (Wagner et al., 1994; Webster et al., 2003). Additionally, key barcoding targets such as the ribosomal RNA gene can vary greatly in copy number between taxa, causing a taxonomically biased representation of community structure (e.g., Klappenbach, Saxman, Cole, & Schmidt, 2001; Wang et al., 2017). Alongside a rarity of metagenomic studies to investigate *sed*aDNA, no standard bioinformatic pipeline currently exists that permits adequate quality control and DNA fragment-size and/or damage analyses to authenticate relative abundance data.

In this study, we applied seven different *sed*aDNA extraction and library-preparation methods to marine sediment core samples collected off Maria Island, Tasmania, Australia. Our overarching goal was to design a laboratory protocol and quality-control bioinformatic pipeline that would (a) maximize total *sed*aDNA yield; (b) retain short (≤500 bp) *sed*aDNA fragments, crucial for *ancient*-sediment research; and (c) capture the broadest eukaryote diversity possible. We used a shotgun sequencing approach to enable a quantitative statistical analysis of eukaryote DNA abundance and diversity,

providing a benchmark for the use of metagenomics in marine *sed*aDNA research.

2 | MATERIALS AND METHODS

2.1 | Sampling and sample details

We collected an eleven cm long seabed sediment core using a KC Denmark multicorer at a water depth of 104 m off Maria Island, Tasmania (148.240E; 42.845S) during the *RV Investigator* voyage IN2018_T02 (20 May 2018). The core was immediately stored at 10°C in the dark for 6 weeks, then sliced into 1–1.5 cm sections in a separate non-DNA laboratory at the Institute for Marine and Antarctic Studies (IMAS), Hobart, Tasmania, Australia. For this study we used three of these core sections (top = 0–1.5 cm, middle = 5–6 cm, bottom = 8.5–10 cm, Figure 1). To minimize contamination, we wiped working benches and washed cutting knives with ethanol, wore new gloves for the cutting of each core section, and removed the outer peripheral 2 cm layer of each of the latter. The core sections (refrigerated)

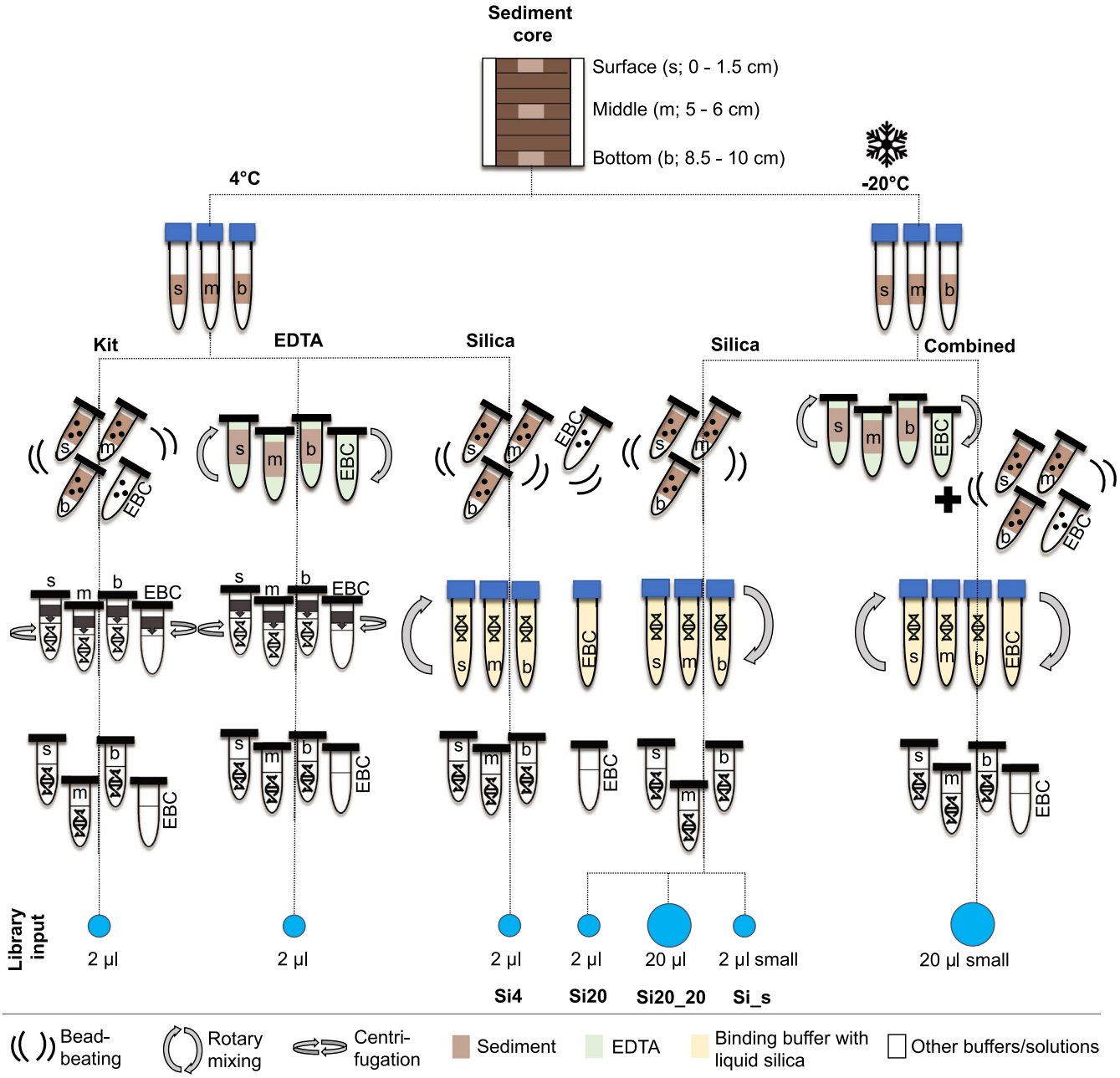

FIGURE 1 Schematic overview of the seven extraction and library preparation techniques for marine *sed*aDNA. Subsamples were taken from the centre of three core sections (s, surface; m, middle; b, bottom) split and stored at 4°C and −20°C. Refrigerated material was utilised for the Kit, EDTA and Si4 treatments, frozen material for the Si20, Si20_20, Si_s and Com treatments. Either 2 or 20 µl template DNA was used as library input. Size-selection was performed on Si_s and Com. Extraction blank controls (EBCs) are indicated as empty tubes, one EBC was shared between Si4 and Si20 [Colour figure can be viewed at wileyonlinelibrary.com]

were hand-carried to the Australian Centre for Ancient DNA (ACAD), Adelaide, in July 2018. The age of the oldest section (8.5–10 cm) was estimated at ~150 years (Pb^{210} data, not shown), which enabled our protocols to be tested on marine-sourced sediments expected to contain relatively large amounts of *sed*aDNA.

2.2 | DNA extractions

2.2.1 | Sediment processing and pretreatment

Core section processing, *sed*aDNA extractions and sequencing library preparations took place at ACAD's ultraclean forensic facilities following aDNA decontamination standards (Willerslev & Cooper, 2005). We placed the three sediment core sections into zip-lock bags sterilised with UV light and manually homogenized them for ~5 min. From each section, two 1 cm^3 subsamples were transferred into two separate 15 ml centrifuge tubes using a sterile disposable spatula. One subsample was kept at 4°C and the other at −20°C for one month. The samples were prepared for the different extraction methods in a glove box decontaminated (3% bleach) between consecutive subsamples. The different *sed*aDNA extraction and sequencing library preparation techniques are detailed below and schematically shown in Figure 1. The final steps of the sequencing library preparations (post IS7/IS8 amplification; see Section 2.3.1) were undertaken at ACAD molecular laboratories physically distant from the forensics facilities above.

2.2.2 | Bead-beating + spin column (DNeasy PowerLyzer PowerSoil Kit, Qiagen; "Kit")

We applied this technique to 0.25 g of the three sediment subsamples stored at 4°C, guided by the manufacturer's protocol with the following specifications. After transferring the sediment into individual bead-tubes using a disposable, sterile spatula, bead-beating was applied in three runs of 20 s with 5 s breaks using a Precellys 24 homogenizer (Bertin Instruments, France), followed by centrifugation (3 min, 10,319 rpm). We retained the optional 5 min incubation steps at 4°C as per the kit's protocol. For consistency in subsequent extractions, DNA was eluted in 80 µl of Buffer EB (Qiagen), instead of the customary C6 solution, and ultimately frozen at −20°C. To monitor laboratory contamination, we used extraction blank controls (EBCs) by treating one empty bead-tube with the same protocol. Consequently, the Kit method provided a total of four extracts (Figure 1).

2.2.3 | Bead-beating + liquid silica in QG Buffer ("Si4" and "Si20")

We applied this lysis process to 0.25 g of the subsamples stored at both 4°C and −20°C (i.e., six extracts). We followed the same protocol as described in Section 2.2.2 down to step 10 of the manufacturer's instructions (addition of solution C3 and centrifugation). After this step, we transferred the supernatant into 15 ml centrifuge tubes containing a DNA-binding buffer consisting of 100 µl silica-solution (Sigma Aldrich), 3 ml modified Buffer QG (2.7 ml Buffer QG Qiagen, 46 µl H_2O, 39.08 µl Triton X-100, and 24.66 mM NaCl, 164.5 mM NaOAc (Brotherton et al., 2013). After stirring on a rotary mixer (1 hr, room temperature), and centrifugation (5 min, 4,500 rpm), the supernatant was removed and the pellet resuspended in 900 µl DNA-binding buffer. We recentrifuged (1 min, 14,000 rpm), removed the supernatant, and washed the pellet twice in 80% EtOH before drying (15 min, 37°C) and resuspended it in 80 µl Buffer EB (Qiagen). Following incubation (10 min, 50°C), we centrifuged (1 min, 14,000 rpm) and stored the supernatant (free of silica) in a sterile Lo-bind tube (Eppendorf) at −20°C. We included one EBC, resulting in a total of seven extracts (Figure 1).

2.2.4 | EDTA + MinElute ("EDTA")

We applied this technique to 0.25 g of the three sediment subsamples stored at 4°C following Slon et al. (2017) with minor modifications. Briefly, we added 1 ml of ethylenediaminetetraacetic acid (EDTA) to the sediment in a 2 ml screw-cap tube and placed it on a rotary mixer (~25 rpm, room temperature) overnight. After centrifuging (3 min, 13,000 rpm), we purified the DNA using the MinElute kit (Qiagen) as per the manufacturer's instructions. After DNA binding using the kit's spin column, we eluted the DNA in 80 µl of Buffer EB. We also included one EBC, making four extracts in total (Figure 1).

2.2.5 | EDTA + bead-beating + liquid silica in QG Buffer ("Combined", or "Com")

In addition to the established DNA extraction techniques described in Sections 2.2.2–2.2.4, we applied a new technique which combined all those methods. Thus, we first incubated 0.25 g of the three frozen sediment subsamples in EDTA overnight as in Section 2.2.4, except that only 0.75 ml EDTA was used to keep volumes consistent with a subsequent step (see below). After centrifugation (3 min, 13,000 rpm), the supernatant was kept at 4°C, and the pellet was processed separately using bead-beating and DNA purification as in Section 2.2.3 (see Figure 1). The resulting 0.75 ml DNA-solution purified from the pellet (step 10 of the DNeasy kit protocol) was then recombined with the 0.75 ml EDTA supernatant (making 1.5 ml in total) and added to 6 ml modified QG buffer with 100 µl liquid silica, before proceeding as described in Section 2.2.3. We ultimately eluted the DNA in 100 µl Buffer EB (Qiagen). This method generated four extracts (three sediment extracts and one EBC).

2.3 | Shotgun sequencing library preparation

2.3.1 | Libraries from DNA extracts using the Kit, Si (Si4 and Si20), and EDTA techniques

Libraries containing 2 µl input DNA ("Si4", "Si20")

We prepared libraries from all DNA extracts (i.e., a total of 15) after Weyrich et al. (2017) with the following modifications. A 2 µl aliquot of DNA was repaired (15 min, 25°C) in a 40 µl reaction using T4 DNA polymerase (New England Biolabs). After purifying the DNA (MinElute Reaction Cleanup Kit, Qiagen), a ligation step followed (T4 DNA ligase, Fermentas) in which truncated Illumina-adapter sequences containing two unique five base-pair (bp) barcodes were attached to the double-stranded DNA (60 min, 22°C; Meyer & Kircher, 2010). An additional DNA purification (MinElute Reaction Cleanup Kit, Qiagen) step was performed, followed by a fill-in reaction with adapter sequences (Bst DNA polymerase, New England Biolabs; 30 min, 37°C, with polymerase deactivation for 10 min, 80°C). We then used 3 µl of the reaction-product as input for a 25 µl PCR (eight replicates per extract) with the primers IS7 and IS8 (Meyer & Kircher, 2010). Each PCR reaction included 14.2 nuclease-free H_2O, 2.5 µl 10× Gold Buffer, 2.5 µl 25 mM $MgCl_2$, 0.25 µl 25 mM dNTPs, 1.25 IS7, 1.25 IS8, and 0.1 µl AmpliTaq Gold Polymerase (Applied Biosystems). Thermal cycling specifications were as follows: 6 min at 94°C, 13 cycles of 30 s denaturation at 94°C, 30 s annealing at 60°C, 40 s extension at 72°C, and 10 min of final extension. We purified PCR products using AxyPrep magnetic beads (Axygen Biosciences; 1:1.8 library:beads) and eluted the DNA in Buffer EB (Qiagen) with 0.05% Tween 20 (Sigma Aldrich).

Libraries containing 20 µl input DNA ("Si20_20")

In order to test whether relatively low or high amounts of input DNA affected library quality and sequencing results (particularly, whether PCR inhibition might become problematic, or whether larger DNA input volumes would benefit sequencing), we prepared four additional libraries based on the DNA extracts derived from the Si20 extraction (three sediment extracts and one EBC; see Section 2.2.3). For these libraries we used 10× DNA input, totalling 20 µl per sample, then followed Section 2.3.1.1 (Figure 1).

Libraries containing 20 µl input DNA and small DNA fragments only ("Si_s")

To target relatively small DNA fragments typical of aDNA, we performed a 'reverse AxyPrep' DNA purification post-library preparation. For this, we prepared four additional libraries as described in Section 2.3.1.2, except that we added AxyPrep beads in a ratio 1:0.4 library:beads in the first clean-up step (aiming to retain fragments ≤500 bp). After 15 min of incubation and 15 min on a magnetic rack to separate beads from the supernatant, we transferred the supernatant into a new sterile 1.7 ml tube. AxyPrep beads were added in a ratio 1:1.2 library:beads. Following two washes with 80% EtOH, we eluted the DNA in 30 µl Buffer EB with 0.05% Tween 20.

All libraries prepared in Sections 2.3.1.1–2.3.1.3 (23 in total) were quantified using the Qubit dsDNA HS Assay (Molecular Probes). Based on the Qubit results, we designed a sequencing pool containing 4 ng of DNA per library. As the EBCs contained only traces of DNA, we also prepared a second sequencing pool with 10 µl per EBC library. We then ran a PCR on these two sequencing pools (eight replicates with 25 µl reactions each) following Section 2.3.1, except that we used Indexing Primer IS4 and the GAII index 1 (Meyer & Kircher, 2010) and 13 cycles. Post-PCR quantifications of both pools were conducted using Qubit and merged into a third sequencing pool. To avoid over-diluting the latter sequencing pool, we used 4 ng DNA from pool 1 and 0.4 ng DNA from pool 2. The DNA quality and quantity of the third sequencing pool was assessed through TapeStation (Agilent Technologies). AxyPrep clean-up (at a ratio of 1:1.1 library:beads) and TapeStation checks were repeated twice to generate a high-quality sequencing pool (i.e., a minimal amount of primer-dimer, DNA concentration = 6.36 ng/µl). We submitted the final pool to the Australian Cancer Research Foundation Cancer Genomics Facility & Centre for Cancer Biology, Adelaide, Australia (hereafter ACRF) for Illumina NextSeq sequencing (2 × 75 bp cycle).

2.3.2 | Libraries prepared from DNA extracts using the Combined ("Com") technique

We prepared four libraries as described in Section 2.3.1.2 down to the step of first post-PCR AxyPrep clean-up and Qubit quantification. Due to the low number of libraries, and to avoid over-diluting the final sequencing pool, we combined the four libraries to achieve a concentration of 4 ng/µl per sediment-derived library and a concentration of 0.4 ng µl for the EBC. We then ran GAII indexing PCR (Index 8), two TapeStation checks and one AxyPrep clean-up as in Section 2.3.1.1. Aiming to retain only the smallest DNA fragments (≤500 bp) from the full content of DNA detected by TapeStation, we added two rounds of the AxyBeads purification as in Section 2.3.1.3, with an initial library:beads ratio of 1:0.4 in round 1 and 1:0.6 in round 2 (Urich, Nery, Lister, Schmitz, & Ecker, 2015). As HMW DNA appeared to persist, we prepared a 1:10 dilution of the pool and repeated the AxyBeads size-selection one more time with an initial library:beads ratio of 1:0.6. We submitted the final pool (nondetectable primer-dimer, DNA concentration between 147 and 347 bp = 1.48 ng/µl) to ACRF for Illumina NextSeq sequencing (2 × 75 bp cycle).

2.4 | Data analyses

2.4.1 | Sequencing data

We ran a FastQC quality control analysis on the raw sequencing reads followed by demultiplexing and adapter trimming using ADAPTERREMOVAL v. 2.1.7-foss-2016a software (--adapterlist –barcodelist; Schubert, Lindgreen, & Orlando, 2016). Specifications included removal of consecutive stretches of low-quality bp and N's (--trim-qualities --trimns) and allowing for a barcode mismatch of 1 bp (--barcode-mm), discarding reads < 25 bp (--minlength) and merging

them (--collapse) into.gz output files. We removed low-complexity sequences using the software Komplexity (--threshold 0.55; Clarke et al., 2019) and duplicated sequences using the dedupe tool in BBMAP version 37.36. FASTQC (version 0.11.5-Java-1.8.0_101; Babraham Bioinformatics) and MULTIQC (version 1.0.dev0; Ewels, Magnusson, Lundin, & Käller, 2016) were run on all merged reads before and after filtering low-complexity sequences, and after deduplication. We used SILVA 132 SSURef NR99 (Quast et al., 2012) as the reference database to build a MALT index and aligned our sequences using MALT (semiglobal alignment; Herbig et al., 2016). The SILVA SSU database was chosen due to the extensive use of the 18S gene in the marine eukaryote taxonomy literature (e.g., De Vargas et al., 2015; Brown et al., 2018). The resulting .blastn files were converted to .rma6 format using the Blast2RMA tool in MEGAN6 (version 6_15_1; Huson et al., 2016) with the default settings except for a minimum support percent of zero ("off"; as we expected a low number of reads to be assigned to eukaryotes), a minimum bit score of 50, an *E*-value of 0.01, and a minimum percent identity of 95%. We avoided subsampling (rarefying) our data to retain maximum numbers of reads throughout the complete data-processing procedure and instead used relative abundances as a means of normalization before downstream statistical analyses. Our complete bioinformatic pipeline is further detailed in the Appendices S1–S6.

2.4.2 | MEGAN6 analysis and subtractive filtering

To visualize taxonomic classifications, we imported rma6-files into MEGAN6 Community Edition (Huson et al., 2016) using the "compare" function including absolute read counts and ignoring unassigned reads. First, we only considered EBCs ranked by species in order to determine contaminant taxa (detailed in Table S1). After selecting all species present in EBCs, we subtracted these from our sediment-derived samples by applying the invert function in MEGAN6 (hereafter, the term "samples" refers to sediment-derived data post EBC subtraction). Next, the assigned reads were ranked by phylum, whereas we collapsed nodes of nontarget taxa comprising bacteria, archaea and fungi. We exported the assigned reads (counts) and converted them to relative abundances for downstream analyses. We combined MEGAN6-classified "Environmental sample Stramenopiles" and "unclassified Stramenopiles" into "Other Stramenopiles". Unassigned sequences were not considered in the statistical analysis.

2.4.3 | DNA fragment-size analysis

We used DNA fragment length as an indication of which treatment maximized the yield of aDNA, with shorter fragments more likely to represent authentic *sed*aDNA. We downloaded the sequence size-frequency report from our MultiQC output (collapsed, filtered and de-duplicated reads; Ewels et al., 2016), generating the number of sequences occurring in intervals of 5 bp from 27 to 122 bp and 3 bp from 122 to 125 bp per treatment. We assessed variation in log-transformed abundances of 21 read-length classes across samples and treatments using covariance-based principal component analysis (PCA; Jolliffe, 2002).

2.4.4 | Comparison of read numbers among taxa and treatments

We considered seven different "extraction treatments", namely Si4, Si20, Si20_20, Si_s, Com, EDTA and Kit (see Sections 2.3.1. and Figure 1). We used a stacked approach of global and pairwise comparisons to test (i) which of the four silica treatments (Si4, Si20, Si20_20, Si_s) performed best, i.e., maximising reads across a wide variety of eukaryotes (see Section 2.4.5), enabling us to select a 'best' silica treatment for (ii); and (ii) which of the different extraction treatments ("best" silica treatment, EDTA, Kit and Com) performed best. In both (i) and (ii) the "best" performing extraction treatments were determined using the weighting procedure described below.

For each taxon separately, we contrasted two linear models whether the number of reads (*count*) (i) varied (treatment model: *count ~treat*) or (ii) not (null model: *count ~ 1*) according to treatment (*treat*), with *treat* being a categorical variable coding for each of the four treatments. We ranked model support through the Akaike's Information Criterion adjusted to finite sample size, AIC_c (Sugiura, 1978), resulting in one probability per model ($wAIC_c$ scaled to a 0 to 1 interval across the two models). We then estimated evidence ratios as the $wAIC_c$ of the treatment model to the $wAIC_c$ of the intercept model (Burnham & Anderson, 2002). We assumed Gaussian errors (corroborated in the saturated model [*count ~ treat*] using Q-Q and fit-vs.-residual plots) and discarded heteroscedasticity effects based on Bartlett's tests (Bartlett, 1937). We used the 'lm' function in the *R* package 'stats' to run the linear models whereas we plotted relative abundances per taxon and treatment in 'graphics' (R Core Team, 2019).

2.4.5 | Treatment efficiency

A treatment was regarded as "efficient" when it maximized the number of reads and appeared unbiased towards the detection of any given taxa (i.e., contributing little to statistical differences in the mean number of reads of a given taxon among treatments). To quantify treatment efficiency, we contrasted our treatment and intercept models (see Section 2.4.4) for each of the six pairwise comparisons between four treatments across taxa (i.e., Si4/Si20/Si20_20/Si_s or Com/EDTA/Kit/best silica treatment with 39 and 41 taxa contributing >1 read, respectively), and calculated a "treatment weight", where a high weight reflects high treatment efficiency. Each treatment was weighted as follows:

$$\text{treatment weight} = \sum_{j=1}^{j=t} \sum_{i=1}^{i=3} \text{null\_wAIC}_c \times \left(\frac{N_1}{SE_1} + \frac{N_2}{SE_2} \right)$$

where in the formula above, null_wAIC$_c$ is the probability of the intercept model, and *N* and *SE* are the mean and the standard error of the number of reads of a given taxon among the three samples per treatment, *j* represents each of *t* taxa, and *i* indexes the triplet of pairwise comparisons involving a treatment. For instance, Si_s can be compared with three other silica treatments (Si4, Si20, Si20_20) in six possible pairwise comparisons, so each pairwise comparison contributes the mean and standard error of each taxon to the Si_s weight. We standardized *N* by *SE* to balance the contributions by more or less abundant taxa. For all taxa, we summed up the weights contributed by each taxon to each set of six pairwise comparisons to obtain the overall "treatment weight". We calculated the confidence intervals of any given treatment weight over 10,000 bootstrapped samples. We standardized the raw and bootstrapped treatment weights on a 0–1 scale by dividing (i) each raw treatment weight by the highest raw weight scored by a treatment, and (ii) each bootstrapped treatment weight by the highest $\text{null\_wAIC}_c \times ((N_1/SE_1)+(N_2/SE_2))$ across bootstrapped values. While our analyses focused on eukaryotes, treatment weights and relative abundance barplots are also included for Bacteria and Archaea for completeness.

3 | RESULTS

3.1 | General statistics

After removing adapters (de-multiplexing, adapter-trimming and collapsing), we retained between 0 (EBCs processed using the Kit and Com methods) and 11.4 M (million) collapsed reads (Com, mid-depth sediment). After filtering (low-complexity removal and deduplication), we retained 143.2 M reads in total, with between 0 (most EBCs) and 11.2 M sequences (Com, mid-depth sediment) per sediment subsample. The lowest number of sequences retained for a sediment subsample was 1.6 M (Si4, surface sediment). MultiQC revealed mean quality scores between 31 and 41 (Phred scores). The sequence-length distribution ranged from 27 to 125 bp with an average of 66 bp (69 bp and 52 bp across all sediment subsamples and EBCs, respectively). GC content averaged 53% for samples and 52% for EBC's, and final duplication levels averaged 0.18% across samples and 2.26% across EBCs.

3.2 | Qualitative analysis of the taxonomically assigned reads (MEGAN6)

In total, across all samples, 41,845 reads (0.28% of total reads acquired) were assigned a taxonomic identification using SILVA SSU as the reference database (plus 17 unassigned reads, excluded from statistical analysis). Of this total, 39,838 reads were assigned to Bacteria (95.16%), 1,395 to Archaea (3.35%), and 569 to eukaryotes (1.37%). The latter spanned 44 eukaryote phyla, with the highest number of reads assigned to Streptophyta (112 reads in total across all samples), followed by Apicomplexa (71 reads) and Dinophyceae (48 reads). However, minimum bit scores in MEGAN (a measure of how well a query sequence aligns to a reference sequence) for Streptophyta were frequently close to the scores for Cyanobacteria and Sulfidobacter, thus the category Streptophyta should be evaluated with caution. Polycystinea, Kinetoplastida, Bacillariophyta, Chordata, and Nematoda were assigned less than 30 reads each. Labyrinthulomycetes, Haplosporidia, Arthropoda, Haptophyceae, Foraminifera and Chrysophyceae were assigned less than 20 reads each, whereas all other taxa were assigned less than 10 reads each. Within the Dinophyceae, most reads for Gonyaulacales and Peridiniales were obtained from extraction methods using DNA-binding in silica-solution, whereas Suessiales and Syndiniales were also identified using the Kit method. Within the Bacillariophyta, different treatments recovered different taxonomic groups (Bacillariophyceae, Coscinodiscophyceae, Mediophyceae), except for method Si4 (no Bacillariophyta reads). Within the Haptophyceae, few reads were recovered for Isochrydiales in individual samples using either extraction method, whereas Prymnesiales were only detected by EDTA. For further details on read yield for individual taxa by the different treatments see Figures S1 and S2.

3.3 | Quantitative analysis of taxonomically assigned reads per treatment

3.3.1 | Principal component analysis (PCA) for 21 DNA read-length classes

The PCA broadly separated the samples from the silica-based treatments to the right (PC1 = 67.6% variation explained) and top (PC2 = 24.3%) of the ordination (total variability in DNA size-distribution explained = 92%; Figure 2). EDTA, Kit and Si4 showed the highest variability (dispersion) in the range of fragment-lengths captured. Principal coefficients indicated that the abundance of read-length classes 27–102 bp increased towards the right of PC1 (Figure 2a,b) and classes 57–125 bp increased towards the bottom of the PC2 including most EDTA and Kit samples (Figure 2a,c). As such, Com, Si_s, Si20, Si20_20 and, to a lesser degree, Si4 achieved the most efficient extraction of the smallest DNA fragments (samples located towards the top right of Figure 2a), whereas EDTA and especially Kit selectively retained the largest DNA fragments (samples located towards the bottom left of Figure 2a). We did not find a clear separation of surface, middle and bottom subsamples within each treatment according to fragment length in the PCA ordination (Figure 2a).

3.3.2 | Global linear-model contrasts (across treatments)

For the four different silica treatments, the "treatment model" (see Section 2.4.4) received more support than the "intercept [*null*] model" in 33 of the 37 eukaryotic taxa and in Bacteria and Archaea (Table S2), indicating that different treatments resulted in different numbers of assigned reads for most taxa. Across all eukaryote

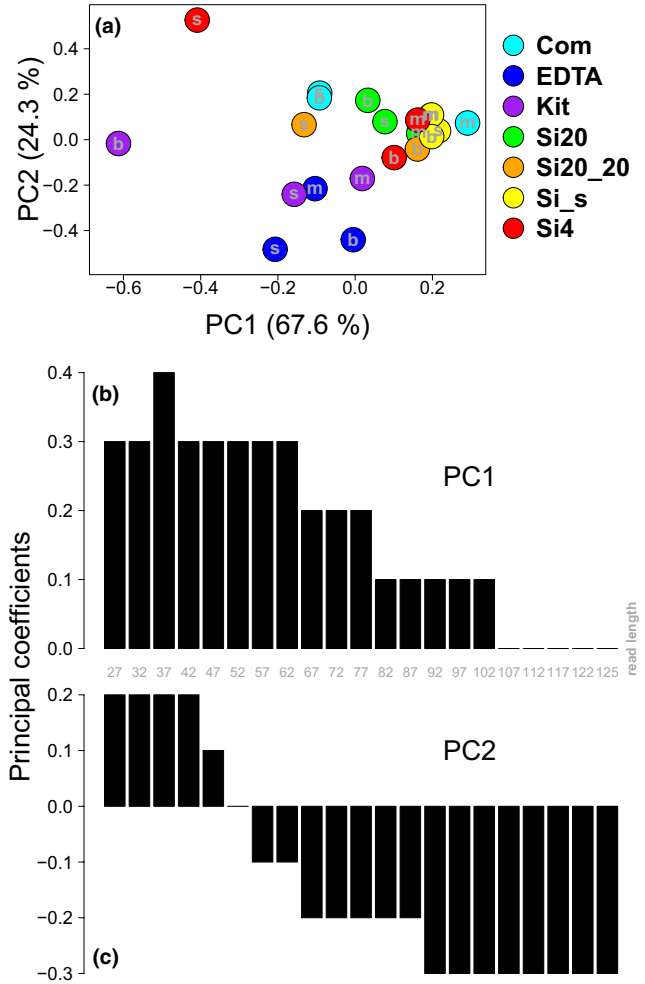

FIGURE 2 (a) Covariance-based principal component analysis (PCA) capturing variation in log-transformed abundances of 21 different read-length classes across triplets of samples within each of the seven extraction methods of marine *sed*aDNA off Tasmania. The histograms show the principal coefficients of each read-length class for (b) PC1 and (c) PC2. The higher the absolute value of a given coefficient, the higher the contribution of the corresponding read-length class to each principal-axes gradient, while a positive (or negative) sign indicates whether the abundance of a read-length class increases (or decreases) towards the right (PC1) or the top (PC2) of the ordination. Symbol insets represent whether read-length data originated from the surface (s), middle (m) or bottom (b) core section [Colour figure can be viewed at wileyonlinelibrary.com]

taxa, the treatment model was almost two times more likely than the intercept model (mean evidence ratio (ER) = 1.9 ± 0.9*SE*, Table S2, i.e., on average over all taxa, it was 1.9 times more likely that treatments differed on read yield than that they did not), and particularly so for Bacillariophyta (ER = 4.6). Eukaryote groups for which treatment and intercept models were equally likely (ERs ≈ 1) were Dinophyceae, Diplonemida, Arthropoda, Cnidaria, Sticholonchida, and Other Stramenopiles. Taxa for which no reads were yielded in the silica treatments were excluded from this analysis.

Within the four final treatments, the treatment model was more likely than the intercept model in 35 of 39 eukaryotic taxa. Across all eukaryote taxa, the treatment model was almost two times more likely than the intercept model (mean ER = 1.9 ± 1.3*SE*, Table S2), and particularly so for Amoebozoa (ER = 6.1). Eukaryote groups for which treatment and intercept models were equally likely (ERs ≈ 1.0) were Ciliophora, Cnidaria and Haplosporidia. For Bacteria and Archaea, the two models were also equally likely (Table S2). Again, taxa for which no reads were obtained were excluded from this analysis.

3.3.3 | Treatment efficiency

For eukaryotes in silica-based treatments, the Si20_20 treatment consistently yielded the highest number of reads across a wide variety of eukaryote taxa (i.e., contributed least to statistical differences in read yield among treatments, see Section 2.4.5), thus, this was the most efficient Si treatment (Figure 3a,c). Si20_20 extraction efficiency was closely followed by that of Si_s, then Si4 and Si20, in decreasing order (Figure 3a,c). For Bacteria, efficiency was ranked highest for Si20 among silica-treatments, closely followed by that of Si_s, then Si20_20 and Si4 (Figure 3e,g). For Archaea, efficiency decreased from Si_s to Si20, Si20_20 and Si4 (Figure 3i,k).

We selected Si20_20 as the "best" silica treatment for eukaryotic taxa (see Section 2.4.5) and compared it against the Com, EDTA and Kit extraction methods. EDTA scored the highest treatment efficiency (Figure 3b,d), closely followed by Com, then Si20_20, whereas Kit scored the lowest (Figure 3b,d). For Bacteria, Com performed best, followed by Kit, EDTA and (best silica-based treatment) Si20 (Figure 3f,h). For Archaea, Kit was most efficient and EDTA, Com and Si_s (best silica-based treatment) followed in decreasing order (Figure 3j,l).

Pairwise comparisons of the silica treatments across all eukaryote taxa were primarily used to determine the treatment efficiencies and, as such, are consistent with the above described performance of the silica and final treatments. For details on the resulting statistics from all pairwise comparisons see Table S3.

4 | DISCUSSION

4.1 | An optimized technique for the extraction of eukaryote DNA from marine sediments

Our analyses showed that all *sed*aDNA extraction treatments performed well for different aspects of the extraction process and that existing protocols need optimization to maximize the yield of ancient sequences from a broad range of eukaryotic taxa. To date, most marine *sed*aDNA studies targeting marine eukaryotes have used the Kit method (e.g., De Schepper et al., 2019; Lejzerowicz et al., 2013; More et al., 2018; Pawłowska et al., 2015), occasionally with slight modifications (Shaw et al., 2019), or a version of the same kit tailored to larger sediment volumes (up to 10 g DNA; e.g., Coolen et al., 2009; Epp et al., 2019; Klouch et al., 2016; Szczuciński et al., 2016) and alternative kit-based methods (e.g., Coolen et al., 2006; Hou et al., 2014; Orsi et al., 2017; More et al., 2018). The EDTA and silica-based techniques have been used in environmental aDNA research

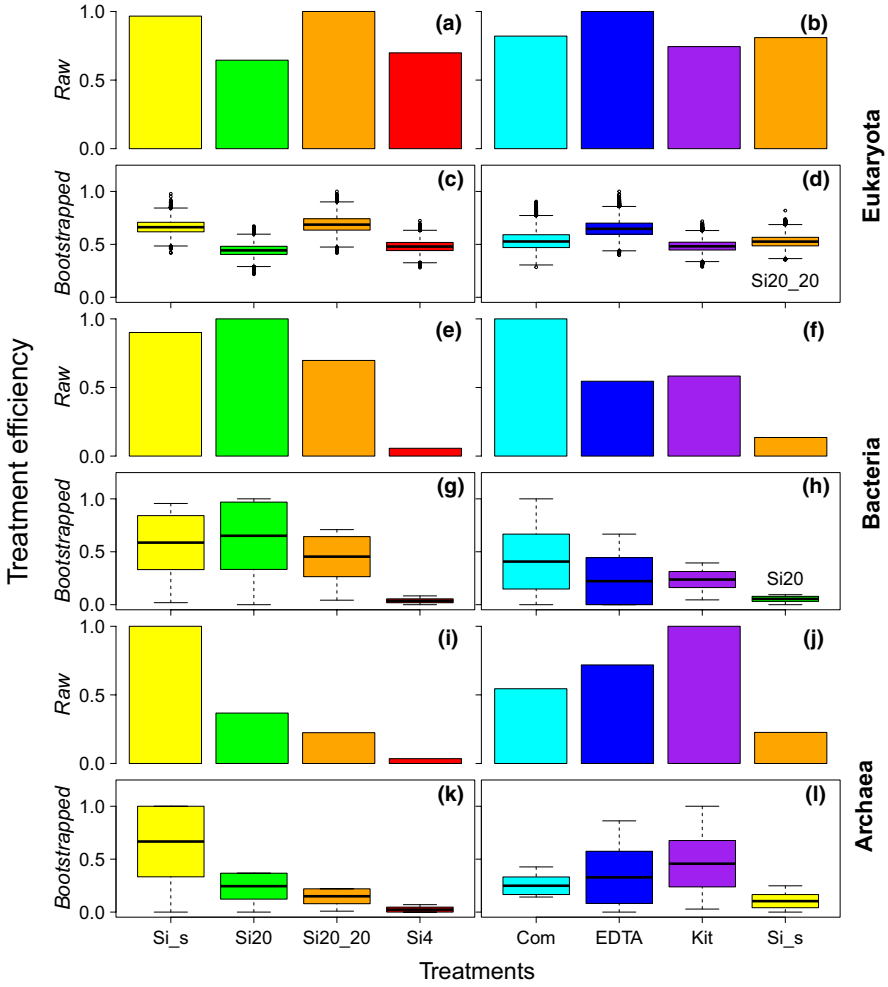

FIGURE 3 Treatment efficiency for the extraction of marine *sed*aDNA off Tasmania among silica-based (Si_s, Si20, Si20_20, Si4) and final (Com, EDTA, Kit along with best performing silica-treatment per taxonomical group) treatments across Eukaryota (a-d), Bacteria (e-h) and Archaea (i-l). The higher the bar (treatment weight on a normalized scale of 0–1), the more efficiently reads were recovered across study taxa (efficiency = broad taxonomic coverage and low bias towards particular taxa). See Methods for methodological description of raw versus bootstrapped values, while treatment efficiency lacks units (see formula in Section 2.4.5) [Colour figure can be viewed at wileyonlinelibrary.com]

previously (Slon et al., 2017; Frisia et al., 2017; Weyrich et al., 2017, respectively) and were applied here to marine sediments for the first time, alongside our novel combined approaches technique.

Across our *sed*aDNA extraction methods, EDTA scored the highest efficiency (high genetic yield across eukaryote taxa, low bias towards particular taxa). Both EDTA and Kit were the best methods to retain the largest DNA fragments, probably due to the use of DNA-binding spin columns, which have been shown previously to favour large DNA fragments (Rohland et al., 2018). In contrast, we found that techniques employing DNA binding to silica in solution outcompeted other treatments in the recovery of the smallest fragments, making it well-suited for highly fragmented and degraded sequences characteristic of *sed*aDNA. Within the silica-based treatments, Si20_20 was the most efficient, suggesting that frozen material and undiluted DNA templates are beneficial for the recovery of *sed*aDNA from marine eukaryotes. Indeed, cold temperatures have been shown to drive cellular lysis and breakage in soil samples (Roh, Villatte, Kim, & Schmid, 2006), and this mechanism may also be relevant for marine eukaryote *sed*aDNA. Si_s' efficiency was on a par with Si20_20, indicating it is possible to remove HMW DNA before sequencing without losing taxonomic resolution. Our bioinformatic pipeline considers collapsed reads only, therefore excludes fragments ≥126 bp, and also filters out modern contaminant sequences. Thus, a size-selection step in the library preparations is optional, yet would be recommended to prevent interference of HMW DNA with aDNA sequencing (i.e., increasing the sensitivity for aDNA molecules during the sequencing run).

Considering that the combined 'Com' extraction protocol was designed using components of the EDTA, Kit and Silica techniques presented here, the high efficiency was expected. The Com method combines frozen material to increase cell lysis (Roh et al., 2006), EDTA incubation to enhance eukaryote DNA isolation (Slon et al., 2017), bead-beating to retrieve DNA from paleo-deposited resting spores and cysts frequently occurring in marine sediments (Klouch et al., 2016; Shaw et al., 2019), and DNA-binding in silica-solution to capture variably-sized DNA fragments (see below). The use of undiluted library DNA as template (20 µl) promoted read yield, and the optional bead-based size-selection to remove HMW fragments that can interfere with sequencing and/or originate from contaminant organisms proved robust. Overall, Com represents a streamlined method highly amenable for *sed*aDNA studies given its efficiency to maximise the broadest eukaryote-diversity signal in marine sediments. A schematic overview of our new extraction method is provided in Figure 4.

FIGURE 4 Flowchart of *sed*aDNA extraction procedure and bioinformatic pipeline optimized for marine eukaryotes. For details on reagent composition and volumes, incubation times and centrifugation settings, as well as parameter settings in the bioinformatic workflow, see text [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Metagenomics in marine *sed*aDNA research

Eukaryotes comprised 1.37% (on average across all samples) of the identifiable community in sediment samples collected at 104 m water depth off Tasmania. This low proportion was expected for several reasons. First, by targeting total DNA we included Archaea and Bacteria, with the latter being highly abundant and active in marine sediments (e.g., Fry, Parkes, Cragg, Weightman, & Webster, 2008) even 2.5 km below the seafloor (Inagaki et al., 2015). Second, we mapped our reads against the SILVA-SSU rRNA database (an extensive, curated database of the SSU rDNA gene, the most widely used taxonomic marker for marine eukaryotes (see De Vargas et al., 2015; Guo, Sui, Zhang, Ren, & Liu, 2015; Wang et al., 2017; Carradec et al., 2018), which means other genetic material potentially holding taxonomic information was disregarded (e.g., plastid DNA). And finally, we applied a stringent bioinformatic filtering approach (Figure 4) to avoid artefacts from exogenous contamination and misassignment of low-complexity reads. As a general rule, the low proportions of eukaryote target DNA are the main reason previous marine *sed*aDNA studies have employed amplicon-based approaches, with PCR amplification of a specific genetic region (e.g., Coolen et al., 2006; Lejzerowicz et al., 2013; Shaw et al., 2019). Unless combined with other paleo-community proxies (such as biomarkers and/or microfossils, as performed in the latter studies), amplicons are unsuitable for *sed*aDNA research as they are biased by a range of confounding factors such as random amplification in the first few PCR cycles, selective amplification of contaminant sequences and fragment-length size dependency (Armbrecht et al., 2019; Taberlet et al., 2012; Ziesemer et al., 2015). Moreover, amplicon data is very difficult to authenticate as "ancient", because the distribution of fragment-sizes within a sample cannot be reliably determined and damage patterns such as cytosine deamination are no longer detectable using bioinformatic tools (Ginolhac, Rasmussen, Gilbert, Willerslev, & Orlando, 2011). Our study based on Tasmanian seafloor material showed, for the first time, that the average fragment length for marine *sed*aDNA (0–10 cm depth) can be very short (~69 bp), confirming the suitability of metagenomics for the quantification of highly fragmented and degraded marine *sed*aDNA. Whether this latter figure (69 bp) might have also been influenced by potential preferential ligation of small DNA fragments to the flow cell during

sequencing (as discussed in Head et al., 2014) is unknown; however, if this was the case such an artefact would apply to all samples and EBCs, and as such, should not impact our overall results.

Using a comprehensive set of different extraction techniques, we identified a total of 45 eukaryotic groups (44 phyla and fungi), including key marine plankton indicator groups such as Bacillariophyta (diatoms), Dinophyceae (dinoflagellates) and Haptophyceae (haptophytes, including coccolithophores). These phyla are the most abundant photosynthetic eukaryotes at our study location off the east coast of Tasmania (Brown et al., 2018; Hallegraeff et al., 2010) and globally (De Vargas et al., 2015). Within Bacillariophyta, we found Bacillariophyceae, Coscinodiscophyceae and Mediophyceae to be the represented taxa. Interestingly, no reads were obtained for diatoms after applying the Si4 extraction protocol, suggesting that sediment freezing prior to aDNA extraction promotes cell lysis and improves subsequent DNA isolation efficiency. A low representation of diatoms in *sed*aDNA has been noted previously from Maria Island inshore sediments, even when using an amplicon approach targeting the 18S gene region (Shaw et al., 2019). Critically, diatom silica frustules may potentially compete with silica-based spin columns and dissolved silica for DNA-binding sites, a phenomenon that requires further investigation. Within the Dinophyceae, most reads from Gonyaulacales, Peridiniales, Suessiales and Syndiniales were retained after applying the Silica and Kit extraction protocols. These techniques comprised an initial bead-beating step that is perhaps more efficient at breaking cellular membranes before DNA extraction, hence leading to more effective DNA capture. Using microscopy, *Protoperidinium* and *Gonyaulax/Protoceratium* have been identified as the dominant dinoflagellate cyst types in our samples (unpublished data), consistent with our *sed*aDNA analyses. For Haptophyceae, it appears that only the gentleness of the EDTA approach allowed detection of Prymnesiales (in addition to Isochrydiales). However, the detection of this group (Prymnesiales) was due to a single read assigned to *Chrysochromulina*, thus further testing is required to confirm that the EDTA extraction approach consistently improves read-yields for Prymnesiales specifically. While our study presents a method that optimizes DNA extractions across eukaryotes and retains small DNA fragments, differences in read yield for specific taxa within the eukaryotes should be considered when the aim is to target a particular taxon or group of taxa. To increase the yield of specific taxa in future studies, we therefore suggest adding target-capture techniques commonly applied in aDNA research (Horn, 2012) as they maximize the quantity and detection of DNA from targeted taxa (Taberlet et al., 2012).

4.3 | Contamination assessment

Our extraction protocols detected 19 bacterial and four eukaryotic taxa in EBCs (Table S1). Bacterial contamination is a known phenomenon in aDNA studies and demands the inclusion of EBCs at all stages from extraction and library preparation to sequencing (Salter et al., 2014). To our knowledge, this is the first study reporting on laboratory contaminants in the marine *sed*aDNA literature, providing a starting point for contaminant tracking in this research field. The primary eukaryote group detected in our EBCs were Thecofilosea within the phylum Cercozoa, especially in the EBC run with the EDTA extractions. Given that Thecofilosea are widespread marine protozoans within the Rhizaria, it is important to consider cross-contamination of reagents or from a sample during our EDTA extraction. However, cross-contamination should have introduced a species diversity similar to that detected in the EDTA-extracted samples, rather than just one taxon, therefore, we find cross-contamination quite unlikely. An alternative explanation might be that Thecofilosea sequences were a PCR artefact as they were consistently short (mostly ~ 27 bp) and just above the minimum bit score of 50 allowed in our alignments (scores between 51 and 54). Either way, because we subtracted all taxa found in EBCs from our samples, contaminant sequences (like those of Thecofilosea) did not impact our statistical analyses. However, their detection emphasizes the importance of stringent contamination controls and data filtering for *sed*aDNA studies (Pedersen et al., 2015). We strongly recommend keeping a detailed record of contaminant taxa in the laboratory for internal contamination tracing, and to report contaminant taxa identified alongside *sed*aDNA results to ensure authenticity—a practice widely used in microbiome research, which also deals with low microbial biomass (Eisenhofer et al., 2018).

4.4 | Extraction of Bacteria and Archaea from marine sediments

Our analyses focused on finding the best extraction method for marine eukaryote *sed*aDNA. However, our data also allowed some assessment of extraction-treatment effects on co-occurring Bacteria and Archaea. Briefly, all extraction treatments except Si4 performed similarly well in yielding bacterial *sed*aDNA, with Com being the most efficient. For Archaea, the kit showed the highest treatment efficiency. However, since both kit and EDTA (second most efficient treatment) included silica spin columns, Com should be considered an alternative if the aim is to retain small, and potentially authentic, DNA fragments. We only dealt with Bacteria and Archaea at the domain-level in our analyses, and the effects of different treatments on individual bacterial and archaeal taxa within these two domains should be the subject of future research.

In conclusion, we used seven different *sed*aDNA extraction techniques for eukaryote DNA from marine sediments. We aimed to optimize read yield, retain small and degraded DNA fragments, and maximize taxonomic coverage to improve the study of biological change in deep seafloor sediments over geological time scales. The optimized combined (Com) technique worked efficiently with only 0.25 g of sediment and so stands as the most cost-effective for future marine *sed*aDNA studies. Com can be completed in ~1 day and lays the foundation for future methodological refinements, particularly in the analysis of *sed*aDNA damage, diatom aDNA retrieval, and use of hybridisation enrichment

target-capture techniques to maximize eukaryote read yield while preserving taxonomic representation of the original communities. Our bioinformatic pipeline for processing metagenomic data incorporates stringent data-filtering criteria to maximize the yield of uncontaminated DNA material. In addition to providing optimized *sed*aDNA extraction and bioinformatic protocols, we contribute a quantitative, metagenomic *sed*aDNA assessment of past eukaryote abundance and diversity from Tasmanian marine sediments. Our study should facilitate future investigations into marine paleo-eukaryote community dynamics from Tasmanian sediments as well as other regions.

ACKNOWLEDGEMENTS

We thank Laura Weyrich, Kieren Mitchell, Holly Heiniger, Nicole Moore and Steve Johnson for their help and advice during extractions and library preparations. We are grateful to Yassine Souilmi and Yichen Liu for their expert guidance on bioinformatic analyses. We thank the Marine National Facility,the crew of *RV Investigator* voyage IN2018_T02 and the scientific team for their support during field work (2018 MNF Grant H0025318), and Prof. Andrew McMinn and Dr Craig Woodward for assistance with collection of the cores. This study was funded through the 2017 Australian Research Council (ARC) Discovery Project DP170102261. SHP and RE were funded by the ARC Discovery Project DP170104665 and the ARC Centre of Excellence for Australian Biodiversity and Heritage CE170100015, respectively, while AC was funded by the ARC Laureate Fellowship FL140100260.

AUTHOR CONTRIBUTIONS

L.A. conceived the idea and designed this research, carried out laboratory work, bioinformatic analyses and wrote the first draft of the manuscript. S.H.P. did the statistical analyses and contributed to the writing of the first draft of the manuscript. R.E. contributed guidance on the bioinformatic analyses. G.H.and C.B. collected and provided the core samples. G.H. and C.B. assisted with data interpretation and A.C. provided guidance on ancient DNA analyses. G.H., C.B. and A.C. secured funding for this project. All coauthors provided comments and feedback on manuscript drafts, and edited the final manuscript submitted for publication.

DATA AVAILABILITY STATEMENT

The demultiplexed data that support the findings of this study are openly available in the NCBI Sequence Read Archive at http://www.ncbi.nlm.nih.gov/bioproject/558683, BioProject ID PRJNA558683 (BioSample accessions SAMN12496346 - SAMN12496352). A metadata table identifying each sediment sample and applied extraction/library preparation technique against NCBI accession numbers is provided with the Supporting Information. Furthermore, the Supporting Information contains a description of our complete bioinformatic pipeline, two Supporting Information figures (proportion of reads assigned to individual taxa), and three Supporting Information tables (list of contaminant taxa and statistical results). The R-script used to generate the statistical results of this study and the associated data files are also provided with this publication.

ORCID

Linda Armbrecht https://orcid.org/0000-0002-1213-1257
Salvador Herrando-Pérez https://orcid.org/0000-0001-6052-6854
Raphael Eisenhofer https://orcid.org/0000-0002-3843-0749
Gustaaf M. Hallegraeff https://orcid.org/0000-0001-8464-7343
Christopher J. S. Bolch https://orcid.org/0000-0001-9047-2391
Alan Cooper https://orcid.org/0000-0002-7738-7851

REFERENCES

Armbrecht, L., Coolen, M. J. L., Lejzerowicz, F., George, S. C., Negandhi, K., Suzuki, Y., … Weyrich, L. S. (2019). Ancient DNA from marine sediments: Precautions and considerations for seafloor coring, sample handling and data generation. *Earth-Science Reviews*, *196*, 102887. https://doi.org/10.1016/j.earscirev.2019.102887

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, *160*, 268–282.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., & Adler, C. J., … The Genographic Project (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications*, *4*, 1764. https://doi.org/10.1038/ncomms2656

Brown, M. V., Van De Kamp, J., Ostrowski, M., Seymour, J. R., Ingleton, T., Messer, L. F., … Nelson, T. M. (2018). Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative. *Scientific Data*, *5*, 180130. https://doi.org/10.1038/sdata.2018.130

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference*. Business Media LLC: Springer Science.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., … Kirilovsky, A. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*, *9*, 373. https://doi.org/10.1038/s41467-017-02342-1

Clarke, E. L., Taylor, L. J., Zhao, C., Connell, A., Lee, J. J., Fett, B., … Bittinger, K. (2019). Sunbeam: An extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome*, *7*, 46. https://doi.org/10.1186/s40168-019-0658-x

Coolen, M. J., Boere, A., Abbas, B., Baas, M., Wakeham, S. G., & Sinninghe Damsté, J. S. (2006). Ancient DNA derived from alkenone-biosynthesizing haptophytes and other algae in Holocene sediments from the Black Sea. *Paleoceanography*, *21*, PA1005. https://doi.org/10.1029/2005PA001188

Coolen, M. J., Orsi, W. D., Balkema, C., Quince, C., Harris, K., Sylva, S. P., … Giosan, L. (2013). Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 8609–8614. https://doi.org/10.1073/pnas.1219283110

Coolen, M. J., Saenz, J. P., Giosan, L., Trowbridge, N. Y., Dimitrov, P., Dimitrov, D., & Eglinton, T. I. (2009). DNA and lipid molecular stratigraphic records of haptophyte succession in the Black Sea during the Holocene. *Earth and Planetary Science Letters*, *284*, 610–621. https://doi.org/10.1016/j.epsl.2009.05.029

De Schepper, S., Ray, J. L., Skaar, K. S., Sadatzki, H., Ijaz, U. Z., Stein, R., & Larsen, A. (2019). The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. *The ISME journal*, *13*, 2566–2577.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., … Carmichael, M. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*, 1261605. https://doi.org/10.1126/science.1261605

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2018). Contamination in low microbial biomass microbiome studies: Issues and recommendations. *Trends in Microbiology*, *27*, 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Epp, L. S., Zimmermann, H. H., & Stoof-Leichsenring, K. R. (2019). Sampling and extraction of ancient DNA from sediments. In *Ancient DNA* (pp. 31–44). New York, NY: Humana Press.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MULTIQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Frisia, S., Weyrich, L. S., Hellstrom, J., Borsato, A., Golledge, N. R., Anesio, A. M., … Cooper, A. (2017). The influence of Antarctic subglacial volcanism on the global iron cycle during the Last Glacial Maximum. *Nature communications*, *8*, 1–9.

Fry, J. C., Parkes, R. J., Cragg, B. A., Weightman, A. J., & Webster, G. (2008). Prokaryotic biodiversity and activity in the deep subseafloor biosphere. *FEMS Microbiology Ecology*, *66*, 181–196. https://doi.org/10.1111/j.1574-6941.2008.00566.x

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). MAPDAMAGE: Testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*, 2153–2155. https://doi.org/10.1093/bioinformatics/btr347

Guo, L., Sui, Z., Zhang, S., Ren, Y., & Liu, Y. (2015). Comparison of potential diatom 'barcode' genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 1369–1380. https://doi.org/10.1099/ijs.0.000076

Hallegraeff, G. M., Bolch, C. J. S., Hill, D. R. A., Jameson, I., Leroi, J.-M., McMinn, A., & Saunders, K. (Eds.) (2010). *Algae of Australia: Phytoplankton of temperate coastal waters*. Canberra and Melbourne: ABRS and CSIRO Publishing.

Hamm, C. E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K., Smetacek, V. (2003). Architecture and material properties of diatom shells provide effective mechanical protection. *Nature*, *421*, 841. https://doi.org/10.1038/nature01416

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*, 61–77. https://doi.org/10.2144/000114133

Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J., & Huson, D. H. (2016). MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *BioRxiv*, 050559.

Horn, S. (2012). Target enrichment via DNA hybridization capture. In B. Shapiro, & M. Hofreiter (Eds.), *Ancient DNA*, *methods and protocols* (pp. 177–188). New York, Dordrecht, Heidelberg, London: Springer.

Hou, W., Dong, H., Li, G., Yang, J., Coolen, M. J., Liu, X., … Lian, B. (2014). Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Scientific Reports*, *4*, 6648.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., & Tappu, R. (2016). MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Computational Biology*, *12*, e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Inagaki, F., Hinrichs, K. U., Kubo, Y., Bowles, M. W., Heuer, V. B., Hong, W. L., … Kaneko, M. (2015). Exploring deep microbial life in coalbearing sediment down to ~2.5 km below the ocean floor. *Science*, *349*, 420–424.

Jolliffe, I. T. (2002). *Principal component analysis*. New York, NY: Springer.

Klappenbach, J. A., Saxman, P. R., Cole, J. R., & Schmidt, T. M. (2001). RRNDB: The ribosomal RNA operon copy number database. *Nucleic Acids Research*, *29*, 181–184. https://doi.org/10.1093/nar/29.1.181

Klouch, K. Z., Schmidt, S., Andrieux-Loyer, F., Le Gac, M., Hervio-Heath, D., Qui-Minet, Z. N., … Siano, R. (2016). Historical records from dated sediment cores reveal the multidecadal dynamic of the toxic dinoflagellate *Alexandrium minutum* in the Bay of Brest (France). *FEMS Microbiology Ecology*, *92*, 7.

Krampitz, G., Drolshagen, H., Häusle, J., & Hof-Irmscher, K. (1983). Organic matrices of mollusc shells. *Biomineralization and biological metal accumulation* (pp. 231–247). Dordrecht: Springer.

Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., … Pawlowski, J. (2013). Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology Letters*, *9*, 20130283. https://doi.org/10.1098/rsbl.2013.0283

Lloyd, K. G., MacGregor, B. J., & Teske, A. (2010). Quantitative PCR methods for RNA and DNA in marine sediments: Maximizing yield while overcoming inhibition. *FEMS Microbiology Ecology*, *72*, 143–151. https://doi.org/10.1111/j.1574-6941.2009.00827.x

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbour Protocols*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Morard, R., Lejzerowicz, F., Darling, K. F., Lecroq-Bennet, B., Pedersen, M. W., Orlando, L., … Kucera, M. (2017). Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. *Biogeosciences*, *14*, 2741–2754. https://doi.org/10.5194/bg-14-2741-2017

More, K. D., Orsi, W. D., Galy, V., Giosan, L., He, L., Grice, K., & Coolen, M. J. (2018). A 43 kyr record of protist communities and their response to oxygen minimum zone variability in the Northeastern Arabian Sea. *Earth and Planetary Science Letters*, *496*, 248–256. https://doi.org/10.1016/j.epsl.2018.05.045

Nayfach, S., & Pollard, K. S. (2016). Toward accurate and quantitative comparative metagenomics. *Cell*, *166*, 1103–1116. https://doi.org/10.1016/j.cell.2016.08.007

Netzel, H., & Duerr, G. (1984). The dinoflagellate cell cortex. In D. Spector (Ed.), *Dinoflagellates* (pp. 34–105). Orlando, FL: Academic Press.

Neugebauer, E., Bykowski, P., Neugebauer, E., & Zjednoczenia, A. (1986). The krill chitin and some aspects of metals transport in Antarctic sea water. *Polish Polar Research*, *7*, 371–376.

Orsi, W. D., Coolen, M. J., Wuchter, C., He, L., More, K. D., Irigoien, X., … Galy, V. (2017). Climate oscillations reflected within the microbiome of Arabian Sea sediments. *Scientific Reports*, *7*, 6040. https://doi.org/10.1038/s41598-017-05590-9

Pawłowska, J., Zajączkowski, M., Łącka, M., Lejzerowicz, F., Esling, P., & Pawlowski, J. (2015). Palaeoceanographic changes in Hornsund Fjord (Spitsbergen, Svalbard) over the last millennium: New insights from ancient DNA. *Climate of the past*, *11*, 3665–3698. https://doi.org/10.5194/cpd-11-3665-2015

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C. D., Haile, J., Hellstrom, M., … Schnell, I. B. (2015). Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B*, *370*, 20130383. https://doi.org/10.1098/rstb.2013.0383

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roh, C., Villatte, F., Kim, B. G., & Schmid, R. D. (2006). Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Applied Biochemistry and Biotechnology*, *134*, 97–112. https://doi.org/10.1385/ABAB:134:2:97

Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols*, *13*, 2447. https://doi.org/10.1038/s41596-018-0050-5

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology*, *12*(1), 87.

Sarras, M. P., Jr., Madden, M. E., Zhang, X., Gunwar, S., Huff, J. K., & Hudson, B. G. (1991). Extracellular matrix (mesoglea) of *Hydra vulgaris*: I. Isolation and characterization. *Developmental Biology*, *148*, 481–494.

Schubert, M., Lindgreen, S., & Orlando, L. (2016). ADAPTERREMOVAL v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88. https://doi.org/10.1186/s13104-016-1900-2

Shapiro, B., & Hofreiter, M. (2012). Introduction. In *Ancient DNA – Methods and protocols*. New York, Dordrecht, Heidelberg, London: Springer.

Shaw, J. L. A., Weyrich, L. S., Hallegraeff, G., & Cooper, A. (2019). Retrospective eDNA assessment of potentially harmful algae in historic ship ballast tank and marine port sediments. *Molecular Ecology*, *28*, 2476–2485.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., De la Rasilla, M., Lalueza-Fox, C., … Stewart, J. R. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*, 605–608. https://doi.org/10.1126/science.aam9695

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods*, *A7*, 13–26.

Szczuciński, W., Pawłowska, J., Lejzerowicz, F., Nishimura, Y., Kokociński, M., Majewski, W., … Pawlowski, J. (2016). Ancient sedimentary DNA reveals past tsunami deposits. *Marine Geology*, *381*, 29–33. https://doi.org/10.1016/j.margeo.2016.08.006

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., & Ecker, J. R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protocols*, *10*, 475. https://doi.org/10.1038/nprot.2014.114

Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., … Pendleton, J. (1994). Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology*, *43*, 250–261. https://doi.org/10.1093/sysbio/43.2.250

Wang, C., Zhang, T., Wang, Y., Katz, L. A., Gao, F., & Song, W. (2017). Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: Impact of copy number variation and experimental error. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170425.

Webster, G., Newberry, C. J., Fry, J. C., & Weightman, A. J. (2003). Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: A cautionary tale. *Journal of Microbiological Methods*, *55*, 155–164. https://doi.org/10.1016/S0167-7012(03)00140-4

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., … Farrell, M. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*, 357–361.

Willerslev, E., & Cooper, A. (2005). Ancient DNA. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 3–16.

Ziesemer, K. A., Mann, A. E., Sankaranarayanan, K., Schroeder, H., Ozga, A. T., Brandt, B. W., … Aldenderfer, M. (2015). Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports*, *5*, 16498. https://doi.org/10.1038/srep16498

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Armbrecht L, Herrando-Pérez S, Eisenhofer R, Hallegraeff GM, Bolch CJS, Cooper A. An optimized method for the extraction of ancient eukaryote DNA from marine sediments. *Mol Ecol Resour*. 2020;20:906–919. https://doi.org/10.1111/1755-0998.13162