Current Protocols

WILEY

Isolation and Analyses of DNA from Historical and Ancient Plant Tissues

Journal:	*Current Protocols*
Manuscript ID	Draft
Wiley - Manuscript type:	Protocol
Date Submitted by the Author:	n/a
Complete List of Authors:	Latorre, Sergio; Max-Planck-Institute for Developmental Biology, Department of Molecular Biology Lang, Patricia; Stanford University, Department of Biology Burbano, Hernán; University College London, Department of Genetics Gutaker, Rafal; Royal Botanic Gardens Kew, Biodiversity Informatics and Spatial Analysis ; New York University, Center for Genomics and Systems Biology
Keywords:	ancient DNA, genomics, herbarium specimens, archaeological remains, plant
Abstract:	The ability to sequence DNA retrieved from ancient and historical material (aDNA) plays a crucial role in reinforcing evolutionary and anthropological inference. While the focus of the field is largely on analyzing DNA from ancient hominids and other animals, we have also learned from plant aDNA, in particular, about human farming practices, crop domestication, environment management, species invasion, and adaptation to various environmental conditions. In this protocol, we outline best practices for plant aDNA isolation, preparation for sequencing, bioinformatic processing and authentication. We describe the process all the way from processing of archaeological or historical plant material to characterizing and authenticating sequencing reads. In alternative protocols, we include modifications to this process that are tailored to strongly degraded DNA. Throughout, we stress the importance of precautionary measures to successfully analyze aDNA. Finally, we discuss the evolution of the archaeogenomics field and the development of new methods, which both shaped this protocol

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Template for a Protocol Article

Please submit this Word document to ScholarOne – every section must be filled for the manuscript to be reviewed.

Article Title

Extraction, library preparation and bioinformatic analysis of historical and ancient plant DNA

AUTHOR(S) AND CONTACT INFORMATION:

Sergio M. Latorre[1], Patricia L. M. Lang[2], Hernán A. Burbano[1,3], Rafal M. Gutaker[4,5]*

[1]Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany.

[2]Department of Biology, Stanford University, Stanford, CA, USA.

[3]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London, UK.

[4]Center for Genomics and Systems Biology, New York University, New York, NY, USA.

[5]Royal Botanic Gardens, Kew, Richmond TW9 3AE. London, UK.

*Corresponding author: rafal.gutaker@gmail.com

SIGNIFICANCE STATEMENT:

Historical and ancient remains hold detailed information about the cultural context and evolutionary history of a species. The DNA preserved within old plant specimens (aDNA) has fruited in key insights into crop domestication and farming, species invasion, and adaptation to changing environments. The success of these studies was possible owing to special precautions and molecular methods adapted to the properties of historical plant remains and their DNA. Here, we detail best practices, and protocols for plant aDNA isolation, processing for high-throughput genome sequencing, and the authentication of sequenced aDNA.

ABSTRACT:

The ability to sequence DNA retrieved from ancient and historical material (aDNA) plays a crucial role in reinforcing evolutionary and anthropological inference. While the focus of the field is largely on analyzing DNA from ancient hominids and other animals, we have also learned from plant aDNA, in particular, about human farming practices, crop domestication, environment management, species invasion, and adaptation to various environmental conditions. In this protocol, we outline best practices for plant aDNA isolation, preparation for sequencing, bioinformatic processing and authentication. We describe the process all the way from processing of archaeological or historical plant material to characterizing and authenticating sequencing reads. In alternative protocols, we include modifications to this process that are tailored to strongly degraded DNA. Throughout, we stress the importance of precautionary measures to successfully analyze aDNA. Finally, we discuss the evolution of the archaeogenomics field and the development of new methods, which both shaped this protocol.

Basic Protocol 1: Isolation of aDNA

Alternate Protocol 1: Isolation of ultra-short DNA

Support Protocol 1: Preparation of PTB-based mix

Support Protocol 2: Preparation of Binding buffer

Basic Protocol 2: Preparation of genomic libraries

Alternate Protocol 2: Preparation of genomic libraries with uracil removal

Basic Protocol 3: Bioinformatic processing and authentication of aDNA

KEYWORDS:

ancient DNA, plant, genomics, herbarium specimens, archaeological remains.

INTRODUCTION:

The inclusion of DNA sequences retrieved from ancient and historical remains (aDNA) plays a crucial role in reinforcing evolutionary and anthropological inference. In the last decade, the rapid development of next generation sequencing (NGS) and bioinformatic methodologies have increased our ability to retrieve, analyze, and most importantly authenticate genomic DNA from historical and prehistorical samples (Orlando et al., 2015). Before the advent of NGS, aDNA analysis was limited to the analysis of a handful of loci retrieved by direct PCR (Svante Pääbo et al., 2004). PCR-based methods have numerous shortcomings, many of which are problematic in the context of aDNA. They have low sensitivity, require large amounts of input DNA, are limited to amplification of relatively long DNA fragments (> 50bp), and have high rates of false-positive amplification, most frequently primer-dimers. In addition to such technical limitations, the authentication of DNA relied exclusively on the reproduction of PCR results in at least one independent laboratory, compromising the reproducibility of aDNA research. Thus, although direct PCR

was the standard before 2006, this technique is not a good fit for degraded aDNA. The shift from PCR to ligation-based methods alleviated most of those issues. In line with previous suggestions, we recommend to completely forgo the use of PCR-based sequencing for degraded DNA (Gutaker & Burbano, 2016; Prüfer & Meyer, 2015). Using ligation-based genomic libraries, the first draft genome was sequenced for a wooly mammoth (Poinar et al., 2006). Since 2010, the archaeogenetics field has blossomed as evidenced by over a thousand ancient human genomes sequenced, which were published in hundreds of papers (Skoglund & Mathieson, 2018). Throughout, new methods and tools were created that have increased the pace of data generation and inspired further investigation of degraded DNA from other animal and plant tissues, deposits of eukaryotic and prokaryotic DNA such as dental calculus and coprolites, as well as DNA found in sediments.

Here, we compile methods that were originally created for animal degraded DNA and were tailored for use with degraded plant materials for the purpose of genomic sequencing (Fig. 1). In this protocol article, we outline how to isolate ancient DNA from degraded plant tissue, construct genomic libraries, and process sequencing data from NGS platforms, while taking into account specific characteristics of ancient DNA. Throughout, we emphasize the risks of contaminating aDNA with exogenous DNA, and the importance of authentication of bona fide endogenous DNA.

STRATEGIC PLANNING:

Due to the lack of a priori information regarding the magnitude of DNA degradation in plant archaeological and historical samples, it is reasonable to assume that we are handling DNA that is damaged, fragmented, and contaminated with environmental DNA (see Comments for details). Typically, aDNA libraries are first screened by being sequenced in low-throughput mode. These libraries are prepared without the inclusion of any aDNA-specific repair steps, which permits the authentication of DNA. The output of the screening should be at least 10,000 DNA fragments per sample, which are used for characterizing the magnitude of DNA degradation, for authentication, as well as for the estimation of endogenous DNA content to plan high-throughput sequencing.

In general, the lower the percentage of endogenous plant DNA relative to the total DNA in a sample, the higher the risk of exogenous contamination. Degraded DNA can be easily contaminated with DNA from other samples, environmental DNA, human DNA, and residual DNA from laboratory spaces, equipment and reagents. For those reasons it is recommended to use certain standard practices when handling degraded plant DNA, similar to those used when working with degraded human DNA. Firstly, a strict physical separation of pre- and post-amplification laboratory spaces; secondly, maintaining an extremely clean pre-amplification environment; and thirdly, working with small experimental batches, while using extraction and library negative controls.

Plant material selected for DNA sequencing can be classified in three contamination risk categories depending on the percentage of endogenous DNA usually found in different types of samples:

Category 1.: very low risk - fresh material and recent herbarium material (up to one year old). This category will not be discussed here since it resembles fresh plant material (Weiß et al., 2016).

Category 2.: low risk - herbarium material that is between 1 and 100 years old.

Category 3.: high risk - herbarium material that is 100 years or older, herbarium material of any age when used to analyze pathogenic and non-pathogenic microorganisms, and archaeological material independent of its age.

All pre-amplification steps (Basic Protocol 1: Parts I and II, and Basic Protocol 2: Parts I-IV) should be carried out in isolated pre-amplification (pre-Amp) lab facilities (Fulton & Shapiro, 2019). Library preparation, including the preparation of the amplification mixture, should be carried out in the pre-Amp facilities. DNA amplification itself should never be carried out in any of the pre-Amp facilities. The amplification mixture, already combined with the extracted DNA, should be moved to the post-amplification (post-Amp) laboratory and amplified there (Basic Protocol 2: Part V). Samples taken out of the pre-Amp facilities should never go back inside the pre-Amp laboratory. Scientists who have moved from the pre-Amp to the post-Amp laboratory should not go back to the pre-Amp laboratory on the same day, unless they have showered and changed clothes. The workflow of Basic Protocols 1 and 2 work very well when distributed among three to four consecutive days. DNA extractions carried out in Basic Protocol 1 start on the afternoon of the first day (Part I), and continue on the second day (Part II). Library preparations throughout Basic Protocol 2 should be conducted in the pre-Amp facilities in the morning (Part I-IV), and continued in the post-Amp facility in the afternoon or, if necessary, on the next day (Part V). Sequencing, bioinformatic processing, and authentication (Basic Protocol 3) can be carried out without spatial and temporal constraints.

Keeping the pre-Amp facilities extremely clean, decontaminated, and sterilized is the cornerstone of the aDNA field (Fulton & Shapiro, 2019; Svante Pääbo et al., 2004). When entering the pre-Amp facilities, researchers should not have been in other laboratory facilities on that day, should wear clean clothes and don protective gear. Typically, protective gear includes: shoe overalls, body overalls with a hood, hairnet, glasses/goggles, face mask, and long nitrile gloves (upon entering the pre-Amp facilities researchers should don an extra pair of gloves that will be frequently changed). This procedure requires a small antechamber room, where researchers can gear-up and decontaminate incoming reagents. The pre-Amp facilities should ideally consist of two separate rooms, one where samples are stored, destructive sampling and documentation is carried out, and a second one, where aDNA is extracted and converted into genomic libraries. The antechamber and the two laboratory rooms should have fixed germicidal UV lamps that remain switched on at all times when no researcher is inside. An additional tabletop UV chamber is essential for directed decontamination of plasticware and some reagents. All rooms should be frequently cleaned with bleach solution and 70% Ethanol solution. Contamination-sensitive operations, such as those described in Basic Protocols 1 and 2, should be conducted in one of three transparent cabinets with additional UV lamps (vertical laminar-flow cabinets are recommended, but in some cases, dead-air boxes are sufficient). Cabinet 1 should be used exclusively for sampling, Cabinet 2 exclusively for handling DNA extracts, and Cabinet 3 exclusively for aliquoting reagents and preparing plasticware. Within Cabinets, researchers should be using filter tips. To avoid contamination, we recommend to dispose of excess liquids from column clean-ups with filter tips as well. Nothing should be deposited on flat surfaces (neither within nor outside the cabinets) without laying out a sheet of tin foil first. After the use of flat surfaces, the tin foil should be discarded and surfaces cleaned with bleach solution first, and then with 70% ethanol. After each use, any tools, pipettes, centrifuges, or any vessels should be cleaned with bleach solution, rinsed with sterile water and UV-treated in a UV chamber for at least 30 minutes at 120 mJ/cm2.

The third and equally important precaution for processing aDNA is keeping the experimental batches small, and including negative controls that will help in determining if pre-Amp facilities are sufficiently clean to avoid contamination. Based on our experience, we suggest to process a maximum of 10 samples in parallel throughout Basic Protocols 1 and 2. We include one negative control for DNA extraction (Basic

Protocol 1, extraction blank), which includes the PTB-based Mix (lysis buffer) without any plant tissue. Throughout library preparation (Basic Protocol 2), we process 10 samples with one extraction blank and add an extra negative control (library blank) that includes sterile water instead of DNA extract, ultimately generating 12 sequencing libraries. Inclusion of those two "blanks" is essential for confirming the lack of DNA contamination in the laboratory, and helps to troubleshoot when contamination is detected.

BASIC PROTOCOL 1

ISOLATION of aDNA

Due to the biochemical characteristics and degraded nature of aDNA, it is necessary to use DNA extractions protocols that efficiently release DNA trapped within or crosslinked with other organic products. Additional challenge lies in the retention and retrieval of very short DNA molecules. Here, we present a protocol that uses an extraction buffer that contains N-Phenacylthiazolium bromide (PTB) and DL-Dithiothreitol (DTT) (Support Protocol 1), compounds that release DNA trapped or crosslinked with different types of organic compounds. We couple the DNA extraction with DNA binding to a silica column (Gutaker et al., 2017; Kistler, 2012; Wales & Kistler, 2019). Additionally, we provide Alternate Protocol 1 that aims at recovering ultra-short DNA molecules by using the same extraction buffer, but binding DNA to silica with a modified binding buffer (Dabney et al., 2013; Gutaker et al., 2017). We recommend the use of Basic Protocol 1 for young herbarium samples (Category 2), whereas we advocate to use Alternate Protocol 1 for old herbarium samples (more than 100 years) and archaeological remains (Category 3).

Materials:

- Nitrile gloves

- Weighing paper

- Photographic camera (optional)

- Tweezers

- Metallic pestles

- Scale

- Parafilm

- Set of micropipettes

- Filter tips

- Rotor

- Incubator

- Microcentrifuge suitable for 1.5 and 2.0-mL microcentrifuge tubes

- Centrifuge with adaptor suitable for 50-mL conical tubes

- 200-µL PCR tubes (1 set per sample)

- 1.5-mL siliconized microcentrifuge tubes (1 set per sample)

- 2-mL microcentrifuge tubes (2 sets per sample)

- 5-mL tubes (1 set per sample)

- 50-mL conical centrifuge tubes (1 set per sample)

- 15-mL conical reservoir (Zymo Research) [CAT: C1031-25] (1 set per sample)

- PTB-based Mix (Support Protocol 1)

- Materials and reagents from the DNeasy® Plant Mini Kit (QIAGEN) [CAT: 69104 / 69106]:

  - Buffer P3

  - Buffer AW1

  - Buffer AW2

  - Buffer AE

  - QIAshredder Spin Columns

  - Collection tubes (2 mL)

Part I [about 3 hours]:

This section will be performed in a pre-Amp facility

1.   Sterilize material with UV light at 120 mJ/cm$_2$ for at least 30 minutes: 2-mL microcentrifuge tubes (2 sets per sample), 1.5-mL siliconized microcentrifuge tubes, 15-mL conical reservoirs, 5 mL tubes, 50-mL conical centrifuge tubes and 200-µL PCR tubes.

Consider extra material as a precaution measure.

2.   Place the sample on weighing paper and document the information about your sample: weight, dimensions, description and additional relevant information.

Consider photographic records when possible.

3.   Subsample the desired amount of tissue and with the help of tweezers, place it inside a labeled 2-mL microcentrifuge tube.

A range from 2 to 20 mg (optimally ~10 mg) of dry plant tissue is recommended.

4.   Using a sterile metal pestle, grind your sample.

Part of the success of the DNA isolation depends on increasing the surface area of the particles. Thus, a finer grain is desirable.

5.   Add 1.2 mL of PTB-based Mix (See Support Protocol 1.1).

6.   Close and seal the microcentrifuge tube with Parafilm.

7.   Place the sealed tube onto a rotor and set it at the slowest possible motion.

8.   Place the rotor within an incubator and set the temperature to 37°C.

9.   Incubate for at least 8 hours.

Overnight incubation is recommended.

Alternatively, if the chosen extraction is 'Dabney et al. modification', continue with Alternate Protocol 1.

Part II [about 6 hours]:

This section should be performed in a pre-Amp facility.

10.   Centrifuge the microtubes with the lysate for 10 minutes at 16.000 x *g* (or maximum speed).

11.   With a pipette, transfer the supernatant (around 1 mL) into a new, sterile and labeled 2-mL microcentrifuge tube.

12.   Add 0.325 volumes (around 325 µL) of Buffer P3 (DNeasy®Ⓡ, QIAGEN), mix by pipetting and incubate for 5 min on ice.

13.   Centrifuge the lysate for 5 min at 16.000 x *g* (or maximum speed).

14.   Pipette a maximum volume of 650 µL of the lysate into a QIAshredder spin column (lilac) placed in a labeled 2-mL collection tube (DNeasy®Ⓡ, QIAGEN).

15.   Centrifuge for 2 min at 16.000 x *g* (or maximum speed).

16.   Without disturbing the pellet, if present, transfer the flow-through into a new, sterile and labeled 5-mL conical tube.

17.   Pipette the remaining lysate (or a maximum volume of 650 µL) into the QIAshredder spin column (lilac) placed in a 2-mL collection tube (same as in 14.) and centrifuge for 2 min at 16.000 x *g* (or maximum speed).

18.   Without disturbing the pellet, if present, transfer the flow-through into a sterile and labeled 5-mL tube from step 16.

All the lysate should be processed at this stage.

19.   Add 1.5 volumes (around 1.8 mL) of Buffer AW1 (DNeasy®⬚, QIAGEN) and gently mix by pipetting.

20.   Fit a labeled DNeasy Mini spin column (colorless) (DNeasy®⬚, QIAGEN) in a 15-mL conical reservoir (Zymo research) and secure it with sterile masking tape. Place the conical reservoir with the assembled column in a labeled 50-mL conical tube.

Avoid touching the inner part of the conical reservoir by holding it from the edges (Fig. 2).

21.   Transfer the whole volume from 19. into the pre-assembled conical reservoir from 20.

22.   Centrifuge for 10 min at 300 x *g*.

Centrifuge for 5 more minutes if the complete content has not flowed through.

23.   Carefully, remove the DNeasy Mini spin column (colorless) (DNeasy®⬚, QIAGEN) from the conical reservoir and place it into a new labeled 2-mL collection tube (DNeasy®⬚, QIAGEN).

At this point, the DNA should be bound to the spin column. However, as precaution, keep the flow-through at -20°C until your sample has been processed.

24.   Add 500 µL of Buffer AW2 (DNeasy®⬚, QIAGEN) into the column, centrifuge for 1 min at 16.000 x *g* (or maximum speed) and discard the flow through.

25.   Repeat step 24.

26.   Dry-spin for 1 min at 16.000 x *g* (or maximum speed).

27.   Rotate the tube 180 degrees and repeat the step above.

28.   Transfer the column to a new labeled 1.5-mL siliconized microcentrifuge tube.

29.   Add 50 µL of Buffer AE (DNeasy®⬚, QIAGEN) directly onto the column, incubate for 5 min at room temperature and centrifuge for 1 min at 16.000 x *g* (or maximum speed).

Higher volume of the elution buffer might increase the total eluted DNA, but will lower its concentration.

30.   Repeat step 29.

31.   In a 20-µL PCR tube, pipette 3 µL for quantification analyses. Keep the remaining DNA extract at -20°C.

We suggest the use of a Qubit fluorometer because of its minimal volume requirement. Usually, we observe between 0.1 – 10 ng/μL of DNA at this stage.

ALTERNATE PROTOCOL 1

ISOLATION OF ULTRASHORT DNA (DABNEY MODIFICATION)

This protocol uses the same extraction buffer as Basic Protocol 1 but employs an alternative method to bind DNA to a silica column, which includes a different type and amount of binding buffer (Support Protocol 2). These modifications increase the recovery of ultra-short DNA molecules (<50bp) (Dabney et al., 2013; Gutaker et al., 2017). We recommend the use of Alternate Protocol 1 for old herbarium specimens and archaeological remains.

Materials:

- Same materials as Basic Protocol 1

- Sodium Acetate (3M) [Thermo Fisher CAT: 17888]

- Binding Buffer (Support Protocol 2)

- Materials and reagents from the MinElute®️ PCR Purification Kit (QIAGEN) [CAT: 28004 / 28006]

    - MinElute Spin Columns

    - Buffer PE

Part I:

Note: Part I corresponds to the Part I of the Basic Protocol 1.

Part II [about 7 hours]:

This section should be performed in a pre-Amp facility.

10.   Centrifuge the microtubes with the lysate for 10 minutes at 16.000 x *g* (or maximum speed).

11.   With a pipette, transfer the supernatant (around 1 mL) into a new, sterile and labeled 2-mL microcentrifuge tube.

12.   Add 0.325 volumes (around 325 µL) of Buffer P3 (DNeasy®️, QIAGEN) and incubate for 5 min on ice.

13.   Centrifuge the lysate for 5 min at 16.000 x *g* (or maximum speed).

14.   Pipette a maximum volume of 650 µL of the lysate into a QIAshredder spin column (lilac) placed in a labeled 2-mL collection tube (DNeasy®️, QIAGEN).

15.   Centrifuge for 2 min at 16.000 x *g* (or maximum speed).

16.   Without disturbing the pellet, if present, transfer the flow-through into a new, sterile and labeled 15-mL conical tube.

17.   Mix 400 µL of 3M Sodium Acetate with 10 mL of Binding Buffer (Support Protocol 2).

18.   Add the mixture to the 15-mL conical tube. Gently, invert the tube several times.

19. Fit a labeled MinElute Spin column (lilac) (MinElute®⬚, QIAGEN) in a 15-mL conical reservoir (Zymo research) and secure it with sterile masking tape. Place the conical reservoir with the assembled column in a labeled 50-mL conical tube.

Avoid touching the inner part of the conical reservoir by holding it from the edges (Fig. 2).

20. Transfer the whole volume from 18. into the pre-assembled conical reservoir from 19.

21. Centrifuge for 10 min at 300 x *g*.

Centrifuge for 5 more minutes if the complete content has not flowed through.

22. Carefully, remove the MinElute Spin column (lilac) (MinElute®⬚, QIAGEN) and place it into a new labeled 2-mL collection tube (DNeasy®⬚, QIAGEN).

At this point, the DNA should be bound to the spin column. However, as precaution, keep the flow-through at -20°C until your sample has been processed.

23. Dry-spin for 1 min at 3.100 *g*.

24. Add 750 µL of Buffer PE (MinElute®⬚, QIAGEN) and centrifuge for 1 min at 3.100 *g*.

25. Repeat step 24.

26. Dry-spin for 1 min at 16.000 x *g* (or maximum speed).

27. Rotate the tube 180 degrees and repeat the step above.

28. Transfer the column to a new labeled 1,5-mL siliconized microcentrifuge tube.

29. Add 50 ul of Buffer AE (DNeasy®⬚ Plant Mini Kit, QIAGEN), incubate for 5 min at room temperature and centrifuge for 1 min at 16.000 x *g* (or maximum speed).

Higher volume of the elution buffer might increase the total eluted DNA, but will lower the concentration.

30. Repeat step 29.

31. In a 20-µL PCR tube, pipette 3 µL for quantification analyses. Keep the remaining DNA extract at -20°C.

We suggest the use of a Qubit fluorometer because of its minimal volume requirement. Usually, we observe between 0.1 – 10 ng/μL of DNA at this stage

SUPPORT PROTOCOL 1

PREPRATION OF PTB-BASED MIX

Releasing DNA trapped or crosslinked to organic compounds is a key step in aDNA extraction. This protocol describes the preparation of the extraction buffer, which aims at releasing DNA trapped in or crosslinked

with different types of organic compounds (Gutaker et al., 2017; Jaenicke-Després et al., 2003; Kistler, 2012; Wales & Kistler, 2019).

Materials:

- Scale

- UV crosslinker

- Tris-HCl, pH 8.0, RNase free, sterile (1M) [CAS: 1185-53-1]

- Ethylenediaminetetraacetic acid (EDTA) (0.5M), pH 8.0, sterile [CAS: 60-00-4]

- Sodium chloride (NaCl) (5M), sterile [CAS: 7647-14-5]

- N-Phenacylthiazolium bromide (PTB) [CAS: 5304-34-7]

- DL-Dithiothreitol (DTT), >= 99.0% [CAS: 3483-12-3]

- Proteinase K [Invitrogen CAT: FEREO0491; CAS: 9001-92-7]

- Sodium dodecyl sulfate (SDS) (10%), sterile [CAS: 151-21-3]

- Ethyl alcohol, pure, 200 proof, molecular biology grade [CAS: 64-17-5]

1.   Prepare a PTB precursor stock solution 1 (4X) following the volumes and concentrations shown in Table 1.1.

Notice that the calculated volumes are based on the indicated starting concentrations. In case your starting concentrations differ, adjust the volumes accordingly.

We suggest a stock solution 1 of a final volume of 50 mL. Adjust the final volume based on your experiment size.

2.   Sterilize with UV light for 30 minutes and store at room temperature.

Lid must be open to permit direct flux of the UV radiation and should be closed as soon as the treatment is finished. Hereafter, the solution should be opened only inside sterilized cabinets.

3.   Prepare the amount of PTB-based Mix needed according to the number of reactions (Table 1.2).

Consider extra volume as precaution.

Do not store the PTB-based Mix (PTB should always be freshly added).

SUPPORT PROTOCOL 2

PREPARATION OF BINDING BUFFER

This protocol describes the preparation of a binding buffer optimized for the recovery of ultra-short DNA molecules. For an efficient recovery of ultra-short molecules it is important to use this buffer in excess, hence we recommend preparing it in large quantities. Alternatively, buffers with similar composition can be obtained commercially.

Materials:

- Guanidine hydrochloride [CAS: 50-01-1; Sigma-Aldrich CAT: G3272]

- Isopropyl alcohol [CAS: 67-63-0]

1. Prepare the desired volume depending on the number of reactions following Table 1.3.

   Consider extra volume as precaution.

BASIC PROTOCOL 2

PREPARATION OF GENOMIC LIBRARIES

Preparing genomic libraries for Illumina-based sequencing is a key step in the production of high-quality aDNA genomic datasets. In Basic Protocol 2, we generate double-stranded double indexed genomic libraries (Fig. 3)(Kircher et al., 2012; Meyer & Kircher, 2010). The double index system increases the number of samples that can be sequenced in a single sequencing pool and almost completely eliminates the possibility of sample misidentification. Basic Protocol 2 does not include enzymatic steps aimed at removing aDNA-associated DNA misincorporations. Consequently, libraries produced using this protocol can be used to screen and authenticate aDNA. We also present Alternate Protocol 2, which includes additional enzymatic steps that reduce the impact of aDNA-associated misincorporations in downstream data analysis.

Materials:

- Nitrile gloves

- Set of micropipettes

- Filter tips

- Vortex

- Microcentrifuge suitable for 200 µL and 1.5 mL microcentrifuge tubes

- Thermal cycler

- 5-mL tube (1 set per sample)

- 1,5-mL siliconized microcentrifuge tubes (4 sets per sample)

- 200-µL PCR tubes / PCR strip tubes (6 sets per sample)

- HPLC Water

- Elution buffer TET (TE buffer pH7.4, 0.05% Tween 20)

- Elution buffer EB (MinElute® PCR Purification Kit; QIAGEN)

- Materials for the blunt-end repair reaction:

    - NEBuffer™ 2.1 (10x) [CAT: B7202S]

    - Adenosine 5'-Triphosphate (ATP) (10mM) (NEB) [CAT: P0756S]

    - Bovine Serum Albumin (BSA) (NEB) [CAT: B9000S]

    - Deoxynucleotide (dNTP) Solution Mix (NEB) [CAT: N0447S].

    - T4 Polynucleotide Kinase (NEB) [CAT: M0201S]

    - T4 DNA Polymerase (NEB) [CAT: M0203S]

- Materials for the adapter ligation reaction:

    - Quick Ligase Buffer (Part of Quick Ligation™ Kit; NEB) [CAT: M2200S]

    - Adapter Mix (10 µM) (Fig. 3)

- Materials for the fill-in reaction:

    - Bst 2.0 DNA Polymerase (NEB) [CAT: M0537S]

    - Deoxynucleotide (dNTP) Solution Mix (NEB) [CAT: N0447S]

    - Isothermal Amplification Buffer (NEB) [CAT: B0537S]

- Materials for the indexing PCR reaction:

    - *Pfu* reaction buffer (Promega) (CAT: M776A)

    - *Pfu* DNA Polymerase (Promega) (CAT: M774A)

    - P5 indexing primers (Fig. 3)

    - P7 indexing primers (Fig. 3)

- Materials and reagents from the MinElute®⬜ PCR Purification Kit (QIAGEN) [CAT: 28004 / 28006]

  - MinElute Spin Columns

  - Buffer PB

1.   Sterilize material with UV light at 120 mJ/cm^2 for at least 30 minutes: 200-µL PCR tubes (6 sets per sample), 1.5-mL siliconized microcentrifuge tubes (4 sets per sample), 5 mL tubes (1 set per sample).

Consider extra material as a precaution measure.

Part I. Blunt-end repair [about 1.5 hours]:

This section would be performed in a pre-Amp facility.

2.   In a 200-µL PCR tube, prepare the Blunt-end repair reaction following Table 2.1.

We suggest the use of PCR strip tubes when handling multiple samples.

Consider a master mix when handling multiple samples.

Adjust the volumes depending on the number of samples and the initial concentrations. Consider extra volume as precaution.

3.   Place the PCR tubes in a Thermal Cycler, and run a blunt-end program with the following specifications:

Step 1: 15°C for 15 minutes;

Step 2: 25°C for 15 minutes

4.   Add 200 µL of buffer PB (MinElute®⬜ PCR Purification Kit; QIAGEN) to each reaction, mix by pipetting and transfer the mix into a labeled MinElute Spin column (MinElute®⬜ PCR Purification Kit; QIAGEN).

5.   Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

6.   Add 600 µL of buffer PE buffer (MinElute®⬜ PCR Purification Kit; QIAGEN) to the column and discard the flow-through.

7.   Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

8.   Discard the flow-through and centrifuge the column for 60 s at 16.000 x *g* (or maximum speed).

9.   Place the column into a 1.5-mL siliconized microcentrifuge tube and add 18 µL of buffer EB.

To increase the DNA binding, make sure to aim to the center of the column membrane.

10.   Incubate for 60 s at room temperature and centrifuge for 60 s at 16.000 x *g* (or maximum speed).

Part II. Adapter ligation [about 1.5 hours]:

This section should be performed in a pre-Amp facility.

11.   In a 1.5-mL microcentrifuge tube, prepare Adapter Ligation master Mix following Table 2.2.

Adjust the volumes depending on the number of samples and the initial concentrations. Consider extra volume as precaution.

In order to prevent undesirable ligation products, make sure that the ligase is added last, separately to each reaction.

12.   Add the ligation mix to the samples and incubate for 20 min at room temperature.

13.   Add 200 µL of buffer PB (MinElute® PCR Purification Kit; QIAGEN) to each reaction, mix by pipetting and transfer the mix into a MinElute Spin column (MinElute® PCR Purification Kit; QIAGEN).

14.   Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

15.   Add 600 µL buffer PE (MinElute® PCR Purification Kit; QIAGEN) to the column and discard the flow-through.

16.   Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

17.   Discard the flow-through and centrifuge the column for 60 s at 16.000 x *g* (or maximum speed).

18.   Transfer the column into a 1.5-mL siliconized microcentrifuge tube and add 20 µL of buffer EB (MinElute® PCR Purification Kit; QIAGEN).

19.   Incubate for 60 s at room temperature and centrifuge for 60 s at 16.000 x *g* (or maximum speed).

Part III. Fill-in reaction [about 1 hour].

This section should be performed in a pre-Amp facility.

20.   In a 200-µL PCR tube, prepare the fill-in reaction following Table 2.3.

We suggest the use of PCR strip tubes when handling multiple samples.

Consider a master mix when handling multiple samples.

Adjust the volumes depending on the number of samples and the initial concentrations. Consider extra volume as precaution.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

21.   Add the adapter-ligated DNA to the fill-in reaction mix, place the PCR tubes in a Thermal Cycler, and run a fill-in program with the following specifications:

Step 1: 37°C for 20 minutes

Step 2: 80°C for 20 minutes

No clean-up step is required since the enzyme is heat-inactivated.

22.   Pipette 37 µL of the library to a 1.5-mL siliconized microcentrifuge tube. Keep the remaining 3 µL for qPCR-based DNA quantification.

We recommend carrying out a qPCR-based quantification as a quality control checkpoint and to help in the estimation of the number of DNA molecules per library.
Do not run the qPCR quantification inside the pre-Amp facilities. Amplified DNA fragments are a source of contamination.

Part IV. Indexing [about 1 hour].

This section should be performed in a pre-Amp facility.

23.   Prepare the indexing PCR reaction following Table 2.4.

Increasing the volume by splitting the total reaction in four tubes prevents the saturation of subsequent amplification reactions, i.e. the plateau stage of PCR in Part V. If a qPCR-based quantification has been carried out, divide the number of estimated DNA molecules by $2e^{10}$. If the result is higher than 4.0, consider more reactions per sample (see next step).

Notice that the calculated volumes are based on the indicated starting concentrations. In case your starting concentrations differ, adjust the volumes accordingly.

24.   Dispense the 400 µL of total volume in 4 different 200-µL PCR tubes.

If more reactions per sample have been considered, then split the total volume in more 200-µL PCR tubes.
Use PCR stripes when convenient.

25.   Seal the tubes and move to a different facility where the indexing PCR reaction is going to take place.

Do not perform any PCR reaction inside the pre-Amp facilities. Amplified DNA fragments are a source of contamination.

Part V. Amplification and cleaning [about 2 hours]:

This section should be performed in a post-Amp facility.

26.   Shake and spin the tubes.

27.   Perform an indexing PCR with the following specifications:

Step 1: Initial denaturation: 98°C for 30 s

Cycle x 10*:

Step 2.1: Denaturation: 98°C for 10 s

Step 2.2: Annealing: 60°C for 20 s

Step 2.3: Elongation: 72°C for 20 s

Step 3: Final extension: 72°C for 10 min

Avoid an excessive number of cycles because it decreases the DNA complexity during sequencing.

28.　　In a 5-mL tube, pool all reactions from the same sample.

29.　　Add 2 mL of buffer PB (MinElute® PCR Purification Kit; QIAGEN) and mix by pipetting.

30.　　Pipette 600 µL into the MinElute Spin column (MinElute® PCR Purification Kit; QIAGEN). Centrifuge the column for 1 min at 16.000 x *g* (or maximum speed) and discard the flow-through.

31.　　Repeat step 30 until the whole volume has been processed.

32.　　Pipette 750 µL of buffer PE (MinElute® PCR Purification Kit; QIAGEN) into the column and centrifuge for 1 min at 16.000 x *g* (or maximum speed).

33.　　Discard the flow-through and centrifuge for 1 min at 16.000 x *g* (or maximum speed).

34.　　Transfer the column into a new 1.5-mL siliconized microcentrifuge tube.

35.　　Add 20 µL of buffer EB (MinElute® PCR Purification Kit; QIAGEN) and incubate for 5 min at room temperature.

Consider the use of TET buffer for long-term DNA storage.

36.　　Centrifuge for 1 min at 16.000 x *g* (or maximum speed).

When assayed with RT-qPCR the number of DNA molecules in solution should range between 1.0e+08 – 1.0e+11. At this point the genomic libraries should be completely indexed and ready for equimolar pooling. Consult your sequencing provider to adjust the final concentration of genomic libraries.

ALTERNATE PROTOCOL 2

PREPARATION OF GENOMIC LIBRARIES WITH URACIL REMOVAL

This library preparation protocol is a modification of Basic Protocol 2 and includes additional enzymatic steps to reduce the magnitude of aDNA-associated nucleotide misincorporation in downstream data analysis (Briggs et al., 2010; Meyer & Kircher, 2010). This protocol increases the confidence of the

segregating sites identified in aDNA-derived libraries. Its use is recommended for deep sequencing of samples originally screened using Basic Protocol 2.

Materials:

- Same materials as Basic Protocol 2

- USER® enzyme (NEB) (CAT: M5505S)

Part I. Blunt-end repair [about 4.5 hours]:

This section should be performed in a pre-Amp facility.

1. Sterilize material with UV light at 120 mJ/cm$_2$ for at least 30 minutes: 200-µL PCR tubes (6 sets per sample), 1.5-mL siliconized microcentrifuge tubes (4 sets per sample), 5 mL tubes (1 set per sample).

Consider extra material as a precaution measure.

2. Prepare the UDG blunt-end repair reaction following Table A2.1.

We suggest the use of PCR strip tubes when handling multiple samples.

Consider a master mix when handling multiple samples.

Adjust the volumes depending on the number of samples and the initial concentrations. Consider extra volume as precaution.

3. Place the PCR tubes in a Thermal Cycler, and incubate at 38°C for 3 hours.

4. Add 0.4 µL of T4 DNA Polymerase (3,000 U/mL) per sample, place the PCR tubes in the Thermal Cycler, and run a blunt-end program with the following specifications:

Step 1: 25°C for 30 minutes

Step 2: 10°C for 5 minutes

5. Add 200 µL of buffer PB (MinElute®️ PCR Purification Kit; QIAGEN) to each reaction, mix by pipetting and transfer the mix into a labeled MinElute Spin column (MinElute®️ PCR Purification Kit; QIAGEN).

Since the total volume of reaction is larger than the volume of the tube, while mixing, always keep some volume in the pipette tip.

6. Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

7. Add 600 µL buffer PE buffer (MinElute®Ⓡ PCR Purification Kit; QIAGEN) to the column and discard the flow-through.

8. Centrifuge the column for 30 s at 16.000 x *g* (or maximum speed).

9. Discard the flow-through and centrifuge the column for 60 s at 16.000 x *g* (or maximum speed).

10. Place the column into a labeled 1.5-mL siliconized microcentrifuge tube and add 18 µL of buffer TET.

11. Incubate for 60 s at room temperature and centrifuge for 60 s at 16.000 x *g* (or maximum speed).

12. Continue the protocol at Part II: Adapter Ligation from Basic Protocol 2 and subsequent Parts III, IV, and V.

When assayed with RT-qPCR the number of DNA molecules in solution should range between 1.0e+08 – 1.0e+11. At this point the genomic libraries should be completely indexed and ready for equimolar pooling. Consult your sequencing provider to adjust the final concentration of genomic libraries.

BASIC PROTOCOL 3

BIOINFORMATIC PROCESSING AND AUTHENTICATION OF ADNA

We present here a bioinformatics protocol for screening aDNA-derived libraries using Unix-based (Linux, Mac, etc) computer systems. The screening will assess the quality of the libraries and the sequencing run, the magnitude of DNA degradation, and the percentage of endogenous DNA in each sample. Additionally, the screening will serve to authenticate the libraries, since it will quantify the amount of aDNA-associated nucleotide misincorporation present in the libraries that were prepared using Basic Protocol 2 (non-UDG treated). We assume that the libraries have been sequenced using the Illumina platform in paired-end mode and that the sequenced reads have been already demultiplexed and assigned to a given sample (a service commonly provided by sequencing centers). While the processing steps described in this protocol can in principle be performed with a variety of software, we here recommend software that in our experience has worked well for aDNA-derived libraries.

Materials:

Hardware

Computer with at least 4 CPU cores and 4 - 16GB of RAM (depending on the size of the reference genome).

Software

- UNIX-based operating system

- AdapterRemoval v2.3.1 (Schubert et al., 2016) [https://github.com/mikkelschubert/adapterremoval]

- BWA v. 0.7.17 (Li, 2013) [https://github.com/lh3/bwa]

- FastQC v. 0.11.9 (Andrews, n.d.) [https://github.com/s-andrews/FastQC]

- samtools v. 1.10 (Li et al., 2009) [https://github.com/samtools/samtools]

- DeDup v. 0.12.6 (Peltzer et al., 2016) [https://github.com/apeltzer/DeDup]

- mapDamage v. 2.2.1 (Jónsson et al., 2013) [https://github.com/ginolhac/mapDamage]

Files

- Demultiplexed Illumina reads in *fastq* format

- Reference genome in *fasta* format

Note: All the commands are displayed using the following format:

```
$ example command
```

Strings and options that might be changed by the user at convenience are displayed as underlined text between angled brackets:

```
$ example command <change_this_text>
```

A demo version of this protocol can be found at: https://gitlab.com/smlatorreo/plant-adna-pipeline

1.   Open a terminal and create different directories where your data and analyses will be stored.

```
$ mkdir -p 1_initial_data/reference_genome 2_trimmed_merged 3_quality_control 4_mapping 5_aDNA_characteristics
```

2.   Remove the adaptors (trim) and merge the pair mates from the raw Illumina reads with AdapterRemoval v2.

```
$ AdapterRemoval --file1 <reads.P1.fastq.gz> --file2 <reads.P2.fastq.gz> --basename 2_trimmed_merged/<prefix> --collapse --gzip --threads <number_of_threads>
```

Because aDNA molecules are particularly small, most of the pair-mate reads are expected to be merged.

3.   Control the quality of the trimmed and merged and the trimmed and non-merged reads with FastQC.

```
$ fastqc -t <number_of_threads> -o 3_quality_control/ 2_trimmed_merged/<prefix>.collapsed.gz
2_trimmed_merged/<prefix>.pair1.truncated.gz 2_trimmed_merged/<prefix>.pair2.truncated.gz
```

A careful inspection of the "Per base sequence quality", "Sequence Length Distribution" and "Adapter content" together with the "Overrepresented sequences" sections should give a sense of the overall quality of the pre-processed reads (Fig. 4A). As a general rule, aDNA merged reads are expected to have overall good quality scores, a short distribution of fragment lengths (median length <100 bp), and minimal adapter content. In contrast, non-merged reads are expected to have retained some adapter sequences and overall lower quality scores.

4.   Place the reference genome inside a specific folder.

$ mv <reference_genome.fasta> 1_initial_data/reference_genome/.

Choose the phylogenetically closest as well as the latest possible version of the reference genome for the studied species.

Hint: explore EnsemblPlants [http://plants.ensembl.org] or Phytozome [https://phytozome.jgi.doe.gov]

5.   Create indices for the reference genome with both BWA and samtools.

$ bwa index 1_initial_data/reference_genome/<reference_genome.fasta>

$ samtools faidx 1_initial_data/reference_genome/<reference_genome.fasta>

6.   Map the trimmed and merged reads to the indexed reference genome using BWA aln.

$ bwa aln -t <number_of_threads> -l 1024 -f 4_mapping/<prefix>.collapsed.sai
1_initial_data/reference_genome/<reference_genome.fasta> 2_trimmed_merged/<prefix>.collapsed.gz

The option -l 1024 inactivates the seed and as a consequence allows a relaxed alignment at the read ends where aDNA misincorporations are expected more often.

7.   Convert the mapped reads into a standard alignment format (SAM).

$ bwa samse -r @RG\\tID:<sampe_name>\\tSM:<sample_name> -f 4_mapping/<prefix>.sam
1_initial_data/reference_genome/<reference_genome.fasta> 4_mapping/<prefix>.collapsed.sai
2_trimmed_merged/<prefix>.collapsed.gz

The use of the flag -r, which specifies the read group, is optional yet advised for further downstream analyses.

8.   The resulting file 4_mapping/<prefix>.sam contains uncompressed mapped and unmapped reads. Use samtools flagstat to calculate the proportion of mapped reads as a proxy for endogenous DNA.

$ samtools flagstat -t <number_of_threads> 4_mapping/<prefix>.sam > 4_mapping/<prefix>_flagstats.log

Inspect the produced file and estimate the endogenous DNA percentage as the proportion of mapped reads (Fig. 4B).

9.   Keep the mapped reads and create a compressed BAM file using samtools.

$ samtools view -@ <number_of_threads> -F 4 -Sbh -o 4_mapping/<prefix>.mapped.bam 4_mapping/<prefix>.sam

10.   Sort the BAM file by chromosome and position using samtools.

$ samtools sort -@ <number_of_threads> -o 4_mapping/<prefix>.mapped.sorted.bam 4_mapping/<prefix>.mapped.bam

11.   Remove the SAM and unsorted BAM files.

```
$ rm 4_mapping/<prefix>.sam 4_mapping/<prefix>.mapped.bam
```

To optimize disk storage space, always work with BAM sorted files.

12.   Filter out PCR optical duplicates with deDup.

```
$ dedup -i 4_mapping/<prefix>.mapped.sorted.bam -m -o 4_mapping/
```

The output file will have the name <prefix>.mapped.sorted_rmdup.bam.

deDup is optimized to remove merged paired-end PCR duplicates, which makes it more suitable for aDNA reads than other similar software tools. Consider the use of MarkDuplicates from Picard Tools ("Picard Toolkit," 2019) for single-end sequenced or non-merged data.

Inspect the log file generated by dedup and take note of the duplication rate.

13.   Analyze ancient DNA associated characteristics with mapDamage.

```
$ mapDamage -i 4_mapping/<prefix>.mapped.sorted_rmdup.bam -r
1_initial_data/reference_genome/<reference_genome.fasta> -d 5_aDNA_characteristics
```

Consider the use of the option `--merge-reference-sequences` *when working with reference genomes with a large set of chromosomes or contigs.*

Consider the use of the option `--no-stats` *when no specific MCMC analyses are required.*

Inspect the produced plots and figures located in the folder 5_aDNA_characteristics (Fig. 4C)

COMMENTARY

BACKGROUND INFORMATION:

Archaeogenetics is the field in which we analyze DNA material derived from *post-mortem* degraded tissues of living organisms (for example old bones, sediments, seeds, shells) for the purpose of their genetic characterization. Due to temporal provenance of such samples, archaeogenetics have been an extraordinary tool for studying evolutionary histories of species, mechanisms of evolution in general (Gutaker & Burbano, 2016; Kistler et al., 2020; B. Shapiro & Hofreiter, 2014), as well as biological anthropology and archaeology. The early successes of archaeogenetic studies through molecular cloning (Higuchi et al., 1984; S. Pääbo, 1985) were soon overshadowed by a high number of irreproducible PCR-based results (Svante Pääbo et al., 2004). Starting in 2006, during the genomic revolution and at the advent of next generation sequencing (NGS), the whole field took a rapid turn towards high-throughput sequencing. In consequence, we observed the emergence of *archaeogenomics* that promised to resolve the technical issues that had formerly been an obstacle in PCR-based archaeogenetics. With an array of technological and methodological developments (Orlando et al., 2015), NGS allowed DNA from ancient and historical material to be utilized in a more responsible and falsifiable way. Biochemical characteristics of aDNA have been described in detail; aDNA is damaged (subject to spontaneous C-to-U deaminations), fragmented (subject to depurination of DNA backbone), and usually contaminated with exogenous DNA (low percentage of endogenous aDNA) (Briggs et al., 2007; Brotherton et al., 2007). These characteristics, originally defined based on aDNA retrieved from human or animal remains, accurately describe the degraded DNA from historical and archaeological plant material as well (Gutaker & Burbano, 2017).

The milestones in archaeogenomics development can be summarized with a number of key methodological inventions. First was the repurposing of the enrichment methods for aDNA capture. Nucleic acids isolated from historical and archaeological material are low in content of *bona fide*, endogenous DNA. In extreme cases, it is less than 1% of the extracted DNA, and hence it was vital to increase endogenous DNA content for efficient sequencing. Such enrichment can be carried out by hybridizing isolated aDNA to specifically designed DNA probes that are either fixed to a microarray (Burbano et al., 2010; Hodges et al., 2009) or bound to magnetic beads in solution (Avila-Arcos et al., 2011; Fu et al., 2013; Maricic et al., 2010). The second, very important development was the biochemical removal of the product of deamination, uracil, which accumulated at the ends of DNA fragments of old tissues (Briggs et al., 2007). Most common DNA polymerases substitute uracil residues with thymines (Wales et al., 2014), confounding the readout of aDNA sequencing. The use of a special enzyme, Uracil-DNA Glycosylase (UDG), removes most of those deaminated residues before amplification (Briggs et al., 2010). An alternative approach is to leverage the presence of uracils at the ends of degraded DNA molecules to selectively enrich them (Gansauge & Meyer, 2014). This methodology combines the use of UDG with single-stranded library preparation, which has been shown to increase the retrieval of degraded DNA molecules (Gansauge et al., 2017; Gansauge & Meyer, 2013). While we have not observed a beneficial effect of single-stranded library preparation on sequencing aDNA isolated from herbarium material (Gutaker et al., 2017), this method is worthwhile to explore for very old archaeological material (Wales et al., 2015).

Although multiple methods specific to isolating aDNA have been developed, they were tailored mainly to handling molecules isolated from animal remains. While the characteristics of DNA in animals and plants are very similar, the tissues from which we isolate DNA differ crucially (e.g. presence of cell walls) and thus make modified molecular methods necessary. The combination of methodologies for isolating DNA from fresh plant tissues with alterations suitable for aDNA perform very well (Gutaker et al., 2017; Kistler, 2012; Wales et al., 2012; Wales & Kistler, 2019). We coupled them with purification procedures that are commonly used with animal aDNA for the purpose of high-throughput genomic sequencing. In this protocol, we suggest constructing genomic libraries from double-stranded DNA, following a well established protocol (Meyer & Kircher, 2010) with modifications suggested for aDNA, i.e. replacement of magnetic bead-based enzymatic purification with column-based purification (Meyer et al., 2012). In line with suggestions for animal aDNA, we too recommend double indexing of genomic libraries to minimize the risk of read misclassification (Kircher et al., 2012).

Variations of our protocol have been successfully used to sequence the genomes of 6,000-year-old barley seeds (Mascher et al., 2016), 2,000-year-old maize cobs (Swarts et al., 2017), and over a century-old herbarium specimens of *Arabidopsis thaliana* (Durvasula et al., 2017). Our protocol, coupled with targeted enrichment on microarrays, allowed us to sequence DNA from over 350-year-old leaves of potato to over 50x coverage (Gutaker et al., 2019). Finally, this protocol was successful in sequencing endophytic plant pathogens, such as the potato-infecting oomycete, *Phytophthora infestans* (Yoshida et al., 2013).

We would like to acknowledge a series of alternative protocols for isolation of DNA from degraded plant material. In our experience with herbarium material, the PTB-based DNA isolation suggested in this protocol allows the most comprehensive extraction of DNA fragments of all sizes (Gutaker et al., 2017). There are however, other protocols that perform very well, in particular the method by Wales and Kistler (Wales & Kistler, 2019). They default to the use of a lysis buffer for DNA isolation without PTB, recommending a PTB-based buffer only for special cases. This protocol has been successfully used to sequence DNA from ancient seeds of maize (Ramos-Madrigal et al., 2016) and grapes (Ramos-Madrigal et al., 2019). Furthermore, there is a popular CTAB-based isolation of plant aDNA (Palmer et al., 2009), which

has been successfully used for sequencing of aDNA of ancient maize cobs (Kistler et al., 2018) and ancient sorghum seeds (Smith et al., 2019). Finally, we would also like to recognize alternative library preparation methods (Carøe et al., 2018) and bioinformatic pipelines, in particular, EAGER (Peltzer et al., 2016) and PALEOMIX (Schubert et al., 2014), which have been used to process sequencing data for degraded plant and animal DNA.

PRECAUTIONS:

Our protocol, just like any other recognized aDNA protocol, should be carried out with precautionary measures, and in facilities that minimize the risk of sample contamination (for details, see Strategic Planning). There are three important properties of aDNA that make it prone to cross-contamination between samples, and to contamination from contemporary environmental DNA: its damage, fragmentation, and endogenous content. Typically, fresh plant material contains undamaged high-molecular weight DNA, of which over 95% are endogenous. The remaining, "exogenous" DNA usually derives from microbes hosted by the plant. In contrast, herbarium plant DNA has been subject to DNA fragmentation. Fragmentation is a function of conserved tissue, storage conditions and age. It occurred about six times faster in herbarium specimens when compared to Moa bones from New Zealand (Allentoft et al., 2012; Weiß et al., 2016). In consequence, most DNA molecules in hundred-year-old specimens are expected to be no longer than ~80 bp. The endogenous DNA content for potato herbarium specimens vary greatly (35%-95%). The excess of exogenous DNA derives from *post-mortem* microbial colonization, and environmental DNA accumulated during storage. Due to this high fragmentation and low endogenous content, it has been suggested to apply precautions for processing herbarium DNA similar to those applied when working with archaeological material (Shepherd & Perrie, 2014). Archaeological material also varies substantially in both DNA fragmentation and endogenous DNA content, in case of ancient maize cobs it is 30bp - 55bp and 0.1%-85% respectively (Swarts et al., 2017).

A sample with low endogenous DNA content can be easily contaminated with DNA from other samples (even if they also exhibit low endogenous DNA content). Cross-contamination is problematic especially when originating from closely related species or conserved orthologous loci independent of their phylogenetic relationship; in such cases, it is difficult to detect, and almost impossible to correct bioinformatically. While cross-contamination among samples is most problematic, other sources of contamination should be taken into account. The most obvious source of external contamination is the human operator who sheds hair and skin; human DNA has also been detected in saliva aerosols and breath. This source of contamination has been described at the early stage of the archaeogenetics field (Malmström et al., 2005), and while it is more problematic when studying degraded DNA from hominids, it still effectively reduces the amount of sequencing information that is obtained from plant aDNA. Perhaps even more alarming is the possibility of sample contamination from sources that are associated with the human environment. Not only microbes, but also food that has been consumed by humans could be a source of DNA contamination. Finally, it is well understood that many sterile chemicals and reagents are frequently contaminated with bacterial DNA (Glassing et al., 2016). While alarming, the phenomenon of contamination can be minimized or detected (for details see the Strategic Planning section).

UNDERSTANDING RESULTS:

The very same characteristics of aDNA that increase the risks of contamination can be used to authenticate its historical or ancient nature. Even though there are many factors impacting DNA degradation through damage and fragmentation, these two characteristics correlate with time (Allentoft et al., 2012; Sawyer et al., 2012; Weiß et al., 2016). Thus, they indirectly inform us about the old

provenance of authentic samples. Additionally, characterizing DNA damage, fragmentation, and endogenous DNA-content are fundamental to planning how to proceed with degraded plant samples for sequencing. For example, in case of very low endogenous DNA content, it might be a better idea to conduct enrichment experiments before sequencing. In this section, we describe how to authenticate and characterize DNA from degraded plant samples, and how to proceed with high-throughput sequencing.

As outlined in the previous section, DNA is subjected to *post mortem* modifications. The most recognisable landmark of aDNA is its damage pattern. It is the effect of spontaneous changes of cytosines to uracils through deamination, which occurs at higher frequencies at the ends of DNA fragments (Briggs et al., 2007). Uracils are incorporated as thymines by many DNA polymerases during PCR amplification, which is why the deamination-generated changes generally are referred to as C-to-T substitutions. When compared to other substitutions, the excess of C-to-T can be observed at the 5' end of sequencing reads, taking a shape similar to a negative exponential curve (Fig. 4C). While an excess of C-to-T substitutions is also observed at the 3' end of sequencing reads derived from single-stranded DNA libraries, if more commonly used double-stranded DNA genomic libraries are analyzed, an excess of G-to-A substitutions is expected (Fig. 4C). The excess of C-to-T and/or G-to-A substitutions can be interrogated using the MapDamage2 (Jónsson et al., 2013) output obtained in Basic Protocol 3. All undisputed aDNA results exhibit these 'damage patterns' and hence we recommend to forfeit the high-throughput sequencing and analyzing of samples that do not match this criterion. In extreme cases, even more than 50% of cytosines get converted into thymines at the first 5' base, but as little as 1% excess is sufficient to authenticate herbarium material from the last century. In cases with less than 2-4% of C-to-T substitutions, we recommend to carry out high-throughput sequencing for non-UDG-treated libraries, and remove those artefacts during bioinformatic processing (for example by removing the first two bases from sequenced reads). In cases where C-to-T substitutions are more abundant, we recommend preparing new UDG-treated libraries for high-throughput sequencing. In extreme cases, we recommend to bioinformatically completely remove any C-to-T/G-to-A substitutions from downstream analyses.

Another important *post mortem* modification characteristic for aDNA is its fragmentation. One of the biochemical processes causing this fragmentation is depurination. Therefore, we frequently observe an excess of purines (G's and A's) at DNA breakpoints in authentic aDNA (Fig. 4C). These patterns can again be interrogated from the graphical output of MapDamage2 (Jónsson et al., 2013), Basic Protocol 3. In addition, to use this fragmentation as another authenticity criterion, we recommend the sequencing of screening libraries in paired-end mode. As the median aDNA fragment length is typically under 100 bp, frequently closer to 50-70bp, the resulting paired forward and reverse reads of the same fragment share substantial overlap. Their merging thus allows precise quantification of the length of the sequenced aDNA reads and improves the quality of base-pair readout. Plotting of the fragment sizes shows a distribution that is most informatively summarized by its median (Fig. 4B). Since the minimum size of bioinformatically meaningful DNA fragments (i.e. those that we can confidently map to a single position in the reference genome) is 30 bp, we recommend to forfeit high-throughput sequencing of genomic aDNA libraries with median fragment size smaller than 30bp. There are bioinformatic methods that increase the confidence in mapping small DNA fragments to plant genomes and avoid incorrect mapping of metagenomic contaminants (de Filippo et al., 2018). However, when sequencing aDNA libraries with low median fragment sizes (30-50bp) one should also consider the economic factor. As reads only use (less than) half the base-pairs sequenced in standard genomic libraries (between 75 to 150 bp), more than the double amount of sequencing is necessary to obtain the same genome coverage.

Finally, the quantification of the amount of *bona fide* endogenous DNA is very important for deciding how to proceed with an aDNA sample. Endogenous DNA is often accompanied with DNA of *pre-* and *post*

mortem colonization of microbes, and of environmental contaminants. Low endogenous content in the samples of interest is not necessarily a consequence of contamination. To test for contamination, the negative controls, extraction and library 'blanks' (see Strategic planning for details) should be interrogated. In our protocol, we suggest calculating a proxy for endogenous DNA content by dividing the total number of merged reads by the subset of merged reads that mapped to the reference genome. High amount (>5%) of unique endogenous DNA reads in negative controls is a reliable marker for cross-contamination among samples. The average endogenous DNA content in samples of herbarium specimens is about 80%, but can vary greatly; in the case of *Arabidopsis thaliana* herbarium specimens, it ranges between 15%-85% (Durvasula et al., 2017). Such high numbers, coupled with high library complexity (high proportion of unique reads), merit direct deep sequencing. Again, one should keep economic hurdles in mind: 20% endogenous DNA means that we will have to sequence five times more DNA molecules to obtain the same coverage as for standard genomic libraries with 100% endogenous DNA. For samples with low endogenous DNA content or low proportion of unique reads, we thus recommend to couple our protocol with targeted enrichment procedures, such as aDNA hybridization on microarrays (Burbano et al., 2010; Hodges et al., 2009) or 'in solution' (Fu et al., 2013; Maricic et al., 2010).

ACKNOWLEDGEMENTS:

LITERATURE CITED:

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, M. T. P., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., & Bunce, M. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B-Biological Sciences, 279(1748), 4724–4733.

Andrews, S. (n.d.). FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved August 5, 2020, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Avila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., Fordyce, S. L., Montiel, R., Vielle-Calzada, J.-P., Willerslev, E., & Gilbert, M. T. P. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Scientific Reports, 1, 74.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Pruefer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Paeaebo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America, 104(37), 14616–14621.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research, 38(6), e87.

Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J., & Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Research, 35(17), 5717–5728.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L. F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., & Pääbo, S. (2010). Targeted investigation of the Neandertal genome by array-based sequence capture. Science, 328(5979), 723–725.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution / British Ecological Society, 9(2), 410–419.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences, 110(39), 15758–15763.

de Filippo, C., Meyer, M., & Prüfer, K. (2018). Quantifying and reducing spurious alignments for the analysis of ultra-short ancient DNA sequences. BMC Biology, 16(1), 121.

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., & Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America, 114(20), 5213–5218.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. Methods in Molecular Biology , 1963, 1–13.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., & Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proceedings of the National Academy of Sciences of the United States of America, 110(6), 2223–2227.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A., & Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Research, 45(10), e79.

Gansauge, M.-T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols, 8(4), 737–748.

Gansauge, M.-T., & Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Research, 24(9), 1543–1549.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathogens, 8, 24.

Gutaker, R. M., & Burbano, H. A. (2016). Reinforcing plant evolutionary genomics using ancient DNA. Current Opinion in Plant Biology. Submitted.

Gutaker, R. M., & Burbano, H. A. (2017). Reinforcing plant evolutionary genomics using ancient DNA. Current Opinion in Plant Biology, 36, 38–45.

Gutaker, R. M., Reiter, E., Furtwängler, A., Schuenemann, V. J., & Burbano, H. A. (2017). Extraction of ultrashort DNA molecules from herbarium specimens. BioTechniques, 62(2), 76–79.

Gutaker, R. M., Weiß, C. L., Ellis, D., Anglin, N. L., Knapp, S., Luis Fernández-Alonso, J., Prat, S., & Burbano, H. A. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. Nature Ecology & Evolution, 3(7), 1093–1101.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature, 312(5991), 282–284.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., & Hannon, G. J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature Protocols, 4(6), 960–974.

Jaenicke-Després, V., Buckler, E. S., Smith, B. D., Gilbert, M. T. P., Cooper, A., Doebley, J., & Pääbo, S. (2003). Early allelic selection in maize as revealed by ancient DNA. Science, 302(5648), 1206–1208.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics , 29(13), 1682–1684.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research, 40(1), e3.

Kistler, L. (2012). Ancient DNA extraction from plants. Methods in Molecular Biology , 840, 71–79.

Kistler, L., Bieker, V. C., Martin, M. D., Pedersen, M. W., Ramos Madrigal, J., & Wales, N. (2020). Ancient Plant Genomics in Archaeology, Herbaria, and the Environment. Annual Review of Plant Biology, 71, 605–629.

Kistler, L., Maezumi, S. Y., Gregorio de Souza, J., Przelomska, N. A. S., Malaquias Costa, F., Smith, O., Loiselle, H., Ramos-Madrigal, J., Wales, N., Ribeiro, E. R., Morrison, R. R., Grimaldo, C., Prous, A. P., Arriaza, B., Gilbert, M. T. P., de Oliveira Freitas, F., & Allaby, R. G. (2018). Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. Science, 362(6420), 1309–1313.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In arXiv [q-bio.GN]. arXiv. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics , 25(16), 2078–2079.

Malmström, H., Storå, J., Dalén, L., Holmlund, G., & Götherström, A. (2005). Extensive human DNA contamination in extracts from ancient dog bones and teeth. Molecular Biology and Evolution, 22(10), 2040–2047.

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PloS One, 5(11), e14004.

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., Korol, A., David, M., Reiter, E., Riehl, S., Schreiber, M., Vohr, S. H., Green, R. E., Dawson, I. K., Russell, J., Kilian, B., Muehlbauer, G. J., Waugh, R., Fahima, T., … Stein, N. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. Nature Genetics, 48(9), 1089–1093.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010(6), db.prot5448.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science, 338(6104), 222–226.

Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. Nature Reviews. Genetics, 16(7), 395–408.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. Nature, 314(6012), 644–645.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annual Review of Genetics, 38, 645–679.

Palmer, S. A., Moore, J. D., Clapham, A. J., Rose, P., & Allaby, R. G. (2009). Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. PloS One, 4(7), e6301.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology, 17, 60.

Picard toolkit. (2019). In Broad Institute, GitHub repository. Broad Institute. http://broadinstitute.github.io/picard/

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science, 311(5759), 392–394.

Prüfer, K., & Meyer, M. (2015). Anthropology. Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans" [Review of Anthropology. Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans"]. Science, 347(6224), 835.

Ramos-Madrigal, J., Runge, A. K. W., Bouby, L., Lacombe, T., Samaniego Castruita, J. A., Adam-Blondon, A.-F., Figueiral, I., Hallavant, C., Martínez-Zapater, J. M., Schaal, C., Töpfer, R., Petersen, B., Sicheritz-Pontén, T., This, P., Bacilieri, R., Gilbert, M. T. P., & Wales, N. (2019). Palaeogenomic insights into the origins of French grapevine diversity. Nature Plants, 5(6), 595–603.

Ramos-Madrigal, J., Smith, B. D., Víctor Moreno-Mayar, J., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M. T. P., & Wales, N. (2016). Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. Current Biology: CB, 26(23), 3195–3201.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PloS One, 7(3), e34131.

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols, 9(5), 1056–1082.

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes, 9, 88.

Shapiro, B., & Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. Science, 343(6169), 1236573.

Shepherd, L., & Perrie, L. (2014). Genetic analyses of herbarium material: Is more care required? Taxon, 63(5), 972–973.

Skoglund, P., & Mathieson, I. (2018). Ancient Genomics of Modern Humans: The First Decade. Annual Review of Genomics and Human Genetics, 19, 381–404.

Smith, O., Nicholson, W. V., Kistler, L., Mace, E., Clapham, A., Rose, P., Stevens, C., Ware, R., Samavedam, S., Barker, G., Jordan, D., Fuller, D. Q., & Allaby, R. G. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. Nature Plants, 5(4), 369–379.

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., Kruse-Peeples, M., Lepak, N., Prim, L., Romay, M. C., Ross-Ibarra, J., Sanchez-Gonzalez, J. de J., Schmidt, C., Schuenemann, V. J., Krause, J., Matson, R. G., Weigel, D., Buckler, E. S., & Burbano, H. A. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science, 357(6350), 512–515.

Wales, N., Alberto Romero-Navarro, J., Cappellini, E., & Gilbert, M. T. P. (2012). Choosing the Best Plant for the Job: A Cost-Effective Assay to Prescreen Ancient Plant Remains Destined for Shotgun Sequencing. PloS One, 7(9). https://doi.org/10.1371/journal.pone.0045644

Wales, N., Andersen, K., Cappellini, E., Avila-Arcos, M. C., & Gilbert, M. T. P. (2014). Optimization of DNA Recovery and Amplification from Non-Carbonized Archaeobotanical Remains. PloS One, 9(1). https://doi.org/10.1371/journal.pone.0086827

Wales, N., Carøe, C., Sandoval-Velasco, M., Gamba, C., Barnett, R., Samaniego, J. A., Madrigal, J. R., Orlando, L., & Gilbert, M. T. P. (2015). New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. BioTechniques, 59(6), 368–371.

Wales, N., & Kistler, L. (2019). Extraction of Ancient DNA from Plant Remains. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), Ancient DNA: Methods and Protocols (pp. 45–55). Springer New York.

Weiß, C. L., Schuenemann, V. J., Devos, J., Shirsekar, G., Reiter, E., Gould, B. A., Stinchcombe, J. R., Krause, J., & Burbano, H. A. (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. Open Science, 3(6), 160239.

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the Phytophthora infestans

lineage that triggered the Irish potato famine (vol 2, e00731, 2013). eLife, 2. https://doi.org/10.7554/eLife.01108

INTERNET RESOURCES:

URLs for important sites relevant to the method. Each must be accompanied by a short description of the subject of the site.

A demo version of the bioinformatic processing protocol can be found at:
https://gitlab.com/smlatorreo/plant-adna-pipeline

FIGURE LEGENDS:

Figure 1. Schematic overview of Basic Protocols. Purple wavy lines show degraded DNA that is targeted for isolation, teal lines represent ligated adapters, while yellow lines - indices with remaining adapter sequence. Yellow ellipses represent enzymes. Blue bars represent base qualities.

Figure 2. Column assembly described in Basic Protocol 1.20. (A.) Force a DNeasy Mini spin column (DNeasy®⬜) onto the conical reservoir (Zymo research). **(B.)** Secure the junction with a piece of sterile masking tape. **(C.)** Place the conical reservoir and column inside a 50-mL conical tube.

Figure 3. Construction of oligonucleotide adapters throughout Basic Protocol 2 (A.) Blunt-ending of double-stranded aDNA (purple) targeted for sequencing. **(B.)** Ligation of double-stranded adaptors (teal). **(C.)** Filling adapters in the enzymatic reaction. **(D.)** Amplification with long single-stranded primers that each carry a unique 7 bp index (yellow) for each genomic library.

Figure 4. The output of the bioinformatic pipeline characterizing aDNA properties. A. Comparison between a set of trimmed and merged reads versus a set of non-merged reads (rows). DIfferences can be attributed to important quality measures (columns). Plots were produced by *FastQC*. **B.** Example information displayed by *samtools flagstat* ran on a SAM (BAM) formatted file. The percentage of mapped reads (red arrow) is a proxy for the endogenous DNA content. **C.** Analyses reported by *mapDamage2* show an excess of Adenine (blue) and Guanine (black) residues at the breakpoints in aDNA backbone as evidence of typical depurination process in ancient-DNA molecules. At the bottom, red and blue lines indicate the excess of C-to-T and G-to-A substitutions, respectively, at the 5' and 3' ends of aDNA sequencing reads. Plots were produced by *mapDamage2*.

TABLES:

Table 1.1. PTB precursor stock solution 1 (4X).

Reagent	Final Concentration	Vol. for 1 mL	Vol. for 50 mL
SDS (10%)	4 %	0.04 g	2 g

Tris (1M)	40 mM	40 µL	2 mL
EDTA (0.5M)	40mM	80 µL	4 mL
NaCl (5M)	20 mM	4 µL	200 µL
Water	-	Until reaching 1 mL total volume	Until reaching 50 mL total volume

Table 1.2. PTB-based Mix.

Reagent	Final Concentration	Vol. for 1 reaction (1.2 mL)
PTB precursor stock solution 1 (4X)	1X	300 µL
Proteinase K (10mg/mL)	0.4 mg/mL	48 µL
PTB (284.18g/mol)	2.5 mM	0.8526 mg
DTT (154.2g/mol)	50 mM	9.252 mg
Water	-	852 µL
Total	-	1.2 mL

Table 1.3. Binding Buffer.

Reagent	Final Concentration	Vol. for 1 RX (10 mL)

Guanidine hydrochloride (100%)	5M		4.776 g
Water	-		Up to 6 mL
Isopropanol (100%)	40 %		4 mL

Table 2.1. Blunt-end repair reaction.

Reagent (initial concentration)	Final Concentration	Vol. per sample
Water	-	11.6 µL
BSA (10 mg/mL)	0.8 mg/mL	4 µL
NEB buffer 2.1 (10X)	1X	5 µL
dNTP (2.5mM)	100 µM	2 µL
ATP (10 mM)	1 mM	5 µL
T4 Polynucleotide Kinase (10,000 U/mL)	400 U/mL	2 µL
T4 DNA Polymerase (3,000 U/mL)	24 U/mL	0.4 µL
DNA	-	20 µL
TOTAL:		50 µL

Table 2.2. Adapter Ligation reaction.

Reagent (initial concentration)	Final Concentration	Vol. per sample
Quick Ligase Buffer (2X)	1X	20 μL
Adapter Mix (10 μM)	250 nM	1 μL
DNA [Blunt-end repaired]	-	18 μL
Quick Ligase (5 U/mL)	0.125 U/mL	1 μL
TOTAL:		40 μL

Table 2.3. Fill-in reaction.

Reagent (initial concentration)	Final Concentration	Vol. per sample
Water		12 μL
Isothermal Amplification Buffer (10 X)	1 X	4 μL
dNTP (2.5 mM)	125 nM	2 μL
Bst 2.0 DNA Polymerase (8 U/mL)	0.4 U/mL	2 μL
DNA [With adapters]	-	20 μL
TOTAL:		40 μL

Table 2.4. Indexing PCR reaction.

Reagent (initial concentration)	Final Concentration	Vol. per Sample
DNA	-	37 µL
Water	-	293 µL
Pfu Buffer (10X)	1X	40 µL
dNTP (25 mM)	250 µM	4 µL
BSA (10 mg/mL)	100 µg/mL	6 µL
Pfu Polymerase (2.5 U/µL)	0.25 U/µL	4 µL
P7 index primer (10 µM)	200 nM	8 µL
P5 index primer (10 µM)	200 nM	8 µL
TOTAL:		400 µL

Table A2.1. UDG blunt-end repair reaction.

Reagent (initial concentration)	Final Concentration*	Vol. per sample
Water	-	8.5 µL
BSA (10 mg/mL)	0.1 mg/mL	0.5 µL
NEB buffer 2.1 (10X)	1X	5 µL

Current Protocols   Page 36 of 40

ATP (10 mM)	1 mM	5 µL
dNTP (2.5 mM)	300 µM	6 µL
T4 Polynucleotide Kinase (10000 U/mL)	400 U/mL	2 µL
USER enzyme (1 U/mL)	0.06 U/mL	3 µL
DNA	-	20 µL
TOTAL:		50 µL

John Wiley & Sons

Basic Protocol 1:
DNA isolation

Sampling

Grinding

Lysis

Shredding

Purification

Elution

Basic Protocol 2:
Genomic libraries

Blunting

Washing

Ligation

Washing

Adapter fill-in

Indexing

Washing

Basic Protocol 3:
Bioinformatic processing

Sequencing

ATGC

Merging

./adapter removal

Validation

./FastQC

Mapping

./bwa

Authentication

./map Damage2

T
C   G

A Blunt-ending

<aDNA> NNN

NNN <aDNA>

B Ligation

ACACTCTTTCCCTACACGACGCTCTTCCGATCT <aDNA> AGATCGGAAGAGC

CGAGAAGGCTAGA <aDNA> TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

C Adapter fill-in

ACACTCTTTCCCTACACGACGCTCTTCCGATCT <aDNA> AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA <aDNA> TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

D Amplification with indexing primers

P7

TGTGCAGACTTGAGGTCAGTG<INDEX>TAGAGCATACGGCAGAAGACGAAC

ACACTCTTTCCCTACACGACGCTCTTCCGATCT <aDNA> AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA <aDNA> TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

AATGATACGGCGACCACCGAGATCTACAC<INDEX>ACACTCTTTCCCTACACGACGCTCTT

P5