Extraction of ultrashort DNA molecules from herbarium specimens

Rafal M. Gutaker¹, Ella Reiter², Anja Furtwängler², Verena J. Schuenemann^{2,3}, and Hernán A. Burbano¹

¹Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany, ²Institute of Archaeological Sciences, University of Tübingen, Tübingen, Germany, and ³Senckenberg Center for Human Evolution and Paleoenvironment, University of Tübingen, Tübingen, Germany

BioTechniques 62:76–79 (February 2017) doi 10.2144/000114517

Keywords: ancient DNA; herbarium; DNA extraction; paleogenomics; next-generation sequencing

Supplementary material for this article is available at www.BioTechniques.com/article/114517.

DNA extracted from herbarium specimens is highly fragmented; therefore, it is crucial to use extraction protocols that retrieve short DNA molecules. Improvements in extraction and DNA library preparation protocols for animal remains have allowed efficient retrieval of molecules shorter than 50 bp. Here, we applied these improvements to DNA extraction protocols for herbarium specimens and evaluated extraction performance by shotgun sequencing, which allows an accurate estimation of the distribution of DNA fragment lengths. Extraction with N-phenacylthiazolium bromide (PTB) buffer decreased median fragment length by 35% when compared with cetyl-trimethyl ammonium bromide (CTAB); modifying the binding conditions of DNA to silica allowed for an additional decrease of 10%. We did not observe a further decrease in length for single-stranded DNA (ssDNA) versus double-stranded DNA (dsDNA) library preparation methods. Our protocol enables the retrieval of ultrashort molecules from herbarium specimens, which will help to unlock the genetic information stored in herbaria.

Since ancient DNA (aDNA) is highly fragmented, it is particularly important to employ extraction protocols that retrieve ultrashort DNA fragments (<50 bp). A recently developed extraction protocol for animal remains efficiently recovers those shorter molecules (1), which has allowed sequencing of highly fragmented hominin (2) and cave bear remains (1) that

are hundreds of thousands of years old. DNA retrieved from herbarium specimens is also highly fragmented. An analysis of 86 herbarium specimens spanning the last 300 years showed that DNA in those samples decayed 6 times faster than in a spatially constrained bone assemblage (3). Consequently, DNA from century-old herbarium specimens is as short as that

of animal remains thousands of years old. To take full advantage of the genetic information stored in those samples it is important to optimize the extraction of ultrashort molecules from desiccated plant tissues.

We assessed the impact of extraction and library preparation methods on the distribution of DNA fragment lengths in 20 *Arabidopsis thaliana* herbarium specimens, which were collected between 1839 and 1898 (Supplementary Table S1). We used a hierarchical experimental design that includes two different phases due to the limited quantity of tissue in each sample (Figure 1). In Phases 1 and 2a, we used 10 *A. thaliana* samples (~20 mg of leaf tissue each), which were subjected to 2 different extraction protocols (~10 mg of tissue per treatment), followed by double-stranded DNA (dsDNA) library preparation. To compare the performance of dsDNA and single-stranded DNA (ssDNA) library preparation methods, in Phase 2b, we applied a single-stranded library preparation method to DNA extracts produced by the most efficient DNA extraction protocol in Phase 2a. In each phase, we evaluated the performance of the different methods by sequencing the genomic libraries on the Illumina MiSeq platform (Illumina, San Diego, CA) (Supplementary Table S2).

Extraction buffers used for ancient bones and teeth are commonly composed predominantly or exclusively of EDTA and proteinase K (4), reagents that are not optimal for DNA extraction from plant tissue. Hence, in the first phase we tested two commonly used DNA extraction buffers for historical plant specimens, which contain either cetyl-trimethyl ammonium bromide (CTAB), or a mixture of N-phenacylthiazolium bromide (PTB) and dithiothreitol (DTT) (5) (Figure 1). CTAB is a strong detergent that under high salt concentrations binds to polysaccharides and aids their removal from the solution (6). Although CTAB is frequently used in DNA extractions from modern plants, it does not have a detectable effect when applied to non-carbonized archaeobio-

METHOD SUMMARY

We optimized the extraction procedure for isolating ultrashort DNA fragments from herbarium specimens through the use of N-phenacylthiazolium bromide (PTB) lysis buffer along with modifications previously used for DNA extraction from ancient bones.

tanical remains (7). PTB cleaves glucose-derived protein crosslinks (8), and it can help to release DNA trapped within sugar-derived condensation products (9); it has been effectively used to retrieve DNA from archaeobotanical remains (10). DTT reduces disulfide bonds, releasing thiolated DNA from crosslinked complexes (11). For better comparison between the CTAB and PTB protocols, we replaced the ethanol precipitation step of the CTAB method with DNA binding to a silica column (12) provided with the DNeasy Plant Mini Kit. Subsequently, libraries were prepared using a dsDNA library protocol (13).

Based on qPCR measurements on unamplified libraries, the PTB protocol recovered a higher number of unique library molecules than the CTAB protocol (paired *t*-test, $P = 0.007$) (Figure 2F and Supplementary Table S3). We found that PTB decreased the median fragment length by 35% (from 88 bp to 57 bp) (paired *t*-test, $P = 2.8 \times 10^{-6}$) when compared with CTAB (Figure 2, A and B). This decrease in length was also manifested as a higher proportion of damaged sites (lambda) (paired *t*-test, $P = 1.3 \times 10^{-6}$) (Figure 2C), which represented the fraction of bonds broken in the DNA backbone (14,15), and was used here as a summary statistic for fragment length distribution. DNA molecules extracted with PTB buffer showed more cytosine (C) to thymine (T) substitutions at the 5' end (paired *t*-test, $P = 1.2 \times 10^{-6}$) (Figure 2G). C-to-T substitutions are typical damage patterns seen in aDNA and result from spontaneous deamination of C to uracil (U), which directs the polymerases used during various steps of library preparation to incorporate adenine (A). As a consequence, the regenerated strand serves as a template for T instead of C (16,17). It is possible that shorter and more damaged fragments of DNA were released after crosslinks were resolved by PTB and DTT since there was a strong negative correlation between median fragment length and C-to-T substitutions at the first base ($R^2 = 0.44$; $P = 1.5 \times 10^{-7}$; $n = 50$) (Supplementary Figure S4). Alternatively, the observed variation in fragment length distribution could be explained by unknown chemical incompatibilities of the lysis and binding buffer (i.e., certain reagents could, in principle, reduce the DNA binding to the silica membrane).

Figure 1. Experimental design for testing the effects of different DNA extraction and library preparation protocols on the properties of sequenced DNA libraries from herbarium specimens. Experiments were conducted in two phases. In Phase 1, we subjected 10 herbarium specimens of *Arabidopsis thaliana* to DNA extraction with 2 different lysis buffers and compared the sequencing results. In Phase 2, we tested 2 DNA-binding methods on a second set of 10 *A. thaliana* specimens (Phase 2a) and then compared the properties of single- and double-stranded libraries constructed with the extracted DNA (Phase 2b).

Figure 2. Effects of different DNA extraction and library preparation protocols on properties of DNA sequencing libraries. Figure shows results from the experimental phases depicted in Figure 1. (A) Density of fragment lengths for merged reads mapped to the *Arabidopsis thaliana* reference genome. The y-axis shows kernel density estimates. (B–G) Distributions are represented as box and whisker plots; medians are depicted by thick black lines; boxes represent data between quartiles Q1 and Q3; whiskers extend to 1.5x the interquartile range between Q1 and Q3; and points represent outliers. Comparisons within experiments resulting in significant differences in a paired *t*-test are connected with black lines (** P -value < 0.01 ; *** P -value < 0.001) (B) Fragment length medians. (C) Proportion of broken DNA fragments (lambda). (D) Proportion of GC content. (E) Proportion of endogenous DNA (proportion of reads mapped to *A. thaliana* reference genome). (F) Number of unique molecules per base of the *A. thaliana* reference genome (molecule coverage of DNA extract) calculated from qPCR measurements of unamplified libraries. (G) Percentage of cytosine to thymine substitutions at the first base on the 5' end.

Finally, in the CTAB protocol we applied a chloroform-isoamyl alcohol wash, which could also have reduced recovery of short molecules.

In Phase 2a, to further increase the recovery of short fragments we used PTB/DTT, which was the more successful extraction buffer in Phase 1, and evaluated two systems for binding DNA to silica. We tested DNeasy Mini Spin Columns (Qiagen, Venlo, The Netherlands), in combination with the binding buffer used in the Plant Mini Kit, and MinElute silica spin columns (Qiagen), in combination with a binding buffer optimized for the recovery of short molecules from animal remains (1) (Figure 1). We found that the latter method decreased the median fragment length by 10% (from 60 bp to 54 bp) (paired *t*-test, $P = 1.9 \times 10^{-4}$), which shows that it also is suitable for recovery of very short sequences from herbarium specimens (Figure 2, A and B). The frequency of C-to-T substitutions at the first base differed significantly between the 2 DNA binding methods (paired

t-test, $P = 3.3 \times 10^{-3}$) (Figure 2G), with a decrease in median fragment length again accompanied by an increase in C-to-T substitutions.

To investigate whether library preparation has an effect on fragment length distribution, in Phase 2b we produced ssDNA libraries (18,19) using the extracts from the modified PTB/DTT extraction (Phase 2a) and compared the ssDNA libraries to the dsDNA libraries previously constructed in Phase 2a (Figure 1). We did not observe a significant decrease in the median fragment length distribution in ssDNA libraries (paired *t*-test, $P = 0.44$) (Figure 2, A and B). Instead, the shape of the distribution changed toward larger numbers of longer and shorter molecules at the cost of intermediate-size molecules, which was reflected in a decreased lambda (Figure 2C) and agrees with previous findings (19). Similarly to Gansauge and Meyer (19), we also detected a reduction of GC content in ssDNA libraries when compared with dsDNA (Figure 2D). This

phenomenon can be attributed to a known bias in dsDNA libraries toward molecules with higher GC content (20,21). We detected uniform GC content across the distribution of fragment lengths, which suggests that the ssDNA library preparation protocol excels in reducing those biases (Supplementary Figure S8). In contrast to previous reports (19), the ssDNA library method did not produce an increase in the proportion of endogenous DNA (Figure 2E, Supplementary Figures S1 and S6). However, it has been suggested that the increase in the proportion of endogenous DNA occurs only when the initial content of endogenous DNA is <10% (22,23). Our *A. thaliana* samples had endogenous DNA contents of 16%–94%, which could explain why we did not detect a gain in endogenous DNA.

Here, we demonstrate that the choice of extraction buffer has a great impact on the length distribution of molecules recovered from herbarium specimens. Ultrashort molecules were most efficiently retrieved using a combination PTB/DTT buffer mixture for DNA extraction and the buffers and conditions suggested by Dabney et al. (1) for DNA binding. We recommend this method for extraction of aDNA from herbarium and archaeobotanical specimens because it preserves most of the information stored in both short and long DNA molecules. In situations where the number of recovered unique library molecules significantly exceeds the intended sequencing coverage, it is possible to optimize sequencing through size selection. The two library preparation methods tested here appear to be equally efficient in retaining short DNA fragments; however, while the single-stranded method reduces GC bias in libraries, it also decreases the fraction of endogenous DNA. Our DNA extraction protocol increases the recovery of short fragments and, thus, the accessibility of precious herbarium specimens for genetic analyses.

Author contributions

R.M.G., V.J.S., and H.A.B designed the experiments. R.M.G., E.R., and A.F. performed the experiments. R.M.G analyzed the data. R.M.G and H.A.B wrote the manuscript with contributions from all authors.

DETECT CRISPR GENE EDITING EVENTS Using Agarose Gels

Streamline gene editing workflows with a simple and accurate enzymatic assay.

AccuCleave™ T7 Kit

- T7 based enzyme kit for mutation frequency detection of CRISPR events.

AccuCleave™ C2 Control DNA Kit

- Contains 2 Control DNA fragments - intact and -2 deletion fragment.

Start now at
www.aati-us.com/AccuCleave

Acknowledgments

We thank curators from the Missouri Botanical Garden, University of Illinois, National Museum of Natural History, West Virginia University, Harvard University, and New York Botanical Gardens for kindly providing samples for this study; Marco Thines, Charles B. Fenster, and Matthew T. Rutter for sampling the herbarium specimens; Daniel Koenig for advice on the experimental design; Patricia Lang, members of the Research Group for Ancient Genomics and Evolutions, and especially Matthias Meyer for comments on the manuscript. This work was funded by the Presidential Innovation Fund of the Max Planck Society.

Competing interests

The authors declare no competing interests

References

- Dabney, J., M. Knapp, I. Glocke, M.T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* 110:15758-15763.
- Meyer, M., J.L. Arsuaga, C. de Filippo, S. Nagel, A. Aximu-Petri, B. Nickel, I. Martinez, A. Gracia, et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* 531:504-507.
- Weiß, C.L., V.J. Schuenemann, J. Devos, G. Shirsekar, E. Reiter, B.A. Gould, J.R. Stinchcombe, J. Krause, and H.A. Burbano. 2016. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *R. Soc. Open Sci.* 3:160239.
- Rohland, N. and M. Hofreiter. 2007. Comparison and optimization of ancient DNA extraction. *BioTechniques* 42:343-352.
- Kistler, L. 2012. Ancient DNA extraction from plants. *Methods Mol. Biol.* 840:71-79.
- Rogers, S.O. and A.J. Bendich. 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol. Biol.* 5:69-76.
- Wales, N., K. Andersen, E. Cappellini, M.C. Avila-Arcos, and M.T. Gilbert. 2014. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS One* 9:e86827.
- Vasan, S., X. Zhang, X. Zhang, A. Kapurniotu, J. Bernhagen, S. Teichberg, J. Basgen, D. Wagle, et al. 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature* 382:275-278.
- Poinar, H.N., M. Hofreiter, W.G. Spaulding, P.S. Martin, B.A. Stankiewicz, H. Bland, R.P. Evershed, G. Possnert, and S. Paabo. 1998. Molecular coproscopy: dung and diet of the extinct ground sloth *Nototheriops shastensis*. *Science* 281:402-406.
- Jaenicke-Després, V., E.S. Buckler, B.D. Smith, M.T. Gilbert, A. Cooper, J. Doebley, and S. Paabo. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* 302:1206-1208.
- Gill, P., A.J. Jeffreys, and D.J. Werrett. 1985. Forensic application of DNA 'fingerprints'. *Nature* 318:577-579.
- Palmer, S.A., J.D. Moore, A.J. Clapham, P. Rose, and R.G. Allaby. 2009. Archaeogenomic evidence of ancient Nubian Barley evolution from six to two-row indicates local adaptation. *PLoS One* 4:e6301.
- Meyer, M. and M. Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010:prot5448.
- Deagle, B.E., J.P. Eveson, and S.N. Jarman. 2006. Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Front. Zool.* 3:11.
- Allentoft, M.E., M. Collins, D. Harker, J. Haile, C.L. Oskam, M.L. Hale, P.F. Campos, J.A. Samaniego, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* 279:4724-4733.
- Briggs, A.W., U. Stenzel, P.L. Johnson, R.E. Green, J. Kelso, K. Prufer, M. Meyer, J. Krause, et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* 104:14616-14621.
- Brotherton, P., P. Endicott, J.J. Sanchez, M. Beaumont, R. Barnett, J. Austin, and A. Cooper. 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 Res.* 35:5717-5728.
- Meyer, M., M. Kircher, M.T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222-226.
- Gansauge, M.T. and M. Meyer. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8:737-748.
- Green, R.E., J. Krause, A.W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710-722.
- Briggs, A.W., J.M. Good, R.E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, et al. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318-321.
- Bennett, E.A., D. Massilani, G. Lizzo, J. Daligault, E.M. Geigl, and T. Grange. 2014. Library construction for ancient genomics: single strand or double strand? *BioTechniques* 56:289-300.
- Wales, N., C. Caroe, M. Sandoval-Velasco, C. Gamba, R. Barnett, J.A. Samaniego, J.R. Madrigal, L. Orlando, and M.T. Gilbert. 2015. New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. *BioTechniques* 59:368-371.

Received 20 September 2016; accepted 23 December 2016.

Address correspondence to Hernán A. Burbano, Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology. E-mail: hernan.burbano@tuebingen.mpg.de

To purchase reprints of this article, contact: biotechniques@fosterprinting.com

Premium labware for research and discovery

Conical Tubes

- Ethanol / methanol-resistant print
- Reliable, secure sealing
- Many style and packaging options

Serological Pipettes

- Individual paper / plastic or bulk packaging
- 1 ml, 2 ml, 5 ml, 10 ml, 25 ml, and 50 ml volumes
- Universal fit

www.sarstedt.com

customerservice@sarstedt.us