Received: 1 April 2022 | Revised: 21 April 2022 | Accepted: 10 May 2022

DOI: 10.1111/1755-0998.13631

MOLECULAR ECOLOGY RESOURCES WILEY

PERSPECTIVE

Recovering the genomes hidden in museum wet collections

Irune Ruiz-Gartzia[1] | Esther Lizano[1,2] | Tomas Marques-Bonet[1,2,3,4] | Joanna L. Kelley[5]

[1]Institute of Evolutionary Biology (UPF-CSIC), Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

[2]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

[3]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain

[4]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[5]School of Biological Sciences, Washington State University, Pullman, Washington, USA

Correspondence
Joanna L. Kelley, School of Biological Sciences, Washington State University, Pullman, WA, USA.
Email: joanna.l.kelley@wsu.edu

Funding information
María de Maeztu, Grant/Award Number: Mobility_Fellowship

Handling Editor: Shawn R Narum

Abstract

Natural history museums hold vast collections of biomaterials. The collections in museums, often painstakingly sampled, are largely unexplored treasures that may help us better understand biodiversity on the planet. Museum collections can provide a unique window into the past of species long gone or currently declining due to human activity. From a molecular perspective, however, many museum samples are stored under conditions that hasten the damage of DNA, RNA and proteins. For example, samples in wet collections are those stored in liquid preservatives, typically ethanol. These ethanol-preserved tissues are often, although not always, formalin-fixed prior to storage, which may damage DNA. In this and recent issues of *Molecular Ecology Resources*, Straube et al (2021), O'Connell et al (2021) and Hahn et al (2022) explore different types of specimens from museum wet collections as new sources of DNA for scientific studies. All three articles found that for wet museum collections, overall specimen condition mattered most for recovering high-quality genomic DNA.

Museum samples are challenging to work with due to the heterogeneity of sample types and the wide variety of conditions under which they are stored. Samples also often lack detailed preservation information, such as initial storage conditions (temperature and light conditions), fixatives used and post-mortem intervals. DNA preservation is highly variable among specimen types, preservation methods and storage conditions (Pääbo, 1989). In recent decades, DNA has been successfully retrieved from dried soft tissues, dry eggshells, bones and teeth (Grealy et al., 2019; Raxworthy & Smith, 2021). Recent *Molecular Ecology Resources* articles by Straube et al. (2021), O'Connell et al. (2022), and Hahn et al. (2022) focus on wet collection samples (also known as spirit-preserved specimens), which are rarely used for molecular studies, despite the fact that they represent a large portion of museum collections (Hahn et al., 2022). One of the major concerns when working with wet collections is the use of formalin to fix and store samples. This fixative causes DNA damage in numerous ways: intra- and intermolecular crosslinking, disruption of base-pairing, promotion of denaturation, and methylol adducts which inhibit DNA amplification (Do & Dobrovic, 2015). The lack of information about fixation and preservation of museum samples is common, which makes it difficult to predict possible DNA damage.

Advances in DNA sequencing approaches have facilitated the acquisition of genomic data from museum specimens. Improvements in high-throughput sequencing platforms have gone hand in hand with

continuous developments in DNA extraction (Dabney et al., 2013) and library preparation (Carøe et al., 2018). Specifically, given the highly fragmented nature and low quantities of DNA usually recovered from museum specimens, particular caution must be taken to maximize the recovery of short DNA fragments. In their articles, Straube et al. (2021) and Hahn et al. (2022) test different DNA extraction techniques to select the most appropriate method for formalin-fixed and ethanol-stored samples. Hot alkaline lysis, which was originally used for formalin-fixed paraffin-embedded clinical biopsy samples, was reported to better retrieve DNA when compared with proteinase K digestion, and in particular when working with low-quality samples (Hahn et al., 2022).

DNA-sequencing library preparation methods are also known to affect sequencing outcomes. Methods developed specifically for degraded DNA such as the blunt-end single-tube (BEST) protocol (Carøe et al., 2018), produce libraries that have more unique DNA molecules (higher complexity). By performing the protocol in a single tube, the number of molecules lost during purification steps is greatly reduced. In contrast to double-stranded protocols, recently developed single-stranded approaches maximize the chances of recovering DNA preserved as either single- or double-stranded DNA because each DNA strand present in an extraction is ligated with sequencing adaptors (Kapp et al., 2021). While the single-stranded method proposed in Straube et al. (2021) performed well, there were no other library methods tested. When testing single-stranded DNA and the BEST method, Hahn et al. (2022) found that the library preparation method did not substantially affect sequencing quality. It is possible that DNA from wet collections is not as degraded as ancient DNA. Therefore, the library preparation method might not have a significant role in sequencing quality if the protocol is optimized for degraded DNA.

Museum specimens are often manipulated and stored under conditions that increase the chance of contamination. For this reason, one of the major challenges when working with museum specimens is contamination. Contamination levels are often highly variable among samples, as was shown in Straube et al. (2021) where contamination values varied from 2.9% to 65.7%. When working with historical samples, DNA damage varies notably depending on the preservation and storage conditions, which makes it difficult to differentiate endogenous from contaminant DNA present in the sample (Straube et al., 2021). This is in direct contrast to ancient DNA, which shows specific damage patterns that help differentiate it from contamination (Skoglund et al., 2014).

The recovery of endogenous mitochondrial and nuclear genomes from formalin-fixed samples is a major leap forward in accessing museum samples for genetic studies. Complete mitochondrial genomes were recovered from wet collections in Straube et al. (2021) and Hahn et al. (2022). Moreover, nuclear whole genomes were also sequenced for some samples in both articles. O'Connell et al. (2022) sequenced nuclear DNA using restriction-site associated DNA (RAD) capture. Nevertheless, some biases were present in the dataset due to the capture method, indicating that it is important to consider how biases may influence interpretation before selecting library preparation and sequencing methodology.

Despite the methodological improvements and technical advances, not all samples are favourable for DNA retrieval. Moreover, extracting DNA from samples requires destruction of a portion or all of a sample, depending on size. This means that for extremely rare

TABLE 1 Decision table for assessing possible sequencing options. Modified from figure 5 in Hahn et al. (2022)

Specimen details			Inferred specimen quality	Sequencing options[a]
Well-preserved?	Viscera present?	Preservation media		
Yes	Yes	Ethanol	Good	All sequencing options available
Yes	Yes	Formalin, formaldehyde concentration < 10,000 mg per L, media pH > 6	Moderate	Amplicon, capture-based and whole mitochondrial sequencing Whole genome sequencing with variation in the likelihood of success
Yes	No	Ethanol	Moderate	Amplicon, capture-based and whole mitochondrial sequencing Whole genome sequencing with variation in the likelihood of success
Yes	Yes	Formalin, formaldehyde concentration > 10,000 mg per L, media pH < 6	Poor	Only amplicon sequencing recommended Capture-based and whole mitochondrial sequencing with variation in the likelihood of success
Yes	No	Formalin	Poor	Only amplicon sequencing recommended Capture-based and whole mitochondrial sequencing with variation in the likelihood of success
Decomposed	Inconsequential	Inconsequential	Very poor	Successful sequencing unlikely

[a]The performance of moderate- and poor-quality specimens can be highly variable. This table and the figure in Hahn et al. (2022) are guides, not guarantees.

specimens and sparsely sampled species, it is necessary to consider whether DNA extraction is feasible and the best use of a sample. The prescreening of museum specimens to optimize the probability of success, especially when DNA extraction methods are destructive of the sample, is highly recommended (Table 1).

Museum collections contain various types of samples, including many from which DNA can be extracted, as has been done with wet collections. O'Connell et al. (2022) highlight extracting DNA from allozyme supernatant for genomic applications. Using wet collections and allozyme supernatants allows the study of unique specimens, increasing the number of individuals of some populations, and in some cases, samples from populations that no longer exist or are difficult to access.

Future efforts should focus on optimizing methodologies to extract DNA, as has been done in Straube et al. (2021) and Hahn et al. (2022). Although several new approaches have been developed to increase the final amount and complexity of DNA for genomic sequencing libraries, molecular crosslinking, DNA damage and PCR inhibitors still pose problems when dealing with DNA from historical samples. Applying new methodologies to reverse crosslinking and repair DNA damage will improve the quality of the extracted DNA. One of the problems is that many samples fail during library amplification due to the presence of DNA inhibitors after extraction, and therefore it is necessary to eliminate or neutralize inhibitors prior to DNA amplification. Finally, as these samples contain small amounts of highly fragmented DNA, the purification steps are critical when preparing sequencing libraries. For this reason, improving the efficiency of DNA purification s or implementing revised library preparation protocols that have only one purification step should increase the amount of DNA.

While choosing the best samples optimizes DNA extraction success (Table 1), the amount of contamination is highly variable (as seen in Straube et al., 2021). A better understanding of the sources of contamination and the development and implementation of contamination assessments will improve the prescreening of museum specimens. For now, the availability of genome references coupled with appropriate computational tools to remove contamination is helping to detect and remove contamination in sequencing reads during data analysis. While reference genomes are becoming increasingly available, the unique samples in museum collections may require additional reference genomes, especially from extinct and endangered species. These three articles demonstrate the suitability of museum wet collections to extracting and sequencing DNA hidden within them. We are at the tip of the iceberg in terms of accessing museum samples for the genomic revolution.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Tomas Marques-Bonet https://orcid.org/0000-0002-5597-3075
Joanna L. Kelley https://orcid.org/0000-0002-7731-605X

REFERENCES

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*, *9*, 410–419.

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

Do, H., & Dobrovic, A. (2015). Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. *Clinical Chemistry*, *61*, 64–71.

Grealy, A., Bunce, M., & Holleley, C. E. (2019). Avian mitochondrial genomes retrieved from museum eggshell. *Molecular Ecology Resources*, *19*, 1052–1062.

Hahn, E. E., Alexander, M. R., Grealy, A., Stiller, J., Gardiner, D. M., & Holleley, C. E. (2022). Unlocking inaccessible historical genomes preserved in formalin. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13505

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity*, *112*, 241–249.

O'Connell, K. A., Mulder, K. P., Wynn, A., de Queiroz, K., & Bell, R. C. (2022). Genomic library preparation and hybridization capture of formalin-fixed tissues and allozyme supernatant for population genomics and considerations for combining capture- and RADseq-based single nucleotide polymorphism data sets. *Molecular Ecology Resources*, *22*, 487–502.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *86*, 1939–1943.

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*, 1049–1060.

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 2229–2234.

Straube, N., Lyra, M. L., Paijmans, J. L. A., Preick, M., Basler, N., Penner, J., Rödel, M.-O., Westbury, M. V., Haddad, C. F. B., Barlow, A., & Hofreiter, M. (2021). Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens. *Molecular Ecology Resources*, *21*, 2299–2315.

How to cite this article: Ruiz-Gartzia, I., Lizano, E., Marques-Bonet, T., & Kelley, J. L. (2022). Recovering the genomes hidden in museum wet collections. *Molecular Ecology Resources*, *00*, 1–3. https://doi.org/10.1111/1755-0998.13631