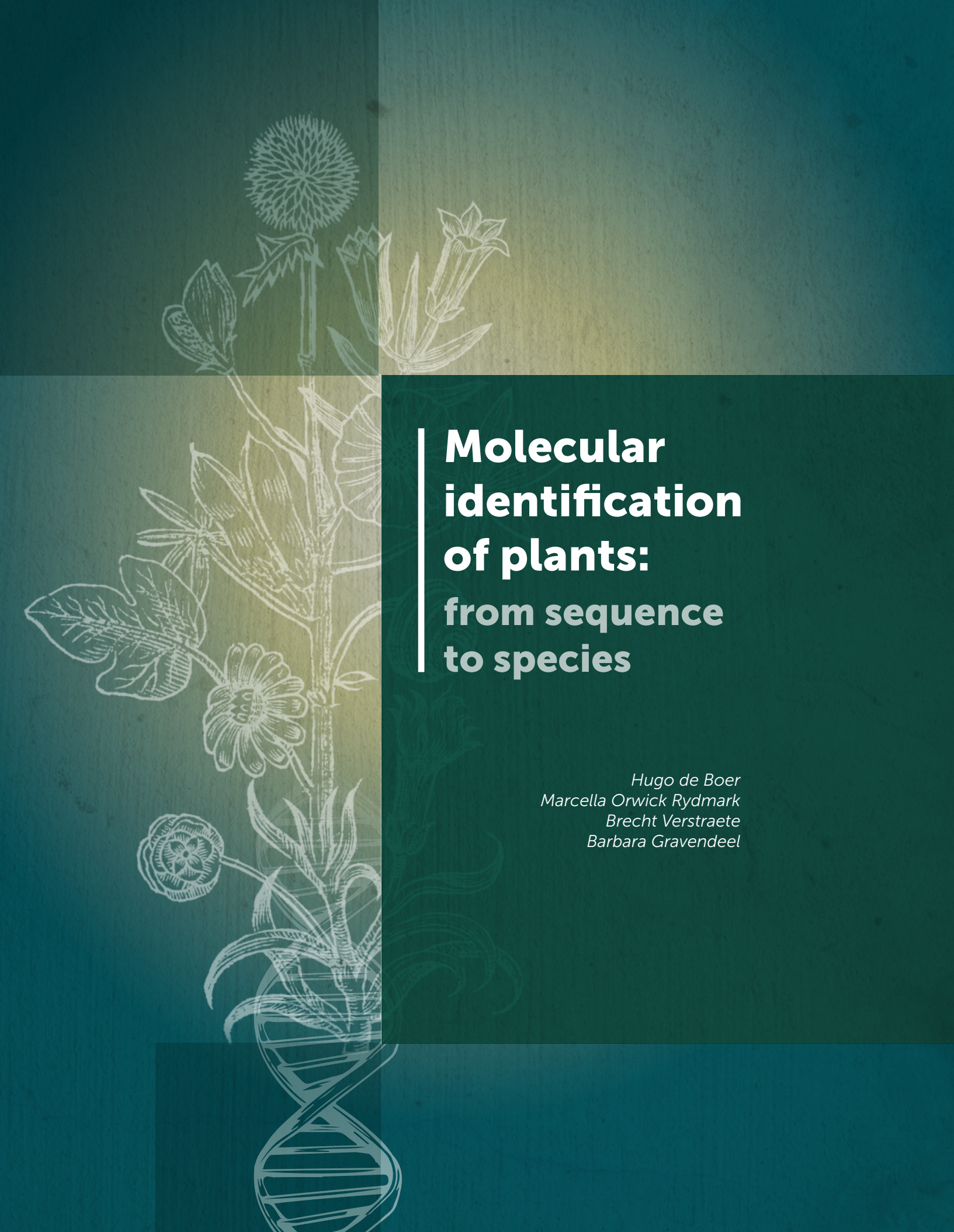

Molecular identification of plants:
from sequence to species

Hugo de Boer
Marcella Orwick Rydmark
Brecht Verstraete
Barbara Gravendeel

Table of contents

Chapter 2
DNA from museum collections

Nataly Allasi Canales[1,2], Andrew C. Clarke[3], Mark Nesbitt[2], Rafal Gutaker[2]

1 Natural History Museum of Denmark, University of Copenhagen, Denmark
2 Royal Botanic Gardens, Kew, United Kingdom
3 Future Food Beacon of Excellence & School of Biosciences, University of Nottingham, United Kingdom

Nataly O. Allasi Canales **allasicanales@gmail.com**
Andrew C. Clarke **andrew.clarke1@nottingham.ac.uk**
Mark Nesbitt **m.nesbitt@kew.org**
Rafal Gutaker **r.gutaker@kew.org**

Citation: Canales NA, Clarke AC, Nesbitt M, Gutaker R (2022) Chapter 2. DNA from museum collections. In: de Boer H, Rydmark MO, Verstraete B, Gravendeel B (Eds) Molecular identification of plants: from sequence to species. Advanced Books. https://doi.org/10.3897/ab.e68634

Introduction

Museum collections of plant origin include herbaria (pressed plants), xylaria (woods), and economic botany (useful plant) specimens. They are not only places of history and display, but also of research, and contain rich repositories of molecules, including DNA. Such DNA, retrieved from historical or ancient tissue, carries unique degradation characteristics and regardless of its age is known as ancient DNA (aDNA). Research into aDNA has developed rapidly in the last decade as a result of an improved understanding of its biochemical properties, the development of specific laboratory protocols for its isolation, and better bioinformatic tools. Why are museum collections useful sources of aDNA? We identify three main reasons: 1) specimens can play a key role in taxonomic and macroevolutionary inference when it is difficult to sample living material, for example, by giving us snapshots of extinct taxa (Van de Paer et al. 2016); 2) accurate identification of specimens that were objects of debate or scientific mystery, as exemplified by misidentified type specimens of the watermelon's progenitor (Chomicki and Renner 2015); 3) specimens can provide us with 'time machines' to study microevolutionary processes and diversity changes over decades- to millennia-long timeframes (Gutaker and Burbano 2017; Pont et al. 2019). In all three cases, specimens are often associated with evidence of their occurrence in space and time. For further examples see Chapter 20 Museomics, and the Glossary.

However, extracting DNA does mean the destruction of a part of the specimen. Museum curators therefore face challenges in balancing the conservation of specimens for future research with the rising demand for aDNA analysis. Increasingly, curators are also considering legal and ethical issues in sampling (Austin et al. 2019; Pálsdóttir et al. 2019). Close collaboration between the aDNA researcher and the curatorial staff of museums is therefore essential for appropriate management of these issues (Freedman et al. 2018).

Ethical and legal aspects

With few exceptions, plant material found in museums originally grew on lands tended or owned by people for many millennia (Ellis et al. 2021). Some specimens, such as artefacts or seeds of domesticated crops have an even more direct connection to human activities. Plant specimens, along with other living things, are therefore not simply assemblages of chemical compounds such as DNA, but also embody spiritual beliefs, diverse forms of ownership, traditional knowledge, and past histories of colonialism and other forms of harm (Anderson et al. 2011; Das and Lowe 2018; Pungetti et al. 2012). The implications of this are still being worked out in dialogues between museums and affected communities, often within a decolonising framework (McAlvay et al. 2021). There are, however, immediate steps that researchers and curators can take to ensure that the use of specimens is both legal and ethical.

A first consideration is whether the plant species or artefacts (such as baskets or wooden objects) are of special significance (e.g., sacred) to the source community. Examples of sacred material include *Banisteriopsis caapi*, used to make ayahuasca in South America (Rivier and Lindgren 1972), or *Duboisia hopwoodii* (pituri), used as tobacco in Australia (Ratsch et al. 2010). An online literature search or consultation with relevant experts will give a rapid pointer, which can be followed up with source communities in the study region. Collaboration with communities and scientists in source countries is essential for acknowledging the rights to plant material (even if not legally enshrined), and can be furthered by publication of results in local languages

and media. These communities also hold significant expertise on plants that will improve the quality and relevance of research (Gewin 2021).

There are international conventions that usually apply when accessing, researching, and moving plant material between institutions and countries. Researchers must also be aware of country-specific laws that may require further permits and inspections, e.g., for plants that produce controlled substances, require phytosanitary checks, or are considered invasive species. Legal elements of the Convention on Biological Diversity (CBD), Nagoya Protocol, and Convention on Trade in Endangered Species (CITES) are covered in Chapter 27 Legislation and policy as well as in other published works (e.g. McManis and Pelletier 2014, Iob and Botigué 2021). While the CBD applies to specimens received by museums from 1992, in ethical terms (and under some implementations of the Nagoya Protocol) its principles, such as benefit-sharing, also apply to pre-1992 specimens (cf. Sherman and Henry 2020).

Sampling museum collections

Locating collections and specimens

Botanical gardens hold living specimens and distribute seeds of these via seed lists (Index Seminum). Their global collections can be searched via PlantSearch, hosted by Botanic Gardens Conservation International. Gene banks hold seeds, and sometimes also tissue and living plants. While they originally focused on crop plants and their wild relatives, many have now broadened in scope to include wild plants, such as Royal Botanic Gardens Kew's Millennium Seed Bank. Many gene bank collections can be searched via Genesys. Herbaria hold dried plant specimens and can be located via Index Herbariorum. Although many herbaria are incompletely recorded in databases, substantial data can already be found in the Global Biodiversity Information Facility (GBIF) (Bieker and Martin 2018). Plants are present in abundance in almost all forms of human activity, and it is therefore not surprising that plant material can also be found outside the confines of herbaria, including in economic botany or ethnobotany collections (Salick et al. 2014), agricultural museums, and anthropology collections. Increasing awareness of the importance of biological collections, their uses, conservation efforts and crosslinks among them, is leading to important initiatives that integrate all digitised natural science collections from natural history museums, universities, and botanic gardens (Bakker et al. 2020).

There are a number of pitfalls when searching online catalogues. It may be necessary to search for accepted names and common synonyms: the same species may appear under different botanical names in a single collection, and accuracy of specimen identification varies. In general, herbarium specimens are the most reliable, as they bear diagnostic criteria such as flowers on which taxonomists rely. Garden material and seeds are often misidentified, or become confused in labelling, or are hybridised during repeated cultivations. Their identifications should be confirmed, for example growing on the seeds or by using morphological criteria (Nesbitt et al. 2003). Additionally, data may be missing, unspecific, or incorrectly transcribed or presented, in derived databases, for example in the case of georeferencing (Maldonado et al. 2015).

Researcher-curator collaboration

Research projects will benefit enormously from a close collaboration between researcher and curator. Museums should be approached early during a project, with the researcher providing

sufficient detail about its background, aims, methodology, and timetable. Museums are often under-staffed and persistence may be required in making contact. Curators' expertise will be crucial in identifying the most appropriate specimens for analysis, not only in their institutions, but in others with which they are familiar. The curator will also play a key role in assessing the provenance of specimens, using museum archives, and the implications for any of the ethical and legal issues addressed above. Curators often have good links to source communities and can advise on appropriate procedures.

After preliminary discussions, the researcher will usually need to fill in a 'destructive sampling' form. This acts as a permanent record of the justification for sampling, and allows the museum to make a detailed check on the aims and methodology of the project (see for example, British Museum form and policies). Requests that have unclear research aims or which employ inappropriate methodologies are unlikely to be approved. Researchers will likely need to sign a Material Transfer Agreement (MTA) or Material Supply Agreement (MSA) with the museum which sets out their legal responsibilities.

Sampling may be carried out by the researcher or the curator. If feasible, it is worthwhile for the researcher to carry out the sampling, as it allows for the investigation of the context of the specimen and for flexibility in choosing the samples. It may also speed up the process of obtaining samples, especially if a large number is required. It also allows samples to be safely hand-carried to the researcher's laboratory. Where materials must be sent, it is safest to use a courier service, with specimens marked "Scientific specimens of no commercial value".

It should be agreed with the museum whether, after sampling, surplus material should be returned or securely retained. Museums can require that they are informed about results and that they check manuscripts before publication. This is in any case good practice to ensure accurate reporting of sample details. Museum policies on co-authorship vary, and this topic should be discussed early. Significant contribution by the curator on the choice of appropriate samples, provenance research, or in technically complex sampling, merits co-authorship. Unless agreed otherwise, DNA sequencing data should be submitted to NCBI GenBank or other public repositories, taking care to give the correct specimen identifier. At a minimum, the museum's unique catalogue number (if one exists), and the name of the museum should be cited. This allows the DNA sequence data to be linked directly with the specimen or object. Other museum and laboratory information may be included with the DNA sequence data or in publications (e.g., the collector name, collection number, dates, locations, and laboratory extraction numbers). Additionally, most museum collections will require that vouchers are annotated in a way that links them to DNA sequencing data (see below). Some museums have also started to permanently store DNA isolates, and we encourage researchers to share their stocks on request. Integrated data management and accessibility of the raw data and results will ultimately bolster curatorial practices, develop a more ethical science, and safeguard collections for future generations (Schindel and Cook 2018). Useful guidance on documentation issues is available from the Global Genome Biodiversity Network (GGBN).

Choice of specimens and sampling

Sampling decisions will be determined both by the research design and the nature of the specimens, in addition to the legal and ethical factors mentioned above. Changes to agreed sampling lists are often necessary once specimens have been examined, for example when they are lost, in poor condition, inadequately annotated or georeferenced, present in small quantities, or of rare taxa. Bulk raw material is usually easy to sample, while objects are usually not subjected to destructive sampling unless the results will inform the history and significance of the

object. For herbarium specimens, preserving the morphological features, especially those that are diagnostic, for future research, is critical. Sampling should be targeted towards tissue types or organs at a given developmental state that are most numerous. For example, if there are many flowers and few leaves, it may be preferable to sample a petal. Or if there are few cauline and many rosette leaves, it may be preferable to sample a rosette leaf.

Different parts of a specimen may yield varying amounts, quality, and types of DNA. Wood, husks, and other tissues that were undergoing senescence at the time of preservation may yield less DNA. Young, immature leaves will have higher cell densities, and therefore are expected to yield more DNA. Seeds are often excellent sources of nuclear DNA, although the genotype of the seed will differ from the parent plant and might be of inconsistent ploidy. It may be necessary to extract DNA from individual seeds or to remove maternal tissue such as the testa. Some herbarium sheets will contain multiple individuals and, in most cases, it is better to sample individuals rather than mixed material. If individuals are pooled for DNA extraction, it may complicate downstream analyses that depend on individual genotypes.

The method of specimen preservation is another consideration for DNA isolation. Desiccation has been shown to preserve plant DNA remarkably well, while charring or ethanol preservation destroys plant DNA almost completely (Forrest et al. 2019; Nistelberger et al. 2016). Although not commonly used for aDNA analysis, ancient waterlogged (saturated with water) specimens have a potential for high endogenous contents as they are usually preserved in cold temperatures (Wagner et al. 2018; Wales et al. 2014).

Before sampling begins, the specimen's identifying data, such as its herbarium ID, should be recorded with great care, and double-checked on both the sample label and typed list of specimens. Additionally, the museum may require that vouchers are annotated with the sampling date, tissue type, sample identifier, and information about the researchers. The voucher, including any labels, should be photographed, ideally before and after sampling. Digital links between herbarium vouchers, imaging, and DNA sequences are very useful; they can be included in herbarium and nucleotide databases.

For desiccated leaves, the most commonly sampled tissue, the process is usually straightforward. Using forceps and a scalpel or scissors one can make a precise cut and remove 1 cm^2 or less of tissue. Generally, between 2 and 10 mg of dry leaf tissue is sufficient for the isolation of complex mixtures of genomic DNA fragments. It is preferable that leaves of lesser value are targeted, for example damaged, folded, or hidden, avoiding possible contamination by mould, lichen, or fungi. The sampling of detached "pocket" material should be conducted with caution, and only if the researcher and curator are confident that the detached material truly belongs to the voucher. For other tissue types, such as wood, researchers may need to develop tailored sampling methods on contemporary material first. After sampling, material should immediately be sealed in a labelled tube or envelope and packaged for transport.

Surface contamination

Potential contamination of the sample, specimen, or wider collection with exogenous DNA is an important consideration. For most museum collections, there will inevitably already be surface DNA contamination of specimens. Ask the curator about adhesives (e.g., wheat starch) and preservatives that were used with the specimen of interest. Curatorial staff and other users of the collections may not routinely wear gloves or, if they do, may not change them between specimens. In most cases, there is unlikely to be any benefit from the person undertaking sampling wearing protective equipment (e.g., face masks, hair nets) that is beyond that normally used by users of the collection. Contamination control is only as good as the weakest link.

Extra precautions may be taken for equipment that is used directly in the sampling process, for example, disposable scalpels that are changed between samples, or wiping of scalpel blades with bleach and ethanol. This will reduce the risk of cross-contamination between specimens. Further precautions may be beneficial if internal tissue is being sampled (e.g., inside a seed). In these cases, surface decontamination (see section below on pre-processing) followed by sampling with DNA-free equipment and while wearing personal protective equipment may be appropriate. In some cases where specialistic equipment such as microdrill is required, it may be beneficial for sampling to be undertaken within an ancient DNA laboratory, where contamination controls can be better implemented, however bringing large amounts of plant material into the laboratory should be limited as it is an additional contamination source.

Contamination of specimens and collections by 'modern' DNA and especially amplified DNA is perhaps the greatest risk, potentially compromising future research. Researchers are likely to have been using molecular laboratories, and steps should be taken to prevent the inadvertent transfer of modern DNA to museum collections. These precautions can include not visiting a collection directly from a modern laboratory, cleaning items that must move between modern laboratories and collections (e.g., clothes, phones, cameras), and using sampling equipment (scalpels, tubes, pens) that has not been taken from a modern laboratory.

Laboratory work with historical samples

Understanding aDNA traits

Before starting any experiments with historical and ancient plant samples, it is important to recognize challenges arising from the degraded nature of aDNA. Unlike DNA isolated from fresh samples, DNA from preserved specimens is fragmented, damaged, and contaminated post mortem (Gutaker and Burbano 2017), that includes even recently collected herbarium specimens (Weiß et al. 2016) and contamination with exogenous DNA (Bieker et al. 2020). Fragmentation describes the accumulation of breaks in the DNA backbone, leading to shorter DNA molecules. Breaks occur more often next to guanine or adenine bases, and this can be visualised in sequencing data with dedicated software (Jónsson et al. 2013). The median expected fragment length for aDNA from herbarium specimens is between 30-90 base pairs (bp) in unheated recent Arabidopsis extractions (Bakker 2019; Weiß et al. 2016). It is important to recognise that fragments shorter than 35 bp might generate spurious alignments due to microbial mismapping (Prüfer et al. 2010). The short length of aDNA fragments calls for special molecular methods that allow the retention of short molecules, as well as conservative bioinformatic settings during data processing.

aDNA is also affected by "damage", post mortem substitutions that convert cytosine to uracil residues through deamination (uracils are read by insensitive DNA polymerases as thymine, hence the commonly used term "C-to-T substitutions'') (Hofreiter et al. 2001). This process occurs preferentially at the ends of DNA molecules (Briggs et al. 2007), particularly with single-stranded DNA overhangs (Overballe-Petersen et al. 2012). Consequently, in the population of sequenced molecules, an elevated number of C-to-T substitutions are observed at the 5′ end, and complementary G-to-A substitutions at the 3′ end. Typically, herbarium-isolated DNA has around from 1 to 6% (in older samples) of cytosine residues converted to thymine (Durvasula et al. 2017; Gutaker et al. 2019; Weiß et al. 2016), while in archaeological material this number might be as high as 30%. These post mortem substitutions should be removed before downstream analyses.

Finally, it is important to recognize that aDNA from plants is in fact a mixture of bona fide endogenous DNA, exogenous DNA introduced pre mortem, (e.g., from endophytic microbes),

and exogenous DNA introduced post mortem (e.g., from microbes involved in decomposition, human-associated collection and museum practices; see above) (Pääbo et al. 2004). Quantification of contamination is commonly done by dividing the number of sequence reads that map to the target reference genome by the total number of sequenced reads from the museum sample. In fresh material, the ratio is often around 0.98; in degraded material it can vary from 0 to 0.95 (Gutaker et al. 2017). Several examples of aDNA successfully obtained from plants are illustrated in Table 1.

Table 1. Examples of selected successfully isolated and sequenced DNA from plant material. *BP: before present.

Species	Tissue	Age BP*	Endogenous DNA	Fragment length (bp)	Damage at 5′ end	Source
Thale cress (*Arabidopsis thaliana*)	Leaf	184	83%	~62	0.026	Durvasula et al. 2017
Potato (*Solanum tuberosum*)	Leaf	361	87%	~45	0.047	Gutaker et al. 2019
Maize (*Zea mays*)	Cobs	1863	80%	~52	0.052	Swarts et al. 2017
Wheat (*Triticum durum*)	Chaff	3150	40%	~53	0.095	Scott et al. 2019
Barley (*Hordeum vulgare*)	Seeds	4988	86%	~49	0.138	Mascher et al. 2016

Recommended working practices for aDNA

Given the characteristics of aDNA (Dabney et al. 2013) and the fact that it is very prone to contamination at any stage, guidelines have been proposed to facilitate the authentication process, and minimise potential contamination before, during, and after DNA extraction (Pääbo et al. 2004). We strongly recommend following gold-standard precautions when working with aDNA (Fulton and Shapiro 2019; Latorre et al. 2020).

The isolation and pre-amplification manipulation of aDNA should be carried out in a dedicated laboratory that is physically separated from labs where post-amplification steps are carried out. Ideally the aDNA laboratory should be supplied with HEPA-filtered air under positive pressure. Users should not move from a 'modern' laboratory (where amplified DNA is handled) to the aDNA laboratory on the same day. Reagents and materials in an aDNA lab should be DNA-free, disposable where possible, and never taken out of the clean lab. Surfaces should be cleaned before and after every experiment with 3–10% bleach, 70% ethanol, and overnight UV-C irradiation. To minimise contamination and ensure a DNA-free laboratory environment, users should wear full body suits, foot protectors, slippers, facemasks, sleeves, and double gloves (Fulton and Shapiro 2019). Together, these precautions limit cross-contamination from amplified and unamplified DNA.

Material preparation is an essential step before DNA can be isolated. Optional pre-processing of dirty samples can be done by gently cleaning the surface with a very low concentration (~3%) of bleach, and rinsing twice with ddH_2O (Cappellini et al. 2010). When handling waterlogged, fragile, or permeable material, avoid using bleach and carry out ddH_2O treatment only. To help identify contamination that might be introduced in the laboratory, samples should always be processed alongside negative controls, including for DNA isolation and library preparations. To reduce the likelihood of cross-contamination, small batches of up to 12 samples at a time are preferable (Latorre et al. 2020).

DNA extraction methods for different tissues should be considered. While plant materials tend to contain inhibitory substances like polyphenols, proteins, and polysaccharides, ancient

plant materials can additionally be rich in humic acids and salts. This set of macromolecules might prevent successful DNA amplification (Wales et al. 2014) by affecting polymerase activity (Schrader et al. 2012). To reduce this inhibitory effect, smaller amounts of sample can be extracted in parallel, and the resulting DNA pooled to achieve a sufficient yield (Wagner et al. 2018).

Here we will cover the basics of recovering the highest quality of DNA from ancient plant tissues. Using a two-day extraction protocol will greatly increase the recovery of endogenous DNA. The first day consists of grinding the plant material. Tissue can be disrupted by: grinding dry, grinding flash-frozen, or grinding material soaked in lysis buffer. In all cases, grinding to finer particles increases the recovery of aDNA. Ground tissue is incubated in a fresh lysis buffer. Three commonly used buffers include CTAB (Kistler 2012), DTT (Wales and Kistler 2019), or PTB mixtures (Latorre et al. 2020). The second day is dedicated to isolating DNA from the lysate. Initial removal of non-DNA particles can be achieved by centrifugation with a shredding column (Latorre et al. 2020) or phenol/chloroform mixture (Kistler 2012; Wales and Kistler 2019; Wagner et al. 2018). In all methods, DNA is then captured in various DNA-binding silica columns (for example QIAgen MinElute columns) and purified (Dabney et al. 2013). Elution from silica columns produces the final, isolated aDNA.

By contrast to primed amplification approaches, even low amounts of isolated DNA can be used for genomic library preparation (Staats et al. 2013) and hence we recommend that a genomic library is constructed using a well-established method (Carøe et al. 2017; Kircher et al. 2012; Meyer and Kircher 2010; Meyer et al. 2012). Quantification of genomic DNA before sequencing using RT-qPCR allows the number of amplification cycles for each sample to be adjusted, in turn allowing the complexity of sequenced DNA fragments to be maximised. Bioinformatic pre-processing is an essential part of aDNA analyses, and is summarised in three available pipelines (Latorre et al. 2020; Peltzer et al. 2016; Schubert et al. 2014). Authentication is another crucial step in bioinformatic analyses that can currently be best achieved with mapDamage software (Jónsson et al. 2013).

Choosing and authenticating aDNA samples

To help decide which sampled material is most promising for further DNA analyses it is necessary to obtain good estimates for fragmentation, damage, and contamination. This can be achieved through sequencing genomic libraries in low-throughput mode (about 10,000 DNA reads per sample), commonly referred to as "screening" and bioinformatic analyses that produce relevant summary statistics. Promising samples will contain aDNA with a median fragment length over 50 bp and endogenous content over 0.2. For samples of particular interest, mapping the accuracy for short aDNA reads can be improved with specialised procedures (de Filippo et al. 2018), and endogenous content can be increased by targeted enrichment on hybridization arrays (Hodges et al. 2009) or 'in solution' (Maricic et al. 2010). Finally, one should pay attention to the frequency of C-to-T substitutions at the ends of the sequenced reads. Samples with 2–6% C-to-Ts can be corrected bioinformatically (by trimming ends or filtering transitions), while a higher percentage of C-to-Ts can be remedied through more effective enzymatic removal of uracil (Briggs et al. 2010).

Characterising DNA fragmentation and damage is very useful for authentication and establishing historical provenance of degraded plant samples. DNA degradation advances with time (Weiß et al. 2016), although its rate is highly modulated by intrinsic and environmental factors. Old samples should be considered authentic only if they exhibit fragmentation and damage patterns congruent with their age, tissue type, and storage conditions. In contrast to library-based approaches, primer-based sequencing (such as Sanger sequencing) does not allow quantification of these characteristics and should not be used with aDNA (Gutaker and Burbano 2017).

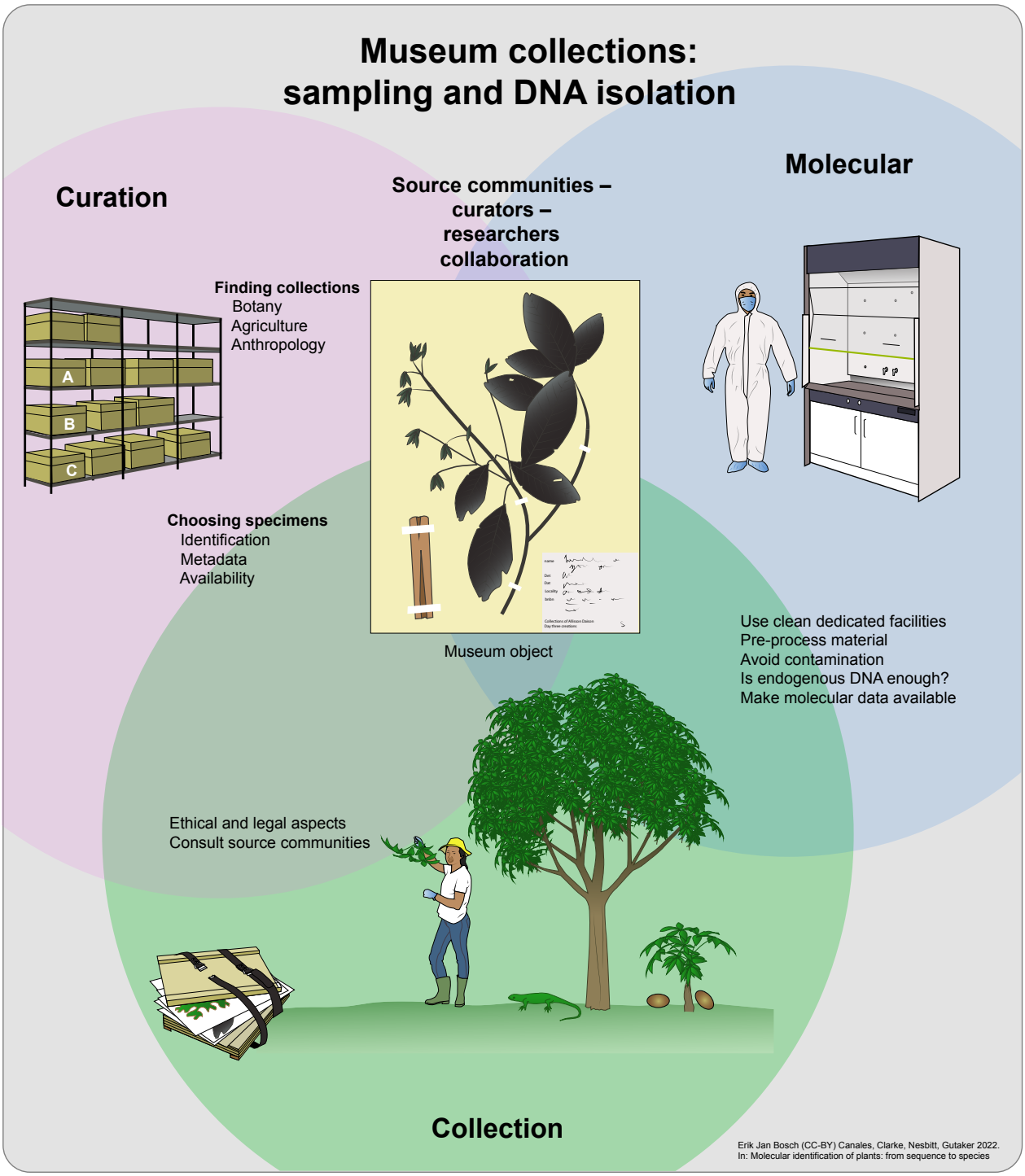

Figure 1. Chapter 2 Infographic: Overview of sampling and obtaining DNA from museum collections. An team effort of communities, curators and researchers (1) Collection of botanical material should have detailed consideration of its ethical and legal aspects and the consultation of source communities in advance, in accordance with CITES, CBD and Nagoya legal and ethical frameworks. (2) Curated botanical samples can be found in different types of museums that include botanic gardens, ethnobotany and anthropological collections. The next step is to find relevant specimens with preferably rich metadata, e.g. species identification, collection place and date. (3) Once the specimens have been identified, they should undergo molecular analyses in clean facilities. Where they will be pre-processed according to their traits, avoiding contamination with other samples, "modern" specimens, and amplicons. Then, it is crucial to identify samples that failed and passed quality controls for endogenous DNA. Finally, the data produced should be linked to their respective vouchers and made available in public repositories like NCBI and BOLD.

Responsible lab use for aDNA

Library-based methods assist with the responsible use of collections, as they preserve the total (non-selective) DNA and 'immortalise' it for future use. Immortalisation only has value if the DNA that has been amplified is truly historical/ancient and devoid of contemporary contamination and hence all the aforementioned precautions are necessary when working with aDNA. We recommend that extracts or library builds are precisely annotated with the methods used and are properly archived.

Questions

1. Name three legal considerations and their related ethical main issues that should be taken into account for aDNA research using museum material.
2. Why is it important to process herbarium samples in a dedicated clean lab?
3. Name three benefits of getting curators involved in the early stages of research using collections.

Glossary

aDNA – Ancient DNA, DNA that exhibits biochemical characteristics typical for DNA from old degraded material, i.e., damage and fragmentation, regardless of age.
Artefact – An object made by humans that is of historical or cultural importance, examples include: clothing, ornaments, utensils.
Authentication – Bioinformatic analyses that quantify damage and fragmentation of sequenced DNA to help rule out that DNA is derived from contemporary contamination.
Collection – Repository of curated biological material arranged in a systematic fashion.
Contamination – Introduction of alien tissue or DNA to a specimen or DNA isolate, examples include: microbial colonisation, human epithelium, plant-based foods, etc.
Curator – Custodian of a collection with expert knowledge about specimens, their organisation, and preservation.
Destructive sampling – Permanent removal of a fragment of a specimen of any size that will be irretrievable after biochemical characterization.
DNA damage – Typically conversion of cytosine to uracil in DNA through deamination, which accumulates with time. During sequencing, uracil is replaced with thymine, hence the common synonym, C-to-T substitutions.
Endogenous DNA – Authentic DNA from targeted individuals of a species, in contrast to exogenous DNA from associated microbes and contemporary plant and human DNA contamination.
Fragmentation – Breaks in the DNA backbone, most frequently caused by depurination, leading to shorter DNA fragments with time.
Immortalization – Molecular manipulation of DNA, for example the attachment of DNA adapters, that allows infinite re-amplification of the original DNA from a biological specimen.
Type specimen – Preserved individual plant that has defining features of that taxon that is used for the first taxonomic description of a species. This permanent feature-specimen link is recognized in a publication.

Voucher – Preserved botanical specimen kept in permanent collection and cited by research project. Vouchers will have been expertly identified and are usually annotated with collection time, place, and collector details.

References

Anderson EN, Pearsall D, Hunn E, Turner N (2011) Ethnobiology. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9781118015872

Austin RM, Sholts SB, Williams L, Kistler L, Hofman CA (2019) Opinion: To curate the molecular past, museums need a carefully considered set of best practices. Proc Natl Acad Sci USA 116, 1471–1474. https://doi.org/10.1073/pnas.1822038116

Bakker FT, Antonelli A, Clarke JA, Cook JA, Edwards SV, Ericson PGP, Faurby S, Ferrand N, Gelang M, Gillespie RG, Irestedt M, Lundin K, Larsson E, Matos-Maraví P, Müller J, von Proschwitz T, Roderick GK, Schliep A, Wahlberg N, Wiedenhoeft J, Källersjö M (2020) The Global Museum: natural history collections and the future of evolutionary science and public education. PeerJ 8, e8225. https://doi.org/10.7717/peerj.8225

Bakker FT (2019) Herbarium genomics: plant archival DNA explored, in: Lindqvist, C., Rajora, O.P. (Eds.) Paleogenomics: Genome-Scale Analysis of Ancient DNA, Population Genomics. Springer International Publishing, Cham, pp. 205–224. https://doi.org/10.1007/13836_2018_40

Bieker VC, Martin MD (2018) Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. Botany Letters 165, 1–10. https://doi.org/10.1080/23818107.2018.1458651

Bieker VC, Sánchez Barreiro F, Rasmussen JA, Brunier M, Wales N, Martin MD (2020) Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. Mol. Ecol. Resour. 20, 1206–1219. https://doi.org/10.1111/1755-0998.13174

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S (2007) Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci USA 104, 14616–14621. https://doi.org/10.1073/pnas.0704665104

Briggs AW, 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 Res. 38, e87. https://doi.org/10.1093/nar/gkp1163

Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, Ashton PD, Ashford DA, Arthur P, Campos PF, Kool J, Willerslev E, Collins MJ (2010) A multidisciplinary study of archaeological grape seeds. Naturwissenschaften 97, 205–217. https://doi.org/10.1007/s00114-009-0629-3

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP (2017) Single-tube library preparation for degraded DNA. Methods Ecol. Evol. 9, 410–419. https://doi.org/10.1111/2041-210X.12871

Chomicki G, Renner SS (2015) Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics. New Phytol. 205, 526–532. https://doi.org/10.1111/nph.13163

Dabney J, Meyer M, Pääbo S (2013) Ancient DNA damage. Cold Spring Harb. Perspect. Biol. 5, a012567. https://doi.org/10.1101/cshperspect.a012567

Das S, Lowe M (2018) Nature read in black and white: decolonial approaches to interpreting natural history collections. Journal of Natural Science Collections 6, 1–14.

de Filippo C, Meyer M, Prüfer K (2018) Quantifying and reducing spurious alignments for the analysis of ultra-short ancient DNA sequences. BMC Biol. 16, 121. https://doi.org/10.1186/s12915-018-0581-9

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C, Hancock AM (2017) African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 114, 5213–5218. https://doi.org/10.1073/pnas.1616736114

Ellis EC, Gauthier N, Klein Goldewijk K, Bliege Bird R, Boivin N, Díaz S, Fuller DQ, Gill JL, Kaplan JO, Kingston N, Locke H, McMichael CNH, Ranco D, Rick TC, Shaw MR, Stephens L, Svenning J-C, Watson JEM (2021) People have

shaped most of terrestrial nature for at least 12,000 years. Proc Natl Acad Sci USA 118, e2023483118. https://doi.org/10.1073/pnas.2023483118

Forrest LL, Hart ML, Hughes M, Wilson HP, Chung K-F, Tseng Y-H, Kidner CA (2019) The limits of Hyb-Seq for herbarium specimens: impact of preservation techniques. Front. Ecol. Evol. 7, 439. https://doi.org/10.3389/fevo.2019.00439

Freedman J, van Dorp L, Brace S (2018) Destructive sampling natural science collections: An overview for museum professionals and researchers. Journal of Natural Science Collections.

Fulton TL, Shapiro B (2019) Setting up an ancient DNA laboratory. Methods Mol. Biol. 1963, 1–13. https://doi.org/10.1007/978-1-4939-9176-1_1

Gewin V (2021) How to include Indigenous researchers and their knowledge. Nature 589, 315–317. https://doi.org/10.1038/d41586-021-00022-1

Gutaker RM, Burbano HA (2017) Reinforcing plant evolutionary genomics using ancient DNA. Curr. Opin. Plant Biol. 36, 38–45. https://doi.org/10.1016/j.pbi.2017.01.002

Gutaker RM, Reiter E, Furtwängler A, Schuenemann VJ, Burbano HA (2017) Extraction of ultrashort DNA molecules from herbarium specimens. BioTechniques 62, 76–79. https://doi.org/10.2144/000114517

Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Luis Fernández-Alonso J, Prat S, Burbano HA (2019) The origins and adaptation of European potatoes reconstructed from historical genomes. Nat. Ecol. Evol. 3, 1093–1101. https://doi.org/10.1038/s41559-019-0921-3

Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, Richard McCombie W, Hannon GJ (2009) Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat. Protoc. 4, 960–974. https://doi.org/10.1038/nprot.2009.68

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. Nat. Rev. Genet. 2, 353–359. https://doi.org/10.1038/35072071

Iob A, Botigué L (2021) Crop archaeogenomics: a powerful resource in need of a well-defined regulation framework. Plants, People, Planet 4, 44–50. https://doi.org/10.1002/ppp3.10233

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40, e3. https://doi.org/10.1093/nar/gkr771

Kistler L (2012) Ancient DNA extraction from plants. Methods Mol. Biol. 840, 71–79. https://doi.org/10.1007/978-1-61779-516-9_10

Latorre SM, Lang PLM, Burbano HA, Gutaker RM (2020) Isolation, library preparation, and bioinformatic analysis of historical and ancient plant DNA. Curr. Protoc. Plant Biol. 5, e20121. https://doi.org/10.1002/cppb.20121

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A (2015) Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? Glob. Ecol. Biogeogr. 24, 973–984. https://doi.org/10.1111/geb.12326

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE 5, e14004. https://doi.org/10.1371/journal.pone.0014004

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, Schreiber M, Vohr SH, Green RE, Dawson IK, Russell J, Kilian B, Muehlbauer GJ, Waugh R, Fahima T, Krause J, Stein N (2016) Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. Nat. Genet. 48, 1089–1093. https://doi.org/10.1038/ng.3611

McAlvay AC, Armstrong CG, Baker J, Elk LB, Bosco S, Hanazaki N, Joseph L, Martínez-Cruz TE, Nesbitt M, Palmer MA, Priprá de Almeida WC, Anderson J, Asfaw Z, Borokini IT, Cano-Contreras EJ, Hoyte S, Hudson M, Ladio AH, Odonne G, Peter S, Rashford J, Wall J, Wolverton S, Vandebroek I (2021) Ethnobiology phase VI: decolonizing institutions, projects, and scholarship. J. Ethnobiol. 41, 170–191. https://doi.org/10.2993/0278-0771-41.2.170

McManis CR, Pelletier JS (2014) Legal aspects of biocultural collections, in: Salick, J., Konchar, K., Nesbitt, M. (Eds.) Presented at the Curating biocultural collections: a handbook, Royal Botanic Gardens, Kew, Kew, pp. 229–243.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Pääbo S (2012) A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222–226. https://doi.org/10.1126/science.1224344

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

Nesbitt M, Colledge S, Murray MA (2003) Organisation and management of seed reference collections. Environmental Archaeology 8, 77–84. https://doi.org/10.1179/env.2003.8.1.77

Nistelberger HM, Smith O, Wales N, Star B, Boessenkool S (2016) The efficacy of high-throughput sequencing and target enrichment on charred archaeobotanical remains. Sci. Rep. 6, 37347. https://doi.org/10.1038/srep37347

Overballe-Petersen S, Orlando L, Willerslev E (2012) Next-generation sequencing offers new insights into DNA degradation. Trends Biotechnol. 30, 364–368. https://doi.org/10.1016/j.tibtech.2012.03.007

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. Annu. Rev. Genet. 38, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Pálsdóttir AH, Bläuer A, Rannamäe E, Boessenkool S, Hallsson JH (2019) Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. R. Soc. Open Sci. 6, 191059. https://doi.org/10.1098/rsos.191059

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K (2016) EAGER: efficient ancient genome reconstruction. Genome Biol. 17, 60. https://doi.org/10.1186/s13059-016-0918-z

Pont C, Wagner S, Kremer A, Orlando L, Plomion C, Salse J (2019) Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biol. 20, 29. https://doi.org/10.1186/s13059-019-1627-1

Prüfer K, Stenzel U, Hofreiter M, Pääbo S, Kelso J, Green RE (2010) Computational challenges in the analysis of ancient DNA. Genome Biol. 11, R47. https://doi.org/10.1186/gb-2010-11-5-r47

Pungetti G, Oviedo G, Hooke D (2012) Sacred species and sites: advances in biocultural conservation. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139030717

Ratsch A, Steadman KJ, Bogossian F (2010) The pituri story: a review of the historical literature surrounding traditional Australian Aboriginal use of nicotine in Central Australia. J. Ethnobiol. Ethnomed. 6, 26. https://doi.org/10.1186/1746-4269-6-26

Rivier L, Lindgren J-E (1972) "Ayahuasca," the South American hallucinogenic drink: an ethnobotanical and chemical investigation. Econ. Bot. 26, 101–129. https://doi.org/10.1007/BF02860772

Salick J, Konchar K, Nesbitt M (2014) Curating Biocultural Collections: A Handbook. Royal Botanic Gardens, Kew, Kew.

Schindel DE, Cook JA (2018) The next generation of natural history collections. PLoS Biol. 16, e2006125. https://doi.org/10.1371/journal.pbio.2006125

Schrader C, Schielke A, Ellerbroek L, Johne R (2012) PCR inhibitors - occurrence, properties and removal. J. Appl. Microbiol. 113, 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, Martin MD, 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. Nat. Protoc. 9, 1056–1082. https://doi.org/10.1038/nprot.2014.063

Scott MF, Botigué LR, Brace S, Stevens CJ, Mullin VE, Stevenson A, Thomas MG, Fuller DQ, Mott R (2019) A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. Nat. Plants 5, 1120–1128. https://doi.org/10.1038/s41477-019-0534-5

Sherman B, Henry RJ (2020) The Nagoya Protocol and historical collections of plants. Nat. Plants 6, 430–432. https://doi.org/10.1038/s41477-020-0657-8

Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE 8, e69189. https://doi.org/10.1371/journal.pone.0069189

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Romay MC, Ross-Ibarra J, Sanchez-Gonzalez J de J, Schmidt C, Schuenemann VJ, Krause J, Matson RG, Weigel D, Buckler ES, Burbano HA (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357, 512–515. https://doi.org/10.1126/science.aam9425

Van de Paer C, Hong-Wa C, Jeziorski C, Besnard G (2016) Mitogenomics of Hesperelaea, an extinct genus of Oleaceae. Gene 594, 197–202. https://doi.org/10.1016/j.gene.2016.09.007

Wagner S, Lagane F, Seguin-Orlando A, Schubert M, Leroy T, Guichoux E, Chancerel E, Bech-Hebelstrup I, Bernard V, Billard C, Billaud Y, Bolliger M, Croutsch C, Čufar K, Eynaud F, Heussner KU, Köninger J, Langenegger F, Leroy F, Lima C, Orlando L (2018) High-Throughput DNA sequencing of ancient wood. Mol. Ecol. 27, 1138–1154. https://doi.org/10.1111/mec.14514

Wales N, Andersen K, Cappellini E, Avila-Arcos MC, Gilbert MTP (2014) Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. PLoS ONE 9, e86827. https://doi.org/10.1371/journal.pone.0086827

Wales N, Kistler L (2019) Extraction of ancient DNA from plant remains. Methods Mol. Biol. 1963, 45–55. https://doi.org/10.1007/978-1-4939-9176-1_6

Weiß CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J, Burbano HA (2016) Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. R. Soc. Open Sci. 3, 160239. https://doi.org/10.1098/rsos.160239

Answers

1. Legal: CITES (restriction in international trade of endangered species), Nagoya Protocol (ownership and other significance to indigenous peoples), and Drug Act (controlled substances).
2. The decay of DNA from historical plant material makes it very susceptible to contamination with exogenous modern DNA.
3. Curators can contribute (1) high-quality metadata such as collection dates and provenance, (2) knowledge of collections in-house and elsewhere, (3) knowledge of source communities and ethical and legal issues, (4) advice on choice of specimens most suitable for sampling.