American Museum of Natural History

Center for Biodiversity and Conservation
Network of Conservation Educators and Practitioners

Applications of Museum Collections and Genomics to Biodiversity Conservation

Author(s): Anna Penna, Lauren T. Clark, Alexander T. Salis, Suzanne K. Macey, and Mary E. Blair

Source: Lessons in Conservation, Vol. 14, Issue 1, pp. 107-147

Published by: Network of Conservation Educators and Practitioners, Center for Biodiversity and Conservation, American Museum of Natural History

DOI: https://doi.org/10.5531/cbc.linc.14.1.7

This article is featured in ***Lessons in Conservation***, the official journal of the Network of Conservation Educators and Practitioners (NCEP), a program of the American Museum of Natural History's Center for Biodiversity and Conservation.

NCEP modules are designed for undergraduate and professional level education, but we encourage users to adapt the materials to their need and context.

Note to educators: this article is available in modifiable format and free to download from our Module Collection on our website, ncep.amnh.org. To access supplementary materials (e.g., presentations, teaching notes, exercise solutions, and associated files), please register as an educator.

Supplementary materials:
https://doi.org/10.5531/cbc.ncep.0190

Applications of Museum Collections and Genomics to Biodiversity Conservation

Anna Penna[i,ii,iii], Lauren T. Clark[iv,v], Alexander T. Salis[vi], Suzanne K. Macey[vii], and Mary E. Blair[vii]

[i]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA; [ii]Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA; [iii]Department of Biology, Lund University, Lund, Sweden; [iv]Institute of Comparative Genomics, American Museum of Natural History, New York, NY, USA; [v]Centre for Paleogenetics, Stockholm University, Stockholm, Sweden; [vi]Department of Herpetology, American Museum of Natural History, New York, NY, USA; [vii]Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA

DOI: https://doi. org/10.5531/cbc.linc.14.1.7 | Supplementary: http://doi.org/10.5531/cbc.ncep.0190

ABSTRACT

To address the challenges of sampling endangered or extinct species in the field, many studies have turned to historically underutilized sources of genetic material: natural history museums. Despite the fact that DNA from specimens collected decades or even hundreds of years ago is often fragmented and degraded, research has shown that historical DNA can still be used effectively to infer phylogenetic relationships and intra-specific patterns of population genetic structure. This synthesis aims to provide students and conservation practitioners with a solid understanding of the methodological strategies needed to apply genetic tools to natural history museum specimens. Specifically, we offer clear definitions and essential considerations for designing a conservation genomics project that includes both modern and historical samples. We recommend that instructors use this synthesis to introduce the foundational knowledge required for two companion exercises: "The application of conservation museomics approaches to the protection of the Iberian Lynx (*Lynx pardinus*)" and "Designing a conservation genomics project incorporating DNA from museum specimens."

INTRODUCTION

Natural history museums and herbaria collections provide a unique spatiotemporal record of life on Earth. These collections house millions of specimens derived from animals, plants, fungi, and other organisms, often consisting of entire preserved organisms or their body parts, commonly referred to as "voucher specimens." Since the sixteenth century, naturalists have preserved these specimens for long-term use in morphology-based taxonomic work and comparative anatomy analyses. Recent advancements in laboratory techniques for studying DNA damage, coupled with the advent of high-throughput sequencing technologies, have transformed these century-old museum specimens into valuable sources of DNA, giving rise to the field of "museomics" (Raxworthy & Smith, 2021; Blair, 2024; Fong et al., 2023). From historical cabinets to modern molecular laboratories, museum collections are now integrating novel technologies that significantly enhance the scientific value of these specimens.

To obtain DNA for genetic analysis, researchers start with a small fragment of biological tissue (typically 50–200 milligrams of bone, teeth, ear punches, hair, etc.) from an organism. The DNA isolated from these samples is categorized based on the time since collection (modern, historical, or ancient) and the preservation process of the tissue source (Table 1). Any DNA recovered from a specimen intentionally preserved for long-term storage in a museum collection is referred to as "archival DNA" (Raxworthy & Smith, 2021). This archival DNA is often exposed to chemicals during preservation, leading to various types of DNA decay beyond the environmental factors that can damage naturally preserved specimens. Additionally, these chemicals can inhibit enzymatic reactions during DNA extraction protocols. In contrast, naturally-preserved specimens, such as those found in

Table 1: DNA can be obtained from samples of various ages (modern, historical, ancient) and collected from different types of biological material that was intentionally prepared for preservation (archival DNA) or has been preserved under certain natural conditions.

	Modern	**Historic**	**Ancient**
Preservation/ age	Typically younger than 30 years old	Between 30* and ~200 years old	Usually >200 years
Archival	Biological tissue preserved for long-term use in genetic research, following a methodology that minimizes the chances of DNA damage (e.g., frozen, or preserved in ethanol or buffer).	Tissue collected from a voucher specimen preserved in natural history collection for long-term morphological research (e.g., dry study skin, formalin-fixed, pinned specimens). Preservation protocols did not account for the risk of DNA damage, often resulting in genetic material that shows some level of damage.	Biological material collected from artifacts intentionally treated to maximize its lifespan but without considering the risk of DNA damage (e.g., the leather from a book cover from medieval times). However, the utility of ancient archival specimens in genetic studies is rarely explored due to the low probability of yielding amplifiable DNA.
Natural preservation	Biological tissue found exposed to the elements (e.g., forensics samples).	Biological material collected from ethnographic collections or archaeological sites.	Biological material collected from archaeological and/or paleontological sites.

*Samples younger than that of 30 years may also demonstrate DNA degradation patterns consistent with historic tissues. This is typical of formalin-fixed specimens or soft tissues that have been exposed to sunlight, heat, fluctuations in temperature, or humidity for long periods of time.

archaeological or paleontological sites, are not typically subjected to chemical treatments and are preserved due to specific environmental conditions (e.g., dry and cold environments). These naturally preserved samples, typically dated to be around 200 years old or older, are labeled as "ancient DNA" (Wandeler et al., 2007; Billerman & Walsh, 2019). The field specializing in ancient DNA samples is known as "paleogenomics."

Most specimens currently housed in museum collections and herbaria were collected less than 200 years ago and are classified as "historical DNA" (Wandeler et al., 2007; Billerman & Walsh, 2019). The majority of these specimens were preserved before the rise of genetic research in non-model organisms in the late 1970s. In recent times, genetic research has become central to biological studies, prompting researchers to collect small tissue samples before preparing voucher specimens for long-term preservation ("modern DNA"). This practice helps protect potential DNA sources from various factors that can cause DNA damage.

Museomics, the study of genetic information acquired from museum specimens, leverages thousands of historic and ancient specimens collected worldwide to provide a time series of genetic

diversity and biodiversity monitoring that was unavailable just 20 years ago (Blair, 2024). The rapid advancements in the ability to access genetic information to monitor diversity throughout time from extinct taxa of the past or endangered taxa of the present comes at a crucial time as the warming climate and human activity continues to accelerate the loss of biodiversity. The past few hundred years offer a clear picture of this loss, with species extinctions, shifts in species distributions, and biodiversity declines occurring at rates 100 to 1,000 times greater than the natural background extinction rate throughout geological time (Ceballos et al., 2015; De Vos et al., 2015).

This biodiversity crisis not only impacts ecosystems and species distributions but also affects genetic variation within species and populations. A recent study estimated a six percent loss of genetic variation in wild populations since the Industrial Revolution (Leigh et al., 2019). To better understand these changes in biodiversity, natural history collections provide a unique resource of DNA data for many rare, endangered, and locally or globally extinct species at different time points. By integrating genetic data from field studies and museum collections, museomics allows researchers to include more species in comparative analyses, establish baselines of genetic diversity for declining species, and reduce the cost and time associated with fieldwork. Museum samples have also enabled us to study the emergence of zoonotic diseases like Lyme Disease (Marshall et al., 1994) and avian influenza (Fanning et al., 2002). In unique cases, museomics may also contribute to ongoing conservation efforts (Blair, 2024; Fong et al., 2023; Raxworthy & Smith, 2021). Though the implementation of conservation efforts is not always possible or relevant to all research questions, the generation of genetic information from modern and historic specimens provides an excellent resource for future research and adds to the growing field of conservation genomics.

What is Conservation Genetics?

Conservation genetics is a field of biology that intersects with ecology, taxonomy, molecular biology, and population genetics. It applies genetic principles and tools to the conservation and management of species, populations, or evolutionarily significant units (ESUs), particularly those that are endangered or at risk of extinction, with the goal of studying their health and viability (Desalle & Amato, 2024). Key parameters in population genetics—genetic relationships, genetic diversity, and population size—are crucial for understanding the evolutionary dynamics and long-term survival of species. Genetic relationships reveal how individuals within a population are related, helping to assess inbreeding levels and genetic drift. Genetic diversity, or the variety of traits within a population, is essential for adaptability, as it increases a population's ability to respond to environmental changes and resist diseases. Low diversity can lead to inbreeding depression and a reduction in fitness. Population size influences genetic variation, with small populations more prone to genetic drift and loss of diversity, while larger populations tend to maintain greater genetic variation, supporting long-term resilience. Conservation genetics focuses on maintaining genetic diversity to ensure species' adaptability and survival, helping wildlife managers predict a population's genetic health and mitigate risks such as inbreeding, loss of diversity, and hybridization.

Scientists working in conservation genetics often study endangered or threatened species to understand the genetic factors that influence their survival. Key questions typically include: What has caused the genetic erosion (i.e., loss of genetic diversity) in this population? How can conservation managers mitigate this and improve the population's genetic health? These questions help identify the root causes of reduced genetic diversity, such as habitat loss, loss of connectivity between populations, over-exploitation, or climate change and inform strategies to increase population resilience.

Prior to the revolution of genetics in the early 2000s and genomics in the 2010s, the information

typically available to protect at-risk populations was limited to observations of populations in the field (Box 1). Now, with improved genomic techniques and the ability to access historic and ancient data, conservation geneticists can better understand genetic relationships across time and geographic ranges. This expanded toolkit allows them to more accurately trace the causes of genetic decline and devise more effective conservation strategies.

When examining fluctuations in genetic diversity, it is essential to revisit the fundamental mechanisms of evolution, such as mutation, gene flow, genetic drift, and natural selection. If comfortable with these concepts, proceed to the next section. If you need a more substantial overview of these concepts (and more!), we suggest you read the NCEP module "Conservation genetics"(https://doi.org/10.5531/cbc.ncep.0123).

Mutations serve as the primary source of genetic diversity, arising from stochastic changes that typically occur as molecular 'errors' during cellular replication. Genetic diversity refers to the range of genetic differences among individuals within a population. When these differences approach zero—indicating that individuals within a population, species, or evolutionarily significant units possess nearly identical genetic information across protein-coding regions—the population becomes increasingly vulnerable to genetic diseases and the spread of harmful alleles. This genetic uniformity can severely impair a population's ability to adapt to changing environmental conditions. If a crucial gene associated with fitness is compromised, the overall health, survival, and adaptability of the group can rapidly decline.

Low genetic diversity not only increases the risk of disease but also elevates the likelihood of extinction. Populations with limited genetic variation may also struggle to cope with environmental changes, or invasive species. This lack of adaptability can be particularly detrimental in the face of rapid climate change, habitat destruction, and other anthropogenic pressures, which are now more prevalent than ever. The inability to adapt can lead to population declines and eventual extinction, thereby triggering downstream ecological effects that disrupt entire ecosystems. This can result in extinction cascades, where the decline or extinction of one species negatively impacts others that rely on it for vital ecosystem services, such as pollination, nutrient cycling, or habitat provisioning.

Box 1: Conservation genetics vs. conservation genomics

Conservation genetics and conservation genomics are both key terms to describe how molecular DNA can be applied to questions critical fields in species conservation, and these two terms are used to signal differences in but they differ in scope and technological approach. Conservation genetics focuses on studying genetic diversity within and between populations by analyzing a few specific genes or markers, such as mitochondrial DNA and microsatellites. It helps assess inbreeding, population structure, and gene flow and is useful in managing small or endangered populations. However, it is limited by the small number of markers studied, providing a less comprehensive view of genetic health. Conservation genomics, on the other hand, utilizes advanced sequencing technologies like whole-genome sequencing to analyze the entire genome, offering a more in-depth understanding of genetic variation, including both neutral and adaptive traits. This broader approach allows for better insights into how species adapt to environmental changes and threats like habitat loss or climate change, making it a powerful tool for long-term conservation strategies. While conservation genetics lays the foundation, conservation genomics provides a more detailed, genome-wide perspective that can enhance species management and adaptation efforts.

Additionally, genetic diversity plays a crucial role in maintaining the resilience of ecosystems. Populations with greater genetic variation are more likely to contain individuals with traits that can withstand environmental stresses, ensuring the survival of species during adverse conditions. For instance, a genetically diverse plant population may produce individuals that can survive drought, while a uniform population may not have any individuals with such adaptive traits. This resilience contributes to the stability and functioning of ecosystems, benefiting not just the species involved but also the human communities that depend on these ecosystems for resources and services.

Once genetic diversity is established, these evolutionary mechanisms interact with it in ways that vary based on the population's size and connectivity. It's important to recognize that these mechanisms do not operate in isolation; they are influenced by demographic and geographical factors. For example, if two populations of the same species become physically isolated from one another, their gene flow—the exchange of genetic material through mating—can be significantly reduced. This decrease in gene flow can lead to the fragmentation of the species into smaller, more isolated populations, where the effects of genetic drift become pronounced. Over time, if these isolated populations remain separate, the resultant loss of genetic diversity is not solely due to diminished gene flow; it is also a consequence of genetic drift and/or inbreeding depression acting more strongly on smaller populations. Consequently, understanding the interplay between these evolutionary mechanisms is crucial for assessing and managing genetic diversity in conservation efforts.

Conservation geneticists use a variety of tools to assess the genetic diversity and health of populations or species. Before collecting samples for genetic or genomic analysis, researchers first define the population and the geographic area of interest, focusing primarily on species that are threatened, vulnerable, or endangered. They may then analyze known phenotypes within the species and identify the species' closest living relatives.

With these foundational questions answered, researchers can develop hypotheses that require biomolecular data, such as genetic, genomic, transcriptomic, or proteomic information. This data is integrated into models along with previously gathered phenotypic and ecological data. These models can provide insights into how adaptable a species or population may be to changing environmental conditions, such as climate and habitat alterations, and help predict future impacts on the species' viability or on specific evolutionarily significant units.

Incorporating Historical and Ancient Samples in Conservation Genetics Studies

A pioneering study by Higuchi et al. (1984) demonstrated the feasibility of recovering authentic DNA from dried muscle tissue obtained from a 140-year-old mounted taxidermy specimen of a quagga (*Equus quagga quagga*; Figure 1). The quagga, a subspecies of the African plains zebra, was hunted to extinction in the 19th century. This groundbreaking study paved the way for researchers to obtain DNA from extinct and rare species housed in museum collections. Since then, numerous studies have incorporated historical and ancient samples into genetic analyses, enriching both the temporal and geographical dimensions of genetic research (Mitchell & Rawlence, 2021; Raxworthy & Smith, 2021).

One common application of museomics, as exemplified by the quagga study, is sequencing DNA from extinct species, such as the iconic dodo (*Raphus cucullatus*; Shapiro et al., 2002), the Tasmanian tiger (*Thylacinus cynocephalus*; Feigin et al., 2018; 2022; Mármol-Sánchez et al., 2023), the wooly mammoth (*Mammuthus primigenius*; van der Valk et al., 2021), and the blue antelope (*Hippotragus leucophaeus*; Plaxton et al., 2023; Hempel et al., 2021). These sequences help confirm the

Figure 1. A museum specimen of the quagga (*Equus quagga quagga*) prepared for display at the Natural Science Museum of Bamberg (Germany) and later sampled for genetic analysis. Image credit: Reinhold Möller via Wikimedia Commons/CC BY-SA 4.0.

taxonomic identity and determine the phylogenetic placement of globally or locally extinct species or populations. Better phylogenies lead to more informed and accepted taxonomies; taxonomic resolution is critical to accurately set species priorities for conservation efforts (Mace, 2004). With the refinement of laboratory protocols to isolate endogenous DNA from preserved specimens, museomics has expanded beyond analyses involving only one or a few individuals, enabling population-level studies.

Natural history collections also offer an alternative to disturbing sensitive ecosystems and sampling protected live organisms, species in remote or inaccessible regions (e.g., areas experiencing conflict or war), or broadly distributed taxa for which collection permits may be difficult to obtain (e.g., Blair et al., 2023; Islam et al., 2024; Penna et al., 2024). By providing researchers access to samples of species and populations that would otherwise be challenging and expensive to collect in the wild, museomics can expand the geographical and taxonomic coverage of genetic studies. These historical samples have been applied to study biogeography, delimit species boundaries, clarify species diversity, and characterize hybrid zones (Raxworthy & Smith, 2021). With the global acceleration of human impacts on biodiversity and habitat loss in recent decades, arguably, the collection of voucher specimens for future research is becoming increasingly important (Rocha et al., 2014). Natural history collections and museomics can offer many benefits to conservation, but it is important to consider the ethics and social/historical context of collections and data derived from collections (Box 2).

The temporal information encoded in ancient and historical specimens allows researchers to address a broader range of questions. By combining modern and museum samples, it is possible to study the dynamics of genetic diversity over time. While historical specimens offer a more limited temporal reach compared to ancient specimens (hundreds of years versus thousands to over a million), the extensive collection of specimens amassed over the last ~200 years provides an unparalleled resource for studying the spatiotemporal records of species and populations globally, spanning multiple generations (Raxworthy & Smith, 2021). Because many historical samples were collected

Box 2: Museomics and ethics

Advances in museomics hold great promise to transform research and practice in biodiversity conservation. At the same time, the increasing use of these technologies and data provokes key ethical questions, some of which relate to general ethical questions in science and in -omics, such as cloning and artificial intelligence, while others relate to the ethics of natural history collections themselves. There are several mechanisms such as formal laws, institutional frameworks and policies, and journal review procedures that are relevant to and that inform ethical decisions made around the curation of and research allowed on natural history collections. Importantly, the ethics and strategies guiding the collection of biological specimens for natural history museums have been renewed and updated as the purpose of museums has evolved (Arengo et al., 2018). Yet, criticism and ethical debates continue to surround biological specimen collection for scientific studies, with some claiming that collection plays a role in species extinction (Minteer et al., 2014), and others arguing that the practice is no longer necessary given new technologies (Byrne, 2023). Collection strategies and institutional policies for collecting do adhere to strict permitting and ethical boundaries, for example by aiming to collect well below levels that would affect demography (Collar, 2000; Winker et al., 2010; ICOM, 2013; Rocha et al., 2014). Others argue that policies and strategies for collecting can and should be grounded in partnerships and agreements with Indigenous guardians and other local resource stewards towards co-stewardship of biodiversity and environmental sustainability (e.g., Blair, 2024).

Thus, conservation museomics, especially as it expands and interacts with a range of disciplines including Indigenous science, anthropology, archaeology, environmental humanities, and museum studies (Blair, 2024), can benefit from broader discussions about data ethics in many fields ranging from remote sensing (e.g., York et al., 2023) to Indigenous data sovereignty. For example, while the incorporation of cultural treasures in museomics is likely to advance the conservation relevance of the field, especially regarding explorations of human relationships with nature and biodiversity, such work must privilege the rights and wishes of Indigenous Peoples and descendant communities to determine how their cultural material and knowledge are used (Kreps, 2008; Simpson, 2009). Further, conservation museomics should consider whether data derived from specimens previously collected from Indigenous lands should be managed in collaboration with descendant communities following the CARE principles for Indigenous Data Governance (Carroll et al., 2020; 2021). Following these principles, Indigenous knowledge holders must be actively involved in any additional collection, as well as in the stewardship and governance of data to ensure ethical reuse of data in specific defined ways. The processes outlined in the CARE principles aim to ensure that Indigenous ethics inform access to data in a way that minimizes harm, maximizes benefits, and allows for future use that promotes the well-being of the descendant community.

when habitat coverage was significantly different, museomics can reveal the impact of climatic events, environmental changes, or human activities. This cross-generational analysis can help disentangle contemporary declines in genetic variation from historically low values driven by ancient population crashes, species-specific traits, and demographic dynamics (Schmid, 2018; Jensen et al., 2022; Brasil et al., 2023). Together, the advantages of population-level and cross-generational analysis allow for a more accurate characterization of changes in historical population size, distribution, connectivity, and relationships, which are crucial for developing conservation and management plans (Leonard, 2008). Genetic data obtained from museum specimens can also potentially help study other biological phenomena, such as biological invasions and species interactions that extend beyond the organisms' DNA, for instance, giving insights into microbiomes, diets, diseases, and parasites.

From the Field to the "Shelf"

According to a 2020 report by the National Academies of Sciences, Engineering, and Medicine, natural history museums in the US alone house between 800 million to 1 billion specimens (National Academies of Sciences, Engineering, and Medicine, 2020). These specimens are organized into dedicated collections based on major biological and taxonomic groups. For example, vertebrate zoology collections include fishes, amphibians, reptiles, birds, and mammals; invertebrate zoology collections encompass insects and related groups (e.g., spiders and crabs), soft-bodied animals (e.g., snails and worms), and various marine organisms (e.g., coral and starfish); and herbaria contain collections of plants, algae, and fungi. Specimens are collected from the field, prepared, preserved, stored, and sampled in various ways for genetic studies.

Collection of Fresh Tissue Samples

In the field, researchers collect blood and/or tissue from various body parts (muscle, spleen, liver, ear-clip), depending on the study question and organism. To collect blood or tissue samples from wild specimens, researchers often use invasive sampling procedures, such as syringes for blood or scissors for internal organs (muscle, liver, or spleen) and external structures (ear, toe, nail). The invasive sampling of internal organs is strictly regulated and requires humane euthanasia, often requiring prior approval from the Institutional Animal Care and Use Committee (IACUC), while the sampling of external structures can often be done without euthanasia. Alternatively, researchers can noninvasively sample biological material left behind by animals in the wild, such as shed hair or feathers, eggshells, and scat.

Including noninvasive sampling for further genetic analysis can help avoid unnecessary destructive sampling of tissue from limited specimens and reduce the bureaucratic burden (e.g., approval of invasive sampling permits) and other operational costs associated with sampling (e.g., setting up multiple traps). However, isolating DNA from noninvasive samples can be challenging due to high levels of DNA damage, lower amounts of DNA, the potential presence of DNA from other organisms (such as environmental microorganisms or microbiota), or the absence of genomic sequences in reference databases, such as GenBank (Locatelli et al., 2020). To minimize these challenges, noninvasive samples should be collected as fresh as possible, and specific laboratory procedures for DNA extraction should be used.

Preparation, Preservation, and Storage of Fresh Tissue

Once fresh tissue samples are collected in the field, they must be preserved and stored individually. Noninvasive samples, such as hair, feathers, or nails, can be kept dry at room temperature or preserved in silica desiccant if they contain moisture (e.g., scat). Fresh tissue, however, is usually preserved in liquid media and kept cold. These samples are often stored in ethanol or another buffer (e.g., RNA later, DNA/RNA Shield) and kept at 4°C, or even -20°C, when possible. If the study requires the preservation of high molecular weight DNA, it is generally argued fresh samples should be flash-frozen immediately after collection and processed for DNA extraction and sequencing as soon as possible (Wong et al., 2012; Dahn et al., 2022). However, flash-freezing in the field can present logistical challenges, especially in remote locations, and can potentially lead to DNA degradation, where putting into ethanol or other buffers may be better (Salis et al., 2025), especially if tissues samples are to be used for multiple projects and stored for future use (for example archived into a "cryo-collection").

Many museums also maintain "cryo-collections" for the long-term preservation of frozen tissue samples collected during modern expeditions, enabling the subsampling of these tissue samples in subsequent genetic research (Radin, 2015). In this case, fresh tissue samples are transferred to -80°C freezers or liquid nitrogen vapor freezers (Figure 2). Storing DNA at such a low temperature and minimizing the

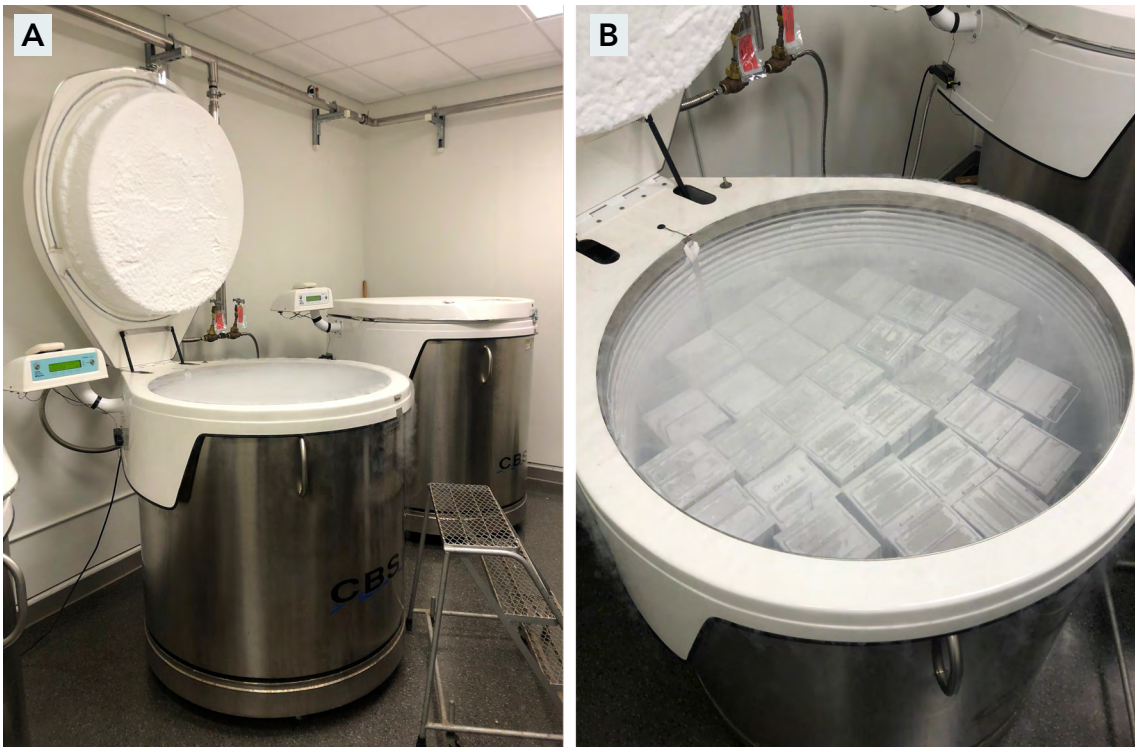

Figure 2. Liquid nitrogen vapor freezers used for long-term storage of biological samples. (A) Shows the tanks from the outside, and (B) shows the interior of a tank filled with 29 vertical stainless steel freezer racks that each fit 9 boxes of 96 samples. Image credit: Anna Penna.

number of thawing cycles helps slow down or halt nucleic acid degradation. These best practices minimize the chances of degrading DNA and other biomolecules for future use. Although some tissue samples stored in cryo-collections were first preserved in the early years of genetic research, they exhibit minimal DNA damage and can still be considered "modern DNA" (Raxworthy & Smith, 2021).

Preservation of Voucher Specimens

The method of preparation used for long-term preservation of voucher specimens varies depending on the organism, tissue composition, and the type of information researchers commonly seek from these specimens (Raxworthy & Smith, 2021). For instance, the color and morphology of bird feathers and mammal hair are crucial for species identification. Therefore, bird and mammal specimens are traditionally preserved as dried skins using a combination of salts and cornstarch. Insects, which have a hard exoskeleton made of chitin, usually maintain their original shape and appearance once dried. However, because these small creatures can be fragile to the touch, they are often preserved using pins inside cardboard boxes.

Aquatic animals, such as marine vertebrates and invertebrates, tend to become too distorted if dried, so they are typically preserved in fluid inside glass jars. Fluid-preserved specimens are also common in herpetological collections (amphibians and reptiles) because these animals have very thin skins that lose their general appearance when dried. Historically, fluid preparations involved an initial step of fixing tissues with formalin to slow down decomposition, followed by transfer to ethanol. More recently, researchers have started preserving specimens in ethanol only, rather than formalin, although this practice varies significantly from country to country (Hahn et al., 2022). For more details on the extensive variety of chemical compounds employed by different institutions across various taxa over the last centuries, you can refer to Simmons (2014) comprehensive catalog.

Table 2 summarizes the most common sources of DNA for museum specimens of different types of organisms, and Figure 3 illustrates some of the different specimen preservation methods.

Table 2: Summary of major types of voucher specimen preservation for different taxonomic groups that can be used as a source of biological tissue for genetic analysis.

Taxonomic group	Preparation type used as a source of tissue
Mammals, birds	Dried skin, bone, teeth, osteocrusts (dried tissue attached to bone, such as muscle or brain), occasionally fluid-preserved specimens (usually stored in ethanol 70%, can contain formalin).
Fishes, amphibians, reptiles	Fluid-preserved specimen (usually stored in ethanol 70%, can contain formalin), bone.
Some invertebrates (e.g., mollusks, arachnids)	Fluid-preserved specimen (usually stored in ethanol 70%, can contain formalin).
Insects, plants, algae, fungi	Dried specimen, pressed in paper, or pinned.

DNA Damage

Routinely obtaining DNA from historical and ancient sources has become possible in the last few decades thanks to recent advances in laboratory protocols for DNA extraction and library preparation (described in more detail in the sections below) and the advent of next-generation sequencing (Chen & Nedoluzhko, 2023). These protocol modifications aim to overcome the challenges of dealing with the differences in size and structural composition of these DNA molecules due to the accumulation of damage.

DNA replication is at the core of the most commonly used sequencing technologies. During replication, a double-stranded DNA molecule is copied to produce two identical molecules. DNA polymerases (a group of enzymes that catalyze DNA synthesis during replication) can duplicate the sequence of nucleotides found in DNA molecules by adding nucleotides at the 3'-OH group of another nucleotide, thus extending the DNA strand being copied. So, if the original DNA molecules present some type of structural or compositional damage, the DNA polymerase will not be able to properly attach to the DNA molecule or extend the nucleotide sequence. Therefore, to better understand the rationale behind the choice of laboratory protocols that should best suit the type of museum specimen available for the project, the researcher must first consider the consequences of these damages to DNA synthesis.

The DNA molecules inside living organisms' cells suffer damage that is repaired by intricate cell machinery. After death, these cellular repair mechanisms stop functioning and the DNA molecules are exposed to numerous factors that threaten its stability. These include digestion by intracellular nucleases and microorganisms, oxidative and hydrolytic damage that results in DNA shearing (that can result in how a nucleotide base is called by a sequencing machine), as well as attachment to proteins and other DNA molecules (crosslink). Consequently, DNA breakages and damages accumulate over time, at rates that can vary depending on the preservation method. That is why researchers usually separate non-modern samples into two categories (i.e., historical and ancient), and differentiate archival from naturally-preserved DNA.

DNA damage can be divided into three main categories (Figure 4): (1) Fragmentation: lesions that lead to a shortening or "fragmentation" of the DNA molecule; (2) Blocking Lesions: lesions that prevent the replication of the DNA molecule by blocking the action of the polymerases; and (3) Miscoding Lesions: lesions that result in the incorrect incorporation of nucleotides during DNA

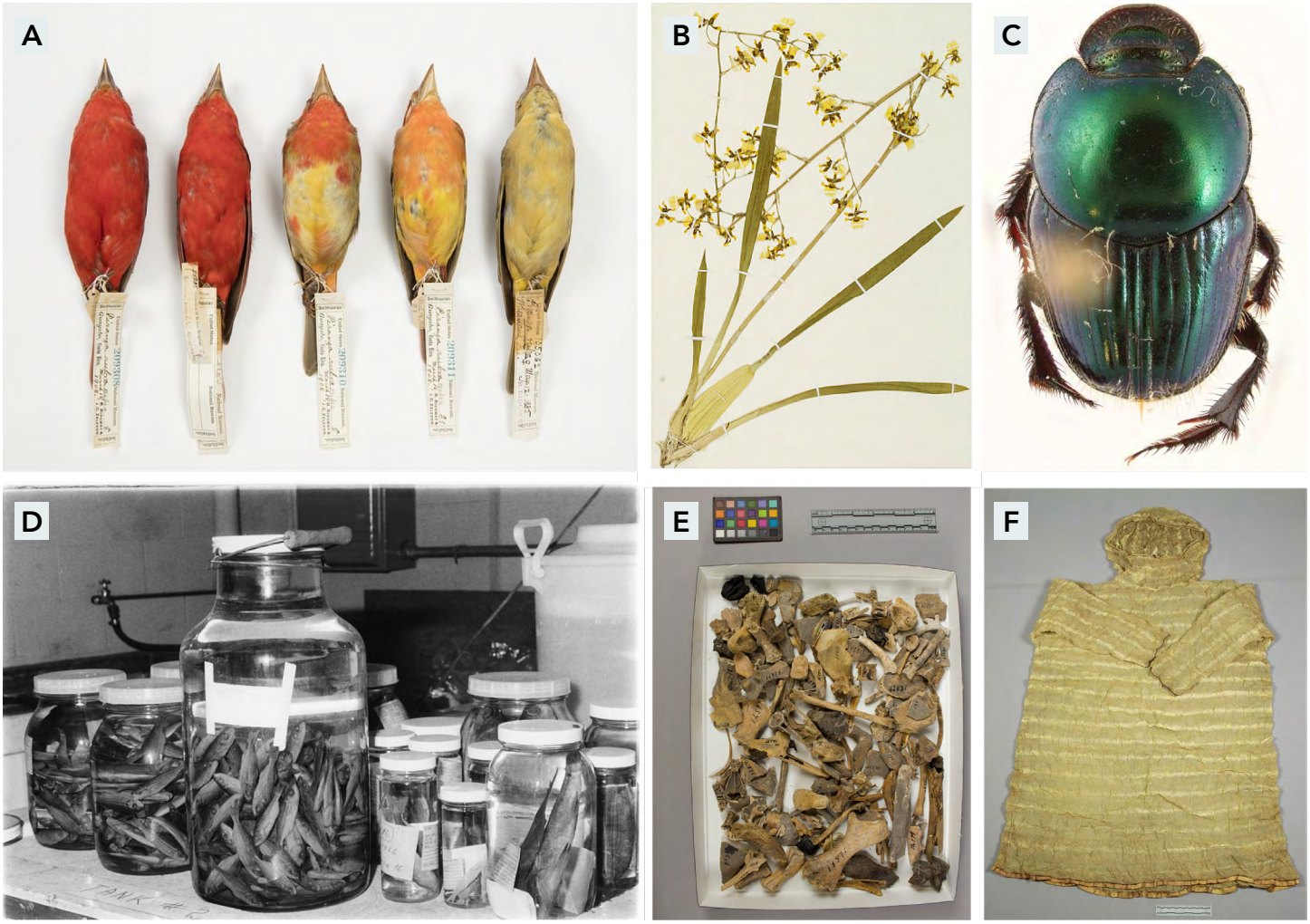

Figure 3. Some examples of specimen preparations often found in natural history collections. (A) Birds and mammal specimens are usually prepared as dry study skins. (B) Plant and algae specimens are dried and pressed on paper. (C) Insects are often preserved as pinned dried specimens. (D) Bats, aquatic organisms (e.g., fish, amphibians, mollusks), and reptiles (e.g., snakes, lizards) are commonly preserved in spirits (e.g., ethanol). Specimens with all internal organs (e.g., for anatomical studies) are also preserved in spirits. Small-sized specimens are most often stored in glass jars, whereas large-bodied specimens are stored in large aluminum cases. (E) Bones are a common find in archaeological collections and can be a great source of DNA. (F) Ethnographic collections often have material objects made of animal and plant parts, such as this waterproof parka made of sea lion intestines by the Aleuts from the North Pacific in the early 1800s. Image credit: (A) "Summer Tanager (*Piranga rubra rubra*), Indiana, USA, 1885" via Smithsonian Institution National Museum of Natural History Division of Birds. (B) "Popcorn Orchid" via Smithsonian Institution National Museum of Natural History. (C) "African elephant dung beetle, Mpala, Kenya, 2001" via Smithsonian Institution National Museum of Natural History Department of Entomology. (D) "Marine specimens in a jar" via Smithsonian Institution National Museum of Natural History Division of Fishes. (E) "Unidentified assemblage of bones collected in an archaeological site in the Dominican Republic in 1872" via Smithsonian Institution National Museum of Natural History Department of Anthropology. (F) "Waterproof parka made of the peritoneal coat of the intestines of sea lions, Aleuts from the Pacific Northwestern Coast, USA, acquired in 1829–1830" via Smithsonian Institution National Museum of Natural History Department of Anthropology.

replication, which are called "missense" nucleotides. Compared to the DNA present inside the cells of living organisms, the DNA of preserved specimens is often sheared into smaller fragments, cytosine bases are deaminated to uracil bases that occur frequently at the end of DNA fragments and degraded DNA molecules often crosslink to other DNA molecules or proteins.

DNA fragmentation leads to shorter DNA fragments either when both strands of the DNA double helix are broken or when only one strand is shortened, such as in single-strand overhangs. Damage

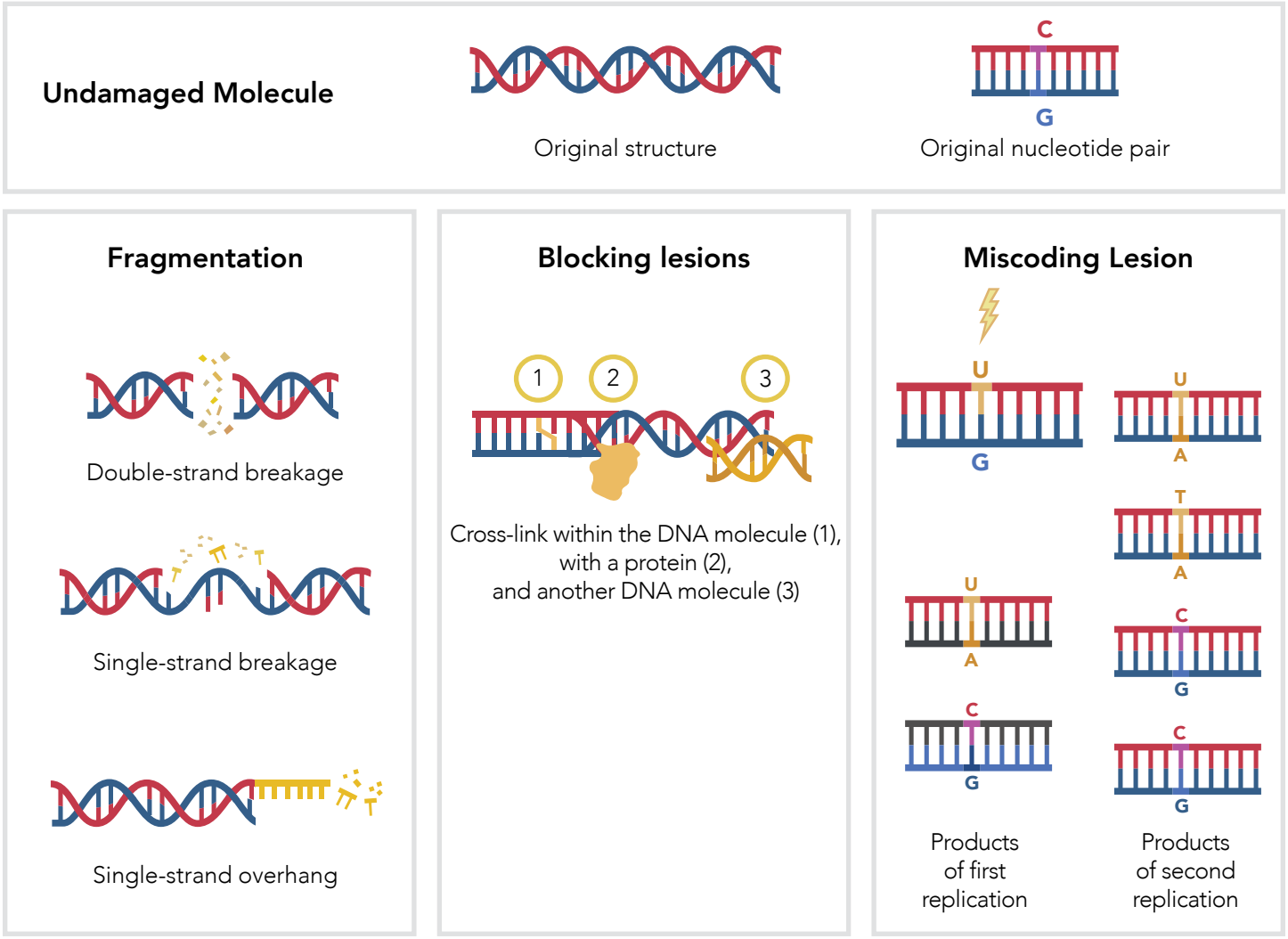

Figure 4. Examples of the three main types of lesions found in DNA molecules: DNA fragmentation, blocking, and miscoding lesions. Image credit: Anna Penna.

to a single strand can also lead to the formation of nicks, where the phosphodiester bond between two adjacent nucleotides is broken. These nicks prevent the extension of the nucleotide chain during replication, potentially resulting in shorter fragments if not repaired. Examples of blocking lesions include the formation of crosslinks within the DNA molecule or between the DNA molecule and proteins or other DNA fragments in the solution. Blocking lesions are common in formalin-fixed specimens and can interfere with DNA replication.

Miscoding lesions lead to the deletion or incorrect incorporation of nucleotides during replication. For instance, one of the most common miscoding lesions found in ancient DNA is the deamination of cytosine to uracil. In this case, DNA polymerases will incorporate an adenine (A) opposite the uracil (U) and, subsequently, a thymine (T) opposite the adenine, causing apparent guanine (G) to adenine (A) and cytosine (C) to thymine (T) substitutions, depending on the strand sequenced. Typically, these miscoding lesions accumulate at the ends of fragments.

As described above, there are multiple sources of ancient and historical tissues from museum specimens that are suitable for genetic analysis. Differences in the types of tissues sampled, the age of the sample, environmental conditions (humidity, temperature, salinity, and pH) and preservation conditions (use of chemicals by museum staff to aid in morphological preservation) can all contribute

to the significant variability observed in the preservation of DNA from museum specimens. It is, thus, important to consider these factors when selecting the appropriate wet lab techniques to ensure that the highest quality endogenous DNA can be isolated from each specimen.

The next section will highlight these key lab techniques in an ancient DNA workflow—from working with museums to destructively sample tissues to extraction and library preparation to sequencing and bioinformatic analysis. These wet lab and bioinformatic techniques have been optimized to isolate the small and damaged DNA fragments typical of historic and ancient tissues from museums. Before sampling, wet lab, or bioinformatics work takes place, researchers need to define their research question. While we do not discuss this process in detail here, we recommend this module's companion exercise "Designing a conservation genomics project incorporating DNA from museum specimens" (https://doi. org/10.5531/cbc.linc.14.1.9), for more guidance on developing a research question.

A PRACTICAL GUIDE FOR INCORPORATING MUSEUM SAMPLES IN GENETIC STUDIES

Designing the Sampling of Genetic Material According to the Project Needs and Resources Already Available

When designing a conservation genomics (or genetics) study, the number of individuals sampled per population or species can vary based on the specific research question. Biological samples are the main source of genetic material. However, acquiring the necessary number and diversity of samples—whether from wild populations, museum collections, or archaeological sites—can be time-consuming, expensive, or logistically complex. Therefore, before starting the tissue sampling phase in the field or from museum collections, it is crucial to carefully consider the study's goals and constraints. Researchers should first assess 1) the temporal or spatial dimensions of the research question; 2) the statistical requirements of the analysis they plan to conduct, and 3) the taxonomy, ecology, and natural history of the study organism. For example, if the goal is to assess decline in genetic diversity over time, it is important to ensure that enough individuals are sampled from multiple historical periods. If the research focuses on population connectivity across the landscape, the sampling should encompass individuals from geographically distinct locations that reflect the species' range, spanning potential barriers to gene flow.

These considerations are directly linked to the broader types of questions asked in conservation genomics. By leveraging genome-wide data, researchers can explore patterns of genetic diversity, detect inbreeding or local adaptation, and model population structure and gene flow—all of which require thoughtful sampling strategies. In addition, sampling design is key for answering phylogenetic questions central to conservation, such as identifying cryptic species, resolving evolutionary relationships among threatened lineages, or determining the closest relatives of recently extinct taxa. Integrating samples from both modern and historical sources, including museum specimens and ancient DNA, enables scientists to uncover evolutionary patterns that may otherwise remain hidden, ultimately providing deeper insights into species boundaries, adaptive potential, and conservation priorities. Careful planning at the sampling stage lays the foundation for robust genetic inferences, guiding effective conservation decisions and enhancing our understanding of biodiversity across space and time.

So how many samples do you need to answer your question? The figures provided here are general guidelines and should be considered as suggestions rather than rigid rules. The aim here is to help you think critically about the challenges involved in sampling and designing a study that will provide the necessary data to answer key conservation questions.

For population-level studies, such as estimating genetic diversity, detecting inbreeding, or identifying signals of local adaptation, researchers typically aim for 20–30 individuals per population to ensure robust allele frequency estimates (Hale et al., 2012; Nazareno et al., 2017). Interestingly, research has been showing that large sample sizes at low sequencing depth are desirable to achieve high accuracy of estimates of genetic variation (Fumagalli et al., 2013; Lou et al., 2021). Studies of population connectivity or landscape genomics require spatially distributed samples, with 10–20 individuals per site across multiple locations, ideally capturing environmental or habitat variation (Meirmans, 2015; Rellstab et al., 2015). In contrast, phylogenomic studies and questions involving cryptic species or evolutionary relationships often prioritize taxonomic breadth over dense population sampling. These projects may only require 1–5 representative genomes per lineage, especially when high-coverage genomes are available (Funk et al., 2012; Lou et al., 2021). However, for species delimitation or defining evolutionarily significant units, sampling 5–10 individuals per suspected lineage can provide the resolution needed to detect subtle genetic differentiation (Carstens et al., 2013). Finally, studies involving temporal sampling—such as using museum specimens or ancient DNA—must balance historical coverage with the technical limitations of degraded DNA, often working with fewer samples but drawing on a wider time span (Fumagalli, 2013). Aligning sample size and sampling design with the specific research objectives ensures efficient use of resources and maximizes the analytical power of conservation genomic studies (McMahon et al., 2014; Shafer et al., 2015).

Before sampling tissue from live specimens in the field, one should start by checking what has already been sampled and stored long-term in museum biobanks for the target species or populations. These biobanks are collections of biological samples and associated information organized systematically for research purposes. Since the proliferation of molecular biology studies, modern cryogenic collections have also become a common practice among natural history museums. These collections house frozen tissues that can be requested as loans for application in genetic research.

You might also want to check what genetic data has already been generated for the focal species/group and is publicly available for download. There are several online repositories that host this type of digital data, such as the National Center of Biotechnology Information (NCBI), American GenBank, European Nucleotide Archive (ENA), Global Genome Biodiversity Network (GGBN), DNA Data Bank of Japan (DDBJ), Chinese Genbank, and Genomes on a Tree, to name a few. These archives contain different data types (raw reads, mapped reads, assemblies, metadata files) or can point to where data were deposited) at various levels of data processing.

Lastly, zoos, botanical gardens, or captive breeding programs often have individuals of rare or endangered animals and plants. These institutions can provide access for sampling the focal species of interest. Sampling from live or recently deceased focal species in these institutions can be a cost-efficient way to obtain high-quality tissue that can proceed directly to library preparation and sequencing. This strategy can be particularly helpful for generating genetic information at high resolution (e.g., annotated chromosome-level reference genomes) and has been successfully implemented by the DNA Zoo consortium (https://www.dnazoo.org/assemblies).

To plan the sampling of historical specimens, the first step is to check museum collections catalogs. Most natural history collections make their catalog databases available online, allowing anyone to search for specimens and download the data. Downloading individual databases can be cumbersome, so using an aggregator that mirrors multiple collections catalogs is the most efficient way to start by simultaneously searching across multiple collections. Examples include the iDigBio portal (https://portal.idigbio.org/) and GBIF (https://www.gbif.org/). There are also clade-

specific aggregators, such as VertNet for vertebrates (http://portal.vertnet.org/search, no longer updated) and JAQ (https://www.jacq.org/) for herbarium specimens. If you are interested in sampling specimens from a particular collection that does not have an online database, directly contact the curators and collection managers to inquire about the specimens registered in their catalog and the sampling procedure. Establishing a relationship with museum specialists can help you determine the best specimens and type of tissues to sample from their collections. Discussing the scientific objectives of your study with the curators and collection managers is an important step in arranging permits for destructive sampling.

Once you have planned your field, zoo, or museum sampling strategy, you will need to arrange all the required permits to collect and transport tissues, and if relevant, handle or observe vertebrate animals. Be aware that tissue collection regulations vary from country to country, and multiple agencies may control these regulations. A good starting point is to check the The Nagoya Protocol on Access and Benefit-sharing (https://www.cbd.int/abs/default.shtml), as well as the degree of protection of your target species under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), an international agreement aimed at ensuring that international trade in specimens of wild animals and plants does not threaten their survival. For further information specific to US laws, the U.S. Fish and Wildlife Service can provide more details. Additionally, you should check with your local Institutional Animal Care and Use Committee (IACUC) for specific details about your institution's regulations regarding the usage and handling of vertebrate samples.

Some questions to guide this step:
1. Does the sampling cover the temporal/geographic/taxonomic range required to answer your question?
2. How many samples does your study require to have enough statistical power to answer your question?
3. How much of your budget do you need to allocate for sampling?
4. What permits are required to sample in the area and to transport biological material across borders?

Collecting Tissue Samples from Voucher Specimens for Genetic Analysis

Before sampling a museum specimen, it is always recommended to inspect it for signs of mold (in the specimen, as well as in the box where it is stored or in the liquid media in which it is preserved), lumps of salt, and the level of tissue disintegration, as these characteristics are often predictive of low DNA extraction and sequencing success. These external debris and other microorganisms can interfere with enzymatic reactions, or even decrease the proportion of endogenous DNA obtained in the final extraction. Ideally, genetic material should be obtained from tissue sampled directly from the preserved specimen, but practices can vary depending on the study organism. For instance, for large-bodied specimens, it is preferable to avoid relying on debris or fragmented material disconnected from the voucher specimen. For herbarium specimens, which include envelopes of leaf tissue on the specimen sheets, the best practice is to take from these envelopes rather than break more of the specimen on display. It is important to note also that for some types of specimens, it may not be possible to subsample. For example, studying small insects often requires the entire specimen to be used. Pinned specimens are often difficult to subsample as they are so brittle and easily destroyed. Because voucher specimens are invaluable materials in multiple types of biological research (e.g., comparative anatomy, morphology, systematics, ontogeny), it is also important to avoid destroying non-bilateral structures that are relevant in taxonomic research (e.g., sampling a limb is preferred over sampling the thorax).

These precautions are crucial for preserving the integrity of both historical and ancient specimens,

but ancient samples present their own unique challenges. Ancient samples are most often limited to tissue collected from sturdy structures, such as fibers, dry bone, and teeth that have been preserved over long periods of time, with occasional exceptions of preserved soft tissues found in optimal (cold, dry, and anoxic) conditions. Compared to historical-level samples, extractions from ancient samples typically have much lower DNA concentrations and the average size of DNA fragments acquired is shorter (35–250 bp; Pääbo, 1989; Dabney et al., 2013; Dabney et al., 2019), and contain DNA from endogenous and exogenous sources. Once collected, ancient specimens are also stored in museum collections. Since there is no standard of how ancient-level specimens must be stored among museums worldwide, they may be stored alongside historical specimens.

When collecting tissue samples for genetic studies, it is essential to maintain tidy laboratory conditions to avoid contamination from external sources or cross-contamination between samples. Therefore, sterilizing all sampling tools and surface areas is crucial when sampling specimens in the field and museum collections. Given that historical and ancient samples yield much lower DNA concentrations than modern samples, contamination from modern sources of DNA can severely impact the results of a museomics study. Contamination from modern human DNA can be particularly problematic for studies involving humans and other primates. Due to the high levels of genetic similarity from close phylogenetic relationships, modern DNA can go undetected and bias bioinformatics results. Therefore, working with these samples requires extremely stringent laboratory protocols (Llamas et al., 2017). Appropriate personal protective equipment (PPE) including gloves and face masks may also be important to utilize during both sample collection and processing to protect project staff from any potentially hazardous chemicals (e.g., arsenic or pesticides) that may have been used during specimen preparation and conservation.

Switching pre-sterilized disposable blades after collecting each tissue sample is ideal to avoid cross-contamination between samples. Alternatively, metal forceps can be sterilized by washing with RNAaway or DNAaway, followed by an ethanol wash, or heating in a high-temperature clean flame, such as a Bunsen burner or alcohol burner, and finally setting the tool in a UV crosslinker for upwards of 20–30 minutes. The ethanol wash ensures that the DNAaway/RNAaway does not corrode the metal tool. If using a bunsen burner or alcohol burner, wait until the metal surface cools down before touching the tissue.

Once you have collected the tissue, ensure each sample is identified with a unique label and stored in a correctly labeled tube. Tubes must be labeled on the cap and the side of the vial with a resistant marker (i.e., that can stand ethanol, water, and changes in temperature). These are archival best practices that can help in the long-term preservation of the collected tissue. Besides collecting the tissue, it is also essential to record all metadata associated with the sample. Without this ancillary data, the archived tissue loses its scientific value. This ancillary information includes physical features and measurements, the sex and age of the individual, the date and locality it was collected, the type of environment it was found in, who collected it, where it was deposited, and more detailed information such as the presence of parasites and habitat details. In conservation, accurate specimen data is fundamental when data from historical specimens inform conservation translocation decisions (Verry et al., 2019).

While all samples should be kept away from UV light, other storage requirements depend on the type of sample and the type of genetic material required for the study. For instance, fresh blood, muscle, or liver samples must be stored cold, whereas some buffers enable storage at room temperature (e.g., DNA/RNA Shield). Hair, feathers, and blood stored on specialized paper can also tolerate room

temperature for longer. Note, a recent study (Colella et al., 2020) demonstrated that storing tissue in ethanol can decrease the molecular weight of DNA molecules, which can impact the average fragment length of the DNA molecules and reduce the sample's adequacy for specific analyses (e.g., long-read sequencing). Ensuring that your samples are stored properly for the desired study will minimize DNA degradation and contamination.

Some questions to guide this step:
1. What tools do you need to minimize the chances of external and cross-contamination when sampling biological material?
2. How are you labeling your tubes?
3. What are the most appropriate storage conditions for the type of samples you will be collecting?
4. How are you keeping track of any associated metadata?

Isolating DNA from Museum Specimens

Specimen preparation techniques, most of which were developed centuries ago, aim to maximize the preservation of specimen body parts and their utility for traditional anatomy-based research. However, these techniques significantly damage and impair access to chemical components, including DNA molecules. Until recently, obtaining genetic material from museum specimens was challenging and prone to failure because the remaining DNA is in low quantity, highly fragmented, and often attached to proteins. These difficulties have been overcome by recent developments in DNA extraction protocols and the emergence of sequencing technologies that perform well with short DNA fragments (commonly called next-generation DNA sequencing). While these methods provide opportunities to access novel data types from hard-to-access species, they also raise some concerns. Notably, compared to modern tissue, samples obtained from preserved specimens tend to have lower DNA yields. Therefore, they should be handled with care to minimize external contamination from sources other than the target specimen, such as bacteria, fungi, and even humans (Knapp et al., 2012). Non-modern samples should ideally be processed in laboratories dedicated to historical or ancient DNA work, which are physically isolated from modern DNA facilities that follow stringent disinfection routines and enforce the use of personal protective equipment (PPE, such as double layers of gloves, full-body suit, face mask, hair, and beard net). To control for external contamination, every step of laboratory work (DNA extraction, purification, and library preparation) should include negative controls (i.e., blank samples with no DNA present; Knapp et al., 2012). A common concern, especially among museum collection staff, is that sampling tissue from preserved specimens always involves some extent of destructive sampling. To address these challenges, scientists have developed protocols that maximize DNA yields and quality while reducing contamination and damage to the voucher specimen. Different solutions have been attempted to accommodate the diversity of specimen preparation types, with variable success (see Box 3 for more details).

The choice of lab work protocol to isolate genetic material from biological samples depends on the type of tissue at hand. This is not unique to museum specimens and applies to fresh tissue collected in the field as well. Different tissues and organisms require specific DNA extraction protocols. For instance, plants have a cellular wall around the cell membrane and tend to accumulate secondary metabolites that can interfere with downstream enzymatic reactions in the extraction process. Therefore, cell lysis must be preceded by a mechanical step (e.g., grinding in liquid nitrogen with a sterile mortar and pestle) to break down the cell wall. Because plants have different secondary metabolites that can inhibit some enzymatic reactions, extra purification steps may be necessary to ensure these unwanted molecules are removed from the solution.

Over the past decades, scientists have optimized DNA extraction protocols to maximize the yield of genetic data recovered from preserved specimens. Because the amount of DNA that can be recovered varies among parts of the specimen, preservation techniques, and the specimen's age, it is essential to choose the correct protocol that suits your study. Box 3 provides an overview of some of the main challenges to extracting DNA from the most commonly available preparation types and references to help you prepare for this step of the laboratory work. Given that fresh tissue samples tend to be less degraded, numerous well-established protocols and commercial kits have been developed for different tissue types and organisms. Therefore, we won't cover the laboratory protocols for fresh samples here.

Box 3: Overview of advantages and challenges to isolating DNA from different specimen preparation types

Dry study skins

- Examples: Mammals and birds.
- Advantages: Fast and simple to collect samples. Toe pads typically have high amounts of endogenous DNA.
- Challenges: Skins are often treated with chemicals (e.g., arsenic).
- *Examples of protocols*:
  - McDonough, M. M., Parker, L. D., Rotzel McInerney, N., Campana, M. G., & Maldonado, J. E. (2018). Performance of commonly requested destructive museum samples for mammalian genomic studies. Journal of Mammalogy, 99(4), 789–802. https://doi.org/10.1093/jmammal/gyy080
  - Campos, P. F., & Gilbert, M. T. P. (2019). DNA extraction from keratin and chitin. Ancient DNA: Methods and protocols, 57–63. https://doi.org/10.1007/978-1-4939-9176-1_7
  - Penna, A., Blair, M. E., Lui, H. L., Peters, E., Kistler, L., & Pozzi, L. (2024). Overcoming challenges to extracting and sequencing historical DNA to support primate evolutionary research and conservation, with an application to Galagos. International Journal of Primatology, 45, 1375–1403. https://doi.org/10.1007/s10764-024-00429-3
  - Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2020). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. Molecular Ecology Resources, 20(5), 1220–1227. https://doi.org/10.1111/1755-0998.13077

Bone and teeth

- Examples: Vertebrates.
- Advantages: Hard tissue that can be found in most ancient specimens, also skeletons are often preserved in historical vertebrate collections.
- Challenges: It is difficult to reach the organic layer, which usually requires more destructive sampling techniques, such as drilling.
- *Examples of protocols*:
  - Dabney, J., & Meyer, M. (2019). Extraction of highly degraded DNA from ancient bones and teeth. Ancient DNA: methods and protocols, 25–29. https://doi.org/10.1007/978-1-4939-9176-1_4
  - Dehasque, M., Pečnerová, P., Kempe Lagerholm, V., Ersmark, E., Danilov, G. K., Mortensen, P., Vartanyan, S., & Dalén, L. (2022). Development and optimization of a silica column-based extraction protocol for ancient DNA. Genes, 13(4), 687. https://doi.org/10.3390/genes13040687
  - Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth, and sediments for high-throughput sequencing. Nature Protocols, 13(11), 2447–2461. https://doi.org/10.1038/s41596-018-0050-5

Fluid-preserved specimens

- Examples: Fish, amphibians, reptiles, aquatic invertebrates, larvae, spiders, scorpions, birds, and bats.
- Advantages: Specimens in high numbers or with lots of tissue. The media can be used as a source of DNA but can have high levels of damage and contamination.
- Challenges: High chances of the specimen being fixed in formalin, which creates crosslinks within the DNA molecule, or between protein and other DNA molecules present in the media. Even samples that are preserved only in ethanol, will yield fragmented DNA.
- *Examples of protocols:*
  - Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+-year-old intractable natural history specimens. Molecular Ecology Resources, 17(5), 1003–1008. https://doi.org/10.1111/1755-0998.12655
  - Straube, N., Lyra, M. L., Paijmans, J. L. A., Preick, M., Basler, N., Penner, J., Rödel, M., 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(7), 2299–2315. https://doi.org/10.1111/1755-0998.13433

Eggshells

- Examples: Birds and reptiles.
- Advantages: Excellent preservation of DNA and other biomolecules, common specimen type in avian and some other collections (e.g., invertebrates and herpetology). Often found in archaeological and paleontological sites. Noninvasive sampling of wild specimens.
- Challenges: Often delicate in nature and prone to species misassignment (species can have eggs that look extremely similar).
- *Examples of protocols*:
  - Oskam, C. L., Haile, J., McLay, E., Rigby, P., Allentoft, M. E., Olsen, M. E., Bengtsson, C., Miller, G. H., Schwenninger, J., Jacomb, C., Walter, R., Baynes, A., Dortch, J., Parker-Pearson, M., Gilbert, M. T. P., Holdaway, R. N., Willerslev, E., & Bunce, M. (2010). Fossil avian eggshell preserves ancient DNA. Proceedings of the Royal Society B: Biological Sciences, 277(1690), 1991–2000. https://doi.org/10.1098/rspb.2009.2019
  - Grealy, A., Langmore, N. E., Joseph, L., & Holleley, C. E. (2021). Genetic barcoding of museum eggshells improves data integrity of avian biological collections. Scientific Reports, 11(1), 1605. https://doi.org/10.1038/s41598-020-79852-4
  - van der Meij, S. E. T., & Nieman, A. M. (2016). Old and new DNA unweave the phylogenetic position of the eastern Atlantic gall crab Detocarcinus balssi (Monod, 1956) (Decapoda: Cryptochiridae). Journal of Zoological Systematics and Evolutionary Research, 54(3), 189–196. https://doi.org/10.1111/jzs.12130

Pinned dried specimens

- Examples: Non-aquatic insects.
- Advantages: One of the more widespread collection types, the number of specimens can often vastly outnumber other collections, making population-level analysis possible.
- Challenges: Destructive sampling can destroy important morphological structures, given that specimens are often very small and collections can have few collected specimens.
- *Examples of protocols*:
  - Lalonde, M. M., & Marcus, J. M. (2020). How old can we go? Evaluating the age limit for effective DNA recovery from historical insect specimens. Systematic Entomology, 45(3), 505–515. https://doi.org/10.1111/syen.12411
  - Orr, R. J., Sannum, M. M., Boessenkool, S., Di Martino, E., Gordon, D. P., Mello, H. L., Obst, M., Ramsfjell, M. H., Smith, A. M. & Liow, L. H. (2021). A molecular phylogeny of historical and contemporary specimens of an under-

studied micro-invertebrate group. Ecology and Evolution, 11(1), 309–320. https://doi.org/10.1002/ece3.7042
- Campos, P.F., Gilbert, M.T.P. (2019). DNA Extraction from Keratin and Chitin. In: Shapiro, B., Barlow, A., Heintzman, P., Hofreiter, M., Paijmans, J., Soares, A. (eds) Ancient DNA. Methods in Molecular Biology, vol 1963. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-9176-1_7

Herbaria specimens

- Examples: Plants, fungi, algae, lichen, moss.
- Advantages: One of the more common and widespread collection types. Fast and simple to collect samples.
- Challenges: Extraction requires some type of mechanical lysis step, and depending on the species, secondary metabolites can interfere with enzymatic reactions.
- *Examples of protocols*:
  - Hart, M. L., Forrest, L. L., Nicholls, J. A., & Kidner, C. A. (2016). Retrieval of hundreds of nuclear loci from herbarium specimens. Taxon, 65(5), 1081–1092. https://doi.org/10.12705/655.9
  - Folk, R. A., Kates, H. R., LaFrance, R., Soltis, D. E., Soltis, P. S., & Guralnick, R. P. (2021). High-throughput methods for efficiently building massive phylogenies from natural history collections. Applications in Plant Sciences, 9(2), e11410. https://doi.org/10.1002/aps3.11410
  - Gutaker, R. M., E. Reiter, A. Furtwängler, V. J. Schuenemann, and H. A. Burbano. (2017). Extraction of ultrashort DNA molecules from herbarium specimens. BioTechniques, 62, 76–79. https://doi.org/10.2144/000114517
  - Løken, S. B., Skrede, I., & Schumacher, T. (2020). The *Helvella corium* species aggregate in Nordic countries–phylogeny and species delimitation. Fungal systematics and evolution, 5(1), 169–186. https://doi.org/10.3114/fuse.2020.05.11
  - Cubero, O. F., Crespo, A. N. A., Fatehi, J., & Bridge, P. D. (1999). DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. Plant Systematics and Evolution, 216, 243–249. https://doi.org/10.1007/BF01084401

Some questions to guide this step:

1. What preservation method is more commonly used for the organism you are interested in studying?
2. How would you identify unsuitable specimens for which there might be low chances of obtaining genetic material (e.g., presence of mold, salt crystals, level of decomposition of internal organs)?
3. What protocols best suit your target species given the type of preservation method most commonly found in natural history collections?
4. What success rate do you expect for the DNA extraction step?
5. Are there any toxic chemicals you will be exposed to? How should you protect yourself to minimize contact and exposure to them?
6. How much of each laboratory reagent and supplies will you need to obtain genetic material for all the samples you have collected?

From Genetic Material to Sequences

Once the DNA from biological specimens has been isolated, it is time to determine the order of genetic information contained in these molecules through sequencing. Sequencing involves determining the exact order of nucleotide bases (A, T, C, and G) that compose each DNA molecule. Because genetic sequencing requires highly specialized machines and trained personnel, this step is usually outsourced to sequencing companies. Once the DNA molecules are sequenced, they are stored in a digital file and referred to as "reads." In the field of ancient and historical DNA, the ability to go further back in time has improved thanks to the development of laboratory protocols for isolating DNA from old and preserved specimens, as well as advancements in sequencing technologies.

In the era of massively parallel sequencing, it only takes a few hours to simultaneously sequence

billions of reads from multiple individuals. In the final step of DNA extraction, DNA molecules are in solution. Before sequencing, these molecules must undergo a series of laboratory procedures to ensure that all DNA molecules (1) can be traced back to a particular individual and (2) are ready to be sequenced. These are the two major goals of laboratory work, which together are called "library preparation." This library preparation step precedes sequencing on high-throughput sequencing (or "next-generation sequencing" NGS) platforms.

We have already covered the main differences in the DNA molecules obtained from fresh and preserved tissue. To fully understand the rationale behind some of the library preparation options for ancient and historical samples, it is important to understand some basic prerequisites for DNA sequencing on different platforms. In the following sections, we provide a brief overview of the main sequencing technologies currently used and the types of genetic data they can generate. We then discuss the main differences between library preparation methods developed for historical and ancient samples. This practical guide concludes with an overview of the applications of different sequencing strategies to conservation genetics.

Overview of Sequencing Technologies

Before the broad availability of parallel sequencing technologies, Sanger sequencing was one of the most commonly used methods. Developed in the mid-1970s, Sanger sequencing became popular due to its low error rates and straightforward laboratory and bioinformatics processes (and because it really was the only sequencing technique available at the time). This method continues to be effective for characterizing specific and informative genomic regions, such as individual genes or short sequences. Its well-established reliability and affordability make Sanger ideal for small-scale sequencing endeavors that demand precise and accurate results without needing extensive data. Furthermore, Sanger sequencing eliminates the need for intricate library preparation, providing a simple and direct workflow. The only requirement is the development of primers (short nucleotide sequences that bind to the specific genomic region of interest). These primers, combined with a series of enzymatic reactions, are used to produce several copies of the fragment of interest through a process called PCR amplification. This amplification step remains in more recent sequencing technologies, although it does not require the development of specific primers.

Currently, the most popular high-throughput sequencing technologies are Illumina, PacBio, and Oxford Nanopore Technologies (ONT). These technologies use different mechanics and principles, with each company offering a variety of machine models (or "platforms") that can deliver different amounts of data (number of reads and total bases sequenced) per run. Each technology has a unique set of strengths tailored for specific applications, and they can be combined to maximize analytical power. Here, we will expand on the major differences between these technologies in terms of (1) the maximum fragment length sequenced, (2) error rate, and (3) cost (see Table 3 for a summary.) Consequently, the application of these different technologies can vary greatly depending on the project needs and the characteristics of the DNA molecules obtained at the end of the extraction step (such as fragment length and degree of damage).

Fragment Length

In Illumina sequencing, single-end sequencing reads the DNA fragment from one end to the other, providing a single sequence of nucleotides. Paired-end sequencing reads from both ends of the DNA fragment, allowing scientists to obtain two sequences per fragment, one from each strand. Paired-end reads provide more information about the DNA and are helpful for detecting structural variations or overlapping sequences. This extra data makes paired-end sequencing more accurate and useful

for complex DNA analysis. However, when assembling short reads to determine the entire genome sequence, it is common to end up with gaps and incomplete assemblies, known as draft genomes.

More recent long-read strategies produce reads that are tens of kilobases in length, creating overlaps that allow for the generation of complete genome assemblies. PacBio and Oxford Nanopore Technologies (ONT) are the most commonly used long-read sequencing technologies today. PacBio relies on the binding of a slow polymerase to a single molecule of circularized DNA, which is sequenced in real-time multiple times. In ONT platforms, a helicase opens the double strand of the DNA molecule, forcing one of the strands to pass through a very small pore, so the changes in the ionic charge of the different nucleotides can be detected. While ONT can sequence much longer fragments and much faster than PacBio, its accuracy is lower. The requirement of long fragments makes long-read sequencing unrealistic for most museum specimens, and even in archival cryo-collections DNA may exhibit significant degradation that may hinder long-read sequencing success (Salis et al., 2025). Despite the challenges associated with using these technologies on museum specimens successful attempts have been made in rare cases (Quatela et al., 2023; Bein et al., 2025).

Error Rates and Accuracy

Determining the nucleotide order in DNA molecules is the main goal of sequencing. Yet, how can we be sure if the nucleotides being represented are, in fact, correct? There are two types of accuracy: consensus accuracy (is there enough support to determine which nucleotide is present in this position?) and read accuracy (has the machine detected the correct nucleotide in this position?). Understanding some aspects of the sequencing technologies is important to ensure you can distinguish actual biological information from sequencing errors.

The higher the number of unique molecules sequenced for a given sample, the higher the chances of assembling the entire genetic information present in the organism, which is commonly described as high "library complexity." Here, the term library refers to the entire pool of molecules sequenced (the reads) for each processed sample (the libraries). The number of unique reads that overlap in a given genomic position determines the coverage or "sequencing depth" for that position. Coverage of a genomic region can be increased with high library complexity, a higher number of reads sequenced for each unique sample, and by sequencing both strands of the DNA molecule (paired-end sequencing). The higher the coverage, the higher the confidence in determining the exact nucleotide present in that specific genetic position, a bioinformatic step usually referred to as "consensus calling." Consensus calculation will get more computationally demanding with increase in sequencing effort, but it cannot account for errors emerging from the sequencing machine.

The sequencing technologies available today have different inherent error rates for individual measurements. Furthermore, accuracy differs not only between technologies but also across genomic regions, as some stretches of the genome are inherently more difficult to read (e.g., palindromic sequences, AT- and CG-rich regions, or very repetitive regions such as telomeres and centromeres). In such cases, it doesn't matter how many times you sequence a particular region; if there are systematic errors and the sequencing platforms consistently make the same mistake, it will not be improved with higher coverage or more computationally demanding bioinformatics analysis. This is particularly true if the fragments sequenced are short and map to ambiguous positions in the reference genome (see the glossary and "Sequencing strategy" section for more information on reference genomes), such as repetitive regions. One strategy to overcome this problem is to combine short-read with long-read sequencing, which continuously read long DNA fragments multiple times.

In conclusion, our level of confidence in determining the nucleotide order from our sequencing data comes from both isolating potential sequencing errors that the machine makes and ensuring that the reads used to determine the sequence for that particular position can be confidently placed at a given genomic region. These confidence steps are performed during quality control steps in the bioinformatics pipeline, a series of computer-based command-line steps that precede the genetic analyses.

Cost

There is considerable variation in sequencing costs per country, given the availability of trained personnel and sequencing equipment, and the cost of importing reagents can change dramatically. Therefore, before planning a conservation genetics project, it is vital to first obtain quotes from laboratory and sequencing companies and evaluate the available budget.

A major concern when deciding the sequencing effort (how many reads you need to reach the desired coverage) is whether the library constructed for that particular sample has enough endogenous DNA (number of reads that map to a reference genome) and a sufficient number of unique DNA molecules (library complexity). For taxa with large genomes and samples with sufficient DNA yield, if a library lacks complexity and has a high number of duplicate DNA molecules (clonality), focusing on hybridization capture protocols (see Box 4) can be considered a strategy to reduce the cost of sequencing non-desired or highly duplicated molecules.

The cost of sequencing is rapidly changing with advances in sequencing platforms, especially due to the increase in output data (number of reads) that can be generated in a single sequencing cycle. To fully take advantage of the sequencing outputs, a commonly used strategy to minimize costs is to combine the DNA molecules from multiple individuals into a single sequencing cycle, referred to as multiplexed sequencing. This strategy adds minimal labor and costs during the laboratory and bioinformatics steps. Briefly, it requires the addition of unique identifiers in the library preparation step (see next section), which will later be used to bioinformatically sort out the provenance of each molecule ("de-multiplexing" step in the bioinformatics pipeline).

When preparing a budget, it is also crucial to consider the computational resources necessary for data storage and analysis. This includes evaluating the costs associated with acquiring and maintaining hardware, such as servers and storage devices, as well as the expenses related to software licenses and cloud services. As the number of raw reads increases, the size of the files storing these reads also grows significantly. Additionally, the size of intermediate files generated during data processing can escalate quickly, further increasing storage requirements. The budget should also account for the memory needed to store and run analyses efficiently. This involves ensuring that there is sufficient RAM and processing power to handle large datasets and complex computations. Furthermore, specialized personnel may be required to manage and analyze the data, ensuring that the infrastructure can handle the volume and complexity of the data efficiently.

Some questions to guide this step:

1. Given the quality of the DNA obtained, would you be able to use long-read sequencing technologies?
2. How can you minimize the sequencing costs for your project?
3. How many samples can you combine in a single lane if you are multiplex sequencing? Think about the output of different sequencing platforms and the desired coverage per sample.
4. Is there a reference genome that you can use in your study?
5. What is the genome size of your focal species?

Table 3: Summary of the main differences, advantages, and challenges of the different sequencing technologies most commonly used in conservation genetics studies. Units: bp = base pair, Kbp = kilo-base pair, Mbp = megabase pair. (* "2x" denotes pair-end sequencing)

		High Throughput Sequencing	
	Sanger	**Short-read sequencing (Illumina)**	**Long-read sequencing (PacBio; ONT)**
Fragment length	500–1000 bp	2x150 bp or 2x300 bp (*)	10–25 Kbp (PacBio); 1–3 Mbp (ONT)
Error rate	0.01%	0.1–0.5%	As of 2024, 0.1% for PacBio HiFi sequencing, and errors are a bit higher for ONT
Advantages	More affordable and extremely efficient if you are focusing on a single or few genetic regions of interest. Because it has the lowest error rate, it manages repetitive sequences well, but can be challenging to determine the number of bases in long tails. Very easy to analyze the data, given that there is well-established software and community support.	Extremely cost-effective, especially when sequencing multiple individuals. Sequencing at higher coverage allows high sensitivity and confidence in determining the presence of rare and low-frequency loci. Because it has been used by the community for so long, it has extremely well-established laboratory and bioinformatics pipelines.	Longer fragments have a greater chance of containing enough unique information to anchor them properly (position and direction) in the genome. Can resolve repetitive regions. Some ONT platforms are extremely fast and portable, allowing sequencing in the field.
Challenges and disadvantages	Requires developing multiple primers and performing multiple PCR amplifications. Lowest scalability when compared to high-throughput sequencing strategies. Consequently, has a much higher price per base pair sequenced if you need to sequence a lot.	The fragment size is a limiting factor in assembling complete reference genomes. Given that most historical/ancient molecules have some level of breakage, this is not such an issue. Struggles to determine the precise number of identical consecutive nucleotides.	It requires extremely high-quality DNA (unfragmented), and the cost/sample can be extremely expensive. Pipelines are not well-established yet. Bioinformatics is computationally demanding and can be laborious.

Choosing a Library Preparation Protocol that Suits Your Samples

Compared to modern samples, the DNA molecules recovered from historical and ancient samples tend to be much shorter. Consequently, short-read technologies (e.g., Illumina) are a more appropriate choice than long-read sequencing. Given that Illumina sequencing is the most commonly used technology for short reads, this section will focus on some details of the library preparation protocol options available for sequencing on this platform (see Table 4 for a summary.)

One of the major goals of library preparation is to prepare the DNA molecules to be read by the sequencing machines. In Illumina platforms, this preparation involves adding adaptors to each DNA molecule, allowing them to bind to a glass flow cell where bridging amplification occurs. Most projects maximize the sequencing output of Illumina machines by opting for multiplexed sequencing of libraries from various individuals. In this case, the reads from each individual library must also be associated with its sample identity. To achieve this, a unique combination of oligos (6–8 nucleotide

sequences, sometimes referred to as unique "barcodes") are added to the end of every DNA molecule of each individual. These barcodes are sequenced together with the DNA molecules they are bound to, and thanks to their unique combination of nucleotides, every read from a particular individual can be isolated bioinformatically (in a process called "demultiplexing"). Combining multiple individual libraries in multiplex sequencing is of course constrained by the desired coverage, the size of the genome, DNA extraction yield, and sequencing strategy. For instance, DNA extraction from a single small-sized insect tends to yield much less DNA than that from tissue of a large mammal. In this case, the library preparation must account for the low DNA yield.

When working with historical and ancient samples, the DNA yield, size of fragments, and the level of degradation present in the DNA molecules will determine the best choice for library preparation. Most library preparation methods optimized for historical and ancient samples have modifications that account for the fact that these samples usually result in (1) much lower input DNA concentrations, (2) shorter and damaged fragments, and in many cases, (3) DNA that is either single or double-stranded. To overcome the challenge of dealing with low-input DNA, most modified library preparation protocols include a few rounds of PCR amplification after the library preparation to increase the number of molecules that can be sequenced. However, this strategy comes with the cost of dramatically increasing the degree of clonality in the library, meaning that most DNA molecules are exact copies of other molecules (or duplicates). While this strategy can help you proceed from library preparation to sequencing, it does not increase the power of completing the entire genomic sequence of an organism. For recommendations on the number of cycles, see Meyer & Kircher (2010).

As for fragment size, the DNA molecules found in high-quality, modern samples are almost intact. If researchers seek to sequence them on a short-read sequencing platform, shearing to the ideal fragment size (~150 or 300 bp) by physical fragmentation (via sonication, or sound waves) or enzymatic fragmentation is the first step of the library preparation. Because the DNA recovered from ancient and historical samples is already extremely fragmented (35–250 bp), they do not require fragmentation prior to library preparation. Moreover, the two strands of shorter DNA molecules separate from each other (a process called "denaturation") at lower temperatures than longer DNA molecules. Therefore, the enzymes used in these protocols have activation temperatures different from those of traditional library preparation protocols. Lastly, many DNA fragments can be lost during laboratory preparation. This can be addressed with protocol optimizations such as using lo-bind tubes (which minimize the affinity of DNA molecules to bind to the inside walls of the tubes), minimizing the number of tube changes (i.e., preparing multiple reactions inside the same tube), adapting the purification step to increase the ratio of binding buffer, and using silica columns that retain shorter fragments (e.g., Qiagen QiaQuick and Qiagen MinElute columns).

Unlike DNA recovered from modern tissues, ancient and historical DNA contains a series of damage signatures, such as the deamination of cytosine that is read by the sequencing machine as uracil, and the formation of single-strand overhangs. Some library protocols can remove the most commonly found miscoding damage (i.e., removing the deoxy uracils using uracil-DNA glycosylase (UDG) and endonuclease VIII (Briggs et al., 2009)); while others preserve these damage patterns using full or partial UDG methods (Rohland et al., 2018), which can be extremely useful to authenticate DNA sequences from human remains, for instance. Lastly, some DNA extraction protocols involve a hot-treatment step (e.g., Hot Alkali Treatment, specifically designed to remove blocking lesions from formalin-preserved specimens). This treatment leads to the separation of the two strands of DNA (a process called "denaturation"), so in this case, a single-strand DNA library preparation method may be required. In general, historical samples from other types of preservation methods tend to respond

Table 4: Comparison of double-stranded (ds) and single-stranded (ss) DNA library protocols currently available for Illumina sequencing of historical and ancient libraries.

	dsDNA libraries	ssDNA libraries
Fragment size	40–500 bp	30–120 bp
Initial amount of input DNA	Some protocols can work with low-input	Ideal for very low input
Allows input of denatured DNA?	no	yes
Keep the information in overhangs?	some protocols	yes
Cost and time consuming	moderate	high
Protocol intricacy	moderate	high

well to double-strand DNA library preparations, but more fragmented samples perform better with a single-stranded library preparation protocol.

Here are some of the most commonly used library preparation protocols for ancient and historical samples:

- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010(6), pdb-prot5448. https://doi.org/10.1101/pdb.prot5448
- Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S., 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(2), 410–419. https://doi.org/10.1111/2041-210X.12871
- 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. https://doi.org/10.1038/nprot.2013.038
- 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(3), 241–249. https://doi.org/10.1093/jhered/esab012

Here are some of the common uracil-DNA glycosylase (UDG) treatment methods:

- Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2009). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research, 38(6), e87. https://doi.org/10.1093/nar/gkp1163
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1660), 20130624. https://doi.org/10.1098/rstb.2013.0624

Some questions to guide this step:

1. What is the average size of fragments obtained after the DNA extraction?
2. After the DNA extraction, are the DNA molecules denatured?
3. Would your project benefit from the information found in the original information from the unpaired nucleotides often found in degraded molecules?

Sequencing Strategy

In any project design, it is vital to ensure that the research question can be answered with the data in hand. When choosing a sequencing strategy for a conservation genomics project, researchers must consider four major factors: (1) genomic resolution, (2) coverage level, (3) number of individuals/samples required to answer the research question, and (4) cost of sequencing. Research projects are heavily constrained by budget limitations, making the cost of sequencing a critical factor that can limit the possibilities of a genomic study. The differences between the main sequencing strategies in terms of genomic resolution achieved, number of individual samples usually employed, and cost of sequencing are summarized in Figure 5.

The highest level of genomic resolution available today is provided by reference genomes. The quality of a reference genome increases with the level of assembly and annotation. A reference genome can be generated using single or multiple tissues from one individual or by combining data from multiple individuals. Most questions addressed using reference genomes focus on understanding gene functions and the evolution of genomic structure at a comparative level, which relies on well-assembled and annotated references. To increase the accuracy of these references

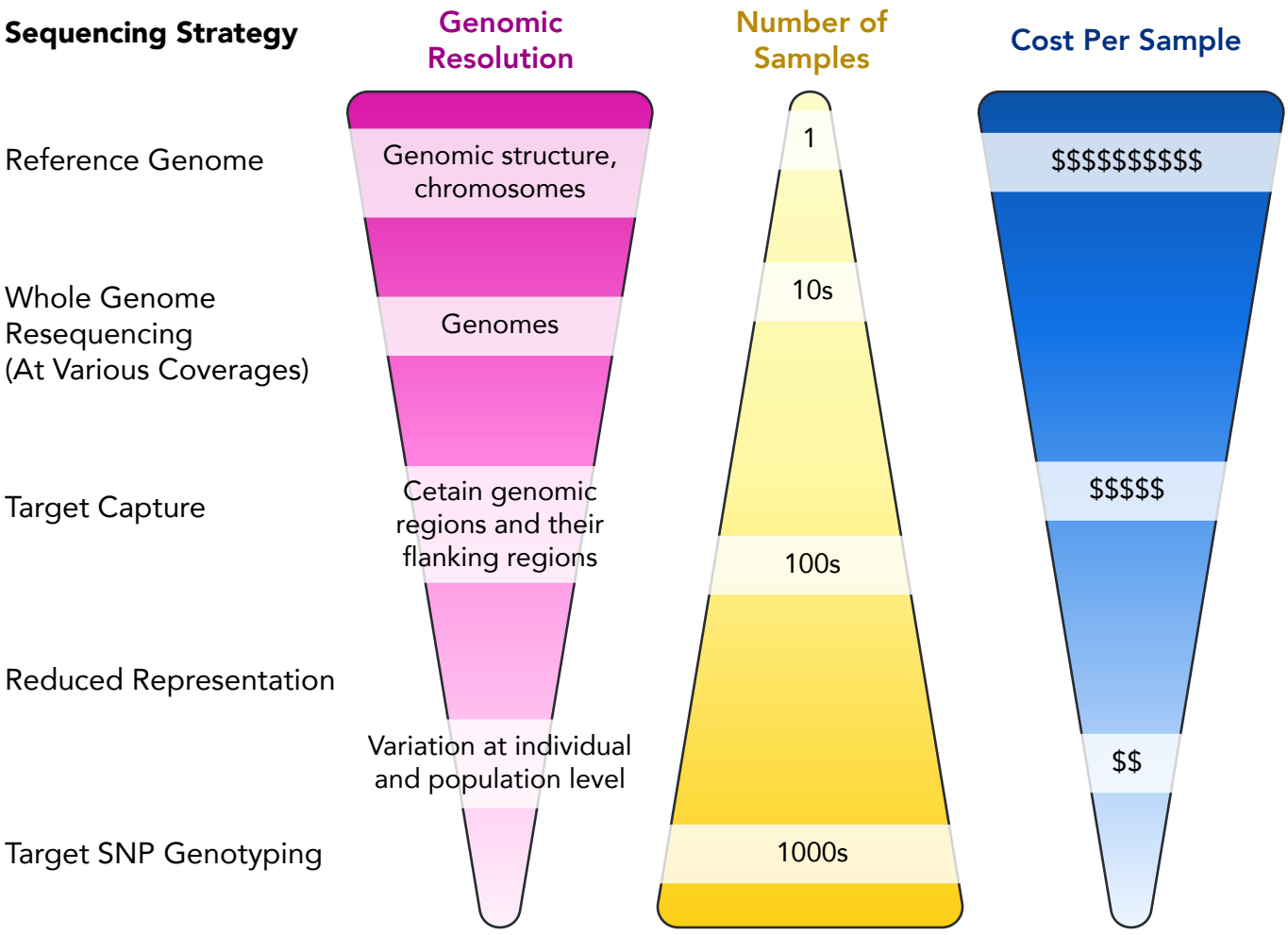

Figure 5. Sequencing strategies can have drastically different costs per sample. This simplified diagram indicates the differences among the sequencing strategies most commonly used in conservation genetic studies when it comes to genomic resolution, number of individual samples, and cost per sample. Image credit: Anna Penna. Figure modified from Threatened Species Initiative Training Module 5.1 (available at https://threatenedspeciesinitiative.com/training-modules/).

(determining sequences, presence of variation, number of copies, relative position, and directionality of challenging genomic regions), researchers usually combine different sequencing strategies (i.e., short and long reads). Consequently, generating high-quality reference genomes can be extremely costly, laborious, and computationally demanding. Very few conservation genetics projects have the necessary funds to generate a well-annotated reference genome. However, some conservation and population genetic analyses indirectly rely on the availability of a good reference genome, given that other sequencing strategies rely on these references for mapping. One way to minimize the costs and labor of generating a reference genome is to incorporate publicly available data, such as the resources available on NCBI and DNAZoo. International collaborative efforts, such as sequencing consortia, are also improving the applications of genetic tools beyond basic research by combining expertise and resources to generate high-quality sequencing data.

As mentioned above, the choice of sequencing strategy depends on the question the project seeks to answer, the required number of samples, and budget availability (Figure 5). The level of DNA degradation in the samples available is another factor that can limit the applicability of some sequencing strategies. For instance, if DNA fragments are too short, restriction-enzyme approaches to generate reduced representations of the genome might result in shorter fragments that won't pass size selection. Even moderately fragmented data can jeopardize the success of long-read sequencing approaches. DNA degradation is strongly related to sample age, but also to the preservation method (what type of chemicals it was exposed, at what buffer and temperature it was stored in, how many times it was thawed, and so on).

In conservation genetics, most questions revolve around the levels of genetic diversity at a population level. Therefore, the required level of genomic resolution is much lower, whereas the number of individuals sequenced needs to be much greater. In this case, approaches such as whole genome resequencing (including at lower coverage), reduced representation or target SNP genotyping are more appropriate. Most reduced representation approaches (e.g., restriction site-associated DNA sequencing and target capture) do not require a reference genome in the bioinformatics pipeline, whereas whole genome resequencing approaches rely on de novo assembly or read mapping to a reference, especially when not sequencing at higher coverage. Therefore, if you opt for whole genome resequencing, it is important to first check the availability of same-species (or phylogenetically close-enough) reference genomes in public repositories (e.g., NCBI).

Another cost-effective approach that remains useful in conservation studies is the sequencing of organellar genomes (mitochondrial DNA or chloroplast DNA) as well as genotyping microsatellites (2-6 base-pair long tandemly repeated DNA sequences found across the genome). These markers can be amplified using traditional PCR methods and are relatively inexpensive compared to large-scale genomic representations. Organelle DNA in particular is found in abundance in the cells, so the protocols require smaller amounts of DNA, making them suitable for studies with limited funding or poor-quality samples. Organelles can be used to identify maternal lineages and infer historical demographic events, whereas microsatellites offer high variability for detecting fine-scale population structure, kinship, and inbreeding. However, both organelle and microsatellites represent only a small fraction of the genome and may not capture genome-wide patterns of variation or selection, which can limit their resolution in some evolutionary or conservation questions. Another disadvantage of microsatellites is that the development of new markers for non-model species can be costly and time-consuming and the amplification success is considerably reduced for degraded material.

Population-level questions (e.g., population structure, gene flow, inbreeding, population size, etc.) often

require good measures of allele frequencies in the populations analyzed. For such a detailed analysis, genetic information from dozens to hundreds of individuals is required. Traditional studies have used a combination of PCR-amplified markers (from nuclear DNA or organelles, such as the mitochondrion and chloroplast) to assess the levels of genetic diversity of the population. With the advent of methods that can track the genealogical history of thousands of loci (e.g., coalescent-based methods), some of the challenges of sequencing numerous individuals can be overcome by using information from thousands of loci from one or a few individuals. Recent and rapid advancements in sequencing strategies have allowed researchers to characterize the frequency and history of genetic diversity by obtaining genealogical histories from more genomic regions by using reduced representation or whole genome resequencing approaches. Conversely, other projects seeking to resolve phylogenetic relationships between species or genera, for example, can use genetic data from much fewer samples (e.g., as little as one sample per taxonomic group). A common strategy in phylogenomic projects that rely on museum specimens to characterize the genetic information of rare or hard-to-sample taxa is to apply hybridization capture to amplify genetic data from conserved regions of the genome and their flanking regions (see Box 4). This capture strategy has been a successful and cost-effective method of answering both population genetic and evolutionary questions.

Box 4 provides a summary of what sequencing strategies of various genomic outputs and resolutions can add to a conservation genomic study, including a few advantages, disadvantages, and the technologies most commonly used.

Box 4: Sequencing strategies commonly used in conservation projects

Reference genome

- A fully assembled genome that can be used as a representation of an individual, population, species, or other taxonomic group. The assembly level of a reference genome can vary a lot depending on the sequencing technology used and the annotation quality. To improve the quality of the reference genome, researchers combine short and long-read sequencing and employ a series of bioinformatics analyses that can predict the function of genomic regions (i.e., functional and structural annotation of assemblies).
- Technologies used: Often requires a combination of short read with deep sequencing (coverage >60x), long-read (PacBio, Hi-C), followed by genome annotation.
- What can be done with it?
  - Identify genes or specific mutations associated with specific traits or diseases.
  - When analyzed at a comparative level (i.e., between reference genomes), it can be useful to determine the genetic basis of certain traits/adaptations.
  - Compare the evolution of genomic structure (e.g., structural variants, number, and direction of gene copies, relative positioning of genes in chromosomes) across the tree of life.
  - A useful resource for reference-based mapping (e.g., whole-genome resequencing)
- Disadvantages:
  - Still extremely expensive and laborious in the laboratory further and bioinformatic analyses.

Whole genome resequencing (WGrS; also known as genome skimming or WGS)

- In this approach, the entire genome of an individual is sequenced at a low (>5X), medium (5–30X), or high coverage (30X or higher), depending on the question and budget available. The reads are then mapped against and compared to sequences of a known reference genome.
- Technologies used: short-read sequencing.
- What can be done with it?

  - Can provide a good resolution of genetic diversity among individuals at various genomic regions.
  - Can be used to obtain SNPs, insertions, deletions, gene rearrangements, and calculate important conservation genetics parameters, such as heterozygosity (genetic diversity), runs of homozygosity (levels of inbreeding), gene flow (connectivity between populations), and genetic load (levels of genetic mutations).
- Disadvantages:
  - Analyzing whole genome data for multiple individuals can be computationally demanding (time of analysis and memory required to store and analyze the data). It can yield many continuous stretches of DNA sequence that are ordered and oriented using paired-end or long-read data, but may still contain gaps. Can cause difficulties and not be cost effective for samples that have low endogenous content of the target species (e.g., samples that are highly contaminated in the case of ancient DNA and historical DNA, or samples where multiple species may be present such as fecal samples and other non-invasive samples), where the majority of sequenced DNA will be from off target species.
- Warning:
  - Assembly can be done *de novo* if the sequences are not fragmented, have extremely high quality, and are sequenced at high coverage. Reads can also be mapped to a reference genome. If the divergence between the reference and the reads is too high (i.e., due to phylogenetic distance, for instance), mapping biases may be present in downstream analyses.

Reduced-representation sequencing approaches

- The sequencing of a subset of the genome by using a combination of restriction enzymes that cut the DNA molecules at specific sequence combinations, followed by a size selection step to get more uniform fragments in the whole genome. Because only digested fragments are sequenced, this approach greatly reduces the complexity of the genome.
- Technologies used: various combinations of restriction site enzymes followed by short-read sequencing of multiple libraries with unique identifiers combined ("multiplexed sequencing") usually in Illumina platforms.
- What can be done with it?
  - A cost-effective way to generate population screens (SNPs) for hundreds of specimens.
  - Can be used to characterize genetic structure, kinship, relatedness and inbreeding levels, and gene flow, all useful in delimiting conservation units and assessing population connectivity across the landscape.
- Disadvantages:
  - Does not sequence variants across the entire genome, and coding regions might be sequenced only partially. Mapping to a reference genome can increase accuracy, but it is also possible to use de novo approaches.
- Warning:
  - Combining reads obtained from different digestion experiments (i.e., that used different enzymes or fragment selection) can lead to excessive missing data that might bias phylogenetic and population clustering results.

Hybridization capture approaches

- The sequencing of a subset of the genome using artificially designed RNA baits. These baits are small oligos that hybridize with specific regions of the genome and are then captured using magnets
- Technologies used: qPCR, bait hybridization followed by capture.

- What can be done with it?
  - Baits can either be designed for a specific project (e.g., target-SNP genotyping) to detect specific genetic variants used in disease screening, detection of pedigrees and hybrids; target specific genetic regions that are transcribed into proteins and some associated regulatory regions (e.g., whole exome sequencing); or target highly conserved regions of the genome that allow comparisons across higher phylogenetic scales (e.g., Ultra Conserved Elements, UCE).
  - Can also be applied in noninvasive and museum samples to amplify specific genomic regions at a desired coverage, reducing the chances of sequencing DNA molecules from exogenous sources.
  - Can be used on highly contaminated samples and samples with multiple species present to specifically target and enrich the DNA of the target species.
- Disadvantages:
  - Requires pre-existing knowledge of the genome to bioinformatically develop the baits.
- Warning:
  - Indirectly, a well-annotated reference genome will be required to use bioinformatics to design the baits. Designing and acquiring the baits can be expensive, which might limit the number of samples sequenced.

Decision Tree to Guide a Museomics Study Design

Now that we covered some of the main steps between the planning of the tissue sampling and the generation of genetic sequences, Figure 6 provides a decision tree that can help the logical reasoning process required to design a museomics study. This decision tree should make the overall planning process more organized and transparent, for instance, when selecting potential source of data and appropriate methods. Use this decision tree to identify potential challenges and alternatives at each step, anticipate problems and plan solutions.

Bioinformatic Considerations

The analysis of sequencing data from ancient and historical samples shares many commonalities with the analysis of data from modern samples, including processing of raw sequencing data (quality control, demultiplexing, adapter trimming, and quality filtering), mapping of reads against reference genomes or databases, post-processing (removal of duplicates, quality filtering of map reads), and variant calling. However, some extra steps and modifications in the bioinformatics pipelines must be included due to challenges arising from the lower sequencing depth, and higher levels of fragmentation, degradation, and contamination found in the sequencing reads obtained from DNA of non-modern samples (Orlando et al., 2021).

In the bioinformatic analysis of ancient and historical DNA, degradation, fragmentation, and contamination are critical factors that must be meticulously addressed. Degradation results in highly fragmented DNA, which can impact the performance of some mapping algorithms. Specialized alignment tools and protocols can more accurately map these shorter reads to the reference. Aligning reads from historical and ancient DNA to a reference genome can introduce reference bias, where reads containing reference alleles are preferentially mapped over those with alternative alleles. These degraded DNA reads also exhibit specific post-mortem damage patterns, such as cytosine deamination, which must be accounted for during analysis. Tools like MapDamage (Jónsson

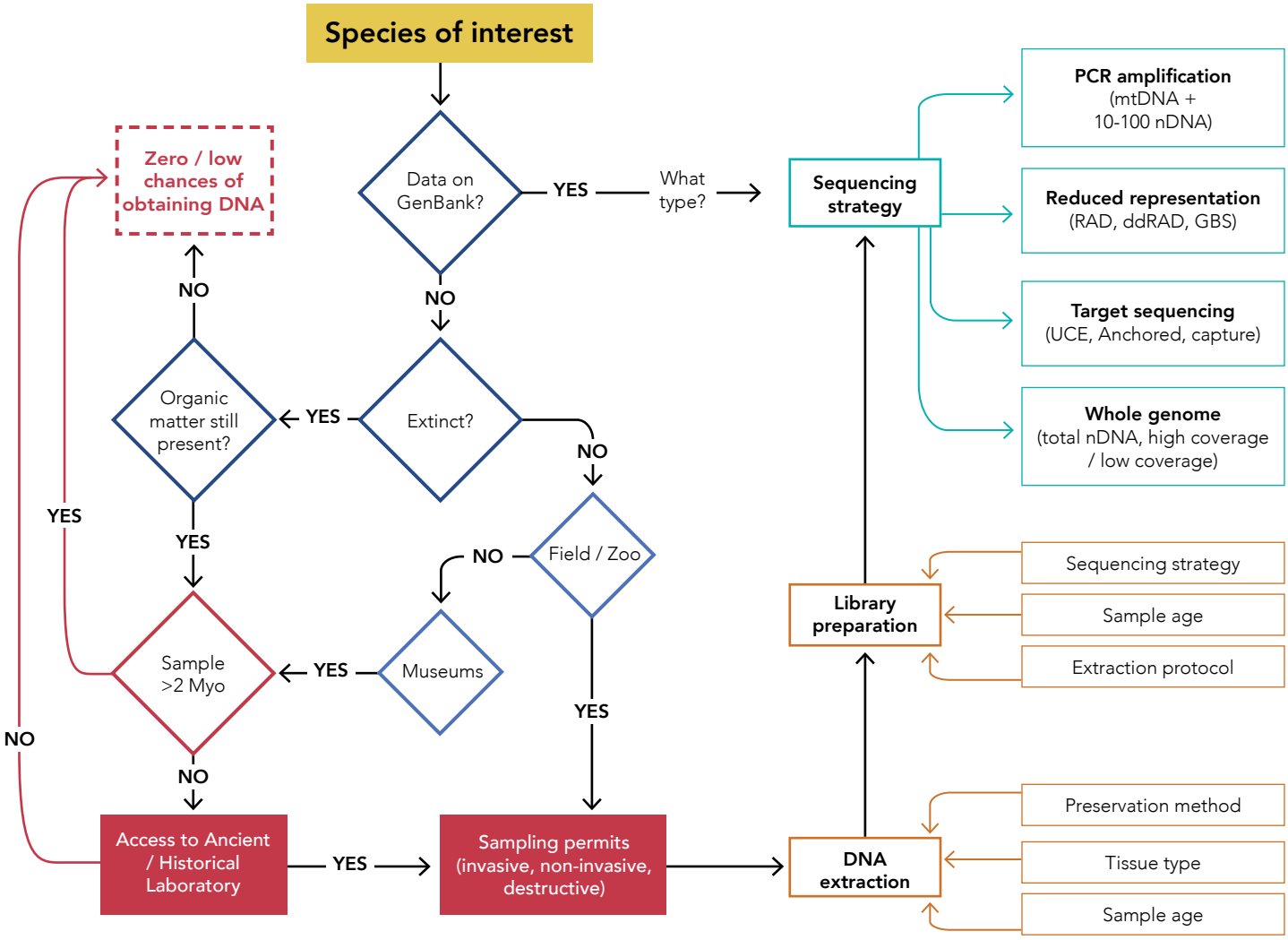

Figure 6. Decision tree to assist in the experimental design of genetic study that incorporates tissue from museum specimens. Image credit: Anna Penna.

et al., 2013) and DamageProfiler (Neukamm et al., 2021) can be used to identify and correct for characteristic post-mortem damage patterns, such as cytosine deamination, which can also serve as a marker to authenticate ancient DNA by distinguishing it from modern contaminants. Contamination from environmental sources and modern DNA requires stringent control measures and computational methods to ensure the authenticity of the ancient DNA. These steps are essential to ensure the reliability and accuracy of genetic data derived from ancient and historical samples, allowing researchers to draw valid conclusions about past populations and species.

The main specialized bioinformatics pipelines currently available for the analysis of genetic data from degraded DNA include EAGER (Peltzer et al., 2016) and PALEOMIX (Schubert et al., 2014), which are specifically tailored for ancient DNA analysis and offer a series of tools to perform the main steps of the bioinformatics data quality control and downstream analysis, including mapping reads to a reference genome, removing contaminants, damage pattern analysis, and various other downstream analyses. Often, the amount of DNA recovered from museum and ancient samples is insufficient to generate high-depth sequencing, either due to the low initial volume of DNA template or due to contamination from exogenous sources. Therefore, the analysis of non-modern samples can also benefit from pipelines designed to maximize the utility of low-coverage, such as ANGSD (Korneliussen et al., 2014) and loco-pipe (Zhou et al., 2024) which incorporate probabilistic frameworks

and best practices for handling genotype uncertainty and other challenges (Lou et al., 2021). These pipelines also have a versatile tool-set for population genetics analysis used to investigate population structure. Another pipeline useful for genetic diversity analysis is GenErode (Kutschera et al., 2022), which is designed to investigate genome erosion in endangered and extinct species and can process reads from modern and ancient samples.

The downstream results generated from these bioinformatics pipelines—such as estimates of genetic diversity, gene flow, and relatedness—are essential for answering key research questions in conservation genomics, evolutionary biology, and phylogenetics. For example, measures of genetic diversity can reveal whether a population is at risk of inbreeding or loss of adaptive potential, while gene flow analyses can help determine whether isolated populations are genetically connected or fragmented. Patterns of relatedness between specimens—especially when integrated with spatial and temporal data—can provide insight into historical population structure, range shifts, or even undocumented extinctions. These insights are critical for understanding how species and populations have responded to past environmental changes, informing conservation strategies for threatened taxa today. Moreover, in phylogenetics and biodiversity studies, historical and ancient genomes can help resolve the evolutionary relationships of extinct or elusive species, refine species boundaries, and provide a deeper temporal perspective on diversification processes.

In Summary

Museomics leverages genetic material from museum specimens, allowing researchers to obtain DNA from a wide range of organisms, including those that are rare, endangered, or extinct. Advances in DNA extraction and sequencing technologies have enabled the recovery of genetic information from historical and ancient samples, despite challenges like low DNA concentrations and fragmentation. By combining different sequencing platforms and samples from modern and historical sources, researchers can obtain valuable insights into the genetic diversity, population structure, and evolutionary history of various species. These insights may then be applied to species conservation by identifying at-risk populations, informing evidence-based management strategies, tracking genetic health, mitigating inbreeding, and enhancing the resilience of species to environmental changes. Consider reading the NCEP Module, "Conservation genetics" (https://doi.org/10.5531/cbc.ncep.0123), for more on the foundational concepts and applications of conservation genetics.

GLOSSARY

- Admixture: The cross between two individuals of a separate species or population whose offspring could result in a new hybrid population. This is not to be confused with introgression/hybridization in which backcrosses among individuals of separate populations are continuous. Admixture, on the other hand, occurs as a one-time event or through multiple generations and increases genetic diversity through the acquisition of new mutations.
- Allele(s): Alternative forms of a gene. These typically refer to changes in protein-coding regions (genes) of the genome that differ and can thus change the phenotype of the individual.
- Carrying capacity: The maximum number of individuals within a population that an environment can sustain.
- Bioinformatics pipeline: A series of computer instructions (scripts) that are deployed to execute several tasks by different programs. In Genomics, these computer programs are specifically designed to implement a set of complex algorithms to process sequence data.
- Chromosome: A single molecule of DNA and associated proteins (e.g., histones that coordinate the condensation of the chromatin, and other proteins responsible for the binding of transcription

factors). Convenient packaging of genetic information. The DNA molecules are really long, so they must be compacted in a way to fit inside the cell. This level of organization is crucial during cell division (ensures the genetic material will be divided in a proper way), and to minimize the exposure of the DNA molecules, unpacking it only in regions needed for cellular processes.

- Coverage: The number of unique reads that overlap in a given genomic position will determine the "coverage" or "sequencing depth" for that position. The higher the coverage, the higher the confidence to determine the exact nucleotide present in that specific genetic position or the "identity" of that position, in a step that is usually referred to as "consensus calling."
- Deleterious mutations/alleles: Alleles, or alternate forms of a gene, that are typically recessive. An accumulation of deleterious alleles in the genome can decrease the fitness of individuals within the population and, thus, can reduce the effective population size within a population or species over time.
- Deamination/deaminated sequences: The loss of an amine group that typically results in the misidentification of a cytosine base as a uracil base. This is a common source of ancient DNA damage that can also be used to authenticate genetic information produced.
- De novo approaches: A technique used to sequence and assemble the complete genome of an organism without the use of a reference genome. This approach is particularly useful for studying species with no prior genomic information, as it reconstructs the genome sequence by iteratively aligning the reads in the dataset to each other.
- Endogenous DNA (Exogenous DNA): The term is used to differentiate the genetic material of the target species (endogenous) from that belonging to parasites, fungi, bacteria, other microorganisms, and contaminant DNA from sources such as researchers and laboratory reagents (exogenous). Determining the identity of the DNA molecules obtained after extraction requires sequencing and bioinformatic analysis.
- Evolutionarily significant units (ESU): A population or group of organisms that is defined as distinct for the purposes of conservation. This term is typically used to more broadly classify a distinct group that cannot be classified as a subspecies or population, but still garners the need for conservation protection.
- Exome: The portion of the genome that consists of the coding regions of genes responsible for producing proteins. Whole exome sequencing is a technique used to sequence all the exons in a genome, allowing the identification of genetic variations that may be linked to rare genetic disorders and understanding the genetic basis of complex traits. It provides an efficient way to focus on the most functionally important parts of the genome.
- Fitness (biological): The ability of an individual (or sometimes in reference to a population or species) to produce viable offspring and thus pass on its genetic material. This can be referred to either with respect to an individual's genotype or phenotype.
- Fixed mutations/alleles: A gene variant that has reached a frequency of 100% in a population has become 'fixed.' Whether the gene variant will be present in all members of the population (again, fixation) or lost entirely, meaning no members of the population possess this allele, depends on how that mutation is selected.
- Gene: A region of the genome that is transcribed into RNA plus the associated regulatory elements required for that transcription. Some examples are protein-coding genes, regulatory genes, tRNA, mRNA, or rRNA genes.
- Genetic load: A decrease in the fitness of individuals in a gene pool on average due to the accumulation of deleterious alleles.
- Genetic rescue: A conservation strategy aimed at increasing the genetic diversity within a population by introducing new genetic variation (individuals with unique genomes).
- Genetic differentiation/genetic distance: A measure of the difference in the frequency of alleles

between different populations due to structure. Often measured using the fixation index (Fst), which ranges from 0, which means complete sharing of genetic material between populations, and 1, no sharing of genetic material.

- Genome: One complete set of chromosomes (sex chromosomes and autosomes), including the sequences of all genes, introns (non-coding), exons (coding regions), and organelle genetic material (mitochondria and chloroplasts). In humans, the entire genome has around 3.3 billion base pairs (bp), and the mitochondria has around 16 thousand bp.
- Genome assembly: The process of putting together the entire order of nucleotides that constitute the genome of a given individual, Ideally, all the reads that were sequenced but do not belong to the organism's own genome should be discarded. This process can be done using a reference genome, though mapping the reads sequenced against a reference sequence (see below); or de novo, in which the entire genome is assembled by finding overlaps between the multiple reads sequenced. De novo approaches are good to detect structural variants, but require very high-quality raw read data and can be very computationally demanding. The level of assembly of the raw sequences varies from contigs (continuous overlapping regions) to chromosomes (when the reads can be spatially attributed to different DNA molecules); and can include annotation information (see below).
- Genome annotation: Obtaining quantitative and qualitative information from the whole genome sequenced. Some of the quantitative metrics include the proportion of GC content, the contiguity between fragments, the average coverage, number of genes or gaps identified. The qualitative information are specificities about what the genomic regions represent (genes, regulatory regions, telomeres, repetitive elements, and so on).
- Genomic/genetic erosion: The collective damage to a species' genome or gene pool due to lack of genetic diversity that are characterized by an increase in maladaptation, genetic introgression, and/or increased genetic load.
- Heterozygosity: Likelihood that a pair of randomly selected alleles will be different (i.e., the proportion of individuals that are heterozygotes in a population at a particular gene or region of the genome.)
- Inbreeding depression: A reduction in the biological fitness of an organism due to an accumulation of (mostly recessive) deleterious mutations that results from the offspring of parents that are too closely related to one another. This term is commonly used in reference to a population or species in which there are so few total individuals in which inbreeding and, thus, loss of genetic diversity presents itself. The term can be used in contrast with outbreeding depression, which is a reduction in biological fitness when individuals from separate species, strains of locally adapted populations cross.
- Introgression/introgressive hybridization: Typically referred to as introgression, introgressive hybridization, or hybridization. This refers to backcrosses between populations that are distinct but gene flow between populations is present.
- Mapping (reads): A bioinformatics step that consists of aligning short DNA sequences (called "reads") obtained from sequencing to a known reference genome. This step helps identify the placement of each read in the genome, allowing researchers to detect variations like mutations or structural differences.
- Microsatellites: Repetitive regions throughout the genome that are well-characterized and used to study the genetic variation of closely related species.
- Multiplexed sequencing: When DNA libraries from multiple samples are combined and sequenced together in a single run, using unique molecular "barcodes" to distinguish between the samples. This approach saves time and costs, allowing researchers to analyze many samples simultaneously without losing track of which data belongs to which sample.

- Nucleotide diversity: Average number of pairwise differences between all different pairs of DNA sequences in the population or sample.
- Reference genome: An entire assembled genome that can be used as a representation of the genome of an individual or species. The references can have different levels of assembly and annotation quality. Used to map raw reads against.
- Shearing (DNA): Breaking down long strands of DNA into smaller fragments. This step is necessary for short-fragment sequencing technologies (e.g., Illumina) of fresh tissue samples. Often, DNA shearing is done by sonication, a method that uses high-frequency sound waves to create vibrations that disrupt the DNA molecules, resulting in random shearing into fragments of varying sizes. This step must be completed before library preparation.
- Size selection: A lab method in which a specific DNA size fragment (e.g., 250 bp) is isolated from genomic DNA and other unwanted fragment sizes. This is typically performed as a part of library preparation, a process by which samples are fragmented and uniquely "labeled" in order to be successfully processed by NGS technologies and high throughput sequencing. Size selection is typically achieved using magnetic beads, instruments like the BluePippin, and gel extraction.
- SNPs: SNPs (or single nucleotide polymorphisms) are variations among individuals in a single nucleotide at a specific position within a genome. These are used to study phylogenetic relationships between individuals of different populations or species.
- Target SNP genotyping: A technique used to identify and analyze specific single nucleotide polymorphisms (SNPs) in a genome. Instead of sequencing the entire genome, this method focuses on particular SNPs of interest, which are often associated with traits, diseases, or evolutionary markers, allowing for a more cost-effective and efficient genetic analysis.
- Ultra Conserved Elements (UCE): Regions of the genome that show little to no variation across different species. UCEs are often found in noncoding regions of the genome but can play important roles in gene regulation and development. Although their exact role is not fully understood, these elements and their flanking regions (which are more variable) are useful for studying evolutionary relationships and tracing the genetic history of organisms.
- Whole genome sequencing: The process of generating the sequences (or raw reads) that will be used to assemble a given genome. Through the combination of different sequencing technologies (short and long reads), it is possible to better characterize the different parts of the genome. The higher the uniqueness of the DNA molecules and the number of reads used to assemble the genome, the higher the confidence in determining the sequence and relative positioning of the DNA molecules that compose the genome.
- Whole genome resequencing: When the entire genome of an organism is sequenced, typically to identify genetic variations such as single nucleotide polymorphisms (SNPs), insertions, deletions, or structural variants. Unlike de novo sequencing, where the genome is sequenced for the first time, resequencing compares the sequenced genome to a reference genome to detect differences.

BIBLIOGRAPHIC RESOURCES

The following readings provide more background on the field of museomics and applications of genetic tools to conservation.

General Introduction to Museomics

- Barnett, R. (2018). Zooarchaeology and ancient DNA. In The Encyclopedia of Archaeological Sciences. https://doi.org/10.1002/9781119188230.saseas0623
- Blair, M. E. (2024). Conservation museomics. Conservation Biology, 38(3) e14234. https://doi.org/10.1111/cobi.14234
- Buerki, S., & Baker, W. J. (2016). Collections-based research in the genomic era. Biological Journal of the Linnean Society, 117(1), 5–10. https://doi.org/10.1111/bij.12721

- Card, D. C., Shapiro, B., Giribet, G., Moritz, C., & Edwards, S. V. (2021). Museum genomics. Annual Review of Genetics, 55(1), 633–659. https://doi.org/10.1146/annurev-genet-071719-020506
- Nakahama, N. (2021). Museum specimens: An overlooked and valuable material for conservation genetics. Ecological Research, 36(1), 13–23. https://doi.org/10.1111/1440-1703.12181
- Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: Advances and challenges in museomics. Trends in Ecology & Evolution, 36(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Applications of Genetics and Genomics to Conservation

- Hogg, C. J. (2024). Translating genomic advances into biodiversity conservation. Nature Reviews Genetics, 25(5), 362–373. https://doi.org/10.1038/s41576-023-00671-0
- McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? Evolutionary applications, 7(9), 999–1007. https://doi.org/10.1111/eva.12193
- Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. Genome Biology, 19, 1–12. https://doi.org/10.1186/s13059-018-1520-3
- Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukeiwickz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., … & Zammit, G. (2023). How genomics can help biodiversity conservation. Trends in Genetics, 39(7), 545–559. https://doi.org/10.1016/j.tig.2023.01.005
- Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M., & Hoffmann, A. A. (2022). Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. Proceedings of the National Academy of Sciences, 119(1), e2105076119. https://doi.org/10.1073/pnas.2105076119

Classic Examples Studies Using a Museomics Approach

- 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. https://doi.org/10.1038/312282a0
- Kjaer, K. H., Pedersen, M. W., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavic, S., Ruter, A. H., Schmidt, A. M. A., Kjeldsen, K. K., Tesakov, A. S., Snowball, I., Gosse, J. C., Alsos, I. G., Wang, Y., Dockter, C., Rasmussen, M., Jorgensen, M. E., … Willerslev, E. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature (London), 612(7939), 283–291. https://doi.org/10.1038/s41586-022-05453-y
- Mármol-Sánchez, E., Fromm, B., Oskolkov, N., Pochon, Z., Kalogeropoulos, P., Eriksson, E., Biryukova, I., Sekar, V., Ersmark, E., Andersson, B., Dalén, L., & Friedländer, M. R. (2023). Historical RNA expression profiles from the extinct Tasmanian tiger. Genome Research, 33(8), 1299–1316. https://doi.org/10.1101/gr.277663.123
- Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature, 499(7456), 74–78. https://doi.org/10.1038/nature12323
- van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., … Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature, 591(7849), 265–269. https://doi.org/10.1038/s41586-021-03224-9

ACKNOWLEDGMENTS

We would like to thank Mckenna Santiago Coyle, Melina Giakoumis, and Stephen Gaughran. A.P. was funded by the American Society of Mammalogy (ASM), Society for Systematic Biology (SSB), Systematics Association, Linnean Society of London, and National Science Foundation of the United States (NSF; Award No. BSC-2120691); A.P. was also funded by the Smithsonian Institution Graham Funds in Human Origins and the University of Texas at San Antonio (UTSA). A.T.S was funded by the U.S. National Science Foundation (DBI 2029955). L.T.C. was funded, in part, by the U.S. National Science Foundation (DBI 2029955). M.E.B. and S.K.M. were funded by the National Science Foundation of the United States (Awards No. BCS-1926105/1926215). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

REFERENCES

Amato, G., DeSalle, R., Ryder, O. A., & Rosenbaum, H. C. (Eds.). (2009). Conservation genetics in the age of genomics. Columbia University Press. https://doi.org/10.7312/amat12832

Arengo, F., Porzecanski, A. L., Blair, M. E., Amato, G., Filardi, C., & Sterling, E. J. (2018). The future of natural history museums. In E. Dorfman (Ed.), The future of natural history museums (1st ed., pp. 82–100). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315531892-6/essential-role-museums-biodiversity-conservation-felicity-arengo-ana-porzecanski-mary-blair-george-amato-christopher-filardi-eleanor-sterling

Bein, B., Chrysostomakis, I., Arantes, L. S., Brown, T., Gerheim, C., Schell, T., Schneider, C., Leushkin, E., Chen, Z., Sigwart, J., Gonzalez, V., Wong, N. L. W. S., Santos, F. R., Blom, M. P. K., Mayer, F., Mazzoni, C. J., Böhne, A., Winkler, S., Greve, C., & Hiller, M. (2025). Long-read sequencing and genome assembly of natural history collection samples and challenging specimens. Genome Biology, 26(1), 25. https://doi.org/10.1186/s13059-025-03487-9

Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: A review of methods, challenges, applications, and future directions. Molecular Ecology Resources, 19(5), 1115–1130. https://doi.org/10.1111/1755-0998.13066

Blair, M., Cao, G., López-Nandam, E., Veronese-Paniagua, D., Birchette, M., Kenyon, M., Md-Zain, B., Munds, R., Nekaris, K., Nijman, V., Roos, C., Thach, H., Sterling, E., & Le, M. (2023). Molecular phylogenetic relationships and unveiling novel genetic diversity among slow and pygmy lorises, including resurrection of *Xanthonycticebus intermedius*. Genes, 14(3), 643. https://doi.org/10.3390/genes14030643

Blair, M. E. (2024). Conservation museomics. Conservation Biology, 38(3) e14234. https://doi.org/10.1111/cobi.14234

Brasil, S. N. R., Kelemen, E. P., & Rehan, S. M. (2023). Historic DNA uncovers genetic effects of climate change and landscape alteration in two wild bee species. Conservation Genetics, 24(1), 85–98. https://doi.org/10.1007/s10592-022-01488-w

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2009). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research, 38(6), e87. https://doi.org/10.1093/nar/gkp1163

Byrne, A. Q. (2023). Reimagining the future of natural history museums with compassionate collection. PLOS Biology, 21(5), e3002101. https://doi.org/10.1371/journal.pbio.3002101

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for indigenous data governance. Data Science Journal, 19, 43. https://doi.org/10.5334/dsj-2020-043

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data, 8(1), 108. https://doi.org/10.1038/s41597-021-00892-0

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. Molecular Ecology, 22(17), 4369–4383. https://doi.org/10.1111/mec.12413

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances, 1(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Chen, N., & Nedoluzhko, A. (2023). Ancient DNA: The past for the future. BMC Genomics, 24(1), 309. https://doi.org/10.1186/s12864-023-09396-0

Colella, J. P., Tigano, A., & MacManes, M. D. (2020). A linked-read approach to museomics: Higher quality de novo genome assemblies from degraded tissues. Molecular Ecology Resources, 20(4), 856–870. https://doi.org/10.1111/1755-0998.13155

Collar, N. J. (2000). Collecting and conservation: Cause and effect. Bird Conservation International, 10(1), 1–15. https://doi.org/10.1017/S0959270900000010

Dabney, J., Meyer, M., & Paabo, S. (2013). Ancient DNA damage. Cold Spring Harbor Perspectives in Biology, 5(7), a012567. https://doi.org/10.1101/cshperspect.a012567

Dabney, J., & Meyer, M. (2019). Extraction of highly degraded DNA from ancient bones and teeth. Ancient DNA: Methods and Protocols, 25–29. https://doi.org/10.1007/978-1-4939-9176-1_4

Dahn, H. A., Mountcastle, J., Balacco, J., Winkler, S., Bista, I., Schmitt, A. D., Pettersson, O. V., Formenti, G., Oliver, K., Smith, M., Tan, W., Kraus, A., Mac, S., Komoroske, L. M., Lama, T., Crawford, A. J., Murphy, R. W., Brown, S., Scott, A. F., … Fedrigo, O. (2022). Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. GigaScience, 11. https://doi.org/10.1093/gigascience/giac068

De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R., & Pimm, S. L. (2015). Estimating the normal background rate of species extinction. Conservation Biology, 29(2), 452–462. https://doi.org/10.1111/cobi.12380

Fanning, T. G., Slemons, R. D., Reid, A. H., Janczewski, T. A., Dean, J., & Taubenberger, J. K. (2002). 1917 avian influenza virus sequences suggest that the 1918 pandemic virus did not acquire Its hemagglutinin directly from birds. Journal of Virology, 76(15), 7860–7862. https://doi.org/10.1128/JVI.76.15.7860-7862.2002

Feigin, C. Y., Newton, A. H., Doronina, L., Schmitz, J., Hipsley, C. A., Mitchell, K. J., Gower, G., Llamas, B., Soubrier, J., Heider, T. N., Menzies, B. R., Cooper, A., O'Neill, R. J., & Pask, A. J. (2017). Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. Nature Ecology & Evolution, 2(1), 182–192. https://doi.org/10.1038/s41559-017-0417-y

Feigin, C., Frankenberg, S., & Pask, A. (2022). A chromosome-scale hybrid genome assembly of the extinct Tasmanian tiger (*Thylacinus cynocephalus*). Genome Biology and Evolution, 14(4), evac048. https://doi.org/10.1093/gbe/evac048

Fong, J. J., Blom, M. P. K., Aowphol, A., McGuire, J. A., Sutcharit, C., & Soltis, P. S. (2023). Editorial: Recent advances in museomics: revolutionizing biodiversity research. Frontiers in Ecology and Evolution, 11. https://doi.org/10.3389/fevo.2023.1188172

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. PLOS ONE, 8(11), e79667. https://doi.org/10.1371/journal.pone.0079667

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution, 27(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

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, 22(6), 2130–2147. https://doi.org/10.1111/1755-0998.13505

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLOS ONE, 7(9), e45170. https://doi.org/10.1371/journal.pone.0045170

Hempel, E., Bibi, F., Faith, J. T., Brink, J. S., Kalthoff, D. C., Kamminga, P., Paijmans, J. L. A., Westbury, M. V., Hofreiter, M., & Zachos, F. E. (2021). Identifying the true number of specimens of the extinct blue antelope (*Hippotragus leucophaeus*). Scientific Reports, 11(1), 2100. https://doi.org/10.1038/s41598-020-80142-2

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. https://doi.org/10.1038/312282a0

International Council of Museums. (2013). ICOM code of ethics for natural history museums. ICOM. https://icom.museum/wp-content/uploads/2018/07/nathcode_ethics_en.pdf

Islam, S., Peart, C., Kehlmaier, C., Sun, Y.-H., Lei, F., Dahl, A., Klemroth, S., Alexopoulou, D., Del Mar Delgado, M., Laiolo, P., Carlos Illera, J., Dirren, S., Hille, S., Lkhagvasuren, D., Töpfer, T., Kaiser, M., Gebauer, A., Martens, J., Paetzold, C., & Päckert, M. (2024). Museomics help resolving the phylogeny of snowfinches (Aves, Passeridae, *Montifringilla* and allies). Molecular Phylogenetics and Evolution, 198, 108135. https://doi.org/10.1016/j.ympev.2024.108135

Jensen, E. L., Díez-del-Molino, D., Gilbert, M. T. P., Bertola, L. D., Borges, F., Cubric-Curik, V., De Navascués, M., Frandsen, P., Heuertz, M., Hvilsom, C., Jiménez-Mena, B., Miettinen, A., Moest, M., Pečnerová, P., Barnes, I., & Vernesi, C. (2022). Ancient and historical DNA in conservation policy. Trends in Ecology & Evolution, 37(5), 420–429. https://doi.org/10.1016/j.tree.2021.12.010

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. https://doi.org/10.1093/bioinformatics/btt193

Knapp, M., Clarke, A. C., Horsburgh, K. A., & Matisoo-Smith, E. A. (2012). Setting the stage – Building and working in an ancient DNA laboratory. Annals of Anatomy - Anatomischer Anzeiger, 194(1), 3–6. https://doi.org/10.1016/j.aanat.2011.03.008

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics, 15(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kreps, C. (2008). Indigenous curation, museums, and intangible cultural heritage. In L. Smith & N. Akagawa (Eds.), Intangible heritage (1st ed., p. 16). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203884973-16/indigenous-curation-museums-intangible-cultural-heritage-christina-kreps

Kutschera, V. E., Kierczak, M., Van Der Valk, T., Von Seth, J., Dussex, N., Lord, E., Dehasque, M., Stanton, D. W. G., Khoonsari, P. E., Nystedt, B., Dalén, L., & Díez-del-Molino, D. (2022). GenErode: A bioinformatics pipeline to investigate genome erosion in endangered and extinct species. BMC Bioinformatics, 23(1), 228. https://doi.org/10.1186/s12859-022-04757-0

Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. Evolutionary Applications, 12(8), 1505–1512. https://doi.org/10.1111/eva.12810

Leonard, J. A. (2008). Ancient DNA applications for wildlife conservation. Molecular Ecology, 17(19), 4186–4196. https://doi.org/10.1111/j.1365-294X.2008.03891.x

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A., & Haak, W. (2017). From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. STAR: Science & Technology of Archaeological Research, 3(1), 1–14. https://doi.org/10.1080/20548923.2016.1258824

Locatelli, N. S., McIntyre, P. B., Therkildsen, N. O., & Baetscher, D. S. (2020). GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. Proceedings of the National Academy of Sciences, 117(51), 32211–32212. https://doi.org/10.1073/pnas.2007421117

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. Molecular Ecology, 30(23), 5966–5993. https://doi.org/10.1111/mec.16077

Mace, G. M. (2004). The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359(1444), 711–719. https://doi.org/10.1098/rstb.2003.1454

Mármol-Sánchez, E., Fromm, B., Oskolkov, N., Pochon, Z., Kalogeropoulos, P., Eriksson, E., Biryukova, I., Sekar, V., Ersmark, E., Andersson, B., Dalén, L., & Friedländer, M. R. (2023). Historical RNA expression profiles from the extinct Tasmanian tiger. Genome Research, 33(8), 1299–1316. https://doi.org/10.1101/gr.277663.123

Marshall, W. F., Telford, S. R., Rys, P. N., Rutledge, B. J., Mathiesen, D., Malawista, S. E., Spielman, A., & Persing, D. H. (1994). Detection of *Borrelia burgdorferi* DNA in museum specimens of *Peromyscus leucopus*. Journal of Infectious Diseases, 170(4), 1027–1032. https://doi.org/10.1093/infdis/170.4.1027

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? Evolutionary applications, 7(9), 999–1007. https://doi.org/10.1111/eva.12193

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. Molecular Ecology, 24(13), 3223–3231. https://doi.org/10.1111/mec.13243

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

Minteer, B. A., Collins, J. P., Love, K. E., & Puschendorf, R. (2014). Avoiding (re)extinction. Science, 344(6181), 260–261. https://doi.org/10.1126/science.1250953

Mitchell, K. J., & Rawlence, N. J. (2021). Examining natural history through the lens of palaeogenomics. Trends in Ecology & Evolution, 36(3), 258–267. https://doi.org/10.1016/j.tree.2020.10.005

National Academies of Sciences, Engineering, and Medicine. (2020). Biological collections: Ensuring critical research and education for the 21st century (p. 25592). National Academies Press. https://doi.org/10.17226/25592

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. Molecular Ecology Resources, 17(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654

Neukamm, J., Peltzer, A., & Nieselt, K. (2021). DamageProfiler: Fast damage pattern calculation for ancient DNA. Bioinformatics, 37(20), 3652–3653. https://doi.org/10.1093/bioinformatics/btab190

Ochoa, A., Onorato, D. P., Roelke-Parker, M. E., Culver, M., & Fitak, R. R. (2022). Give and take: Effects of genetic admixture on mutation load in endangered Florida panthers. Journal of Heredity, 113(5), 491–499. https://doi.org/10.1093/jhered/esac037

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. Nature Reviews Methods Primers, 1(1), 14. https://doi.org/10.1038/s43586-020-00011-0

Pääbo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences, 86(6), 1939–1943. https://doi.org/10.1073/pnas.86.6.1939

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: Efficient ancient genome reconstruction. Genome Biology, 17(1), 60. https://doi.org/10.1186/s13059-016-0918-z

Penna, A., Blair, M. E., Lui, H.-L., Peters, E., Kistler, L., & Pozzi, L. (2024). Overcoming challenges to extracting and sequencing historical DNA to support primate evolutionary research and conservation, with an application to galagos. International Journal of Primatology, 45(6), 1375–1403. https://doi.org/10.1007/s10764-024-00429-3

Plaxton, L., Hempel, E., Marsh, W. A., Portela Miguez, R., Waurick, I., Kitchener, A. C., Hofreiter, M., Lister, A. M., Zachos, F. E., & Brace, S. (2023). Assessing the identity of rare historical museum specimens of the extinct blue antelope (*Hippotragus leucophaeus*) using an ancient DNA approach. Mammalian Biology, 103(6), 549–560. https://doi.org/10.1007/s42991-023-00373-4

Quatela, A.-S., Cangren, P., Jafari, F., Michel, T., De Boer, H. J., & Oxelman, B. (2023). Retrieval of long DNA reads from herbarium specimens. AoB PLANTS, 15(6), plad074. https://doi.org/10.1093/aobpla/plad074

Radin, J. (2015). Planned Hindsight: The vital valuations of frozen tissue at the zoo and the natural history museum. Journal of Cultural Economy, 8(3), 361–378. https://doi.org/10.1080/17530350.2015.1039458

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. Trends in Ecology & Evolution, 36(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Molecular Ecology, 24(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rocha, L. A., Aleixo, A., Allen, G., Almeda, F., Baldwin, C. C., Barclay, M. V. L., Bates, J. M., Bauer, A. M., Benzoni, F., Berns, C. M., Berumen, M. L., Blackburn, D. C., Blum, S., Bolaños, F., Bowie, R. C. K., Britz, R., Brown, R. M., Cadena,

C. D., Carpenter, K., … Witt, C. C. (2014). Specimen collection: An essential tool. Science, 344(6186), 814–815. https://doi.org/10.1126/science.344.6186.814

Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth, and sediments for high-throughput sequencing. Nature protocols, 13(11), 2447–2461. https://doi.org/10.1038/s41596-018-0050-5

Salis, A.T., Watson, Z., Forcellati, M., Ali, N., Karmakar, S., Kaur, J., Rabibisoa, N., Raxworthy, C.J. and Smith, B.T. (2025), Unrecognised DNA Degradation in Flash-Frozen Genetic Samples in Natural History Collections. Molecular Ecology Resources e14121. https://doi.org/10.1111/1755-0998.14121

Schmid, S., Neuenschwander, S., Pitteloud, C., Heckel, G., Pajkovic, M., Arlettaz, R., & Alvarez, N. (2018). Spatial and temporal genetic dynamics of the grasshopper *Oedaleus decorus* revealed by museum genomics. Ecology and Evolution, 8(3), 1480–1495. https://doi.org/10.1002/ece3.3699

Schubert, M., Ermini, L., Sarkissian, C. D., 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. https://doi.org/10.1038/nprot.2014.063

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution, 30(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G. M., Bininda-Emonds, O. R. P., Lee, P. L. M., & Cooper, A. (2002). Flight of the dodo. Science, 295(5560), 1683–1683. https://doi.org/10.1126/science.295.5560.1683

Simmons, J. E. (2014). Fluid preservation: A comprehensive reference. Rowman & Littlefield Publishers. https://doi.org/10.5771/9781442229662

Simpson, M. (2009). Museums and restorative justice: Heritage, repatriation and cultural education. Museum International, 61(1–2), 121–129. https://doi.org/10.1111/j.1468-0033.2009.01669.x

Van Der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., … Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature, 591(7849), 265–269. https://doi.org/10.1038/s41586-021-03224-9

Verry, A. J. F., Scarsbrook, L., Scofield, R. P., Tennyson, A. J. D., Weston, K. A., Robertson, B. C., & Rawlence, N. J. (2019). Who, where, what, wren? Using ancient DNA to examine the veracity of museum specimen data: a case study of the New Zealand rock wren (*Xenicus gilviventris*). Frontiers in Ecology and Evolution, 7, 496. https://doi.org/10.3389/fevo.2019.00496

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: Museum specimens in population genetics. Trends in Ecology & Evolution, 22(12), 634–642. https://doi.org/10.1016/j.tree.2007.08.017

Winker, K., Michael Reed, J., Escalante, P., Askins, R. A., Cicero, C., Hough, G. E., & Bates, J. (2010). The Importance, effects, and ethics of bird collecting. The Auk, 127(3), 690–695. https://doi.org/10.1525/auk.2010.09199

Wong, P. B., Wiley, E. O., Johnson, W. E., Ryder, O. A., O'Brien, S. J., Haussler, D., Koepfli, K.-P., Houck, M. L., Perelman, P., Mastromonaco, G., Bentley, A. C., Venkatesh, B., Zhang, Y., Murphy, R. W., & G10KCOS. (2012). Tissue sampling methods and standards for vertebrate genomics. GigaScience, 1(1), 2047-217X-1–8. https://doi.org/10.1186/2047-217X-1-8

York, N. D. L., Pritchard, R., Sauls, L. A., Enns, C., & Foster, T. (2023). Justice and ethics in conservation remote sensing: Current discourses and research needs. Biological Conservation, 287, 110319. https://doi.org/10.1016/j.biocon.2023.110319

Zhou, Z. T., Owens, G. L., Larson, W. A., Lou, R. N., & Sudmant, P. H. (2024). loco-pipe: An automated pipeline for population genomics with low-coverage whole-genome sequencing. Bioinformatics Advances, 4(1), vbae098. https://doi.org/10.1093/bioadv/vbae098