The use of museum specimens with high-throughput DNA sequencers

Andrew S. Burrell [a, *], Todd R. Disotell [a, b], Christina M. Bergey [a, b]

[a] Center for the Study of Human Origins, Department of Anthropology, New York University, 25 Waverly Place, New York, NY 10003, USA
[b] New York Consortium in Evolutionary Primatology, USA

ABSTRACT

Natural history collections have long been used by morphologists, anatomists, and taxonomists to probe the evolutionary process and describe biological diversity. These biological archives also offer great opportunities for genetic research in taxonomy, conservation, systematics, and population biology. They allow assays of past populations, including those of extinct species, giving context to present patterns of genetic variation and direct measures of evolutionary processes. Despite this potential, museum specimens are difficult to work with because natural postmortem processes and preservation methods fragment and damage DNA. These problems have restricted geneticists' ability to use natural history collections primarily by limiting how much of the genome can be surveyed. Recent advances in DNA sequencing technology, however, have radically changed this, making truly genomic studies from museum specimens possible. We review the opportunities and drawbacks of the use of museum specimens, and suggest how to best execute projects when incorporating such samples. Several high-throughput (HT) sequencing methodologies, including whole genome shotgun sequencing, sequence capture, and restriction digests (demonstrated here), can be used with archived biomaterials.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

Natural history collections are unique repositories of biodiversity. As human pressure on wild populations has increased over the past 200 years, the specimens deposited in museums have acquired greater value as samples of the biological past. Genetic studies incorporating natural history collections allow recently extinct species and populations to be surveyed. For example, Cooper et al. (1992) used museum specimens to sequence a portion of the 12S rRNA gene in moas, showing that the extinct bird is not closely related to the kiwi, New Zealand's other endemic flightless bird, suggesting that flightlessness may have evolved independently in the two lineages. Historical samples also enable direct assessment of changes in the genetics of populations over time. Such time series studies have been able to document selective sweeps of alleles associated with insecticide or herbicide resistance after introduction of chemicals (Hartley et al., 2006; Délye et al., 2013), or changes over time in genetic variation within populations in response to recent climate change (Rubidge et al., 2012; Bi et al., 2013; Kuhn et al., 2013). They also allow measures of evolutionary processes such as lineage sorting (Mende and Hundsdoerfer, 2013), and even the effects of pathogens on populations, such as the extinction of the Christmas Island rat after the introduction of a trypanosome parasite (Wyatt et al., 2008).

Human remains housed in museums are overwhelmingly archaeological, forensic, or fossil in nature and therefore exposed to the environment for long periods of time. This leads to different types of DNA damage than is found in intentionally collected and preserved specimens (Sawyer et al., 2012), making human ancient DNA (aDNA) studies beyond the scope of this review. However, primate museum specimens have been used to probe a number of questions, primarily in taxonomy and systematics (e.g., Geissmann et al., 2004; Monda et al., 2007; Boubli et al., 2008; Haus et al., 2013). Guschanski et al. (2013) used museum specimens of guenons, a speciose group of African monkeys, to assemble the best taxonomic sampling of that radiation to date and then inferred a phylogeny based on whole mitochondrial genomes. Markolf et al. (2013) investigated the taxonomy of the brown lemur complex, using morphological and molecular data from museum samples, and the same as well as vocalization data from living animals, to infer species limits. Such integrative studies hold great promise for taxonomic research, especially when type specimens can also be included.

In addition to taxonomic research, primates in museums have been used for biogeographic reconstructions and demographic

---

* Corresponding author.
  E-mail address: andrew.burrell@nyu.edu (A.S. Burrell).

inference. Alfaro et al. (2012) used mitochondrial data from museum skins to infer a rapid late Pleistocene expansion of robust capuchins (*Sapajus*) into the range of gracile capuchins (*Cebus*), leading to widespread current sympatry. Thalmann et al. (2011) exploited the temporal sampling available in natural history collections to document (via microsatellite genotypes) a sharp decline ($60\times$) in the effective population size of Cross River gorillas (*Gorilla gorilla diehli*).

Museums have sample collections of considerable size and geographic breadth. For many species, it may currently be more feasible to sample populations from museums than from the wild, given the expense and bureaucratic difficulty of organizing collecting expeditions. Localities currently too remote, dangerous, or expensive to visit may have been sampled in the past. Many museum collections are extensive. For example, the Natural History Museum in London has over 12,000 primate specimens; the American Museum of Natural History is not far behind at ~10,000. Historical and geographical factors often influence what species were collected and from where they were sampled. The Smithsonian's Natural History Museum has ~10,200 primate samples. Of these, ~5,300 are of New World monkeys (52.3%), with some 385 specimens of howler monkeys (*Alouatta*) from Panama alone (3.8% of the entire primate collection). In contrast, there are just 172 samples of lemuroids, one of the main primate radiations, representing 1.7% of the collection. This difference no doubt reflects the relative strength of historical and geographical links between the United States and Panama. In contrast, the Muséum National d'Histoire Naturelle in Paris, with its colonial ties to Madagascar, has ~1,020 lemuroids out of ~4,400 digitally cataloged specimens (23.4%) while it has only one howler from Panama. When designing projects, museums will have to be targeted based on their history and the geographic distribution of the taxon of interest, and it is likely that collections from multiple museums will need to be sampled. However, these ample museum collections can allow well-designed genetic studies with large population sizes and dense geographic sampling.

The use of museum specimens in evolutionary genetics

The use of museum specimens in modern evolutionary genetics began in the 1980s, when Higuchi et al. (1984) cloned and sequenced 229 base pairs (bp) of the mitochondrial genome of *Equus quagga quagga*, an extinct type of zebra. Since that pioneering study, natural history collections have been increasingly used to address a wide range of questions. Early topics investigated commonly included phylogeny and taxonomy (e.g., Higuchi et al., 1984; Thomas et al., 1989; Cooper et al., 1992), assays of genetic variability over time (e.g., Thomas et al., 1990; Taylor et al., 1994; Nielsen et al., 1997; Bouzat et al., 1998), and changes in population size (Miller and Kapuscinski, 1997).

While the earliest studies used cloning, the vast majority have obtained short sequences of DNA sequence data and/or nuclear allele frequencies via polymerase chain reaction (PCR) and Sanger sequencing or DNA fragment length analyses. Unlike other aDNA approaches using materials exposed to the environment for hundreds or thousands of years, researchers utilizing museum specimens have been able to incorporate nuclear loci in their work since the mid-1990s (e.g., Roy et al., 1994; Taylor et al., 1994; Mundy et al., 1997) in the form of microsatellite allele frequency data. Improved laboratory methods led to the use of nuclear DNA (nDNA) sequences in the early 2000s (Li et al., 2000; Vallianatos et al., 2002; de la Herrán et al., 2004), and the number of studies using multiple loci has increased since then. Research using many museum samples (>100) has also become more common (e.g., Godoy et al., 2004; Miller et al., 2006). Polymerase chain reaction-based studies of natural history collections, however, still almost always

use fewer than 20 loci and commonly use just one, especially when working with relatively large sample sizes. A search on Thomson Reuters Web of Science for papers published in 2013 using genetic data from museum specimens resulted in 29 matches, 15 of which used only mitochondrial DNA (mtDNA). Thirteen papers used more than one locus, but only seven of these used more than two loci. Almost all of the studies with >2 loci used microsatellites. The two museum-specimen papers from 2013 that generated truly genomic data (Bi et al., 2013; Staats et al., 2013) used second generation, high-throughput (HT) sequencing technologies that avoid initial PCR amplifications.

Problems of PCR-based approaches with museum specimens

There are several reasons why studies using PCR-based approaches are limited to relatively low numbers of loci, but the most significant is DNA fragmentation. DNA appears to become fragmented quickly after death (Bär et al., 1988). The cause of post-mortem fragmentation is unclear: for older samples, DNA breakage appears to be due largely to depurination (Lindahl, 1993; Overballe-Petersen et al., 2012), while in younger samples (<100 years) breaks tend to occur 3′ to adenosine residues, suggesting autolytic enzymes may be responsible (Sawyer et al., 2012). Regardless of the cause, the relationship between specimen age and DNA fragmentation is not linear (Pääbo, 1989; Wandeler et al., 2003; Zimmermann et al., 2008; Allentoft et al., 2012; Sawyer et al., 2012), and even relatively recently collected samples (<20 years) can have highly fragmented DNA (Sawyer et al., 2012). The quality and amplifiability of DNA recovered from museum specimens may be due more to the sample preservation methods used (Cooper, 1994; Hall et al., 1997; Mason et al., 2011), storage conditions (Mason et al., 2011), the tissues targeted (Cooper et al., 1992; Casas-Marce et al., 2010), or how quickly the sample was desiccated (Pääbo, 1989) than to the age of the specimen itself. As a result, fragmentation is a problem for most museum specimens.

There are two basic negative consequences of DNA fragmentation when using PCR-based approaches: contamination and small amplicon size. If relatively intact exogenous DNA is co-extracted with the fragmented DNA from a museum or archaeological sample, the exogenous DNA may be preferentially amplified during PCR, even if it is present at low concentration (Pääbo, 1989). This contamination is common, and its prevention requires extensive precautions. The steps taken to avoid contamination in studies of natural history collections are similar to those taken for studies of archaeological or fossil materials (Cooper and Poinar, 2000; Hofreiter et al., 2001; Pääbo et al., 2004). They are time consuming and cumbersome and will not always ensure that exogenous DNA will not be amplified along with or instead of endogenous DNA. They include sampling specimens using personal protective equipment such as gloves, coats, and masks; treating the outside of the samples with chemicals, enzymes, or ultraviolet (UV) light to degrade or cross-link exogenous DNA; targeting DNA sources, such as the interior of bones or teeth, that are freer of foreign DNA; using separate lab facilities for the extraction and amplification of DNA; employing negative controls during extraction and PCR; using UV light and/or 10% bleach solution to clean working surfaces and implements; and performing independent replication in a separate laboratory (Willerslev and Cooper, 2005; Wandeler et al., 2007).

The second main negative effect of DNA fragmentation on PCR-based methods is the inability to amplify long stretches of DNA. Degraded DNA extracts often have average molecule sizes of 100–200 bp, and so PCR reactions have to target regions often <300 bp (Pääbo, 1989). This can be particularly problematic when trying to amplify nDNA for sequencing. Because nuclear loci tend to have less sequence variation than equivalent lengths of mtDNA

(Moore, 1995), longer sequences may be needed to generate sufficient data. As amplicons need to be small, this often means that multiple, overlapping PCRs have to be performed to obtain the full sequence of a nuclear locus. This makes sequencing nuclear loci time consuming and expensive, and is a major reason why few PCR-based studies of museum specimens contain multilocus sequence data. In addition, the need to avoid allelic dropout (the failure of one allele to detectably amplify during PCR) forces the use of small amplicons. Rates of dropout from museum specimens >30 years old can reach 50% for loci over 200 bp (Wandeler et al., 2003) and are higher for longer amplicons than shorter ones. Allelic dropout can thus be mitigated by designing primer pairs that target short amplicons <200 bp.

Extreme DNA fragmentation can limit the type and number of loci that can be amplified. In some samples, nuclear DNA may be too fragmented for even short targets to amplify. Under these circumstances only mitochondrial DNA may be recovered (Cooper and Wayne, 1998). This is because the mitochondrial genome is present within cells in high copy number compared with the nuclear genome (Pikó and Taylor, 1987) and so there may be some mitochondrial fragments still long enough to amplify even when no long stretches of nDNA are present. Fortunately for researchers using natural history collections, the overall amount of DNA recovered from museum specimens is generally higher than that recovered from archaeological artifacts (Sawyer et al., 2012), making amplification of nDNA more feasible. However, the quality and quantity of DNA recovered from museum specimens can vary greatly and is not guaranteed to be better than that from archaeological samples.

Museum specimens—like other aDNA sources—can suffer from low DNA quantities as well as fragmentation (Wandeler et al., 2007). Unlike archaeological samples, museum specimens have generally not been exposed to water or mineralization. Rather, DNA quantity, like degree of fragmentation, varies and often is mostly a function of sample preservation method (Cooper, 1994; Hall et al., 1997; Zimmermann et al., 2008) and/or type of tissue targeted (Casas-Marce et al., 2010). Low DNA quantities reduce the chances of successful PCR and increase allelic dropout (Taberlet et al., 1996). To address this, DNA concentrations need to be carefully quantified (Morin et al., 2001), and it may be necessary to perform multiple amplifications to make sure that both chromosomes have been amplified at a given locus (Taberlet et al., 1996).

Apart from fragmentation and low DNA concentrations, museum specimen DNA can suffer other types of damage, primarily crosslinks between DNA strands or other molecules, oxidation, and hydrolysis (Pääbo, 1989; Lindahl, 1993; Pääbo et al., 2004; Willerslev and Cooper, 2005) that affect PCR and sequencing. It is not clear how much effect preservation method has on these types of damage, although unbuffered formalin is especially bad for DNA preservation (Miething et al., 2006; Zimmermann et al., 2008). Crosslinks and oxidative nucleotide damage can lead to nonamplification of DNA (Höss et al., 1996) or PCR errors (Pääbo et al., 2004; Willerslev and Cooper, 2005). Hydrolytic damage causes nucleotide misincorporations. In particular, hydrolytic deamination of cytosine to uracil and adenine to hypoxyxanthine can occur. These deaminated bases then get converted to thymine or guanine by polymerases, leading to artifactual C > T and G > A substitutions during PCR (Pääbo et al., 2004; Briggs et al., 2007). These misincorporations can lead to incorrect sequence data that can bias inferences or (if in the priming site) PCR failure (Rowe et al., 2011). Both oxidative and hydrolytic damage appear to be less of a problem in museum than in archaeological specimens, probably because they accumulate more slowly over time than DNA fragmentation damage (Pääbo, 1989; Sawyer et al., 2012). In addition, some types of damage, such as cytosine deamination to uracil, can be enzymatically repaired prior to library preparation (Briggs et al., 2010).

These difficulties can make obtaining multilocus nuclear sequence or allele frequency data via PCR from museum samples prohibitive. Truly genomic studies using natural history collections have therefore not been possible when taking a PCR-based methodological approach.

High-throughput (HT) sequencing and museum specimens

The advent of 'next generation' HT DNA sequencers has radically altered our ability to obtain multilocus genetic data from natural history collections. The primary reason is that most HT sequencing platforms require short stretches of DNA as their template (Mardis, 2013). This obviates the basic problem afflicting PCR-based studies of museum specimens: the fragmented nature of their DNA. The use of short DNA templates greatly increases the number of loci that can be sequenced and greatly reduces the problem of contamination. Although DNA damage and low DNA quantities can still prove problematic for HT sequencing, and some special steps have to be taken when sequencing museum specimens, HT library preparation is usually conveniently similar to that when using high-quality DNA (Table 1).

Data collection approaches

A range of HT sequencing approaches is feasible with museum specimens if available DNA is of appropriate size, quantity, and quality. The ones with the most promise are whole genome sequencing, sequence capture, and restriction digest assays. Not all

Table 1
HT sequencing workflow when working with museum specimens

Step	Actions when working with museum specimens
Determine sample requirements	Does project need temporal sampling, extinct populations or species, or a wide geographic and/or taxonomic sampling? If yes, use archival samples.
Identify samples	Check online databases. Many museums have searchable ones.
Obtain permission	Collections managers/curators want written applications that describe the study, its merit, the samples requested, and the sampling protocol to be used.
Determine what types of sample to target	Different tissues and preservation methods affect DNA quality and quantity. See Table 3.
Sample preparation for extraction	Rehydrate in buffer, remove hair from skins, decontaminate surface of sample. Follow prudent aDNA sample handling protocols, but contamination issues are less problematic than for PCR-based approaches.
DNA extraction	Extraction methods yield different quantities of DNA. See Table 3.
Size selection and quantification	Knowledge of the quantity and fragment size range of your DNA is necessary. Small DNA fragments (<70 bp) may need to be removed; DNA quantity needs to be measured after removal. If necessary and possible, pool multiple size-selected extracts to get sufficient DNA for library preparation.
Sequencing approach	Several different approaches to data collection are available. See Table 3.
Library preparation	Generally follow standard procedures after making sure DNA extractions are of appropriate fragment size and quantity. Fragmentation of DNA via sonication or other method is likely unnecessary.
Data processing	Standard data filtering and mapping approaches apply. However, one may need to remove C > T and G > A changes and to take a single-end mapping approach to paired-end sequences. If a reference genome evolutionarily close to the study taxon is not available for mapping, it may be necessary or helpful to sequence a whole genome of the study taxon.

HT methods may work with museum specimens, however. It is highly unlikely, for example, that RNA sequencing would be possible.

Whole genome sequencing

Shotgun sequencing of whole genomes is feasible with museum specimens if DNA of appropriate size, quantity, and quality is available. This approach has already been used to spectacular effect on DNA extracted from ancient hominins (e.g., Green et al., 2010; Meyer et al., 2012), and it has also been demonstrated to work on museum specimens. Rowe et al. (2011) made libraries from a rat's toe (including both skin and bone) and molar. They achieved low average read depth for these libraries ($0.64\times$ and $0.38\times$, respectively), with 31%–46% of the genome represented by at least one read. Staats et al. (2013) got much higher average sequencing depth ($12-38\times$) and genomic coverage (81–98%) for nuclear genomes prepared from *Arabidopsis* and fungal specimens. Both studies reported mapping rates of about 40%, about half that expected from high-quality sources of DNA. The low mapping may be a result of postmortem DNA modifications (Rowe et al., 2011). Most recently, Hung et al. (2014) sequenced the genomes of four passenger pigeons using DNA extracted from the toe pads of skin specimens, obtaining sequencing depths of $5-20\times$. Reads were mapped to the domestic pigeon draft genome sequence, with mapping rates of 57–75% after filtering, closer to the mapping rates achieved using high-quality DNA.

Because of the high per-cell copy number and small size of the mitochondrial genome, whole genome shotgun sequencing can also be used to recover mitochondrial sequences, even if the coverage of the nuclear genome is too low for reliable SNP calling. Multiple studies have taken this approach (Miller et al., 2009; Rowe et al., 2011; Menzies et al., 2012; Hung et al., 2013; Staats et al., 2013). Average read depths are generally high and about an order of magnitude larger than for nuclear loci (Table 2).

Sequence capture

Sequence capture (or 'target enrichment') approaches have also been used successfully for sequencing DNA from museum samples. This method involves hybridizing genomic DNA to DNA or RNA probes or 'baits' present either in a solution (e.g., Gnirke et al., 2009) or on an array (e.g., Albert et al., 2007; Okou et al., 2007) and then washing away unbound, non-target DNA. The result is a DNA solution enriched for specific targets that can then be sequenced using HT platforms. Some prior knowledge of the target genome (or the genome of a closely related organism) may be needed to design baits (McCormack et al., 2013), although baits are tolerant of a fair amount of sequence variation (e.g., human exome baits can capture rhesus macaque genomic DNA; Vallender, 2011).

Because only a small fraction of the genome is assayed, multiple individuals can be sequenced simultaneously via a 'multiplex' approach, unlike whole genome sequencing of organisms with large genomes. Multiplexing involves adding unique sequence barcodes or indexes to the DNA fragments of each individual during library preparation. This enables multiple barcoded or indexed individuals to be sequenced simultaneously and then 'demultiplexed' bioinformatically afterwards by clustering reads that all have the same unique sequence barcode/index. Bi et al. (2013) used a multiplexed target enrichment technique to sequence ~12,000 exons in 40 samples of modern and 90-year-old museum specimen Alpine chipmunks in order to detect any genetic effects of climate-related population decline (Moritz et al.,

Table 2
Studies using HT methods with museum specimens

Study	Taxon	Locus/Loci	Sequencing approach	% Mapping	% Contamination	Read depth	DNA source
Bi et al., 2013	Alpine chipmunk (*Tamias alpinus*)	~12,000 exons, mt genome, SRY, 7 autosomal genes	Sequence capture	~40%	<0.3% human and bacterial	18.9× avg	Dried skins
Guschanski et al., 2013	Guenons (Cercopithicini)	Mt genomes	Sequence capture	3.9% avg; (0%–62%)	0.09% (human)	91× avg: (0.01×–953×)	Bone, dried tissue, skin, teeth
Hung et al., 2013	Passenger pigeon (*Ectopistes migratorius*)	Mt genomes	Whole genome shotgun sequencing	15–16%	~4.4%, ~17.8% (human, fungal, bacterial)	98× and 188×	Toe pad
Hung et al., 2014	Passenger pigeon (*Ectopistes migratorius*)	Whole genomes	Whole genome shotgun sequencing	57–75%	N/A	5× to 20× avg	Toe pad
Mason et al., 2011	Colugos (*Galeopterus variegatus*)	Mt genomes	Sequence capture	77%	~0.5% (human)	~979×	Dried tissue adherent to bone skin
Menzies et al., 2012	Tasmanian tiger (*Thylacinus cynocephalus*)	Portions of the mt genome; numts	Whole genome shotgun sequencing	54.2%	<10% overall, (1.7% human)	Not reported	Hair shaft
Miller et al., 2009	Tasmanian tiger (*Thylacinus cynocephalus*)	Mt genomes	Whole genome shotgun sequencing	1.1–12.1%	4.3–8.9% (human)	49.7×–66.8×	Hair shaft
Rowe et al., 2011	Rat (*Rattus norvegicus*)	Whole genomes, mt genomes	Whole genome shotgun sequencing	21.8–38.0%	<0.01% (human)	0.38×–0.64× (nuclear); 14.2×–82× (mt)	Molar, toe (skin and bone)
Staats et al., 2013	Various plants, fungi, and insects	Whole genomes, mt genomes, chloroplast genomes	Whole genome shotgun sequencing	0.002–64.0%	Proportions not reported	12.1×–35.8× (nuclear); 2.3×–176× (mt)	Dried plant material; head, leg, thorax of insects
Tin et al., 2014	Drosophila, ants (*Camponotus, Pogonomyrmex, Linepithema*)	Whole genomes, 1,000s of loci	RAD-Seq, Whole genome shotgun sequencing	19–74%	Proportions not reported	0.37× (avg for whole genome shotgun sequencing)	Insect bodies
This study	Baboon (*Papio anubis*)	>50,000 200 bp loci	RAD-Seq	~72%	<10%	2.7×–3.2× (avg)	Dried skin

2008). Despite using ~1,600 single-nucleotide polymorphisms (SNPs), no significant loss of genetic diversity was observed, although modern populations did appear to be more highly structured than past populations. Similar capture methods have also been used to sequence whole mitochondrial genomes of museum specimens of colugos and guenons (Mason et al., 2011; Guschanski et al., 2013). Average capture efficiency (in terms of percentage of reads that mapped to reference genomes) in all of these studies ranged widely, from an average of 3.9% (0.01—62.40%; Guschanski et al., 2013) to 46.0% (Bi et al., 2013) to ~77.0% (Mason et al., 2011). Baits were capable of capturing target DNA even at mismatch rates of ~10% (Mason et al., 2011). Sequence capture methods can also be used to enrich samples for whole genome sequencing of aDNA samples suffering from excessive contamination (Carpenter et al., 2013). However, this is unlikely to be a necessary approach for most researchers using museum samples as contamination levels are generally low (see 'Contamination and DNA damage in HT studies' below).

Restriction digest methods

Smaller portions of the genome can be targeted for sequencing using restriction enzymes. These methods—such as reduced representation libraries (RRLs; van Tassell et al., 2008), restriction-site associated DNA sequencing (RAD-Seq; Baird et al., 2008), complexity reduction of polymorphic sequences (CRoPS; van Orsouw et al., 2007), and multiplexed shotgun genotyping (MSG; Andolfatto et al., 2011)—digest total genomic DNA with one or more restriction enzymes, creating a population of sequences containing one or more short sequence motifs. Various techniques are then employed to ensure that only fragments with the cut-site motif(s) are sequenced. As with sequence capture, multiple individuals can be multiplexed. These methods have been commonly used for population genomics and systematics (Davey et al., 2011; McCormack et al., 2013), and can be adapted for low quality templates by increasing the amount of starting DNA (Etter et al., 2011). So far only RAD-Seq has been used with museum specimens.

RAD-Seq from museum samples

The basic RAD-Seq protocol begins with the digestion of genomic DNA by a restriction enzyme. A P1 adapter with an overhang appropriate for the restriction enzyme cut motif is then ligated on to the digested DNA. The P1 adapter consists of a PCR priming site, an Illumina sequencing priming site, and a barcode to allow identification of sequences of specific individuals during multiplex sequencing. The DNA from multiple individuals is now pooled. If using high-quality DNA, the pooled DNA is sonicated to have an average fragment length <1,000 bp. The DNA from museum skins is probably already fragmented, so no sonication may be required. Small and large DNA fragments unsuitable for sequencing are excluded via gel extraction. A P2 adapter is then ligated on to size-selected fragments. This is a Y-adapter, meaning that it has a complementary sequence for only a portion of its length. The mismatched portion includes the sequence of a reverse PCR amplification primer. The complementary sequence for the reverse amplification primer is only created after a round of PCR fills it in. This ensures that only fragments with both the P1 and P2 adapters get amplified and therefore sequenced. Since only fragments with a cut site can have P1, this also ensures that only DNA associated with cut sites is ultimately sequenced (Baird et al., 2008; Etter et al., 2011).

Tin et al. (2014) present a RAD-Seq protocol that modifies a single-strand approach to library preparation developed for aDNA samples by Gansauge and Meyer (2013). Five >50 year-old ant specimens were subjected to a non-destructive DNA extraction that yielded 5—766 ng of total genomic DNA per specimen with average fragment lengths of ~50 bp but ranging from ~25 to 160 bp. Relatively small amounts of DNA were then used for library preparation, ranging from 14 to 220 ng. Fifty-six to seventy-six percent reads mapped to a reference genome after bioinformatic filtering. The mean post-filter sequenced fragment size was ~30 bp, and the authors report that many reads had to be discarded because they were too short to map to reference genomes. They therefore recommend a size selection step to minimize the wasted sequencing of very short fragments, and also suggest several other modifications to the protocol to improve overall yields of mappable reads.

Another pilot RAD-Seq study at New York University's Molecular Anthropology Laboratory presented here suggests that the technique can be used on museum specimens. Five samples were taken from different parts of a dry, NaCl-preserved olive baboon (*Papio anubis*) hide roughly 40 years old (collected in 1973 by F. L. Brett at Metahara, Ethiopia). Any hair was removed with a scalpel and the skin samples were incubated for 24 h in $1\times$ TE, with the buffer removed and replaced once during incubation. After rehydration, samples ranged from 15 to 70 mg in mass. DNA was extracted using a Qiagen DNEasy Blood and Tissue Kit following Rowe et al. (2011). DNA concentration was quantified using a Qubit 2.0 fluorometer and a dsDNA BR assay, and fragment size was determined using an Agilent Bioanalyzer 2100 with a DNA1000 chip. DNA concentrations were generally high, between 16 and 111 ng/μl in 50 μl volumes. Fragment sizes averaged between 290 and 510 bp. Two samples with the highest DNA concentrations and average fragment lengths were selected for RAD-Seq. Because the templates contained degraded DNA, library preparation began with ~600 ng of DNA of each sample, $3\times$ more than used in prior multiplexed library preparations (Bergey et al., 2013). Library preparation using the enzyme *Psp*XI (New England Biotech) followed Etter et al. (2011), except for the use of Agencourt SPRI-select beads for size selection and AMPure XP beads for library clean-up. Both skin sample DNA preps were given barcoded adapters, as they were added to a pool of prepared DNA from 10 high-quality samples prior to 100 cycles of paired-end sequencing on an Illumina MiSeq platform. Reads were mapped to the baboon draft genome (https://www.hgsc.bcm.edu/non-human-primates/baboon-genome-project) and filtered for quality. Analyses were completed with the RAD-Seq-primate pipeline (https://code.google.com/p/rad-seq-primate/) that makes use of the programs BWA (Li and Durbin, 2009), Picard (http://picard.sourceforge.net), BamTools (Barnett et al., 2011), and GATK (DePristo et al., 2011). Statistics on the success of the sequencing were computed with SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010).

Simulated digests estimate that *Psp*XI cuts the baboon genome into ~55,000 fragments, resulting in ~110,000 paired-end sequenced sites (Bergey et al., 2013). After sequencing and data filtration, the skin samples had ~72,000 to ~77,000 loci with at least one mapped read, so roughly 65% and 70% of the ~110,000 possible loci were sequenced to a $1\times$ depth or greater (Table 2). Average read depths across all loci for the two skin samples were $2.6\times$ to $3.2\times$. The proportions of reads that mapped to *Psp*XI cut sites in the baboon reference genome were 90.4% to 91.0%. While not all cut sites were sequenced, a very high proportion of reads clearly mapped to cut sites, verifying that the technique does work in preserved skin specimens.

The reason only 65—70% of the possible cut sites were sequenced most likely has to do with the amount of initial DNA of appropriate fragment size used for multiplexed library preparation, and could likely be improved with a minor modification. While the preserved skin samples were quantified prior to ligation, a size selection was not performed, meaning that many adapters were

possibly ligated to fragments <300 bp long, which later were removed during the size selection step of library preparation. The two skin samples were also multiplexed with 10 other baboon libraries made from high-quality DNA and sequenced on an Illumina MiSeq. Per-locus read depths were therefore low, since the skin sample libraries likely had considerably less DNA in the targeted size range than the high-quality libraries. This highlights the need to remove very small fragments from DNA extracts of museum skins before quantifying the extract prior to library construction and to take care when multiplexing DNA from both low- and high-quality sources. Higher locus coverage and read depth can likely be obtained by increasing the amount of museum sample used for library preparation and by increasing the number of sequencing reads per sample, either by reducing the number of multiplexed samples per run and/or by using a platform with higher output, such as the Illumina HiSeq. Locus coverage and read depth can also be affected by the number of cut sites generated by the restriction enzyme used, but in this example we used a very rare cutter (Bergey et al., 2013), so read depth could not be improved significantly by using another enzyme. If possible, researchers should conduct simulated *in silico* digests of potential restriction enzymes on target or closely related genome sequences in order to select an enzyme with an appropriate number of cut sites. The ideal number depends upon several factors, including the number of individuals to be multiplexed and the number of sites desired. For many applications, rare cutters may be best, but some, like association studies, might need common cutters.

These two pilot studies do suggest that restriction enzyme approaches can work with museum specimens if certain initial conditions can be met. Because these methods cut up already fragmented DNA, the starting total genomic DNA cannot be composed largely of short segments. A minimum amount of starting DNA is also likely necessary to ensure that as many of the cut sites as possible are present in a digestable and mappable form (i.e., without lesions in the cut site, and in fragments long enough to be mappable after digestion and sequencing). A realistic condition for successful RAD-Seq, for example, might be beginning library preparation with 50 ng or more of DNA that is >70 bp in length (it must be longer than your P1 adapter). Many museum specimens will not yield DNA of this quality and quantity, so researchers must

be aware of the condition of their DNA extracts before attempting restriction enzyme based sequencing. Extract conditions can be determined using double-stranded DNA quantifiers such as a Qubit fluorometer (Invitrogen) and DNA analyzers like the Agilent 2100 Bioanalyzer.

Amplicon sequencing

Specific loci can be targeted for HT sequencing via PCR. Because of the high output of HT sequencers, large numbers of barcoded PCR products from many individuals can be sequenced at once (Meyer et al., 2008), making this approach attractive for some population-level questions. For example, the parallel sequencing of multiple mitochondrial genomes has commonly been done via amplicon sequencing (e.g., Chan et al., 2010; Morin et al., 2010; Gunnarsdottir et al., 2011; Sosa et al., 2012). As the total number of base pairs of PCR product to be sequenced will likely be much smaller than the sequencing capacity of most HT platforms, many loci from many individuals can potentially be multiplexed in a single sequencing run, reducing per-sample costs. However, as these approaches are explicitly based on PCR, they can be expensive, time consuming, and subject to contamination when working with degraded DNA. Like sequence capture methods, they also require some knowledge of the genome of the taxon of interest in order to design primers. Amplicon sequencing is therefore most likely an option when only a few loci need to be sequenced in large numbers of individuals.

Pros and cons of different sequencing approaches

The various HT sequencing methods have strengths and weaknesses (Table 3). Whole genome shotgun sequencing results in data suitable for systematics, certain kinds of demographic inference such as the pairwise sequentially Markovian coalescent (PSMC; Li and Durbin, 2011), and assays for the effects of selection. However, the per-sample costs can be high in organisms with large genomes (such as primates) and therefore limit sample sizes. Also, whole genome sequencing can require more starting DNA than other methods, and so may be impractical when DNA quantities are very low (although whole genome amplification might rectify this).

Table 3
Options available when working with museum specimens

Step	Pros	Cons
Sample tissue: soft tissues (skins, connective tissues, etc.)	Easy to sample, easier to get permission to sample	DNA quality and quantity highly variable, depending on curation history
Sample tissue: hard tissues (teeth, bones, antlers, etc)	More likely to have higher quality DNA	Hard to sample (either invasively or noninvasively) and harder to get permission to sample
DNA extraction: invasive	Easy to do (soft tissues); may yield higher amounts of DNA	Destroys whole or part of sample; may be harder to get sampling permission
DNA extraction: noninvasive	Keeps specimen intact; facilitates obtaining sampling permission	Hard to do, especially at a museum; DNA yield may be lower than invasive extraction
Library prep: whole genome shotgun sequencing	Large amounts of sequence data; methodologically simple	Often low coverage depth (except mt genome); individuals with large genomes generally cannot be multiplexed, so sample size is lower; may be more data than necessary
Library prep: sequence capture	Specific loci can be targeted; multiplexing possible, so per individual cost can be low after investment in baits; length of sequenced loci determined by researcher	Potentially high upfront cost for baits; limits on how evolutionarily divergent taxa can be from baits and still have capture; technically more difficult than other methods; some knowledge of the target genome (or close relative) may be needed
Library prep: restriction digests	Large number of SNPs can be assayed; multiplexing simple, so per individual cost can be low; no prior knowledge of target genome needed	Variation in sequences of cut sites creates null alleles; limits on how evolutionarily divergent taxa can be and still collect orthologous loci; only get SNP data; requires relatively good quality and quantity of DNA, so may not be suitable for many specimens
Library prep: amplicon sequencing	Can multiplex many individuals; per sample costs potentially quite low	PCR-based; not realistically capable of generating as much data as other methods

Restriction enzyme approaches result in large amounts of SNP data, which can be used for detecting population structure, hybridization, the systematics of relatively recent radiations (if using a concatenation approach), and association studies. Also, as only a fraction of the entire genome is being sequenced, multiplexing makes large sample sizes possible. However, it may be necessary to have relatively large amounts of DNA >70 bp long for reliable sequencing, especially with RRLs and RAD-Seq. This may often not be possible, especially with small organisms where little tissue can be sacrificed for DNA extraction or for specimens where DNA fragmentation is pronounced. Direct assays of the quantity and length of DNA from museum specimens must be conducted to determine whether restriction enzyme approaches are appropriate. Sequence capture allows specific loci to be targeted for sequencing, and the researcher can determine how long the targeted regions should be. As with restriction digests, multiplexing is possible and large sample sizes can be used. Sequence capture could be used for a wide variety of applications, including most that can be done with whole genome sequencing or restriction digests. Sequence capture can be more expensive than restriction digests and requires knowledge of the sequence of targeted genomic regions prior to bait construction. This may not be possible for some non-model organisms, although researchers could generate de novo whole genome sequences of the organism(s) being studied in order to create appropriate baits.

It is likely that sequence capture will be the most appropriate course of action for many researchers, although restriction digest approaches will be a possibility for some if enough samples yield DNA of sufficient quality, and amplicon sequencing may be desirable if only a few loci need to be sequenced in a large number of individuals. Whole genome sequencing can be done if large blocks of sequence data are needed from only a few individuals. Also, researchers may want to sequence a whole genome of the taxon of interest in addition to generating more limited sequence datasets in many samples via a sequence capture, restriction digest, or amplicon sequencing approach in order to construct baits or to aid in contig assembly and mapping. However, we also note that, as sequencing costs fall and throughput increases, whole genome sequencing will likely become standard even for population level studies involving many individuals.

Contamination and DNA damage in HT studies

Contamination levels in the HT studies have generally been low. Rowe et al. (2011) found <0.1% of reads mapped to the human or bacterial genomes, while Bi et al. (2013) reported <0.3%. Miller et al. (2009) found somewhat higher levels of human contamination in Tasmanian tiger hairs, ~4.3% to 8.9%. The higher proportion of contamination in the hairs may be due to their exposed nature and relatively low levels of endogenous DNA. Menzies et al. (2012) had the highest reported contamination, at 10.1%. Compared with archaeological samples, where exogenous DNA can often represent ~99% of extracted DNA (Carpenter et al., 2013), museum samples generally appear to have a high enough proportion of endogenous DNA for conventional whole genome shotgun sequencing.

Nucleotide damage was generally low, with C > T misincorporations at <0.45% (Staats et al., 2013) and 0.5–0.6% (Miller et al., 2009) and both C > T and G > A misincorporations at <0.6% (Bi et al., 2013) and <0.27% (Rowe et al., 2011). The similar C > T and G > A rates in Bi et al. (2013) may in part be due to a proofreading enzyme that fails at uracil residues, reducing the amount of C > T misincorporations. Despite low rates, Bi et al. (2013) still recommend removal of all C > T and G > A changes from datasets, as even low levels of damage can affect demographic inferences and other population level analyses (Axelsson et al., 2008).

Working with museum specimens: choosing tissue type, extracting DNA, and bioinformatics

DNA yields vary between tissue types. There is a consensus that hard tissues such as teeth and bone provide the most intact DNA because they were less exposed than soft tissues like skins to chemical preservatives or other environmental agents that may fragment DNA (Wandeler et al., 2007). Claws (and possibly nails) also contain usable DNA (Hedmark and Ellegren, 2005), as do antlers (Hoffmann and Griebeler, 2013). Specimens stored in fluids can also be used as sources of DNA (Garrigos et al., 2013). However, it is unclear whether hard tissues yield more or less DNA overall compared with hides, as different studies have had conflicting results (e.g., Casas-Marce et al., 2010; Rowe et al., 2011). Preservation method is likely the cause for this variance; it seems probable that skins will have greater DNA when relatively non-damaging preservatives have been used. Choosing which tissue types to sample can therefore be difficult, especially if preservation methods have not been reported or are incorrect (Hall et al., 1997; see Table 3).

Hides are often preferred for destructive sampling by museum collections managers, who understandably wish to minimize harm to hard tissues useful for morphological or other research. However, noninvasive DNA extractions of bones and teeth that yield amplifiable nuclear as well as mitochondrial DNA appear to be possible (Bolnick et al., 2012; Garrigos et al., 2013; using modifications of: Rohland et al. [2004] and; Rohland and Hofreiter [2007]). These all bathe a specimen in a buffer, either proteinase K and EDTA (Bolnick et al., 2012), or a guanidine-based solution (GuSCN; Garrigos et al., 2013) in order to recover DNA from the specimen. For the GuSCN approach, DNA concentrations averaged 26 ng/μl (0.11–120.00 ng/μl) in 50 μl elution volumes (Garrigos et al., 2013), levels appropriate for HT sequencing. Similar 'bathing' techniques have also been used to successfully extract DNA from insects (e.g., Gilbert et al., 2007; Hunter et al., 2008; Mikheyev et al., 2009).

DNA extraction from invasively-collected samples can be performed using either commercially available kits, such as the Qiagen DNeasy Blood and Tissue kits (e.g., Rowe et al., 2011; Bi et al., 2013), or via column-based protocols developed for other aDNA extractions (e.g., Guschanski et al., 2013, following Rohland et al. [2010]). Both approaches appear to yield sufficient amounts of DNA, and the quantity and quality of DNA recovery is likely more dependent on the curation history of the sample than the extraction method used (Table 3). For dried samples, incubation in a buffer such as $1\times$ TE is recommended in order to rehydrate the sample (Moraes-Barros and Morgante, 2007; Bi et al., 2013). If DNA is available in sufficient quantity, size selection prior to library preparation—and specifically adapter ligation—may be helpful in order to remove fragments that are too short for mapping after sequencing.

Sequences generated from museum specimens do need some special bioinformatic processing. In 'paired-end' sequencing, each DNA fragment is read twice, once from each end. However, for museum specimens it may be prudent to discard the second read and only include the first in analyses, i.e., follow a 'single-end' strategy (Rowe et al., 2011). This may be necessary if fragment chimeras (two or more biological fragments that are artificially joined into one) form during library preparation. When the two ends of a fragment chimera are sequenced via paired-end sequencing, the two reads will come from two separate loci, making mapping difficult if a single locus is expected. Also, as noted before, C > T and G > A changes may need to be removed entirely from the dataset, as they may be the result of postmortem DNA damage (Bi et al., 2013).

New sample collection and preservation

DNA damage, especially fragmentation, appears to occur quickly after cells begin to die. The best evidence suggests that it is enzymatic processes (Sawyer et al., 2012) or sample preservation method (Cooper, 1994; Hall et al., 1997) that are primarily responsible for DNA damage or enzymatic inhibition in museum specimens. Because both the autolytic and hydrolytic processes that fragment DNA or modify nucleotides are dependent upon water, quick desiccation is recommended (Pääbo, 1989; Pääbo et al., 2004). Preservation methods that promote DNA degradation or damage, such as unbuffered formalin, should be avoided (Ferrer et al., 2008). Newly collected bones and teeth should be cleaned without harsh chemical treatments such as bleaching. Hides should be desiccated quickly with simple salts, such as NaCl, if possible. Such preservation methods also do not introduce compounds that inhibit protease and polymerase activity (Hall et al., 1997).

As new specimens are collected and deposited in museums, collections managers are increasingly trying to store small tissue or other samples from them specifically for use in future genetic studies (Prendini et al., 2002). The best ways to preserve and store such small samples include cryopreservation in liquid nitrogen, storage in >70% ethanol or a supersaturated salt solution, such as RNAlater, or desiccation (Nagy, 2010; Jackson et al., 2012). Cryopreservation is the only method known to keep DNA in good condition over long periods of time (as long as multiple freeze–thaw cycles are avoided), but it is expensive, requires considerable maintenance, and is often not available under field conditions. Ethanol needs to be replaced over time and does not preserve molecules other than DNA, such as proteins or RNA (Jackson et al., 2012). Desiccated samples need to be monitored to make sure they remain dry, and drying agents, such as silica, need to be replaced periodically (Nagy, 2010; Jackson et al., 2012). Fluid samples can be desiccated and fixed on special papers, such as FTA cards. Supersaturated salt solutions may be best for low-maintenance long-term storage, especially if frozen, as they need little if any monitoring or maintenance if kept in well-sealed containers. These salt solutions are also easy to work with in the field as they keep DNA (and even RNA) stable at room temperature for weeks or months. Their main downside is that they can be expensive (Nagy, 2010).

Conclusions

The HT sequencing technologies that require short DNA fragments as their template have made it possible to generate large genomic datasets from biological samples archived in natural history collections. With relatively minor modifications, protocols using museum specimens as DNA sources can follow those using high-quality DNA. Whole genome shotgun sequencing, sequence capture, and possibly restriction digest methods can be applied to many archived materials. Museums house large numbers of samples, often from a wide range of geographic localities. Researchers working with hard to sample, non-model organisms now have the opportunity to develop projects that are large both in terms of numbers of individuals and populations sampled and in the number of loci assayed.

Acknowledgments

References

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., Gibbs, R.A., 2007. Direct selection of human genomic loci by microarray hybridization. Nature Methods 4, 903–906.

Alfaro, J.W.L., Boubli, J.P., Olson, L.E., Di Fiore, A.D., Wilson, B., Gutierrez-Espeleta, G.A., Chiou, K.L., Schulte, M., Neitzel, S., Ross, V., Schwochow, D., Nguyen, M.T.T., Farias, I., Janson, C.H., Alfaro, M.E., 2012. Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys. J. Biogeogr. 39, 272–288.

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

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T., Stern, D.L., 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21, 610–617.

Axelsson, E., Willerslev, E., Gilbert, M.T.P., Nielsen, R., 2008. The effect of ancient DNA damage on inferences of demographic histories. Mol. Biol. Evol. 25, 2181–2187.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.

Bär, W., Kratzer, A., Mächler, M., Schmid, W., 1988. Postmortem stability of DNA. Forensic Sci. Int. 39, 59–70.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., Marth, G.T., 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691–1692.

Bergey, C.M., Pozzi, L., Disotell, T.R., Burrell, A.S., 2013. A new method for genome-wide marker development and genotyping holds great promise for molecular primatology. Int. J. Primatol. 34, 303–314.

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R., Moritz, C., 2013. Unlocking the vault: next-generation museum population genomics. Mol. Ecol. 22, 6018–6032.

Bolnick, D.A., Bonine, H.M., Mata-Míguez, J., Kemp, B.M., Snow, M.H., LeBlanc, S.A., 2012. Nondestructive sampling of human skeletal remains yields ancient nuclear and mitochondrial DNA. Am. J. Phys. Anthropol. 147, 293–300.

Boubli, J.P., da Silva, M.N.F., Amado, M.V., Hrbek, T., Pontual, F.B., Farias, I.P., 2008. A taxonomic reassessment of *Cacajao melanocephalus* Humboldt (1811), with the description of two new species. Int. J. Primatol. 29, 723–741.

Bouzat, J.L., Lewin, H.A., Paige, K.N., 1998. The ghost of genetic diversity past: historical DNA analysis of the greater prairie chicken. Am. Nat. 152, 1–6.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. 104, 14616–14621.

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

Carpenter, M.L., Buenrostro, J.D., Valdiosera, C., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., Gilbert, M.T., Willerslev, E., Greenleaf, W.J., Bustamante, C.D., 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am. J. Hum. Genet. 93, 852–864.

Casas-Marce, M., Revilla, E., Godoy, J.A., 2010. Searching for DNA in museum specimens: a comparison of sources in a mammal species. Mol. Ecol. Resour. 10, 502–507.

Chan, Y.-C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C.-C., Pei, K.J.-C., Vigilant, L., 2010. Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates* gibbons. PLoS One 5, e14419.

Cooper, A., 1994. DNA from museum specimens. In: Herrmann, B., Hummel, S. (Eds.), Ancient DNA. Springer-Verlag, New York, pp. 149–165.

Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. Science 289, 1139.

Cooper, A., Wayne, R., 1998. New uses for old DNA. Curr. Opin. Biotech. 9, 49–53.

Cooper, A., Mourer-Chauviré, C., Chambers, G.K., von Haeseler, A., Wilson, A.C., Pääbo, S., 1992. Independent origins of New Zealand moas and kiwis. Proc. Nat. Acad. Sci. 89, 8741–8744.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.

de la Herrán, R., Robles, F., Martinez-Espin, E., Lorente, J.A., Rejon, C.R., Garrido-Ramos, M.A., Rejon, M.R., 2004. Genetic identification of western Mediterranean sturgeons and its implication for conservation. Conserv. Genet. 5, 545—551.

Délye, C., Deulvot, C., Chauvel, B., 2013. DNA analysis of herbarium specimens of the grass weed *Alopecurus myosuroides* reveals herbicide resistance pre-dated herbicides. PLoS One 8, e75117.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491—498.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A., 2011. Molecular Methods for Evolutionary Genetics, vol. 772. Springer, New York, pp. 1—19.

Ferrer, I., Martinez, A., Boluda, S., Parchi, P., Barrachina, M., 2008. Brain banks: benefits, limitations and cautions concerning the use of post-mortem brain tissue for molecular studies. Cell Tissue Bank. 9, 181—194.

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

Garrigos, Y.E., Hugueny, B., Koerner, K., Ibanez, C., Bonillo, C., Pruvost, P., Causse, R., Cruaud, C., Gaubert, P., 2013. Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei: Cyprinodontidae). Zootaxa 3640, 373—394.

Geissmann, T., Groves, C.P., Roos, C., 2004. The Tenasserim Lutung, *Trachypithecus barbei* (Blyth, 1847) (Primates: Cercopithecidae): Description of a live specimen, and a reassessment of phylogenetic affinities, taxonomic history, and distribution. Contrib. Zool. 73, 271—282.

Gilbert, M.T., Moore, W., Melchior, L., Worobey, M., 2007. DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One 2, e272.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotech. 27, 182—189.

Godoy, J.A., Negro, J.J., Hiraldo, F., Donázar, J.A., 2004. Phylogeography, genetic structure and diversity in the endangered bearded vulture (*Gypaetus barbatus*, L.) as revealed by mitochondrial DNA. Mol. Ecol. 13, 371—390.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A draft sequence of the Neandertal genome. Science 328, 710—722.

Gunnarsdottir, E.D., Li, M., Bauchet, M., Finstermeier, K., Stoneking, M., 2011. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 21, 1—11.

Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T., Lenglet, G., Mayer, F., Savolainen, V., 2013. Next-generation museomics disentangles one of the largest primate radiations. Syst. Biol. 62, 539—554.

Hall, L.M., Wilcox, M.S., Jones, D.S., 1997. Association of enzyme inhibition with methods of museum skin preparation. Biotechniques 22, 928—934.

Hartley, C.J., Newcomb, R.D., Russell, R.J., Yong, C.G., Stevens, J.R., Yeates, D.K., La Salle, J., Oakeshott, J.G., 2006. Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. Proc. Natl. Acad. Sci. 103, 8757—8762.

Haus, T., Akom, E., Agwanda, B., Hofreiter, M., Roos, C., Zinner, D., 2013. Mitochondrial distribution and diversity of African green monkeys (*Chlorocebus* Gray, 1870). Am. J. Primatol. 75, 350—360.

Hedmark, E., Ellegren, H., 2005. Microsatellite genotyping of DNA isolated from claws left on tanned carnivore hides. Int. J. Legal Med. 119, 370—373.

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, 282—284.

Hoffmann, G.S., Griebeler, E.M., 2013. An improved high yield method to obtain microsatellite genotypes from red deer antlers up to 200 years old. Mol. Ecol. Resour. 13, 440—446.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S., 2001. Ancient DNA. Nat. Rev. Genet. 2, 353—359.

Höss, M., Jaruga, P., Zastawny, T.H., Dizdaroglu, M., Pääbo, S., 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res. 24, 1304—1307.

Hung, C.-M., Lin, R.-C., Chu, J.-H., Yeh, C.-F., Yao, C.-Y., Li, S.-H., 2013. The de novo assembly of mitochondrial genomes of the extinct passenger pigeon (*Ectopistes migratorius*) with next generation sequencing. PLoS One 8, e56301.

Hung, C.-M., Shaner, P.-J.L., Zink, R.M., Liu, W.-C., Chu, T.-C., Huang, W.-S., Li, S.-H., 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proc. Natl. Acad. Sci. 111, 10636—10641.

Hunter, S.J., Goodall, T.I., Walsh, K.A., Owen, R., Day, J.C., 2008. Nondestructive DNA extraction from blackflies (Diptera: Simuliidae): retaining voucher specimens for DNA barcoding projects. Mol. Ecol. Resour. 8, 56—61.

Jackson, J.A., Laikre, L., Baker, C.S., Kendall, K.C., 2012. Guidelines for collecting and maintaining archives for genetic monitoring. Conserv. Genet. Resour. 4, 527—536.

Kuhn, K., Schwenk, K., Both, C., Canal, D., Johansson, U.S., van der Mije, S., Töpfer, T., Päckert, M., 2013. Differentiation in neutral genes and a candidate gene in the pied flycatcher: using biological archives to track global climate change. Ecol. Evol. 3, 4799—4814.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754—1760.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493—496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., the 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078—2079.

Li, J., Liao, X., Yang, H., 2000. Molecular characterization of a parasitic tapeworm (*Ligula*) based on DNA sequences from formalin-fixed specimens. Biochem. Genet. 38, 309—322.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362, 709—715.

Mardis, E.M., 2013. Next generation sequencing platforms. A. Rev. Anal. Chem. 6, 287—303.

Markolf, M., Rakotonirina, H., Fichtel, C., von Grumbkow, P., Brameier, M., Kappeler, P.M., 2013. True lemurs…true species — species delimitation using multiple data sources in the brown lemur complex. BMC Evol. Biol. 13, 233.

Mason, V.C., Li, G., Helgen, K.M., Murphy, W.J., 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome Res. 21, 1695—1704.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526—538.

Mende, M.B., Hundsdoerfer, A.K., 2013. Mitochondrial lineage sorting in action — historical biogeography of the *Hyles euphorbiae* complex (Sphingidae, Lepidoptera) in Italy. BMC Evol. Biol. 13, 83.

Menzies, B.R., Renfree, M.B., Heider, T., Mayer, F., Hildebrandt, T.B., Pask, A.J., 2012. Limited genetic diversity preceded extinction of the Tasmanian tiger. PloS One 7, e35433.

Meyer, M., Stenzel, U., Hofreiter, M., 2008. Parallel tagged sequencing on the 454 platform. Nat. Protocols 3, 267—278.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M.V., Derevianko, A.P., Patterson, N., Andrés, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222—226.

Miething, F., Hering, S., Hanschke, B., Dressler, J., 2006. Effect of fixation to the degradation of nuclear and mitochondrial DNA in different tissues. J. Histochem. Cytochem. 54, 371—374.

Mikheyev, A.S., Bresson, S., Conant, P., 2009. Single-queen introductions characterize regional and local invasions by the facultatively clonal little fire ant *Wasmannia auropunctata*. Mol. Ecol. 18, 2937—2944.

Miller, C.R., Waits, L.P., Joyce, P., 2006. Phylogeography and mitochondrial diversity of extirpated brown bear (*Ursus arctos*) populations in the contiguous United States and Mexico. Mol. Ecol. 15, 4477—4485.

Miller, L.M., Kapuscinski, A.R., 1997. Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. Genetics 147, 1249—1258.

Miller, W., Drautz, D.I., Janecka, J.E., Lesk, A.M., Ratan, A., Tomsho, L.P., Packard, M., Zhang, Y., McClellan, L.R., Qi, J., Zhao, F., Gilbert, M.T.P., Dalén, L., Arsuaga, J.L., Ericson, P.G.P., Huson, D.H., Helgen, K.M., Murphy, W.J., Götherström, A., Schuster, S.C., 2009. The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). Genome Res. 19, 213—220.

Monda, K., Simmons, R.E., Kressirier, P., Su, B., Woodruff, D.S., 2007. Am. J. Primatol. 69, 1285—1306.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation — mitochondrial-gene trees versus nuclear-gene trees. Evolution 49, 718—726.

Moraes-Barros, N., Morgante, J.S., 2007. A simple protocol for the extraction and sequence analysis of DNA from study skin of museum collections. Genet. Mol. Biol. 30, 1181—1185.

Morin, P.A., Chambers, K.E., Boesch, C., Vigilant, L., 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Mol. Ecol. 10, 1835—1844.

Morin, P.A., Archer, F.I., Foote, A.D., Vilstrup, J., Allen, E.E., Wade, P., Durban, J., Parsons, K., Pitman, R., Li, L., Bouffard, P., Nielsen, S.C.A., Rasmussen, M., Willerslev, E., Gilbert, M.T.P., Harkins, T., 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res. 20, 908—916.

Moritz, C., Patton, J.L., Conroy, C.J., Parra, J.L., White, G.C., Beissinger, S.R., 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science 322, 261—264.

Mundy, N.I., Unitt, P., Woodruff, D.S., 1997. Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. Auk 114, 126—129.

Nagy, Z.T., 2010. A hands-on overview of tissue preservation methods for molecular genetic analyses. Org. Divers. Evol. 10, 91–105.

Nielsen, E.E., Hansen, M.M., Loeschcke, V., 1997. Analysis of microsatellite DNA from old scale samples of Atlantic salmon Salmo salar: a comparison of genetic composition over 60 years. Mol. Ecol. 6, 487–492.

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., Zwick, M.E., 2007. Microarray-based genomic selection for high-throughput resequencing. Nature Methods 4, 1109–1112.

Overballe-Petersen, S., Orlando, L., Willerslev, E., 2012. Next-generation sequencing offers new insights into DNA degradation. Trends Biotech. 30, 364–368.

Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. 86, 1939–1943.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., Hofreiter, M., 2004. Genetic analyses from ancient DNA. A. Rev. Genet. 38, 645–679.

Pikó, L., Taylor, K.D., 1987. Amounts of mitochondrial-DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Dev. Biol. 123, 364–374.

Prendini, L., Hanner, R., DeSalle, R., 2002. Obtaining, storing and archiving specimens and tissue samples for use in molecular studies. In: DeSalle, R., Giribet, G., Wheeler, W. (Eds.), Techniques in Molecular Systematics and Evolution. Birkhaeuser, Basle, pp. 176–248.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Rohland, N., Hofreiter, M., 2007. Comparison and optimization of ancient DNA extraction. BioTechniques 42, 343–352.

Rohland, N., Siedel, H., Hofreiter, M., 2004. Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. BioTechniques 36, 814–816, 818–821.

Rohland, N., Siedel, H., Hofreiter, M., 2010. A rapid column-based ancient DNA extraction method for increased sample throughput. Mol. Ecol. Resour. 10, 677–683.

Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K., Moritz, C.C., 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. Mol. Ecol. Resour. 11, 1082–1092.

Roy, M.S., Girman, D.J., Taylor, A.C., Wayne, R.K., 1994. The use of museum specimens to reconstruct the genetic variability and relationships of extinct populations. Experientia 50, 551–557.

Rubidge, E.M., Patton, J.L., Lim, M., Burton, A.C., Brashares, J.S., Moritz, C., 2012. Climate-induced range contraction drives genetic erosion in an alpine mammal. Nat. Climate Change 2, 285–288.

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

Sosa, M.X., Sivakumar, I.K.A., Maragh, S., Veeramachaneni, V., Hariharan, R., Parulekar, M., Fredrikson, K.M., Harkins, T.T., Lin, J., Feldman, A.B., Tata, P., Ehret, G.B., Chakravarti, A., 2012. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. PLoS Comp. Biol. 8, e1002737.

Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., Bakker, F.T., 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PloS One 8, e69189.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P., Bouvet, J., 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res. 24, 3189–3194.

Taylor, A.C., Sherwin, W.B., Wayne, R.K., 1994. Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat Lasiorhinus krefftii. Mol. Ecol. 3, 277–290.

Thalmann, O., Wegmann, D., Spitzner, M., Arandjelovic, M., Guschanski, K., Leuenberger, C., Bergl, R.A., Vigilant, L., 2011. Historical sampling reveals dramatic demographic changes in western gorilla populations. BMC Evol. Biol. 11, 85.

Thomas, R.H., Schaffner, W., Wilson, A.C., Pääbo, S., 1989. DNA phylogeny of the extinct marsupial wolf. Nature 340, 465–467.

Thomas, W.K., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. 101–112.

Tin, M.M.-Y., Economo, E.P., Mikheyev, A.S., 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. PLoS One 9, e96793.

Vallender, E.J., 2011. Expanding whole exome resequencing into non-human primates. Genome Biology 12, R87.

Vallianatos, M., Lougheed, S.C., Boag, P.T., 2002. Conservation genetics of the loggerhead shrike (Lanius ludovicianus) in central and eastern North America. Conserv. Genet. 3, 1–13.

van Orsouw, N.J., Hogers, R.C., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., Schneiders, H., van der Poel, H., van Oeveren, J., Verstegen, H., van Eijk, M.J., 2007. Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One 14, e1172.

van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods 5, 247–252.

Wandeler, P., Smith, S., Morin, P.A., Pettifor, R.A., Funk, S.M., 2003. Patterns of nuclear DNA degeneration over time – a case study in historic teeth samples. Mol. Ecol. 12, 1087–1093.

Wandeler, P., Hoeck, P.E., Keller, L.F., 2007. Back to the future: museum specimens in population genetics. Trends Ecol. Evol. 22, 634–642.

Willerslev, E., Cooper, A., 2005. Ancient DNA. Proc. R. Soc. B 272, 3–16.

Wyatt, K.B., Campos, P.F., Gilbert, M.T.P., Kolokotronis, S.-O., Hynes, W.H., DeSalle, R., Ball, S.J., Daszak, P., MacPhee, R.D.E., Greenwood, A.D., 2008. Historical mammal extinction on Christmas Island (Indian Ocean) correlates with introduced infectious disease. PLoS One 3, e3602.

Zimmermann, J., Hajibabaei, M., Blackburn, D.C., Hanken, J., Cantin, E., Posfai, J., Evans, T.C., 2008. DNA damage in preserved specimens and tissue samples: a molecular assessment. Frontiers in Zool. 5, 18.