Conservation Genet Resour
DOI 10.1007/s12686-011-9545-x

TECHNICAL REVIEW

Guidelines for collecting and maintaining archives for genetic monitoring

Jennifer A. Jackson · Linda Laikre · C. Scott Baker · Katherine C. Kendall · The Genetic Monitoring Working Group

Received: 20 April 2011 / Accepted: 20 September 2011

Abstract Rapid advances in molecular genetic techniques and the statistical analysis of genetic data have revolutionized the way that populations of animals, plants and microorganisms can be monitored. Genetic monitoring is the practice of using molecular genetic markers to track changes in the abundance, diversity or distribution of populations, species or ecosystems over time, and to follow adaptive and non-adaptive genetic responses to changing external conditions. In recent years, genetic monitoring has become a valuable tool in conservation management of biological diversity and ecological analysis, helping to illuminate and define cryptic and poorly understood species and populations. Many of the detected biodiversity declines, changes in distribution and hybridization events have helped to drive changes in policy and management. Because a time series of samples is necessary to detect trends of change in genetic diversity and species composition, archiving is a critical component of genetic monitoring. Here we discuss the collection, development, maintenance, and use of archives for genetic monitoring. This includes an overview of the genetic markers that facilitate effective monitoring, describes how tissue and DNA can be stored, and provides guidelines for proper practice.

Keywords Conservation · Museum DNA · Biodiversity · Molecular markers · Biological collections

The NCEAS/NESCent Working Group on Genetic Monitoring (GeM) is chaired by F. W. Allendorf, University of Montana, USA, and M. K. Schwartz, USDA Forest Service, USA. The other members are: C. S. Baker (Oregon State University, USA), D. P. Gregovich (University of Alaska, USA), M. M. Hansen (Aarhus University, Denmark), J.A. Jackson (Oregon State University, USA), K. C. Kendall (US Geological Survey, USA), L. Laikre (Stockholm University, Sweden), K. McKelvey (USDA Forest Service, USA), M. C. Neel (University of Maryland, USA), I. Olivieri (Université de Montpellier II, France), N. Ryman (Stockholm University, Sweden), R. Short Bull (University of Montana, USA), J.B. Stetz (University of Montana, USA), D. A. Tallmon (University of Alaska, USA), C. D. Vojta (US Forest Service, USA), D. M. Waller (University of Wisconsin, USA) and R. S. Waples (National Marine Fisheries Service, USA).

J. A. Jackson
Marine Mammal Institute, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA

Present Address:
J. A. Jackson (✉)
British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
e-mail: Jennifer.Jackson@bas.ac.uk

L. Laikre
Division of Population Genetics, Department of Zoology, Stockholm University, 10691 Stockholm, Sweden

C. S. Baker
Marine Mammal Institute and Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA
e-mail: scott.baker@oregonstate.edu

C. S. Baker
School of Biological Sciences, University of Auckland, Auckland, New Zealand

K. C. Kendall
US Geological Survey, Northern Rocky Mountain Science Center, Glacier National Park, West Glacier, MT, USA

Since the origins of human societies, marine and terrestrial plant and animal populations have been subject to a variety of anthropogenic environmental impacts including loss of habitat, direct exploitation and encroachment of introduced

Published online: 16 October 2011

species (including pathogens). Impacts on biodiversity include changes in species abundance and distribution and loss of genetic diversity (Frankham 2005; Wright et al. 2008). Methods for assessing and monitoring these types of changes are necessary to assure conservation and sustainable use of our remaining biodiversity. Various molecular genetic techniques are now becoming affordable and reliable approaches in this respect (Schwartz et al. 2007).

Genetic monitoring has been defined as "quantifying temporal changes in population genetic metrics or other population data generated using molecular markers" (Schwartz et al. 2007). Integral to genetic monitoring is the interpretation of individual and population genetic data in the context of ecological and evolutionary processes, particularly in human impacted environments. In addition, genetic monitoring can provide valuable baseline information to evaluate population responses to future global environmental changes, such as global warming. Genetic monitoring can be used to monitor population processes in elusive and cryptic species that cannot be directly counted, e.g., by using DNA obtained from feces, shed hair, feathers, skin and scales (Proctor et al. 2005; Boulanger et al. 2004; Piggott et al. 2006; Prugh et al. 2005), or from hunter kills, market products or incidental mortality (e.g., fisheries by-catch, road-kills) (Pichler and Baker 2000; Bellinger 2003; Baker 2008). "Resurrection" analyses of museum-collected specimens or other artifacts (e.g., Austin and Melville 2006; Kelley Thomas et al. 1990; Groombridge et al. 2000) can also provide an historical baseline for comparison with current estimates of species abundance.

Genetic monitoring projects require a time series of archived genetic data, either in the form of specimen tissue, extracted DNA, or records of previously obtained genetic information (e.g., DNA sequences or genotypic data). For the purposes of genetic monitoring, DNA and tissue archives are spatially and temporally explicit, intentional collections of individuals from the population of interest, with multiple samples obtained from each period of collection. In contrast, the temporal or spatial spread of traditional historical archives is often sporadic or unknown.

Although methods of genetic monitoring have received considerable attention, less attention has been given to the archiving of samples required to detect trends over time. Here we review the techniques available for generating and maintaining tissue and DNA archives for genetic monitoring, discuss challenges facing archivists, geneticists, and managers, and present a series of guidelines for archiving material in order to facilitate genetic monitoring of wild plants and animals. We consider archives that span both genetic monitoring categories defined by Schwartz et al. (2007). Category I encompasses the use of genetic markers as *identifiers* of individuals, populations and species for traditional population monitoring purposes. At the individual level, genetic identification can enable estimation of population abundance and vital rates within the framework of mark-recapture models. Genetic species identification can be used to monitor changes in distribution through occupancy modeling that incorporates detection probability (MacKenzie et al. 2006). Category II represents the use of genetic markers to monitor changes in population genetic parameters, e.g., amount of genetic variation, degree of population divergence, rate of gene flow, and effective population size (N_e). We focus on the use of both modern and historical DNA archives for genetic monitoring but do not discuss the use of ancient DNA, as this has been covered elsewhere (e.g., Leonard 2008; Pääbo et al. 2004).

Construction of archives

The ability of archival time series to detect changes depends on a combination of factors, including the generation time of the species of interest relative to the age of the archival data, the number and distribution of individuals sampled at each time point, the preservation of material for DNA analysis, and the genetic marker types employed to perform the analysis. The success of genetic monitoring thus crucially hinges on the quality, age, and size of the genetic archive available for the species or population of interest.

Marker types and their utility for genetic monitoring

Genetic monitoring schemes that estimate abundance or monitor population changes require variable genetic markers that allow identification of individuals or population level diversity, in order to identify changes in the abundance or diversity of individuals or populations through time. Commonly used genetic marker types include mitochondrial and chloroplast DNA, nuclear introns, microsatellites, single nucleotide polymorphisms (SNPs) and amplified fragment length polymorphisms. Attributes of these markers and their utility in molecular ecology have been discussed elsewhere (Sunnucks 2000; Morin et al. 2004; Selkoe and Toonen 2006).

The utility of different types of molecular markers for genetic monitoring depends on the quality and quantity of DNA available for analysis. For example, large amounts of good quality DNA are required to amplify microsatellites and screen for SNPs. These markers can be used to estimate population size, bottlenecks and kinship and determine sex and identity (Selkoe and Toonen 2006). Microsatellite loci are prone to amplification errors when DNA quantity or quality is low, biasing amplification toward shorter fragment lengths (Taberlet et al. 1999) or to

just one allele at a locus (i.e., allelic dropout). Furthermore, microsatellite data are often difficult to compare between laboratories and studies when results have not been calibrated to published size designations, or when individuals use different size bins for classifying loci. Inferred allele sizes can vary from true allele sizes for a variety of reasons (Morin et al. 2010). Some technical fixes have been proposed to surmount these difficulties (McKelvey and Schwartz 2005). For example, when datasets are coordinated among laboratories, allelic size bins can be based on the same reference dataset, fragment size names can be standardized, and all names can refer back to the smallest fragment described by the designated laboratory and shared among the reference datasets in all laboratories (e.g., Stephenson et al. 2009). Many publications detail optimal practice protocols to avoid these common problems (e.g., Kendall et al. 2009; Selkoe and Toonen 2006; Morin et al. 2010; Paetkau 2003; Roon et al. 2005; Stephenson et al. 2009).

While the use of SNPs has until recently been limited by the high cost of screening and sequencing, SNPs are considered to be easier to standardize across laboratories and platforms and can therefore provide a more robust comparison of population genetic characteristics over studies and time (Morin et al. 2004).

Mitochondrial and chloroplast markers are frequently used in genetic monitoring projects to describe population origins, hybrid introgression, geographical distribution of species and population mixing (e.g., Wirgin et al. 1997; Bowen et al. 2007; Petit et al. 1998). These markers evolve too slowly to permit estimation of individual identity and abundance (Category I monitoring). As they are clonally inherited, they only represent the evolutionary history of a single locus, therefore a limited inference of the organism genetic history. However, they can provide estimates of changes in species abundance and distributions through changes in diversity and haplotype richness, as well as mixed-stock approaches (Manel et al. 2005). They also have the practical advantage of being present in large copy numbers in cells, and are thus often the only markers available for PCR amplification of highly degraded samples, (e.g., those up to 47,000 years old, Hagelberg et al. 1989). We summarize the comparative advantages of different marker types in Table 1.

Intentional genetic archives

Tissue archives

In recent years, collection of animal tissues for the dedicated purpose of genetic sampling has increased, motivated by the difficulty of obtaining adequate genetic material from museum collections and a growing demand for genetic analyses to complement traditional taxonomic approaches to species identification. Such archives are collected for a variety of purposes, most recently for DNA taxonomic studies (e.g., the Barcode of Life project, Hajibabaei et al. 2005), management of protected species (e.g., sea turtles and cetaceans) and for monitoring and experimental manipulation of domesticated animals and plants (Baxevanis 2003).

For genetic archives there is a direct relationship between the cost of specimen preservation and the quality of the preserved DNA. A summary of common preservation techniques is given in Table 2. One of the most widespread and affordable means of storing tissue is in a high-concentration ethanol solution (>70%) or in a lysis buffer containing EDTA. Concentrated ethanol minimizes water content (which can hydrolyze DNA if it is acidic), while lysis buffer is thought to protect DNA from degradation during DNA extraction by chelating the metal ions that are required for most DNA degradation processes (Kilpatrick 2001). Storage of tissue in ethanol preserves DNA for future monitoring purposes, but does not preserve other components of the living biochemistry of tissues, including RNA and proteins. Storage at −80°C preserves good quality DNA indefinitely by halting DNA degradation, but permanent frozen storage presents logistical disadvantages. For example, accessing samples requires the DNA-degrading practice of thawing and sometimes re-freezing. Thawing can also occur during power outages or freezer malfunctions. A combination of low temperature and ethanol storage helps prevent degradation during thaws.

Cryogenic storage (−196°C) is traditionally the 'gold standard' for long-term tissue storage. These archives preserve tissue biochemistry as well as DNA, providing the maximum amount of information on species taxonomy, geography, genetic history, life history, diseases, parasites and the state of the species' environment. As such they provide a valuable insight into the status of the species in its environment at the time of collection and represent the ideal storage method for monitoring the biochemistry and genetics of wildlife populations. However, this approach, like others, preserves a finite amount of tissue, which can be depleted through repeated sampling for analysis (see "Access protocols and guidelines").

The rewards of cryogenically archiving tissue are high in the long term, as they provide good quality genomic material for an indefinite period. However, these archives are labor intensive to set up, take substantial space and energy to maintain, and require a backup to prevent loss of data during power failure. Funding limitations prevent most institutions and agencies from creating to maintaining such archives (e.g., Edwards et al. 2005) despite their obvious advantages.

Table 1 Comparison of various marker types of use in genetic monitoring

Marker type	Variability	# Loci	Ease of optimization	Comparability between studies	Range of available analyses	References
Mitochondrial and chloroplast						
DNA sequence, e.g., control region, cytochrome oxidase 1	+	+	+++	+++	A1, A2, A3 B1, B2	Sunnucks (2000), Schwartz et al. (2007)
Multilocus nuclear						
AFLP: Amplified Fragment Length Polymorphisms	++	++	++	+	A1 B1, B2, B4, B5, B6	Mariette et al. (2002), Sunnucks (2000), Schwartz et al. (2007)
Microsatellite arrays: Hyper-variable, co-dominant markers	+++	++	+	+	A1 B1, B2, B4, B5, B6, B7	Schwartz et al. (2007), Selkoe and Toonen (2006), Sunnucks (2000)
SNP: Co-dominant Single nucleotide polymorphisms	+++	++	++	++/+++	A1 B1, B2, B4, B5, B6, B7	Morin et al. (2004), Schwartz et al. (2007), Sunnucks (2000)
Genomic sequences or Reduced Representation Shotgun Sequencing	+++	+++	N/A	+++	A1, A2, A3 B1, B2, B3, B4, B5, B6, B7	Allendorf et al. (2010)
Allozyme: Single locus nuclear protein	+	++	+	+++	A1, A2 B1, B2, B4, B5	Jorde and Ryman (1996), Sunnucks (2000), Palm et al. (2003), Schwartz et al. (2007)

+ (low), ++ (medium), +++ (high). Range of available analyses for genetic monitoring. (A) Species level: hybridization (A1), changes in distribution (A2), identification of pathogens or parasites (A3). (B) Population and individual level: genetic diversity (B1), mixture proportions (B2), measurement of adaptive change (B3), effective population size (B4), population structure and dispersal (B5), population abundance (B6), vital rates, e.g., survival (B7)

Most animal and plant tissues may also be preserved by freeze drying using liquid nitrogen (Leboeuf et al. 2008; Murphy et al. 2000), or by using silica beads or some other desiccant (e.g., blotting blood or tissue onto filter paper) at room temperature (Murphy et al. 2000; Wasser et al. 1997). Rapid freezing can disrupt the cellular ultra-structure of the tissue, while silica beads are often easier to transport in the field than liquid nitrogen. Optimal drying methods vary with field conditions, the species of interest and the planned use of the sample (see Prendini et al. 2002). DNA can be effectively and inexpensively preserved by drying tissue onto a fixation matrix: a number of these are now commercially available (usually with salts added to bind PCR inhibitors, e.g., Smith and Burgoyne 2004; Makowski et al. 1998). Archiving dried tissues holds distinct advantages where the number of samples collected is large relative to available archival space or energy. Some fixation methods possess advantages for field collection where access to specimens is time-restricted (i.e., samples can be quickly rubbed onto paper). However, the long-term (>25 years) prospects for DNA amplification of dried tissue are unknown, and its use for other biochemical analyses (e.g., RNA or toxicology) is limited.

An unlimited source of genetic material for analysis can be provided through cellular tissue cultures. Cell culture techniques are routinely used in cancer and drug discovery studies (e.g., ATCC Global Bioresource Center). However, the utility of such approaches for genetic monitoring projects is currently limited by the technical challenge of culturing, along with the cost of subsequent cryogenic storage. Presently only a few institutions routinely use this approach to preserve cells from endangered wildlife (e.g., the 'Frozen Zoo'[1] of the San Diego Zoo).

DNA archives

By comparison to tissue storage, there are few definitive studies regarding the optimal storage conditions for extracted DNA, as optimum conditions vary with each laboratory environment and species product being archived. There are however, a number of suggested best practice guidelines (e.g., Prendini et al. 2002; Morin et al. 2010). DNA for genetic monitoring is commonly stored dry, or in a neutral pH buffer with chelating agents such as EDTA which sequesters excess positively charged ions, and is often kept at low temperatures (e.g., −20°C). Dry storage requires that DNA be protected from heat, moisture

[1] http://www.sandiegozoo.org/conservation/science/at_the_zoo/the_frozen_zoo/.

Table 2 Merits of archival preservation for genetic monitoring research

Tissue preservation	Markers	DNA quality	Energy efficiency (space/electricity)	Ease of use in field	Repeat use impact on sample quality	Notes
Formalin	M, short μsats	−	+++	N/A	N/A	Special protocols required
Drying (air, silica)	M, C, μsats for some species	+	+++	N/A	N/A	Historical specimens available, some parts preserve DNA reasonably well (e.g., fish scales, bone, hair). Access to existing collections limited
>70% Ethanol	M, C, N, μsats, SNP	++	+++	++	++	DNA slowly becomes acidified unless ethanol is replaced or buffered, other biochemical information from tissue degrades (i.e., enzymes, proteins, pathogens)
Saturated salt solution with Dimethyl sulfoxide (DMSO)	M, C, N, μsats, SNP	++	+++	++	++	Good DNA preservation reported for birds, mammals and invertebrates. Variable among taxa and studies
Freezing (−20 to −80°C)	M, C, N, μsats, SNP	++	++	+	+	
Nitrogen preservation (−196°C)	M, C, N, μsats, SNP	+++	+	−	+	
DNA preservation						
DNA extraction in EDTA/TE buffer, storage at −20°C	M, C, N, μsats, SNP	+++	++	+	+	
DNA extraction, dried, bound to nuclease-inhibiting substrate	M, C, N, μsats, SNP	++	+++	+++	++	DNA may degrade if improperly preserved, exposed to heat or UV
WGA in EDTA/TE buffer, storage at −80°C	M, C, N, μsats, SNP	++	+	+	+	
WGA dried	M, C, N, μsats, SNP	++	+++	++	++	DNA degrades if improperly preserved, exposed to heat or UV

− (poor), + (low), ++ (medium), +++ (high). Markers: M (mtDNA), C (cpDNA), μsat (microsatellites), N (nuclear sequences), SNP (Single nucleotide polymorphism). *WGA* whole genome amplification, *mtDNA* mitochondrial DNA, *cpDNA* chloroplast DNA, *SNP* single nucleotide polymorphism, *TE* Tris–EDTA, *EDTA* ethylenediaminetetraacetic acid

and UV light. Protection is enhanced by the presence of the sugar trehalose (Smith and Morin 2005), which is thought to form hydrogen bonds, enhancing DNA stability and thus slowing degradation. If these conditions can be maintained, it is an affordable approach that requires minimum storage space and energy requirements to maintain. Lower temperatures slow the rate of buffered DNA degradation. However, the long-term preservation potential for DNA extractions is not yet well understood, and more studies of the effects of preservation types on DNA quality will surely benefit the field.

"Endless DNA"—whole genome amplification

Some of the problems of access and limitations of long-term tissue storage can be overcome by whole genome amplification (WGA). An archive of WGA can provide a long-term source of 'endless DNA', which can be sampled repeatedly with no depletion risk. WGA acts to increase the quantity of DNA using a high-processively DNA polymerase with multiple-strand displacement amplification (e.g., Dean et al. 2002) or through ligation-mediated PCR. This latter technique is preferred when available samples are small, and/or degraded since it is specialized to amplify DNA from short fragments (Hughes 2005). There are no data regarding the long-term quality of such archives, but storage at −20°C in a protective EDTA buffer, or dry and protected from light, are currently preferred methods. Disadvantages of this approach depend on the amplification protocol employed, but can include slight degradation of the genetic information (e.g., SNP error rates slightly increased, Teo et al. 2008) and short length amplified

fragments. However, this approach can produce high molecular weight DNA (>10 kilo-bases in length) from genomic samples of size ~1 ng (although a minimum of 10 ng is recommended). As synthetic DNA, WGA is not subject to the Convention on International Trade in Endangered Species (CITES) restrictions on international exchange (Bowen and Avise 1994). This facilitates genetic monitoring projects where the populations or species of interest are internationally distributed, or when research laboratories are not within the endemic range (Janecka et al. 2006; Baker 2008). However the Nagoya Protocol on Access to Genetic Resources, recently adopted to the Convention on Biological Diversity (Secretariat of the CBD 2011; http://www.cbd.int), will most likely affect the international transfer and use of WGA.

Collection of archives

Sampling strategies and timeframes

Sampling strategies reflect a trade-off between efforts to collect tissue samples from the maximum possible number of individuals across the geographic range of the population, and the accessibility of these individuals and samples for collection. Collection methods vary greatly depending on the species of interest and its general accessibility. Non-invasive field collections, e.g., hair, scat or shed feathers, can be developed in the absence of sightings of the species in question, which can be advantageous where the species is rare, cryptic or dangerous to handle. Although these collections are low-cost relative to invasive sampling strategies, they require careful storage in the field, as they are subject to rapid degradation after separating from the animal. Storage measures should minimize moisture, exposure to UV radiation, and may also be preserved by addition of preservative agents. Minimally invasive field collections of e.g., blood or biopsy samples, provide larger quantities of higher quality DNA, but are usually more limited in number and distribution due to the additional costs of collection. When planning a comprehensive survey, it is recommended that an initial pilot project be carried out to identify potential problems with handling and preservation (Morin et al. 2010).

In general, the more individuals that can be sampled and the more regions surveyed, the better for subsequent analysis of abundance. Exhaustive surveys are rarely possible in the wild, so it is usually more useful to know instead what the minimum survey sizes should be in order to provide the degree of precision required for the parameters being monitored. Simulations suggest that genetic samples of 10–60% of the total population may be required to detect biologically realistic changes in abundance and effective population size (an increase or decrease of 10% per generation) for populations of 100–500 individuals, although this number will vary with organism life history, the extent of population change and the time period over which monitoring is performed (Tallmon et al. 2010). Appropriate time frames for sample collection depend on organism life history regimes, most notably the degree of generational overlap and average generation time. Typically, sampling timeframes are influenced by ease of collection which may vary by season, geographical range, type of sampling and availability of funding. A web-based initiative to provide tailored genetic monitoring advice for managers can be found at http://www.alaska.fws.gov/gem/mainPage_1.htm; Stetz et al. (in press).

Development and maintenance of archives

Techniques of laboratory inventory management

Laboratory inventory management systems (LIMS) are a framework for recording comprehensive inventories of archival holdings. This framework facilitates archiving and accessing samples, associates samples with ancillary data (e.g., sample type, collection times and locations), and monitors their progress through analytical stages via regular user input. Many LIMS are commercially available, and a few are open source. The most successful to date are those which can allow networking between inventories from other institutions and holdings without substantial time and investment in re-coding data, e.g., the universal online LIMS integrated into the Barcode of Life project (http://www.barcodinglife.org/). This is an important consideration when choosing LIMS software.

Access protocols and guidelines

Access protocols are commonly used, as they enable resource-limited archivists to balance the cost of protecting collections from depletion and preparing and sending material to individuals requesting access, against the benefit of providing access to further the advancement of biological science. Many institutions and museums only allow institutionally based researchers access to specimens, and usually only those with funded research projects of demonstrable scientific value. In addition to these minimum requirements, restrictions are based on the number, rarity and replace-ability of the samples requested, and availability of alternative sources of material to the researcher. Implementing the Nagoya Protocol on Access to Genetic Resources (Secretariat of the CBD 2011; http://www.cbd.int) will also require further development of national and international legal arrangements with

respect to access to genetic and other biological diversity (Krusar 2011).

Ease of access depends on the archive type and storage conditions; for example it might be possible to sub-sample from dried to ethanol-preserved samples without affecting the quality of the archived tissue, while repeatedly accessing frozen samples may increase tissue and DNA degradation as a result of repeated thawing and re-freezing (e.g., Ohsako et al. 1997). Curators usually consider these factors and make access decisions on a case-by-case basis, particularly when samples are rare and/or depleted. Nearly all archives request that acknowledgements be published in literature arising from use of the collections, and that genetic sequence data be submitted to a public-access repository such as GenBank.

An excellent example of a Material Transfer Agreement protocol was developed by the Southwest Fisheries Science Center Cetacean and Marine Turtle Archive (La Jolla, USA, http://www.swfsc.noaa.gov/textblock.aspx?Division=PRD&ParentMenuId=229&id=12498). The facility has established a loans committee to discuss requests for genetic samples. This considers written proposals from qualified investigators and bases lending decisions on the importance of the study objectives, quality of experimental design, technical feasibility and sample availability. Once loans have been agreed to, the recipients must confirm in writing that they will provide all sequence or genotype data once manuscripts have been accepted for publication. They must also provide annual reports on research progress, cite accession numbers in all publications, provide the center with reprints of manuscripts, use specimens only for the proposed research, and not distribute these samples to others. In this way a balance is struck between the logistic difficulties of providing access to archival data, and the scientific and conservation benefits of enabling public access to collections. It should be noted that since most access protocols are not yet legally binding, they are often subject to non-compliance after samples have been sent. The content of access protocols may change when countries adopt the Convention on Biological Diversity′s Nagoya Protocol as it requires parties to take 'appropriate, effective and proportionate legislative, administrative or policy measures' to ensure that genetic resources utilized within their jurisdiction have been accessed in accordance with the legislation and requirements of the party that provided them. This agreement creates a need for national legislation within both the provider and the user country, but as yet only a few countries have drafted national legislation on such access and benefit sharing (Morgera and Tsioumani 2011).

Data ownership

An archive is often developed through the efforts of many individual sample collectors. We advise clarifying the property rights of such collectors with regard to subsequent usage of these samples, either through a signed agreement or memorandum of understanding when the specimen is added to the archive. Otherwise, subsequent usage of archived information may require lengthy consultation with multiple collectors. The delays associated with negotiating permissions and acknowledgements may prevent inclusion of some samples/data in projects with limited time frames. Large-scale projects (such as Genome 10 K, http://www.genome10k.org) therefore usually attempt to negotiate this in advance (e.g., Haussler et al. 2009). These considerations are particularly pertinent to genetic monitoring projects, as analyses of archival data are expected to occur for many years post collection, when members of the original research team may have retired or died and details of collection rights are less clearly recalled.

Institutional, public, and research archives: retiring your collection?

DNA archives are normally subject to the intellectual and physical property rights of the collectors, or the institutions they work for. Many individual researchers have amassed invaluable private collections of samples, the fate of which is uncertain. A few central repositories have been established for private collections: two prominent examples include the ATCC global bio-resource center (a centralized collection of cell lines) and the Ocean Genomic Legacy (a repository for donated tissue samples, htpp://www.oglf.org). Museums often accept donations of collections if they are well labeled and preserved, but may not have appropriate facilities for storage or resources for loans and distribution. Large repositories that have been archived on-line and are able to offer reasonable access protocols represent a good target for donations, as they can provide the best infrastructure for making samples available to other researchers wishing to conduct genetic monitoring.

Archives of genetic metadata

Primary genetic datasets arising from genetic monitoring efforts and analyses (e.g., microsatellite allele sizes and bins, DNA sequence data and chromatograms, and unphased allelic data) are necessary for accurate comparisons between studies and to provide a baseline for future monitoring efforts. Most journals publishing articles describing new DNA sequence data require that these data are made available on the open-access GenBank database (htpp://www.ncbi.nlm.nih.gov). We recommend that any genetic monitoring data lodged in this repository provide detailed information on geographic collection localities (not currently a mandatory requirement of sequence submission on

Table 3 General guidelines for the collection and maintenance of an archive for genetic monitoring

Project Stage	Considerations	Recommendations	References
Study design	Species distribution; longevity; accessibility; timeframe of survey	Careful planning of sample strategy in order to survey as widely as possible across the geographic distribution of species. Consider sample sizes in relation to the type of genetic monitoring to be performed. Consider sampling times in relation to the average generation time of the species of interest	Felsenstein (2006), Ryman et al. (2006), Larsson et al. (2009), Tallmon et al. (2010), Morin et al. (2010)
Sample collection	Type of sample; size of sample; storage and transport in field	Ethanol storage for tissue samples. If drying samples, keep free from moisture, contamination and UV light. Use preservatives where appropriate, e.g., for scat	Prendini et al. (2002). Also see Table 1
Sample archiving	Type of sample; storage space and funding; markers of interest; time-frame of study	Store tissue as appropriate for marker types of particular interest, e.g., in liquid nitrogen (maximal usage of samples), stored at −80°C, stored in ethanol, dried. Consider degrading impacts of freeze-thawing in storage protocol. Consider Whole Genome Amplification for samples which will be regularly used and shared with collaborators. Store DNA dry or in a neutral pH buffer. Provide Memoranda of Understanding and access protocols, in order to clarify relationship of collectors to the archive collection, and to control usage of materials by interested researchers	Prendini et al. (2002), Morin et al. (2010), Hughes (2005)
Data generation	Type of survey e.g., abundance, population structure, species ID	Thoroughly evaluate error at each stage of data generation. Make markers independently replicable and directly comparable between studies where possible e.g., DNA sequence data. If microsatellites, provide open access to raw data, and calibrate allelic size bins by comparison with a widely used reference dataset	Morin et al. (2010). Also see Table 2
Data archiving	Permanently associated with sample archive in order to allow for future replication or updates. Comparable between independent studies and over time	Database data with a laboratory inventory management system. Lodge DNA data in GenBank and primary data as open access datasets in appropriate repository e.g., DRYAD and DataOne	See http://www.barcodinglife.org, http://www.ncbi.nlm.nih.gov and http://www.dataone.org

GenBank) and ecological habitat data, where available, to facilitate comparison between studies and subsequent monitoring. It is also extremely useful to make available primary data generated for microsatellites or SNPs. Many evolutionary genetic journals and funding bodies now require that these data be made available upon publication (e.g., UK Research Councils, the US National Science Foundation and National Institutes of Health, Costello 2009; Howe et al. 2008). Public archiving of genetic metadata is facilitated by multi-disciplinary data repositories such as DRYAD (http://www.datadryad.org) and DataOne (https://www.dataone.org). Several major evolutionary and ecological journals have agreed to a Joint Data Archiving Policy whereby data used in each article submitted for publication must be archived with sufficient detail that the results can be re-created elsewhere (http://www.datadryad.org/jdap, Whitlock et al. 2010). Exceptions are granted with discretion, since there are some cases where detailed geo-spatial information on wildlife abundance could increase the threats to that species or population (e.g., through hunting and poaching).

The value of archives—into the future

The proper archiving of geographic and temporal data associated with each archival sample is crucial for current and future monitoring efforts (Table 3). Furthermore current efforts to extract DNA can provide only one of the many pieces of information we are able to obtain from samples in the future; RNA, proteins and other cellular material are all potentially available for the future if the best preservation methods are employed today. How we choose our samples, store them and allow access to them now dictates the standard of the science that we are able to achieve in the future (e.g., Haussler et al. 2009). It also determines our power to detect changes in population distribution and size for endangered and threatened species, and ultimately to provide the best conservation management advice for protecting those species. Bridging the gap between history and the present will continue to become easier with the development of protocols to non-invasively collect DNA, repair damaged DNA and amplify whole genomes from small samples. Future genomic and conservation initiatives are critically dependent on the archives of today.

Acknowledgments This work was conducted as part of the Genetic Monitoring (GeM) Working Group jointly supported by the National Evolutionary Synthesis Center (NSF #EF-0423641) and the National Center for Ecological Analysis and Synthesis, a center funded by NSF (NSF #EF-0553768), the University of California, Santa Barbara, and the State of California. L.L. also acknowledges the European Community's Seventh Framework Programme (FP/2007-2013) under grant agreement No 217246 made with BONUS, the joint Baltic Sea research and development programme. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Allendorf F, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709

Austin JJ, Melville J (2006) Incorporating historical museum specimens into molecular systematic and conservation genetics research. Mol Ecol Notes 6:1089–1092

Baker CS (2008) A truer measure of the market: the molecular ecology of fisheries and wildlife trade. Mol Ecol 17:3985–3998

Baxevanis AD (2003) The molecular biology database collection: 2003 update. Nucleic Acids Res 31:1–12

Bellinger MR (2003) Loss of genetic variation in greater prairie chickens following a population bottleneck in Wisconsin, USA. Conserv Biol 17(3):717–724

Boulanger J, Himmer S, Swan C (2004) Monitoring of grizzly bear population trends and demography using DNA mark-recapture methods in the Owikeno Lake area of British Columbia. Can J Zool 82:1267–1277

Bowen BW, Avise JC (1994) Conservation research and the legal status of PCR products. Science 266:713

Bowen BW, Grant WS, Hills-Starr Z, Shaver DJ, Bjornadal A, Bolten AB, Bass AL (2007) Mixed stock analysis reveals the migrations of juvenile hawksbill turtles (*Eretmochelys imbricate*) in the Caribbean sea. Mol Ecol 16:49–60

Costello MJ (2009) Motivating online publication of data. Bioscience 59(5):418–427

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci USA 99:5261–5266

Edwards SV, Birks S, Brumfield RT, Hanner R (2005) Future of avian genetic resources collections: archives of evolutionary and environmental history. Auk 122(3):979–984

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol Biol Evol 23(3):691–700

Frankham R (2005) Genetics and extinction. Biol Conserv 126(2): 131–140

Groombridge JJ, Jones CG, Bruford MW, Nichols RA (2000) 'Ghost' alleles of the Mauritius kestrel. Nature 403:616

Hagelberg E, Sykes B, Hedges R (1989) Ancient bone DNA amplified. Nature 342:485

Hajibabaei M, de Waard JR, Ivanova NV et al (2005) Critical factors for assembling a high volume of DNA barcodes. Philos Trans R Soc Lond Ser B Biol Sci 360:1959–1967

Haussler D et al (2009) Genome 10 K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered 100: 659–674

Howe D, Costanzo M, Fey P et al (2008) The future of biocuration. Nature 455:47–50

Hughes S (2005) Whole genome amplification. Methods express series. Scion Publishing Limited, Oxfordshire

Janecka JE, Grassman LI, Derr JN, Honeycutt RL, Eiadthong W, Tewes ME (2006) Rapid whole genome amplification of DNA from felids: applications for conservation genetics. Wildl Soc Bull 34(4):1134–1141

Jorde PE, Ryman N (1996) Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. Genetics 143:1369–1381

Kelley Thomas W, Pääbo S, Villablanca FX, Wilson AC (1990) Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J Mol Evol 31(2):101–112

Kendall KC, Stetz JB, Boulanger J, Macleod AC, Paetkau D, White GC (2009) Demography and genetic structure of a recovering grizzly bear population. J Wildl Manage 73(1):3–17

Kilpatrick CW (2001) Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. Biochem Genet 40(1/2):53–62

Krusar TA (2011) What are the implications of the Nagoya Protocol for research on biodiversity? Bioscience 61(4):256–257

Larsson LC, Charlier J, Laikre L, Ryman N (2009) Statistical power for detecting genetic divergence—organelle versus nuclear markers. Conserv Genet 10(5):1255–1264

Leboeuf C, Ratajczak P, Zhao W et al (2008) Long-term preservation at room temperature of freeze-dried human tumor samples dedicated to nucleic acids analysis. Cell Preserv Technol 6:191–198

Leonard JA (2008) Ancient DNA applications for wildlife conservation. Mol Ecol 17(19):1–11

MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, San Diego

Makowski GS, Davis EL, Hopfer SM (1998) Amplification of Guthrie card DNA: effect of guanidine thiocyanate on binding of natural whole blood PCR inhibitors. J Clin Lab Anal 11(2):87–93

Manel S, Gaggiotti O, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. Trends Ecol Evol 20:136–142

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. Mol Ecol 11:1145–1156

McKelvey KS, Schwartz MK (2005) DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. Mol Ecol Notes 5:716–718

Morgera E, Tsioumani E (2011) Yesterday, today and tomorrow: looking afresh at the Convention on Biological Diversity. University of Edinburgh School of Law Working Paper Series No 2011/21

Morin PA, Luikart G, Wayne RK, SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19(4):208–216

Morin PA, Martien KK, Archer FI, Cipriano F, Steel D, Jackson JA, Taylor BL (2010) Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. J Hered 101(1):1–10

Murphy MA, Waits LP, Kendall KC (2000) Quantitative evaluation of fecal drying methods for brown bear DNA analysis. Wildl Soc Bull 28(4):951–957

Ohsako S, Nagano R, Sugimoto Y, Goto K (1997) Comparison of the nuclear DNA stability against freezing-thawing and high temperature treatments between spermatozoa and somatic cells. J Vet Med Sci 59(11):1085–1088

Pääbo S, Poinar H, Serre D et al (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38:645–679

Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. Mol Ecol 12(6):1375–1387

Palm S, Laikre L, Jorde PE, Ryman N (2003) Effective population size and temporal genetic change in stream resident brown trout (*Salmo trutta*, L.). Conserv Genet 4:249–264

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12(4):844–855

Pichler FB, Baker CS (2000) Loss of genetic diversity in the endemic Hector's dolphin due to fisheries-related mortality. Proc R Soc Lond Ser B Biol Sci 267:97–102

Piggott MP, Banks SC, Stone N, Banffy C, Taylor AC (2006) Estimating population size of endangered brush-tailed rock-wallaby (*Petrogale penicillata*) colonies using faecal DNA. Mol Ecol 15:81–91

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 WC (eds) Techniques in molecular evolution and systematics. Birkhaeuser Verlag AG, Basel, pp 176–248

Proctor MF, McLellan BN, Strobeck C, Barclay RMR (2005) Genetic analysis reveals demographic fragmentation of grizzly bears yielding vulnerably small populations. Proc R Soc Lond Ser B Biol Sci 272:2409–2416

Prugh LR, Ritland CE, Arthur SM, Krebs CJ (2005) Monitoring coyote population dynamics by genotyping faeces. Mol Ecol 14(5):1585–1596

Roon DA, Waits LP, Kendall K (2005) A simulation test of the effectiveness of several methods for error-checking non-invasive genetic data. Anim Conserv 8:203–215

Ryman N, Palm S, André C, Carvalho G, Dahlgren T, Jorde PE, Laikre L, Larsson L, Palmé A, Ruzzante D (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. Mol Ecol 15(8):255–261

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22(1):25–33

Secretariat of the Convention on Biological Diversity (2011) Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological Diversity: text and annex. Montreal, Canada

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615–629

Smith LM, Burgoyne LA (2004) Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper. BMC Ecol 4(4):1–11

Smith S, Morin PA (2005) Optimal storage conditions for highly dilute DNA samples: a role for Trehalose as a preserving agent. J Forensic Sci 50(5):1101–1108

Stephenson JJ, Campbell MR, Hess JE et al (2009) A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conserv Genet 10:1145–1149

Stetz JB, Kendall KC, Vojta CD et al Genetic monitoring for managers: a new online resource. J Fish Wildl Manag (in press)

Sunnucks P (2000) Efficient genetic markers for population biology. Trends Ecol Evol 15:199–203

Taberlet P, Waits LP, Luikart G (1999) Non-invasive genetic sampling: look before you leap. Trends Ecol Evol 14:323–327

Tallmon D, Gregovich D, Waples RS et al (2010) When are genetic methods useful for estimating contemporary abundance and detecting population trends? Mol Ecol Res 10(4):684–692

Teo YY, Inouye M, Small KS et al (2008) Whole genome-amplified DNA: insights and imputation. Nat Methods 5(4):279–280

Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. Mol Ecol 6:1091–1097

Whitlock MC, McPeek MA, Rausher MD, Rieseberg L, Moore AJ (2010) Data archiving. Am Nat 175(2):145–146

Wirgin I, Maceda L, Stabile J, Mesing C (1997) An evaluation of introgression of Atlantic coast striped bass mitochondrial DNA in a Gulf of Mexico population using formalin-preserved museum collections. Mol Ecol 6:907–916

Wright LI, Tregenza T, Hosken DJ (2008) Inbreeding, inbreeding depression and extinction. Conserv Genet 9:833–843