Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

Museums are biobanks: unlocking the genetic potential of the three billion specimens in the world's biological collections

David K Yeates[1], Andreas Zwick[1] and Alexander S Mikheyev[2]

Museums and herbaria represent vast repositories of biological material. Until recently, working with these collections has been difficult, due to the poor condition of historical DNA. However, recent advances in next-generation sequencing technology, and subsequent development of techniques for preparing and sequencing historical DNA, have recently made working with collection specimens an attractive option. Here we describe the unique technical challenges of working with collection specimens, and innovative molecular methods developed to tackle them. We also highlight possible applications of collection specimens, for taxonomy, ecology and evolution. The application of next-generation sequencing methods to museum and herbaria collections is still in its infancy. However, by giving researchers access to billions of specimens across time and space, it holds considerable promise for generating future discoveries across many fields.

Addresses
[1] Australian National Insect Collection, CSIRO National Research Collections Australia, PO Box 1700, Canberra, ACT 2601, Australia
[2] Ecology and Evolution Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Kunigami-gun 904-0412, Japan

Corresponding author: Yeates, David K (david.yeates@csiro.au)

Current Opinion in Insect Science 2016, **18**:83–88

This review comes from a themed issue on **Insect phylogenetics**

Edited by **Gregory W Courtney** and **Brian M Wiegmann**

For a complete overview see the Issue and the Editorial

Available online 19th October 2016

http://dx.doi.org/10.1016/j.cois.2016.09.009

Introduction

The molecular revolution has profoundly impacted the biological sciences. In the past decades biobanks and biorepositories have been developed to store tissue or DNA samples appropriate for genetic and genomic research [1], with storage in liquid nitrogen being the 'gold standard'. These biobanks were initiated to store human and model organism accessions, but increasingly their use has been broadened to include a wider range of taxa, and efforts are underway to join them in a virtual network, the Global Genome Biodiversity Network, GGBN [2]. At the time of writing, the GGBN had 49 members around the world storing just over 250,000 tissue samples of only 32,000 species (www.ggbn.org). While this is impressive, these numbers pale in comparison to the estimated 3 billion specimens from 2 million species stored by the world's museums and herbaria [3]. This includes samples of all the nearly 2 million described species and all their synonyms, as well as samples of the about 20,000 species newly described each year [4,5]. Assuming an average genome size of 0.5 Gb, these specimens contain zettabases (10^{21}) of sequence data, on the scale of total available hard drive storage capacity [6]. These museum and herbarium (together termed collections below) samples also span an incredible geographical range across all biomes in all continents and a temporal range extending back prior to the industrial revolution. In addition to their phenotype, what if we could sample the genotype of these specimens?

Increasingly we can do this, with careful extra work. Standard museum and herbarium specimens are often stored at or near room temperature, either air dried or in preservative liquids such as ethanol and formalin. Earlier experimental protocols employing Sanger sequencing technology established that genetic sequences from museum and herbarium specimens were possible to obtain, but success was patchy and limited to genes found in high copy number, such as those from cellular organelles [7]. Similar results were obtained from ancient DNA where genetic sequences were recovered from environmental samples [8]. These early successes highlighted the enormous potential of old DNA samples to answer compelling biological questions, especially about the evolutionary history of extinct species and the nature and trajectory of biological change through time [9–12].

In recent years high throughput sequencing (HTS) technology has greatly expanded and synergised the genetic and genomic potential of biological collections. This is largely because degraded DNA in collection samples is a much more tractable starting point for HTS than previous sequencing technologies, producing greater data yields and the assembly of sequences from a greater variety of genes [13,14••]. The power of new sequencing technology to unlock the genetic and genomic potential of museum and herbarium specimens is so great that it has blurred the distinction between biobank and collection, especially when DNA sequence is the target data. Obviously, there are cases where high-quality biobank tissues are

essential, such as for studies of RNA and complete genome assemblies. Conversely, molecular studies linked to biodiversity benefit more from the taxonomic breadth of biological collections than from a limited number of high quality samples in biobanks. Putting special applications aside, this review focuses on the landscape of opportunities natural history collections offer if looked upon as vast storehouses of genomic DNA, and on important sequencing strategies that open up this rapidly developing field.

Challenges posed by museum specimens

Museum and herbarium specimens pose a number of unique challenges, which require the development of novel molecular and analytic approaches to dealing with them. Below we outline some considerations and proposed solutions.

Damage to the specimen

By definition, museum specimens are irreplaceable, but DNA extraction often results in damage to the specimens. Fortunately, a number of approaches suitable for invertebrates and vertebrates minimize damage to the specimen, while producing adequate DNA yields [15•,16–18].

Fragmentation

The DNA of museum and herbarium specimens will almost certainly be fragmented by a number of processes that begin after death such as DNA hydrolysis through nucleases in the body itself, chemicals used as killing and/or fixing agents, preservatives such as ethanol and formalin [19], and chemicals used to protect against pest attack in the collection such as dichlorvos [20]. Fragmentation is generally not an obstacle for HTS methods because they require short lengths of DNA template. However, extremely short fragments may not carry enough information to be useful, and may need to be filtered out either bioinformatically, or ideally during library construction.

Contamination

Many museum samples contain not just endogenous DNA, but also DNA from bacterial, fungal and other contaminants that have grown in the sample post-mortem. In addition, there is possible contamination from other material that was stored together in the same tray, or vial, or that was brought into contact during specimen handling. Contamination is a major problem, because amplification by polymerase chain reaction (PCR) can bias towards longer, more intact fragments resulting in the overrepresentation of non-endogenous DNA. At best, this consumes sequencing capacity, which can increase the expense of sequencing by over an order of magnitude. At worst, contamination can yield erroneous data. As a result, when working with museum specimens, it is worth following best practices developed for the study of human specimens, such as decontamination with UV-light and physical separation between areas used for DNA extraction and amplification [21].

DNA degradation

DNA breaks down over time, which causes a range of miscoding lesions and can lead to erroneous sequence reads. Depurination, especially in guanosine residues, leads to strand breaks, and deamination of cytosine residues into uracil also occurs [22,23]. Both depurination and deamination can lead to GC→AT sequencing errors. In addition, interstrand cross-linking may occur post-mortem, particularly in formalin-fixed specimens, preventing polymerase bypass, and blocking DNA denaturation [24].

Sequencing strategies

The sequencing strategies are outlined in Figure 1 focus on preparing sequencing libraries from the DNA extract, and, optionally, enriching them for endogenous DNA and evenly targeted sequences.

Direct sequencing of museum samples

While it is certainly possible to sequence museum samples directly, this may not be cost-effective for large numbers of samples, or for organisms with large genomes. In addition, with the exception of the strategy outlined in Figure 1b′, direct sequencing does not eliminate contaminant DNA, which may substantially waste sequencing capacity. However, since unbiased genome representation is often lost during enrichment, direct sequencing may be the best approach for low-input samples. PCR-free libraries are a solution for direct sequencing, which greatly minimize the risk of contamination [25••]. Alternatively, extremely low-coverage whole genome shotgun sequencing ('genome skimming') permits the sequencing of highly abundant DNA (e.g. ribosomal genes, mitochondrial and chloroplast genomes) for a large number of samples, sufficient for many phylogenetic questions [15•,26].

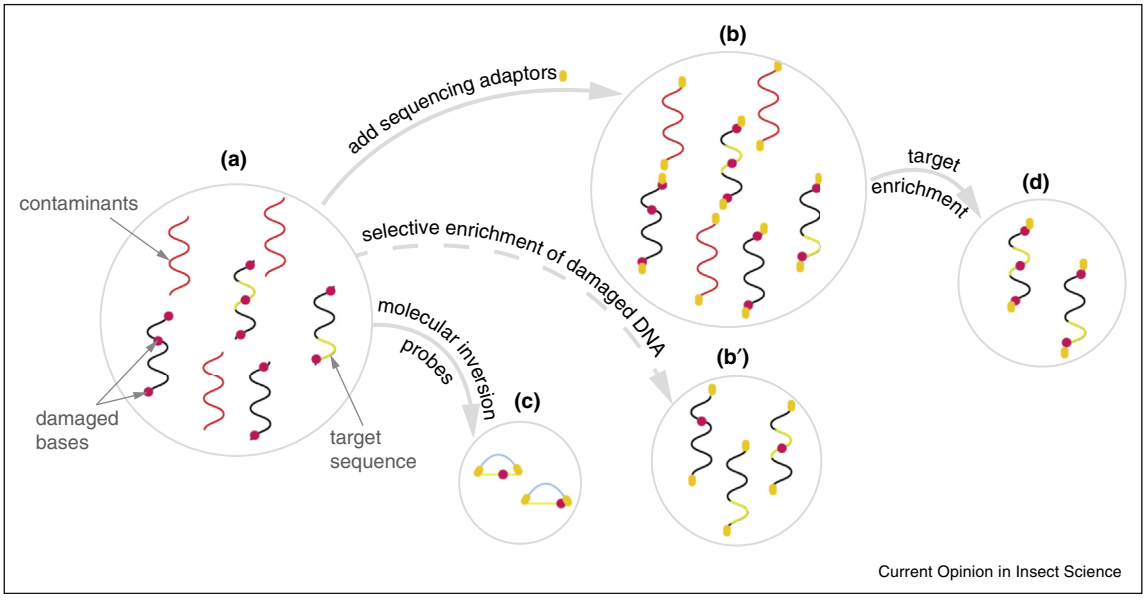
Targeted reduced genome representation

These methods involve the selective capture of genomic regions prior to NGS [27]. Sequence capture methods are technically demanding, require construction of libraries prior to hybridization, and do not scale well [28•]. However, they are cost-effective when dealing with large numbers of samples.

Hybrid enrichment

This method involves hybridizing genomic DNA to DNA probes or 'baits' and then washing away the non-target DNA (e.g. [29–31]). The resulting enriched DNA can be sequenced using various HTS platforms. Bait design requires some knowledge of the target genome, so may require a transcriptome or genome sequence within or adjacent to the target group. The standard probe designs usually work for closely related species with 10–15% divergence, but baits can be designed around targets with

Figure 1



Processing of museum specimen DNA generally requires multiple steps. **(a)** The initial pool of DNA contains fragments from the target organism (black), contaminants (red), and sites where the template has undergone sequence-altering damage (red circles). Target sequences of particular interest for the investigator are highlighted in yellow. The goal is to maximize the yield of endogenous DNA in the sequencing libraries, and possibly select a subset of the genome for targeted sequencing. **(b)** A common first step is to generate a sequencing library by attaching synthetic adaptors to both ends of the DNA fragment. **(b′)** An ingenious protocol by Gansauge and Meyer [50••] takes advantage of high rates of cytosine deamination in single-stranded overhangs in ancient DNA to enrich for endogenous material. Libraries prepared in steps (b) or (b′) can be used for whole-genome sequencing, though without enrichment for endogenous DNA, it is likely that many reads, potentially the overwhelming majority, will be wasted on contaminant sequences. Alternatively, molecular inversion probes can capture target sequence, while simultaneously making constructs that can be used directly for sequencing **(c)**. For many applications, whole genome sequencing is not cost-effective, and a variety of enrichment approaches exist, which capture just a subset of the target genome **(d)**.

different phylogenetic scope (e.g. family versus order). Because they work on fragmented DNA, these standard hybrid probes will work with museum specimens, but the success of hybrid capture is greatest with more recently collected material [32].

Alternatively, ultraconserved elements (UCE) and other very slowly evolving regions could be used to anchor DNA fragments to hybrid probes (e.g. [29,33]). These techniques rely on the identification of short 50–200 bps genomic regions that are highly conserved and act as anchors for genetic markers. Variable sequence is recovered flanking the anchors that can be used in phylogenetic analysis. An additional benefit is that at least UCEs appear to have little overlap with known paralogous genes. In the context of museum specimens, the more degraded the DNA, the less flanking sequence will be captured; as variability tends to increase with distance to conserved areas, old museum specimens might still be useful to capture variation between distantly related species, but not for closely related species. Depending on the application, using anchors for degraded DNA may be problematic, since not enough variably flanking sequence may exist, though workable data may still be obtained [34•,35].

Molecular inversion probes

Molecular inversion probes (MIPs) are single-stranded DNA molecules containing on their ends sequences complementary to two conserved regions flanking the target of up to several hundred base pairs. Following hybridization to the target, gap-filling and ligation result in circularized DNA molecules containing the target sequence with adaptors and barcodes ready for downstream analyses [36]. Long used for model systems, they have a number of advantages, such as low input template amount — in species with genomes 1 Gb or smaller, 20–50 ng of genomic DNA should be sufficient [28•]. MIPs do not require the development of genomic libraries, are relatively inexpensive and software for the design and analysis of MIP markers is available. Markers can be designed to have a broad phylogenetic target. Because the technique targets short DNA fragments, the technique should be amenable for degraded museum material, though the synthesis of the probes themselves is expensive.

Random reduced genome representation

Restriction site associated DNA (RAD-seq) analysis [37,38] involves sequencing regions of the genome associated with conserved cutting sites of restriction enzymes.

This technique is low-cost and allows many samples to be processed simultaneously. It was primarily developed for population genetics, but RAD-seq data are also used in low-level phylogenetics, with the completeness and usefulness of the data matrix decreasing with evolutionary distance. In principle, one can apply RAD-seq to museum samples without modification, but given the already fragmented nature of the DNA, many of the resulting fragments are too short to provide meaningful information [15•]. In addition, many libraries have low complexity and consist of many PCR duplicates [39]. As a result, a number of techniques have been developed, combining the benefits of hybridization with RAD-seq. One such technique is RAPTURE (RAD Capture; [40]), which involves pooling of RAD-seq data, followed by joint capture to increase specificity and reduce capacity needed for sequencing, that is, permitting the sequencing of more samples in a run. An alternative called hyRAD (hybridization RAD; [41••]) uses RAD-seq fragments of fresh samples as cheap probes to capture degraded fragments from museum specimens. These techniques show considerable promise for non-model systems, since they do not require extensive genomic information prior to analysis.

What can we achieve by sequencing museum specimens?

Perhaps the most obvious application of museum and herbarium specimens is in keeping with the traditional role of collections, namely for taxonomic research. Molecular markers have already been used to link type specimens to specific members of morphologically indistinguishable species groups [42]. Likewise, molecular analysis of collection specimens can be used for species delimitation and phylogenetics [15•,43]. In addition, museums and herbaria contain many unidentified specimens, many of which represent new species [44]; molecular analysis can speed up the species discovery process.

Just as powerfully, collection specimens can be used to study ecological and evolutionary processes. For example, museum specimens have been used to reconstruct the history of biological invasions [45–47]. They have also been useful in examining changes in effective population size [7], and connectivity between populations [48]. Finally, by providing a window into the past, museum specimens can permit comparisons between time points, permitting direct measurement of allelic changes across time, and thus direct measures of selection [25••,49].

Driven by advances in next-generation sequencing technology, the study of museum and herbarium specimens is in its infancy. With time, we are sure that the few applications outlined above will prove to be just a small sampling of what is possible. There remain many avenues for future research, both in the perfection and development of molecular tools, and in the analysis of the downstream data. Hopefully, the growing appreciation of the genetic value of collection specimens and the discoveries made through their collections will not only spur on research, but also result in the allocation of resources to the collections themselves, which provide ongoing stewardship and management of a wealth of samples that cannot be recreated.

Acknowledgements

David Yeates receives funding from the Australian National Insect Collection Schlinger Trust, and he and Andreas Zwick are also supported by CSIRO National Research Collections Australia.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Seberg O, Droege G, Barker K, Coddington JA, Funk V, Gostel M, Petersen G, Smith PP: **Global Genome Biodiversity Network: saving a blueprint of the tree of life – a botanical perspective**. *Ann Bot* 2016 http://dx.doi.org/10.1093/aob/mcw121.
2. Droege G, Barker K, Astrin JJ, Bartels P, Butler C, Cantrill D, Coddington J, Forest F, Gemeinholzer B, Hobern D *et al.*: **The Global Genome Biodiversity Network (GGBN) data portal**. *Nucleic Acids Res* 2014, **42**:D607-D612.
3. Wheeler QD, Knapp S, Stevenson DW, Stevenson J, Blum SD, Boom BM, Borisy GG, Buizer JL, Carvalho MRD, Cibrian A *et al.*: **Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity**. *Syst Biodivers* 2012, **10**:1-20.
4. Chapman AD: *Numbers of Living Species in Australia and the World*. Department of the Environment, Water, Heritage and the Arts; 2009.
5. IISE: *State of Observed Species*. International Institute for Species Exploration; 2011:. http://www.esf.edu/species/documents/sos2011.pdf [retrieved August 2016].
6. Hilbert M, López P: **The world's technological capacity to store, communicate, and compute information**. *Science* 2011, **332**:60-65.
7. Wandeler P, Hoeck PEA, Keller LF: **Back to the future: museum specimens in population genetics**. *Trends Ecol Evol* 2007, **22**:634-642.
8. Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M: **Genetic analyses from ancient DNA**. *Annu Rev Genet* 2004, **38**:645-679.
9. Buerki S, Baker WJ: **Collections-based research in the genomic era**. *Biol J Linn Soc Lond* 2015, **117**:5-17.
10. Holmes MW, Hammond TT, Wogan GOU, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM *et al.*: **Natural history collections as windows on evolutionary processes**. *Mol Ecol* 2016, **25**:864-881.
11. Linderholm A: **Ancient DNA: the next generation – chapter and verse**. *Biol J Linn Soc Lond* 2016, **117**:150-160.
12. Morozova I, Flegontov P, Mikheyev AS, Bruskin S, Asgharian H, Ponomarenko P, Klyuchnikov V, ArunKumar G, Prokhortchouk E, Gankin Y *et al.*: **Toward high-resolution population genomics using archaeological samples**. *DNA Res* 2016, **23**:295-310.
13. Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT: **Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens**. *PLoS ONE* 2013, **8**:e69189.

14. •• Burrell AS, Disotell TR, Bergey CM: **The use of museum specimens with high-throughput DNA sequencers**. *J Hum Evol* 2015, **79**:35-44.
This extensive review on the topic outlines recent literature, reviewing opportunities and drawback to using museum specimens.

15. • Tin MM-Y, Economo EP, Mikheyev AS: **Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics**. *PLoS ONE* 2014, **9**:e96793.
This study presents a minimally-destructive extraction protocol that works with a range of insects, and shows that genome skimming at low coverage (<1×) can be used to reconstruct accurate phylogenies.

16. Gilbert MTP, Moore W, Melchior L, Worobey M: **DNA extraction from dry museum beetles without conferring external morphological damage**. *PLoS ONE* 2007, **2**:e272.

17. Alvarado Bremer JR, Smith BL, Moulton DL: **Shake and stew: a non-destructive PCR-ready DNA isolation method from a single preserved fish larva**. *J Fishscicom* 2014, **84**:267-272.

18. Hofreiter M: **Nondestructive DNA extraction from museum specimens**. In *Ancient DNA: Methods and Protocol.* Edited by Shapiro B, Hofreiter M. Springer Science + Business Media, LLC; 2012:93-100.

19. Hykin SM, Bi K, McGuire JA: **Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing**. *PLOS ONE* 2015, **10**:e0141579.

20. Espeland M, Irestedt M, Johanson KA, Akerlund M, Bergh J-E, Källersjö M: **Dichlorvos exposure impedes extraction and amplification of DNA from insects in museum collections**. *Front Zool* 2010, **7**:2.

21. Poinar HN, Cooper A: **Ancient DNA: do it right or not at all**. *Science* 2000, **289**:1139.

22. Zimmermann J, Hajibabaei M, Blackburn DC, Hanken J, Cantin E, Posfai J, Evans TC Jr: **DNA damage in preserved specimens and tissue samples: a molecular assessment**. *Front Zool* 2008, **5**:18.

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

24. Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, Meyer M: **Road blocks on paleogenomes – polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA**. *Nucleic Acids Res* 2010, **38**:e161.

25. •• Mikheyev AS, Tin MMY, Arora J, Seeley TD: **Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite**. *Nat Commun* 2015, **6**:7991.
This study introduces a PCR-free library preparation protocol for whole genome shotgun sequencing, and uses it to measure genetic changes between two populations, comparing museum samples with modern-day samples. This study bring us important issues about the downstream analysis of data from museum specimens, particularly the effect of DNA degradation on estimates of population genetic parameters.

26. Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE *et al.*: **Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline**. *Biol J Linn Soc Lond* 2016, **117**:33-43.

27. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ: **Target-enrichment strategies for next-generation sequencing**. *Nat Methods* 2010, **7**:111-118.

28. • Niedzicka M, Fijarczyk A, Dudek K, Stuglik M, Babik W: **Molecular inversion probes for targeted resequencing in non-model organisms**. *Sci Rep* 2016, **6**:24051.
This study introduces molecular inversion probes (MIPS) for medium-scale (hundreds or thousands of targets) targeted resequencing in non-model organisms. The use of these probes looks like a promising avenue in museum specimen research.

29. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC: **Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales**. *Syst Biol* 2012, **61**:717-726.

30. Lemmon AR, Emme SA, Lemmon EM: **Anchored hybrid enrichment for massively high-throughput phylogenomics**. *Syst Biol* 2012, **61**:727-744.

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

32. Young AD, Lemmon AR, Skevington JH, Mengual X, Ståhls G, Reemer M, Jordaens K, Kelso S, Lemmon EM, Hauser M *et al.*: **Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae)**. *BMC Evol Biol* 2016, **16**:143.

33. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT: **Applications of next-generation sequencing to phylogeography and phylogenetics**. *Mol Phylogenet Evol* 2013, **66**:526-538.

34. • Blaimer BB, Lloyd MW, Guillory WX, Brady SG: **Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens**. *PLoS ONE* 2016, **11**:e0161531.
Using specimens ranging in age from 2 to 121 years this study carried out UCE-based phylogenetic analysis. This study shows that UCE contig lengths are dramatically shorter for specimens more than 20 years old, but some phylogenetic data is nonetheless recoverable.

35. McCormack JE, Tsai WLE, Faircloth BC: **Sequence capture of ultraconserved elements from bird museum specimens**. *Mol Ecol Resour* 2016, **16**:1189-1203.

36. Hardenbol P, Banér J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U *et al.*: **Multiplexed genotyping with sequence-tagged molecular inversion probes**. *Nat Biotechnol* 2003, **21**:673-678.

37. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA: **Rapid SNP discovery and genetic mapping using sequenced RAD markers**. *PLoS ONE* 2008, **3**:e3376.

38. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE: **Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species**. *PLoS ONE* 2012, **7**:e37135.

39. Tin MMY, Rheindt FE, Cros E, Mikheyev AS: **Degenerate adaptor sequences for detecting PCR duplicates in reduced representation sequencing data improve genotype calling accuracy**. *Mol Ecol Resour* 2015, **15**:329-336.

40. Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR: **RAD capture (RAPTURE): flexible and efficient sequence-based genotyping**. *Genetics* 2016, **202**:389-400.

41. •• Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, Pajkovic M, Ronikier M, Alvarez N: **Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens**. *PLoS ONE* 2016, **11**:e0151651.
The hyRAD method is described, which is specifically developed for museum specimens, and combines the benefits of RAD-seq for selecting a subset of loci without a genomic reference with hybridization selectively enrich them for museum material. This method should be extremely powerful for population genetics, and for phylogenetics of relatively recent radiations.

42. Price BW, Henry CS, Hall AC, Mochizuki A, Duelli P, Brooks SJ: **Singing from the Grave: DNA from a 180 year old type specimen confirms the identity of Chrysoperlacarnea (Stephens)**. *PLoS ONE* 2015, **10**:e0121127.

43. Maddison DR, Cooper KW: **Species delimitation in the ground beetle subgenus Liocosmius (Coleoptera: Carabidae: Bembidion), including standard and next-generation sequencing of museum specimens**. *Zool J Linn Soc* 2014, **172**:741-770.

44. Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB *et al.*: **Herbaria are a major frontier for species discovery**. *Proc Natl Acad Sci U S A* 2010, **107**:22169-22171.

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

46. Russello MA, Avery ML, Wright TF: **Genetic evidence links invasive monk parakeet populations in the United States to the international pet trade**. *BMC Evol Biol* 2008, **8**:217.

47. Marsico TD, Burt JW, Espeland EK, Gilchrist GW, Jamieson MA, Lindström L, Roderick GK, Swope S, Szűcs M, Tsutsui ND: **Underutilized resources for studying the evolution of invasive species during their introduction, establishment, and lag phases**. *Evol Appl* 2010, **3**:203-219.

48. Rollins LA, Woolnough AP, Wilton AN, Sinclair R, Sherwin WB: **Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia**. *Mol Ecol* 2009, **18**:1560-1573.

49. Hartley CJ, Newcomb RD, Russell RJ, Yong CG, Stevens JR, Yeates DK, La Salle J, Oakeshott JG: **Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance**. *Proc Natl Acad Sci U S A* 2006, **103**:8757-8762.

50. •• Gansauge M-T, Meyer M: **Selective enrichment of damaged DNA molecules for ancient genome sequencing**. *Genome Res* 2014, **24**:1543-1549.

This protocol takes advantage of the fact that ancient and historic DNA has high levels of cytosine deamination at the exposed ends, to specifically enrich for them. Although moderately labor-intensive, this protocol provides a 10× enrichment of the target samples.