https://doi.org/10.1038/s41576-024-00801-2

Exploring biodiversity through museomics

Charles C. Davis & Sandra Knapp

Natural history collections are unparalleled resources for studying biodiversity across multiple dimensions. As multi-omic tools are being applied to museum specimens, the emerging field of museomics is undergoing a rapid and exciting transformation, with new opportunities, challenges and confrontations with past legacies.

Natural history collections (NHCs) collectively preserve samples and document records of populations of plants, animals, fungi and microorganisms that span the history and diversity of life on Earth. Recent advances in ancient DNA (aDNA) techniques, including experimental and computational methods for sequence analysis of fragmentary and highly degraded samples, are enabling the genomic analysis of preserved museum specimens. Although the fragmented nature of DNA from such specimens currently limits de novo assembly of complete genomes, new sequencing methods can generate genome-scale data and detect fine-scale genetic bases of adaptation from genome-wide scans. Parallel advances in -omic methods have opened new research directions across comparative biology, ushering in the era of museomics – the application of multi-omic tools to NHCs. Already, these collections have provided specimens and data for basic and applied research, including to characterize long-extinct species and populations, trace the effect of humans on genomic and phenotypic changes, and investigate the effects of ongoing climate change. These recent developments are providing opportunities and challenges for genomic research using NHCs.

Enhancing visibility of the global metamuseum

The online mobilization of digitized collections has transformed the use of NHCs. The emergent global metamuseum, like the 'global metaherbarium'[1], is a common, digitally interlinked and open-access resource that will stimulate large-scale and novel science. Readily accessible today in online portals such as the Global Biodiversity Information Facility (GBIF) and Integrated Digitized Biocollections (iDigBio), this platform is ideal for viewing, searching and connecting physically scattered biodiversity data, and provides much more content than is available from physical specimens alone. An 'extended digital specimen' in the metamuseum minimally includes an image of the physical specimen and a transcribed label, but may also include its genomic sequence, field images, inferred species distribution models, literature citations and other data that can be linked to the original organism. The visibility and accessibility of extended specimens stimulates increased visitation to NHCs by diverse scholars, including geneticists and genomicists, who use physical collections in their research.

Developing a comprehensive tree of life

Herbarium specimens, many of which are more than 200 years old, have been critical to broaden the taxonomic scope of phylogenomic investigations – for example, characterizing the explosive early diversification of the first flowering plants[2]. Genomic sequences derived from NHCs are also essential for establishing accurate DNA barcoding libraries and for cutting-edge environmental DNA (eDNA) assessments. These assessments have been important for a range of applications, from authenticating plant-derived medicines, inventorying fish in the megadiverse Amazon River basin, to studying insect herbivores on plants. Habitat- or locality-specific short-read and long-read sequences of specimens in NHCs can facilitate the recognition of new taxa in eDNA studies, speeding up analyses and connecting new results with historical surveys.

Recording human-mediated changes

NHCs serve to establish historical biodiversity baselines in the same way that climate records serve as vital references for global climate models. Time-stamped specimens can be used to determine how species, including crops and invasive species, move, adapt and diversify. Genomic analysis of historical collections of *Amaranthus tuberculatus* in North America demonstrated that herbicides used since the 1960s exerted exceptionally strong selective pressure that contributes to notable shifts in allele frequencies, driving herbicide resistance and increasing the success of this weed in contemporary agricultural environments[3].

Collections also capture early indicators of genetic change in response to climate change. For example, a time series including specimens from NHCs revealed genomic variations in the migratory yellow warbler (*Setophaga petechia*) that were linked to climate across its breeding range[4]; populations experiencing notable declines lacked the allele frequency shifts necessary to adapt to future climatic conditions. Moreover, genomic data from herbarium specimens of *Arabidopsis thaliana* spanning two centuries revealed that key transcription factors related to stomatal development are evolutionarily stable, but key regulatory genes show local adaptation[5]. As stomatal density in plant leaves decreases with increased carbon dioxide levels, this finding is suggestive of adaptive evolutionary changes to environmental conditions.

Understanding extinction

Research with NHCs has challenged the idea that population instability is always a major cause of rapid extinction, demonstrating that natural selection can affect genomes unpredictably and idiosyncratically. Genomic analysis of museum specimens from extinct and living Australian rodents (Hydromyini) showed no evidence of reduced genetic diversity prior to extinction; instead, populations declined precipitously after European colonization[6]. By contrast, despite its large population size, the extinct passenger pigeon (*Ectopistes migratorius*) – once the most abundant bird in North America – had surprisingly low genetic diversity that facilitated rapid adaptive evolution and purging of harmful mutations, resulting in significant loss of neutral genetic diversity[7]. Nonetheless, in other cases, more expected and predictable changes in genetic diversity have accompanied range size reduction and extinction. Estimates of genomic variation from samples of more than 10,000 individuals from 20 plant and animal species, many from NHCs, suggest that over 10% of genetic diversity might already be lost for many species, exceeding the United Nations' post-2020 targets for genetic preservation[8].

Exploring new ecological communities

NHCs contain an almost completely unexplored set of ecological communities of bacteria, fungi, viruses and other microorganisms that have an outsized effect on our planet, and thus act as windows into their interactions. Researchers working with NHCs are applying highly sensitive metagenomic approaches to characterize genomes of various taxa and their associated organisms, including microbiomes, parasites and pathogens.

Genomic data from NHCs can profoundly change our understanding of epidemiology. For example, detection of the Lyme disease-causing bacterium spirochaete in specimens of white-footed mice collected long before the first cases were diagnosed in the USA provided evidence against an introduction of this pathogen from Europe[9]. New sequencing techniques, coupled with the emergence of host and pathogen genome data, enable even more powerful uses of NHCs, including tracking the evolution of the malaria pathogen *Plasmodium* in mosquito specimens, investigating coronaviruses in bat specimens, or monitoring emerging plant diseases.

Ensuring reproducibility with vouchers

NHCs are priceless repositories that preserve specimens permanently, making them available for loan and study. NHCs adhere to standardized archival protocols and retain taxonomic information, permits and other data as part of the extended specimen. Voucher specimens held permanently in NHCs are linked to materials sequenced for genomic investigation to ensure that published results are traceable, verifiable, repeatable and reputable. Unfortunately, references to vouchers are often omitted in genomic publications and associated genomic repositories such as GenBank[10]. Without such permanently accessible specimens, reproducibility is greatly compromised, and errors are likely to propagate in the literature, especially as taxonomic revisions occur. Reference to catalogued and curated specimens with proper documentation not only supports accurate taxonomy but also provides the best evidence that specimens and their derivatives were collected legally.

Giving credit and attribution

The absence of detailed sampling information reduces inclusivity by ignoring the contribution to the work of all involved, including curators, collectors, local communities, and field scientists[11]. Museums and the people who work in them provide crucial material and context for analyses, so it is surprising that genome sequencing papers typically include many hundreds of authors who have provided sequences or material, but genomic studies using NHCs rarely do so. Regular use of the Contributor Role Taxonomy (CRediT) would better and more inclusively acknowledge contributions and provide guidance for co-authorship for material providers such as NHCs in genetic and genomic studies[11]. However, good practice is becoming more common as genomic studies advance[12].

Grappling with the colonial past

The legacy of NHCs in the global north and the colonial periods in which they were amassed are receiving new attention and calls for accountability[13]. The Nagoya Protocol, nested within the Convention on Biological Diversity, aims to address the blurred boundaries between historical colonialism, continued racism and the benefits of collecting scientific specimens. Its mission is to ensure that the bounties reaped from genetic resources, including from NHCs, are shared fairly and equitably, thereby promoting both conservation and sustainable use of Earth's incredible biodiversity. Nagoya continues to evolve; considerable revisions weave Indigenous and local perspectives and wisdom into partnerships based on equity and respect, amplifying global movements to decolonize scientific and academic practices.

> "[Natural history collections] have an increasingly important role in bringing together genomic studies with biodiversity science"

NHCs have an increasingly important role in bringing together genomic studies with biodiversity science. Increasing and amplifying conversations and collaborations between genomicists and biodiversity scientists will benefit both enormously and help us answer the biggest questions about the origin, evolution and maintenance of the diversity of life on Earth.

Charles C. Davis [1] & **Sandra Knapp** [2]

[1]Department of Organismic and Evolutionary Biology, Harvard University Herbaria, Cambridge, MA, USA. [2]Natural History Museum, London, UK.

e-mail: cdavis@oeb.harvard.edu; s.knapp@nhm.ac.uk

References

1. Davis, C. C. The herbarium of the future. *Trends Ecol. Evol.* **38**, 412–423 (2022).
2. Zuntini, A. R. et al. Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).
3. Kreiner, J. M. et al. Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science* **378**, 1079–1085 (2022).
4. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83–86 (2018).
5. Lang, P. L. M. et al. Century-long timelines of herbarium genomes predict plant stomatal response to climate change. *Nat. Ecol. Evol.* **8**, 1641–1653 (2024).
6. Roycroft, E. et al. Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement. *Proc. Natl Acad. Sci. USA* **118**, e2021390118 (2021).
7. Murray, G. G. R. et al. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* **358**, 951–954 (2017).
8. Exposito-Alonso, M. et al. Genetic diversity loss in the Anthropocene. *Science* **377**, 1431–1435 (2022).
9. Marshall, W. F. et al. Detection of *Borrelia burgdorferi* DNA in museum specimens of *Peromyscus leucopus*. *J. Infect. Dis.* **170**, 1027–1032 (1994).
10. Buckner, J. C. et al. The critical importance of vouchers in genomics. *eLife* **10**, e68264 (2021).
11. Davis, C. C. et al. Guidelines for the effective and ethical sampling of herbaria. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-024-02544-z (2024).
12. Gardner, E. M. et al. Engagement with indigenous people preserves local knowledge and biodiversity alike. *Curr. Biol.* **32**, R511–R512 (2022).
13. Park, D. S. et al. The colonial legacy of herbaria. *Nat. Hum. Behav.* **7**, 1059–1068 (2023).

Acknowledgements

The authors thank the curators of NHCs for opening their collections for new and exciting uses, and the institutions and governments that support them.

Competing interests

The authors declare no competing interests.

Related links

Contributor Role Taxonomy: https://credit.niso.org/
GBIF: https://www.gbif.org/
iDigBio: https://www.idigbio.org/
Nagoya Protocol: https://www.cbd.int/abs/text