*Molecular Ecology Resources*

WILEY

MOLECULAR ECOLOGY RESOURCES

OPINION

# Revisiting Genetic Data Stewardship Practices in Aotearoa New Zealand: A Call to Action on Integrating Māori Data Sovereignty

Manpreet K. Dhami[1,2] | Paige Matheson[3] | Starsha Bird[3] | Leilani Walker[4,5] | Holden Hohaia[1] | Angela McGaughran[3]

[1]Manaaki Whenua Landcare Research, Lincoln, New Zealand | [2]School of Biological Sciences, Waipapa Taumata Rau—University of Auckland, Auckland, New Zealand | [3]Te Aka Mātuatua/School of Science, University of Waikato, Hamilton, New Zealand | [4]Auckland University of Technology, Auckland, New Zealand | [5]Natural Sciences, Auckland Museum, Auckland, New Zealand

**Correspondence:** Manpreet K. Dhami (dhamim@landcareresearch.co.nz)

**Received:** 14 October 2024 | **Revised:** 7 May 2025 | **Accepted:** 16 July 2025

**Handling Editor:** Andrew P. Kinziger

**Funding:** This work was funded by New Zealand Biological Heritage National Science Challenge (https://bioheritage.nz/) Strategic Investment Priority grant 'National DNA Reference Library', contract number 2223-44-010 to M.K.D. and H.H.

**Keywords:** FAIR and CARE data principles | metadata practices | reference databases

## ABSTRACT

Genetic data, including environmental DNA (eDNA), are regularly used to monitor escalating biodiversity concerns globally. In Aotearoa New Zealand, biodiversity is unique and cherished—many species are taonga (treasured) and cared for by kaitiaki (guardians with customary responsibilities), specifically mana whenua with custodial rights (Māori; the Indigenous people of New Zealand). Discussions are currently underway regarding the development of a reference DNA barcode database for biodiversity in Aotearoa New Zealand to improve outcomes for biosecurity surveillance and biodiversity assessment. A priority of these discussions is that the database development and eventual implementation accords with Te Tiriti o Waitangi (The Treaty of Waitangi). Here, we evaluate current practices for storing genetic data from samples collected in Aotearoa New Zealand by examining two major public data repositories—the National Centre for Biotechnology Information (NCBI) GenBank and the Barcode of Life Data System (BOLD). We find that current database practices limit opportunities for Māori data sovereignty, with DNA from many taonga species uploaded to public repositories with no associated restrictions or guidelines over use. This is an important finding that will help shape the development of a future DNA reference database for Aotearoa New Zealand that integrates the rights and interests of Indigenous communities.

## 1 | Introduction

Biodiversity is one of Earth's most extraordinary and central features, providing a range of values from environmental to social and economic. The ongoing and projected loss of global biodiversity has significant impacts on ecosystems and the services they provide. At the core of this concern, up to 90% of biodiversity on Earth is undescribed (Moura and Jetz 2021); thus, many newly discovered species are already at risk of extinction (Liu et al. 2022). In Aotearoa New Zealand (AoNZ hereafter), over 80% of vascular plants and insects are endemic (DOC 2016), with a significant proportion of these also threatened or at risk of extinction (Hare et al. 2019). The unique taiao (natural environment) of AoNZ contributes greatly to the identity of its people; thus, there is a collective and increasing interest in conserving endemic biodiversity and managing invasive threats (Kahui and Cullinane 2019). To achieve this, we urgently require baseline data that catalogues existing and newly discovered species.

Genetic data have the ability to address key conservation and biosecurity (i.e., species preservation, and prevention of the introduction and spread of harmful species, respectively) issues—from detecting and monitoring species to providing information about their genetic diversity and adaptive potential (Theissinger et al. 2023). Environmental DNA (eDNA) metabarcoding is one tool that shows great promise for contributing to biodiversity and conservation outcomes, relying on identification of DNA from an environmental sample (e.g., water, soil, air) against reference sequences (Blackman et al. 2022) to identify species that reside in or interact with the sampled environment.

In AoNZ, recent studies have exemplified the use of eDNA to monitor endemic species (e.g., kākahi/freshwater mussel *Echyridella menziesii*; Steiner et al. 2022), detect invasive species (e.g., koi carp *Cyprinus rubrofuscus*; Collins et al. 2022), and monitor ecosystem health (e.g., the 'Taxon-Independent Community Index', Wilkinson et al. 2024; Lakes380 lake health monitoring, https://ourlakesourfuture.co.nz/). Various initiatives, such as Wai Tuwhera o te Taiao Open Waters Aotearoa (https://www.epa.govt.nz/community-involvement/open-waters-aotearoa/), work to involve community groups, hapū (subtribes), and kura (schools) in this mahi (work) by funding a range of community-driven eDNA projects that provide baseline testing and education to inform restoration programs. However, despite the increasing value of genetic data for applied environmental outcomes, the reliance on reference databases to match unknown genetic sequences against is a limiting factor because these databases have a high degree of missing data (Hotaling et al. 2021). For example, one three-year plankton metabarcoding assay could only assign 50% of OTUs to genus or species level (Berry et al. 2023). Lack of reference data routinely hampers efforts to identify and accurately understand trends in biodiversity change, including the detection of new species and of species declines. Given these limitations, wānanga (discussions) were recently held between biodiversity, systematics, and genetics researchers (Māori and non-Māori), research providers, database and infrastructure specialists, data sovereignty experts, Māori scholars and kaumātua (Māori community elders) and the wider public, with a focus towards developing a reference database for biodiversity in AoNZ that will fill DNA data gaps to enhance nationwide biosecurity and conservation outcomes—including those based on eDNA data (Dhami and Hohaia 2023).

In te ao Māori (the Māori worldview), whakapapa refers, in simple terms, to a genealogical framework (Carter 2005; Hikuroa 2017). Moreover, it is the interconnectedness between the 'seen and unseen, humans and more-than humans, and the natural and spiritual world' (Clapcott et al. 2018; Harmsworth and Awatere 2013). Whakapapa is embodied within DNA at the physical and spiritual level (Hudson et al. 2016; Collier-Robinson et al. 2019). Thus, any DNA data obtained from taonga species (typically pre-European contact species present in AoNZ or those that arrived with Māori; see Taiuru 2022 for a detailed evaluation of taonga species definition), are considered taonga in their own right, and are therefore subject to tikanga (Māori customs, protocols) (Collier-Robinson et al. 2019). As such, a critical priority of the national reference database initiative is that the rights and expectations of Māori (the Indigenous people of New Zealand) are incorporated alongside the stewardship best practice that emerges from a Te Tiriti o Waitangi (The Treaty of Waitangi; signed 1840) guided approach, that is, one which would result in a partnership of science-derived tools and Indigenous values to create an enduring resource that contributes broadly to the kaitiakitanga (guardianship) of taiao and its associated taonga.

While recent efforts have been made globally to ascertain opportunities for Indigenous data sovereignty, the vast majority of data systems and scientific repositories are open access, owing to the 'open data movement'—an initiative that promotes availability and accessibility of data to enable greater transparency, collaboration, and innovation (Hudson et al. 2020). AoNZ is uniquely positioned to reassess how a future practice of data generation, storage, and sharing may integrate novel tools to enable Indigenous rights and sovereignty, while delivering the data resources for conservation and biodiversity outcomes. To honour a Te Tiriti-led approach, a national DNA reference library for AoNZ must be guided by Indigenous rights and expectations around data sovereignty and stewardship. As a first step, we here evaluate current practices for storing genetic data from samples that were collected in AoNZ and deposited to either NCBI GenBank or BOLD public data repositories to: (A) determine what the current practice is for storing potentially taonga data; and (B) identify shortcomings and barriers in data stewardship to help fine-tune the best practice for genetic databases moving forward.

## 2 | Methods

### 2.1 | Genetic Sequence Databases

Barcode of Life Database (BOLD) is a free-to-access, cloud-based storage platform for DNA barcode data, that is, short DNA sequences derived from the mitochondrial (mt) cytochrome *c* oxidase I (COI) gene—a common and highly distinguishable genetic marker across a range of taxa often used for species discrimination (Deagle et al. 2014), developed at the Centre for Biodiversity Genomics in Canada (https://boldsystems.org/index.php), and widely considered the gold standard for reference databases. We extracted all DNA barcode records that were classified as collected in AoNZ by searching 'New Zealand' in the public data portal search bar. We removed all records that were classified as 'mined from Genbank', to reduce overlap between databases, yielding a total of 79,932 records (as of April 2025). All records were downloaded directly from BOLD as a tab-separated value (TSV) file (Supporting Information). Metrics of interest (see below) were curated using the python package pandas version 2.0.1 (McKinney 2010). Hereafter, 'BOLD records' refers to the records we extracted from BOLD that were listed as being collected from AoNZ.

The National Centre Biotechnology Information (NCBI) GenBank repository (https://www.ncbi.nlm.nih.gov/genbank/) contains the largest collection of molecular biological and genetic data available. Within the nucleotide database (nt), we extracted records that had been listed as collected in AoNZ and by searching 'country = New Zealand[WORD]' in the search bar. We only considered animals and plants from NCBI due to the exceptionally large number of microbial records. We also removed all records from one study (188,573 records; focused

on the European hedgehog *Erinaceus europaeus*) to prevent extremely skewed results, leaving a total of 289,422 records for analysis (as at April 2025). Records were downloaded from NCBI in GenBank file format and parsed using BioPython version 1.81 (Cock et al. 2009) to create a comma-separated values (CSV) file (Supporting Information). Data was curated using pandas version 2.0.1 (McKinney 2010) to count and rank metrics of interest (below). Hereafter, 'NCBI records' refers to the records we extracted from NCBI that were listed as being collected from AoNZ.

### 2.2 | Current Practice in DNA Data Storage

DNA data are typically deposited in repositories by researchers who are largely conducting publicly funded research. Submitting primary data to an open-access data repository is often a mandatory requirement imposed both by the funding agencies and by journals where the results of the research may be published. The peer review process that is, a critical evaluation to ensure the veracity of the research undertaken, also often requires access to the data by expert reviewers.

The process of data deposition typically involves providing scientifically accurate metadata, including organismal details (e.g., Latin name, taxonomic placement, tissue type, host), environment and collection details (e.g., location, biome, collector(s), institutional affiliations), scientific purpose (e.g., study design, associated journal publication) and data-specific metadata (e.g., molecule type, gene, sequencing methodology), among other information (Clark et al. 2016). Post-deposition, the integrity of submitted data and metadata is checked for accuracy and completion against metadata standards set by the database, prior to inclusion and hosting of the data.

### 2.3 | Analysis of General Trends

We first examined broad trends across BOLD and NCBI databases. All BOLD records consisted of mtDNA COI data, while NCBI contained a much broader range of genetic data types. To understand the type of genetic data being deposited in NCBI for AoNZ samples, we first extracted metadata from the fields: 'organelle' (e.g., plastid, mitochondrion), 'molecule type' (mRNA, genomic DNA), and 'gene' (e.g., COI, ribulose biphosphate carboxylase—rbcL, a chloroplast gene often used as a marker to investigate photosynthetic organisms; Bell et al. 2017, etc.). We then classified records as: (A) 'mitochondrion' or 'plastid' if they had either of these terms in the organelle field; (B) 'mRNA', 'tRNA', 'rRNA' or 'Genomic RNA' if this was listed in the molecule type field, regardless of any further data in the gene or organelle fields; (C) 'genomic DNA + gene' if the record had both 'genomic DNA' in the molecule type field and a gene listed in the gene field; or (D) 'genomic DNA' if the record stated this in the molecule type field, but no further data was present in the gene or organelle fields.

Next, to understand general taxonomic trends or biases in the databases, we extracted metadata from the 'phylum' and 'species' fields on BOLD and the 'organism' field on NCBI (which provides the entire phylogenetic classification of the organism). To investigate temporal trends, we extracted data from the 'event date' field on BOLD and the 'collection date' field on NCBI, as well as the date the data was submitted to NCBI (BOLD does not have an option to extract deposition date), and examined changes in the number of records deposited over time.

Finally, to understand whether DNA data was being deposited by researchers based in AoNZ or internationally, we extracted metadata from the 'collectors' and 'institution storing' fields for BOLD, and the 'author' (i.e., authors of the cited article) and 'collected by' field attributes for NCBI.

### 2.4 | Analysis of Species Trends

Our second major focus was to understand whether certain species were researched more intensively than others. Thus, we first determined the proportion of records from each database that were sequenced to species level (i.e., data was accurate and complete enough to identify sequences to species) based on whether data was provided in the species field. We then ranked the top one hundred species in each database with respect to the number of records with species level data and searched the literature to classify the status of each species (endemic = only found in AoNZ; non-endemic = found elsewhere, including native, invasive, exotic and naturalised species), focusing specifically on whether species were endemic or non-endemic due to the ambiguity of definitions of native/exotic status, particularly of lesser known species. We were able to further distinguish the top ten species into 'endemic', 'native' or 'non-native' due to these species being more common in the literature.

To gain further insight about database practices from researchers who actively use and upload genetic data, we selected all endemic species from the top ten species with DNA data uploaded to BOLD or NCBI and sent a brief questionnaire regarding study intentions, responsibilities, funding, opportunities, and engagement with iwi Māori (Supporting Information) to the corresponding author/collector of the associated publication. Across the two databases, enough information could be obtained to trace samples back to an author or collector with a relevant email address for six endemic species (*Hydropysche fimbrata*, *Zelandobius confusus*, piriwai/spiny gilled mayfly *Coloburiscus humeralis*, tuangi/cockle *Austrovenus stutchburyi*, toheroa *Paphies ventricosa* and titipounamu/rifleman *Acanthisitta chloris*) and five lead authors. Note that in this manuscript, the Latin names of species are provided upon first mention, with the Māori and common names also provided where possible; Latin or common names are then used thereafter. Also see Table 1 for a glossary of te reo Māori terminology and some technical terms used here.

Finally, to assess whether database practices and researcher acknowledgement of iwi contributions were aligned, we conducted a literature search via the Web of Science Core Collection on 1 May 2025, using the following parameters: Funding Text = ('iwi' or 'Māori' or 'maori' or 'hapū' or 'hapu' or 'rūnanga' or 'runaka' or 'whanau' or 'ngāti' or 'ngāi') AND (ALL = [gene* or genom* or DNA]) NOT (title = [human]), and filtered the results by Countries/Regions = New Zealand. Temporal trends of resulting literature were analysed by plotting them alongside data

**TABLE 1** | Given this manuscript is at the intersection of data sovereignty and molecular technology, we provide a glossary of terms below to ensure clarity to audiences from a cross-section of these fields.

| Relevance | Term | Definition |
|---|---|---|
| Data sovereignty | Aotearoa | Māori name for New Zealand |
| | Hāpori | Māori community |
| | Iwi/hapū | Tribe/subtribe |
| | Iwi kaimahi | Workers or employees affiliated with an iwi/tribe |
| | Kaitiaki/−tanga | Custodian/−ship, guardian/−ship |
| | Kaumātua | Māori community elders |
| | Kaupapa Māori | Māori approach, customary practice, institution or ideology |
| | Kura | School, typically with lessons in te reo Māori |
| | Mātauranga | Māori knowledge system |
| | Māori | Indigenous people of Aotearoa New Zealand |
| | Mahi | Work, practice |
| | Mana | Authority, influence, power |
| | Mana whenua | Territorial rights, authority or jurisdiction over land |
| | Motu | Country, lands |
| | Ngā atua | Ancestors, gods |
| | Taiao | Environment, nature |
| | Taonga | Treasure |
| | Taonga species | Typically pre-European contact species present in Aotearoa New Zealand or those that arrived with Māori |
| | Te ao Māori | Māori worldview |
| | Te reo Māori | The Māori language |
| | Te Tiriti o Waitangi | The Treaty of Waitangi—founding document of Aotearoa New Zealand signed by the Crown and Māori leaders (1840) that enshrines the rights of Indigenous peoples in relation to the declaration of New Zealand as a British colony. However, since then, significant breaches of these rights occurred and many are still being resolved via the Waitangi Tribunal |
| | Tikanga | Cultural practices, customs, protocols |
| | Tino rangatiratanga | Self-determination, sovereignty, autonomy |
| | Wānanga | Conversations, discussions, workshops |
| | Whakapapa | Genealogy |
| | Whenua | Land |
| Molecular technology | Barcoding | Sequencing a small, conserved gene or locus for species identification. For example, cytochrome oxidase 1 gene for invertebrates and internal transcribed spacer 1 (non-coding region) for fungi are commonly used barcodes. Typically, between 600 and 1500 base pairs, these universal regions represent $<0.0000001\%$ of an organism's genome |
| | DNA | Deoxyribonucleic acid—molecule that carries an organism's genetic information |
| | eDNA | Environmental DNA—DNA derived from environmental samples such as water, soil, air |
| | Genome sequencing | Sequencing the total genomic content of an organism, typically representing a near complete profile of all nuclear and organelle genomes |
| | Metabarcoding | Barcoding of mixed species samples to reveal community composition. For example, water samples can be processed with metabarcoding to identify all the plants or insects present in the local environment |
| | OTU | Operational taxonomic units, represents an equivalent to a unique 'species' derived from metabarcoding sequencing data from mixed species communities |
| | Reference sequences | High quality DNA sequence representing a taxonomically identified, often vouchered specimen of a known species |

deposition trends for the NCBI database to assess the relationship between the two.

## 3 | Results

Broadly, neither of the two major assessed repositories requires evidence of consultation with Indigenous communities or allows for Indigenous metadata inclusion; indeed, the process of data deposition has been optimised to minimise barriers for data submission. The primary mechanism for enabling Indigenous data sovereignty on open access public databases thus remains through data stewardship, whereby only accurate representation of the data held within databases (specifically collection location, species metadata, and collector or institutional information) can allow Indigenous communities to access and connect with the genetic data.

### 3.1 | General Trends

BOLD repository stores primarily mitochondrial markers, thus as expected all records were mtDNA (Figure 1A). We examined the type of data that was deposited to NCBI and found that the majority of records (72%; $n = 207,403$) were labelled as 'genomic DNA' with no further data in the gene or organelle fields. Sixteen percent ($n = 45,537$) of the records were derived from mRNA, 5% ($n = 15,838$) from 'genomic DNA' (with additional information on the specific gene or locus sequenced), 5% ($n = 14,983$) from mitochondrial genes, and 2% ($n = 5013$) from plastid genes. The remaining 648 records (< 1%) were derived from other data types, such as tRNA, rRNA and genomic RNA (Figure 1B).

Examining taxonomic trends/biases of submissions, we found that the majority of BOLD records were from AoNZ-collected Arthropoda (90%; $n = 71,987$), followed by Chordata (4%; $n = 3510$), Mollusca (2%; $n = 1652$), and all other phyla (2%; $n = 1808$). One percent of BOLD records ($n = 975$) did not specify the phylum (Figure 1C). In contrast to BOLD, arthropods comprised just 8% ($n = 23,849$) of NCBI records, which were instead dominated by Chordata (59%; $n = 169,729$) and Nematoda (28%; $n = 82,201$; Figure 1D). Tracheophyta was the only plant-related phylum to appear in either database's top phyla, making up 2% of NCBI records ($n = 5290$; Figure 1D). All other phyla ($n = 19$) comprised 3% of remaining NCBI records ($n = 8353$).

The year of collection for samples deposited in both BOLD and NCBI peaked from 2011 to 2015 (Figure 1E,F). However, while the majority of BOLD records provided the date that the sample was collected (93%; $n = 74,084$), over 92% of NCBI records did not provide temporal information ($n = 266,872$). NCBI was the only database that provided information on the date of submission (i.e., the date data was uploaded to the database), with two peaks in data submissions occurring in 2014 (34% of total submissions; $n = 98,518$) and 2017 (41% of total submissions; $n = 118,297$; Figure 2).

Finally, the majority of BOLD records provided the names of the collectors (91%; $n = 72,371$; Figure 1G), whereas only 2% of all NCBI records provided collector information ($n = 5530$). Instead, the majority of NCBI records included named authors of the publication presenting the data (63%; $n = 181,162$), and a further 37% ($n = 107,317$) provided no names associated with the record (however, 65% [$n = 69,970$] of these no-author submissions belonged to a single whole genome shotgun sequencing project for rifleman; accession: PRJNA253841; Figure 1F). Virtually all (99.9%; $n = 77,489$) BOLD records named the institution that stored the DNA data, with the Centre for Biodiversity Genomics, University of Guelph, Canada, storing the most records ($n = 54,717$; 68%). Seven of the top ten institutions were based in AoNZ ($n = 17,985$ records in total; Table 2). NCBI provided no metadata field to identify the depositing institution.

### 3.2 | Temporal Trends in Data Deposition and Literature Attributions

Data deposition to NCBI database remained steady throughout the studied period, with the exception of 2014 and 2017 (Figure 2), which coincided with the deposition of two whole genome shotgun sequencing projects: rifleman (accession: PRJNA212877) and *Teladorsagia circumcincta* (accession: PRJNA72569) respectively. In contrast, we observed a steady increase in the acknowledgement of iwi Māori in genetic studies conducted in AoNZ over the study period (Figure 2). This persistent increase followed the establishment of the first formal advocacy group of Māori researchers and practitioners, Te Mana Rauranga (Māori Data Sovereignty Network; https://www.temanararaunga.maori.nz/) in 2015. This group launched the Charter of Māori Data Sovereignty in 2016, outlining the principles for Māori data control, access, and use, followed by a consolidated model for system-wide governance of Māori data in 2023 (Kukutai et al. 2023). Throughout this period of increased advocacy in AoNZ and internationally, tools and approaches for the implementation of indigenous data sovereignty have also grown, including the 2017 establishment of Genomics Aotearoa (the national genomics platform in AoNZ) (https://www.genomics-aotearoa.org.nz/), the introduction and popularisation of BioCultural (BC) and Traditional Knowledge (TK) Notices by Local Contexts (https://localcontexts.org/labels/biocultural-labels/) in collaboration with Māori to guide use of indigenous knowledge and genomic data in digital data repositories, the establishment of CARE principles ((C)ollective benefit, (A)uthority to control, (R)esponsibility, and (E)thics; https://www.gida-global.org/care) in 2019, and the utilisation of BC and TK labels in large genome sequencing projects such as the Earth BioGenome Project and the Darwin Tree of Life (Mc Cartney et al. 2024).

### 3.3 | Species Trends

The majority of the BOLD records lacked species-level data (77%; $n = 61,729$; Figure 3A) compared to NCBI, where 98% ($n = 282,226$) of records were available at species-level (Figure 3B). Ranking the top one-hundred species with species-level data (with respect to number of records), 39% of BOLD species and 66% of NCBI species were recognised as endemic to AoNZ (Figure 3C,D). The top ten species from BOLD represented 3% of the total BOLD records ($n = 2487$; Figure 3E) and included six endemic species (*Hydropsyche fimbriata*,


**FIGURE 1** | Legend on next page.

**FIGURE 1** | General trends in current DNA data storage: (A and B) data type of BOLD and f NCBI records respectively; (C and D) proportion of phyla represented in BOLD and NCBI records respectively (see Supporting Information for all phyla); (E and F) date of collection for BOLD and NCBI records, respectively; (G) proportion of records that had/did not have information about who collected the data for BOLD; and (H) proportion of records that had information for author, collector, both, or neither for NCBI.

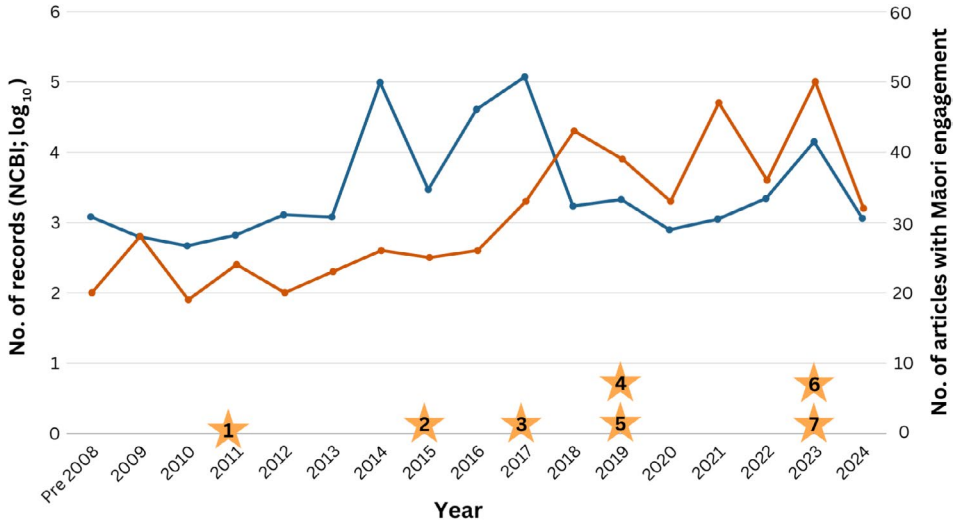


**FIGURE 2** | Temporal trends in the submission of genetic data to NCBI (blue line; main axis) and acknowledgement of indigenous communities in publications (orange line; secondary axis). Stars indicate important milestones in the advocacy of Māori data sovereignty: (1) Wai 262 report released (Ko Aotearoa Tēnei); (2) establishment of Te Mana Raraunga—the Māori Data Sovereignty Network; (3) establishment of Genomics Aotearoa; (4) introduction of traditional knowledge (TK) and biocultural (BC) labels; (5) introduction of CARE principles for indigenous data governance; (6) expansion of TK and BC labels within biodiversity genomics; and (7) publication of the Te Mana Raraunga model for system-wide governance of Māori data.

tuangi/cockle, *Zelandobius confusus*, *Coloburiscus humeralis*, *Archichauliodes diversus*, and *Austrosimulium australense*) and four non-endemic species (*Psychoda sigma*, *Trialeurodes vaporariorum*, *Lycoriella sativae*, and *Opogona omoscopa*; Figure 3E). In contrast, the top ten species list from NCBI represented 90% of the total NCBI records ($n = 261{,}357$; Figure 3F), and was largely dominated by the endemic rifleman (33% of total records; $n = 96{,}915$), followed by the non-native platyhelminthe *Teladorsagia circumcincta* (28% of total records; $n = 81{,}759$; Figure 3F). Similarly to BOLD, the top ten NCBI species included many non-endemic species (*T. circumcincta*, European hedgehog, *Microctonus aethiopoides*, *Microctonus hyperodae*, and dunnock *Prunella modularis*). However, while BOLD had zero native species, NCBI had four (kūaka/bar-tailed godwit *Limosa lapponica*, red damselfly *Xanthocnemis zealandica*, *Atriophallophorus winterbourni*, and New Zealand mud snail *Potamopyrgus antipodarum*).

We received replies from two authors in response to our enquiries regarding their research and database experiences. Both authors indicated that the data was collected and used for PhD research projects, with one relating to freshwater insects including caddisflies, stoneflies, and mayflies (*H. fimbriata*, *Zelandobius confucius*, and *C. humeralis*, respectively), and the other relating to rifleman. The purpose of the former project was to investigate dispersal patterns and population connectivity of stream insects (funded by National Institute of Water and Atmospheric Research, NIWA; and the University of Waikato, including its Environmental Research Institute, ERI), while the latter was to understand the position of rifleman in the avian phylogeny and its potential possession of genomic signatures common to vocal learning species (funded by The Royal Society Te Āparangi Marsden fund and Birds NZ). Both projects required Department of Conservation (DOC) ethics and export permits to collect samples from protected lands and send them offshore for DNA sequencing. Furthermore, both indicated that contact was made with tangata whenua (people of the land) to obtain permissions, with one author noting that regular updates were emailed to tangata whenua throughout the research process. Both authors noted that no funding or specific resources were available to support collaboration with iwi, and one author (for rifleman) further highlighted that meaningful collaboration was limited by the availability of iwi kaimahi (iwi workers) and distrust by iwi of some researchers.

## 4 | Discussion

### 4.1 | Current Database Infrastructures and Missing Metadata Limit Opportunities for Māori Data Sovereignty

We evaluated records of genetic data from organisms that had been collected in AoNZ and stored in either BOLD or NCBI databases to investigate the current practice of storing potentially taonga data. We discovered that current practices limit Māori sovereignty/custodial rights over such data: many endemic taonga species have genetic data uploaded to public repositories

**TABLE 2** | The top ten institutions listed as storing DNA data on BOLD, the country where the institution was located, and the number of records deposited (*n*). A full record of institutions is available in Supporting Information.

| Institution | Country | *n* |
|---|---|---|
| Centre for Biodiversity Genomics | Canada | 54,717 |
| Landcare Research Arthropod Collection | AoNZ | 7435 |
| University of Waikato | AoNZ | 2, 823 |
| National Institute of Water and Atmospheric Research, Wellington | AoNZ | 2325 |
| Museum of New Zealand Te Papa Tongarewa | AoNZ | 1907 |
| National Institute of Water and Atmospheric Research, Auckland | AoNZ | 1497 |
| National Institute of Water and Atmospheric Research, Hamilton | AoNZ | 1366 |
| Canterbury Museum | AoNZ | 632 |
| Senckenberg Natural History Collections Dresden, Museum of Zoology | Germany | 542 |
| Canadian National Collection of Insects, Arachnids and Nematodes | Canada | 506 |
| Total | | 77,489 |

with no restrictions or guidelines regarding who has authority over the data and how it can be used/shared. Furthermore, current database infrastructure limits the ability to achieve data stewardship due to poor standards of metadata accompanying sequences and a large degree of missing information relevant for understanding whether the data was collected and uploaded ethically and accurately. This is an important finding that will help shape the development of a future reference DNA database for AoNZ that proposes a partnership model where the rights and interests of Indigenous communities are valued and prioritised.

We extracted over 350,000 publicly available records of genetic data from over 3000 different species across the two public repositories. However, the true number of records from AoNZ could be much higher, considering that 88% of sequence data in publicly available databases is not accompanied by a specified country of origin and thus would not have been picked up by our search terms (Mc Cartney et al. 2022). Indeed, missing metadata was a common thread among the general trends of our study, particularly NCBI, where a large proportion of records were missing information regarding collection date. Species biases also dominated our results, with the majority of AoNZ-collected species data coming from arthropods (90% in BOLD) and Chordata (59% in NCBI) and plants massively under-represented in both databases—collectively highlighting important taxonomic gaps that require urgent attention. Temporal trends appeared to indicate a drop-off in deposition of AoNZ-collected data following increased deposition during the mid 2010s, but it remains unclear whether this trend is accurate for the NCBI database as 92% of data lodged with NCBI lacked temporal information.

When examining species trends, we found that the majority of the data deposited for the top one-hundred species across both databases represented endemic species. For example, rifleman is New Zealand's smallest bird—a taonga species treasured by Māori for being a messenger to Tāne, the atua (god) of the forest (Kendrick et al. 2020). Yet, despite its cultural significance, many of the records for rifleman came from a project to sequence its entire genome ($n = 69{,}970$ records), and were deposited to NCBI without any information in the author field. Although we were eventually able to identify author information at the level of the BioProject accession number for these records, this exemplifies the lack of accuracy and completeness in metadata deposition. Indeed, NCBI had a large proportion of records (98%) with blank author and collector fields, though most BOLD records named the sample collector (87%) and institution (99.9%) that deposited the data.

Depositing data to public repositories, which is often compulsory for many publishing journals, is important for fostering transparency and establishing rigour of research and has been an extremely productive way of maximising outcomes from genetic data. However, current open-access policies appear to be unintentionally facilitating negligent stewardship standards, which may limit the ethical value, transparency, and thus reproducibility of the research (Hudson et al. 2020; Mc Cartney et al. 2023). Exacerbating this problem, neither NCBI nor BOLD currently provide a metadata field stipulating who has cultural authority for governing the data, nor whether consents were obtained to collect and share the data. However, we are aware that BOLD developers are actively exploring the incorporation of such metadata fields into a future iteration of BOLD in collaboration with indigenous data sovereignty experts (pers. comm. Ratnasingham, S., Centre for Biodiversity Genomics).

### 4.2 | Towards the Development of Enduring Genetic Resources for Improved Indigenous Data Sovereignty and Research Outcomes

We contend that improving data stewardship practices for genetic databases can not only improve research outcomes (Toczydlowski et al. 2021; Vaughan et al. 2025), but is the first step in embedding indigenous data sovereignty. For Māori, it is important that DNA sequence data for taonga species represents an enduring resource, produced for and benefiting future generations (Hudson et al. 2021). Poor metadata standards, such as missing attribution for collector and collection location and a lack of cultural authority attribution, run counter to this aspiration. While tension between the global benefits of open access data on the one hand and protection of the cultural rights and sovereignty of Indigenous communities on the other persists, improving data stewardship practices remains a shared goal (Carroll et al. 2020). Finding this common ground and embedding perspectives beyond those that are largely technical or infrastructural is essential in solving this complex issue.


**FIGURE 3** | (A and B) the proportion of records that were available at 'species level' for BOLD and NCBI, respectively; (C and D) the proportion of the top one-hundred species with species-level data that were endemic or non-endemic to AoNZ for BOLD and NCBI, respectively; and (E and F) the top ten species lists, showing the number of records (*n*), and whether species were determined to be endemic, native, or non-native. Māori and common names have been used where possible, otherwise, Latin names are provided.

Indigenous scholars and scientists, in AoNZ and internationally, have proposed a number of frameworks and opportunities to improve metadata standards, create more equitable access to data and culturally responsive practice, and establish prior informed consent or engagement with Indigenous communities (e.g., Hudson et al. 2021, 2016). Incorporating these practices successfully remains challenging; however, recent efforts have been promising. For example, the Atlas of Life in Australia (https://atlasoflife.org.au/) and the Rakeiora project in AoNZ (https://www.genomics-aotearoa.org.nz/our-work/health-projects/rakeiora-pathfinder-genomic-medicine) both prioritise Indigenous rights through co-development and leveraging of

Indigenous Knowledge (IK) to improve conservation and health outcomes.

In line with these efforts, a national reference database for AoNZ should commit to strict metadata standards that, alongside standard metrics (e.g., location, date of collection, sex, etc.), include a record of engagement with mana whenua, ethics approvals if required, contributors and collaborators (including authors, collectors, mātauranga contributors), and publication outputs (Forsdick et al. 2021). Declaration of rights and sovereignty via territorial bounds of iwi/hapū should also be applied through globally recognised approaches. BioCultural (BC) and Traditional Knowledge (TK) Notices could be utilised to this end, with Indigenous groups able to add BC Labels to data to signal rules that allow mana whenua to govern its appropriate and consensual use (Liggins et al. 2021; Wu et al. 2022; see an example for bilberry here: https://www.vaccinium.org/bio_data/1085390). Another way to potentially ensure appropriate governance is to specify that the submitter of the primary material (DNA or specimens), as opposed to the database, owns the data, a condition that Wilderlab—an eDNA laboratory based in Wellington, New Zealand—recently updated in their terms (5.1: https://www.wilderlab.co.nz/terms). Finally, progress on implementing Indigenous data sovereignty across a variety of data types, repositories and institutions can be appraised through The Collaboratory for Indigenous Data Governance, which advocates for action beyond recognition of Indigenous rights towards tangible change in institutional policies and practices (https://indigenousdatalab.org/).

### 4.3 | Researcher Acknowledgement of Indigenous Contributions Rise, Despite Funding and Database Infrastructure Barriers

Likely in recognition of the growing advocacy by indigenous scholars in AoNZ and internationally, we found that the acknowledgement of indigenous communities in genetics research publications has grown over the past two decades. While a lack of iwi attribution in data deposition across public databases persists due to lack of metadata requirements, many researchers have utilised acknowledgement sections in journal publications, where nuanced details about iwi engagement could be recorded, as an alternative avenue. Similar trends have been noted in other fields of biological research, such as ecology (Wehi et al. 2019) and systematics (Veale et al. 2019), where researchers are finding unique ways of highlighting the importance of iwi engagement and their contributions through co-design, co-development, co-authorship, and the use of indigenous languages, such as te reo Māori in naming new species. These examples broadly represent progress in the right direction, yet researchers continue to face barriers in engaging meaningfully with iwi. Such barriers can include a limited number of iwi kaimahi who are 'spread too thin', making it difficult to know who to contact, the hesitance of Māori who may be sceptical of the motives of non-Māori (see McAllister 2022), and the tight timelines and open data requirements of funders. Addressing these issues will require better resourcing (including financial, which both authors in our questionnaire identified as not being available) to foster genuine engagement between Māori and researchers to reach more equitable outcomes from which Māori communities can clearly benefit. Enabling the development of resilient relationships between Māori and non-Māori working with taonga species will also foster effective identification of appropriate points of contact for such engagement. Overcoming well-founded distrust in this space is also necessary, and will require academic leadership based on kaupapa Māori—where science relating to taonga is led by Māori, positioned in te ao Māori (including mātauranga, te reo Māori and tikanga), and underpinned by Te Tiriti o Waitangi (Moko-Painting et al. 2023). Facilitating this, guidelines for conducting genomic research on taonga species based on kaupapa Māori and at the intersection of genomics and te ao Māori (e.g., Collier-Robinson et al. 2019) have recently been published (Hudson et al. 2021). Increased support of kaupapa Māori and mātauranga at the systemic and institutional level will further enhance Māori leadership and engagement for such research (Haar and Martin 2022; Wehi et al. 2019).

## 5 | Concluding Remarks

Transformative change is urgently needed to protect and conserve the remaining biodiversity on Earth (Leadley et al. 2022). Reference databases are an important asset to this goal by providing baseline data, cataloguing and monitoring biodiversity, and allowing a benchmark for conservation outcomes (Kress et al. 2015; Dopheide et al. 2025). In AoNZ, a national reference DNA library could aid the safeguarding of taonga species data and contribute to the conservation of the taiao (environment).

It is encouraging to see that the majority of AoNZ-collected samples in the current study were collected by researchers at AoNZ institutions. However, as demonstrated here, current practices of poor metadata storage and an absence of attribution of data to indigenous communities who may have custodial rights undermine the sovereignty of Māori. Now is the time to consider how to best design these databases moving forward, and key to ensuring the best scientific and cultural outcomes is the prioritisation of responsible, ethical, equitable access and use of samples and their DNA (Mc Cartney et al. 2023). By acknowledging and incorporating Indigenous perspectives, involving Indigenous communities and researchers within the scientific process can foster a more equitable and empowering relationship between researchers and Indigenous communities.

#### Acknowledgements

The authors would like to thank Maui Hudson (University of Waikato), Steve Pawson and Shaun Ogilvie (University of Canterbury), and Melanie Mark-Shadbolt (Bioheritage National Science Challenge) for support and discussions about this work. The authors also acknowledge two anonymous reviewers for their constructive feedback.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

No new data was generated in this study. Full records of database entries downloaded from NCBI GenBank and BOLD and analysed in this study are deposited on DRYAD, doi: 10.5061/dryad.j6q573npw.

## References


Bell, K. L., V. M. Loeffler, and B. J. Brosi. 2017. "An *rbcL* Reference Library to Aid in the Identification of Plant Species Mixtures by DNA Metabarcoding." *Applications in Plant Sciences* 5: 1600110. https://doi.org/10.3732/apps.1600110.

Berry, T. E., M. L. Coghlan, B. J. Saunders, et al. 2023. "A 3-Year Plankton DNA Metabarcoding Survey Reveals Marine Biodiversity Patterns in Australian Coastal Waters." *Diversity and Distributions* 29: 862–878. https://doi.org/10.1111/ddi.13699.

Blackman, R. C., J.-C. Walser, L. Rüber, et al. 2022. "General Principles for Assignments of Communities From eDNA: Open Versus Closed Taxonomic Databases." *Environmental DNA* 5: 326–342. https://doi.org/10.1002/edn3.382.

Carroll, S. R., I. Garba, O. L. Figueroa-Rodríguez, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science* 19, no. 43: 1–12. https://doi.org/10.5334/dsj-2020-043.

Carter, L. 2005. "Naming to Own Place Names as Indicators of Human Interaction With the Environment." *AlterNative: An International Journal of Indigenous Peoples* 1: 6–24. https://doi.org/10.1177/117718010500100102.

Clapcott, J., J. Ataria, C. Hepburn, et al. 2018. "Mātauranga Māori: Shaping Marine and Freshwater Futures." *New Zealand Journal of Marine and Freshwater Research* 52: 457–466. https://doi.org/10.1080/00288330.2018.1539404.

Clark, K., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2016. "GenBank." *Nucleic Acids Research* 44: D67–D72. https://doi.org/10.1093/nar/gkv1276.

Cock, P. J., T. Antao, J. T. Chang, et al. 2009. "Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics." *Bioinformatics* 25: 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

Collier-Robinson, L., A. Rayne, M. Rupene, C. Thoms, and T. Steeves. 2019. "Embedding Indigenous Principles in Genomic Research of Culturally Significant Species." *New Zealand Journal of Ecology* 43: 1–9. https://doi.org/10.20417/nzjecol.43.36.

Collins, M. A., C. M. Browne, T. L. Edwards, et al. 2022. "How Low Can They Go: A Comparison Between Dog (*Canis familiaris*) and Environmental DNA Detection of Invasive Koi Carp (*Cyprinus rubrofuscus*)." *Applied Animal Behaviour Science* 255: 105729. https://doi.org/10.1016/j.applanim.2022.105729.

Deagle, B. E., S. N. Jarman, F. P. Coissac, and P. Taberlet. 2014. "DNA Metabarcoding and the Cytochrome *c* Oxidase Subunit I Marker: Not a Perfect Match." *Biology Letters* 10: 20140562. https://doi.org/10.1098/rsbl.2014.0562.

Dhami, M. K., and H. Hohaia. 2023. "Te Tiriti-Guided National DNA Reference Library Wānanga Series. Bioheritage National Science Challenge, Strategic Investment Priority." https://www.landcaresearch.co.nz/events/national-dna-database-webinar-series/.

DOC. 2016. "New Zealand Biodiversity Action Plan. Department of Conservation." https://www.doc.govt.nz/documents/conservation/new-zealand-biodiversity-action-plan-2016-2020.pdf.

Dopheide, A., T. Brav-Cubitt, A. Podolyan, et al. 2025. "Fast-Tracking Bespoke DNA Reference Database Generation From Museum Collections for Biomonitoring and Conservation." *Molecular Ecology Resources* 25: e13733. https://doi.org/10.1111/1755-0998.13733.

Forsdick, N. J., C. I. Adams, A. Alexander, et al. 2021. *Current Applications and Future Promise of Genetic/Genomic Data for Conservation in an Aotearoa New Zealand Context (Science for Conservation 337)*. Department of Conservation. https://www.doc.govt.nz/globalassets/documents/science-and-technical/sfc337entire.pdf.

Haar, J., and W. J. Martin. 2022. "He Aronga Takirua: Cultural Double-Shift of Māori Scientists." *Human Relations* 75: 1001–1027. https://doi.org/10.1177/00187267211003955.

Hare, K. M., S. B. Borrelle, H. L. Buckley, et al. 2019. "Intractable: Species in New Zealand That Continue to Decline Despite Conservation Efforts." *Journal of the Royal Society of New Zealand* 49: 301–319. https://doi.org/10.1080/03036758.2019.1599967.

Harmsworth, G. R., and S. Awatere. 2013. "Indigenous Māori Knowledge and Perspectives of Ecosystems." In *Ecosystem Services in New Zealand—Conditions and Trends*, edited by J. R. Dymond and J. R. Dymond, 274–286. Manaaki Whenua Press.

Hikuroa, D. 2017. "Mātauranga Māori—The Ūkaipō of Knowledge in NEW ZEALAND." *Journal of the Royal Society of New Zealand* 47: 5–10. https://doi.org/10.1080/03036758.2016.1252407.

Hotaling, S., J. L. Kelley, and P. B. Frandsen. 2021. "Toward a Genome Sequence for Every Animal: Where Are We Now?" *PNAS* 118: e2109019118. https://doi.org/10.1073/pnas.2109019118.

Hudson, M., A. Beaton, M. Milne, et al. 2016. *Te Mata Ira: Guidelines for Genomic Research With Māori*. Te Mata Hautū Taketake—Māori & Indigenous Governance Centre, University of Waikato.

Hudson, M., N. A. Garrison, R. Sterling, et al. 2020. "Rights, Interests and Expectations: Indigenous Perspectives on Unrestricted Access to Genomic Data." *Nature Reviews Genetics* 21: 377–384. https://doi.org/10.1038/s41576-020-0228-x.

Hudson, M., A. Thompson, P. Wilcox, et al. 2021. *Te Nohonga Kaitiaki Guidelines for Genomic Research on Taonga Species*. Te Kotahi Research Institute, University of Waikato.

Kahui, V., and A. Cullinane. 2019. "The Ecosystem Commons." *New Zealand Journal of Ecology* 43: 1–5. https://doi.org/10.20417/nzjecol.43.28.

Kendrick, J., A. Jonathan, and M. Black. 2020. "Te Ahu Turanga Cultural Impact Assessment. Ngāti Kahungunu ki Tāmaki Nui-A-Rua." https://www.horizons.govt.nz/HRC/media/Media/Consent-Documents-Folder/Te-Ahu-a-Turanga-Manawatu-Tararua-Highway-Construc/Ngati%20Kahungunu%20ki%20Tamaki%20nui-a-Rua%20CIA%201%20March%202020.pdf.

Kress, W. J., C. Garcia-Robledo, M. Uriarte, and D. L. Erickson. 2015. "DNA Barcodes for Ecology, Evolution and Conservation." *Trends in Ecology & Evolution* 30: 25–35. https://doi.org/10.1016/j.tree.2014.10.008.

Kukutai, T., K. Campbell-Kamariera, A. Mead, et al. 2023. *Māori Data Governance Model*. Te Kāhui Raraunga.

Leadley, P., A. Gonzalez, D. Obura, et al. 2022. "Achieving Global Biodiversity Goals by 2050 Requires Urgent and Integrated Actions." *One Earth* 5: 597–603. https://doi.org/10.1016/j.oneear.2022.05.009.

Liggins, L., M. Hudson, and J. Anderson. 2021. "Creating Space for Indigenous Perspectives on Access and Benefit-Sharing: Encouraging Researcher Use of the Local Contexts Notices." *Molecular Ecology* 30: 2477–2482. https://doi.org/10.1111/mec.15918.

Liu, J., F. Slik, S. Zheng, and D. B. Lindenmayer. 2022. "Undescribed Species Have Higher Extinction Risk Than Known Species." *Conservation Letters* 15: e12876. https://doi.org/10.1111/conl.12876.

Mc Cartney, A. M., J. Anderson, L. Liggins, et al. 2022. "Balancing Openness With Indigenous Data Sovereignty: An Opportunity to Leave No One Behind in the Journey to Sequence All of Life." *Proceedings of the National Academy of Sciences* 119: e2115860119. https://doi.org/10.1073/pnas.2115860119.

Mc Cartney, A. M., G. Formenti, A. Mouton, et al. 2024. "The European Eference Genome Atlas: Piloting a Decentralised Approach to Equitable Biodiversity Genomics." *NPJ Biodiversity* 3: 28. https://doi.org/10.1038/s44185-024-00054-6.

Mc Cartney, A. M., M. A. Head, K. S. Tsosie, et al. 2023. "Indigenous Peoples and Local Communities as Partners in the Sequencing of Global Eukaryotic Biodiversity." *NPJ Biodiversity* 2: 8. https://doi.org/10.1038/s44185-023-00013-7.

McAllister, T. 2022. "50 Reasons Why There Are No Māori in Your Science Department." *Journal of Global Indigeneity* 6: 1–10.

McKinney, W. 2010. "Data Structures for Statistical Computing in Python." *Proceedings of the 9th Python in Science Conference* 445: 51–56. https://doi.org/10.25080/Majora-92bf1922-00a.

Moko-Painting, T., L. Hamley, D. Hikuroa, et al. 2023. "(Re)Emergence of Pūtaiao: Conceptualising Kaupapa Māori Science." *Environment and Planning F* 2: 11–37. https://doi.org/10.1177/26349825231164617.

Moura, M. R., and W. Jetz. 2021. "Shortfalls and Opportunities in Terrestrial Vertebrate Species Discovery." *Nature Ecology & Evolution* 5: 631–639. https://doi.org/10.1038/s41559-021-01411-5.

Steiner, K., T. Drinan, A. Zaiko, et al. 2022. "Heterogenous Distribution of Kākahi (Freshwater Mussel *Echyridella menziesii*) Environmental DNA in 5 New Zealand Lakes of Differing Size and Geomorphology." *Freshwater Science* 42: 1–12. https://doi.org/10.1086/722968.

Taiuru, K. 2022. *Tikanga Tawhito Tikanga Hou Kaitiaki Guidelines for DNA Research, Storage and Seed Banks With Taonga Materials*. Karaitiana Taiuru. https://ndhadeliver.natlib.govt.nz/delivery/DeliveryManagerServlet?dps_pid=IE81094224.

Theissinger, K., C. Fernandes, G. Formenti, et al. 2023. "How Can Genomics Help Biodiversity Conservation." *Trends in Genetics* 39: 545–559. https://doi.org/10.1016/j.tig.2023.01.005.

Toczydlowski, R. H., L. Liggins, M. R. Gaither, et al. 2021. "Poor Data Stewardship Will Hinder Global Genetic Diversity Surveillance." *Proceedings of the National Academy of Sciences* 118: e2107934118. https://doi.org/10.1073/pnas.2107934118.

Vaughan, A. L., E. Parvizi, P. Matheson, A. McGaughran, and M. K. Dhami. 2025. "Current Stewardship Practices in Invasion Biology Limit the Value and Secondary Use of Genomic Data." *Molecular Ecology Resources* 25, no. 5: 1–13. https://doi.org/10.1111/1755-0998.13858.

Veale, A. J., P. de Lange, T. R. Buckley, et al. 2019. "Using te reo Māori and ta re Moriori in Taxonomy." *New Zealand Journal of Zoology* 43: 3388–3398. https://doi.org/10.20417/nzjecol.43.30.

Wehi, P. M., J. R. Beggs, and T. G. McAllister. 2019. "Ka Mua, ka Muri." *New Zealand Journal of Ecology* 43: 1–8. https://doi.org/10.20417/nzjecol.43.40.

Wilkinson, S. P., A. A. Gault, S. A. Welsh, et al. 2024. "TICI: A Taxon-Independent Community Index for eDNA-Based Ecological Health Assessment." *PeerJ* 12: e16963. https://doi.org/10.7717/peerj.16963.

Wu, C., C. Deng, E. Hilario, et al. 2022. "A Chromosome-Scale Assembly of the Bilberry Genome Identifies a Complex Locus Controlling Berry Anthocyanin Composition." *Molecular Ecology Resources* 22: 345–360. https://doi.org/10.1111/1755-0998.13467.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** men70021-sup-0001-Supinfo.docx.