Current applications and future promise of genetic/genomic data for conservation in an Aotearoa New Zealand context

Natalie J. Forsdick, Clare I.M. Adams, Alana Alexander (Ngāpuhi: Te Hikutu), Anna C. Clark, Levi Collier-Robinson (Ngāi Tahu, Ngāti Porou, Te Whānau-ā-Apanui, Ngāti Apa ki te Rā Tō), Ilina Cubrinovska, Max Croll Dowgray, Eddy J. Dowle, Laura Duntsch, Stephanie J. Galla, Lucy Howell, Molly Magid, Aisling Rayne, Alexander J.F. Verry, Jana R. Wold and Tammy E. Steeves
## Contents

### Abstract

Kia taiaro ora, kia tangata ora

<table>
<thead>
<tr>
<th>1. Conducting conservation genetic/genomic research in Aotearoa New Zealand</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Introduction</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Considerations for taonga species</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Selecting an appropriate conservation genetics/genomics tool</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Technical considerations for conservation genetic/genomic research</td>
<td>8</td>
</tr>
<tr>
<td>1.4.1 Best-practice sample collection and storage for genetic/genomic research</td>
<td>8</td>
</tr>
<tr>
<td>1.4.2 Metadata collection and management</td>
<td>9</td>
</tr>
<tr>
<td>1.4.3 Data sovereignty</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Genetic/genomic data types and conservation applications</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Pedigree data</td>
<td>12</td>
</tr>
<tr>
<td>2.1.1 Best practice for pedigree establishment and maintenance</td>
<td>13</td>
</tr>
<tr>
<td>2.1.2 Challenges of pedigrees and genetic solutions</td>
<td>13</td>
</tr>
<tr>
<td>2.2 Genetic data</td>
<td>14</td>
</tr>
<tr>
<td>2.2.1 Nuclear genetic data</td>
<td>14</td>
</tr>
<tr>
<td>2.2.2 Mitochondrial data</td>
<td>14</td>
</tr>
<tr>
<td>2.2.3 DNA profiling</td>
<td>15</td>
</tr>
<tr>
<td>2.2.4 Limitations of genetic data</td>
<td>15</td>
</tr>
<tr>
<td>2.3 Reduced-representation genomic data</td>
<td>15</td>
</tr>
<tr>
<td>2.3.1 RAD-seq data</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2 Microarrays</td>
<td>16</td>
</tr>
<tr>
<td>2.3.3 Limitations of reduced-representation approaches</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Whole-genome sequencing</td>
<td>17</td>
</tr>
<tr>
<td>2.4.1 Reference genomes</td>
<td>17</td>
</tr>
<tr>
<td>2.4.2 Population-level resequencing</td>
<td>19</td>
</tr>
<tr>
<td>2.5 Ancient DNA</td>
<td>19</td>
</tr>
<tr>
<td>2.5.1 Challenges and limitations associated with aDNA</td>
<td>20</td>
</tr>
<tr>
<td>2.6 Environmental DNA</td>
<td>20</td>
</tr>
<tr>
<td>2.6.1 Conservation applications of environmental DNA</td>
<td>21</td>
</tr>
<tr>
<td>2.6.2 Limitations of environmental DNA</td>
<td>21</td>
</tr>
<tr>
<td>2.7 Transcriptomics</td>
<td>22</td>
</tr>
<tr>
<td>2.7.1 Conservation applications of transcriptomics</td>
<td>22</td>
</tr>
<tr>
<td>2.7.2 Challenges and limitations for conservation transcriptomics</td>
<td>23</td>
</tr>
<tr>
<td>2.8 Microbiomes</td>
<td>24</td>
</tr>
<tr>
<td>2.8.1 Methodological considerations associated with microbiome analysis</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Future conservation genetic/genomic research tools and directions</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 PangENomes and genome graphs</td>
<td>26</td>
</tr>
<tr>
<td>3.2 Chromosomics and structural variants</td>
<td>26</td>
</tr>
<tr>
<td>3.3 Characterising adaptive variation</td>
<td>27</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>3.4</td>
<td>Hologenomics</td>
</tr>
<tr>
<td>3.5</td>
<td>Genome editing</td>
</tr>
<tr>
<td>3.6</td>
<td>Biobanking</td>
</tr>
<tr>
<td>3.7</td>
<td>Multi-species genomic pipelines</td>
</tr>
<tr>
<td>3.8</td>
<td>Improved facilitation of genetics/genomics research to enhance conservation outcomes</td>
</tr>
<tr>
<td>4.</td>
<td>Glossary</td>
</tr>
<tr>
<td>5.</td>
<td>Acknowledgements</td>
</tr>
<tr>
<td>6.</td>
<td>References</td>
</tr>
</tbody>
</table>
Current applications and future promise of genetic/genomic data for conservation in an Aotearoa New Zealand context

Natalie J. Forsdick1,2, Clare I.M. Adams3, Alana Alexander3 (Ngāpuhi: Te Hikutu), Anna C. Clark3, Levi Collier-Robinson1 (Ngāi Tahu, Ngāti Porou, Te Whānau-ā-Apanui, Ngāti Apa ki te Rā Tō), Iлина Cubrinovska1, Max Croll Dowgray1, Eddy J. Dowlle3, Laura Duntsch4, Stephanie J. Galla1,5, Lucy Howell1,6, Molly Magid1, Aisling Rayne1, Alexander J.F. Verry7, Jana R. Wold1 and Tammy E. Steeves1

1 School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140.
2 Manaaki Whenua – Landcare Research, Private Bag 92170 Auckland Mail Centre, Auckland 1142.
3 Department of Anatomy, University of Otago, 270 Great King Street, Dunedin, 9016.
4 School of Biological Sciences, University of Auckland, 3 Symonds Street, Auckland 1010.
5 Department of Biological Sciences, Boise State University, Idaho, USA, 83706.
6 Gateway Antarctica, School of Earth and Environment, University of Canterbury, Private bag 4800, Christchurch 8140.
7 Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9016.

Corresponding author: Natalie Forsdick, natalieforsdick@gmail.com or forsdickn@landcareresearch.co.nz

Abstract

The field of conservation biology has a long history of incorporating diverse disciplines into its ‘toolbox’ for improved outcomes. One such discipline is conservation genomics, which has experienced fast-paced growth and development over the last decade and offers exciting opportunities to help achieve the vision outlined in Aotearoa New Zealand’s national strategy for biodiversity – Te Mana o te Taiao. However, integrating these emerging methodologies into meaningful conservation practice has proven challenging, mostly due to uncertainty around the utility of these data and effective allocation of limited funding. This report addresses these challenges by outlining potential strategies for utilising genetic/genomics in conservation from the perspective of predominantly early-career conservation researchers working as Te Tiriti o Waitangi partners. It is intended to initiate discussion among conservation practitioners and researchers, mana whenua and local communities. To support practitioners in identifying appropriate and cost-effective genetic/genomic tools, their associated costs and benefits for informing conservation management are presented. Because conservation genetic/genomic data generated for – and associated with – taonga (treasured) species are also taonga, the report emphasises the need for collaborative research partnerships that centre the needs, aspirations and expertise of mana whenua, and highlights key aspects of data management and sovereignty. A transdisciplinary approach to conservation that includes genetics/genomics is recommended.

Keywords: conservation genetics, conservation genomics, molecular markers, Indigenous data sovereignty, threatened species management, New Zealand, genetic diversity, genome sequencing and assembly.

© Copyright August 2022, Department of Conservation. This paper may be cited as:
Kia taiao ora, kia tangata ora
by Isobel Joy Te Aho-White

Cover illustration details

*Kia taiao ora, kia tangata ora* by Isobel Joy Te Aho-White combines the principles of science through a Western lens with Mātauranga Māori understandings of whakapapa. The structures of DNA and a phylogenetic tree are treated with traditional Māori design to bring them into a contemporary, multicultural context. Here the DNA is envisaged as pātiki kōwhaiwhai, a pattern seen on the rafters of wharenui, to represent manaakitanga (hospitality), our connection to the environment, and how nourishment of the environment in turn nourishes the people. Behind is an abstracted phylogenetic tree with a weaving pattern treatment based on raukūmara and poutama designs, to represent connections/whakapapa and learning respectively.

About the artist

Isobel Joy Te Aho White (Ngāti Kahungunu ki Wairoa, Kāi Tahu, Pākehā ki te Cornwall me te Denmark) is an illustrator based in Wellington, with a Bachelor of Design from Massey University, majoring in Illustration. Much of Isobel’s work incorporates Māori design and plant life that is native to Aotearoa New Zealand, and is inspired by her ancestral lineages of healers on both Māori and Pākehā sides.
1. Conducting conservation genetic/genomic research in Aotearoa New Zealand

1.1 Introduction

The national strategy for biodiversity, Te Mana o te Taiao (DOC 2020), challenges Aotearoa New Zealand to restore and enhance biodiversity for future generations. Many taonga (treasured) animals and plants are threatened by habitat loss, disease, invasive species, incidental bycatch, direct hunting, and climate change. To achieve the vision of protection, restoration and sustainable use of biodiversity outlined in Te Mana o te Taiao, we must bring together diverse ways of knowing and seeing the world – each of which bring their own unique toolboxes, including the use of genetic/genomic data (Rayne et al. 2020).

Genetic and genomic data provide a lens for exploring the interconnections, histories and future of populations through DNA. Such knowledge is critical to better understand the present state of our biodiversity and to co-develop robust, evidence-based management strategies for threatened species. For example, many populations face challenges associated with inbreeding, loss of genetic diversity and, ultimately, reduced capacity to respond to future change (i.e. adaptive potential; de Villemereuil et al. 2019). Strategies such as conservation breeding programmes and conservation translocations (i.e. moving plants or animals to promote gene flow and enhance diversity for existing populations or establishing new populations; Seddon 2010) can play an important role in minimising inbreeding, increasing genome-wide diversity and enhancing adaptive potential (Mable 2019).

Recent technological advances and increased capacity and capability in the global genomics community further enable researchers and practitioners to ask new questions and revisit old concepts. For instance, the shift away from using a handful of neutral genetic markers toward whole-genome resequencing allows investigation of adaptive (maladaptive) variation and has reignited interest in the role of structural variants (large-scale rearrangements within the genome; e.g. Lamichhaney et al. 2016; Weissensteiner et al. 2020). Similarly, developing research highlights the potential role of gene drives (a genome editing technology that increases the likelihood of an allele with a known beneficial or detrimental effect being inherited) in managing pest species or reintroducing critical genetic variation into threatened species (Dearden et al. 2018; Phelps et al. 2020).

However, the rapid expansion of conservation genetics/genomics puts researchers and practitioners under pressure to keep up to date with an increasingly complex toolbox (Taylor et al. 2017a). Here, genetic data refers to the use of a relatively small subset of variable loci assumed to be representative of the diversity present within the genome (the entire complement of DNA of an organism or species). Genomic data refers to data generated with high-throughput DNA sequencing methods to characterise genome-wide variation across many thousands of loci. Both data types have benefits and challenges associated with development and analysis that make them suited to specific applications. Beyond new tools (e.g. reduced-representation sequencing, whole-genome resequencing, genome editing technologies), the resulting datasets and their potential applications can be numerous and confusing. Further, existing knowledge, capabilities and aspirations vary widely across taxa. For example, while kākāpō (Strigops habroptilus) recovery is informed by a dedicated species recovery team and an extensive long-term dataset including a high-quality

1 See Glossary for definitions of technical terms presented in bold font at first mention.

To navigate genetic/genomic technologies and co-develop conservation management approaches that (i) enhance biodiversity and (ii) empower all individuals and groups involved, scientists and conservation practitioners must prioritise clear communication and genuine partnership (Jarvis et al. 2020). Namely, the framing, co-development and application of genetic/genomic research should be determined by the species’ needs, as well as the needs, aspirations and expertise of the individuals and groups involved, especially mana whenua (tribal group(s) with customary rights over a defined area of land or territory, including the air and water; Collier-Robinson et al. 2019). In Aotearoa New Zealand, Te Tiriti o Waitangi (1840) provides such a framework for partnership between Māori and non-Māori. Further, the WAI 262 claim and the subsequent Waitangi Tribunal report (Jones 2012) provide a clear mandate for mana whenua to maintain kaitiakitanga (stewardship) over data or resources arising from taonga species (Waitangi Tribunal 2011).

We are aware that conservation practitioners may have limited opportunities to delve into the capabilities and promise of genetic/genomic tools for conservation, particularly when they may not encounter such tools in day-to-day operations. In this review, we – a cross-institutional team of predominantly early-career researchers using genetic/genomic tools for conservation applications – provide an overview of existing genetic/genomic methodologies and the current and aspirational applications of such data for conservation. Many of these technologies may have been previously considered with regard to Aotearoa New Zealand’s bioheritage (Inwood et al. 2020), but here we focus on those aspects of greatest relevance to Te Mana o te Taiao in a conservation management context. We present a table of current attributes for existing genetic/genomic tools to assist conservation practitioners in identifying appropriate tools that can be further discussed with conservation genetic/genomic researchers. We also highlight the importance of data sovereignty and data management considerations, identify future tools and applications for these data, and consider ways in which we can enhance conservation outcomes by better facilitating such research in Aotearoa New Zealand.

1.2 Considerations for taonga species

As conservation genetic/genomic researchers in Aotearoa New Zealand, our research primarily involves species which are taonga to iwi, hapū and whānau. Through whakapapa (genealogy), data associated with these species are taonga in their own right, including data obtained through genetic/genomic methods (Collier-Robinson et al. 2019). Thus, one of the most critical aspects in the application of genetic/genomic data to Aotearoa New Zealand conservation is ensuring that research is undertaken in a manner that upholds Te Tiriti o Waitangi, recognises the rangatiratanga (authority) of mana whenua, the mauri of the species and ecosystems in question, and prioritises Māori research leadership (Harmsworth & Awatere 2013). Trusted research partnerships that centre the needs, aspirations and expertise of mana whenua will enhance the capacity of Māori and non-Māori research partners to restore and enhance biocultural diversity (Rayne et al. 2020; Wehi et al. 2020a). Through iterative engagement, clear data management plans can be co-created that realise Indigenous data sovereignty and ensure benefit-sharing (see 1.4.2 Metadata collection and management). Although this contribution focuses on taonga species, we consider the research and data management practices described here to be broadly applicable across all Aotearoa New Zealand conservation genetic/genomic research.
While there is a growing scholarship that describes the need to engage with mana whenua when seeking to apply genetic/genomic tools to modern samples (Collier-Robinson et al. 2019; Hudson et al. 2020; Rayne et al. 2020) as well as identifying the whakapapa of human remains (e.g. Knapp et al. 2012b), there is relatively limited scholarship in regards to engagement relating to the use of historical samples originating from non-human organisms, whether held in museums or private collections (but see Wehi et al. 2021). In addition, the legal, ethical and social ramifications associated with environmental DNA (eDNA) research in an Aotearoa New Zealand context are rarely discussed in the peer-reviewed literature, despite the potential for unforeseen impacts on Indigenous communities where such data may identify (or fail to identify) links between Indigenous communities and the land (see Handsley-Davis et al. 2021 for details). Further, emerging microbiome studies in conservation consider not just the genome of a focal taxon but also the genomes of its associated microbial communities (see 2.8 Microbiomes), which also contribute to the mauri of species and ecosystems. Researchers should be applying the same process of iterative engagement for genetic/genomic research pertaining to modern samples to the use of historical specimens, fossils, and environmental and microbiome samples, whether the focal taxa are extinct or extant, as such data is taonga (Collier-Robinson et al. 2019; Wehi et al. 2021). It is important to ensure that mana whenua have kaitiakitanga and rangatiratanga over such taonga data outputs, and that benefit-sharing is established, and we expand on related scholarship in 1.4.3 Data sovereignty.

1.3 Selecting an appropriate conservation genetics/genomics tool

As we describe in Chapter 2, the wide array of genetic/genomic methodologies can make choosing the appropriate data type(s) for specific conservation applications a daunting task. Conservation researchers and practitioners must clearly define research questions and conservation goals to facilitate this decision-making. The selected genetics/genomics approach may also be influenced by additional factors including ethical considerations regarding sampling strategies, timeliness of the method to inform management decisions and the potential for future applications of the data to answer other research questions.

In terms of day-to-day management queries such as genetic sexing, existing partnerships with researchers may provide the platform for carrying out this work. For example, University of Canterbury researchers in the Conservation, Systematics and Research Team provide genetic sexing to recovery programmes free of charge when associated with active postgraduate research projects. There are also other opportunities for similar infrequent or intermittent projects on genetic sexing, species identification and wildlife forensics to be carried out by commercial service providers (e.g. EcoGene®), with larger genomic projects involving genomic sequencing for conservation delivered by commercial sequencing providers (e.g. The Elshire Group Ltd., Auckland Genomics, Otago Genomics Facility). In our experience as conservation genetic/genomic researchers, funding for conservation genetic/genomic projects is often obtained from external funding bodies (e.g. research grants on local, national or international scales) in combination with Department of Conservation (DOC) support. In these circumstances, excluding any in-kind contributions, DOC’s primary contribution to research projects is not financial, but rather the logistical support and institutional knowledge supplied throughout the development and implementation of a given research project (e.g. assistance with obtaining samples and critical metadata). As such, the financial cost of such research may not be the primary determinant in DOC’s decisions to establish a research project. However, a potential future challenge of the current funding system is that external research funding (particularly large national grants) is increasingly driven by the development of novel approaches and may become less accessible for routine conservation genetic applications using standard tools. Obtaining conservation genetic/genomic data should be considered a valuable investment as it represents a long-term resource for ongoing management.
Table 1. Attributes associated with the establishment and implementation of various genetic/genomic tools for conservation applications, as summarised in Chapter 2.

For each tool presented, we provide an example research question that could be answered. We then provide current (May 2022) estimates of the various costs associated with using the presented tool to answer the research question. Costs include sampling invasiveness (which may have ethical and permitting impacts); time (from outset of research through to implementing conservation management actions based on research results); financial (including those associated with staff hours, laboratory consumables); and resources required to generate, analyse, and store the data (encompassing technical skill, laboratory facilities (lab type and equipment), analysis software).

As this is a rapidly evolving field, costs are similarly changing, and so we recommend periodic review of this table (e.g. by an advisory group such as that recommended in 3.8 Improved facilitation of genetics/genomics research to enhance conservation outcomes). Time, costs and resource requirements for data generation, analysis and storage are, in part, dependent on genome size and complexity of the focal organism (with the smallest and lowest complexity genomes (e.g. microbes, some insects, birds) having the lowest requirements compared with those for larger and more complex genomes (e.g. mammals, amphibians, and some plants, insects and fish).

SNPs = single-nucleotide polymorphisms, kB = kilobytes, MB = megabytes, GB = gigabytes, TB = terabytes.

<table>
<thead>
<tr>
<th>TOOL</th>
<th>Example research question</th>
<th>Invasiveness of sampling</th>
<th>Time from initiation to implementation of management</th>
<th>Financial ($NZ)</th>
<th>Laboratory resources for data generation</th>
<th>Computational resources for data analysis and storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedigree</td>
<td>Which individuals would make good pairings in a captive breeding programme?</td>
<td>Observational sampling</td>
<td>&gt; 3 years</td>
<td>&lt; 10,000</td>
<td>Low</td>
<td>Low (on the order of MB)</td>
</tr>
<tr>
<td>Genetic markers (e.g. microsatellite panel)</td>
<td>Has neutral genetic diversity been maintained following population establishment at a new site?</td>
<td>Blood or tissue sampling</td>
<td>1–2 years</td>
<td>&lt; 10,000</td>
<td>Moderate</td>
<td>Low (kB–MB)</td>
</tr>
<tr>
<td>Reduced-representation genomic sequencing</td>
<td>What is the founder representation within a captive breeding population?</td>
<td>Blood or tissue sampling</td>
<td>1–2 years</td>
<td>&lt; 30,000</td>
<td>Moderate</td>
<td>Moderate (on the order of MB)</td>
</tr>
<tr>
<td>Genome resequencing</td>
<td>What is the genetic basis of disease susceptibility?*</td>
<td>Blood or tissue sampling</td>
<td>1–2 years</td>
<td>&lt; 50,000**</td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td>Reference genome</td>
<td>Is a particular cryptic species present at particular locations?</td>
<td>Environmental sampling</td>
<td>1–3 years</td>
<td>&lt; 20,000</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>eDNA</td>
<td>How genetically diverse was this species prior to human arrival?</td>
<td>Use of pre-existing samples (e.g. from museum collections) or fossil excavation</td>
<td>1–3 years</td>
<td>&lt; 20,000</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>aDNA</td>
<td>How will a species respond to a climate change?</td>
<td>Blood or tissue sampling, lethal sampling (whole organism, organ collection, gut content)</td>
<td>1–5 years</td>
<td>&lt; 20,000</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Transcriptomes</td>
<td></td>
<td>Blood or tissue sampling, skin swabs, fecal samples, lethal sampling (whole organism, organ collection, gut content)</td>
<td>1–3 years</td>
<td>&lt; 20,000</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Microbiomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
### Additional considerations

| Pedigree validation may be required using genetic or genomic markers | Estimates derived from neutral genetic markers may not reflect functional diversity | Typically used alongside a reference genome to improve accuracy of estimates | Typically used alongside a reference genome to improve accuracy of estimates | Typically used alongside population-level sequencing data to provide population-level estimates of conservation-relevant metrics | Cannot be used to confirm species absence | aDNA methods are often used in combination with genetic markers or SNPs generated from modern samples for comparison | Typically used in combination with a reference genome | Hologenomics combines genome and microbiome sequencing, so as such is more expensive, and is in an exploratory phase for conservation applications |

### Additional benefits

| Can readily integrate additional samples downstream | Can readily integrate additional samples downstream | Can readily integrate additional samples downstream | Can readily integrate additional samples downstream, and be applied to answer additional research questions | Can readily be applied to answer additional research questions | Can readily include additional sampling locations | Provides a temporal aspect to genetic studies | Potential to identify SNPs from transcriptomic data | Can readily be applied to answer additional research questions |

### Specific resources required

| Method to identify individuals (e.g. individual banding), spreadsheets and/or pedigree management software | Genetics lab, including reagents and equipment; analysis software | Genetics lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; large data storage capacity | Genetics lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; large data storage capacity | Genetics lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; reference databases; large data storage capacity | eDNA clean lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; reference databases; large data storage capacity | aDNA clean lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; reference databases; large data storage capacity | RNA clean lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; reference databases; large data storage capacity | Genetics lab, including reagents and equipment; sequencing provider; high-performance computing system; analysis software; reference databases; large data storage capacity |

*We acknowledge that this question cannot be answered with a reference genome alone. However, in a conservation context, reference genomes will always be generated as a resource, and never as the sole tool informing management.

**Costs are highly dependent on the genome size and complexity of the focal species and the desired reference genome quality – lower quality (i.e. ‘good enough’) genomes may require low-end expenditure, while high quality (i.e. ‘platinum’) genomes requiring multiple sequencing types may substantially increase costs.*
With rapid technological advancements and broad overlap between many tools, there is no simple decision-making tool to help conservation practitioners select the methodologies that best fit the question at hand. In Table 1 we attempt to bridge this gap by providing estimates of costs, benefits and considerations when implementing the various tools described in Chapter 2. This table can be used as a starting point when considering implementation of genetic or genomic research to support conservation management. As sequencing costs continue to decline and new analytical methodologies are developed, many of the technical challenges associated with these tools will be reduced over time. Further, additional unforeseen benefits may arise. Other considerations for practitioners may include the extent to which routine conservation management may need to be altered to incorporate sample or (meta)data collection, or the potential downstream impacts of results on current management practice. Currently, genetic tools may appear more cost-effective than genomic tools across all measures described here and may still be sufficient to answer the question of interest. However, these efficiencies must be weighed against the known limitations of genetic data, particularly as we move towards assessment of adaptive (or maladaptive) variation (see 3.3 Characterising adaptive variation).

One key technical challenge associated with the shift towards genomic methodologies is the increasing scale of the computational requirements (and consequently, additional financial costs) associated with analysis and storage of large genomic datasets. These requirements vary depending on the scale of the data, with population-level genomic analyses requiring access to high-capacity computational systems that can process and analyse data up to the scale of terabytes. While many institutions may provide local computing infrastructure, national and international services are also available (e.g. the New Zealand eScience Infrastructure www.nesi.org.nz, cloud computing services such as Amazon Web Services aws.amazon.com or, more locally, Catalyst Cloud catalystcloud.nz). Such services may incur additional expenses to projects and require specific data security considerations (see 1.4.3 Data sovereignty).

1.4 Technical considerations for conservation genetic/genomic research

Clearly defining research questions and selecting appropriate sampling protocols and data management are key components of any research project. Data management plans are essential and should encompass the short- and long-term management (including curation, storage and access) of samples, raw data and associated metadata, and processed data and outputs. Such plans are particularly critical for large and/or long-term projects where many different people will be involved over the lifetime of the project.

1.4.1 Best-practice sample collection and storage for genetic/genomic research

To ensure preservation for genetic/genomic purposes, where possible high-quality samples (e.g. blood, tissue) should be collected once per individual in accordance with local tikanga and stored in a manner appropriate for the intended downstream application (Fig. 1). For example, while storing samples in ethanol for downstream genetics/genomics is common and may be the most practicable method for population genomic research of widespread species, it impedes the use of the samples for transcriptome analysis. Collecting samples in ways that ensure usage for a range of potential downstream applications reduces the need for resampling, minimising stress on sampled individuals. These samples will act as a resource for future genetic or genomic applications, particularly those requiring high-molecular-weight DNA (e.g. long-read sequencing requires very high-quality DNA; Amarasinghe et al. 2020). Sampling, storage and DNA extraction protocols differ for eDNA and ancient DNA (aDNA) research, and relevant protocols should be followed (Hofreiter et al. 2001; Jarman et al. 2018).
In the absence of a centralised national repository, we encourage researchers to follow best practice for collection and curation of samples. Samples should be recorded in laboratory databases to minimise the potential for unnecessary resampling and to ensure sample usage can be tracked. Sample databases should capture individual identification, laboratory identification, type and quantity of raw material, and use of the material for analysis, along with all relevant metadata associated with the samples to provide the necessary context for downstream analyses (including data sovereignty details; see 1.4.2 Metadata collection and management). We encourage researchers and practitioners to collect other metadata (e.g. phenotypic data, ecological data) simultaneously with sampling where possible, providing broader downstream data applications.

Figure 1. Sampling and storage requirements for various genetic/genomic tools currently in use. These should be considered the minimum requirements and are best considered in a broader context that includes an assessment of current and future use, as well as feasibility and cost.

Note: aDNA = ancient DNA, eDNA = environmental DNA, RRS = reduced-representation sequencing, 3C = 3C sequencing technologies.

1.4.2 Metadata collection and management

Metadata refers to ‘data about the data’. For example, metadata associated with blood samples taken from individual birds may include information such as collection date, GPS location, species, individual identifiers (e.g. band/tag numbers), photographs of individuals and/or sampling locations, age/class, sex, pedigree (parents, siblings, offspring). Metadata associated with raw (unprocessed) genomic sequence data could include (but is not limited to) the genomic library preparation details, sequencing provider and platform, date of sequencing, individual or project barcode sequences, individual or location identifiers (where appropriate), and sequence quality scores. Metadata facilitates research reproducibility and data re-use (Duntsch et al. 2021; Toczydlowski et al. 2021). For processed data this may include details of analysis tools and software version numbers, reference genomes used for sequence alignment, or details of databases from which additional data was collected, or output data deposited. Metadata may also encompass records of consultation with mana whenua, ethics approvals and sampling permit numbers, lists of collaborators and contributors, sources of funding and publication outputs (e.g. student theses, journal articles, DOC internal reports). We recommend that relevant metadata are captured and stored alongside raw and processed data to ensure correct interpretation of the data (e.g. at minimum in README files alongside any reference genomes, spreadsheets capturing pedigree, phenotype or monitoring data, or captured in purpose-built databases such as that of the Genomic Observatories Metadatabase (GEOME); Riginos et al. 2020). An example of metadatabasing in Aotearoa New Zealand is that of the Ira Moana Project (sites.massey.ac.nz/iramoana/), aiming to aggregate metadata associated with primarily marine genetic/genomic studies based on international metadatabasing standards (Riginos et al. 2020).
There is extensive scholarship regarding data sovereignty pertaining to human genetic/genomic data and growing scholarship pertaining to culturally significant species (e.g. Claw et al. 2018; Collier-Robinson et al. 2019; Caron et al. 2020; Hudson et al. 2020; Koia & Shepherd 2020; Potenski 2020; Walter et al. 2020; Handsley-Davis et al. 2021; McCartney et al. 2022; Rayne et al. 2022). In addition, multiple initiatives have been established to operationalise these concepts, such as the CARE principles for Indigenous data governance to complement the FAIR principles for data management (Wilkinson et al. 2016; Carroll et al. 2020, 2021); Te Nohonga Kaitiaki (www.genomics-aotearoa.org.nz/projects/te-nohonga-kaitiaki; Hudson et al. 2021), Tikanga Tawhito: Tikanga Hou Kaitiaki Guidelines (https://www.taiuru.maori.nz/guidelines-for-dna-research-storage-and-seed-banks-with-taonga-materials/; Taiuru 2022); Biocultural Labels and Notices (localcontexts.org/labels/biocultural-labels/; Anderson & Hudson 2020; Liggins et al. 2021). For brevity, we do not reproduce existing scholarship here, but we urge researchers and practitioners to explore the resources referenced above and to centre mana whenua needs, aspirations and expertise in all aspects of research and practice.

Below, we draw attention to four important, although not exhaustive, points for researchers and practitioners to consider:

i. To date, there have been disparities in the levels of protection given to different genetic/genomic data types, where genomic data are generally given greater protections than genetic data. However, engaging with scholarship such as Handsley-Davis et al. (2021) will better enable researchers and practitioners to communicate the risks and benefits associated with these data to mana whenua.

ii. Discussion of project-specific considerations regarding data security and accessibility during analysis and long-term storage should be initiated at the outset of the project (Box 1). Many researchers will have access to centralised repositories maintained by their institutes (e.g. crown research institutes (CRIs) and universities), where data can be securely held on behalf of mana whenua, as existing international genetic/genomic databases may not yet be capable of meeting data sovereignty requirements. Existing data repositories can adapt to meet these needs through the inclusion of Biocultural Labels and Notices to indicate data provenance, associated ethics, permits, and expectations around appropriate use and to connect data back to Indigenous communities (Anderson & Hudson 2020; Liggins et al. 2021). Indigenous needs may conflict with the current global emphasis on open-access data availability to facilitate reproducibility of research, but these positions are not diametrically opposite (as discussed at length in international scholarship including Carroll et al. 2020, 2021; McCartney et al. 2022).

iii. Where mana whenua hold sovereignty over data, these data are not inaccessible for re-use. Instead, the onus is on researchers to engage with mana whenua to request access, providing transparency in the intended use, applications and potential downstream impacts of these data.

iv. Similarly, international research journals must adapt to be responsive to Indigenous views and values. There has been concern among researchers around the ability to publish research when data sovereignty agreements may require restricted data access; however, international journals are beginning to respond to these needs (Potenski 2020; Liggins et al. 2021). Indeed, genomic research relating to Aotearoa New Zealand taonga has been successfully published with data hosted on behalf of mana whenua on password-protected local repositories (e.g. Galla et al. 2019, 2020; Rayne et al. 2022) and on the Aotearoa Genomic Data Repository (https://data.agdr.org.nz/; e.g. Oliphant et al. 2020; Miller et al. 2022). Until clear frameworks for data sovereignty are better
established, researchers and practitioners must continue to navigate open-access culture – particularly for methods such as eDNA which rely upon accessible databases – in ways that uphold the responsibilities mandated in Te Tiriti o Waitangi (Box 1).

v. As researchers, practitioners and mana whenua continue to grow trusted relationships – in part, through meaningful engagement around data access and storage – we foresee opportunities to co-create more impactful research, including through local knowledge and benefit-sharing (e.g. Polfus et al. 2016; Bowles et al. 2020; Gros-Balthazard et al. 2020; Henson et al. 2021; Rayne et al. 2022). Shared benefits may include opportunities to protect or enhance taonga, to grow tribal capacity – whether in Western science or through the growth of tribal knowledge, practices or processes – and/or to realise other diverse aspirations. We are encouraged by recent efforts to increase researcher accountability around benefit-sharing, e.g. through the introduction of mandatory benefit-sharing statements in some international journals.

Box 1. Key questions to consider during iterative engagement with mana whenua when developing conservation genetic/genomic research projects.

<table>
<thead>
<tr>
<th>Key questions to consider when engaging with mana whenua to develop conservation genetic/genomic research projects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engagement</strong></td>
</tr>
<tr>
<td>Which iwi/hapū/whānau should researchers engage with? What opportunities will there be for benefit sharing? What information and koha (contributions) are required to support cultural expertise of mana whenua for any research to be considered?</td>
</tr>
<tr>
<td>Scientists should not presume that all iwi/hapū/whānau will have the same knowledge, priorities and concerns, and careful engagement with all relevant parties will be key to reducing inequities relating to data management.</td>
</tr>
<tr>
<td><strong>Data generation</strong></td>
</tr>
<tr>
<td>What will the sampling strategy involve (e.g. sample type, number of individuals, locations)? Where will samples be processed? What method of data generation will be used (e.g. microsatellites, whole genome resequencing)? Where will data be generated (e.g. local/overseas sequencing)? Where will the data be analysed (e.g. by local/international researchers on local/overseas computing platforms)?</td>
</tr>
<tr>
<td><strong>Data security</strong></td>
</tr>
<tr>
<td>Where will the data be stored – including samples, raw and processed data and analysis outputs – before and after publication? How will genetic/genomic data be accessed by external researchers, before and after publication (e.g. on local servers, national/international repositories)? How will associated metadata be managed? Is there a need for metadata anonymisation (e.g. to limit sample identification and protect sample locations and individual privacy) and how will this be implemented? If individuals, whānau, hapū or iwi contribute mātauranga, how will this be explicitly recognised and protected?</td>
</tr>
<tr>
<td>To ensure preservation and protection of data associated with taonga species, a clear data management plan needs to be developed and implemented before initiation of research. It must be responsive to the needs, aspirations and expertise of mana whenua.</td>
</tr>
</tbody>
</table>
2. Genetic/genomic data types and conservation applications

In this chapter we describe the genetic/genomic tools (presented in loose chronological order of development) currently used to inform conservation management and highlight opportunities and challenges for each of these tools. Each section is intended to be a stand-alone description that can be referred to on an as-needed basis by conservation researchers and practitioners to support decision-making (along with Table 1). As this is a dynamic field, we recommend periodic review of these tools, their applications and limitations, along with those included in 3.5 Future conservation genetic/genomic research tools and directions.

2.1 Pedigree data

Pedigrees – family trees showing genealogical relationships between individuals – are a long-standing tool used in biological sciences (Wright 1922). While pedigrees are not a molecular tool as such, we classify them as a genetic tool as they are used to monitor and understand relationships and variation between individuals. Over the past 40 years, pedigrees have become a staple of conservation management, allowing practitioners to manage the genetics of small populations by strategically pairing or translocating individuals to minimise inbreeding and maximise adaptive potential (Ballou & Lacy 1995; Ivy & Lacy 2012; Galla et al. 2022). Using genealogical relationships, pedigrees can be used to produce estimates of kinship (i.e. relatedness) and individual inbreeding. Conservation breeding programmes currently prioritise pairing individuals with low mean kinship (i.e. relatedness between an individual and all others in a population). This paradigm can minimise genetic drift by maintaining the representation of the individuals that started the population (i.e. founders) to minimise loss of genetic diversity, inbreeding and adaptation to captivity (Frankham 2008; Lacy 2009). There are decades of empirical research and simulation studies that support this management approach to maximise neutral variation (e.g. Ballou & Lacy 1995; Rudnick & Lacy 2008); however, no standardised approaches to date have incorporated adaptive variation, which has only recently been able to be quantified (see 3.1 Characterising adaptive variation). Further, selection for/against individuals with adaptive/maladaptive traits (e.g. low hatching success or low immunocompetence) may inadvertently reduce the ability of populations to have sufficient evolutionary potential to adapt to future selection pressures.

Pedigrees represent an accessible tool for conservation management, as collecting and analysing pedigree data can be readily incorporated into routine management practices for most captive populations. Researchers in the zoo and aquaria communities have developed tools for studbook management (SPARKS, PopLink and ZIMS; Ballou et al. 2010; Faust et al. 2019, www.species360.org) and pedigree analysis (PMx; Lacy et al. 2012), which has increased the uptake of this approach. While these management approaches have often been applied to captive or ex situ populations, their use is increasing for wild or semi-wild populations (Pemberton 2008). A reliable pedigree is an asset for creating pairing or translocation recommendations but can also be used to evaluate heritability of specific traits (Randolph et al. 1981), understand the fitness and contributions of individuals to a population over time (Hunter et al. 2019) and model population growth and viability (Lacy 2000).
2.1.1 **Best practice for pedigree establishment and maintenance**

There are four key practices when establishing an effective pedigree management system:

- using a robust system to identify individuals to ensure correct assignment of relationships (Allen et al. 2019)
- collecting genetic data from founding individuals to evaluate relatedness (Bergner et al. 2014; Hogg et al. 2019)
- using monitoring technologies (e.g. cameras, RFID tags; Bonter & Bridge 2011) to ensure accurate breeding records, particularly when working with wild populations
- collecting tissue or blood samples from all individuals for downstream genetic analyses in the case of any uncertainties or known errors (Ryder & Feistner 1995; Frasier et al. 2009). This can also facilitate downstream pedigree evaluation to assist with error detection and correction.

2.1.2 **Challenges of pedigrees and genetic solutions**

While pedigrees are an intuitive tool with diverse applications in conservation, they also have limitations (Fig. 2). First, pedigrees are unlikely to be developed for species other than those that are most threatened and/or geographically restricted, where individual identification and monitoring over multiple generations is both necessary and feasible. For most pedigrees, founding individuals are assumed to be unrelated, as their relationships are typically unknown. However, for threatened species that have experienced severe population bottlenecks, it is unlikely that the founders are completely unrelated (Bergner et al. 2014; Hogg et al. 2019). This ‘founder effect’ is exacerbated when pedigrees are shallow (< 5 generations recorded; Balloux et al. 2004; Pemberton 2004; Rudnick & Lacy 2008). In addition to the assumption regarding founder relatedness, pedigrees often struggle with missing data. This is particularly challenging for wild populations, where there may be difficulties in correctly identifying putative parents and offspring.

Estimates generated from pedigrees are only as accurate as the pedigree itself, so accurate individual identification and knowledge of the breeding system of the focal species are essential for creating a robust pedigree. Incorrect identification of relationships between individuals resulting from extra-pair paternity (Ewen et al. 1999; Castro et al. 2004; Forsdick et al. 2021a) or nest parasitism (Overbeek et al. 2017) may impact conclusions drawn from pedigrees in the absence of molecular genetic data (Reid et al. 2014). Such missing or erroneous data can be verified and corrected using genetic or genomic markers (Overbeek et al. 2020).

Where long-term intensive population monitoring is not feasible, post-hoc pedigrees can be generated from genetic or genomic data (Flanagan & Jones 2019). Further, relatedness estimates derived from pedigrees are probability-based and may not capture the exact extent of shared genomic variation (Bérénos et al. 2014; but see also Nietlisbach et al. 2017). Pedigree data used in conjunction with genomic data may provide the most precise estimates of relatedness to support conservation breeding programmes (e.g. genomic data can be used to estimate relatedness among founders; Hogg et al. 2018; Galla et al. 2020; Wright et al. 2021).
2.2 Genetic data

2.2.1 Nuclear genetic data

Genetic data typically comprise a small number of short regions of the genome, assumed to be representative of the neutral variation of the whole genome. Common nuclear genetic markers include allozymes, AFLPs or RFLPs (amplified/restriction fragment length polymorphisms) and microsatellites. Such markers are usually derived from repetitive regions of DNA that can be highly variable between individuals, making them ideal for investigating diversity and differentiation within and between species (Forsdick et al. 2017), parentage (Castro et al. 2004; Taylor et al. 2008; Overbeek et al. 2017), relatedness (Carroll et al. 2012) and interspecific hybridisation (Steeves et al. 2010; Cubrinovska et al. 2016), and in assessing the outcomes of conservation management actions (e.g. translocation outcomes; Heber et al. 2013). Genetic data also support wildlife forensics, tracking the illegal trade of wildlife and identifying the use of protected species in commercial products (Baker et al. 1996; Gentile et al. 2013; Ferreira et al. 2015) and disease screening (e.g. beak and feather disease; Sarker et al. 2014).

Genetic data can also be used to assess gene regions that may be under selection, such as those associated with immune function and mate choice (Miller & Lambert 2004; Kamiya et al. 2014; Grueber et al. 2015; Lillie et al. 2015; Sutton et al. 2015). Putatively adaptive loci (sites in the genome associated with genes) can be targeted and characterised through comparative genomics and species-specific primers (Alcaide & Edwards 2011; Grueber et al. 2015). These data can help both with the management of detrimental alleles (Hedrick 2001) and with the maintenance of adaptive variation at specific loci within a population (Amos & Balmford 2001; Kohn et al. 2006).

2.2.2 Mitochondrial data

Mitochondrial DNA (mtDNA) differs from nuclear data in that it represents a short (15,000–20,000 base pairs (bp)), circular sequence of DNA specific to the mitochondria (known as the mitogenome) that is maternally inherited in vertebrates and found as 10s–1000s of copies in each cell (O’Hara et al. 2019). These properties make mtDNA useful independently or in combination with nuclear genetic markers for broad-scale comparisons of population-
level diversity and differentiation (Chapple et al. 2012; Alexander et al. 2016; Mischler et al. 2018), taxonomic delimitation and phylogenetic inferences (Boon et al. 2000; Banker et al. 2017; Rosenbaum et al. 2017), and for investigating potential drivers of extinction, including in ancient DNA analysis (Allentoft et al. 2014; see 2.5 Ancient DNA). Taxonomic delimitation is of particular importance for prioritisation of conservation efforts, and complete or partial mtDNA data may be sufficient for such assessments when combined with nuclear genetic/genomic, morphological and behavioural data (but see Rubinoff & Holland 2005; Dincă et al. 2019; Pedraza-Marrón et al. 2019).

2.2.3 DNA profiling

DNA profiles consisting of genotypes constructed from multiple nuclear markers, sex-linked markers (those occurring on sex chromosomes that can be used to infer individual sex), and/or mtDNA markers (to confirm patterns of maternal relatedness) can also be used for genetic monitoring of species, including population demographic and genetic diversity estimates using repeated temporal samples (Carroll et al. 2018). Based on these samples, ‘recaptures’ of DNA profiles of individuals can be used with mark-recapture models to estimate population abundances (Taberlet et al. 1999). These approaches can also be combined with parentage analysis in a gametic-mark-recapture framework, where genotypes of individuals can be ‘recaptured’ in offspring to estimate both abundance and population connectivity (Garrigue et al. 2004; Carroll et al. 2012). These approaches are particularly useful for estimating the abundance of rare or cryptic species (i.e. from feathers, fur or faeces; Bañuelos et al. 2019) and for species where photo-identification has limited applicability due to a low instance of natural markings and/or where tagging/banding is not possible (e.g. some cetacean species such as Hector’s and Māui dolphins; Baker et al. 2013; Hamner et al. 2014a, b, 2017). Individual-based DNA profiles can additionally be used to identify rare immigration/emigration events (Hamner et al. 2014a) and for monitoring genetic erosion (Leroy et al. 2018).

2.2.4 Limitations of genetic data

Neutral genetic markers may be a poor proxy for adaptive variation (Marsden et al. 2013; Grueber et al. 2015), especially for highly variable genes like those of the major histocompatibility complex associated with immune function (Sommer 2005). The relatively small numbers of markers used (e.g. as for microsatellites) may lack the resolution required for accurate estimation of relatedness and inbreeding in genetically-depauperate species (Taylor 2015). With rapid technological advances in genome sequencing in the past decade, genome-wide analyses are now possible that in many cases can provide improved resolution and accuracy compared with genetic approaches (Supple & Shapiro 2018; Galla et al. 2020).

Specific limitations associated with mtDNA arise from its nature as a single non-recombining, maternally inherited genetic locus, meaning that it cannot be used to detect male-mediated gene flow and may be impacted by incomplete lineage sorting resulting from rapid diversification events (Paijmans et al. 2013). Furthermore, even the use of complete mitogenomes lacks the power and resolution of multiple unlinked nuclear loci (e.g. microsatellites, single nucleotide polymorphisms (SNPs); Teske et al. 2018). Thus, mtDNA is best applied in tandem with nuclear genetic or genomic data to inform species-specific conservation management.

2.3 Reduced-representation genomic data

Reduced-representation sequencing (RRS) involves sequencing a subset of the genome to identify a set of genomic variants (e.g. SNPs) across all sequenced individuals. The two primary approaches are restriction site associated DNA sequencing (RAD-seq) and microarrays, although there are other approaches such as Genotyping-in-Thousands by
sequencing (GT-seq; Campbell et al. 2015b; Schmidt et al. 2020). These approaches have similar conservation applications as genetic markers but by identifying thousands or millions of variable sites provide much greater power and resolution in analyses, increasing confidence in the accuracy of estimates (Lemopoulos et al. 2019).

2.3.1 RAD-seq data

RAD-seq encompasses a range of approaches (including Genotyping By Sequencing (GBS), reduced-representation libraries (RRL) and double-digest RAD-seq (ddRAD-seq)) that use restriction enzymes to target subsets of genome-wide loci (Andrews et al. 2016). RAD-seq in conjunction with a reference genome can provide robust estimates of genetic diversity and population structure while requiring relatively inexpensive sequencing and low computational resources. As a result, RAD-seq represents a cost-effective entry-point for conservation genomics where resources (both genomic and economic) may be limited (Andrews et al. 2016). This approach has particular advantages for species with large (> 3 Gb) or complex genomes, or where existing knowledge and/or resources are limited.

RAD-seq approaches have been used to estimate genetic diversity (Zhang et al. 2019), population demographics (Kleinman-Ruiz et al. 2017; Marandel et al. 2020), parentage assignment and relatedness estimation (Thrasher et al. 2018), interspecific hybridisation (Colston-Nepali et al. 2019; Forsdick et al. 2021b; Attard et al. 2022) and population structure and gene flow (Dierickx et al. 2015; Lavretsky et al. 2019; Rexer-Huber et al. 2019; Rick et al. 2019). Bioinformatic advances are enabling new and creative ways to leverage RRS to address a range of conservation questions (Dorant et al. 2020).

2.3.2 Microarrays

Microarrays are used to simultaneously genotype thousands of SNPs within and among populations at relatively low cost per individual. SNP-chips are one such microarray and are known for their low genotyping error rates and low rates of missing data. Development of a SNP-chip first requires identifying genome-wide variation. Most commonly, a reference genome is generated, against which resequencing data from 10–20 individuals is aligned for SNP detection (see 2.4 Whole-genome sequencing). A subset of SNPs is then selected for inclusion on the SNP-chip. SNP-chips usually only represent a small fraction of the genome (e.g. a SNP-chip comprising 50,000 SNPs from a bird with a 1.1 Gb genome represents < 0.01% of the genome). Microarrays typically require hundreds of samples for inclusion in each sequencing batch to be cost effective and so may not be the most feasible method for genotyping individuals from threatened species. Microarrays have been employed to investigate genotype-phenotype associations, trait heritability, population demographics and signatures of inbreeding in livestock and wild populations (Angeloni et al. 2012; Toro et al. 2014; Duntsch et al. 2020; Latch 2020).

2.3.3 Limitations of reduced-representation approaches

It is important to consider that RAD-seq and microarrays remain reduced-representation approaches, whereby only a small fraction of the diversity of a species’ genome can be explored. As such, RRS approaches will be superseded by genomic resequencing (see 2.4.2 Population-level resequencing) for species with relatively small genomes (< 3 Gb) as sequencing costs continue to decline. RRS provides limited utility for characterising adaptive variation and limited ability to detect other important types of diversity such as structural variants. While RAD-seq approaches are relatively cost-effective, initial high development costs and large minimum sample sizes (hundreds or even thousands of individuals) required for microarrays limit their use for conservation in the absence of large-scale consortia involvement (where 10s–100s of researchers work towards a common goal, and substantial funding can be sourced) or long-term commitments to continued research, with more feasible applications.
for human health and commercially significant species (e.g. sheep, Kijas et al. 2014; cattle, Harris & Johnson 2010). Additional challenges arise from ascertainment bias with microarrays (McTavish & Hillis 2015), or batch effects with RAD-seq approaches, whereby data generated from one sequencing batch may produce data of vastly different quality than that from another (Leigh et al. 2018).

### 2.4 Whole-genome sequencing

#### 2.4.1 Reference genomes

Despite rapid developments in DNA sequencing technologies, it is not yet possible to sequence the entire complement of DNA of an organism in one piece. Thus, genomes must be assembled from many shorter sequences (analogous to puzzle pieces). Sequencing developments to date have increased the scale of sequencing to not only span more of the genome (sequence coverage), but also to do so many times (sequence depth; Fig. 3). Increased sequence depth increases the number of sequences produced that overlap with one another, allowing more accurate assembly of sequences into genomes. These high-quality assembled genomes can be used as reference genomes to guide alignments of population-level reduced-representation or resequencing data (see 2.4.2 Population-level resequencing) for intraspecific comparisons of diversity and differentiation, or for direct interspecific comparisons.

![Figure 3. Visualisation of genome sequencing and assembly concepts using the analogy of a genome as a puzzle made of many puzzle pieces. For example, the size of the kēkēwai/freshwater crayfish (Paranephrops zealandicus) genome is 2.71 billion (2.71 Gb) base pairs, or puzzle pieces. To be confident about the arrangement of puzzle pieces, each is sequenced many times to create overlapping sequences (sequencing depth; here ranging from 3–5×, but typically approx. 40× for short-read sequencing for the purpose of genome assembly) from which sequencing and/or assembly errors can be identified and corrected. Coverage is the proportion of the genome that is sequenced. Credit: AR.](image)

Initial conservation genomics research in Aotearoa New Zealand has been heavily biased towards birds. Reference genomes have been used to inform conservation management actions including breeding recommendations for kākī (black stilt, *Himantopus novaeseelandiae*) and kākāriki karaka (orange-fronted parakeets, *Cyanoramphus malherbi*; Galla et al. 2020), assessment of adaptive potential in hihi (*Notiomystis cincta*; de Villemereuil et al. 2019) and research currently in progress aims to characterise the underlying basis of inbreeding depression and improve breeding outcomes for kākāpō (Guhlin et al. unpubl. data).

A reference genome acts as the foundation for population-level genomic analysis, and so the quality of a reference genome dictates its utility for downstream analyses. However, with the wide range of sequencing platforms, read lengths and computational pipelines for genome assembly, it is important to be aware that not all reference genomes are created equally. Substantial effort is required to produce high-quality reference genomes that can be used to address a wide breadth of conservation challenges. A fundamental requirement for sequencing and assembling high-quality genomes is a high-quality sample, obtained through best-practice sample collection and storage (see 1.4.1 Best-practice sample collection and storage for genetic/genomic research). Other requirements include a high level of technical knowledge and access to high-performance computational resources, particularly when working with large genomes (> 3 Gb).

Although sequencing costs are decreasing, these combined costs remain high and, as a result, there will usually only be one high-quality reference genome produced per species. The increasing number of reference genomes available has been supported by the efforts of large consortia that aim to assemble genomes across a wide range of taxa and/or geographic locations (e.g. the Vertebrate Genomes Project, Genome 10K Community of Scientists 2009; Koepfli et al. 2015; Bat 1K, Teeling et al. 2018; the Earth BioGenome Project, Lewin et al. 2018; the Cetacean Genomes Project, Morin et al. 2020).

The majority of genomic data in Aotearoa New Zealand to date have been generated with short-read sequencing approaches, producing millions of DNA sequences (‘reads’) typically < 500 bp in length. High sequence depth (> 40-fold) is required to produce accurate genome assemblies from short-read data alone, but high-complexity repetitive genomic regions may remain challenging to accurately assemble even with high depth short-read sequencing.

With improving technologies and decreasing costs, long-read sequencing technologies are becoming more accessible. Long-read sequencing can produce reads tens of thousands of base pairs long and is considered essential for assembly of large or complex genomes, such as for those species with polyploid or highly repetitive genomes (Scott et al. 2020). Compared to short-read sequencing alone, long reads act as a foundation to dramatically improve the ability to assemble high-quality genomes and ensure a greater proportion of the genome can be assembled more accurately (Fig. 4; Morin et al. 2020). To further improve assembly accuracy, existing short-read data may be leveraged to ‘polish’ a long-read genome assembly. Although highly beneficial for a range of applications, relatively high costs and limited availability of such long-read sequencing platforms in Aotearoa New Zealand, along with technical challenges associated with the extraction of high-molecular-weight DNA, may limit their use in the near future.

Figure 4. Visualisation of the difference between short- and long-read sequencing, using the 2.71 Gb kēkēwai (freshwater crayfish; Paranephrops zealandicus) genome as an example. When using short-read sequencing, large numbers of reads (analogous to small puzzle pieces) are required to cover the kēkēwai genome. Such short-read genomes typically have many gaps (regions of unknown sequence), particularly due to sequencing challenges associated with repetitive DNA regions. In comparison, long-read sequencing (represented by large puzzle pieces) requires fewer reads to cover the same genome, and these reads are better able to span repetitive regions, resulting in fewer gaps. Long-reads also facilitate improved scaffolding of genomes, where short-reads can be used to correct misassemblies and fill gaps (represented by the small puzzle pieces overlaid in the image for long-read sequencing. Credit: AR.
Once considered unattainable for species of conservation concern, gold- or platinum-quality genome assemblies (i.e. ultra-high-quality genome assemblies, such as that for the kākāpō reference genome from the individual known as Jane) that represent complete or near-complete chromosomes are becoming more common. Such high-quality genome assemblies require not only short-read sequencing technologies but also long-read sequencing and more recent sequencing developments such as the ability to capture the spatial structure of DNA within the nucleus (known as 3C sequencing, encompassing Hi-C and relatives, Belton et al. 2012; Ulahannan et al. 2019, preprint), and optical mapping (an image-based approach to sequencing to produce site-genomic distance data; Lam et al. 2012; Teo et al. 2015). These emerging technologies provide additional context on the landscape of genes and regulatory elements within chromosomes, and enhance genome assembly quality and completeness (Lieberman-Aiden et al. 2009; Ghurye et al. 2019; Yuan et al. 2020). Stand-alone reference genomes can also be accompanied by a reference transcriptome, identifying the complement of genes encoded in the genome that characterise the phenotype of an organism, and that enable assessment of responses to environmental change or disease (see 2.7 Transcriptomics).

### 2.4.2 Population-level resequencing

Generating data to the level of resolution required for a reference genome at population-level scale remains prohibitively expensive. Resequencing data in combination with a reference genome can provide an affordable means to investigate diversity at the population level. In contrast with a reference genome where sequence data may be sourced from multiple platforms, resequencing data consists of short-read sequencing data of individuals at low–moderate coverage (<30-fold). These short reads act like puzzle pieces that can then be aligned against the complete picture that is the reference genome. Sequences can then be compared against the reference and between individuals to identify genomic variants (e.g. SNPs) throughout the genome, allowing conservation-relevant metrics to be estimated with much greater accuracy than is possible using fewer genetic markers (Galla et al. 2020), including comparisons of genomic diversity and relatedness (Galla et al. 2020), population differentiation and structuring (Lado et al. 2020) and introgression resulting from interspecific hybridisation (Leroy et al. 2020), all of which can inform conservation management including translocations and conservation breeding programmes.

Additional applications of resequencing genomes include the ability to investigate adaptive variation (Brandies et al. 2019; see 3.3 Characterising adaptive variation). This has broad and significant implications for understanding the genomic basis of traits important for management, such as those associated with adaptation (to the environment and/or captivity) and reproductive fitness (Angeloni et al. 2012; Hoelzel et al. 2019). To this end, leveraging genomic resequencing data beyond assessments of patterns of genetic diversity requires the use of data other than SNPs (see 3.5 Future conservation genetic/genomic research tools and directions).

### 2.5 Ancient DNA

Ancient DNA (i.e. DNA isolated from old biological material, aDNA; Leonard 2008) provides a window into the past via the retrieval of DNA from sources including museum specimens, subfossils, sediment cores and coprolites. Ancient DNA can provide a useful tool for conservation managers as knowledge of past genetic diversity, geographic range expansions/contractions and the factors that lead to population declines or extinctions is important for informing management decisions (Leonard 2008; Grealy et al. 2017). Much of Aotearoa New Zealand’s endemic biota has been driven to extinction or reduced to relictual distributions following human arrival, and so aDNA provides a means to examine the genetic composition of such species prior to anthropogenic impacts.
Examples of the use of aDNA to inform the conservation and management of endemic Aotearoa New Zealand species includes quantifying temporal declines in genetic diversity (Grueber & Jamieson 2008; Tracy & Jamieson 2011; Dussex et al. 2015; Bergner et al. 2016), determining the origin of contemporary genetic structure (Tracy & Jamieson 2011), clarifying past distribution of species and populations/genetic lineages (Shepherd & Lambert 2008; Wilmshurst et al. 2014; Verry et al. 2019; Scarsbrook et al. 2021), and characterising unsustainable harvesting rates that have driven species towards extinction (Rawlence et al. 2016).

### 2.5.1 Challenges and limitations associated with aDNA

The degraded nature of aDNA requires the use of dedicated clean laboratory spaces and specialised laboratory and bioinformatic techniques (Knapp et al. 2012b). Care must be taken when generating and analysing aDNA as DNA degradation and/or modern DNA contamination can bias results and lead to erroneous conclusions. Furthermore, aDNA studies are often limited by small sample sizes, with available samples in museum collections/subfossil deposits unlikely to be representative of past populations. While much aDNA work to date has relied on mitogenome data due to the small mitogenome size and high copy-number making it relatively easy to characterise, decreasing sequencing costs mean it is becoming more feasible to generate nuclear genomic data from degraded samples. Primarily applicable to very well-preserved sources of aDNA (e.g. museum skins), sequencing of complete ancient nuclear genomes enables the direct comparison of past and present genetic diversity within populations of threatened species. This could be directly applied to some of Aotearoa New Zealand’s most threatened species (e.g. takahē and kākāpō), with well-preserved specimens collected from the 1800s onward present in museum collections (Grueber & Jamieson 2008; Dussex et al. 2018).

### 2.6 Environmental DNA

Environmental DNA (eDNA) has recently gained conservation interest due to its ability to detect rare, cryptic and invasive taxa on broad scales via non-invasive environmental sampling from water, soil and air (Taberlet et al. 2012, 2018). Methodologies can be species-specific (targeted eDNA) or have broad multi-species applications (eDNA metabarcoding), with sensitivity often equal to or greater than traditional species detection and monitoring methods using netting, electrofishing or underwater videos (Lodge et al. 2012; Lacoursière-Roussel et al. 2016; Olds et al. 2016; Evans et al. 2017; Stat et al. 2019; Goutte et al. 2020). Non-invasive sampling minimises disturbance and physical harm, which could be critical in studies of threatened species such as hoiho (yellow-eyed penguin, Megadyptes antipodes; Ellenberg et al. 2007, 2013; Young et al. 2020).

While eDNA methodologies can be as or more cost effective than traditional sampling and monitoring (Shaw et al. 2016; Evans et al. 2017; Lugg et al. 2018), patterns of eDNA dispersal in the environment must be understood in order to characterise the presence of taxa (Jane et al. 2015; Barnes & Turner 2016). Studies of eDNA dispersal through marine water have shown that macrofaunal signals do not travel far (<1 km) and may remain stratified within water layers, particularly within marine environments and other large, slow-moving bodies of water (Eichmiller et al. 2014; Jeunen et al. 2020). Thus, eDNA can be applied to detect species presence or composition in specific sites within water bodies. In comparison, lotic bodies of water such as rivers may carry eDNA far downstream (>10 km), influencing species detection far from the source (Deiner & Altermatt 2014; Carraro et al. 2018). Increasingly, air eDNA methods are being explored (Clare et al. 2021; Clare et al. 2022; Lynggaard et al. 2022) along with other novel approaches (Gregorič et al. 2022).
2.6.1 Conservation applications of environmental DNA

Conservation applications of eDNA include biosecurity (Pochon et al. 2017), site occupancy modelling (Muha et al. 2017) and detection of cryptic pest species in managed areas (Ramón-Laca et al. 2014). Detecting species presence or composition using eDNA methods may supersede current species monitoring methods and facilitate assessment of environmental stressors on ecosystems, such as those resulting from primary production (Laroche et al. 2018; Macher et al. 2018).

Environmental DNA can also be used to assess wildlife and ecosystem health (Hall et al. 2016; Strand et al. 2019; Schadewell & Adams 2021) and understand species interactions (Farrell et al. 2000; Bleijswijk et al. 2014; Nichols et al. 2015). Within Aotearoa New Zealand, analysis of faecal eDNA has been used to infer the diets of kekeno (New Zealand fur seals, *Arctocephalus forsteri*; Emami-Khoyi et al. 2016), kororā (little blue penguin, *Eudyptula minor*; Murray et al. 2011), hoiho (Young et al. 2020) and Bryde’s whales (*Balaenoptera edeni brydei*; Carroll et al. 2019). Environmental DNA-based diet analysis is especially useful for analysing soft-bodied prey, which are otherwise difficult to identify. For example, establishing the earthworm diet of the endangered endemic carnivorous land snail (*Powelliphanta augusta*) can inform habitat restoration and site suitability for translocations (Waterhouse et al. 2014).

Environmental DNA also has potential as a tool for monitoring species abundance and measuring population genetic diversity and differentiation. There is increasing interest in the correlation between eDNA abundance and species biomass with applications for tracing migratory patterns or spawning activity (Doi et al. 2015; Laramie et al. 2015; Bylemans et al. 2017; Thalinger et al. 2019; Yates et al. 2019), although analytical challenges remain regarding the impacts of factors such as size, age and cell/naked DNA shedding rates (Iversen et al. 2015; Klymus et al. 2015; Vasselon et al. 2018). There is also growing evidence that eDNA can be used to estimate conservation-relevant metrics including genetic diversity and population structure in species that are challenging to sample (Parsons et al. 2018; Adams et al. 2019; Stepien et al. 2019; Tsuji et al. 2020; Adams et al. 2022, preprint).

2.6.2 Limitations of environmental DNA

Environmental DNA is best used in tandem with traditional biodiversity surveying methods, as open-access sequence databases may be depauperate of target taxa due to regional biases or data-access limitations (Porter & Hajibabaei 2018; Sato et al. 2018). Additional limitations arise from the high sensitivity of eDNA studies that can increase the risk of false positive or false negative presence identification results due to faecal deposits by mobile predators, extreme weather events, data-deficient databases, insufficient sampling and field or laboratory contamination (Merkes et al. 2014; Goldberg et al. 2016; Dickie et al. 2018; Staley et al. 2018; Furlan et al. 2020; Ragot & Villemur 2022). As with aDNA methodologies, dedicated clean laboratories are required to minimise the potential for contamination (Goldberg et al. 2016).

With the growing use of eDNA tools in Aotearoa New Zealand and abroad, the applications for these tools are diversifying. While the metabarcoding of water and soil samples is increasingly utilised, alternative substrates, such as air (Johnson et al. 2019; Clare et al. 2021, 2022; Lynggaard et al. 2022; Roger et al. 2022), remain less explored. Further, there is variability in species identification databases (Corfe-Tan et al. 2019; Othman et al. 2020; Banerjee et al. 2021; Ito et al. 2022, preprint). eDNA techniques rely on access to accurate, complete databases of reference sequences from known taxa, making such databases a critical component for eDNA analyses. Widespread sampling is needed to generate baseline eDNA data and populate reference databases, and can be delivered through community-driven science initiatives (e.g. the Environmental Protection Agency’s Wai Tūwhera o te Taiao – Open Waters Aotearoa programme [https://www.epa.govt.nz/community-involvement/open-waters-aotearoa/]). Increasingly, companies such as WilderLab Ltd. are contributing to community science efforts.
to census local biodiversity through the provision of sampling kits, sequencing and analysis. Currently in Aotearoa New Zealand however, reference sequences for many species are absent from such databases. Nevertheless, eDNA techniques hold great promise for species and ecosystem monitoring in Aotearoa New Zealand, and such research will flourish through partnerships with mana whenua.

2.7 Transcriptomics

Although every cell within an organism contains the same set of genes, not every gene is active in every cell. The transcriptome represents the complete set of RNA transcripts reflecting the expression of genes produced from the genome of a cell, tissue or organism at a specific development stage or physiological condition (Wang et al. 2009). The function and composition of these transcripts is essential for our understanding of an organism’s phenotype. Two key aims of transcriptomics relevant to conservation are to (i) quantify changing gene expression during development or stress and (ii) annotate a genome by cataloging all transcripts. Various technologies have been developed for transcriptomic research, with RNA sequencing (RNA-seq) propelling transcriptomics beyond clinical biology applications (Wang et al. 2009; Todd et al. 2016). Unlike whole-genome sequencing where reads are aligned to an assembled reference to improve confidence and resolution, RNA-seq does not require any prior knowledge of the transcriptome, making it particularly useful for the study of novel traits in non-model organisms (organisms that have not been used for extensive research, unlike those used in clinical applications such as mice, rats, etc.; Alvarez et al. 2015; Todd et al. 2016).

Transcriptomics facilitates genome annotation, identifying features such as gene coding regions that enhance the utility of reference genomes (Yandell & Ence 2012). Accurate genome annotation enables an understanding of gene expression changes, and identification of genes associated with disease and other traits that may be relevant for conservation (Videvall et al. 2015; Connon et al. 2018). Many of these annotations can be predicted using existing databases or inferred from closely related species (Ekblom & Wolf 2014; Aken et al. 2016; Dominguez Del Angel et al. 2018), but transcriptome sequencing may be required for accurate annotation of novel gene models in non-model species (Trapnell et al. 2010).

As gene expression within a cell can be affected by stage of development, environment or stress (Wang et al. 2009), transcriptome sequencing for genome annotation often requires sequencing of various tissues, sexes and life stages to accurately represent the diversity of gene expression in a species. Gene expression changes in response to stress or across development can be profiled in an individual cell, specific tissue or the whole organism (Todd et al. 2016; Kulkarni et al. 2019). These gene expression changes are frequently referred to as the up- or down-regulation of a gene (Costa-Silva et al. 2017). Historically, gene expression changes were often assessed in individual genes; however, with the rise of RNA-seq whole transcriptome profiling has now become possible for almost any organism (Alvarez et al. 2015; Todd et al. 2016). Population-level transcriptomic data can also be used to identify genomic SNPs, thereby enabling acquisition and comparison of two data types from a single sequencing method (Lopez-Maestre et al. 2016).

2.7.1 Conservation applications of transcriptomics

Transcriptome profiling is the most efficient way to acquire a comprehensive snapshot of an organism’s physiological state and has the potential to have an immense impact on understanding wildlife health (Fig. 5), particularly for understanding the processes of and response to disease in non-model species (Field et al. 2015; Videvall et al. 2015; Campbell et al. 2018). Transcriptome profiling differs from genome annotation in that it is typically restricted to a single tissue type. Combining transcriptomics with other epidemiological
data can explain why populations differ in their functional response to a disease. Although assessments of wild populations can be challenging, a recent study on lethal viral infections in amphibians successfully used infected and non-infected wild populations to understand how a history of disease alters a population’s gene expression profile (Campbell et al. 2018). Greater understanding of disease history and population diversity can facilitate management of adaptive variation relating to immunity and development in wild populations.

![Figure 5](image)

Figure 5. Two examples of the use of transcriptomics to inform conservation management. Example A compares hypothetical immune gene expression changes between healthy and diseased samples across populations. Gene expression changes are displayed as the log-fold change between samples, where each box represents the log fold-change in expression of various genes between diseased and non-diseased samples from three populations (X, Y and Z). Genes that are up-regulated (green) have higher rates of expression in diseased samples. Due to connectivity between genes, up-regulation in one gene can result in down-regulation of others (brown), and vice versa. In this example a stronger immune response is seen in population Z. Example B compares hypothetical stress responses in an insect population reared in current and future climate conditions. Different genes display different levels of regulation across stages of development. In this example, log-fold changes are greater in populations reared in future conditions than in current conditions. Combined with ecological and physiological data, these types of studies can provide information on the adaptive potential of a species. Credit: EJD.

With climate change increasingly affecting environments, there is a growing need to understand organisms’ resilience to change (Somero 2010; Moritz & Agudo 2013). When incorporated with ecological and physiological data, comparative transcriptomic assessments across environments can infer a population’s physiological capacity to respond to changing environmental conditions (Seebacher et al. 2015; Kelly 2019; Anderson & Song 2020). Such studies typically involve moving wild individuals into controlled laboratory environments where they can be exposed to predicted future conditions (Narum & Campbell 2015; DeBiasse & Kelly 2016; Riddell et al. 2019). For example, research into the adaptive potential of an Afrotropical butterfly (Bicyclus anynana) has shown that despite expressing two distinct phenotypes in dry and wet seasons, this species has limited adaptability in the face of environmental change (Oostra et al. 2018). Research in this area can support conservation practitioners in making proactive management decisions regarding climate change.

### 2.7.2 Challenges and limitations for conservation transcriptomics

Transcriptomics has made significant contributions to our understanding of physiology, evolutionary biology and ecology; being particularly useful in gaining understanding of responses to disease or stress. However, it is still under-utilised in conservation biology (Alvarez et al. 2015; Todd et al. 2016; Connon et al. 2018), mainly because of the technique’s requirement for destructive sampling of individual tissues under tightly controlled conditions. To detect meaningful gene expression differences requires the ability to distinguish true differential expression from background noise (Todd et al. 2016). For example, to understand response to disease it would be beneficial to compare multiple samples of both infected and uninfected individuals from the same population (i.e. genetic background), sex and developmental stage in identical environmental conditions. Such experimental protocols are
often difficult to enact in wild populations of threatened species and as a result, conservation transcriptomic studies may remain limited (Connon et al. 2018). However, for species adversely impacted by disease or rapid environmental change, researchers and practitioners may need to consider whether the knowledge potentially obtainable from transcriptomic approaches outweighs the costs of lethal sampling. Research into the adaptability of species to environmental change can facilitate proactive conservation in the face of climate change and could be prioritised to focused on high-risk species identified via climate change vulnerability assessments (Wheatley et al. 2017), while considering species interactions (Hance et al. 2007; Memmott et al. 2007).

2.8 Microbiomes

So far, this discussion of conservation genetics and genomics has been largely restricted to species-specific approaches. However, the microbiome – the microorganisms that reside on or within the tissues of a host species, including bacteria, fungi, and viruses – has the potential to offer insights into key conservation questions (West et al. 2019). Although initial research into host-associated microbiomes was largely based on human health and model organisms (Gilbert et al. 2018; Davidson et al. 2020), here we review some ways in which microbiome studies may inform conservation.

Just as conservation-relevant processes such as population bottlenecks leave an impact on the genome, the microbiome can also be affected. For example, Asian tiger mosquitoes (Aedes albopictus) introduced to Italy show lower microbiomic diversity than mosquitoes from within the native range (Rosso et al. 2018). Additional processes that can impact microbiome diversity include captivity (e.g. Tasmanian devils (Sarcophilus harrissii); Cheng et al. 2015; and other species as summarised by West et al. 2019), poor physiological condition of individuals (e.g. fasting humpback whales, Megaptera novaeangliae; Vendl et al. 2020) and the presence of pathogens (Van Cise et al. 2020). Reductions in microbiome diversity are important as they are associated with negative health outcomes (Vangay et al. 2018). Many host organisms rely on symbionts for defence against pathogens (Vorburger & Perlman 2018; McLaren & Callahan 2020), and the microbiome can directly impact behaviour and memory in some species (Davidson et al. 2020). In addition, perturbation of the microbiome could have other consequences, as it appears to have an important role in local adaptation (Suzuki et al. 2019) and adapting to a changing world (Cunning & Baker 2020; Voolstra & Ziegler 2020), which may be important to consider when planning translocations.

Aspects of the microbiome could also be useful for individual and/or population monitoring. In humans, the skin microbiome can predict the age of an individual to within approximately 4 years (Huang et al. 2020b). When applied to species of conservation concern, similar analyses may allow for previously unknown individuals to be aged, which may be valuable when correcting pedigrees. Rapidly evolving microbes can help uncover patterns not evident in host genomes, especially when the host has low genetic diversity due to recent population bottlenecks (Wirth et al. 2005). For example, genetic investigation of feline immunodeficiency virus revealed previously uncharacterised population structure and demography in cougars (Puma concolor; Biek et al. 2006). Similarly, extension of current genomic analyses for demographic inference to microbiomes could increase temporal resolution to conservation-relevant time scales. Finally, monitoring the microbiome of individuals released from captivity could be used as a measure of translocation success as the microbiome shifts towards that seen in wild conspecifics (e.g. Tasmanian devils, Chong et al. 2019). Despite microbiome analyses for conservation being a relatively new field of interest, these methods have already been applied to threatened Aotearoa New Zealand species, including the critically endangered kākāpō (Waite et al. 2014; Perry et al. 2017). Such exploratory microbiome studies represent the first steps towards actively incorporating ‘microbial rescue’ into conservation strategies (West et al. 2019; Mueller et al. 2020).
2.8.1 Methodological considerations associated with microbiome analysis

Careful sampling design is essential when undertaking microbiome studies to inform conservation (Knight et al. 2018). Considerations should include which tissue types the microbiome will be sampled across (e.g. gut, faeces, skin, oral), and whether sampling will be sufficient to control for individual age and sex, seasonal differences and other confounding effects. Metadata should be collected on all of the above and any other factors that could influence microbiome composition, as microbiome data is only as useful as the metadata that accompanies it (Goodrich et al. 2014; Knight et al. 2018).

Following the design of the microbiome sampling experiment, sample collection can commence. Because microorganisms are present in most environments, sampling blanks must also be collected at various stages to control for the presence of background environmental or laboratory contamination (Knight et al. 2018; Karstens et al. 2019). Correct sample storage is also key to ensuring the extraction of ultra-high-quality DNA (as per the requirements for reference genome sequencing). Some microorganisms are resistant to standard DNA extraction techniques and may require specialised protocols. It is good practice to also extract DNA from a ‘mock community’ with a known species composition to evaluate biases in the quantities of DNA extracted from different microorganism species.

Finally, the profiling approaches to be taken after extracting microbiome DNA must be determined. The most common current approach is to target small nuclear regions of genomes using PCR amplification so that microbial species (e.g. bacteria, fungi) can be identified (Knight et al. 2018). The advantage of a PCR-based approach is that multiple samples can be included in a single sequencing run, reducing costs. However, this may be dependent on PCR primer choice, as not all species will be amplified equally, and so the results may not reflect the extracted DNA. In addition, PCR-based approaches can amplify background contamination (Karstens et al. 2019). An alternative is a more expensive metagenomic resequencing approach (similar to that mentioned in 2.4.2 Population-level resequencing) to sequence the genomes of the microorganisms present in the sample. This approach can also be used to identify the entire microbial community and characterise adaptive variation present in the sequenced DNA.

Additional challenges of microbiome analysis include preventing DNA of the host organism from overwhelming microbial DNA during sequencing (Knight et al. 2018). As downstream analyses generally include comparing microbial community richness and composition between groups of interest (e.g. captive versus wild animals), microbiome research is limited by the availability of microbial sequences in reference databases.
3. Future conservation genetic/genomic research tools and directions

In Chapter 2 of this report we summarised the current state of play for genetics and genomics in Aotearoa New Zealand. We highlighted the dynamic nature of the presently available tools for conservation applications. In addition to a growing number of conservation genetics and genomics research projects utilising these tools, existing approaches are being combined to create new tools; for example, the combination of aDNA and eDNA to investigate ancient environmental DNA (aeDNA) is applicable to restoration ecology as eDNA binds to substrates and therefore can provide information on past species’ presence (Wilmshurst et al. 2014; Buxton et al. 2017; Hofman & Rick 2018). In addition, emerging techniques promise to expand the conservation genetic/genomic toolbox. In the following text we identify aspirational applications for existing genetic/genomic tools and for those in development and suggest potential strategies to further support this developing landscape.

The majority of tools discussed here are still in development for use in well-characterised systems (i.e. model organisms, with relevance to human health or primary industry), and their utility in a conservation context is yet to be proven. It would be disingenuous to overpromise on the conservation applications of these emerging tools, and so we have taken a cautious approach throughout to describe both the opportunities and challenges of these tools for conservation in an Aotearoa-specific context. Nevertheless, these developments all contribute to building the foundations for a bright future in the genetics/genomics research space with strong relevance for conservation.

3.1 Pangeneomes and genome graphs

In the future, we anticipate that genomics will be applied more broadly to capture population- or species-level diversity with pangenomes. Pangeneomes incorporate multiple individual reference genomes to capture the entire complement of diversity within a species, enabling researchers to differentiate between ‘core’ (genes and gene regions fixed in all individuals) and ‘accessory’ (genes and gene regions that are variable) genomic regions (Tettelin et al. 2005). Pangeneomes can be represented by genome graphs, allowing visualisation of multiple genomes simultaneously by representing core genomic regions as a single structure and accessory regions as alternative ‘paths’ for read alignment (summarised by Ameur 2019).

The application of pangeneomes and genome graphs promises to be a significant advancement in the field of conservation genomics, as highly complex and rare traits may be characterised in individuals of interest (Gao et al. 2019; Bayer et al. 2020), including those that hinder species recovery.

3.2 Chromosomics and structural variants

Chromosomics integrate cytogenetics and whole-genome sequencing to study chromosomal diversity (Potter & Deakin 2018; Deakin et al. 2019). This discipline goes beyond characterising (SNPs) to include analysis of structural variants (SVs). SVs represent a source of genomic variation likely to have large phenotype effects (e.g. Dorant et al. 2020; Huang et al. 2020a; Derežanin et al. 2022): SVs are large rearrangements (> 50 bp) within the genome that impact the form and structure of chromosomes, and include regions of the genome that are inverted, translocated from one location to another, inserted, or even lost entirely. Recent genomic research from human clinical studies and primary industry indicates that SVs are a significant source of genomic variation, as they have been found to intersect with gene coding regions more often than SNPs and impact a greater proportion of the genome overall.
(Chiang et al. 2017; Catanach et al. 2019). As such, the high likelihood that complex traits (e.g. reproductive traits in birds; Huynh et al. 2011; Küpper et al. 2016; Kim et al. 2017; Knief et al. 2017) are determined by SVs make them of interest for conservation. The large and complex nature of SVs cannot be adequately captured by short-read sequencing (see 2.4 Whole-genome sequencing), and so approaches that combine cytogenetics and long-read sequencing are currently being developed to reliably characterise SVs in threatened species. Moving forward, we anticipate that combining pangeneomic and transcriptomic approaches will facilitate research into adaptive variation (see 3.3 Characterising adaptive variation; Alongs et al. 2020; Bayer et al. 2020; Golicz et al. 2020; Liu et al. 2020), with potential applications including modelling the ability of species to adapt to future climate change or novel diseases. Collaboration with primary industry researchers has proven fruitful for conservation genomic research, due to the overlap in research interests pertaining to small populations (Galla et al. 2016). As much of the work regarding pangeneomes and SVs to date has focused on crop species, continued collaborations with primary industry researchers will further enhance conservation research in these areas.

3.3 Characterising adaptive variation

Traditional population genetics approaches investigate neutral variation to answer demographic questions (e.g. population size, inbreeding, connectivity). However, the need to conserve and monitor adaptive or maladaptive genetic diversity is increasingly emphasised (Hoelzel et al. 2019; Mable 2019; Teixeira & Huber 2021). Adaptive variation refers to the genetic basis of phenotypic variation, i.e. traits that influence survival and reproductive fitness of individuals. Maladaptive variation refers specifically to genetic variation that reduces survival and/or reproductive fitness of individuals. While efforts have been made to characterise adaptive variation in model organisms, the broad application of these methods for species of conservation concern remains limited by challenges including small sample sizes inherent to threatened populations, limited availability of genomic resources (i.e. reference genomes and annotations) for non-model species, and the complexities of polygenic traits, i.e. those determined by combinations of multiple loci (Fig. 6 but also see Batley et al. 2019; Brandies et al. 2020).

For example, SVs (see 3.2 Chromosomes and structural variants) are increasingly recognised as a pervasive source of adaptive and maladaptive variation in threatened species (e.g. Wold et al. 2021). Despite their significance, SVs remain poorly understood and characterised for most species of conservation concern (but see Cayuela et al. 2021). Even as the field of chromosomics develops, however, the resources needed to reliably characterise adaptive or maladaptive SVs – including cyrogenetic and long-read sequence data, well-characterised measures of fitness and selection pressures and a high-quality, annotated reference genome – are likely to remain challenging to produce for many threatened species.

In the face of such challenges, we prefer to avoid overpromising on the role of adaptive variation in conservation management in the near future. However, we are optimistic that creative approaches to leveraging available resources (e.g. Cayuela et al. 2020; Dorant et al. 2020; Huang et al. 2020a) will enable researchers and practitioners to make informed decisions for enhancing genome-wide diversity, which necessarily includes present, and/or future, adaptive variation. In some cases, these may include management actions that target putatively adaptive or maladaptive variation.
Figure 6. Reproduced from Parker et al. (2022, preprint): a framework for assessing key criteria for characterising adaptive variation in threatened species, including whether (i) populations are sufficiently large and genetically diverse to differentiate between selection and genetic drift; (ii) differential selection pressures are well characterised; (iii) fitness measures – or suitable proxies – are well characterised; (iv) a high-quality reference genome is available; (v) population genomic data adequately captures genome-wide diversity; (vi) comprehensive sampling is representative of relevant locally adapted populations. The further each coloured section extends toward the dark green circle reflects how well that criterion is met. Overall image design after Suding et al. (2015).

For example, in a primary production context – where specific traits are targeted in breeding programmes – pedigrees are used to inform trait mapping approaches to identify quantitative trait loci (QTLs), or genomic regions associated with specific phenotypic traits. Analytical advancements now facilitate statistical tests to characterise such loci (e.g. genome-wide association studies (GWAS); Mable 2019). Such pedigree-informed approaches, combined with comprehensive long-term datasets capturing fitness-related traits, can be used to inform conservation management (Galla et al. 2022), including the identification of populations that have inadvertently become selected to captivity (Grueber et al. 2017). In the future, these approaches could be extended to include selective breeding, but extensive scholarship around social license, ethical and cultural perspectives is needed before moving into practical considerations. Further, a mechanistic understanding of how genetic diversity influences phenotype generally requires ‘omics’ data to understand how DNA methylation, gene expression and protein composition influence phenotype (Mable 2019). However, for the purposes of understanding the response of species of conservation concern to specific challenges (e.g. climate, disease), a statistical association between genetic and phenotypic evidence may be sufficient to guide management decisions (e.g. reintroduction of disease-resistant individuals to areas where disease has previously extirpated the species; Epstein et al. 2016; Hubert et al. 2018).
3.4 Hologenomics

An extension of microbiomics relevant to conservation biology is **hologenomics** (Carthey et al. 2020). As evolution acts on both the host organism and its microbiome (the combined assemblage of which is termed the ‘holobiont’), this causes changes in the ‘hologenome’, or genomic content of the entire holobiont (Morris 2018). Thus, hologenomics goes beyond analysing genome and microbiome data separately (Rosenberg & Zilber-Rosenberg 2018). Greater variation within the hologenome allows species to more rapidly adapt to local environments than can be achieved through genomic changes alone (Rosenberg & Zilber-Rosenberg 2018). Hologenomics could allow conservation-relevant issues such as disease susceptibility/resistance (Postler & Ghosh 2017) and population sizes/connectivity (Wirth et al. 2005) to be assessed more broadly, and with greater resolution than investigating either the host genome or microbiome alone. While the substantial analytical challenges of this approach have thus far limited research to model species (Snijders et al. 2017), techniques derived from eDNA sampling could be used to understand interactions between the host and the environment, along with the ‘microbial appropriateness’ of environments at potential translocation sites (Koskella & Bergelson 2020). Similarly, metatranscriptomics approaches can be used to characterise gene regulation of the microbiome community and its potential application to host health and local adaptation (Knight et al. 2018).

3.5 Genome editing

Genome editing is the modification of an organism’s DNA by adding, removing or otherwise altering genetic material to produce targeted effects (Jinek et al. 2012; Mali et al. 2013). The development of the CRISPR-Cas9 genome engineering system is particularly relevant to conservation, as it provides a rapid, accurate and efficient method for producing such targeted changes (Doudna & Charpentier 2014). The primary biologically-feasible conservation application of genome editing in Aotearoa New Zealand is for pest control (Campbell et al. 2015a; Dearden et al. 2018). While this application represents a departure from that of other tools reviewed here that are implemented in a species conservation manner, we include genome editing, as its use in pest control is both topical and, if implemented successfully, would shift the focus of conservation actions in Aotearoa New Zealand.

Invasive species management currently employs direct culling, trapping and poisoning. Ecological variation, off-target impacts and ethical concerns create challenges for existing invasive species management strategies and have prompted research into alternative pest control solutions (Russell 2014; Latham et al. 2015; Kirk et al. 2020; MacDonald et al. 2020). Gene drives present one such method of non-lethal pest control, utilising the CRISPR-Cas9 genome editing system to target pest fertility (Esvelt et al. 2014). Gene drives alter inheritance mechanisms so that all offspring inherit the gene drive, ‘driving’ the technology through the target pest population (Fig. 7). Gene drives can be used to disrupt the fertility of one sex, while the other sex continues to propagate the gene drive to subsequent generations by reproducing with wild, non-gene drive individuals (Prowse et al. 2017).
Gene drive technology is of significant international interest for potential eradication applications (Hammond et al. 2016; Hammond & Galizi 2017). Beyond eradication, gene drives have been proposed as a tool for engineering thermal tolerance to mitigate the widespread ecological impacts of climate change (Anthony et al. 2017). Other proposed conservation applications of genome editing include engineering the genomes of endangered species to build novel resistance to emerging pathological and environmental threats, or to restore lost genomic variation (Phelps et al. 2020). Its use for de-extinction (the resurrection of extinct species using exact replicas, or through incorporation of lost phenotypic traits to create functional proxies of such species; Seddon et al. 2014; IUCN/SSC 2016) of keystone historical species has also been proposed. However, substantial technical advancements are still required, along with careful consideration of potential benefits (e.g. biodiversity gain) against the risks (e.g. redirection of funding away from existing threatened species conservation programmes; Bennett et al. 2017; see 3.6 Biobanking for further discussion). Any project aiming to restore genomic diversity through genome editing techniques will require accurate characterisation of the extent of diversity lost (with challenges relating to the age and potential degradation of samples), its potential adaptive or maladaptive impacts and precise genomic location, and substantial experimental work prior to any in situ conservation application (Phelps et al. 2020).

Technical challenges associated with genome editing for conservation applications include the complexity of gene function and local adaptation processes, particularly in wild populations (Kardos & Shafer 2018). High-quality annotated reference genomes are essential for accurate and rigorous characterisation of target genes and to develop a comprehensive understanding of phenotypic expressions of genotypes in various genetic backgrounds (Johnson et al. 2016). Further, laboratory and field trials must be carefully designed to be representative of real-world impacts and ecosystem settings, as there may be differences in implementation and effects between captive laboratory populations and wild populations due to local behavioural adaptation, gene-environment interactions (e.g. epigenetic variation; Kardos & Shafer 2018) or other indirect ecological effects (Tompkins & Veltman 2006; Russell et al. 2009; Mazza et al. 2020). These data can then inform predictive models assessing the efficacy of a genome editing application in target populations under variable conditions (including environmental change and conservation management; Champer et al. 2021).

In addition, social science research is required to assess social perspectives on genome editing, particularly those underpinning Indigenous values (Hudson et al. 2019; Palmer et al. 2020; Palmer et al. 2021). Such considerations are crucial, as the fundamental impacts of genome editing concern inheritance, therefore strongly implicating whakapapa, and genome editing may have implications for the mauri of the species. For example, kiore (Pacific rat, *Rattus exulans*) is an introduced species that is considered taonga (McClelland 2002). Furthermore, invasive species or conservation management has direct relevance to mana whenua as kaitiaki (guardians) of taonga species that may benefit from these measures. Iterative engagement regarding the potential...
uses of genome editing technologies, both for pest control and to enhance biodiversity, will be required to encompass the broad range of values and perspectives (Hudson et al. 2019).

Alongside the technical and social challenges, legislative challenges have been exacerbated by rapid technological advances (Royal Society Te Apārangi 2019). The current legal definition of genome editing in Aotearoa New Zealand limits the potential for research and funding. Consequently, data evaluated in an international context may be misinterpreted in the absence of local knowledge and may not adequately capture mana whenua values in data use and applications. Despite real-world application of genome editing technologies in Aotearoa New Zealand being unlikely in the near future, the technology is advancing rapidly and rigorous transdisciplinary evaluation to integrate local social and cultural values is required if this tool is to support ambitious projects such as Predator Free 2050 in Aotearoa New Zealand.

### 3.6 Biobanking

Biobanking (also known as cryopreservation) is a method used to preserve DNA, gametes, embryos, somatic cells, blood or tissue samples for assisted reproduction (Leon-Quinto et al. 2009; for plants this typically involves seed banking; O’Donnell & Sharrock 2017; Walters & Pence 2020; but also see Wyse et al. 2018). Biobanks are routinely used in assisted reproduction for agricultural breeding programmes, and can act as repositories of samples for future genetic/genomic research. They have been touted for their potential conservation application as, combined with genome editing or cloning techniques, samples stored in biobanks could be used to enhance genetic diversity of remnant populations to increase population sizes or to resurrect extinct species (Strand et al. 2020). Biobanked samples were recently used to produce a clone of a deceased black-footed ferret (*Mustela nigripes*) with the goal of reintroducing lost genetic diversity into this endangered species (https://reviverrestore.org/projects/black-footed-ferret/; see Sandler et al. 2021 for discussion relating to the associated ethical considerations).

Various biobanks exist overseas (e.g. the San Diego Frozen Zoo® https://science.sandiegozoo.org/resources/frozen-zoo; the Frozen Ark https://www.frozenark.org/; CryoArks https://www.cryoarks.org/), and a local project, Tāpui Aotearoa (https://www.nextfoundation.org.nz/investment/tapui-aotearoa/) aims to initiate discussion around potential establishment of a biobank for endemic fauna in Aotearoa New Zealand. We foresee that with applications similar to those of genome editing and a focus on threatened taonga species, such discussions must emphasise data sovereignty and accurate collection and storage of metadata alongside biobanked materials to adequately capture local contexts, knowledge and protect mana whenua interests (see sections 1.4.2 Metadata collection and management and 1.4.3 Data sovereignty).

Advancements in cryopreservation techniques and expansion of biobanking facilities may promote de-extinction or other cloning-based techniques aimed at increasing population size. There has been extensive discourse around potential risks, benefits, and conservation relevance of de-extinction (Bennett et al. 2017; Iacona et al. 2017; Sandler 2017; Valdez et al. 2019; Genovesi & Simberloff 2020), including scholarship specific to the Aotearoa New Zealand context (Taylor et al. 2017b) and from a conservation genetics perspective (Steeves et al. 2017). Even if the goals associated with Predator Free 2050 are achieved and sufficiently large areas of high-quality habitat were to become available to support the resurrection of a species, many of the ethical, legal, financial and technical challenges associated with genome editing remain applicable to both de-extinction and cloning more broadly. While we consider any such applications to be in the distant future in Aotearoa New Zealand, associated research developments overseas may contribute to conservation of extant species.
3.7 Multi-species genomic pipelines

The global disconnect between conservation research and practice is widely recognised (also known as the research-implementation gap; Jarvis et al. 2020; Kadykalo et al. 2021). We envision an increasingly collaborative interface, where researchers and practitioners co-develop targeted research to inform conservation management actions (e.g. Tasmanian devil conservation, Hogg et al. 2017; conservation efforts for a range of threatened birds in Aotearoa New Zealand, Galla et al. 2022). Establishment of broad-scale genomic data generation and analysis pipelines may represent one pathway to enhance such collaborations. These initiatives are established to build repositories of genomic data/resources for multiple threatened species management programmes, provide analysis tools for use by researchers and practitioners and result in the accelerated implementation of genomics-informed conservation management actions across multiple species/ecosystems simultaneously. For example, the Australian Threatened Species Initiative (TSI) was initiated to support conservation decision making by establishing a national repository of genomic data accessible via an online web tool for analysis (Hogg et al. 2022), while the California Conservation Genomics Project (CCGP) was designed as a state-wide multi-species landscape genomics project to facilitate conservation prioritisation across California (Shaffer et al. 2022). In Aotearoa New Zealand, independent analysis pipelines are becoming available (e.g. those of Genomics for Aotearoa New Zealand (GFANZ), https://genomics.nz/page/bioinformatics-platform and AgResearch, https://www.agresearch.co.nz/doing-business/products-and-services/bioinformatics/).

However, there are various limitations associated with the development and implementation of such pipelines in Aotearoa New Zealand. Despite the high proportion of threatened species in Aotearoa New Zealand (Bradshaw et al. 2010) there are relatively few threatened species management programmes, and so the application of a large-scale multi-species pipeline for conservation is somewhat limited. Development and optimisation of a pipeline similar to the TSI or CCGP would require substantial resourcing (e.g. US$12 million funding across projects involving 235 species for the CCGP, AUS$1.4 million funding for projects involving 61 species for the TSI; Hogg et al. 2022; Shaffer et al. 2022). Multi-species pipelines require standardised data inputs to facilitate robust, comparable analysis (e.g. sequencing types and depth for assembling reference genomes and for population-level resequencing, and reference genome assembly methods and quality; see 2.4.1 Reference genomes, 2.3 Reduced-representation sequencing, 2.4.2 Population-level resequencing, and the technical standards described in Shaffer et al. 2022), which may limit the re-use of existing genomic resources. Data sovereignty concerns will need to be addressed to determine the accessibility of any such data, particularly as the computational scale and data storage requirements for multi-species analyses may grow to exceed those available via existing platforms in Aotearoa New Zealand (e.g. NeSI, Catalyst Cloud). There is also the potential for data to become divorced from the local context, so early stages of pipeline development must emphasise the need for collecting and using metadata to support interrogation of input data and accurate interpretation of results. To maximise pipeline efficiency, analyses are likely to be limited to characterisation of genome-wide diversity and population structure as opposed to assessment of high-complexity adaptive variation or SVs with associated impacts on population fitness (e.g. Hogg et al. 2022; Shaffer et al. 2022; also see 3.2 Chromosomes and structural variants). Even within primary industry, where species biology is well-understood and genomic resources are well-characterised, few broad-scale pipelines exist (but see research across Eucalyptus spp., Silva-Junior et al. 2015).

While the TSI aims to produce an online tool to empower conservation practitioners to analyse genomic data as needed, we recognise that practitioners are already time-poor (Taylor et al. 2017). While we do not presume to speak for all practitioners, many of whom may be interested in broadening their skills and having greater hands-on input in the genomics space, this may not be widely practicable/feasible. If such a pipeline were to be developed for conservation in Aotearoa New Zealand, ensuring co-development by researchers and practitioners to best
meet the needs and capabilities of both parties could fall under the purview of a conservation genetics/genomics advisory group (see 3.8 Improved facilitation of genetics/genomics research to enhance conservation outcomes).

3.8 Improved facilitation of genetics/genomics research to enhance conservation outcomes

For the genetics/genomics toolbox to effectively inform conservation management decisions, particularly in light of rapid technological developments, we advocate for a well-resourced DOC genetics/genomics advisory group. This group should comprise both internal and external practitioners and researchers, across a range of career stages, and include members of, or exist in partnership with, Kahui Kaupapa Atawhai. Such a group should centre the principles of Te Tiriti o Waitangi to facilitate the establishment of genetic/genomic research by advising on the most appropriate research strategies to meet conservation needs, while balancing current and future uses of samples and data against feasibility and costs. This group could also be tasked with developing data management guidelines encompassing sample collection and curation, management of data and metadata to support downstream research, and facilitating collaborations with institutes with the necessary skill, experience and resources to implement specific research. Further, the establishment of such a group may facilitate connections among practitioners, researchers, local mana whenua and species-specific research and conservation recovery groups to enhance conservation outcomes across Aotearoa New Zealand. However, we stress that researchers and practitioners must be proactive in establishing and maintaining trusted relationships, which may then lead to larger and/or long-term collaborative research projects. As predominantly early career researchers, we argue that senior researchers are best placed to maintain a clear and consistent line of communication between practitioners, mana whenua and their own research groups. In addition to better facilitating the permitting process, doing so will create the opportunities for early career researchers to gain an understanding of research best-practice within the limited time frames available to them (e.g. three years of funding for a PhD).

Related to this, a major concern identified by early career researchers and worth highlighting here relates to DOC’s research permitting processes which are subject to lengthy delays and inconsistencies among taxa, type of data generated and messaging around data management. With samples and data persisting beyond the timeframes of permits, and the increasing inter-generationality of research projects, further concerns arise from the absence of periodic reviews once permits are granted. Addressing potential solutions to these issues is beyond the scope of this report, but could fall under the mandate of a DOC genetics/genomics advisory group.

Given these challenges and opportunities, we foresee conservation genetic/genomic research becoming increasingly transdisciplinary in nature. Indeed, to achieve the vision outlined in Te Mana o te Taiao, both species-specific and ecosystem-based approaches will need to include genetic/genomic, microbial, ecological, physiological and environmental data, alongside mātauranga Māori where mana whenua wish to contribute this. Thus, to produce more ‘winners’ – including species, ecosystems, and people – benefitting from conservation actions (Nelson et al. 2019), we encourage researchers and practitioners to focus on developing collaborative and iterative communicative practices that incorporate a wide array of disciplines and perspectives.
4. Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3C sequencing</strong></td>
<td>Chromosome conformation capture techniques such as Hi-C, Omni-C and Pore-C used to characterise the spatial structure of DNA to identify genomic interactions.</td>
</tr>
<tr>
<td><strong>adaptive potential</strong></td>
<td>The ability of a population to adapt to immediate environmental change, typically estimated by the extent of genomic diversity present and quantified by relative reproductive fitness. Contributes to overall evolutionary potential, which refers to a species’ capacity to respond to environmental change through time.</td>
</tr>
<tr>
<td><strong>adaptive (maladaptive) variation</strong></td>
<td>Genomic variation that results in phenotypic variation and has some effect on individual fitness.</td>
</tr>
<tr>
<td><strong>ancient DNA (aDNA)</strong></td>
<td>DNA extracted from historical museum skins, subfossils or fossils.</td>
</tr>
<tr>
<td><strong>base pairs (bp)</strong></td>
<td>Individual nucleotides that code the DNA.</td>
</tr>
<tr>
<td><strong>chromosomics</strong></td>
<td>A discipline that integrates cytogenetics and whole-genome sequencing to study chromosome-level diversity.</td>
</tr>
<tr>
<td><strong>coverage</strong></td>
<td>Can have two distinct meanings: 1) when used in a reference genome context, it can relate to the proportion of the genome represented by the reference; and 2) when used in a population-level sequencing context, it refers to the amount of sequencing depth supporting a variant call.</td>
</tr>
<tr>
<td><strong>cytogenetics</strong></td>
<td>The study of the form and structure of DNA within the nucleus of a cell.</td>
</tr>
<tr>
<td><strong>depth</strong></td>
<td>The number of times the genome is sequenced.</td>
</tr>
<tr>
<td><strong>environmental DNA (eDNA)</strong></td>
<td>DNA extracted from environmental samples such as water, soil or air.</td>
</tr>
<tr>
<td><strong>Gb</strong></td>
<td>Gigabase pairs; one billion base pairs of DNA.</td>
</tr>
<tr>
<td><strong>genetic data</strong></td>
<td>Data representative of a subset of the genome, typically comprising tens of loci.</td>
</tr>
<tr>
<td><strong>genetic drift</strong></td>
<td>The fluctuation in allele frequencies between generations due to stochastic processes.</td>
</tr>
<tr>
<td><strong>genome</strong></td>
<td>The full complement of DNA characterising an individual.</td>
</tr>
<tr>
<td><strong>genomic data</strong></td>
<td>Data representative of the genome, comprising hundreds to millions of loci. These data are generated with high-throughput sequencing techniques.</td>
</tr>
<tr>
<td><strong>genome graph</strong></td>
<td>Graph structure used to represent genomic variation detected among multiple individuals simultaneously.</td>
</tr>
<tr>
<td><strong>hologenomics</strong></td>
<td>The study of the genomic interactions between a host organism and its microbiome.</td>
</tr>
<tr>
<td><strong>iterative engagement</strong></td>
<td>A process of continuous or sequential discussions between parties throughout the research process from initial planning through to final public dissemination of results. For examples of such approaches, see Cisternas et al. 2019; Rayne et al. 2022.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>long-read sequencing</td>
<td>DNA sequencing using platforms such as Oxford Nanopore Technologies or PacBio SMRT that can produce sequence reads in excess of 10,000 bp in length.</td>
</tr>
<tr>
<td>microarray</td>
<td>A genomic method used to genotype large numbers of loci at population-scale, simultaneously.</td>
</tr>
<tr>
<td>microbiome</td>
<td>The microorganisms that reside on and/or within the tissues of a host species, including bacteria, fungi and viruses.</td>
</tr>
<tr>
<td>mitochondrial DNA (mtDNA)</td>
<td>The DNA specific to the mitochondrial organelle, with a short circular structure and high copy number within an individual cell.</td>
</tr>
<tr>
<td>neutral variation</td>
<td>Genomic variation that does not impact fitness. Also known as neutral diversity.</td>
</tr>
<tr>
<td>pangenome(s)</td>
<td>Multiple high-quality genome assemblies that capture all of the genomic diversity within a species. Pangenomes may eventually supersede single individual genomes for reference purposes.</td>
</tr>
<tr>
<td>restriction-site associated DNA sequencing (RAD-seq)</td>
<td>A method using restriction enzymes to target subsets of loci throughout the genomes of all sequenced individuals.</td>
</tr>
<tr>
<td>reduced-representation sequencing (RRS)</td>
<td>Sequencing methods such as restriction-site associated methods (RAD-seq, ddRAD-seq, GBS), where a reduced subset of the genome is sequenced and assumed to be representative of the diversity throughout the complete genome.</td>
</tr>
<tr>
<td>reference genome(s)</td>
<td>A representation of the genome of a species that can be used alone for interspecific comparisons or as a reference against which population-level resequencing or RRS data can be aligned for intraspecific comparisons.</td>
</tr>
<tr>
<td>reproductive fitness</td>
<td>Capacity of individuals in a population to propagate their genes to subsequent generations. Quantified through estimates of relative fertility and mortality.</td>
</tr>
<tr>
<td>short-read sequencing</td>
<td>DNA sequencing conducted using platforms such as Illumina MiSeq, HiSeq and NovaSeq that produce short (&lt; 500 bp) sequence reads.</td>
</tr>
<tr>
<td>single-nucleotide polymorphisms (SNPs)</td>
<td>The most common form of variation in the genome. SNPs have low mutation rates and are often biallelic, with known characteristics making analysis relatively straightforward.</td>
</tr>
<tr>
<td>SNP-chip</td>
<td>A type of genomic microarray used to generate single-nucleotide polymorphism data for large numbers of individuals within a species.</td>
</tr>
<tr>
<td>structural variants (SVs)</td>
<td>A diverse class of genomic variation impacting the form and structure of chromosomes. Common types of structural variants include copy number variants, deletions, duplications, insertions, inversions and translocations &gt; 50 bp.</td>
</tr>
<tr>
<td>transcriptome</td>
<td>The array of RNA transcripts that are expressed by an organism, determining individual phenotype.</td>
</tr>
</tbody>
</table>
5. Acknowledgements

We wish to thank Sharon Alderson and Richard Maloney at the Department of Conservation for the invitation to contribute this review. We thank Rod Hitchmough and Bruce McKinlay and an anonymous external reviewer for their helpful feedback.

NJF is grateful to the early-career researchers involved here for contributing their expertise, to TES, Thomas Buckley, Manaaki Whenua Landcare Research, Genomics Aotearoa, the University of Canterbury’s (UC) Conservation, Systematics and Evolution Research Team (ConSERT), the University of Otago’s (UO) Knapp/Matisoo-Smith lab group, and the Kakī Recovery Programme for their support, and to Julia Allwood and Talia Brav-Cubitt for helpful discussions relating to topics included here.

CIMA wishes to thank the Knapp Lab and Waters Lab, the UO Departments of Zoology, Anatomy, and Marine Science, and iwi partnerships for their support.

AA wishes to thank the Gemmell Lab, the UO Department of Anatomy, Genomics Aotearoa, Predator Free 2050, Phil Wilcox, Ben Te Aika, Ian Shapcott, and iwi/hapū collaborative partnerships for their support.

ACC would like to thank Neil Gemmell and the Gemmell lab, project collaborators, the UO Department of Anatomy, members and supporters of SING (Summer internship for Indigenous peoples in Genomics), the Ministry for Pacific Peoples, and BLAKE New Zealand for their continued support and contribution.

IC is grateful to all research partners, all past and present members of the UC’s ConSERT team, the Shore Plover Recovery Group, the Isaacs Conservation and Wildlife Trust, especially Anne Richardson and Leigh Percasky and Rose Collen for their amazing support and insight.

ED thanks the Gemmell lab, the UO Department of Anatomy and the Marsden Fund for their support.

LD thanks Anna Santure for her support and for helpful insights relating to topics included here.

SJG thanks her research partners (Kakī Recovery Programme, Kākāriki Karaka Recovery Programme), UC’s ConSERT members, her PhD committee members (TES and Anna Santure), and chair of her PhD oral exam (Carolyn Hogg) for their mentorship regarding pedigrees and conservation breeding programmes. SJG is grateful to UC and Boise State University for their support.

LH would like to thank Sarah Flanagan and Michelle LaRue, as well as Christchurch City Council and Gateway Antarctica for their support.

MM wishes to thank the UC’s ConSERT and Fulbright New Zealand for their support.

AR is grateful to her research partners, UC’s ConSERT and FERG (Freshwater Ecology Research Group) members, New Zealand’s Biological Heritage National Science Challenge and the University of Canterbury for their support.

TES extends her gratitude for the early career researchers across Aotearoa New Zealand who so graciously shared their enthusiasm and expertise in the development of this contribution. She is also eager to acknowledge past, present and future members of UC’s ConSERT, and the team’s many research partners, who work tirelessly to enhance the recovery of some of Aotearoa New Zealand’s most threatened taonga species.
6. References


DOC (Department of Conservation) 2020: *Te Mana o Te Taiao: Aotearoa New Zealand Biodiversity Strategy 2020*.


Dussex, N.; Rawlence, N.J.; Robertson, B.C. 2015: Ancient and contemporary DNA reveal a pre-human decline but no population bottleneck associated with recent human persecution in the kea (*Nestor notabilis*). *PLoS ONE* 10(2): e0138966. https://doi.org/10.1371/journal.pone.0138966


Faust, L.J.; Bergstrom, Y.; Thompson, S.; Bier, L. 2019: *PopLink* (2.5) [Computer software]. Lincoln Park Zoo.


Grueber, C.E.; Reid-Wainscoat, E.E.; Fox, S.; Belov, K.; Shier, D.M.; Hogg, C.J.; Pemberton, D. 2017: Increasing generations in captivity is associated with increased vulnerability of Tasmanian devils to vehicle strike following release to the wild. *Scientific Reports* 7(1): 2161. [https://doi.org/10.1038/s41598-017-02273-3](https://doi.org/10.1038/s41598-017-02273-3)


Hedrick, P.W. 2001: Conservation genetics: where are we now? *Trends in Ecology & Evolution* 16(11): 629–636. [https://doi.org/10.1016/S0169-5347(00)02282-0](https://doi.org/10.1016/S0169-5347(00)02282-0)


Forsdick et al. 2022–Genetic/genomic data for conservation in Aotearoa New Zealand


McLaren, M.R.; Callahan, B.J. 2020: Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1808), 20190592. https://doi.org/10.1098/rstb.2019.0592


Sandler, R. 2017: De-extinction: costs, benefits and ethics. *Nature Ecology & Evolution* 1(4): 1–2. [https://doi.org/10.1038/s41559-017-0105](https://doi.org/10.1038/s41559-017-0105)


Scott, A.D.; Zimin, A.V.; Puiu, D.; Workman, R.; Britton, M.; Zaman, S.; Caballero, M.; Read, A.C.; Bogdanove, A.J.; Burns, E.; Neale, D.B. 2020: A reference genome sequence for giant sequoia. *G3: Genes, Genomes, Genetics* 10(11): 3907–3919. [https://doi.org/10.1534/g3.120.401612](https://doi.org/10.1534/g3.120.401612)


Seebacher, F.; White, C.R.; Franklin, C.E. 2015: Physiological plasticity increases resilience of ectothermic animals to climate change. *Nature Climate Change* 5(8): 61–66. [https://doi.org/10.1038/nclimate2457](https://doi.org/10.1038/nclimate2457)


