







# Threatened Species Initiative: Empowering conservation action using genomic resources

Carolyn J. Hogg<sup>a,1</sup> , Kym Ottewell<sup>b</sup> , Peter Latch<sup>c</sup>, Maurizio Rossetto<sup>d</sup> , James Biggs<sup>e</sup>, Andrew Gilbert<sup>f</sup> , Sarah Richmond<sup>f</sup>, and Katherine Belov<sup>a</sup>

Edited by Harris Lewin, Evolution and Ecology and The Genome Center, University of California, Davis, CA; received September 2, 2021; accepted October 25, 2021

Globally, 15,521 animal species are listed as threatened by the International Union for the Conservation of Nature, and of these less than 3% have genomic resources that can inform conservation management. To combat this, global genome initiatives are developing genomic resources, yet production of a reference genome alone does not conserve a species. The reference genome allows us to develop a suite of tools to understand both genome-wide and functional diversity within and between species. Conservation practitioners can use these tools to inform their decision-making. But, at present there is an implementation gap between the release of genome information and the use of genomic data in applied conservation by conservation practitioners. In May 2020, we launched the Threatened Species Initiative and brought a consortium of genome biologists, population biologists, bioinformaticians, population geneticists, and ecologists together with conservation agencies across Australia, including government, zoos, and nongovernment organizations. Our objective is to create a foundation of genomic data to advance our understanding of key Australian threatened species, and ultimately empower conservation practitioners to access and apply genomic data to their decision-making processes through a web-based portal. Currently, we are developing genomic resources for 61 threatened species from a range of taxa, across Australia, with more than 130 collaborators from government, academia, and conservation organizations. Developed in direct consultation with government threatened-species managers and other conservation practitioners, herein we present our framework for meeting their needs and our systematic approach to integrating genomics into threatened species recovery.

endangered | genomes | reduced representation sequencing | applied conservation | management

An estimated 37,470 animal, plant, and fungi species are now listed as threatened (vulnerable, endangered, critically endangered) by the International Union for the Conservation of Nature (IUCN) Red List (downloaded August 2021) with most known species (72%) still to be assessed (1). Species listing on the IUCN Red List is rigorous, with multiple assessments, reviews, and consistency checks to ensure robustness of the global list (1). However, global biodiversity is not evenly spread across the globe, with just 17 megadiverse countries home to 60 to 80% of all life on earth (2). As a result, the responsibility of conserving much of the world's

biodiversity tends to fall upon these few nations, 15 of which are classified as “developing economies” by the United Nations (3). The range of threats contributing to the global biodiversity crisis (4) are broad, including habitat loss and fragmentation, invasive pest species, disease, and climate change (5). As the human population continues to increase and encroach on the natural world, a 10-year program has commenced (6)—“The United Nations Decade of Ecosystem Restoration 2021–2030”—to help slow biodiversity loss. Fragmentation and modification of habitat reduces population size and connectivity for many species and threatened species

<sup>a</sup>School of Life & Environmental Science, University of Sydney, Sydney, NSW 2006, Australia; <sup>b</sup>Conservation Science Centre, Department of Biodiversity, Conservation, & Attractions, Kensington, WA 6151, Australia; <sup>c</sup>Australian Government Department of Agriculture, Water & Environment, Canberra, ACT 2600, Australia; <sup>d</sup>Research Centre for Ecosystem Resilience, Australian Institute of Botanical Science, The Royal Botanic Garden Sydney, Sydney, NSW 2000, Australia; <sup>e</sup>Zoo and Aquarium Association Australasia, Mosman, NSW 2088, Australia; and <sup>f</sup>Bioplatforms Australia, North Ryde, NSW 2113, Australia

Author contributions: C.J.H., K.O., and P.L. designed research; C.J.H., M.R., J.B., A.G., S.R., and K.B. performed research; C.J.H. and S.R. analyzed data; and C.J.H., K.O., P.L., M.R., J.B., A.G., S.R., and K.B. wrote the paper.

Competing interest statement: All authors are members of the Threatened Species Initiative Steering Committee.

This article is a PNAS Direct Submission.

This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: Carolyn.hogg@sydney.edu.au.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115643118/-/DCSupplemental>.

Published January 18, 2022.

are typically found in small, isolated populations susceptible to genetic risks and other stochastic processes (7). Conservation practitioners are more frequently using conservation translocations as a restoration tool for maintaining populations of threatened fauna and flora (8, 9). Yet, translocations can further entrench small population risks because when managing a species in a fragmented landscape, behind a fence, or on an island, natural gene flow is reduced (7). As a result, genetic management is becoming integral to the conservation of an ever-greater number of species.

Genomes, and their associated downstream applications, are powerful tools for discovery of new knowledge around species behavior and biology. They can improve our understanding of species' taxonomy, provide information regarding past and future evolutionary processes, and complement current ecological survey and study methods (10). In 2018, of the ~13,500 animal species on the IUCN Red List, less than 0.8% of species had published genomes on the National Center for Biotechnology Information (11); in the past 3 y this has increased slightly to 2.4% of the 15,521 listed threatened species. Although there is an increase in global genome consortia, such as the Earth Bio-genome Project (10, 12), the Vertebrate Genome Project (13), and the Global Invertebrate Genomics Alliance (14), that are creating genomes for nonmodel species, genomic resources for some of our most critically endangered species are still lacking. Furthermore, developing reference genomes for species does not impact their conservation on their own, but rather it is the downstream applications and tools that use reference genomes that can significantly improve species' conservation.

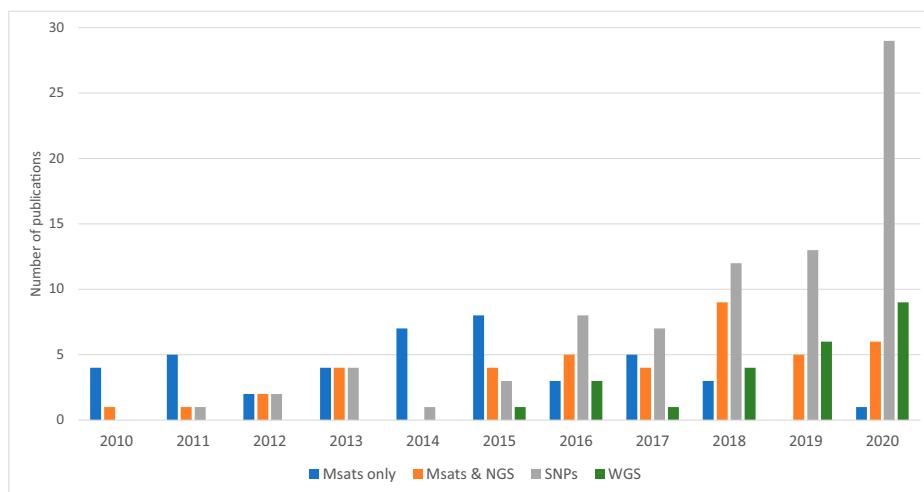
A recent review by Supple and Shapiro (15) highlighted that the "transition to genomic technologies is only just beginning" and that there needs to be an expansion in the available datasets so researchers can ask different questions applicable to conservation. Here, we reviewed the conservation-focused peer-reviewed literature to explore the trends in increasing use of genomic data in studies regarding the management of threatened or endangered species (see *SI Appendix* for methodological details). We identified a total of 498 papers containing a variety of sequencing methods and types of studies: 263 (52.8%) used either microsatellites, SNPs, or whole genome

data, to address population genetics/genomics; and 89 (17.9%) were some form of review (*SI Appendix, Table S1*). Of the 212 papers that used nuclear DNA to address population genetics/genomics, there has been a marked decrease in the use of microsatellites and an increase in the use of SNPs since 2010 (Fig. 1). As expected, with genome technologies becoming more prominent in nonmodel species after 2010, there was an increase in using next-generation sequencing to improve the development of microsatellite markers (2015–2020) and an increased use of thousands of SNPs to improve genome-wide diversity studies (Fig. 1). More recently (since 2017) there has been a steady increase in the number of studies using resequenced whole genomes (Fig. 1). Although this is not a fully comprehensive search of all the conservation genomics/genetics works currently published, we find that even in the absence of available reference genomes for threatened species, there has been a sustained uptake of other genomic approaches in conservation genetic studies of threatened species, with many leading to explicit conservation recommendations (see refs. 15–17 for more comprehensive reviews).

As Supple and Shapiro (15) (and others) point out, the suite of genomic tools available to researchers to understand both genome-wide and functional diversity within and between species and populations, can be greatly expanded when reference genome information is available, enabling more precise targeting of conservation measures (11, 15, 16). Indeed, we know that conservation practitioners use genetic information in their decision-making (*SI Appendix, Table S2*), particularly when it comes to managing threatened species in small populations within fragmented landscapes (18). However, the use of "big data" genomic approaches presents challenges for practitioners to access and interpret the available information.

### Threatened Species Initiative

Australia is one of the 17 megadiverse nations. Separating from other continents over 42 to 53 million y ago (19, 20) means many of the species in Australia are unique, with 87% of mammals, 45% birds, 93% reptiles, 94% amphibians, and 92% of plants endemic to the island continent (21). However, many Australian species have seen marked declines since European



**Fig. 1.** Number of peer-reviewed publications using nuclear DNA for threatened/endangered species conservation management using microsatellite markers (msats only), those using a combination of microsatellites and next-generation sequence (msats and NGS) data (transcriptomes, RADseq, whole-genomes, SNPs), SNPs only, and WGS between 2010 and 2020.

**Table 1. Environmental genome initiatives that have been supported by Bioplatforms Australia that have produced genomic resources for Australian wildlife and plant species**

Initiative	Start date	More information
Biome of Australia Soil Environments (BASE)	2011	<a href="https://bioplatforms.com/projects/soil-biodiversity/">https://bioplatforms.com/projects/soil-biodiversity/</a>
Great Barrier Reef	2012	<a href="https://bioplatforms.com/projects/great-barrier-reef/">https://bioplatforms.com/projects/great-barrier-reef/</a>
DNA Barcoding	2013	<a href="https://bioplatforms.com/projects/dna-barcoding/">https://bioplatforms.com/projects/dna-barcoding/</a>
Marine Microbe	2015	<a href="https://bioplatforms.com/projects/marine-microbes/">https://bioplatforms.com/projects/marine-microbes/</a>
Koala Genome	2015	<a href="https://bioplatforms.com/projects/koala-genome/">https://bioplatforms.com/projects/koala-genome/</a>
Oz Mammal Genomics	2016	<a href="https://bioplatforms.com/projects/oz-mammals-genomics/">https://bioplatforms.com/projects/oz-mammals-genomics/</a>
Australian Microbiome	2018	<a href="https://bioplatforms.com/projects/australian-microbiome/">https://bioplatforms.com/projects/australian-microbiome/</a>
Genomics for Australian Plants (GAP)	2018	<a href="https://bioplatforms.com/projects/genomics-for-australian-plants/">https://bioplatforms.com/projects/genomics-for-australian-plants/</a>
Australian Amphibian and Reptile Genomics Initiative	2020	<a href="https://bioplatforms.com/projects/reptiles-and-amphibians/">https://bioplatforms.com/projects/reptiles-and-amphibians/</a>
Threatened Species Initiative	2020	<a href="https://bioplatforms.com/projects/threatened-species/">https://bioplatforms.com/projects/threatened-species/</a>
Native Grasses	2020	<a href="https://bioplatforms.com/projects/australian-grasslands/">https://bioplatforms.com/projects/australian-grasslands/</a>

For details on these initiatives, alongside others, please see <https://bioplatforms.com/projects>.

settlement in 1788, with 1,774 species (480 animals; 1,294 plants, as of 2016) listed as threatened under the Australian Environment Protection and Biodiversity Conservation Act (22). Various recovery and other conservation plans have been put in place by the Australian, State, and Territory Governments with actions to address threats and support the long-term recovery of these species. Globally, Australia has the worst record of mammal extinctions in the world. Multiple species have faced population declines of over 90% in the past two decades (23). The loss of Australian mammal species is largely due to predation by introduced species and changes to fire regimes (23, 24), with our first mammal extinction attributed to anthropogenic climate change declared in 2016 (25). Apart from managing species in often increasingly fragmented landscapes, to address the challenges of rapidly declining populations, many threatened species are increasingly being managed in large, fenced areas, in zoological/botanic garden insurance populations, and on offshore islands. Consequently, genetic diversity and gene flow are reduced for many species and this needs to be accounted for in ongoing management actions.

Conservation biologists and practitioners have a range of technological tools at their disposal to address the various challenges of conserving biodiversity (26). However, for many conservation practitioners there is often an implementation gap between research and development of new tools and their application in conservation practice (27). One such research implementation gap that has been widely discussed is the use of genomics and associated tools for conservation of threatened species (28–30). Although recent reviews (see refs. 15 and 31–33) discuss the value of genomes for conservation and protection of biodiversity, as sequencing technology improves, there are increasing requirements around genome quality, bioinformatic knowledge, and handling of big data. This creates an ever-widening research–implementation gap between the creation of genomic resources by genome biologists and bioinformaticians and the application of these resources in conservation management by conservation practitioners.

Bioplatforms Australia (Bioplatforms), a nonprofit organization that supports Australian Life Science research by investing in state-of-the-art infrastructure and expertise in genomics, proteomics, metabolomics, and bioinformatics, has invested in a number of genome initiatives over the past 10 y, producing genomic resources for Australian species (Table 1). The focus of many of these initiatives has been on reference genome

production, comparative genomics, and phylogenomics to resolve species taxonomy for conservation application. Building on the success of these programs, the mission of the Threatened Species Initiative (TSI), launched in May 2020, is to bridge the implementation gap between the production of genomic resources and their application in conservation management (<https://threatenedspeciesinitiative.com/>). From the outset, TSI has been developed in direct consultation with governmental threatened species managers and other conservation practitioners, around their needs and knowledge gaps (*SI Appendix, Table S2*). It brings together genome biologists, population biologists, bioinformaticians, population geneticists, and ecologists with conservation agencies across Australia, including government, zoos, botanic gardens, and nongovernment organizations (NGOs). Our objective is to create a foundation of genomic data to advance our understanding of representative Australian threatened species, in addition to fast-tracking genomic information to conservation end-users through online resources and open-access data. We aim to empower conservation practitioners to leverage genomic information to tackle critical biological and conservation issues, including genetic data to inform translocations, captive breeding, seed banking, and ongoing population management.

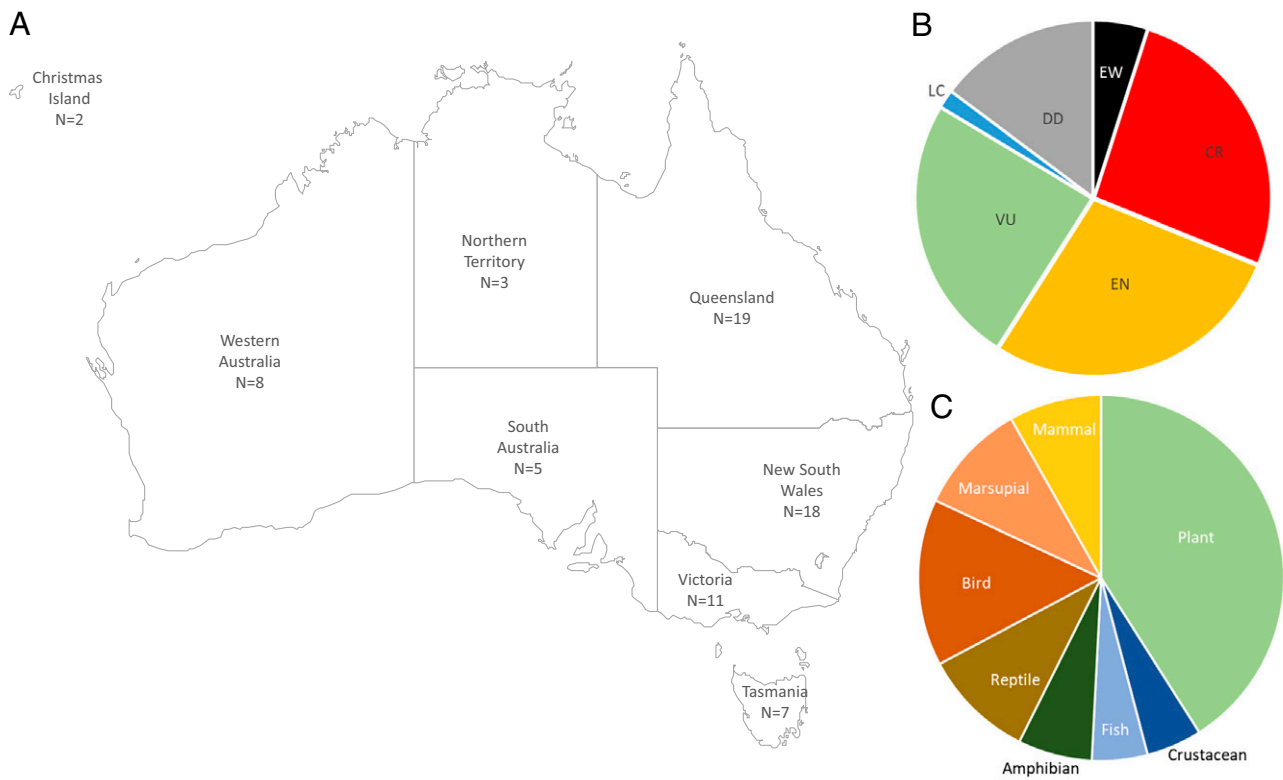
Studies from New Zealand/Aotearoa (28) and Australia (34) show that conservation practitioners know the value of using genetic data in conservation decision-making, but access to easily interpretable information is lacking. In Australia, projects such as Devil Tools & Tech (34) and Restore & Renew (35) have shown that by creating partnerships between academic researchers and conservation practitioners, the latest genome technologies and techniques can be applied in real-time to conservation decision-making. It was the success of these programs with specific species and their philosophy of open access to the latest research data that led to the development of the TSI. TSI's goal is to undertake applied research that has direct management applications, while ensuring the research is innovative and novel for peer-review publication and to attract competitive research funding.

Our approach to engineering and building a bridge for the current genomic research–implementation gap is threefold: 1) use genome sequencing technologies that meet the needs of the conservation end-users while maximizing the limited conservation resources available (both funding and sample access), so genomic data can be developed for as many threatened species as possible; 2) develop an on-line interface where TSI project

teams can obtain protocols and use a set of established bioinformatic tools and workflows to provide genetic outputs in a standardized reporting format for conservation practitioners; and 3) open-data access, where genomic data will be open access but other related metadata may be restricted due to threatened species and indigenous sensitivities (36). To ensure seamless delivery of the larger project, a pilot phase was commenced in August 2020, to test and bed down workflows and pipelines to ensure outputs were fit-for-purpose for conservation management and decision-making. Eight species (two birds: eastern bristlebird, *Dasyornis brachypterus* and orange-bellied parrot, *Neophema chrysogaster*; two marsupials: eastern barred bandicoot, *Perameles gunnii*, and western barred bandicoot, *Perameles bougainville*; two mammals: ghost bat, *Macroderma gigas* and Hastings River mouse, *Pseudomys oralis*; one fish: swan galaxias, *Galaxias fontanus*; and one plant: native guava, *Rhodomyrtus psidioides*) were selected for the pilot phase through consultation with the Australian, State, and Territory threatened species managers. Note, the Australian Amphibian and Reptile Genomics (AusARG) project commenced at the same time as the TSI and is undertaking similar activities for reptiles and amphibians, so these taxa were not included in the initial TSI pilot phase. The species were grouped into five scenarios to enable comprehensive testing of the different stages of the TSI conservation genomics pipeline: 1) the species has no reference genome, no population genetic data; 2) the species has closely related species with a reference genome, but no population genetic data; 3) the species has no reference genome, and population genetic data exists; 4) the species has a reference genome, or conspecific genome, some

population genetic data, and is subject to conservation action which mixes genetically distinct populations; and 5) the species has no reference genome, but short-read data exists, and some population genetic data exists.

This pilot phase was followed by a “Request for Partnership” round in early 2021, and with a second scheduled for early 2022. In the Request for Partnership academic researchers are encouraged to select species from a preselected list of threatened species, which has been prioritized by the Australian Federal, State, and Territory government agencies. Initially it was anticipated that the current TSI funding (AUD\$1.4M) would be able to provide genomic resources for between 40 and 50 threatened plant, animal, and invertebrate species over its 3-y lifespan. In 2021, this goal was superseded, with 61 species currently supported by the program from across Australia (Fig. 2 and *SI Appendix, Table S3*), representing “extinct in the wild” ( $n = 3$ ), “critically endangered” ( $n = 16$ ), “endangered” ( $n = 17$ ), “vulnerable” ( $n = 15$ ), and “data-deficient species” ( $n = 9$ ). Note, one “least concern” species is supported to investigate its value as a genetic rescue surrogate for a critically endangered species. Participating project teams are encouraged to leverage other funding opportunities using TSI resources as “seed” funding; this will see a multiplier effect from the base investment and provide genomic resources for more species. Of the 61 species’ projects, there are over 130 project team members representing government (46%), academia (35%), and nongovernment/conservation organizations (19%). All participating project teams are encouraged to work with local Aboriginal nations where possible and provide tangible on-ground conservation outcomes as part of their projects.



**Fig. 2. Species involved in the TSI by: (A) geographical location, noting some species are found in more than one State or Territory; (B) IUCN threat status: extinct in the wild (EW), critically endangered (CR), endangered (EN), vulnerable (VU), least concern (LC), data deficient (DD); and (C) taxa. Base Australia map by Free Vector Maps (<https://freevectormaps.com/>).**

**Table 2. Assembly features of Tasmanian devil (38), koala (39) and Hi-C scaffolded koala genomes (dnazoo.org)**

Genome assembly (publication year)	Tasmanian devil (2012)	Koala (2018)	Hi-C koala (2020)
Scaffold genome size (Gb)	3.17	3.19	3.19
Contig genome size (Gb)	2.93	3.19	3.19
No. scaffolds	35,975	1,907	1,318
No. contigs	237,292	1,907	1,935
Scaffold N50 (Mb)	1.85	11.59	480.11
Contig N50 (Mb)	20.14kb	11.59	11.4
Gaps (%)	7.66	0	0.01

**Genome Sequencing.** There are more than 30 genomes of Australian species, with ~40 draft genomes in development through the Bioplatforms Australia initiatives. These genomes have used a variety of sequencing technology over the years, including whole-genome shotgun approach with Sanger sequencing [e.g., Tamar wallaby, *Macropus eugenii* (37)]; Illumina platform [e.g., Tasmanian devil, *Sarcophilus harrisii* (38)]; PacBio RS II platform with Illumina HiSeq [e.g., koala, *Phascolarctos cinereus* (39)], and 10X Genomics linked-read sequencing on NovaSeq. 6000 [e.g., brown antechinus, *Antechinus stuartii* (40)]. Some of these genomes may be now classified as low-quality by today's genome standards, but their conservation application has been significant. For example, the original 2012 Tasmanian devil genome (38) (Table 2) has been used with much success for the management of both wild and captive populations of this endangered species (see full review, ref. 11)]. The Tasmanian devil genome allowed for the development of conservation-based tools, such as species-specific microsatellite markers, characterization of immune gene families, blocking primers for use in metagenomics studies, as a few examples (11). The 2018 koala genome (39) (Table 2), is permitting a large-scale genomic survey of the species to understand both genome-wide and functional diversity in light of the recent Australian megafires, which saw more than 126,000 km<sup>2</sup> of habitat burned (41). This genomic survey will inform potential future management actions around habitat restoration and translocations for a globally recognized species. Other draft genomes for the woylie [*Bettongia penicillate ogilbyi* (42)] have been used in real-time (as the genome was assembled) to inform management actions and translocation success for both the woylie (43), and other cogenetic species (16), such as the boodie (*Bettongia lesueur*). It should be noted that most of these genomes are not chromosome length assemblies, although the recently released koala chromosome assembly ([https://www.dnazoo.org/assemblies/Phascolarctos\\_cinereus](https://www.dnazoo.org/assemblies/Phascolarctos_cinereus), January 2021) has improved the 2018 assembly (Table 2). During the 17 y between the human genome being published (44, 45) and the chromosome-scale, haplotype-resolved assembly being released (46), the original genome exponentially changed human medicine and our understanding of *Homo sapiens*. As a result, the TSI Steering Committee has opted to fund long-read genome data [HiFi reads of PacBio Sequel II system (47)] with associated species-specific transcriptome data for more species to meet conservation needs, rather than focusing on producing chromosome-length assemblies for a few species. Project teams are encouraged to seek funding to facilitate chromosome-length assemblies in the future using HiC (48) technology. Appropriately collected and stored tissue samples are being archived where possible within Australian museum collections to ensure future assemblies use the same specimen (49).

Sampling requirements for high-quality genomes can be extremely difficult to meet for threatened species, particularly those that are listed as critically endangered (49, 50). Many long-read technologies require nonfragmented DNA, which is most easily obtained from tissue samples that are flash-frozen or freshly collected. While relatively large amounts of fresh, preferably young, leaves are required for the high molecular weight DNA extraction needed for assembling a plant genome, collecting leaf tissue for genotype by sequencing is less stringent and requires significantly smaller amounts of silica-dried tissue (and can even work from herbarium specimen). Given the static nature of plants, and the small population size of many of the most threatened species, sometimes most living individuals can be sampled (51). For animal species, however, collecting fresh tissue samples that need to be flash-frozen from cryptic species is more problematic. It is also impractical in a large geographic country like Australia ~7.69 million km<sup>2</sup>, with a relatively small human population (~25.4 million), where access to liquid nitrogen in remote locations is logistically challenging and transport networks from remote locations are limited, resulting in difficulties transporting samples to laboratory facilities in a timely manner. Furthermore, many Australian animal species are small, and so blood volumes greater than 100 to 500 µL may not be achievable.

Although sequencing costs in the United States, Europe, and China are relatively low, the nature of distance and small turnover in other parts of the world means that discounted sequencing costs tend not to be available for many. Of the 17 megadiverse nations (2), the United States has the cheapest sequencing. To ensure the full value of genomic resources for the conservation of global biodiversity, it is important to invest in local conservation communities and empower them to develop resources within country. For many threatened and endemic species, sending samples to the United States, Europe, and China may be also be constrained by international (e.g., CITES) and national (e.g., United States Endangered Species Act; Australian Environment Protection and Biodiversity Conservation Act) biosecurity, trade regulations, and permit requirements. Furthermore, for many Indigenous and First Nations peoples the natural world, and their affiliation with it, holds cultural significance, meaning that movement of samples, or even extracted DNA, across international borders is often restricted. This brings to the fore potential issues with sampling and the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization (52, 53). Globally we need to embed indigenous principles into genomic research (36, 54), and be able to facilitate genome projects within nations, often where sequencing is not cheap. This requires us to rethink what kinds of genomes we are seeking to produce to effect change in conservation practice



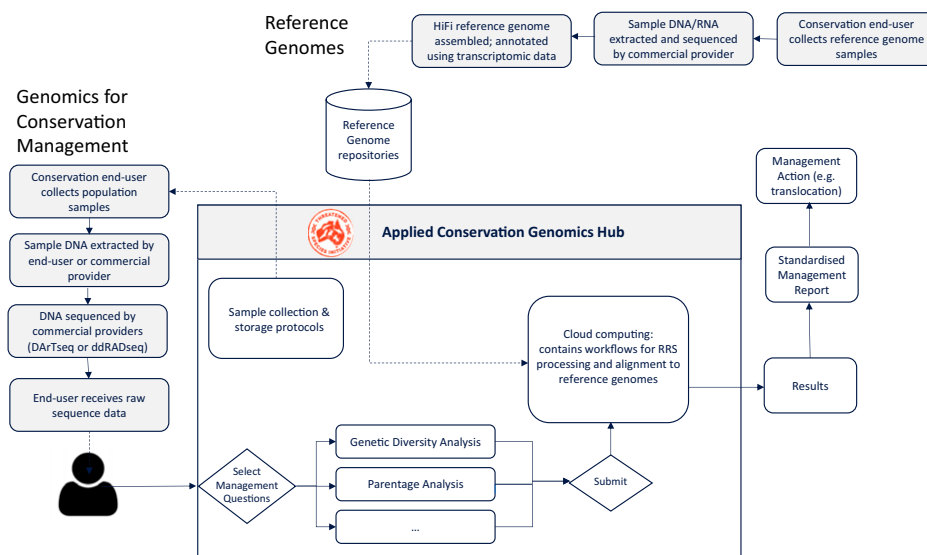
and ensure the genomic resources, and associated downstream tools that are created, are utilized to their full potential (50).

TSI is also producing supporting population genetic data (for up to 190 individuals) for species that require it to inform conservation management action (Fig. 3). This will not cover all the population genetic data that will be required for some species, but rather is a launchpad for coinvestment into using genetic data for conservation management. Reduced representation sequencing (RRS) has been selected for population genetic data, although it does have limitations for some population analyses, such as runs of homozygosity (RoH), identification of alleles within species genes, or effective population sizes. For these analyses, whole-genome resequencing (WGS) is needed but is also currently costly for many taxa with larger genomes (e.g., mammals, amphibians). Either double-digest RADseq (55) (ddRAD) or Diversity Arrays Technology (56) (DArTseq), have been selected as the sequencing methods of choice for TSI population genetics, as both are readily available within Australia from commercial providers and will ensure that the bioinformatic workflows are useful across the range of taxa to be undertaken in this project. Our current workflow can either align RRS data to a reference genome or be used de novo (57). Using species-specific transcriptome data to annotate the genomes allows for conservation managers to have access to functional data, particularly around gene families that are not conserved between species, such as the immune genes.

To facilitate the long-term uptake of genetic data into population monitoring and management, TSI is also trialing the use of low-density SNP arrays, where reduced subsets of informative SNP loci identified through the above WGS and population genomic approaches are selected and optimized for high-throughput automated genotyping. SNP arrays can be flexibly designed to contain loci targeted to specific conservation applications: for example, to ascertain population structure and monitor neutral and adaptive genetic diversity (58–60), assess parentage and kinship (61, 62), and monitor introgression/hybridization (63). Besides the initial investment in SNP discovery and multiplex primer design, downstream genotyping costs are highly affordable (e.g., MassARRAY iPLEX system ~AUD\$11

per sample per 50-plex) with minimal requirements for data analysis, making the routine genetic analysis of populations accessible to a wider array of end-users. Furthermore, SNP genotyping systems, such as MassARRAY, are suitable for application with noninvasive samples (scats, hair) (64), expanding the utility of the method in wildlife monitoring scenarios. We advocate for developing arrays and calling SNPs against reference genomes to ensure future use of the data as SNP locations will be known. As more high-quality reference genomes become available and sequencing costs reduce, WGS will become the norm. In the interim however, using RRS data aligned to a draft reference genome can permit a wide-range of conservation actions for a species [see Brandies et al. (11)].

**Applied Conservation Genomics Hub.** A key aim of the TSI is to develop an online platform, an applied conservation genomics hub, to empower nongeneticists to be able to use these genomic resources in their conservation decision-making. The TSI is committed to developing such a platform (Fig. 3). The Hub will host protocols for sample collection and storage, in addition to a suite of existing analytical pipelines and workflows [e.g., STACKS (57), dartR (65), Sequoia (66)] with a user-friendly interface that has point-and-click options, rather than a command-line interface. The outputs from these workflows can be used to answer some of the most common conservation management questions (*SI Appendix, Table S2*). Users will be able to manipulate their data for their specific species, but the output report will be standardized, with different modules for different management questions. The report will be in a simple, consistent format to ensure that conservation practitioners are receiving the same information for their species in a standardized way so they can become familiar with summary methods for genetic data. Reports will include standard genetic metrics (such as heterozygosity, inbreeding, relatedness) in addition to an appendix with sequencing methods used, number of filtered SNPs, filtering used, and compute requirements for the datasets. Standardizing the reporting will assist with reproducibility over time. Users who are creating the reports will also have the option to add more outputs/variables if they so desire. By standardizing



**Fig. 3. Components and the interoperable framework of the TSI. Currently, smaller working groups are supporting the development of workflows and protocols for sample collection and storage, bioinformatics, and standardized reporting.**

the output report, we aim to further promote the education of the conservation practitioners in the use of genetic data in the management practice and encourage the uptake of longer-term genetic monitoring in-line with the Convention of Biological Diversity targets (67, 68). This is perhaps TSI's biggest innovation, because while techniques can change and initial interpretations might be complex, once baseline genomic information is developed and there is standardized management reporting, cheap, effective, long-term monitoring tools can become a reality.

We fully recognize that this online platform and associated standardized reporting will not be a simple task to achieve, as there are many nuances in the interpretation of genetic data for management purposes. However, with the ever-widening gap between genome biologists and conservation practitioners, we need to develop solutions to bridge this divide. Not knowing how to interpret and use the information, nor how it is generated or who to contact, are a few of the reasons that have been flagged by conservation practitioners for why they are not routinely using genetic data in their management practice (28). The platform will be a living, iterative system, which we anticipate will start small and grow with time, use, need, and technological development. TSI has recognized that we need to start to fill this niche, as the gap between the genome biologists and the conservation practitioners is widening each year as the costs of sequencing reduce, bioinformatics becomes more challenging, and the need for genomic resources for conservation management increases.

## Conclusion

Genomes are a powerful tool. Their downstream applications are far reaching and can assist in answering complex biological questions, evolutionary processes, and complement current ecological studies. Yet, there is a gap between the

genome biologists and bioinformaticians creating reference genomes for nonmodel species and those conservation practitioners, ecologists, and policy makers for which the data are extremely useful. Here we have presented the Threatened Species Initiative, a new project that will generate genomes and associated genomic resources, in addition to providing conservation practitioners with the tools to apply the data in their conservation actions, decision-making, management, and policy. Over time we aim to collaborate with other international consortia to utilize the tools developed for applied conservation management. We hope that the framework that underpins the TSI will persist for as long as others wish to use a multidisciplinary approach when managing threatened species.

**Data Availability.** All study data are included in the main text and [SI Appendix](#).

## Acknowledgments

We thank the traditional custodians of the land upon which we work, and pay respects to their elders, past, present, and emerging; the team of the Australian Wildlife Genomics Group (University of Sydney), who have contributed to the development of the Threatened Species Initiative over the years; Luke Silver for his assistance in collating Tasmanian devil and koala information; Adele Gonsalvez for collating the International Union for the Conservation of Nature genome information; and the collaborators of the Threatened Species Initiative who provided samples, contribute their time to developing protocols around workflows, and undertake downstream analysis of genetic data for management purposes. The Threatened Species Initiative is supported by funding from Bioplatforms Australia, through the Australian Government National Collaborative Research Infrastructure Strategy and the Australian Department of Agriculture, Water, and Environment. Institutional support is provided to C.J.H., K.O., and P.L. by the University of Sydney, the Western Australian Department of Biodiversity, Conservation, and Attractions, and the Australian Government Department of Agriculture, Water, and the Environment, respectively. We also acknowledge the ongoing support of Amazon Web Services and RONIN cloud services in the development of the on-line cloud computing portal.

- 1 S. P. Bachman *et al.*, Progress, challenges and opportunities for Red Listing. *Biol. Conserv.* **234**, 45–55 (2019).
- 2 R. A. Mittermeier, C. G. Mittermeier, P. R. Gil, *Megadiversity: Earth's Biologically Wealthiest Nations* (CEMEX, Arlington, VA, 1997).
- 3 UN/DESA, *Country Classification—Data Sources, Country Classifications and Aggregation Methodology* (United Nations, 2014).
- 4 G. Ceballos, P. R. Ehrlich, R. Dirzo, Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl. Acad. Sci. U.S.A.* **114**, e6089–e6096 (2017).
- 5 G. Ceballos *et al.*, Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015).
- 6 United Nations, *The United Nations Decade on Ecosystem Restoration 2021–2030* (United Nations, New York, 2020).
- 7 R. Frankham *et al.*, *Genetic Management of Fragmented Animal and Plant Populations* (Oxford University Press, 2017).
- 8 IUCN/Species Survival Commission, *Guidelines for Reintroductions and Other Conservation Translocations. Version 1.0* (IUCN Species Survival Commission, Gland, Switzerland, 2013).
- 9 L. E. Commander *et al.*, *Guidelines for the Translocation of Threatened Plants in Australia* (Australian Network for Plant Conservation Inc., 2018).
- 10 H. A. Lewin *et al.*, Earth BioGenome project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4325–4333 (2018).
- 11 P. Brandies, E. Peel, C. J. Hogg, K. Belov, The value of reference genomes in the conservation of threatened species. *Genes (Basel)* **10**, 846 (2019).
- 12 H. A. Lewin, S. Richards; The Earth Biogenome Project Working Group, The earth BioGenome project 2020: Starting the clock. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115635118 (2022).
- 13 Genome 10K Community of Scientists, Genome 10K: A proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J. Hered.* **100**, 659–674 (2009).
- 14 H. Bracken-Grissom *et al.*; GIGA Community of Scientists, The Global Invertebrate Genomics Alliance (GIGA): Developing community resources to study diverse invertebrate genomes. *J. Hered.* **105**, 1–18 (2014).
- 15 M. A. Supple, B. Shapiro, Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, 131 (2018).
- 16 S. J. Galla *et al.*, Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes (Basel)* **10**, 9 (2018).
- 17 B. J. McMahon, E. C. Teeling, J. Höglund, How and why should we implement genomics into conservation? *Evol. Appl.* **7**, 999–1007 (2014).
- 18 K. Ralls *et al.*, Call for a paradigm shift in the genetic management of fragmented populations. *Conserv. Lett.* **11**, e12412 (2018).
- 19 J. Veevers, M. McElhinny, The separation of Australia from other continents. *Earth Sci. Rev.* **12**, 139–143 (1976).
- 20 R. Hall, "Sundaland and Wallacea, geology, plate tectonics and palaeogeography" in *Biotic Evolution and Environmental Change in Southeast Asia*, D. J. Gower *et al.*, Eds. (Cambridge University Press, Cambridge, UK, 2012), pp. 32–78.
- 21 A. D. Chapman, *Numbers of Living Species in Australia and the World* (Australian Biological Research Study, 2nd Ed., 2009).
- 22 W. J. Jackson *et al.*, "Overview: State and trends" in *Australia State of the Environment 2016* (Australian Government Department of the Environment and Energy, Canberra, Australia, 2016).
- 23 J. C. Z. Woinarski, A. A. Burbidge, P. L. Harrison, Ongoing unraveling of a continental fauna: Decline and extinction of Australian mammals since European settlement. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4531–4540 (2015).

- 24 M. J. Lawes *et al.*, Small mammals decline with increasing fire extent in northern Australia: Evidence from long-term monitoring in Kakadu National Park. *Int. J. Wildland Fire* **24**, 712–722 (2015).
- 25 G. R. Fulton, The Bramble Cay melomys: The first mammalian extinction due to human-induced climate change. *Pac. Conserv. Biol.* **23**, 1–3 (2017).
- 26 J. J. Lahoz-Monfort *et al.*, A call for international leadership and coordination to realize the potential of conservation technology. *Bioscience* **69**, 823–832 (2019).
- 27 A. T. Knight *et al.*, Knowing but not doing: Selecting priority conservation areas and the research-implementation gap. *Conserv. Biol.* **22**, 610–617 (2008).
- 28 H. R. Taylor, N. Dussex, Y. van Heezik, Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* **10**, 231–242 (2017).
- 29 S. J. Galla *et al.*, Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. *Mol. Ecol.* **25**, 5267–5281 (2016).
- 30 M. Britt, S. E. Haworth, J. B. Johnson, D. Martchenko, A. B. A. Shafer, The importance of non-academic coauthors in bridging the conservation genetics gap. *Biol. Conserv.* **218**, 118–123 (2018).
- 31 M. P. Tan *et al.*, Applications of next-generation sequencing technologies and computational tools in molecular evolution and aquatic animals conservation studies: A short review. *Evol. Bioinform. Online* **15**, 1176934319892284 (2019).
- 32 S. J. Luo, Y. C. Liu, X. Xu, Tigers of the world: Genomics and conservation. *Annu. Rev. Anim. Biosci.* **7**, 521–548 (2019).
- 33 T. Marques-Bonet, C. Hvilsom, Genomic variation of the great apes and the application to conservation. *Int. Zoo Yearb.* **52**, 25–33 (2018).
- 34 C. J. Hogg *et al.*, "Devil tools & tech": A synergy of conservation research and management practice. *Conserv. Lett.* **10**, 133–138 (2017).
- 35 M. Rossetto *et al.*, Restore and renew: A genomics-era framework for species provenance delimitation. *Restor. Ecol.* **27**, 538–548 (2019).
- 36 A. M. Mc Cartney *et al.*, Balancing openness with Indigenous data sovereignty: An opportunity to leave no one behind in the journey to sequence all of life. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115860119 (2022).
- 37 M. B. Renfree *et al.*, Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol.* **12**, R81 (2011).
- 38 E. P. Murchison *et al.*, Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* **148**, 780–791 (2012).
- 39 R. N. Johnson *et al.*, Adaptation and conservation insights from the koala genome. *Nature Genetics* **50**, 1102 (2018).
- 40 P. A. Brandies, S. Tang, R. S. P. Johnson, C. J. Hogg, K. Belov, The first *Antechinus* reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. *GigaByte*, 10.46471/gigabyte.7 (2020).
- 41 B. A. Wintle, S. Legge, J. C. Z. Woinarski, After the megafires: What next for Australian wildlife? *Trends Ecol. Evol.* **35**, 753–757 (2020).
- 42 E. Peel, L. Silver, P. Brandies, C. J. Hogg, K. Belov, A reference genome for the critically endangered woylie, *Bettongia penicillata ogilbyi*. *GigaBytes*, in press.
- 43 K. A. Farquharson *et al.*, Metapopulation management of a critically endangered marsupial in the age of genomics. *Glob. Ecol. Conserv.* **31**, e01869 (2021).
- 44 E. S. Lander *et al.*; International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- 45 International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- 46 S. Garg *et al.*, Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat. Biotechnol.* **39**, 309–312 (2021).
- 47 D. Lang *et al.*, Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultra-long reads of Oxford Nanopore. *GigaScience* **9**, g1aa123 (2020).
- 48 E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- 49 M. K. N. Lawniczak *et al.*, Standards recommendations for the earth BioGenome project. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115639118 (2022).
- 50 E. Peel *et al.*, Best genome sequencing strategies for annotation of complex immune gene families. *Genome Biol.*, in press.
- 51 M. Rossetto *et al.*, A conservation genomics workflow to guide practical management actions. *Glob. Ecol. Conserv.* **26**, e01492 (2021).
- 52 M. Y. Teran, The Nagoya protocol and indigenous peoples. *Int. Indig. Policy J.*, 10.18584/iipj.2016.7.2.6 (2016).
- 53 M. Buck, C. Hamilton, The Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity. *Rev. Eur. Community Int. Environ. Law* **20**, 47–61 (2011).
- 54 L. Collier-Robinson, A. Rayne, M. Rupene, C. Thoms, T. Steeves, Embedding indigenous principles in genomic research of culturally significant species: A conservation genomics case study. *New Zeal. J. Ecol.* **43**, 3389 (2019).
- 55 B. Peterson, J. Weber, E. Kay, H. Fisher, H. Hoekstra, Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135 (2012).
- 56 C. P. Sansaloni *et al.*, A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* **6**, 16 (2010).
- 57 B. Wright *et al.*, From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics* **20**, 453 (2019).
- 58 J. B. Stetz *et al.*, Discovery of 20,000 RAD-SNPs and development of a 52-SNP array for monitoring river otters. *Conserv. Genet. Resour.* **8**, 299–302 (2016).
- 59 B. Dunker *et al.*, A set of 156 SNP markers for teal (*Tectona grandis* Linn. f.). *Conserv. Genet. Resour.* **12**, 205–207 (2020).
- 60 S. Mathur, J. M. Tomeček, A. Heniff, R. Luna, J. A. DeWoody, Evidence of genetic erosion in a peripheral population of a North American game bird: The Montezuma quail (*Cyrtonyx montezumae*). *Conserv. Genet.* **20**, 1369–1381 (2019).
- 61 S. A. Kaiser *et al.*, A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Mol. Ecol. Resour.* **17**, 183–193 (2017).
- 62 S. Roques, E. Chancerel, C. Boury, M. Pierre, M. L. Acolas, From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecol. Evol.* **9**, 7017–7029 (2019).
- 63 B. Nussberger, P. Wandeler, G. Camenisch, A SNP chip to detect introgression in wildcats allows accurate genotyping of single hairs. *Eur. J. Wildl. Res.* **60**, 405–410 (2014).
- 64 E. L. Carroll *et al.*, Genetic and genomic monitoring with minimally invasive sampling methods. *Evol. Appl.* **11**, 1094–1119 (2018).
- 65 B. Gruber, P. J. Unmack, O. F. Berry, A. Georges, darr: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699 (2018).
- 66 J. Huisman, Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.* **17**, 1009–1024 (2017).
- 67 S. P. Flanagan, B. R. Forester, E. K. Latch, S. N. Aitken, S. Hoban, Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evol. Appl.* **11**, 1035–1052 (2017).
- 68 S. Hoban *et al.*, Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biol. Conserv.* **248**, 108654 (2020).