INSIGHTS

PERSPECTIVES

GENOMICS

Reference genomes for conservation

High-quality reference genomes for non–model species can benefit conservation

By **Sadye Paez, Robert H. S. Kraus, Beth Shapiro, M. Thomas P. Gilbert, Erich D. Jarvis, the Vertebrate Genomes Project Conservation Group**

As of 2022, the International Union for Conservation of Nature (IUCN) Red List estimates that more than 32% of fungal, plant, and animal species are threatened with extinction. This sixth mass extinction is caused by the activities and expanding biomass of humans, necessitating a distinct name for this geological epoch—the Anthropocene (*1*). Human population growth and the vertebrate extinction rate (*2*) have been linearly correlated over the past 500 years (see the figure). For some species of conservation concern, documenting, informing, and mitigating this biodiversity loss has been helped by powerful genomic tools, including a reference assembly (*3*). Yet, currently, only a small fraction (<1%) of the ~35,500 species assessed as threatened with extinction have an available genome assembly, and to date, most are in draft form. It is proposed that conservation efforts can be enhanced by the production of high-quality reference genome assemblies.

Conservation genomics leverages genetic data, from individual loci to genomic-scale datasets, to aid preservation of species and population-level biodiversity. This includes using genomic data to measure effective population sizes, demographic history, and genetic diversity and to perform genetic manipulations pre- or postextinction. Many of these efforts have been conducted using first- and second-generation genome sequencing and assembly technologies with short reads, leading to sequence errors, structural errors, and missing sequences. Now, third-generation genome technologies—with improvements in longer read lengths, nucleotide accuracy, chromosomal maps, and assembly algorithms—have led to more complete assemblies (*4*). These high-quality assemblies have 10- to 200-fold improvements in quality metrics, including the amount of sequence assignable to chromosomes, genes fully assembled, and recovery of GC-rich regulatory regions (*4*). Method developments are also underway for generating complete and error-free genome assemblies (telomere to telomere) (*5*) of both maternal and paternal haplotypes. Given the extra computational and financial costs that such improved genomes incur, a legitimate question often asked is, what is the added value of these high-quality assemblies, beyond current draft genomes, for conservation?

Species must maintain a certain level of genetic diversity to adapt to various environmental changes and/or population decreases, whether natural or human driven. Genetic diversity or other genomic health assessments have historically drawn on polymerase chain reaction (PCR)–generated sequence data from DNA microsatellites, which are tandem repetitive sequences that tend to diverge at a higher rate compared with single-nucleotide variants. Typically, the greater the diversity in microsatellites among individuals, the healthier the population. Genomic health of a population can also be assessed by identifying changes in mutational load (the population frequency of deleterious alleles) and estimating lengths of runs of homozygosity (ROHs) (*6*). The accumulation of ROHs in small and inbred populations can fix or drive deleterious alleles to high frequencies. Identifying

Author affiliations and Vertebrate Genomes Project Conservation Group authors and affiliations are listed in the supplementary materials. Email: spaez@rockefeller.edu

PHOTO: STEPHEN BELCHER/MINDEN PICTURES

There were once ~30,000 kākāpō (*Strigops habroptilus*) on mainland New Zealand, but there are now only ~200 on nearby islands. High-quality reference genomes are aiding conservation breeding programs of these critically endangered parrots.

these and other signs of poor genomic health can provide a warning that a species is becoming critically endangered or identify specific populations that are fragile and deserving of focused conservation efforts. Although draft short read–based reference genomes have been successfully used for such analyses, third-generation reference genomes would lead to a comprehensive identification of microsatellites, mutational load, ROHs, and diverse segments of heterozygous variants (*6*) (see table S1).

Two examples of third-generation genome assemblies providing key information for conservation are those of the kākāpō (*Strigops habroptilus*) and vaquita (*Phocoena sinus*), which are both critically endangered (*4*) (see table S2). The kākāpō is a parrot endemic to New Zealand whose population once comprised ~30,000 individuals on the mainland. Human colonization circa 1360 CE and again in the 1800s reduced the population to 18 birds by 1977, but it is now recovering, with ~200 living on nearby islands. Analyses of a third-generation kākāpō genome and second-generation resequenced genomes from 49 individuals representing both extant and historical populations revealed that the surviving island population has had low genomic heterozygosity in long ROHs for the past 10,000 years, whereas the now-extinct mainland population did not (*6*). These findings affect conservation decision-making, whereby closely related individuals can now be bred with less concern for deleterious mutations, allowing a small population an opportunity to thrive.

The vaquita is a small porpoise endemic to the Gulf of California, Mexico, and is, at present, the world's most endangered marine mammal. Fewer than 19 individuals survive today, which is a reduction from a historical effective population size (N_e) of ~5000 (*7*) caused by bycatch in gillnets for shrimp and finfish over the past century. Inferred historical population analyses based on a third-generation vaquita genome assembly revealed that the species has had low genomic heterozygosity and a small N_e for the past ~250,000 years (*7*). This suggests that the vaquita, like the kākāpō, may have survived recent population declines because of effective long-term purging of deleterious mutations in the wild (*8*). Broadly, these findings indicate that low heterozygosity in a population will not always be detrimental if deleterious mutations have been purged and that introduction of individuals with higher heterozygosity but more deleterious alleles into a population with less deleterious alleles needs to be cautiously considered.

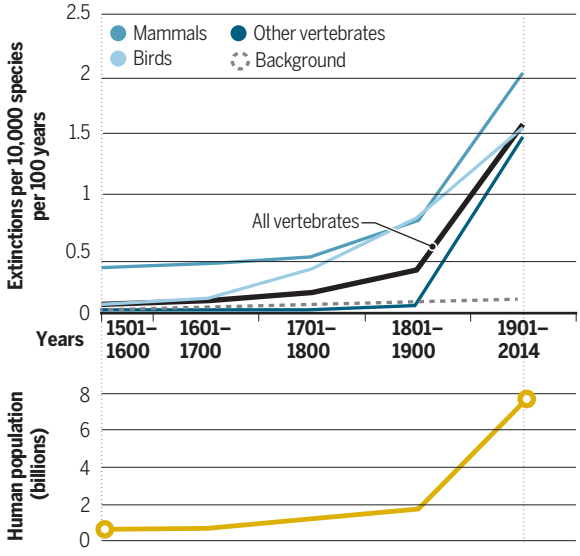
Genomic diversity can also be measured by counting and comparing structural variants (SVs), such as indels (insertions and deletions), inversions, chromosomal fusions, copy-number variations, and transposable elements. Identification of SVs is more straightforward in third-generation assemblies (*4*). SVs are increasingly appreciated as rich sources of adaptive polymorphism (*9*), representing conservation-relevant biological adaptations. For example, some SVs are adaptations to certain diseases, and thus selective breeding of individuals with these variants could potentially enhance population resistance to environmental changes.

Developing and implementing conservation management strategies often requires delineation of populations or species to distinguish between subspecies and cryptic species and, consequently, to differentiate conservation strategies. This was originally determined with DNA barcodes—that is, small fragments of DNA that are divergent between species, such as fragments of the organellular cytochrome oxidase subunit 1 and 16*S* and 12*S* ribosomal RNA genes—although this is increasingly complemented with analysis of whole-organelle sequences. The long-read approaches used in third-generation genome assemblies simplifies the reconstruction of entire organelle genomes (*4*). These organelle genomes have revealed repeat regions and gene duplications that were not assembled properly or were missed entirely in first- and second-generation genome assemblies (*4*). Additionally, these organelle genome assemblies, sometimes generated as a single-molecule sequence read, can be used for refined species delineation, phylogeography, and population studies; they also reduce problems that arise when, for example, mitochondrial genomes are confused with nuclear mitochondrial sequences (NUMTS).

Genetic rescue—which includes genetically informed translocations of a species from one geographical region to another, other breeding strategies, and more extreme interventions such as gene editing—aims to increase diversity or prevent the fixation of deleterious alleles by facilitating gene flow from one population to another. Although applications of gene editing have been mainly limited to agriculture, for example, to augment disease resistance in crops (*10*), proposed applications to conservation include improving a species' resistance to viral and bacterial infections or toxins and a species' capacity to adapt to anthropogenic and natural changes to their habitats, such as changes in temperature, salinity, or precipitation. Although these approaches are in the early stages of development and additional research is needed, third-generation genome assemblies may be critical to such efforts, for example, by

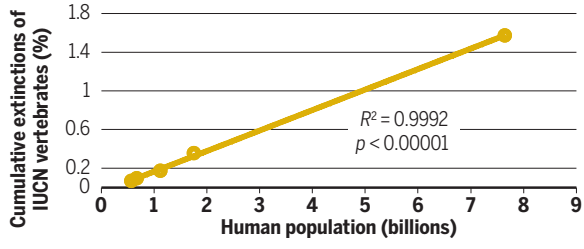
Vertebrate extinction rate and human population growth

The extinction rate of vertebrates, calculated according to historical and current records of the International Union for Conservation of Nature (IUCN) animal extinction list, is shown. All vertebrates as well as different species combinations are shown. The expected cumulative background extinction is based on geological extinction estimates between the fifth mass extinction (~65 million years ago) and 10,000 years ago. Graphs are modified from (*2*).



Human population correlates with extinction

There is a correlation between human population size and cumulative extinction of all vertebrates in the past 500 years; each point is a 100-year mean from the extinction data in the graph above.



GRAPHIC: N. DESAI/*SCIENCE*; (HUMAN POPULATION DATA) ECOLOGY.COM

better assembling translocations and having nearly all available sequences to determine potential off-target sites of genome editing.

Another potential benefit of third-generation genome assemblies to conservation will be for deextinction, such as resurrecting extinct traits in a living species or proxies of extinct species, for example, creating cold-adapted elephants using genomic diversity that evolved in the woolly mammoth (*11*). One approach is cloning using somatic cell nuclear transfer (SCNT), whereby nuclei of cells from extinct sublineages are transferred into enucleated oocytes that are then transplanted into a female. Preliminary reports indicate that this was successfully done for the Przewalski's horse and black-footed ferret with decades-old cryobanked cells, with the resulting clones still living in captivity (*12*). But this approach requires preserved living cells, which limits its application, and it is also not straightforward for egg-laying species such as birds and fishes. In these cases, gene editing of cells before SCNT or of early-stage embryos before egg formation might work better. But this approach requires knowledge of which edits to make. Contiguous and nearly complete genomes provide greater resolution to identify species-specific coding and regulatory sequences for gene editing.

In the absence of frozen cells, complete genome sequence data could also be used to create synthetic chromosomes and place them into viable cells, as was achieved by the Yeast 2.0 Project, which synthesized the entire genome of *Saccharomyces cerevisiae* (*13*). Although yeast genomes are 3 to 10% of the size of vertebrate genomes and the technology does not yet exist to synthesize larger genomes, this highlights the potential power for synthetic biology in deextinction efforts. In multicellular organisms, synthesized chromosomes could be placed in enucleated oocytes of another species, similar to the SCNT approach.

As genomic analyses and synthetic biology become components of conservation management, there are challenges to overcome, including developing approaches that consider other complex genome organizations, such as species with germline cells that have germline-specific chromosomes (e.g., lamprey and songbirds), and rearranged chromosomes during different developmental stages (e.g., single-cell ciliate protists) (*14*, *15*). Multicellular organisms also rely on microbial symbionts, some of which are inherited. High-quality genome assemblies for symbiotic microbes, such as those being developed by the Earth HoloGenome Initiative, are the crucial first step in incorporating this information into conservation management plans.

Addressing biodiversity loss is a complex problem that requires multifaceted solutions. Genomics can be an important component of conservation management. It is urgent that high-quality reference genome assemblies and cryopreserved cells be produced for endangered species now and, eventually, for all species. Waiting for technology improvements, policy changes, or outcomes of nongenomic efforts places too many species in peril. Generating high-quality genome assemblies from poorly preserved tissue or fossil remains (DNA can be extracted from samples up to 1 million years old in permafrost) is impossible because of the short lengths of surviving DNA molecules, highlighting the need for optimized cryopreservation of cells and tissues. When sex chromosomes exist, sequencing the heterogametic sex (e.g., males in mammals and females in birds) is preferable. Also, material should be preserved from multiple individuals so that information about population genetic diversity can be obtained. A notable and continuing challenge lies with the ethical, legal, and moral implications of translating genomic data to conservation. Coordination between scientists and other stakeholders is important, especially for access and benefit sharing of samples and the resulting digital sequence information with Indigenous Peoples and local communities. Genome assemblies by themselves, even if complete and error free, cannot fully address the ongoing sixth mass extinction. But high-quality reference genome assemblies are advantageous for pre- and postconservation management and monitoring with other strategies, such as preserving land, forest, and water reserves, and with other protections to the environment. ■

REFERENCES AND NOTES

1. C. N. Waters *et al.*, *Science* **351**, aad2622 (2016).
2. G. Ceballos *et al.*, *Sci. Adv.* **1**, e1400253 (2015).
3. P. Brandies, E. Peel, C. J. Hogg, K. Belov, *Genes* **10**, 846 (2019).
4. A. Rhie *et al.*, *Nature* **592**, 737 (2021).
5. S. Nurk *et al.*, *Science* **376**, 44 (2022).
6. N. Dussex *et al.*, *Cell Genom.* **1**, 100002 (2021).
7. P. A. Morin *et al.*, *Mol. Ecol. Resour.* **21**, 1008 (2021).
8. J. A. Robinson *et al.*, *Science* **376**, 635 (2022).
9. D. Weetman, L. S. Djogbenou, E. Lucas, *Curr. Opin. Insect Sci.* **27**, 82 (2018).
10. M. P. Phelps, L. W. Seeb, J. E. Seeb, *Conserv. Biol.* **34**, 54 (2020).
11. B. Shapiro, *Genome Biol.* **16**, 228 (2015).
12. Revive & Restore, Projects (2021); https://reviverestore.org/projects/.
13. I. S. Pretorius, J. D. Boeke, *FEMS Yeast Res.* **18**, foy032 (2018).
14. C. M. Kinsella *et al.*, *Nat. Commun.* **10**, 5468 (2019).
15. J. J. Smith *et al.*, *Nat. Genet.* **50**, 270 (2018).

ACKNOWLEDGMENTS

S.P. and R.H.S.K. contributed equally to this work.
See supplementary materials for full acknowledgments.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm8127

10.1126/science.abm8127

PLANETARY SCIENCE

In the glare of the Sun

Searches during twilight toward the Sun have found several asteroids near Venus' orbit

By **Scott S. Sheppard**

Asteroid surveys generally operate at night, mostly finding objects beyond Earth's orbit. This creates a blind spot because many near-Earth objects (NEOs) could be lurking in the sunlight interior to Earth's orbit. New telescopic surveys are braving the Sun's glare and searching for asteroids toward the Sun during twilight. These surveys have found many previously undiscovered asteroids interior to Earth, including the first asteroid with an orbit interior to Venus, 'Ayló'chaxnim (2020 AV2), and an asteroid with the shortest-known orbital period around the Sun, 2021 PH27 (*1*, *2*).

NEOs are classified into different dynamical types (see the figure). Starting from the most distant are the Amors, which approach Earth but do not cross Earth's orbit. Apollos cross Earth's orbit but have semimajor axes greater than that of Earth. Atens also cross Earth's orbit but have semimajor axes less than that of Earth. Atiras (also called Apohele) have orbits completely interior to Earth, and Vatiras have orbits completely interior to Venus, with 2020 AV2 being the first known.

NEOs have dynamically unstable orbits of ~10 million years. A reservoir must exist that replenishes the NEOs because their numbers have been in a steady state over the past few billion years (*3*). Most NEOs are likely dislodged objects from the main belt of asteroids between Mars and Jupiter (*4*–*6*). Physical observations show that NEOs are similar to main belt asteroids (MBAs), with a small fraction being dormant comets from the outer Solar System (*7*).

MBAs with orbital periods near whole number ratios with Jupiter's period are depleted, which indicates that these areas are dynamically unstable. Small MBAs continually move into these unstable regions

Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC 20015, USA.
Email: ssheppard@carnegiescience.edu

Reference genomes for conservation

Sadye PaezRobert H. S. KrausBeth ShapiroM. Thomas P. GilbertErich D. Jarvis, and Farooq Omar Al-Ajli, and Gerardo Ceballos, and Andrew J. Crawford, and Olivier Fedrigo, and Rebecca N. Johnson, and Warren E. Johnson, and Tomas Marques-Bonet, and Phillip A. Morin, and Ralf C. Mueller, and Oliver A. Ryder, and Emma C. Teeling, and Byrappa Venkatesh

Science, 377 (6604), • DOI: 10.1126/science.abm8127