Science

AAAS

Supplementary Materials for

Reference genomes for conservation

Sadye Paez *et al.*

Corresponding author: Sadye Paez, spaez@rockefeller.edu

The PDF file includes:

Tables S1 and S2
Full list of authors and affiliations
Acknowledgments
References

Supplementary Materials

Conservation pipeline	Task or question	Draft-quality assemblies	Reference assemblies	T2T-quality assemblies	Available algorithms and technology
Samples, assembly & annotation	Acquisition and extraction of ultra-high molecular weight DNA required for assembly	NO	YES	YES	YES
	Fully assemble and annotate all genes	NO	YES	YES	YES
Biodiversity and species delineation	Fully characterize biodiversity	NO	YES	YES	YES
	Fully assemble mitochondrial DNA	NO	YES	YES	YES
	Redefine & reclassify cryptic species	POSSIBLE	YES	YES	NO
	Identify all structural variants	NO	NO	YES	NO
	Fully characterize coding variation	NO	YES	YES	YES
	Infer species evolutionary history	YES	YES	YES	YES
	Predict future speciation events	YES	YES	YES	NO
	Identify sex-specific allele fixation in sex-linked genes	YES	YES	YES	NO
Genomic health assessment	Fully assess allelic diversity	NO	NO	YES	YES
	Determine effective population size	YES	YES	YES	YES
	Analyze inbreeding coefficient	YES	YES	YES	YES
	Analyze outbreeding depression	YES	YES	YES	YES
	Quantify historical hybridization	POSSIBLE	YES	YES	POSSIBLE
	Quantify current hybridization events	YES	YES	YES	POSSIBLE
	Compute runs of homozygosity (to assess inbreeding)	NO	YES	YES	YES
	Identify purging of harmful mutations	YES	YES	YES	YES
Evolution inference	Infer evolution of full genome structure	NO	YES	YES	YES
	Describe historical demography	YES	YES	YES	YES
Heterozygosity	Generate fully phased haplotypes	NO	MAYBE	YES	YES
	Assess linkage disequilibrium/reproductive isolation	YES	YES	YES	YES
Species rescue	Genetic rescue via gene editing	POSSIBLE	YES	YES	YES
	Genetic rescue via breeding programs	YES	YES	YES	YES
	Reintroduction	YES	YES	YES	YES
	Restore all long-extinct alleles	NO	YES	YES	YES
	Resurrect an extinct species	NO	UNLIKELY	POSSIBLE	NO

Supplementary Table 1 | Qualitative assessment of the assembly quality and new algorithms needed to address specific tasks in conservation associated with a mass extinction. The specific metrics for each genome assembly level are defined in Rhie et al 2021 (ref 4 of main text). The first two columns (purple) indicate the task or question that needs to be addressed. The next columns indicate the genome assembly level (green) and/or algorithms (blue) required to complete/answer those tasks/questions. NO indicates that data type and/or algorithms are not sufficient to address the tasks/questions; YES indicates that data type and/or algorithms are sufficient to address the tasks/questions; POSSIBLE indicates that it is theoretical to address the tasks/questions with current data type and/or algorithms but that this has not yet been well tested. These are qualitative assessments based on the authors' experiences and general knowledge.

Common name	Latin name	Core question	Genomic regions	Technologies	Findings	Ref
Crested ibis	*Nipponia nippon*	Population genomics	SNPs, genes	Reference assembly and WGS resequencing (Illumina)	Population bottleneck with inbreeding, following by increased heterozygosity	(1)
**Kākāpō	*Strigops habroptilus*	Population genomics and mutational load	SNPs and genome rearrangement	VGP pipeline 1.5 reference assembly (PacBio) and WGS resequencing (Illumina)	Purging of deleterious mutations in endangered population, long term homozygosity	(2)
Killer whale	*Orcinus orca*	Phylogeography, taxonomy, evolution, and demographic diversity	SNPs	Reference assembly combined with Low-coverage WGS (Illumina)	Global divergence from niche colonization, genome-culture co-evolution, recurrent admixture	(3, 4)
Finless porpoise	*Neophocaena asiaeorientalis*	Taxonomy, adaptation to fresh water	SNPs	Moderate coverage WGS (Illumina)	Critically endangered Yangtze River population is a distinct species	(5)
**Vaquita	*Phocoena sinus*	Population genomics and mutational load	SNPs	Reference assembly (PacBio) combined with high-coverage WGS of other individuals	Long term small population size, low mutation load	(6, 7)
Chimpanzee	*Pan troglodytes*	Phylogeography, population genomics, selection, and gene flow	SNPs	High-coverage WGS (Illumina), genomic capture	Genetically identified confiscated animals from wild and captive populations	(8-10)
Mountain gorilla	*Gorilla beringei beringei*	Population genomics and mutational load	SNPs	High-coverage WGS (Illumina)	Low genome-wide heterozygosity, purging of deleterious variants	(11)
Orangutan	*Pongo abelii*, *P. pygmaeus*	Adaptive evolution, population demographic history, and phylogeography	Genes under selection, SNPs	Low to moderate coverage Illumina WGS	Significant adaptive differences in Sumatra and Borneo animals	(12, 13)
Island fox	*Urocyon littoralis*	Recent demographic history	SNPs, genes	High-coverage WGS (Illumina)	Long-term small population sizes, purging deleterious alleles	(14)
Sea otter	*Enhydra lutris*, *Pteronura brasiliensis*	Historical demography, adaptation, and mutational load	SNPs, genes under selection	High-coverage WGS (Illumina, Chicago Hi-C)	Low genome-wide heterozygosity, different burden of deleterious variants	(15)
Lion	*Panthera leo*	Species-complex including population structure and admixture	Whole genomes, mtDNA/autosomal markers	Illumina and BGISeq sequencing	Cave and modern lions did not hybridize following divergence	(16)
Tasmanian devil	*Sarcophilus harrisii*	Population diversity and immune genes to inform captive breeding programs and disease treatment	Microsatellites, immune gene diversity	Low to medium coverage WGS (454, Illumina)	Historically low genetic diversity likely contributed to devastating impact of facial tumor disease.	(17-21)

SNPs: single nucleotide polymorphisms

WGS: whole genome sequencing

**: high-quality reference genome assemblies

Supplementary Table 2 | Examples of how genomics has helped conservation of vertebrate species. Listed are 12 examples of endangered vertebrate species for which whole genome assemblies have enabled researchers to identify key genomic regions that show the effects of dwindling population size and/or human interventions to

prevent it. The examples are sorted taxonomically (birds, cetaceans, primates, carnivores, and a marsupial). Several of these examples (ibis, kākāpō, and vaquita) are discussed in the main text. Additional examples exist in the literature for vertebrates, insects, plants and other organisms.

Full list of authors and affiliations

Sadye Paez[1]*, Robert H. S. Kraus[2]†*, Beth Shapiro[4,5]*, M. Thomas P. Gilbert[6,7]*, Erich D. Jarvis[1,3,8]*

Vertebrate Genomes Project Conservation Group authors: Farooq Omar Al-Ajli[3,9,10], Gerardo Ceballos[11], Andrew J. Crawford[12], Olivier Fedrigo[3], Rebecca N. Johnson[13], Warren E. Johnson[14], Tomas Marques-Bonet[15,16,17,18], Phillip A. Morin[19], Ralf C. Mueller[20,21], Oliver A. Ryder[22], Emma C. Teeling[23], Byrappa Venkatesh[24]

†These authors contributed equally to this work.
*Corresponding author. Email: spaez@rockefeller.edu (S.P.); rkraus@ab.mpg.de (R.H.S.K.); beth.shapiro@gmail.com (B.S.); tgilbert@sund.ku.dk (M.T.P.G.); ejarvis@rockefeller.edu (E.D.J.)

[1]Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY 10065, USA.
[2]Department of Migration, Max Planck Institute of Animal Behavior, 78315 Radolfzell, Germany.
[3]Vertebrate Genome Laboratory, The Rockefeller University, New York, NY 10065, USA.
[4]Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA.
[5]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA.
[6]Center for Evolutionary Hologenomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
[7]NTNU University Museum, Trondheim 7012, Norway.
[8]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA.
[9]Monash University Malaysia Genomics Facility, School of Science, Selangor Darul Ehsan, Malaysia.
[10]Tropical Medicine and Biology Multidisciplinary Platform, Monash University Malaysia, Selangor Darul Ehsan, Malaysia.
[11]Instituto de Ecologia, Universidad Nacional Autónoma de Mexico, CU, Coyoacán, 04510 Ciudad de México, Mexico.
[12]Department of Biological Sciences, Universidad de los Andes, Bogotá 111711, Colombia.
[13]Smithsonian National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012, USA.
[14]The Walter Reed Biosystematics Unit and Smithsonian Conservation Biology Institute, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746, USA.
[15]Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain.
[16]Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010 Barcelona, Spain.
[17]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Baldiri i Reixac 4 08028, Spain.
[18]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columnes s/n, 08193 Cerdanyola del Vallès, Spain.
[19]Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla, CA, USA.
[20]Department of Migration, Max Planck Institute of Animal Behavior, 78315 Radolfzell, Germany.
[21]Department of Biology, University of Konstanz, 78464 Konstanz, Germany.
[22]San Diego Zoo Wildlife Alliance, Beckman Center, Escondido, CA 92027, USA.
[23]School of Biology and Environmental Science, University College, Dublin, Ireland.
[24]Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673, Singapore.

Acknowledgments

References and Notes

1. S. H. Feng *et al.*, The Genomic Footprints of the Fall and Recovery of the Crested Ibis. *Current Biology* **29**, 340-349.e347 (2019).
2. N. Dussex *et al.*, Purging of genetic load in the highly inbred and critically endangered kākāpō. *Cell Genomics*, **1**, 100002 ((2021).
3. A. D. Foote *et al.*, Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology* **28**, 3427–3444 (2019).
4. A. D. Foote *et al.*, Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications* **7**, 12 (2016).
5. X. M. Zhou *et al.*, Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nature Communications* **9**, 8 (2018).
6. P. A. Morin *et al.*, Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Mol Ecol Resour*, **21**, 1008-1020 (2021).
7. J. A. Robinson *et al.*, The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science* **376**, 635-639 (2022).
8. J. Prado-Martinez *et al.*, The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. *BMC Genomics* **14**, 363 (2013).
9. M. de Manuel *et al.*, Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**, 477-481 (2016).
10. P. Frandsen *et al.*, Targeted conservation genetics of the endangered chimpanzee. *Heredity (Edinb)* **125**, 15-27 (2020).
11. Y. Xue *et al.*, Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242 (2015).
12. A. Nater *et al.*, Morphometric, Behavioral, and Genomic Evidence for a New Orangutan Species. *Current Biology* **27**, 3487-3498.e3410 (2017).
13. M. P. Mattle-Greminger *et al.*, Genomes reveal marked differences in the adaptive evolution between orangutan species. *Genome Biology* **19**, 13 (2018).
14. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology* **28**, 3487-+ (2018).
15. A. C. Beichman *et al.*, Aquatic adaptation and depleted diversity: a deep dive into the genomes of the sea otter and giant otter. *Molecular Biology and Evolution*, (2019).
16. M. de Manuel *et al.*, The evolutionary history of extinct and living lions. *Proc Natl Acad Sci U S A* **117**, 10927-10934 (2020).
17. C. E. Grueber, E. Peel, R. Gooley, K. Belov, Genomic insights into a contagious cancer in Tasmanian devils. *Trends in Genetics* **31**, 528-535 (2015).
18. B. Wright *et al.*, Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. *BMC Genomics* **16**, 11 (2015).
19. B. Wright *et al.*, Variants in the host genome may inhibit tumour growth in devil facial tumours: evidence from genomewide association. *Scientific Reports* **7**, 6 (2017).
20. W. Miller *et al.*, Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences* **108**, 12348-12353 (2011).

21.   E. P. Murchison *et al.*, Genome Sequencing and Analysis of the Tasmanian Devil and Its Transmissible Cancer. *Cell* **148**, 780-791 (2012).