How genomics can help biodiversity conservation

European Reference Genome Atlas (ERGA) Consortium; Dennis, Alice

Published in:
Trends in Genetics

DOI:
10.1016/j.tig.2023.01.005

Publication date:
2023

Document Version
Version created as part of publication process; publisher's layout; not normally made publicly available

Trends in Genetics

CellPress OPEN ACCESS

Review

How genomics can help biodiversity conservation

Kathrin Theissinger,[1,36] Carlos Fernandes,[2,3,36] Giulio Formenti,[4,36] Iliana Bista,[5,6] Paul R. Berg,[7,8,9] Christoph Bleidorn,[10] Aureliano Bombarely,[11] Angelica Crottini,[12,13,14] Guido R. Gallo,[15] José A. Godoy,[16] Sissel Jentoft,[9] Joanna Malukiewicz,[17] Alice Mouton,[18] Rebekah A. Oomen,[8,9] Sadye Paez,[4] Per J. Palsbøll,[19,20] Christophe Pampoulie,[21] María J. Ruiz-López,[16,22] Simona Secomandi,[15] Hannes Svardal,[23] Constantina Theofanopoulou,[4,24] Jan de Vries,[25] Ann-Marie Waldvogel,[26] Guojie Zhang,[27,28,29] Erich D. Jarvis,[4] Miklós Bálint,[1] Claudio Ciofi,[30] Robert M. Waterhouse,[31,32] Camila J. Mazzoni,[33,34] Jacob Höglund,[35,*] and The European Reference Genome Atlas Consortium[37]

The availability of public genomic resources can greatly assist biodiversity assessment, conservation, and restoration efforts by providing evidence for scientifically informed management decisions. Here we survey the main approaches and applications in biodiversity and conservation genomics, considering practical factors, such as cost, time, prerequisite skills, and current shortcomings of applications. Most approaches perform best in combination with reference genomes from the target species or closely related species. We review case studies to illustrate how reference genomes can facilitate biodiversity research and conservation across the tree of life. We conclude that the time is ripe to view reference genomes as fundamental resources and to integrate their use as a best practice in conservation genomics.

Highlights

Genomics provides effective tools to characterize biodiversity, but the full implementation of genomic techniques in practical conservation is still limited. We review some of the main approaches in biodiversity genomics available to conservationists and genomicists.

High-quality, long-read sequencing and bioinformatic technologies facilitate genome sequencing and assembly for any species. We summarize how reference genomes, in conjunction with population genomic data, can contribute to biodiversity monitoring, conservation, and restoration efforts.

Over the past decade, many initiatives to generate reference genomes spanning the tree of life have emerged worldwide. We call for increased integration of reference genomes and population genomics data into interdisciplinary conservation efforts to fully unlock the potential of genomics in safeguarding global biodiversity.

The value of integrating genomics into conservation

We are in the midst of the sixth mass extinction, a biodiversity crisis with devastating consequences on ecosystem functioning and health, evolutionary heritage, and the adaptive potential of species, ultimately posing a major threat to humanity [1,2]. Although genetic diversity has long been recognized as fundamental to all levels of biological organization (individuals, populations, species, communities, and ecosystems), genomics is often neglected in biodiversity assessments and conservation efforts [3]. This gap between our ability to generate data and study genomic diversity and the use of genomics in conservation may be due to several factors [4]. These include prioritizing limited funds to address anthropogenic threats [5], limited knowledge transfer and collaboration between genomicists and conservation practitioners, stakeholders, and politicians [6], and a general lack of explicit references to measuring, monitoring, and preserving genetic diversity in the most relevant international regulations, such as those issued by the Convention for Biological Diversity [7] (although the situation may have recently started to change, https://www.cbd.int/article/cop15-cbd-press-release-final-19dec2022).

Biodiversity preservation critically depends on addressing key conservation issues. These include taxonomic identification and biodiversity monitoring associated with ecosystem protection and restoration (e.g., for invasive species management). At the same time, human activities exert significant demographic pressures on habitats and endangered species. This requires managing small populations, restoring and increasing genetic diversity of target species/populations, and supporting species adaptation to a changing environment. Genomic data can help tackle these issues as they allow us to characterize and monitor genetic diversity through a wide array of

[1]LOEWE Centre for Translational Biodiversity Genomics, Senckenberg Biodiversity and Climate Research Centre, Georg-Voigt-Str. 14-16, 60325 Frankfurt/Main, Germany
[2]CE3C - Centre for Ecology, Evolution and Environmental Changes & CHANGE - Global Change and Sustainability Institute, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

emerging tools. The novel insights that can be obtained from genomic data have led to the formation of several national and international initiatives aiming to expand the genomic resources available for non-model species (Table S1 in the supplemental information online). At the same time, scientists and practitioners are collaborating to: (i) standardize protocols for detecting and monitoring species genetic diversity and their adaptive potential, and (ii) integrate genetic and evolutionary knowledge into conservation planning [8]. These actions are critical to promoting transboundary management to ensure the persistence of populations and species and, ultimately, the continued provision of nature-based ecosystem services. Recently, in a forum article [9] we argued for a new era in conservation genomics underpinned by **reference genomes** (see Glossary), also taking advantage of the increasing ease and widespread interest in generating them. Here, we expand this perspective in more detail, illustrated with recent examples. We also provide an overview on how genomics can help biodiversity conservation, which should be helpful to both conservationists who are not experts in genomics and to genomicists without a conservation biology background interested in engaging with conservation genomics research.

Genomic approaches in biodiversity research

High-throughput genomic sequencing technologies have lately evolved from mainly generating short (50–300 bp) to much longer (>10 000 bp) DNA sequencing reads. Genetic and genomic approaches commonly used in biodiversity research include DNA barcoding/metabarcoding, reduced representation DNA techniques, transcriptome sequencing (RNA-Seq), and whole-genome (re-)sequencing (Figure 1). Successful execution of these approaches depends on the quantity and quality of the available biological material, laboratory and bioinformatic skills, feasibility and costs, and on the quality and completeness of the available reference genome database. We review these common approaches, including their specific merits and limitations (Table 1). Since the quality of the starting biological material is a major consideration in applying genomic approaches, we also review approaches that use noninvasive sampling.

DNA barcoding and metabarcoding

DNA barcoding has become a standard, efficient genetic approach for species identification and biodiversity monitoring [10]. DNA barcoding sequences informative DNA loci with universal or taxon-specific primers that anneal to conserved flanking regions. Initially, DNA barcoding was based on the mitochondrial cytochrome *c* oxidase subunit I (COI) gene and mainly focused on animals, particularly invertebrates. Over the years, additional DNA loci have been utilized to barcode vertebrates (e.g., 12S, 16S, and *Cytb* in mtDNA), plants (*rbcL* and *matK* in cpDNA), fungi (ITS in rDNA), protists and nematodes (18S in rDNA), and bacteria (16S in rDNA) [10]. DNA metabarcoding combines the principles of DNA barcoding with next-generation sequencing (NGS), enabling the analysis of complex samples containing a mixture of specimens and/or species [11]. Metabarcoding has been widely used in biodiversity assessment and monitoring (e.g., species turnover during ecosystem restoration [12], mapping of ecological networks [13], or detection of invasive species [14]). A key advantage of metabarcoding is bulk sampling and sequencing, circumventing costly sorting and processing of samples into individual specimens (Table 1), thereby enabling high-throughput ecosystem-wide assessments and monitoring in most environments [15]. Moreover, metabarcoding is an appropriate approach to sequence environmental DNA, often degraded into short fragments in an environmental medium [16]. However, the short length of the DNA regions targeted in barcoding and metabarcoding can often limit accurate characterization of the genetic and taxonomic diversity in a community, failing to discern closely related taxa or taxa with introgressed nuclear genes or organellar genomes. **Genome skimming** circumvents some of the experimental biases in metabarcoding, potentially allowing more accurate metagenomic estimates of biodiversity [17,18] and wildlife forensic investigations [19]. Genome skimming has been conducted *in situ*, enabling rapid, field-based

[3]Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal
[4]The Rockefeller University, 1230 York Ave, New York, NY 10065, USA
[5]Naturalis Biodiversity Center, Darwinweg 2, 2333, CR, Leiden, The Netherlands
[6]Wellcome Sanger Institute, Tree of Life, Wellcome Genome Campus, Hinxton, CB10 1SA, UK
[7]NIVA - Norwegian Institute for Water Research, Økernveien, 94, 0579 Oslo, Norway
[8]Centre for Coastal Research, University of Agder, Gimlemoen 25j, 4630 Kristiansand, Norway
[9]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, PO BOX 1066 Blinderm, 0316 Oslo, Norway
[10]University of Göttingen, Department of Animal Evolution and Biodiversity, Untere Karspüle, 2, 37073, Göttingen, Germany
[11]Universita degli Studi di Milano, Via Celoria 26, 20133, Milan, Italy
[12]CIBIO/InBio, Centro de Investigação em Biodiversidade e Recursos Genéticos, Rua Padre Armando Quintas, 7, 4485-661, Portugal
[13]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal
[14]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[15]Department of Biosciences, University of Milan, Milan, Italy
[16]Estación Biológica de Doñana, CSIC, Calle Americo Vespucio 26, 41092, Sevillle, Spain
[17]Primate Genetics Laborator, German Primate Center, Kellnerweg 4, 37077, Göttingen, Germany
[18]InBios - Conservation Genetics Lab, University of Liege, Chemin de la Vallée 4, 4000, Liege, Belgium
[19]Groningen Institute of Evolutionary Life Sciences, University of Groningen, Nijenborgh, 9747, AG, Groningen, The Netherlands
[20]Center for Coastal Studies, 5 Holway Avenue, Provincetown, MA 02657, USA
[21]Marine and Freshwater Research Institute, Fornubúðir, 5,220, Hanafjörður, Iceland
[22]CIBER de Epidemiología y Salud Pública (CIBERESP), Spain
[23]Department of Biology, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Antwerp, Belgium
[24]Hunter College, City University of New York, NY, USA

assessments of non-model organisms (e.g., [20]) (Table 1). Use of genome skimming can also provide increased information content compared with barcoding and can be applied for sequencing of challenging specimens, such as museum specimens whose DNA may be too degraded and challenging for Sanger sequencing [21].

Reduced genomic representation

Reduced representation DNA sequencing (RRS) approaches are predominant in studies aimed at non-model species [22]. RRS approaches provide genome-wide data in large sample sizes at a comparably reduced cost (Table 1). A small fraction of the genome is reproducibly targeted in each specimen, either using **restriction endonucleases** and size selection (e.g., RADseq and related methods [23]) or captured by hybridization using probes (e.g., ultra-conserved elements or custom baits) or a combination of both (e.g., HyRAD [24], Rapture [25]). Although RRS approaches only capture a small portion of the genome, they provide sufficient genome-wide data to estimate genetic diversity, **inbreeding**, **effective population size**, population structure and assignment, gene flow, phylogeographic patterns, and phylogenetic relationships [22,26]. However, combining RRS-derived data from different studies necessitates identical experimental protocols, which limits replicability. Moreover, RRS approaches utilizing restriction endonucleases (and PCR) may be subject to **allelic dropout** [23]. Although RRS approaches can be conducted without a reference genome, alignment to a reference genome improves inferences obtained from RRS data [27]. Furthermore, a reference genome provides genome coordinates for most SNPs, thereby facilitating the identification of linked loci, which is key to many population genetic inferences [28].

Gene expression

Gene expression data (usually RNA-Seq) have given rise to a new conservation framework by characterizing genetic variation in natural populations through functional variation [29] and rapid responses of individuals or populations to environmental change [30]. Differences in gene expression have been linked to life history traits and population dynamics, aiding in the identification of candidate genes potentially affecting eco-evolutionary processes [31]. Gene expression data have provided insights into responses to pesticide exposure [32] and susceptibility or resistance to diseases [33]. Gene expression data have also been used to predict range shifts and identify vulnerable populations or adaptive phenotypes [34]. Further, gene expression data have been used to identify entire gene networks in the absence of prior knowledge of the genes involved [35] (Table 1). Analyzing gene expression data is more challenging than DNA-based analyses, because RNA is more susceptible to degradation than DNA and because transcription varies across cell and tissue types, sex, age, physiological and life stage, and even according to circadian rhythms. Consequently, gene expression studies require careful consideration of confounding biological factors [36]. As for RRS approaches, gene expression data are most informative when aligned against a reference genome [37], which provides an independent functional annotation to support the assignment of reads to genes [38,39].

In parallel with genomic assessments of DNA and RNA, epigenomics is being increasingly applied in biodiversity conservation research [40]. Epigenetic modifications may play a key role in phenotypic and biological innovations during major ecological transitions, thus potentially enabling rapid evolutionary responses [41,42]. Identifying the epigenetic modifications that modulate phenotypic variation can therefore potentially identify species that are vulnerable to environmental changes and be informative for monitoring the evolution of invasive species [43]. These approaches rely entirely on the availability of a reference genome to identify epigenetic signatures.

Whole-genome sequencing

Whole-genome sequencing (WGS) data offer unparalleled power and resolution in analyses of demographic history, **admixture** and **introgression**, recombination and **linkage disequilibrium**,

[25]University of Goettingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goettingen Center for Molecular Biosciences (GZMB), Campus Institute Data Science (CIDAS), Goldschmidtstr. 1, 37077, Goettingen, Germany
[26]Institute of Zoology, University of Cologne, Zülpicherstrasse 47b, D-50674, Cologne, Germany
[27]Evolutionary & Organismal Biology Research Center, Zhejiang University School of Medicine, Hangzhou, 310058, China
[28]Villum Center for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Denmark
[29]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China
[30]University of Florence, Department of Biology, Via Madonna del Piano 6, Sesto Fiorentino, (FI) 50019, Italy
[31]University of Lausanne, Department of Ecology and Evolution, Le Biophore, UNIL-Sorge, 1015 Lausanne, Switzerland
[32]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[33]Leibniz Institute for Zoo and Wildlife Research (IZW), Alfred-Kowalke-Str 17, 10315 Berlin, Germany
[34]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Koenigin-Luise-Str 6-8, 14195 Berlin, Germany
[35]Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75246, Uppsala, Sweden
[36]The authors are co-first authors.
[37]The ERGA Consortium.

*Correspondence:
jacob.hoglund@ebc.uu.se (J. Höglund).

genetic load, natural selection, and species diversification [44–47]. Elucidating evolutionary processes may require analyses of linkage disequilibrium between many physically distant loci [48]. WGS data enable detection of genomic regions under selection, mutations in regulatory elements, rare variants and **structural variation**, and investigating the genetic basis and architecture of phenotypic traits (e.g., disease susceptibility/resistance in endangered species) [49]. WGS data can also provide unique insights into phylogenetic relationships, evolutionary history, extinct and cryptic taxa, and ancestral intraspecific genetic diversity and structure, which are valuable in assessing temporal **genomic erosion** and guiding conservation and restoration efforts [50,51]. Since samples of extinct or endangered species usually are rare, WGS data maximize the genomic information obtained from each sample. The highly fragmented DNA in museum and subfossil specimens (often <100 bp) requires the availability of reference genomes for WGS read mapping and downstream evolutionary analyses [52].

A cost-efficient approach for population genomic studies is Pool-Seq (i.e., deep sequencing of multiple pooled individuals) [53] (Table 1). Pool-Seq has been applied to investigate genome-wide patterns of genetic diversity, signatures of adaptation, population differentiation, and genotype–environment associations [54–56]. In some cases, Pool-Seq is the only possible or practical approach, for example, with microscopic organisms. It facilitates surveys of multiple populations and conditions, with large numbers of individuals (>50) sampled in each, and can significantly reduce costs in species with a small to intermediate-sized genome (<1 Gb). Despite the cost-saving advantages, Pool-Seq entails some analytical and statistical challenges [53]. The robustness and accuracy of results directly depend on the pool size [53] and can be reduced in complex systems with patterns of weak population structure [57].

Noninvasive genomic sampling

Noninvasive or minimally invasive sampling of biological material (e.g., from feces, feather, or hair) [58,59] is commonly used in wildlife monitoring as the main source of genetic material to provide insights into the ecology of endangered species [60]. Noninvasive samples typically contain low amounts of poor-quality DNA, often contaminated with exogenous DNA [59]. The technical challenges specific to noninvasive samples have been addressed through approaches that maximize the amount of retrieved endogenous DNA, such as: (i) SNP genotyping arrays [61]; (ii) multiplex PCR amplicon sequencing [62]; or (iii) target sequence capture using RNA/DNA oligonucleotides [63], which may be developed from existing information obtained using RRS approaches or reference genomes generated from high-quality DNA specimens.

What can reference genomes bring to biodiversity conservation?

Since reference genomes are a prerequisite or enhance many of the analyses outlined earlier, their availability is desirable for any biodiversity genomics project [64]. Indeed, reference genomes are becoming commonplace across the entire tree of life. This is due to recent technological developments (Box 1) that enable the generation of accurate, chromosome-level genome assemblies for non-model organisms [9], and to multiple concerted international initiatives devoted to systematically generating reference genomes at scale (Table S1 in the supplemental information online). Next, we summarize how the availability of reference genomes substantially improves the use of genomics in biodiversity conservation.

The key to the whole trove of genomic information

Reference genomes provide the backbone and framework to map and annotate intra- and interspecific genomic variation [65–67]. They also provide a context to genetic variants, regions of low heterozygosity, patterns of linkage disequilibrium, and **site frequency spectra (SFS)** along chromosomes. In turn, these can be used to identify candidate genes associated with

Glossary

Accessory genome: portion of the pangenome of a species containing the genes that are not present in all lineages.
Admixture: production of new genetic combinations in hybrid populations through recombination.
Allelic dropout: experimental failure to detect allele(s) at a heterozygous locus.
Chromosome-level assembly: the process of generating a contiguous sequence of all chromosomes of a genome, often aided by genetic maps or proximity ligation techniques (3C-seq, Hi-C); term also used to refer to the resulting genome sequence.
Core genome: portion of the pangenome of a species containing the genes present in all lineages.
***De novo* assembly:** the process of generating a genome sequence from individual sequencing reads without the use of an existing reference; the term is also used to refer to the resulting genome sequence.
Ecotypes: intraspecific phenotypes adapted to different ecological niches.
Effective population size: N_e; key parameter that reflects the rate of genetic drift and inbreeding; the size of an idealized population with the observed rate of genetic drift.
Genetic load: actual or potential reduction in mean population fitness due to genetic causes (deleterious mutations, genetic drift, inbreeding, migration, recombination).
Genome skimming: low-depth whole-genome sequencing typically used to recover high-copy DNA regions, such as mitochondrial and chloroplast DNA and repetitive nuclear loci (e.g., satellite DNA, **transposable elements**).
Genomic erosion: loss and degradation of genome-wide diversity due to inbreeding, genetic drift, introgression, and selection in small or isolated populations.
Hybridization: interbreeding of individuals from genetically distinct populations.
Inbreeding: production of offspring from the mating of closely related individuals. Inferred more formally from the inbreeding coefficient (F), which denotes the probability that two gene copies at a locus are identical by descent.
Inbreeding depression: reduced fitness in offspring as a result of inbreeding.

phenotypes and adaptations and elucidate complex genomic features such as satellite DNA, gene families, segmental duplications, or structural variation. Structural variation, such as large inversions that can limit recombination, may be linked to adaptation to environmental change and may have important monitoring and management implications (Box 2: Case study 1 on the Atlantic cod) [68]. Chromosomal-level genome assemblies improve detection of structural variants of different sizes, including some spanning megabases across multiple gene families (Box 2: Case study 2 on the horseshoe crabs). Reference genomes can reveal integration, recombination, or deletion of **mobile elements**, which may play a role in genomic and taxonomic diversification [69]. Reference genomes also enable the rapid and efficient *in silico* development of custom SNP or STR assays targeting specific functional variants of interest, which facilitates the analysis of noninvasive, forensic or museum DNA samples [70].

Functional and adaptive genetic variation

Reference genomes facilitate the identification of functional genetic variation underlying phenotypic traits, fitness variation, and adaptive potential [71–73]. Such analyses are mostly successful at identifying variants with large effects [71,74,75], likely representing a small proportion of the adaptive functional variation. However, large-effect variants can underlie traits that are responsible for resilience to anthropogenic threats such as diseases or climate change. Such traits, identified, for example, in genome-wide association studies (GWAS), can be selected for in captive breeding or translocation programs (Box 2: Case study 3 on the ash dieback). Those programs must then carefully consider issues such as maintaining genome-wide diversity and evolutionary potential [76–78] and genetic adaptation to captivity [79]. Most phenotypic traits are polygenic (i.e., under the control of multiple interacting loci with small effects). These loci are, in turn, likely subject to gene regulatory networks difficult to fathom with DNA-based tests aimed at detecting adaptive variation at individual loci. To this end, statistical tools have been developed to detect contributions from multiple loci across the entire genome that may underlie the function and heritability of trait variation [80], as well as more direct approaches such as massively parallel reporter assays [81]. Reference genomes facilitate the estimation of the relative fitness of alleles and genotypes in threatened species [82], which may not be amenable to experimental approaches, for example, by WGS comparisons of individuals with different phenotypes [83]. Notwithstanding the analytical challenges, WGS data, combined with phenotypic and fitness data, enable a thorough understanding of adaptive variants [74] that drive responses (or the lack thereof) to current and future anthropogenic threats [84], such as ongoing environmental change [85]. Incorporating insights of adaptive potential into species distribution models helps modeling climate change effects on endangered populations [86,87], such as genotype–phenotype associations underlying species resilience, which can guide assisted breeding programs [88] (Box 2: Case study 4 on the European beech). Identifying adaptive variation may also aid in defining conservation units that ensure the preservation of evolutionary heritage and adaptive potential [89].

Inbreeding and genetic load

Assessing detrimental fitness effects associated with inbreeding (**inbreeding depression**) is a central topic in conservation genetics [90,91]. The availability of annotated reference genomes is invaluable for characterizing inbreeding and the genetic architecture of inbreeding depression, in particular the number of loci involved, their effect size, and the contributions of deleterious recessive alleles [92,93]. This knowledge can guide the selection of founders, breeders, and individuals for reintroduction in *ex situ* conservation management and help design purging strategies, or inform monitoring of deleterious variants and genomic erosion in the wild [94]. **Runs of homozygosity (ROH)** analyses are useful for distinguishing historical and contemporary inbreeding, pinpointing candidate loci contributing to inbreeding depression (particularly loci with large effects), and selecting captive breeders with different ROH to minimize inbreeding

Introgression: gene flow between hybridizing populations or species by backcrossing of hybrids with one or both parental species.
Linkage disequilibrium: nonrandom association between alleles at different loci.
Mobile elements: DNA sequences that are able to integrate into new sites within a genome, sometimes also transferred among species. They are one of the major drivers of genome evolution.
Outbreeding: mating between individuals from genetically distinct lineages (e.g., populations, subspecies, or species).
Outbreeding depression: reduced fitness of offspring from matings between genetically divergent individuals.
Pangenome: the entire set of DNA sequences (or genes) of a species represented by the core genome and the accessory genome.
Reference genome: contiguous and accurate genome assembly representative of a species with coordinates of genes and other important features annotated.
Restriction endonuclease: an enzyme produced by certain bacteria that has the property of cleaving DNA molecules at or near a specific sequence motif.
Runs of homozygosity (ROH): genome regions that are homozygous due to identity-by-descent. ROH arise when two copies of an ancestral haplotype are brought together in the same individual. The number and size of ROHs are the result of various processes, in particular nonrandom mating, inbreeding, and population demographic history.
Site frequency spectrum (SFS): distribution of allele frequencies of a given set of loci (often SNPs) in a population sample (aka 'allele frequency spectrum').
Structural variation: regions of a chromosome presenting structural changes such as insertion, deletion, inversion, or translocation of DNA.
Transposable elements: see 'Mobile elements'.

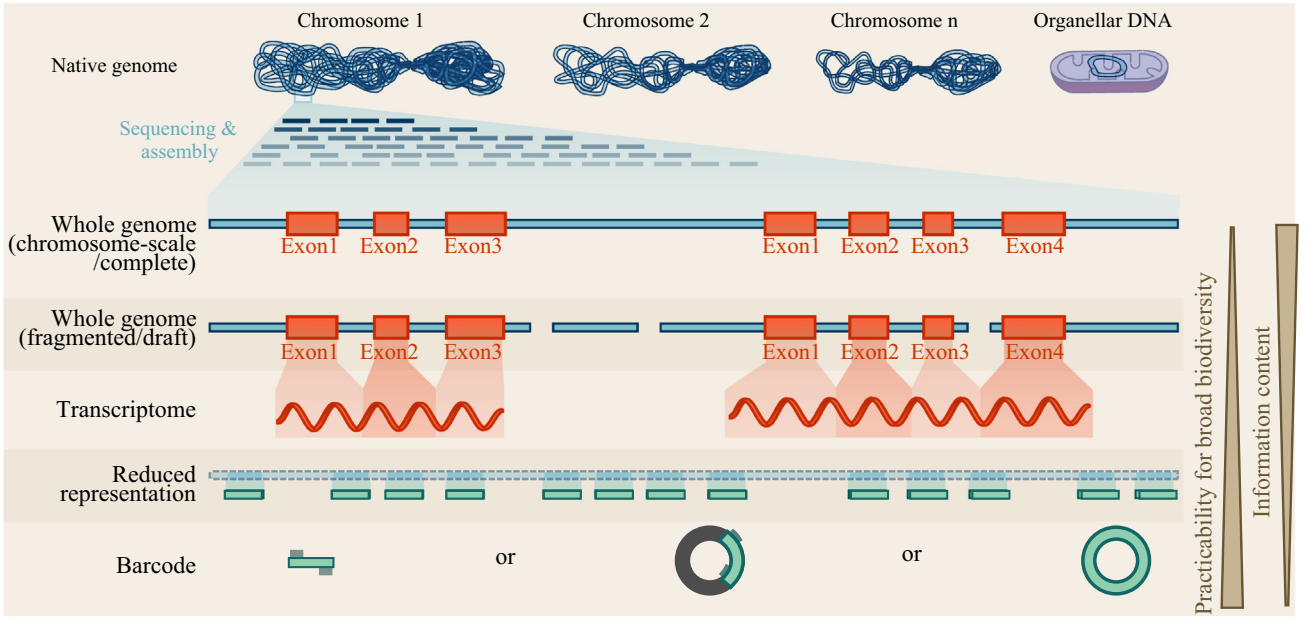

Figure 1. Genomic approaches for biodiversity research. In a eukaryotic cell, DNA is natively packed and allocated over different chromosomes, ranging from a few to, more rarely, hundreds in the nucleus, as well as over organellar genomes (mitochondrial DNA and, in plants and algae, also plastid DNA), making up the genome. Current technology cannot sequence entire chromosomes longer than the read lengths (usually 10–100 kbp). Such reads yield short pieces of sequence information. The shortest pieces used for biodiversity research are the amplification and/or selective sequencing of specific regions of the genomes: multiple loci ('multilocus') or single marker regions ('barcodes'). Reduced representation samples low copy regions of the DNA. Transcriptomic data, in particular RNA-Seq at a genome-wide scale, can be assembled *de novo* or mapped onto a genomic backbone. Finally, reads are assembled into highly contiguous chromosomes (chromosome scale/complete) or less contiguous (fragmented/draft) pseudomolecules. On the right, arrows give an indication of (i) the ease of covering a broad taxonomic diversity with a given technique, and (ii) the amount of biodiversity or population information produced by a technique.

depression [90,91,95]. Reference genomes facilitate predicting the putative functional significance and fitness consequences of specific allelic variants that have drifted to high frequency or fixation in small wild populations and estimating the accumulated genetic load and dynamics of deleterious variation [51,96–98]. Such in-depth reference genome-based analyses are being carried out in an increasing number of endangered species (e.g., Box 2: Case study 5 on the Iberian lynx, the crested ibis [99], and the Sumatran rhinoceros [92]). In perspective, reference genomes combined with WGS data provided invaluable insights into the long-term management of translocated populations (Box 2: Case study 6 on the Florida panther).

Outbreeding and hybridization

Outbreeding can also result in a reduction of fitness, known as **outbreeding depression**, due to chromosomal or genic incompatibilities, epistatic interactions, disruption of interactions between coadapted genes, or introduction of variants that are maladaptive to local environmental conditions. In risk assessments of outbreeding depression, WGS and reference genomes can be invaluable in enabling detailed analyses of adaptive divergence and structural variation across the genome [46,82]. This knowledge is important for reducing the chances of outbreeding depression when selecting source populations and individuals for translocations to reinforce a population threatened with extinction [100]. Human-mediated **hybridization** and introgression have increased dramatically worldwide. In association with the perturbation and homogenization of environments, hybridization can become a major conservation concern, threatening biodiversity and evolutionary heritage. Genome-wide data make it possible to characterize introgression patterns, dynamics, and admixture proportions more effectively than traditionally possible with

Table 1. List of genomic approaches with application and comparison of raw sequencing costs (i.e., costs of sample collection, researcher time, and analysis are not included)

	DNA barcoding/ metabarcoding	Genome skimming	Reduced representation DNA sequencing	Transcriptome sequencing	Whole-genome resequencing
What genome do you get?	None	Organelle, *k*-mer representation of nuclear	None	Coding regions only, variable fragmentation	Nonrepetitive genome, depends on coverage
Cost in dollars (as of date)[a]	$5 per sample (Sanger/NGS)	$50	$50	$100–$400	$100–$600
What type of samples are needed	Fresh tissue samples, museum specimens, noninvasive samples	Fresh tissue samples, museum specimens	Fresh tissue samples, museum specimens, noninvasive samples	Tissue-specific, live/fresh, flash frozen/in RNA buffer	Fresh tissue samples, museum specimens
Genetic diversity[b]	Yes, but limited	Yes, but limited	Yes	Yes	Yes
Population structure[c]	Yes, but weak to detect shallow/cryptic genetic structure; economical for detailed spatial sampling	Yes, typically organelle based	Yes	Yes	Yes
Phylogenetic information	Yes, but barcode based	Yes, typically organelle based	Yes	Yes	Yes
Introgression event	No	No	Yes, but no individual genes	Yes, but limited detection power	Yes
QTL mapping	No	No	Yes, but low resolution	Yes, expression QTL (eQTL)	Yes
Natural selection signal detection	No	Yes, on organelle genes	Yes	Yes	Yes
Gene structure study	No	Yes, on organelle genes	Potentially, if reference genome available	No	Yes, if reference genome available
Gene family analyses	No	Yes, on organelle genes	Potentially, if reference genome available	Yes	Yes, if reference genome available
Genome rearrangement study	No	Yes, on organelle	Potentially, if reference genome available	No	Yes, if reference genome available
Functional genomic study[d]	No	Yes, on organelle genes	No	Yes	No
Genome size estimation	No	Yes, typically organelle genome	No	No	Depending on coverage
Linkage disequilibrium	No	No	Yes, usually, not always	Limited	Yes
Demographic reconstructions (MSMC)[e]	No	No	No	No	Yes
Demographic reconstructions from SFS	No	No	Yes	No	Yes
GWAS	No	No	Yes, but low resolution	Yes, T(ranscriptome) WAS	Yes

Whole-genome sequencing			
	Short reads	Linked reads (10X/stLFR/Tell-Seq)	Long reads + scaffolding technologies
What genome do you get?	Very fragmented, contig N50 <10 kb	Contig N50 >200 kb, scaffold N50 >10 Mb	Reference genome
Cost in dollars (as of date)[a]	$100–$600	$2000	$5000
What type of samples are needed	Fresh tissue samples, museum specimens	Frozen upon collection	Large quantity of high molecular weight DNA

(continued on next page)

Table 1. (continued)

Whole-genome sequencing			
	Short reads	Linked reads (10X/stLFR/Tell-Seq)	Long reads + scaffolding technologies
Genetic diversity[b]	Yes	Yes	Yes
Population structure[c]	Yes	Yes	Yes
Phylogenetic information	Yes	Yes	Yes
Introgression event	Yes, but limited detection power	Yes	Yes
QTL mapping	Yes, but low resolution	Yes	Yes
Natural selection signal detection	Partially, if reference genome available	Yes	Yes
Gene structure study	Partially, if reference genome available	Yes, miss few regions	Yes
Gene family analyses	No	Yes	Yes
Genome rearrangement study	No	Yes, but fragmented karyotype	Yes, can trace karyotype evolution
Functional genomic study[d]	Yes, but missing many regions	Yes, but missing some regions	Yes
Linkage disequilibrium	Yes	Yes	Yes
Demographic reconstructions (MSMC)	Yes	Yes	Yes
Demographic reconstructions from SFS	Yes	Yes	Yes
GWAS	Yes	Yes	Yes

[a]Based on a 2 Gb vertebrate genome.
[b]Noting that single genomes deliver estimates from two haplotypes only.
[c]Noting that single specimens do not deliver to this aim well.
[d]Integrated with ChIP-seq, Hi-C, RNAseq for regulatory element annotation.
[e]Multiple sequentially Markovian coalescent.

only dozens of markers [101]. Introgression rates in different genomic regions can vary. Whole-genome data mapped to reference genomes are useful to identify admixture (or migrant) tracts along individual genomes and thus to understand local effects of introgression and the age of admixture events [102,103].

Genome editing: engineering adaptation, gene drives, and de-extinction

Genome-editing tools, such as CRISPR-Cas9 [104], allow precise modification of genes and genomes of living organisms. Genome editing can be applied as a tool to mediate locus-specific genetic rescue in endangered species threatened with high frequencies of deleterious

Box 1. Genomes, population genomics, and pangenomes

Reference genomes, point representations of the structure and organization of a species' genome, are expected to revolutionize conservation genomics [9]. Currently, a combination of single-molecule long-read sequencing (SMRT or nanopore sequencing) or linked reads (e.g., TELL-seq or stLFR) for contigging, optical maps, and/or proximity ligation reads (e.g., 3C-seq or Hi-C) for scaffolding appears to be a general strategy capable of generating phased chromosome-scale reference genomes across the tree of life [124,125] (see Table 1 in main text). When sufficient high molecular weight DNA for long-read sequencing remains problematic to obtain (e.g., with small organisms, albeit solutions are rapidly becoming available [126]), an alternative cost-effective strategy is represented by using linked reads only [127]. To gain more accurate insights into a species' genomic diversity, the genomes of large numbers of individuals can be resequenced. Such multiple genomes do not have to be assembled *de novo* (***de novo* assembly**) and may be derived from cheaper and quicker short-read sequencing and subsequent mapping to an existing assembly [128]. Additionally, the availability of multiple reference genomes per species provides a more complete genome characterization [129,130] and allows estimating a species **pangenome** [131]. Individuals of a species are part of a pangenomic architecture, sharing a larger proportion of their genes as a core set, whilst differing in a smaller proportion of dispensable genes [132]. The analysis of such data helps to differentiate adaptive and neutral evolutionary processes, to define populations, **ecotypes**, and species, and identify **core** and **accessory genomic** regions, gene flow, hybridization, and incomplete lineage sorting [133].

Box 2. Case studies

1. Atlantic cod

The Atlantic cod (*Gadus morhua*; Figure IA) comprises several cryptic ecotypes differing in feeding, spawning, and migratory behaviors and, consequently, in population dynamics relevant for fisheries management [134]. The Atlantic cod was one of the first non-model organisms with a chromosome-anchored genome assembly [135]. Combined with population genomic analyses, the genome assembly enabled identification of four large chromosomal inversions that discriminate between the migratory and nonmigratory ecotypes in Norwegian and Icelandic waters and are associated with adaptation to environmental conditions [134,136]. Each of the four inversion regions contains hundreds of genes with a high level of divergence between the inverted and noninverted haplotypes. The four inversions originated independently up to 1.7 million years ago and have been maintained due to strong selection [136,137].

2. Horseshoe crabs

Horseshoe crabs (Figure IB) belong to the chelicerate order Xiphosura. This enigmatic order comprises only four extant species (*Limulus polyphemus*, *Tachypleus gigas*, *Tachypleus tridentatus*, and *Carcinoscorpius rotundicauda*), all considered 'living fossils' since their morphology is virtually identical to that of their Triassic ancestors. Recent studies have generated **chromosome-level assemblies** of three of these species, and analyses suggest three rounds of whole-genome duplication [138,139]. Several gene families, particularly those involved in innate immunity, have undergone extensive tandem duplication. These expanded gene families may be important components of the innate immune system of horseshoe crabs, whose amebocytes are presently exploited for detecting endotoxin contamination [139]. These genomic resources are of value for breeding programs and conservation [138].

3. Ash dieback

In the mid-1990s, European ash trees (*Fraxinus excelsior*; Figure IC) started dying across Europe from a new disease, ash dieback, which spread rapidly and was determined to be caused by *Hymenoscyphus fraxineus*, a fungal pathogen new to Europe. Genome sequencing of this fungus along with isolates from East Asia confirmed that the European outbreak was likely caused by the introduction of imported Asian *Fraxinus mandshurica.* Importantly, the fungal genomes showed that at most two independent introductions of *H. fraxineus* had occurred [140]. The availability of the fungal genomes will aid surveillance for new introductions. While no ash trees appeared to be fully resistant, threatening extinction of this iconic species, analysis of Danish forestry germplasm identified individuals that had reduced susceptibility. An international consortium sequenced the European ash genome [141] and used it to map polygenic traits associated with reduced susceptibility [142], promising to assist in directed breeding to restock ash forests across the continent.

4. European beech

In the summers of 2018 and 2019, around two-thirds of European beech trees (*Fagus sylvatica*; Figure ID) were damaged or killed by extreme drought. It is critical to keep beech woods healthy, as they are home to over 6000 other species of animals and plants. However, not all the trees in each forest responded in the same way, with severely damaged trees often sitting less than 5 m from fully healthy ones. This suggests that the genetic makeup of a tree, rather than its local environment, determines how well it copes with drought. After the beech genome was assembled [143], over 400 European beech trees from pairs of neighboring trees that had responded differently to the droughts were analyzed [88]. More than 80 regions of the genome differed between healthy and damaged trees. Subsequently, a genetic test was developed which can quickly and inexpensively predict how well an individual beech tree might cope with drought. Now this genetic test can be used to select and reproduce trees that are better adapted to drought.

5. Iberian lynx

One of the first whole genomes sequenced for a highly endangered species was that of the Iberian lynx (*Lynx pardinus*; Figure IE). Whole-genome resequencing analyses revealed that the species underwent a history of serial population declines, leading to one of the lowest genomic diversities ever reported [65]. Historical genetic data allowed the quantification of recent diversity loss and validated the mixing of the two differentiated remnant populations [144]. Subsequent genome-wide studies focused on functional variation have shown a higher genetic load in the more bottlenecked lynx populations [145], but also the purging of highly deleterious variants in Iberian lynx with respect to the more abundant Eurasian lynx [146]. These studies have also generated a catalog of potentially deleterious variation that can now be tested for association with reduced fitness and genetic diseases. The access to genome-wide variation also allowed the selection of a set of highly efficient and informative SNPs [147], which is now being extensively applied in combination with noninvasive samples to monitor and manage genetic diversity across *ex situ*, remnant, and reintroduced populations. These concerted conservation efforts aided by genetic information succeeded in avoiding the imminent extinction and in reverting the negative population trends of the species by the turn of the century.

6. Florida panther

In the early 1990s, isolated Florida panthers, a population of puma (*Puma concolor*; Figure IF) in the Big Cypress National Preserve, showed low reproductive success and multiple signs of inbreeding. Therefore, pumas from Texas were translocated to support the isolated Florida population. Genetic diversity and reproductive success increased and the consequences of severe inbreeding, such as undescended testicles, heart failures, cowlick, and kinked tails, vanished [148]. A recent study explored the genomic consequences of the long-term isolation of this and other puma populations using a reference genome and a geographically broad panel of resequenced individuals [95]. The genome of a Florida panther descended from translocated Central American individuals had long tracts of homozygosity despite recent outbreeding. This suggests that sustaining diversity in small and isolated populations will require either repeated translocations or restoration of landscape connectivity.

7. Freshwater sponge

Relatively few chromosome-level reference genomes are available for non-model invertebrate species, which hampers the implementation of conservation or biodiversity genomics approaches in a group of species that spans more than 95% of the animal tree of life. The freshwater sponge *Ephydatia muelleri* (Figure IG) is a common species in Palearctic lakes and freshwater streams that can even colonize drinking water systems, is transported by migratory birds and plays important ecological roles where it

grows. Sponges provide fundamental ecosystem services and can serve as biomonitors for pollution and natural samplers for biodiversity [149]. The recent publication of the chromosome-level genome of *E. muelleri* [150] has opened new avenues to investigate selection and adaptation in sponges, as well as enabled accurate SNP calling to assess genetic diversity, demographic events, and gene flow. This important resource will revolutionize the way we approach conservation in this group of non-model invertebrates and will aid in assessing and monitoring their conservation status, which is an inherently difficult task for poorly studied invertebrates.

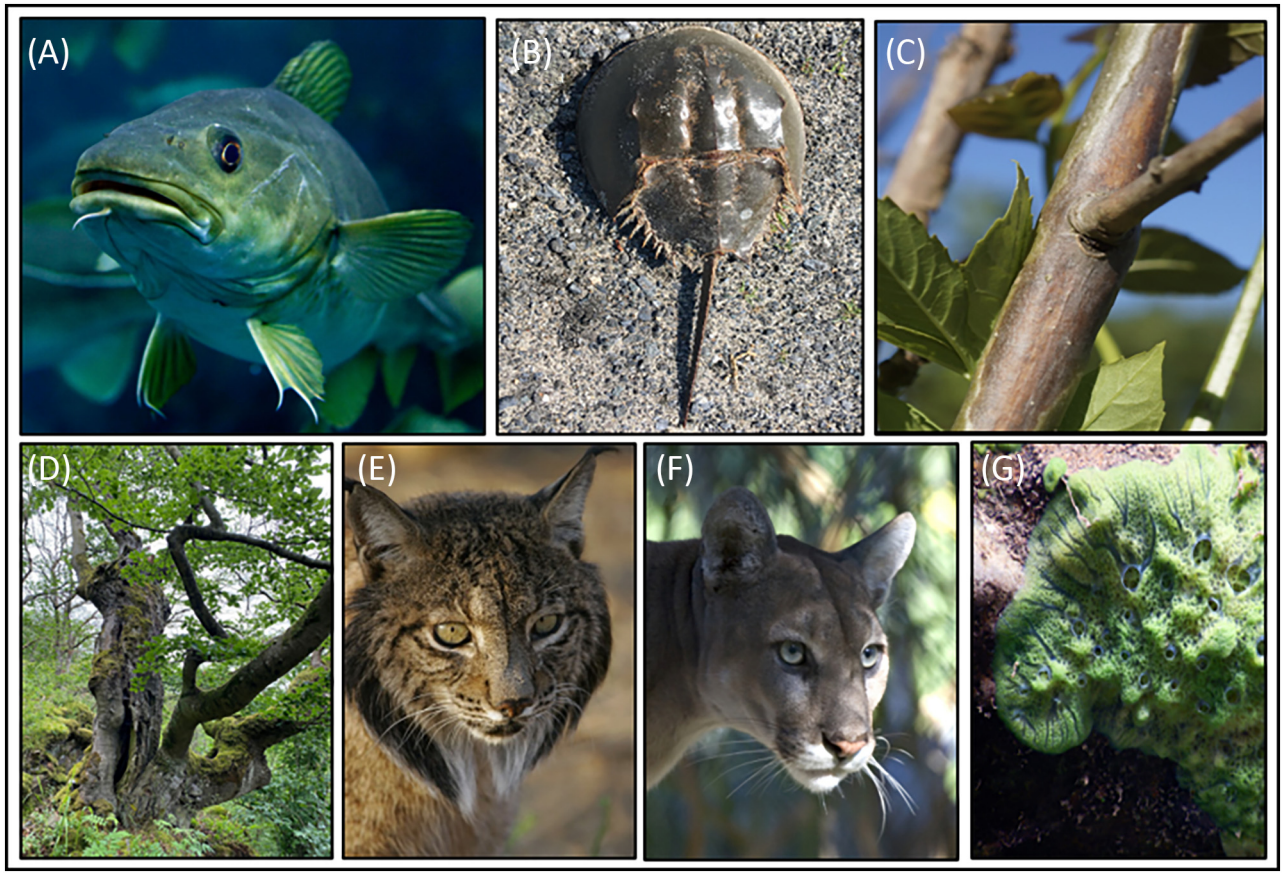

Figure I. Images of case study species. (A) The Atlantic cod (*Gadus morhua*) provided by WaterFrame/Alamy Stock Photo. (B) The Atlantic Horsehoe crab (*Limulus polyphemus*). (C) The European ash (*Fraxinus excelsior*). (D) The European beech (*Fagus sylvatica*). (E) The Iberian lynx (*Lynx pardinus*). (F) The Florida panther (*Puma concolor*). (G) The Mueller's freshwater sponge (*Ephydatia muelleri*).

mutations or to increase resistance to infectious diseases and resilience to anthropogenic environmental change [105]. Modifications could even involve gene drives to assist the spread of deleterious mutations through invasive populations [106,107], although this is still controversial due to concerns of possible unintended outcomes [108,109] and lack of an international regulatory framework for safe and responsible use [110]. Reference genomes are necessary for both the editing and gene drive approaches. An even more technically challenging prospect is resurrecting extinct species, or more accurately, functional proxies of extinct species [111]. Species driven to extinction by humans may be ethically justifiable candidates for de-extinction, especially keystone species for extant ecosystems with beneficial cascading effects [112]. De-extinction is controversial since the causes of extinction may still be present [113] or because they could divert funding and public attention from ongoing, critical conservation

efforts of threatened species [114]. Cloning via somatic cell nuclear transfer (SCNT) could allow de-extinction of recently extinct mammal species where well-preserved tissues have been cryopreserved [111]. Likewise, SCNT can be used to resurrect extinct genetic variation to facilitate genetic rescue in highly inbred, endangered species [115]. Obviously, for the vast majority of species driven extinct by humans, what remains at most is degraded and fragmented DNA. In such cases, reference genomes from extant relatives are crucial for attempts at reconstructing a genome sequence of the extinct species from short-read WGS data [111].

Beyond single species: structure and function of communities

There is increasing interest in genomics tools usable at the community level [116]. Reference genomes are particularly helpful in metagenomic and metatranscriptomic analyses, where the DNA or RNA from an entire community sample is sequenced. Metagenomics targets community composition and its functional potential [117], whereas metatranscriptomics provides a temporal snapshot of community activity through gene expression [118]. Both approaches depend on the availability of annotated reference genomes. Metagenomics and metatranscriptomics are frequently applied to microbial community samples, benefitting from a more extensive reference genome availability, in particular for known bacteria, archaea, and fungi [119]. Indeed, in the absence of reference genomes, most non-bacterial sequences remain unidentified. The accuracy and information content of the genomes is highly relevant for metagenomic assignment [120]. Reference genomes will allow a more complete characterization of communities when applying metagenomic or metatranscriptomic approaches [121], facilitating monitoring and managing taxonomic and functional diversity in entire ecosystems.

Concluding remarks

Accelerating global biodiversity loss due to rapid environmental change and other anthropogenic impacts drives the need for urgent conservation efforts. Genome assemblies provide a fundamental framework to interpret and protect biodiversity. While many questions remain (see Outstanding questions), we have reviewed how biodiversity conservation can benefit from genomic data built upon reference genomes. We presented case studies that, among many others,

Box 3. Connecting genomic initiatives: from genomes to conservation actions

Genomics can help address main ecological questions that are central to understanding and maintaining biodiversity and ecosystem services. In Europe, in recent years there have been three main political initiatives to support biodiversity research and conservation: the Biodiversity Strategy for 2030, the European Partnership on Biodiversity, and the European Green Deal. These initiatives all advocate for the importance of genetics in ecosystem preservation and restoration. To address this, several initiatives have started to bring genomics applications for biodiversity to a European perspective. Notable examples of such initiatives are the European Reference Genome Atlas (ERGA) and the Genomic Biodiversity Knowledge for resilient Ecosystems (G-BiKE). ERGA is a consortium involving nearly 50 EU countries and international partners, recently funded through a Horizon Europe call on Biodiversity and Ecosystem Services, with the aim of generating high-quality, nearly error-free reference genomes representing the genetic makeup of European eukaryotic biodiversity, ranging from endangered species, species of importance for ecosystem function and stability, to key species for agriculture, forestry, and fisheries, but also non-model organisms from under-represented taxa that make up a huge proportion of biodiversity. Besides a diverse yet coordinated production pipeline, strong components of ERGA are the application of reference genomes to conservation genomics endeavors, ethical, legal, and social implications on the use of genomic resources, as well as citizen science initiatives. G-BiKE is a network funded by the European Cooperation in Science and Technology (COST) program involving more than 110 researchers and practitioners from 39 European countries with the aim of establishing the use of genomic data as a standard tool for monitoring and managing wild and *ex situ* populations of plants and animals. This is achieved by organizing workshops, training schools, short-term scientific missions, and virtual mobility grants. The ultimate goal of G-BiKE is to integrate genetic diversity monitoring into EU policy and planning on biodiversity conservation, including the missions of the European Green Deal, the Biodiversity Strategy for 2030, the Natura 2000 network, and Habitats Directive. Recently, G-BiKE has enlarged its scope by playing an active role in the process leading to the post-2020 Global Biodiversity Framework of the Convention on Biological Diversity. A joint venture between ERGA and G-BiKE is anticipated to advocate for the incorporation of genomic data into European biodiversity and ecosystem services protection programs.

Outstanding questions

A considerable fraction of biodiversity is endangered, but some taxa are more accessible and studied than others. How do we obtain high-quality DNA and generate genomic information from very rare and small (e.g., flies, meiobiota, gastrotrichs) or challenging taxa (e.g., conifers due to large genome size, or uncultured fungi)?

How can we promote both taxonomic and geographic coverage to contribute to the global endeavor of a balanced conservation of biodiversity?

We are now able to estimate many genomic aspects relevant to conservation, but we still need to integrate them with geographical and environmental data. How do we combine these factors and translate them into adaptive potential and extinction risk in a predictive quantitative manner that can inform conservation efforts?

While various efforts already exist to address ethical, legal, and social issues, there are still legitimate concerns about the rights of local human communities over biodiversity resources. How can we provide open access to reference genomes, while also ensuring benefit sharing for those communities?

Biodiversity protection and climate change mitigation are strongly co-dependent. How can biodiversity research be directed towards sustainable generation and analysis of genomic data?

Compiling reference genomes and 'omics data (e.g., transcriptomic, epigenomic) for endangered species creates 'digital Noah's arks' that complement the more fragile and perishable frozen zoos. How can we create momentum for the concerted construction of these multi-omics digital arks that may be vital for de-extinction efforts in the future?

Reference genomes and genome editing will increasingly allow identification and modification of genomic regions to increase fitness, persistence, and adaptive potential of endangered species. How can we

demonstrate how genomic resources enrich and enhance integrated conservation efforts. Biodiversity conservation must explicitly consider genomic diversity to optimize strategies that minimize the loss of fitness, thus aiding in maintaining population viability, and to preserve evolutionary potential for adaptive responses to environmental change and diversification in general. To this end, we advocate for promoting reference genome-based approaches in conservation research and encourage knowledge transfer between the research community, conservation practitioners, and society (Box 3) [122]. As the Human Genome Project strongly impacted the biomedical sciences over the past two decades [123], we anticipate that the availability of reference genomes across the tree of life will provide a solid, quantitative, and comparable foundation for biodiversity assessments, conservation, and restoration.

strike the right balance in including genome editing in the conservation toolbox?

Acknowledgments

We thank Cecilia Helmerson, Antonio Rivas, Nelson Lauzon, and Sally Leys, for the pictures of the Atlantic cod, the Iberian lynx, the European beech, and the freshwater sponge, respectively. We also thank Fabien Condamine, Love Dalén, Richard Durbin, Bruno Fosso, Roderic Guigó, Marc Hanikenne, Alberto Pallavicini, Olga Vinnere Pettersson, Xavier Turon, and Detlef Weigel for their contributions to the manuscript, as well as the whole ERGA community for making this possible.

Declaration of interests

No interests are declared.

Supplemental information

Supplemental information associated with this article can be found online https://doi.org/10.1016/j.tig.2023.01.005.

References

1. Ceballos, G. *et al.* (2020) Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proc. Natl. Acad. Sci. U. S. A.* 117, 13596–13602
2. Cowie, R.H. *et al.* (2022) The Sixth Mass Extinction: fact, fiction or speculation? *Biol. Rev. Camb. Philos. Soc.* 97, 640–663
3. Stange, M. *et al.* (2020) The importance of genomic variation for biodiversity, ecosystems and people. *Nat. Rev. Genet.* 22, 89–105
4. Shafer, A.B.A. *et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* 30, 78–87
5. Wootton, J.T. and Pfister, C.A. (2013) Experimental separation of genetic and demographic factors on extinction risk in wild populations. *Ecology* 94, 2117–2123
6. Taylor, H.R. *et al.* (2017) Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* 10, 231–242
7. Hoban, S. *et al.* (2021) Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: current actions and indicators are insufficient. *Biol. Conserv.* 261, 109233
8. Nielsen, E.S. *et al.* (2022) Molecular ecology meets systematic conservation planning. *Trends Ecol. Evol.* 38, 143–155
9. Formenti, G. *et al.* (2022) The era of reference genomes in conservation genomics. *Trends Ecol. Evol.* 37, 197–202
10. Kress, W.J. *et al.* (2015) DNA barcodes for ecology, evolution, and conservation. *Trends Ecol. Evol.* 30, 25–35
11. Deiner, K. *et al.* (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895
12. Theissinger, K. *et al.* (2019) Mosquito control actions affect chironomid diversity in temporary wetlands of the Upper Rhine Valley. *Mol. Ecol.* 28, 4300–4316
13. Cuff, J.P. *et al.* (2022) Overcoming the pitfalls of merging dietary metabarcoding into ecological networks. *Methods Ecol. Evol.* 13, 545–559
14. Klymus, K.E. *et al.* (2017) Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS One* 12, e0177643
15. Arribas, P. *et al.* (2021) Connecting high-throughput biodiversity inventories: opportunities for a site-based genomic framework for global integration and synthesis. *Mol. Ecol.* 30, 1120–1135
16. Ruppert, K.M. *et al.* (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* 17, e00547
17. Dodsworth, S. (2015) Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* 20, 525–527
18. Bista, I. *et al.* (2018) Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. *Mol. Ecol. Resour.* 18, 1020–1034
19. Johri, S. *et al.* (2019) "Genome skimming" with the MinION hand-held sequencer identifies CITES-listed shark species in India's exports market. *Sci. Rep.* 9, 4476
20. Franco-Sierra, N.D. and Díaz-Nieto, J.F. (2020) Rapid mitochondrial genome sequencing based on Oxford Nanopore Sequencing and a proxy for vertebrate species identification. *Ecol. Evol.* 10, 3544–3560
21. Raxworthy, C.J. and Smith, B.T. (2021) Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.* 36, 1049–1060
22. Hohenlohe, P.A. *et al.* (2021) Population genomics for wildlife conservation and management. *Mol. Ecol.* 30, 62–82
23. Andrews, K.R. *et al.* (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92
24. Suchan, T. *et al.* (2016) Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One* 11, e0151651
25. Ali, O.A. *et al.* (2016) RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202, 389–400

26. McCormack, J.E. *et al.* (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66, 526–538
27. Rochette, N.C. *et al.* (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28, 4737–4754
28. Baetscher, D.S. *et al.* (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Mol. Ecol. Resour.* 18, 296–305
29. Alvarez, M. *et al.* (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol. Ecol.* 24, 710–725
30. Oomen, R.A. and Hutchings, J.A. (2017) Transcriptomic responses to environmental change in fishes: insights from RNA sequencing. *Facets* 2, 610–641
31. Green, D.A., II and Kronforst, M.R. (2019) Monarch butterflies use an environmentally sensitive, internal timer to control overwintering dynamics. *Mol. Ecol.* 28, 3642–3655
32. Fraser, D. *et al.* (2018) Genome-wide expression reveals multiple systemic effects associated with detection of anticoagulant poisons in bobcats (*Lynx rufus*). *Mol. Ecol.* 27, 1170–1187
33. Campbell, L.J. *et al.* (2018) A novel approach to wildlife transcriptomics provides evidence of disease-mediated differential expression and changes to the microbiome of amphibian populations. *Mol. Ecol.* 27, 1413–1427
34. Trego, M.L. *et al.* (2019) Tracking transcriptomic responses to endogenous and exogenous variation in cetaceans in the Southern California Bight. *Conserv. Physiol.* 7, coz018
35. Bianco, L. *et al.* (2017) Pathway Inspector: a pathway based web application for RNAseq analysis of model and non-model organisms. *Bioinformatics* 33, 453–455
36. Wang, B. *et al.* (2019) Reviving the transcriptome studies: an insight into the emergence of single-molecule transcriptome sequencing. *Front. Genet.* 10, 384
37. Payá-Milans, M. *et al.* (2018) Comprehensive evaluation of RNA-seq analysis pipelines in diploid and polyploid species. *GigaScience* 7, giy132
38. Conesa, A. *et al.* (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13
39. McCormick, R.F. *et al.* (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354
40. Rey, O. *et al.* (2020) Linking epigenetics and biological conservation: towards a conservation epigenetics perspective. *Funct. Ecol.* 34, 414–427
41. Rey, O. *et al.* (2016) Adaptation to global change: a transposable element–epigenetics perspective. *Trends Ecol. Evol.* 31, 514–526
42. Lea, A.J. *et al.* (2017) Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nat. Ecol. Evol.* 1, 1074–1083
43. Gatzmann, F. *et al.* (2018) The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. *Epigenetics Chromatin* 11, 57
44. Gautier, M. *et al.* (2016) Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. *Mol. Biol. Evol.* 33, 2801–2814
45. Malinsky, M. *et al.* (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* 2, 1940–1955
46. Leitwein, M. *et al.* (2020) Using haplotype information for conservation genomics. *Trends Ecol. Evol.* 35, 245–258
47. Chueca, L.J. *et al.* (2021) Whole-genome re-sequencing data to infer historical demography and speciation processes in land snails: the study of two Candidula sister species. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 376, 20200156
48. Marques, D.A. *et al.* (2019) A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34, 531–544
49. Fuentes-Pardo, A.P. and Ruzzante, D.E. (2017) Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Mol. Ecol.* 26, 5369–5406
50. Díez-Del-Molino, D. *et al.* (2018) Quantifying temporal genomic erosion in endangered species. *Trends Ecol. Evol.* 33, 176–185
51. van der Valk, T. *et al.* (2019) Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Curr. Biol.* 29, 165–170
52. de Manuel, M. *et al.* (2020) The evolutionary history of extinct and living lions. *Proc. Natl. Acad. Sci. U. S. A.* 117, 10927–10934
53. Schlötterer, C. *et al.* (2014) Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15, 749–763
54. Waldvogel, A.-M. *et al.* (2018) The genomic footprint of climate adaptation in *Chironomus riparius*. *Mol. Ecol.* 27, 1439–1456
55. Willoughby, J.R. *et al.* (2018) Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Mol. Ecol.* 27, 4041–4051
56. Morales, H.E. *et al.* (2019) Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Sci. Adv.* 5, eaav9963
57. Dorant, Y. *et al.* (2019) Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: the American lobster (*Homarus americanus*) as a case study. *Ecol. Evol.* 9, 6606–6623
58. Carroll, E.L. *et al.* (2018) Genetic and genomic monitoring with minimally invasive sampling methods. *Evol. Appl.* 11, 1094–1119
59. von Thaden, A. *et al.* (2020) Applying genomic data in wildlife monitoring: development guidelines for genotyping degraded samples with reduced single nucleotide polymorphism panels. *Mol. Ecol. Resour.* 20, 662–680
60. Ruiz-Lopez, M.J. *et al.* (2016) A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity* 116, 167–176
61. Harmoinen, J. *et al.* (2021) Reliable wolf-dog hybrid detection in Europe using a reduced SNP panel developed for non-invasively collected samples. *BMC Genomics* 22, 473
62. Hayward, K.M. *et al.* (2022) Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: a new panel to enable ongoing monitoring of Canadian polar bear populations. *Mol. Ecol. Resour.* 22, 1906–1918
63. Hernandez-Rodriguez, J. *et al.* (2018) The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Mol. Ecol. Resour.* 18, 319–333
64. Paez, S. *et al.* (2022) Reference genomes for conservation. *Science* 377, 364–366
65. Abascal, F. *et al.* (2016) Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* 17, 251
66. Brandies, P. *et al.* (2019) The value of reference genomes in the conservation of threatened species. *Genes* 10, 846
67. Zoonomia Consortium (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature* 587, 240–245
68. Mérot, C. *et al.* (2020) A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol. Evol.* 35, 561–572
69. Storer, J.M. *et al.* (2022) Recently integrated Alu elements in capuchin monkeys: a resource for *Cebus/Sapajus* genomics. *Genes* 13, 572
70. Banes, G.L. *et al.* (2020) Genomic targets for high-resolution inference of kinship, ancestry and disease susceptibility in orang-utans (genus: *Pongo*). *BMC Genomics* 21, 873
71. Hoban, S. *et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am. Nat.* 188, 379–397
72. Manel, S. *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol. Ecol.* 25, 170–184
73. Capblancq, T. *et al.* (2020) Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annu. Rev. Ecol. Evol. Syst.* 51, 245–269
74. Kardos, M. *et al.* (2016) Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Mol. Ecol. Resour.* 16, 727–741

75. Prince, D.J. *et al.* (2017) The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* 3, e1603198
76. Kardos, M. and Shafer, A.B.A. (2018) The peril of gene-targeted conservation. *Trends Ecol. Evol.* 33, 827–839
77. Flanagan, S.P. *et al.* (2018) Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evol. Appl.* 11, 1035–1052
78. Scott, P.A. *et al.* (2020) Individual heterozygosity predicts translocation success in threatened desert tortoises. *Science* 370, 1086–1089
79. Willoughby, J.R. *et al.* (2017) Inbreeding and selection shape genomic diversity in captive populations: implications for the conservation of endangered species. *PLoS One* 12, e0175996
80. Forester, B.R. *et al.* (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* 27, 2215–2233
81. Romero, I.G. and Lea, A.J. (2022) Leveraging massively parallel reporter assays for evolutionary questions. *arXiv* Published online April 12, 2022. https://doi.org/10.48550/arXiv.2204.05857
82. Wold, J. *et al.* (2021) Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol. Ecol.* 30, 5949–5965
83. Batley, K.C. *et al.* (2021) Whole genomes reveal multiple candidate genes and pathways involved in the immune response of dolphins to a highly infectious virus. *Mol. Ecol.* 30, 6434–6448
84. Gignoux-Wolfsohn, S.A. *et al.* (2021) Genomic signatures of selection in bats surviving white-nose syndrome. *Mol. Ecol.* 30, 5643–5657
85. Sanchez-Donoso, I. *et al.* (2022) Massive genome inversion drives coexistence of divergent morphs in common quails. *Curr. Biol.* 32, 462–469
86. Razgour, O. *et al.* (2019) Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl. Acad. Sci. U. S. A.* 116, 10418–10423
87. Waldvogel, A.-M. *et al.* (2020) Evolutionary genomics can improve prediction of species' responses to climate change. *Evol. Lett.* 4, 4–18
88. Pfenninger, M. *et al.* (2021) Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife* 10, e65532
89. Hendricks, S.A. *et al.* (2019) Conservation genomics illuminates the adaptive uniqueness of North American gray wolves. *Conserv. Genet.* 20, 29–43
90. Hedrick, P.W. and Garcia-Dorado, A. (2016) Understanding inbreeding depression, purging, and genetic rescue. *Trends Ecol. Evol.* 31, 940–952
91. Kardos, M. *et al.* (2016) Genomics advances the study of inbreeding depression in the wild. *Evol. Appl.* 9, 1205–1218
92. von Seth, J. *et al.* (2021) Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nat. Commun.* 12, 2393
93. Wang, P. *et al.* (2021) Genomic consequences of long-term population decline in brown eared pheasant. *Mol. Biol. Evol.* 38, 263–273
94. Leroy, G. *et al.* (2018) Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evol. Appl.* 11, 1066–1083
95. Saremi, N.F. *et al.* (2019) Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nat. Commun.* 10, 4769
96. Grossen, C. *et al.* (2020) Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat. Commun.* 11, 1001
97. Bertorelle, G. *et al.* (2022) Genetic load: genomic estimates and applications in non-model animals. *Nat. Rev. Genet.* 23, 492–503
98. Alvarez-Estape, M. *et al.* (2022) Insights from the rescue and breeding management of Cuvier's gazelle (*Gazella cuvieri*) through whole-genome sequencing. *Evol. Appl.* 15, 351–364
99. Feng, S. *et al.* (2019) The genomic footprints of the fall and recovery of the crested ibis. *Curr. Biol.* 29, 340–349
100. Bell, D.A. *et al.* (2019) The exciting potential and remaining uncertainties of genetic rescue. *Trends Ecol. Evol.* 34, 1070–1079
101. Medugorac, I. *et al.* (2017) Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* 49, 470–475
102. Rogers, J. *et al.* (2019) The comparative genomics and complex population history of *Papio* baboons. *Sci. Adv.* 5, eaau6947
103. Stroupe, S. *et al.* (2022) Genomic evaluation of hybridization in historic and modern North American bison (*Bison bison*). *Sci. Rep.* 12, 6397
104. Knott, G.J. and Doudna, J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science* 361, 866–869
105. Piaggio, A.J. *et al.* (2017) Is it time for synthetic biodiversity conservation? *Trends Ecol. Evol.* 32, 97–107
106. Hammond, A. *et al.* (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* 34, 78–83
107. Bier, E. (2022) Gene drives gaining speed. *Nat. Rev. Genet.* 23, 5–22
108. Esvelt, K.M. and Gemmell, N.J. (2017) Conservation demands safe gene drive. *PLoS Biol.* 15, e2003850
109. Noble, C. *et al.* (2018) Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *eLife* 7, e33423
110. Webber, B.L. *et al.* (2015) Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? *Proc. Natl. Acad. Sci. U. S. A.* 112, 10565–10567
111. Shapiro, B. (2017) Pathways to de-extinction: how close can we get to resurrection of an extinct species? *Funct. Ecol.* 31, 996–1002
112. McCauley, D.J. *et al.* (2017) A mammoth undertaking: harnessing insight from functional ecology to shape de-extinction priority setting. *Funct. Ecol.* 31, 1003–1011
113. Donlan, J. (2014) De-extinction in a crisis discipline. *Front. Biogeogr.* 6, 25–28
114. Bennett, J.R. *et al.* (2017) Spending limited resources on de-extinction could lead to net biodiversity loss. *Nat. Ecol. Evol.* 1, 53
115. Wisely, S.M. *et al.* (2015) A road map for 21st century genetic restoration: gene pool enrichment of the black-footed ferret. *J. Hered.* 106, 581–592
116. Breed, M.F. *et al.* (2019) The potential of genomics for restoring ecosystems and biodiversity. *Nat. Rev. Genet.* 20, 615–628
117. Hartman, W.H. *et al.* (2017) A genomic perspective on stoichiometric regulation of soil carbon cycling. *ISME J.* 11, 2652–2665
118. Cristescu, M.E. (2019) Can environmental RNA revolutionize biodiversity science? *Trends Ecol. Evol.* 34, 694–697
119. Rodriguez-R, L.M. *et al.* (2018) The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* 46, W282–W288
120. Tu, Q. *et al.* (2014) Strain/species identification in metagenomes using genome-specific markers. *Nucleic Acids Res.* 42, e67
121. Gelsinger, D.R. *et al.* (2020) Regulatory noncoding small RNAs are diverse and abundant in an extremophilic microbial community. *mSystems* 5, e00584-19
122. Sherkow, J.S. *et al.* (2022) Ethical, legal, and social issues in the Earth BioGenome Project. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115859119
123. Gates, A.J. *et al.* (2021) A wealth of discovery built on the Human Genome Project — by the numbers. *Nature* 590, 212–215
124. Scott, A.D. *et al.* (2020) A reference genome sequence for giant sequoia. *G3* 10, 3907–3919
125. Rhie, A. *et al.* (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746
126. Kingan, S.B. *et al.* (2019) A high-quality de novo genome assembly from a single mosquito using PacBio sequencing. *Genes* 10, 62
127. Etherington, G.J. *et al.* (2020) Sequencing smart: *de novo* sequencing and assembly approaches for a non-model mammal. *Gigascience* 9, giaa045
128. Hazzouri, K.M. *et al.* (2015) Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat. Commun.* 6, 8824
129. Monnahan, P.J. *et al.* (2020) Using multiple reference genomes to identify and resolve annotation inconsistencies. *BMC Genomics* 21, 281
130. Valiente-Mullor, C. *et al.* (2021) One is not enough: on the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Comput. Biol.* 17, e1008678

131. Llamas, B. *et al.* (2019) A strategy for building and using a human reference pangenome. *F1000Res.* 8, 1751
132. Gerdol, M. *et al.* (2020) Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol.* 21, 275
133. Song, J.-M. *et al.* (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45
134. Berg, P.R. *et al.* (2016) Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* 6, 23246
135. Star, B. *et al.* (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477, 207–210
136. Berg, P.R. *et al.* (2017) Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity* 119, 418–428
137. Matschiner, M. *et al.* (2022) Supergene origin and maintenance in Atlantic cod. *Nat. Ecol. Evol.* 6, 469–481
138. Nong, W. *et al.* (2021) Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication. *Commun. Biol.* 4, 83
139. Shingate, P. *et al.* (2020) Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat. Commun.* 11, 2322
140. McMullan, M. *et al.* (2018) The ash dieback invasion of Europe was founded by two genetically divergent individuals. *Nat. Ecol. Evol.* 2, 1000–1008
141. Sollars, E.S.A. *et al.* (2017) Genome sequence and genetic diversity of European ash trees. *Nature* 541, 212–216
142. Harper, A.L. *et al.* (2016) Molecular markers for tolerance of 1European ash (*Fraxinus excelsior*) to dieback disease identified using associative transcriptomics. *Sci. Rep.* 6, 19335
143. Mishra, B. *et al.* (2018) A reference genome of the European beech (*Fagus sylvatica* L.). *Gigascience* 7, giy063
144. Casas-Marce, M. *et al.* (2017) Spatiotemporal dynamics of genetic variation in the Iberian lynx along its path to extinction reconstructed with ancient DNA. *Mol. Biol. Evol.* 34, 2893–2907
145. Lucena-Perez, M. *et al.* (2021) Bottleneck-associated changes in the genomic landscape of genetic diversity in wild lynx populations. *Evol. Appl.* 14, 2664–2679
146. Kleinman-Ruiz, D. *et al.* (2022) Purging of deleterious burden in the endangered Iberian lynx. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2110614119
147. Kleinman-Ruiz, D. *et al.* (2017) Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics* 18, 556
148. Johnson, W.E. *et al.* (2010) Genetic restoration of the Florida panther. *Science* 329, 1641–1645
149. Mariani, S. *et al.* (2019) Sponges as natural environmental DNA samplers. *Curr. Biol.* 29, R401–R402
150. Kenny, N.J. *et al.* (2020) Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nat. Commun.* 11, 3676