
This is the **submitted version** of the review article:

Pegueroles Queralt, Maria Cinta; Pascual, Marta; Carreras, Carles. «Going beyond a reference genome in conservation genomics». *Trends in ecology & evolution*, Vol. 39, Num. 1 (January 2024), p. 13-15 DOI 10.1016/j.tree.2023.11.009, PMID 38040545

This version is available at <https://ddd.uab.cat/record/325107>

under the terms of the  ^{IN}COPYRIGHT license.

Title: no more than 8 words

Going beyond a reference genome in conservation genomics

Cinta Peguerols^{1,2}, Marta Pascual^{1,2}, Carlos Carreras^{1,2}

1. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), Av. Diagonal 643. E08028, Barcelona, Spain
2. Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Av. Diagonal 643. E08028, Barcelona, Spain

Twitter handler: @EvoGeneticsLab

Keywords: Include 4-5 descriptive keywords (maximum of six).

Genome diversity, inbreeding, biodiversity, population structure, demographic history

Abstract No more than 50 words only one reference

The current biodiversity crisis demands scientifically based management. The generation of reference genomes is crucial in conservation genomics, but not enough to capture the species diversity. By incorporating whole genome sequencing at the population level Nigenda-Morales et al. provide key genomic information for the conservation of the Pacific fin whale.

Main text

Maximum of 1200 words (this limit does not include text in the boxes, abstract, references, tables or figure legends)

☐ Maximum of one element per article (i.e. a Figure, Table or Box)

☐ 12 references total

☐ Include both page & line numbers to aid the reviewing process

Use concise logical subheadings and provide clear links between sections.

Global reports on biodiversity based on scientific research indicate unprecedented rates of biodiversity loss and suggest that we are facing the sixth mass extinction [1]. This current trend demands urgent conservation actions to stop and revert the situation by protecting species, habitats and ecosystems. However, conservation planning needs to be supported by solid scientific research to match the proposed actions with the biological needs of the species [2]. In this context, genomics can greatly assist biodiversity conservation and management by providing key information on population structuring, inbreeding, adaptation or mutation load among others [3]. The progress in high-throughput sequencing enabled large-scale generation of reference genomes [4]. As a consequence, many initiatives are now on the way to provide reference genomes with the ultimate goal to sequence all earth biodiversity [5]. This is going to be without doubt one of the major milestones of humankind in the current era, but the generation of a single reference genome per species does not provide all genomic information needed for conservation. Although a reference genome is an invaluable centrepiece of conservation genomics [4], the genome of a single individual is unable to capture all the intraspecific genome diversity, which is crucial for the management of wild populations.

No single genome can represent the diversity of a species [6] as structural and sequence polymorphisms may differentially drive the future conservation actions of the populations.

Consequently, genomics needs to look beyond the generation of reference genomes as the logical next step for a more effective species conservation (Figure 1). Thus, targeted population-specific diagnosis can be achieved by integrating whole-genome sequencing data from multiple individuals across the species' distribution range with the reference genome. This approach yields comprehensive insights on population structure to delineate conservation units, detect recent and ancient signals of inbreeding, infer the demographic history of the analysed populations or assess their mutation load (Figure 1). Additionally, generating high quality and density of genomic markers may allow performing genome base simulations to assess the potential effectiveness of different conservation strategies. By using this approach, Nigela-Morales *et al.* [7] went beyond the reference genome and provided key genomic information for the conservation of the Pacific fin whale (*Balaenoptera physalus*).

Conservation of species at population level

The species is generally the taxonomic unit used for conservation by most management bodies such as the IUCN (<https://www.iucnredlist.org>). However, since the foundation of disciplines like conservation genetics in the second half of the 20th century, growing evidence indicate that the processes within populations are crucial to understand the survival potential of the species [8]. Historical or current habitat fragmentation and gene flow reduction imply limited connections among populations, thus requiring independent assessments. Furthermore, the intensity of threats frequently has uneven intensity across populations, leading to overexploited populations and pristine relicts within the same species. Finally, effective population sizes can also be different along the species distribution and this is a crucial parameter in conservation genomics with a high impact on genome diversity. For all these reasons, population specific conservation units have been defined for many species, sometimes with populations assigned to different conservation categories depending on their particularities [9]. The need of population specific assessments for conservation is evidenced in Nigela-Morales *et al.* [7], as this study included fin whales from two areas with divergent threat intensity and historical population sizes. The large population from Eastern North Pacific (ENP) was intensively exploited by industrial whaling during the last century, while the reduced population of the Gulf of California (GOC) has remained unspoiled. Genomic data supported the existence of these two isolated management units with little admixture, therefore reinforcing the need for independent conservation strategies.

Demographic inference and the footprint of inbreeding

Genomic data revealed the demographic history of both whale populations. The ancestral stock of whales split into the current two groups between 16 and 25 kya and, after this divergence, they experienced different demographic processes. The population from Eastern North Pacific hosted a large effective population size of tens of thousands of individuals before suffering a severe bottleneck due to human hunting that reduced its size by 99% only a few generations ago. Conversely, the population of the Gulf of California was not affected by anthropogenic activity, but remained small for thousands of years, with only a few hundred of individuals and little incoming gene flow. These contrasting demographic patterns, including a natural and a human-induced bottleneck, shaped the genome differentially in the populations. The individuals from the Gulf of California showed clearer signals of inbreeding with large fragments and a large accumulated portion of the genome without polymorphism. By combining these results, the authors demonstrated that a sustained reduced effective population size can be more devastating for the genome diversity within the populations than a recent drastic harvesting, at least for long-longevity species. However, the recent reduction of individuals in the Eastern north Pacific may have a huge impact on genome diversity in the future if no action is taken.

Genetic load and conservation scenarios simulations

While the examination of overall genetic diversity is unable to predict the risk of extinction effectively [10], the annotation of reference genomes enables a functional analysis of the polymorphisms found and their adaptive potential. Consequently, the analysis of genome-wide genetic load provides some insights into the potential fitness of the genomes resulting from the different demographic trajectories [11]. In the case of the Pacific fin whales, the population of the Gulf of California showed a significantly higher proportion of moderately deleterious alleles and a higher proportion of derived homozygote positions in coding regions. This result indicates that the demographic history of this population has not only eroded its genome variation, but has also increased the risk of extinction due to a high genetic load. Simulations of different conservation scenarios provided some clues about what conservation strategies could be more effective in preserving the genome diversity and fitness of the Pacific fin whales. Promoting a fast recovery of population effective sizes in the North East Pacific resulted in a significant improvement in the following next generations. Additionally, although the levels of migration into the Gulf of California were low, they proved to be crucial in the maintenance of the genomic health of this population.

In summary, Nigela-Morales *et al.* showcased the potential of whole genome sequencing to aid conservation at population level. This study has important messages for future research on conservation genomics and their need for translation into biodiversity policy and practice [12]. The first one is that historical demography can be more impactful on the genomic health of the populations than an intense harvest or recent population reduction. Thus, population-based diagnosis at genome level can be fundamental to evaluate the current and future perspectives of the populations of endangered species. Furthermore, it shows that some pristine populations genomic potential could be compromised, even without human interactions, indicating that population assessments and actions at the whole distribution range level are needed. In our opinion, going beyond the reference genome by incorporating whole genome information across the species distribution, will be the new standard in conservation genomics.

Acknowledgements

The authors want to thank Gisela Marín-Capuz for greatly improving the graphical design of the Figure. The authors are supported by the projects MarGech (PID2020-118550RB funded by MCIN/ AEI/10.13039/501100011033), and GenoMarTur (CNS2022-135205 funded by MCIN/ AEI/10.13039/501100011033 and the 'European Union NextGenerationEU/PRTR'). The authors are members of the research group SGR2021-01271 funded by the 'Generalitat de Catalunya' and are part of the 'Catalan Initiative for the Earth BioGenome Project' (CBP).

References (max 12)

- IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In (Díaz, S. et al., eds), IPBES secretariat, Bonn, Germany. 56
- Hohenlohe, P.A. *et al.* (2021) Population genomics for wildlife conservation and management. *Mol Ecol* 30, 62–82
- Theissinger, K. *et al.* (2023) How genomics can help biodiversity conservation. *Trends in Genetics* 39, 545–559
- Formenti, G. *et al.* (2022) The era of reference genomes in conservation genomics. *Trends Ecol Evol* 37, 197–202

- 135 Lewin, H.A. *et al.* (2018) Earth BioGenome Project: Sequencing life for the future of life. *Proc*
 136 *Natl Acad Sci U S A*, 115, 4325–4333
- 137 Liao, W.W. *et al.* (2023) A draft human pangenome reference. *Nature* 2023 617 312–324
- 138 Nigenda-Morales, S.F. *et al.* (2023) The genomic footprint of whaling and isolation in fin whale
 139 populations. *Nature Communications* 14, 1–18
- 140 Willi, Y. *et al.* (2022) Conservation genetics as a management tool: The five best-supported
 141 paradigms to assist the management of threatened species. *Proc Natl Acad Sci U S A* 119,
 142 e2105076119
- 143 Wallace, B.P. *et al.* (2010) Regional management units for marine turtles: a novel framework
 144 for prioritizing conservation and research across multiple scales. *PLoS One* 5, e15465
- 145 Teixeira, J.C. and Huber, C.D. (2021) The inflated significance of neutral genetic diversity in
 146 conservation genetics. *Proc Natl Acad Sci U S A* 118, e2015096118
- 147 Kardos, M. *et al.* (2021) The crucial role of genome-wide genetic variation in conservation. *Proc*
 148 *Natl Acad Sci U S A* 118, e2104642118
- 149 De León, L.F. *et al.* (2023) Harnessing the omics revolution to address the global biodiversity
 150 crisis. *Curr Opin Biotechnol* 80, 102901

151

152

153 **Figure 1. Beyond a reference genome in conservation genomics.** A) The genome diversity of a
 154 species is driven by the structure, adaptation and gene flow of populations across its range of
 155 distribution, by the differential threats they can face and the different population sizes and
 156 demographic histories. B) Genome diversity can be assessed by combining a good quality
 157 reference genome with whole genome sequencing data. Multiple populations, fair number of
 158 individuals and high sequence coverage are needed to provide population specific diagnosis on
 159 genome diversity. C) As a result, key information for conservation and management can be
 160 extracted including population structuring and gene flow, signals of past or recent inbreeding,
 161 demographic history and admixture or mutation load. Additionally, genome-based simulations
 162 can provide useful insights on the potential effectiveness of different conservation strategies.