Received: 2 June 2019 | Revised: 4 June 2019 | Accepted: 5 June 2019

DOI: 10.1111/mec.15141

NEWS AND VIEWS

Perspectives

MOLECULAR ECOLOGY WILEY

Excavating ghost footprints and tangled trees from modern genomes

Arne Jacobs | Nina Overgaard Therkildsen

Department of Natural Resources, Cornell University, Ithaca, NY, USA

Correspondence
Nina Overgaard Therkildsen, Department of Natural Resources, Cornell University, Ithaca, NY, USA.
Email: nt246@cornell.edu

Funding information
Division of Ocean Sciences, Grant/Award Number: OCE-1756316

Due to pervasive gene flow and admixture, simple bifurcating trees often do not provide an accurate representation of relationships among diverging lineages, but limited resolution in the available genomic data and the spatial distribution of samples has hindered detailed insights regarding the evolutionary and demographic history of many species and populations. In this issue of *Molecular Ecology*, Foote et al. (2019) combine a powerful sampling design with novel analytical methods adopted from human genetics to describe previously unrecognized patterns of recurrent vicariance and admixture among lineages in the globally distributed killer whale (*Orcinus orca*). Based on sequence data from modern samples alone, they discover clear signatures of ancient admixture with a now extinct "ghost" lineage, providing one of the first accounts of archaic introgression in a nonhominid species. Coupling a cost-effective sequencing strategy with novel analytical approaches, their paper provides a roadmap for advancing inference of evolutionary history in other nonmodel species, promising exciting times ahead for our field.

KEYWORDS
admixture, archaic ancestry, evolutionary history, ghost populations

Simple bifurcating trees, which traditionally have played a central role in descriptions of diversification patterns within and among species, fail to account for complex scenarios that include genetic exchange among diverging lineages at various points in time. Such gene flow appears to be pervasive in nature (Arnold, 2015), but despite vast improvements in analytical methods and data availability, it remains nontrivial to interpret patterns of allele sharing and genetic similarity between separate populations. Genetic similarity can reflect recent divergence, but may also, for example, mask more ancient splits followed by historical or ongoing erosion of genetic differentiation through admixture after secondary contact, or retention of ancestral diversity due to large effective population sizes. Our understanding of the evolutionary history of many taxa thus remains rudimentary.

In this issue of *Molecular Ecology*, Foote et al. (2019) illustrate how comprehensive geographical and genomic sampling coupled with novel analytical techniques, which so far have seen little application outside the human genetics literature, reveal previously unrecognized complexity in the evolutionary history of killer whales (Figure 1). This species has a global distribution, but in some regions shows pronounced fine-scale geographical variation, for example in diet, morphology and ecology. Despite extensive prior genetic analysis, the relationships between killer whale lineages had remained somewhat unresolved due to spatial gaps in sampling combined with pervasive admixture and incomplete lineage sorting. With powerful analysis tools and 47 resequenced genomes that almost fully represent the known global geographical and genetic diversity of this species, Foote et al. (2019) show that the distinctiveness of different ecotypes is shaped by multiple deep ancestral splits coupled with more recent gene flow among specific lineages. They reason that the recurrent vicariance events and subsequent admixture after secondary contact have probably been driven by the cyclical expansion and contraction of high-latitude habitat during multiple consecutive glacial cycles.

A particularly intriguing result from Foote et al.'s (2019) paper is the detection of genomic tracts introgressed from a now extinct

FIGURE 1 Killer whale (*Orcinus orca*) photographed off Shetland (photo credit Andy Foote)

(or otherwise unsampled) deeply divergent "ghost" population into Antarctic killer whale morphs. These archaic ancestry segments have up to four-fold older coalescence time with low-latitude populations than the genome-wide average and thus probably originated from a lineage that split off to colonize southern habitat that became available during an earlier interglacial period (>200,000 years ago), while the rest of the genome suggests the current Antarctic group split off during the most recent glacial cycle (~60,000 years ago).

Until now, the study of archaic admixture has mostly been limited to hominids, focusing in particular on introgression from Neanderthals and Denisovans into modern humans (Dannemann & Racimo, 2018; Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015) and often taking advantage of reference genomes generated from ancient remains of these extinct lineages. With the increasing sophistication of palaeogenomic tools and sequencing techniques, such analytical approaches are also starting to be applied in other species with a strong palaeontological record, such as horses (Fages et al., 2019). Yet, we are unlikely to be able to recover genomic reference data for all relevant extinct lineages in many species.

While there are well-established frameworks for taking gene flow from "ghost populations" (populations for which we have no data) into account in estimation of population sizes and migration rates, Foote et al. (2019) push beyond this by actually identifying specific genomic regions and variants originating from archaic introgression. They do this with an exciting new method that utilizes the distribution of private alleles across the genome of a potentially admixed population compared to a nonadmixed outgroup to identify regions likely to be introgressed from a highly divergent source (Skov et al., 2018). Being able to identify and characterize such introgressed tracts directly in modern genome sequences without the need for any *a priori* knowledge of archaic lineages is an important step forward for several reasons. First, not accounting for variation in ancestry and contrasting evolutionary histories of different parts of the genome can inflate divergence times estimated based on average genome-wide patterns. Second, the ability to identify introgressed sequences opens entirely new opportunities to examine whether such ancestry tracts contain variation that is beneficial,

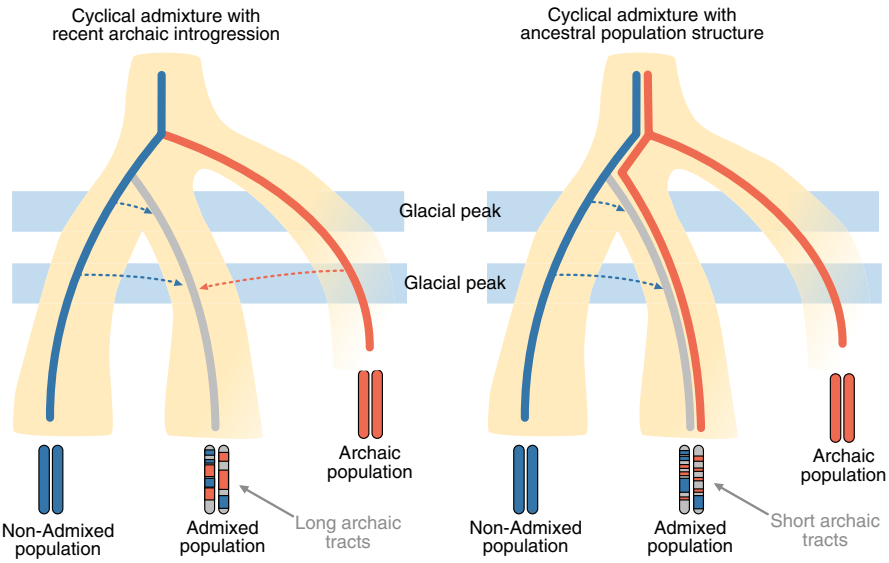

FIGURE 2 Conceptual illustrations of different admixture histories with a ghost population. Left panel: evolutionary history of a population with cyclical admixture from a nonadmixed population during glacial peaks and recent introgression from an archaic population. The different cycles of admixture from the nonadmixed population (during glacial peaks when less high-latitude habitat would have been available, thus bringing divergent lineages into secondary contact) leads to introgression tracts of variable lengths. The more recent admixture from the archaic population is represented in longer archaic tracts in the genome. Right panel: evolutionary history of a population that has experienced cyclical admixture during glacial peaks from a nonadmixed population and shows signatures of ancestral population structure involving the archaic lineage. In contrast to introgression, the ancestral population structure and sorting of archaic ancestry is represented by shorter archaic tracts across the genome of the admixed population because the two lineages have been in contact for longer, giving recombination more time to break down the tracts

neutral or deleterious for the recipient lineage, as has been explored for hominid admixture (e.g., Racimo et al., 2015). The ability to examine evidence for these contrasting scenarios across a broad range of organisms will substantially improve our understanding of the role admixture plays in shaping evolutionary histories.

Foote et al. (2019) infer that up to ~20% of the genome in Antarctic killer whales potentially originated from archaic admixture. The length of these archaic tracts and the density of private alleles within them provide important insights into the time since the introgression took place and the degree of divergence between the introgressing and receiving lineages (Figure 2; Racimo et al., 2015). As recombination will break down linkage among private alleles over time, longer archaic tracts indicate more recent introgression (as there has been less time for recombination to break them down), while shorter tracts suggest more ancient admixture. Foote et al. (2019) find that the introgressed archaic tracts in Antarctic killer whales are about an order of magnitude shorter than those seen non-African humans and suggest that they are more likely to result from ancestral population structure and ancient admixture than recent introgression (Figure 2).

Thus, by utilizing novel analytical methods, Foote et al. (2019) provide one of the first detailed investigations into archaic tracts in nonhominid genomes. Two of the very few other examples so far come from European sea bass (Duranton et al., 2019) and bonobo (Kuhlwilm, Han, Sousa, Excoffier, & Marques-Bonet, 2019). Yet, we suspect that ancient admixture dynamics are likely to be quite common in nature. It is only now that we have the genomic tools and the ability to relatively cost-effectively generate genomic data of sufficient resolution in nonmodel species to excavate such archaic ancestry from modern samples, and we will probably soon see this type of analysis shedding new light on the evolutionary history of many species.

Foote et al.'s (2019) novel insights into the evolutionary history of killer whales were underpinned by two key aspects of their data set: (a) having fully resequenced genomes that revealed multimodal patterns of ancestry within individuals and (b) having global coverage in sampling that together provide a comprehensive overview of the current diversity within the species and illustrate that gene flow has occurred across great distances and after long periods of isolation between populations. So far, only a few similar global whole genome data sets exist for nonmodel species, examples including the brown rat (Puckett & Munshi-South, 2019) and yellowfin tuna (Barth, Damerau, Matschiner, Jentoft, & Hanel, 2017). However, with sequencing and library preparation costs continuing to decline, such data are now becoming within reach for many research programmes, even with relatively modest budgets (Therkildsen & Palumbi, 2017). Many population genomic studies in nonmodel species use relatively large sample sizes from a restricted number of locations and limit the per-sample cost by only sequencing a subset of the genome (e.g., with RAD-seq or target capture). In contrast, the strategy used by Foote et al. (2019) was to spread their sampling across just a few individuals from each of many locations, and use low- to medium-coverage whole-genome resequencing analysed in a probabilistic framework. While costly high-coverage sequencing would have provided more certain individual genotypes, it would have limited the number of individuals they could include. Although large sample sizes and confident genotype calls are needed for robust inference in some types of analysis, there are other types of analysis for which the large number of genetically independent loci represented in whole genome data provide sufficient precision in estimation of population genetic parameters even when based on just a single or a few individuals (e.g., Li & Durbin, 2011). Furthermore, for analysis requiring multi-individual data, simulation studies have demonstrated that sampling many individuals at low read depth often provides more precise estimates of population parameters than higher read depth for fewer individuals (e.g., Buerkle & Gompert, 2013; Fumagalli, 2013).

Foote et al. (2019) illustrate how a powerful sampling design and analytical approaches borrowed from human genetics allowed unprecedented insights into complex evolutionary patterns, providing a roadmap for future efforts to reconstruct the evolutionary and demographic history of nonmodel organisms. As the methods for detecting archaic tracts without ancient reference sequences have emerged only over the past few years, Foote et al.'s (2019) study also highlights how the time lag between application of novel analytical tools in humans and nonmodel species is substantially shortening, promising exciting times ahead for our field.

AUTHOR CONTRIBUTION

The authors wrote the paper together.

ORCID

Arne Jacobs https://orcid.org/0000-0001-7635-5447
Nina Overgaard Therkildsen https://orcid.org/0000-0002-6591-591X

REFERENCES

Arnold, M. L. (2015). *Divergence with genetic exchange*. Oxford, UK: Oxford University Press.

Barth, J. M. I., Damerau, M., Matschiner, M., Jentoft, S., & Hanel, R. (2017). Genomic differentiation and demographic histories of Atlantic and Indo-Pacific yellowfin tuna (*Thunnus albacares*) populations. *Genome Biology and Evolution*, *9*(4), 1084–1098. https://doi.org/10.1093/gbe/evx067

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22*(11), 3028–3035. https://doi.org/10.1111/mec.12105

Dannemann, M., & Racimo, F. (2018). Something old, something borrowed: Admixture and adaptation in human evolution. *Current Opinion in Genetics & Development*, *53*, 1–8. https://doi.org/10.1016/j.gde.2018.05.009

Duranton, M., Allal, F., Valière, S., Bouchez, O., Bonhomme, F., & Gagnaire, P.-A. (2019). The contribution of ancient admixture to reproductive isolation between European sea bass lineages. *bioRxiv*. https://doi.org/10.1101/641829

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., … Orlando, L. (2019). Tracking five millennia of horse

management with extensive ancient genome time series. *Cell*, *177*, 1419–1435. https://doi.org/10.1016/j.cell.2019.03.049

Foote, A. D., Martin, M. D., Louis, M., Pacheco, G., Robertson, K. M., Sinding, M.-H., … Morin, P. A. (2019). Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Molecular Ecology*, 28, 3247–3444. https://doi.org/10.1111/mec.15099

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS ONE*, *8*(11), e79667. https://doi.org/10.1371/journal.pone.0079667

Kuhlwilm, M., Han, S., Sousa, V. C., Excoffier, L., & Marques-Bonet, T. (2019). Ancient admixture from an extinct ape lineage into bonobos. *Nature Ecology & Evolution*, *3*(6), 957–965. https://doi.org/10.1038/s41559-019-0881-7

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496.

Puckett, E. E., & Munshi-South, J. (2019). Brown rat demography reveals pre-commensal structure in eastern Asia before expansion into Southeast Asia. *Genome Research*, *29*(5), 762–770. https://doi.org/10.1101/gr.235754.118

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, *16*(6), 359–371. https://doi.org/10.1038/nrg3936

Skov, L., Hui, R., Shchur, V., Hobolth, A., Scally, A., Schierup, M. H., & Durbin, R. (2018). Detecting archaic introgression using an unadmixed outgroup. *PLoS Genetics*, *14*(9), e1007641. https://doi.org/10.1371/journal.pgen.1007641

Therkildsen, N. O., & Palumbi, S.R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources*, *17*(2), 194–208. http://doi.org/10.1111/1755-0998.12593

How to cite this article: Jacobs A, Therkildsen NO. Excavating ghost footprints and tangled trees from modern genomes. *Mol Ecol*. 2019;28:3287–3290. https://doi.org/10.1111/mec.15141