RESEARCH ARTICLE

Correspondence
nicolas.galtier@umontpellier.fr

Lacking conservation genomics in the giant Galápagos tortoise

Etienne Loire [ID],[1] and Nicolas Galtier [ID],[2]

Abstract

Conservation policy in the giant Galpagos tortoise, an iconic endangered animal, has been assisted by genetic markers for 15 years: a dozen loci have been used to delineate thirteen (sub)species, between which hybridization is prevented. Here, comparative reanalysis of a previously published NGS data set reveals a conflict with traditional markers. Genetic diversity and population substructure in the giant Galpagos tortoise are found to be particularly low, questioning the genetic relevance of current conservation practices. Further examination of giant Galapagos tortoise population genomics is critically needed.

[1]Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR CMAEE and C2B, Montpellier, France, [2]ISEM - Univ Montpellier - CNRS - IRD, Montpellier, France

Contents

Introduction

Biological conservation implies collaboration between scientists, biodiversity managers and policy makers. We can only applaud when scientific progress influences policies and practices in a way that is beneficial to nature, and ultimately to humans. A recent Comment in Nature (Nichols, 2015) highlights a spectacular conservation initiative geared towards an iconic endangered species – the giant Galápagos tortoise.

The Galápagos Islands host a number of fascinating endemic animals, such as Darwin's finches, Galápagos iguanas and Galápagos penguins and their preservation is clearly a major issue. These Islands are a paramount conservation priority. Remarkable efforts have been made by the Charles Darwin Research Station and the Galápagos National Park Service to protect indigenous species while eradicating invasive ones (e.g. Marquez et al., 2013). The Galapagos tortoise, *Chelonoidis nigra*, is probably the most emblematic of all species of the archipelago. These large tortoises have been heavily impacted by human activities over the centuries. Since the mid-1960s they have been the focus of a specific long-term conservation program, which has been fully successful in terms of population density recovery (e.g. Aguilera et al., 2015; Gibbs et al., 2014). The program mainly involved bringing eggs or adults from natural sites to rearing/breeding centers, and releasing in the field juveniles of sufficient size, thus escaping predation by rats (Cayot, 2008).

Since the 2000's, population genetics research has been instrumental in the management plan of Galápagos tortoises. Thirteen giant Galápagos tortoise species (Poulakakis, Edwards, et al., 2015) have been described, or confirmed, using genetic data – specifically, mitochondrial DNA and 9-12 microsatellite loci (Caccone, Gentile, Gibbs, et al., 2002; Ciofi et al., 2002). Based on this information, population densities have been monitored in a way that aims at preserving, or even restoring, the integrity of the distinct gene pools (e.g. Edwards, Garrick, et al., 2014; Garrick, Benavides, Russello, Gibbs, et al., 2012; Milinkovitch, Monteyne, Russello, et al., 2007). Milinkovitch, Monteyne, Russello, et al. (2007), for instance, identified one "hybrid" individual among the thousands that had been repatriated to the Española island after captive breeding. This was removed from the field. Garrick, Benavides, Russello, Gibbs, et al. (2012) and Edwards, Benavides, et al. (2013) discuss the "conservation value" of individuals based on their genotypes. The most recent implementation of this strategy is an attempt to recreate the genome of the famous Lonesome George tortoise, the last member of the *C. (nigra) abingdoni* species, which died in 2012. The plan is to cross individuals from another species that carry *abingdoni* alleles – hence the recent transfer by helicopter and boat of 32 tortoises from their current habitat to the island of Santa Cruz (Nichols, 2015).

Some problems, however, still hamper the task of defining *C. nigra* taxonomic and conservation units. For instance, mitochondrial DNA and microsatellite data do not fully agree (Poulakakis, Russello, et al., 2012). Published levels of mitochondrial coding sequence divergence between named species of giant Galápagos tortoises (<1%, Caccone, Gentile, Burns, et al., 2004) are not higher than typical amounts of within-species mitochondrial polymorphism in reptiles (Bazin et al., 2006). Importantly, individuals carrying so-called 'hybrid' genotypes or unexpected mitochondrial haplotypes, given their geographic origin, are common in the field (e.g. Garrick, Benavides, Russello, Gibbs, et al., 2012). Because they are interpreted as being the result of non-natural causes (Poulakakis, Glaberman, et al., 2008; Russello, Beheregaray, et al., 2007), such individuals have been removed from published analyses (Garrick, Benavides, Russello, Hyseni, et al., 2014;

Poulakakis, Edwards, et al., 2015), thus artificially inflating measured levels of genetic differentiation between entities. In spite of this bias, estimated FST as low as 0.1 were considered sufficient to delineate species (Poulakakis, Edwards, et al., 2015), in absence of evidence for hybrid depression or local adaptation. Indeed, very little is known regarding the link between genetic variation, morphological variation and fitness in *C. nigra*. The conspicuous "saddleback" morphotype, for instance, is observed in several unrelated species that carry various combinations of mtDNA haplotypes and microsatellite alleles (e.g. Poulakakis, Russello, et al., 2012), and the heritability of this trait is so far unknown (Chiari et al., 2009). Whether gene flow, if any, is recent and human mediated, or ancient and "natural", is critical with respect to conservation policy in *C. nigra*.

Analysis

Four years ago, we published the blood transcriptome of five giant Galápagos tortoises which, based on mtDNA, were assigned to three distinct species – *becki*, *vandenburghi*, and *porteri* Loire et al. (2013). To our surprise, our analysis of 1000 coding loci did not reveal any evidence of genetic structure in this sample, i.e. the homozygosity of the five individuals was not higher than expected under the hypothesis that they belonged to a single panmictic population. The total genetic variation detected in the sample was also quite low – heterozygosity was around 0.1%, consistent with earlier findings Caccone, Gentile, Burns, et al. (2004).

Here we looked at the situation from a comparative standpoint and plotted the inbreeding coefficient F as a function of log-transformed synonymous diversity, π_S, in a dataset of 53 species of animals for which the transcriptome of at least four individuals sampled from different populations was available (Perry et al., 2012; Romiguier, Gayral, et al., 2014, Table S1). The synonymous diversity π_S measures heterozygosity at (presumably neutral) third codon positions; π_S is an unbiased estimate of the $\theta = 4.N_e.m$ parameter, where N_e is the effective population size and m is the per-base, per-generation mutation rate. Inbreeding coefficient F measures the departure from Hardy-Weinberg equilibrium by comparing the expected and observed numbers of heterozygous genotypes, here summed across individuals and loci. F is expected to equal zero in absence of population structure, and to take positive values when individuals are more homozygous, on average, than expected under the assumption of panmixy. Data on all species in this dataset were analysed via the same procedure (Ballenghien et al., 2017; Gayral et al., 2013), thus ensuring optimal comparability between data points. Figure 1 indicates that *C. nigra* harbours lower genetic polymorphism and population structuring than most vertebrates or other animal species, which casts serious doubts on the existence of several differentiated gene pools in this taxon. Interestingly, the data point closest to *C. nigra* in this figure is *Homo sapiens* – a species in which very few people would argue the need to define (sub)species.

The above analysis is based on small samples – from five to eleven individuals per species – which in most cases precluded definition of sub-populations and calculation of the F_{ST} statistics – but note that a significantly positive F_{ST}, if any, should imply a significantly positive F, and that F in this analysis was calculated using the method of Weir and Cockerham (1984), which accounts for small sample size. Our analysis might be affected by, e.g., sequencing errors, genotyping errors, hidden paralogy, cross-contamination or other experimental artefacts (Ballenghien et al., 2017). To control for possible sources of noise and confirm that our approach has the power to identify strong genetic structure when it exists, we similarly analysed samples made of individuals from two congeneric species (Galtier, 2016, Table S1). Distinctively high values of F were obtained with the two-species samples, compared to the single-species ones (Figure 1, open circles), casting further doubts on the hypothesis that our *C. nigra* sample comprises individuals from more than one species.

It was suggested from mtDNA and microsatellite data that some *C. nigra* individuals are of "hybrid origin", i.e., F1 from "purebred" parents or backcrosses (e.g. Garrick, Benavides, Russello, Hyseni, et al., 2014; Russello, Poulakakis, et al., 2010). Among the four individuals of our sample for which microsatellite data are available, two were consistently assigned to the same population by the three methods used by Russello, Poulakakis, et al. (2010), whereas two obtained conflicting assignments. Russello, Poulakakis, et al. (2010) analyzed 156 individuals of unknown
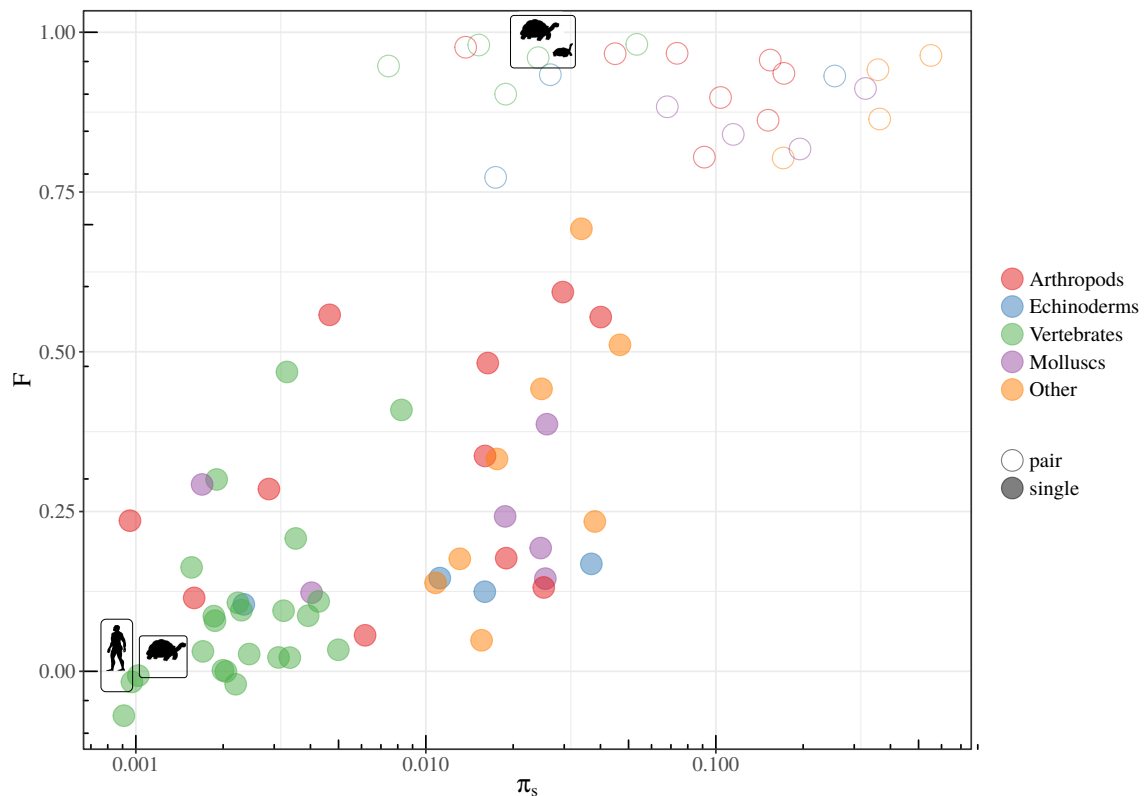
Figure 1 – Each filled circle corresponds to a single species of animals in which the transcriptome of at least four individuals was sequenced (Perry et al., 2012; Romiguier, Gayral, et al., 2014). Harvest ant *Messor barbarus* was excluded due to its highly specific mating system (Romiguier, Fournier, et al., 2017). Each open circle corresponds to a pair of congeneric species of animals in which the transcriptome of at least five individuals was sequenced (Galtier, 2016). Genotypes were called using the improved, contamination-aware algorithm of Ballenghien et al. (2017), parameter γ being set to 0.2. X-axis (log scale) reflects the amount of neutral genetic variation within the sample. This is measured by the synonymous-site heterozygosity, π_S, for single species, and by the proportion of fixed synonymous differences for species pairs. Y-axis reflects the detected amount of population substructure, measured by the F statistics. The positions in the graph of *C. nigra*, *Homo sapiens* (single species) and of the *C. nigra/Chelonoidis carbonaria* pair are indicated by thumbnails.

geographic origin. Re-analysing their Table S1, we found that 109 of these (70%) were assigned to two or three distinct populations by the three methods they used. Our sample, therefore, is not particularly enriched in individuals of "hybrid origin". Still, to control for a possible impact of hybrids in our analysis, we calculated the maximal F across the ten possible pairs of individuals from our sample. If any two individuals from our sample belonged to distinct species, this statistics should take a high value. We rather found that the maximal pairwise F equals a low 0.12, indicating that the absence of homozygote excess in *C. nigra* (Figure 1) is not explained by the presence of a few "hybrids" in our sample, but rather reflects an overall lack of genetic differentiation.

Discussion

The sample of Loire et al. (2013) is small in terms of the number of individuals and obviously insufficient to provide a definitive answer regarding population genetic structure in *C. nigra*. Our sample, however, is large with regard to the number of loci. Distinct loci independently sample the coalescent process, and therefore provide small but many pieces of information on population history. The strength of even small-sized samples to reconstruct population processes,

provided that a large number of loci is available, has been demonstrated in a number of recent publications (e.g. Abascal et al., 2016; Der Sarkissian et al., 2015). The popular PSMC method (Li and Durbin, 2011), for instance, estimates former fluctuations in effective population size based on a single diploid genome. Roux et al. (2016) specifically demonstrated the potential of the Romiguier, Gayral, et al. (2014) data set to inform on species boundary and the history of gene flow between populations.

Our results, if confirmed, are highly relevant for the conservation of Galápagos tortoises. They could imply that the current practice whereby only individuals from the same "species" are crossed is not only pointless – in the absence of truly differentiated gene pools – but also potentially harmful to the species due to the risk of inbreeding depression. The risk is particularly high when large numbers of closely related individuals are simultaneously released in a closed area (Milinkovitch, Monteyne, Gibbs, et al., 2004). In the same vein, physical removal from the field of naturally occurring 'hybrids' (Milinkovitch, Monteyne, Russello, et al., 2007) and sterilization of individuals to avoid undesired crosses (Hunter et al., 2013) might well actively participate in the decline in the mean fitness of local populations (Frankham et al., 2014).

Of note, even the most recent publications on this issue, including the description of a new species (Poulakakis, Edwards, et al., 2015) have focused on the same limited set of markers developed some 15 years ago. We suggest that, nowadays in the genomic era, the conservation policy regarding Galápagos tortoises should no longer be based entirely on just a dozen loci, given the demonstrated potential of genome-wide data to enhance population genetics research in other groups of threatened animals (e.g. Abascal et al., 2016; Der Sarkissian et al., 2015; Dobrynin et al., 2015; Lamichhaney et al., 2015; Liu et al., 2014; Moore et al., 2014). Knowing the importance of inbreeding depression as a threat to conservation of endangered species (e.g. Allendorf et al., 2010; Frankham, 2015; Walling et al., 2011) we suggest that forced inbreeding and sterilization interventions in Galápagos tortoises be delayed until the genetic architecture of the species is clarified via the analysis of genome-wide data in a sufficiently large sample, thus resolving the discrepancy between existing datasets.

There is some controversy and subjectivity in the definition of relevant conservation units and strategies (Edmands, 2007; Funk et al., 2012; Shafer et al., 2015; Taylor et al., 2017). Giant Galapagos tortoises are maximally protected throughout the archipelago. The definition of conservation units in *C. nigra*, therefore, is not relevant to the preservation of local populations. Rather, genetic data have been used to make decisions on which individuals should be crossed with which, and where they should be located. This is only meaningful if the distinct entities are sufficiently isolated, genetically and geographically, such that gene flow is naturally low. This is why species boundaries matter: current conservation goals in *C. nigra* are dependent on the validity of the existing population model and taxonomy.

It is likely that human intervention during the last centuries has not only dramatically affected *C. nigra* population density, but also favoured migration of individuals across islands. It is also quite likely, based on existing data, that gene flow between islands and/or local extinction/recolonization was substantial before human intervention, preventing speciation in *C. nigra* – otherwise hybrids would not be so common, total genetic diversity/divergence would be higher, population substructure would be stronger and more obviously correlated to morphology. In other words, *C. nigra* might well be better described as a perturbed metapopulation than a collection of endangered species. If this was true then the goals of resurrecting ancestral "lineages" and preventing crosses between genetic entities would just be nonsense. Only a genome-wide analysis has the power to discriminate between ancient and recent gene flow, clarify the link between genotype and phenotype, and inform conservation policies. Current management aims at preserving/restoring particular combinations of alleles at 12 specific loci; relevance to phenotypes, fitness, population history and species barriers in *C. nigra* is anything but clear.

Data availability

Supplementary Table S1 can be downloaded from `https://zenodo.org/record/5571874`

Conflict of interest disclosure

The authors declare that they have no financial conflict of interest with the content of this article. Nicolas Galtier is a recommender for PCI Evolutionary Biology and PCI Genomics.

References

Abascal F et al. (2016). *Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biology* **17**, 251. https://doi.org/10.1186/s13059-016-1090-1.

Aguilera WT et al. (2015). *Conservation: Giant tortoises hatch on Galapagos island. Nature* **517**, 271. https://doi.org/10.1038/517271a.

Allendorf FW et al. (2010). *Genomics and the future of conservation genetics. Nature Reviews Genetics* **11**, 697–709. https://doi.org/10.1038/nrg2844.

Ballenghien M et al. (2017). *Patterns of cross-contamination in a multi-species population genomic project: detection, quantification, impact, solutions. BMC Biology* **15**, 25. https://doi.org/10.1186/s12915-017-0366-6.

Bazin E et al. (2006). *Population size does not influence mitochondrial genetic diversity in animals. Science* **312**, 570–572. https://doi.org/10.1126/science.1122033.

Caccone A, Gentile G, Burns CE, et al. (2004). *Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. Molecular Phylogenetics and Evolution* **31**, 794–798. https://doi.org/10.1016/j.ympev.2004.02.004.

Caccone A, Gentile G, Gibbs JP, et al. (2002). *Phylogeography and history of giant Galápagos tortoises. Evolution* **56**, 2052–2066. https://doi.org/10.1111/j.0014-3820.2002.tb00131.x.

Cayot LJ (2008). *The restoration of giant tortoises and land iguana populations in Galápagos. Galápagos Research* **65**, 39–43.

Chiari Y et al. (2009). *Morphometrics parallel genetics in a newly discovered and endangered taxon of Galápagos tortoise. PLoS One* **4**, e6272. https://doi.org/10.1371/journal.pone.0006272.

Ciofi C et al. (2002). *Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises. Molecular Ecology* **11**, 2265–2283. https://doi.org/10.1046/j.1365-294x.2002.01617.x..

Der Sarkissian C et al. (2015). *Evolutionary genomics and conservation of the endangered Przewalski's horse. Current Biology* **25**, 2577–2583. https://doi.org/10.1016/j.cub.2015.08.032.

Dobrynin P et al. (2015). *Genomic legacy of the African cheetah, Acinonyx jubatus. Genome Biology* **16**, 277. https://doi.org/10.1186/s13059-015-0837-4.

Edmands S (2007). *Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Molecular Ecology* **16**, 463–475. https://doi.org/10.1111/j.1365-294X.2006.03148.x.

Edwards DL, Benavides E, et al. (2013). *The genetic legacy of Lonesome George survives: Giant tortoises with Pinta Island ancestry identified in Galápagos. Biological Conservation* **157**, 225–228. https://doi.org/10.1016/j.biocon.2012.10.014.

Edwards DL, Garrick RC, et al. (2014). *Cryptic structure and niche divergence within threatened Galápagos giant tortoises from southern Isabela Island. Conservation Genetics* **15**, 1357–1369. https://doi.org/10.1007/s10592-014-0622-z.

Frankham R (2015). *Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Molecular Ecology* **24**, 2610–2618. https://doi.org/10.1111/mec.13139.

Frankham R et al. (2014). *Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation* **170**, 56–63. https://doi.org/10.1016/j.biocon.2013.12.036.

Funk WC et al. (2012). *Harnessing genomics for delineating conservation units. Trends in Ecology and Evolution* **27**, 489–496. https://doi.org/10.1016/j.tree.2012.05.012.

Galtier N (2016). *Adaptive protein evolution in animals and the effective population size hypothesis.* PLoS Genetics **12**, e1005774. https://doi.org/10.1371/journal.pgen.1005774.

Garrick RC, Benavides E, Russello MA, Gibbs JP, et al. (2012). *Genetic rediscovery of an 'extinct' Galápagos giant tortoise species.* Current Biology **22**, R10–1. https://doi.org/10.1016/j.cub.2011.12.004.

Garrick RC, Benavides E, Russello MA, Hyseni C, et al. (2014). *Lineage fusion in Galápagos giant tortoises.* Molecular Ecology **23**, 5276–5290. https://doi.org/10.1111/mec.12919.

Gayral P et al. (2013). *Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap.* PLoS Genetics **9**, e1003457. https://doi.org/10.1371/journal.pgen.1003457.

Gibbs JP et al. (2014). *Demographic outcomes and ecosystem implications of giant tortoise reintroduction to Española Island, Galapagos.* PLoS One **9**, e110742. https://doi.org/10.1371/journal.pone.0110742.

Hunter EA et al. (2013). *Equivalency of Galápagos giant tortoises used as ecological replacement species to restore ecosystem functions.* Conservation Biology **27**, 701–709. https://doi.org/10.1111/cobi.12038.

Lamichhaney S et al. (2015). *Evolution of Darwin's finches and their beaks revealed by genome sequencing.* Nature **518**, 371–375. https://doi.org/10.1038/nature14181.

Li H, Durbin R (2011). *Inference of human population history from individual whole-genome sequences.* Nature **475**, 493–496. https://doi.org/10.1038/nature10231.

Liu S et al. (2014). *Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears.* Cell **157**, 785–794. https://doi.org/10.1016/j.cell.2014.03.054.

Loire E et al. (2013). *Population genomics of the endangered giant Galápagos tortoise.* Genome Biology **14**, R136. https://doi.org/10.1186/gb-2013-14-12-r136..

Marquez C et al. (2013). *Population response of Giant Galápagos Tortoises to feral goat removal.* Restoration Ecology **2**, 181–185. https://doi.org/10.1111/j.1526-100X.2012.00891.x.

Milinkovitch MC, Monteyne D, Gibbs JP, et al. (2004). *Genetic analysis of a successful repatriation programme: giant Galápagos tortoises.* Proceedings of The Royal Society B Biological Sciences **271**, 341–345. https://doi.org/10.1098/rspb.2003.2607.

Milinkovitch MC, Monteyne D, Russello MA, et al. (2007). *Giant Galápagos tortoises; molecular genetic analyses identify a trans-island hybrid in a repatriation program of an endangered taxon.* BMC Ecology **7**, 2. https://doi.org/10.1186/1472-6785-7-2.

Moore JS et al. (2014). *Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci.* Molecular Ecology **23**, 5680–5697. https://doi.org/10.1111/mec.12972.

Nichols H (2015). *Lonesome George 'look-alike' captured for tortoise-breeding effort.* Nature News. https://doi.org/10.1038/nature.2015.18955.

Perry GH et al. (2012). *Comparative RNA sequencing reveals substantial genetic variation in endangered primates.* Genome Research **22**, 602–610. https://doi.org/10.1101/gr.130468.111.

Poulakakis N, Edwards DL, et al. (2015). *Description of a new Galapagos giant tortoise species (Chelonoidis; Testudines: Testudinidae) from Cerro Fatal on Santa Cruz Island.* PLoS One **10**, e0138779. https://doi.org/10.1371/journal.pone.0138779.

Poulakakis N, Glaberman S, et al. (2008). *Historical DNA analysis reveals living descendants of an extinct species of Galápagos tortoise.* Proceedings of the National Academy of Sciences of the USA **105**, 15464–15469. https://doi.org/10.1073/pnas.0805340105.

Poulakakis N, Russello MA, et al. (2012). *Unravelling the peculiarities of island life: vicariance, dispersal and the diversification of the extinct and extant giant Galápagos tortoises.* Molecular Ecology **21**, 160–173. https://doi.org/10.1111/j.1365-294X.2011.05370.x.

Romiguier J, Fournier A, et al. (2017). *Convergent evolution of social hybridogenesis in Messor harvester ants.* Molecular Ecology **26**, 1108–1117. https://doi.org/10.1111/mec.13899.

Romiguier J, Gayral P, et al. (2014). *Comparative population genomics in animals uncovers the determinants of genetic diversity.* Nature **515**, 261–263. https://doi.org/10.1038/nature13685.

Roux C et al. (2016). *Shedding light on the grey zone of speciation along a continuum of genomic divergence.* Biology **14**, e2000234. https://doi.org/10.1371/journal.pbio.2000234.

Russello MA, Beheregaray LB, et al. (2007). *Lonesome George is not alone among Galápagos tortoises. Current Biology* **17**, R317–318. https://doi.org/10.1016/j.cub.2007.03.002.

Russello MA, Poulakakis N, et al. (2010). *DNA from the past informs ex situ conservation for the future: an "extinct" species of Galápagos tortoise identified in captivity. PLoS One* **5**, e8683. https://doi.org/10.1371/journal.pone.0008683.

Shafer AB et al. (2015). *Genomics and the challenging translation into conservation practice. Trends in Ecology and Evolution* **30**, 78–87. https://doi.org/10.1016/j.tree.2014.11.009.

Taylor BL et al. (2017). *Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans. Marine Mammal Science* **33**, 12–26. https://doi.org/10.1111/mms.12413.

Walling CA et al. (2011). *Inbreeding depression in red deer calves. BMC Evolutionary Biology* **11**, 318. https://doi.org/10.1186/1471-2148-11-318.

Weir BS, Cockerham CC (1984). *Estimating F-statistics for the analysis of population structure. Evolution* **38**, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.