Understanding evolution through genomic analysis of extinct species

Pere Renom i Vilaró

TESI DOCTORAL UPF / 2022

DIRECTOR DE LA TESI

Dr. Carles Lalueza-Fox

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat Pompeu Fabra Barcelona

Cover design:
Elisenda Nogué (Metagràfica)

Cover image:
Frayed string as a metaphor for damage in ancient DNA, the main challenge of paleogenomics.

"*The tree of life should perhaps be called the coral of life, base of branches dead*"

Charles Darwin, (1859)

A les branques més properes
d'aquest meravellós corall,
els meus pares, les meves filles i la Gemma

Agraïments

Inspirat en un filòsof medieval, el físic Isaac Newton va escriure: “Si he vist més lluny és perquè estava assegut damunt les espatlles d’un gegant”. Es referia al fet que tot progrés es fonamenta en les generacions anteriors. Sense les aportacions dels científics que el van precedir, probablement Newton no hauria mai descobert la gravitació universal.

En un àmbit radicalment diferent hi trobem la tradició catalana dels castellers, on es produeix literalment el que Newton assenyalava metafòricament. Gràcies a la col·laboració de centenars de persones, fent pinya i ajudant-se els uns als altres, l’enxaneta s’hi enfila al capdamunt i pot tocar el cel.

Una tesi és també una obra col·lectiva. El resultat final és superior a la suma de les parts. Aquestes pàgines, i aquestes paraules culminen un projecte personal que, amb nombroses i dilatades interrupcions, s’ha anat gestant durant més de 40 anys. Al llarg del llarg camí, he estat acompanyat per moltes persones que m’han recolzat i ajudat, directament, materialment, tècnicament i emocionalment. I voldria expressar la meva més profunda gratitud.

Com que són moltes, les agrupo en colles, com els castelleres. El primer agraïment és pel cap de colla, el director d’aquesta tesi, Carles Lalueza, que ha dirigit amb encert, compromís, generositat i paciència a un individu, sovint excessivament entusiasta i accelerat, que a més, no provenia del món de la genètica. Moltes gràcies Carles per haver confiat en mi i haver-me donat aquesta meravellosa oportunitat. Has aconseguit que es fes realitat un dels meus projectes vitals.

Els naturalistes empordanesos amb els que he descobert mil racons de la costa Brava tant dins com fora de l’aigua, el Cap de Creus, l’Albera, el Canigó, els Aiguamolls..., i també l’anellament d’ocells, la pesca elèctrica, la pesca de tonyines amb el Tio Gel, la mítica aventura de les foques a les paradisíaques illes gregues de Milos, Kimolos i Poliegos..., Gràcies amics! Ponç Feliu (Póntiakor), Mercè Sánchez, Pepe Esteban, Dani Schneider, Jordi Sargatal, Aleix Comas, Toni Llobet, Joan Carles Gimisó i Deli Saavedra.

Als companys de l'Escola del Mar de Badalona, on vaig aprendre a tenir cura dels aquaris i vaig descobrir molts aspectes bàsics de la biologia dels peixos del Mediterrani, Enric Cahner, Gregori Muñoz-Ramos, Xavi Seglar i Francesc Camps. L'aventura d'ensinistrar un pop és molt difícil de superar!

Als companys de la carrera de biologia de la UB amb els que vaig connectar des del primer dia, i encara avui ens veiem de tant en tant, Marc Robert, Montse Roura, Ester Nin, Sara Pérez, Elena Palmero, Xavier de Pedro i Pep Arcos.

A la gent del Departament d'Ecologia de la UB, on m'hi vaig estar 10 anys coordinant la Xarxa de Vigilància dels Alguers, i vaig iniciar una tesi doctoral sobre l'ecologia i la paleoecologia de *Posidonia oceanica,* que va quedar inacabada. Pels dos directors d'aquella tesi Javier Romero i especialment Miguel Ángel Mateo, amb qui he compartit estada a Holanda, incomptables campanyes a les Illes Medes i a Port Lligat, workshop a Còrsega, i el projecte Medcore que ens han portat de la Toscana a Egipte, Tunísia, Marroc i Tabarca. Quins records! Muchas gracias por tu gran amistad estimado Migüel! També a la resta de membres del grup de fanerògames marines Marta Manzanera, Marta Pérez, Olga Invers, Fiona Tomàs i Teresa Alcoverro, a les que conec més vestides amb neoprè que amb roba de carrer. Sense oblidar els altres ecòlegs amb els que vaig compartit despatx, laboratori o campanyes, Biel Obrador, Blanca Rios, Susana Bernal, Mª Ángeles Gallegos, Mireia Vila, Núria Bonada, Sergi Pla, Rosa Casanovas, Tura Puntí, Joan Gomà, Cesca Casadesus, Juan Carlos García, Marta Comerma, Jaume Piera. També a l'Óscar Arribas, un erudit i expert en aconseguir llibres antics o descatalogats. I a l'Adriana Vergés una apassionada de la biologia marina i la música, amb qui vaig compartir workshop en un memorable viatge a Austràlia. Al final he acabat la tesi, ho veus? Encara que d'un altre tema...

Als snobs, un grup peculiar de biòlegs del qual sempre han sorgit converses interessants i enriquidores, Ester Gaya, Miquel de Cáceres, Enrique Navarro i Miquel Àngel Rodríguez. Són graminívors els humans?

Pel grup de pares de filles, encertadament batejat “Pares amb cognom sense continuïtat” i rebatejat “Big data”. En Fede Bartumeus (Von Nietzsche), Rafa Marcé i Simone Mariani. A banda de la passió per la ciència, també compartim sopars i trobades familiars, com la ja tradicional sortida en barca a Cadaqués amb en Fede i família. És un luxe fer-te consultes de la tesi a la Cala Bona, Fede! Menció a part mereix en Simone (Dottore in Alaska, Barelli, Simon-Barr), amb qui sense dubte més he parlat de filosofia de la ciència, padrí de la meva filla i alter ego “tu Simó, a partir d’ara t’anomenaràs Pere...”.

Al companys de TV3 amb els que he compartit mitja vida en la sensacional aventura del programa *Quèquicom*, el millor màster de cultura científica que us pugueu imaginar. La Cari Pardo, la meva estimadíssima Cari (Antoine Merie), que em coneix tant com ma mare, quantes coses hem fet i après plegats!! La Virginia Gran (Immensa) generosa, eficient i afectuosa, en Miquel Piris divertit com pocs, i tota la resta de la tropa Georgina Pujol, Samantha Vall, Roger Caubet, Dani Vallvé, Griselda Guiteras, Marc Boada, Marc Durà, Míriam Martí, Núria Carner, Núria Roig, Toni Galobardes, Mercè Managuerra, Jaume Vilalta, Lluís Baldomero, Dani Gresely. I com no, Dani Jiménez, que m’ha introduït i guiat en el món de la divulgació científica experimental. Hi ha molt I+D darrere de cada experiment a la televisió. Ets un crack connectant amb el públic!

Als tots els companys actuals de TV3 al programa *Tot es mou*, i especialment a la Míriam Riau (Afrodita, Tatiana, Agatòclia) gran subministradora de notícies científiques, també Jordi Gil, Eloi Cordomí, Laia Ferrer, Ester Bertran, Dimas Rodríguez, Montse Tió, Peté, Laura Freixes, Elsa Sabater, Ester Alloza i Albert Aguilera. Gràcies Albert per accedir amablement a que m’agafés el permís laboral, sense el qual no hauria pogut acabar aquesta tesi.

Als filòsofs Cel·la Mondéjar, Cèsar Cabrera i Elisenda Nogué, la qual ha dissenyat les cobertes de dos dels meus llibres i d’aquesta tesi, això ja és una tradició a mantenir!

A l’editorial Cossetània, especialment la Marta Ferré, en Jordi Ferré i en Josep Maria Olivé, per la vostra generositat en l’edició del meu llibre *Ciència sota el focus*. També per haver-me regalat la foto de la coberta d’aquesta tesi.

D'ençà que vaig descobrir l'equitació, la meva felicitat s'ha ampliat sensiblement. Muntar és una activitat climàcica. Els meus entrenadors mereixen ser reconeguts Lluís Matamala, Oliver Singleton i Albert Puig. No he pogut evitar dedicar un capítol de la tesi al cavall…

El petit i exclusiu grup dels "Berberetxus" Guillem Masdeu, i Eva Santana. Amb en Guillem vaig fer natació molts anys, activitat que, per la seva duresa, forja l'esperit i permet suportar qualsevol adversitat vital ulterior, com escriure una tesi. I amb l'Eva vaig compartir un any de seccions setmanals de ciència en anglès al diari Ara amb "Ara kids". Algun dia publicarem un llibre plegats!

Els "Peus a la galleda" un curiós grup integrat per Jordi García i Uri Soler, ens uneix la política, l'antropologia i les tortugues. Quins grans temes!

Al grup "MB2010 & beyond" Pito Costa i Miquel Crisol amb qui he fet molts cims pirinencs i el mític Montblanc, del qual també va sorgir una publicació de ciència.

En l'apartat miscel·lània o calaix de sastre, incloc amics amb els quals he compartit moments important que d'alguna manera també han contribuït en aquesta tesi, Xavier Vallès en natació i epidemiologia, especialment durant l'etapa Covid, Jordi Grífols el meu veterinari de capçalera, especialista en lloros, per la qual cosa ha tingut un paper important en l'article de la cotorra de Carolina. La família Giribet, especialment en Josep, un artista-científic completament heterodox amb el que vaig compartir el projecte de Visita Virtual las Parcs Naturals de Catalunya. A Carles Salas dissenyador de la meva web i d'algunes boniques figures d'aquesta tesi. Lluís Dalmau, Francesc Oliveras, Jordi Soler, Antonio Peña, Violeta Moizé, Robert Flo, Gorka Muñoa, i Cristina Llorens, una font d'inspiració. Carolina Terrisse amb qui vaig fer un documental innovador sobre el VIH. Ricard Martín, el meu professor de biologia a les Escoles Minguella, i el seu germà David, biòleg de l'IBE amb qui he tingut converses molt estimulants. Ignasi Ribas i Rubén Duro dues *rara avis* amb qui també és un plaer conversar de ciència i als quals he dedicat dos reportatges del *Quèquicom*.

“La màfia” és la meva colla d’amics de tota la vida, ens coneixem des que érem espermatozous. Gràcies Joan Sánchez, Lluís Zamora, Josep Ricart, i Jordi Giralt, per tot el que hem compartit. Sou com germans.
Els companys de l’Institut de Biologia Evolutiva (IBE), Toni De Dios, el meu principal benefactor, m’ha ajudat en tot i més, i sempre ho ha fet amablement. Ets un fenòmenu! També a Pablo Carrión, Iñigo Olalde, Claudia Fontserè, Laia Llovera i Pere Gelabert. Casualitats de la vida abans de començar la tesi a l’IBE ja havia fet reportatges del *Quèquicom* amb David Comas sobre genètica humana, Francesc Calafell sobre genètica i genealogia dels Renom, i Tomàs Marquès sobre el gat calicó. Vull també agrair a Joan Carles Senar (amb el qual també vaig fer un report sobre cotorres), Tomàs Marquès, Salvador Carranza, Elena Bosch i Santi Mañosa per haver accedit a formar part del tribunal d’aquesta tesi.
A Isaac Casanovas, Roger Vila, i Juan Carlos Rando, que han revisat diverses seccions d’aquesta tesi amb molt deteniment, sense tenir-ne cap obligació, i han contribuït a que millorés sensiblement.
A Ramon i David Masferrer de Masjoan (Espinelves), on es conserva l’exemplar de l’extingida cotorra de Carolina que va portar dels EUA el seu avantpassat Marià Masferrer. Com l’efecte papallona, el seu redescobriment va alterar inesperadament el curs de la meva vida i la d’altres persones del meu entorn, i ens ha portat fins aquí. I a Martí Boada per haver inventariat i classificat l’exemplar fa dècades.
Quan hem repassat totes les colles castelleres que conformen la pinya, les manilles i el tronc del castell, arribem al pom de dalt, on hi trobem la meva estimada família. Els meus avis (encara recordo l’enuig de la meva àvia paterna, quan li vaig ensenyar uns dibuixos meus de l’evolució, -que encara conservo- perquè com a cristiana els considerava ofensius, o els avis materns, que ens van deixar omplir el seu jardí de tortugues, tórtores, gallines de totes les races, pintades, galls d’indi...), els Urgell, els meus sogres Joaquim Urgell i Marta Velázquez, els meus cunyats Joaquim, Xavi, Martí, Pilar Cucart, Cristina Herrada, Alba Barnet, Violant Molina i Carol

López, els nebots Arnau, Oriol, Òria, Annet, Armand, Emma i Èric. El cosí Lluís Parés, i els oncles Teresa Vilaró i Pere Miquel Parés, mentor durant el meu debut naturalista, i amb qui vaig descobrir l'ornitologia i l'equitació! Els meus germans Albert i Jordi, que m'han patit com ningú.
I finalment arriben els acotxadors, els meus pares Pere i Montse, que m'estimen amb desfici i sempre m'han animat a ser allò que volgués ser, fins i tot quan deia, de ben petit, que faria un doctorat en biologia. Les meves filles Gal·la i Gina, l'exemple viu que l'evolució ha arribat al seu clímax (Darwin estaria desesperat), i la Gemma que m'estima i em fa costat lleialment encara que la ciència li interessi menys que una úlcera d'estómac.
Moltes gràcies a tots per haver contribuït a que aquesta tesi sigui una realitat. Us estimo. Faig l'aleta i me'n torno cap avall, no sigui que la gravetat em faci baixar de cop.

Badalona, 7 de desembre de 2022.

Abstract

According to estimates, over the 3.5 billion years that have elapsed since the appearance of life, the Earth has been inhabited by some 4 billion different species, of which 99.99% have already become extinct. Extinction is a fundamental process in evolution. Traditionally, extinct species have been studied by paleontologists based on fossil remains. With the emergence of paleogenomics, there has been a true revolution in the study of evolution. In this thesis we recover the complete nuclear genome of two extinct species, the Carolina parakeet (*Conuropsis carolinensis*), distributed in the southeastern United States and extinct in 1918, and the butterfly Xerces Blue (*Glaucopsyche xerces*) native to the coastal sand dunes of San Francisco and extinct in 1941. We also recover partial nuclear and mitochondrial genome of the Tenerife giant rat (*Canariomys bravoi*), an endemic species of Tenerife (Canary Islands) that it became extinct after the fourth century BCE. Thanks to the ancient DNA we have been able to carry out precise phylogenies of the three species, analyze possible signs of genetic erosion, explore the functionality of some genes, estimate the evolutionary rate and reconstruct the demographic history, in order to elucidate the dynamics of population decline. In brief, we have answered specific adaptive questions and improved our understanding of human-mediated extinction.

Resum

Segons les estimacions, al llarg dels 3,500 milions d'anys transcorreguts d'ençà de l'aparició de la vida, la Terra ha estat habitada per uns 4,000 milions d'espècies diferents, de les quals el 99.99% ja s'han extingit. L'extinció és un procés fonamental en l'evolució. Tradicionalment les espècies extingides han estat estudiades per paleontòlegs a partir de les restes fòssils. Amb el sorgiment de la paleogenòmica s'ha produit una veritable revolució en l'estudi evolutiu. En aquesta tesi recuperem el genoma nuclear complert de dues espècies extingides la cotorra de Carolina (*Conuropsis carolinensis*), distribuida pel sud est dels EUA i extingida el 1918, i la papallona Xerces Blue (*Glaucopsyche xerces*) originària de les dunes costaneres de San Francisco i extingida el 1941. També recuperem el genoma nuclear i mitocondrial parcial de la rata gegant de Tenerife (*Canariomys bravoi*), una espècie endèmica de Tenerife (Illes Canàries) que es va extingir després del segle IV aC. Gràcies a l'ADN antic hem pogut dur a terme filogènies precises de les tres espècies, analitzar possible signes d'erosió genètica, explorar la funcionalitat d'alguns gens, estimar la taxa evolutiva i reconstruir la història demogràfica, per tal de dilucidar la dinàmica de la davallada poblacional. En definitiva, hem donat resposta a qüestions adaptatives específiques i hem millorat la nostra comprensió de l'extinció mediada pels humans.

Index

1. INTRODUCTION

1.1. Extinction

1.1.1. Why focus on extinction

In the late 18th century, more than 50 years before Charles Darwin's Origin of Species was published (Darwin, 1859), some natural philosophers independently proposed the theory that ontogeny recapitulates phylogeny. After a good initial acceptance among the scientific community, new discoveries made it fall into disrepute and it ended up being abandoned. Today, a non-literal version of that theory has been reclaimed by evolutionary and developmental biology (evo-devo).

Certainly one only has to observe the development of various vertebrates at different stages to realize that there are quite a few similarities. Some common process they must share, even if a complete recapitulation does not take place.

But under a closer inspection, the recapitulation also occurs in another sense. The formation of the organism is indisputably progressive. After syngamy, the diploid zygote begins to divide exponentially and little by little various sets of cells differentiate into tissues and organs. When can we say that an amalgam of cells is already a human, a chicken or a mouse? The border is fuzzy and perhaps even arbitrary. On the contrary, we can say with certainty when a human, a chicken or a mouse ceases to be: when they die. Development is gradual and diffuse, while death is sudden and sharp. And what is true for individual organisms (ontogeny) is also true for species as a whole (phylogeny). Species arise by progressive differentiation from other species, while they disappear suddenly when they become extinct.

Let's see a couple of examples. In the North Atlantic there was a marine bird very similar to penguins called the giant auk (*Pinginus impennis*). Its origin dates back to the Miocene about 7 million years ago, when it was differentiated from the razorbill (*Alca torda*) (Smith & Clarke 2015). After suffering indiscriminate hunting for

centuries by humans, on July 3, 1844, commissioned by a Danish collector and merchant, some fishermen killed the last confirmed couple, who were placidly breeding on Eldey Island, Iceland, oblivious to the tragedy that befalls them (Kjartansdóttir 2019). The species had gradually emerged from other species, reached a population of millions of individuals, and became extinct forever (Figure 1.1).

Even if a species has a huge population, over time (millennia, centuries or decades) it eventually disappears completely. And that dynamic has existed since the origin of life. About 150 million years ago, at the end of the Jurassic, a group of carnivorous dinosaurs named tyrannosaurs emerged and gradually diversified. One of the last of the lineage was the famous *Tyrannosaurus rex*, which lived between 68 and 66 million years ago (note that it arose about 82 million years after the appearance of its family). It weighed between 4 and 7 tons, reached a global population of about 20,000 individuals, persisted for ~127,000 generations, and accumulated a total number of ~2.5 billion of individuals (Marshall et al., 2021), and yet at some point of the Cretaceous, the last specimen died and the species became extinct. It literally disappeared overnight, now it's there, and now it's gone (Figure 1.1).

Figure 1.1. Pair of Great Auks drawn by John James Audubon, in The Birds of America (1827-1838). Note the morphological similarity with the penguins of the southern hemisphere (family *Spheniscidae*) despite not being directly related, as a result of evolutionary convergence. *Tyrannosaurus rex* skull found in 1988 by Kathy Wankel in eastern Montana, and excavated by the paleontologist Jack Horner. The specimen is one of the most complete ever discovered, with 80-85% of the skeleton recovered.

Behind this temporal asymmetry between the appearance (gradual) and disappearance (sudden) of a species is obviously entropy.
In the same way that it can take years to build a skyscraper and tear it down in an instant, the emergence of a species is costly because it requires the investment of energy (anti-entropic), while its extinction occurs spontaneously when the invested energy decays (entropic).
Studying extinction is therefore particularly relevant to understanding evolution since it focuses on a phenomenon, sharp and clearly defined in time. But there are more reasons.
The evolutionary biologist Stephen Jay Gould in his book *Bully for Brontosaurus* (Gould, 1992) analyzed the paradigmatic case of the evolution of the horse. He explained that in 1870 the biologist and anthropologist Thomas Henry Huxley conceived a linear evolution of the horse as the clearest proof to reinforce the theory of his colleague Charles Darwin. Initially he placed it in Europe but when he saw the impressive collection of fossils of the North American paleontologist Othniel C. Marsh, he became convinced that the horse had basically evolved in North America. From the collaboration of both scientists emerged the first schemes showing the well-known triple evolutionary tendency of the horse: increase in body size, probably to reduce the number of predators (see section 1.1.3), reduction in the number of fingers to improve speed, and increased height and complexity of molars to adapt to grazing grasses (Figure 1.2). However, in the 150 years that have passed since then, the number of equid fossils found has increased dramatically and forced a redrawing of the evolution of the horse. Now, it is no longer linear but is structured in a bushy and complex shape with many truncated branches (Figure 1.3). The key to the initial misinterpretation is the lack of sufficient information. If a particular lineage survives to the present day with a single representative, we will tend to classify the related fossils as direct ancestors and order them in a linear evolution.

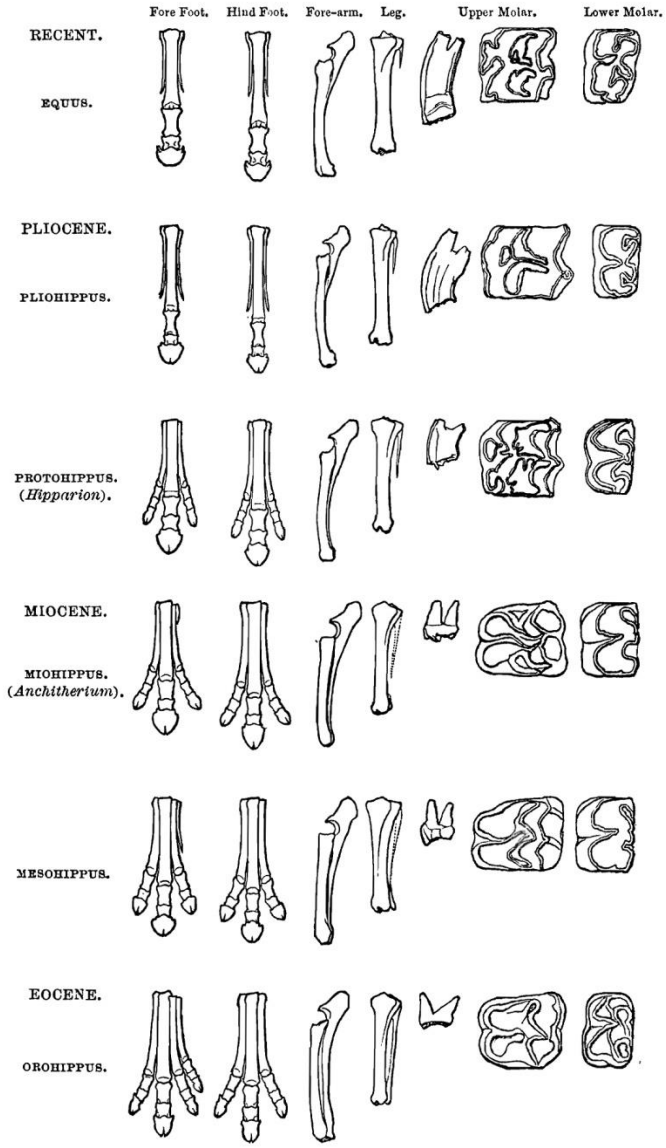


Figure 1.2. Linear evolution of the horse (from bottom to top) based on the progressive loss of the number of fingers and the increase in the complexity of the dental crown [from Marsh, 1879].

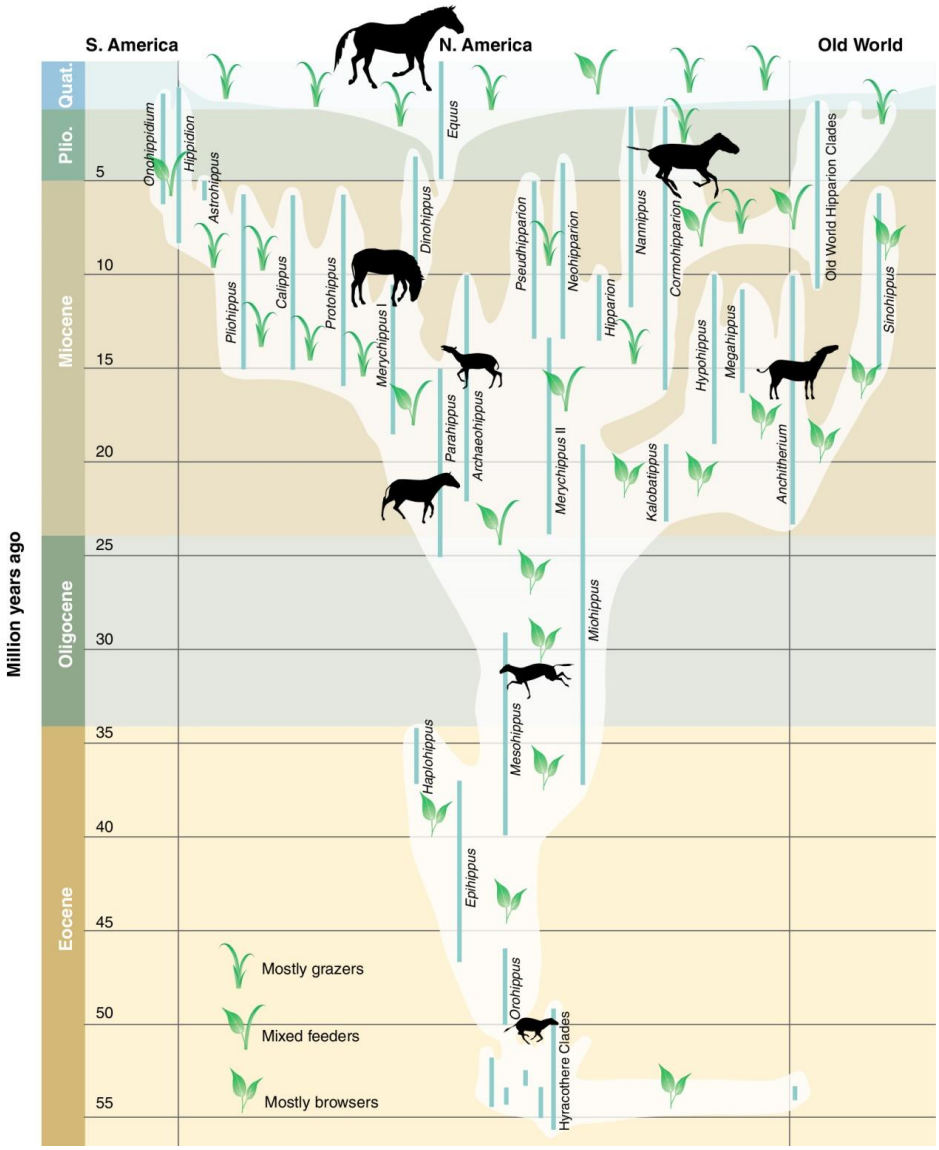


Figure 1.3. “Bushy” evolution the Family Equidae as a result of advances in knowledge from fossils. Phylogeny, geographic distribution, diet, and body sizes over the past 55 My.The vertical lines represent the actual time ranges of equid genera or clades.The first ~35 My (Eocene to early Miocene) of horse phylogeny are characterized by browsing species of relatively small body size.The remaining ~20 My (middle Miocene until the present day) are characterized by genera that are either primarily browsing/grazing or are mixed feeders, exhibiting a large diversification in body size [from MacFadden, 2005].

We have also encountered this problem in the evolutionary study of our own species that we have established in the canonical series: *Australopithecus- Homo habilis- Homo erectus- Homo sapiens* which is associated with the skewed teleological myth of humans located at the top of the evolution As we discover more and more fossils we can begin to understand the diversity of different lineages over time and turn a stunted trunk into a lush bush. Consequently, extinct species allow us to understand the true structure and branching of evolutionary trees.

Another myth that has harmed the study of evolution has been the race for survival, from which we have concluded that living species are the winners, while extinct species are the losers. In reality, it cannot be said that none of the species that populated the Earth were less adapted to their environment than the extant ones are today. All species are winners while they exist and losers when they disappear. The extinction is not necessarily linked to a loss or decrease in fitness, several stochastic phenomena are also involved. Therefore, studying extinct species provides complementary evolutionary information to the study of living species. Like life and death, they are two sides of the same coin.

But there is still one last and powerful reason to study extinct species. The seemingly trivial question of what is the biodiversity of our planet, has no precise or satisfactory answer. Scientists have no way of counting the number of species of living things today simply because it is too large. There have been, however, various estimates. In 1988, evolutionary biologist Robert May observed that the diversity of land animals is negatively correlated with size (May, 1988). He reasoned then that we have probably found most species of large animals, such as mammals and birds, but neglected the vast majority of small animals. So he used the known diversity of the large ones to calculate the diversity of the smaller ones. The estimate he got is between 10 and 50 million species of land animals, a range so wide that it's hard to get it wrong. A more recent estimate adjusts the number to 8.7 million plants and animals, of which only 1.2 million have been identified (Mora et al., 2011). And if we encounter this difficulty in determining current

biodiversity, imagine how unattainable it is to find out precisely the biodiversity of the past.

All taxa do not fossilize at the same rate, nor does a single taxon maintain a constant rate of fossilization over time, with notable exceptions such as foraminifera, which have been accumulating in marine sediments for more than 500 million years with wonderful regularity (Gooday, 1994). We also do not have specific values for extinction or speciation. Therefore, the estimate of past biodiversity has a much greater uncertainty than the estimate of current biodiversity. And what is the relationship between both biodiversities? Again we have to enter the area of estimates.

It is estimated that marine invertebrate species are no more than 5% of those that have existed since the appearance of life on the planet (Jablonski, 1994), and currently living plants and animals are less than 1% of all species that have ever existed (Raup, 1992). Consequently, studying extinct species means focusing interest on more than 99% of the planet's biodiversity. It's not a small thing.

1.1.2. Extinction rates and events

We currently understand extinction as a phenomenon that combines two initially opposing views: cataclysmic extinction events, proposed by Georges Cuvier, and basal or background extinction rates, proposed by Charles Lyell and Charles Darwin (Pievani, 2014).
The life expectancy of a species can vary greatly, but no species lives forever. The fossil record indicates that the average lifespan of an invertebrate species is about 11 million years, while mammal species live around 1 million years (Table 1.1)[1].

Group	Estimated life span (My)
Dinoflagellates	13
All invertebrates	11
Cenozoic bivalves	10
Diatoms	8
Planktonic foraminifera	7
Echinoderms	6
Marine invertebrates	5–10
Marine animals	4–5
Cenozoic mammals	1–2
Mammals	1
All fossil groups	0.5–5

Table 1.1. Estimates of species' life spans, from origination to extinction [from May et al., 1995].

Using a lifespan of 1 to 10 million years (10^6-10^7 years) it can be calculated the average extinction rate of organisms. On that basis, if one followed the fates of 1 million species, one would expect to observe about 0.1-1 extinction per year—in other words, 1 species going extinct every 1-10 years. This reasoning led to the establishment of the number of extinct species (E) per million species-years (MSY) or E/MSY (Primm et al., 1995). The rate

[1] *Homo sapiens* has already existed for about 250,000 years, which means that statistically we have lived ¼ of our total time.

remains a convenient benchmark against which to compare modern extinctions, and allows calculating the background extinction rate for any taxon, for example, the number of modern bird species described is roughly 10,000, thus, one should see one extinction every 100 to 1,000 years. Similarly, there are about 5,000 described species of mammals, so a background extinction rate of 200 to 2,000 years could be expected.

The estimate assumes that all extinctions happened independently and gradually on the periods of calm in Earth's geologic history.

However, when the fossil record is studied, large discontinuities are also observed, identifiable as abrupt changes in the composition of organisms on both sides of certain geological strata. Paleontologists have determined that when more than three quarters of the species that inhabit the planet are lost in a geologically short interval, we are facing a mass extinction. In fact, these mass extinctions have served to divide many of the different periods of the Geologic Time Scale. The Phanerozoic Eon is divided into three epochs: Paleozoic, Mesozoic and Cenozoic (meaning "old life", "middle life" and "recent life"), and they represent the main stages of the macroscopic fossil record. These eras are further divided by other catastrophic extinction boundaries. Five main mass extinction events have been identified, the so called "Big Five": Ordovician, Devonian, Permian, Triassic and Cretaceous (Figure 1.4) (Barnosky et al., 2011).

Ordovician

The Ordovician event ended 443 Ma. Within 3.3 to 1.9 Myr, it is estimated that 57% of genera and 86% of species were lost. Extinction included marine organisms such as some bryozoans, reef-building brachiopods, trilobites, graptolites, and conodonts. The causes were probably an onset of alternating glacial and interglacial episodes that lead to repeated changes in sea level, the uplift and weathering of the Appalachians affecting atmospheric and ocean chemistry and the sequestration of CO_2.

Devonian

The Devonian event ended 359 Ma. Within 29 to 2 Myr, it is estimated that 35% of genera and 75% of species were lost. Extinction included many marine species, such as corals, brachiopods, and single-celled foraminiferans. The causes are not well understood yet, but could be the global cooling (followed by global warming), possibly tied to the diversification of land plants, with associated weathering, paedogenesis, and the drawdown of global CO_2. There is evidence of widespread deep-water anoxia and the spread of anoxic waters by transgressions. The timing and importance of bolide impacts is still debated.

Permian

The Permian event ended 251 Ma. Within 2.8 Myr to 160.000 years, it is estimated that 56% of genera and 96% of species were lost. It was the Earth's largest extinction event, decimating most marine species such as all trilobites, plus insects and other terrestrial animals. Most scientific evidence suggests the causes were global warming and atmospheric changes associated with huge volcanic eruptions in what is now Siberia. There were also the spread of deep marine anoxic waters, elevated H_2S and CO_2 concentrations in both marine and terrestrial realms, and ocean acidification. Evidence for a bolide impact still debated.

Triassic

The Triassic event ended 200 Ma. Within 8.3 Myr to 600.000 years, it is estimated that 47% of genera and 80% of species were lost. Extinction included many marine sponges, gastropods, bivalves, cephalopods, brachiopods, as well as some terrestrial insects and vertebrates. Activity in the Central Atlantic Magmatic Province (CAMP) thought to have elevated atmospheric CO_2 levels, which increased global temperatures and led to a calcification crisis in the world oceans.

Cretaceous

The Cretaceous event ended ~65.5 Ma (Schulte et al., 2010). Within 2.5 Myr to less than a year it is estimated that 40% of

genera and 76% of species were lost. Extinction included many species in both marine and terrestrial habitats including pterosaurs, mosasaurs and other marine reptiles, many insects, and all non-Avian dinosaurs. The scientific consensus is that was caused by environmental consequences from the impact of a large asteroid hitting Earth in the vicinity of what is now Mexico. Preceding the impact, biota may have been declining owing to a variety of causes: Deccan volcanism contemporaneous with global warming; tectonic uplift altering biogeography and accelerating erosion, potentially contributing to ocean eutrophication and anoxic episodes. CO_2 spike just before extinction, drop during extinction.

Although many of the causes of mass extinctions remain controversial, a correlation has been found with a single main cause: continental flood basalts. Magma degassing could be a primary kill mechanism for mass extinctions and other intervals of faunal turnover, which may be related to CO_2, SO_2, Cl, and F release (Green et al., 2022).

While the selectivity of the background extintion may act on factors such as local abundance, reproductive mode, body size, feeding strategy, geographic range at the species level and species richness, the 'big five' mass extinctions appear to fit a model of 'non-constructive selectivity' - not strictly random, but often selecting on features such as clade-level geographic range and thus, are unlikely to reinforce or promote long-term adaptation (Jablonski, 2004).

Strikingly, there are certain species that for multiple reasons avoid extinction and are able to survive along geological periods of hundreds of millions of years, and they do so in two different ways: by progressively transforming into new species (chronospecies), or by preserving their morphology apparently unchanged (living fossils). Chronospecies are difficult to identify from the fossil record and are often a source of debate (Stanley, 1978; Forey et al., 2004). Living fossils, on the other hand, are obvious and illustrative. Perhaps the three most remarkable examples are the ginkgo, the horseshoe crab and the coelacanth.

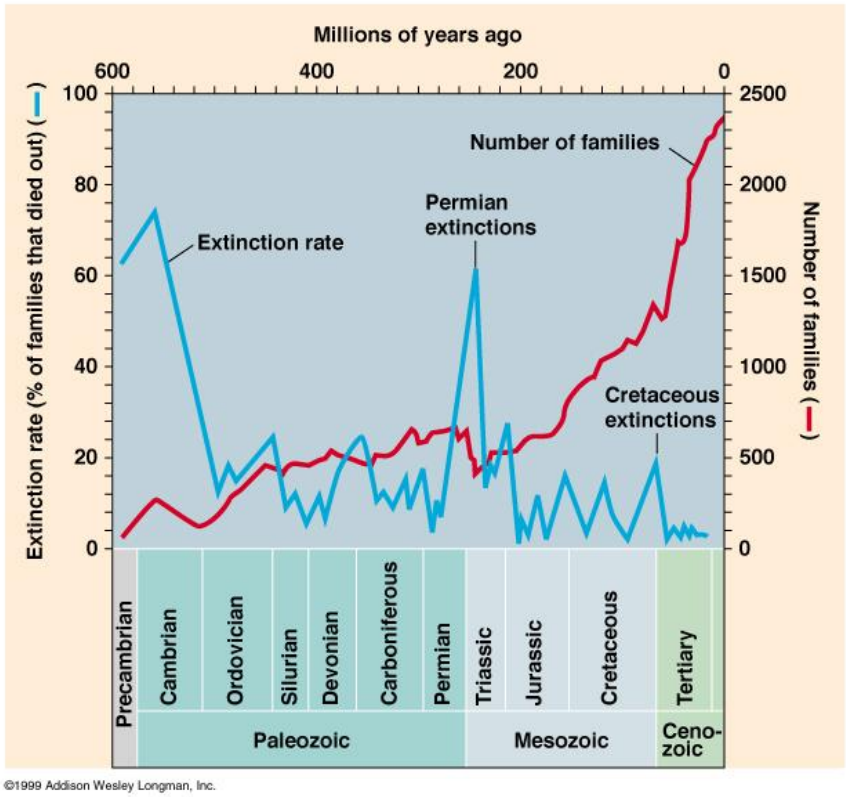


Figure 1.4. Rates of extinction (measured as the percentage of taxonomic families that disappear in the fossil record in each successive period) are highly variable over time. There are "Big Five" major mass extinctions and other 13 minor mass extinctions. The two major episodes took place at the end of Permian, which affected mainly marine invertebrates, and at the end of Cretaceous, which affected terrestrial dinosaurs. Note that, both extinctions determined the end of the Paleozoic and the Mesozoic, respectively. Extinction in the Cambrian Period is as great in terms of percent losses of genera as any of the major events. Even so, it is not considered to be a major mass extinction for two reasons. First, most of the losses are from only one major group of animals, the trilobites. Second, the diversity in the Cambrian Period was far less than at the times of the "Big Five." This means that the total number of extinctions was also much lower in the Cambrian. Despite extinctions, the total taxonomic diversity has generally increased over time (red line).
[From Steven M. Carr. https://www.mun.ca/biology/scarr/Research.html].

The ginkgo (*Ginkgo biloba*) is a tree of the conifer family that has unique deciduous fan-shaped bilobed leaves, instead of the characteristic perennial acicular leaves of its closest relatives, the pines, firs and cedars. It was only known in the fossil record of the Permian period (ca 200 Ma) until it was rediscovered in an isolated region in central China, where it had been maintained as an ornamental plant.

The horseshoe crab is a very singular arthropod, whose fossils date back to the Jurassic period (60 Ma). Currently, four species have been identified, the Atlantic horseshoe crab (*Limulus polyphemus*), found along the Atlantic coast of the United States and the Southeast Gulf of Mexico, and the mangrove horseshoe crab (*Carcinoscorpius rotundicauda*), the Indo-Pacific horseshoe crab (*Tachypleus gigas*), and the tri-spine horseshoe crab (*Tachypleus tridentatus*), found in Southeast and East Asia.

Finally, the coelacanth. (*Latimeria sp.*), is a fish whose specimens were only known in the fossil record as early as the Carboniferous period (300 Ma), and which were rediscovered in a fish market in Madagascar in 1938 (Smith, 1956). Today, two species have been identified, living at a certain depth in the Indian Ocean around Madagascar, South Africa and the Comoros Islands (*Latimeria chalumnae*), and in SE Asia (*L. menadoensis*). One study (Amemiya et al., 2013) indicates that this stable phenotype for so many years corresponds to a very stable genotype. Protein-coding genes evolve significantly more slowly than those of tetrapods. Somehow, they survive much longer than most organisms because their molecular clock ticks at a different rate.

Background extinction, mass extinctions, and extintion avoidance are found throughout the history of life. We can see it all around, and a beautiful example arises when we analyze the spindle diagram of Mesozoic mammals (Figure 1.5; Luo, 2007). Some groups like the Haramiyidians, the Morganucodonts, or the Docodonts are clades whose progressive extinction does not coincide with any great cataclysm. Instead, the Eutricodones, the Multituberculatas or the Spalacotheroids show a relatively sudden extinction that coincides with the end of the Cretaceous, the fifth great mass extinction. The Australosphenidans-monotremes, due

to their distant appearance in the Jurassic and their arrival to the present in the form of a few surviving species, such as the echidnes (*Tachyglossus* and *Zaglossus sp*) and the platypus (*Ornithorhynchus anatinus*), characterized by ancestral features, could be considered examples of living fossils.

As we will see in the next section, the Cenozoic was a period of great diversification of marsupial mammals (Metatherians) and especially of placental mammals (Eutherians). The thread of evolution forms new spindles over and over again.

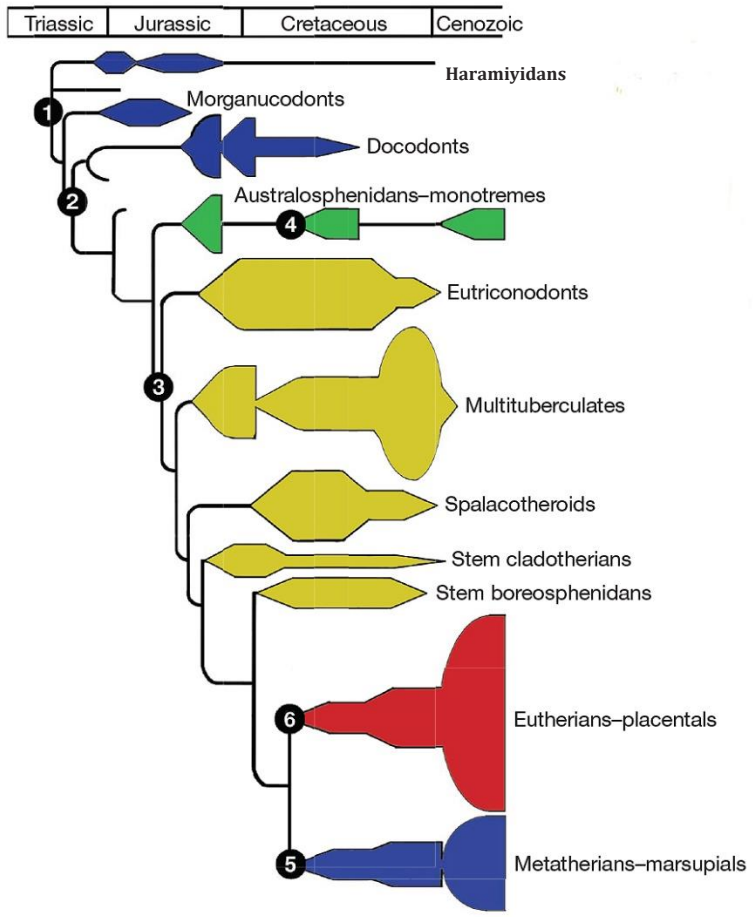


Figure 1.5. Extinction and diversification patterns of the order- or family-level Mesozoic mammal groups [modified from Luo, 2007].

1.1.3. Extinction of the Pleistocene

After the Cretaceous mass extinction, all non-avian dinosaurs disappeared, and with the exception of some ectothermic species such as sea turtles and crocodiles, no tetrapods weighing more than 25 kg survived. All the ecosystems on the planet that until then had housed gigantic animals were empty of large vertebrates. A wide range of new possibilities for evolution then opened up and a race to occupy the free niches began, marked above all by a rapid increase in body size. Land mammals were initially at a disadvantage due to their small size (between 3-5 g and 5-12kg) (Luo, 2007), so non-mammalian vertebrates had a window of opportunity of approximately ten million years (Paleocene) to evolve towards gigantism without much competition (Mitchell et al., 2014). The niches of large carnivores were frequently occupied by terrestrial crocodilians (*Pristichampsus*), large snakes (*Titanoboa*), varanid lizards, or the mythical flightless birds in Europe and North America (*Gastornis*), and in South America (*Paleopsilopterus*) (Smith et al., 2010). The niches of large herbivores were also occupied by flightless birds in America (*Brontornis)* and Madagascar (*Aepyornis*), and there were even large flightless birds probably omnivorous in Australia (*Dromornis*) (Murray & Vickers-Rich, 2004).

Gradually, mammals also experienced an almost exponential increase in body size as they diversified to fill vacant ecological niches, and with the idiosyncrasy of each continent, they ended up replacing the other large vertebrates. For example, on the northern continents, large predatory birds disappeared when large placental carnivores evolved, whereas in isolated South America, large birds were not replaced by local marsupial carnivores and remained dominant until placental carnivores from North America arrived after the two Americas were united through the Isthmus of Panama, during the Pliocene.

Some of these large mammals are well known, for example, among the herbivores we find the mastodons (*Mammut sp*), the woolly mammoth (*Mammuthus primigenius*), the woolly rhinoceros

(*Coelodonta antiquitatis*), or the Irish elk (*Megaloceros giganteus*), and among the carnivores we find the saber-toothed tiger (*Smilodon fatalis*), the cave bear (*Ursus spelaeus*), the cave lion (*Panthera spelaea*), or the bear dog (*Amphicyon sp*), the archetypal representatives of the ice age fauna.

The term "megafauna" often refers to those adult animals greater than a certain mass. 100 pounds (45.3 kg) was initially suggested (Martin, 1973). Other authors have proposed ≥1000 kg for terrestrial megaherbivores (Owen-Smith, 1988) and ≥100 kg for terrestrial megacarnivores.

For herbivores, large size is a generally effective strategy for protection against predators. Adult megaherbivores rarely have natural enemies. Its main limitations are food resources and biomechanics. When they reach a large population they are considered ecological engineers capable of altering vegetation on a landscape scale. In carnivores, the large size allows them to access large but not mega prey (45-999 kg). They are limited by prey abundance and biomechanics. Their activity in the ecosystem regulates herbivore populations either through direct consumption or by inducing behavioral changes (the so-called "ecology of fear") (Zanette & Clinchy, 2019).

The Cretaceous extinction provided the ecological opportunity for mammals to become larger. Land mammals did so in an exponentially decreasing manner, reaching a maximum size in about 40 million years, as the evolutionary possibilities for increasing body size were progressively exhausted and abiotic factors began to limit the upper limit.

Analysis of the variation in maximum body size over the last 40 million years suggests a correlation with planetary temperature, and with the area of continental lands. The colder it is, the more advantageous it is to have a large mass, due to the fact that the body surface/volume ratio is reduced, that is to say, animals have proportionally less body surface when they are bigger, and consequently they do not lose as much heat (Bergmann’s Rule). But in addition, during the glacial periods the polar caps increased in extent and the sea level around the world decreased, so the

emerged surface of the continents increased and offered more resources to animals (Smith et al., 2010; Ashton et al., 2000).

The Pleistocene megafauna is a nice example of evolutionary radiation after a mass extinction, but also played a prominent role in the history of science. In the 1830s, a young Charles Darwin made a stopover in South America on his trip aboard the Beagle, before arriving at the Galápagos Islands. While his colleagues mapped the coast, he made several expeditions in the area and discovered a lot of fossils of extinct mammals. Many of the species Darwin discovered were previously unknown to science, including several giant sloths (*Scelidotherium leptocephalum*, *Mylodon darwini* and *Glossotherium sp.*), glyptodonts, ancient horses (*Equus curvidens*) and gonphoterids ancestors of elephants (*Cuvieronius* and *Notiomastodon*), apart from the very rare *Toxodon platensis*, and *Macrauchenia patachonica* (Fernicola et al., 2009). The skeletons of some of these animals, such as the giant sloth, bore striking similarities to modern sloths, despite the differences in size. Darwin concluded that fossils were strong evidence for the "transmutation" or evolution of species (Darwin, 1859). This evidence was even more compelling because he discovered and excavated the fossils himself and could see the stratigraphy and be aware of the passage of time. Unfortunately, he was unable to find out why these giants became extinct.

The extinction of the Quaternary megafauna led to the loss of up to 90% of large animal species not only from South America explored by Darwin, but from all ice-free continents except for Africa, and represented the most profound faunal transition experienced by Earth's ecosystems during the Cenozoic.

Why did they become extinct? The cause was the climate change or were we, the humans?

The climate hypothesis was dominant during most of the 20th century, but did not consistently explain a number of features of the mass extinction: (1) The climatic events related to the last glacial cycles were planetary, however, the extinctions occurred in completely different periods in different places of the world (Fernandez, 2016). (2) There were more than 30 glacial cycles during the Pleistocene at least as intense as the last one, but

without mass extinctions recorded (Lima-Ribeiro et al., 2012). (3) The extinctions in southern Europe, tropical Asia, Oceania and almost all the islands of the planet did not coincide with glacial changes, but with climatically stable periods (Elias & Schreve, 2016). (4) Extinctions have been systematically more recent on islands, although island populations, with no possibility of emigrating, and with a much smaller number of specimens and a genetic pool than continental ones, should have become extinct earlier (Martin & Steadman, 1999; Araujo et al., 2015). (5) Plants are usually more affected by climate changes than animals, but there was no mass floral extinctions in the Quaternary (Fernandez, 2016). (6) Extinctions present an unprecedented size selectivity in the fossil record, since they almost exclusively affect megafauna species, when the latest studies reveal that the smallest species are most vulnerable to climate disturbances (Barnosky et al., 2004). (7) The fossils of megafauna show a drastic reduction in the age of sexual maturity immediately before the extinction, which has studied and verified on repeated occasions in current populations of mammals subject to high hunting pressure (Lima-Ribeiro et al., 2012).

Instead, does seem to be an increasingly clear correspondence between the dates of the arrival of humans and the extinction of the megafauna in each land mass (Martin,1984; Martin & Steadman, 1999; Barnosky et al., 2004; Edmeades, 2021). Megafauna are generally K-strategist, with extended longevity, low population growth rates, low mortality rates, and few or no predators capable of killing an adult. These characteristics, although not unique to this megafauna, make it highly vulnerable to human overexploitation. After the expansion of humans throughout the planet, the only place that today preserve an important part of the terrestrial megafauna common in most ecosystems during the Pleistocene is sub-Saharan Africa (and to a lesser extent, South Asia) (Figure 1.6). It is believed that sub-Saharan Africa megafauna was able to survive thanks to the fact that endured a long presence of hominin hunters, with whom coevolved and developed instinctive fight or flight responses (Araujo et al., 2015), while in the rest of the world the megafauna had never coexisted with hominids, and would not

have developed any instinctive protection mechanism. This absence of a defensive response is known as ecological tameness or naivety, and has been recorded as an extinction factor on numerous occasions (Diamond, 1984). In fact, it is still currently observed in those endemic species of recently discovered islands and continents such as the Galápagos or Antarctica (Lima-Ribeiro et al., 2012). For example, penguins show a deep instinct to escape from the leopard seal with which they have coevolved, but they are very trusting with humans.

However, in recent years, evidence has accumulated indicating that the extinction of megafauna was probably a consequence of the combined effect between climate change and human overexploitation.

An example is found in the extinction of the woolly mammoth. In central and eastern Europe, the development of the architecture of mammoth bone dwellings is associated with the abundance of mammoth bone beds which were massively exploited by humans in the Upper Paleolithic (Iakovleva, 2015). At the same time, climate and a population models show that the collapse of the climatic niche of the mammoth caused a significant drop in their population size, making woolly mammoths more vulnerable to the increasing hunting pressure from human populations (Nogués-Bravo et al., 2008). In another study, a statistical approach revealed the interaction between climate and humans in the extinction of the megafauna in southeast Australia (Saltré et al., 2019). And a large-scale study analyzed the distribution and timing of all megafaunal extinctions in relation to climatic variables and human arrival on five landmasses, and it unveiled that the pattern of extinctions is best explained by models that combine both anthropogenic and climatic factors (Prescott et al., 2012).

Unfortunately, union is strength even for extinction.

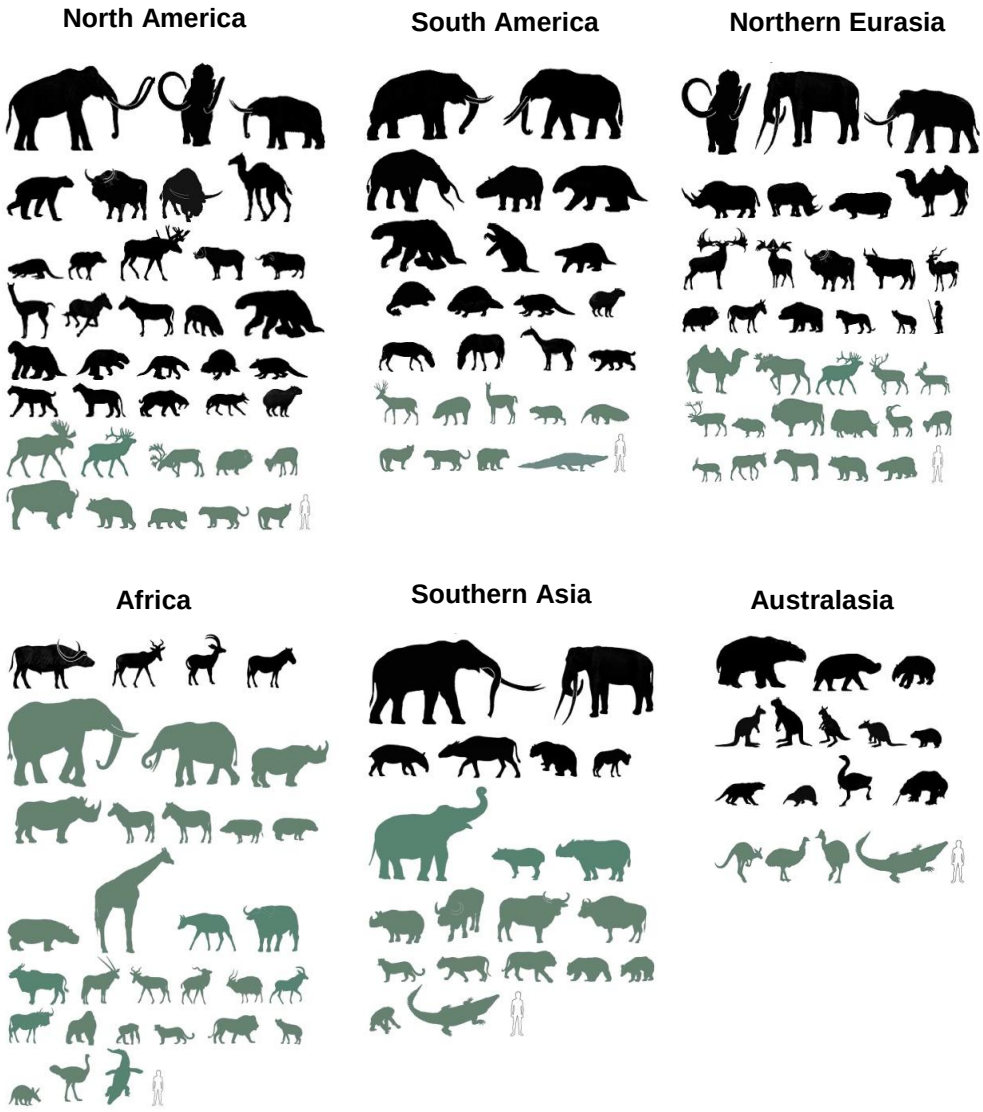


Figure 1.6. Selected extinct (black) and extant (grey) megafauna in each subcontinent. Note that *Homo neanderthalensis* is also included in Northern Eurasia. Outline *Homo sapiens* gives approximate scale [from Stuart, 2015].

1.1.4. Extinction of the Anthropocene

If we compare the rates and amounts of extinction during geologic mass extinction with the amounts of species losses over the past few centuries and millennia, we see a similar trend. According to a group of authoritative evolutionists like Edward O. Wilson and Niles Eldredge, we have evidence that humans are now causing the so called ‘‘Sixth Mass Extinction’’ (Pievani, 2014). But the analogy has even deeper connotations. The 1995 Nobel Prize in Chemistry Paul J. Crutzen popularized the term “Anthropocene”, to describe what he regarded as a new epoch, characterized by human dominance of biological, chemical and geological processes on Earth. The exponential growth of humankind associated with the increase in our technological capacity has devastating consequences on the Planet: fragmentation of habitats, chemical pollution, invasive species, spread of pathogens, overexploitation of resources in hunting and fishing and changing global climate (Pievani, 2014). The synergy of all these factors results in a drastic reduction of biodiversity.

Among terrestrial vertebrates, 322 species have become extinct since 1500, and populations of the remaining species show 25% average decline in abundance. Invertebrate patterns are equally dire: 67% of monitored populations show 45% mean abundance decline (Dirzo et al., 2014). Nearly half of known vertebrate species, 32% (8,851/27,600), are decreasing in population size and range. In the 177 mammals for which there is detailed data, all have lost 30% or more of their geographic ranges and more than 40% of the species have experienced severe population declines (Ceballos et al., 2017).

A key measure of humanity’s global impact on biodiversity is by how much it has increased species extinction rates. Familiar statements are that these are 100-1000 times pre-human or background extinction levels. But establishing a background rate for comparison is not straightforward, and varies from 0.1 E/MSY (De Vos et al., 2014; Pimm, et al., 2014), ~1 E/MSY (Pimm et al., 1995), 1.8 E/MSY for mammals (Barnosky et al., 2011), up to 2 E/MSY (Ceballos et al., 2015). Both theory and empirical data

indicate that extinction rates changes markedly depending on the length of time over which they are measured. Extrapolating a rate computed over a short time, therefore, will probably yield a rate that is either much faster or much slower than the average million-year rate (Barnosky, *et al* 2011).

In recent decades, various approaches have been used to determine the extinction rate: compare currently measured extinction rates to background rates assessed from fossil record (Pereira et al., 2010), use various modelling techniques, including species-area relationships, to assess loss of species (Dirzo & Raven, 2003), compare rate of expected near-term future losses to estimated background extinction rates (Barnosky et al., 2011), assess extinction in context of long-term clade dynamics (Barnosky et al., 2011), assess percentage loss of species (Pimm et al., 1995), and even use molecular phylogenies to estimate extinction rate (Roelants et al., 2007). Each of these approaches has its limitations, for example, species-area relationships may overestimate extinction (Axelsen et al., 2013).

In spite of the disparity of methods and estimates, the numbers suggest that we have not yet seen the sixth mass extinction, but the current extinction rates are higher than those that caused Big Five extinctions in geological time; they could be severe enough to carry extinction magnitudes to the Big Five benchmark in as little as three centuries (Barnosky et al., 2011).

Selective pressures in mass extinction are not tightly linked to traits favoured during 'normal' times, and thus are unlikely to reinforce or promote long-term adaptation. Perhaps intrinsic factors diminish in importance with increasing perturbation. Those shifts in selectivity, along with the sheer magnitude of the 'big five' events, may explain why postextinction periods are important in opening opportunities for once-marginal groups, for example the expansion of mammals after the dinosaurs' demise (Jablonski, 2004).

In that context, the current reduction of formerly widespread ranges and disproportionate culling of certain kinds of species may be particularly informative in indicating that Anthropocene extinction acquires the characteristics of a mass extinction.

What is the impact of this Sixth Mass extinction? In assessing extinction and the diversity of the remaining biota it is important to look at more than just the number of species lost. Whereas the loss of 'redundant' species may be barely perceptible, more extensive losses of whole populations, groups of related species (clades) or those that share particular morphologies (for example, large body sizes) or functional attributes such as feeding mechanisms, can have profound effects, leading to the collapse of entire ecosystems and the extermination of great evolutionary dynasties(Jablonski, 2004). For example, the extinction of a keystone predator species can be transmitted top down and affect the entire ecosystem, as in the case of the extinction of the sea otter *(Enhydra lutris)* off the coast of California, which represented the disappearance of kelp forests due to excess pressure from herbivores, especially sea urchins, and that in turn led to the extinction of the Steller's sea cow (*Hydrodamalis gigas*) (Estes et al., 2016). The effect of extinction can also spread bottom up, by replacing primary producers or consumers. In this sense, the ecological niche left by the extinction of most of the Pleistocene herbivorous megafauna on many continents has been successfully filled by domestic livestock.
But beyond the loss of ecosystem functionality, extinctions also represent very important losses of the gene pool. Exposito-Alonso et al., (2022), introduce a mathematical framework that bridges biodiversity theory and population genetics to understand the loss of naturally occurring DNA mutations with decreasing habitat. By analysing genomic variation of 10,095 georeferenced individuals from 20 plant and animal species, they show that genome-wide diversity follows a mutations-area relationship power law with geographic area, which can predict genetic diversity loss from local population extinctions. They estimate that more than 10% of genetic diversity may already be lost for many threatened and nonthreatened species.
How biodiversity will be modified after the Sixth Mass extinction?
The most vulnerable species are in fact those having more specialized and restricted habitats, with greater body size, less displacement capacities and adaptive plasticity. It follows that with this trend of extinction and habitat fragmentation in the future more

opportunistic species, smaller in size and able to live with an ubiquitous human presence, will be greatly favoured. We can expect, therefore, a compelling and disproportionate success of pests, rats, crows, gulls, cockroaches and other insects (Pievani, 2014).

We will see the same trend in vegetation. Kress & Krupnick, (2022), classify, 86,592 species of vascular plant species (~30%) as winners or losers with respect to their compatibility with human activities and show that currently losers greatly outnumber winners and that these losers will continue to exceed winners in the future even if species now deemed tentative winners succeed. Furthermore, when mapped, both taxonomically and phylogenetically, although winners and losers are today distributed across nearly all orders of vascular plants, only two of the nine major phylogenetic lineages favor winners over losers (basal angiosperms and early diverging eudicots). Some examples of plant winners in the Anthropocene are *Ginkgo biloba*, *Prunus serotina, Halophila stipulacea* and *Merremia tuberosa*, some losers are *Magnolia ekmanii*, *Ceratozamia kuesteriana*, *Araucaria muelleri* and *Sidalcea stipularis*. Finally, some currently neutral, neither winners nor losers examples are *Cytisus oromediterraneus* and *Carex bullata*.

We might think that, as in the other five mass extinctions, the sixth mass extinction will generate the conditions for a new adaptive radiation. However, recovery of biodiversity will not occur on any timeframe meaningful to people: evolution of new species typically takes at least hundreds of thousands of years, and recovery from mass extinction episodes probably occurs on timescales encompassing millions of years (Barnosky et al., 2011). Davis et al., (2018) use a birth-death tree framework to show that even if extinction rates slow to preanthropogenic background levels, recovery of lost phylogenetic diversity of over 300 mammal species erased by human activity since the Late Pleistocene, will likely take millions of years.

Consequently, if we want to avoid this catastrophic scenario, we must take effective conservation measures as soon as possible. But given that resources and time are limited, we must prioritize.

Taxonomic species richness, the most used metric for measuring biodiversity among researchers, governments, and managers, is insufficient for these purposes because it implicitly treats all species equally (Davis et al., 2018). Functional diversity, a richer metric that captures species' ecological adaptations and contributions to ecosystem function is growing in popularity, but it is difficult to measure and hard to compare between different taxonomic groups. Phylogenetic diversity, the amount of independent evolution within a phylogeny, is a complementary metric that measures lineage history and may be correlated to functional trait diversity and evolutionary potential (Davis et al., 2018).

Alternative prioritization strategies have been proposed to safeguard biodiversity over macroevolutionary time scales. The first prioritizes the most distantly related species -maximizing phylogenetic diversity- in the hopes of capturing at least some lineages that will successfully diversify into the future. The second prioritizes lineages that are currently speciating, in the hopes that successful lineages will continue to generate species into the future (Cantalapiedra et al., 2019). Unfortunately, there are no conclusive results to choose between these two strategies (Cantalapiedra et al., 2019).

The alternative is to protect natural habitat. From the basis of existing protected areas, Key Biodiversity Areas and ecologically intact areas Allan et al., (2022) represent the distribution of 35,561 species of mammals, birds, amphibians, reptiles, freshwater crabs, shrimp, and crayfish scaled to the sizes of their ranges while also capturing samples of all terrestrial ecoregions. Their conclusion is that at least 64 million square kilometers (44% of terrestrial area) would require conservation attention to stop the global biodiversity crisis.

In any case, efforts to conserve biodiversity must be made on all fronts, apart from those mentioned above, it should be added the control of poaching and illegal trade in protected species, and the encouragement of captive breeding of critically endangered species with a view to future reintroductions. Otherwise, we will not succeed.

1.2. Ancient DNA

1.2.1. From palaeontology to palaeogenomics

Extinct species are known from their fossil or subfossil remains. Traditionally, this type of samples have defined the object of study of palaeontology. However, with the progress of sequencing techniques, genetics (the study of genes) has broadened its molecular scope and has given rise to genomics (the study of the genome), and this, in turn, has broadened its temporal scope and originated palaeogenomics (the study of ancient genomes; aDNA). Now fossils continue to be studied morphologically by palaeontology, but they can also be studied molecularly by palaeogenomics (under certain conditions, see next section 1.2.2). Thus, we have two complementary disciplines to better understand evolution.

What are the benefits of paleogenomics?

1. Extends molecular analysis to extinct organisms
2. Identifies species from fossil remains that are not diagnostic enough
3. Provides results unaffected by autapomorphies
4. Provides results unaffected by homoplasies and incipient speciation
5. Fills temporal gaps (to Mid-Pleistocene at most)

To begin with, it is evident that by analyzing extinct species, paleogenomics incorporates, in the evolutionary study, alleles that are not found in the current populations and, therefore, are not accessible to genomics. We could say that it extracts more juice from fossils.

In 2010, palaeogenomics reached a historic milestone, when it led to the identification of a new extinct hominin from the analysis of the complete genome contained in the distal manual phalanx of the fifth digit excavated in Denisova Cave in the Altai Mountains, Russia (Reich et al., 2010). The so-called Denisovans shared a common ancestor with modern humans and Neanderthals 1 Ma.

Additionally, it has been found gene flow from Denisovans into Oceanians (Reich et al., 2010) and from Neanderthals into ancestors of all non-Africans (Green et al., 2010), among other hybridization events (Meyer et al., 2012; Durvasul & Sankararaman, 2020). Meanwhile, palaeontology was not able to lift a finger, because was not diagnostic enough.

When a species has differentiated greatly from its ancestors, it accumulates unique morphological traits that are difficult to compare and relate to other species in order to establish taxonomies and phylogenies. These unique traits are called autapomorphies and are a true challenge for palaeontology. On the contrary, these morphological transformations are often caused by changes in very few genes (Lalueza-Fox, 1999), so in these particular cases palaeogenomics can establish evolutive relationships more easily (relatively) than palaeontology.

There are very illustrative examples of the contribution of palaeogenomics in clarifying taxonomies and phylogenies conditioned by autapomorphies in fossil remains. After the breakup of the supercontinent Gondwana during the late Mesozoic, South America spent most of the Cenozoic as an insular continent whose splendid isolation allowed its fauna to evolve into many forms found nowhere else on Earth, most of which are currently extinct. When Charles Darwin explored Patagonia, he discovered a few species of fossil mammals (see section 1.1.3). Some of these fossils posed a special identification challenge because they showed an unusual mix of morphological traits and it was far from obvious to identify its closest extant relatives. Nor were there massive collections of skeletons for comparison as there are today. The most outstanding case was the species *Macrauchenia patachonica*, described by Richard Owen, one of the most reputable comparate anatomist and palaeontologist of its time, as follows: “A large extinct Mammiferous Animal, referrible to the Order Pachydermata; but with affinities to the Ruminantia, and especially to the Camelidae” (Fernicola et al., 2009). Subsequently, morphological studies of South American ungulates already classified *Macrauchenia* in the order Litopterna but there was controversy over its origin (Cifelli, 1993; de Muizon &, Cifelli, 2000; O’Leary et al., 2013). The situation was clarified by

different molecular approaches based on ancient proteins (Welker et al., 2015) and aDNA. Westbury et al., (2017) recoverded almost the complete mtDNA from bone samples and elaborated a dated phylogenetic tree which situates *Macrauchenia* as a sister taxon to all living Perissodactyla, with a divergence estimate at ~66 Ma, successfully demonstrating that even taxa marked by deep divergence times with no close living relatives are amenable to palaeogenomics. This revelation came after 180 years of research, unfortunately too late to satisfy Darwin's curiosity.

The glyptodon was another mammal of the South American megafauna whose taxonomy and phylogeny also benefited from palaeogenomics. It accumulated different autapomorphic adaptations, like a strong cuirass made of dermal bony plates and usually a large body mass that in some species, such as *Glyptodon clavipes*, was around 2,000 kg (Vizcaíno et al., 2012). The first palaeontological studies already placed it within the cingulate xenarthrans (armadillos and their allies), but its precise phylogenetic position remained somehow controversial until it was possible to obtain mitogenomes. The molecular phylogenetic reconstruction establishes that glyptodonts are in fact deeply nested within the armadillo crown-group from whom diverged no earlier than around 35 Ma, in good agreement with their fossil record (Delsuc et al., 2016). The study also performed a statistical reconstruction of body mass for the last common ancestor of Glytodontinae, Chlamyphorinae and Tolypeutinae, estimated in a mere 6 kg, implying a spectacular increase in glyptodont body mass during the Neogene.

Other environments where autapomorphies are favored throughout evolution are islands. The best known case is the dodo (*Raphus cucullatus*) of Mauritius (see Figure 1.17). The dodo was a bird that, as a consequence of living in an environment without predators, lost the ability to fly and became a giant. The arrival of Europeans on the island caused its extinction in less than a century (ca. 1662). From historical illustrations and various skeletal remains it has been estimated that it weighed about 14 kg (Angst et al., 2011; Louchart & Mourer-Chauviré, 2011) and was about 1 m tall (the equivalent of a goose with the body mass of a farm turkey).

The wings were vestigial and should recall those of the flightless cormorant (*Phalacrocorax harrisi*), it was covered by a filamentous plumage similar to that of kiwis (*Apteryx sp*) and its bill was very long, robust with a hooked tip, reminiscent of an albatross, or raptor. These disparate anatomical features confused early taxonomists which classified it as small ostrich, rail, albatross, or vulture (Hume et al., 2009). In 1842, Danish zoologist Johannes Reinhardt proposed that dodos were ground pigeons, based on studies of a dodo skull he had discovered in the collection of the Natural History Museum of Denmark (Reinhardt, 1842). This view was later supported by English naturalists Hugh Strickland and Alexander Melville in their monograph (Strickland & Melville,1848). However, few more significant advances were made since then until the 21st century, when mtDNA was extracted from dodo remains and a phylogeny was made with 36 species of family Columbidae. We know now that the closest living relative of the dodo is the monotypic Nicobar pigeon (*Caloenas nicobarica*) from the Nicobar Islands and nearby Southeast Asia, and that they diverged in the mid/late Eocene, around 42.6 Ma, long before Mauritius emerged in a series of volcanic events, the earliest of which occurred around 6.8 to 7.8 Ma (Shapiro et al., 2002). The direct ancestor of the dodo that colonised Mauritius is supposed to have disappeared without trace.

Another example of autapomorphies arising in insularity is found in the Balearic Islands. During the Mesinian crisis between 5.97 and 5.33 Ma (Manzi et al., 2013) the Mediterranean Sea practically dried up and the Balearic Islands were connected to the rest of Europe so many animals colonised the islands by land. However, the Atlantic waters found a way through the present Gibraltar Strait and rapidly refilled the Mediterranean 5.33 Ma in an event known as the Zanclean flood (Garcia-Castellanos et al., 2009). Then, the Balearic fauna was isolated and began to differentiate from the continental fauna. Some of those insular animals became extinct with the arrival of humans, such as the bovid *Myotragus balearicus*, of which enough fossil remains have been found to identify its different chronospecies (Bover et al. 2019). Its evolutionary process shows drastic morphological changes since the Pliocene,

including: a decrease in overall body mass, an increase of limb bone robustness, a decrease of bone length, a progressive reduction in number and size of incisiform and premolar teeth, and a reduction of brain size and sense organs (Bover & Tolosa, 2005). In addition, its eyes were not directed towards the sides, as are those of nearly all the herbivorous mammals, but towards the front like nearly all primates and carnivorans, granting them stereoscopic vision (Alcover et al., 1981). These unique autapomorphic changes, made it challenging to resolve its phylogenetic placement within Caprini based on morphological traits alone. However, the recovery of partial mtDNA (Lalueza-Fox et al., 2002) and the complete mitogenome (Bover et al., 2019) allowed establishing a robust phylogeny and placed *Myotragus* in the *Ovis* clade along with sheep and revealed a closer relationship with the takin (*Budorcas taxicolor*) from which it diverged around 7.1 Ma.

It should be noted that there is an almost complete reliance on mtDNA sequences for phylogenetic studies of extinct animals. For species that are not very closely related to each other, this is not problematic because when enough time has passed between speciation events all parts of the genome are expected to show the same phylogeny. However, when closely related species are studied, it is important to remember that the mtDNA represents a single genetic locus that might or might not reflect the overall history of the genome. Caution in the interpretations of the results is therefore necessary (Hofreiter et al., 2001).

But if palaeontology has difficulties in identifying fossils showing autapomorphies, i.e., that incorporate numerous evolutionary novelties with respect to possible reference species, it also has difficulties at the opposite extreme, in identifying fossils that are morphologically too similar, either because they are at an incipient stage of speciation, or worse, because they show homoplasies, that is, traits gained or lost independently in separate lineages throughout evolution.

Homoplasy is the result of evolutionary convergence, parallelism or reversal, and its ubiquity creates difficulties in classical phylogenetic analyses (Wake, 1991). The level of homoplasy varies among taxa, and if is rampant, existing cladistic methods fail

(Wake, 1991). We can also find homoplasy in molecular phylogenies, but there are methods to identify them (Crispell et al., 2019).

An interesting case where palaeogenomics made it possible to resolve a morphological phylogeny probably compromised by homoplasy is found in Caribbean rodents. Because the morphological evolution of rodents is characterized by high levels of homoplasy (Jaeger, 1988), an alternative approach to unravel the origin and diversification of South American hystricognaths is to explore their phylogeny with molecular characters (Huchon & Douzery, 2001). The Capromyidae hystricognaths, also called hutias, are endemic rodents of the Caribbean islands and represent a model of dispersal for non-flying mammals in the Greater Antilles. This family has experienced severe extinctions during the Holocene and its phylogenetic affinities with respect to other caviomorph relatives are still debated as morphological and molecular data disagree. Fabre et al., 2014 constructed a 50-taxon mitochondrial and nuclear multigene dataset comprising five hutias from museums and 25 echimyids sampled across major clades. The results showed that Capromyidae are nested within Echimyidae (spiny rats) and should be considered a subfamily thereof. It was also estimated that the split between hutias and Atlantic forest spiny rats occurred 16.5 Ma, which is more recent than the Greater Antilles and Aves Ridge (GAARlandia) event from the early Oligocene (35-34 Ma), when a land bridge or island chain connected the Greater Antilles to South America.This would suggest that during the Early Miocene, an echimyid-like ancestor colonized the Greater Antilles from an eastern South American source population via rafting. Vicariant separation between islands and dispersal waves associated with glacial cicles contributed to generate the observed diversity of hutias (Fabre et al., 2014).

A final advantage of palaeogenomics over palaeontology is that fills temporal gaps over the last hundreds of thousands of years (Orlando et al., 2013; see next section 1.2.2).

The fossil record is a spectacular archive of extinction, and provides a vital deep-time perspective on factors governing extinction patterns (Jablonski, 2004). However, it is not a

continuous record, but contains many gaps. There are periods of millions of years which are sparsely represented or not represented at all in the fossil record. It is certain that there are thousands of organisms for which no fossil has ever been found and, in most cases, it is very likely that this situation will never be remedied. This is simply because the process of fossilisation is extremely rare. The type of organism and the environment in which they lived are also important considerations. Animals with hard exoskeletons are better preserved and well represented, as are organisms that lived in the shallows around oceans, seas, and lakes. The data is skewed.

Martin (1993) used the number of existing species of primates (200), the origin of this group (65 Ma), and an average individual species duration (1 million years) to infer that about 6,500 extinct primate species preceded the modern ones. This estimate can be used to assess the existing sample of the primate fossil record. Around 1980, a review recognized 250 species of fossil primates. These species represent 3.8% of the estimated total of past primate species (Figure 1.7). Although random distribution of discovery of fossil species within the tree is in fact unlikely, this does not affect the two main points that can be recognized. Firstly, time of origin is likely to be considerably earlier than indicated by first known fossil representatives. And secondly, a very low sampling density may lead to assign inappropriate ancestral positions to known fragmentary fossils (Martin, 1993).

Palaeogenomics can obtain phylogenetic trees by coalescence without gaps. However, it depends on (well-dated) fossils for calibrations. According to Martin (1993) it might be advisable to add 40% to all dates of divergence based on calibration derived from properly identified fossil evidence.

Palaeontology and palaeogenomics should continue to work together and collaborate with each other.

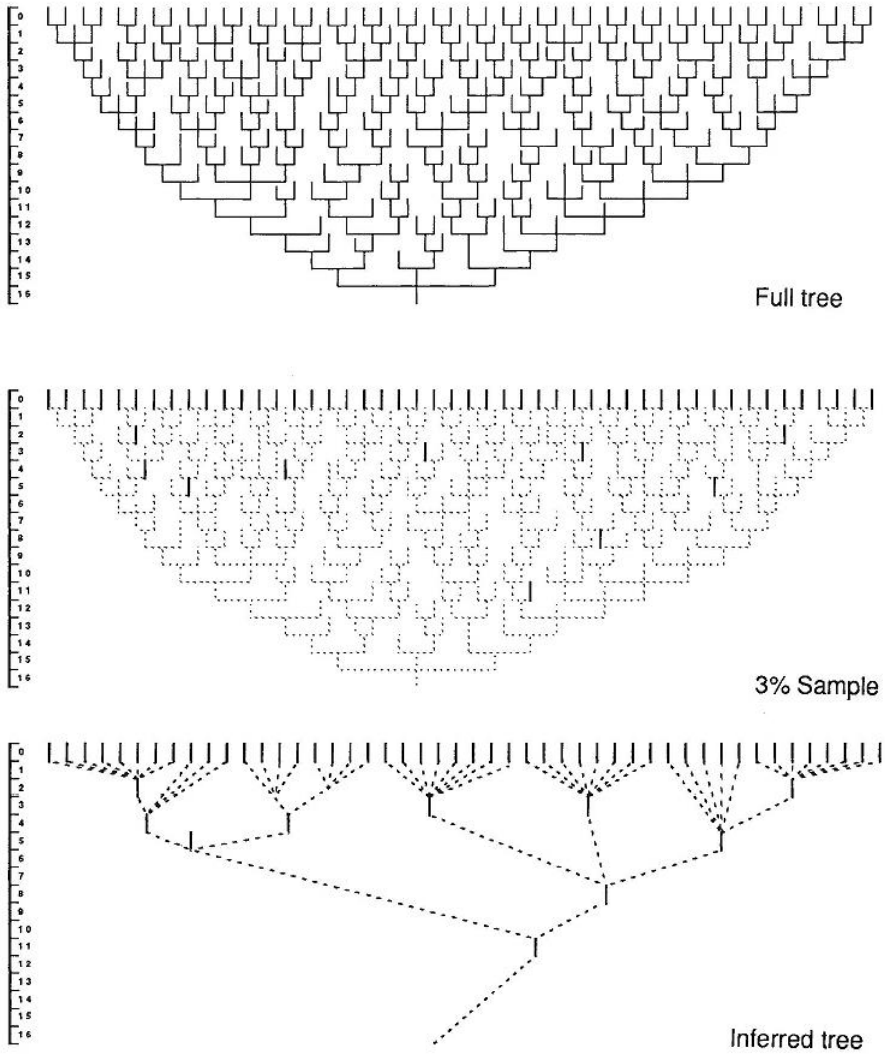


Figure 1.7. Full tree: Model branching tree, with progressive expansion of number of species from 1 to 48 over a period of 16 million years. Each species has been given a standard survival time of 1 million years. 3% Sample: A typical example in which a 3% sample of fossil species (n = 10 species) has been randomly distributed throughout the tree. In this case, the earliest known species is dated at 11 Myr and underestimates the true time of origin of the group by 5 million years (corresponding to a required correction factor of 45%). Inferred tree: An extreme illustration of the kind of tree that might be reconstructed if major gaps in the fossil record are not acknowledged. The time of origin of the group (based on the first known species) is likely to be seriously underestimated and fossil species may be forced into unrealistic ancestral positions [from Martin, 1993].

1.2.2. Biochemistry of ancient DNA

Since its discovery (Watson & Crick, 1953), the DNA molecule, with its elegant double helix geometry and its importance in the genetic coding of all living beings (excluding viruses), has become a symbol not only of the biological sciences but also of humanity as a whole.

It is reasonable to think that if it has endured for 3.8 billion years transmitting genetic information across generations through all living lineages from bacteria to whales, it is because it has advantages over possible alternatives. Before the genes included in DNA were subject to natural selection, the molecule itself was already selected. Exactly how DNA came into existence is still a mystery. Conventional wisdom suggests that RNA-based life eventually switched to DNA (Forterre et al., 2007). Although RNA can serve as an adequate but relatively short-lived and mutation-prone carrier for a limited amount of genetic information, as seen in RNA viruses, the development of autonomously replicating cells probably depended on reduction of the ribose moiety in nucleotides to the unusual sugar, deoxyribose and form deoxyribonucleic acid -DNA-. This event provided genomes of greatly improved chemical stability (Lindahl, 1993a). However, when studied in detail, it is discovered that this chemical stability is continually threatened and is, to some extent, limited (Lindahl, 1993a).

In all living organisms DNA suffers from many environmental stresses, including UV light radiation, and attacks by reactive oxygen species and carcinogens, produced by the cellular metabolic activity itself. The list of endogenous DNA-damaging agents, processes and DNA adducts is long (De Bont, & van Larebeke, 2004). As a result, hydrolysis, oxidation and nonenzymatic methylation of DNA occur at different sites (Figure 1.8), and at significant rates *in vivo*, and are counteracted by specific DNA repair processes (Lindahl, 1993a).

In addition, DNA suffers from intrinsic errors and unusual structures, which are formed during replication or recombination, and they must be corrected by the various repair protein machineries to avoid alterations of the base sequences or

entanglement of the DNA. These repair complexes include nucleases that play important roles in eliminating the damaged or mismatched nucleotides (Nishino & Morikawa, 2002). Nucleases can also digest the DNA in food into nucleotides for use in rebuilding the organism's own DNA, just as proteases digest food proteins into amino acids, and can digest DNA in apoptotic cells as well (Kawane et al., 2014).

Nucleases can be regarded as molecular scissors, which cleave phosphodiester bonds between the sugars and the phosphate moieties of DNA. However, figuratively and literally they are a double-edged sword. Their activity must be perfectly regulated or it could become self-injurious to the cell. For example, in healthy cells, CAD (caspase-activated deoxyribonuclease) is locked in a complex with its inhibitor (ICAD) and is only released when it receives a specific stimulus generated during cell apoptosis (Kawane et al., 2014).

When an organism dies, this delicate chemical balance of repair and control ceases. Then, nucleases, a plethora of other endogenous damaging agents, and microbial activity, degrade the DNA, which is thereafter called ancient DNA (aDNA). Under certain fortunate conditions this genetic material, ranging in age from decades to hundreds of thousands of years old, can be retrieved and analyzed.

Since the beginning of paleogenomics, the main characteristic of recovered aDNA is its high degree of fragmentation (Pääbo, 1989). Some studies have been performed in the laboratory to assess the rate of DNA fragmentation (Lindahl & Nyberg, 1972). However, the kinetics of long term post-mortem DNA decay in the environment is still poorly understood. Allentoft et al., (2012) use a quantitative real-time PCR (qPCR) design to measure relative copy numbers of mitochondrial DNA (mtDNA) fragments from 158 radiocarbon-dated bones of the extinct New Zealand moa (Aves: Dinornithiformes). The results confirm empirically a long-hypothesized exponential decay relationship. The average DNA half-life within this geographically constrained fossil assemblage was estimated to be 521 years for a 242 bp mtDNA. With an effective burial temperature of 13.1ºC, the rate is almost 400 times slower than predicted from

published kinetic data of *in vitro* DNA fragmentation at pH 5 (Figure 1.9). By calculating DNA fragmentation rates on Illumina HiSeq data, the study also show that nuclear DNA has degraded at least twice as fast as mtDNA. However, only 38.6% of the variation in DNA preservation could be explained by the age of the fossils. The rest of variation could be due to physical, chemical and biological factors such as taphonomy, fossil storage, oxygenation, microbial diagenesis, pH and ionic strength, and the presence of cations, humics and humates. The model developed predicts a half-life of 158,000 years for a 30 bp mtDNA fragment in bone at –5ºC. Even rough estimates such as this imply that sequenceable bone DNA fragments may still be present more than 1Myr after deposition in deep frozen environments (Allentoft et al., 2012). Indeed, as sequencing techniques have improved, the age of recovered aDNA samples has increased accordingly. The oldest full genome sequence determined so far is a 1.12-times coverage draft genome from a horse bone recovered from Alaska early Middle Pleistocene permafrost dated to approximately 560–780,000 years BP, almost an order of magnitude older than previous studies (Orlando et al., 2013).

Another study (Kistler et al., 2017) conducted a meta-analysis with 185 paleogenomic datasets largely based on mammal bone samples and compared DNA survival with environmental variables and sample ages, but found no correlation between DNA fragmentation and sample age over the timespans analyzed, even when controlling for environmental variables (Figure 1.10 a, b). In contrast, it was observed a strong association of humidity and thermal fluctuation with DNA fragmentation. The researchers propose a model for ancient DNA decay wherein fragmentation rapidly reaches a threshold, then subsequently slows. The observed loss of DNA over time may be due to a bulk diffusion process in many cases, highlighting the importance of tissues and environments creating effectively closed systems for DNA preservation.

Under the same perspective yet another model has been developed to predict the survival of DNA fragments of different length across the globe for both open and cave sites, based on

environmental temperature, and calibrated to cases where the DNA degradation rate has been independently measured (Hofreiter et al., 2014). The results reflect expected latitudinal climatic variations and indicate the areas of the world where aDNA recovery is most challenging (Figure 1.10 c).

What is the biochemical cause behind the observed DNA fragmentation? Based on *in vitro* experiments using modern DNA (Lindahl & Nyberg, 1972), it has been suggested that fragmentation is owing to hydrolytic depurination and subsequent β elimination resulting in single-strand breaks (Figure 1.11 a) (Dabney et al., 2013). This assumption was later proved correct for aDNA with high-throughput sequencing methods, which rely on the ligation of DNA adaptors to the ends of DNA molecules to construct sequencing libraries. Using this approach and a reference genome to infer the bases immediately adjacent to the aDNA fragments, it was shown that the purines adenine (A) and guanine (G) are overrepresented immediately outside of the DNA fragments extracted from the remains of Neanderthals, mammoths, and cave bears, which were all ~40,000 years old (Briggs et al., 2007). DNA is mainly fragmented through the loss of purines, hence depurination.

But the aDNA is not only fragmented, it also shows a characteristic fragmentation pattern. On average the fragments are less than 70 bp long and have a leptokurtic and right-skewed distribution (Figure 1.11 b). This distribution is explained by biomolecular context. For example, a short-range (~10 bp) periodicity observed in the distribution of fragment lengths is attributed to the period of a complete turn of the DNA double-helix around a histone (Pedersen et al., 2014), which is thought to offer some protection against breakage at histone-adjacent sites. It is necessary to keep in mind that DNA breakdown in colder environments appears to more faithfully reflect cellular architecture and the *in vivo* genome context, whereas breakdown in warmer conditions is much less discriminant (Briggs et al., 2007).

Apart from fragmentation, another typical alteration of aDNA are miscoding lesions, the most common of which is deamination. A primary target of deamination is cytosine. Its product, uracil, will

direct the incorporation of adenine (A) during DNA replication, resulting in apparent C to T or G to A substitutions (depending on the strand sequenced) (Figure 1.12) (Dabney et al., 2013). The highest frequency of these misincorporations are located at the ends of aDNA molecules (Briggs et al., 2007). One possible mechanism underlying this is the presence of single-stranded overhanging ends in the aDNA, because the rate of cytosine deamination is ~2 orders of magnitude higher in single- than in double-stranded DNA (Lindahl, 1993a). An alternative and not mutually exclusive mechanism is ''DNA breathing'' in the ends of molecules, which could cause them to be partially single-stranded and thus more susceptible to deamination (Briggs et al., 2007). Understanding DNA breathing as thermally driven structural fluctuations at physiological temperatures that should reflect the heterogeneity of the double-stranded DNA stability near the melting temperature (von Hippel et al., 2013).

This hypothesis implies that deamination is fragmentation dependent. However, the meta-analysis by Kistler et al., (2017) showed that deamination correlates with both sample age and temperature, while fragmentation only correlates with temperature. (Figure 1.13). In addition, the study carried out a set of simulations to test the effects of varying the fragmentation rate on the accumulation of deaminated residues, showing that a time dependent fragmentation process is incongruent with the observation of total cytosine deamination in single stranded overhangs ($\delta s = 1$). In conclusion, the processes that influence the rate of cytosine deamination are less well understood than those of DNA fragmentation, with a wide variation occurring between samples (Kistler et al., 2015).

Apart from fragmentation and deamination, aDNA undergoes some modifications that obstruct the movement of DNA polymerases along a template strand, preventing their amplification and sequencing. Such blocking lesions occur in the form of nucleotide modifications and crosslinks, which can form either between DNA strands, between different DNA fragments, or between DNA and other molecules (Dabney et al., 2013).

The retrieval of ancient DNA sequences is far from routine, as the researcher has to contend with the fact that very little and often no endogenous DNA survives in ancient tissues, whereas contemporary DNA is pervasive in the environment, both inside and outside the laboratory. Therefore, great precautions need to be taken to avoid the presence of exogenous DNA (Hofreiter et al., 2001). Making a virtue out of necessity, the limitations associated with the aDNA such as fragmentation pattern and deamination, have become authentication criteria to ensure that recovered DNA sequences from samples are ancient instead of modern contamination.

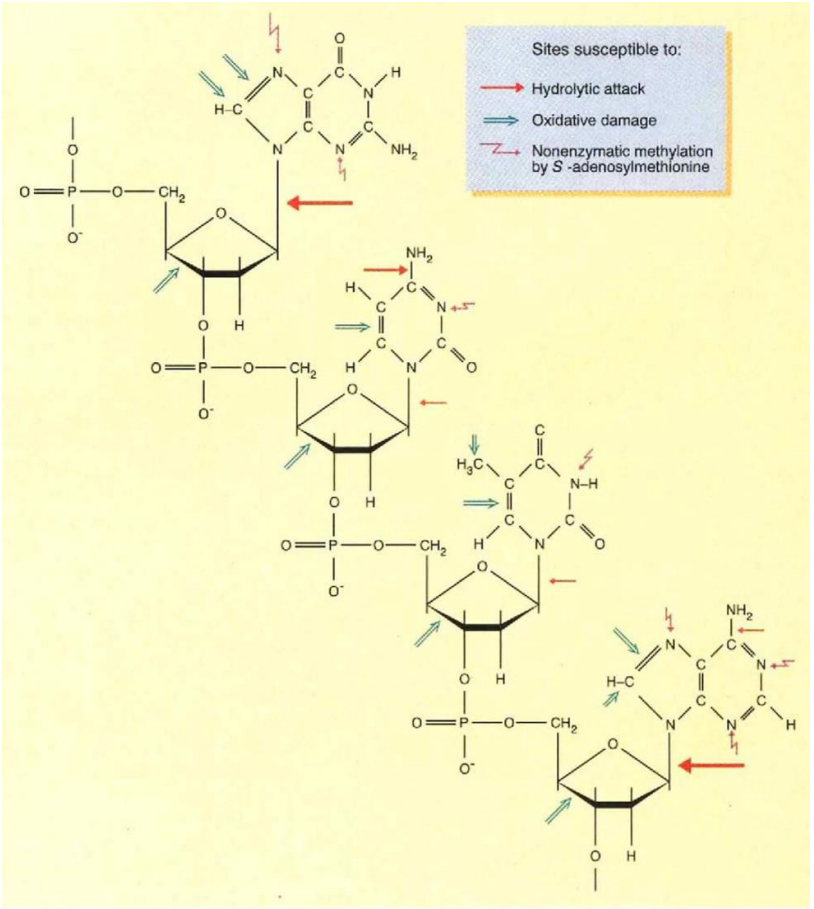


Figure 1.8. Target sites for intracellular DNA decay. A short segment of one strand of the DNA double helix is shown with the four common bases (G, guanine; C, cytosine; T, thymine; A, adenine). Sites susceptible to hydrolytic attack are indicated by solid red arrows, oxidative damage by open green arrows, and nonenzymatic methylation by 5-adenosylmethionine as zig-zagged purple arrows. Major sites of damage are indicated by the large arrows. Hydrolytic and oxidative damage, but not methylated residues would accumulate in ancient DNA [from Lindahl, 1993a].

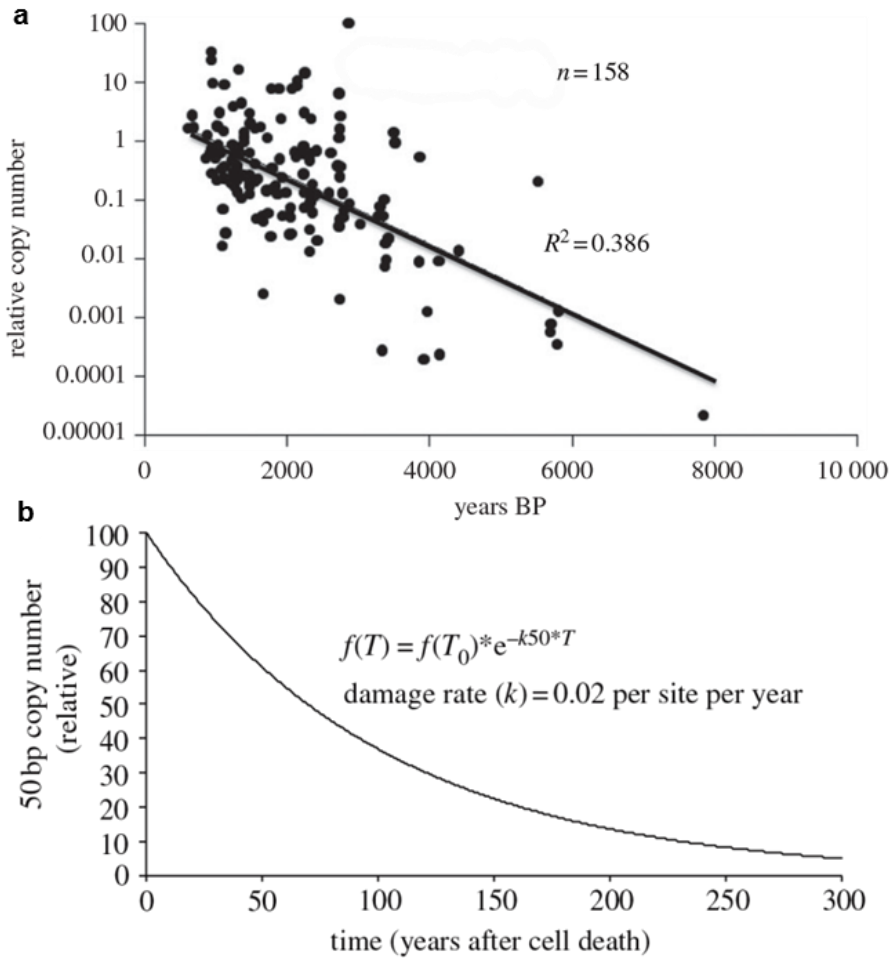


Figure 1.9. (a) Correlations between age and DNA preservation. Relative mtDNA copy numbers in moa (Aves: Dinornithiformes) bone plotted against age for all 158 fossils. (b) A hypothetical signal of temporal DNA decay. The model assumes that the observed damage fraction (λ) can be converted to a rate of decay (k) when the age (T) of a sample is known. It implies that the number of DNA copies of a given length (L) will decline exponentially with time, hence the notion that DNA has a half-life [from Allentoft et al., 2012].

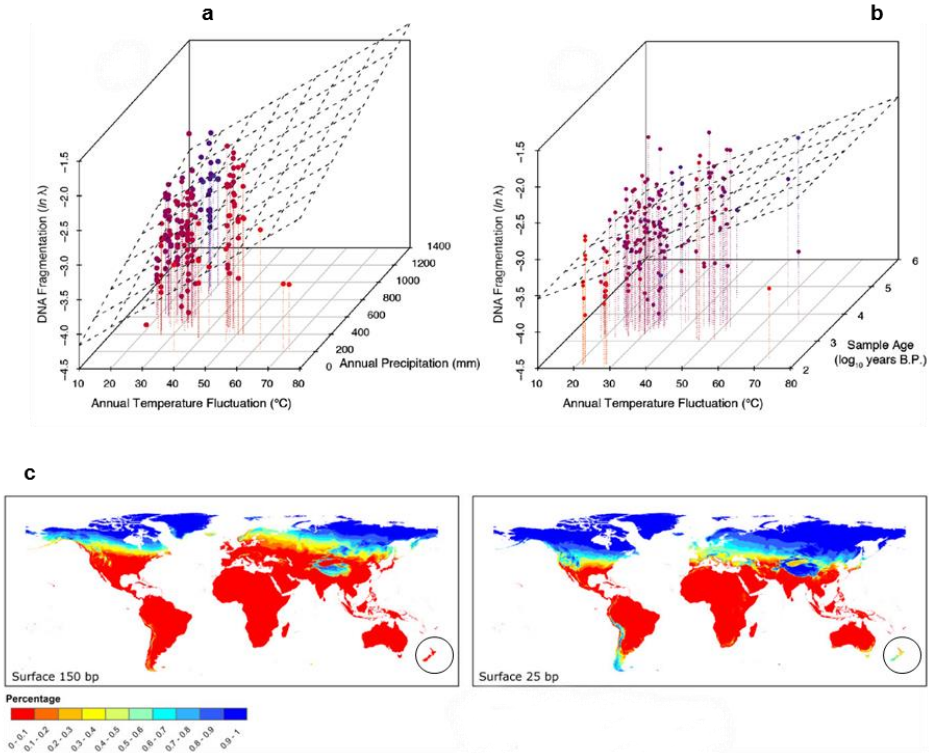


Figure 1.10. Relationships between DNA degradation parameters and environmental variables. (a) DNA fragmentation is correlated with thermal fluctuation and precipitation. (b) DNA fragmentation is correlated with thermal fluctuation but is not influenced by sample age. Coloring is used to enhance the *z*-axis variation: red points are the nearest and blue are the most distant. (c) Estimation of DNA survival. Expected survival of DNA after 10,000 years, for 150 bp fragments (left) and 25 bp fragments (right). While it is clear that the chances of any DNA surviving in desert and tropical regions is minimal for any fragment length, in temperate regions the recovery of short fragments is much more likely than that of longer ones. Note the case of New Zealand from where the moa samples from the previous figure were obtained (circles) [from Kistler et al., 2017 (a, b); Hofreiter et al., 2014(c)].

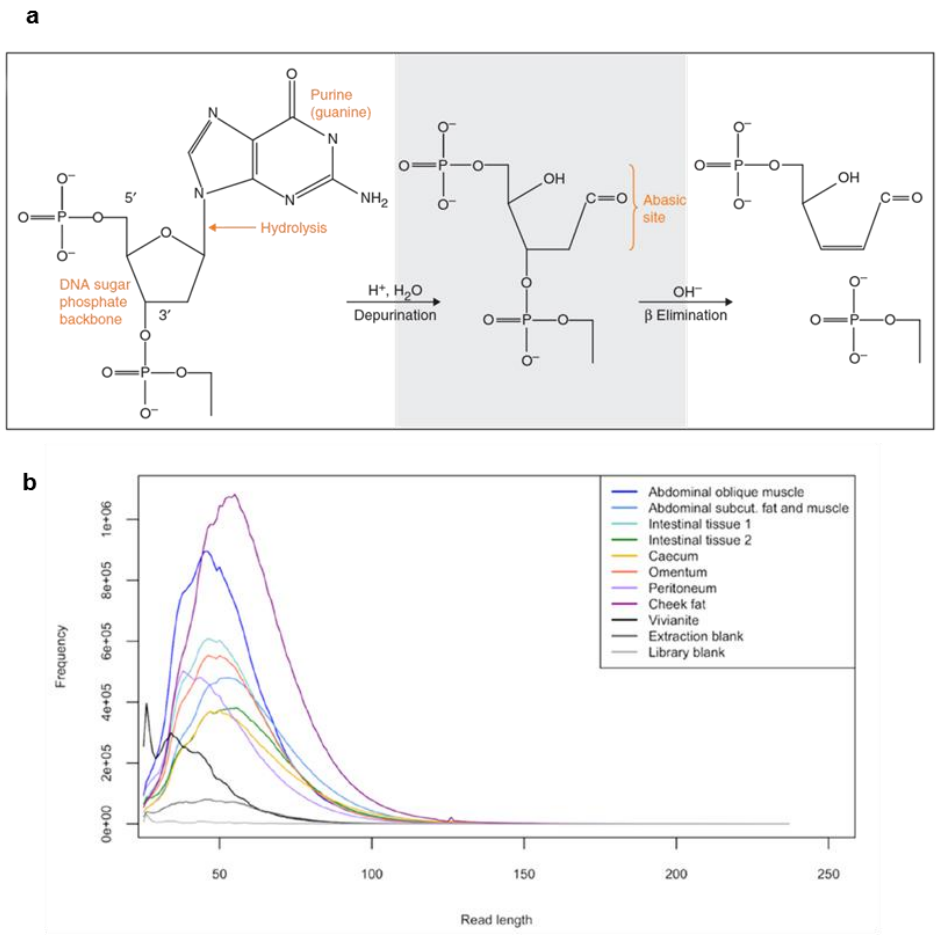


Figure 1.11. Ancient DNA fragmentation. (a) A likely cause of fragmentation in ancient DNA is depurination, in which the N-glycosyl bond between a sugar and an adenine or guanine residue is cleaved, resulting in an abasic site. The DNA strand is then fragmented through β elimination, leaving 3'-aldehydic and 5'-phosphate ends. (b) Read length distribution of samples recovered from a 42,000-year-old permafrost-preserved woolly mammoth calf. Note the characteristic leptokurtic and right-skewed distributions. The amount of DNA recovered depends on the tissue, while the length distribution of reads is similar. Average read length ranges from 34 to 56 bp [from Dabney et al., 2013 (a); Ferrari et al., 2018 (b)].

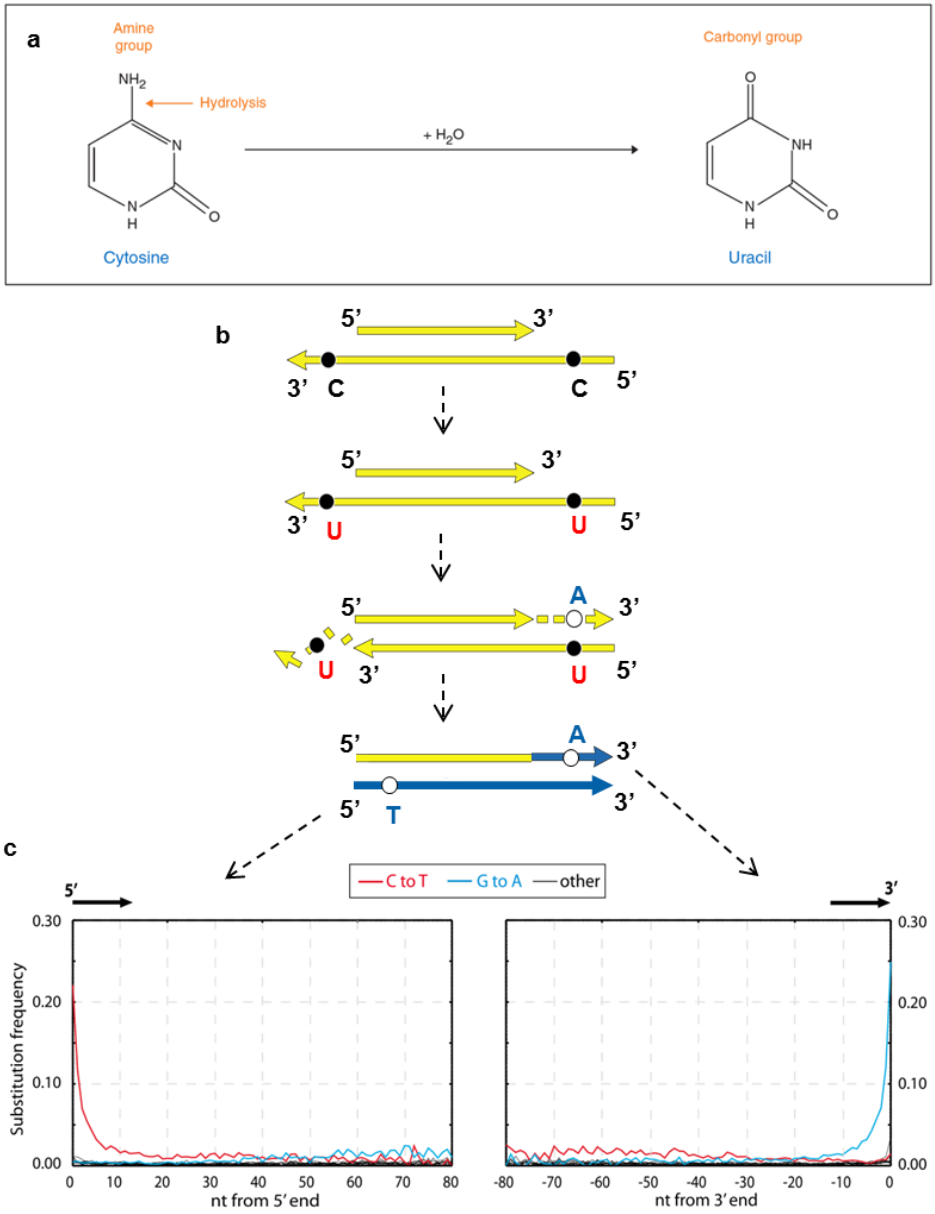


Figure 1.12. Ancient DNA miscoding lesions. (a) Deamination of cytosine to uracil is the major mechanism leading to miscoding lesions. (b) Blunt end repair with T4 DNA polymerase will remove 3' overhangs and fill in 5' overgangs incorporating an A across from the U, and in turn a T across from the A, causing apparent G to A and C to T substitutions. (c) Misincorporation patterns in Neandertal DNA sequences. The frequencies of the 12 possible mismatches are plotted as a function of distance from 5' and 3'-ends [from Dabney et al., 2013 (a); modified from Briggs et al., 2007 (b, c)].

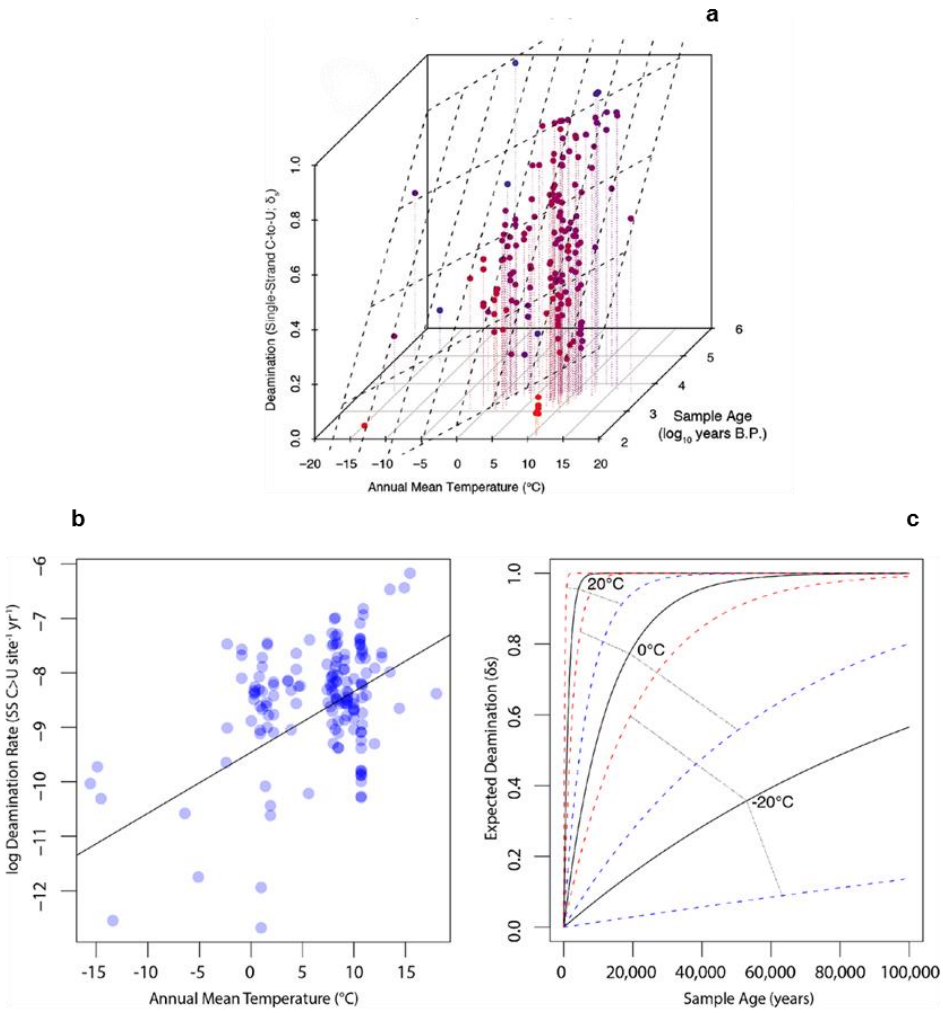


Figure 1.13. (a) Deamination is a thermal age parameter, strongly associated with both age and temperature. Coloring is used to enhance the z-axis variation: red points are the nearest and blue are the most distant. (b) Linear regression of temperature and log deamination rate. (c) Expected deamination (δs values) over a 100,000 year time span. The expected deamination levels are shown for samples from -20, 0 and 20ºC contexts (solid lines), along with upper (red) and lower (blue) 95% confidence bounds. Note that the uncertainty of predictions increases greatly with time [from Kistler et al., 2017].

1.2.3. Straddling failure and success

The Austrian Nobel Prize in Physiology or Medicine Konrad Lorenz wrote in 1949 the book "Auf den Hund kommen" translated into English as "*Man Meets Dog*", where in a very entertaining and amusing language he analyzes the relationship between humans and dogs and speculates on the domestication process, motivated by the common interest in hunting. If instead of being an ethologist, the author had been an evolutionary biologist, the book would surely have been entitled "*Man Meets Horse*". Certainly the domestication of the horse is undoubtedly the most extravagant we humans have ever undertaken, since instead of being based on a predator-prey trophic relationship (cow, sheep, goat, pig...), or predator-predator (dog and cat), it is based on a ludic relationship. We probably domesticated horses to ride them, although we also took advantage of their meat and traction power, and used them as a means of transportation, or in warfare (Levine, 1999). For some 6,000 years (Goodwin, 2007) the horse has been part of human history, but for the last 150 years it has also acquired a relevant role in evolutionary biology. The first clear example of an evolutionary process based on the fossil record was the transformation of the horse (Marsh, 1879; section 1.1.1). This model, although modified over time, has become archetypal and is still used today in scholarship and academia to explain evolution.

A century later, fate brought an equid back to play a foundational role in evolutionary biology. The new discipline of paleogenomics began with the recovery of aDNA from a quagga, an extinct type of zebra, preserved for 120 years in a German natural history museum (Higuchi et al., 1984). Henceforth, the aDNA moved on the back of a horse, straddling failure and success.

The history of aDNA over the last 40 years fits very well with a model developed by the Gartner company, used in the assessment of new technologies and their marketing: the hype cycle (Figure 1.14), which in turn is inspired by the Amara's law:

"We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run" (Ratcliffe, 2016).

Shortly after the publication of the quagga paper, another paper was published in which aDNA was extracted from ancient Egyptian mummies (Pääbo, 1985). Obtaining aDNA was very laborious and the results were often not reproducible. It seemed that paleogenomics would die young. But the discovery of the polymerase chain reaction (PCR) by Kary Mullis revolutionized molecular biology, and earned him the Nobel Prize in chemistry in 1993.

From then on, paleogenomics experienced exponential growth. The phase of inflated expectations began (Figure 1.14). aDNA sequences were recovered from the Tasmanian wolf (Thomas et al., 1989), the New Zealand moa (Cooper et al., 1992) and the woolly mammoth (Hagelberg et al., 1994). aDNA was also recovered from plants, such as maize (Rollo et al., 1988). And of course from humans, on this occasion, however, aDNA was isolated not from mummified tissues but from bones, the most abundant tissue among fossil and subfossil remains and, consequently, the most promising source (Hagelberg et al., 1989).

But as is often the case in any human endeavor, paleogenomics was also affected by the competitive germ. A race began to recover older and older aDNA. Perhaps inspired by Michael Crichton's novel Jurassic Park (1990), one paper claimed to have obtained aDNA from a bee of 25-40 million years ago trapped in amber (Cano et al., 1992). Another one, claimed to have recovered aDNA from 25-30 million year old termites (DeSalle et al., 1992). It also appeared that aDNA had been recovered from magnolia from 17-20 million years ago (Golenberg et al., 1990). And finally arrived the paroxysm with the recovery of aDNA from a 120-135 million year old weevil also trapped in amber (Cano et al., 1993) and the astonishing recovery of part of the cytochrome b (mtDAN) from an 80 million year old dinosaur (Woodward et al., 1994). *Science* and *Nature*, the most prestigious science journals in the world, published these phenomenal findings. Science had reached the Mesozoic horizon and a Cretaceous Park could now be envisioned.

But from the peak of inflated expectations paleogenomics plunged to the trough of disillusionment (Figure 1.14). The biochemist Thomas Lindahl stepped in to bring order among so much genetic

madness and studied the processes of DNA degradation *in vitro*. He observed that hydrolysis, oxidation and methylation rapidly destroy DNA post-mortem, and calculated aDNA survival of no more than 100,000 years under the best environmental conditions (Lindahl, 1993a) (see previous section 1.2.2). The coup de grâce was witnessed by two molecular biologists when they elaborated a phylogenetic tree of cytochrome b that included several tetrapods and the dinosaur of Woodward et al., (1994), which instead of being related to birds or reptiles, was placed as a sister taxon to humans (Hedges & Schweitzer, 1995). The conclusion was inevitable: studies that claimed to have recovered aDNA millions of years old had actually sequenced modern DNA contaminations. This result invalidated all publications with aDNA older than one million years, named antediluvian DNA (Lindahl, 1993b). And it even invalidated one of the foundational articles of the discipline, the recovery of aDNA from Egyptian mummies of Pääbo (1985).

From that moment on, paleogenomics revised and reformulated the protocols for obtaining aDNA and practically draconian measures were proposed (Cooper & Poinar, 2000). Many studies focused on extinct animals, where contamination by modern human DNA was easy to detect and eliminate (Lalueza-Fox, 1999). Others, however, led by Svante Pääbo, preferred to study ancient human and Neanderthal populations, adopting many measures to avoid contamination.

The development of the next-generation sequencing (NGS), was by far the most transformative technology in the history of aDNA research, profoundly affecting wet-laboratory and dry-laboratory activities alike (Orlando et al., 2021) and marked the entry into the slope of enlightement (Figure 1.14). At its most basic level, the success of NGS lies in its ability to accommodate the sequencing and analysis of millions of loci in parallel from minute amounts of ultrashort DNA fragments.

Orlando et al., 2021 provides a careful review of the aDNA with concepts and state-of-the-art methods. It also presents the analytical milestones in aDNA research (Figure 1.15) that have allowed paleogenomics to enter the last phase of the hype cycle: the plateau of productivity (Figure 1.14). At this stage, high quality

aDNA papers are published on topics that were initially unimaginable. It has been possible to recover the complete aDNA of several extinct species such as the mammoth, the dodo, the cave bear, or the aurochs (see section 1.3.2) and several ancient microbes. The genome of human populations from Vikings to Paleo-Inuit to Neanderthals has also been sequenced, a fact that, by the way, was recognized with the Nobel Prize in Physiology or Medicine 2022 to Svante Pääbo.

And horses were not far behind. From that initial mitochondrial genome of the quagga, it has been possible to sequence its complete nuclear genome (Jónsson et al., 2014). It has also been possible to reconstruct the complex phylogenetic tree of equids (Vilstrup et al., 2013; Barrón-Ortiz et al., 2017; Gaunitz, et al., 2018), understand the process of horse domestication (Fages et al., 2019), discover the horse coats most prized by medieval horsemen (Wutke et al., 2016), verify that genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art (Pruvost et al., 2011), and sequence a 1.12-times coverage draft genome from a horse bone recovered from permafrost dated to approximately 560–780,000 years BP. the oldest full genome sequence determined so far by almost an order of magnitude (Orlando et al., 2013). We will not be able to extend the time horizon much further, but we will continue to expand aDNA applications and recoveries to samples that have seemed impossible until now.

Hopefully, the horse will continue to spur paleogenomics ahead.

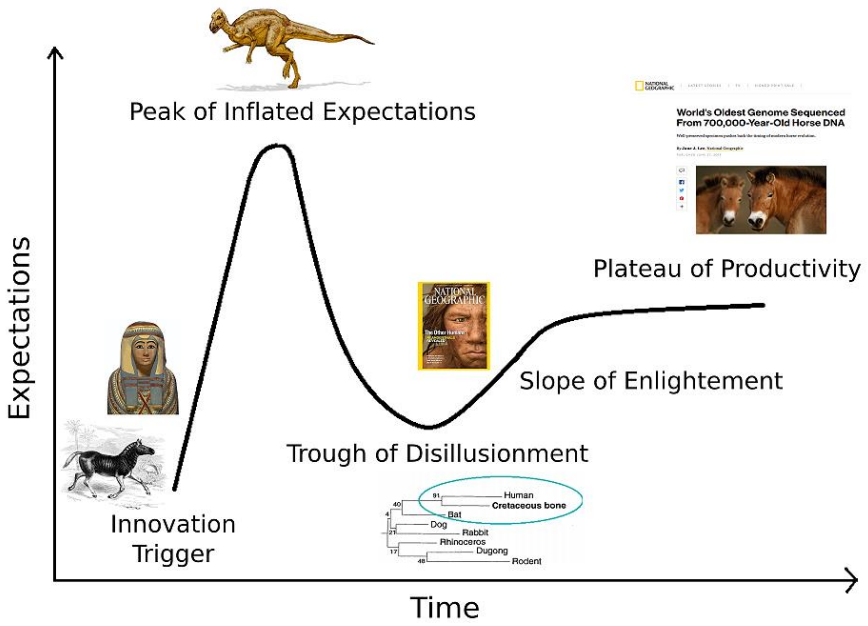


Figure 1.14. The Hype Cycle is a graphical tool developed by Gartner, an information technology research and advisory company based in Connecticut (USA). Depicts how expectations on a new technology change over time, and identifies five phases. Paleogenomics fits this cycle surprisingly well. Illustrations representative of each phase are included (see main text) [from: Patrícia Chrzanová Pečnerová, 2017 https://www.molecularecologist.com/].

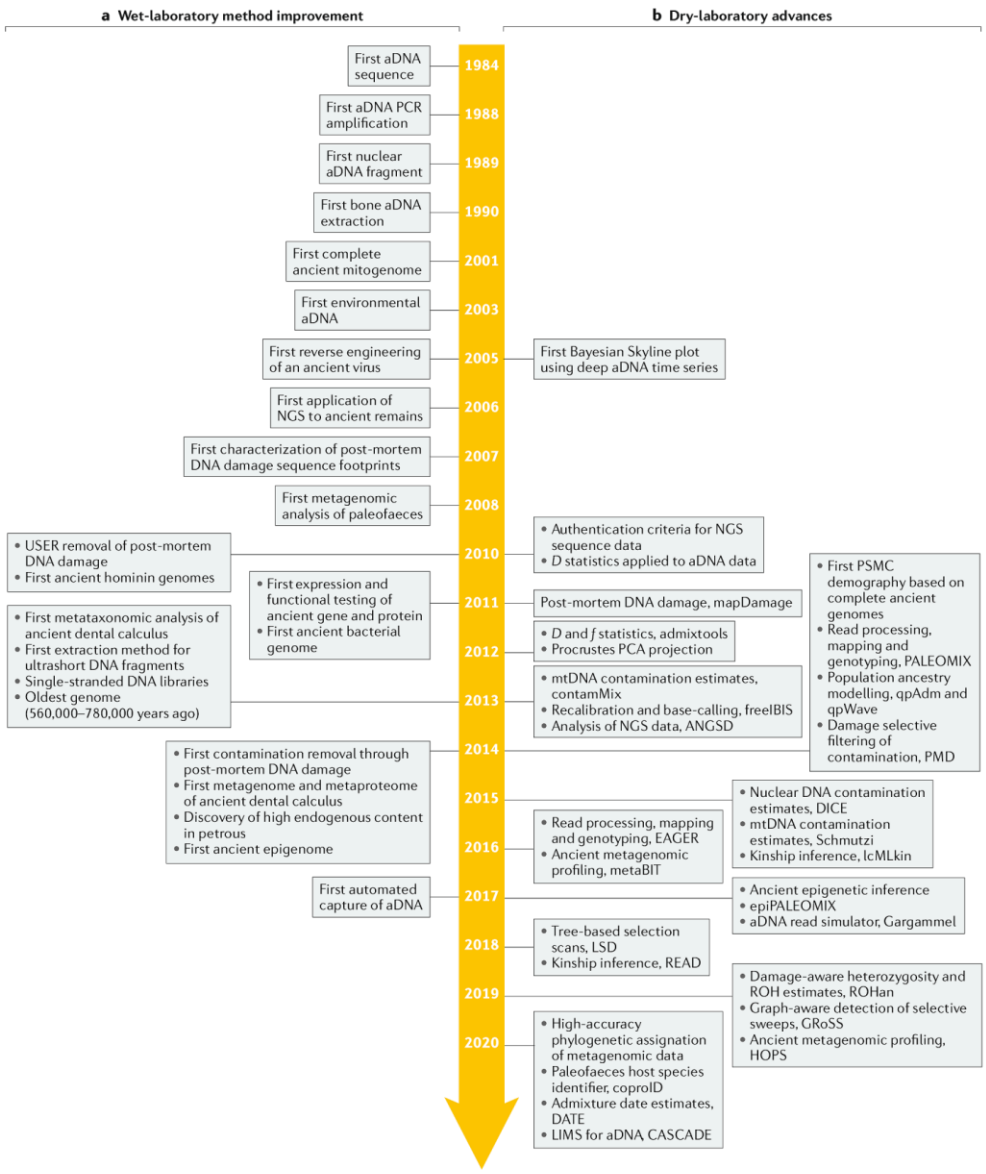


Figure 1.15. Analytical milestones in aDNA research. Key milestones pertaining to wet-laboratory method improvement (part a) or dry-laboratory advances (part b). aDNA, ancient DNA; LIMS, laboratory information management systems; mtDNA, mitochondrial DNA; NGS, next-generation sequencing; PCA, principal component analysis; PSMC, pairwise sequentially Markovian coalescent; ROH, runs of homozygosity; USER, uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) (New England Biolabs) [from Orlando et al., 2021].

1.3. Extinct species

1.3.1. Selecting extinct species

A person dies twice: when he passes on, and when his family forgets him, which usually happens in the fourth generation (with whom we rarely coexist). Very few people remember the full names of their great-grandparents. Genealogy is documentary research that allows us to resurrect the memory of sadly forgotten family members. It provides the name, the most relevant dates (birth, marriage and death), the profession, the heritage and the link with relatives of the same age, ancestry and descent.

It happens analogously with the species. An animal becomes extinct twice: when the last individual disappears and when the collective memory is lost. But there are many particularities (Figure 1.16).

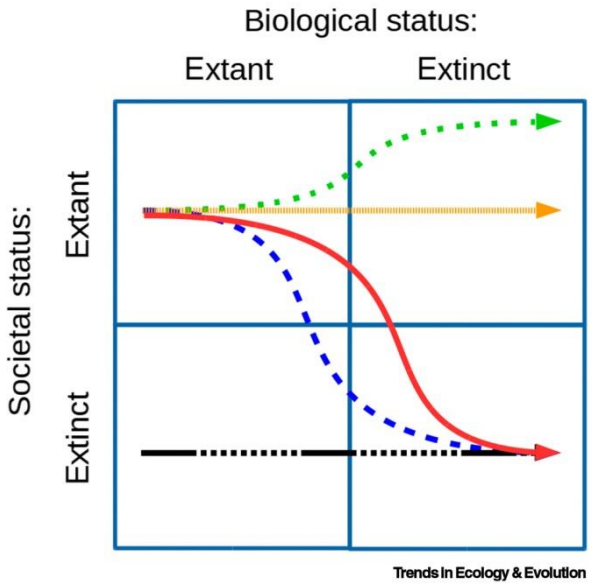


Figure 1.16. Diagram outlining the main conceptual trajectories between biological and societal species status (extant or extinct). The figure presents only the major types of scenarios, while there are more complex combinations of trajectories. It also does not present transient peaks in societal salience, such as those that often follow extinction reports, nor the trajectories characterized by extinction 'reversal', for example, due to species rediscovery, reintroduction, or de-extinction [from Jarić et al., 2022].

The standard situation is oblivion shortly after extinction (red line), we would find many examples such as the Ebro river sturgeon (†*Acipenser sturio*, 1970), the Mediterranean monk seal in the Balearic Islands (†*Monachus monachus*, 1958), the Barbary lion in North Africa (†*Panthera leo leo*, 1922), or the auroch in Europe (†*Bos primigenius*, 1627). Oblivion can also occur before extinction (dotted blue line) such as the naiads, a river mussel that was eaten in some parts of Spain and its shells used, until it was no longer exploited although still today some population survives. It may also happen that the social relevance of a species is not affected by its extinction (dotted orange line), such as the wolf (†*Canis lupus*), which remains alive in the popular imagionary after a century of beeing absent in most of Europe (we find it in the tale of Little Red Riding Hood, in the fable of Peter and the Wolf, in countless sayings and proverbs, in coat of arms, toponyms and surnames[2]), or the North Atlantic right whale († *Eubalaena glacialis*), virtually exctinct in the 19th century[3] due to overfishing in the eastern population, but still present in Basque popular culture (Salvador & Nores, 2017).

Another case occurs when a certain species has no social relevance either before or after extinction (black dotted line), such as most invertebrates ("bugs"), amphibians, or fish that are continuously going extinct without even having been described by science. But perhaps the most curious case is the one that increases the social relevance of a species, after or even long after extinction (dotted green line). In this last group there are many iconic or emblematic species such as dinosaurs, the woolly mammoth (*Mammuthus primigenius*), the saber-toothed tiger (*Smilodon fatalis*), the Irish elk (*Megaloceros giganteus*), the dodo (*Raphus cucullatus*) (Figure 1.17), the marsupial wolf (*Thylacinus cynocephalus*), the migratory pigeon (*Ectopistes migratorius*), the giant auk (*Pinginus impennis*), or the moas (Dinornithiformes)

[2] The Spanish surname *López* comes from Lope, which in turn derives from the Latin *lupus*, wolf. In 2006 it was the fifth most common surname in Spain and the most common in the province of Lugo. The Catalan surname *Llop* is the most common in the region of Terra Alta. The surname *Wolf* is widespread in Europe, especially in Germany.

[3] During the 20th century, there was only one confirmed sighting in 1977 in the Cantabrian Sea (Aguilar, 1981; Notarbartolo di Sciara *et al.* 1998).

among many others. These are generally species, of hair and feathers, which, as the ecologist Ramon Margalef said, we feel particularly connected to because of the phylogenetic proximity. In fact, going further, the ethologist Konrad Lorenz already pointed out that the disproportionately large heads and eyes, rounded snouts or short beaks of puppies and chicks remind us of our own babies and awaken a certain parental instinct. Instinct that, by the way, the illustrator Walt Disney knew how to exploit when he created his cartoon characters based on young mammals and birds with anthropomorphic features.

However, even extinct species that survive in the collective memory gradually transform over time, and often become imprecise, stylized or simplified, and get detached from the actual species.

The sequencing of the genome of extinct species serves to recover the collective memory of those species already forgotten or even never known by society, but it also helps to maintain the memory of those that still retain some social relevance. The FASTA file is the *in silico* recovery of the species. It could be considered a first de-extinction. The return to life of this species is the second de-extinction.

And why is it important to preserve the collective memory? To avoid the shifting baseline syndrome, a psychological and sociological phenomenon by which people constantly lower their thresholds of accepted environmental conditions. In the absence of past information or historical experience, each new generation accepts as normal the increasingly impoverished situation in which they are born and raised. Although it is hard to believe, thousands of years ago in Europe there were large concentrations of tens of thousands of herbivores like the ones we can still see today in the African savannahs, and any rocky coastline in the Mediterranean was covered with luxuriant corals and inhabited for large fish such as those currently found in the Medes Islands Natural Park (Girona, Spain) or in the Cabrera National Park (Balearic Islands, Spain).

Not being aware that species were there and have since gone extinct can produce a false perception of the severity of threats to biodiversity, leading us to underestimate true extinction rates, and

reduce our will to pursue ambitious conservation goals. For example, it could reduce public support for rewilding efforts, especially if such species are no longer present in our memory as natural parts of the ecosystem (Jarić et al., 2022).

There is a long list of species that are in decline or going extinct silently, invisible to people and societies. Let's sequence them! We must improve our knowledge about evolution, and we need icons to conserve biodiversity.

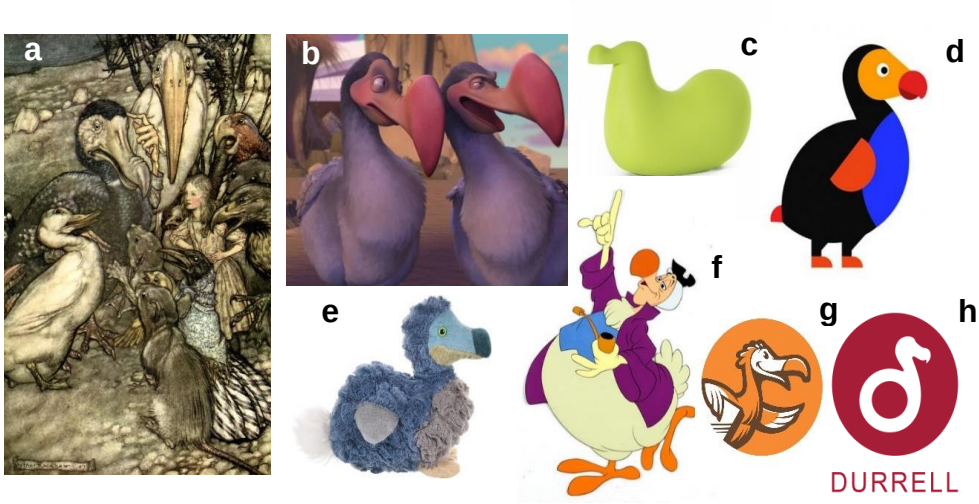


Figure 1.17. The dodo is certainly the most popular icon of species extinction and conservation. Although its rise to celebrity was fortuitous (Turveya & Cheke, 2008) received a major boost when Charles Dodgson included it in Alice's Adventures in Wonderland largely because he identified himself with the bird; he had adopted 'Dodo' as a nickname for himself that reflected his stammer ('Do-Do-Dodgson'; Gardner, 2000). a) Alice's Adventures in Wonderland (Arthur Rackham, 1907). b) Dodos in the computer-animated comedy film Ice Age (2002), c) Design object (Konsepti, Oiva Toikka, 2022), d) The dodo, animalist Youtube channel, e) Dodo plush, f) Mr. Dodo in Alice in Wonderland (Walt Disney, 1951), g) Dodo pizza brand, and h) Durrell Wildlife Conservation Trust.

1.3.2. Selected extinct species

Despite it may seem routine nowadays, the number of whole genomes of extinct species published or known to be underway is really small. It is not surprising because there are even very few genomes for the nearly 35,000 current endangered species (Paez et al., 2022). When we review the list of extinct paleogenomes (Table 1.2), we see that it only includes mammals and birds, with the exception of one invertebrate, and that most of the species are emblematic or iconic. It should also be noted that it incorporates everything from historical specimens preserved in museums that are a century or less old, to fossils or subfossils that are tens of thousands of years old. High coverage generally corresponds to the most recent samples, with the notable exception of the Woolly Rhino (*Coelodonta antiquitatis*). All fossil samples come from temperate or cold latitudes.

When we analyze the topics of the articles referenced in the table, we see that the sequencing of the complete nuclear genome allows us to elaborate more careful phylogenies than mtDNA, to know the evolutionary and demographic history of the species before extinction, to discover where a species came from, if it disappeared slowly, or suddenly, determine the possible causes of its extinction, uncover the functionality of some genes and quantify the degree of inbreeding, and gene flow or hybridization with nearby species. In short, it contributes to a better understanding of evolution.

On a more applied level, the sequencing of complete nuclear paleogenomes also allows population genomics of species on the verge of extinction (and thus, not included in the table), comparing the current diversity with the historical diversity of specimens preserved in museus, as in the case of the kakapo (*Strigops habroptilus*), a flightless parrot endemic to New Zealand, of which only 201 individuals remain today (Dussex et al., 2021).

But it goes further: it opens the door to the dream of de-extinction. For example, the cave bear, despite being extinct tens of thousands of years ago, survives, in part, inserted into the genome of the current brown bear (Barlow et al., 2018). In the same way, the auroch genome survives spread across several British breeds

of cattle (Sinding & Gilbert, 2016). This molecular survival raises the possibility to recover, de-extinct, the species through selective breeding. But with the improvement of gene editing techniques, especially with the advent of CRISPR, it is not even necessary to find the genome of an extinct species in living lineages. It is theoretically possible to obtain the extinct genome by modifying the genome of a nearby living species (see section 5.4).
Consequently, paleogenomes contributes to understand evolution and improve conservation.

Species	Common name	Extinction date	Highlights	Reference
Equus quagga quagga	Quagga Zebra *	1883 CE	Phylogeny of Equids Demographic history Funcionality studies Gene flow in all lineages despite chromosomal numbers	Jónsson *et al*. 2014
Ectopistes migratorius	Passenger Pigeon *	1914 CE	Demographic history Dramatic population fluctuations increases vulnerability	Hung *et al*. 2014
Camelops hesternus	Camelops	ca 13,000 yr BP	Phylogeny with camels and dromedaries Corrects phylogeny based on fossils Usefulness of paleogenomics in distant lineages	Heintzman *et al*. 2015
Mammuthus primigenius	Woolly Mammoth *	ca 3,900 yr BP	Demographic history Bottleneck during Middle & Early Pleistocene Genomic erosion in Wrangle Is. population	Palkopoulou *et al*. 2015
Tympanuchus cupido	Heath Hen	1932 CE	Evolutionary history De-extinction	Johnson *et al*. 2015
Raphus cucullatus	Dodo	Late 17th century CE	De-extinction Evolutionary history Funcionality studies	Shapiro 2016; unpublished
Bos taurus primigenius	Auroch *	1627 CE	Aurochs & cattle cross-breed after domestication De-extinction, from genome conserved in different cow breeds	Sinding and Gilbert 2016
Ursus spelaeus	Cave Bear	ca 24,000 yr BP	Hybridization with brown bears during the Pleistocene Directional & temporal model of gene flow Survival of genes from extinct species	Barlow *et al*. 2018
Thylacinus cynocephalus	Thylacine *	1936 CE	Demographic history Convergenent evolution with canids Phylogeny of marsupial carnivores	Feigin *et al*. 2018
Mammuthus columbi	Columbian Mammoth	ca 10,900 yr BP	Phylogeny & evolution of elephantids Hybridization of Columbian & Woolly Mammoth latitudinally Isolation between jungle & savannah elephant	Palkopoulou *et al*. 2018
Mammut americanum	Mastodon	ca 10,000 yr BP		
Anomalopteryx didiformis	Little Bush Moa	15th century CE	Wing loss study Funcionality studies	Cloutier *et al*. 2018
Coelodonta antiquitatis	Woolly Rhino *	ca 14,100 yr BP	Demographic history No signs of inbreeding suggest abrupt extinction Identifications of artic adaptations like Woolly Mammoth	Lord *et al*. 2020
Pinguinus impennis	Great Auk	1844 CE		Unpublished ???
Conuropsis carolinensis	Carolina Parakeet *	1918 CE	Demographic history Phylogeny within Psittaciformes No signs of inbreeding suggest abrupt extinction Evidence for potential adaptation to toxic diet	Gelabert *et al*. 2020
Rattus macleari	Christmas Is. Rat *	1908 CE	De-extinction Evolutionary divergence limits extinct species genomes Funcionality studies	Lin *et al*. 2022
Glaucopsyche xerces	Xerces Blue *	1941 CE	Demographic history Phylogeny within Polyommatinae subfamily Genomic erosion suggests long population decline	De Dios *et al*. 2022 (in prep.)

Table 1.2. List of paleogenomes from extinct non-human species/subspecies chronologically arranged. (*) Paleogenomes sequenced at high coverage (>10-12x); the maximum coverage ever achieved corresponds to the Christmas Island Rat (60x). Highlighted in gray the paleogenomes sequenced in the present work (12% of the total). Note that all species are mammals or birds, except a butterfly.

1.3.3. Carolina parakeet (*Conuropsis carolinensis*)

The Carolina parakeet (*Conuropsis carolinensis*) is a species that had a biological and social status evolution different from those shown in Figure.1.16. Its decline basically began with the arrival of Europeans in North America in the 17th and 18th centuries, equipped with firearms and with an imperative need to shoot. It was probably initially hunted as entertainment, and because it was considered a pest for crops. But as it became rarer throughout the 19th century, its popularity increased among naturalists, who illustrated it in numerous treatises on ornithology and hunted or captured it with a Victorian collecting spirit. This popularity pushed the pursuit to the last populations located in Florida and explains why today there are more than 400 specimens naturalized and preserved in different museums of the world (Burgio et al., 2018). Therefore, in this case the social relevance of the species appears shortly before the extinction and precipitates it, although there are also other causes that would have influenced it such as the loss of habitat and perhaps the involuntary introduction of some infectious disease from poultry (McKinley, 1980; Snyder, 2004). Today, almost exactly a century after extinction, the popularity of the Carolina parakeet has continued to rise to the point where it has become a true icon of American biologists, naturalists and conservationists and is well known among the general population. Outside the USA, on the other hand, there are a few dozen specimens distributed by some European museums such as The Museum of Zoology (University of Cambridge), the National Museums of Scotland, or the Natural History Museum at Tring (UK), and the species is only known among the circles of ornithologists and specialists. That is why it was a real surprise to locate a specimen preserved in a private collection in Espinelves (Girona, Spain).

During the first week of December 2015 I went to Espinelves on the occasion of the 35th Fir Fair in order to make a live connection at *Divendres*, a magazine of Catalan Television. Since I had to explain the characteristics of the conifers, I arrived earlier to visit a spectacular botanical garden called the *Arboretum*, which has the

largest concentration of monumental trees in Catalonia, such as the giant sequoia, coast redwood and California incense-cedar, more than 40 meters high. Many of these trees were planted by Marià Masferrer Rierola between the years 1860 and 1911. Born in 1856, his status as a second son, without the responsibilities of an heir, but with enough money to live on an income, allowed him to travel and devote his live to studying nature until the end of his days. In fact, he was one of the pioneers of naturalism in Catalonia. Marià also gathered an important collection of more than 250 birds and other vertebrates. He had a lot of experience stuffing animals and collaborated with the Martorell Museum, precursor of the Museum of Natural Sciences of Barcelona. Most of the species are native to the area, but there are also American ones, such as the legendary quetzal (*Pharomachrus sp*), the cardinal (*Cardinalis cardinalis*), and a specimen of Carolina parakeet, nicely preserved.

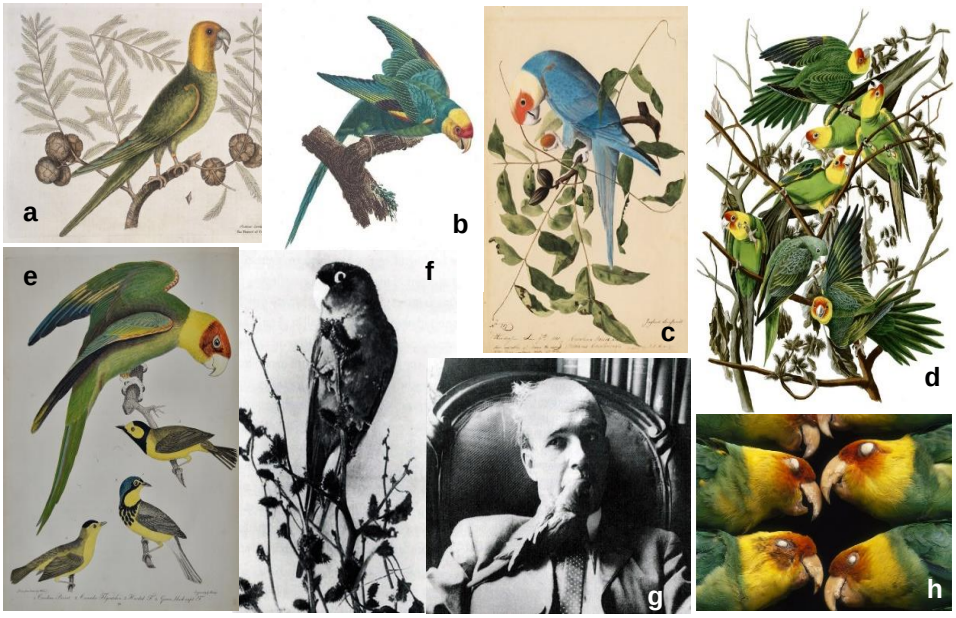


Figure 1.18. The naturalists of the 18th and 19th centuries represented the Carolina parakeet in all its splendor. From the illustrations, we were able to learn that the easternmost population had a bluish hue, that the main food was the toxic cocklebur, or that the juvenile individuals had a green head. One of the last living specimens, bred in captivity, was photographed in black and white. The species was losing its life and color. a) Mark Catesby (ca 1722-1726), b) Jaques Barraband (1801), c-d) John J. Audubon (1811 and 1833), e) Alexander Wilson (1808-1814), f) Robert W. Shufeldt (ca 1900) g) 1906 h) Skins of museum collection.

Seeing the specimen watching me from inside the glass case with the vacant look of glass eyes, I was somewhat distressed to think that never again would anyone be able to see it fly freely, or form noisy flocks over the swamps of the American southeast. It was another victim of the tireless exterminating activity of Humanity. Would there be any way to get this beautiful bird back? Is de-extinction really feasible? To answer these questions, I contacted the aDNA specialist at the Institute of Evolutionary Biology, Dr. Carles Lalueza Fox.

The first step in de-extinction of a species is to sequence its entire nuclear genome. And to our surprise, despite the abundance of specimens in North American museums and despite its popularity, no one had done it.

Thus, we had the free path, to undertake an adventure that culminated in the recovery of the complete genome, contributed to a better understanding of evolutionary processes, and was reflected in a report on the science program of channel 33 *Quèquicom* (https://www.pererenom.com/desextincio-reviure-una-especie/), a scientific article with notable media coverage, and the present doctoral thesis. The milestone of recovering the parakeet *in silico* was fruitful.

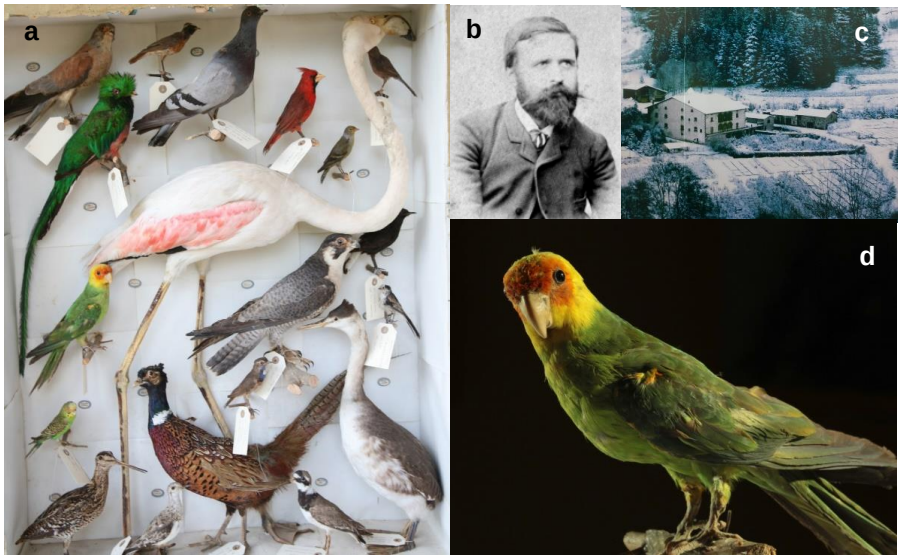


Figure 1.19. a) Bird collection located in Masjoan, b) Marià Masferrer i Rierola (1856-1923) collector of bird specimens, c) Masjoan (Espinelves, Spain), d) speciem of Carolina parakeet analysed [photos Pere Renom (a) and Marc Durà (d)].

1.3.4. Tenerife giant rat (*Canariomys bravoi*)

The next step in the de-extinction of a species is to achieve it *in vitro*. Genetic editing with the CRISPR technique makes it possible to modify the genome of a living species to transform it into the genome of the extinct species (see later section 5.4). But despite the fact that both species are phylogenetically closely related, thousands of modifications must be introduced, which turns genetic editing into a pharaonic and impractical task. The solution is to identify and introduce only the relevant genetic differences between both species, therefore, it is necessary to have an accurate genetic annotation. In birds, the best annotated genome is that of the chicken, phylogenetically too far from the parakeet. The most well-annotated species are the most commonly used in the laboratory: the mouse and the rat. Consequently, the best candidates for eventual de-extinction should be rodents. Because they also have short life cycles, breed easily, and have been cloned with relative success, something that has never been achieved with a bird (cloning is the last phase of de-extinction, see section 5.4).

When choosing a rodent species, it better be iconic, have some unique trait, or help us understand some evolutionary process. And of course it should be easy to sample. A species that fits all these criteria quite well is the Tenerife giant rat (*Canariomys bravoi*). Although practically unknown around the world (even among the inhabitants of Tenerife), it shows the characteristic insular gigantism experienced by many other iconic species such as tortoises in the Galapagos, moas in New Zealand, or the dodo in island of Mauritius, among many others.

The first remains of *C. bravoi* were recovered by Telesforo Bravo in the 1950s in the Acantilado de Martiánez (Martiánez cliff) and other fossil localities in the north coast of Tenerife. The material was later described by Crusafont-Pairó and Petter (1964) who erected the species *C. bravoi*. These authors reported material from several sites (Martiánez, Santa Úrsula, Montaña de Gesa and Callado de Fanabe) but did not provide a detailed description for the new genus and species. More than 20 years later López-Martínez and López-Jurado (1987) provided a diagnosis for the genus

Canariomys (Figure 1.21). These studies are based on the morphology of fossils and subfossils.

Today there are several collections of the giant rat, kept at the Museo de la Naturaleza y el Hombre in Santa Cruz de Tenerife and at the Catalan Institute of Paleontology Miquel Crusafont (ICP).

This rat became extinct in dates close to the arrival of the first human settlers, the Guanches, in the Canary Islands. It could have disappeared by direct hunting, predation by dogs and cats introduced by humans, or even by competition with the common mouse (*Mus musculus*), also introduced.

With the recovery of the genome, it is possible to find out if insular gigantism is a universal process, reached by the different species through the same genetic mutations, or if it involves different genes in each particular case (or a mixture of both hypotheses). It also makes it possible to confirm or rule out the known morphological phylogeny and to understand which was the most likely colonization route. Finally, combined with dating, it allows calculating the speed of insular evolution.

Until now, no molecular study of the giant rat had been carried out, since the climate at the latitude where the Canary Islands are located is too hot and humid to allow good DNA conservation. For this reason, the recovery of mitochondrial and especially nuclear DNA has been a very costly task, which we have only partially managed to accomplish. The merit has been obtaining interesting results where, from the outset, it seemed impossible.

In order to be a good candidate for de-extinction, apart from the criteria mentioned above, it would therefore be necessary to add a temperate or cold geographical origin of the samples.

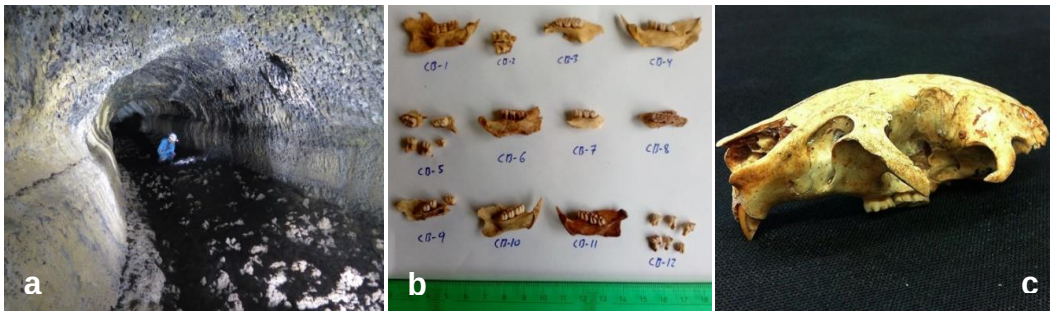


Figure 1.20. a) Cueva del Viento (Tenerife, Canary Islands), a volcanic tube where several remains of the giant rat have been found, b) giant rat samples analysed in the present work c) giant rat skull preserved at the Catalan Institute of Paleontology Miquel Crusafont (ICP) [photos P. Oromí (a), Juan Carlos Rando (b) and Pere Renom (c)].

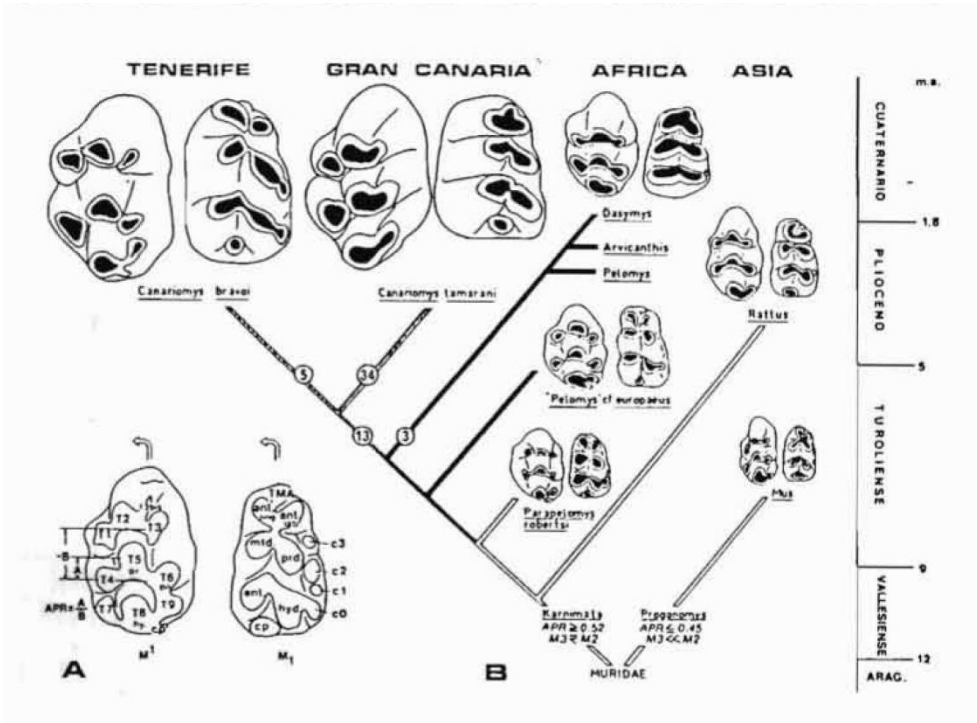


Figure 1.21. Phylogeny of the giant rats of Tenerife and Gran Canaria (Canary Islands) and their African relatives based on dental morphology [from López-Martínez & López-Jurado, 1987].

1.3.5. Xerces Blue (*Glaucopsyche xerces*)

Some insects, particularly butterflies, seem to have all the necessary traits to become ideal candidates for de-extinction. Despite being "bugs", they have a colouration, shapes and flight that we precive as aesthetic (what is the adaptive function of aesthetic perception?). In addition, the vast majority of species show conspicuous differences that we can easily identify, name and classify. Many butterflies had vernacular names long before Linnaean taxonomy arrived. Consequently, it is relatively easy that they become iconic species. Secondly, there are two species that, for very different reasons, have been intensively studied and genetically characterized: the monarch butterfly (*Danaus plexippus*) (Zhan et al., 2011; Zhan & Reppert, 2013) and the silkworm (*Bombyx mori*) (Mita et al., 2004; Goldsmith et al., 2005). To these, we should add yet another insect, the common fruit fly (*Drosophila melanogaster*), which, despite belonging to another order (Diptera), has been the invertebrate most used in genetic studies practically since the foundation of the discipline. Therefore, we have very well sequenced and annotated reference genomes. Butterflies are also relatively easy to breed in captivity, have very short life cycles (annual or biennial), and cloning should not be particularly complex. Finally, regarding the recovery of aDNA, although it is extremely difficult for butterflies to fossilize, some nocturnal species (moths) often rest on tree trunks[4] and can be trapped in resin and preserved for a long time in amber (Poinar, 1993). But also, because of their beauty, butterflies have been collected and kept in museums or private collections since the 17th century. These historical specimens are often in good condition and are a very interesting resource for recovering aDNA from species or populations that are gone (Short et al., 2018).

[4] Perhaps the best known species is the peppered moth (*Biston betolaria*), resting on the birch trunk, it has starred -together with 70 other species- in the phenomenon known as industrial melanism, a paradigmatic example of change by natural selection, which has been widely used in the development and teaching of population genetics and evolution in general.

Beyond de-extinction, it is important to incorporate an insect in the study of evolution through extinct species. Insects make up a huge percentage of Earth's biodiversity. Therefore, any generalization based solely on vertebrates would be somewhat naïve. Extinction dynamics may be particularly different for insects (r-strategists), where population numbers are gigantic compared to vertebrates (most K-strategists).

The species we chose is The Xerces Blue (*Glaucopsyche xerces*) native to the coastal sand dunes of San Francisco (USA) and disappeared in 1941. It is the first well-documented butterfly that has been extinct because of humans. There are other extinct species from remote places where only very old specimens are known, but it is never confirmed if they are still present somewhere, or what exactly happened. In the case of the Xerces, shortly before it disappeared, quite a few specimens were captured for collections, but it seems clear that the ultimate cause of its extinction was the destruction of its habitat (Downey & Lang, 1956; Figure 1.22). Once extinct, it became very popular among American entomologists, who even founded the Xerces Society (https://www.xerces.org/), today a very prestigious entomological association.

This butterfly practically lived with another morphologically quite similar species, the Silvery Blue (*G. lygdamus*), which has survived to the present day. The vicariance between the two species and the great phenotypic variability of Xerces Blues raised doubts about its species status (Downey & Lang, 1956). Initially, one of the objectives of the present work was to elucidate the phylogenetic relationships of both species based on the recovery of aDNA from specimens in museum collections. It happens often that science turns into a race, and in 2021 a team of American scientists published an article in the journal Biology Letters in which this phylogeny was made (Grewe et al., 2021). The results obtained from complete mitochondrial genomes indicate that Xerces Blue was a distinct species that diverged from the Silvery Blue lineage at least 850,000 years ago. However, although mtDNA has led to an unprecedented rise in the identification of cryptic species, it is widely acknowledged that nuclear DNA sequence data are also

necessary to properly define species boundaries. Sometimes even mtDNA and nuclear DNA provide contradictory results (Hinojosa et al., 2019). With these premises we continued the research to definitively clarify the taxonomy and phylogeny of Xerces Blue, and to answer other questions that can only be addressed from the recovery of the nuclear genome. What was the dynamics of Xerces Blue population before extinction? Have Xerces and Silvery Blue followed similar population trends? If they were (almost) sympatric, what was the reproductive barrier that allowed speciation? How do species survive evolutionary bottlenecks? What characteristics have allowed Silvery Blue to survive? The answers to these questions will undoubtedly contribute to the main goal of the Xerces Society: the invertebrate conservation. Although the Xerces Blue is an ideal candidate for de-extinction, the study of its genome should guarantee that its sad fate will not happen again.

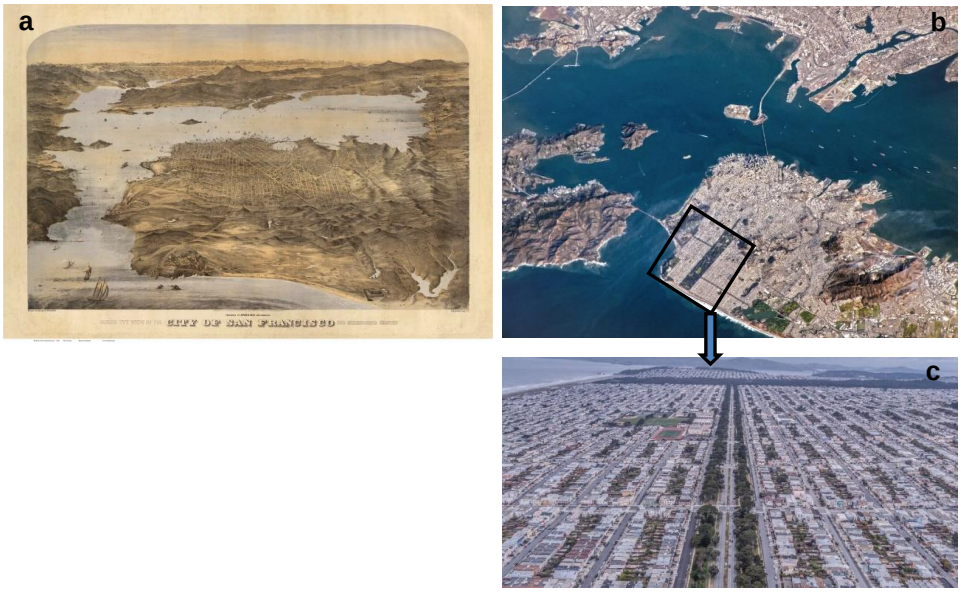


Figure 1.22. a) Bird's eye view of the City of San Francisco and the surrounding contry, drawn by George H. Goddard in 1868. Note the extension of coastal sand dunes habitat of the Xerces Blue, b) Photo from International Space Station (ISS) of San Francisco Bay taken by an astronaut on December 3, 2020. c) Aerial image of the area framed in the upper photo. Note the extension of the urbanized area and the lack of any original habitat.

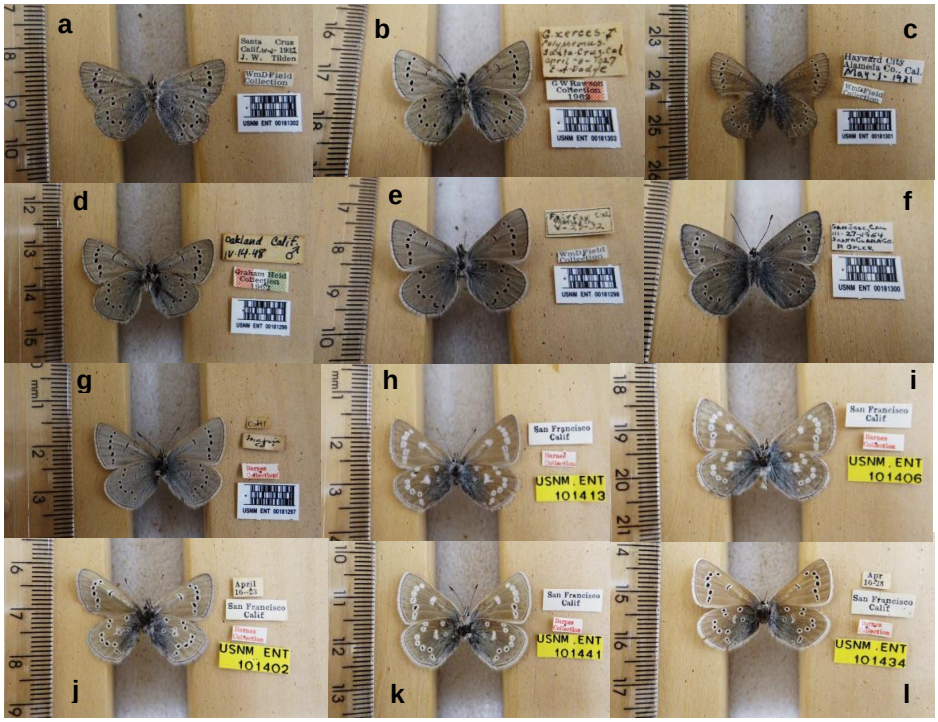


Figure 1.23. Ventral wing surfaces of butterflies analysed in the present work. a-g) Specimens of Silvery Blue (*Glaucopsyche lygdamus*), h-l) Specimens of Xerces Blue (*Glaucopsyche xerces*). Note the pattern of white spots, often without black ocelli, in the later [photos Bob Robbins].

2. METHODS

Materials and methods are described in detail in the corresponding sections of the results, especially in the supplementary sections of each article (section 4). A brief general summary is presented here.

Sampling and DNA extraction

Samples from which aDNA is extracted are very diverse (Figure 2.1), although bones work best. In this thesis we have worked with three different types of samples. From the Carolina parakeet (*Conuropsis carolinensis*) two different samples of about 100 mg were obtained, one from the femur and one from toepads, with the help of a Dremel machine. From the Tenerife giant rat (*Canariomys bravoi*), approximately 30 mg of tooth cementum was obtained by drilling at low speed (5000 r.p.m.) with a new engraving cutter. And from the butterflies Xerces Blue (*Glaucopsyche xerces*) and Silvery Blue (*Glaucopsyche lygdamus*) DNA extraction was performed from 12 abdominal samples.

DNA extraction from bone, teeth or abdomen powder was performed following the method proposed by Dabney et al., (2013). The samples were digested, the solution was then resuspended by vortexing and was incubated in a rotating plate overnight. Digested samples were purified, and DNA was isolated following a combination of Phenol/Chloroform and column purification.

Strict protocols were followed to minimize the amount of human DNA in the ancient DNA laboratory, including wearing a full-body suit, sleeves, shoe covers, clean shoes, facemask, hair net and double gloving.

Library construction

Most current sequencing techniques require modification of the DNA extract before sequencing. This modification generally consists of attaching artificial DNA segments (adapters) to both ends of the DNA molecules in order to be recognised by the sequencing platform. This process is called “library construction”.

In order to attach the adapters, it is necessary that the DNA fragments are blunt-end, but the vast majority of aDNA fragments

have overhanging ends. There are therefore three essential steps to follow in order to prepare the libraries: end-repair, adapter ligation, and nick fill-in; following the BEST protocol (Carøe et al., 2018). First, during the end-repair step, 3' overhangs are removed and 5' overhangs are filled in to create blunt ends. Next, short adapters are ligated to the ends of double stranded DNA molecules.

In severely degraded aDNA samples, surviving molecules may be single-stranded, may contain nicks or end modifications that inhibit adapter ligation. These molecules are lost during the preparation of double-stranded libraries (Gansauge & Meyer, 2013), which impairs the correct characterisation of aDNA. To overcome this limitation, another method has been developed that allows the incorporation of both single-stranded and double-stranded molecules into DNA libraries (Meyer et al., 2012; Gansauge & Meyer 2013). First, double-stranded DNA fragments are converted into single-stranded DNA by heat denaturation. The single-stranded DNA fragments are then immobilised on streptavidin-coated beads, the adapters are attached and used to copy the template strand, which is converted to double-stranded. The library molecules are released from the beads by thermal denaturation and used for PCR amplification.

Amplification

The amounts of DNA recovered from the samples are always very small, so they need to be amplified, i.e. increased in number. It is used the polymerase chain reaction (PCR), a technique that is based on the enzyme DNA polymerase, capable of building a molecule of DNA by sequentially adding on free nucleotides according to the instructions of a template. It is also necessary to use *primers*, which are short single strand DNA fragments that are complementary to the 3' ends. It should be remembered that DNA polymerase can only bind to and elongate from a double-stranded region of DNA. The function of the primers is to provide a double-stranded starting point for the single-stranded DNA, where DNA polymerase will initiate replication. Primers are designed and selected according to the target region of the DNA to be amplified.

PCR is a cyclic process consisting of three steps. The initial step is the denaturation, or separation, of the two strands of DNA. This is achieved by heating the material to a temperature of about 95°C. In the second step, the temperature is reduced to about 55°C so that the primers can anneal to the templates (both strands). In the third step, the temperature is raised again to about 72°C and the DNA polymerase starts adding nucleotides from the primers, synthesising the complementary strand to the template. The whole cycle takes about five minutes. At the end, the temperature is raised again and the process is repeated. The DNA generated is then used as a template for further replication, resulting in a chain reaction that exponentially amplifies the original DNA molecule (2^n) (Figure 2.2).

Targeted enrichment

Most ancient samples have very low proportions of endogenous DNA, the amount of sequencing required to obtain significant information is unaffordable for many research groups. This problem can be overcome by using targeted capture methodologies that increase the endogenous content of aDNA libraries, either during library construction (Gansauge & Meyer, 2014), or afterwards (Haak et al., 2015). In this thesis, an enrichment approach has been applied in the samples of Tenerife giant rat (*Canariomys bravoi*): mtDNA capture.

After library preparation, a sample was enriched for mtDNA sequences with the use of commercially biotinylated probes for mouse mtDNA (MYbaits). Two consecutive hybridizations were conducted with the myBaits. During incubation mtDNA library fragments hybridized to the baits. Library molecules that did not hybridize were exogenous DNA and were thus washed away, while the mtDNA library pool was kept and extracted. After PCR amplification, the enriched library was ready for sequencing.

Sequencing

Different sequencing techniques are available. In this thesis we have used HiSeq400 of Illumina platform (Illumina, USA) in Macrogen, Inc. biotechnology company. All the process of the

Illumina technology occurs in a flow cell, which is usually partitioned in different lanes. The surface of the flow cells is densely covered with forward and reverse primers, complementary to the adapters introduced during library construction (Buermans & den Dunnen 2014). To load the library into the flow cells it is necessary to denature the double-stranded fragments into single-stranded molecules, which are able to hybridise at one end to surface primers. Then, for each original library fragment, a cluster of copies is created using bridge amplification (Figure 2.3, a). In short, the free 3' end of the fragments bends over and hybridizes to a complementary primer on the surface, forming a fixed structure that enables the synthesis of the complementary strand. Next, the double-stranded bridge is denatured and the amplification cycle starts again. Once the process is finished, a flow cell contains millions of spatially separated clusters, each of them with about a thousand copies of an original library template. This amount generates sufficient signal intensity to be detected during the next phase. Sequencing on the Illunina platform is based in the four-colour cyclic reversible termination method (Figure 2.3, b), and comprises four steps: nucleotide incorporation, wash, imaging, cleavage dye and wash. First, DNA polymerase initiates DNA synthesis by incorporating a fluorescently labelled nucleotide, complementary to the base of the template strand. Nucleotides transport a base-unique fluorescent label. In the second step, unincorporated nucleotides are rinsed. During the third step, image is acquired to determine which nucleotide was incorporated. Finally, in the last step the fluorescent dye is removed and a new cycle can begin. Note that only one nucleotide is incorporated in each cycle, therefore the number of cycles sets the length of the resulting reads (Figure 2.3, c). Sequencing on Illumina platform can be performed in both single-end and paired-end modes. In the first case, only one end of the DNA fragment is sequenced using the primer site present in one of the adapters. In the later case, once the sequencing from one end is finished, the complementary strand is synthesized from the primer site in the second adapter (Olalde, 2016).

Bioinformatics processing

Sequencing generates millions of short fragments (reads), which must be processed computationally. The discipline involved is bioinformatics. Raw reads are stored in a specific file format known as FastQ. The first task to be undertaken is to remove the adaptor sequences at the ends of each read, as they could compromise downstream processes. Many different programs exist for removing adapter sequences from high-throughput sequencing data (Kircher 2012; Lindgreen 2012). We mainly used AdapterRemoval that is able to trim from single end and pair end FastQ reads (Needleman & Wunsch, 1970). Only the reads with a final length $\geq$ 30 bases were kept for the subsequent analysis to avoid ambiguous results.

Once the adapters have been removed, the reads can be assembled to reconstruct the whole genome. However, reads from aDNA are extremely short and do not allow for *de novo* assembly. Consequently, palaeogenomics is forced to use a reference genome on which reads can be aligned in a process called mapping (Figgure 2.4). For mapping we used the software Burrows Wheeler Aligner (BWA) (Li & Durbin, 2009). In particular BWA-backtrack, adapted to aDNA, which performs efficiently with short reads, and is capable to perform a global alignment of the read against the reference. This procedure has provided very good results, but it has its limitations. For example, it is necessary that the reference genome belongs to a species phylogenetically close to the palaeogenome to be recovered, otherwise many reads do not have a homologous sequence and are lost. Also, highly repetitive areas of the genome, usually with structural functions such as telomeres or centromeres, cannot be recovered.

When starting the alignment of reads with the reference genome, it is also necessary to remove duplicate reads, which are inevitably generated during the PCR amplification process. Otherwise they would generate a fictitiously high coverage (number of reads recovered for a given position in the reference genome), complicating the detection of polymorphisms. The most practical way to detect duplicates is from the outer alignment coordinates. Reads with identical coordinates are duplicates to be removed.

Finally, it is necessary to ensure that the sequenced reads are authentically ancient, and not the result of contamination with modern DNA. The usual authentication procedures are analysis of the deamination pattern at the read ends, and read length distribution (see section 1.2.2). The software used to characterise aDNA damage patterns are PMDtools (Skoglund et al., 2014) and mapDamage2.0 (Jónsson et al., 2013).

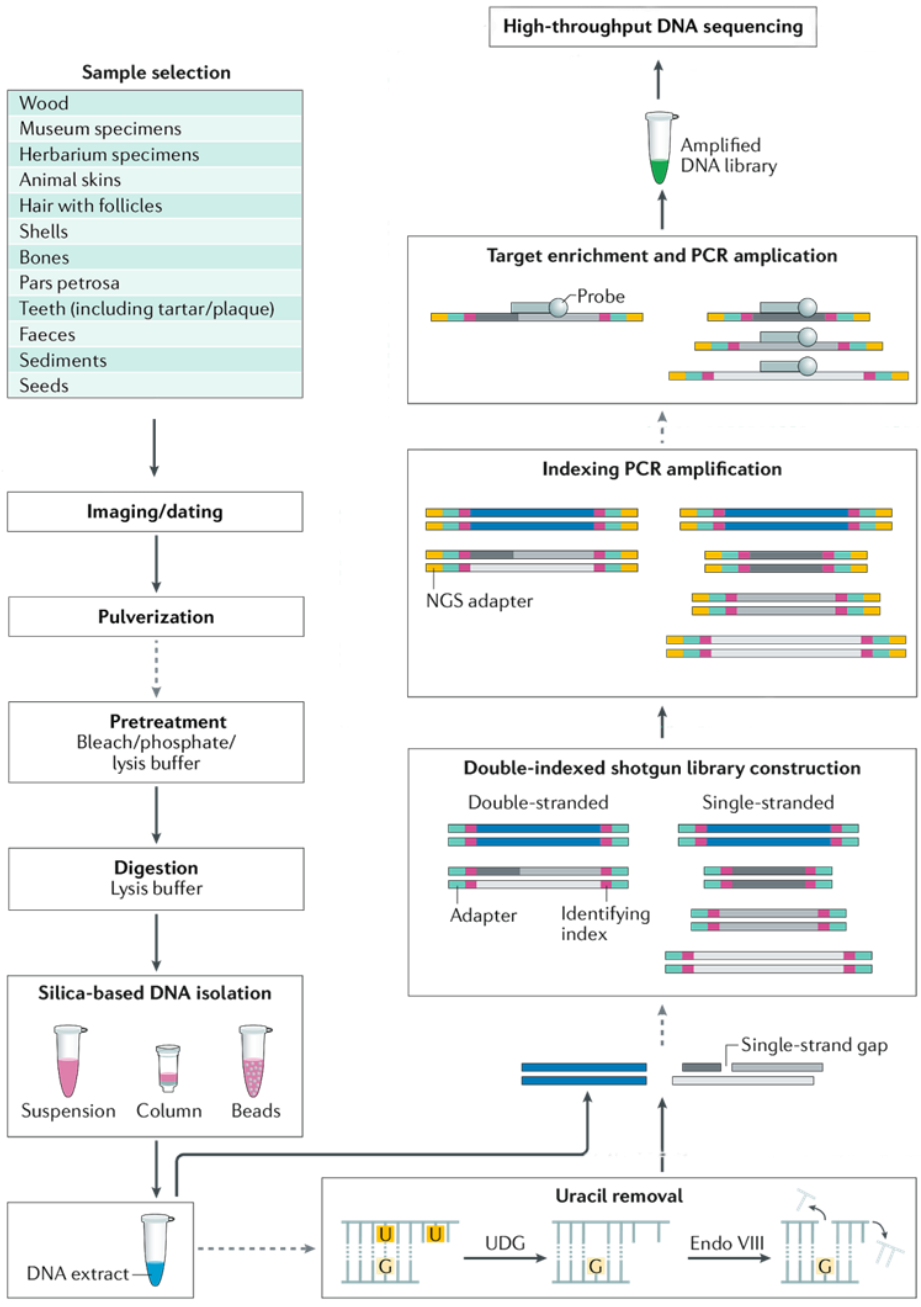


Figure 2.1. Experimental process. A wide range of remains are amenable to ancient DNA analysis. The diferent wet-laboratory procedures include sample preparation, DNA extraction, DNA library construction, target enrichment and PCR amplification. Uracil elimination is performed with uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII). Mandatory step (solid arrow) and optional step (dotted arrow) are indicated [modified from Orlando et al., 2021].

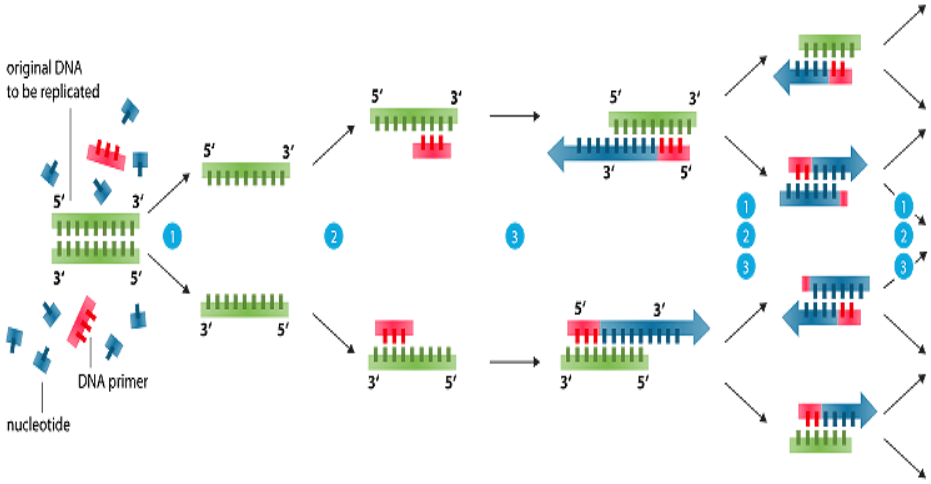


Figure 2.2. Polymerase chain reaction (PCR). A three-step process is carried ou at diferent temperatures: (1) Denaturation (95°C), or separation, of the two strands of the DNA molecule. (2) Annealing (55°C), the primers can anneal to the template. (3) Elongation (72°C), the DNA polymerase begins adding nucleotides onto the ends of the annealed primers. Note that DNA polymerase (not shown) elongates only from 3' to 5' ends. The DNA generated is itself used as a template for new replications, thus the number copies are doubled in each cycle [from wiki commons].

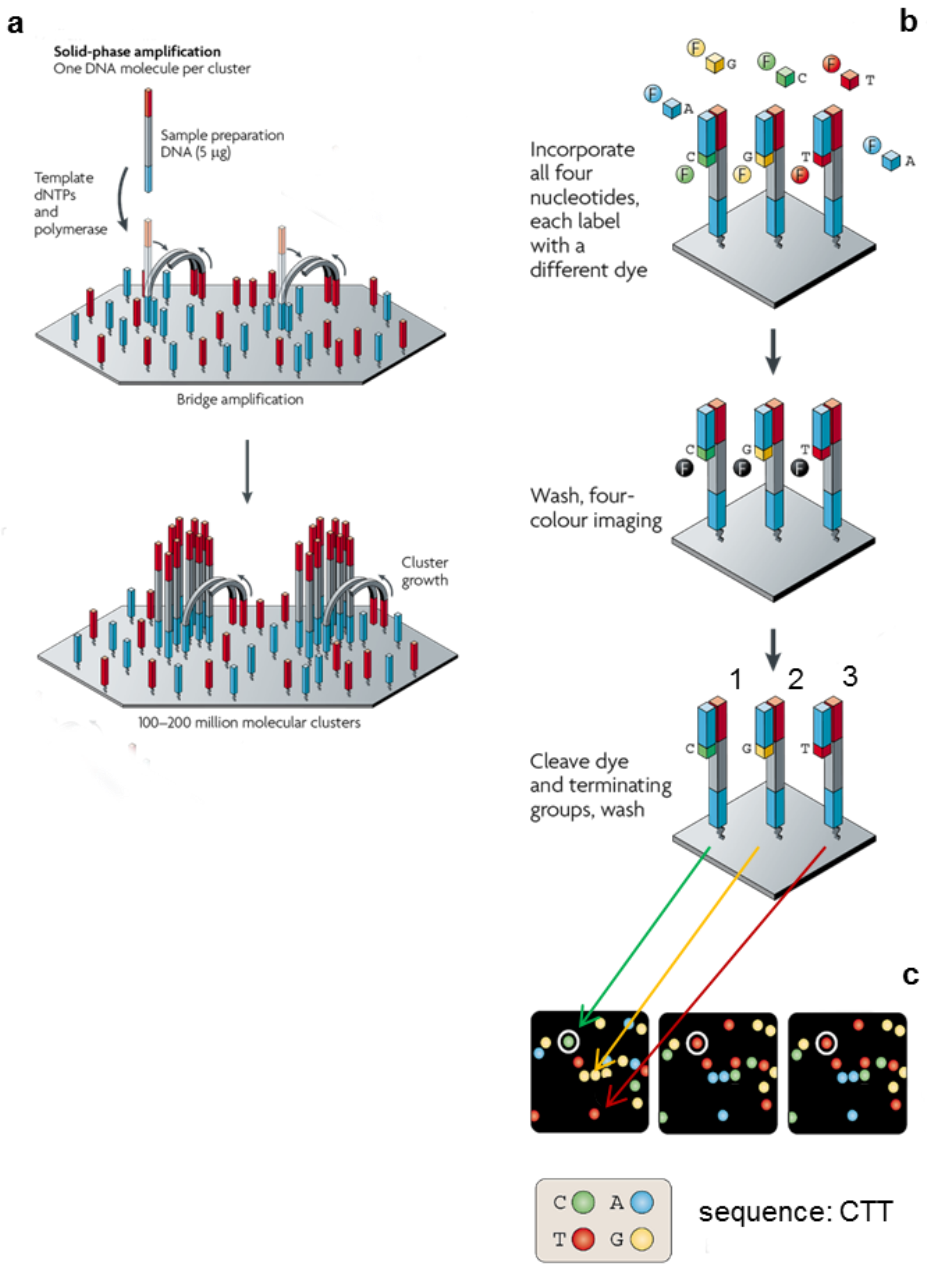


Figure 2.3. Sequencing process a) Template immobilization in solid-phase for cluster generation on the Illumina sequencing platform. b) Sequencing on the Illunina platform by the four-colour cyclic reversible termination method. One cycle is shown. The steps are: nucletide incorporation, wash, imaging, cleavage dye and wash. Clusters are shown as single templates for illustrative purposes c) The four-colour images highlight the sequencing data from clusters. Three cycles are shown for cluster 1 (white circle), and only one cycle is shown for clusters 2 and 3 [modified from Metzker, 2010].

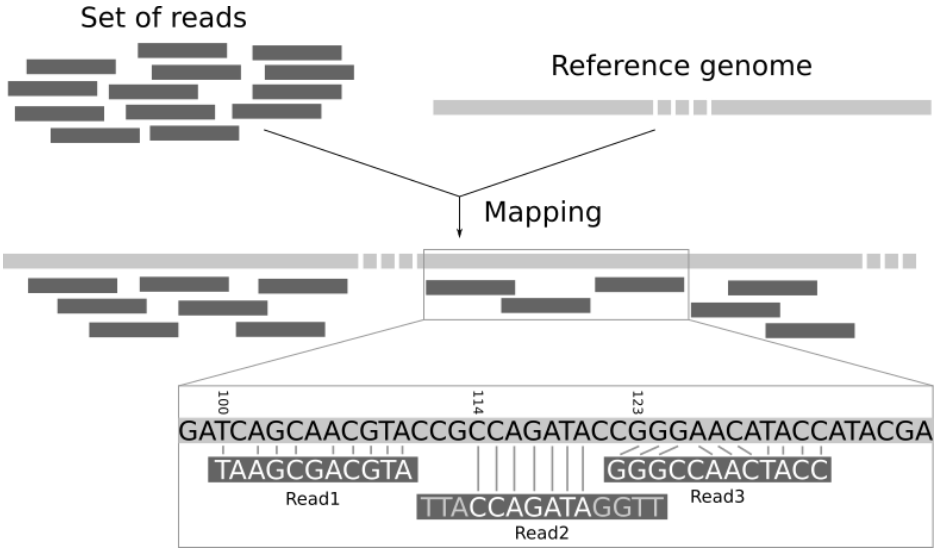


Figure 2.4. Mapping process. Location of the reads on the reference genome. The first read is aligned at position 100 and the alignment has two mismatches. The second read is aligned at position 114. It is a local alignment with clippings on the left and right. The third read is aligned at position 123. It consists of a 2-base insertion and a 1-base deletion [from https://training.galaxyproject.org/].

3. OBJECTIVES

The main objective of the present thesis is to generate genome-wide aDNA of three species of animals extinct during the last centuries or millennia, and belonging to well-differentiated taxa, to try to establish common genomic patterns for declining species, solve several evolutionary questions and find out the causes of their extinction.

The specific objectives are:

(1) Carolina parakeet (*Conuropsis carolinensis*)
- Precisely determine the phylogeny of the Carolina parakeet in the Psittaciform order.
- Explore some gens related to diet.
- Determine de demographic history in order to elucidate the dynamics of population decline.
- Analyze the possible signs of genetic erosion to find out how the extinction occurred.

(2) Tenerife giant rat (*Canariomys bravoi*)
- Determine the continental ancestor of the insular species and the elapsed time since they separated.
- Calculate the evolutionary rate that led it to increase the body mass until it became giant, and compare the value with non-insular mammals.

(3) Xerces Blue (*Glaucopsyche xerces*)
- Definitively establish whether Xerces Blue and Silvery Blue *(G. lygdamus*), are two different species or two regional variants of the same species.
- Place Xerces Blue within the subfamily Polyommatinae tree
- Determine its demographic history in order to elucidate the dynamics of population decline.
- Analyze the possible signs of genetic erosion to find out how the extinction occurred.

4. RESULTS

4.1. Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet

Pere Gelabert, Marcela Sandoval-Velasco, Aitor Serres, Marc de Manuel, Pere Renom, Ashot Margaryan, Josefin Stiller, Toni de-Dios, Qi Fang, Shaohong Feng, Santi Mañosa, George Pacheco, Manuel Ferrando-Bernal, Guolin Shi, Fei Hao, Xianqing Chen, Bent Petersen, Remi-Andre Olsen, Arcadi Navarro, Yuan Deng, Love Dalen, Tomàs Marquès-Bonet, Guojie Zhang, Agostinho Antunes,Thomas P. Gilbert,
and Carles Lalueza-Fox

Current Biology, January 2020; 30: 108-114
https://doi.org/10.1016/j.cub.2019.10.066

Report

Current Biology

Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet

Highlights

- First whole genome from the extinct Carolina parakeet and the sun parakeet
- Divergence time between *Conuropsis* and *Aratinga* around 3 mya
- Evidence for potential adaptation to toxic diet in two extremely conserved proteins
- No signs of inbreeding in the Carolina parakeet suggest the extinction was abrupt

Authors

Pere Gelabert,
Marcela Sandoval-Velasco,
Aitor Serres, ..., Agostinho Antunes,
M. Thomas P. Gilbert,
Carles Lalueza-Fox

Correspondence

tgilbert@bio.ku.dk (M.T.P.G.),
carles.lalueza@upf.edu (C.L.-F.)

In Brief

Gelabert et al. report the whole genome of the extinct Carolina parakeet and provide evidence of its phylogeny, adaptation to a toxic cocklebur diet, and demographic history. The lack of signs of recent inbreeding typically found in endangered species suggests its abrupt extinction was human mediated.

Gelabert et al., 2020, Current Biology *30*, 1–7
January 6, 2020
https://doi.org/10.1016/j.cub.2019.10.066

CellPress

Please cite this article in press as: Gelabert et al., Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.10.066

Current Biology
Report

Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet

Pere Gelabert,[1,2] Marcela Sandoval-Velasco,[3] Aitor Serres,[1] Marc de Manuel,[1] Pere Renom,[1] Ashot Margaryan,[3] Josefin Stiller,[4] Toni de-Dios,[1] Qi Fang,[5] Shaohong Feng,[5] Santi Mañosa,[6] George Pacheco,[3] Manuel Ferrando-Bernal,[1] Guolin Shi,[7] Fei Hao,[7] Xianqing Chen,[8] Bent Petersen,[3,9] Remi-André Olsen,[10] Arcadi Navarro,[1,11,12] Yuan Deng,[5] Love Dalén,[13] Tomàs Marquès-Bonet,[1,11,12,14] Guojie Zhang,[4,15,16,17] Agostinho Antunes,[18,19] M. Thomas P. Gilbert,[3,20,*] and Carles Lalueza-Fox[1,21,*]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Dr. Aiguader 88, 08003 Barcelona, Spain
[2]Department of Evolutionary Anthropology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria
[3]Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark
[4]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
[5]BGI-Shenzhen, Beishan Industrial Zone, Building 11, Shenzhen 518083, China
[6]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Facultat de Biologia, Avinguda Diagonal 643, 08028 Barcelona, Spain
[7]Center of Special Environmental Biomechanics & Biomedical Engineering, School of Life Sciences, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an 710072, China
[8]Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an 710072, China
[9]Centre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), Faculty of Applied Sciences, AIMST University, Semeling Road, 08100 Kedah, Malaysia
[10]ScieLifeLab, Department of Biochemistry and Biophysics, Stockholm University, Frescativägen 40, SE-17121 Solna, Sweden
[11]Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
[12]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08036 Barcelona, Spain
[13]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Frescativägen 40, 10405 Stockholm, Sweden
[14]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, c. de les Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain
[15]China National GeneBank, BGI-Shenzhen, Jinsha Road, Shenzhen 518120, China
[16]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China
[17]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China
[18]CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450–208 Porto, Portugal
[19]Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
[20]NTNU University Museum, Erling Skakkes gate 47c, 7012 Trondheim, Norway
[21]Lead Contact
*Correspondence: tgilbert@bio.ku.dk (M.T.P.G.), carles.lalueza@upf.edu (C.L.-F.)
https://doi.org/10.1016/j.cub.2019.10.066

SUMMARY

As the only endemic neotropical parrot to have recently lived in the northern hemisphere, the Carolina parakeet (*Conuropsis carolinensis*) was an iconic North American bird. The last surviving specimen died in the Cincinnati Zoo in 1918 [1]. The cause of its extinction remains contentious: besides excessive mortality associated to habitat destruction and active hunting, their survival could have been negatively affected by its range having become increasingly patchy [2] or by the exposure to poultry pathogens [3, 4]. In addition, the Carolina parakeet showed a predilection for cockleburs, an herbaceous plant that contains a powerful toxin, carboxyatractyloside, or CAT [5], which did not seem to affect them but made the birds notoriously toxic to most predators [3]. To explore the demographic history of this bird, we generated the complete genomic sequence of a preserved specimen held in a private collection in Espinelves (Girona, Spain), as well as of a close extant relative, *Aratinga solstitialis*. We identified two non-synonymous genetic changes in two highly conserved proteins known to interact with CAT that could underlie a specific dietary adaptation to this toxin. Our genomic analyses did not reveal evidence of a dramatic past demographic decline in the Carolina parakeet; also, its genome did not exhibit the long runs of homozygosity that are signals of recent inbreeding and are typically found in endangered species. As such, our results suggest its extinction was an abrupt process and thus likely solely attributable to human causes.

RESULTS

The Carolina Parakeet and the Sun Parakeet Genomes

Given that *de novo* genome assembly is impractical with the typically short and degraded DNA found in historic and ancient

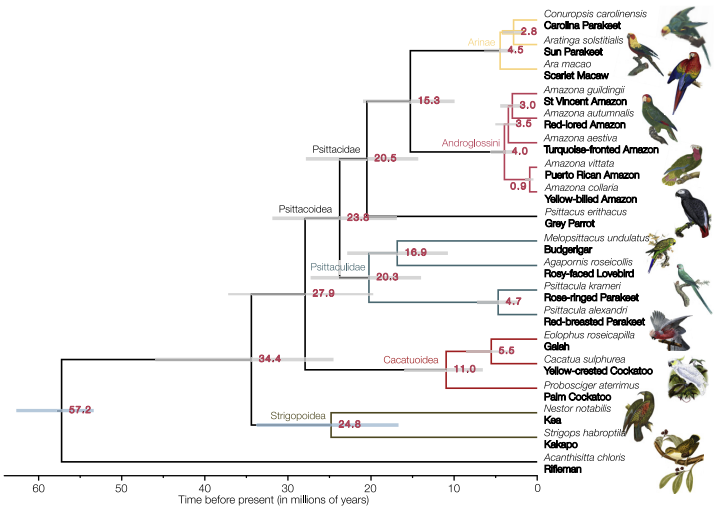


Figure 1. Phylogenetic Relationships of the Carolina Parakeet

Calibrated phylogeny built with BEAST2 based on 50 nuclear UCE loci from 18 species of parrots and a passerine (common names in bold). The analysis was constrained to a topology obtained from maximum likelihood analysis of 4,988 nuclear loci (9,864,148 bp), in which all nodes had 100 bootstrap support. Node ages were estimated using two fossil calibrations (highlighted in blue HPD intervals); gray bars indicate 95% HPD intervals of unconstrained nodes. Clade names follow a recent nomenclature revision [23]. See also Figure S2.

specimens [6, 7], we chose to generate a *de novo* assembly genome of the species' closest extant relative (*Aratinga solstitialis*; the sun parakeet), against which we could subsequently map and call full-genome variants from the sequenced Carolina parakeet (STAR Methods). Previous analyses of the phylogenetic relationship of the *Conuropsis* genus to extant parrots have been assessed based on morphology [8] and a short (876-nucleotide) fragment of the mtDNA genome retrieved from the toes of six specimens [9]. Both studies concluded that *Conuropsis* falls in a clade as a sister group to three *Aratinga* species [9]. Guided by this information, we generated a *de novo* reference genome of *Aratinga solstitialis* from a bird's breeder specimen. This genome was assembled and annotated using the B10K consortium pipelines [10] to render it consistent with previously published avian nuclear genomes during subsequent analyses. The genome was based on Illumina reads from three long-range Nextera libraries of different insert sizes and assembled with ALLPATH to a final N50 scaffold measure of 19.5 Mbp.

Following subsequent whole-genome shotgun sequencing of the Carolina parakeet DNA extract using the BGISeq-500 platform, which has been demonstrated to be effective for ancient DNA [11], we were able to map 209,887,920 unique reads from *C. carolinensis* against the 1,168,990,576 bp of *A. solstitialis* genome, covering 93% of the genomic positions, with a mean depth of 13.4× (STAR Methods). We also recovered the entire mtDNA genome to 150× depth of coverage. The reads exhibit characteristic ancient DNA deamination pattern at their ends [12], with a value close to 5% (Figure S1) that is consistent with our sample being just 100 years old [13]. We determined which positions were derived in *C. carolinensis* or *A. solstitialis* using the chicken *G. gallus* as outgroup. A total of 28,348 missense and 152 nonsense mutations were identified between *Conuropsis* and *Aratinga*. Of the former, 502 mutations were predicted to be deleterious mutations using SIFT software [14]. The Carolina parakeet transition/transversion (Ts/Tv) value is 2.309, the ratio of the non-synonymous to synonymous substitutions in all genes (dN/dS) is 0.48, and values similar to these have been reported in other bird genomes [15–17].

Sex determination of the specimens is difficult using morphological observations alone, as they have been described as being alike in coloration [4]. Females are the heterogametic sex (ZW) in birds; using genetic data, we were able to conclude that our specimen was a female because it showed about half the average depth coverage on the sex chromosomes (e.g., 13.4× genome-wide versus 7.11× at the *DMRT1* gene that is located in the Z chromosome).

Phylogenetic Relationships

In order to investigate the phylogenetic placement of *Conuropsis* within Psittaciformes and estimate its divergence time, we used 4,988 nuclear loci (ultraconserved elements [UCEs], comprising 9,864,148 bp) extracted from the genomes of *C. carolinensis*, 17 extant parrots, and the rifleman *Acanthisitta chloris* (Passeriformes) as an outgroup. Individual gene trees summarized into a coalescent species tree were congruent with concatenated datasets and supported by maximum local posterior probability in all but one node. Gene trees suffered from the few and unusually short loci of one of the samples (*Strigops habroptila*; 77% of loci missing and 95.81% gaps across the concatenated alignment), which resulted in a low supported relationship with Psittacoidea (support = 0.23). This sample was also problematic in coalescent-based analyses in the original study that generated the data [18–21]. Concatenated analyses of all loci and of 95% and 100% completeness were congruent and had maximum bootstrap support for *Strigops+Nestor* as the sister to all other parrots, as found before [18–21]. All other relationships and the placement of *Conuropsis* were entirely congruent between analyses suppress unambiguous and highly supported. *Conuropsis* was consistently placed as the sister group of *Aratinga*, which in turn is sister group to the macaw *Ara* within Arinae (macaws, conures, and allies) (Figure 1). We also used the complete coding region of the mitochondrial (mtDNA) genome sequence to investigate the placement of *Conuropsis* against a greater sampling within the Arini and found the same placement as with nuclear data (Figure S2). Molecular clock analysis employing two fossil calibrations [21] suggests that the *Aratinga-Conuropsis* split occurred around 2.8 mya (1.6–

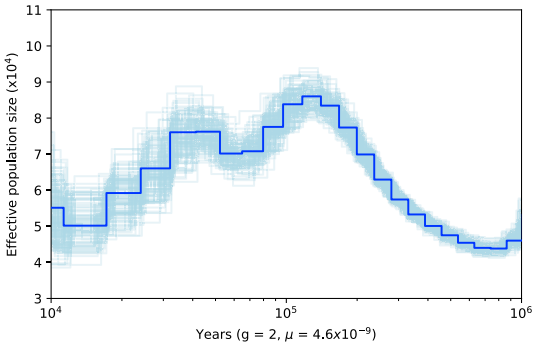


Figure 2. Demographic History of the Carolina Parakeet
Pairwise sequentially Markovian coalescent (PSMC) plot shows *Conuropsis carolinensis* population history. We performed 100 bootstrap repetitions. The PSMC plot shows demographic fluctuations of the parakeet population size starting with the beginning of the Last Glacial Period.
See also Figure S3.

4.4; 95% highest posterior density [HPD] interval) from nuclear genome data and around 3.8 mya (2.73–5.05; 95% HPD interval) from mtDNA data. Both dates roughly coincide with the widely recognized date of 3 mya for the final closure of the Panama Isthmus [22]. It seems plausible, therefore, that the dispersal to North America occurred after the North and South American landmasses were continuous.

Demographic History

We used the pairwise sequentially Markovian coalescent (PSMC) algorithm [24] to evaluate the past demographic evolution of *Conuropsis* and *Aratinga* species. We found that the Carolina parakeet population experienced an increase in effective population size (N_e) during the Middle Pleistocene, followed by demographic fluctuations that started during the Last Glacial Period (~110 kya) and a subsequent population decline that continued until recent times (Figure 2). In contrast, the PSMC of the endangered *Aratinga solstitialis* shows a stronger and continuous population decline and a longer period of lower effective population size than *Conuropsis* (Figure S3).

We then profiled both the overall heterozygosity across the genome and the distribution of long runs of homozygosity (RoHs) (Figure S4). The former is a measure of overall genetic diversity, whereas RoHs arise when identical chromosomal fragments are inherited from a recent common ancestor. Thus, significantly reduced heterozygosity is typical of populations that have been small and isolated for long periods, although elevated levels of RoH are usually observed in inbred populations [25]. Both may therefore be typical of endangered species. We found that *Conuropsis* had a heterozygosity slightly below the average across bird genomes [10] but clearly does not appear to be an outlier (Figure 3) (the low level of heterozygosity of our *Aratinga* specimen could be influenced by the fact that it was an individual bred in captivity). In addition, 188 total RoHs were detected for *Conuropsis* (9 of them >1 Mb), although for *Aratinga*, the number is much higher (611 total RoHs; 85 >1

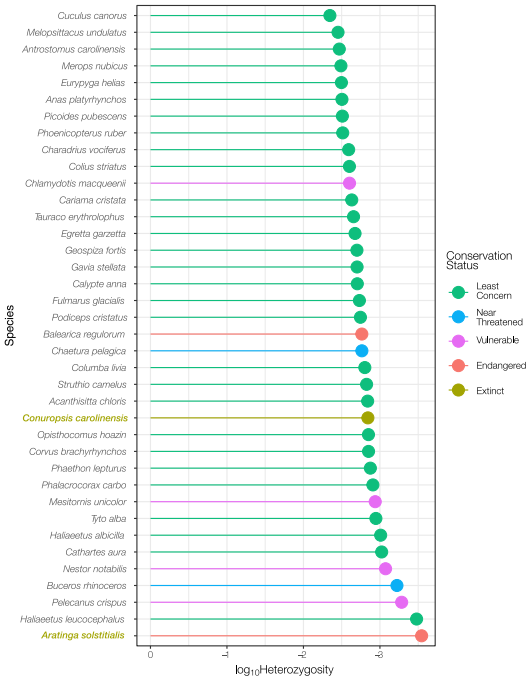


Figure 3. Genetic Diversity among Birds
Logarithm of the average genome heterozygosity for most published avian genomes. All species belong to different taxonomic orders except for *Conuropsis carolinensis* and *Aratinga solstitialis*, which are both Psittaciformes. Samples are colored by IUCN Conservation status.
See also Figure S4.

Mb) (Figure S4). Nevertheless, we report the presence of a single, long run of homozygosity of 7.15 Mb, which is suggestive of recent inbreeding in the ancestors of the Espinelves specimen. It cannot be discarded, in fact, that our specimen was originally bred in captivity. Alternatively, having only a single long RoH could be indicative of some selective sweep in that particular region.

Adaptation to Toxic Diet

We leveraged our data to explore the genomic basis of a curious behavior of this species, relating to its dietary habits. The Carolina parakeet consumed a variety of fruits, seeds, and to a lesser extent, buds and flowers, but most remarkably, it showed a predilection for cocklebur (*Xanthium strumarium*). This is unusual, as cocklebur contain significant levels of a diterpenoid glucoside, the carboxyatractyloside or CAT [5], a lethal toxin that inhibits mitochondrial energy production [26]. In a collection of 99 feeding observations of *Conuropsis*, the highest plant intake (n = 17) corresponded in fact to cocklebur [4]. CAT inhibits the function of four mitochondrial ATP transporters (ANT1, ANT2, ANT3, and ANT4; encoded by *SLC25A4*, *SLC25A5*, *SLC25A6*, and *SLC25A31*, respectively), which is lethal [26, 27]. We next explored these genes further by

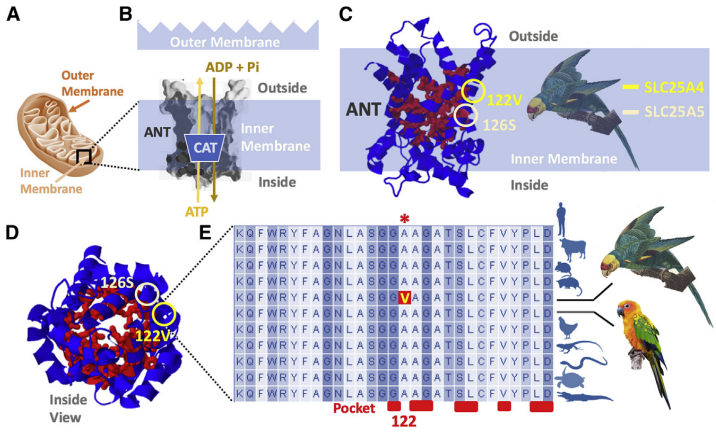


Figure 4. Adaptation to Toxic Diet

(A) Mitochondria representation of the outer and inner membrane (zoom in to B).

(B) Cartoon of the bovine ANT protein X-ray crystallographic structure (approximate location of the inhibition by CAT blocking the flux of ATP and ADP + Pi).

(C) Three-dimensional modeled structure of the SLC25A4 in *Conuropsis* with variable positions of the SLC25A4 in yellow and SLC25A5 in light orange (both modeled protein structures are quite similar—so only one was represented for simplification). The red region of the protein corresponds to the pocket.

(D) Inside view of the 3D modeled structure of the SLC25A4 in *Conuropsis* (positions labeled as in C).

(E) Sequential depiction of the amino acids around the position A122V of *SLC25A4* in *Conuropsis* (comparison bottom-down: human; cow; mouse; opossum; *C. carolinensis*; *A. solstitialis*; chicken; anoles; python; green turtle; and crocodile) and indication of the pocket sites (in red) in the protein segment represented.

comparing them against their orthologs in other available avian genomes, including one recently generated dataset representing 363 species spanning nearly all avian families [28]. First, we found that *SLC25A6* and *SLC25A31* genes are not present in that dataset, presumably due to the annotation pipeline used. We did, however, find that the *Conuropsis SLC25A4* and *SLC25A5* genes carry two non-synonymous amino acid changes with respect to the *Aratinga* annotation: A122V in *SLC25A4* (a C to T substitution covered by 14 DNA reads) and T126S in *SLC25A5* (an A to T substitution covered by 13 DNA reads). An additional variant in this gene, V227A, is shared with 24 other species from different orders. The two *SLC25A4* and *SLC25A5* substitutions found are conserved in a diverse dataset of vertebrates, in 37 previously published avian genomes [10], and in the newly available avian genome dataset (Figure 4). Among the large avian dataset, additional non-synonymous substitutions in the four codons preceding and opposite these two positions have only been found in one single species (*Pomatorhinus ruficollis*). The two sites are located in a helix of the protein and are flanking pocket sites, likely influencing the functionality of both proteins. Therefore, it is possible that these mutations conferred the species with a unique adaptive mechanism for dealing with the toxic CAT present in its diet, although we do not know whether they could be shared with other *Aratinga* species (besides *A. solstitialis*).

DISCUSSION

The extinct Carolina parakeet's genome could provide evidence for specific adaptive peculiarities of this species and also help answer questions related to the population history and extinction dynamics of this paradigmatic bird.

Taking advantage of having eighteen available parrot genomes, we have generated the first Psittaciformes genome-wide phylogeny, which showed that the divergence time for *Conuropsis* evolutionary lineage and its subsequent colonization of the North American subcontinent took place around 3 mya. Considering that the time to the most recent common ancestor of all Psittaciformes is at least ten times larger (about 34.4 mya), we can conclude that the evolutionary history underlying the Carolina-parakeet-specific adaptations is a rather recent process within this order of birds.

We also uncovered evidence of a past population history of expansions and contractions with low effective population size but no dramatic signals of widespread, recent inbreeding that interestingly were discernable in *Aratinga*. This suggests that, despite the perception of high parakeet abundance based on observations of large and noisy flocks, this species had experienced population contractions that were likely associated to past climatic oscillations. However, scarce evidence of inbreeding indicates that it suffered a very quick extinction process that left no traces in the genomes of the last specimens. In fact, the bird's final extinction was likely accelerated by collectors and trappers when it became evident that it was extremely rare [8].

We found evidence that the Carolina parakeet was adapted to the cocklebur's toxin, but we caution that this feeding behavior is not exclusive of *Conuropsis*; parrots in general ingest fruits and seeds known to be toxic to other vertebrates [29]. It has been proposed that some species could neutralize them by consuming clay from river banks, which would have a toxin-absorbing function [30], although other physiological detoxification mechanisms cannot be discarded. Nevertheless, it would be interesting in the future to functionally test the two variants detected in the *SLC25A4* and *SLC25A5* genes using avian cell lines.

Other potential factors for *Conuropsis* extinction, such as the exposure to poultry pathogens, will likely require a metagenomic screening of at least several parakeet specimens; however, preliminary results from our sample do not show a significant presence of bird viruses.

The potential adaptation to the CAT toxin and the lack of evidence for a dramatic long-term decline and widespread inbreeding suggests that no additional factors contributed to the extinction process. Therefore, the abrupt disappearance of the Carolina parakeet seems to be directly attributable to human pressures.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Conuropsis carolinensis specimen
  - Aratinga solstitialis specimen
- METHOD DETAILS
  - *Conuropsis* DNA extraction and sequencing
  - *Aratinga solstitialis* DNA extraction
  - *Aratinga solstitialis* sequencing and assembly
  - *Aratinga solstitialis* annotation
  - *Conuropsis carolinensis* mapping and variant calling
  - Sex determination
  - Ultraconserved Elements (UCE) phylogenetic tree
  - Mitochondrial phylogenetic tree
  - *Conuropsis* population history
  - *Conuropsis* average genome heterozygosity
  - *Conuropsis* Runs of Homozygosity (RoHs)
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.10.066.

ACKNOWLEDGMENTS

We are grateful to Arboretum Masjoan (Espinelves, Barcelona) for allowing us to sample the Carolina parakeet of their collection, Antoni Ballester for the radiographic exploration of the specimen, Jordi Grífols and Josep Maria Ramió for the *Aratinga* sample, and Wen Wang (Kunming Institute of Zoology, Yunnan, China) for his help and advice. The authors acknowledge the Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and the Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with sequencing of the *Aratinga* genome as well as *de novo* assembly of the data and access to the UPPMAX computational infrastructure. This work was supported by ERC Consolidator Grant 681396 ''Extinction Genomics'' to M.T.P.G. and M.S.-V., Obra Social ''La Caixa'' and Secretaria d'Universitats i Recerca (GRC2017-SGR880) to T.M.-B. and C.L.-F., BFU2017-86471-P and PGC2018-101927-B-I00 from MINECO/FEDER, UE to T.M.-B, and PGC2018-095931-B-100 from MINECO/FEDER, UE to C.L.-F. T.M.-B. is also supported by a U01 MH106874 grant and Howard Hughes International Early Career and CERCA Programme (Generalitat de Catalunya). A.A. is partially supported by the Strategic Funding UID/Multi/04423/2019 through national funds provided by the Portuguese Foundation for Science and Technology (FCT) and the European Regional Development Fund—program PT2020, by the European Structural and Investment Funds—Competitiveness and Internationalization Operational Program-COMPETE 2020, and by National Funds through the FCT under the project PTDC/CTA-AMB/31774/2017(POCI-01-0145-FEDER/031774/2017).

AUTHOR CONTRIBUTIONS

P.R., C.L.-F., and M.T.P.G. conceived the project. P.R., S.M., and C.L.-F. studied and sampled the specimens. M.S.-V., G.P., G.S., F.H., and X.C. performed experimental work. P.G., A.S., A.M., J.S., T.d.-D., Q.F., S.F., M.F.-B., B.P., A.N., Y.D., and L.D. undertook different computational analyses. M.P.T.G., C.L.-F., L.D., T.M.-B., and G.Z. coordinated different computational teams. A.A. performed the *in silico* protein analysis. J.S. and A.A. worked in visualization. P.G., M.P.T.G., and C.L.-F. wrote the manuscript with input from all coauthors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 5, 2019
Revised: October 3, 2019
Accepted: October 30, 2019
Published: December 12, 2019

REFERENCES

1. Elphick, C.S., Roberts, D.L., and Reed, J.M. (2010). Estimated dates of recent extinctions for North American and Hawaiian birds. Biol. Conserv. *143*, 617–624.
2. Burgio, K.R., Carlson, C.J., and Tingley, M.W. (2017). Lazarus ecology: Recovering the distribution and migratory patterns of the extinct Carolina parakeet. Ecol. Evol. *7*, 5467–5475.
3. Snyder, N.F.R. (2004). The Carolina Parakeet: Glimpses of a Vanished Bird (Princeton University Press).
4. Snyder, N.F.R., and Russel, K. (2002). Carolina parakeet (*Conuropsis carolinensis*), version 2.0. In The Birds of North America, A.F. Poole, and F.B. Gill, eds. (Cornell Lab of Ornithology).
5. Stuart, B.P., Cole, R.J., and Gosser, H.S. (1981). Cocklebur (*Xanthium strumarium*, L. var. *strumarium*) intoxication in swine: review and redefinition of the toxic principle. Vet. Pathol. *18*, 368–383.
6. Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., and Green, R.E. (2010). Computational challenges in the analysis of ancient DNA. Genome Biol. *11*, R47.
7. Shapiro, B., and Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. Science *343*, 1236573.
8. Forshaw, J.M. (1989). Parrots of the World Third Edition (Lansdowne Editions).
9. Kirchman, J.J., Schirtzinger, E.E., and Wright, T.F. (2012). Phylogenetic relationships of the extinct Carolina Parakeet (*Conuropsis carolinensis*) inferred from DNA sequence data. Auk *129*, 197–204.
10. Zhang, G., Li, B., Li, C., Gilbert, M.T., Jarvis, E.D., and Wang, J.; Avian Genome Consortium (2014). Comparative genomic data of the Avian Phylogenomics Project. Gigascience *3*, 26.
11. Mak, S.S.T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.S., Kuderna, L.F.K., Zhang, W., Fu, S., Vieira, F.G., et al. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. Gigascience *6*, 1–13.
12. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.
13. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE *7*, e34131.
14. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452-7.
15. Jonker, R.M., Zhang, Q., Van Hooft, P., Loonen, M.J.J.E., Van der Jeugd, H.P., Crooijmans, R.P.M.A., Groenen, M.A.M., Prins, H.H.T., and Kraus, R.H.S. (2012). The development of a genome wide SNP set for the Barnacle goose *Branta leucopsis*. PLoS ONE *7*, e38412.
16. Kraus, R.H.S., Kerstens, H.H.D., Van Hooft, P., Crooijmans, R.P.M.A., Van Der Poel, J.J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H.H.T., and

Groenen, M.A. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). BMC Genomics *12*, 150.

17. van Bers, N.E.M., van Oers, K., Kerstens, H.H.D., Dibbits, B.W., Crooijmans, R.P.M.A., Visser, M.E., and Groenen, M.A.M. (2010). Genome-wide SNP detection in the great tit *Parus major* using high throughput sequencing. Mol. Ecol. *19* (*Suppl 1*), 89–99.
18. de Kloet, R.S., and de Kloet, S.R. (2005). The evolution of the spindlin gene in birds: sequence analysis of an intron of the spindlin W and Z gene reveals four major divisions of the Psittaciformes. Mol. Phylogenet. Evol. *36*, 706–721.
19. Tokita, M., Kiyoshi, T., and Armstrong, K.N. (2007). Evolution of craniofacial novelty in parrots through developmental modularity and heterochrony. Evol. Dev. *9*, 590–601.
20. Wright, T.F., Schirtzinger, E.E., Matsumoto, T., Eberhard, J.R., Graves, G.R., Sanchez, J.J., Capelli, S., Müller, H., Scharpegge, J., Chambers, G.K., and Fleischer, R.C. (2008). A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the cretaceous. Mol. Biol. Evol. *25*, 2141–2156.
21. Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J., et al. (2019). Earth history and the passerine superradiation. Proc. Natl. Acad. Sci. USA *116*, 7916–7925.
22. O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., de Queiroz, A., Farris, D.W., Norris, R.D., et al. (2016). Formation of the Isthmus of Panama. Sci. Adv. *2*, e1600883.
23. Joseph, L., Toon, A., Schirtzinger, E.E., Wright, T.F., and Schodde, R. (2012). A revised nomenclature and classification for family-group taxa of parrots (Psittaciformes). Zootaxa *3205*, 26–40.
24. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
25. Prado-Martinez, J., Hernando-Herraez, I., Lorente-Galdos, B., Dabad, M., Ramirez, O., Baeza-Delgado, C., Morcillo-Suarez, C., Alkan, C., Hormozdiari, F., Raineri, E., et al. (2013). The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. BMC Genomics *14*, 363.
26. Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézéguet, V., Lauquin, G.J.M., and Brandolin, G. (2003). Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature *426*, 39–44.
27. Maldonado, E.N., DeHart, D.N., Patnaik, J., Klatt, S.C., Gooz, M.B., and Lemasters, J.J. (2016). ATP/ADP turnover and import of glycolytic ATP into mitochondria in cancer cells is independent of the adenine nucleotide translocator. J. Biol. Chem. *291*, 19642–19650.
28. China National GeneBank (2016). The Bird 10,000 Genomes (B10K) Project. https://b10k.genomics.cn/index.html.
29. Gilardi, J.D., and Toft, C.A. (2012). Parrots eat nutritious foods despite toxins. PLoS ONE *7*, e38293.
30. Gilardi, J.D., Duffey, S.S., Munn, C.A., and Tell, L.A. (1999). Biochemical functions of geophagy in parrots: detoxification of dietary toxins and cytoprotective effects. J. Chem. Ecol. *25*, 897–922.
31. van der Zwan, H., van der Westhuizen, F., Visser, C., and van der Sluis, R. (2018). Draft de novo genome sequence of *Agapornis roseicollis* for application in avian breeding. Anim. Biotechnol. *29*, 241–246.
32. Wirthlin, M., Lima, N.C.B., Guedes, R.L.M., Soares, A.E.R., Almeida, L.G.P., Cavaleiro, N.P., Loss de Morais, G., Chaves, A.V., Howard, J.T., Teixeira, M.M., et al. (2018). Parrot genomes and the evolution of heightened longevity and cognition. Curr. Biol. *28*, 4001–4008.e7.
33. Oleksyk, T.K., Pombert, J.F., Siu, D., Mazo-Vargas, A., Ramos, B., Guiblet, W., Afanador, Y., Ruiz-Rodriguez, C.T., Nickerson, M.L., Logue, D.M., et al. (2012). A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. Gigascience *1*, 14.
34. Seabury, C.M., Dowd, S.E., Seabury, P.M., Raudsepp, T., Brightsmith, D.J., Liboriussen, P., Halley, Y., Fisher, C.A., Owens, E., Viswanathan, G., and Tizard, I.R. (2013). A multi-platform draft de novo genome assembly and comparative analysis for the Scarlet Macaw (*Ara macao*). PLoS ONE *8*, e62415.
35. Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA *108*, 1513–1518.
36. Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res. *19*, 1117–1123.
37. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience *1*, 18.
38. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
39. She, R., Chu, J.S.C., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. *19*, 143–149.
40. Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. *14*, 988–995.
41. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
42. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.
43. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
44. Broad Institute (2019). Picard. http://broadinstitute.github.io/picard/.
45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
46. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
47. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
48. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.
49. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.
50. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.
51. GitHub (2015). Bam Util. https://github.com/statgen/bamUtil.
52. Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., and Ellegren, H. (2015). Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. Curr. Biol. *25*, 1375–1380.
53. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
54. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
55. R Development Core Team (2017). R: A language and environment for statistical computing (Version 3.4. 2) (R Foundation for Statistical Computing).

56. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

57. Schreiber, J. (2017). Pomegranate: fast and flexible probabilistic modeling in Python. J. Mach. Learn. Res. *18*, 1–6.

58. Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., and Glenn, T.C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. *61*, 717–726.

59. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

60. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

61. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

62. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*Suppl 6*), 153.

63. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650.

64. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904.

65. Luther, D. (1986). Die ausgestorbenen Vögel der Welt, Fourth Edition (Westarp-Wissenschaften).

66. McKinley, D., and Hardy, J.W. (1985). The Carolina Parakeet in Florida (Florida Ornithological Society).

67. Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

68. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

69. Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A., Willerslev, E., Krogh, A., and Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. BMC Genomics *13*, 178.

70. Bull, J.J. (1985). Sex determining mechanisms: an evolutionary perspective. Experientia *41*, 1285–1296.

71. Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., and Sinclair, A.H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature *461*, 267–271.

72. Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., and Glenn, T.C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. *61*, 717–726.

73. Miller, M.A., Pfeiffer, W, Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE) IEEE. https://www.phylo.org/sub_sections/portal/sc2010_paper.pdf.

74. Nabholz, B., Lanfear, R., and Fuchs, J. (2016). Body mass-corrected molecular rate for bird mitochondrial DNA. Mol. Ecol. *25*, 4438–4449.

75. Smeds, L., Qvarnström, A., and Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. Genome Res. *26*, 1211–1218.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Conuropsis carolinensis tissue samples	Arboretum Masjoan private collection at Espinelves (Girona, Spain)	N/A
Aratinga solstitialis	Blood sample from *Aratinga* breeder	N/A
Deposited Data		
Whole genome data of one historical *Conuropsis carolinensis* sample	This paper	PRJEB33130
Aratinga solstitialis assembly	This paper	PRJEB33135
Aratinga solstitialis annotation data	This paper	PRJEB33153
Taeniopygia guttata 3.2.4 annotated proteins from Ensembl	N/A	Release 85
Gallus gallus 5.0 annotated proteins from Ensembl	N/A	Release 85
Whole-genome sequencing data of *Acanthisitta chloris*	[10]	NCBI Project ID: PRJNA212877
Whole-genome sequencing data of *Agapornis roseicollis*	[31]	NCBI Project ID: PRJNA355979
Whole-genome sequencing data of *Amazona aestiva*	[32]	NCBI Project ID: PRJNA294082
Whole-genome sequencing data of *Amazona collaria*	N/A	NCBI Project ID: PRJNA490036
Whole-genome sequencing data of *Amazona guildingii*	[28]	Ultraconserved Elements (UCEs) extracted from whole genome data
Whole-genome sequencing data of *Amazona vittata*	[33]	NCBI Project ID: PRJNA171587
Whole-genome sequencing data of *Ara macao*	[34]	NCBI Project ID: PRJNA175470
Whole-genome sequencing data of *Eolophus roseicapilla*	[28]	Ultraconserved Elements (UCEs) extracted from whole genome data
Whole-genome sequencing data of *Melopsittacus undulatus*	N/A	NCBI Project ID: PRJNA72527
Whole-genome sequencing data of *Nestor notabilis*	[10]	NCBI Project ID: PRJNA212900
Whole-genome sequencing data of *Probosciger aterrimus*	[28]	Ultraconserved Elements (UCEs) extracted from whole genome data
Whole-genome sequencing data of *Psittacula krameri*	N/A	NCBI Project ID: PRJNA377329
Targeted capture ultraconserved elements for *Conuropsis carolinensis*	This paper	N/A
Targeted capture ultraconserved elements for *Aratinga solstitialis*	This paper	N/A
Targeted capture ultraconserved elements for *Amazona autumnalis*, *Cacatua sulphurea*, *Psittacula alexandri*, *Psittacus erithacus*, *Strigops habroptila*	[21]	https://doi.org/10.5061/dryad.2vd01grFile: passerines.unaligned.uce.contigs.tar.gz
4,988 aligned UCE loci	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1
4,988 gene trees from UCE loci and summary tree from ASTRAL-III	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1
Concatenated dataset of 4,988 aligned UCE loci and resulting IQTREE tree	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1
Concatenated dataset of 2,755 aligned UCE loci which are present in 95% of taxa and resulting IQTREE tree	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1
Concatenated dataset of 893 aligned UCE loci which are present in 100% of taxa and resulting IQTREE tree	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BEAST input files for 2 sets of 50 loci with calibration constraints and resulting trees	This paper	https://doi.org/10.17632/ p4wt7jc9 dw.1
Commands used for bioinformatic processing and phylogenetic analysis	N/A	https://doi.org/10.17632/ p4wt7jc9 dw.1
SOFTWARE AND ALGORITHMS		
ALLPATHS-LG	[35]	v.52485
AbySS	[36]	v.1.3.5
SOAPdenovo	[37]	v.2
TblastN	[38]	v.2.2.2
genBlastA	[39]	v.1.0.4
GeneWise	[40]	v.2.4.1
MUSCLE	[41]	v.3.8.31
cutadapt	[42]	v1.9.1
BWA	[43]	v.0.7.1
Picard	[44]	v2.0.1
Samtools	[45]	v.1.6
mapdamage2	[46]	2.7.12
GATK	[47]	3.7
ClustalOmega	[48]	1.2.1
BEAST	[49]	v.1.8.4 (v.2)
JmodelTest	[50]	v.2.1.10
BamUtil	[51]	v1.0.13
PSMC	[52]	v1.0
SNPeff	[53]	v4.3
PLINK	[54]	v1.9b
R	[55]	v3.5.1
SIFT	[14]	v6.2.1
Schmutzi	[56]	v1.5.4
Pomegranate	[57]	python v3
bcftools	N/A	v.1.9
PHYLUCE	[58]	v.1.6.6
IQTREE (incl. ModelFinder and UFBoot)	[59–61]	v.1.6.10
ASTRAL-III	[62]	v.5.6.3
BEAST2	[63]	v.2.6.0
Tracer	[64]	v.1.7

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Pere Gelabert (peregelabertx@gmail.com). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Conuropsis carolinensis specimen

At least 720 skins and 16 Carolina parakeet skeletons are preserved in museum collections globally [65, 66]. We sampled one such specimen with the intention of generating the first near-complete whole genome information of the species. The specimen is preserved in a private collection in the village of Espinelves (Girona, Spain), and was collected at the beginning of the 20th century by the Catalan naturalist Marià Masferrer i Rierola (1856-1923).

The Carolina parakeet is believed to have consisted of two subspecies: *Conuropsis carolinensis carolinensis*, that was principally distributed in Florida and along the Southeast coast of United States, and *Conuropsis carolinensis ludovicianus* that was distributed

across the central states of the country [2]. Both subspecies could be differentiated by morphological features such as coloration and body size. The wing, bill, and tail lengths of all adult *C. c. ludovicianus* significantly averaged more than in all adult *C. c. carolinensis* [4]. The wing and tarsal lengths (Figure S1), as well as the general color pattern of the Espinelves specimen, indicate it belongs to *C. c. carolinensis*.

Aratinga solstitialis specimen

A sample of blood was obtained *in vivo* from a female specimen from an official *Aratinga* breeder.

METHOD DETAILS

Conuropsis DNA extraction and sequencing

Two different samples of about 100 mg were obtained, one from the femur (leg bones were preserved inside the naturalized specimen) and one from toepads, with the help of a Dremel machine.

The two samples were digested using 2mL of extraction buffer containing 10mM Tris-HCL (pH 8), 10mM NaCl, 5mM $CaCl_2$, 2.5mM EDTA, 1% SDS, 1% Proteinase K and 40mM DTT. The solution was resuspended by vortexing and was incubated in a rotating plate overnight at 55°C. Digested samples were purified, and DNA was isolated following a combination of Phenol/Chloroform and column purification, as outlined below.

After incubation, digested samples were centrifuged for 3 min at 3000 × g and the supernatant was collected and mixed with 1X volume of Phenol. The sample solution was incubated on a rotor for 5min at RT. After, it was centrifuged for 3 min at 5000 × g and the upper aqueous layer was collected in new low-bind Eppendorf tube. The collected aqueous layer was mixed with 1X volume of Chloroform and the process was repeated. Again, the upper aqueous layer was collected in new tube and mixed with 10X volume of binding buffer prepared as previously described [67].The sample solution mixed with the binding buffer was poured into a binding apparatus constructed by fitting an ZymoV extension reservoir in a MinElute column and set inside a 50mL Falcon tube (as in [68]. Samples were centrifuged at 300 × g until all the liquid had passed through. The MinElute column was then separated from the reservoir and set into a new 2ml low-bind collection tube. The column was washed with 730μL of QIAGEN buffer PE, centrifuged at 3,300 × g, flow-though was discarded and the MinElute column was dry-spun for 1 min at 6000 × g in a bench-top centrifuge. DNA was eluted in a final volume of 50μL by adding twice 25μL of QIAGEN EB buffer and incubating for 5 min at 37°C between each elution. Samples were centrifuged at 6000 × g and extracted DNA was quantified using a Qubit instrument.

Following extraction, 15 μL of DNA extract was built into blunt-end libraries using both the NEBNext DNA Sample Prep Master Mix Set 2 (Cat No. E6070) and the BEST protocol using BGI adapters (as in [11]. Two libraries were built for each method. For the NEB protocol, the libraries were prepared according to manufacturer's instructions, only skipping the initial nebulization step.

The resulting DNA library for each method was then amplified and indexed in 4 PCR reactions of 50μL each with 16μL of DNA template on each, mixed with 25 μL 2X KAPA U+ Buffer, 1.5 μL of BGI amplification primer (10 μM)(sequences described in [11]. Thermocycling conditions were 3 min at 98°C, followed by 22 cycles of 20 s at 98°C, 30 s at 60°C and 30 s at 72°C, and a final 7 min elongation step at 72°C. The amplified library was purified with PB buffer on QIAGEN MinElute columns, before being eluted in 30 μL EB. Negative library controls, constructed with H_2O, were included, as well as libraries constructed on the negative extraction controls; both subsequently yielded no DNA sequences.

Amplified libraries were quantified using a TapeStation instrument (Agilent) and two sequencing pools were created by merging the amplified libraries for each method and sequenced on 2 lanes of a BGISEQ-500 sequencing instrument using 100SR chemistry. Libraries prepared from tibia bone powder exhibited longer DNA reads in comparison with the toe tissue ($x = 83$bp versus $x = 61$ bp, $p < 0.001$). NEB libraries yielded longer DNA reads than BEST libraries ($x = 84.86$ and $x = 63.45$ versus $x = 75.39$ and $x = 60.31$ in tibia and toe respectively, $p < 0.001$). NEB libraries were also the ones that yielded higher endogenous DNA content as well as lower clonality.

Aratinga solstitialis DNA extraction

Parallel genomic DNA extractions were performed on blood from a single *Aratinga solstitialis* female individual using the DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA) following manufacturer's instructions. The resulting DNA extracts were quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) with no modification of its standard protocol. To check for molecular integrity, each DNA extract was run on the 2200 TapeStation (Agilent Technologies, Santa Clara, CA) following manufacturer's protocol.

Aratinga solstitialis sequencing and assembly

Using the high molecular weight (HMW) DNA extracts, a short PCR-free insert library with 180 bp inserts was prepared using TruSeq DNA kit (Illumina, CA, USA) according to the manufacturer's instructions. In addition, three different mate-pair libraries were built using the Nextera protocol (Illumina, CA, USA). These comprised one 3 kb mate-pair library, one 5 kb mate-pair library, and one 20kb mate-pair library. All libraries were indexed to enable de-multiplexing after sequencing. The libraries were subsequently sequenced on the Illumina HiSeq X platform (using 2 × 150 bp reads), where the first lane was used for the 180 bp insert library. For the second lane, the three mate-pair libraries were pooled in equimolar ratios prior to sequencing.

In order to generate a *de novo* assembly, three different assemblers were used: ALLPATHS-LG v.52485, ABySS v.1.3.5 and SOAPdenovo. Out of the three assemblers, ALLPATHS-LG gave the best result, with an N50 scaffold measure of 19.5Mbp.

Aratinga solstitialis annotation

We used a homology-based method to annotate the protein-coding genes in the *Aratinga* genomes by using Ensembl gene sets (release 85) of chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*) and human (*Homo sapiens*), and genes derived from published avian transcriptomes. The protein sequences of the reference gene set compiled above were used as references for homology-based gene prediction.

We aligned reference protein sequences to the genome by TBLASTN with an E-value cut-off of 1e-5. We linked the hits into candidate gene loci with genBlastA and removed candidate loci with a homologous block length shorter than 30% of length of query protein. We extracted genomic sequences of candidate gene loci and 2,000bp upstream/downstream sequences as input for GeneWise to predict gene models in the genome. Then we translated the predicted coding regions into protein sequences, and ran MUSCLE for each pair of predicted protein and reference protein. We filtered out the predicted proteins with length of < 30aa or percent identity < 40%, as well as the pseudogenes (genes containing > 2 frameshifts or pre-mature stop codons) and retrogenes. The output of GeneWise could include redundant gene models, which overlap at the same genome regions. Hierarchical clustering was applied to the output of GeneWise to build a non-redundant gene set. Genes that overlapped in > 40% of their coding sequence were clustered and kept the sequence with the highest identity to the reference genes. We removed the highly duplicated genes (frequency of duplications > 10) in two conditions: 1) with a single exon; 2) with > 70% repeat sequences in coding region.

Conuropsis carolinensis mapping and variant calling

The ancient DNA reads were clipped using cutadapt; sequencing adapters were removed. Only reads longer than 30bp were kept. Filtered reads were mapped against the *A. solstitialis* assembly with Burrows-Wheeler Aligner (BWA) [69], setting no trimming, disabling seed, increasing stringency for edit distance, and allowing opening of 2 gaps. Duplicated reads were removed using Picard-tools MarkDuplicates. Mapped reads with mapping quality below 30 were removed using Samtools. The resulting reads were examined with mapdamage2 to assess the degradation rate of the data, which is a sign of authenticity. We detected the presence of typical aDNA-damaged bases at the end of reads. To avoid problems in the next steps, we trimmed 2 nt from each read end using BamUtil trimbam.

Genotypes were estimated using GATK UnifiedGenotyper. We removed calls with base quality below 30 (-mbq), and we set the rest of parameters as default. The average depth of coverage of the sample was 13.4X. To prevent variant calling errors in repetitive or complex regions, we used GATK SelectVariants to exclude the calls with depths of coverage below 10x and above 35X. Afterward we also used GATK SelectVariants and GATK FilterVariants to exclude from the call-set InDels and heterozygous calls in allele frequencies below 0.2 and above 0.8. We subsequently used the *A. solstitialis* assembly annotations to build a SNPeff database and used *Gallus* annotations to determine derived alleles.

Sex determination

The *Aratinga* genome -which we knew was a female- showed, as expected, half of coverage in the ZW chromosomes [70]. We plotted the depth of coverage distribution for each scaffold of the Carolina parakeet using Samtools and found identical coverage distribution. We subsequently searched for the *DMRT1* gene [71] to confirm the *Aratinga* Z chromosome scaffold.

Ultraconserved Elements (UCE) phylogenetic tree

For phylogenetic analysis, we targeted 5,060 UCE loci from 14 species with whole genome sequences (including the two new genomes presented here) and from 5 parrots that were included in a previous UCE capture study [21]. The Tetrapods-UCE-5Kv1 bait set [72] was applied to 18 parrots and the outgroup *Acanthisitta chloris*. A total of 4,988 UCE sequences were identified and extracted with the flanking 1000bp to both sides, aligned and trimmed using PHYLUCE (commands in 10.17632/p4wt7jc9dw.1). *Strigops habroptila* from the targeted capture study had significantly fewer and shorter loci than all other samples (1,648 loci, 269bp length on average compared to 757bp on average across samples from [21]) but we kept the sample because of its significance for fossil calibration.

We used coalescent and concatenation approaches to infer phylogenetic relationships. First, we constructed maximum likelihood gene trees for all 4988 alignments using IQTREE with 1000 ultrafast bootstrap replicates after determining the most appropriate nucleotide substitution model with ModelFinder. The resulting gene trees were summarized into a coalescence-based species tree using ASTRAL-III. Second, we concatenated all loci (9,864,148bp), the 2,755 loci that were present in 95% of all species (5,561,275bp) and the 893 loci that were present in 100% species (1,840,245bp) and analyzed them as above.

For calibration analyses, we drew two random samples of 50 loci that had all taxa and had the same substitution model (HKY+F+G4, the most common model across all loci). For both random samples, we executed two MCMC chains (100 million generations, sampled every 5,000 generations) in BEAST2 on the CIPRES Science Gateway [73]. Each analysis was performed on the topology from concatenation, employing a birth-death model, a relaxed clock model with lognormal distribution on the rate prior and HKY+F+G4 as the substitution model. The age of two nodes was constrained with lognormal distributions following the thorough published fossil justifications [21]. First, a lognormal prior was placed on the root of the tree (Passeriformes+Psittaciformes, *Eozygodactylus americanus*) with an offset of 51.81 Mya and a 97.5% quantile encompassing 66.5 Mya. Second, a lognormal prior was placed on

the MRCA of *Strigops+Nestor* (*Nelepsittacus minimus*) with an offset of 15.9 Mya and a 97.5% quantile at 66.5 Mya. Replicate runs were checked for convergence in Tracer, combined and annotated after a burning of 30% with LogCombiner and TreeAnnotator.

Mitochondrial phylogenetic tree

Trimmed DNA reads (209,887,920) were mapped against *A. solstitialis* mtDNA genome (JX441869). The mtDNA consensus sequence of *Conuropsis* was obtained by using schmutzi endoCaller and aligned with Clustal Omega to 11 other Arini mtDNA genomes and *Amazona ventralis* as outgroup. The obtained alignment of 13 sequences of 18,731bp in total length was dated using BEAST based on a fixed clock rate of 0.0042 substitutions/site/MY for all coding regions, which was previously determined for the brown-throated Parakeet *Eupsittula (*formerly *Aratinga) pertinax* [74]. The number of polymorphic sites of *Conuropsis* mtDNA genome in the alignment was 4,369. We used the GTR+I+G nucleotide substitution model selected by jModelTest with the Akaike Information Criterion.

Conuropsis population history

We used the Pairwise Sequentially Markovian Coalescent (PSMC) model to explore the demographic history of *C. carolinensis*. We obtained a fastq sequence of *C. carolinensis* for autosomal regions in scaffolds longer than 100Kbps. Only positions with a depth of coverage above 8X and below 50X were kept. Posteriorly a PSMC was built using the following parameters: -N25, -t15, -r5, -p "4+25*2+4+6." We used age of sexual maturity (1 year) [3], multiplied by a factor of two as a proxy for generation time, following the same approach as in a previous study of PSMC in 38 avian species [52] and a mutation rate of $2.3x10^{-9}$, estimated from bird pedigree information [75].

Conuropsis average genome heterozygosity

To identify regions of the *C. carolinensis* genome that shows signs of homozygosity we plotted the distribution of heterozygous positions across the genome sequence. We examined the scaffolds counting the number of heterozygous positions in windows of 50Kb with 10Kb of overlap. We define the average genome heterozygosity as the proportion of heterozygous sites genome-wide divided by the total number of callable bases. We kept only SNV sites applying the following filtering criteria: Read Depth > 10, Genotype Quality > 20, Allele Balance 0.2 < AB < 0.8 (hypergeometric distribution 0.95 CI [0.233-0.766]). All variable repeats, indels and multiallelic sites were removed. Non-variable sites were considered callable if their read depth was larger than 10. Additional heterozygosity values for other bird species were extracted from published avian genomes [10].

Conuropsis Runs of Homozygosity (RoHs)

RoHs were called based on the density of heterozygous sites in the genome using a Hidden Markov Model (HMM) for segmentation: First, the *Aratinga* reference genome was partitioned into 50Kb windows guided by the *Conuropsis* callability mask, namely, uncallable *Conuropsis* sites were omitted in the window tally. Heterozygosity values were calculated for each window as described above. Next, an HMM (python3 pomegrenate package) was fitted to the data. Emissions were modeled based on the empirical window heterozygosity distribution with a two/three component Gaussian Mixture Model (GMM). The first component of the GMM was reserved to extremely small heterozygosity values in order to capture the RoH variability, while the second component was allowed to vary freely. If necessary, a third mixture component was added to capture outliers. The transition probabilities were trained using the Baum-Welch algorithm.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details of experiments can be found at the STAR Methods. The phylogenetic tree in Figure 1 was performed with BEAST 2 (https://www.beast2.org). The pattern of post-mortem damage in Figure S1 was generated with mapdamage2 and the contamination estimates at the mtDNA was done with Schmutzi program. Adaptors from the DNA reads were removed with cutadapt. Genetic differences between *Conuropsis* and *Aratinga* were explored with SIFT software and the prediction of effects of some polymorphisms was done with SNPeff software.

DATA AND CODE AVAILABILITY

The accession numbers for the *Conuropsis* and *Aratinga* genomes reported in this paper are in the European Nucleotide Archive (ENA): PRJEB33130 and PRJEB33153, respectively.

Current Biology, Volume *30*

Supplemental Information

Evolutionary History, Genomic Adaptation to Toxic Diet, and Extinction of the Carolina Parakeet

Pere Gelabert, Marcela Sandoval-Velasco, Aitor Serres, Marc de Manuel, Pere Renom, Ashot Margaryan, Josefin Stiller, Toni de-Dios, Qi Fang, Shaohong Feng, Santi Mañosa, George Pacheco, Manuel Ferrando-Bernal, Guolin Shi, Fei Hao, Xianqing Chen, Bent Petersen, Remi-André Olsen, Arcadi Navarro, Yuan Deng, Love Dalén, Tomàs Marquès-Bonet, Guojie Zhang, Agostinho Antunes, M. Thomas P. Gilbert, and Carles Lalueza-Fox

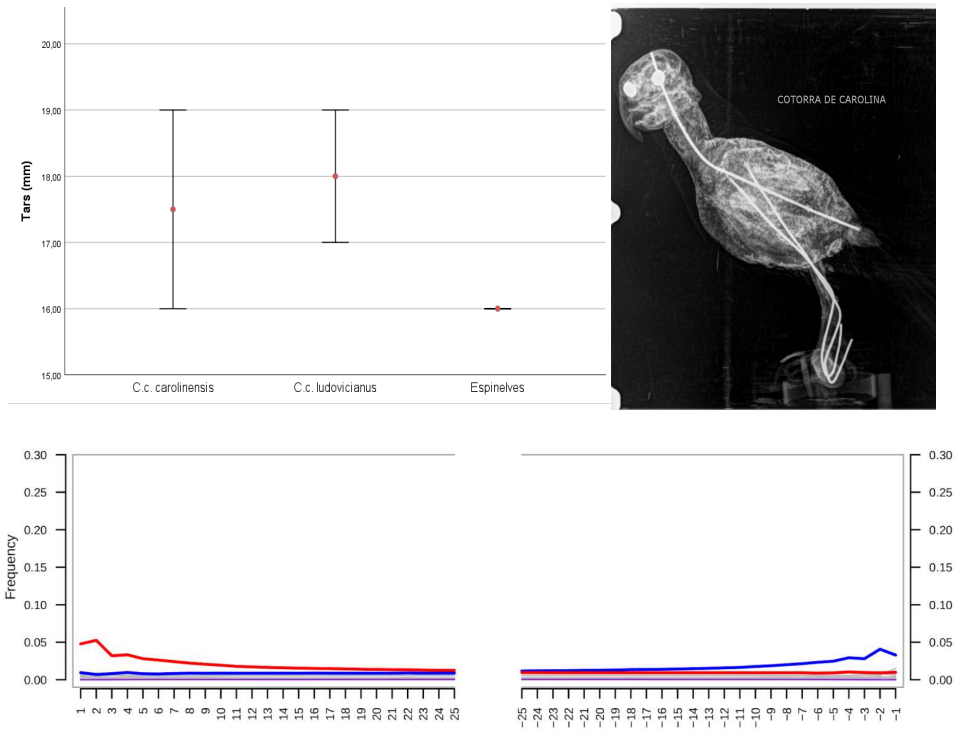


Figure S1: Morphological and post-mortem DNA damage analysis. Related to the STAR Methods. Above left: tarsal measurements of the Espinelves female studied compared with average and range measurements for adult females from the literature [S1]; the values indicate the specimen belongs to *Conuropsis carolinensis carolinensis* subspecies. Above right: X-ray image used to measure the tarsal bones. Below: nucleotide deamination pattern at the end of the Carolina parakeet reads, deriving from cytosine deamination in ancient sequences [S2]. The ratio (about 4-5% of the reads) is consistent with the age of the sample.

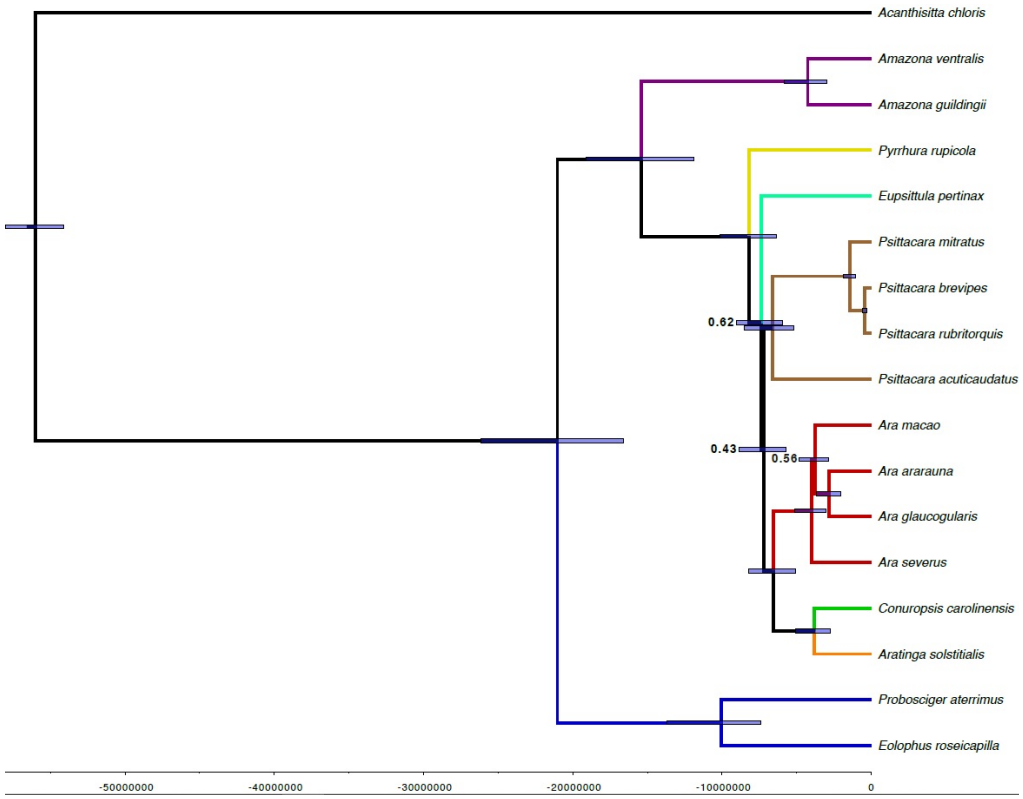


Figure S2: Calibrated phylogeny built with BEAST based on mitochondrial genomes of 13 species of Neotropical parrots. Related to Figure 1. A 0.0042 substitutions/site/MY molecular rate was used to calibrate the phylogeny. 95% HPD intervals are shown in the nodes in addition to bootstrap values below 1. The *Aratinga*-*Conuropsis* split occurred around 3.8 Mya.

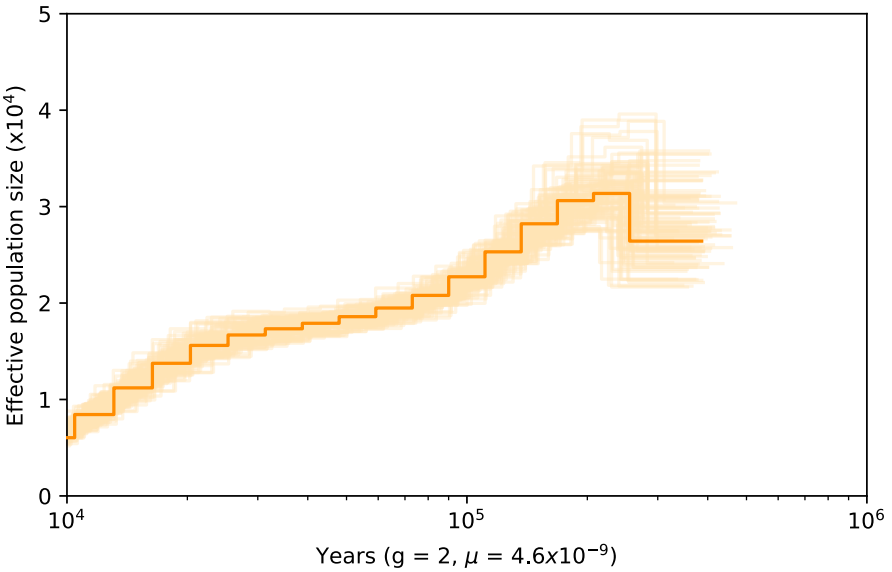


Figure S3: Pairwise Sequentially Markovian Coalescent (PSMC) plot of *Aratinga solstitialis* population history. Related to Figure 2. We performed 100 bootstrap repetitions.

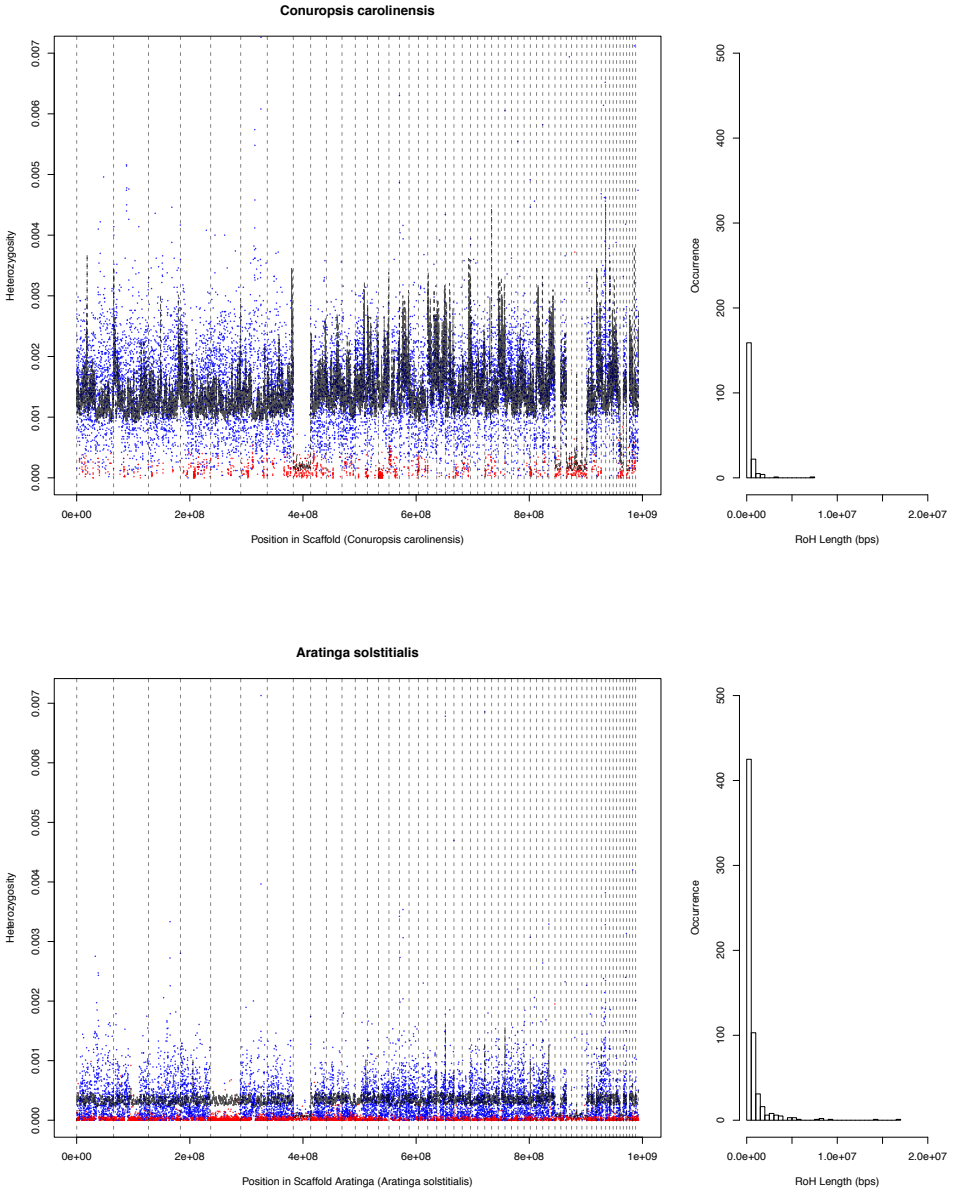


Figure S4: Heterozygosity across scaffolds (vertical doted lines represent contigs in the scaffold) and histogram of Runs of Homozygosity (RoH). Related to Figure 4. Length count for *Conuropsis carolinensis* (above) and *Aratinga solstitialis* (below), at the same scale. Only one long RoH of >7Mb was observed in the former species.

Supplemental References

S1. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P., Orlando, L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bionformatic 29 (13): 1682-1684.

S2. Snyder, N.F.R., Russel, K. 2002. Carolina Parakeet (*Conuropsis carolinensis*),version 2.0. In: Poole, A.F., Gill, F.B. (Eds.), The Birds of North America. Cornell Lab of Ornithology, Ithaca.

4.2. Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives

Pere Renom, Toni de-Dios, Sergi Civit, Laia Llovera, Alejandro Sánchez-Gracia, Esther Lizano, Juan Carlos Rando, Tomàs Marquès-Bonet, Gael J. Kergoat, Isaac Casanovas-Vilar and Carles Lalueza-Fox

Biology Letters. December 2021; 17:20210533
https://doi.org/10.1098/rsbl.2021.0533

BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Research

Cite this article: Renom P *et al*. 2021 Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives. *Biol. Lett.* **17**: 20210533.
https://doi.org/10.1098/rsbl.2021.0533

Received: 14 October 2021
Accepted: 26 November 2021

Subject Areas:
evolution, taxonomy and systematics

Keywords:
ancient DNA, body mass, gigantism, insular evolution, molecular phylogeny, rodents

Authors for correspondence:
Isaac Casanovas-Vilar
e-mail: isaac.casanovas@icp.cat
Carles Lalueza-Fox
e-mail: carles.lalueza.fox@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5744039.

THE ROYAL SOCIETY PUBLISHING

Evolutionary biology

Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives

Pere Renom[1], Toni de-Dios[1,2], Sergi Civit[3], Laia Llovera[1], Alejandro Sánchez-Gracia[3], Esther Lizano[1,4], Juan Carlos Rando[5], Tomàs Marquès-Bonet[1,4,6,7], Gael J. Kergoat[8], Isaac Casanovas-Vilar[4] and Carles Lalueza-Fox[1]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona 08003, Spain
[2]University of Tartu, Institute of Genomics, Estonian Biocentre, Tartu 51010, Estonia
[3]Departament of Genetics, Microbiology and Statistics-Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona 08028, Spain
[4]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona 08193, Spain
[5]Departamento de Biología Animal, Edafología y Geología, Universidad de La Laguna, La Laguna 38206, Tenerife, Spain
[6]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona 08010, Spain
[7]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona 08036, Spain
[8]CBGP, INRAE, IRD, CIRAD, Institut Agro, Univ Montpellier, Montpellier, France

Td-D, 0000-0001-9260-8846; SC, 0000-0002-2511-4579; LL, 0000-0002-5836-593X; AS-G, 0000-0003-4543-4577; EL, 0000-0003-3304-9807; JCR, 0000-0001-7397-869X; TM-B, 0000-0002-5597-3075; GJK, 0000-0002-8284-6215; IC-V, 0000-0001-7092-9622; CL-F, 0000-0002-1730-5914

Evolution of vertebrate endemics in oceanic islands follows a predictable pattern, known as the island rule, according to which gigantism arises in originally small-sized species and dwarfism in large ones. Species of extinct insular giant rodents are known from all over the world. In the Canary Islands, two examples of giant rats, †*Canariomys bravoi* and †*Canariomys tamarani*, endemic to Tenerife and Gran Canaria, respectively, disappeared soon after human settlement. The highly derived morphological features of these insular endemic rodents hamper the reconstruction of their evolutionary histories. We have retrieved partial nuclear and mitochondrial data from †*C. bravoi* and used this information to explore its evolutionary affinities. The resulting dated phylogeny confidently places †*C. bravoi* within the African grass rat clade (*Arvicanthis niloticus*). The estimated divergence time, 650 000 years ago (95% higher posterior densities: 373 000–944 000), points toward an island colonization during the Günz–Mindel interglacial stage. †*Canariomys bravoi* ancestors would have reached the island via passive rafting and then underwent a yearly increase of mean body mass calculated between 0.0015 g and 0.0023 g; this corresponds to fast evolutionary rates (in darwins (d), ranging from 7.09 d to 2.78 d) that are well above those observed for non-insular mammals.

1. Introduction

The Canary Islands are located northwest off the coast of Africa, with their nearest island (Fuerteventura) being only separated from the continent by about 100 km (figure 1*b*). Although this volcanic archipelago was never connected to the mainland by any land bridge or island chain, colonization of

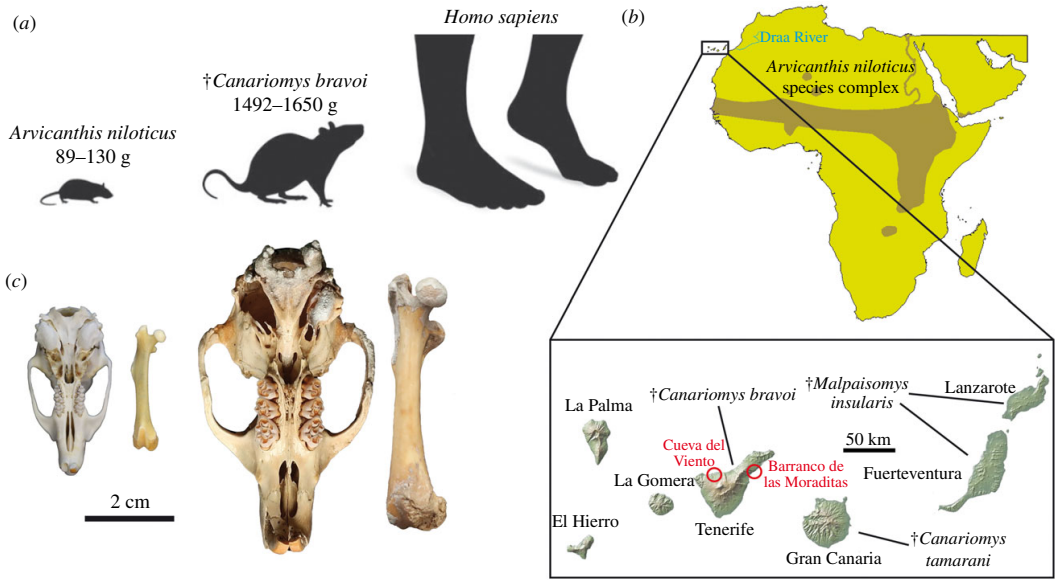


Figure 1. (*a*) Outline drawings and body mass for *A. niloticus* and †*C. bravoi*, along with a human reference. (*b*) Current distribution of *A. niloticus* in Africa (data from IUCN, 2008). The inset shows the Canary Islands and the distribution of extinct endemic rodents. Sampling sites are indicated for Tenerife. (*c*) Size differences between a typical *A. niloticus* representative (left) and †*C. bravoi* (right) as illustrated by their crania and femora (specimens are curated in IMEDEA and DZUL collections with numbers 12758 (*A. niloticus*) and 3199 (†*C. bravoi*), respectively). The latter has not been subjected to DNA analysis.

terrestrial organisms from the mainland was favoured by dominant oceanic currents. This archipelago offers a unique opportunity to study the colonization and diversification of multiple groups of organisms, such as birds, reptiles or small mammals [1]. Among the latter, there are three known striking examples of gigantism: the lava mouse of Fuerteventura and Lanzarote (†*Malpaisomys insularis*) and the extinct giant rats of Tenerife and Gran Canaria (†*Canariomys bravoi* and †*Canariomys tamarani*, respectively).

The Tenerife giant rat was described by Crusafont-Pairó & Petter [2] after the discovery of numerous specimens in Quaternary sites. Subsequent studies explored its diet, ecology, body mass and extinction causes, and also tentatively assessed its phylogenetic affinities based on dental traits (e.g. [3,4]). †*Canariomys bravoi* shows a set of traits characteristic of insular rodents, including gigantism, a robust skeleton and high-crowned teeth (figure 1*c*). It became extinct after the fourth century BCE, likely in relation to the arrival of Canarian indigenous people [4]. †*Canariomys tamarani* also became extinct soon after the arrival of the first settlers while †*M. insularis* survived until the beginning of the fourteenth century, when Europeans reached the archipelago [5]. Ancient mitochondrial DNA (mtDNA) from †*M. insularis*, showed close affinities to the extant genus *Mus* and pointed to a 6.9 Ma divergence date (genetic data were obtained by means of the traditional polymerase chain reaction method).

Hot and humid thermal conditions hamper the retrieval of ancient genetic data [6]. Without this information, it is difficult to unravel the affinities of highly modified extinct species such as †*C. bravoi* to their mainland smaller relatives. Here, we managed to retrieve partial nuclear and mtDNA data from two †*C. bravoi* specimens; we subsequently used this information to provide divergence age estimates and phylogenetic relationships for this lineage and determine the rate of increase in body size of this insular rodent.

2. Material and methods

(a) The samples

We performed DNA extraction from 12 mandibles: two from Barranco Moraditas and 10 more from Cueva del Viento (figure 1*b*). Specimens used for extractions were deposited in the Vertebrate collection (DZUL) of Departamento de Biología Animal, Edafología y Geología de la Universidad de La Laguna (Tenerife) with the following inventory numbers: CB-1 (DZUL 3200); CB-2 (DZUL 3201); CB-3 (DZUL 3202); CB-4 (DZUL 3203); CB-5 (DZUL 3204); CB-6 (DZUL 3205); CB-7 (DZUL 3206); CB-8 (DZUL 3207); CB-9 (DZUL 3208); CB-10 (DZUL 3209); CB-11 (DZUL 3210) and CB-12 (DZUL 3211).

Cueva del Viento site is a 17 km-long system of volcanic lava tubes formed 0.17–0.13 Ma [7] and situated in the north side of Tenerife at 700 m above sea level. The animals went into the cave through a small pit fall that acted like a trap. Bones were found in connection, showing the absence of transport after deposition. Previous calibrated radiocarbon ages of †*C. bravoi* samples from this site are between 17 300 and 2150 cal BP [4,8]. The samples from Barranco de las Moraditas were recovered from a small cave infilling in basaltic materials of Quaternary age at the east of Tenerife [7]. The median age reported for another †*C. bravoi* sample from this site is 2310 cal BP [4].

(b) DNA extraction, mitochondrial DNA capture and library preparation

All DNA extraction and initial library preparation steps were performed in a dedicated clean laboratory, physically isolated from the laboratory used for post-PCR analyses; no previous work on extinct or extant rodents was ever conducted in our laboratory. Strict protocols were followed to minimize the amount of human DNA in the ancient DNA laboratory, including wearing a full-body suit, sleeves, shoe covers, clean shoes, facemask, hair net and double gloving.

First, teeth samples were UV irradiated (245 nm) for 10 min and the outermost surface of the teeth was scraped off with a drill engraving cutter, followed by another UV irradiation in order to exclude the surface DNA contamination. Second, approximately 30 mg of tooth cementum was obtained by drilling at low speed (5000 r.p.m.) with a new engraving cutter.

DNA extraction from teeth powder was performed following the method of Dabney *et al*. [9]. The teeth powder samples, including an extraction blank, were added to 1 ml of extraction buffer (final concentrations: 0.45 M EDTA, 0.25 mg ml^{-1} proteinase K, pH 8.0), resuspended by vortexing and incubated at 37°C overnight on rotation. The remaining tooth powder was then pelleted by centrifugation in a bench-top centrifuge at maximum speed (16 100*g*). The supernatant was added to 10 ml of binding buffer (final concentrations: 5 M guanidine hydrochloride, 40% (vol/vol) isopropanol, 0.05% Tween-20 and 90 mM sodium acetate (pH 5.2)) and purified on a High Pure Extender column (Roche). DNA samples were eluted with 45 µl of EDTA TE buffer (pH 8.0). However, 10 samples failed to yield quantifiable DNA after extraction and only two from Cueva del Viento (CB-4 and CB-10) were further selected for library building.

A total of 35 µl of each DNA extract was used for library preparation in three sequential reactions: end-repair, adapter ligation, and nick fill-in; following the BEST protocol [10]. DNA extract from CB-4 was used for DNA-library preparation prior to Illumina sequencing; the resulting library was amplified by PCR with two uniquely barcoded primers and used for shotgun sequencing. Both libraries were purified with a 1× AMPure clean (Beckman Coulter) and eluted in 25 µl of low EDTA TE buffer (pH 8.0). Library size and concentration were determined with the Agilent DNA 7500 Kit on the 2100 BioAnalyzer instrument. The DNA libraries were sequenced using HiSeq400 of Illumina platform (Illumina, USA) in Macrogen, Inc. biotechnology company.

After library preparation, sample CB-10 was enriched for mtDNA sequences with the use of commercially biotinylated probes for mouse mtDNA (MYbaits). Prior to hybridization, the DNA library (approx. 500 ng) was brought to 7 µl using a Speedvak concentrator. Two consecutive hybridizations were conducted with the myBaits Capture Kit (Arbor Biosciences) according to the manufacturer's manual v. 4.01. The hybridization reaction was carried out at 65°C for 24 h in a final volume of 30 µl. Captured targets were recovered with Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen), followed by bead–bait binding and washing according to the manufacturer's recommendations. After the first round of enrichment, post captured amplification was performed using PCR primers IS5 and IS6. All of the captured material was concentrated to 7 µl and used for the second round of hybridization. The second hybridization was performed under the same conditions and the final captured pool was amplified with P5 and P7 indexed primers compatible for Illumina sequencing [11]. Sample CB-10 was radiocarbon dated to 2800 ± 30 years BP (Beta-598676).

(c) Phylogenetic analysis

All resulting DNA reads from samples CB-4 and CB-10 were mapped (edit distance equal to 0.0001) to *Mus musculus* (MN964117.1), *Rattus rattus* (NC_012374.1) and *Arvicanthis niloticus* (CM022273.1) mtDNA genomes. Mapped reads were subsequently blasted and only reads specific to rodent mitogenomes were retained. Additionally, DNA reads from CB-4 were mapped against the *A. niloticus* nuclear genome (NCBI: txid61156) with standard aDNA edit distance (0.01).

The authenticity of the generated sequences was confirmed with the observation of the typical post-mortem ancient DNA damage at the end of the DNA reads (electronic supplementary material, figure S2) and length fragmentation pattern (electronic supplementary material, figures S3 and S4). We further validated it with PMD tools [12], a statistical tool designed to isolate endogenous from contaminant DNA sequences; the PMD score distribution obtained is shifted toward positive values (electronic supplementary material, figures S5 and S6), which is characteristic of ancient samples. Several precautions were taken to account for the low coverage and the existence of single DNA reads: the ends of the reads were trimmed to eliminate potential post-mortem damage and C–T and G–A substitutions were only considered when they were shared with other rodent species.

We further inferred a time-calibrated Bayesian phylogenetic tree, relying on three secondary calibration points based on the results of Aghová *et al*. [13]. Bayesian inference (BI) was used to estimate the phylogenetic relationships and node ages using the BEAST v. 2.6.5 package [14]. The multiple sequence alignment was built using MAFFT software [15]. We first aligned arvicanthin sequences and then we added to this alignment the sequence of †*C. bravoi* using the –add option in MAFFT. Best fit model of nucleotide substitution for this alignment was elected with jModelTest [16] based on the Bayesian information criterion.

To infer the time-calibrated phylogeny, we used the Bayesian uncorrelated lognormal relaxed clock (ULRC) model implemented in BEAST v. 2 [17]. We used a coalescent model tree prior with a constant population size [18]. We set three palaeontological calibration points at different nodes of the tree: a *Mus*/*Rattus* divergence between 11.6 and 13.8 Ma [13,19], an *Arvicanthis*/*Lemniscomys* divergence between 6.1 and 8.5 Ma [20,21] and a basal node of the Arvicanthini of 8.5–9.2 Ma [13,20].

The Tree Model was set to a birth–death speciation process [22] to account more accurately for extinct and missing lineages. We used BEAST 2 to run 180 million generations of the model to sample trees from the posterior distribution (each 5000 generations). After examining effective sample sizes (ESS) and the traces for posterior, prior and likelihood with the tool Tracer [23], we discarded the first 20% of trees from the analysis (burn-in proportion). Finally, we generated the tree with median age estimates and 95% higher posterior densities (95% HPD bars) using tree annotator tool (distributed with the BEAST v. 2 package). Convergence of runs was assessed by examining the ESS of parameters, using the recommended threshold of 200 [17].

(d) Evolutionary rates

We calculated evolutionary rates for the body mass (in grams) increase from *A. niloticus* to †*C. bravoi*. Mean body mass for *A. niloticus* (sexes combined) is taken from Monadjem *et al*. [24], while estimated mean body mass for †*C. bravoi* is taken from Moncunill-Solé *et al*. [25] and is based on multiple regressions considering dental, cranial and postcranial measurements.

Evolutionary rates are calculated using the simple classic equation by Haldane [26]:

$$r = \left(\ln(x1) - \frac{\ln(x2)}{\Delta t}\right), \tag{2.1}$$

where, r is the rate of change (in darwins, d); $x1$ and $x2$ are the initial and final value of the analysed variable, respectively; and Δt is the amount of time elapsed. The calculations are carried for the minimum and maximum divergence dates between *A. niloticus* and †*C. bravoi*. The age of the oldest †*C. bravoi* fossils (17 300 cal BP [4,8]) is taken as the endpoint of the size increase trend. Calculated evolutionary rates are compared to those derived for other mammals (e.g. [27,28]) as well as to those for †*M. insularis*, which is included because it represents another case of murid gigantism in the same archipelago. Body mass for †*M. insularis* is estimated from published cranial and postcranial measurements [29], applying described equations [25]. Estimated mean body mass for †*M. insularis* is compared to

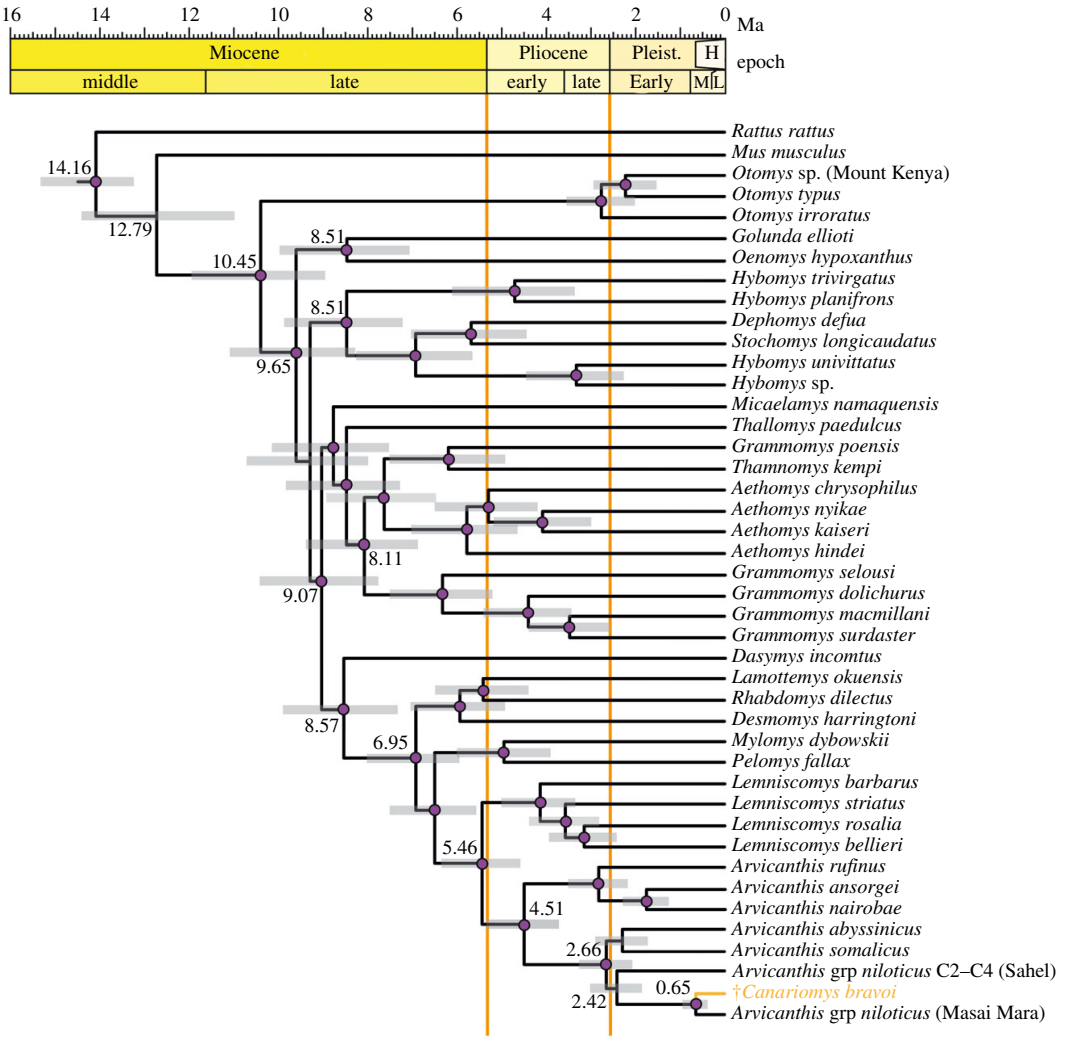


Figure 2. Molecular phylogenetic tree of the Muridae with the mitochondrial DNA data. Median ages are indicated at the nodes while error bars (grey shading) at nodes correspond to the 95% highest posterior density (HPD) intervals of age estimates. Purple circles at nodes indicate posterior probabilities greater than 95%. NCBI codes for each mitogenome can be found at electronic supplementary material, table S1.

Table 1. Mapping statistics of †*C. bravoi* mitochondrial and nuclear DNA reads. CB-4 corresponds to shotgun sequencing and CB-10 to mtDNA capture and sequencing. mtDNA reads were mapped following the procedure described in the Methods section; nuclear reads from CB-4 were mapped against the *Arvicanthis niloticus* nuclear genome (NCBI:txid61156).

specimen (mtDNA)	sequenced read pairs	mapped reads	unique q20 reads	BLAST reads	mapped bases	reference recovered
CB-4	119 102 486	281	104	41	1826	10.78%
CB-10	3 746 833	10 745	55	45	3387	7.21%
specimen (nuclear DNA)	**sequenced read pairs**	**mapped reads**	**unique reads**	**unique q20 reads**	**mapped bases**	**average coverage**
CB-4	119 102 486	111 084	40 434	35 593	1 616 176	0.0006X

that of its closest relatives, *Mus* (*Coelomys*) *pahari* and *Mus* (*Coelomys*) *crociduroides* from Southeastern Asia [30] with calculations formulated for minimum and maximum divergence dates between †*M. insularis* and *Mus* (*Coelomys*) spp. The oldest †*Malpaisomys* fossils have been dated to 32 000 cal BP [30], which is taken as the age for the end of the size increase trend.

3. Results and discussion

A total of 1 616 176 nucleotides mapped to *Arvicanthis* nuclear genome, representing only a 0.0006× depth of coverage but proving that genomic retrieval of †*C. bravoi* is a possible—albeit a challenging—task.

A total of 2627 mapped mtDNA nucleotides were aligned to a large dataset of rodent mitogenomes (table S1) and subsequently used for the phylogenetic analysis (table 1). The resulting dated phylogeny confidently places †*C. bravoi* within the *Arvicanthis* genus, in the *A. niloticus* species complex [30] (figure 2); it is closely related to a specimen from Masai Mara (Kenya) and is more distinctly related to a specimen belonging to the C2–C4 lineage that is distributed across the Sahel (both *Arvicanthis* specimens correspond to cryptic species in the *A. niloticus* species complex [31]). This unexpected placement parallels that of *M. insularis*, which was found to cluster between members of the genus *Mus* [32], despite its uncommon dental and skeletal traits. †*Canariomys* and †*Malpaisomys* belong to different murine tribes (Murini and Arvicanthini), thus indicating that their origin must be traced to different ancestors and likely different colonization events (the former being much more recent).

The expansion of *Arvicanthis* species through North Africa was heavily influenced by Pleistocene climatic fluctuations [33,34]. When environmental conditions changed and the Sahara region dried up, different *Arvicanthis* populations were cornered in areas of grassland and savannah habitats far apart each other. The current patchy distribution of members of the *A. niloticus* species complex includes the Nile River up to the great African lakes, the whole strip of the Sahel and some isolated populations surviving in Pleistocene refuges such as the Hoggar mountains (southern Algeria) (figure 1*b*). A previous molecular study indicated that the *A. niloticus* species complex likely originated in eastern Africa as early as 2 Ma and differentiated in genetically distinct lineages from east to west as a result of Pleistocene climatic cycles [33].

The divergence time between †*C. bravoi* and its closest *Arvicanthis* relative is estimated at 650 000 years ago (95% HPD intervals: 373 000–944 000 years ago) (figure 2). This interval basically includes the Günz and Mindel glaciations, as well as the Günz–Mindel interglacial. This interglacial appears as the most probable period for the colonization of Tenerife by †*Canariomys* ancestors. Interglacial periods altered the monsoon regime and increased rainfall across Africa. Satellite images of the Draa ancient river bed, which drains the anti-Atlas and flows right in front of the Canary Islands, suggests it must have been a river with a flow of more than 400 $m^3 s^{-1}$ that period and probably dragged logs and masses of vegetation to the sea on which some †*Canariomys* ancestors might have drifted away.

This relatively recent split date points to a rapid evolutionary process associated with gigantism. Body mass of *A. niloticus* ranges between 89 and 130 g (mean 114 g) [24], whereas the estimated weight of †*C. bravoi* was 1492–1650 g (mean 1571 g) [25], which is almost 14 times heavier (we are assuming that the body size of the *A. niloticus* ancestor that originated the *Canariomys* lineage had the same weight than the extant *A. niloticus*). This size increase is comparable to that of the extinct Sicilian giant dormouse †*Leithia melitensis* (13.5 heavier than its putative most closely related species, the garden dormouse *Eliomys quercinus* [11]) and generally well above values inferred for Pleistocene giant murines from several Mediterranean islands (usually 2–3 times heavier than their mainland ancestors [34,35]). It is also far greater than that calculated for †*Malpaisomys*, which is almost four times heavier (90 g) than its mainland relatives (24 g) [30]. It is difficult to assess when the process of size increase was achieved because the age of the oldest †*C. bravoi* fossils is not well constrained and their earliest date for the start of gigantism is necessarily after the splitting from a mainland *Arvicanthis* lineage. According to our data, the resulting temporal range would suppose a yearly increase of mean body mass of just between 0.0015 g and 0.0023 g (considering maximum and minimum splitting dates, respectively). This corresponds to evolutionary rates (in darwins (d)), ranging from 7.09 d to 2.78 d that are well above those observed for non-insular mammals (usually less than 1 d). By contrast, calculated evolutionary rates for †*Malpaisomys* are in the range of those of mainland mammals (0.22–0.16 d). We must remark that our †*Canariomys* evolutionary rate estimates represents a minimal estimate and assumes a constant rate of change. However, large body size may have not been achieved at a constant rate but soon after colonization and stabilized from then on. This would imply even a faster initial growing-size rate for †*Canariomys*.

However, only the retrieval of additional †*Canariomys* nuclear genome-wide data would further clarify its evolutionary history and would also allow the identification of the genomic regions under selection that might be responsible for the conspicuous physical differences observed between this extinct lineage and its living relatives.

Data accessibility. Genetic data generated are deposited in the Dryad database (project https://doi.org/10.5061/dryad.9s4mw6mhn).

Authors' contributions. P.R. and C.L.-F. conceived the project; J.C.R. collected and studied the skeletal samples; L.L. and E.L. performed experimental work; T.d.-D., P.R., S.C. and A.S.-G. performed computational work; G.J.K., T.M.-B. and I.C.-V. helped in the interpretation of the results; T.M.-B. and C.L.-F. coordinated experimental and computational teams; P.R., C.L.-F., G.J.K. and I.C.-V. wrote the manuscript with inputs from all co-authors. All authors agree to be held accountable for the content therein and approve the final version of the manuscript.

Competing interests. The authors declare no conflict of interest.

Funding. C.L.-F. is supported by a PGC2018-0955931-B-100 grant (MCIU/AEI/FEDER, UE) of Spain; T.M.-B. is supported by funding from the European Research Council (ERC) (grant agreement no. 864203), BFU2017-86471-P (MINECO/FEDER, UE), 'Unidad de Excelencia María de Maeztu', funded by the AEI (CEX2018-000792-M), Howard Hughes International Early Career and Generalitat de Catalunya (CERCA Programme and 2017 SGR 880); I.C.-V. is supported by grant nos. I+D+i PID2020-117289GBI00 funded by MCIN/AEI/10.13039/501100011033/, the MINECO (RYC-2013-12470) and the Generalitat de Catalunya (CERCA Programme and 2017 SGR 116).

References

1. Juan C, Emerson BC, Oromí P, Hewitt GM. 2000 Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends Ecol. Evol.* **5**, 104–109. (doi:10.1016/s0169-5347(99)01776-0)
2. Crusafont-Pairo M, Petter F. 1964 Un muriné géant fossile des îles Canaries: *Canariomys bravoi* gen. nov., sp. nov. (Rongeurs, Muridés). *Mammalia* **28**, 608–611. (doi:10.1515/mamm.1964.28.4.607)
3. Michaux J, Hautier L, Hutterer R, Lebrun R, Guy F, García-Talavera F. 2012 Body shape and life style of the

extinct rodent *Canariomys bravoi* (Mammalia, Murinae) from Tenerife, Canary Islands (Spain). *C. R. Palevol.* **11**, 485–494. (doi:10.1016/j.crpv.2012.06.004)

4. Rando JC, Alcover JA, Galván B, Navarro JF. 2014 Reappraisal of the extinction of *Canariomys bravoi*, the giant rat from Tenerife (Canary Islands). *Quat. Sci. Rev.* **94**, 22–27. (doi:10.1016/J.QUASCIREV.2014.04.013)
5. Rando JC, Alcover JA, Navarro JF, García-Talavera F, Hutterer R, Michaux J. 2008 Chronology and causes of the extinction of the lava mouse, *Malpaisomys insularis* (Rodentia: Muridae) from the Canary Islands. *Quat. Res.* **70**, 141–148. (doi:10.1016/j.yqres.2008.04.012)
6. García-Garcerà M, Gigli E, Sanchez-Quinto F, Ramirez O, Calafell F, Civit S, Lalueza-Fox C. 2011 Fragmentation of contaminant and endogenous DNA in ancient samples determined by shotgun sequencing; prospects for human palaeogenomics. *PLoS ONE* **6**, e24161. (doi:10.1371/journal.pone.0024161)
7. Ancochea E, Fuster JM, Ibarrola E, Cendrero A, Coello J, Hernán F, Cantagrel JM, Jamond C. 1990 Volcanic evolution of the island of Tenerife (Canary Islands) in the light of new K-Ar data. *J. Volcanol. Geotherm. Res.* **44**, 231–249. (doi:10.1016/0377-0273(90)90019-C)
8. Crowley BE, Yanes Y, Mosher SG, Rando JC. 2019 Revisiting the foraging ecology and extinction history of two endemic vertebrates from Tenerife, Canary Islands. *Quaternary* **2**, 10. (doi:10.3390/quat2010010)
9. Dabney J *et al.* 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15 758–15 763. (doi:10.1073/pnas.1314445110)
10. Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2018 Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419. (doi:10.1111/2041-210X.12871)
11. Kircher M, Sawyer S, Meyer M. 2012 Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3. (doi:10.1093/nar/gkr771)
12. Skoglund P, Northoff BH, Shunkov MV, Derenvianko AP, Pääbo S, Krause J, Jakobsson M. 2014 Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234. (doi:10.1073/pnas.1318934111)
13. Aghová T, Kimura Y, Bryja J, Dobigny G, Granjon L, Kergoat GJ. 2018 Fossils know it best: using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). *Mol. Phylogenetics Evol.* **128**, 98–111. (doi:10.1016/j.ympev.2018.07.017)
14. Bouckaert R *et al.* 2019 BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650. (doi:10.1371/journal.pcbi.1006650)
15. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/Mst010)
16. Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
17. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)
18. Heled J, Drummond AJ. 2009 Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580. (doi:10.1093/molbev/msp274)
19. Jacobs LL, Flynn LJ. 2005 Of mice again: the Siwalik rodent record, murine distribution, and molecular clocks. In *Interpreting the past: essays on human, primate, and mammal evolution in honor of David Pilbeam* (eds DE Lieberman, RJ Smith, J Kelley), pp. 63–80. Leiden, The Netherlands: Brill Academic Publishers, Inc.
20. Geraads D. 2001 Rongeurs du Miocène supérieur de Ch'orora, Ethiopie: Murinae, Dendromurinae et conclusions. *Palaeovertebrata* **30**, 89–109.
21. Manthi FK. 2007 A preliminary review of the rodent fauna from Lemudong'o, southwestern Kenya, and its implication to the late Miocene paleoenvironments. *Kirtlandia* **56**, 92–105.
22. Gernhard T. 2008 The conditioned reconstructed process. *J. Theor. Biol.* **253**, 769–778. (doi:10.1016/j.jtbi.2008.04.005)
23. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904. (doi:10.1093/sysbio/syy032)
24. Monadjem A, Taylor PJ, Denys C, Cotterill FPD. 2015 *Rodents of Sub-Saharan Africa. A biogeographic and taxonomic synthesis*. Berlin, Germany: De Gruyter.
25. Moncunill-Solé B, Jordana X, Marín-Moratalla N, Moyà-Solà S, Köhler M. 2014 How large are the extinct giant insular rodents? New body mass estimations from teeth and bones. *Integr. Zool.* **9**, 197–212. (doi:10.1111/1749-4877.12063)
26. Haldane JBS. 1949 Suggestions as to quantitative measurement of rates of evolution. *Evolution* **3**, 51–56. (doi:10.1111/j.1558-5646.1949.tb00004.x)
27. Gingerich PD. 1983 Rates of evolution: effects of time and temporal scaling. *Science* **222**, 159. (doi:10.1126/science.222.4620.159-a)
28. MacFadden BJ. 1986 Fossil horses from '*Eohippus*' (*Hyracotherium*) to *Equus*: scaling, Cope's Law, and the evolution of body size. *Paleobiology* **12**, 355–369. (doi:10.1017/S0094837300003109)
29. Hutterer R, López-Martínez N, Michaux J. 1998 A new rodent from quaternary deposits of the Canary Islands and its relationships with Neogene and recent murids of Europe and Africa. *Palaeovertebrata* **18**, 241–262.
30. Denys C, Taylor PJ, Burgin CJ, Aplin KP, Fabre P-H, Haslauer R, Woinarski JCZ, Breed WG, Menzies JI. 2017 . Species accounts of Muridae. In *Handbook of the mammals of the world. 7. Rodents II, handbook of mammals of the world* (eds DE Wilson, TE Lacher Jr, RA Mittermeier), pp. 598–884. Barcelona, Spain: Lynx Edicions.
31. Bryja J, Colangelo P, Lavrenchenko LA, Meheretu Y, Šumbera R, Bryjová A, Verheyen E, Leirs H, Castiglia R. 2019 Diversity and evolution of African rats (Muridae: *Arvicanthis*)—from radiation in East Africa to repeated colonization of northwestern and southeastern savannas. *J. Zool. Syst. Evol. Res.* **57**, 970–988. (doi:10.1111/jzs.12290)
32. Pagès M, Chevret P, Gros-Balthazard M, Hughes S, Alcover JA, Hutterer R, Rando JC, Michaux J, Hänni C. 2012 Paleogenetic analyses reveal unsuspected phylogenetic affinities between mice and the extinct *Malpaisomys insularis*, and endemic rodent of the Canaries. *PLoS ONE* **7**, e311123. (doi:10.1371/journal.pone.0031123)
33. Dobigny G, Tatard C, Gauthier P, Ba K, Duplantier JM, Granjon L, Kergoat GJ. 2013 Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and Sub-Saharan open habitats Pleistocene history. *PLoS ONE* **8**, e77815. (doi:10.1371/journal.pone.0077815)
34. Lomolino MV, van der Geer AA, Lyras GA, Palombo MR, Sax DF, Rozzi R. 2013 Of mice and mammoths: generality and antiquity of the island rule. *J. Biogeogr.* **40**, 1427–1439. (doi:10.1111/jbi.12096)
35. Gingerich PD. 2001 Rates of evolution on the time scale of the evolutionary process. In *Microevolution rate, pattern, process* (eds AP Hendry, MT Kinnison), pp. 127–144, Dordrecht, The Netherlands: Springer.

Supplemental Information

Genetic data from the extinct giant rat from Tenerife (Canary Islands) points to a recent divergence from mainland relatives

Pere Renom, Toni de-Dios, Sergi Civit, Laia Llovera, Alejandro Sánchez-Gracia, Esther Lizano, Juan Carlos Rando, Tomàs Marquès-Bonet, Gael J. Kergoat, Isaac Casanovas-Vilar, Carles Lalueza-Fox

	Sequence (5' -> 3')
P7 indexing primer:	CAAGCAGAAGACGGCATACGAGATNNNNNNNGTGACTGGAGTTCAGACGTGT
P5 indexing primer:	AATGATACGGCGACCACCGAGATCTACACNNNNNNNACACTCTTTCCCTACACGACGCTCTT

	Index sequence (5' -> 3')	
	i7	i5
Sample CB-4	AAGACGA	GCTAATC
Sample CB-10	TTGAGCG	CAACTAT

Table S1. Adaptors sequences and samples indexes used in the study.

Rodent	NCBI CODE	Rodent	NCBI CODE
Rattus rattus	NC_012374.1	*Grammomys selousi*	NC_053798.1
Mus musculus	MN964117.1	*Grammomys selousi*	MN807582.1
Otomys sp.	MN807604.1	*Grammomys dolichurus*	MN807581.1
Otomys typus	NC_053811.1	*Grammomys dolichurus*	NC_053797.1
Otomys typus	MN807603.1	*Grammomys macmillani*	MN807580.1
Otomys irroratus	MK166028.1	*Grammomys surdaster*	NC_053796.1
Otomys irroratus	NC_049116.1	*Grammomys surdaster*	MN807579.1
Otomys irroratus	MN964114.1	*Dasymys incomtus*	MN807606.1
Golunda ellioti	NC_053815.1	*Dasymys incomtus*	NC_053812.1
Golunda ellioti	MN807614.1	*Lamottemys okuensis*	NC_053804.1
Oenomys hypoxanthus	MN807605.1	*Rhabdomys dilectus*	MN807596.1
Hybomys trivirgatus	MN807602.1	*Desmomys harringtoni*	NC_053805.1
Hybomys trivirgatus	NC_053810.1	*Desmomys harringtoni*	MN807595.1
Hybomys planifrons	MN807601.1	*Mylomys dybowskii*	NC_053803.1
Hybomys planifrons	NC_053809.1	*Mylomys dybowskii*	MN807593.1
Dephomys defua	NC_053808.1	*Pelomys fallax*	MN807592.1
Dephomys defua	MN807600.1	*Lemniscomys barbarus*	NC_053800.1
Stochomys longicaudatus	MN807599.1	*Lemniscomys barbarus*	MN807585.1
Stochomys longicaudatus	NC_053807.1	*Lemniscomys striatus*	MN807586.1
Hybomys univittatus	MN807598.1	*Lemniscomys rosalia*	NC_053799.1
Hybomys univittatus	NC_053806.1	*Lemniscomys rosalia*	MN807584.1
Hybomys sp.	MN807597.1	*Lemniscomys bellieri*	MN807583.1
Micaelamys namaquensis	MN807615.1	*Arvicanthis rufinus*	MN964125.1
Micaelamys namaquensis	NC_053816.1	*Arvicanthis rufinus (1)*	NC_053802.1
Thallomys paedulcus	NC_053814.1	*Arvicanthis rufinus (1)*	MN807591.1
Thallomys paedulcus	MN807613.1	*Arvicanthis nairobae*	MN807590.1
Grammomys poensis	MN807611.1	*Arvicanthis abyssinicus (2)*	MN807589.1
Thamnomys kempi	NC_053813.1	*Arvicanthis somalicus*	NC_053801.1
Thamnomys kempi	MN807610.1	*Arvicanthis somalicus*	MN807588.1
Aethomys chrysophilus	MN807612.1	*Arvicanthis niloticus (3)*	MN807587.1
Aethomys nyikae	MN807609.1	*Arvicanthis niloticus (4)*	CM022273.1
Aethomys kaiseri	MN807608.1	*†Canariomys bravoi*	
Aethomys hindei	MN807607.1		

Table S2. Relation of NCBI codes of *Muridae* included in the molecular phylogenetic tree (listed in the same order). Discrepant species are indicated according to Bryja et al. 2019. (1) *Arvicanthis ansorgei,* (2) *Arvicanthis blicki*, (3) C2-C4 and (4) Masai Mara.

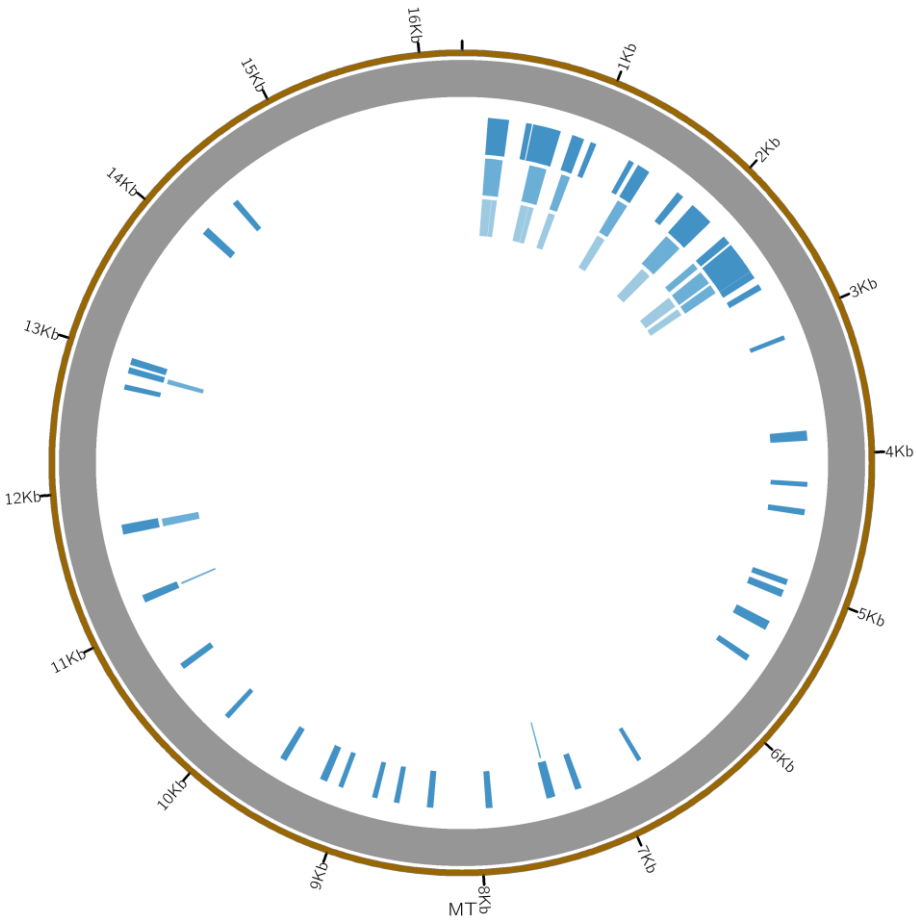


Figure S1. Validated reads distribution across *Arvicanthis niloticus* mitochondrial DNA genome. The outer grey circumference represents the mappability of the genome for 50bp reads (100%). The inner most histograms represent positions covered by at least one reads (dark blue), two reads (blue), and three or more reads (light blue).

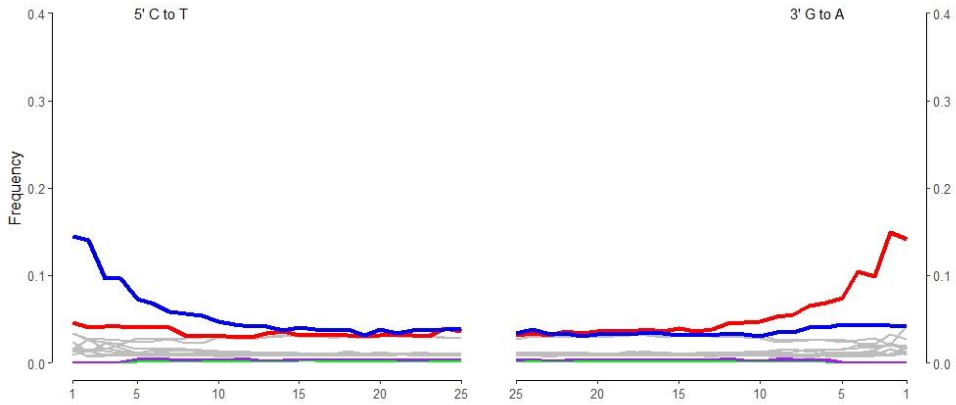


Figure S2. Deamination pattern at the end of mapped nuclear reads. Frequency of 5' C to T substitutions (blue) and 3' G to substitutions (red) are over-represented at the last bps of the reads in comparison with other type of substitutions (gray), as expected from a putatively ancient sample.

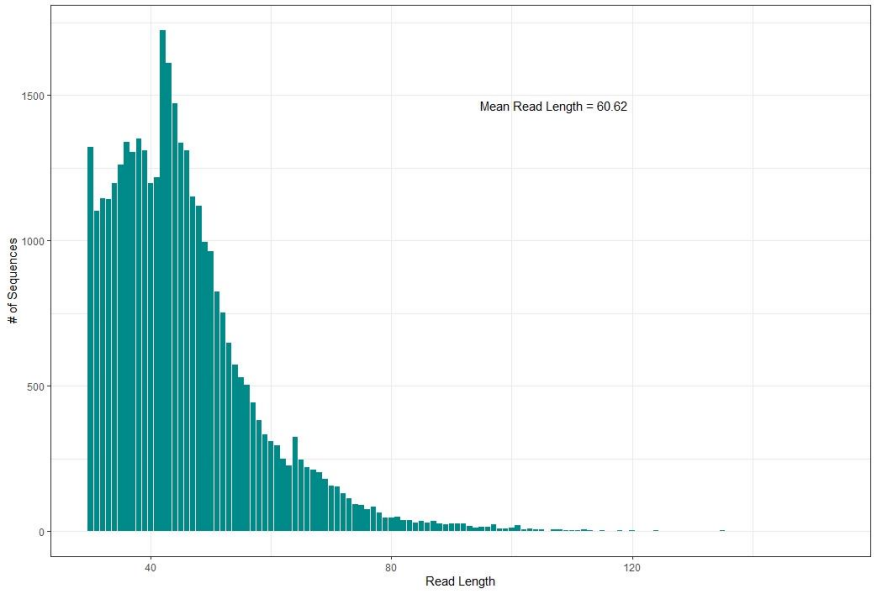


Figure S3. Read length histogram of mapped nuclear DNA reads. Their short length agrees with the sequences being ancient.

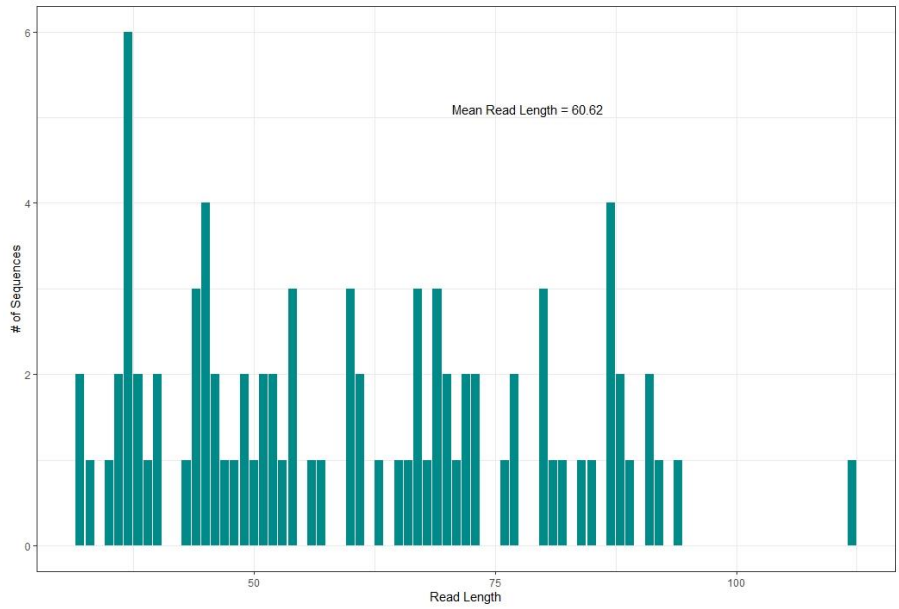


Figure S4. Read length histogram of validated mitochondrial DNA reads. Their short length agrees with the sequences being ancient.

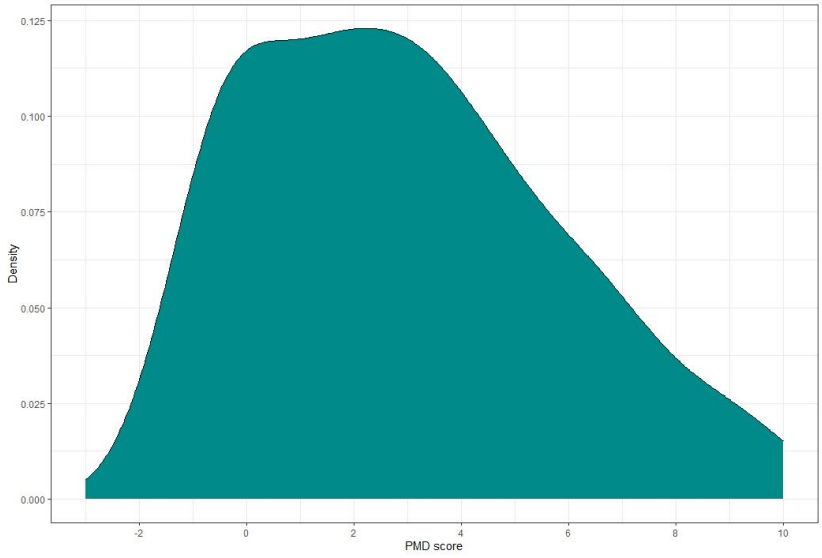


Figure S5. PMD score density of the validated mitochondrial DNA reads. A distribution shifted towards positive values supports the sequences being likely ancient.

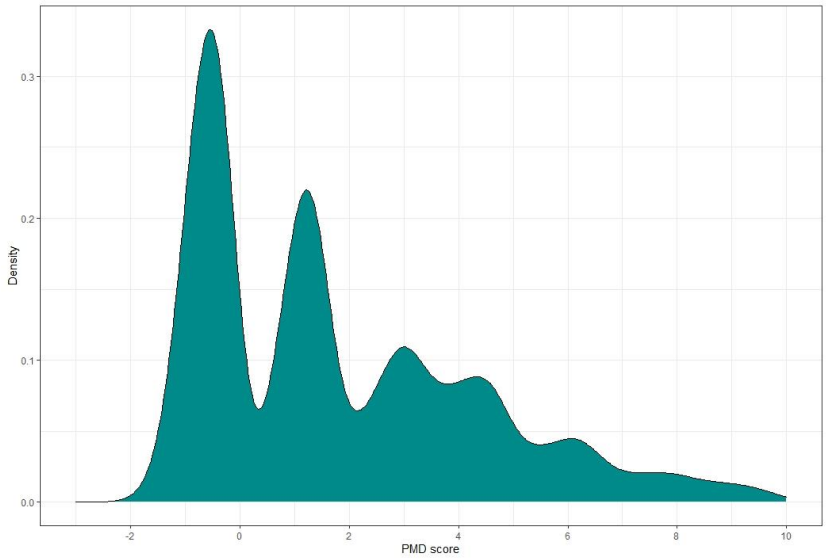


Figure S6. PMD score density of mapped nuclear DNA reads. A distribution shifted towards positive values supports the sequences being likely ancient.

4.3. Whole-genomes from the extinct Xerces Blue butterfly can help identify declining insect species

Toni de-Dios, Claudia Fontsere, Pere Renom, Josefin Stiller, Laia Llovera, Marcela Uliano-Silva, Alejandro Sánchez-Gracia, Charlotte Wright, Esther Lizan, Berta Caballero, Arcadi Navarro, Sergi Civit, Robert K. Robbins, Mark Blaxter, Tomàs Marquès-Bonet, Roger Vila and Carles Lalueza-Fox

In preparation

Whole-genomes from the extinct Xerces Blue butterfly can help identify declining insect species

Toni de-Dios[1,2+], Claudia Fontsere[1,3+], Pere Renom[1+], Josefin Stiller[3], Laia Llovera[1], Marcela Uliano-Silva[4], Alejandro Sánchez-Gracia[5], Charlotte Wright[4], Esther Lizano[1,6], Berta Caballero[7], Arcadi Navarro[1,8], Sergi Civit[9], Robert K. Robbins[10], Mark Blaxter[4], Tomàs Marquès-Bonet[1,6,8,11*], Roger Vila[1*], Carles Lalueza-Fox[12,1*]

[1] Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), 08003 Barcelona, Spain
[2] Institute of Genomics, University of Tartu, Riia 23B, 51010 Tartu, Estonia
[3] Section for Evolutionary Genomics, The Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen K, Denmark
[3] Centre for Biodiversity Genomics, University of Copenhagen, DK-2100, Denmark
[4] Wellcome Sanger Institute, Hinxton, Saffron Walden CB10 1RQ, UK
[5] Departament of Genetics, Microbiology and Statistics-Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona 08028, Spain
[6] Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain
[7] Museu de Ciències Naturals de Barcelona, 08019, Barcelona, Spain
[8] Catalan Institution of Research and Advanced Studies (ICREA), 08010 Barcelona, Spain
[9] Departament of Genetics, Microbiology and Statistics-Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona 08028, Spain
[10]Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012, USA
[11]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08036 Barcelona, Spain

[12]Museu de Ciències Naturals de Barcelona (MCNB), 08019 Barcelona, Spain
+ equally contributed
*Corresponding Authors
Carles Lalueza-Fox
Tomas Marquès-Bonet
Roger Vila

Keywords: Extinction, Ancient Genomics, Butterflies, Population Genomics, Xerces Blue; Genomic Erosion

Abstract
The Xerces Blue (*Glaucopsyche xerces*) is one of the first butterfly to become extinct at global scale in historical times. We sequenced the genomes of four 80 to 100-year-old Xerces Blue specimens and seven historic and one modern specimens of its closest relative, the Silvery Blue (*G. lygdamus*). We compared these to a novel annotated genome of the Green-Underside Blue (*G. alexis*). Phylogenetic analysis indicate that Xerces Blue diverged from the Silvery Blue lineage at least 850,000 years ago. We show that both species experienced population growth during the Marine Isotopic Stage 7 interglacial period, but that the Xerces Blue decreased to a very low effective population size subsequently. Runs of homozygosity in the Xerces Blue were significantly greater than in the Silvery Blue, suggesting a higher incidence of inbreeding. The genomic patterns of population decline observed in Xerces Blue can be used to monitor other, endangered insects.

Introduction
The Xerces Blue butterfly (*Glaucopsyche xerces*) was native to the coastal sand dunes of San Francisco in association the common Deerwood (*Acmispon glaber*), which was the preferred food source for larval stage (Tilden 1956). It was notable for its iridescent blue colouration on the dorsal (upper) wing surface, and conspicuous, variable white spots on the ventral surface (Downey 1956). With the growth of San Francisco and the destruction of sand dune habitats, the Xerces Blue became restricted to a few sites in what

is now Golden Gate National Recreation Area. The last specimens were reportedly collected by entomologist W. Harry Lange on March 23, 1941, and the Xerces blue has never been seen flying (Downey 1956). It is considered the first butterfly to have been driven to extinction by human activities (Downey 1956).

The Xerces Blue and the closely related Silvery Blue (*Glaucopsyche lygdamus*) were recently proposed to be distinct species based on mtDNA data from a Xerces Blue museum specimen (Grewe et al., 2021). However, two nuclear genes analysed (ribosomal 28S and histone H3) were invariable and genome-wide data were unavailable for the Xerces Blue, hampered by the inherent difficulties of retrieving genome-wide data from historical insect specimens (Thomsen et al., 2009; Staats et al., 2013) and the absence of a suitable reference genome. The genus Glaucopsyche consists of 18 extant species distributed across the temperate regions of the northern hemisphere. To provide a relevant reference, we generated an annotated genome from the Palearctic Green-Underside Blue butterfly *Glaucopsyche alexis* (Hinojosa Galisteo and Vila 2021). Using DNA extracted from five Xerces Blue and seven Silvery Blue (*Glaucopsyche lygdamus*) historical specimens from the vicinity of San Francisco, and also from a modern Silvery Blue from Canada, we generated whole genome resequencing data for both species and investigated their relationships and historical population genetics.

Methods

Historical butterfly specimens

The Xerces Blue specimens analysed belong to the Barnes collection deposited at the Smithsonian National Museum of Natural History (Figure S1). Two of them were collected on April 26th, 1923. The Silvery Blue specimens were mostly collected between 1927 and 1948, in Haywood City, Santa Cruz, Oakland, San José, Fairfax and Marin County (these locations surround San Francisco Bay) (Table S1).

DNA extraction and sequencing of Xerces Blue and Silvery Blue specimens

All DNA extraction and initial library preparation steps (prior to amplification) were performed in a dedicated clean lab, physically-isolated from the laboratory used for post-PCR analyses. Strict protocols were followed to minimize the amount of human DNA in the ancient DNA laboratory, including the wearing a full body suit, sleeves, shoe covers, clean shoes, facemask, hair net and double gloving, as well as frequent bleach cleaning of benches and instruments. DNA extraction was performed from 12 abdominal samples of historical Xerces Blue and Silvery Blue, as well as a modern Silvery Blue specimen from Canada. Experimental procedures are described in detail in the Supplementary Material.

***Glaucopsyche alexis* genome sequencing and annotation**

Glaucopsyche alexis was chosen as a congeneric reference to compare the demographic histories of both the Xerces Blue and the Silvery Blue (Fig. 1). We generated a G. alexis reference genome from a male specimen collected in Alcalá de la Selva in Teruel (Spain). Its genome has a sequence length of 619,543,730 bp on 24 chromosomes – including the Z sex chromosome – and the mitochondrial genome. The genome sequence is biologically complete (BUSCO Lepidoptera completeness 97.1%). The G. alexis genome was sequenced at the Sanger Institute as part of the Darwin Tree of Life Project following the extraction, sequencing and assembly protocols developed for Lepidoptera (Hinojosa Galisteo and Vila 2021).

Xerces Blue and Silvery Blue mapping and variant calling

The ancient DNA reads were clipped using AdapterRemoval2, and only reads longer than 25bp were kept. Filtered reads were mapped against the G. alexis assembly with Burrows-Wheeler Aligner (BWA), with parameters optimised for the analysis of aDNA (Supplementary Materials). Basic mapping statistics were generated using Qualimap2 (Table S2). We used bedtools to assess genome coverage across the reference, using windows of 1mbp for the nuclear fraction of the genome (Figure S2 – S4).

Authenticity of the sequences was assessed by characterising aDNA damage patterns with pmdtools and MapDamage2 (Figure S5 and S6).

We used snpAD, a program for genotype calling in ancient specimens. The mapped sequences were transformed from bam-format into snpAD-format files, priors for base composition estimated, and genotypes were called using standard settings. The VCFs were combined and concatenated with CombineVariants and GatherVcfs from GATK and filtered with vcftools to keep only sites within the mappable fraction of the genome previously obtained with minimum read depth of 2, max read depth of 30, genotype quality > 30, maximum missingness of 0.6, minor allele frequency of 5% and excluding indels and multiallelic sites.

Genotype likelihoods were obtained with ANGSD using the GATK model with the following parameters for all the samples: uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 10 -C 50 -baq 1 -minInd 5 -skipTriallelic 1 -GL 2 -minMapQ 30 -doGlf 2 -doMajorMinor 1 -doMaf 2 -minMaf 0.05 -SNP_pval 1e-6.

Mitochondrial phylogenetic tree and divergence dating

Haploid variants were called using bcftools with a ploidy of 1 and filtering low quality indels and variants and a consensus sequence was exported. We downloaded 14 complete mitochondrial genomes for Polyommatinae from NCBI (Table S3).

All mitochondrial genomes were annotated with MitoFinder using Shijimiaeoides divina as the reference. The 11 protein-coding genes were aligned with the codon-aware aligner MACSE and the 2 ribosomal rRNAs were aligned with MAFFT l-ins-i. We first investigated phylogenetic relationships among five G. xerces and eight G. lygdamus individuals, with G. alexis as the outgroup. We used IQ-TREE2 to select the best fitting nucleotide substitution model for each partition and merge similar partitions (Kalyaanamoorthy et al., 2017), build a maximum likelihood tree and assess support with 1000 ultrafast bootstrap replicates (Hoang et al., 2018).

To infer a time-calibrated phylogenetic hypothesis, we selected one individual of Xerces Blue (L003) and Silvery Blue (RVcoll10-B005)

and analysed together with 13 other *Polyommatinae* species. As above, we used IQ-TREE2 to merge partitions. We used BEAST2 with the bModelTest package to perform phylogenetic site model averaging for each of the merged partitions. Because there is no accepted molecular clock rate for butterflies and no fossils to apply in this part of the phylogeny, we used two strategies to apply time constraints to the analysis. First, we used two published molecular clock rates for the mitochondrial COX1 gene (1.5% divergence/Ma estimated for various invertebrates (Quek et al., 2004), and the 'standard' insect mitochondrial clock 2.3% divergence/Ma (Van Zandt Brower 1994)). We applied a strict clock with a normal prior set up to span the 1.5-2.3% range with the 95% HPD interval (mean=1.9%, sigma=0.00119). Second, we borrowed the age of the most recent common ancestor of our sampled taxa from fossil-calibrated analyses across butterflies, which has been estimated to ~33 Ma (Chazot et al., 2019; Wiemers et al., 2020). We fixed the root age to 33 Ma and allowed the remaining node ages to be estimated using a strict clock. Analyses were run twice from different starting seeds for 10 million MCMC generations and trees were sampled every 1000 generations. Runs were checked for convergence with Tracer and all effective sample size (ESS) values were >200. Runs were combined with the BEAST2 package LogCombiner, after removing the first 10% of topologies as burn-in, and a maximum credibility tree was generated with TreeAnnotator. Phylogenetic analyses were performed on the National Life Science Supercomputing Center - Computerome 2.0 (www.computerome.dk).

Xerces Blue and Silvery Blue population histories

We used the Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011) to explore the demographic history of both butterfly species. We obtained a consensus fastq sequence of the mappable fraction of the genome for each autosomal chromosome (total of 22 chromosomes of *G. alexis* assembly). Only positions with a depth of coverage above 4X and below 15X were kept. Posteriorly a PSMC was built using the following parameters: -N25 -t15 -r5 -p "28*2+3+5". We used 1 year for the

generation time and a mutation rate of 1.9x10-9, estimated in Heliconius Melpomene (Martin et al., 2016). Considering that calling consensus sequences from low coverage samples (< 10x) can underestimate heterozygous sites (Keightley et al., 2015), and given the different coverage between samples, we corrected by False Negative Rate the samples with coverage lower than the coverage of L005 (for Xerces Blue) and L013 (for Silvery Blue), as recommended by the developers of the software, so that all samples are comparable with each other. However, since in our dataset we do not reach a coverage > 20x, we acknowledge that we are not capturing the whole diversity and thus our PSMC might infer lower historical effective population sizes.

Population stratification and average genome heterozygosity

Principal Component Analysis (PCA) was performed using PCAngsd after obtaining genotype likelihoods with ANGSD including all individuals. To assess global levels of heterozygosity, the unfolded SFS was calculated for each sample separately using ANGSD and realSFS with the following quality filter parameters: -uniqueOnly 1 - remove_bads 1 -only_proper_pairs 1 -trim 10 -C 50 -baq 1 -minMapQ 30 -minQ 30 -setMaxDepth 200 - doCounts 1 -GL 2 -doSaf 1.

Runs of Homozygosity (RoH)

RoH were called based on the density of heterozygous sites in the genome using the implemented Hidden Markov Model (HMM) in bcftools roh with the following parameters: -G30 --skip-indels --AF-dflt 0.4 --rec-rate 1e-9 from the mappable fraction of the genome with the filtered VCF file. We kept the RoH with a phred score > 85. We divided the RoH into different size bins: very short RoH (<100 kb), short RoH (100-500 kb), intermediate RoH (500 kb-1Mb) and long (1-5 Mb or > 5Mb). Short RoH reflect LD patterns, intermediate size RoH describe background inbreeding due to genetic drift and long RoH appear in the case of recent inbreeding (Ceballos et al., 2018).

Deleterious load and fixed mutations

We used the *G. alexis* annotations to create a SNPeff database that we used to annotate our callings. Using SNPeff again and the set of variants discovered by angsd, we predicted the putative effect of those variant in the analysed individuals (Table S2). In addition to wide genome mutations, we specifically focused on mutations present in homozygosis, heterozygosis and the previously annotated RoH.

Results

The Xerces Blue and the Silvery Blue are recovered as sister taxa with high support (posterior probability=1); the modern Silvery Blue from Canada clusters basal to the historical specimens from the same species. Our phylogenetic analysis yielded an origin of this subgroup of *Polyommatinae* at 12.4 Ma (7.16-13.76 Ma 95% HPD interval) and divergence of the Xerces Blue from the Silvery Blue at 900,000 years ago (560,895-1,159,512 years 95% HPD interval, Fig. 1B). A second estimate based on larger-scale fossil-based calibrations (Espeland et al., 2018) placed the origin of the subgroup at ca. 33 Ma (Chazot et al., 2019) with the subsequent divergence of the Xerces Blue and Silvery Blue at 2.40 Ma (2.14-3.08 Ma 95% HPD interval, Figure 1B). The recent speciation of Xerces and Silvery Blue is not obviously due to infection with the Wolbachia, as no evidence of infection of the sampled specimens with this alpha-proteobacterium was detected in the raw read data.

The PCA also supports the relationships among them; the historical specimens are equally distant to *G. alexis* in the first PC, explaining 52.81% of the variance (Fig. S7). The second PC separates the Xerces Blue from the Silvery Blue specimens.

We first explored the demographic history of both butterfly species, first using the two specimens with highest coverage (L05 and L13) (Fig. 2). We found an increase in effective population size in both species that is roughly coincident with the interglacial Marine Isotopic Stage 7 (approximately from 240,000 to 190,000 years ago (Batchelor et al., 2019)). After this timepoint the trends differ; there is a continuous decrease in Xerces Blue population size in parallel to the Wisconsin Glacial Episode, which started about 75,000 years

ago. However, both the modern and the historical Silvery Blue does not appear to have been similarly affected by this event (Fig. S3), suggesting different adaptive strategies to cope with cooling temperatures and/or food plant availability.

Second, we generated PSMC curves from the remaining lower-coverage individuals and down-sampled data from specimen L05 to 50% and 75% of the total coverage to explore the effects of coverage on estimation of heterozygous sites. Although there was a reduction in the effective population size estimates, as expected, the temporal trajectories in lower-coverage individuals were similar to their respective, higher-coverage Xerces Blue and Silvery Blue references (Fig. S8).

We subsequently explored the heterozygosity of each individual and found that Xerces Blue had 22% less heterozygosity on average than the Silvery Blue historical samples, a difference that is statistically significant (T-test; p=0.0072) (Fig. S9, Table S2). We searched for runs of homozygosity (RoH) that can indicate the existence of inbreeding in a dwindling population. The total fraction of the genome presenting RoH, although limited, is much higher in Xerces Blue (up to 6% of the genome) than in Silvery Blue, especially in short RoH of size between 100 and 500 kb (Fig. 3 and Fig. S9), consistent with background inbreeding. The limited presence of long RoH discards consanguinity as a common scenario in Xerces Blue.

We identified amino acid-changing alleles that may be suggestive of a deleterious genetic load associated with long-term low population numbers in the Xerces Blue. The average Ka/Ks ratio is higher in Xerces Blue than in Silvery Blue; the former also carries a higher fraction of nonsense and functionally high-to-moderate effect variants in homozygosis and RoHs with an increased concentration of high-to-moderate effect variants (Fig. 4), as predicted with a functional prediction toolbox, SnpEff (Coon et al., 2012).

Discussion

We have used a modern reference genome and ancient DNA genome sequence data from museum specimens to explore the

relationships and historical population genetic history of an extinct butterfly, the Xerces Blue; this is the first ancient genome ever generated from an extinct insect. Based upon a near-complete mtDNA genome from a Xerces Blue specimen, Grewe et al., (2021) proposed that the Xerces Blue and the Silvery Blue were distinct species. We confirm this finding using full mitochondrial genomes and extensive nuclear genomic data from multiple specimens. Given the lack of evidence for Wolbachia infection, a detailed analysis of genomic architectures could help identify barriers to introgression between these species.

Our analyses indicate that the Xerces Blue had experienced a severe demographic decline for tens of thousands of years, likely associated with changing climatic factors. Thus, the historical destruction of the Xerces Blue habitat by humans was likely the final blow in the extinction process. We provide evidence for low population size in Xerces Blue, correlated with low genetic variation, a higher proportion of runs of homozygosity and increased frequency of deleterious, amino acid-changing alleles (Spielman et al., 2004; Szpiech et al., 2013; Palkopoulou et al., 2015). However, there was no genetic evidence of recent inbreeding.

Inbreeding genetic signals in the form of long chromosomal sections with no variation sometimes occur in critically endangered species (Díez-del-Molino et al., 2018; van der Valk et al., 2019) and in extinct species such as the last Mammoths from Wrangel Island (Rogers and Slatkin 2017) or the Altai Neanderthal (Prüfer et al., 2014). The PSMC shows a continuous low effective population size for Xerces Blue; demographic declines are also seen in some extinct species, including Wrangel Mammoths (Palkopoulou et al., 2015) but not in others such as the Woolly Rhino that showed a pre-extinction demographic stability and relatively low inbreeding signals (Lord et al., 2020). In many endangered species there is little concordance between genome diversity, population sizes and conservation status (Díez-del-Molino et al., 2018); this decoupling was also observed in the genomes of the extinct passenger pigeon that despite being one of the world's most numerous vertebrates, showed a surprisingly low genetic (Murray et al., 2017). Despite

being notoriously abundant, insects, and in particular butterflies, are very sensitive to climate fluctuations; therefore, we suggest that insects with observations of demographic traits indicative of long-term low effective population size such as those found in Xerces Blue should be considered to be especially vulnerable to extinction events. Irrespective of this, our study further demonstrates the value of ancient DNA in museum specimens for evolutionary studies at a population scale.

Acknowledgments

C.L.-F. is supported by a PGC2018-0955931-B-100 grant (MCIU/AEI/FEDER, UE) of Spain; T.M.-B. is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 864203), BFU2017-86471-P (MINECO/FEDER, UE), "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M), Howard Hughes International Early Career, Secretaria d'Universitats i Recerca and CERCA Programme (Generalitat de Catalunya, GRC 2017-SGR-880); R.V. is supported by grant PID2019-107078GB-I00, funded by MCIN/AEI/ 10.13039/501100011033, and by GRC 2017-SGR-991 (Generalitat de Catalunya). We are grateful to the SCIENCE Faculty at University of Copenhagen for free access to Computerome 2.0. This research was funded in whole, or in part, by the Wellcome Trust Grants 206194 and 218328 (MU, MB). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Competing Interest Statement: The authors declare no competing interest.

Author Contributions: C.L.-F., R.V., R.R. and P.R. conceived the project. R.R. studied and sampled the specimens. L.L. and E.L. performed experimental work. T.d-D., C.F., J.S., A.S.G., M.U.-S., C.W., S.C. and P.R. undertook different computational analyses. C.L.-F., T.M.-B., A.N. and M.B. coordinated different computational

teams. J.S. worked in visualization. C.L.-F. and R.R. wrote the manuscript with input from all coauthors.

Data Accessibility

The genetic data generated is publicly available; the accession numbers for the Xerces Blue and Silvery Blue genomes reported in this study are in the European Nucleotide Archive (ENA): PRJEB47122. Data on G. alexis are available in INSDC under BioProject PRJEB43798 and genome assembly accessions GCA_905404095.1 (primary haplotype) and GCA_905404225.1 (secondary, alternate haplotype).

References

Batchelor CL, Margold M, Krapp M, Murton DK, Dalton AS, Gibbard PL, Stokes CR, Murton JB, Manica A. 2019. The configuration of Northern Hemisphere ice sheets through the Quaternary. Nat. Commun. [Internet] 10:3713. Available from: https://doi.org/10.1038/s41467-019-11601-2

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. 19:220–234.

Chazot N, Wahlberg N, Freitas AVL, Mitter C, Labandeira C, Sohn JC, Sahoo RK, Seraphim N, De Jong R, Heikkilä M. 2019. Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. Syst. Biol. 68:797–813.

Coon M, Wang L, Cingolani P, Nguyen T, Wang LL, Land SJ, Ruden DM, Lu X, Platts A. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 6:80–92.

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. 2018. Quantifying Temporal Genomic Erosion in Endangered Species. Trends Ecol. Evol. 33:176–185.

Downey JC. 1956. Analysis of Variation in a recently extinct polymorphic lycaenid butterfly, Glaucopsyche Xerces. Bull. So. Calif. Acad. Sci. 55:153–170.

Espeland M, Breinholt J, Willmott KR, Warren AD, Vila R, Toussaint EFA, Maunsell SC, Aduse-Poku K, Talavera G, Eastwood R, et al. 2018. A Comprehensive and Dated Phylogenomic Analysis of Butterflies. Curr. Biol. [Internet] 28:770-778.e5.
https://www.sciencedirect.com/science/article/pii/S0960982218300940

Grewe F, Kronforst MR, Pierce NE, Moreau CS. 2021. Museum genomics reveals the Xerces blue butterfly (*Glaucopsyche xerces*) was a distinct species driven to extinction . Biol. Lett. 17:20210123.

Hinojosa Galisteo JC, Vila R. 2021. The genome sequence of the green-underside blue, Glaucopsyche alexis (Poda, 1761). Wellcome Open Res. 6:274.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. [Internet] 35:518–522.
https://doi.org/10.1093/molbev/msx281

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–589.

Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD. 2015. Estimation of the Spontaneous Mutation Rate in Heliconius melpomene. Mol. Biol. Evol. 32:239–243.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature [Internet] 475:493–496. https://doi.org/10.1038/nature10231

Lord E, Dussex N, Kierczak M, Díez-del-Molino D, Ryder OA, Stanton DWG, Gilbert MTP, Sánchez-Barreiro F, Zhang G, Sinding MHS, et al. 2020. Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros. Curr. Biol. [Internet] 30:3871-3879.e7
https://doi.org/10.1016/j.cub.2020.07.046

Martin SH, Möst M, Palmer WJ, Salazar C, McMillan WO, Jiggins FM, Jiggins CD. 2016. Natural selection and genetic diversity in the butterfly heliconius melpomene. Genetics [Internet] 203:525–541. Available from: https://www.genetics.org/content/203/1/525

Murray GGR, Soares AER, Novak BJ, Schaefer NK, Cahill JA, Baker AJ, Demboski JR, Doll A, Da Fonseca RR, Fulton TL, et al. 2017. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. Science. 358:951–954.

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, et al. 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Curr. Biol. [Internet] 25:1395–1400. https://www.sciencedirect.com/science/article/pii/S0960982215004200

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49.

Quek SP, Davies SJ, Itino T, Pierce NE. 2004. Codiversification in an ant-plant mutualism: Stem texture and the evolution of host use in Crematogaster (Formicidae: Myrmicinae) inhabitants of Macaranga (Euphorbiaceae). Evolution (N. Y). 58:554–570.

Rogers RL, Slatkin M. 2017. Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet. [Internet] 13:e1006601. Available from: https://doi.org/10.1371/journal.pgen.1006601

Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. Proc. Natl. Acad. Sci. U. S. A. [Internet] 101:15261–15264. https://pubmed.ncbi.nlm.nih.gov/15477597

Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic Treasure Troves: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. PLoS One [Internet] 8:e69189–e69189. https://pubmed.ncbi.nlm.nih.gov/23922691

Szpiech ZA, Xu J, Pemberton TJ, Peng W, Zöllner S, Rosenberg NA, Li JZ. 2013. Long Runs of Homozygosity Are Enriched for Deleterious Variation. Am. J. Hum. Genet. [Internet] 93:90–102. https://www.sciencedirect.com/science/article/pii/S0002929713002164

Thomsen PF, Elias S, Gilbert MTP, Haile J, Munch K, Kuzmina S, Froese DG, Sher A, Holdaway RN, Willerslev E. 2009. Non-destructive sampling of ancient insect DNA. PLoS One 4:e5048.

Tilden JW. 1956. San Francisco's vanishing butterflies. Lepid. News 10:133–1445.

van der Valk T, Díez-Del-Molino D, Marques-Bonet T, Guschanski K, Dalén L. 2019. Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. Curr. Biol. 29:165-170.e6.

Van Zandt Brower A. 1994. Phylogeny of Heliconius Butterflies Inferred from Mitochondrial DNA Sequences (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. [Internet] 3:159–174. https://www.sciencedirect.com/science/article/pii/S1055790384710189

Wiemers M, Chazot N, Wheat CW, Schweiger O, Wahlberg N. 2020. A complete time-calibrated multi-gene phylogeny of the european butterflies. Zookeys 2020:97–124.

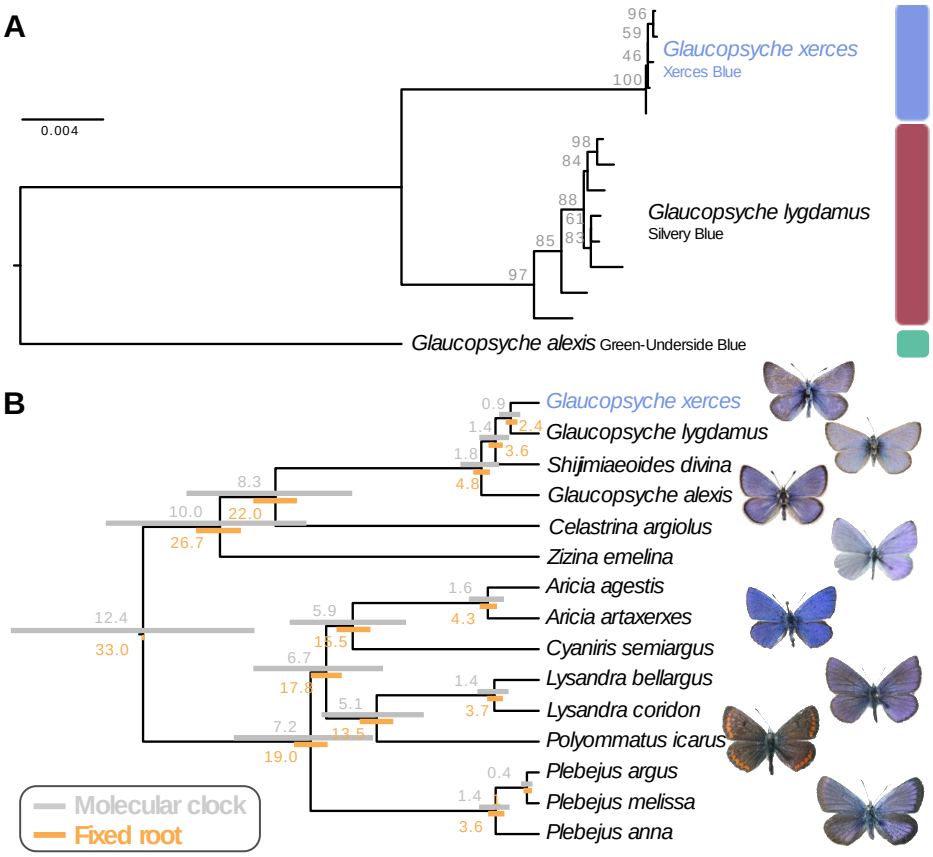


Fig. 1. Phylogenetic placement of the Xerces Blue. A: Maximum likelihood tree from whole mitochondrial genomes of Xerces Blue, Silvery Blue and Green-Underside Blue. Node labels are bootstrap support values. B: Time-calibrated phylogeny from Bayesian inference using mitochondrial protein-coding genes of Xerces Blue and related butterflies. Node values show median age estimates from dating analysis with a molecular clock (above nodes) or from fixing the age of the root (below nodes). Bars are 95% HPD intervals for node ages.

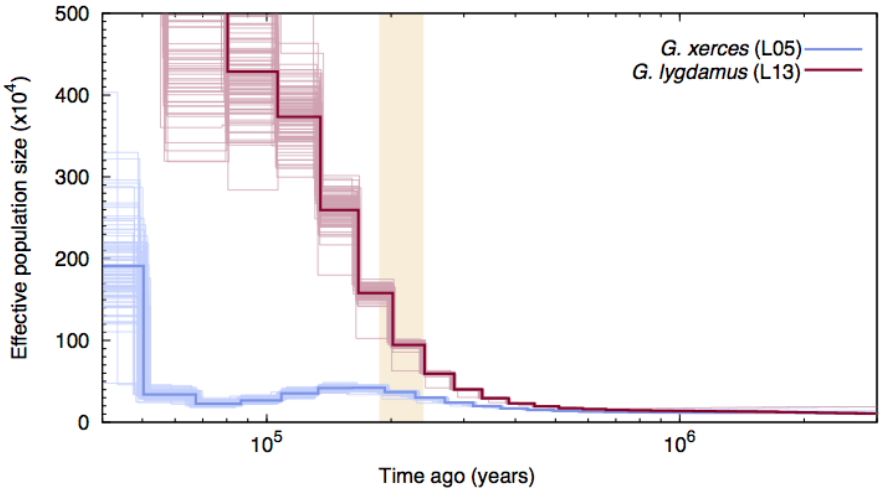


Fig. 2. PSMC plot of one Xerces Blue (*Glaucopsyche xerces*) (L05) specimen and one Silvery Blue specimen (*Glaucopsyche lygdamus*). The two historical samples are those with higher average coverage. Individual PSMC plots were bootstrapped 100 times each (lighter lines). One year of generation time and a mutation rate of $\mu=1.9\times10^{-9}$ were used. The peak of the Marine Isotopic Stage 7 interglacial is marked in yellow.

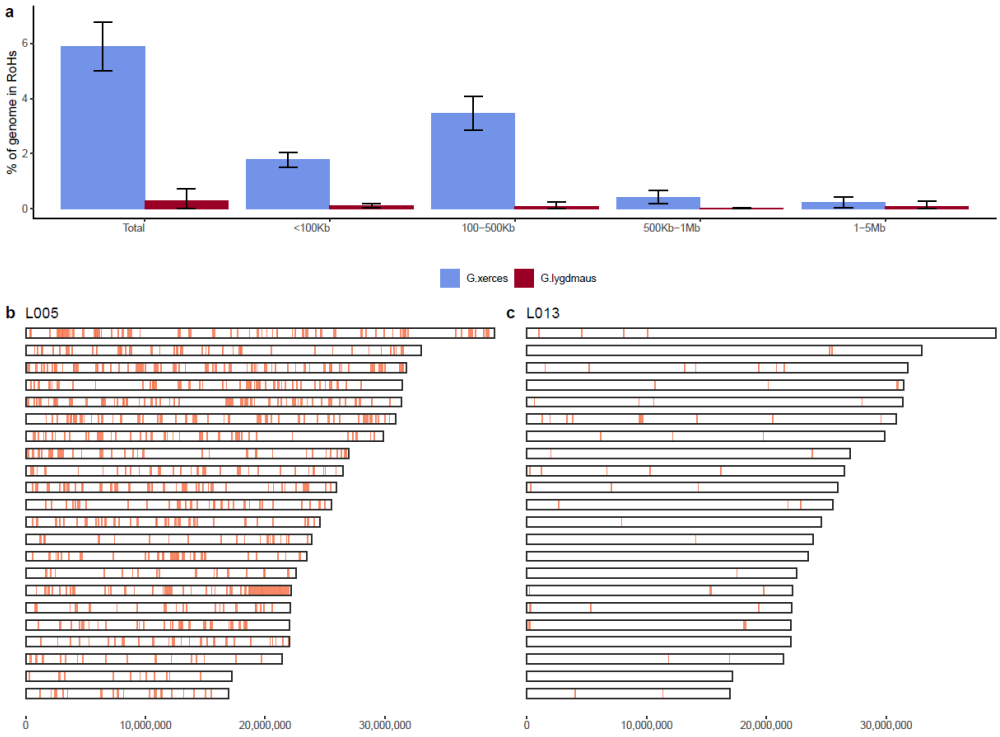


Fig. 3. Runs of Homozygosity (RoH) in the genomes of Xerces Blue and Silvery Blue (modern and historical). a: Percentage of the autosomal genome in RoH by size bins: very short RoH (<100 Kb), short RoH (100-500 Kb), intermediate RoH (500Kb-1Mb) and long (1-5Mb). Short RoH reflect LD patterns, intermediate size RoH describe background inbreeding due to genetic drift and long RoH appear in the case of very recent inbreeding due to consanguinity. Error bars show the standard deviation. b: Distribution of RoH in the autosomal genome of a Xerces specimen, L05 c) Distribution of RoH in the autosomal genome of a Silvery specimen L13.

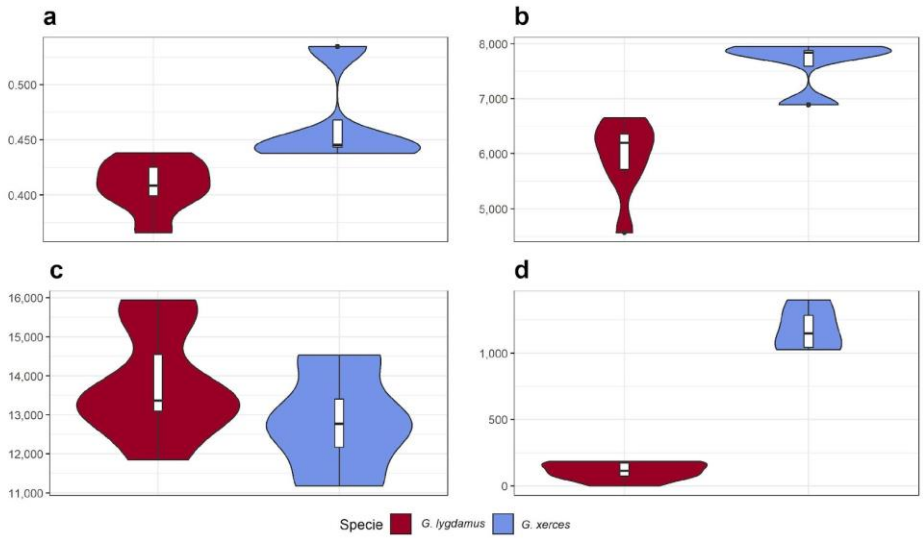


Fig. 4. Functional effect prediction on the fixed amino acid-changing alleles observed in Xerces Blue and Silvery Blue. a: Wide genome Ka/Ks ratio comparison. b: High-to-moderate effect variant comparison in homozygous sites. c: High-to-moderate effect variant comparison in heterozygous sites. d: Presence of high-to-moderate variants in regions of the genome in RoH. Error bars show the standard deviation.

Supplementary Information

This file includes:

Materials and Methods

Figs. S1 to S9

Tables S1

Materials and Methods

One ml of digestion buffer (final concentrations: 3 mM CaCl2, % SDS, 40 mM DTT, 0.25 mg/ml proteinase K, 100 mM Tris buffer pH 8.0 and 100 mM NaCl) was added to each crushed butterfly residue, including an extraction blank, and incubated at 37 °C overnight (24h) on rotation (750-900 rpm). Next, DNA extraction was continued following the method proposed by Dabney et al., (2013). Remaining butterfly sample was then pelleted by centrifugation in a bench-top centrifuge for 2 min at maximum speed (16,100 × g). The supernatant was added to 10 mL of binding buffer (final concentrations: 5 M guanidine hydrochloride, 40% (vol/vol) isopropanol, 0.05% Tween-20, and 90 mM sodium acetate (pH 5.2)) and purified on a High Pure Extender column (Roche). DNA extracts were eluted with 45 µL of low EDTA TE buffer (pH 8.0) and quantified using a Qubit instrument.

Following extraction, the DNA extract was converted into Illumina sequencing libraries following the BEST protocol(26). Each library was amplified by PCR using two uniquely barcoded primers, prior to being purified with a 1.5x AMPure clean (Beckman Coulter) and eluted in 25 µl of low EDTA TE buffer (pH 8.0). One Xerces Blue sample did not yield detectable DNA in two independent extractions. For each of the successful extracts we prepared a single library which was shotgun sequenced on the HiseqX Illumina platform. We mapped 124,101,622 and 184,084,237 unique DNA reads of Xerces Blue and Silvery Blue, respectively, against the *G. alexis* reference genome (Table S2). The DNA reads exhibited typical ancient DNA features, such as short mean read length (ranging from 47.55 to 67.41 bases on average, depending on the specimen (Fig. S1)) and post-mortem deamination patterns at the 5' and 3' ends (Table S2) (Fig. S2). The historical genomes covered 49.3% (Xerces Blue) and 55.2% (Silvery Blue) of the *G.*

alexis reference genome. To estimate the mappable fraction of this reference, we randomly fragmented it to 50 to 70 nucleotides and mapped the generated fragments back to the complete genome. An average of 57.8% of the *G. alexis* genome was covered with these read lengths (Fig. S3). We suggest that reduced coverage from the historical specimens may be due to genomic divergence of *G. xerces* and *G. lygdamus* from the *G. alexis* reference and also the presence of unmappable, repetitive regions (Fig. S4).

The sex of the specimens was determined by differential coverage of the Z chromosome (females are the heterogametic sex in the Lepidoptera and show reduced coverage on the Z chromosome). As listed in the original museum records, we found one Silvery Blue and two Xerces Blue females (Table S2). Inter-individual comparisons suggested no close kinship link among the studied individuals.

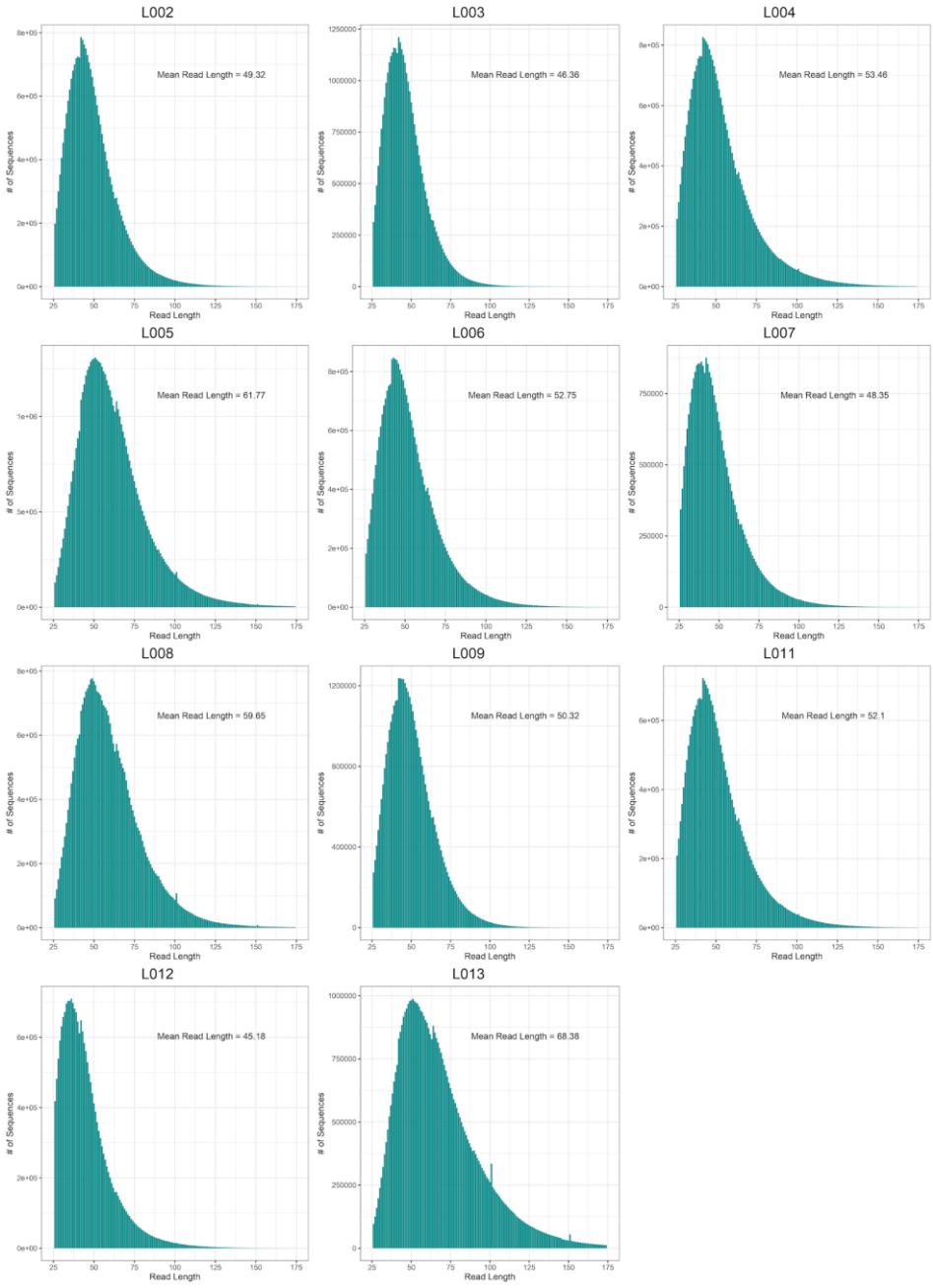


Fig. S1. Read length distribution across the 11 analysed samples.

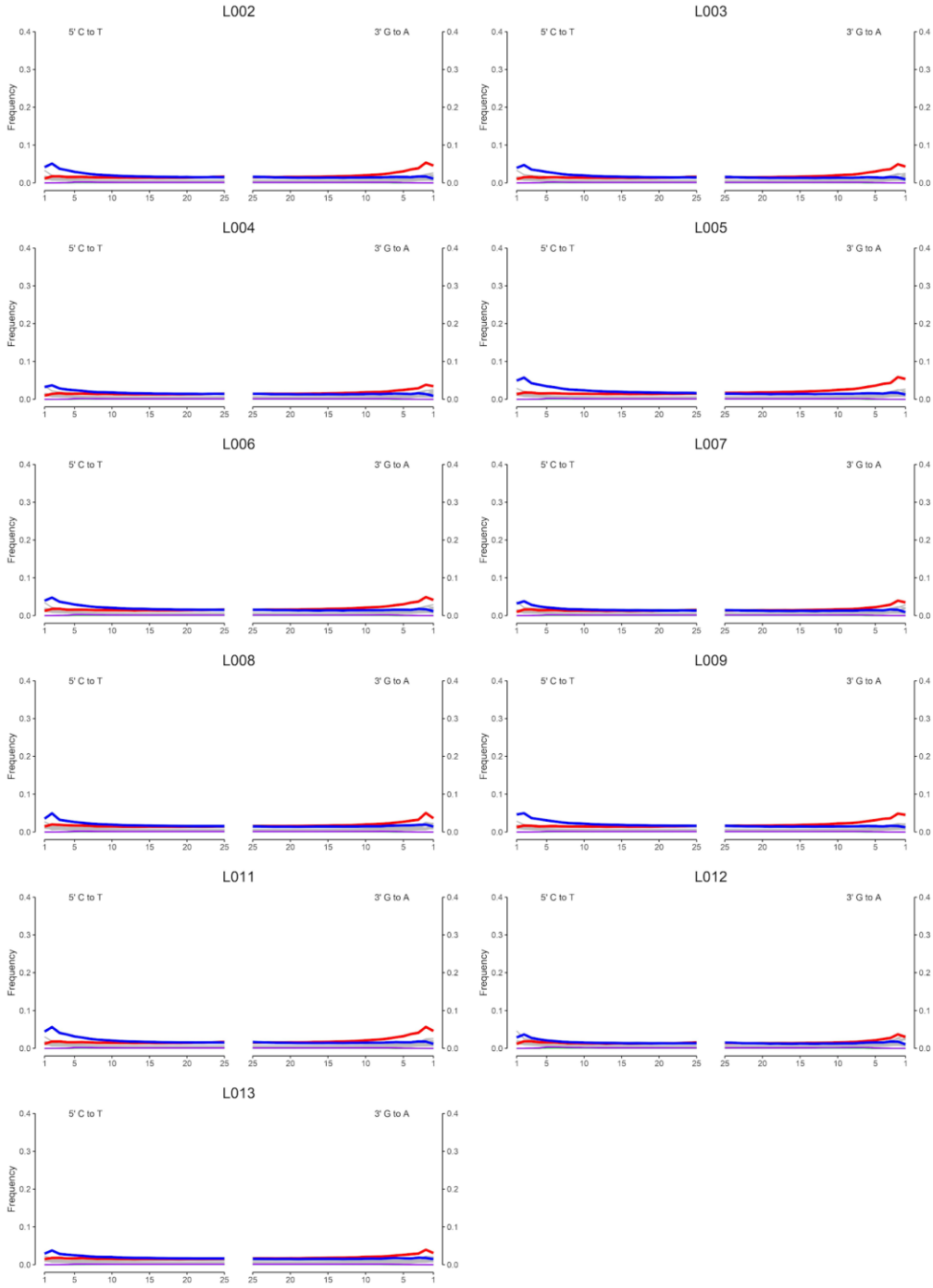


Fig. S2. Post-mortem damage patterns of the analysed samples. Frequency of C to T (blue) and G to A (red) substitutions are displayed across the last 25 bases of each end.

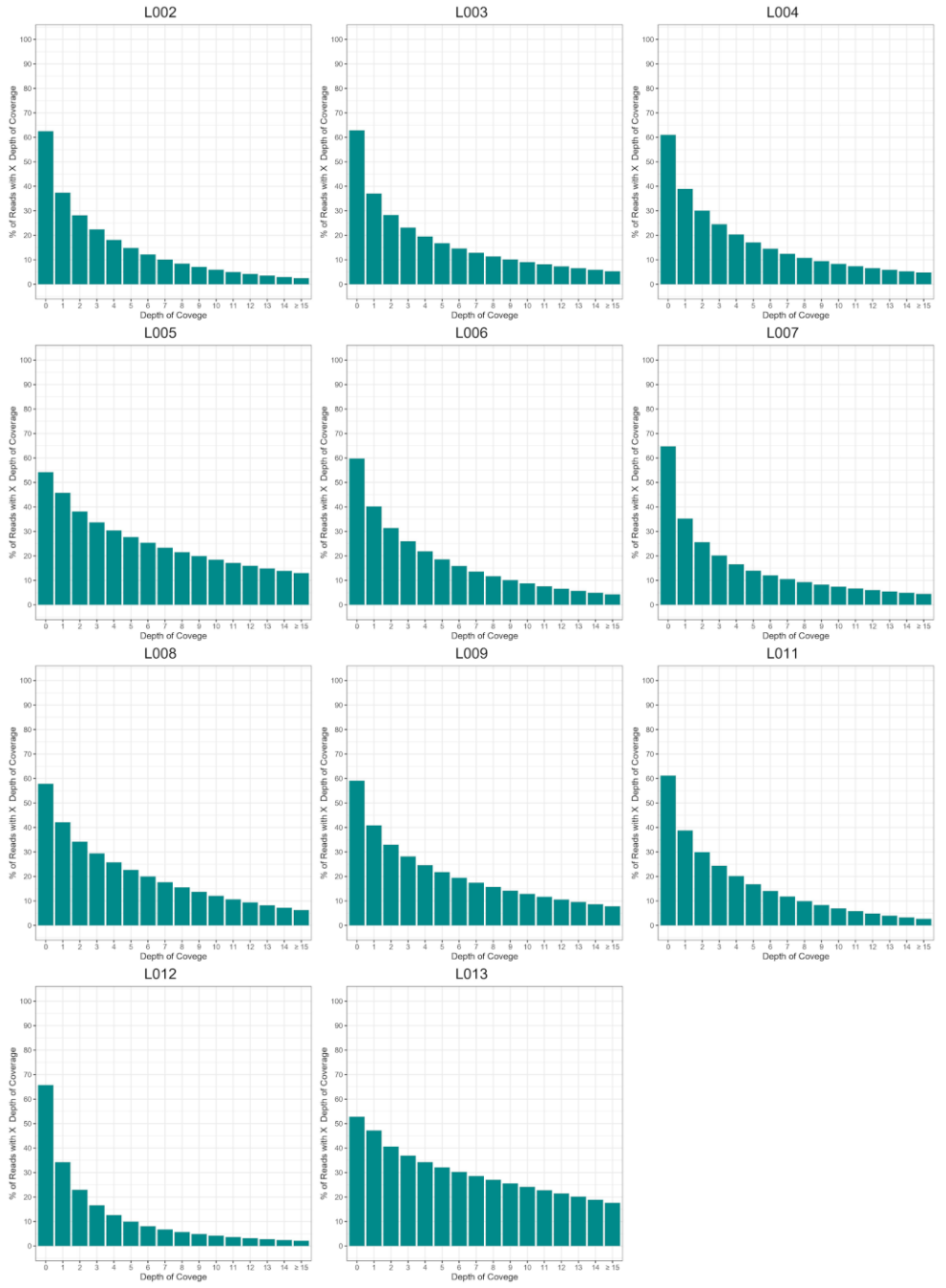


Fig. S3. Depth of coverage distribution per position.

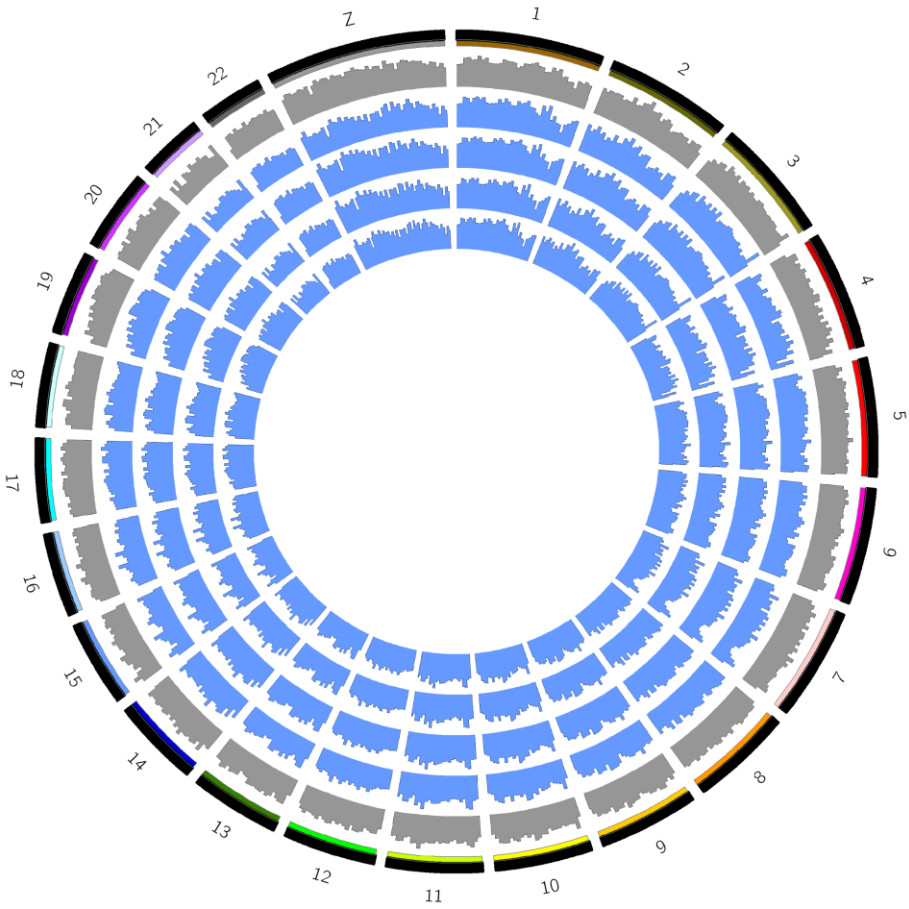


Fig. S4A. Coverage distribution of *G. xerces* samples across the reference. Reference mappability is represented by the outermost grey histogram. Blue histograms represent the coverage of *G. xerces* samples across the *G. alexis* reference in windows of 1mbp. The sample represented are, from outer to inner rings: L003, L005, L007 and L009.

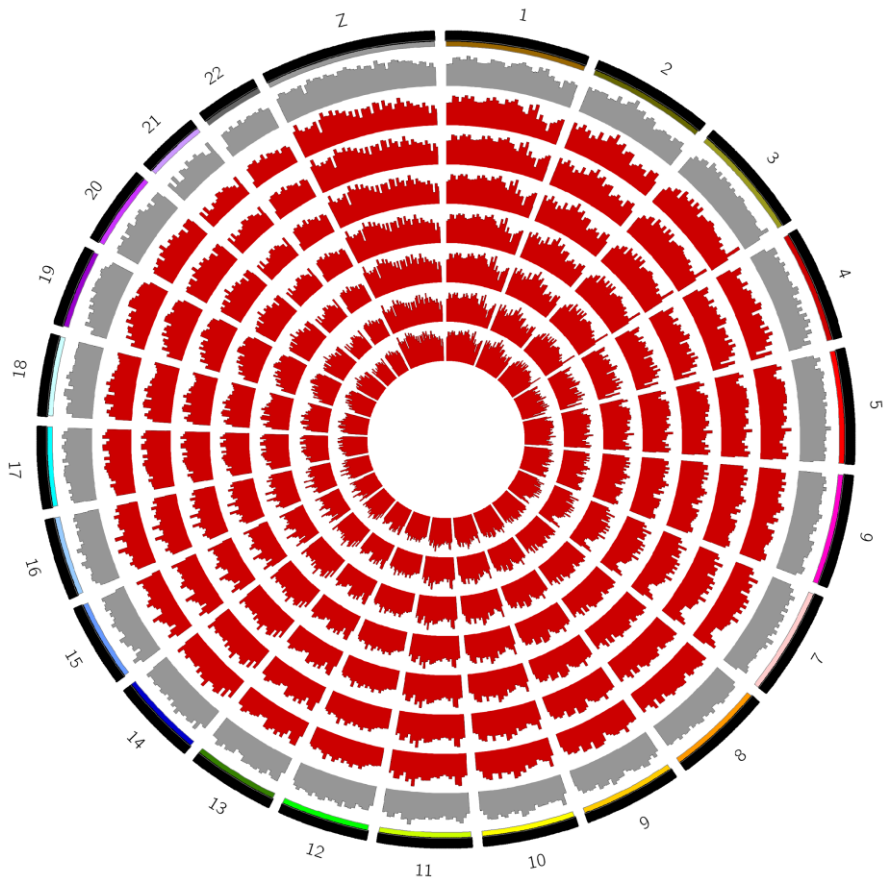


Fig. S4B. Coverage distribution of *G. lygdamus* samples across the reference. Reference mappability is represented by the outermost grey histogram. Red histograms represent the coverage of *G. lygdamus* samples across the *G. alexis* reference in windows of 1mbp. The sample represented are, from outer to inner rings: L002, L004, L006, L008, L011, L012 and L013.

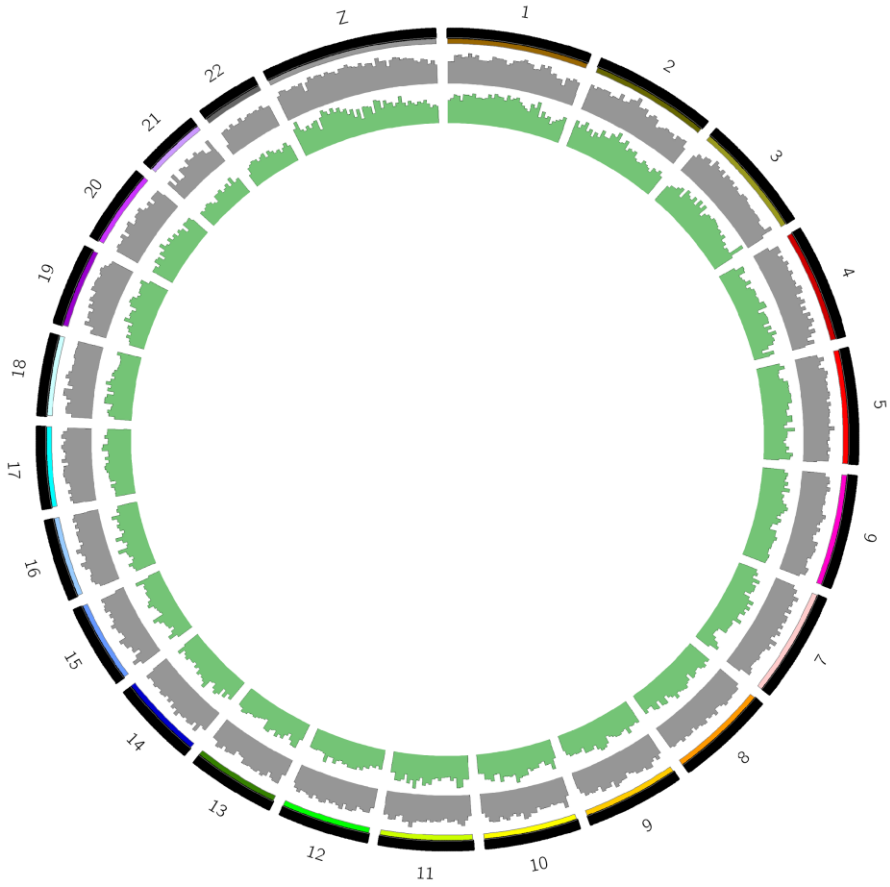


Fig. S4C. Coverage distribution of *G. lygdamus* samples across the reference. Reference mappability is represented by the outermost grey histogram. Green histograms represent the coverage of *G. lygdamus* samples across the *G. alexis* reference in windows of 1mbp. The sample represented are, from outer to inner rings: RVcoll10-B005 (Canadian specimen).

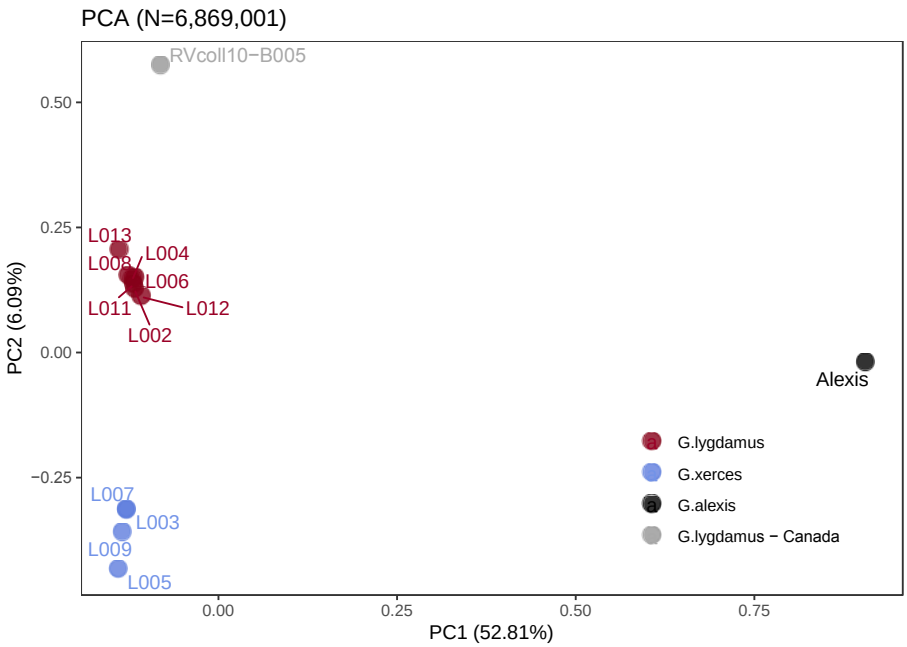


Fig. S5. Plotting of PC1 and PC2 of the Principal Component Analysis (PCA). The PCA was generated with nuclear DNA data (N=6,682,591 SNPs) from eleven historical butterfly specimens (4 *G. xerces* and 7 *G. lygdamus*), a modern *G. lygdamus* from Canada (RVcoll10-B005) and a modern *G. alexis* reference genome. The PCA shows a clear separation of both historical species and the reference in the first PC (explaining 52.81% of the variance), and separation of *G. xerces* and *G. lygdamus* by the second PC (explaining 6.09% of the variance), supporting they are separated lineages.

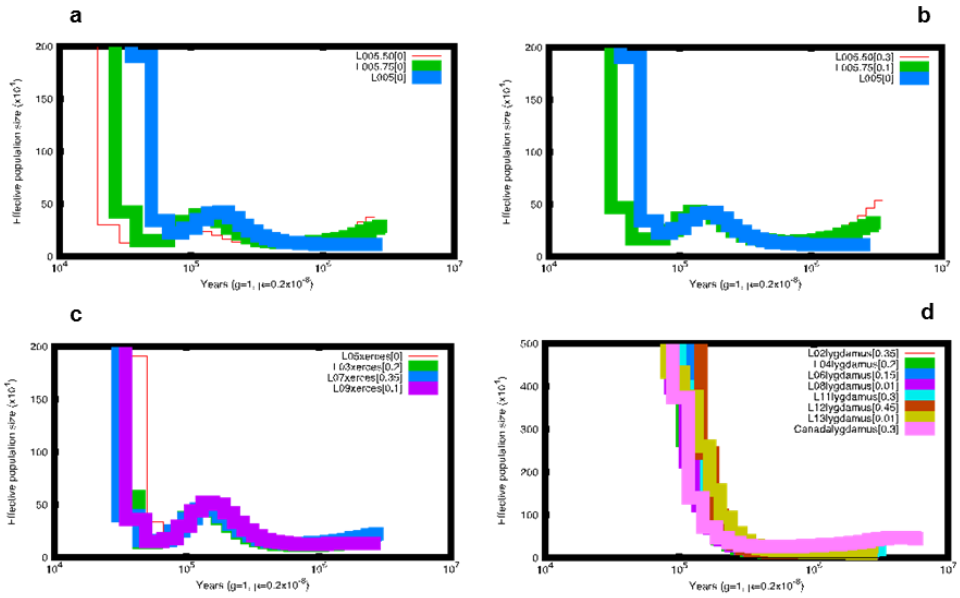


Fig. S6. Pairwise Sequentially Makovian Coalescent (PSMC) plots of Xerces Blue and Silvery Blue. a) PSMC of Xerces Blue L05 down sampled to half (red) and 75% of coverage (green). b) PSMC of down sampled Xerces Blue L05 corrected. Lower coverage results in underestimation of heterozygote site and thus lower historical effective population sizes. This situation can be corrected assuming a False Negative Rate (FNR) by visually adjusting the curves using the psmc_plot.py program from the PSMC package. c) PSMC of Xerces Blue L03, L05, L07 and L09 corrected assuming FNR. d) PSMC of historical Silvery Blue L02, L04, L006, L08, L11, L12 and L13 and modern Silvery Blue from Canada (RVcoll10-B005) corrected assuming FNR. Despite current differences in coverage, individuals from each species follow the same trajectory.

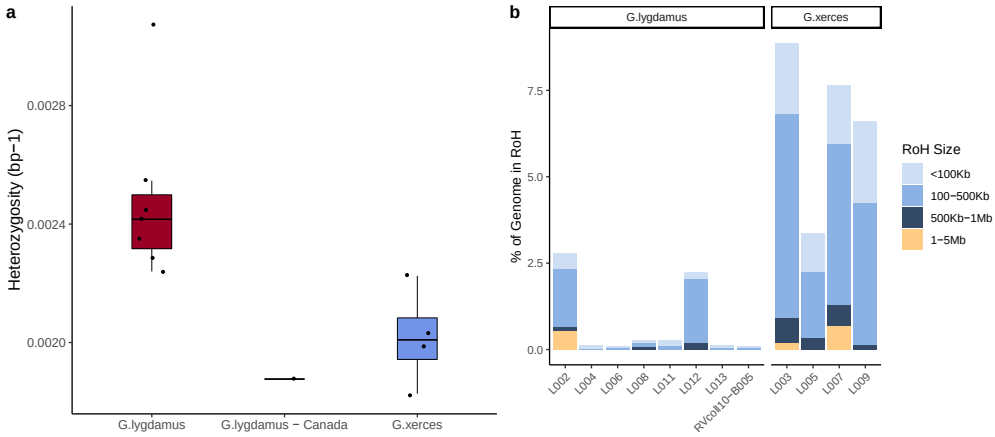


Fig. S7. a) Heterozygosity estimates for low coverage data from both Silvery Blue (historical and modern) and Xerces Blue butterflies. b) ROHs for each individual classified by length.

Genome #	Genus	Species	Subspecies	State	Locality	Date	Collection
USNMENT101413	Glaucopsyche	xerces		California	San Francisco		Barnes
USNMENT101402	Glaucopsyche	xerces		California	San Francisco	April 16-23	Barnes
USNMENT101441	Glaucopsyche	xerces		California	San Francisco		Barnes
USNMENT101406	Glaucopsyche	xerces		California	San Francisco		Barnes
USNMENT101434	Glaucopsyche	xerces		California	San Francisco	April 16-23	Barnes
USNMENT00181297	Glaucopsyche	lygdamus	incognitus	California	Marin County		Barnes
USNMENT00181298	Glaucopsyche	lygdamus	incognitus	California	Fairfax	27-maig-32	Wm D Field
USNMENT00181299	Glaucopsyche	lygdamus	incognitus	California	Oakland	14-abr-48	Graham Heid
USNMENT00181300	Glaucopsyche	lygdamus	incognitus	California	San Jose	27-març-64	P. Opler
USNMENT00181301	Glaucopsyche	lygdamus	incognitus	California	Haywood City	1-maig-31	Wm D Field
USNMENT00181302	Glaucopsyche	lygdamus	incognitus	California	Santa Cruz	1-abr-32	JW Tilden/Field
USNMENT00181303	Glaucopsyche	lygdamus	incognitus	California	Santa Cruz	8-abr-27	GW Rawson

Table S1. List of historical specimens analysed in this study.

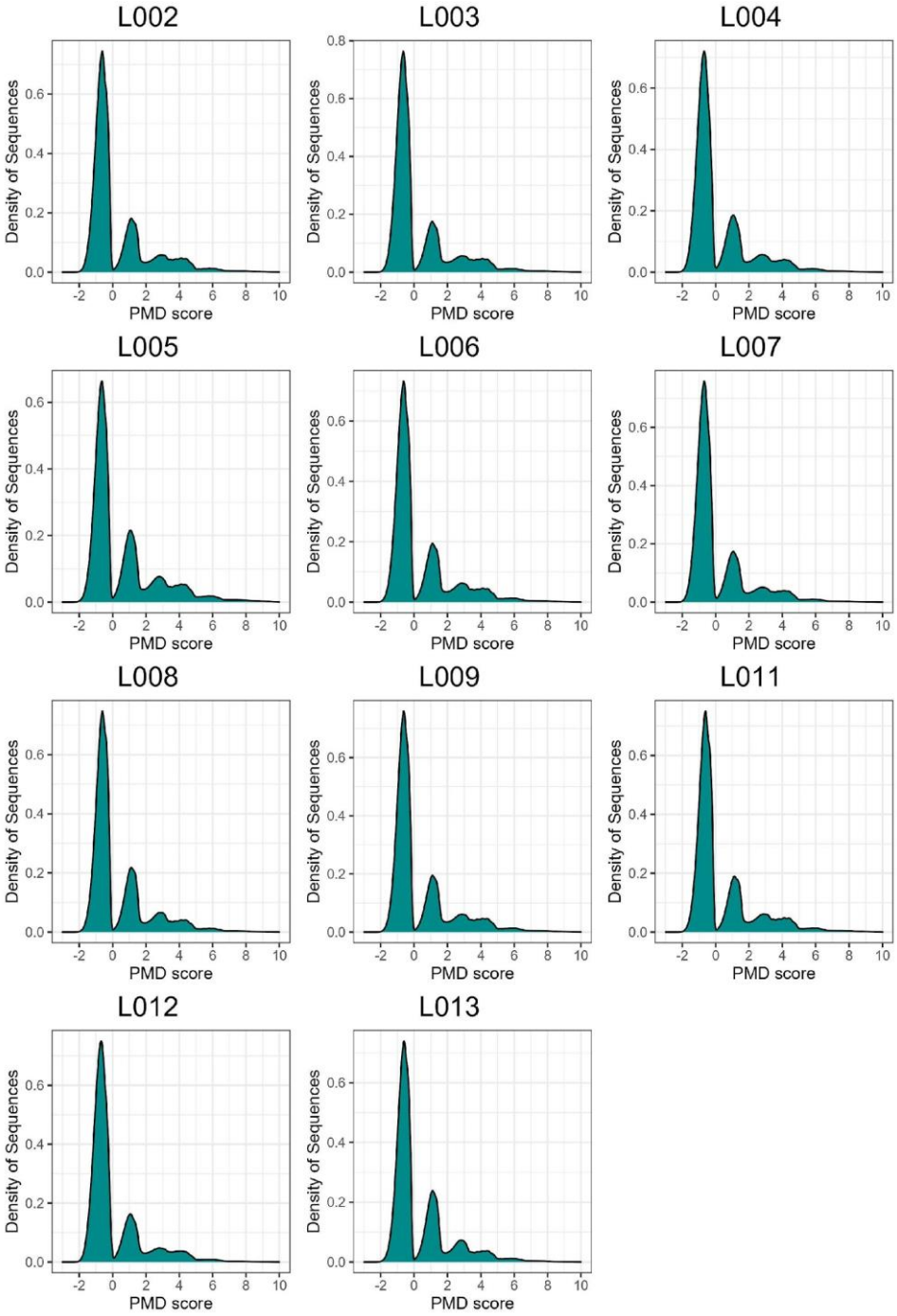


Fig. S8. PMD score distribution of the analysed samples.

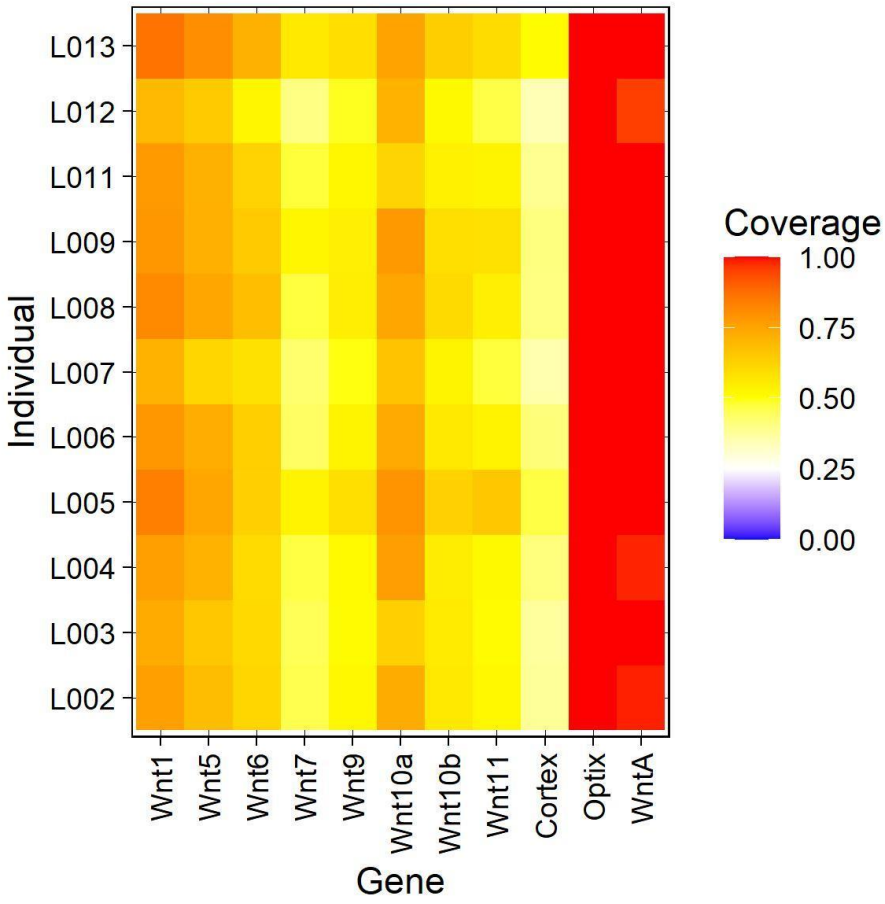


Fig. S9. Coverage Heatmap of Genes associated to change in butterfly colour patterns. Positions were considered as covered if at least one read was overlapping with them.

5. DISCUSSION

5.1. Carolina parakeet (*Conuropsis carolinensis*)

5.1.1. Causes of extinction

The extinction of the Carolina parakeet is shrouded in mystery and speculation. Historical accounts are sparse, scattered, not very detailed, more literary than scientific, misquoted, and sometimes even contradictory (McKinley, 1979). We will probably never know exactly what happened; even so, we do not have to give up on formulating reasonable hypotheses that bring us closer to the truth. It is often said that the study of the past allows us to avoid mistakes in the future. But as human beings are the only animals that trip over the same stone twice, we can also reverse the approach and consider that the current problem of parrot conservation is a reflection of the problem that existed some centuries ago. By seeing what causes the extinction of parrots in the present, we will understand what made them extinct in the past.

According to the International Union for Conservation of Nature (IUCN) criteria, of the world's 398 extant parrot species (Psittaciformes), 111 (28%) are classified as threatened (Olah et al., 2016). The psittacine group is one of the most vulnerable in the order of birds. Its great specialization favors that its status deteriorates easily. It is estimated that between 1600 and 1980, 18 taxa became extinct (Collar, 1997).

Current causes of decline and extinction are: capture to supply the exotic pet trade, reduced natural distribution, and habitat loss.

Today the value of a parrot sold as a pet is much higher than any other exploitation (although as always the benefit is poorly distributed; about $5 per captured parrot is paid to local populations, while this same animal is sold to the final buyer in Europe or in the USA for $500). Two centuries ago, there was virtually no trade in exotic species, so the Carolina parakeet was hunted indiscriminately. The current catch would be conceptually equivalent to the hunt at that time.

People shot Carolina parakeets for sport, food, feathers, scientific collections, and to protect crops (Snyder & Russell 2002). Beyond anecdotes about widespread shooting, like Audubon's account (1831) and Gedney's (1877) observation of hundreds of imported individuals for the pet trade, there is no direct evidence to suggest overexploitation as the main driver of the Carolina parakeet's decline; in fact, McKinley (1980) points out that populations may have already declined substantially before trapping and shooting began in earnest (Burgio et al., 2021).

Instead of overexploitation, both McKinley (1980) and Snyder (2004) suggested disease as the most likely factor leading to the Carolina parakeet's extinction, despite there being only anecdotal evidence of parrots exhibiting any symptoms associated with illness, and most of these observations were of captive birds (Burgio et al., 2021). Snyder and Russell (2002) suggested that Carolina parakeets' affinity for foraging on farms may have brought them into close contact with domesticated poultry, which may have contributed to disease transmission. However, there are opposing arguments. First, South America has also been exposed to domestic poultry and no epidemic has been detected among the numerous and diverse populations of parrots, nor any extinction for this reason. And second, the analysis of the genome of the Carolina parakeet provided by this thesis has not detected any pathogen, although it is not ruled out that it was a specimen bred in captivity, and therefore, isolated from a possible epidemic.

The second cause of extinction among parrots is a reduced natural distribution. Of the 18 taxa extinct between 1600 and 1980, 16 were island endemic species. 68% of currently threatened parrots have restricted distributions (Collar, 1997). As is well known, island animals evolve in predator-free environments and are especially exposed to introduced species such as dogs, cats, or rats, aside, of course, from direct hunting/capture by humans. The most notable historical cases of island extinctions are the Antilles and the Pacific islands.

In the Antilles the following species became extinct: the Cuban macaw (*Ara tricolor*), the Guadeloupe amazon (*Amazona violacea*), the Guadeloupe parakeet (*Psittacara labati*), the Jamaican red

macaw (*Ara gossei*), the Lesser Antillean macaw (*Ara guadeloupensis*), the Martinique amazon (*Amazona martinicana*), the Martinique macaw (*Ara martinicus*), the Puerto Rican parakeet (*Psittacara maugei*), and the St. Croix macaw (*Ara autocthones*). A few naturalized specimens of some of these species are preserved in museums, or fossil remains have been found, but many are only known from the descriptions made by the first Europeans (Clark, 1905; Williams & Steadman, 2001).

It is also explained that the Polynesians ventured on their transoceanic voyages spurred to find parrots and capture (extinct) them for their feathers, and introduced some species beyond their natural range such as the Kuhl's lorikeet (*Vini kuhlii*) in the Cook Islands (Collar, 1997).

The only two species of parrots extinct in recent centuries that had wide continental distributions are the Glaucous Macaw (*Anodorhynchus glaucus*), and the Carolina parakeet. Unfortunately, this list will be increased with the foreseeable and imminent extinction of the Spix macaw (*Cyanopsitta spixii*). All three species maintain a remarkable parallelism. Glaucous Macaw was native to the palm groves located between north Argentina, south Paraguay, Bolivia, northeast Uruguay, and Brazil, and went extinct in the 19th century, likely due to the loss of habitat and pet-trade demands from Europe (Yamashita & Valle, 1993). Spix macaw is associated with mature stands of caraiba (*Tabebuia caraiba*), a tree that grew on the banks of the São Francisco River, and its minor tributaries (Brazil), until massive logging almost made it disappear, in addition, there was also a strong capture pressure to supply the demand of collectors (Collar, 1997). The Carolina parakeet was also targeted for collecting in the final phase of its decline, and was also closely associated with a particular habitat.

First range maps for this species were generated by drawing a polygon encompassing all of the most distant areas in which the Carolina parakeet had been reported (Hasbrouck, 1891; Snyder & Russell, 2002). The result was an area that practically occupied the eastern half of the USA. Peers et al., (2016), modeled the distribution based on historic presence from freely accessible databases and published material, and related presence with

historic climate data (period 1901-1930). And Burgio et al., (2017) carried out further distribution modeling based on a larger data set, filtered out suspected vagrants and incorrect or highly uncertain georeferences, and related them to a larger climate data set (period 1895-1924). The results were increasingly narrower distributions. Surprisingly, however, these models do not take into account the preferred habitat of the species (Figure 5.1, a). When we look at the raw data that fed these models, located on the map, we find that both sightings and collections of parakeets are associated with river courses and wetlands (Burgio et al., 2018) (Figure 5.1, b). In principle, the association with the river courses could be explained in two ways: either the American explorers moved preferentially through the rivers and, therefore, only sighted the parakeets that accidentally flew through that area, or the preferred habitat of the parakeets was the riverside vegetation. The dilemma can be resolved on the basis of the specimens collected in Florida, a terrain made up mostly of wetlands, difficult for explorers to access. If the parakeet preferred wetlands, it is reasonable to think that it also preferred riverbanks covered by the same vegetation. And how can we explain the sightings far from rivers and wetlands?

In general, the shape of a bird's wings indicates its type of life. Birds with long, narrow wings tend to fly long distances, while birds with short, rounded wings tend to fly shorter distances. The Carolina parakeet had proportionately long and narrow wings, so it could leave the safety of the forest to feed on human crops, move regionally in search of fruit, and even migrate seasonally to avoid the harsh winters of Midwest (McKinley, 1977; Burgio et al., 2017). Sightings under these conditions should be considered migrants or vagrants outside preferred habitat.

The dominant tree in the riparian vegetation and wetlands of the American Southeast at elevations <30 m above sea level is the bald cypress (*Taxodium distichum*) (Wilhite & Toliver, 1990). Carolina parakeets were frequently associated with this tree, as an important food source and nesting (Snyder & Russell, 2002), since older trees (centenarians or millennials) were eventually infected by fungi and developed holes that birds used for roosting and

breeding. Overall, the native range of the bald cypress overlaps almost completely with the range of eastern subspecies of Carolina parakeet (*C. c. carolinensis*) and the year-round portion of the range of western subspecies (*C. c. ludovicianus*) (Burgio et al., 2021) (Figure 5.2, a-e). However, the distribution of bald cypress is highly discontinuous in all areas reflecting the presence of suitable riparian or wetland habitat. Therefore, in reality the original distribution of the parakeet was more restricted than estimated or modelled. Neither the parakeet, nor Glaucus macaw, nor Spix macaw originally had the large ranges that would be expected for a continental species (although relative to the distribution of island parrots it is at least an order of magnitude larger). Consequently, they were more vulnerable than previously thought.

The last cause of extinction of parrots is the loss of habitat. The bald cypress forest is one of the most heavily altered forest cover types in the United States. Mattoon (1915) in a monumental work describes the entire timber industry developed around the bald cypress and quantifies the thousands of tons that were extracted for the construction of houses. The consequence is that currently only three virgin bald cypress forests with commercially valuable timber are known to remain: the National Audubon Society sanctuaries at Four Holes Swamp, South Carolina, and Corkscrew Swamp, Florida, along with a private parcel of old growth bald cypress in southwestern Arkansas. These three tracts only total 24 km^2 out of the 162.000 km^2 of virgin bald cypress forest estimated to have originally existed in the South (Stahle et al., 2006).

The Carolina parakeet was not a feeding specialist, but showed a predilection for cockleburs (*Xanthium strumarium*) (Snyder & Russell, 2002), which contain a lethal toxin inhibiting mitochondrial energy production. In particular it inhibits four mitochondrial ATP transporters. In this thesis we explored the genes coding for these four proteins (SLC25A4, SLC25A5, SLC25A6 and SLC25A31) and detected mutations involving significant structural changes in two of them (see section 4.1). Therefore, it is possible that these mutations conferred the species with a unique adaptive mechanism for dealing with the toxin present in its diet.

With the arrival of the Europeans, and the progressive deforestation, the Carolina parakeet was forced to expand its diet and in addition to consuming weed and tree seed, it included cultivated fruits, which made it odious to farmers. Therefore, the reduction of habitat should not affect its food resources. On the contrary, without old trees, the bird lost roosting and breeding sites and this must have caused a population decline. It has been said that by sharing habitat with the Ivory woodpecker (*Campephilus principalis*), which survived fast longer, it has been concluded that habitat loss cannot be the sole cause of the parakeet's extinction. Surely it was not, but this argument ignores the fact that the woodpecker drills its own nests while the parakeet cannot. Nor does it have the alternative of other parakeets, such as *Aratinga canicularis*, of burrowing its own nests in large tropical termite mounds (soft soil), when there are not enough trees with holes (Collar, 1997). However, it appears that it may also have been negatively affected by another insect. The honey bee (*Apis mellifera*), introduced by Europeans, often uses holes in trees to establish a new colony, therefore, a competition for space aggravated by deforestation should be established (McKinley 1980; Snyder & Russell 2002). It has been observed that the densities of some species of tropical parrots are twice as high in primary forest as in secondary forest, and twice as high in secondary forest as in converted agricultural areas (Collar, 1997). Perhaps the Carolina parakeet followed the same progression.

Once the decline started, what signals should it leave in the DNA? The analysis of the genome of the Carolina parakeet carried out in this thesis has not detected significant signs of genomic erosion. Neither heterosygosity nor the distribution of long runs of homozygosity (RoHs) show values far from the mean of the birds. This suggests that the population decline was fast enough to leave no trace in the genome. It would be interesting to extend the genomic erosion analysis to more psittacine species, and especially to include Glaucus macaw and Spix macaw, whose fates have paralleled those of the Carolina parakeet. However, some studies indicate that the genetic diversity of a species is poorly predicted by its population size (Leffler et al., 2012), and others go

even further and affirm that, in certain circumstances such as in the case of the passenger pigeon (*Ectopistes migratorius*), analyses such as the pairwise sequentially Markovian coalescent (PSMC) are unlikely to reliably inform us of demographic history (Murray et al., 2017), contrary to what has been previously established for the same species (Hung et al., 2014).

In this thesis we also used PSMC algorithm to evaluate the past demographic evolution of *Conuropsis* and *Aratinga* species. We found that the Carolina parakeet population experienced an increase in effective population size (Ne) during the Middle Pleistocene, followed by demographic fluctuations that started during the Last Glacial Period (~110 kya) and a subsequent population decline that continued until recent times. In contrast, the PSMC of the endangered *Aratinga solstitialis* shows a stronger and continuous population decline and a longer period of lower effective population size than *Conuropsis*.

Unfortunately, the resolution of this approach is insufficient on a time scale of less than 10,000 years, and therefore does not allow us to detect human influence on extinction. But we have an alternative way: dendrochronology. The bald cypress, which is the habitat type of the Carolina parakeet, forms, like any conifer, growth rings that can be correlated with annual climatic variables, and it is also one of the longest-lived trees in the world, which allows long time series to be obtained. For this reason it has been intensively used to perform paleoclimatic reconstructions (Stahle et al., 1988; 2012; Stahle & Hawks, 2016). When we analyze the climate reconstruction for the southeastern U.S. over the last 1600 years, we see that there has been no trend attributable to sustained climate change (Stahle & Hawks, 2016; Figure 5.2, f-h). On the contrary, alternating cycles of droughts and rainy periods are observed every 30 years. Consequently, climatic variables during the last millennium do not explain the decline of the Carolina parakeet, and reinforce the importance of anthropogenic causes of extinction. These causes would be hunting and, above all, habitat destruction acting synergistically in a population with a smaller original distribution than previously thought, although stochastic

events could become important when the population went down a certain level.

Additionally, an intrinsic factor in parrots that makes them vulnerable to extinction is that they generally are K-strategists, characterized by low fecundity, slow growth, late maturation and longevity.

Low fecundity is evident in the fact that only a fraction of sexually mature adults breed. In the Monk parakeet (*Myiopsitta monachus*) only 37-60% of adults breed each year, in macaws (*Ara sp*) only 1/5 of adults breed each year. Although there are exceptions among the small parrots, such as the lovebirds *(Agapornis sp*), which mature sexually at 80 days (37 days after fledging), or the budgerigars (*Melopsittacus undulatus*), which mature at 6 months, they can breed throughout the year, and lay two clutches per year (Collar, 1997).

Slow growth implies a long learning process for young individuals, accompanying and observing adults (macaws, for example, need years of learning in order to survive). Hence, many parrots flock together and maintain close social relationships reinforced by allopreening behaviors and complex vocalizations (Martella & Bucher, 1990). Some parrots flock in single species flocks (African Greys) and other birds operate within mixed species flocks (Amazons, Conures and Macaws). Flocks are essential for survival, since they protect their members from predators and facilitate the location of food. As a species begins a population decline, the flocks become smaller and smaller, which has a feedback effect: survival is reduced and in turn the flock size is reduced. In the final stages of a parrot's extinction, only pairs or solitary individuals are observed.

It has been pointed out that K-strategists, together with the arboreal life, the extraordinary combined manipulative ability of legs, beak and tongue, their vocal repertoire, the age-hierarchical social structure and their remarkable cognitive abilities are a noteworthy example of evolutionary convergence with primates (Collar, 1997). Perhaps that is why we humans have connected so well with them. We have had parrots as pets for 2000 years and are fascinated by their beauty. Etymologically, "parrot" comes from the French given

name Pierre and its diminutives Pierrot or Perrot. In the Catalan and Spanish we also find the equivalent term "perico". In other words, we have established a sufficiently strong link to call them as one of our own. In addition, they bring us important therapeutic benefits since, as with dogs, we can communicate verbally.

Although the Carolina parakeet can no longer speak, we should listen to its message and not repeat the same mistakes with the magnificent parrot species that still inhabit the planet.

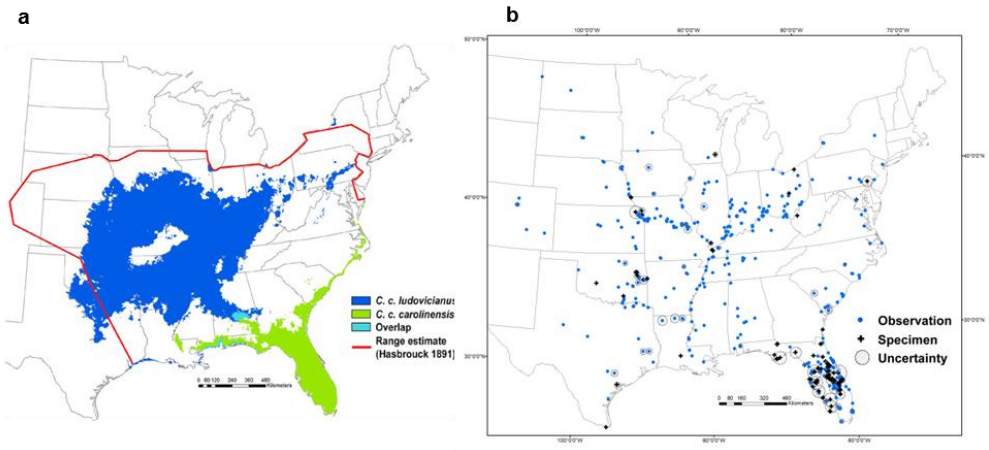


Figure 5.1. (a) Estimated original distribution of Carolina parakeet (*Conuropsis carolinensis*) generated by drawing a polygon encompassing all of the bird historical reports (red outline), and by modelling filtered historical and climatic variables sorted in western (*C. c. ludovicianus*; blue) and eastern subspecies (*C. c. carolinensis*; green). (b) Historic sightings of the Carolina parakeet (blue dots) and specimen collection (black crosses) in the period 1564-1944. Circles with diagonal lines show estimates of uncertainty associated with each point. Note that most of specimens were taken from Florida, and most of the sightings were associated to river courses, indicative of the birds' dependence on riparian and wetland vegetation. Therefore, the real distribution was presumably much smaller than any estimation [modified from Burgio et al., 2017 (a); 2018 (b)].

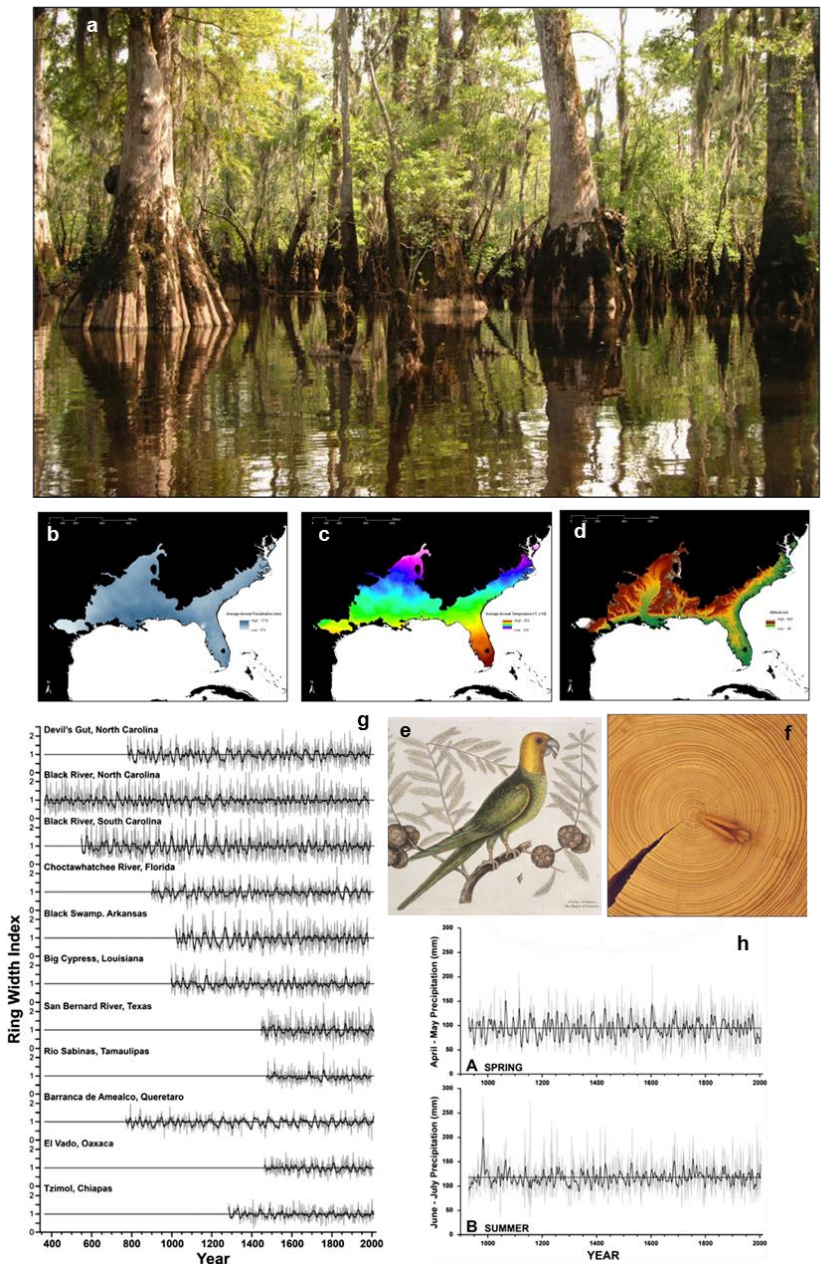


Figure 5.2. (a) Pristine forest of old-growth bald cypress (*Taxodium distichum*) at Black River, North Carolina (USA). (b) Average annual precipitation (mm), (d) annual temperature (ºC) and altitude (m) along the distribution area of bald cypress. This range overlaps with the estimated range of Carolina parakeet (*Conuropsis carolinensis*). (e) Illustration of Carolina parakeet by Mark Catesby (ca 1722-1726), eating seeds of bald cypress. This draw was used by Carl Linnaeus to classify the species. (f) The annual rings of bald cypress can be exactly dated and used to infer past climate conditions. (g) Bald cypress chronologies from the south-eastern United States (top) and Mexico (bottom). The time series represent nonclimatic detrended and standardized annual values of total ring width (gray) and a smoothed version highlighting 20-year (black). Note that specimen from Black river is 1600 years old. (h) Spring and summer precipitation reconstructions (gray) for the south-eastern United States, both include a smoothed version to highlight decadal variability (black). No climatic trend is observed in the studied period [from Stahle & Hawks, 2016 (a, e); Davis, 2016 (b-d); Stahle et al., 2012 (f-g)].

5.1.2. Biogeographic puzzle

The progress of science often makes it possible to resolve long-debated questions, but at the same time raises new ones. Sometimes, it also reopens issues that seemed to be solved.

After having been classified by Linnaeus in the monophyletic genus *Conuropsis*, most specialists who studied the phylogeny of the Carolina parakeet suggested that it was closely related to the conures of the genus *Aratinga*. They based this inference on shared morphological traits, such as a long, pointed tail and wings, feathered cheeks and lores, and a comparatively broad and heavy bill (Forshaw, 1989; Snyder, 2004).

The first molecular phylogeny of the Carolina parakeet based on mtDNA confirmed, years later, this hypothesis (Kirchman et al., 2012) and established that the sister taxa were *Aratinga auricapillus*, *A. solstitialis* and *Nandayus nenday*, but raised a new problem: the genus *Aratinga* appeared as polyphyletic and the authors recommended revising its taxonomy. Numerous subsequent molecular studies have proposed different taxonomies and phylogenies (Remsen et al., 2013; Schweizer et al., 2014). Some authors even showed that molecular phylogenies in this clade vary markedly depending on the methods and mitochondrial markers used, and suggested that this uncertainty may arise from taxon sampling, rapid evolution of the taxa under study and a short time interval between the divergences of particular lineages (Urantowka et al., 2017).

One of the results of this thesis has been the recovery of 93% of the nuclear genome and 100% of the mtDNA of *C. carolinensis*, with an average depth of 13.4x and 150x, respectively. This has allowed the elaboration of two phylogenetic trees. In the first, we used 4,988 nuclear loci (ultraconserved elements [UCEs], comprising 9,864,148 bp) extracted from the genomes of *C. carolinensis*, 17 extant parrots, and the rifleman *Acanthisitta chloris* (Passeriformes) as an outgroup, in order to investigate the phylogenetic placement of *C. carolinensis* within Psittaciformes. In the second tree, we used the complete coding region of the mtDNA genome sequence to investigate the placement of *C. carolinensis*

against a greater sampling within de la tribu Arini. In addition, we were able to estimate, for the first time, the divergence time of *C. carolinensis* in each of the two trees.

The nuclear phylogenetic tree obtained provides an interesting synopsis of the situation of *C. carolinensis* within the Psittaciformes order and does not raise doubts (Figure 1, section 4.1). In contrast, the mitochondrial phylogenetic tree provides a new result that does not quite fit in detail with the other phylogenies of the Arini group (Figure S2, section 4.1). It also confirms that the species most closely related to *C. carolinensis* is *Aratinga solstitialis*. This tropical species is currently located in an area between Guyana, Suriname and northern Brazil (Collar, 1997), and is therefore much more distant geographically from the original distribution of *C. carolinensis* than other members of the Arini, like the genus *Psittacara*, whose species *Psittacara holochlora* was parapatric with *C. carolinensis*.

Molecular clock analysis employing two fossil calibrations (Oliveros et al., 2019) suggests that the *Aratinga-Conuropsis* split occurred around 2.8 Ma (1.6- 4.4; 95% highest posterior density [HPD] interval) from nuclear genome data and around 3.8 Ma (2.73–5.05; 95% HPD interval) from mtDNA data. We therefore have an uncertainty range of 1 million years.

Consequently, a challenging biogeographical puzzle is posed: (1) what was the relationship between *Conuropsis carolinensis* and its geographical neighbours, the Central American psittacines? and (2) when did the dispersal to North America actually occur?

When was the dispersal to North America

Few fossil records of *Conuropsis carolinensis* exist. Several bones from archeological sites in central Illinois were identified (Parmalee, 1958; 1967), as well as three bones from an Amerindian burial site in south-western Ontario (Prevec, 1984), both corresponding to the Holocene. In addition, an humerus was found at Dickerson Coquina Pit located in Saint Lucie County, Florida (Kilmer &. Steadman, 2016), corresponding to middle Pleistocene (late Irvingtonian Land Mammal Age ~1.9/1.72 to 0.15 Ma). Finally, an humerus from the upper Miocene of Nebraska was classified as a

new species *Conuropsis fratercula* (Wetmore, 1926). Although the validity of this generic placement is questionable (Olson, 1985), its placement within the Psittacidae was not (Kirchman et al., 2012), but it should be, given the abundant molecular evidence (Schweizer et al., 2010; Wright et al., 2008).

It would be convenient to try to recover aDNA from these fossils to establish their phylogenetic relationships with the historical samples of *Conuropsis carolinensis* and to obtain a more precise date of divergence. Absolute dates could also be carried out with C^{14}. In any case, the date is not critical to understand the process of dispersion towards North America. 3 Ma it is the widely recognized date for the final closure of the Panama Isthmus (O'Dea et al., 2016). Despite the dispersal could have occurred after the North and South American landmasses were continuous, it could also have happened earlier. As it has seen, the evolution of many psittacids, including the Arini, is marked by radiations and trans-oceanic dispersals between different subcontinents of Gondwana (Schweizer et al., 2010).

What was the relationship with the neighboring psittacines

It has been established that the origin of the Psittaciformes is located in Australasia during the Cretaceous (Wright et al., 2008). The absence of primates on that continent allowed their ecological niche as consumers of fruit and tree seeds to be occupied by a great diversity of bats, arboreal marsupials and birds, including parrots (Collar, 1997). Subsequently, continental drift isolated the large groups of parrots and the Arinae subfamily evolved between Antarctica and South America (Wright et al., 2008).

When we analyze the current diversity of parrots in South America we see that the highest values are concentrated in the Amazon basin and decrease with latitude (Figure 5.3, a).The staggering contrast in biotic diversity between equatorial and polar latitudes is one of Earth's most salient biological characteristics. Although this phenomenon has been recognized since the 19th century, the proximate and ultimate causes of species richness gradients continue to galvanize scientific debate and drive hypothesis testing in macroecology and biogeography. Research efforts during the

past decade have winnowed the number of potential hypotheses to a few: (1) energy availability (evapotranspiration); (2) evolutionary time; (3) habitat heterogeneity; (4) area (tropical regions had a greater geographical extent until relatively recently; 30-40 million years ago, when temperate zones increased in size); and (5) geometric constraints (Wiens & Donoghue, 2004).

Rahbek and Graves (2001) examined bird diversity in South America and found that the areas that exhibited the highest avian diversity (>650 species) were restricted to Andean Ecuador, south-eastern Peru and southern Bolivia. Species richness in neotropical birds seems to be linked directly to habitat diversity, which is correlated with topographic heterogeneity. The species-poor zone in central Amazonia overlapped 5-16 distinctive ecosystems, whereas species-rich zone in Andean-Amazonia overlapped 16-24 ecosystems. Parrots show the same pattern of maximum diversity associated with the Andean-Amazon borders, but also associated to lowland jungle with the least (relative) diversity of ecosystems. This suggests that there should be other mechanisms that contribute to their diversification (Figure 5.3, a).

Burgio et al., (2022), modelled the specific, functional, and phylogenetic diversity of psittacines in South America to develop an Integrated Biodiversity Index (IBI), in order to identify global biodiversity hotspots for parrots, and improve its conservation. Functional traits in the analysis include: body size and shape characteristics, diet, foraging strategy, foraging location and range size. When we compare the specific diversity (S) with the functional diversity of parrots we see that there is no coincidence (Figure 5.3, b-f).

The conclusion is that beyond the ecosystem diversity analyzed by Rahbek and Graves (2001), and the functional (adaptive) diversity analyzed by Burgio *et al*., (2022), there must be other selection mechanisms that also contribute to the diversity of parrots.

One of these mechanisms is the clan behaviour of most parrots, and their ability to recognize flock members. Behavioural and socio-cultural traits are recognized in the restriction of gene flow in species with high cognitive capacity and complex societies. For example, a study on red-fronted macaw (*Ara rubrogenys*) from

Bolivia, found no evidence of geographical and ecological barriers, owing to the high dispersal ability, nesting and foraging habits between genetic clusters. The lack of genetic intermixing despite long-distance foraging and seasonal movements suggests recruitment in natal colonies and other social factors reinforcing philopatry-related genetic structure (Blanco et al., 2021).

This result was also found in passerines. A study used mate-choice experiments in siskins (*Carduelis spinus*) to analyse how familiarity and patterns of ornamentation (i.e. the size of wing patches) interact to influence mating success. Results show that females clearly prefer familiar individuals when choosing between familiar and unfamiliar males with similar-sized wing patches. Furthermore, when females were given the choice between a highly ornamented unfamiliar male and a less ornamented familiar male, half of the females still preferred the socially familiar birds as mates (Senar et al., 2013).

An analogy with humans could even be drawn. A study examines unions between individuals with non-Western immigrant origins and those from the native majorities in six North American and Western European countries. The analysis shows that certain deep social cleavages, involving African ancestry in the United States and Muslim religion in Western Europe, hinder the formation of mixed unions (Alba & Foner, 2015). Culture and familiarity are invisible barriers that contribute to the isolation of populations and favour sympatric speciation.

And what are the genetic consequences of these barriers? According to Joly (2011), the existence of a certain species is the result on a metastable equilibrium between inbreeding and outbreeding. A moderate degree of inbreeding ensures the expression and further fixation of recessive genes with adaptive value, reduces the recombination load, fight off the accumulation of recessive deleterious mutations, reduces the cost of sex and promotes population fragmentation, which can, in turn, promote collaborative or altruistic behaviour.

This dynamic of inbreeding and outbreeding could leave certain signals in the genome of some psittacids, such as conures (†*Conuropsis*, *Aratinga, Enicognathus, Eupsittula, Guaruba,*

Leptosittaca, Ognorhynchus, Psittacara, Pyrrhura and *Thectocercus*) hindering the taxonomy and molecular phylogeny based on a few genetic markers. At the same time, highlights the importance of carrying out new phylogenies based on a greater number of genetic markers to definitively clarify the taxonomy and evolution of this group.

Remarkably, the key to solving the puzzle of Carolina parakeet biogeography is provided by a popular proverb: *Birds of feather flock together*.

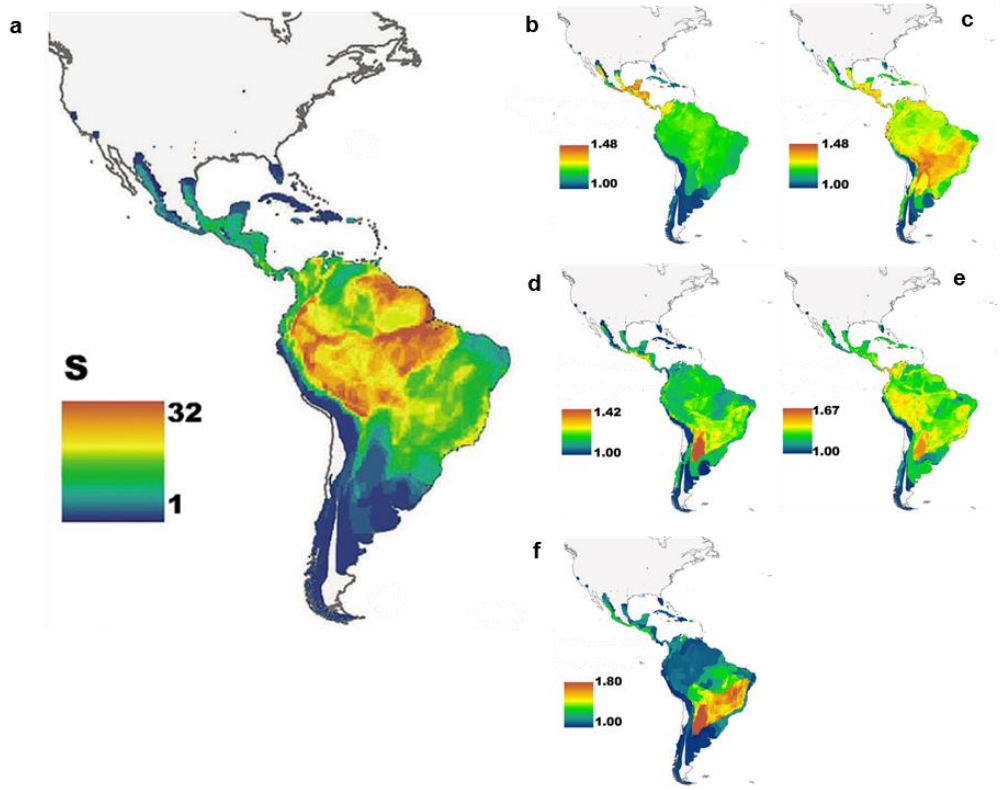


Figure 5.3. (a) America map of Psittaciforms (parrots) species richness (S). Note that if the distribution of the Carolina parakeet (*Conuropsis carolinensis*) was included, the species richness between both ends of the continent would be more symmetrical; (b-f) America map of functional diversity (Rao's Q, based on Hill numbers). Functional traits in the analysis include: (b) body size and shape characteristics, (c) diet, (d) foraging strategy, (e) foraging location and (f) range size. Note that none of the functional traits seems to be associated with species richness, and except for body size and shape, the rest of the traits show some kind of association among them [modified from Burgio et al., 2022].

5.2. Tenerife giant rat (*Canariomys bravoi*)

5.2.1. Causes of extinction

The Tenerife giant rat (*Canariomys bravoi*) lived adapted to its small stable world and literally isolated from the rest of the world, until it became extinct. When an extinction occurs, it is reasonable to think that there has been a change. If nothing changes, there is no reason for a species that had been living on the island for 650,000 years to become extinct.

The first step in finding out the cause of an extinction is determining when it occurred. Once the moment is known, it is necessary to determine what happened at that moment, or some time before, since no matter how fast an extinction may be, it is never instantaneous. When we have a correlation between a date and an event we can establish causality.

But this procedure is usually more complex than it seems. To begin with, it is not easy to determine the exact date of extinction of a species. The various ^{14}C dates of the *C. bravoi* bones indicate a wide range. Michaux et al., (1996) obtains the first dating of 12,230 ± 140 years BP, which calibrated are 12,880-11,810 cal BC, Bocherens et al., (2006) obtains values between 1397-1118 cal BC and 4826- 4552 cal BC, Rando et al., (2014) obtains values from 400-231 cal BC to 4338-4172 Cal BC and Crowley et al., (2019) obtains values between 16,965 Cal BC and 2315 Cal BC. That is, the samples are between 2,400 and 19,000 years old, so extinction was therefore a later event than the year 400 cal BC.

What happened after that year? The two usual hypotheses that are contrasted to explain extinctions are a climatic modification with ecological consequences, or the arrival of humans on the island.

Climate hypothesis

Pollen analysis at Lake Tigalmamine (1650 m asl in the Middle Atlas) in Morocco (Cheddadi et al., 1998) shows a strong shift in environmental conditions between the Middle and Late Holocene. This climatic shift is characterized by a lowering of temperatures,

especially the winter ones. Furthermore, a core of marine sediment located off the African coast (Cap Blanc, Mauritania) shows a strong increase in aeolian dust during the Holocene since ca 6000–5000 BP, as a result of the rapid aridization of the Sahara (deMenocal et al., 2000).

Therefore, there was indeed a major climate change in the Canary Islands region. The analysis of the stable isotopes (δ13C & δ15N) of the remains of *Canariomys* show a variation of the values during this period compatible with a change in environmental temperature. It shouldn't be critical though, since the rat still survived 2000 or 3000 years more (Bocherens et al., 2006).

Human hypothesis

Discarded the climatic hypothesis, we analyze the human one. It is also not easy to verify, since there is considerable uncertainty in the estimation of the arrival of humans in the Canary Islands. Analysis of ^{14}C dates on archaeological materials in the archipelago indicates a human settlement that began within the Current Era (Velasco et al., 2020). It has been suggested that the extinction of *C. bravoi* was probably an event linked to the first human presence on the island of Tenerife (Rando et al., 2014).

One way to confirm this hypothesis is through comparison with similar situations. The total number of recorded insular rodent species extinction around the globe is at least 80 (Woods, 1989; Flannery, 1995; Alcover et al., 1998; Wilson & Reeder, 2005). Of these, at least 10 insular rodent species have become extinct during the last 150 years (Amori et al., 2008). The islands of the Neotropical region, particularly the West Indies, have suffered the highest losses during the last 7000 years at the genus level, with extinction of 14 insular rodent genera and at least 50 species (Woods, 1989; Alcover et al., 1998; MacPhee & Fleming, 2003). In general, all these extinctions occur symptomatically shortly after the arrival of humans on the islands. Americans often talk about the "smoking gun". We didn't see how the murder took place, but when we got to the scene there was a smoking gun...

If we assume, then, that behind the extinction of *Canariomys* is the arrival of humans in Tenerife, the next question that arises is what was the ultimate cause of this extinction?

The first colonizers were Neolithic peoples, who arrived on the island with their domestic animals (dogs, goats, sheep and pigs - the cat arrived later), commensal animals (house mouse, *Mus musculus*), and an agricultural and pastoral culture with a certain hunting suplement (Michaux et al., 1996, among many others). One or more of these elements could have influenced the extinction of *Canariomys*.

Hunting

Hunting has an important component of choice. The hunter chooses the species he wants to hunt, and he does so based on several parameters. There are prey more or less easy to catch, prey more or less dangerous, prey more or less abundant, prey more or less large and prey more or less tasty. And these relationships can vary over time, space and from one hunter to another. Behind these relationships is the optimal foraging theory. A predator tends to maximize profits and minimize losses. There are a lot of studies that determine the prey-predator relationship in different ecosystems and different trophic levels (Barnes et al., 2010; Nakazawa et al., 2013; Tsai et al., 2016). It has been seen, for example, that Pleistocene humans tended to hunt large herbivores with a higher fat content (Ben-Dor & Barkai, 2021). In fact, the extinction of the Pleistocene megafauna is clear evidence of that.

Canariomys became giant when we consider the size of the mouse (*Arvicanthis*) from which it came, but it was actually a relatively normal-sized rat. *Sensu stricto* we should call it a giant mouse. This size makes it especially vulnerable to continental predators but it is only small game for humans. If aboriginal hunted rats it is unlikely that this activity led them to chase the rats to the last corner of the island and cause the extinction. There is also frequent mention of the naivety of island animals, which have evolved without predators and are very confident, facilitating their hunting and extermination (Diamond, 1984). However, this behavior is not observed today

among insular endemic rodents, which continue to be as elusive as continental ones. An additional argument is that the Canarian aborigines had their own herds, therefore game meat was not essential but only complementary to the main diet. Finally, it has also been observed that prey subjected to strong hunting pressure advance sexual maturity to reproduce earlier (Proaktor et al., 2007). Several species of giant lizards have survived in the Canary Islands, cornered on cliffs inaccessible to predators. Their current maximum sizes are significantly smaller than their subfossil remains (Rando, 2003; Mateo & López-Jurado, 1992). This process has been documented for reptiles in several archipelagos (Pregil, 1986). It would be worthwhile to analyze the abundant bones of *Canariomys* in order to determine the body size at which the bony epiphyses close (the moment in which the animal stops investing energy in growth to dedicate it to reproduction). If a progressive reduction in body size were detected over time, we would be faced with proof that hunting could have caused its extinction.

Habitat destruction

Although the dentition indicates that *C. bravoi* was a herbivorous rodent (Firmat et al., 2010), stable carbon and nitrogen isotope analyzes indicate that it was omnivorous and often foraged in the forest canopy (Crowley et al., 2019). It had a very wide distribution throughout the island of Tenerife since its remains appear from Anaga to Teno, and from sea level to Las Cañadas, being especially abundant in some volcanic tubes.

When the first human settlers arrived, they transformed the island's landscape into crops and pastures. They began in the areas with the smoothest orography, which corresponds to the thermophilic forests, and little by little they deforested entirely. From the remains of coal it has been detected how the tree species of the thermophilic forests were the first to be exploited, followed by the shrub species and finally the laurel forest located in the culminating parts and steep slopes (Machado et al., 1997). Today, only a few enclaves with the original laurel forest vegetation are preserved, and have been declared protected spaces, such as the Anaga

peninsula. Two species of pigeon endemic to the Canary Islands have survived in this ecosystem, the Laurel pigeon (*Columba junoniae*) and the Bolle's pigeon (*Columba bollii*) (Martín et al., 2000). If *Canariomys* was ubiquitous, as humans deforested the island, it would have been able to take refuge in the laurel forest and survive, just as the pigeons did. Therefore, it seems unlikely that habitat loss was a determining cause of their extinction.

Dog predation

The dog has been man's best friend since the Paleolithic (some studies estimate domestication at over 15,000 years). The first settlers of the Canary Islands brought it to guard their flocks and their homes. It is not unusual, however, for some dogs to become feral, or have a weak bond with humans, so that they periodically leave the company of their owners, or the herd, and may devote themselves to exploring the territory and hunting. The return to the wild of some domestic mammals such as dogs is advocated for the extinction of endemic rodent species in the Caribbean region (Morgan & Woods, 1986).

The dog is an excellent rat hunter, in fact, even better than the cat. Michaux et al., (1996) consider the giant rat's size to make it good prey for medium-sized dogs. There are even current breeds of dogs selected to hunt rats, such as terriers.

Several historical chronicles corresponding to the period of the Castilian conquest (15th century) point out that there were many feral dogs on the island of Tenerife and that they even ate badly buried human corpses, or forced the islanders who slept in the open field, to climb up trees for not being attacked. There are thousands of human and herd bones with signs of predation by dogs found in various sites, especially in Tenerife and Gran Canaria, which confirm these descriptions. Some of these remains date back to the fourth century (Alberto-Barroso, 1999; Alberto-Barroso et al., 2017; 2021a; 2021b). If this situation had already occurred some centuries before, it would be reasonable to think that the hungry feral dogs could have been a cause of the extinction of the giant rat.

Interaction with the house mouse

When the house mouse (*Mus musculus*) is introduced to islands where it has neither predators nor natural competitors, the population increases to surprising densities and then exerts a very important impact on the entire island biota. For example, on Gough Island (South Atlantic Ocean) it was seen that inexplicably often dead albatross chicks appeared in the nest. When night surveillance cameras were installed, it was discovered that during the night, while the adult birds were at sea fishing, several mice would bite and devour the chick alive (often starting with the anus, free of down). The chick ended up dying of blood loss (Caravaggi et al., 2019).

As a commensal species, the house mouse was accidentally introduced to Tenerife with the first settlers. The common rat (*Rattus rattus*) was not introduced until centuries later, so the mouse had no predators and only a single supposed competitor. Perhaps competition between both rodents for food and burrows, or mouse attacks on giant rat adults or young could have caused the extinction.

Parasites and infectious diseases

There is no data or evidence that *Canariomys* could have become extinct as a result of any parasite or infectious disease introduced by humans or their partner animals. If their remains had aDNA well preserved, paleogenomics could help us detect the presence of pathogenic microorganisms. It would also reveal whether the extinction was sudden or progressive.

Unfortunately, identifying the ultimate cause of the extinction is very difficult and it was possibly not unique, but rather a combination of some of the factors discussed above, acting separately or synergistically. The disappearance of the Tenerife giant rat will remain a mystery. We lost one (more) species, and we don't know how.

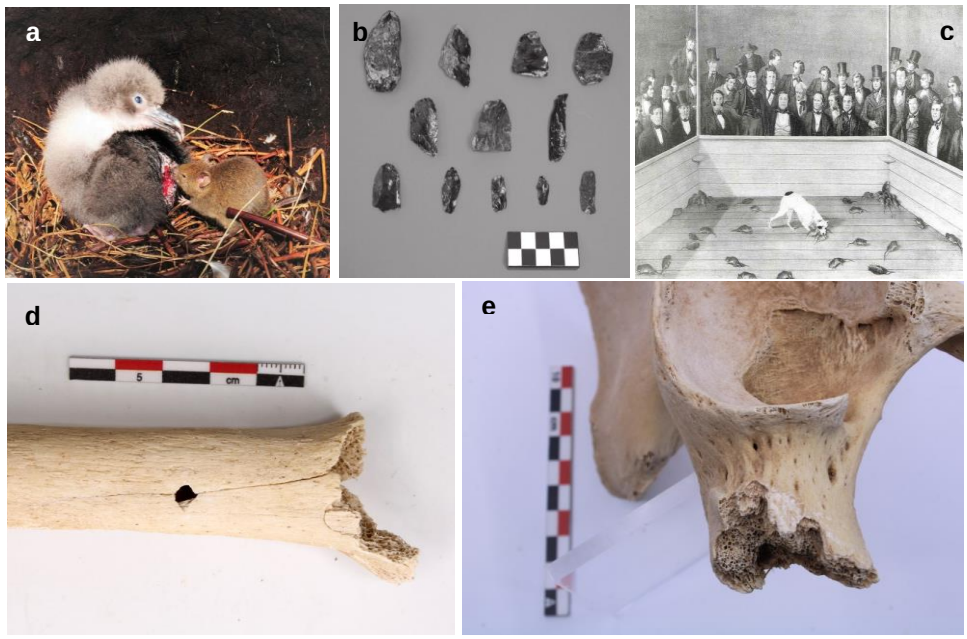


Figure 5.4. Three possible causes of the extinction of the Tenerife giant rat (*Canariomys bravoi*). (a) The house mouse (*Mus musculus*) can become a predator of much larger animals, such as this albatross chick on Gough Island (South Atlantic Ocean). (b) The aborigines of the Canary Islands did not have metals and made their tools and hunting weapons with volcanic stone, like these obsidian points found in the site of El Barranco de San Blas (Tenerife). (c) Rat-bating was a very common practice in London during the 19th century, some breeds of dogs such as terriers were selected for their effectiveness as rat hunters. (d, e) A human humerus with a canine hole and human pelvis bitten by a dog, from funerary cave of El barranco de Guayadeque, 8th-10th centuries AD. (Gran Canaria) [from (a) Gough Island Restoration Team; (b) Chávez et al., 2005; (c) Wikicommons; (d,e) Verónica Alberto Barroso, Tibicena, Arqueología y Patrimonio].

5.2.2. African Pleistocene rodents

With paleogenomics we have been able to determine that the African grass rat (*Arvicanthis niloticus*) colonized Tenerife to become a giant rat (*Canariomys bravoi*), and we know that this occurred about 650,000 years ago. We already know the fate of the giant rat: extinction, but what happened to the African grass rat since then? What remains of that ancestral population?

The dominant marine currents on the northwestern African coast tend to move from north to south (Arístegui et al., 2009), therefore, it may be expected that the river that dragged to the sea the raft carrying the ancestor of the giant rat was located north of the Canary Islands, and we hypothesize that it was most likely the Draa River. However, when we look at the current distribution of the African grass rat we see that it is located in a belt that crosses Africa from east to west below 15º latitude N. The closest population to the Canary Islands is thousands of kilometres away to the south, separated by the insurmountable ecological barrier of the Sahara desert. It is clear that when Tenerife was colonised tens of thousands of years ago, the climate, and therefore the landscape, of North Africa were very different. Some models have reconstructed what the precipitation was during the interglacial periods and from the isohyets it was modelled how the vegetation was distributed (Larrasoaña et al., 2013). It can be seen (Figure 5.5) that where today there is only sand, millennia ago there was a large expanse of wooded grassland, with numerous rivers and large lakes, ecosystems that are very conducive to the expansion of *Arvicanthis*.

Several studies (Sicard et al., 2004; Dobigny et al., 2013; Bryja et al., 2019) have analysed in depth the different current populations of *Arvicanthis* and have confirmed that the Pleistocene climatic cycles determined the expansions and retractions of this rodent and are the main factor of the present speciation. They have also located the geographical origin of the group in East Africa about 2 Ma, and have detected the existence of various cryptic species grouped (provisionally) under the *Arvicanthis niloticus* complex.

It would be expected that once the last retraction of *Arvicanthis* occurred in the Holocene, the current populations geographically closest (or rather, less distant) to the Canary Islands should also be the genetically closest to *Canariomys*.

Ideally, the genome of *Canariomys* could be included in the phylogenetic tree of the genus *Arvicanthis* elaborated by Bryja et al*.,* (2019), to see which group is more closely related and confirm or rule out the hypothesis of geographical proximity. Unfortunately, the phylogeny of *Arvicanthis* was carried out with some genetic markers (cytb, two nuclear exons -IRBP, RAG- and four nuclear introns -WLS, DHCR, TRPV, SMO-) that we were unable to recover from *Canariomys*. The *Arvicanthis niloticus* specimen that we used as a reference to recover the *Canariomys* aDNA came from a US laboratory breeding colony, originated from 29 individuals trapped at the Masai Mara National Reserve in south-western Kenya in 1993 by the research team headed by Professor Laura Smale at Michigan State University (McElhinny et al., 1997; Refinetti, 2004). For this reason, when our *Arvicanthis* of reference was included in the phylogenetic tree of Bryja et al., (2019) it was placed next to the Masai Mara specimens (see section 4.2). Therefore, reference specimen used would not be the closest to *Canariomys* neither geographically nor genetically.

Given the extreme difficulty of extracting more aDNA from *Canariomys*, an exhaustive sampling of the less studied or more isolated current populations of *Arvicanthis* could be carried out (for example, the Hoggar mountains in southern Algeria), to detect if any of them has acted as a Pleistocene refuge and preserves ancestral traits shared with *Canariomys.* But there is a second no less harsh possibility: to sample Pleistocene *Arvicanthis* directly.

Both the common rat (*Rattus rattus*) and the house mouse (*Mus musculus*) have been commensal rodents of humans for centuries and have travelled with us around the world. However, millennia ago, humans had other species of commensal rodents. In the Balearic Islands, for example, the arrival of humans at the beginning of prehistory (2350–2150 cal BC) involved the involuntary introduction of the wood mouse (*Apodemus sylvaticus*) and the dormouse (*Eliomys quercinus*). It seems that the spread of

the house mouse during Roman times would have displaced both species to the completely wild life form they have today (Cucchi et al., 2005; Alcover, 2008).

Something similar happened in Africa. Before the arrival of commensal murids of Asian origin (rat and mouse), there were endemic rodents that occupied the niche of commensal species of humans, at least since *Homo erectus*. The main commensal rodent species were *Mastomys*, some gerbils and especially *Arvicanthis* (Geraads, 2012; Stoetzel, 2017). In fact, commensalism is so evident in this last species that its fossil remains are often used as indirect evidence to infer human presence in a site. Its expansion throughout the African continent from the area of Ethiopia, where it originated, clearly parallels human expansion (Stoetzel, 2017).

Classically, commensal rodents are associated with the Neolithic because they are linked to sedentarisation and agriculture, but they can also be attracted by food waste and to seek protection from predators and climatic hazards (Cucchi et al., 2005). Therefore, these rodents are in some way associated with the Palaeolithic as well.

Several sites containing *Arvicanthis* fossils associated with humans have been identified across Africa. Its first appearance in North Africa dates from the Early Pleistocene and since then remains have been found until the beginning of the Holocene, when the aridisation of the Sahara made this rodent disappear from the region (Table 5.1).

The most interesting site for this thesis is, without a doubt, El Mnasra Cave (Témara region, Morocco), located a few tens of kilometres north of the Draa river (Figure 5.5) (Stoetzel, 2017). This is the Pleistocene population of *Arvicanthis* closest to the Canary Islands and from where colonisation most likely occurred.

The big challenge here is to recover aDNA in good conditions. The location of all the North African sites is, according to Hofreiter et al., 2014, within the world regions with the lowest aDNA survival rate. Despite the fact that many sites are located relatively close to the Atlas, they do not have enough altitude to guarantee low temperatures that significantly improve the probability of preservation. We can only hope that the conditions inside the

caves, under the layers of sediment, have allowed relative preservation. In fact, these sites are geographically closer to the zones of intermediate aDNA conservation than the Canary Islands, where we managed to recover *Canariomys* aDNA.

Woods et al., (2017) makes a compilation of published studies from 1990 to 2017 in which ancient or degraded DNA was utilised to study micromammal fauna and finds 42. It is a small proportion in relation to the efforts devoted to mammalian megafauna and ancient humans, but proves that these studies are viable.

The present situation is that on the one hand we have palaeontology studying fossil remains of *Arvicanthis* associated with humans, and on the other we have molecular biology studying the complexity of current *Arvicanthis* populations. If it were possible to apply palaeogenomics, we would combine the benefits of both disciplines and we could achieve a very interesting perspective on the evolution of *Arvicanthis* throughout the Pleistocene and perhaps even follow the trace of this rodent from its East African origin, its expansion and diversification across the continent, until its arrival to the North African coast, from where it colonised the Canary Islands and became a giant.

Site	Species	Epoch	References
Jebel Ressas, Tunisia	*Arvicanthis sp.*	Early Pleistocene (~1-1.5 Ma)	(1)
Tighenif, Algeria	*A. arambourgi*	Middle Pleistocene (~700-800 ka)	(2)(3)
Aïn Mefta, Algeria	*A. niloticus*	Middle Pleistocene	(2)(4)
Tadjera, Algeria	*A. arambourgi*	Middle Pleistocene	(2)
Irhoud Derbala Virage, Morocco	*A. niloticus irhoudae*	Middle Pleistocene	(2)
El Mnasra, Moroco	*Arvicanthis sp.*	Late Pleistocene (~110 ka)	(5)
Chrafate, Moroco	*A. niloticus*	Late Pleistocene	(6)
Egypt	*A. niloticus*	Late Pleistocene	(7)
Palestine	*A. ectos*	Late Pleistocene	(8)
Ti'n Torha, Libya	*A. niloticus*	early Holocene	(9)
Ez Zarka, Morocco	*A. niloticus*	early Holocene	(6)

Table 5.1. Sites in North Africa where fossils of *Arvicanthis* have been found, chronologically arranged. References: (1) Mein & Pickford, 1992; (2) Jaeger,1975; (3) Denys et al., 1987; (4) Ameur, 1976; (5) Stoetzel et al., 2014; (6) Ouahbi et al*.,* 2003; (7) Kowalski et al. 1989, 1993; (8) Tchernov, 1968; (9) Masseti, 2010. Holocene samples date to the epoch before the great aridity crisis in the Sahara desert around 4 ka BP [own elaboration from Stoetzel, 2017].

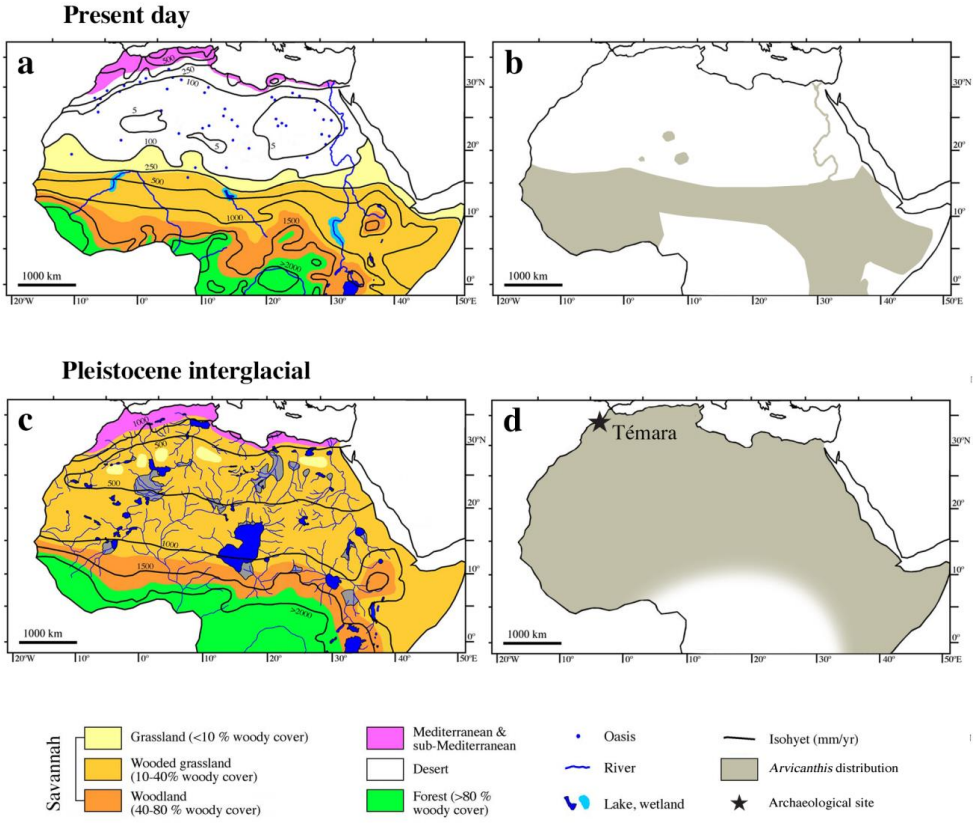


Figure 5.5. (a) Present day vegetation and mean annual precipitation (MAP) in North Africa. Note the latitudinal vegetation belts (Mediterranean and sub-Mediterranean, desert, grassland, woodland, and forest) (b) Current distribution of the *Arvicanthis* genus (dark gray), data from www.iucnredlist. Note that it is restricted to wooded grassland and woodland. (c) Reconstruction of North African vegetation during past green Sahara periods according to estimated and reconstructed MAP for the Eemian GSP (122-128 kyr BP) (d) Estimated distribution of the *Arvicanthis* genus (dark gray) during interglacial periods; (star) location of El Mnasra Cave (Témara region, Morocco) with fossil remains of *Arvicanthis* (~110 ka BP) [modified by Carles Sales from Larrasoaña et al., 2013].

5.2.3. Island colonizers

The African grass rat (*Arvicanthis niloticus*) has a unique characteristic among rodents: a mainly diurnal activity. Furthermore, it also possesses many advantageous traits that make it an ideal laboratory model for biomedical research: small body size, easy breeding, short life cycle, relatively insensitive to photoperiod variations, lack of hibernation, and unlike the standard laboratory rat (*Rattus rattus*), reaches maximum body mass shortly and does not show marked sexual dimorphism. It even exhibits a good temperament that makes it easy to handle (Refinetti, 2004). For these reasons it was imported to the USA in the 1990s where breeding colonies were established as a diurnal model for sleep and circadian rhythm studies (Refinetti, 2004). However, it was soon found that when it was fed with dry kibble it rapidly developed type 2 diabetes. Most of the rodent models commonly used to study diabetes have three drawbacks: (1) they are genetically or chemically modified to produce diabetes; (2) unlike humans, most require a high-fat diet; and (3) they take too long to develop the disease. In contrast, the African grass rat develops diabetes spontaneously and rapidly (8-10 weeks) on a high-carbohydrate diet, similar to humans. Therefore, it has become an ideal model not only for the study of sleep, but also for physiological-nutritional studies and especially type 2 diabetes (Noda et al., 2010; Chaabo, et al., 2010; Leow et al., 2016; Subramaniam et al., 2018; 2019). This medical relevance surely stimulated the sequencing of its genome in the year 2020, and the subsequent identification of some genes associated with type 2 diabetes and metabolic dysfunctions (Toh et al., 2022), an eventuality that happily benefited this thesis, since it provided a reference genome very close to the extinct Tenerife giant rat (*Canariomys bravoi*) and allowed the partial recovery of its nuclear and mitochondrial DNA, since the genomes of the rat (*Rattus rattus*) and the mouse (*Mus musculus*) that we had initially used as a reference turned out to be unsuccessful.

This physiological resemblance between the African grass rat and humans presents another interesting analogy. It has been found

that some human groups in the United States show a much higher incidence of obesity and type 2 diabetes than the American average. Some of these groups are the Hawaiians of Polynesian origin and the Pima Indians of Arizona and Mexico (Bogardus et al., 1986; Ravussin & Bogardus, 1990; Schulz & Chaudhari, 2015).

In 1962 J.V. Neel proposed the "thrifty gene" hypothesis (Neel, 1962) according to which, in the past, populations that had experienced periods of famine were more likely to survive if they were metabolically sparing and stored calories efficiently. Their survival gave them an adaptive advantage and this feature spread through the population. Based on his hypothesis, however, what had historically been a benefit became detrimental in present day society with high-calorie diets (especially in North America). The genetic ability to store calories efficiently in our current environment becomes a risk factor for type 2 diabetes and obesity. Some more recent studies indirectly confirm the existence of this thrifty gene (or set of genes) (Bogardus et al., 1986; Ravussin & Bogardus, 1990).

Similarly, the African grass rat could be adapted to withstand food shortages with thrifty genes, and just as the Polynesians were able to cross the Pacific Ocean and colonize the most remote islands on the planet, surely surviving long periods of food shortage, the rat would also be metabolically prepared to endure the journey from Africa to the Canary Islands by rafting dispersal. These crossings are usually very demanding energetically because they can last weeks. Hence, they can only be tolerated by reptiles (mainly geckos, skinks, lizards, and turtles) with a particularly resistant physiology, and small mammals, especially rodents, that contain heterothermic taxa, and in scarcity of food and water are able to a temporary reduction of metabolism and body temperature (Ali & Vences, 2019).

Of the nearly 2,300 species of rodents in the world, at least 388 extant species are restricted to islands (17.5% of all extant rodent species, or almost 1 in 5) (Amori et al., 2008). There is no doubt that this is a group successful in island colonization. But when taxonomy is analysed in more detail we see that there is a very uneven representation. Of the 34 families of extant rodents only 7 have insular endemic forms (Table 5.2). The family Muridae

presents 63% of the island genera, followed by the family Capromidae (~12%), Nesomyidae (~9%) and Sciuridae (~7%).

We could think that these values reflect the diversity and abundance of continental rodents close to the islands. When we calculate the diversity of each of the rodent families (Figure 5.6) we see that indeed the Muridae are the ones with the largest number of species (816), but the second family with the most species is Cricetidae (765) and on the contrary it represents less than a 3% of insular rodents. Therefore, there must be other factors that determine island representativeness. Murids have a wide global distribution that includes the areas of the planet with more islands relatively close to the continents (Palearctic, Oriental and Australia), consequently, they also have zoogeography in their favour.

However, it is surprising to note that island colonizers occur in phylogenetically clustered families, even if they have very distant geographical distributions (Figure 5.6). Perhaps they share common traits that would preadapt them to island colonization. These traits could be physiological and ethological, such as the aforementioned thrifty gene, the plasticity to adapt to very diverse environments, the reduced size, or the habit of hiding in logs that end up being washed away by torrential rains and transported to the sea.

The fact is that in the Canary Islands, apart from the Tenerife giant rat (*Canariomys bravoi*), there were two other insular endemisms, the Gran Canaria giant rat (*Canariomys tamarani*), which according to morphological analyses is placed in the same genus (López-Martínez & López-Jurado, 1987) and the lava mouse (*Malpaisomys insularis*) from Fuerteventura and Lanzarote, closely related to the genus *Mus* (Hutterer, et al., 1988; Pagès et al., 2012). All three species belong to the family *Muridae*.

It would be interesting to determine if there are genes shared by these families of island colonizing rodents, thanks to which they have sailed farther than any other terrestrial mammal (with the exception, indeed, of our own species).

Family	Genera			Zoogeographical region
	Extant	Extinct	%	
Muridae	59	6	63,1	Paleartic, Australian & Oriental
Capromyidae	5	7	11,7	Neotropical
Nesomyidae	9		8,7	Afrotropical
Sciuridae	7		6,8	Oriental
Heptaxodontidae †		5	4,9	Neotropical
Cricetidae	1	2	2,9	Neotropical
Gliridae	1	1	1,9	Paleartic
Total	82	21	100	

Table 5.2. Rodent families with insular endemisms both extant and extinct during Late Pleistocene and Holoce. They are ordered by the number and percentage of genera. † Extinct family. If we also consider the entire Pleistocene, several genera of insular Gliridae (dormice) are known from Mediterranean islands (Van de Geer et al., 2010) [own elaboration from Amori et al., 2008].

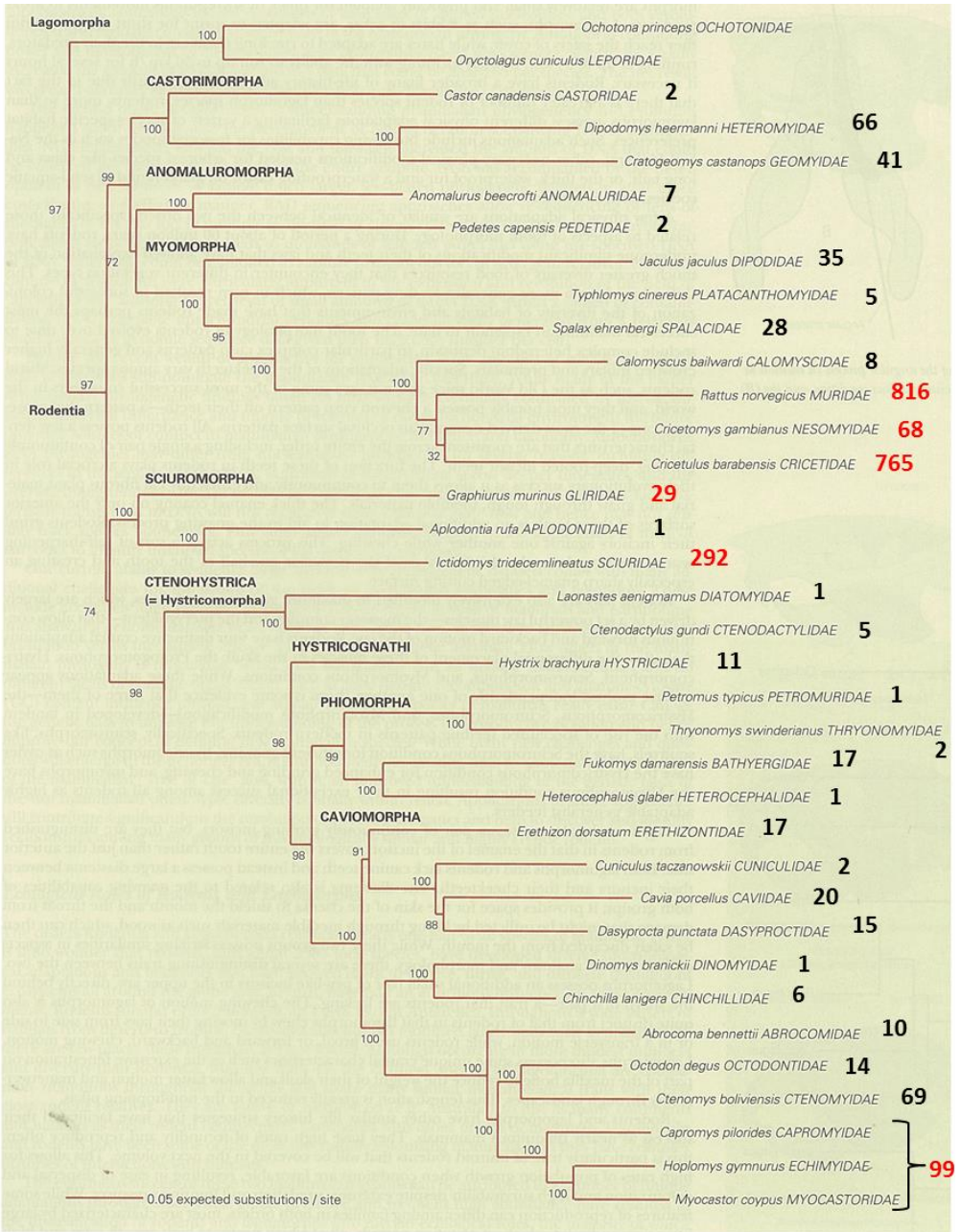


Figure 5.6. Phylogenetic tree depicting the evolutionary relationships between families of Rodentia and Lagomorpha based on a maximum-likelihood (ML) analysis of 31 genes and 39,099 nucleotide base pairs. The number of species is indicated next to each family. Red indicates families that have extant or extinct endemic island rodents. Note that in general the most diverse families are those that colonize the islands, but a certain kinship bias is also intuited. The last three families are usually grouped into a single one [modified from Upham et al., 2016].

5.2.4. Gigantic gigantism

When in the 17th century the Irish clergyman Jonathan Swift wrote his famous book of adventures "Gulliver's Travels", he could not imagine that the island of Lilliput, populated by tiny humans, and the island of Brobdingnag, populated by gigantic humans, would somehow anticipate, by about 200 years, one of the most astonishing findings in evolutionary biology: the island rule. In 1964 biologist and naturalist J. Bristol Foster published an article in the journal Nature entitled "The evolution of mammals on islands" in which he studied 116 insular species of mammals and compared them to their continental counterparts (Foster, 1964). After observing changes in body size, he proposed that small continental species become giants on islands due to the absence of predators, while large continental species become dwarfs on islands as a consequence of food limitation. This rule aroused great controversy in the scientific world, it was studied in depth, questioned, contradicted and nuanced (Lomolino,1985; Michaux et al., 2002; Meiri et al., 2008; Köhler et al., 2008; Lomolino et al., 2012; van der Geer et al., 2013; Durst & Roth, 2015). Today, almost 60 years after its publication, it is still being studied. In fact, it has been so extensively analysed that meta-analyses can now be performed. Surely the most recent and exhaustive is published by Benítez-López et al., (2021). The results (Figure 5.7) confirm that the island rule is a generalized pattern among vertebrates, but it depends on a series of ecological and environmental pressures. This meta-analysis is based on a formidable number of studies, islands and living species, but does not include any extinct species.

Knowing that the Tenerife giant rat (*Canariomys bravoi*) was a worthy representative of the island rule, we plotted it in Benítez-López et al., (2021) graph. For the island-mainland comparison, the authors calculated the log response ratio (lnRR) as the natural logarithm of the ratio between the mean body size of individuals from an insular population Mi and that of mainland relatives Mm (lnRR = log(Mi/Mm)). They also calculated the natural logarithm of Mm for the graphical representation.

It has been possible to estimate the body mass of *C. bravoi* (Montcunill-Solé et al., 2014) at 1492-1650 g (mean 1571 g). In addition, molecular phylogeny has clearly revealed that the African grass rat (*Arvicanthis niloticus*) is its mainland closest relative (see section 4.2), with a body mass of 89-130g (mean 114g; Monadjem et al., 2015). Therefore, the lnRR of *C. bravoi* is 1.14 and the natural logarithm of *Arvicanthis* body mass is 2.057. Unexpectedly, this point appears as an outlier on the mammal graph of Benítez-López *et al*., 2021 (Figure 5.7, a).

How should the regression be modified to include this point? There are three possibilities. (1) The line could greatly increase the slope and maintain the intercept point with X-axis. The result would be a magnification of changes in body mass, gigantism and dwarfism would increase (Figure 5.7, b). (2) The line could increase the slope and shift the intercept point with X-axis to the right. This would imply that gigantism would increase but dwarfism would not change (Figure 5.7, c). Finally (3) the line could move in parallel, keeping the slope and moving the intercept point with X-axis further to the right. The result would be, once again, an increase in gigantism, but now dwarfism would practically disappear (Figure 5.7, d). Note that by any of the three pathways gigantism is increased to include the exceptional size of *C. bravoi*.

What could modify this regression? Likely the isolation time of island animals.

Many of the islands included in the work of Benítez-López et al., 2021 have been connected to the continent by relatively recent land bridges; consequently the phenotypic differences have not had time to accumulate. In addition, the relative proximity of these islands allows regular dispersal between mainland and island populations, promotes gene flow, and counteracts body-size modification trends.

The importance of the isolation time was already recognized and studied by Lomolino et al., 2013. In their study, these authors compiled data on 63 island mammals belonging to 7 orders (including rodents), extinct during the Pleistocene and early Holocene and compared them to their putative continental ancestors. As predicted, the slope of the regression for extinct

mammals is steeper (more negative) than for extant mammals (difference between slopes = 7.47, P < 0.0001, d.f. = 405). The change in slope is achieved by a combination of pathways 1 and 2 discussed above: it steepens the values at both ends of the regression line and shifts the intercept point with X-axis slightly to the right. The increase in slope is basically produced by incorporating very extreme cases of gigantism and especially of dwarfism, represented by dwarf elephants (Proboscidea) and dwarf hippos (Artiodactyla), which somehow "pull" the regression downwards. The shift in the intercept point with X-axis could be due, in part, to the difference in the total number of cases included in the regression calculations (n=52 for extinct insular mammals and n=357 for extant insular mammals). But, in part, it could also be because there is probably no species located exactly at the intercept point with X-axis, meaning that it would not change its body size at all when colonizing an island. Animals with an intermediate body mass of between 2.5 and 3.5 g on the logarithmic scale (Figure 5.8), which with the antilogarithm represent between 300 and 3000 grams, can evolve towards gigantism or dwarfism depending on the ecological and environmental conditions. Many rodents present this range of body mass and it has already been observed that they can evolve in both directions (Benítez-López et al., 2021). The same patterns have been described for lagomorphs as well.

When we add *C. bravoi* to the graph by Lomolino et al., (2013) we see that it groups well with the rest of the extinct insular rodents, although it is still the highest value, together with the extinct Sicilian giant dormouse (*Leithia melitensis*) (Figure 5.8). But the dispersion of rodent values is such that in isolation they do not show any relationship with the size of continental ancestors, that is, they do not seem to fit the island rule. This fact is easily explained because the island rule is a macroevolutionary trend that is only observed with large number of samples and especially when body masses are distributed over the entire range, while the extinct rodents included in the analysis are few (14) and cover only a small part of the range.

On the other hand, Benítez-López et al., (2021), observed that the slope of the regression increased for smaller and more remote islands, that is, the further away and smaller an island is, the more gigantism or dwarfism it shows. Physiography generates in current species the same evolutionary trends as the time of isolation in extinct species. We could then think that if we combine physiography and time of isolation we could explain the differences in gigantism observed in the species of extinct insular rodents. But when we plotted the body mass of the extinct rodents against the surface area of each of the islands, we found no relationship. It should be noted that except for one case (*Agathaeromys praeuniversitatis*), all the extinct rodents included in the study by Lomolino et al., (2013) inhabited different Mediterranean islands (Table 5.3). During the Messinian Salinity Crisis, between 5.97 and 5.33 Ma ago (Manzi et al., 2013), the closure of the Strait of Gibraltar practically caused the Mediterranean Sea to dry up and as a consequence the Mediterranean islands were connected to the continent (Hsü et al., 1977). Therefore, rodents colonized them by land. Instead, the island of Tenerife is truly oceanic since it has never been connected to the African continent. The ancestor of the giant rat had to colonize the island by natural rafting. The colonization process in one and the others has exerted very different selective pressures that make comparisons difficult. In fact, Benítez-López et al., (2021) determined that in mammals, the mainland body mass explains only 11.7% of the variance of island rule, while phylogeny explains up to 20.1% of the variance. This phylogeny can vary greatly depending on the colonization processes. Finally, in all extinct insular rodents, the continental ancestor is inferred based on comparative anatomy, with all the uncertainties that this can entails, while for *C. bravoi* we have determined the ancestor molecularly and we have a reliable baseline to accurately calculate the degree of gigantism. Other works have been carried out with aDNA at both ends of the island rule such as a study of the size increase of the common vole (*Microtus arvalis*) in the Orkney archipelago (Martinkova et al., 2013), and an estimation of the dwarfing rate of an extinct Sicilian elephant (Baleka et al., 2021). Paleogenomics is very useful to

confirm phylogenetic relationships, provide divergence times that allow calculating evolutionary rates and extend the island rule with more extinct animals.

In any case, it seems clear that the island rule is a pattern that is fulfilled for extant species of vertebrates and magnified for extinct ones, which had much more time to evolve and diverge from their continental ancestors. Time pushes island vertebrates into diminutive dwarfism or giant gigantism.

Based on the length of their strides (10 yards, or 9 meters) it has been estimated that Gulliver's giants were 12 times larger than a human. The Tenerife giant rat was 14 times larger than its continental ancestor. Even in the proportions Jonathan Swift got it right.

Insular species	Ancestor	Log10 Ancestral body mass (g)	Si	Island	Surface (km2)
*Agathaeromys praeuniversitatis**	*A. donovani*	1.8921	0.29	Bonaire	288
*Hattomys gargantua**	*H. beetsi*	2.2466	4.96	Gargano	2000
Hypnomys morpheus	*Eliomys quercinus*	1.8692	3.09	Majorca–Minorca	4336
Kritimys catreus	*K. aff. K. kiridus*	2.2046	3.24	Crete	8450
Leithia cartei	*Eliomys quercinus*	1.8692	3.00	Sicily	25711
Leithia melitensis	*Eliomys quercinus*	1.8692	13.53	Sicily	25711
Microtus (Terricola) ex gr. M. savii	*M. (T.) savii*	2.0065	1.26	Sicily	25711
Microtus (Tyrrhenicola) henseli	*M. ruffoi*	2.1064	1.78	Corsica–Sardinia	32812
*Mikrotia magna**	smallest *Mikrotia* (Biancone 1)	1.6254	9.48	Gargano	2000
*Mikrotia middle-sized lineage**	smallest *Mikrotia* (Biancone 1)	1.6254	3.32	Gargano	2000
*Mikrotia small-sized lineage**	smallest *Mikrotia* (Biancone 1)	1.6254	1.06	Gargano	2000
Mus minotaurus	*M. musculus*	1.2222	3.23	Crete	8336
Rhagamys orthodon	*Rhagapodemus ballesioi*	1.5336	2.63	Corsica–Sardinia	32812
Canariomys bravoi	*Arvicanthis niloticus*	2.057	13.78	Tenerife	2034

Table 5.3. Variation in insular body size S_i (insular body mass/ mainland body mass) among extinct rodents.*Calculation of S_i based on comparison to early chronospecies on the island. Owing to the dynamics of sea level and the seafloor through the Miocene to Holocene epochs the islands listed here include areas that are currently peninsulas (Gargano, Italy), regions of current islands (northern palaeo-island of Sicily), or separate islands that were previously joined (Corsica-Sardinia; Majorca-Minorca). Island areas are calculated accordingly [modified from Lomolino et al., 2013].

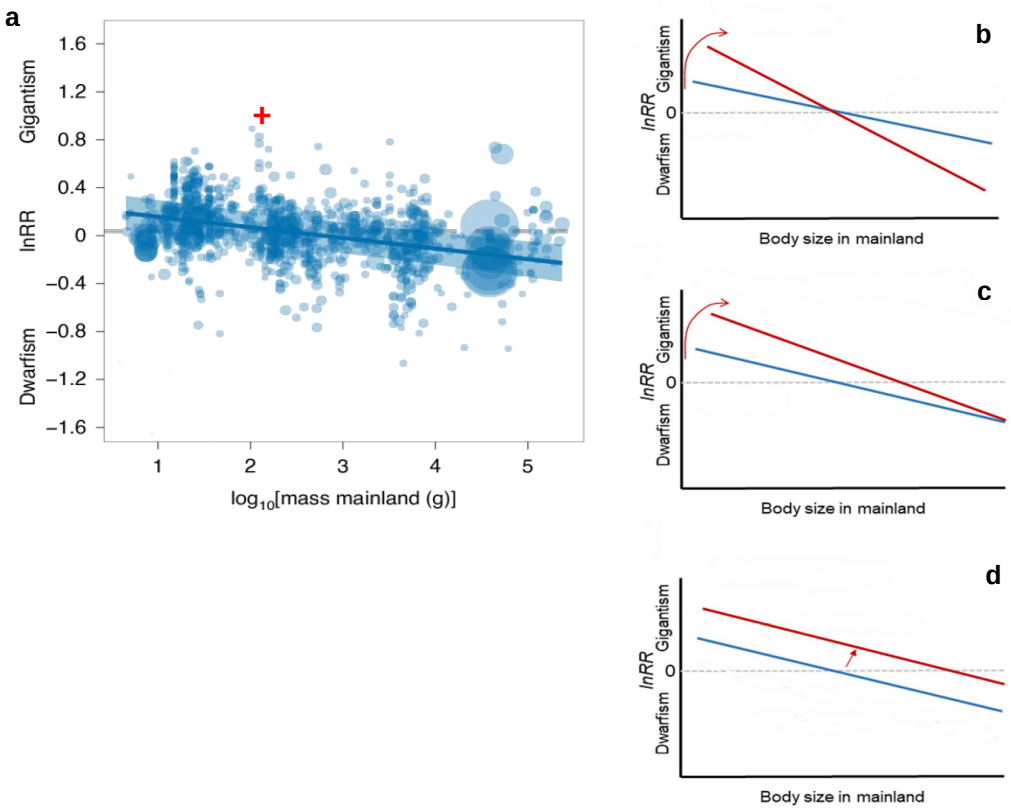


Figure 5.7. 'Island rule' effects in terrestrial extant mammals (N =1,058). (a) Relationship between lnRR (log of ratio between island body mass and mainland body mass) and body mass in the mainland. Model was fitted using phylogenetic multi-level meta-regression. lnRR >0 indicates gigantism, lnRR <0 indicates dwarfism and lnRR = 0 indicates stasis (no shift in body size from mainland to island populations). The size of points represents the inverse of the sampling variance for each paired island-mainland response ratio in the model. Shaded areas represent 95% confidence intervals. The red cross is the value of the extinct Tenerife giant rat (*Canariomys bravoi*). (b-d) Conceptual models depicting the different hypotheses to include the Tenerife giant rat in the regression [modified from Benítez-López et al., 2021].

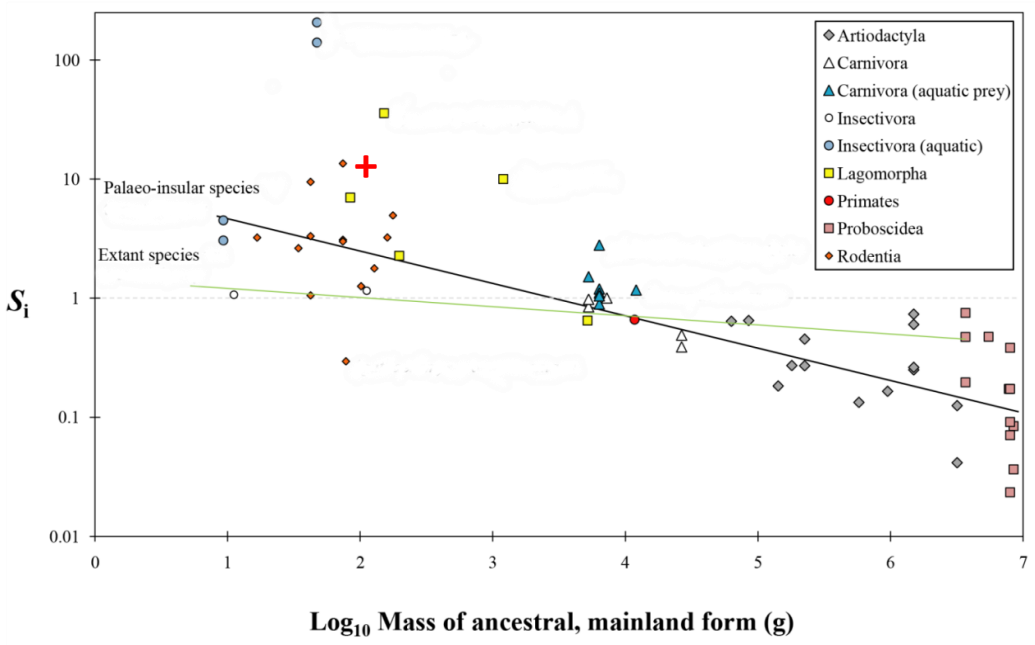


Figure 5.8. Differences in insular body size trends of extinct (black trend line, labelled simbols, n=52) and extant insular mammals (green trend line, n=357). The red cross is the value of the extinct Tenerife giant rat (*Caraniomys bravoi*) [modified from Lomolino et al., 2013].

5.3. Xerces blue (*Glaucopsyche xerces*)

5.3.1. Causes of extinction

For as long as we have records, we know that the Xerces Blue butterfly was always considered a rare or infrequent species, with a single local population (Boisduval, 1852; Tilden, 1956; Downey & Lange, 1956). There are no field data or fossil records that allow us to know what its original distribution was.

This information is confirmed from paleogenomics. The recovery of the complete genome of Xerces Blue, which we have carried out in this thesis, has allowed us to detect signs of inbreeding, such as the accumulation of ROHs and low heterozygosity in relation to its sister rate the Silvery Blue. This indicates that the population had been surviving for some time with a small number of individuals. On the other hand, the analysis of the demographic evolution of the two specimens with the highest coverage, as well as the rest of the specimens with lower coverage, coincide in showing a continuous population decline from about 75,000 years ago, in accordance with the Wisconsin Glacial episode (see section 4.3).

The commonly accepted cause of the extinction of the Xerces Blue is the destruction of its habitat (Downey & Lang, 1956). It is true that it was associated with the coastal dune systems of the San Francisco area, and that when this city grew, its habitat was drastically reduced, which is why it became extinct. The last spot where it lived measured 65 x 140 feet (43 x 20 meters = 860 m^2; Downey & Lange, 1956). But this is the final chapter of his story. What happened before?

Historical descriptions highlight that the butterfly lived associated with the common Deerwood plant (*Acmispon glaber*), but also with *Lupinus arboreus*, *L. micranthus* and *Astragalus menziesii*, since it was the food for its larvae (Downey & Lange, 1956). These plants are part of the floristic group called coastal sage scrub, typical of the Californian chaparral (Steppan, 1991).

The California coast, between latitudes 42º 30' N and 30º N, is characterized by a Mediterranean climate, with hot, dry summers

and mild, moderately rainy winters. The coastal fog generated in summer when the cold sea currents coming from the north cool and condenses the air, contribute to conserving the environmental humidity in the coastline. To the east, the Sierra Nevada and Sierra Cascada mountains limit the Mediterranean strip to about 100-200 km wide. The geographical center of this area is precisely the city of San Francisco (Folch & Camarasa, 1993).

Although there is evidence of climates with Mediterranean characteristics during the last part of the Miocene, 7 million years ago, the proper Mediterranean climate appeared in the Pleistocene, after the first glaciation, 2 million years ago.

When we compare florist diversity between the Mediterranean Basin and Mediterranean California, we see that it is much higher in the latter. This is due to the effect that the glaciation had on both continents. In Western Europe, when the ice cap grew, it pushed thermophilic species south, but the east-west orientation of the Alps and Pyrenees mountain ranges acted as a barrier and many species became extinct. As the ice receded, thermophilic species regained territory, but with depleted diversity. In contrast, in North America the main orientation of mountain ranges is north-south, so when ice caps grew, thermophilic species were able to shift their ranges far to the south, and survived without extinction. During the interglacial they recolonized territory preserving the original diversity (Folch & Camarasa, 1993). Consequently, California was a refuge for many thermophilic species, including the coastal sage scrub and, predictably, the Xerces Blue butterfly.

We can estimate the potential distribution of Xerces Blue from the area occupied by the common Deerwood, as a proxy of the coastal sage scrub (Figure 5.10, a, b). If instead of using only the points of occurrence of the plant, we use its potential distribution, we obtain a very neat potential distribution for Xerces Blue (Figure 5.10, e). Note that the areas where the probability of occurrence is highest (red) are in the vicinity of San Diego, and in two small coastal areas located further north. These areas are Francisco Bay including Point Reyes National Seashore and Golden Gate National Recreation Area (Figure 5.10, c), and Guadalupe-Nipomo Dunes National Wildlife Refuge (Figure 5.10, d), that is, they coincide with

protected areas. The Xerces Blue became extinct in one of the places where its original habitat is still best preserved today. We could therefore think that it disappeared because its habitat was reduced below a certain threshold.

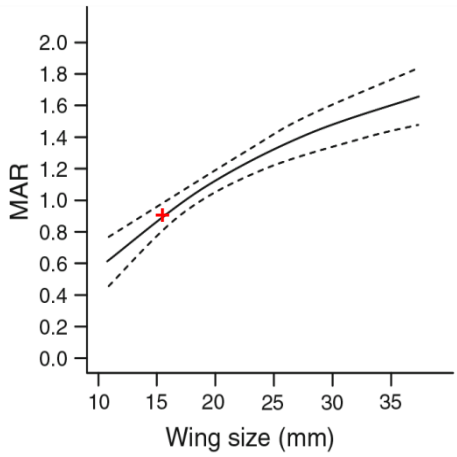
The minimum area requirement (MAR) of a species is the smallest area of suitable habitat needed to sustain a viable population. It is considered a key parameter in conservation planning (Groves et al., 2002). Baguette & Stevens, 2013 estimated the MAR for 142 species of European butterflies and found that the parameters that best predicted it were: body size and four life-history traits (myrmecophily, thermal tolerance, mate searching strategy, and ovigeny). We don't have enough information about these traits in Xerces Blue, but we do know its wing sizes, which correlate with body size. According to Downey & Lange, 1956 the length of the forewing (average male-female) was 16 mm, while the length of the hindwing (same size male-female) was 12.5 mm. When we represent this value in the graph of Baguette & Stevens, 2013 (Figure 5.9, a) we see that it predicts a MAR of between 0.8 and 1. This is equivalent to a minimum area of just under one ha (Figure 5.9, b). The Golden Gate National Recreation Area currently has an area of 81,000 acres, about 328 km^2, equivalent to category 8, well above the category 1 corresponding to Xerces. Although in recent decades the San Francisco area has been the subject of various initiatives to restore coastal ecosystems, it seems that the area available to Xerces was much greater than the minimum necessary to survive (Downey & Lang, 1956). This indicates that probably apart from habitat reduction, there should be other factors that caused its extinction. It is also surprising that the last population of Xerces Blue was not located in the San Diego area where the common Deerwood currently has the widest potential distribution. Perhaps in addition to the plant, Xerces needed the dunes, where it should find other essential factors for its survival, such as ants (Pierce & Eastealt, 1986).

As some authors have already pointed out, species with restricted distributions may be under no particular threat if their distributions are stable, whereas widespread species may be declining fast

(Thomas & Abery, 1995). Behind an extinction there is not only a reduced distribution, there is also a rate of decline.

For Xerces we know its existing status just before extinction, and we can infer what might be its original status from the distribution of its associated plant. In addition, with paleogenomics we have been able to calculate its rate of decline on a time scale of tens of thousands of years. Unfortunately, with this method we have no resolution to know the decline of the last centuries, nor consequently, to discern to what extent the human contribution to the extinction was significant or not.

a



b

Category	Minimum patch area for a population	
1	1 ha	Tiny
2	4 ha	Very small
3	16 ha	Small
4	64 ha	Rather small
5	260 ha	Rather large
6	10 km^2	Large
7	40 km^2	Very large
8	160 km^2	Humongous
9	Unlimited	Infinite

Figure 5.9. (a) Effect of wing size in a linear model for minimum area requirements (MAR) of European butterflies (N= 142 species). Dotted lines illustrate the 95 % confidence interval. Red cross is the value of the Xerces Blue (*Glaucopsyche xerces*). (b) Categories of minimum area requirements (MAR) [from Baguette & Stevens, 2013].

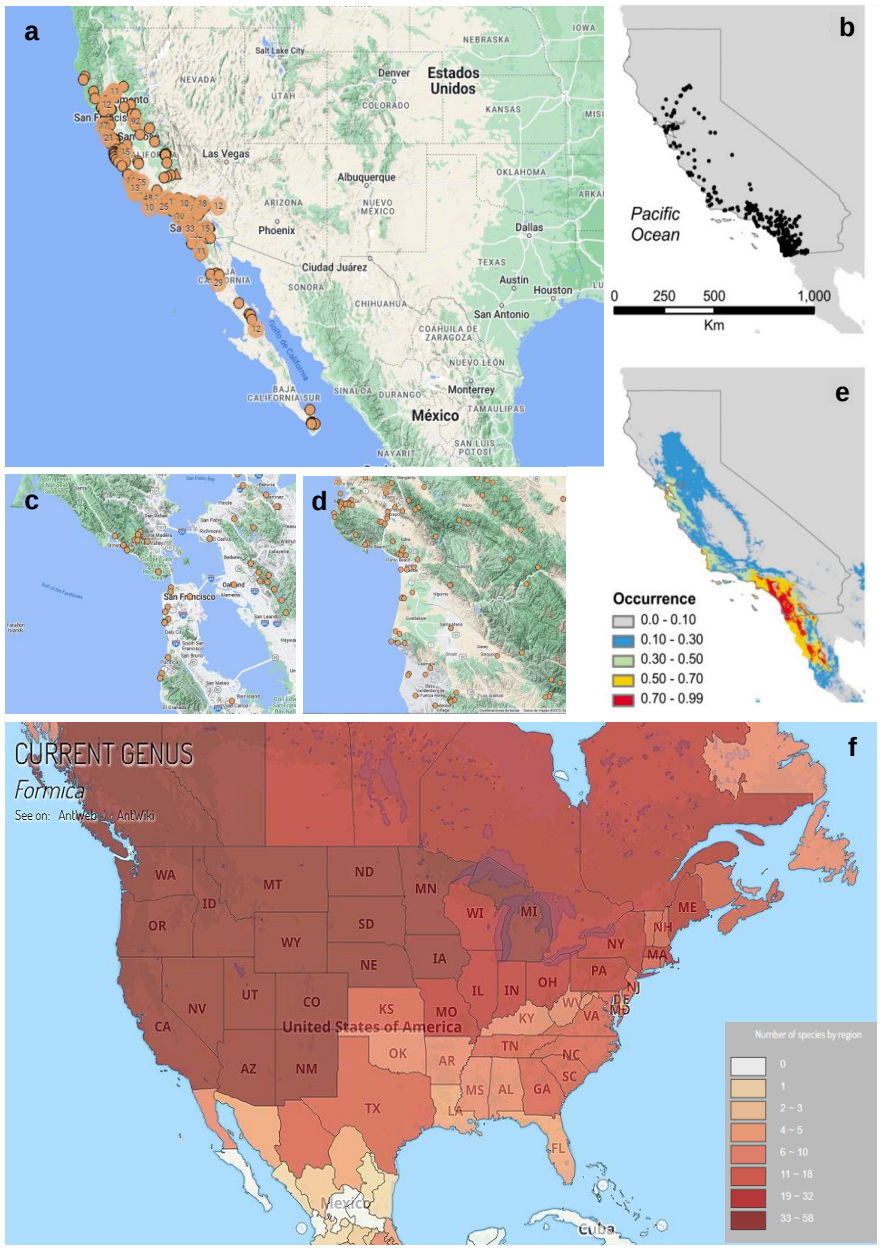


Figure 5.10. (a) Potential distribution of The Xerces Blue butterfly (*Glaucopsyche xerces*) according to the presence of its associated plant, the common Deerwood (*Acmispon glaber*). (b) Occurrence localities of *Acmispon glaber.* (c) Detail of San Francisco Bay including Point Reyes National Seashore and Golden Gate National Recreation Area where some plants are still preserved. (d) Detail of Guadalupe-Nipomo Dunes National Wildlife Refuge with plants as well. (e) Modeled potential distribution of *Acmispon glaber*. (f) Abundance of ant species of the genus *Formica* in North America, with some of which the butterfly establishes mutualism during the larval period [From: Consortium of California Herbaria https://cch2.org/portal/ (a, c, d) Riordan & Philip, 2009 (b, e), and https://antmaps.org/ (f)].

5.3.2. Blues of speciation

One of the basic axioms of evolutionary theory is that speciation occurs, at least initially, in allopatry or peripatry but not in sympatry (Mayr, 1963). There has to be an isolation between two populations of a species so that they can start to diverge until they speciate. The few cases of sympatric speciation identified at present are very problematic to accept, although there are notable works that address it (Busby et al., 2017). What can obviously happen is that two populations of a species diverge allopatrically, and once they have developed a partial barrier to reproduction, they come into contact again and exchange genes. In fact, it was seen that this process is more frequent in evolution than previously thought, and we can find examples even in our own species. Europeans conserve around 2% of Neanderthal genes (Green et al., 2010) and Melanesians conserve 4-6% of Denisovan genes (Reich et al., 2010).

But the casuistry is enormous and we can face situations that are more difficult to interpret. From the results of this thesis we have found that the butterflies Xerces Blue (*Glaucopsyche xerces*) and Silvery Blue (*Glaucopsyche lygdamus*) were two clearly differentiated species that apparently lived in sympatry but without gene flow. Therefore, two questions arise: (1) how did reproductive isolation occur and (2) how, once contact was restored, isolation was maintained.

Using two different approaches, one molecularly calibrated (Espeland et al., 2018) and the other calibrated from fossil remains (Chazot et al., 2019) we have estimated the divergence time of the Xerces Blue from the Silvery Blue. With the first approximation we obtained that the divergence occurred 900,000 years ago (560,895-1,159,512 years 95% HPD interval). The second approximation provided an earlier divergence of 2.40 million years ago (2.14-3.08 Ma 95% HPD interval). What happened in these periods?

If instead of the mean we take into account the range, the first of the two divergence periods coincides with the Chibanian, traditionally designated as the Middle Pleistocene. Its beginning

770,000 years ago is marked by an inversion of the Earth's magnetic field called Brunhes-Matuyama. It was characterized by global cooling, lengthening of the glacial stage, and a change in ocean circulation, which, in fact, is believed to cause a global mass extinction of more than 100 species (20%) of calcareous benthic foraminifera (Kender et al., 2016).

The second estimated range of divergence coincides with an even more interesting geological and evolutionary period: the beginning of the Gelasian 2.5 million years ago. It is about the transition between the Pliocene and the Pleistocene, when the ice sheets in the Northern Hemisphere began to grow, which is seen as the beginning of the Quaternary ice age. It is estimated that during this period the Earth experienced about 20 glacial cycles of varying intensity.

Both periods of divergence therefore coincide with glaciations; climate changes significant enough to completely transform ecosystems, extinguish species and isolate populations that, over time, generate new species. Jennings et al., (2007), for example, analyzes the evolution of ice caps during the Middle Quaternary in North America and shows the way they transform and isolate the landscape.

Although the distribution of Xerces Blue shortly before its extinction was located exclusively in the vicinity of San Francisco Bay, The Silvery Blue has a much wider distribution that reaches Canada and was surely affected by the glaciations. Furthermore, an ice age alters the climate far beyond the areas covered by the ice cap. The pattern of precipitation and vegetation also changes in mid-latitudes, away from the ice. It is reasonable to think that these transformations isolated butterfly populations, especially if we take into account the moderate dispersal capacity of these non-migratory species.

Consequently, the divergence could be caused by the isolation of populations as a direct or indirect result of some glaciation. It would therefore be a case of orthodox allopatric speciation.

Once speciated, the climatic conditions changed, the ecosystems were transformed and the butterflies were able to expand their distribution ranges until they came into secondary contact. What

are the barriers that prevent the mating of two genetically close butterflies (sister taxa) that live in the same region?

When we study butterflies it is easy to experience an anthropocentric bias. We well understand a river, a desert, or a mountain range as physical barriers that are difficult to overcome, and therefore capable of isolating populations of lions or humans, but the physical barriers that separated butterfly populations are often invisible to our eyes. They move on another spatial scale where local conditions must be much more relevant. The wind, for example, is a very important barrier in butterflies. The Strait of Messina, which separates the Italian peninsula from Sicily by only 3 km, is a passage feared by sailors and a practically insurmountable barrier in the gene flow in butterflies (Scalercio et al., 2020).

North America is currently inhabited by two species of butterflies of the *Glaucopsyche* genus: *G. lygdamus* and *G. piasus*, with very wide distributions that, on a large scale, seem overlapping (Figure 5.11). Additionally, each has many described subspecies, some locally extinct or threatened. With the progress of genomics, the taxonomy of this group will surely change and some subspecies may acquire the rank of species, as happened between *G. xerces* and *G. lygdamus*. The current situation in the USA is, therefore, analogous to that which occurred at the beginning of the 20th century in San Francisco, since we have different butterfly populations (at least two species), partially coexisting, in principle, without hybridizing.

The wide distribution of *G. lygdamus* allows its larva to feed on very diverse plants of the genera *Astragalus, Hedysarum Lathyrus, Lotus, Lupinus, Thermopsis, Vicia*, as well as the species *Medicago sativa*, *Melilotus alba*, *Oxytropis lambertii* and *Acmispon glaber* (Scott, 1986), while *G. xerces* seems to have fed mainly on *Acmispon glaber,* but also *Lupinus arboreus*, *L. micranthus* and *Astragalus* menziesii (Downey & Lange, 1956). It is also known that as myrmecophilous species, the *Glaucopsyche* establish mutualistic relationships with ants. It has been observed that *G. lygdamus* has a preference for the species *Formica altipetens* (Pierce & Eastealt, 1986). Although *G. xerces* also maintained mutualism with some ants of the same genus, the fact that some

33-58 species of ants of the genus *Formica* have been identified in California opens the possibility that the reproductive barrier between both butterflies was underground, in the anthills that each one used. Myrmecophily requires a high ecological specialization in the form of chemical and acoustic signalling, including particular organs and behavioural adaptations. Thus, it seems plausible that hybrids with intermediates characteristics, if they existed, would have had a low fitness and reinforcement would have fostered the emergence of pre-mating barriers between the species.

Another possibility is that the reproductive barrier was inside the butterflies' cells. 70% of insects are infected by an intracellular bacterium of the genus *Wolbachia*. Their interactions with hosts are often complex and, in some cases, have evolved from parasitism to mutualism. Some hosts cannot reproduce, or even survive, without *Wolbachia* colonization. It seems that it can become a very effective interspecific reproductive barrier, in the case that one species or population has it and the other does not, or even when both populations are infected by different strains (Telschow et al., 2005). However, since genomic analyzes of both butterflies have revealed that they were not infected, this hypothesis is considered unlikely.

We can also think that the reproductive barrier was in the reproductive system itself, which evolved to become incompatible between both species. A usual mechanism in butterflies is the so called “lock-and-key hypothesis” of the genitalic structures mechanically preventing interspecific mating. Downey & Lange (1956) provide a description of *G. xerces* accompanied by detailed illustrations of its genitalia. The authors indicate that they are very similar to *G. lygdamus*, and that based on this feature alone, both butterflies should be considered the same species.

But perhaps the reproductive barrier between the Xerces Blue and the Silvery Blue is so obvious that it has gone unnoticed. In fact, we even use it to name them: the blue. All species of the genus *Glaucopsyche* stand out for their shades of blue. And the entomologists, shrewd observers, had to specify the descriptions by labelling the colour blue with all kinds of nuances.

Already in the first historical description, the French entomologist Boisduval (1852) says: "Aloe over the violet-blue sea", and adds "above the wings of the male of the same blue as in our Alexis, of which it has the size and the port". His reference in describing the American butterfly *G. xerces* was the European butterfly *G. alexis*, just as curiously we did at the genomic level. Later, Tilden (1956) compares the coloration of Xerces Blue and the Silvery Blue like this: "This association is easily seen to be faulty when specimens of all are at hand, since the soft lavender blue of all forms of xerces is quite different from the cold blue of lygdamus subspecies". And finally Downey & Lange (1956) make another description of *G. xerces*: "The upper surface shade of blue in the males is also quite variable. It is influenced by many factors including the wear of the specimen; fuscous infusion, particularly in distal areas and along wing veins; white scales on wing veins; and light [...] In general it may be stated that the blue of the males [Xerces] is of a lighter shade than that found in adjacent races of *G. lygdamus* ".

Aloe over the violet-blue sea, soft lavender blue, cold blue, lighter shade, silvery blue... these are beautiful descriptions but in reality unscientific to define a colour. The objective way to define it is based on its physical properties.

Generally diurnal butterflies have very different wing coloration between the dorsal and ventral surfaces. The traditional interpretation of this difference is that the dorsal part of the wings is used for intraspecific recognition, is subject to sexual selection and, therefore, shows all kinds of colorations and designs, while the ventral part of the wings would have a function of avoiding predation, either by crypsis or disconcerting predators with false "eyes", that is why they present usually earthy colours and have ocelli. However, in nature there are more exceptions than rules, and the ocelli may also be subject to sexual selection and the dorsal surface of the wings may also show aposematic colours to warn predators of their toxicity. There are even mimetic butterflies that, without being poisonous, imitate the coloration of those that are. In any case, it is accepted that when a species of butterfly shows marked sexual dimorphism in dorsal surface coloration, the colours must play some role in mating. Butterflies of the genus

Glaucopsyche fit this pattern, with males displaying a more intense blue than females.

The blue colour is not very abundant in nature, in fact it is not achieved from a pigment as in the case of green from chlorophyll, yellow and orange from carotenoids, or red from anthocyanins. It is a structural colour that is achieved from micro and nanostructures of the scales (or feathers) of the wings, which only reflect a certain wavelength. That is why it is a colour that changes according to the incidence of light and has an attractive shine.

Unlike butterflies, humans have not evolved to detect subtle colour differences in the blues of the wings, nor can we see in the ultraviolet range, but we have the technology to measure it objectively. With a spectrometer we can measure the reflectance spectrum and carry out comparative studies. This approach was already made by Giraldo et al., (2016), which measured reflectance in various species of butterflies of the *Morpho* genus, known throughout the world for their size and spectacular iridescent blue (Debat et al., 2018). The results obtained confirm that the type of reflectance is correlated with the microscopic architecture of the wings, and this, in turn, with the phylogeny (Figure 5.12). Another study of nine closely related polyommatine butterflies, living in the same habitat, with the males of all species possessing blue dorsal coloration on their wings found both structural and spectral species-specific differences in their blues. In addition, the temporal distribution of the species throughout the year shows that the flight periods of the species with somewhat similar colours were well separated (Bálint et al., 2012).

Thus, perhaps the reproductive barrier that we have not been able to detect between the Xerces Blue and the Silvery Blue lies in a subtle reflectance of the blues on their wing surfaces. Butterflies would be able to distinguish these blues and choose the mate accordingly.

As it is a non-destructive method, a study could be undertaken to measure the reflectance in all the specimens of both butterfly species kept in various collections in the USA (for instance, there are *G. xerces* in Bohart Museum, California Academy of Sciences,

the Harvard Museum of Natural History, The Florida Museum and Smithsonian Institution).

If there were significant differences, the last step would be to identify the genes associated with these micro and nanostructural changes in the wing scales. And in this field there are also precedents (Zhang et al., 2017; Brien et al., 2018; Thayer et al., 2020). Our contribution would be to extend these studies to paleogenomics.

The maculation pattern on the underside of butterflies' wings has been further studied at the genetic level (Zhang & Reed, 2016; Mazo-Vargas et al., 2017; Fenner et al., 2020; Banerjee et al., 2021; among many others). However, it does not seem to be very useful in determining the reproductive barrier between *G. xerces* and *G. lygdamus* for two reasons. In the first place, because the population of *G. xerces* was polymorphic in the reverse wing, with up to 5 different variants being described: from white spots without ocelli to completely ocellated spots passing through all the intermediate forms. This variability is much higher than that found within *G. lygdamus* itself. Second, a historical article explains a field observation in which males and females of *G. xerces* were found copulating with crossed spot patterns: ocellated male and non-ocellated female, and ocellated female and non-ocellated male (Downey & Lang, 1956). This observation could indicate that the ocelli were not a sexually relevant trait.

In summary, blue colour is the most promising trait for identifying the reproductive barrier that kept the Xerces Blue and the Silvery Blue from gene flow. Sadly, one of the species was lost and one of the blues remains silent.

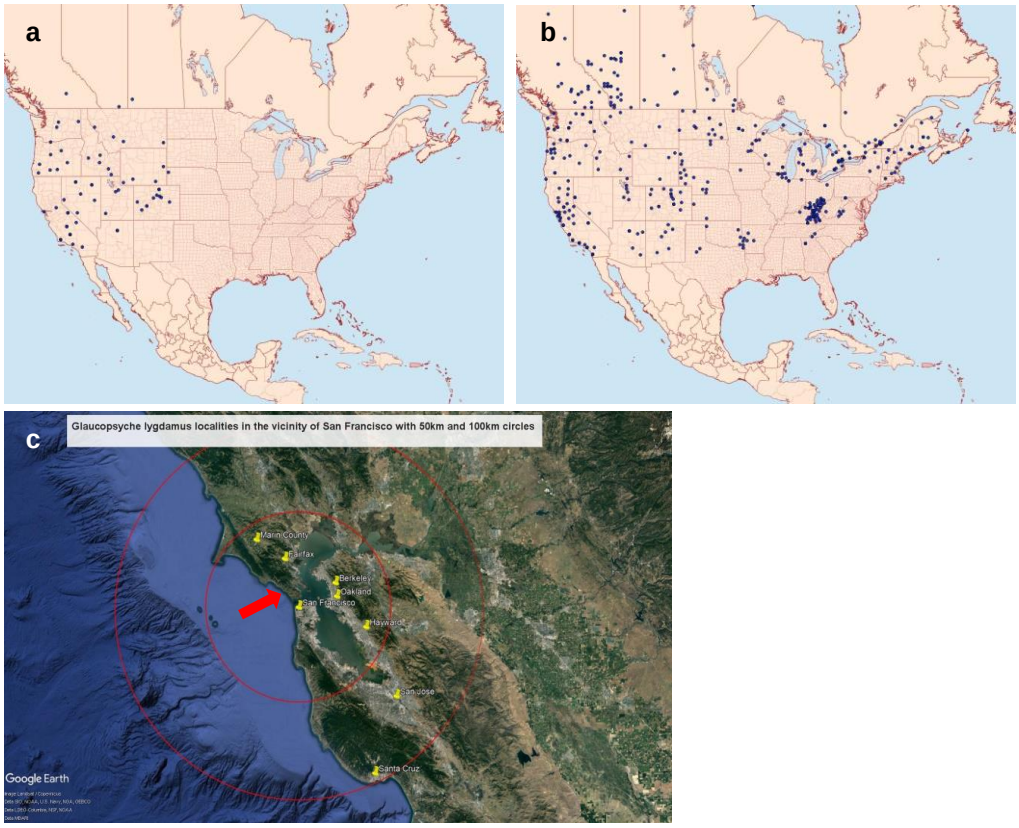


Figure 5.11. (a) Occurrence of *Glaucopsyche piasus.* (b) Occurence of *Glaucopsyche lygdamus.* Note that there is some overlap in the distribution of both butterflies. (c) Detail of San Francisco Bay with the localities where historical samples of *Glaucopsyche lygdamus* where obtained (yellow thumb tacks) and their distances from San Francisco city (red circles). Historical sampling site and last observation of *Glaucopsyche xerces* (red arrow) [from Digital Guide to moth identification, North American Moth Photographers Group, Mississipi Entomological Museum (a, b). Robert K. Robbins, Smithsonian Institution (c)].

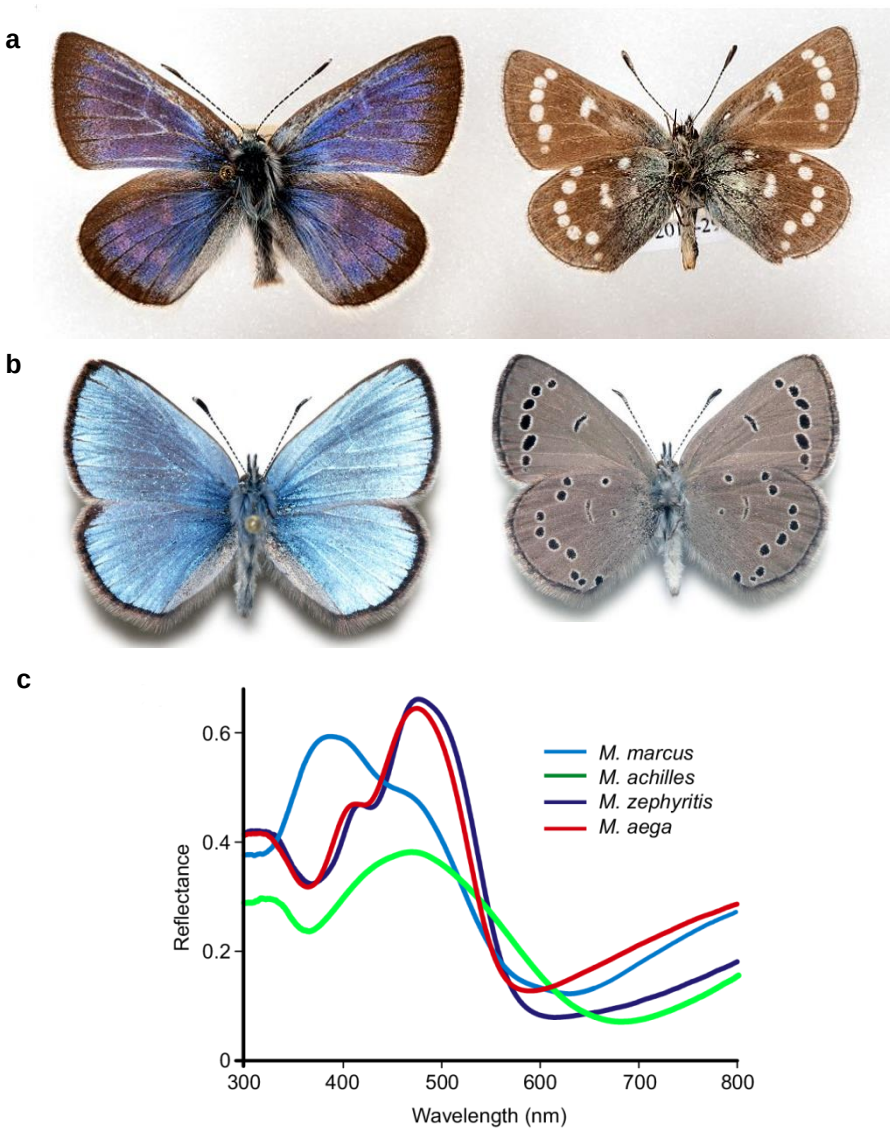


Figure 5.12. (a) Dorsal and ventral wing surfaces of Xerces Blue (*Glaucopsyche xerces*). (b) Dorsal and ventral wing surfaces of Silvery Blue (*Glaucopsyche lygdamus*). Note the pattern of white spots, often without black ocelli, in the first. For the clarity of the comparison we have chosen especially different specimens, but very often the differences in coloration between both species are very subtle. (c) Dorsal wing reflectance spectra of four *Morpho* species measured with a spectrometer and an integrating sphere [from Florida Museum (a), Kim Davis, Mike Stangeland & Andrew Warren; www.butterfliesofamerica.com (b), Giraldo et al., 2016 (c)].

5.3.3. Genetic diversity and population size

Pushed by an unlimited positivism, or perhaps by a hypertrophic vanity, we tend to think that rationality will allow us to access all the secrets of the Universe. Only time and effort are needed for all the questions to be solved. But reality is stubborn and in its progress science occasionally encounters seemingly insurmountable obstacles. The mythical aura of these inexplicable phenomena increases as unsuccessful attempts to find solutions accumulate.

Neutral theory of molecular evolution predicts that genetic diversity increases with population size, yet observed levels of diversity across metazoans vary only 3 orders of magnitude while population sizes vary over 12 (Buffalo, 2021). This unexpectedly narrow range of diversity is known as Lewontin's Paradox of Variation (1974) and remains unresolved through the genomics era (Buffalo, 2021). We fail to understand the processes that connect a central parameter of population ecology, census size (N_c), to a central parameter of population genetics, effective population size (N_e) (Buffalo, 2021). Both disciplines, genetics and ecology, provide true visions of reality, but Truth eludes (Figure 5.13).

Buffalo, 2021 revisit Lewontin's Paradox by integrating several data sets in order to compare the observed relationship between diversity and census size with the predicted relationship under different selection models. The study analyzes 172 metazoan taxa and obtains a weak relationship between pairwise diversity (data from Leffler et al., 2012, Corbett-Detig et al., 2015; Romiguier et al., 2014) and approximate population size (Figure 5.14, a). As expected, when the results are analyzed according to phylum, it is clearly observed that the Chordata have the lowest genetic diversities and population sizes, while the Arthropoda have the highest.

The analysis is performed on a set of extant species and is static (snapshot). But we could also include extinct species, which by definition have a null population size and are, therefore, projected directly on the ordinate axis. One could even use this visualization for a dynamic analysis (film) and see how the relationship between genetic diversity and population size changed over time, both for

species that became extinct and for those that suffered a population bottleneck. In Figure 5.14, c, we show the different theoretical pathways by which a species can become extinct. In principle, the faster the extinction, the less signal it will leave in the genome, and the more progressive it is, the more signals of genetic erosion will be accumulated, such as runs of homozygosity (ROH) (Ceballos et al., 2018).

In the present thesis we have analyzed the genome of the extinct butterfly Xerces Blue (*Glaucopsyche xerces*), compared it with that of its congener Silvery Blue (*G. lygdamus*), and detected clear signs of genetic erosion. We found that Xerces Blue had 22% less heterozygosity on average than the Silvery Blue historical samples, a difference that is statistically significant (T-test; p=0.0072) (see section 4.3). We searched for runs of homozygosity (ROH) that can indicate the existence of inbreeding in a dwindling population. The total fraction of the genome presenting ROH, although limited, is much higher in Xerces Blue (up to 6% of the genome) than in Silvery Blue, especially in short ROH of size between 100 and 500 kb, consistent with background inbreeding. The limited presence of long ROH discards consanguinity as a common scenario in Xerces Blue. Therefore, we would be facing an extinction process of intermediate intensity (speed) (orange arrow in Figure 5.14, c). An example of high intensity (speed) characterized by leaving no traces in the genome can be found in another species analyzed in this thesis, the Carolina parakeet (*Conuropsis carolinensis*) (see section 4.1), which we believe disappeared in a matter of decades (red arrow in Figure 5.14, c). And at the opposite extreme, some species have gone extinct in long processes during which they have accumulated many signs of genetic erosion such as the last Mammoths from Wrangel Island (Rogers & Slatkin, 2017), or the Altai Neanderthal (Prüfer et al., 2014) (yellow arrow in Figure 5.14, c).

When a species experiences a significant reduction in population size and subsequently recovers, we say that it has undergone a population bottleneck. A bottleneck could be considered as a truncated extinction process. In Figure 5.14, d, we show the main itineraries that a species can follow when it undergoes a population

bottleneck. The loss of population size can be more or less accentuated and the impact on the genome is also variable. Once the decline is over, the population recovers, but the final diversity is impoverished compared to the initial situation.

The relationship between genetic diversity and population size has great potential in conservation. Buffalo (2021) represented the pairwise diversity- population size of meatzoans according to their International Union for the Conservation of Nature (IUCN) Red List conservation status (Figure 5.14, b). It is clearly seen that the vast majority of the taxa in the list are Chordata and that the classification is made solely on the basis of population size and does not take into account genetic diversity (margin boxplots). Díez-del-Molino et al., (2018), previously detected little concordance between genome diversity, population sizes and conservation status in a much smaller data set.

But if instead of the static analysis we use the dynamic analysis proposed above, the potential for conservation increases significantly, because we can monitor not only the obvious loss of population size, but also the loss of genetic diversity. However, to perform this type of analysis, references are needed.

These references can be obtained by comparing, for example, average heterozygosities in threatened taxa with those in taxonomically related nonthreatened taxa (Spielman et al., 2004). They can also be determined experimentally, such as a study that exposed the butterfly *Bicyclus anynana* to different degrees of bottlenecks (Saccheri et al., 2001). And finally, references can be obtained from Museomics (Lalueza-Fox, 2022; Díez-del-Molino et al., 2018; Staats et al., 2013), i.e., from the comparison between the genetic diversity of current specimens of threatened species and the diversity of specimens preserved in museums before the onset of the population decline (Figure 5.14, e). In this way it will be possible to know how genetic diversity changes as the species moves from the Least Concern zone (LC) to the Critically Endangered zone (CR) (upper margin boxplots) and act accordingly.

The genomic patterns of population decline of Xerces Blue was obtained by combining two of these procedures: comparison with a

taxonomically related nonthreatened taxa and museomics. And can be used to monitor other endangered insects, which so far have not received much attention.

Of the 69,903 species of vertebrate (mammals, birds, reptiles, amphibians and fishes), 48,101 (69%) have had their risk of extinction assessed via the IUCN Red List process. In contrast, 8,121 out of 1,000,000 (0.8%) known insect species have been through the same process. According to comprehensive review of 73 historical reports of insect declines from across the globe, the rates of decline may lead to the extinction of 40% of the world's insect species over the next few decades (Sánchez-Bayo & Wyckhuys, 2019).

The abundant collections of insects gathered over centuries and housed in museums around the world should serve not only for educational purposes, but also as a valuable source of aDNA. Studying them we will not resolve the Lewontin's paradox but we will transform it into another rhetorical figure, the allegory of biodiversity conservation.

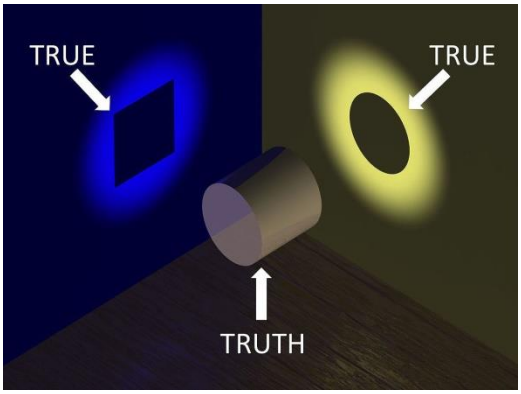


Figure 5.13. The views of reality provided by population genetics and population ecology are true, but the Truth eludes in the form of Lewontin's Paradox of Variation.

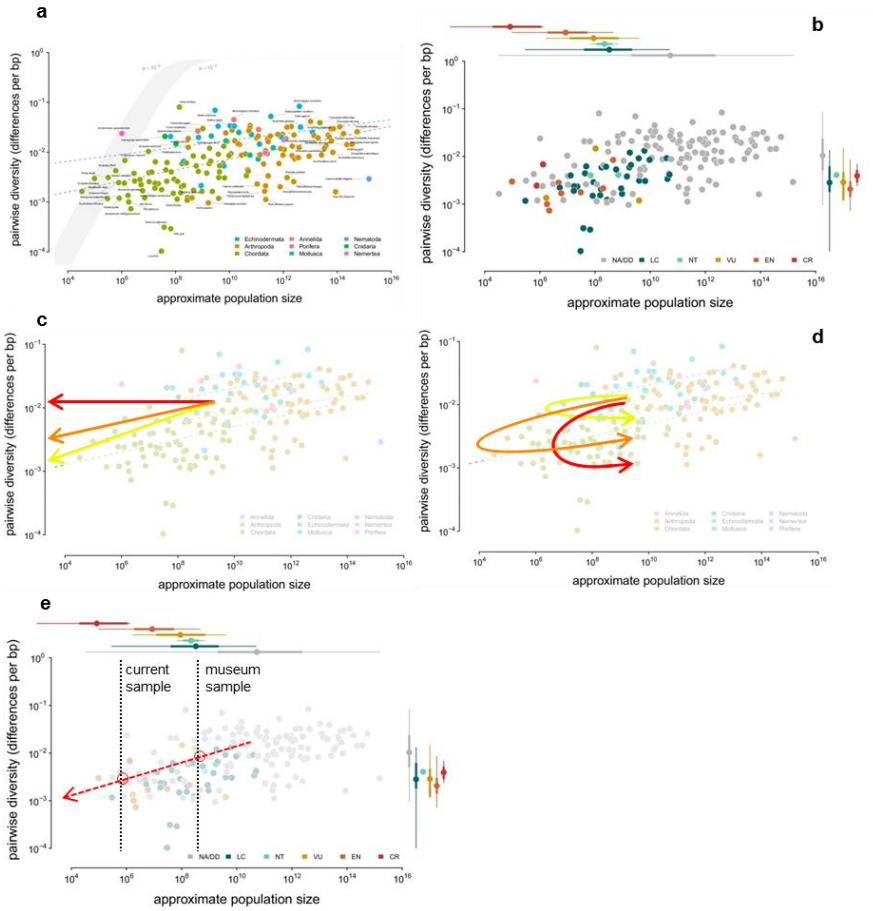


Figure 5.14. (a) Visualization of Lewontin's Paradox of Variation. Pairwise diversity, which varies over three orders of magnitude, shows a weak relationship with approximate population size, which varies over 12 orders of magnitude. The shaded curve shows the range of expected neutral diversity if Ne were to equal Nc, for two mutation rates. The dark gray dashed line is the regression fit, and the blue dashed line is the regression fit using a phylogenetic mixed-effects model. Points are colored by phylum. The most relevant are Chordata (green) and Arthropoda (brown). (b) A version of the previous figure with points coloured by their IUCN Red List conservation status. Margin boxplots show the diversity and population size ranges (thin lines) and interquartile ranges (thick lines) for each category. NA/DD indicates no entry, or Data Deficient; LC is Least Concern, NT is Near Threatened, VU is Vulnerable, EN is Endangered, and CR is Critically Endangered. Note that only some Chordata have been classified by IUCN, and their status is determined by population size and not by genetic diversity. (c) Conceptual diagram of the effect of extinction speed on the genome of a given species; fast (red arrow), intermediate (orange arrow), and slow (yellow arrow). (d) Conceptual diagram of the effect of population bottleneck in the genome of a given species. Only three contrasting situations are illustrated. The population decline can be abrupt and brief with little impact on the genome (yellow arrow), gentle and prolonged with limited impact on the genome (orange arrow), and gentle and brief with strong impact on the genome (red arrow). After population recovery genome diversity remains reduced. (e) Samples preserved in museums allow establishing the decline of endangered species [modified from Buffalo, 2021].

5.4 De-extinction

The Phoenix was, according to Greek mythology, a fantastic bird endowed with immortality. Every 500 years it died consumed by fire (self-ignition), and was reborn from its own ashes (Van der Broek, 1972).

De-extinction is a procedure that aims to reverse the process, in principle irreversible, of the extinction of a species (Richmond et al., 2016). Apparently it defies thermodynamics, which is why it arouses both astonishment and admiration. It would seem that science ventures beyond its limits and acquires the biblical power to raise the dead, like Lazarus, or more precisely, the mythological power to make them reborn from their own ashes. All things considered, the first step in the de-extinction of an animal is the recovery of its genetic material, very often from some ground bone remains. Although the bones in powder are not ashes, it seems sufficiently, both for its aspect and for its composition, mostly mineral.

However, when we analyse the detail of the de-extinction, we discover that there is nothing biblical or mythological about it, and it falls perfectly within the limits of thermodynamics, biochemistry, palaeontology, biology (mainly genetics and embryology) and computing. It is therefore the result of a fruitful and stimulating union of various disciplines, whose base is, at the same time, a synergy: paleogenomics (palaeontology and genomics).

At the beginning of paleogenomics, an excess of enthusiasm in the scientific community led to the belief that all genetic material could be recovered and species as mythical as the dinosaurs could be de-extinct. This boundless optimism inspired Michael Crichton's novel Jurassic Park, which Steven Spilberg later made into a film with overwhelming success. Unfortunately, reality is not always stranger than fiction, and today we know that when an organism dies, its DNA degrades. The warmer the environment, the faster the degradation. In the best of cases, aDNA can be preserved for up to 1 million years (see section 1.2.3). Therefore, this is the horizon of de-extinction. We will have to settle for a Pleistocene Park.

However, even in the best possible environmental conditions, aDNA does not remain intact, but instead undergoes progressive and unstoppable degradation. This fact rules out the first path that we could propose for de-extinction: the cloning of Pleistocene animal cells frozen in the Siberian permafrost. Specimens of mammoths, horses, cave lions or wolves have been recovered, tens or even hundreds of thousands of years old, with an exquisite state of preservation, with hair, whiskers on the snout and delicately closed eyelids as if the animal was sleeping (Fisher et al., 2014; Solly, 2019a; 2019b; Boeskorov et al., 2021). But neither the soft tissues are a good reservoir of aDNA, nor the good reservoir of some bones, prevent the degradation of the genome. The aDNA is always recovered very fragmented and it is necessary to map it, using some closely related species as a reference. This condition already limits from the outset which species are candidates for de-extinction. A species with no close living relative, such as *Macrauchenia patachonica*, a rare Pleistocene South American mammal distantly related to perissodactyls (Chimento & Agnolin, 2020), cannot be accurately mapped, because many genes might not have an exact homologue in a reference DNA, and they go unnoticed.

If all goes well, and it is possible to map the genome of the extinct species (currently it has only been achieved 17 times, see section 1.3.2), it is obtained *in silico* (the FASTA code). Then, it is necessary to obtain it *in vitro* and finally *in vivo*.

There is a particular situation in which the genome of the extinct species apparently survives *in vivo* as part of existing populations of certain closely related variant or subspecies. In such a case, selective breeding would allow the genome of the extinct species to be brought together in a single individual. We highlight two examples with predictably different results. The quagga and the aurochs.

The quagga (*Equus quagga quagga*) was a peculiar zebra that only showed the characteristic striped pattern on the front half of the body, and the background fur was brownish instead of white. It lived in South Africa until it became extinct at the end of the 19th century as a result of excessive hunting (see section 1.2.3). In

1955, the German zoologist Lutz Heck, apparently observing populations of the plains zebra (*Equus quagga bohemi*) that had been parapatric with the quagga, detected a certain variability in the pattern of the stripes, and the brownish background in many specimens, and had the intuition that by selective breeding, the quagga could be recovered (Heck, 1955). Subsequently, in 1984 mDNA was isolated from the quagga and phylogeny confirmed that it was indeed a subspecies of the plains zebra (Higuchi et al., 1984, see section 1.2.3). With this background, in 1987 the South African naturalist Reinhold Rau started a project to recover the quagga through selective breeding. After 35 years of selection, the F4 and F5 specimens are practically indistinguishable from the true quagga (www.quaggaproject.org/). However, there is a high degree of interindividual variability, indicating that the phenotype has not been fixed yet. But there is a serious additional obstacle. In 2014, the complete genome of the quagga with a very detailed phylogeny of equids was published (Jónsson et al., 2014). The study found a gene in the quagga, absent in plains zebra populations, that it seemed to have a relevant adaptive function. Therefore, this gene is impossible to recover through selective breeding. The neoquagga will never be the quagga, no matter how much it looks like it.

The case of the aurochs is different. Initially, the Heck brothers carried out selective breeding programs for various breeds of rustic cows in order to recover the aurochs phenotype (Heck, 1951). The result was the Heck cow obtained in the 1920-1930s, quite similar to what the original aurochs is believed to have looked like (although smaller in body size and horn length). The project was heavily criticized. In 2015, the complete paleogenome of an aurochs from the Britain was sequenced and it was found that it had hybridized with different breeds of British cows (Park et al., 2015). Some authors consider it possible that the complete genome of some population of aurochs survives distributed among several current bovine breeds (Sinding & Gilbert, 2016) and, consequently, this species could become de-extinct through selective breeding, with genomic supervision. In fact, the True Nature Foundation, which is driving the Uruz aurochs back-

breeding project, has stated that it will be using the published aurochs genome as a template for breeding (Stokstad, 2015), or for genome editing of hand picked individuals in the breeding program (True Nature Foundation, 2016).

In fact, the case of the aurochs, if confirmed, would be exceptional. Species that have gone extinct generally diverged enough from their sister taxa to have unique genetic variants in genes.

We can do a couple of thought experiments to further explore selective breeding as a putative de-extinction system. We know that on average European individuals have around 1-3% of the Neanderthal genome and that up to 20% of the Neanderthal genome is conserved among humans (Vernot & Akey, 2014). Even if we knew which Neanderthal genes each human is a carrier of (something highly expensive) and managed to get them to reproduce selectively (something highly immoral), we would only get a modern human with 30% Neanderthal genes. The remaining 70% has been lost forever. On the other hand, it has been sequenced the complete paleogenome of the extinct cave bear (*Ursus spelaeus complex*) and it was found some admixture with the brown bear (*Ursus arctos*), so that the current populations of brown bear conserve between 0.9 to 2.4% of the extinct bear genome (Barlow et al., 2018). It is highly probable that the entire cave bear genome is not conserved among the different current brown bear populations, but even if it were, a program of selective breeding of brown bears to de-extinct the cave bear is unthinkable.

The alternative de-extinction procedure is gene editing. The idea is apparently simple: starting from the reference genome, changes are progressively introduced until it becomes the paleogenome. This procedure was only feasible after the introduction of the CRISPR genetic editing technique, basically discovered by the Alicante microbiologist Francisco J.M. Mojica (Mojica & Montoliu, 2016). But it has limitations. Although it is possible to precisely control genome breakpoints, errors often occur, requiring analysis at each generation to ensure correctness.

The number of differences between the paleogenome and the reference genome can be very large, even between closely related species. For example, between the mammoth and the Asian

elephant, 1.4 million nucleotide changes have been identified. Introducing several tens to several hundred changes in each generation by the CRISPR technique makes gene editing unfeasible (Campa et al., 2019). That is why it is necessary to have the annotation that avoids carrying out genetic editing blindly, and allows it to be directed only to modify genes that encode relevant differences. The mammoth annotation reveals that there are 2020 amino acid changes that affect 1642 proteins (Lalueza, 2018). This number of changes, although still very large, is already beginning to make gene editing manageable. However, it must be taken into account that having a good annotation of a genome is almost as arduous as the gene editing itself.

The artificial synthesis of DNA is another very promising way to achieve the de-extinction of a species. Since in 1979 the Nobel Prize winner HG. Corana synthesized a 207-nucleotide gene in the laboratory, the size of synthetic genomes has been increasing exponentially to reach almost a million nucleotides (Lalueza, 2018), but despite progress, artificial synthesis is still far from being able to produce an entire eukaryotic genome *de nuovo* with corresponding chromosomal architecture.

The editing of the reference genome to transform it into the paleogenome by the CRISPR technique can be carried out in adult cells (somatic cell lines, especially fibroblasts), in gametes, or in embryonic cells. In all three cases it is necessary to analyze the success obtained by different means before obtaining a new generation. In the embryonic cells, depending on the technique used, mosaic individuals or chimera individuals are obtained, and it is necessary to carry out crosses with wild type and genotype to identify the specimens that have incorporated the changes correctly (Garcia & Pujol, 2016; Pujol, 2016). When the edition has been completed after n generations, the specimen that was wanted to be de-extinct is directly obtained. It is evident that if we work with animals with a long life cycle, this approach does not seem adequate. In contrast, gene editing in adult cell lines has the advantage that the generation cycles are very short, and techniques such as fluorescence can be used to identify the cells that have incorporated the changes correctly. However, when the

editing process is complete after n generations it is necessary to clone to obtain the extinct specimen.

Cloning is not a simple procedure either. First of all, it is necessary to enucleate an oocyte of the surrogate species, transfer the cell nucleus with the paleogenome, stimulate its development and transfer the embryo to the reproductive system of the female of the surrogate species. If the development occurs normally, a de-extinct specimen will be born from the surrogate female. Currently, the cloning of mice is the one that has the highest percentage of success and even so it does not exceed 20% (Ogura, 2020).

But de-extinction has an unsolvable basic obstacle. No matter how high the paleogenome recovery percentage is, and no matter how large the coverage is (the number of paleogenome fragments for a given position in the reference genome), it will not be possible to recover 100% of the paleogenome and it will not be identical to the original one. Repetitive sequences can represent 30% or more (Lalueza, 2018), and are impossible to map, but since they do not encode, in principle they do not represent a problem. The real problems are: (1) the interpretation mistakes in the mapping produced by various causes, such as, for example, the discordance of bases in different reads for a certain position in the reference genome, and above all (2) the existence of regions in the paleogenome that have become so differentiated that they cannot be mapped onto the reference genome. Lin et al., (2022), recover the paleogenome of the extinct Christmas Island rat (*Rattus macleari*) in the best imaginable conditions: the animal became extinct recently (1898-1908), therefore, the paleogenome had little time to degrade, moreover, they use as a reference not a single genome, but the genome of several existing rats, with which the Christmas Island rat is so closely related that it shares genus, and finally one of those genomes, that of the Norway brown rat (*R. norvegicus*), is of extremely high-quality, as it is an intensely studied model species. And yet, they fail to recover more than 95% of the paleogenome. The almost 5% of the lost genome does not correspond to repetitive sequences, but to specially differentiated zones of the genome, associated with immune response and olfaction. In other words, the lost zones of the genome are those

that precisely indicate divergent processes of adaptation, and consequently have great evolutionary relevance.

The conclusion is that de-extinction does not manage to recover the disappeared animal, in fact the most it achieves is to produce a transgenic animal, a genetic chimera with a variable degree of approximation to the original paleogenome. Perhaps we should start talking about semi-de-extinction or neo-de-extinction. Prefixes accumulate.

According to some authors (Novak, 2018), there are currently seven active de-extinction projects. However, excluding selective breeding, only two appear to use gene editing techniques: the mammoth (*Mammuthus primigenius*), and the passenger pigeon (*Ectopistes migratorius*). Both projects have little chance of success, since aside from the gigantic work of gene editing, they have not yet resolved the final stage of cloning. Currently, the list of cloned species is short, and does not include proboscideans or any birds.

Are the three species we have studied in this thesis good candidates for de-extinction? From the Tenerife giant rat (*Canariomys bravoi*) we were only able to recover a part of the nuclear and mitochondrial DNA, therefore it is discarded. We have recovered the complete genome of the Carolina parakeet (*Conuropsis carolinensis*), but at the moment, cloning is not feasible, therefore, it is also ruled out. Instead, the Xerces Blue butterfly (*Glaucopyche xerces*) proves to be an ideal candidate. In the Figure 5.15 we show a diagram of what the de-extinction procedure should be, after recovering the complete paleogenome.

Once a specimen has become de-extinct, several would have to be de-extinct as well to guarantee a certain degree of genetic variation in the population. Therefore, the whole process would have to be repeated several times from other samples of the extinct species.

The last step would be the reintroduction, first under controlled conditions, and finally in the natural habitat, with monitoring to ensure the proper adaptation of the species to the ecosystem.

All de-extinction initiatives stress the importance of recovering the original ecosystem through the biological activity of reintroduced species (Jorgensen, 2012; Seddon et al., 2014). The mammoth

with its intensive grazing would transform the unproductive arctic tundra into a highly productive herbaceous steppe, like the one that existed during the Pleistocene. Passenger pigeons would redisperse the seeds of many North American plant species and would contribute with their flammable depositions to generate periodic fires that would renew the forests (Lalueza, 2018). And what would Xerces Blue do? Basically splashing its iridescent blue again on the coastal sage scrub, which fortunately still has a large area (see section 5.3.1). In addition, it would predictably transform the local economy, attracting entomologists and naturalists from around the world, who would visit California to witness a unique event in the human history of biodiversity extermination: the rebirth of a species.

The etymological root of “Phoenix” is shared with “Phoenician”, since the Greeks, and later the Romans, associated this mythological bird with the Phoenician civilization, characterized by trading around the Mediterranean and producing a famous red-purple dye made from the secretion of various marine snails of the genus *Murex* (the compound is 6,6′-dibromoindigo). The extraction of this dye involved tens of thousands of snails and a lot of labor, as a result, the dye was expensive and became a sign of status, used to dye Caesar's toga and later some elements of the Pope's wardrobe. The word Phoenician would mean "those who work with red dyes" and "Phoenix" could mean "the Phoenician bird" or "the purple-red bird” (Van der Broek, 1972). Representations of the Phoenix are diverse, such as an eagle (*Aquila sp*.), the demoiselle crane (*Grus virgo*), or even as the golden pheasant (*Chrysolophus pictus*).

Fortunately, none of these species, neither the molluscs (intensely exploited), nor the birds, have become extinct and, therefore, for the moment, they will not have to be de-extinct from their ashes.

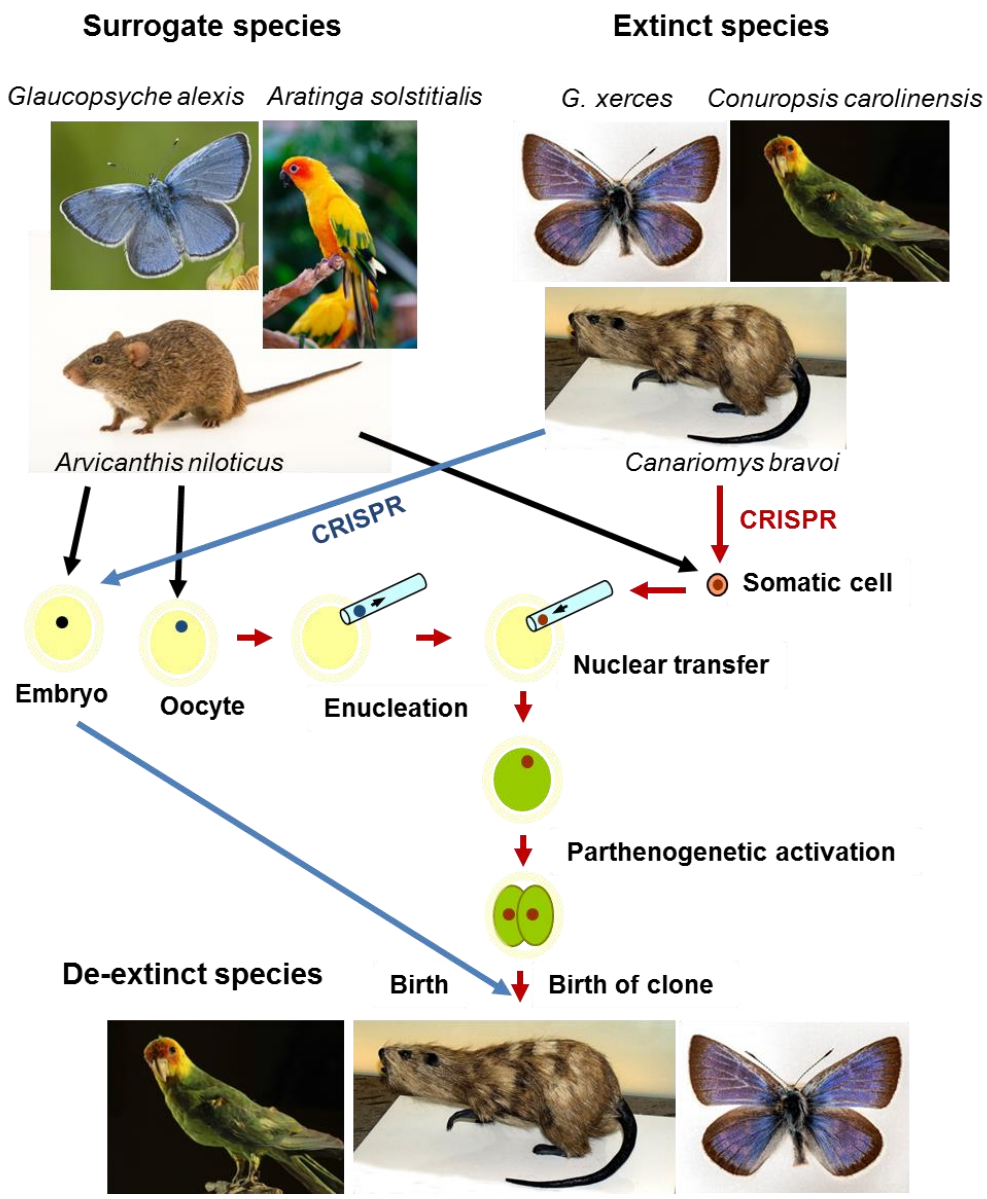


Figure 5.15. Conceptual diagram of the de-extinction of the three species studied in this thesis: Carolina parekeet (*Conuropsis carolinensis*), Tenerife giant rat (*Canariomys bravoi*) and Xerces Blue butterfly (*Glaucopsyche xerces*). Cloning pathway (red arrows) and direct pathway (blue arrows). The previous procedure of paleogenome sequencing, obtained by mapping on the genomes of the three surrogate species, is not shown [photos from Florida Museum, Marc Durà, Joel Sartore, Wikicommons & www.mascotarios.org/].

6. FUTURE LINES OF RESEARCH

Let us imagine all knowledgeable reality as the dial of a clock. What a person knows that he knows is at most (in fact, much less) the equivalent of the area covered by a minute arc. Adjacent to this tiny zone would be a zone of two more minutes, representing what a person knows that he does not know (be aware that we do not know at least twice as much as we know), finally the remaining 57 minutes represent the zone that a person does not even know that he does not know. As we expand what we know, we also expand what we know we don't know. This is how individual knowledge works and this is how collective knowledge, science, works as well. In this framework we can understand Socrates' famous phrase: "I only know that I know nothing".

This thesis has answered some questions, but it has inevitably raised many new ones. Probably, in the scientific activity almost as important as the discoveries are the new lines of research that it raises.

In a general scope, beyond all the evolutionary questions that paleogenomics will resolve over time, there is a basic biochemical question that has not been definitively resolved in 40 years: what mechanisms determine the degradation of ancient DNA? What is the relationship between depurination and deamination? Apart from the essential experimental approaches, it would be pertinent to take advantage of the great abundance of data to carry out more meta-analyses, especially those looking for correlations between the degree of fragmentation and miscoding lesions.

Within the specific scope of this thesis, several evolutionary questions have been raised, which are described below:

(1) Carolina parakeet (*Conuropsis carolinensis*)

- Our phylogeny places the Carolina parakeet related to the sun parakeet (*Aratinga sosltitialis*) located in South America. It would be interesting to recover aDNA of the extinct parrot species of the West Indies from museums and fossil remains and undertake a new phylogeny. There are already some precedents such as the recovery of the complete mitochondrial genome of the extinct

Cuban Macaw (*Ara tricolor*) (Johansson et al., 2018), or the phylogeny of extinct parrots from the Western Indian Ocean islands (Jackson et al., 2015). It would also be convenient to include in the phylogeny the different extant species and subspecies of the Mesoamerican parrots (Martínez-Gómez et al., 2017), especially the green parakeet (*Psittacara holochlorus*) that is parapatric with the original distribution of Carolina parakeet. In this way we would not only better understand the phylogeny but also the biogeography of the Carolina parakeet and reveal whether the dispersal to North America occurred through the Isthmus of Panama or by hopping across the Caribbean islands.

- Few fossil records of Carolina parakeet exist. Several Holocene bones from Illinois (Parmalee, 1958; 1967) and Ontario (Prevec, 1984) have been identified, together with middle Pleistocene bone from Florida (Kilmer &. Steadman, 2016). There is also a controversial upper Miocene bone from Nebraska classified as the chronospecies *Conuropsis fratercula* (Wetmore, 1926). It would be convenient to try to recover aDNA from these fossils to establish their phylogenetic relationships with the historical samples of Carolina parakeet and to obtain a more precise date of divergence, also by absolute dating of C^{14} technique.
- At least 720 skins and 16 Carolina parakeet skeletons are preserved in museum collections globally (Luther, 1986; McKinley & Hardy, 1985), of these, 460 specimens have the locality of collection georeferenced (Burgio et al., 2018). There is also the book by Paul Hahn (1963) “Where is that vanished bird? An index to the known specimens of the extinct and near extinct North American species” by far the most authoritative resource for all of the Carolina Parakeet specimens. Ancient DNA could be extracted from many of these specimens and a population paleogenomics study could be made. Possible pathogens could also be analyzed to definitively rule out or confirm an epidemic as a cause of extinction.
- The genome of the Carolina parakeet shows no evidence of genetic erosion. Neither heterosygosity nor the distribution of long runs of homozygosity (RoHs) show values far from the mean of the birds. This suggests that the population decline was fast

enough to leave no trace in the genome. It would be interesting to extend the genomic erosion analysis to other psittacine species, especially to the extinct Glaucous Macaw (*Anodorhynchus glaucus*) and near-extinct Spix macaw (*Cyanopsitta spixii*), whose fates have paralleled those of the Carolina parakeet.

- The Carolina parakeet was one of the few parrot species distributed in temperate latitudes of the northern hemisphere. The other is the Derbyan parakeet (*Psittacula derbiana*), confined to a small pocket of moist evergreen forest in the hills and mountains of the Indian, Tibet and China (Collar 1997). In the southern hemisphere we also find some species in temperate latitudes such as the monk parakeet (*Myiopsitta monachus*), which shows adaptations to cold climates common with the Carolina parakeet like having a fully feathered cere and roost communally throughout the year (Burgio et al., 2016). The monk parakeet now persists in multiple invasive colonies throughout the former range of the Carolina parakeet (Burgio et al., 2017) and has also occupied other areas of temperate latitude such as the Iberian Peninsula and specially Barcelona, one of the cities in the world with the highest parakeet density (Rodríguez-Pastor et al., 2012), taking advantage of anthropogenic factors (Cardador et al., 2022). Cold adaptations are also present in species such as the Kea (*Nestor notabilis*) and the kakapo (*Strigops habroptilus*) from New Zealand, and parrots adapted to live at high altitude such as the Andean parakeet (*Bolborhynchus orbygnesius*), found in the upland wooded valleys of the central eastern Andes in Bolivia and Peru, at elevations of 1,500 to 5,000 m. It would be interesting to analyze the genome of all these species to determine if they share genes with the Carolina parakeet to survive in cooler climates.

(2) Tenerife giant rat (*Canariomys bravoi*)

- Surely the most costly work of this thesis has been to obtain genetic material from the giant rat of Tenerife. However, after many efforts we have only managed to retrieve partial nuclear and mitochondrial DNA. Taking into account that there are abundant samples of these animals in the Catalan Institute of

Paleontology Miquel Crusafont (Bellaterra), as well as in the vertebrate collection of the Department of Animal Biology, Soil and Geology of the University of La Laguna (Tenerife), it would be very pertinent to try a new extraction of aDNA from new samples, taking advantage of the technical improvements of paleogenomics (Orlando et al., 2021). In parallel, aDNA should also be obtained from the giant rat of Gran Canaria (*Canariomys tamarani*) practically unstudied since it was described (López-Martínez & López-Jurado, 1987), in order to perform a molecular phylogeny and establish if it indeed shares genera with the giant rat of Tenerife and when the divergence between the two species occurred, which would parsimoniously imply a staggered colonization from Africa to Gran Canaria and from there to Tenerife, or on the contrary, if they do not share a close relationship, which would indicate two independent colonization events; three, in fact, if we take into account the lava mouse of Fuerteventura and Lanzarote (*Malpaisomys insularis*), related to the genus *Mus* (Pagès et al., 2012). The handicap in this case is that there are not many specimens of Gran Canaria giant rat.

- Once the genome of the Tenerife giant rat has been sequenced, the genome of other extant and extinct insular giant rodents, such as the Caribbean hutias (Capromyinae), could be sequenced to see if gigantism is achieved by the same evolutionary pathways or by different pathways.
- Obtaining the paleogenome of the Tenerife giant rat would also make it possible to see whether the extinction process was sudden or progressive and would help determine the causes of extinction. It would also be possible to analyze the presence of pathogenic microorganisms.
- It would be interesting to locate the current and Pleistocene populations of *Arvicanthis* genetically most closely related to *Canariomys*. On the one hand, it should be done an exhaustive sampling of the less studied or more isolated current populations of *Arvicanthis* (for example, the Hoggar mountains in southern Algeria), to detect if any of them has acted as a Pleistocene refuge and preserves ancestral traits shared with *Canariomys*. On the other hand, aDNA could be recovered from *Arvicanthis* fossils

of the Pleistocene found in numerous sites (Table 5.1), especially in El Mnasra Cave (Témara region, Morocco), located a few tens of kilometers north of the Draa river. This is the Pleistocene population of *Arvicanthis* closest to the Canary Islands and from where colonization most likely occurred.

- Like many rodents, the African grass rat experiences periodic population outbreaks associated with climate, which in turn have a direct effect on the abundance of food resources. Because of their impact on agriculture, these outbreaks have been studied in the Sahel (Poulet, 1985), in Kenya (Delany & Monro, 2011) and in Uganda (Mayamba et al., 2019). Colonization of the Canary Islands from rafting dispersal involves: 1) the existence of a more humid climate during the Pleistocene that allowed the presence of *Arvicanthis* in North Africa; 2) sporadic torrential rains with the capacity to drag trunks and masses of vegetation to the sea hosting some specimen of African grass rat, and 3) periodic population outbreaks that increase the probability of dragging individuals to the sea. It would be relevant to study the demographic evolution of North African *Arvicanthis* during the Pleistocene from a pairwise sequentially Markovian coalescent (PSMC) analysis to determine if island colonization coincides with any population increase.
- Based on the body mass of the Tenerife giant rat and its continental ancestor, the African grass rat, the divergence time and the earliest dating of Tenerife giant rat, we calculated a rate of increase in body mass to reach gigantism well above those observed for non-insular mammals. It has been reported that the African grass rat shows great plasticity in body sizes among its current populations (Neal, 1981). Was this plasticity the key that allowed the evolution towards gigantism? Russian biologist Dimitri Belyaev began a study of the domestication of foxes in the 1950s, which his disciples still continue today, 70 years later. They observed that when less aggressive foxes are selected, morphological and physiological changes are involuntarily selected as well, such as drooping ears and spotted fur. This analogy with the domestication of dogs would indicate the existence of a universal mechanism of evolutionary

transformation in domestic animals (Trut et al., 2009). Surprisingly, the selected traits became fixed in the fox population in just 10 generations. Based on this experiment, a long-term (10-15 years) captive breeding program of the African grass rat could be undertaken by selecting the largest individuals. The idea would be to reproduce in the laboratory the process of gigantism that it experienced when it colonized the island of Tenerife.

- Rodents are an island colonizing group. It is surprising to note that island colonizers occur in phylogenetically clustered families, even if they have very distant geographical distributions. Perhaps they share common physiological and ethological traits that would predispose them to island colonization. It would be necessary to determine which are the genes responsible for those traits.

(3) Xerces Blue (*Glaucopsyche xerces*)

- We have found that the butterflies Xerces Blue (*Glaucopsyche xerces*) and Silvery Blue (*Glaucopsyche lygdamus*) were two clearly differentiated species that apparently lived in sympatry but without gene flow. It is accepted that when a species of butterfly shows marked sexual dimorphism in dorsal wing surface coloration, like in the case of *Glaucopsyche*, the colours must play some role in mating. To determine what was the reproductive barrier that kept the two species apart, the reflectance spectrum of the blue coloration from the dorsal wing surface should be measured in a comparative study. This approach was successfully made in various species of *Morpho* butterflies (Giraldo et al., 2016), and polyommatine butterflies (Bálint et al., 2012).
- It is known that myrmecophilous butterflies establish mutualistic relationships with ants. Myrmecophily requires a high ecological specialization in the form of chemical and acoustic signalling, including particular organs and behavioural adaptations. It has been observed that *G. lygdamus* has a preference for the species *Formica altipetens* (Pierce & Eastealt, 1986). *G. xerces* also maintained mutualism with some ants of the same genus. Its decline also coincided with the introduction of the Argentine ant (*Linepithema humile*), into the region and it has been proposed

that this invasive ant may have contributed to species loss by outcompeting native ant species (Grewe et al., 2021). However, *L. humile* is known to tend lycaenid larvae in other cases (Trager & Daniels, 2009; Ikenaga et al., 2020). The fact that some 33-58 species of ants of the genus *Formica* have been identified in California opens the possibility to study the unknown myrmecophily relationships of *Glaucopsyche* and determine their effect as an interspecific reproductive barrier, as limiting ecological factor and even as a cause of extinction.

- Once the genome of a species is available, the next step is annotation, i.e, the identification and location of all coding regions of the genome. This is a challenging task in Lepidoptera, because it is a complex and diverse group. It could be done using as a primary guide the genome of the monarch butterfly (*Danaus plexippus*) (Zhan et al., 2011), and as general tool the MonarchBase portal (https://monarchbase. umassmed.edu) (Zhan & Reppert, 2013). It also could be used the silkworm genome (*Bombyx mori*) (Mita et al., 2004) widely studied as a model organism (Goldsmith et al., 2005), and of course the *Drosophila* genome, which is extremely well studied.

Solving all of these questions requires several more doctoral theses. When we have the answers, the list of pending questions will have increased rather than decreased. Far from thinking of the research as the work of Sisyphus, we should see it as the work of Prometheus, who brought fire to mankind. Although most of the findings of science are small scintillation, the whole constitutes an astonishing constellation.

7. BIBLIOGRAPHY

Aguilar, A. (1981). The black right whale, *Eubalaena glacialis*, in the Cantabrian Sea. *Rep. Int. Whal. Commn*., *31*, 457-459.

Alba, R., & Foner, N. (2015). Mixed Unions and Immigrant-Group Integration in North America and Western Europe. *The Annals of the American Academy, AAPSS, 662*, 38-56. DOI: 10.1177/0002716215594611

Alberto Barroso, V. (1999). Los animales en las prácticas funerarias guanches. *Anuario de Estudios Atlánticos* (45), 19-60.

Alberto-Barroso, V., Delgado-Darias, T., Brito-Mayor, A., & Velasco-Vázquez, J. (2021a). The ritualized use of dogs: considerations about their role in the mortuary belief system of the ancient canarians. *Canarias Arqueológica*, (22), 653-660.

Alberto Barroso, V., Moreno Benítez, M., Alamón Núñez, M., Suárez Medina, I., & Mendoza Medina, F. (2017). Estudio zooarqueológico de la restinga (Gran Canaria, España). Datos para la definición de un modelo productivo. *XXII Coloquio de Historia Canario-Americana*, 137, 1-17.

Alberto-Barroso, V., Moreno, M. A., Delgado Darias, T., & Mendoza-Medina, F. (2021b). Postmortem animal interference in an indigenous burial (Lanzarote, Canary Islands): Implications for archaeological research. *Bioarchaeology International, 5*(3-4), 175-187.

Alcover, J. A. (2008). The First Mallorcans: Prehistoric Colonization in the Western Mediterranean. *J. World Prehistory, 21*, 19–84.

Alcover, J. A., Campillo, X., Macias, M., & Sans, A. (1998). Mammal species of the world: additional data on insular mammals. *American Museum Novitates, 3248*, 1–29.

Alcover, J. A., Moya-Sol, S., & Pons-Moy, J. (1981). In: Moll (Ed.), *Les quimeres del passat. Els vertebrats fossils del Plio-Quaternari de les Balears i les Pitiüses*. Palma de Mallorca.

Ali, J. R., & Vences, M. (2019). Mammals and longdistance over-water colonization: The case for rafting dispersal; the case against phantom causeways. *J Biogeogr., 46*, 2632–2636. https://doi.org/10.1111/jbi.13659

Allan, J. R, Possingham, H. P., Atkinson, S. C., Waldron, A., Di Marco, M., Butchart, S. H. M., Adams, V. M., Kissling, W. D., Worsdell, T., Sandbrook, C., Gibbon, G., Kumar, K., Mehta, P., Maron, M., Williams, B. A., Jones, K. R., Wintle, B. A., Reside, A. E., & Watson, J. E. M. (2022). The minimum land area requiring conservation attention to safeguard biodiversity. *Science 376*, 1094–1101. DOI: 10.1126/science.abl9127

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, M. P. T., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., & Bunce, N. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fòssils. *Proc. R. Soc. B. 279*, 4724–4733. doi:10.1098/rspb.2012.1745

Amemiya, C. T., Alföldi, J., Lee, A. P., Fan, S., Philippe, H., MacCallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., Organ, C., Chalopin, D., Smith, J. J., Robinson, M., Dorrington, R. A., Gerdol, M., Aken, B., Biscotti, M. A. et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nature 496*, 311-316. doi:10.1038/nature12027

Ameur, R. (1976). Données nouvelles sur les rongeurs d'âge Pléistocène moyen du gisement d'Aïn Meftah (Willaya de Tlemcen). *Bulletin de la Société d'histoire naturelle d'Afrique du Nord, 67*(1–2), 133–140.

Amori, G., Gippolitib, S., & Helgenc, K. M. (2008). Diversity, distribution, and conservation of endemic island rodents. *Quaternary International, 182*, 6–15.

Angst D., Buffetaut, E., & Abourachid, A. (2011). The end of the fat dodo? A new mass estimate for Raphus cucullatus. *Die Naturwissenschaften, 98*(3), 233–6.

Araujo, B., Oliveira-Santos, L., Lima-Ribeiro, M., Diniz-Filho, J. A., & Fernandez, F. (2015). Bigger kill than chill: The uneven roles of humans and climate on late Quaternary megafaunal extinctions. *Quaternary International, 431*. DOI:10.1016/j.quaint.2015.10.045

Arístegui, J., Barton, E. D., Álvarez-Salgado, X. A., Santos, A. M. P., Figueiras, F. G., Kifani, S., Hernández-León, S., Mason, E., Machú, E., & Demarcq, H. (2009). Sub-regional ecosystem variability in the Canary Current upwelling. *Progress in Oceanography, 83*, 33-48,

Ashton, K. G., Tracy, M. C., & de Queiroz, A. (2000). Is Bergmann's Rule Valid for Mammals? *The American Naturalist, 156*(4), 390-415. doi:10.1086/303400.

Audubon, J. J. (1831). *Ornithological biography*. Volume 1. Edinburgh, UK: Adam & Charles Black.

Axelsen, J. B., Roll, U., Stone, L., & Solow, A. (2013). Species-area relationships always overestimate extinction rates from habitat loss: comment. *Ecology, 94*(3), 761–763.

Baguette, M. & Stevens, V. (2013). Predicting minimum area requirements of butterflies using life-history traits. *J Insect Conserv.* DOI 10.1007/s10841-013-9548-x

Baleka, S., Herridge, V. L., Catalano, G., Lister, A. M., Dickinson, M. R., Di Patti, C., Barlow, A., Penkman, K. E. H., Hofreiter, M., & Paijmans, J. L. A. (2021). Estimating the dwarfing rate of an extinct Sicilian elephant. *Current biology*, *31*(16), 3606-3612.e7

Bálint, Z., Kertész, K., Piszter, G., Vértesy, Z., & Biró, L. P. (2012). The well-tuned blues: the role of structural colours as optical signals in the species recognition of a local butterfly fauna (Lepidoptera: Lycaenidae: Polyommatinae). *J. R. Soc. Interface, 9*, 1745-1756. doi:10.1098/rsif.2011.0854

Banerjee, T. D., Sean, S. K., & Monteiro, A. (2021). optix is involved in eyespot development via a possible positional information mechanism. *bioRxiv.* DOI:10.1101/2021.05.22.445259

Barlow, A., Cahill, J. A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G. G., Paijmans, J. L. A., Rabeder, G., Frischauf, C., Grandal-d'Anglade, A., García-Vázquez, A., Murtskhvaladze, M., Saarma, U., Anijalg, P., Skrbinse, T., Bertorelle, G., Gasparian, B., Bar-Oz, G., Pinhasi, R., Slatkin, M., Dalén, L., Shapiro, B. & Hofreiter, M. (2018). Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol., 2*(10), 1563-1570.

Barnes, C., Maxwell, D., Reuman, D. C., & Jennings, S. (2010). Global patterns in predator–prey size relationships reveal size dependency of trophic transfer efficiency. *Ecology, 91*(1), 222–232.

Barnosky, A., Bell, C., Emslie, S., Goodwin, T., Mead, J., Repenning, C., Scott, E., & Shabel, A. (2004). Exceptional

record of mid-Pleistocene vertebrates helps differentiate climatic from anthropogenic ecosystem perturbations. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 9297-302. 10.1073/pnas.0402592101.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature, 471,* 51-57. doi:10.1038/nature0967

Barrón-Ortiz, C., Rodrigues, A. T., Theodor, J. M., Kooyman, B. P., Yang, D. Y., Speller, C. F., & Orlando, L. (2017). Cheek tooth morphology and ancient mitochondrial DNA of late Pleistocene horses from the western interior of North America: Implications for the taxonomy of North American Late Pleistocene Equus. *PloS one, 12*(8), p.e0183045-e0183045

Ben-Dor, M., & Barkai, R. (2021). Prey Size Decline as a Unifying Ecological Selecting Agent in Pleistocene Human Evolution. *Quaternary, 4,* 7. https://doi.org/10.3390/quat4010007

Benítez-López, A., Santini, L., Gallego-Zamorano, J., Milá, B., Walkden, P., Huijbregts, M. A. J., & Tobias, J. A. (2021). The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. *Nature Ecology & Evolution, 5*, 768–786. https://doi.org/10.1038/s41559-021-01426-y

Blanco, G., Morinha, F., Roques, S., Hiraldo, F., Rojas, A., & Tella, J. L. (2021). Fine-scale genetic structure in the critically endangered red-fronted macaw in the absence of geographic and ecological barriers. *Scientific Reports, 11*, 556. https://doi.org/10.1038/s41598-020-79575-6

Bocherens, H., Michaux, J., Garcia Talavera, F., & Van der Plicht, J. (2006). Extinction of endemic vertebrates on islands: The

case of the giant rat *Canariomys bravoi* (Mammalia, Rodentia) on Tenerife (Canary Islands, Spain). *C. R. Palevol, 5*, 885–891.

Boeskorov, G. G., Plotnikov, V. V., Protopopov, A. V., Baryshnikov, G. F., Fosse, P., Dalén, L., Stanton, D. W. G., Pavlov, I. S., Suzuki, N., & Tikhonov, A. N. (2021). The Preliminary Analysis of Cave Lion Cubs *Panthera spelaea* (Goldfuss, 1810) from the Permafrost of Siberia. *Quaternary, 4*(3), 24. https://doi.org/10.3390/quat4030024

Bogardus, C., Lillioja, S., Ravussin, E., Abbott, W., Zawadzki, J. K., Young, A. et al. (1986). Familial Dependence of the Resting Metabolic Rate. *New Engl J Med., 315*(2), 96–100. doi:10.1056/NEJM198607103150205.

Boisduval, J. A. (1852). *Lépidoptères de la Californie*. Paris. pp 52.

Bover, P., Llamas, B., Mitchell, K. J., Thomson, V. A., Alcover, J. A., Lalueza-Fox, C., Cooper, A., & Pons, J. (2019). Unraveling the phylogenetic relationships of the extinct bovid *Myotragus balearicus* Bate 1909 from the Balearic Islands. *Quaternary Science Reviews, 215*, 185-195. https://doi.org/10.1016/j.quascirev.2019.05.0050277-3791/

Bover, P., Tolosa, F., (2005). The olfactory ability of *Myotragus balearicus*: preliminary notes. In: Alcover, J. A., Bover, P. (Eds.), Proceedings of the International Symposium “Insular Vertebrate Evolution: the Palaeontological Approach”, *Mon. Soc. Hist. Nat. Balears*, 12, pp. 85e94.

Brien, M. N., Enciso-Romero, J., Parnell, A. J., Salazar, P. A., Morochz, C., Chalá, D., Bainbridge, H. E., Zinn, T., Curran, E. M., & Nadeau, N. J. (2018). Phenotypic variation in *Heliconius erato* crosses shows that iridescent structural colour is sex-linked and controlled by multiple genes. *Interface Focus 9,* 20180047. http://dx.doi.org/10.1098/rsfs.2018.0047

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci, 104*(37): 14616-21. doi: 10.1073/pnas.0704665104

Bryja, J., Colangelo, P., Lavrenchenko, L. A., Meheretu, Y., Šumbera, R., Bryjová, A., Verheyen, E., Leirs, H., & Castiglia, R. (2019). Diversity and evolution of African Grass Rats (Muridae: Arvicanthis) -From radiation in East Africa to repeated colonization of northwestern and southeastern savannas. *J Zool Syst Evol Res*., *00*, 1–19. DOI: 10.1111/jzs.12290

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta, 1842*, 1932–1941.

Buffalo, V. (2021) Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's Paradox. *eLife,* 10:e67509. DOI: https://doi.org/10.7554/eLife.67509

Burgio, K. R., Carlson, C. J., & Tingley, M. W. (2017). Lazarus ecology: Recovering the distribution and migratory patterns of the extinct Carolina parakeet. *Ecology and Evolution*, 1-9. DOI: 10.1002/ece3.3135

Burgio, K. R., Carlson, C. J., & Bond, A. L. (2018). Georeferenced sighting and specimen occurrence data of the extinct Carolina Parakeet (*Conuropsis carolinensis*) from 1564-1944. *Biodiversity Data Journal 6*: e25280. https://doi.org/10.3897/BDJ.6.e25280

Burgio, K. R., Carlson, C. J., Bond, A. L., Rubega, M. A., & Tingley, M. W. (2021). The two extinctions of the Carolina Parakeet

Conuropsis carolinensis. *Bird Conservation International, Short Communication*. *Cambridge University Press* 8p. doi:10.1017/S0959270921000241

Burgio, K. R., Davis, K. E., Dreiss, L. M., Cisneros, L. M., Klingbeil, B. T., Presley, S. J., van Rees, C. B., & Willig, M. R. (2022). Integrating multiple dimensions of biodiversity to inform global parrot conservation. *Animal Biodiversity and Conservation, 45*(2), 189–202. https://doi.org/10.32800/abc.2022.45.0189

Burgio, K. R., van Rees, C., Block, K. E., Pyle, P., Patten, M. A., Spreyer, M. F., & Bucher, E. H. (2016). Monk parakeet (*Myiopsitta monachus*). In P. Rodewald (Ed.), *The birds of North America online*. Ithaca: Cornell Lab of Ornithology, https://doi.org/10.2173/bna.322

Busby, R. C., Faynel, C., Moser, A., & Robbins, R. K. (2017). Sympatric Diversification in the Upper Amazon: A Revision of the Eumaeine Genus Paraspiculatus (Lepidoptera: Lycaenidae). *Smithsonian Contributions to Zoology 649*, 65p.

Campa, C. C., Weisbach, N.R., Santinha, A. J., Incarnato, D., & Platt, R. J. (2019). Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods 16*, 887–893.

Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., & Poinar Jr, G. O. (1993). Amplification and sequencing of DNA from a 120–135 million-year-old weevil. *Nature 363*, 536–538. doi:10.1038/363536a0

Cano, R. J., Poinar, H., & Poinar Jr, G. O. (1992). Isolation and partial characterisation of DNA from the bee *Proplebeia dominicana* (Apidae: Hymenoptera) in 25–40 million year old amber. *Med. Sci. Res., 20*, 249–251.

Cantalapiedra, J. L., Aze, T., Cadotte, M. W., Dalla Riva, G. V., Huang, D., Mazel, F., Pennell, M. W., Ríos, M., & Mooers, A. Ø. (2019). Conserving evolutionary history does not result in greater diversity over geological time scales. *Proc. R. Soc. B 286*, 20182896. http://dx.doi.org/10.1098/rspb.2018.2896

Caravaggi, A., Cuthbert, R., Ryan, P. G., Cooper, J., & Bond, A. L. (2019). The impacts of introduced House Mice on the breeding success of nesting seabirds on Gough Island. *Ibis 161*, 648–661. doi: 10.1111/ibi.12664

Cardador, L., Tella, J. L., Louvrier, J., Anadón, J. D., Abellán, P., & Carrete, M. (2022). Climate matching and anthropogenic factors contribute to the colonization and extinction of local populations during avian invasions. *Diversity & distributions*, *28*(9), 1908-1921. DOI: 10.1111/ddi.13591

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods Ecol. Evol., 9*, 410–419. doi:10.1111/2041-210X.12871

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Sci. Adv., 1*, e1400253

Ceballos, G., Ehrlichb, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *PNAS,* E6089–E6096 www.pnas.org/cgi/doi/10.1073/pnas.1704949114

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet., 19*, 220–234.

Chaabo, F., Pronczuk, A., Maslova, E., & Hayes, K. C. (2010). Nutricional correlates and dynamics of diabetes in the Nile rat (*Arvicanthis niloticus*): a novel model for diet-induced type 2 diabetes and the metabolic syndrome. *Nutrition & Metabolism, 7*, 29. http://www.nutritionandmetabolism.com/content/7/1/29

Chávez Álvarez, M. E., Pérez Caamaño, F., Pérez González, E., Soler Segura, J., Goñi Quinteiro, A., & Tejera Gaspar, A. (2005). El barranco de San Blas (San Miguel de Abona, Tenerife). Breve secuencia de un modelo por definir.

Chazot, N., Wahlberg, N., Freitas, A. V. L., Mitter, C., Labandeira, C., Sohn, J. C., Sahoo, R. K., Seraphim, N., De Jong, R., & Heikkilä, M. (2019). Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. *Syst. Biol., 68*, 797–813.

Cheddadi, R., Lamb, H. F., Guiot, J. & Van Der Kaars, S. (1998). Holocene climatic change in Morocco: A quantitative reconstruction from pollen data. *Clim. Dyn. 14*, 883–890.

Chimento, N. R., & Agnolin, F. L. (2020). Phylogenetic tree of Litopterna and Perissodactyla indicates a complex early history of hoofed mammals. *Scientific Reports, 10*, 13280 doi.org/10.1038/s41598-020-70287-5

Cifelli, R.L. (1993). *The phylogeny of the native South American ungulates, in Mammal Phylogeny: Placentals* (eds F.S. Szalay, M.J. Novacek & M.C. McKenna), Springer-Verlag, New York, Berlin, pp. 195–216.

Clark, A. H. (1905). The West Indian parrots. *The Auk, 22*(4), 337–344. doi:10.2307/4069996

Cloutier, A., Sackton, T. B., Grayson, P., Edwards, S. V. & Baker, A. J. (2018). First nuclear genome assembly of an extinct moa

species, the little bush moa (*Anomalopteryx didiformis*). *BioRxiv*. PPR: PPR18857.

Collar, N. J. (1997). Family Psittacidae (Parrots). Pp 280-480 in: del Hoyo, J., Elliot, A. & Sargatal, J. eds. (1997). *Handbook of the Birds of ther World*. Vol 4. Sandgrouse to Cuckoos. Lynx Edicions. Barcelona.

Cooper, A., Mourer-Chauviré, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., & Pääbo, S. (1992). Independent origins of New Zealand moas and kiwis. *Proc. Natl Acad. Sci. USA 89*, 8741–8744. doi:10.1073/pnas.89.18.8741

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do It Right or Not at All. *Science 289*(5482), 1139. doi: 10.1126/science.289.5482.1139b

Corbett-Detig, R.B., Hartl, D.L., & Sackton, T. B. (2015). Natural selection constrains neutral diversity across a wide range of species. *PLOS Biology 13*, e1002112.
DOI: https://doi.org/10.1371/journal.pbio.1002112

Crispell, J., Balaz, D., & Gordon, S. V. (2019). HomoplasyFinder: a simple tool to identify homoplasies on a phylogeny. *Microbial Genomics, 5.*
DOI 10.1099/mgen.0.000245

Crowley, B. E., Yanes, Y., Mosher, S. G., & Rando, J. C. (2019). Revisiting the Foraging Ecology and Extinction History of Two Endemic Vertebrates from Tenerife, Canary Islands. *Quaternary, 2*, 10. doi:10.3390/quat2010010

Crusafont-Pairo, M., & Petter, F. (1964). Un muriné géant fossile des îles Canaries: *Canariomys bravoi* gen. nov., sp. nov. (Rongeurs, Muridés). *Mammalia, 28*, 608–611.
doi:10.1515/mamm.1964.28.4.607

Cucchi, T., Vigne, J. D., & Auffray, J. C. (2005). First occurrence of the house mouse (*Mus musculus domesticus* Schwarz and Schwarz 1943) in the Western Mediterranean: A zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society, 84*, 429–445.

Dabney, J., Meyer, M. & Pääbo, S. (2013). Ancient DNA Damage. *Cold Spring Harb Perspect Biol, 5,* a012567.

Dabney, J. et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA, 110*, 15758–15763. doi:10.1073/pnas.1314445110

Darwin, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. Publisher John Murray. London.

Davis, C. (2016). *Taxodium distichum*. Lafayette College. Campus Tree Project. https://sites.lafayette.edu/campustrees/

Davis, M., Faurby, S. & Svenning, J. C. (2018). Mammal diversity will take millions of years to recover from the current biodiversity crisis. *PNAS, 115* (44), 11262–11267. www.pnas.org/cgi/doi/10.1073/pnas.1804906115

Debat, V., Berthier, S., Blandin, P., Chazot, N., Elias, M., Gomez, D. & Llaurens, V. (2018). Why are Morpho Blue? *Biodiversity and Evolution*, 139-174.

De Bont, R., & van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis,19*(3), 169-185.
DOI: 10.1093/mutage/geh025

Delany, M. & Monro, R. (2011). Population dynamics of *Arvicanthis niloticus* (Rodentia: Muridae) in Kenya. *Journal of Zoology, 209*, 85-103.

Delsuc, F., Gibb, G. C., Kuch, M., Billet, G., Hautier, L., Southon, J., Rouillard, J. M., Fernicola, J. C., Vizcaíno, S. F., MacPhee, R. D. E., & Poinar, H. N. (2016). The phylogenetic affinities of the extinct glyptodonts. *Current Biology 26*, R141–R156.

deMenocal, P. B., Ortiz, J., Guilderson, T., & Sarnthein, M. (2000). Coherent high- and low-latitude climate variability during the Holocene warm period. *Science, 288*, 2198–2202.

de Muizon, C. & Cifelli, R.L. (2000). The 'condylarths' (archaic Ungulata, Mammalia) from the early Palaeocene of Tiupampa (Bolivia): implications on the origin of the South American ungulates. *Geodiversitas, 22*, 47–150.

Denys, C., Geraads, D., Hublin, J. J., & Tong, H. (1987). Méthode d'étude taphonomique des microvertébrés. Application au site Pléistocène de Tighenif (Algérie). *Archeozoologia, 2*, 53–82.

De Vos J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R., & Pimm, S. L. (2014). Estimating the normal background rate of species extinction. *Conservation Biology, 29*(2), 452–462. DOI: 10.1111/cobi.12380

DeSalle, R., Gatesy, J., Wheeler, W. C., & Grimaldi, D. (1992). DNA Sequences from a Fossil Termite in Oligo-Miocene Amber and Their Phylogenetic Implications. *Science (New York, N.Y.), 257*, 1933-6. 10.1126/science.1411508.

Diamond, J. M. (1984). Historic extinctions: a rosetta stone for understanding prehistoric extinctions. *Quaternary extinctions*, 824-862.

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying Temporal Genomic Erosion in Endangered Species. *Trends Ecol. Evol., 33*,176–185.

Dirzo, R. & Raven, P. H. (2003). Global state of biodiversity and loss. *Annu. Rev. Environ. Resour., 28*, 137–167.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. *Science 345*, 401-406.
DOI: 10.1126/science.1251817

Dobigny, G., Tatard, C., Gauthier, P., Khalilou, B., Duplantier, J. M., et al. (2013). Mitochondrial and Nuclear Genes-Based Phylogeography of *Arvicanthis niloticus* (Murinae) and Sub-Saharan Open Habitats Pleistocene History. *PLoS ONE, 8*(11), e77815. doi:10.1371/journal.pone.0077815

Downey, J. C., & Lange JR, W. H. (1956). Analysis of variation in a recently extinct polymorphic lycaenid butterfly, Glaucopsyche xerces (bdv.), with notes on its biology and taxonomy. *Bulletin, So. Calif. Academy of Sciences, 55*, (3), 153-170.

Durst, P. A. P., & Roth, V. L. (2015). Mainland size variation informs predictive models of exceptional insular body size change in rodents. *Proc. R. Soc. B, 282*, 20150239.
http://dx.doi.org/10.1098/rspb.2015.0239

Durvasul, A., & Sankararaman, S. (2020). Recovering signals of ghost archaic introgression in African populations. *Science Advances, 6*(7) eaax5097.
DOI: 10.1126/sciadv.aax5097

Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes D., Joanna D., Lewin, H. A.,

Hastie, A. R., Formenti, G., Fedrigo, O., Guhlin, J., Harrop, T. W. R., Le Lec, M. F., Dearden, P. K., Haggerty, L., Martin, F. J., Kodali, V., Thibaud-Nissen, F., Iorns, D., Knapp, M., Gemmell, N. J., Robertson, F., Moorhouse, R., Digby, A., Eason, D., Vercoe, D., Howard, J., Jarvis, E. D., Robertson, B. C., & Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics, 1*(1), p.100002

Edmeades, B. (2021). *Megafauna: First Victims of the Human-Caused Extinction*. Houndstooth Press. pp 400.

Elias, S. & Schreve, D. (2016). Late Pleistocene Megafaunal Extinctions. 10.1016/B978-0-12-409548-9.10283-0.

Espeland, M., Breinholt, J., Willmott, K. R., Warren, A. D., Vila, R., Toussaint, E. F. A, Maunsell, S. C., Aduse-Poku, K., Talavera, G., Eastwood, R., et al. (2018). A Comprehensive and Dated Phylogenomic Analysis of Butterflies. *Curr. Biol. 28*, 770-778.e5. https://www.sciencedirect.com/science/article/pii/S0960982218300940

Estes, J. A., Burdin, A., & Doak, D. F. (2016). Sea otters, kelp forests, and the extinction of Steller's sea cow. *Proceedings of the National Academy of Sciences - PNAS,113*(4), 880-885.

Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., Lang, P. L. M., Leventhal, L., Nogues-Bravo, D., Pagowski, V., Ruffley, M., Spence, J. P., Toro Arana, S. E., Weiß, C. L., & Zess, E. (2022). Genetic diversity loss in e Anthropocene. *Science, 377,* 1431–1435. DOI: 10.1126/science.abn5642

Fabre, P. H., Vilstrup, J. T, Raghavan, M., Der Sarkissian, C., Willerslev, E., Douzery E. J. P., & Orlando, L. (2014). Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biol. Lett., 10*, 20140266.

http://dx.doi.org/10.1098/rsbl.2014.0266

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Allentoft, M., Anthony, D., Baimukhanov, T., Trixl, S., Tumen, D., Undrakhbold, S., Usmanova, E., Vahdati, A., Valenzuela-Lamas, S., Viegas, C. et al. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell, 177*(6), 1419-1435.e31

Feigin, C. Y., Newton, A. H., Doronina, L., Schmitz, J., Hipsley, C. A., Mitchell, K. J., Gower, G., Llamas, B., Soubrier, J., Heider, T. N., Menzies, B. R., Cooper, A., O'Neill, R. J. & Pask, A. J. (2018).Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nature Ecology and Evolution, 2*, 182–192.

Fenner, J., Benson, C., Rodriguez-Caro, L., Ren, A., Papa, R., Martin, A., Hoffmann, F., Range, R. & Counterman, B. A. (2020). Wnt Genes in Wing Pattern Development of Coliadinae Butterflies. *Front. Ecol. Evol., 8*, 197.

Fernandez, F. (2016). Human Dispersal and Late Quaternary Megafaunal Extinctions: the Role of the Americas in the Global Puzzle.

Fernicola, J. C., Vizcaíno, S. F., & De Iuliis, G. (2009). The fossil mammals collected by Charles Darwin in South America during his travels on board the HMS Beagle. *Revista de la Asociación Geológica Argentina, 64*(1), 147-159.

Ferrari, G., Lischer, H. E. L., Neukamm, J., Rayo, E., Borel, N., Pospischil, A., Rühli, F., Bouwman, A. S., & Campana, M. G. (2018). Assessing Metagenomic Signals Recovered from Lyuba, a 42,000-Year-Old Permafrost-Preserved Woolly Mammoth Calf. *Genes 9*, 436. doi:10.3390/genes9090436

Firmat, C., Gomes Rodrigues, H., Renaud, S., Claude, J., Hutterer, R., Garcia-Talavera, F., & Michaux, J. (2010). Mandible morphology, dental microwear, and diet of the extinct giant rats Canariomys (Rodentia: Murinae) of the Canary Islands (Spain). *Biological Journal of the Linnean Society, 101*, 28–40.

Fisher, D., Shirley, E., Whalen, C., Calamari, Z. Rountrey, A., Tikhonov, A., Buigues, B., Lacombat, E., Grigoriev, S., & Lazarev, P. (2014). X-ray computed tomography of two mammoth calf mummies. *Journal of Paleontology, 88*, 4: 664-675.

Flannery, T., (1995). *Mammals of the South-West Pacific and Moluccan Islands*. Comstock/ Cornell University Press, Ithaca, New York, pp. 1–464.

Folch, R. & Camarasa, J. M. (Eds.) (1993). *Biosfera. Mediterrànies*. Vol. 5 Enciclopèdia Catalana. Barcelona. pp 439.

Forey, P. L., Fortey R. A., Kenrick, P., & Smith A. B. (2004). Taxonomy and fossils: a critical appraisal. *Phil. Trans. R. Soc. Lond*, B359639–653 http://doi.org/10.1098/rstb.2003.1453

Forshaw, J. M. (1989). *Parrots of the World*. Lansdowne Editions. pp 672.

Forterre, P., Filée, J., & Myllykallio, H. (2007). Origin and Evolution of DNA and DNA Replication Machineries - In book: *The Genetic Code and the Origin of Life*. pp.145-168.

Foster, J. B. (1964). The evolution of mammals on islands. *Nature, 202*(4929), 234–235.

Gansauge, M. T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc., 8*, 737–748.

Gansauge, M. T., & Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res., 24*(9), 1543–1549. doi: 10.1101/gr.174201.114

Garcia-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernàndez, M., Vergés, J., & De Vicente, R. (2009). Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature 462*, 778–781. https://doi.org/10.1038/nature08555

Garcia, M. & Pujol, A. (2016). Experimentació en animals: la revolució dels transgènics. *Treballs de la Societat Catalana de Biologia, 66*, 37-42. DOI: 10.2436/20.1501.02.157

Gardner, M. (2000). *The annotated Alice: the definitive edition*. New York: W.W. Norton & Co.; London (UK): Allen Lane, The Penguin Press. 384 p.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A, Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I. J., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A. F., Davoudi, H., et al. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science, 360*(6384), 111-114.

Gedney, C. W. (1877). Foreign cage birds: containing full directions for successfully breeding, rearing, and managing the various beautiful aviary birds imported into this country. Vol. 1. *London, UK: "The Bazaar" Office*.

Geraads, D. (2012). The Faunal Context of Human Evolution in the Late Middle/Late Pleistocene of Northwestern Africa. In: Hublin, JJ., McPherron, S. (eds) *Modern Origins. Vertebrate Paleobiology and Paleoanthropology*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2929-2_4

Giraldo, M. A., Yoshioka, S., Liu, C. & Stavenga. D. G. (2016). Coloration mechanisms and phylogeny of Morpho butterflies. *Journal of Experimental Biology 219*, 3936-3944. doi:10.1242/jeb.148726

Goldsmith, M R., Shimada, T., & Abe. H. (2005). The genetics and genomics of the silkworm, *Bombyx mori*. *Annu Rev Entomol 50*, 71-100.
doi: 10.1146/annurev.ento.50.071803.130456.

Golenberg, E.M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D., & Zurawski, G. (1990). Chloroplast DNA sequence from a Miocene magnolia species. *Nature 344*, 656–658. doi:10.1038/344656a0

Gooday, A. J. (1994). The biology of deep-sea foraminifera: A review of some advances and their applications in paleoceanography. *Palaios. SEPM Society for Sedimentary Geology, 9*(1): 14-31. https://doi.org/10.2307/3515075

Goodwin, D. (2007). Horse Behaviour: Evolution, Domestication and Feralisation. In: Waran, N. (eds) *The Welfare of Horses. Animal Welfare*, vol 1. Springer, Dordrecht. https://doi-org.sare.upf.edu/10.1007/978-0-306-48215-1_1

Gould, S. J. (1992). *Bully for Brontosaurus: Reflections in Natural History*. W. W. Norton & Company. 540 pp.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H. Y., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox,

C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., & Pääbo, S. (2010). A draft sequence of the neandertal genome. *Science 328*, 710–722.

Green, T., Renne, P. R. & Keller, C. B. (2022). Continental flood basalts drive Phanerozoic extinctions. *Proceedings of the National Academy of Sciences – PNAS, 119*(38), p.1-e2120441119. https://doi.org/10.1073/pnas.2120441119

Grewe, F., Kronforst, M. R., Pierce, N.E., & Moreau, C. S. (2021). Museum genomics reveals the Xerces blue butterfly (*Glaucopsyche xerces*) was a distinct species driven to extinction. *Biol. Lett. 17*, 20210123.

Groves, C. R., Jensen, D. B., Valutis, L. L., Redford, K. H., Shaffer, M. L., Scott, J. M., Baumgartner, J. V., Higgins, J. V., Beck, M. W. & Anderson, M. G. (2002). Planning for Biodiversity Conservation: Putting Conservation Science into Practice. *BioScience, 52*(6): 499-512.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S. et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*, 207–211.

Hagelberg, E., Thomas, M. G., Cook, C. E., Sher, A. V., Baryshnikov, G. F., Lister. A. M. (1994). DNA from ancient mammoth bones. *Nature 370*, 333–334. doi: 10.1038/370333b0

Hagelberg, E., Sykes, B., & Hedges, R. (1989). Ancient bone DNA amplified. *Nature 342*, 485. doi:10.1038/342485a0

Hahn, P. (1963). *Where is that vanished bird? An index to the known specimens of the extinct and near extinct North*

American species. Royal Ontario Museum, University of Toronto. pp 347.

Hasbrouck, E. M. (1891). The carolina paroquet (*Conurus carolinensis*). *The Auk, 8*, 369–379.

Heck, H. (1951). The breeding-back of the Aurochs. *Oryx 1*, 117–122. DOI: http://dx.doi.org/10.1017/ S0030605300035286

Heck, L. (1955). Grosswild im Etoshaland.

Hedges, S.B., & Schweitzer, M. H. (1995). Detecting dinosaur DNA. *Science 5214*, 1191-1192.
doi: 10.1126/science.7761839

Heintzman, P. D., Zazula, G. D., Cahill, J. A., Reyes, A. V., MacPhee, R. D. E. & Shapiro, B. (2015). Genomic Data from Extinct North American Camelops Revise Camel Evolutionary History. *Molecular Biology and Evolution 32*, 2433–2440.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature 312*(5991):282–284. doi:10.1038/312282a0

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., & Vila, R. (2019). A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular ecology, 28*(17), 3857-3868.

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2014). The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays, 36*, 1-10.
DOI 10.1002/bies.201400160

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. (2001). Ancient DNA. *Nature Reviews. Genetics, 2*, 353-359.

Hsü, K. J., Montadert, L., Bernoulli, D., Cita, M. B., Erickson, A., Garrison, R. E., Kidd, R. B., Mèlierés, F., Müller, C., & Wright, R. (1977). History of the Mediterranean salinity crisis. *Nature, 267*, 399–403. https://doi.org/10.1038/267399a0

Huchon, D. & Douzery, E. J. P. (2001). From the Old World to the New World: A Molecular Chronicle of the Phylogeny and Biogeography of Hystricognath Rodents. *Molecular Phylogenetics and Evolution, 20*(2), 238–251. doi:10.1006/mpev.2001.0961

Hume, J. P., Cheke, A. S., & McOran-Campbell, A. (2009). How Owen 'stole' the Dodo: academic rivalry and disputed rights to a newly-discovered subfossil deposit in nineteenth century Mauritius. *Historical Biology, 21*(1),33-49. DOI: 10.1080/08912960903101868

Hung, C. M., Shanner, P. J. L., Zink, R. M., Liu, W. C., Chu, T. C., Huang, W. S. & Li, S. H. (2014). Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences USA, 111*, 10636-10641.

Hutterer, R., López-Martínez, N., & Michaux, J. (1988). A new rodent from Quaternary deposits of the Canary Islands and its relationships with Neogene and recent murids of Europe and Africa. *Palaeovertebrata, 18*, 241–262.

Iakovleva, L. (2015). The architecture of mammoth bone circular dwellings of the Upper Palaeolithic settlements in Central and Eastern Europe and their socio-symbolic meanings. *Quaternary International, 359-360*, 324-334.

Ikenaga, N., Touyama, Y., Kameyama, T., & Ito, F. (2020). Effects of Argentine ants (Hymenoptera: Formicidae) on myrmecophilous lycaenid butterfly, Narathura basalus (Lepidoptera: Lycaenidae), in western Japan. *Entomol. Sci., 23*, 69–73. doi:10.1111/ens.12396

Jablonski, D. (1994). Extinctions in the fossil record. *Phil. Trans. R. Soc. Lond. B 344,* 11–17. http://doi.org/10.1098/rstb.1994.0045

Jablonski, AD. (2004). Extinction: past and present. *Nature, 427*, 589.

Jackson, H., Jones, C. G., Agapow, P. M., Tatayah, V., Groombridge, J. J., & Collinson, M. (2015). Micro-evolutionary diversification among Indian Ocean parrots: temporal and spatial changes in phylogenetic diversity as a consequence of extinction and invasion. *Ibis, 157*(3), 496-510.

Jaeger, J. J. (1975). *Evolution des rongeurs du Miocène à l'actuel en Afrique nord-occidentale*. PhD dissertation, Université de Montpellier.

Jaeger, J. J. (1988). Rodent phylogeny: New data and old problems. In "*The Phylogeny and Classification of the Tetrapods*" (M. J. Benton, Ed.), Clarendon, Oxford. pp. 177–199.

Jarić, I., Roll, U., Bonaiuto, M., Brook, B. W., Courchamp, F., Firth, J. A., Gaston, K, J., Heger, T., Jeschke, J. M., Ladle, R. J., Meinard, Y., Roberts, D. L., Sherren, K., Soga, M., Soriano-Redondo, A., Veríssimo, D., & Correia, R. A. (2022). Societal extinction of species. *Trends in Ecology & Evolution, 37*(5), 411-419. https://doi.org/10.1016/j.tree.2021.12.011

Jennings, C. E., Aber, J. S., Balco, G., Barendregt, R., Bierman, P. R., Rovey II, C. W., Roy, M., & Thorleifson, L. H. (2007).

Glaciations. Mid-Quaternary in North America. 10.1016/B978-0-444-53643-3.00124-2.

Johnson, J., Novak, B., Athrey, G., Shapiro, B., Phelan, R., & Brand, S. (2015). Whole genome sequence analysis reveals evolutionary history of extinct Heath Hen. In *Proceedings of the Joint Meeting of the American Ornithologist's Union & Cooper Ornithological Society*, Norman, OK, USA.

Johansson, U. S., Ericson, P. G. P., Blom, M. P. K., & Irestedt, M. (2018). The phylogenetic position of the extinct Cuban Macaw Ara tricolor based on complete mitochondrial genome sequences. *Ibis, 160*(3), 666-672.

Joly, E. (2011). The existence of species rests on a metastable equilibrium between inbreeding and outbreeding. An essay on the close relationship between speciation, inbreeding and recessive mutations. *Biology Direct, 6*,62. http://www.biology-direct.com/content/6/1/62

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics, 29,* 1682–1684.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T.S., Vilstrup, J.T., Lear, T., Myka, J.L., Lundquist, J., Miller, D.C., Alfarhan, A.H., Alquraishi, S.A., Al-Rashied, K.A.S., Statgegaard, J., Strauss, G., Bertelsen, M.F., Sicheritz-Ponten, T., Antczak, D.F., Bailey, E., Nielsen, R., Willerslev, E. & Orlando, L. (2014). Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences USA, 111,* 18655–18660.

Jorgensen, D. (2012). Reintroduction and De-extinction. *BioScience 63*(9), 719-720.

Kawane, K., Motani, K., & Nagata, S. (2014). DNA Degradation and its defects. *Cold Spring Harb Perspect Biol,6,* a016394. doi: 10.1101/cshperspect.a016394

Kender, S., McClymont, E. L., Elmore, A. C., Emanuele, D., Leng, M. J., & Elderfield, H. (2016). Mid Pleistocene foraminiferal mass extinction coupled with phytoplankton evolution. *Natur Communications, 7*, 11970. DOI: 10.1038/ncomms11970

Kilmer, J. A., & Steadman, D. W. (2016). A middle Pleistocene bird community from Saint Lucie County, Florida. *Bulletin of the Florida Museum of Natural History, 55*(1), 1-38.

Kircher, M. (2012). Analysis of High-Throughput Ancient DNA Sequencing Data. In: Shapiro, B., Hofreiter, M., editors. *Ancient DNA: Methods and Protocols*.

Kirchman, J. J., Schirtzinger, E. E., & Wright, T. F. (2012). Phylogenetic relationships of the extinct Carolina Parakeet (*Conuropsis carolinensis*) inferred from DNA sequence data. *Auk, 129,* 197–204.

Kistler, L., Smith, O., Ware, R., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V., & Allaby, R. G. (2015). Thermal age, cytosine deamination and the veracity of 8,000 year old wheat DNA from sediments. *bioRxiv*. doi: http://dx.doi.org/10.1101/032060

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Research, 45*(11), 6310–6320. doi: 10.1093/nar/gkx361

Kjartansdóttir, K. (2019). The changing symbolic meaning of the extinct great auk and its afterlife as a museum object at the Natural History Museum of Denmark. *Nordisk Museologi, 2*, 41–56.

Köhler, M., Moyà-Solà, S. & Wrangham, R. W. (2008). Island rules cannot be broken. *Trends in Ecology and Evolution, 23*(1), 6-7.

Kowalski, K., Szyndlar, Z., Młynarski, M., Bocheński, Z., & Rzebik-Kowalska, B. (1993). Remains of small vertebrates from Bir Tarfawi and their palaeoecological significance. In Wendorf, F., Schild, R., & Close, A. E. (Eds.), *Egypt during the last interglacial*. New York: Plenum Press. pp. 155–204.

Kowalski, K., Van Neer, W., Bocheński, Z., Młynarski, M., Rzebik-Kowalska, B., Szyndlar, Z., Gautier, A., Schild, R., Close, A. E., & Wendorf, F. (1989). A last interglacial fauna from the eastern Sahara. *Quaternary Research, 32*(3), 335–341.

Kress, W. J., & Krupnick, G. A. (2022). Lords of the biosphere: Plant winners and losers in the Anthropocene. *Plants, People, Planet,* 1–17. https://doi.org/10.1002/ppp3.10252

Lalueza-Fox, C. (1999). *Missatges del passat*. Edicions Bromera, S.L. pp 208.

Lalueza-Fox, C. (2018). *De-extinctions*. Tibidabo Publishing, USA.

Lalueza-Fox, C. (2022). Museomics. *Current Biology 32*, R1–R2.

Lalueza-Fox, C. Shapiro, B., Bover, P., Alcover, J. A. & Bertranpetit, J. (2002). Molecular phylogeny and evolution of the extinct bovid *Myotragus balearicus*. *Molecular Phylogenetics and Evolution, 25*, 501–510.

Larrasoaña, J. C., Roberts, A. P., & Rohling, E. J. (2013). Dynamics of Green Sahara Periods and Their Role in Hominin Evolution. *Plos One, 8*(10), e76514.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLOS Biol., 10*, e1001388.

Leow, S., Bolsinger, J., Pronczuk, A., Hayes K. C., & Sambanthamurthi, R. (2016) Hepatic transcriptome implications for palm fruit juice deterrence of type 2 diabetes mellitus in young male Nile rats. *Genes & Nutrition, 11*,29. DOI 10.1186/s12263-016-0545-z

Levine, M. A. (1999). Investigating the origins of horse domestication. *Equine veterinary journal, 31*(S28), 6-14.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press. https://doi.org/10.1111/j.1558-5646.1975.tb00851.x

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25,* 1754–1760.

Lima-Ribeiro, M., Nogués-Bravo, D., Marske, K., Fernandez, F., Araujo, B., Diniz-Filho, J. A., & Diniz-Filho, F. (2012). Human arrival scenarios have a strong influence on interpretations of the late Quaternary extinctions. *Proceedings of the National Academy of Sciences, 109.* 10.1073/pnas.1206920109.

Lin, J., Duchene, D., Caroe, C., Smith, O., Ciucani, M. M., Niemann, J., Richmond, D., Greenwood, A. D., MacPhee, R., Zhang, G., Gopalakrishnan, S., Gilbert, M. T. P. (2022). Probing the genomic limits of de-extinction in the Christmas Island rat. *Current Biology, 32,* 1-7.

Lindahl, T. (1993a). Instability and decay of the primary structure of DNA. *Nature 362*, 709–715. doi:10.1038/362709a0

Lindahl, T. (1993b). Recovery of antediluvian DNA. *Nature, 365*(6448), 700-700.

Lindahl, T. & Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry 11*, 3610-3618.

Lindgreen, S. (2012). AdapterRemoval: easy cleaning of nextgeneration equencing reads. *BMC Res. Notes, 5*, 337.

Lomolino, M. V. (1985). Body Size of Mammals on Islands: The Island Rule Reexamined. *The American Naturalist, 125*(2), 310-316.

Lomolino, M. V., Sax, D. F., Palombo M. R., & van der Geer, A. A. (2012). Of mice and mammoths: evaluations of causal explanations for body size evolution in insular mammals. *Journal of Biogeography, 39*, 842–854.

Lomolino, M. V., van der Geer, A. A., Lyras, G. A., Palombo, M. R., Sax, D. F., & Rozzi, R. (2013). Of mice and mammoths: generality and antiquity of the island rule. *Journal of Biogeography, 40*: 1427–1439.

López-Martínez, N. & López-Jurado, L. F. (1987). Un nuevo murido gigante del Cuaternario de Gran Canaria *Canariomys tamarani* nov. sp. (Rodentia, Mammalia). Doñana. *Acta Vertebrata, 2*, 1–66.

Lord, E., Dussex, N., Kierczak, M., Díez-del-Molino, D., Ryder, O. A., Stanton, D. W. G., Gilbert, M. T. P., Sánchez-Barreiro, F., Zhang, G., Sinding, M. H. S., Lorenzen, E. D., Willerslev, E., Protopopov, A., Shidlovskiy, F., Fedorov, S., Bocherens, H., Nathan, S. K. S. S., Goossens, B., van der Plicht, J., Chan, Y.

L., Prost, S., Potapova, O., Kirillova, I., Lister, A. M., Heintzman, P. D., Kapp, J. D., Shapiro, B., Vartanyan, S., Götherstrom, A. & Dalén, L. (2020). Pre-extinction Demographic Stability and Signatures of Adaptation in the Woolly Rhinoceros. *Current Biology 30*, 3871-3879.

Louchart, A., & Mourer-Chauviré, C. (2011). The dodo was not so slim: leg dimensions and scaling to body mass. *Die Naturwissenschaften, 98*(4): 357–8.

Luo, Z. (2007).Transformation and diversification in early mammal evolution. *Nature, 450,* 1011-1019. doi:10.1038/nature06277

Luther, D. (1986). *Die ausgestorbenen Vögel der Welt*. Fourth Edition (Westarp-Wissenschaften).

MacFadden, B. J. (2005). Fossil Horses--Evidence for Evolution. *Science, 307*, 1728.

Machado Yanes, M. C., del Arco Aguilar, M. C., Vernet, J. L. & Ourcival, J. M. (1997). Man and Vegetation in Northern Tenerife (Canary Islands, Spain), during the Prehispanic Period Based on Charcoal Analyses. Vegetation History and *Archaeobotany, 6*(3), 187–95. http://www.jstor.org/stable/23417333.

MacPhee, R. D. E., & Fleming, C. (2003). A possible heptaxodontine andother caviidan rodents from the Quaternary of Jamaica. *American Museum Novitates, 3422*, 1–42.

Manzi, V., Gennari, R., Hilgen, F., Krijgsman, W., Lugli, S., Roveri, M., & Sierro, F.J., (2013). Age refinement of the Messinian salinity crisis onset in the Mediterranean. *Terra Nova, 25*, 315e327. https://doi.org/10.1111/ter.12038.

Marsh, O. O. (1879). *Polydactyle Horses*, Recent and Extinct.

Marshall, C. R., Latorre, D. V., Wilson, C. J., Frank, T. M., Magoulick, K. M., Zimmt, J. B., & Poust, A. W. (2021). Absolute abundance and preservation rate of Tyrannosaurus rex. *Science, 372*, 284–287.

Martella, M. B., & Bucher, E. H. (1990). Vocalizations of the Monk Parakeet. *Bird Behavior, 8*(2), 101-110. DOI:10.3727/015613890791784290

Martin, P. S. (1973). Discovery of America. *Science, 179*, 969–974.

Martin, P. S. (1984). Prehistoric overkill: The global model. In Martin, P. S., Klein, R. G. (eds.). *Quaternary Extinctions: A Prehistoric Revolution*. University of Arizona Press. pp. 354–403.

Martin, R. D. (1993). Primate origins: plugging the gaps*. Nature, 363,* 223-234.

Martín, A., Hérnández, M. A., Lorenzo, J. A., Nogales, M. & González, C. (2000). *Las palomas endémicas de Canarias*. Consejería de Política y Medio Ambiente del Gobierno de Canarias and SEO/Birdlife.

Martin, P. S. & Steadman, D. W. (1999). Prehistoric extinctions on islands and continents. In MacPhee, R. D. E (ed.). *Extinctions in near time: causes, contexts and consequences*.

Martínez-Gómez, J. E., Matías-Ferrer, N., & Escalante-Pliego, P. (2017). Phylogeny and taxonomy of the Socorro parakeet (*Psittacara holochlorus brevipes*): recent speciation with minor morphological differentiation. *Journal of Ornithology, 158*(4), 965-978. DOI 10.1007/s10336-017-1465-4

Martinkova, N., Barnett, R., Cucchi, T., Struchen, R., Pascal, M., Pascal, M., Fischer, M. C., Higham, T., Brace, S., Simon, Y. W. H., et al. (2013). Divergent evolutionary processes associated

with colonization of offshore islands. *Mol. Ecol., 22*, 5205–5220.

Masseti, M. (2010). Holocenemammals of Libya: A biogeographical, historical and archaeozoological approach. *Journal of Arid Environments, 74*, 794–805.

Mateo, J. A. & López-Jurado, L. F. (1992). Study of dentition in lizards from Gran Canaria island (Canary Islands) and its ecological and evolutionary significance. *Biol. J. Linn. Soc., 4,* 39-48.

Mattoon, W. R. (1915). *The Southern Cypress*. In: USDA Bulletin No. 272 Washington, D.C.

May, R. M (1988). How Many Species are There on Earth? *Science, 241*(4872), 1441-1449.

May, R. M., Lawton, J. H., & Stork, N. E. (1995). *Assessing extinction rates*. In Extinction Rates (J. H. Lawton and R. M. May, Eds.), Oxford University Press, Oxford. pp. 1–24.

Mayamba, A., Modest, R., Makundi, R., Kimaro, D., Isabirye, M., Massawe, A., Kifumba, D., Nakiyemba, A., Leirs, H., Mdangi, M., Rianah, S., & Mulungu, L. (2019). Species composition and community structure of small pest rodents (Muridae) in cultivated and fallow fields in maize-growing areas in Mayuge district, Eastern Uganda. *Ecology and Evolution, 9*. 10.1002/ece3.5371.

Mayr, E. (1963). *Animal Species and Evolution*. Belknap Press. pp

Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R. W. R., Zhang, L., Papador, J. D., Martinez-Najera, D., Jiggins, C. D., Kronforst, M. R., Breuker, C. J., Reed, R. D., Patel, N. H., McMillan, W. O. & Martin, A. (2017). Macroevolutionary shifts of WntA function potentiate butterfly

wing-pattern diversity. *Proc. Natl. Acad. Sci., 114*, 10701–10706.

McElhinny, T.L., Smale, L. & Holekamp, K. E. (1997). Patterns of body temperature, activity, and reproductive behavior in a tropical murid rodent, *Arvicanthis niloticus*. *Physiol. Behav., 62*(1), 91–96.

McKinley, D. (1977). Climatic relations, seasonal mobility, and hibernation in the Carolina Parakeet. *Jack-Pine Warbler, 55*, 107–124.

McKinley, D. (1979). Historical Review of the Carolina Parakeet in the Carolinas. *Brimleyana, 1*, 81-98.

McKinley, D. (1980). The balance of decimating factors and recruitment in extinction of the Carolina Parakeet. Part 1. *Indiana Audubon Q., 58*: 8–18.

McKinley, D., & Hardy, J. W. (1985). The Carolina Parakeet in Florida. *Florida Ornithological Society*.

Mein, P., & Pickford, M. (1992). Gisements karstiques pléistocènes au Djebel Ressas, Tunisie. *Comptes Rendus de l'Académie des Sciences, Paris, Série II, 315*, 247–253.

Meiri, S., Cooper, N., & Purvis, A. (2008). The island rule: made to be broken? *Proc. R. Soc. B.,* 275141–148.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature reviews. Genetics, 11*(1), 31-46.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N, Tandon, A., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M. F.,

Shunkov, M- V., Derevianko, A. P., Patterson, N., Andrés, A. M., Eichler, E. E., Slatkin, M., Reich, D., Kelso, K. & Pääbo, S. (2012). A high-coverage genome sequence from an archaic denisovan individual. *Science, 338*, 222–226.

Michaux, J. R., Goüy De Bellocq, J., Sarà, M., & Morand, S. (2002). Body size increase in insular rodent populations: a role for predators? *Global Ecology & Biogeography, 11*, 427–436.

Michaux, J., López-Martínez, N., & Hernández-Pacheco, J. J. (1996). A 14C dating of *Canariomys bravoi* (mammalia Rodentia), the extinct giant rat from Tenerife (Canary Islands, Spain), and the recent history of the endemic mammals in the archipelago. *Vie Milieu, 46*(3/4): 261-266.

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., Morishita, S., & Sasaki, T. (2004). The genome sequence of silkworm, *Bombyx mori*. *DNA Res, 11*(1), 27-35. doi: 10.1093/dnares/11.1.27.

Mitchell, K. J., Llamas, B., Soubrier, J., Rawlence, N. J., Worthy, T. H., Wood, J., Lee, M. S. Y., & Cooper, A. (2014). Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science, 344*(6186), 898–900. doi:10.1126/science.1251981

Mojica, F.J., & Montoliu, L. (2016). On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals". *Trends in Microbiology, 24*(10), 811–820. doi:10.1016/j.tim.2016.06.005.

Monadjem, A,, Taylor, P. J., Denys, C., Cotterill, F. P. D. (2015). *Rodents of Sub-Saharan Africa. A biogeographic and taxonomic synthesis*. Berlin, Germany: De Gruyter.

Moncunill-Solé, B., Jordana, X., Marín-Moratalla, N., Moyà-Solà, S., & Köhler, M. (2014). How large are the extinct giant insular rodents? New body mass estimations from teeth and bones. *Integr. Zool., 9*, 197–212. doi:10.1111/1749-4877.12063

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biol, 9*(8), e1001127. doi:10.1371/journal.pbio.1001127

Morgan, G. S. & Woods, C. A. (1986). Extinction and the zoogeography of West Indian land mammals. *Biol. J. Linn. Soc., 28*(1-2): 167-203.

Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., Demboski, J. R., Doll, A., Da Fonseca, R. R., Fulton, T. L., Gilbert, M. T. P., Heintzman, P. D., Letts, B., McIntosh, G., O'Connell, B. L., Peck, M., Pipes, M. L., Rice, E. S., Santos, K. M., Sohrweide, A. G., Vohr, S. H., Corbett-Detig, R. B., Green, R. E. & Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science 17, 358*(6365): 951-954. doi: 10.1126/science.aao0960.

Murray, P. F., & Vickers-Rich, P. (2004). *Magnificent Mihirungs: The Colossal Flightless Birds of the Australian Dreamtime*. Indiana University Press. pp. 416.

Nakazawa, T., Ohba, S., & Ushio, M. (2013). Predator–prey body size relationships when predators can consume prey larger than themselves. *Biol Lett, 9*, 20121193. http://dx.doi.org/10.1098/rsbl.2012.1193

Neal, B. R. (1981). Reproductive biology of the unstriped grass rat, Arvicanthis, in East Africa. *Z. Saugetierk, 46*:174-189.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol., 48*, 443–453.

Neel, J. V. (1962). Diabetes mellitus: a thrifty genotype rendered detrimental by "progress"? *Am J Human Genet, 14*: 353-62.

Nishino, T. & Morikawa, K. (2002). Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene, 21*, 9022-9032.

Noda, K., Melhorn, M. I., Zandi, S., Frimmel, S., Tayyari, F., Hisatomi, T., Almulki, L., Pronczuk, A., Hayes, K. C., & Hafezi-Moghadam, A. (2010). An animal model of spontaneous metabolic syndrome: Nile grass rat. *The FASEB Journal, 2*4, 2443-2453. DOI: 10.1096/fj.09-152678

Nogués-Bravo, D., Rodríguez, J., Hortal, J., Batra, P., Araújo, M. B., & Barnosky, A. (2008). Climate change, humans, and the extinction of the woolly mammoth. *PLoS biology, 6*(4), p.e79-e79

Notarbartolo di Sciara, G., Politi, E., Bayed, A., Beaubrun, P. C., Knowlton, A. (1998). A winter survey off southern Morocco, with a special emphasis on right whales. *Report of the International Whaling Commission, 48*: 547–550.

Novak, B.J. (2018). De-Extinction. *Genes (Basel), 9*(11): 548.

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., Collins, L. S., de Queiroz, A., Farris, D. W., Norris, R. D., et al. (2016). Formation of the Isthmus of Panama. *Sci. Adv., 2*, e1600883.

Ogura, A. (2020). How to improve mouse cloning. *Theriogenology, 150*, 215-220 https://doi.org/10.1016/j.theriogenology.2020.01.038

Olah, G., Butchart, S. H., Symes, A., Guzmán, I. M., Cunningham, R., Brightsmith, D. J., & Heinsohn, R. (2016) Ecological and socioeconomic factors affecting extinction risk in parrots. *Biodivers. Conserv., 25*, 205–223.

Olalde Marquínez, Í. (2016). *From the Mesolithic to the Bronze Age: unraveling 5.000 years of European population history with paleogenomics*. Doctoral thesis. Universitat Pompeu Fabra. pp. 195.

O'Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., Goldberg, S. L., Kraatz, B. P., Luo, Z-X., Meng, J., Ni, X., Novacek, M. J., Perini, F. A., Randall, Z. S., Rougier, G. W., Sargis, E. J., Silcox, M. T., Simmons, N. B., Spaulding, M., Velazco, P. M., Weksler, M., Wible, J. R., & Cirranello, A. L. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science, 339,* 662–667.

Oliveros, C. H., Field, D. J., Ksepka, D. T., Barker, F. K., Aleixo, A., Andersen, M. J., Alström, P., Benz, B. W., Braun, E. L., Braun, M. J., et al. (2019). Earth history and the passerine superradiation. *Proc. Natl. Acad. Sci. USA, 116,* 7916–7925.

Olson, S. L. (1985). The fossil record of birds. In *Avian Biology* (Farmer, D. S., King, J. R., & Parkes, K. C. Eds.). Academic Press, Orlando, Florida. pp 79-256.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A., Vogt, J., Szklarczyk, D., Kelstrup, C. D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A. M. V., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G. D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J. F., Weinstock, J., Gregersen,

K., Røed, K. H., Eisenmann, V., Rubin, C. J., Miller, D. C., Antczak, D. F., Bertelsen, M. F., Brunak, S., Al-Rasheid, K. A. S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M. T. P., Kjær, K., Sicheritz-Ponten, T., Jensen, L. J., Olsen, J. V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., & Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature, 499*, 74-78 doi:10.1038/nature12323

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews. Methods Primers, 1,*14. https://doi.org/10.1038/s43586-020-00011-0

Ouahbi, Y., Aberkan, M., & Serre, F. (2003). Recent Quaternary fossil mammals of Chrafate and Ez Zarka. The origin of modern fauna in the Northern Rif (NW Morocco, Northern Africa). *Geologica Acta, 1*(3), 277–288.

Owen-Smith, R. N. (1988). *Megaherbivores: The Influence of Very Large Body Size on Ecology*. Cambridge University Press, New York.

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature, 314*(6012), p.644-645

Pääbo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. USA, 86*, 1939-1943.

Paez, S., Kraus, R. H. S., Shapiro, B., Gilbert, M. T. P., & Jarvis, E. D. (2022). Reference genomes for conservation. *Science, 377*(6604), 364-366. DOI:10.1126/science.abm8127

Pagès, M., Chevret, P., Gros-Balthazard, M., Hughes, S., Alcover, J.A., Rainer, H., Rando, J. C., Michaux, J., & Hänni, C. (2012).

Paleogenetic analyses reveal unsuspected phylogenetic affinities between mice and the extinct *Malpaisomys insularis*, an endemic rodent of the Canaries. *PLoS ONE, 7*(2), e31123. doi:10.1371/journal.pone.0031123

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A. M., To, T. H., Kortschak, D., Raison, J. M., Qu, Z., Chin, T.-J., Alt, K. W., Claesson, S., Dalén, L., MacPhee, R. D. E., Meller, H., Roca, A. L., Ryder, O. A., Heiman, D., Young, S., Breen, M., Williams, C., Aken, B. L., Ruffier, M., Karlsson, E., Johnson, J., Di Palma, F., Alfoldi, J., Adelson, D. L., Mailund, T., Munch, K., Lindblad-Toh, K., Hofreiter, M., Poinar, H. & Reich, D. (2018). A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences USA, 115*(11), E2566-E2574.

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D. & Dalén, L. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology, 25*, 1395–1400.

Park, S. D. E., Magee, D. A., McGettigan, P. A., Teasdale, M. D., Edwards, C. J., Lohan, A. J., Murphy, A., Braud, M., Donoghue, M. T., Liu, Y., Chamberlain, A. T., Rue-Albrecht, K., Schroeder, S., Spillane, C., Tai, S., Bradley, D. G., Sonstegard, T. S., Loftus, B. J. & MacHugh, D. E. (2015). Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol., 16,* 234. DOI: http://dx.doi.org/10.1186/s13059-015-0790-2

Parmalee, P. W. (1958). Remains of rare and extinct birds from Illinois Indian sites. *Auk, 75*, 169–176.

Parmalee, P. W. (1967). Additional noteworthy records of birds from archaeological sites. *Wilson Bulletin, 79*, 155–162.

Pedersen, J.S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S. et al. (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res., 24*, 454–466.

Peers, M. J., Thornton, D. H., Majchrzak, Y. N., Bastille-Rousseau, G., & Murray, D. L. (2016). De-extinction potential under climate change: Extensive mismatch between historic and future habitat suitability for three candidate birds. *Biological Conservation, 197*, 164–170.

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarrés, J. F., Araújo, M. B., Balvanera, P., Biggs, R., Cheung, W. W. L., Chini, L., H. Cooper, H. D., Gilman, E. L., Guénette, S., Hurtt, G. C., Huntington, H. P., Mace, G. M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R. J., Sumaila, U. R., & Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science, 330*, 1496–1501.

Pierce, N, E. & Eastealt, S. (1986). The selective advantage of attendant ants for the larvae of a Lycaenid butterfly, *Glaucopsyche lygdamus*. *Journal of Animal Ecology, 55*(2), 451-462.

Pievani, T. (2014). The sixth mass extinction: Anthropocene and the human impact on biodiversity. *Rend. Fis. Acc. Lincei, 25*:85–93. DOI 10.1007/s12210-013-0258-9

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science, 344*, 1246752.

DOI: 10.1126/science.1246752

Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. (1995). The future of biodiversity. *Science, 269*, 347–350.

Poinar Jr, G. O. (1993). Insects in amber. *Annual review of entomology, 38*(1), 145-159.

Poulet, A. R. (1985). The ecological basis forecasting rodent outbreaks in a Sahelian agrosystem. *Acta Zoologica Fennica, 173*, 107-111.

Pregil, G. K. (1986). Body size of insular lizards: a patern of Holocene dwarfism. *Evolution, 40*, 997-1008.

Prescott, G. W., Williams, D. R., Balmford, A., Green, R. E., & Manica, A. (2012). Quantitative global analysis of the role of climate and people in explaining late Quaternary megafaunal extinctions. *Proceedings of the National Academy of Sciences., 109*(12), 4527–4531. doi/10.1073/pnas.1113875109

Prevec, R. (1984). The Carolina Parakeet - its first appearance in southern Ontario. *Kewa, 84*,4–8.

Proaktor, G., Coulson, T., & Milner-Gulland, E. J. (2007). Evolutionary Responses to Harvesting in Ungulates. *Journal of Animal Ecology, 76*(4), 669–678. http://www.jstor.org/stable/4539172

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., De Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature, 505*, 43–49.

Pruvost, M., Bellone, R., Benecke, N., Sandoval-Castellanos, E., Cieslak, M., Kuznetsova, T., Morales-Muñiz, A., O'Connor, T.,

Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Genotypes of predomestic horses match phenotypes painted in Paleolithic works of cave art. *Proceedings of the National Academy of Sciences – PNAS, 108(*46), 18626-18630.

Pujol, A. (2016). Transgénesis en animales: las técnicas clásicas y las novedades. *Animales de Laboratorio. Revista de la Sociedad Española para las Ciencias del Animal de Laboratorio, 70*, 18-23.

Rando, J. C. (2003). Protagonistas de una catástrofe silenciosa: Los vertebrados extintos de Canarias. *El Indiferente, 14*, 4-15.

Rando, J. C., Alcover, J. A., Galván, B., & Navarro, J. F. (2014). Reappraisal of the extinction of *Canariomys bravoi*, the giant rat from Tenerife (Canary Islands). *Quaternary Science Reviews, 94,* 22-27.

Rahbek, C. & Graves, G. R. (2001). Multiscale assessment of patterns of avian species richness. *PNAS, 98*(8), 4534–4539. www.pnas.orgycgiydoiy10.1073ypnas.071034898

Ratcliffe, S. ed. (2016). *Roy Amara 1925–2007, American futurologist. Oxford Essential Quotations.* Vol. 1 (4th ed.). Oxford University Press. doi:10.1093/acref/9780191826719.001.0001.

Raup, D. (1992). *Extinction: bad genes or bad luck?* Norton, New York. pp. 210.

Ravussin, E. & Bogardus, C. (1990). Energy expenditure in the obese: is there a thrifty gene? *Infusionstherapie, 17*(2), 108-12. doi: 10.1159/000222456.

Refinetti, R. (2004). The Nile Grass Rat as a Laboratory Animal. *Resource, 33,* 9.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. & Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature, 468*, 1053–1060. DOI 10.1038/nature09710

Reinhardt, J. T. (1842). In: Kroyer H, editor. *Noiere Oplysning om det I Kiobenhavn fundne Drontehoved*. Vol. 4. København (Denmark): Naturhistorisk Tidskrift. p. 71.

Remsen, J. V, Schirtzinger, E. E., Ferraroni, A., Silveira, L. F., & Wright, T. F. (2013). DNA-sequence data require revision of the parrot genus Aratinga (Aves: Psittacidae). *Zootaxa, 3641*(3), 296–300.

Richmond, D. J., Sinding, M. H. S. & Gilbert, M. T. P. (2016). The potential and pitfalls of de-extinction. *Zoologica Scripta, 45*(S1), 22-36.

Riordan, E. C., & Philip, W. (2009). Rundel Modelling the distribution of a threatened habitat: the California sage scrub. *Journal of Biogeography, 36*, 2176-2188.

Rodríguez-Pastor, R., Senar, J. C., Ortega, A., Faus, J., Uribe, F. & Montalvo, T. (2012). Distribution patterns of invasive Monk parakeets (*Myiopsitta monachus*) in an urban habitat. *Animal Biodiversity and Conservation, 35*(1), 107–117.

Roelants, K., Gower, D. J., Wilkinson, M., Loader, S. P., Biju, S. D., Guillaume, K., Moriau, L., & Bossuyt, F. (2007).Global patterns of diversification in the history of modern amphibians. *Proc. Natl Acad. Sci. USA, 104*, 887–892.

Rogers, R. L., & Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet, 13,* e1006601. https://doi.org/10.1371/journal.pgen.1006601

Rollo, F., Amici, A., Salvi, R., & Garbuglia, A. (1988). Short but faithful pieces of ancient DNA. *Nature, 335*, 774. doi:10.1038/335774a0

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. A., Weinert, L. A., Belkhir, K., Bierne, N., Glémin, S., et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature, 515*, 261-263 https://doi.org/10.1038/nature13685, PMID: 25141177

Saccheri, I. J., Nichols, R. A., & Brakefield, P. M. (2001). Effects of bottlenecks on quantitative genetic variation in the butterfy *Bicyclus anynana*. *Genet. Res., Camb. 77,* 167-181.

Saltré, F., Chadoeuf, J., Peters, K. J., McDowell, M. C., Friedrich, T., Timmermann, A., Ulm, S., & Bradshaw, C. J. A. (2019). Climate-human interaction associated with southeast Australian megafauna extinction paterns. *Nature Communications, 10*(1), 5311-9.

Salvador, A., & Nores, C. (2017). Ballena de los vascos – *Eubalaena glacialis*. In: *Enciclopedia Virtual de los Vertebrados Españoles*. Salvador, A., Barja, I. (Eds.). Museo Nacional de Ciencias Naturales, Madrid. http://www.vertebradosibericos.org/

Sánchez-Bayoa, F., & Wyckhuysb, K. A. G. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation, 232*, 8–27. https://doi.org/10.1016/j.biocon.2019.01.020

Scalercio, S., Cini, A., Menchetti, M., Vodă, R., Bonelli, S., Bordoni, A., Casacci, L. P., Dincă, V., Balletto, E., Vila, R., & Dapporto, D. (2020). How long is 3 kilometres for a butterfly? Ecological constraints and functional traits explain strong matriline differentiation between Sicily and the Italian Peninsula. *Journal of Animal Ecology, 89*, 2013-2026. DOI: 10.1111/1365-2656.13196

Schulte, P., Alegret, L., Arenillas, I., Arz, J. A., Barton, P. J., Bown, P. R., Bralower, T. J., Christeson, G. L., Claeys, P., Cockell, C. S., Collins, G. S., Deutsch, A., Goldin, T. J., Goto, K., Grajales-Nishimura, J. M., Grieve, R. A. F., Gulick, S. P. S., Johnson, K. R., Kiessling, W., Koeberl, C., Kring, D. A., MacLeod, K. G., Matsui, T., Melosh, J., Montanari, A., Morgan, J. V., Neal, C. R., Nichols, D. J., Norris, R. D., Pierazzo, E., Ravizza, G., Rebolledo-Vieyra, M., Reimold, W. U., Robin, E., Salge, T., Speijer, R. P., Sweet, A. R., Urrutia-Fucugauchi, J., Vajda, V., Whalen, M. T., & Willumsen, P. S. (2010). Chicxulub Asteroid Impact and Mass Extinction at the Cretaceous-Paleogene Boundary. *Science, 327*(5970), 1214-1218.

Schulz, L. O. & Chaudhari, L. S. (2015). High-Risk Populations: The Pimas of Arizona and Mexico. *Curr Obes Rep, 4*, 92–98. DOI 10.1007/s13679-014-0132-9

Schweizer, M., Seehausen, O., Güntert, M., & Hertwig, S. T. (2010). The evolutionary diversification of parrots supports a taxon pulse model with multiple trans–oceanic dispersal events and local radiations. *Molecular Phylogenetics and Evolution, 54*(3), 984–994. Doi: 10.1016/j.ympev.2009.08.021

Schweizer, M., Hertwig, S. T., Seehausen, O., & Ebach, M. (2014). Diversity versus disparity and the role of ecological opportunity in a continental bird radiation. *J Biogeogr, 41*(7), 1301–12.

Scott, J. A. (1986). *The Butterflies of North America: A Natural History and Field Guide*. Stanford University Press. pp. 583.

Seddon, P. J., Moehrenschlager, A. & Ewen, J. (2014). Reintroducing resurrected species: selecting DeExtinction candidates. *Trends in Ecology and Evolution, 29*(3), 140-147.

Senar, J. C., Mateos-Gonzalez, F., Uribe, F., & Arroyo, L. (2013) Familiarity adds to attractiveness in matters of siskin mate choice. *Proc R Soc B 280,* 20132361. http://dx.doi.org/10.1098/rspb.2013.2361

Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G. M., Bininda-Emonds, O. R. P., Lee, P. L. M., & Cooper, A. (2002). Flight of the Dodo. *Science, 295*(5560). DOI: 10.1126/science.295.5560.1683

Short, A.E., Dikow, T., & Moreau, C. S. (2018). Entomological collections in the age of big data. *Annu. Rev. Entomol., 63*, 513–530. doi:10.1146/annurev-ento-031616-035536

Sicard, B., Catalan, J., Ag'Atteynine, S., Abdoulaye, D., & Britton-Davidian, J. (2004). Effects of climate and local aridity on the latitudinal and habitat distribution of *Arvicanthis niloticus* and *Arvicanthis ansorgei* (Rodentia, Murinae) in Mali. *Journal of Biogeography, 31*, 5-18.

Sinding, M. S. & Gilbert, M. T. P. (2016) The draft genome of extinct european Aurochs and its implications for de-extinction. *Open Quaternary, 2*: 1–9.

Skoglund, P. et al. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. USA., 111*, 2229– 2234.

Smith, J. L. B. (1956). *Old Fourlegs: The story of the coelacanth.* Longmans, Green and Co Ltd. London. pp. 260.

Smith, F. A., Boyer, A. G., Brown, J. H., Costa, D. P., Dayan, T., Ernest, S. K. M., Evans, A. R., Fortelius, M., Gittleman, J. L., Hamilton, M. J., Harding, L. E., Lintulaakso, K., Lyons, S. K., McCain, C., Okie, J. G., Saarinen, J. J., Sibly, R. M., Stephens, P. R., Theodor, J., & Uhen, M. D. (2010). The Evolution of Maximum Body Size of Terrestrial Mammals. *Science, 330*(6008), 1216-1219. doi:10.1126/science.1194830

Smith, N. A., & Clarke, J. A. (2015). Systematics and evolution of the Pan-Alcidae (Aves, Charadriiformes). *Journal of Avian Biology, 45*, 125–140.

Snyder, N. F. (2004). *The Carolina Parakeet: Glimpses of a vanished bird.* Princeton: Princeton University Press. pp 168.

Snyder, N. F., & Russell, K. (2002). Carolina Parakeet (*Conuropsis carolinensis*). In A. Poole (Ed.), *Birds of North America Online, Ithaca: Cornell Lab of Ornithology*, https://doi.org/10.2713/bna.667

Solly, M. (2019a). A Perfectly Preserved 32,000-Year-Old Wolf Head Was Found in Siberian Permafrost. *Smithsonian Magazine.* www.smithsonianmag.com/smart-news/perfectly-preserved-32000-year-old-wolf-head-was-found-siberia-180972418/

Solly, M. (2019b). Scientists Extracted Liquid Blood From 42,000-Year-Old Foal Found in Siberian Permafrost. *Smithsonian Magazine.* www.smithsonianmag.com/smart-news/scientists-extracted-liquid-blood-42000-year-old-foal-found-siberian-permafrost-180971979/

Spielman, D., Brook, B. W., & Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A., 101*, 15261–15264. https://pubmed.ncbi.nlm.nih.gov/15477597

Staats, M., Erkens, R. H. J., van de Vossenberg, B., Wieringa, J. J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J. E., & Bakker, F. T. (2013). Genomic Treasure Troves: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. *PLoS One, 8,* e69189–e69189.

Stahle, D. W., Cleaveland, M. K., & Hehr, J. G. (1988). North Carolina Climate Changes Reconstructed from Tree Rings: A.D. 372 to 1985. *Science, 240*(4858): 1517-1519. DOI: 10.1126/science.240.4858.1517

Stahle, D. W., Cleaveland, M. K., Griffin, R. D., Spond, M. D., Fye, F. K., Culpepper, R. B., & Patton, D. (2006). Decadal drought effects on endangered woodpecker habitat. *Eos, Transactions of the American Geophysical Union, 87*(12), 121-125.

Stahle, D. W., Burnette, D. J., Villanueva, J., Cerano, J., Fye, F. K., Griffin, R. D., Cleaveland, M. K., Stahle, D. K., Edmondson, J. R., & Wolff. K. P. (2012). Tree-ring analysis of ancient baldcypress trees and subfossil wood. *Quaternary Science Reviews, 34,* 1-15.

Stahle, D. W., & Hawks, G. G. (2016). *The Ancient Bald Cypress.* Consortium Black River, North Carolina.

Stanley, S. (1978). Chronospecies' longevities, the origin of genera, and the punctuational model of evolution. *Paleobiology, 4*(1), 26-40. doi:10.1017/S0094837300005662

Steppan, S. J. (1991). Geographic distribution of flower morphological traits in subspecies of *Lotus scoparius*. *Journal of Biogeography, 18*, 321-331.

Stoetzel, E. (2017). Adaptations and Dispersals of Anatomically Modern Humans in the Changing Environments of North

Africa: the Contribution of Microvertebrates. *Afr Archaeol Rev*. https://doi.org/10.1007/s10437-017-9272-0

Stoetzel, E., Campmas, E., Michel, P., Bougariane, B., Ouchaou, B., Amani, F., El Hajraoui, M. A., & Nespoulet, R. (2014). Context of modern human occupations in North Africa: Contribution of the Témara caves data. *Quaternary International-Special Issue: Northwest African Prehistory, 320*, 143–161.

Stokstad, E. (2015). Bringing back the aurochs. *Science, 350*, 1144–1147. http://dx.doi.org/10.1126/ science.350.6265.1144

Strickland, H. E., & Melville, A. G. (1848). *The dodo and its kindred, or the history, affinities and osteology of the dodo, solitaire and other extinct birds of the islands Mauritius, Rodriguez and Bourbon*. London. 141 pp.

Stuart, A. J. (2015). Late Quaternary megafaunal extinctions on the continents: a short review. *Geological Journal, 50*: 338-363. DOI: 10.1002/gj.2633

Subramaniam, A., Landstrom, M., Luu, A. & Hayes, K. C. (2018). The Nile Rat (*Arvicanthis niloticus*) as a Superior Carbohydrate-Sensitive Model for Type 2 Diabetes Mellitus (T2DM). *Nutrients, 10*, 235. doi:10.3390/nu10020235

Subramaniam, A., Landstrom, M., & Hayes, K. C. (2019) Genetic Permissiveness and Dietary Glycemic Load Interact to Predict Type-II Diabetes in the Nile rat (*Arvicanthis niloticus*). *Nutrients, 11*, 1538. doi:10.3390/nu11071538

Tchernov, E. (1968). *Succession of rodent faunas during the Upper Pleistocene of Israel.* Verlag Paul Parey: Hamburg.

Telschow, A., Hammerstein, P. & Werren, J. H. (2005). The effect of Wolbachia versus genetic incompatibilities on reinforcement and speciation. *Evolution (N.Y), 59*, 1607–1619.

Thayer, R. C., Allen, F. I., & Patel, N. H. (2020). Structural color in Junonia butterflies evolves by tuning scale lamina thickness. *eLife, 9*, e52187. DOI: https://doi.org/10.7554/eLife.52187

Thomas, C. D. & Abery, J. C. G. (1995). Estimating rates of butterfly decline from distribution maps: the effect of scale. *Biol. Conserv., 73*, 59–65.

Thomas, R. H., Schaffner, W., Wilson, A. C., Pääbo, S. (1989). DNA phylogeny of the ancient marsupial wolf. *Nature, 340*, 465–467. doi:10.1038/340465a0

Tilden, J. W. (1956). San Francisco's vanishing butterflies. *The Lepidopterists' News. San Jose State College, 10* (3-4), 113-115.

Toh, H., Yang, C., Formenti, G., Raja, K., Yan, L., Tracey, A., Chow, W., Howe, K., Bergeron, L. A., Zhang, G., Haase, B., Mountcastle, J., Fedrigo, O., Fogg, J., Kirilenko, B., Munegowda, C., Hiller, M., Jain, A., Kihara, D., Rhie, A., Phillippy, A. M., Swanson, S. A., Jiang, P., Clegg, D. O., Jarvis, E. D., Thomson, J. A., Stewart, R., Chaisson, M. J. P., & Bukhman, Y. V. (2022). A haplotype-resolved genome assembly of the Nile rat facilitates exploration of the genetic basis of diabetes. *BMC Biol 20*, 245. https://doi.org/10.1186/s12915-022-01427-8

Trager, M. D., & Daniels, J. C. (2009). Ant tending of Miami blue butterfly larvae (Lepidoptera: Lycaenidae): partner diversity and effects on larval performance. *Flo. Entomol., 92*, 474–482. doi:10.1653/024.092.0309

True Nature Foundation (2016). Species Restoration (WWW Document). www.truenature foundation.org/wildlife/aurochs

Trut, L., Oskina, I., & Kharlamova, A. (2009). Animal evolution during domestication: the domesticated fox as a model. *Bioessays, 31*, 349-360.

Tsai, C., Hsieh, C. & Nakazawa, T. (2016). Predator–prey mass ratio revisited: does preference of relative prey body size depend on individual predator size? *Functional Ecology, 30,* 1979–1987. doi: 10.1111/1365-2435.12680

Turveya, S. T. & Cheke, A. S. (2008). Dead as a dodo: the fortuitous rise to fame of an extinction icon. *Historical Biology, 20*(2), 149–163. DOI: 10.1080/08912960802376199

Upham, N., Esselstyn, J. & Jetz, W. (2016). In: Wilson, D. E., Lacher, T. E., Jr & Mittermeier, R. A. eds. *Handbook of the Mammals of the Word. Vol. 6 Lagomorphs and Rodents*.

Urantowka, A. D., Kroczak, A., & Mackiewicz, P. (2017). The influence of molecular markers and methods on inferring the phylogenetic relationships between the representatives of the Arini (parrots, Psittaciformes), determined on the basis of their complete mitochondrial genomes. *BMC Evolutionary Biology, 17*, 166. DOI 10.1186/s12862-017-1012-1

Van der Broek, R. (1972).*The Myth of the Phoenix, According to Classical and Early Christian Traditions*. Translation by Seeger, I., & Brill, E. J. Leiden. pp. 487.

Van der Geer, A. A., Lyras, A. G., Lomolino, M. V., Palombo, M. R. & Sax, D. F. (2013). Body size evolution of palaeo-insular mammals: temporal variations and interspecific interactions. *Journal of Biogeography, 40*, 1440–1450. doi:10.1111/jbi.12119

Van der Geer, A., Lyras, G., De Vos, J., Dermitzakis, M., (2010). Evolution of Island Mammals. Adaptation and Extinction of Placental Mammals on Islands. Wiley-Blackwell, Hoboken, New Jersey.

Velasco Vázquez, J., Alberto Barroso, V., Delgado Darias, T., Moreno Benítez, M., Lecuyer , C., & Richardin, P. (2020). Poblamiento, colonización y primera historia de Canarias: el C14 como paradigma. *Anuario de Estudios Atlánticos, 66*, 066-001 http://anuariosatlanticos.casadecolon.com/index.php/aea/article/view/10530/9904

Vernot, B. & Akey, J. M. (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science, 343*(6174), 1017-1021.

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., et al. (2013) Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLoS ONE, 8*(2), e55950. doi:10.1371/journal.pone.0055950

Vizcaíno, S. F., Cassini, G. H., Toledo, N., & Bargo, M. S. (2012). On the evolution of large size in mammalian herbivores of Cenozoic faunes of southern South America. In *Bones, clone and biomes: an 80-million year history of recent Neotropical mammals*, Patterson, B. & Costa, L. eds. (University of Chicago Press: Chicago), pp. 76–101.

von Hippel, P. H., Johnson, N. P., & Marcus, A. H. (2013). Fifty years of DNA "Breathing": Reflections on old and new approaches. *Biopolymers; 99*(12), 923-954.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature, 171,* 737-738.

Wake, D. B. (1991). Homoplasy: The Result of Natural Selection, or Evidence of Design Limitations?". *The American Naturalist, 138*(3), 543–567. doi:10.1086/285234

Welker, F., Collins, M. J., Thomas, J. A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S. T., Reguero, M., Gelfo, J. N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D. A., Ashton, P. D., Rowsell, K., Porter, D. M., Kessler, B., Fischer, R., Baessmann, C., Kaspar, S., Olsen, J. V., Kiley, P., Elliott, J. A., Kelstrup, C. D., Mullin, V., Hofreiter, M., Willerslev, E., Hublin, J. J., Orlando, L., Barnes, I., & MacPhee, R. D. E. (2015). Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature, 522*(7554), 81–84. doi:10.1038/nature14249.

Westbury, M., Baleka, S., Barlow, A., Hartmann, S., Paijmans, J. L. A., Kramarz, A., Forasiepi, A. M., Bond, M., Gelfo, J. N., Reguero, M. A., López-Mendoza, P., Taglioretti, M., Scaglia, F., Rinderknecht, A., Jones, W., Mena, F., Billet, G., de Muizon, C., Aguilar, J. L., MacPhee, R. D. E. & Hofreiter, M. (2017). A mitogenomic timetree for Darwin's enigmàtic South American mammal *Macrauchenia patachonica*. *Nature Communications, 8*, 15951.
DOI: 10.1038/ncomms15951

Wetmore, A. (1926). Descriptions of additional fossil birds from the Miocene of Nebraska. *American Museum Novitates, 211*.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution, 19*, 639-644.

Wilhite, L. P. & Toliver, J. R. (1990). *Taxodium distichum* (L.) Rich. Baldcypress. In R. M. Burns and B. H. Honkala [eds.], *Silvics of North America, 1. Conifers*. U.S. Department of Agriculture, Forest Service, Washington, D.C. USA.

Williams, M. I., & Steadman, D. W. (2001). The historic and prehistoric distribution of parrots (Psittacidae) in the West Indies. In Woods, C. A.; Florence E. Sergile (eds.). *Biogeography of the West Indies: Patterns and Perspectives* (2nd ed.). Florida: CRC Press. pp. 175–189.

Wilson, D. E., & Reeder, D. M. (Eds.), (2005). *Mammal Species of the World: A Taxonomic and Geographic Reference*. Third ed., Vol. 2. John Hopkins University Press, Baltimore, pp. 745–2142.

Woods, C .A. (1989). The biogeography of West Indian rodents. In: Woods, C.A. (Ed.), *Biogeography of the West Indies: Past, Present, and Future*. Sandhill Crane Press, Gainesville, FL, pp. 741–798.

Woods, R., Marr, M. M., Brace, S., & Barnes, I. (2017). The small and the dead: a review of ancient DNA studies analysing micromammal species. *Genes, 8*, 312. doi:10.3390/genes8110312

Woodward, S. R., Weyand, N. J., & Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science, 265*,1229. doi: 10.1126/science.7973705

Wright, T. F., Schirtzinger, E. E., Matsumoto, T., Eberhard, J. R., Graves, G. R., Sanchez, J. J., Capelli, S., Müller, H., Scharpegge, J., Chambers, G. K., & Fleischer, R. C. (2008). A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the cretaceous. *Mol. Biol. Evol., 25*, 2141–2156.

Wutke, S., Benecke, N., Sandoval-Castellanos, E., Döhle, H. J., Friederich, S., Gonzalez, J., Hallsson, J. H., Hofreiter, M., Lõugas, L., Magnell, O., Morales-Muniz, A., Orlando, L., Pálsdóttir, A. H., Reissmann, M., Ruttkay, M., Trinks, A., & Ludwig, A. (2016). Spotted phenotypes in horses lost

attractiveness in the Middle Ages. *Scientific reports, 6*(1), 38548-38548.

Yamashita, C., & Valle, M. D. P. (1993). On the linkage between *Anodorhynchus macaws* and palm nuts, and the extinction of the Glaucous Macaw. *Bull. Brit. Ornithol. Club, 113,* 53–59.

Zanette, L. Y., & Clinchy, M. (2019). Ecology of fear. *Current Biology, 29*(9): R309–R313. doi:10.1016/j.cub.2019.02.042.

Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell, 147*(5): 1171-1185.

Zhan, S. & Reppert, S. M. (2013). MonarchBase: the monarch butterfly genome database. *Nucleic Acid Research, 41*: D758-D763.

Zhang, L., Mazo-Vargas, A. & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proc. Natl. Acad. Sci., 114*, 10707–10712. doi/10.1073/pnas.1709058114

Zhang, L. & Reed, R. D. (2016). Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. *Nat. Commun., 7*, 11769.