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

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

Published in:
Current Biology

DOI:
10.1016/j.cub.2019.10.066

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
https://doi.org/10.1016/j.cub.2019.10.066
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.
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,6 Santi Maniosa,6 George Pacheco,5 Manuel Ferrando-Bernal,1 Guolin Shi,7 Fei Hao,7 Xiangning Chen,8 Bent Petersen,3,9 Remi-André Olsen,10 Arcadi Navarro,1,11,12 Yuan Deng,5 Love Dalén,13 Tomás Marques-Bonet,1,11,12,14 Guojie Zhang,4,15,16,17 Agostinho Antunes,18,19 M. Thomas P. Gilbert,3,20,* and Carles Lalueza-Fox1,21,*

1Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Dr. Aiguardes 88, 08003 Barcelona, Spain
2Department of Evolutionary Anthropology, University of Vienna, Althanstraße 14, 1090 Vienna, Austria
3Section for Evolutionary Genomics, The GLOBE Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark
4Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
5BGI-Shenzhen, Beishan Industrial Zone, Building 11, Shenzhen 518083, China
6Departamento 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
7Center of Special Environmental Biomechanics & Biomedical Engineering, School of Life Sciences, Northwestern Polytechnical University, 127 West Youyi Road, Xi’an 710072, China
8Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, 127 West Youyi Road, Xi’an 710072, China
9Centre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), Faculty of Applied Sciences, AIMST University, Semeling Road, 08100 Kedah, Malaysia
10ScieLifeLab, Department of Biochemistry and Biophysics, Stockholm University, Frescativägen 40, SE-17121 Solna, Sweden
11Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
12CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Baldirí i Reixac 4, 08036 Barcelona, Spain
13Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Frescativägen 40, 10405 Stockholm, Sweden
14Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, c. de les Columnes s/n, 08193 Cerdanyola del Vallès, Barcelona, Spain
15China National GeneBank, BGI-Shenzhen, Jinsha Road, Shenzhen 518120, China
16State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China
17Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 32 Jiaochang Donglu, Kunming 650223, China
18CIIMAR/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
19Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal
20NTNU University Museum, Erling Skakkes gate 47c, 7012 Trondheim, Norway
21Lead 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 with 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
predicted to be deleterious mutations using SIFT software [14].

2.309, the ratio of the non-synonymous to synonymous

Conuropsis missense and 152 nonsense mutations were identified between

using the chicken

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 Psittacoidea 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 the sister genera

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.
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 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 cockleburs (Xanthium strumarium). This is unusual, as cockleburs 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 cockleburs [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 both Picttactiniformes. Samples are colored by IUCN Conservation status. See also Figure S4.
amino acid changes with respect to the Aratinga SLC25A4 and pipeline used. We did, however, find that the not present in that dataset, presumably due to the annotation [28]. First, we found that representing 363 species spanning nearly all avian families avian genomes, including one recently generated dataset comparing them against their orthologs in other available 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 Arboetum Masjoan (Espinelves, Barcelona) for allowing us to sample the Carolina Parakeet of their collection, Antoní Ballester for the radiographic exploration of the specimen, Jordi Grifois and Josep Maria Ramí for the Aratinga sample, and Wen Wang (Kunming Institute of Zoology, Yunnan, China) for his help and advice. The authors acknowledge the Science and Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Technology (FCT) and the European Regional Development Fund—program PTDC/CTA-AMB/31774/2017.

AUTHOR CONTRIBUTIONS


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


### STAR+METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Conuropsis carolinensis</em> tissue samples</td>
<td>Arboretum Masjoan private collection at Espinelves (Girona, Spain)</td>
<td>N/A</td>
</tr>
<tr>
<td><em>Aratinga solstitialis</em></td>
<td>Blood sample from <em>Aratinga</em> breeder</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole genome data of one historical <em>Conuropsis carolinensis</em> sample</td>
<td>This paper</td>
<td>PRJEB33130</td>
</tr>
<tr>
<td><em>Aratinga solstitialis</em> assembly</td>
<td>This paper</td>
<td>PRJEB33135</td>
</tr>
<tr>
<td><em>Aratinga solstitialis</em> annotation data</td>
<td>This paper</td>
<td>PRJEB33153</td>
</tr>
<tr>
<td><em>Taeiopygia guttata</em> 3.2.4 annotated proteins from Ensembl</td>
<td>N/A</td>
<td>Release 85</td>
</tr>
<tr>
<td><em>Gallus gallus</em> 5.0 annotated proteins from Ensembl</td>
<td>N/A</td>
<td>Release 85</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Acanthisita chloris</em></td>
<td>[10]</td>
<td>NCBI Project ID: PRJNA212877</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Agapornis roseicollis</em></td>
<td>[31]</td>
<td>NCBI Project ID: PRJNA355979</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Amazona aestiva</em></td>
<td>[32]</td>
<td>NCBI Project ID: PRJNA294082</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Amazona collaria</em></td>
<td>N/A</td>
<td>NCBI Project ID: PRJNA490036</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Amazona guildingii</em></td>
<td>[28]</td>
<td>Ultraconserved Elements (UCEs) extracted from whole genome data</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Amazona vittata</em></td>
<td>[33]</td>
<td>NCBI Project ID: PRJNA171587</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Ara macao</em></td>
<td>[34]</td>
<td>NCBI Project ID: PRJNA175470</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Eolophus roseicapilla</em></td>
<td>[28]</td>
<td>Ultraconserved Elements (UCEs) extracted from whole genome data</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Meiopittacus undulatus</em></td>
<td>N/A</td>
<td>NCBI Project ID: PRJNA72527</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Nestor notabilis</em></td>
<td>[10]</td>
<td>NCBI Project ID: PRJNA212900</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Probosciger aterrimus</em></td>
<td>[28]</td>
<td>Ultraconserved Elements (UCEs) extracted from whole genome data</td>
</tr>
<tr>
<td>Whole-genome sequencing data of <em>Psittacula krameri</em></td>
<td>N/A</td>
<td>NCBI Project ID: PRJNA377329</td>
</tr>
<tr>
<td>Targeted capture ultraconserved elements for <em>Conuropsis carolinensis</em></td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Targeted capture ultraconserved elements for <em>Aratinga solstitialis</em></td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Targeted capture ultraconserved elements for <em>Amazona autumnalis</em>, <em>Cacatua sulphurea</em>, <em>Psittacula alexandri</em>, <em>Psittacus erithacus</em>, <em>Strigops habroptila</em></td>
<td>[21]</td>
<td><a href="https://doi.org/10.5061/dryad.2vd01grFile">https://doi.org/10.5061/dryad.2vd01grFile</a>: passerines.unaligned.uce.contigs.tar.gz</td>
</tr>
<tr>
<td>4,988 aligned UCE loci</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/p4w17jc9">https://doi.org/10.17632/p4w17jc9</a> dw.1</td>
</tr>
<tr>
<td>4,988 gene trees from UCE loci and summary tree from ASTRAL-III</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/p4w17jc9">https://doi.org/10.17632/p4w17jc9</a> dw.1</td>
</tr>
<tr>
<td>Concatenated dataset of 4,988 aligned UCE loci and resulting IQTREE tree</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/p4w17jc9">https://doi.org/10.17632/p4w17jc9</a> dw.1</td>
</tr>
<tr>
<td>Concatenated dataset of 2,755 aligned UCE loci which are present in 95% of taxa and resulting IQTREE tree</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/p4w17jc9">https://doi.org/10.17632/p4w17jc9</a> dw.1</td>
</tr>
<tr>
<td>Concatenated dataset of 893 aligned UCE loci which are present in 100% of taxa and resulting IQTREE tree</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/p4w17jc9">https://doi.org/10.17632/p4w17jc9</a> dw.1</td>
</tr>
</tbody>
</table>

(Continued on next page)
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 Maria` 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 toe pads, 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 CaCl₂, 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 5 min 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 730mL of QIAGEN buffer PE, centrifuged at 3,300 × g, flow-through 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 H₂O, 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 = 83bp 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.

**Conurops 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 Tetrapsods-UCE-5Kv1 bait set [72] was applied to 18 parrots and the outgroup *Acanthisitta chloirs*. 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.

We subsequently searched for the *DMRT1* gene [71] to confirm the *Aratinga* Z chromosome scaffold.
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.