









# Clinging on the brink: Whole genomes reveal human-induced population declines and severe inbreeding in the Critically Endangered Emirati Leaf-toed Gecko (*Asaccus caudivolvulus*)

Bernat Burriel-Carranza<sup>1,2</sup>  | Gabriel Mochales-Riaño<sup>1</sup>  | Adrián Talavera<sup>1</sup>  | Johannes Els<sup>3</sup>  | Maria Estarellas<sup>1</sup>  | Saleh Al Saadi<sup>4</sup> | Juan Diego Urriago Suarez<sup>5,6</sup>  | Per Olof Olsson<sup>7</sup> | Michael Matschiner<sup>8</sup>  | Salvador Carranza<sup>1</sup> 

<sup>1</sup>Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

<sup>2</sup>Museu de Ciències Naturals de Barcelona, Barcelona, Spain

<sup>3</sup>Breeding Centre for Endangered Arabian Wildlife, Environment and Protected Areas Authority, Sharjah, United Arab Emirates

<sup>4</sup>Environment Authority, Muscat, Oman

<sup>5</sup>Fujairah Research Centre, Fujairah, United Arab Emirates

<sup>6</sup>Foresea, Dubai, United Arab Emirates

<sup>7</sup>Fujairah Genetics Center, Fujairah, United Arab Emirates

<sup>8</sup>Natural History Museum, University of Oslo, Oslo, Norway

## Correspondence

Bernat Burriel-Carranza, Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta, 08003 Barcelona, Spain.  
Email: [bernat.burriel@ibe.upf-csic.es](mailto:bernat.burriel@ibe.upf-csic.es)

## Funding information

Mohamed Bin Zayed Species Conservation Fund, Grant/Award Number: 212527924; MCIN/AEI/10.13039/501100011033 and by ERDF, a way of making Europe, Grant/Award Number: PID2021-128901NB-I00; Field Study for the Conservation of Reptiles in Oman, Ministry of Environment and Climate Affairs, Oman, Grant/Award Number: 22412027; Departament de Recerca i Universitats de la Generalitat de Catalunya, Grant/Award Number: 2021 SGR 00751; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: FPU18/04742, PRE2019-088729 and PRE2022-101473; "la Caixa" doctoral fellowship programme, Grant/Award Number: LCF/BQ/DR20/11790007

**Handling Editor:** Paul A. Hohenlohe

## Abstract

Human-mediated habitat destruction has had a profound impact on increased species extinction rates and population declines worldwide. The coastal development in the United Arab Emirates (UAE) over the last two decades, serves as an example of how habitat transformation can alter the landscape of a country in just a few years. Here, we study the genomic implications of habitat transformation in the Critically Endangered Emirati Leaf-toed Gecko (*Asaccus caudivolvulus*), the only endemic vertebrate of the UAE. We generate a high-quality reference genome for this gecko, representing the first reference genome for the family Phyllodactylidae, and produce whole-genome resequencing data for 23 specimens from 10 different species of leaf-toed geckos. Our results show that *A. caudivolvulus* has consistently lower genetic diversity than any other Arabian species of *Asaccus*, suggesting a history of ancient population declines. However, high levels of recent inbreeding are recorded among populations in heavily developed areas, with a more than 50% increase in long runs of homozygosity within a 9-year period. Moreover, results suggest that this species does not effectively purge deleterious mutations, hence making it more vulnerable to future stochastic threats. Overall, results show that *A. caudivolvulus* is in urgent need of protection, and habitat preservation must be warranted to ensure the species' survival.

## KEYWORDS

conservation genomics, genomics, population genetics – empirical, reptiles

## 1 | INTRODUCTION

Over the last century, unprecedented extinction rates caused by human activities have been reported across many populations and species from a wide spectrum of taxa (Ceballos et al., 2020; Pimm et al., 2014) with habitat loss and fragmentation often being the initial cause of population declines with subsequent genetic impoverishment (Allendorf et al., 2010; Frankham, 2005; Hung et al., 2014). Endangered species, usually confined to small and isolated populations, are especially vulnerable to population declines. In these cases, genetic drift, high levels of inbreeding and loss of genetic variation, may facilitate the expression of harmful recessive alleles, have a negative impact on their fitness and thereby increase their risk of extinction (Kardos et al., 2018; Kleinman-Ruiz et al., 2022).

With the advent of next-generation sequencing and the increasing availability of reference genomes, the field of conservation genomics has greatly benefited from new ways to evaluate and assess genomic impoverishment in wild populations (Escoda & Castresana, 2021; Irizarry et al., 2016; Kardos et al., 2018). The depth of information retrieved from whole genome sequencing (WGS) allows for the effective detection of a wide spectrum of genomic characteristics, including inbreeding and the increase of genetic load, among others, even when only a handful of samples are available (Gaughran et al., 2018; Iannucci et al., 2021; Mochales-Riaño et al., 2023). This is of particular importance when dealing with endangered species, as these are usually elusive and reduced in numbers (McMahon et al., 2014; Morin et al., 2021).

Despite the exponential growth in the availability of reference genome data in recent years, squamate reptiles fall behind other taxonomic groups in terms of reference genome availability. To date, reference genomes are available for less than 1% of squamate reptiles whereas this is the case for 9% and 6% of mammals and birds respectively (Card et al., 2023). Within squamates, geckos are particularly underrepresented with only 10 available reference genomes (Card et al., 2023) for a group of 2291 species (Uetz et al., 2023). The lack of genomic resources for geckos can have profound implications in conservation and management strategies for endangered species within this group.

The Emirati Leaf-toed Gecko (*Asaccus caudivolvulus*) is a medium-sized gecko from the family Phyllodactylidae, thought to be extinct in the wild. Historically spread through the Western Hajar Mountains of Oman and the United Arab Emirates (UAE; Arnold & Gardner, 1994), a taxonomic review based on molecular data restricted *A. caudivolvulus* to two isolated localities in the UAE East Coast, thus making this species the only endemic vertebrate of the UAE (Carranza et al., 2016). However, the persistence of this species was already dubious when Carranza et al. (2016) redescribed it, as the rapid development of the UAE's East Coast had already affected both localities. The type locality (Figure 1: white star in locality 1) has been inaccessible since 2006, being affected by heavy development that destroyed most of the natural coastal habitat of *A. caudivolvulus*. The second one was also found in a critical condition (Figure 1: locality 5) since the small outcrop on which it was situated was ready to be removed to construct a resort as part of the ongoing

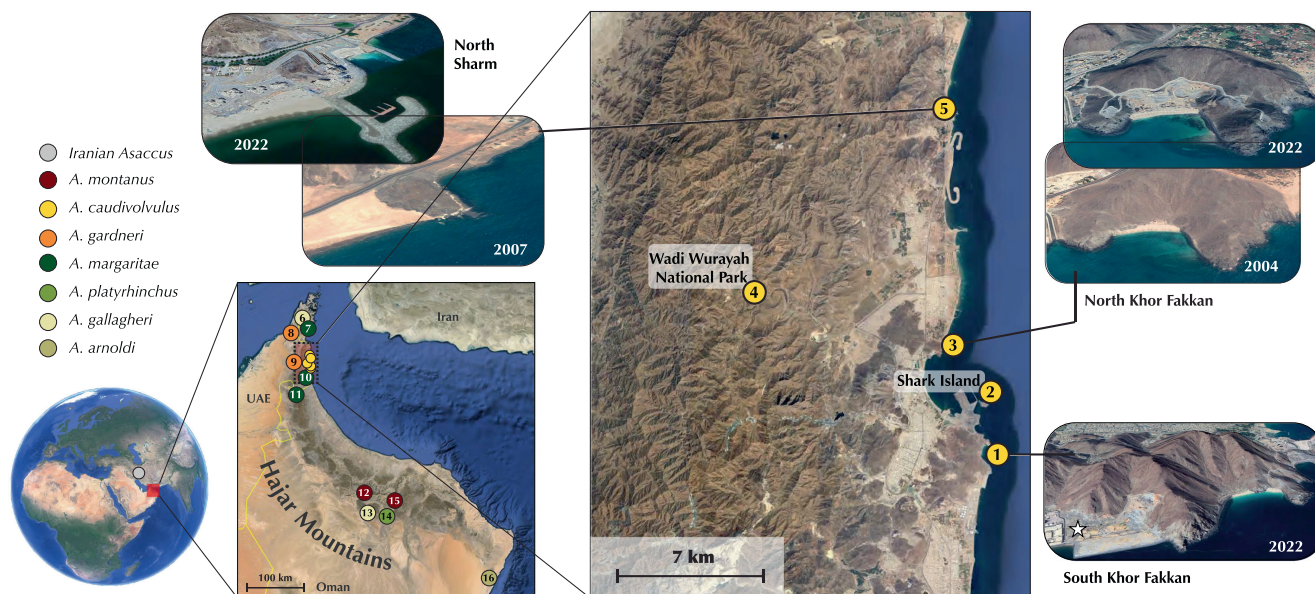
coastal development, and was inaccessible since 2015 (Carranza et al., 2016). This critical situation granted *A. caudivolvulus* the category of Critically Endangered in the IUCN red list of threatened species in 2018 (Carranza & Els, 2021). A sighting of this species was reported in late 2018 in Wadi Wurayah National Park (Farkas et al., 2018). However, the identity of the reported specimen had not been genetically confirmed and in the available photograph, upper-arm tubercles were not clearly distinguishable (i.e. a diagnostic character to identify *A. caudivolvulus* from its sister species *A. gardneri*).

Here, we evaluate the genomic status of *A. caudivolvulus* by performing a state-of-the-art conservation genomics assessment of this Critically Endangered species. We assemble and annotate the reference genome of *A. caudivolvulus* and generate WGS data for nine specimens from all known localities, as well as for 11 individuals representing the other six Arabian *Asaccus* and three Iranian *Asaccus* species. With the genomic data generated herein, we explore the systematics of all Arabian *Asaccus* species to provide a comprehensive understanding on the evolution of this genus, placing special attention to *A. caudivolvulus* and its intraspecific variation. We then focus on evaluating the population structure, demographic history, and genetic diversity of *A. caudivolvulus* and other relatives, to evaluate its current genomic status and the impacts of habitat loss due to the current coastal development in the region.

## 2 | MATERIALS AND METHODS

### 2.1 | *A. caudivolvulus* population discovery

In 2022, in collaboration with the local authorities from Sharjah and Fujairah in the UAE, we set up an expedition to explore whether *A. caudivolvulus* could still be found in the wild. After exhaustive sampling throughout the UAE East Coast, five extant populations were confirmed (Figure 1): Locality 1: South of Khor Fakkan, *A. caudivolvulus* was found north of its type locality in an almost inaccessible, non-developed coastal habitat, enclosed by two outcrops reaching the sea; Locality 2: Few specimens were found on Shark Island, a small islet of approximately 450m length, 400m off the mainland; Locality 3: North of Khor Fakkan, in a promontory currently under development, *A. caudivolvulus* was found in relatively high numbers; Locality 4: We confirmed the presence of *A. caudivolvulus* in Wadi Wurayah National Park; Locality 5: We were able to access the heavily-developed area north of Sharms where *A. caudivolvulus* was discovered by Carranza et al. (2016). Although we confirmed its presence at this location, its habitat had been extremely reduced to a small patch of rocks. It is notable that *A. caudivolvulus* was not found in the mainland-facing slopes of the promontories in which the species was discovered (e.g. locality 1), or in the adjacent mountains further inland (which were separated by roads in localities 3 and 5), thus suggesting their strong relation with highly humid habitats (Carranza et al., 2016). From each of the five localities we sampled between 1 and 3 specimens (see Table 1) for subsequent molecular analyses.



**FIGURE 1** Map of the study area showing the location of the Hajar Mountains in South-eastern Arabia in the first inset map, and the specific geographical region where *A. caudivolvulus* resides in the second inset map. Numbers show the 19 localities from where all samples used in this study were taken. The grey circle in the global map shows the approximate locality of the three species of Iranian *Asaccus* used for this study (see Table 1 for specific coordinates for each locality). Inset pictures for localities 3 and 5 show the development and habitat transformation that these two localities underwent during the last 18 and 15 years respectively. The white star in the picture from locality 1 marks the type locality for the species, inaccessible since 2006. Map: Google Earth 2023 (Image: Landast/Copernicus). Temporal series were done by using the historical imagery tool in Google Earth 2023.

## 2.2 | DNA extraction, library preparation and sequencing

We extracted high molecular weight (HMW) genomic DNA from tail tips of 23 individuals including nine *Asaccus caudivolvulus*, 11 representatives from all other Arabian *Asaccus* species and lineages (Arnold & Gardner, 1994; Burriel-Carranza et al., 2023; Carranza et al., 2016; Simó-Riudalbas et al., 2018; Tamar et al., 2019) and three Iranian *Asaccus* species (Table 1). We followed the manufacturer's protocol of the MagAttract HMW Kit (Qiagen) and sequenced a total of 22 Illumina libraries aiming for a coverage of ~20× (sample S7866 was exceptionally sequenced at ~60× coverage for polishing the reference genome, see below) on a 2×150 NovaSeq6000 lane (Novogene Co). HMW DNA for the reference genome assembly was obtained from muscle tissue of a female *A. caudivolvulus* specimen (Sample code S7866; Table 1), following the same protocol as above. Long read data were obtained for this specimen through sequencing 61.1 Gbp of Oxford Nanopore Technologies (ONT) data. For chromosome-level scaffolding we used liver and muscle tissue samples from another female specimen to prepare a Hi-C library using the Omni-C kit (Dovetail Genomics) following the manufacturer's protocol. This Library was sequenced aiming for a ~60× coverage at 2×150 NovaSeq6000 (CNAG). Full-length transcriptome libraries were obtained by pooling six different tissues (liver, heart, lungs, eyes, kidneys and brain) of the latter specimen and sequencing was performed on a 1/6th of 8M SMRT PacBio HiFi cell using the Sequel II system. Sequencing was conducted by Novogene Co.

## 2.3 | Genome assembly and scaffolding

The reference genome was assembled with NextDenovo assembler (<https://github.com/Nextomics/NextDenovo>). We ran the assembler four different times, discarding reads that were smaller than 1, 3, 5 and 8 kbp respectively. Then we selected the best assembly according to QUAST v.5.1.0 (Gurevich et al., 2013) metrics (contig number, N50 and L50 scores) and proceeded to polish the consensus assembly (Table S1). Since Nanopore reads contain systematic error in homopolymeric regions, we applied three rounds of polishing with nanopore reads as input to the Racon (Vaser et al., 2017) and Medaka (<https://github.com/nanoporetech/medaka>) programs (i.e. two and one round respectively). Then, we applied two rounds of polishing with the Illumina reads as input to the Pilon (Walker et al., 2014) software. Purge dups (Guan et al., 2020) was used to remove haplotypic duplicates from the assembly. Then we scaffolded the draft assembly using Omni-C data with SALSA2 (Ghurye et al., 2019). Manual curation was performed with Pretext (<https://github.com/wtsi-hpag/PretextView>) resulting in a total of 48 scaffolds, of which the first 20 comprised 99.8% of the data. Then, we annotated repetitive elements in the genome with Red (Girgis, 2015). Contamination was assessed with BlobToolKit (Challis et al., 2020) using the NCBI taxdump database, identifying and discarding nine of the smallest scaffolds as putative contaminations. Finally, we identified the mitochondrial genome within our assembly with MitoFinder (Allio et al., 2020; Li et al., 2016), using the mitochondrial genome of *Phyllodactylus unctus* (Yan et al., 2014) as a reference. The mitochondrial genome of *A. caudivolvulus* was identified within

one of the smallest scaffolds (33,387bp). We removed that scaffold and assembled de novo the mitochondrial genome with the 60x Illumina reads. Quality assessment and general metrics for the final assembly were carried out with gfastats (Formenti et al., 2022) while completeness of the genome assembly was assessed with BUSCO v5.3.0 (Simão et al., 2015) against the vertebrata\_odb10 database ( $n=3354$ ). The final assembly contained 20 chromosome level scaffolds (11 macro-chromosomes and nine micro-chromosomes), as well as 17 unplaced scaffolds.

## 2.4 | Genome annotation

Genome annotation was performed based on the integration of three independent annotation approaches: (i) *Annotation based on de novo sequenced full-length transcriptome data (PacBio Iso-Seq)*: Subreads from the PacBio SMRTcell were first converted into circular consensus sequences with the ccs function in pbccs v.6.4 (Pacific Biosciences of California, Inc.), followed by demultiplexing and primer removal with lima v.2.7.1 (Pacific Biosciences of

California, Inc.). Non-chimaeric full-length reads were then obtained by removing concatemers with isoseq3 v.3.8.3 (Pacific Biosciences of California, Inc.) and trimming polyA tails with the python script `tama_flnc_polya_cleanup.py` (<https://github.com/GenomeRIK/tama/wiki>). Resulting reads were mapped to the reference genome with the splice-aware mapper pbmm2 v.1.10 (Pacific Biosciences of California, Inc.). The resulting mapped reads were collapsed into unique isoforms with isoseq3 v.3.8.3 with the `-do-not-collapse-extra-5exons` flag. Collapsed unique isoforms were then used to identify protein-coding regions with GeneMarkS-T (Tang et al., 2015); (ii) *Annotation based on protein evidence from public databases*: We merged the published protein data available for the 'Vertebrata' clade in OrthoDB v.11 ([www.orthodb.org](http://www.orthodb.org)) as well as for 'Sauropsida' in OrthoDB v.10 (<https://v10.orthodb.org/>) as input for the BRAKER2 pipeline (Brůna et al., 2021), using these databases to self-train GeneMark-EP+ and Augustus to make gene predictions (Brůna et al., 2021); (iii) *Annotation based on other annotated reference genomes*: We downloaded annotation files for *Sphaerodactylus townsendi* (Pinto et al., 2022; NCBI BioSample ID: SAMN20179316), *Gekko japonicus* (NCBI BioSample ID: SAMN04157958) and *Pogona*

Species	Sample code	Loc	Lat	Lon	Cov	Heterozygosity (SNPs/Kbp)
<i>A. caudivolvulus</i>	CN21121	1	25.334	56.377	16.2	0.805
<i>A. caudivolvulus</i>	CN21123	1	25.336	56.378	14.7	0.844
<i>A. caudivolvulus</i>	CN21125	1	25.329	56.377	20.7	0.864
<i>A. caudivolvulus</i>	CN21128	2	25.354	56.376	16.0	0.125
<i>A. caudivolvulus</i>	CN21141	3	25.374	56.353	17.5	0.485
<i>A. caudivolvulus</i>	CN21148	3	25.374	56.356	18.9	0.670
<i>A. caudivolvulus</i>	CN21161	4	25.397	56.270	17.1	0.836
<i>A. caudivolvulus</i>	<b>S7866</b>	<b>5</b>	<b>25.478</b>	<b>56.362</b>	<b>56.0</b>	<b>0.649</b>
<i>A. caudivolvulus</i>	CN21156	5	25.479	56.364	15.9	0.314
<i>A. gallagheri</i>	CN8387	6	25.979	56.205	20.6	5.496
<i>A. margaritae</i>	CN7126	7	25.965	56.203	15.4	4.323
<i>A. gardneri</i>	CN21164	8	25.857	56.098	15.8	4.128
<i>A. gardneri</i>	CN21154	9	25.444	56.194	15.8	4.218
<i>A. margaritae</i>	CAS250892	10	25.265	56.307	14.5	2.149
<i>A. margaritae</i>	CN749	11	25.008	56.215	14.7	2.645
<i>A. montanus</i>	CN4045	12	23.295	57.132	10.9	2.804
<i>A. gallagheri</i>	CN2593	13	23.192	57.199	13.4	4.279
<i>A. platyrhynchus</i>	ZFMK84264	14	23.056	57.469	17.9	1.386
<i>A. montanus</i>	CN196	15	23.081	57.670	11.8	2.415
<i>A. arnoldi</i>	CN10790	16	22.107	59.357	13.5	3.713
<i>A. elisae</i>	MVZ234315	17	32.849	48.264	14.7	2.463
<i>A. griseonotus</i>	MVZ234325	18	33.260	47.804	13.7	2.352
<i>A. nasrullahi</i>	MVZ234330	19	32.849	48.264	18.4	0.665

**TABLE 1** Information on all samples used for this study including locality code, coordinates in decimal degrees (WGS84), coverage, and genome-wide heterozygosity in number of heterozygous sites by kbp.

*Note:* The specimen used for the reference genome assembly is marked in bold. Locality column corresponds to the localities shown in Figure 1. Loc: Locality number; Lat: Latitude; Lon: Longitude; Cov: Coverage.



*vitticeps* (NCBI BioSample ID: SAMEA2300447), and used the homology-based gene prediction program GeMoMa v1.9 (Keilwagen et al., 2016, 2018) to predict homologous genes on the reference genome of *A. caudivolvulus*. Finally, we integrated the three annotation approaches into a single file with TSEBRA (Gabriel et al., 2021), keeping all predictions from GeMoMa and GeneMarkS-T. BUSCO scores as well as general metrics (i.e. number of genes, transcripts, or number of isoforms among others) were obtained to evaluate the structural annotation of *A. caudivolvulus*.

## 2.5 | Processing of whole genome sequencing

Raw Illumina reads for each of the 23 samples were first processed with fastp (Chen et al., 2018) applying a sequence quality filtering of 30, trimming poly-G and poly-X tails and performing base correction and adapter trimming. Filtered Illumina reads were then mapped to the de novo assembled reference genome with bwa v.0.7.17-r1188 (Li, 2013), PCR duplicates were marked and removed with PicardTools v.3.0.0 (<https://broadinstitute.github.io/picard/>) and the final alignment files in BAM format were indexed with SAMtools v.1.10 (Li et al., 2009). The per-individual read coverage distribution was determined with BEDTools v.2.27.1 (Quinlan & Hall, 2010) and variant calling was performed independently for each scaffold with GATK's 'HaplotypeCaller', 'GenomicsDBImport' and 'GenotypeGVCFs' tools (McKenna et al., 2010), applying a minimum base quality score of 30. Reads of specimen S7866 were downsampled to a ~20× coverage before the variant calling so that all samples were comparable.

A strict filtering pipeline was then applied to variant calls to ensure high call reliability. Calls were discarded if they had a Phred-scaled *p*-value of Fisher's exact test for strand bias above 20, a quality score normalized by read depth below two, a root mean square of the mapping quality below 40, or if the overall read depth across the 23 samples was either below five or above 550. We also excluded sites with Mann–Whitney–Wilcoxon rank sum test statistic below −0.5 for either mapping quality bias between reference and alternative alleles as well as for site position bias within reads. Indels were normalised and SNPs within 10bp of an indel were also excluded. We further excluded repetitive regions inferred with Red (Girgis, 2015) (see above) and sites within regions of the *Asaccus caudivolvulus* reference genome in which mapping was likely to be ambiguous. These regions were determined using the SNPable pipeline (<http://lh3lh3.users.sourceforge.net/snpable.shtml>), in which the reference genome is split into overlapping fragments of 100bp (overlapping by 99bp), followed by mapping those fragments back to the reference, and estimating the number of fragments mapping correctly at each site. We then excluded those regions in which less than 95 of the 100 fragments mapped back correctly. Finally, we masked individual genotypes with a read depth below five, removed all sites that were no longer polymorphic, and concatenated the filtered scaffolds into a single file in VCF format with VCFtools (Danecek et al., 2011), resulting in a final dataset of 102,621,032 SNPs.

## 2.6 | Phylogenomic reconstructions

To investigate the phylogenomic relationships within the genus *Asaccus*, we performed phylogenomic reconstructions based on genome-wide nuclear SNPs and assembled mitochondrial genome sequences. To reconstruct a maximum-likelihood (ML) phylogeny with nuclear SNPs, we further filtered the 102,621,032 SNP dataset with VCFtools, excluding SNPs with more than 20% missing data and selecting one allele at random at each site. The resulting final dataset contained 96,296,559 SNPs. We split the dataset in non-overlapping windows with a length of 100kbp, and ran IQTREE2 (Nguyen et al., 2015) on each of these independently, applying a GTR+ASC model (accounting for ascertainment bias), 100 starting parsimony tree searches and 1000 bootstrap replicates. Bootstrap convergence was ensured post-hoc by using the *bsconverge* function implemented in raxml-ng v.1.2.0 (Kozlov et al., 2019). The best trees of all analyses were combined and a maximum clade credibility tree across all windows was obtained with TreeAnnotator implemented in BEAST2 (Bouckaert et al., 2019).

One representative from each species (including two highly divergent lineages within *A. montanus* species; Tamar et al., 2019) was then selected to generate a time-calibrated species tree under the multi-species coalescent (MSC) with SNAPP (Bryant et al., 2012), using the program's implementation in BEAST2 v.2.6.4 (Bouckaert et al., 2019) and the SNP-based molecular clock of Stange et al. (2018). We again filtered the initial SNPs dataset only keeping one biallelic SNP and not allowing more than 10% missing data for every window of 10kbp, resulting in a final dataset of 159,370 unlinked SNPs (uSNPs). We then generated the input file of SNAPP with the 'snapp\_prep.rb' script (<https://github.com/mmatschiner/tutorials>), constraining the deepest node of the phylogeny (the divergence between *A. montanus* and all other *Asaccus*) with a normally distributed prior centred at 31.93 million years ago (Ma) and a standard deviation of 4 according to the mean and 95% highest posterior density (HPD) interval estimated in Burriel-Carranza et al. (2023) respectively. We then selected 50,000 uSNPs at random and ran three replicate chains of Markov-chain Monte Carlo, each with a length of  $5 \times 10^6$  generations, sampling every 50 generations. Convergence and stationarity of these chains was checked with Tracer V1.7 (Rambaut et al., 2018). Posterior distributions were combined with LogCombiner v.2.6.4, discarding 10% of the posterior trees as burnin, and a maximum clade credibility tree was obtained with node ages set to median heights in TreeAnnotator v.2.6.4 (both programs are part of the BEAST2 package; Bouckaert et al., 2019). Additionally, to infer the phylogenomic relationships of populations within *A. caudivolvulus*, we reconstructed a time-calibrated species tree under the MSC only with *A. caudivolvulus* samples and *A. gardnerii* as outgroup. We generated a new dataset of uSNPs with the same filter settings as above, resulting in 163,866 uSNPs. We then constrained the deepest node (the split between *A. caudivolvulus* and *A. gardnerii*) to a normal distribution centred at 5.08 Ma (with a standard deviation of 0.8) according to the results from the previously inferred species tree, and ran the analysis with the same specifications as above.

To test for discrepancies between nuclear and mitochondrial evolutionary histories, we extracted a set of 13 proteins coding

genes (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cytb*, *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5* and *ND6*) and two ribosomal genes (*rrnL* and *rrnS*) from the mitochondrial genomes identified with MitoFinder. We aligned sequences of each gene separately with MAFFT v. 7.505 (Katoh & Standley, 2013), and concatenated the resulting alignments. The final dataset had a length of 12,671bp. We then inferred the ML mitochondrial phylogeny with IQTREE2 (Nguyen et al., 2015) from the concatenated alignment, applying a GTR model, 100 starting parsimony tree searches, and 1000 bootstrap replicates. Bootstrap convergence was again ensured using the function *bsconverge* from *raxml-ng* v.1.2.0 (Kozlov et al., 2019).

## 2.7 | Population structure

We analysed the patterns of population structure within *A. caudivolvulus* by means of a principal component analysis (PCA) and through admixture analyses on a reduced dataset containing the *A. caudivolvulus* specimens only. We further filtered the dataset by excluding non-biallelic SNPs with more than 10% missing data, and applying a minor allele frequency of 0.05. We then assessed the decay of linkage disequilibrium with PopLDdecay (Zhang et al., 2019) and kept 1 SNP every 10,000 to obtain uSNPs, resulting in a final dataset of 147,419 uSNPs. The PCA was implemented in Plink v2.00a2.3 (Chang et al., 2015). We then applied ADMIXTURE v.1.3.0 (Alexander et al., 2009; Alexander & Lange, 2011) to detect population structure, assuming *K*, the number of genetic clusters, ranging from *K*=1 to *K*=5, with a total of 15 cross validation replicates for each tested *K*.

## 2.8 | Demographic history

The demographic history of all samples of *Asaccus caudivolvulus* as well as *A. gardneri* and *A. margaritae*, two closely related species to *A. caudivolvulus* presenting similar distribution, habitat requirements and morphologic characteristics, was inferred with the Pairwise Sequential Markovian Coalescent (PSMC) software. Heterozygous positions were inferred from BAM files with SAMtools v.1.10 (Li et al., 2009) discarding samples with low mapping quality (<30) or low base quality (<30). A generation time (*t*) of 3.5 years was assumed for all species (J.E., pers. obs.) and the mutation rate was assumed to match the median squamate mutation rate ( $1.75 \times 10^{-9}$  per site per year; Gemmell et al., 2020), resulting in  $6.125 \times 10^{-9}$  per site per generation. We performed 15 bootstrap replicates for each sample to assess variability in the  $N_e$  estimates using PSMC.

## 2.9 | Heterozygosity and ROH

We calculated genome-wide heterozygosity estimates for each of the 23 *Asaccus* specimens from a VCF file including invariant sites. We generated non-overlapping sliding windows of 100 kbp and

calculated the number of heterozygous sites among the number of callable sites per window (considering callable sites those not excluded by some of the following filters: proximity to indels, repetitive regions, mappability, overall read depth, and read depth per individual; see WGS data processing above). Only those windows with more than 60,000 callable sites were considered. Additionally, we calculated genome-wide heterozygosities per chromosome, calculating the number of heterozygous sites among callable sites (applying the same filtering as above) using the function 'stats' in BCFtools v.1.13 (Li, 2011).

We explored the number of genomic segments that are identical by descent (IBD) and thus result from inbreeding, by determining runs of homozygosity (ROH). A reduced dataset excluding *A. montanus* and all the Iranian samples was filtered discarding non-biallelic SNPs with more than 10% missing data, resulting in a dataset containing 50,064,602 SNPs. ROH were estimated based on density of heterozygous sites, using the Hidden Markov Model implemented in the *roh* function of BCFtools v.1.13 (Li, 2011). Once ROH were inferred, we kept only those with a minimum length of at least 100kbp and a quality Phred score above 70. Results were divided into short (0.1–0.5 Mbp), medium (0.5–1 Mbp) and long (>1 Mbp) ROH depending on sequence length. Then, we performed a correlation test between the number and cumulative length of ROH, and the frequency of long ROH ( $F_{\text{ROH}}$ ) within each sample's genome was calculated.

We further explored the IBD segments classified as IROH by estimating the number of generations (*g*) for each ROH, keeping only IBD segments arising from recent ancestors (i.e. the last 10 generations) and calculating the frequency of the remaining ROH for each sample ( $F_{\text{ROH}10}$ ). Based on the simplifying assumptions that ROH decay exponentially over time and that recombination rate is constant across the *Asaccus* genome, the age of each ROH in numbers of generations was calculated by solving *g* in the equation  $l = 100/2gcM$ , where *l* is the length of the ROH in centimorgans (cM; Thompson, 2013). Since genetic map length data is not well understood in geckos, we followed the approach implemented in Dodge et al. (2023) and used two extremes of reported squamate recombination rates to calculate genetic map distances (*m*) in cM/Mbp by solving the equation  $m = C/G \times 50cM$ , where *C* is the number of crossovers per meiosis and *G* is the haploid genome size in Mbp (Calderón & Pigozzi, 2006). The lower extreme was represented by *Anolis carolinensis* with an average of 23.3 crossovers per meiosis (Lisachov et al., 2017) and a genome size of 1.68Gb (Peterson et al., 1994), while the higher extreme was represented by *Trapelus sanguinolentus* with an average of 38.4 crossovers per meiosis (Lisachov et al., 2019) and a genome size of 1.68Gb (Vinogradov, 1998).

## 2.10 | Mutational load

We estimated the mutational load for coding regions in *Asaccus caudivolvulus* specimens using SNPeff v.4.3 (Cingolani et al., 2012). To avoid reference bias, we excluded all samples that were not equidistant in branch length to the reference genome assembled in the

present study. Moreover, since the reference genome presented in this study is the first available genome in the whole family, we were not able to implement this analysis with another reference. Hence, we were only able to compare the effects of mutational load on the *Asaccus caudivolvulus* specimens from localities 1 to 4. We produced a reduced dataset containing the seven *A. caudivolvulus* and not allowing missing data, which yielded a resulting dataset of 2,299,095 SNPs. We generated a custom database with the annotation files of the reference genome and identified putative deleterious alleles falling into one of the following default categories: (i) Low: variants found to be mostly harmless, unlikely to modify the protein's behaviour; (ii) Moderate: alleles possibly influencing the protein effectiveness but being non-disruptive; (iii) High: variants with most likely high or disruptive effects on the protein, probably causing protein loss of function; (iv) Modifier: Variants with no evidence of impact or difficult-to-predict impact. Then, we counted the number of derived alleles for each of the categories, removing observations with warnings produced by SNPeff. Since we used sites with no missing data, each variable could be independently counted without the need of bootstrapping (Dussex et al., 2021). Finally, we assessed the distribution of mutational load in each population by calculating the number of alleles in homozygosity (the realized load), the number of alleles in heterozygosity (masked load), and the sum of both (total load) for each type of lowly-, moderately- and highly-deleterious alleles (Bertorelle et al., 2022).

### 3 | RESULTS

#### 3.1 | Genome assembly and annotation

Using a combination of Nanopore long read sequencing at a ~40x mean genome coverage, Illumina short-reads sequencing at ~60x mean genome coverage and Omni-C reads sequenced at a ~60x mean genome coverage, we successfully assembled the first

chromosome-level genome for the Phyllodactylidae family and the most contiguous genome among all available gecko genomes (Table 2). The final assembly span was 1.73 Gb, comprising 37 scaffolds with a scaffold N50 of 114.54 Mbp (Table 2). We found that 99.8% of the assembly was comprised by the first 20 scaffolds, including 11 macro- and nine micro-scaffolds (Figure S1). This number of scaffolds matches the chromosome number of the Fan-footed gecko, *Ptyodactylus hasselquisti* (i.e. the closest species to the *Asaccus* genus with known karyotype data;  $2n=40$ ), thus suggesting a chromosome-level assembly (King, 1987). The Emirati Leaf-toed Gecko genome also had high BUSCO scores with a completeness of 96.7% of vertebrate orthologue groups, a GC content of 44.3% and 37.2% of the genome composed of repetitive elements (Figure S2).

Within the Phyllodactylidae family, two sex determination systems have been described so far: temperature-dependent sex determination (TSD) in the genus *Tarentola*, and female heterogamety (ZW) in species from the *Thecadactylus* and *Phyllodactylus* genera (Nielsen et al., 2019). Since the only known chromosome-dependent sex determination system in the family is female heterogamety, we opportunistically sequenced a female Emirati Leaf-toed Gecko. However, we were not able to confidently identify a sex-chromosome pair, with no clear signal of female hemizygosity. Further analyses of individual read depth in the re-sequenced specimens did not show any clear patterns of a pair of sexual chromosomes, which could be indicative of a TSD system in this species.

We successfully annotated the reference genome of *A. caudivolvulus* through the integration of three independent annotation approaches based on full-length transcriptome data, protein evidence from public databases and other annotated reference genomes. This allowed us to identify up to 39,360 genes and a total of 76,795 coding regions with a mean number of 7.3 exons per transcript. BUSCO evaluation on the sauropsida\_odb10 database indicated that we captured most of the core sauropsida genes ( $n=7480$ ), with 95.3% of the genes present and complete for *A. caudivolvulus*, providing

TABLE 2 Reference genome comparison of *Asaccus caudivolvulus* and other published geckotan reference genomes, including genome size, number, and N50 for scaffolds and contigs, as well as the technologies used for the reference genome assembly.

Family	Species	Gm size Gb	Sf n°	Sf N50 Mb	Cg n°	Cg N50 Mb	Technology	Study
Phyllodactylidae	<i>Asaccus caudivolvulus</i>	1.73	<b>37</b>	114.5	290	19.95	ONT + Illumina + Omni-C	Present study
Sphaerodactylidae	<i>Sphaerodactylus townsendii</i>	1.81	1742	133.8	4332	9.46	ONT + Illumina + Hi-C	Pinto et al., 2022
Gekkonidae	<i>Paroedura picta</i>	1.56	4871	109.0	118,732	0.02	Illumina	Hara et al. 2018
Gekkonidae	<i>Gekko japonicus</i>	2.49	191,500	0.71	335,470	0.03	Illumina	Liu et al., 2015
Eublepharidae	<i>Eublepharis macularius</i>	2.24	76	<b>145.6</b>	<b>137</b>	<b>80.1</b>	PacBio HiFi + Hi-C	Pinto et al., 2023
Gekkonidae	<i>Lepidodactylus listeri</i>	2.35	401	119.9	381	72.3	PacBio HiFi + Hi-C	Dodge et al., 2023

Note: In bold, the best result in each column.

Abbreviations: Cg, Contig; Gm, Genome; ONT, Oxford Nanopore Technologies; Sf, Scaffold.

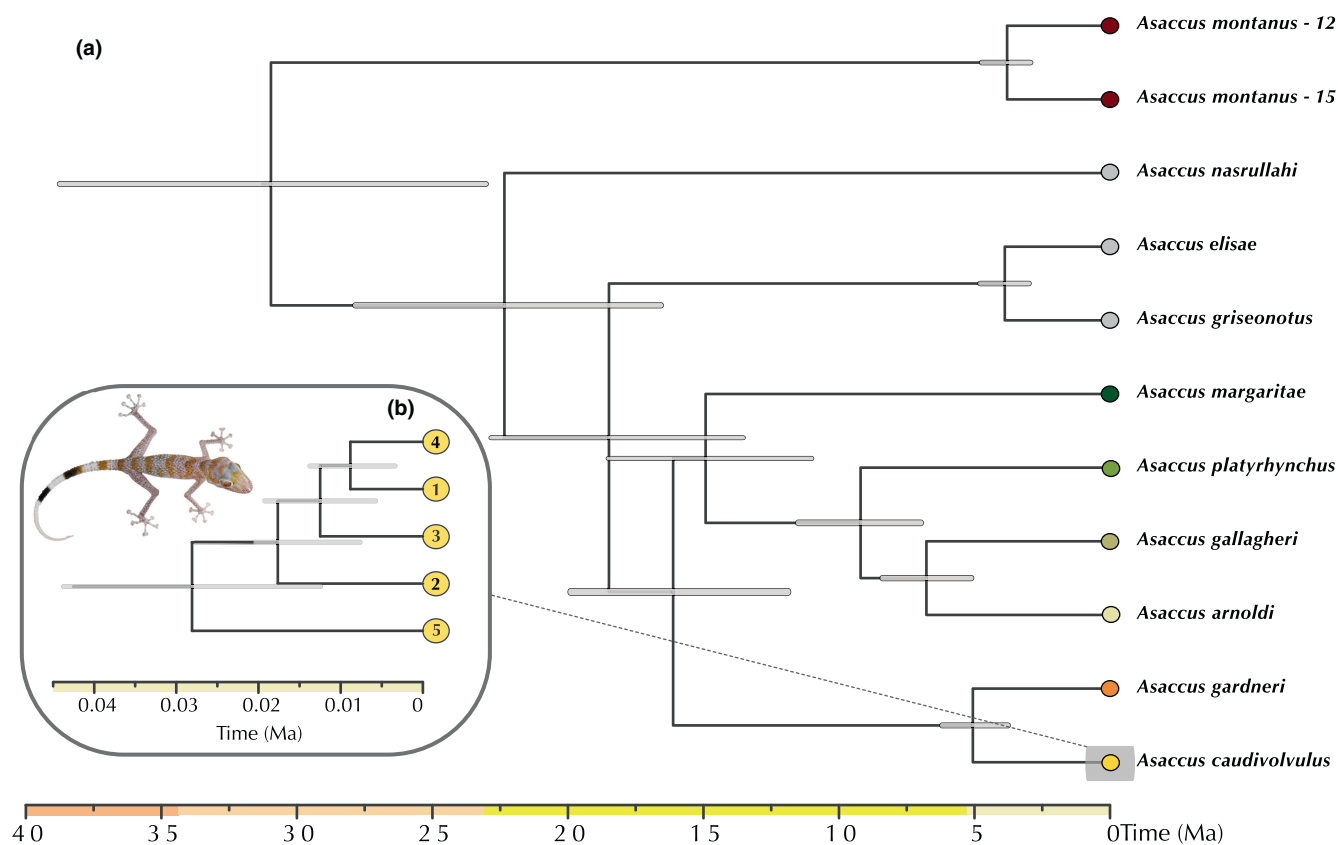
an almost complete annotation with a substantially high number of genes compared to other genomes annotated by the NCBI pipeline (Peel et al., 2022). The final genome assembly and annotation files were stored at NCBI under bioproject code PRJNA1043593 and at Mendeley Data project accessible at doi: [10.17632/bwtznysk9v.2](https://doi.org/10.17632/bwtznysk9v.2) respectively.

We also assembled the mitochondrial genome for the *A. caudivolvulus* reference sample as well as for all other *Asaccus* samples. Although complete circularization was not accomplished, we were able to assemble a 15,997-bp genome for the reference sample, and 13 protein-coding genes, two rRNAs and 22 tRNAs were identified for all samples, except for *A. nasrullahi* sample MVZ234330-19 and *A. caudivolvulus* sample CN21123-1, for which the assembly of the mitochondrial genome was not successful. The assembled mitochondrial genomes are available in NCBI under accession codes PP886534–PP886554.

### 3.2 | Phylogenomics

To explore the evolutionary history of the *Asaccus* genus, we inferred the phylogenomic relationships among species through ML

tree inference as well as Bayesian inference of the time-calibrated species tree under the MSC (Figures S3 and 2 respectively). The ML tree reconstruction shows strong support for the position of *A. margaritae* as sister to the clade formed by *A. platyrhynchus*, *A. arnoldi* and *A. gallagheri*. We also recovered two distinct Iranian clades, in contrast to the reported monophyly of *A. nasrullahi*, *A. griseonotus* and *A. elisae* (Carranza et al., 2016; Fattahi et al., 2020; Simó-Riudalbas et al., 2018). These results are in congruence with a recent phylogenomic reconstruction based on ddRAD sequencing (Burriel-Carranza et al., 2023). The same topology was supported by the time-calibrated species tree inferred under the MSC (Figure 2a), which yielded a mean time of divergence between *A. caudivolvulus* and *A. gardneri* of 5.08 Ma (95% HPD: 3.7–6.3 Ma). The MSC reconstruction of divergences within *A. caudivolvulus* (Figure 2b) showed that the first split within the species occurred approximately 27,000 years ago between the North Sharm population (locality 5) and all others. The internal topology of the clade formed by populations 1–4 should be interpreted with caution, since the ML and MSC species trees were discordant in the relationships of these populations. In the species tree based on ML, the specimen from Wadi Wurayah National Park (locality 4) is sister to all other localities and the specimen from Shark Island (locality 2) is recovered among the South Khor Fakkan



**FIGURE 2** Time-calibrated species trees. (a) Species tree inferred with SNAPP from a dataset of 50,000 uSNPs using one representative of each *Asaccus* species (including two deeply divergent lineages of *A. montanus*). Numbers following the two *A. montanus* tip labels correspond to the locality code from Table 1. (b) Time-calibrated species tree using SNAPP on a 50,000 uSNPs dataset containing all *A. caudivolvulus* specimens used in the present study. Grey bars depict the 95% posterior density intervals for the ages of all nodes. All nodes are supported by a posterior probability of 1. The geological time scale is colour coded according to the Commission for the Geological Map of the World (CGMW), Paris, France.

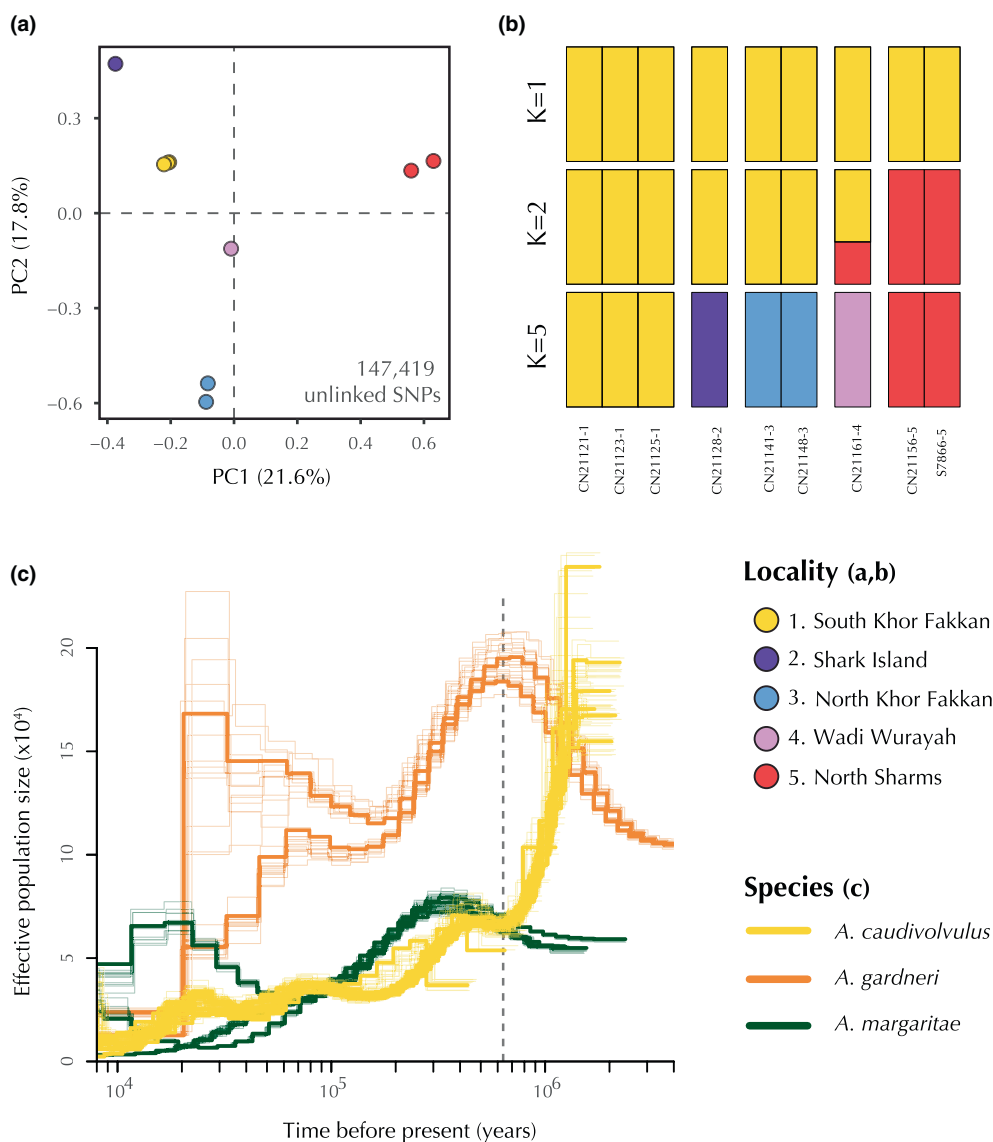


specimens (being sister to the geographically closest mainland specimen; Figure S3a). In the MSC species tree reconstruction, on the other hand, the specimen from Shark Island is recovered as sister to all other localities (localities 1, 3 and 4; Figure 2a).

The mitogenomic phylogeny recovered the same interspecific relationships among *Asaccus* as the nuclear ones, contrasting again with previous published data (see above). Here, however, the monophyly of *A. margaritae* with *A. gallagheri*, *A. arnoldi* and *A. platyrhynchus* was not robustly supported (Figure S3b). The support for the monophyly of the Iranian *Asaccus* could not be tested in the mitochondrial phylogeny since we were not able to recover the mitochondrial genome of *A. nasrullahi*.

### 3.3 | Population structure

We explored the population structure of *A. caudivolvulus* with a dataset of 147,419 uSNPs with no more than 10% missing data. In the PCA, all *A. caudivolvulus* localities were geographically segregated on PC1 (Figure 3a). The most likely number of populations inferred with ADMIXTURE was  $K=1$  (cross-validation=1.26). However, when a  $K=5$  was enforced, all specimens were clustered into their respective locality, with no signs of admixture between them (Figure 3b). In both analyses the *A. caudivolvulus* specimens from north of Sharm (locality 5) were the first to segregate from the rest of the localities, suggesting a greater genetic distance between them and the other specimens.



**FIGURE 3** Population structure and demographic history analyses. (a) First two principal components of a PCA of genetic variation in *A. caudivolvulus* summarizing 147,419 unlinked SNPs. (b) Results of admixture analyses in *A. caudivolvulus* specimens with  $K=1$ ,  $K=2$  and  $K=5$ . The numbers after the last dash in the specimen code represent locality code in Figure 1. (c) Demographic history reconstruction for the nine specimens of *A. caudivolvulus*, two *A. gardneri* and three *A. margaritae*. The generation time was set to 3.5 years and the substitution rate to  $0.6 \times 10^{-8}$  per site per year; the grey dashed line defines the end of the first recorded population decline in *A. caudivolvulus*, coinciding with the maximum effective population size recorded in *A. gardneri*; 15 bootstrap replicates for each sample are shown in the background with paler thinner lines.

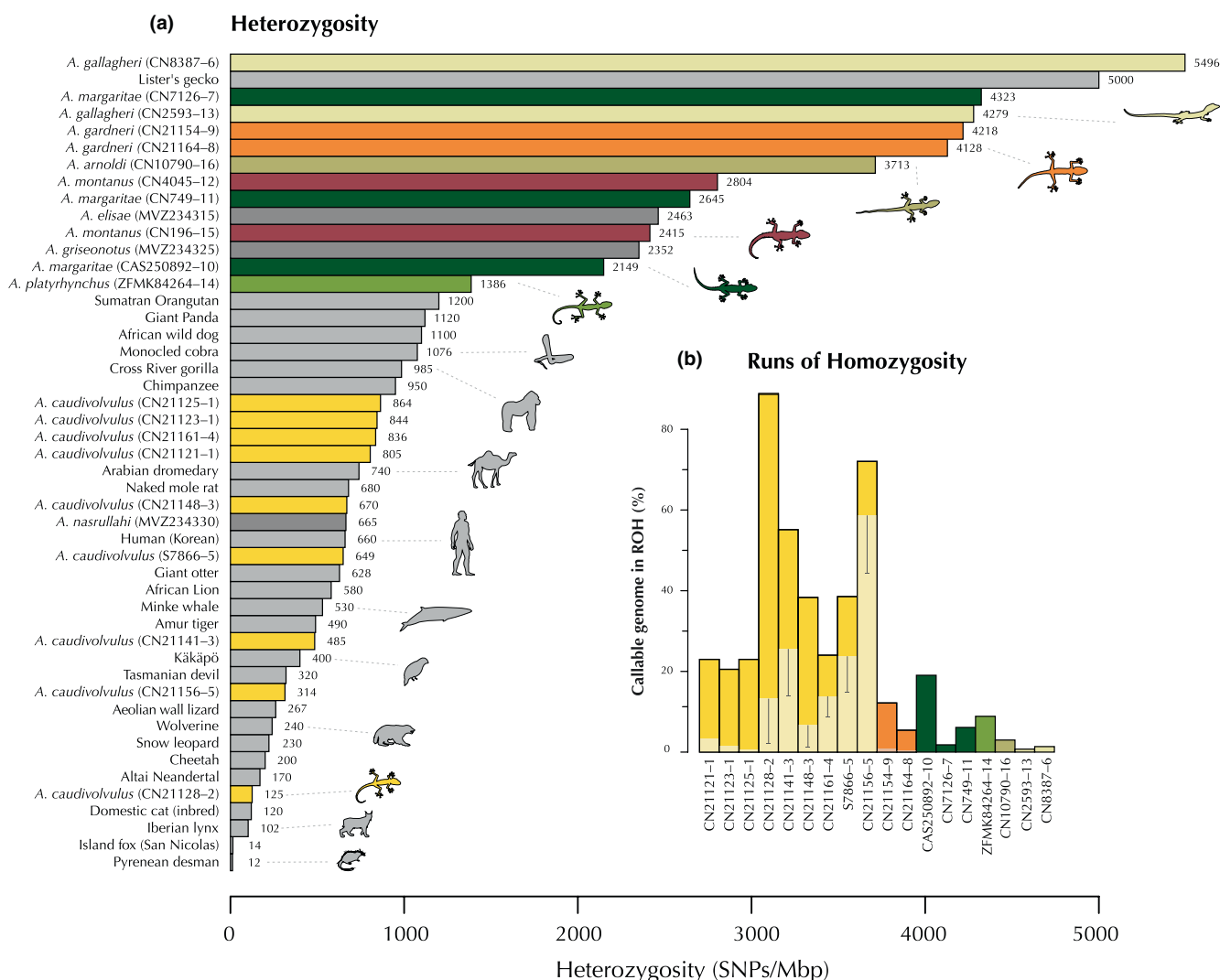
### 3.4 | Demographic history

The demographic history of *A. caudivolvulus*, *A. gardneri* and *A. margaritae* was evaluated through a PSMC analysis, which showed a continuous and sharp decline of *A. caudivolvulus* population sizes through time (Figure 3c), with the steepest drop occurring approximately 1.5–0.7 Ma. Effective population sizes in this species decreased from 150,000 individuals to approximately 60,000. In contrast, *A. caudivolvulus*' sister species, *A. gardneri*, presented the opposite pattern, showing increasing levels of population size in the 1.5- to 0.7-Ma period. A similar pattern was also found with *A. margaritae*, which, despite having already low base population sizes,

showed signs of population increase when population declines of *A. gardneri* occur, especially in the specimen from population 7 (specimen code CN7126; Table 1) which shows a significant population growth between  $10^5$  and  $10^4$  years ago. Moreover, the PSMC analyses allowed us to identify the date of coalescence among the three *A. margaritae* lineages to approximately 80,000 years ago.

### 3.5 | Heterozygosity and ROH

Genome-wide heterozygosity in *Asaccus caudivolvulus* samples was consistently lower than in all other *Asaccus* species (Figure 4a



**FIGURE 4** Assessment of genetic diversity. (a) Genome-wide heterozygosity for several mammal, non-avian reptile and bird species, most of them of conservation concern. For references for all species not sequenced in this study see Table S2. All Arabian leaf-toed geckos sequenced in this work are shown in colour. Iranian *Asaccus* species are highlighted in dark grey. Silhouettes for non-*Asaccus* species were retrieved from Phylopic ([phylopic.org](http://phylopic.org)). (b) Percentage of the callable genome under runs of homozygosity (ROH) for each sample, excluding *A. montanus* and all Iranian species. Paler bars represent the proportion of the genome under ROH accumulated in the last 10 generations ( $F_{ROH10}$ ; which is 0 or negligible in all species except *A. caudivolvulus*) when the lower extreme of the squamate recombination rate is used (see Materials and Methods). Whiskers show the  $F_{ROH10}$  for the higher extreme of the squamate recombination rate. An absence of a whisker means that the higher recombination rate is 0. The number after the last dash in the specimen codes indicates the locality of each specimen as in Table 1.

and Figure S4). Additionally, within *A. caudivolvulus* specimens, those found in non-developed localities presented higher levels of heterozygosity than those found in developed areas, although these had still lower heterozygosity than any other *Asaccus* species (with the exception of *A. nasrullahi*). While genome-wide heterozygosity ranged between 125 and 864 SNPs/Mbp in *A. caudivolvulus*, it was above 1000 SNPs/Mbp in all other *Asaccus* species (excluding *A. nasrullahi*), reaching 5496 SNPs/Mbp in *A. gallagheri* (Figure 4a and Figure S4a). Such low heterozygosity levels in *A. caudivolvulus* are comparable to other Critically Endangered species such as the Iberian lynx (*Lynx pardinus*) in the case of the Shark Island specimen (CN21128-2), or to the kākāpō (*Strigops habroptila*) in the case of the specimen from North Sharm in 2022 (Figure 4a, Table S2).

ROH analyses supported again a high disparity between *A. caudivolvulus* and the other *Asaccus* species (Figure 4b). For some specimens from localities 2, 3 and 5, more than 50% of their genome was in ROH (with sample CN21128-2 having more than 80%), while none of the other *Asaccus* species had more than 20% of their genome in ROH, and generally showed values of less than 10% (Figure 4b). Moreover, when inspecting the frequency of short ROH (sROH: 0.1–0.5 Mbp), medium ROH (mROH: 0.5–1 Mbp) and long ROH (lROH: >1 Mbp), we found that ROH found in most of the evaluated *Asaccus* species (including all *A. caudivolvulus* from locality 1) corresponded to sROH (Figure 5a). Short and medium ROH can be indicative of long-term small effective population sizes (Mochales-Riño et al., 2023) while lROH are consistent with recent inbreeding events. Specimens from localities 2 to 5 contained a high percentage of lROH, thus suggesting a history of recent inbreeding events in those localities (Figure 5a and Figure S4b). Further inspection of  $F_{ROH10}$  (inbreeding resulting from the last 10 generations; Figure 4b), as well as lROH distribution in chromosome 1 (Figure 5b), suggest that the specimen recently collected in the heavily developed locality in North Sharm (CN21156-5) is a product of a very recent inbreeding event, as 44%–58% of its genome (corresponding to the lower and higher ends of the squamate recombination rates used respectively) is in ROH resulting from inbreeding in the last 10 generations (Figure 4b, Table S3). Surprisingly, while CN21128-2 is the specimen with the highest percentage of its genome in ROH (Figure 4b) and presents a high percentage of its genome in lROH (Figure 5a), those lROH seem to be partially fragmented with mean ROH lengths being approximately four times lower than the aforementioned CN21156-5 specimen (Table S3, Figure 4b).

These results were supported by the correlation between the number of ROH (NROH) and the cumulative sum of ROH (SROH; Pearson correlation  $t=2.13$ , correlation value=0.47 and  $p$ -value=0.04). Characteristic for small and inbred populations, *A. caudivolvulus* specimens from localities 3 and 5 presented a small NROH but a high SROH, whereas large and diverse populations tend to present reduced values in both metrics (Figure 5c).

### 3.6 | Mutational load

We compared the mutational load of four populations of the Emirati Leaf-toed Gecko (localities 1–4). Specimens from locality 5 were not included since preliminary analyses showed that they were influenced by reference bias. Although significance could not be confirmed due to the low number of samples from each locality, considerable differences were observed between specimens from locality 2 and the other localities. While the total load remained similar, the realized load (i.e. the number of high-impact deleterious alleles affecting the current generation) observed in the Shark Island specimen was 1.4–2 times higher than any other specimen (Figure 5d and Table S4). Conversely, the masked load was 3.3–5.5 times lower in the Shark Island specimen than in any other specimen (Figure 5d and Table S4). The same pattern was observed when evaluating the number of moderate-impact and low-impact alleles, with the specimen from Shark Island having consistently higher realized loads and lower masked loads than any other specimen (Table S4 and Figure S5).

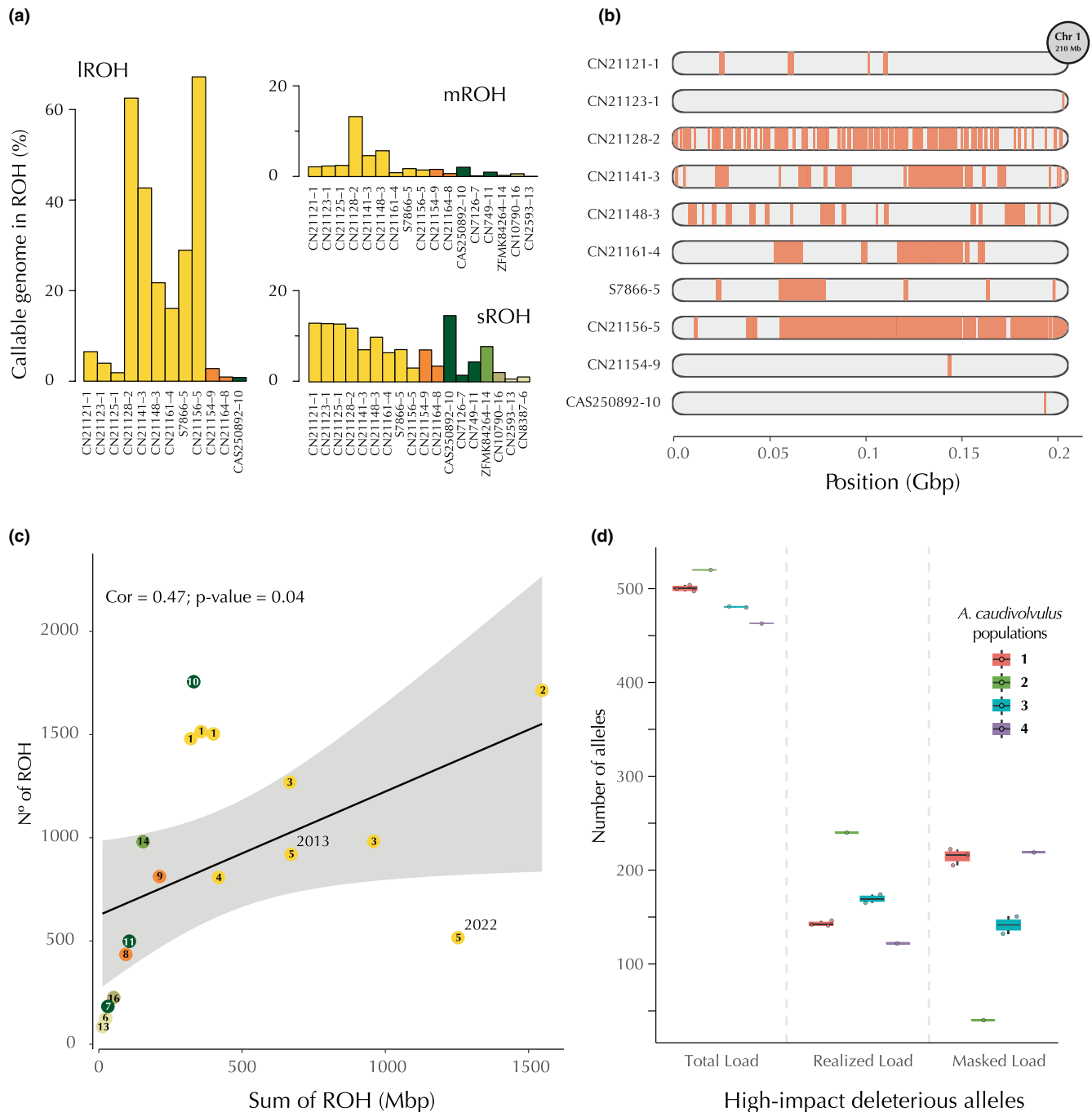
## 4 | DISCUSSION

In the present study, we assembled and annotated a chromosome-level genome for *Asaccus caudivolvulus*, representing the first reference genome for the family Phyllodactylidae. This reference genome fills a gap for the Gekkomorpha group (including Eublepharidae, Sphaerodactylidae, Gekkonidae and Phyllodactylidae), where reference genomes for all the other families have already been produced (Liu et al., 2015; Pinto et al., 2022, 2023), and can thus become a key resource to better understand the history and evolution of sex determination in geckos (Gamble, 2010).

Here, we have generated, for the first time, whole-genome resources for all Arabian leaf-toed geckos, together with representatives from Iran. We have placed a special focus on the Critically Endangered Emirati Leaf-toed Gecko (*A. caudivolvulus*), the only endemic vertebrate of the UAE, which was feared to be extinct, but recently found by our team in five isolated localities across the UAE East Coast (i.e. some of them being heavily developed; Figure 1). The present study has applied an integrative conservation genomics approach, providing detailed insights into the phylogenomics, population structure, demographic history, genomic diversity and mutational load of *A. caudivolvulus*, in order to guide future conservation plans for this species.

### 4.1 | Phylogeography and demographic history of the Emirati Leaf-toed Gecko

*Asaccus caudivolvulus* is part of a clade of leaf-toed geckos that colonized the Hajar Mountains through Iran about 20–10 Ma, coinciding with the uplift of the Zagros Mountains in Iran (Burriel-Carranza et al., 2023). Since then, a species radiation has occurred with some



**FIGURE 5** Distribution of runs of homozygosity (ROH) and mutational load. (a) Proportion of the callable genome under long ( $>1$  Mbp), medium ( $0.5 > ROH < 1$  Mbp) and short ( $0.1 > ROH < 0.5$  Mbp) ROH (IROH, mROH and sROH respectively). Colours represent different *Asaccus* species as in Figure 1. (b) Samples containing IROH (red bars) across chromosome 1, showing the distribution, amount and fragmentation of IROH across the chromosome. Numbers after the last dash in the specimen code show locality numbers. All specimens are of the species *A. caudivolvulus*, with the exception of the last two, which are an *A. gardneri* and *A. margaritae*, respectively. (c) Pearson's correlation between the number (NROH) and cumulative sum of ROH length (SROH). Samples above the line suggest long-term inbreeding and samples below the line support recent inbreeding. Numbers represent the locality code for each specimen. (d) Mutational load is given as the number of high-impact deleterious alleles affecting each *A. caudivolvulus* specimen from populations 1–4. Total load: Sum of Masked load and two times Realized load (counting each deleterious allele); Realized load: Number of deleterious alleles in homozygosity, thus being expressed in the current generation; Masked Load: Number of deleterious alleles being observed in heterozygosity.

species dispersing south throughout the whole mountain range. Traditionally, this clade of *Asaccus* geckos had been divided into two monophyletic groups that can be phenotypically distinguished based

on the coloration of the tail in males: the yellow-tailed clade (composed by *A. platyrhynchus* and the dwarf *Asaccus* species *A. gallagheri* and *A. arnoldi*); and the white-tailed clade (composed by *A. gardneri*,



*A. caudivolvulus* and *A. margaritae*, and restricted to the Western Hajar; Carranza et al., 2016; Simó-Riudalbas et al., 2018). However, the monophyly of these two groups was contested by a recent phylogeny based on ddRADseq data (Burriel-Carranza et al., 2023), which recovered *A. margaritae* as sister to all other *Asaccus* from the yellow-tailed clade. In the present study, we corroborate this new topology with strong and shallow support from the nuclear and mitochondrial phylogenies, respectively. Therefore, we confirm *A. margaritae* to be sister species to the yellow-tailed clade (Figure 2a and Figure S3). These findings are also phenotypically congruent, since *A. margaritae* juvenile specimens also present a yellow–copperish tail (adult males do not), while juveniles from *A. gardneri* and *A. caudivolvulus* always present white tails with one to three dark bands (Burriel-Carranza et al., 2022; Carranza et al., 2021).

The time-calibrated MSC species tree suggested that *A. caudivolvulus* diverged from its sister species *A. gardneri* approximately 5.05 Ma, in the earliest split among described Arabian *Asaccus* species (Figure 2a). In contrast to other Arabian *Asaccus* (e.g. *A. gallagheri* or *A. montanus*; Figure S3), intraspecific divergence within *A. caudivolvulus* appears to be relatively recent, with the first split within the species having occurred around 28,000 years ago, separating North Sharm specimens from the other localities (Figure 2b). Despite the recent split among populations, we did not find clear evidence of gene flow between them. Population structure analyses clearly separated specimens geographically (Figure 3a), with the exception of the specimen from Wadi Wurayah National Park, which was partially assigned to several populations until  $K=5$  (Figure 3b). This population is of particular interest, as it is found in the only locality that is not by the sea (Figure 1). Despite extensive sampling in the area, the specimens found in Wadi Wurayah National Park were exclusively seen in rocks that had been translocated and placed within a parking area at the entrance to the National Park. Additionally, this parking area shows a distinct substrate composition compared to the natural surroundings of the wadi. Together, these observations suggest that this population is probably the product of human introduction. The provenance of this population remains unknown as it clustered independently in the genomic space reconstructed by means of a PCA (Figure 3a). However, given the genetic similarity to the *A. caudivolvulus* specimens from near Khor Fakkan, the geographic proximity between this city and the National Park, and the number of extraction sites surrounding the city, it appears plausible that the origin of this population comes from an unsampled and possibly extinct population that probably occupied former outcrops surrounding Khor Fakkan.

The reconstruction of the demographic history of *A. caudivolvulus* has shown that this species has not only been affected by current anthropogenic coastal development, but also by a long history of population declines and subsequent bottlenecks. In particular, a steep population drop between 1.5 and 0.7 Ma with effective population sizes plummeting from approximately 150,000 specimens to 60,000 was observed in *A. caudivolvulus*. This sudden decline coincides with a population growth of its sister species, *A. gardneri* (Figure 3c). Gardner's Leaf-toed Gecko is the largest Arabian *Asaccus*

species, reaching up to 71 mm of snout-to-vent length, and it can be considerably abundant at low elevations in wadis, caves and outcrops in the Musandam Peninsula, reaching south until the Dibba Line (Carranza et al., 2016, 2021). In the Musandam Peninsula, this species partially shares its distribution with the smallest species of the former white-tailed *Asaccus* clade, *A. margaritae*. In the regions where both species are present (e.g. locality 7; Figure 1), *A. margaritae* only occurs in summits above 1300 m above sea level (asl), beyond the 1000 m maximum elevation that *A. gardneri* reach (Carranza et al., 2021). However, in regions where these two species do not coexist, *A. margaritae* can be found up to 125 m a.s.l. (e.g. locality 11; Figure 1). Considering that both species share similar habitat requirements and diets (Carranza et al., 2021), along with opposite population trends—especially in locality 7—and the fact that *A. gardneri* is larger and more agile than *A. margaritae*, suggests a scenario consistent with the displacement of *A. margaritae* by *A. gardneri* towards mountain summits through competitive interactions. Although currently occurring in allopatry, such displacement through competition could have also occurred with the Emirati Leaf-toed Gecko, which has a similar body size and habits as *A. margaritae* (Burriel-Carranza et al., 2022), and would explain the dramatic population declines observed in this species around 1 mya. However, as an alternative to resource competition with other species, the dramatic population drop in *A. caudivolvulus* could have been driven by the aridification trends of the Arabian Peninsula. In contrast to *A. margaritae* and *A. gardneri*, *A. caudivolvulus* seems to be constrained to habitats with exceptionally high humidity (Carranza et al., 2016) which, consequently, make this species especially susceptible to the ongoing aridification of Arabia and especially to hyper-aridity episodes (Böhme et al., 2021). After this first bottleneck, small fluctuations on effective population sizes in *A. caudivolvulus*, seemingly decoupled from *A. gardneri*'s demographic history, were also observed (Figure 3c). Such fluctuations could have been the result of population declines and growths responding to further Quaternary climatic fluctuations on already displaced or reduced populations (Petraglia et al., 2018; Roberts et al., 2018).

## 4.2 | Signals of recent inbreeding in an already depauperate genome

Long-term population declines have left a genomic footprint in the *A. caudivolvulus* genome in terms of much lower base heterozygosity levels than other congeneric species (with the exception of *A. nasrullahi*; Figure 4a and Figure S4). The loss of diversity in past bottlenecks in *A. caudivolvulus* could have already diminished their genetic toolkit, limiting the capacity of adaptation to new threats for this species. The currently most diverse population (locality 1) has similar heterozygosity levels as the Critically Endangered and long-termed inbred Cross River gorilla (Xue et al., 2015), but an up to seven times lower heterozygosity estimate (125 SNPs/Mbp) was found in the specimen from Shark Island (locality 2), representing the lowest heterozygosity values ever recorded in reptiles using whole genome

resequencing (Gabrielli et al., 2023) and suggesting a scenario of severe recent inbreeding (Figure 4a).

Recent and long-term inbreeding can also be evaluated through the analysis of IBD segments represented by ROH. While sROH and mROH are indicative of old population bottlenecks, IROH suggest a recent history of inbreeding. While all *A. caudivolvulus* samples had a higher percentage of their genome under ROH than any other congeneric species (Figure 4b), the distribution of these into short, medium and long fragments was not even among samples (Figure 5a–c). The composition of ROH from the specimens found south of Khor Fakkan (locality 1) was mostly of sROH, suggesting that this population was most likely not affected by recent inbreeding. In all other localities, all samples presented a higher number of IROH than the sum of mROH and sROH together, supporting recent inbreeding. It is noteworthy that all of the latter localities were affected, up to some extent, by habitat fragmentation, transformation or destruction that could be promoting inbreeding. Specimens from localities 2 and 4 are constrained to very small areas, the first one spanning a 150×60m stretch of rock and the second being restricted to an islet of approximately 450×200m. Specimens from localities 3 and 5 have undergone a great deal of habitat destruction due to coastal development on the UAE East coast (Figure 1).

Specimens from localities 3 and 5 show signals of a recent inbreeding history. Specimen CN21141–3 from north of Khor Fakkan had a maximum value of  $F_{\text{ROH}10}$  of 0.26 (Figure 4b), which suggests that its parents were related at approximately the level of full siblings (Kardos et al., 2018). This is also supported by the correlation between NROH and SROH (Figure 5c), where this specimen is found below the regression line, suggesting a scenario of recent inbreeding. This pattern is not apparent in the other sample from this locality (CN21148–3), which was found slightly away from the construction site (Figure 1). Exceptionally worrying are the results obtained from the two temporal samples from north of Sharms. While specimen S7866–5 (from 2013) already presented considerable levels of recent inbreeding ( $F_{\text{IROH}}=0.28$ ;  $F_{\text{ROH}10}=0.13$ – $0.23$ ), the specimen sampled 9 years later (CN21156–5), after the development of a hotel complex that caused a notable amount of habitat destruction (Figure 1), showed extremely high values of recent inbreeding ( $F_{\text{IROH}}=0.67$ ,  $F_{\text{ROH}10}=0.59$ – $0.45$ ). Additionally, 83.5% of the observed IROH accumulated during the last 10 generations, suggesting a continuous and successive mating between closely related individuals. Since *A. caudivolvulus* seems to be a species with relatively low dispersal capabilities (i.e. as observed by the lack of admixture between populations at less than 3km from each other; Figure 3b), the maintenance of each population's genetic pool is of utmost importance, as it is not expected for migrants to arrive at new localities and increase the genomic diversity. Hence the continuous inbreeding caused by anthropic pressures may lead to reduced fitness and inbreeding depression in this species if no measures are taken into consideration. The case of rapid depauperation of genomic diversity observed in the locality north of Sharms, foreshadows the effects that coastal development may cause in other *Asaccus* populations if no conservation management plans are implemented.

From the four localities with high  $F_{\text{IROH}}$ , the only case that is most probably not anthropologically driven, and therefore might have been occurring for an extended period of time, is that of Shark Island. Long-distance marine colonization in geckos has been reported extensively, with dispersal across up to 6000km in some cases (Carranza et al., 2000). Hence, only being about 400m away from another mainland *Asaccus* population, natural colonization of Shark Island is more than probable to have occurred, even several times. Island populations are expected to have high levels of inbreeding, retaining lower population sizes due to limited resource availability (Frankham, 1998). Despite being relatively close to mainland, the extremely high frequency of ROH in specimen CN21128–2 (Figure 4b) suggests a high degree of inbreeding in the Shark Island population. Surprisingly, although having an  $F_{\text{IROH}}$  of 62% (Figure 5a and 54b), only between 2% and 12% of the genome was found to be under ROH from inbreeding events in the last 10 generations (Figure 4b, Table S3). Since ROH can be easily fragmented when mating of non-related individuals occurs, such disparity could be indicative of occasional migrants from the mainland adding diversity to the island's genetic pool and thus preventing the population's collapse. However, the large fraction of the genome under IROH older than 10 generations suggest a long history of recurrent inbreeding, which could also lead to reduced individual fitness and inbreeding depression. Further studies should focus on analysing a larger sample size of *A. caudivolvulus* from Shark Island to assess the viability of this population.

### 4.3 | Coping with inbreeding

Mutational load, the reduction in individual and mean population fitness caused by recent deleterious mutations, can be partitioned into two categories in diploid organisms: the realized load and the masked load. The realised load (or expressed load) represents those alleles in homozygosity (assuming recessivity) that reduce the fitness of the current generation. The masked load (or potential load) is represented by those alleles in heterozygosity carrying potentially harmful recessive variants (Bertorelle et al., 2022). Therefore, long stretches of the genome being IBD as a result of inbreeding or long demographic bottlenecks tend to leave a footprint on the mutational load of an individual, which should have lower masked loads and higher realized loads than 'healthy' specimens (Bertorelle et al., 2022). Some species seem to be able to avoid inbreeding depression by purging of deleterious mutations and thus reducing their realised load. Although this has been recently reported in several cases of endangered and inbred species (Dusseix et al., 2021; Kleinman-Ruiz et al., 2022; Mochales-Riaño et al., 2023), including some examples of reptiles like the Chinese crocodile lizard (Xie et al., 2022), it does not seem to be the case in *A. caudivolvulus*. The specimen from Shark Island, following the theoretical predictions after a prolonged bottleneck (Bertorelle et al., 2022), had consistently higher realized load and lower masked load than specimens from other localities, together

with a higher total load than in other populations (Figure 5d and Figure S5; Table S4). Such result is congruent with other analyses suggesting a long-term bottleneck in Shark Island but adds another layer of complexity for the survival of this species: Since this result suggests that the species does not have the capability to purge deleterious mutations, it becomes even more susceptible to future threats.

However, as this analysis could not be carried out on *A. caudivolvulus* specimens from population 5, or be compared to any other *Asaccus* species due to the effects of reference bias, further studies are advisable to better understand the effects of mutational load on leaf-toed geckos. Future work should focus on this direction, including more specimens from each locality and using a reference genome from a phylogenetically equidistant relative to all *Asaccus* species (i.e. the Socotran endemic genus *Haemodracon* would represent the closest relative) when alternative reference genomes become available.

## 5 | CONCLUSIONS

Here we provide a comprehensive and exhaustive conservation genomics study on the Critically Endangered Emirati Leaf-toed Gecko (*Asaccus caudivolvulus*), the only endemic vertebrate of the UAE. We assembled a chromosome-level reference genome for the species (constituting the first reference genome for the Phyllodactylidae family) and provided whole genomes for 23 *Asaccus* specimens including all Arabian species and three Iranian *Asaccus*. Despite the discovery of *A. caudivolvulus* in five localities, results show that this species is severely threatened. Signals of ancient and recent population declines, extremely low genetic diversity, several evidences of recent inbreeding, and low capabilities of purging of deleterious mutations make this gecko especially vulnerable to any future stochastic factor, whether demographic, environmental, catastrophic, or genetic. Although some signals seem to be originating from natural processes, the demise of this species has been evidently worsened by the human activities through the UAE East Coast over the last decade, which has undergone a rapid transformation with recurrent destruction of the coastal outcrops that serve as the natural habitat of the Emirati Leaf-toed Gecko.

We strongly advocate for the preservation of this species' habitat, especially considering that relocation further north might not be feasible due to the presence of its sister species, *A. gardneri*, which could pose a threat through competition. Another viable approach would be the implementation of ex situ breeding programmes for this species. With a carefully planned breeding programme of unrelated individuals, the genetic diversity within the species could be bolstered and thereby mitigate the IROH observed. However, further studies including more individuals should be conducted to ensure the viability of such techniques. Altogether, results suggest that this emblematic gecko species might be on the brink of extinction and its survival in the wild might not be warranted unless immediate protection is ensured.

## AUTHOR CONTRIBUTIONS

BB-C, JE, MM and SC conceptualised the study. BB-C, GM-R, AT, JE, JDUS, OO and SC conducted field or laboratory work. BB-C analysed the data. BB-C wrote the original draft. All authors reviewed, edited and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

We would like to thank the Mohamed Bin Zayed Species Conservation Fund for their invaluable support in our conservation efforts by funding project Ref: 212527924 on the Emirati Leaf-toed Gecko. Their partnership through this funded project serves as a beacon of hope for the future of the UAE environment and the preservation of our natural heritage. Special thanks to Nicolas Heard, Kirk Duthler and Conrad Harley from the MBZ Species Conservation Fund for their interest and collaboration. From Sharjah, we would like to thank His Highness Sheikh Dr Sultan bin Mohammed Al Qasimi, Supreme Council Member and Ruler of Sharjah, H. E. Ms. Hana Saif al Suwaidi (Chairperson of the Environment and Protected Areas Authority, Sharjah), Paul Vercammen and Kevin Budd (Breeding Centre for Endangered Arabian Wildlife), and Gary Feulner (Dubai Natural History Group) for their continuous support. This project was supported by the Patronage of H. H. Sheikh Mohammed bin Hamad Al Sharqi, Crown Prince of the Emirate of Fujairah of the U.A.E. We acknowledge his support and inspiration in the initiation and mentoring of this project, without whom this project would not have been possible. We would like to express our gratitude to Ms. Asella Mohalla, the Director of the Fujairah Environment Authority, for her generosity in granting us access permits to the marine reserves. Special thanks to Abdul Nasser Obeidat, the Head Researcher of the Fujairah Environment Authority. His invaluable support and expert guidance during our work in the Wadi Wurayah National Park. We sincerely thank the Address Beach Resort Fujairah for their vital logistical support, which played an invaluable role in our study. From Oman, we would like to thank all the past and present members of the Ministry of Environment and Climate Affairs, MECA, Oman, now the Environment Authority Oman, and especially to Ali Al Kiyumi, Suleiman Nasser Al Akhzami, Thuraya Al Sariri, Ahmed Said Al Shukaili, Ali Alghafri, Sultan Khalifa, Hamed Al Farqani, Salim Bait Bilal, Iman Sulaiman Alzari, Aziza Saud Al Adhoobi, Mohammed Al Shariani, Zeyana Salim Al Omairi and Abdullah bin Ali Al Amri, Chariman of the Environment Authority. We are also very grateful to past members and collaborators including Edwin Nicholas Arnold, Michael D. Robinson, Andrew Gardner, Josep Roca, Philip de Pous, Margarita Metallinou, Raquel Vasconcelos, Jiri Smid, Karin Tamar, Loukia Spilani, Dean Adams, Lidia Escoda and Fabrizia Ronco for their help and support. This work was funded by the Mohamed Bin Zayed Species Conservation Fund (Project 212527924), grants PGC2018-098290-B-I00 (MCIU/AEI/FEDER, UE), PID2021-128901NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF, a way of making Europe, Field Study for the Conservation of Reptiles in Oman, Ministry of Environment and Climate Affairs, Oman (Ref: 22412027), and grant 2021 SGR 00420

from the Departament de Recerca i Universitats de la Generalitat de Catalunya to SC. Nanopore genome sequencing was supported by ORG.one, an initiative by Oxford Nanopore Technologies to support sequencing of endangered species, which provided PromethION flow cells and sequencing reagents. The computations for the assembly of the reference genome were performed on resources provided by Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. BB-C was funded by an FPU grant from Ministerio de Ciencia, Innovación y Universidades, Spain (FPU18/04742). GM-R was funded by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE2019-088729). AT is supported by the 'la Caixa' doctoral fellowship programme (LCF/BQ/DR20/11790007). ME was funded by an FPI grant from the Ministerio de Ciencia, Innovación y Universidades, Spain (PRE 2022-101473). Permits from Oman were issued by the Nature Conservation Department of the Ministry of Environment and Climate Affairs, Oman (Refs: 08/2005; 16/2008; 38/2010; 12/2011; 13/2013; 21/2013; 37/2014; 31/2016; 6210/10/21).

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Reference genome assembly, ONT long reads, Illumina and OmniC raw data used for the reference genome assembly, as well as whole genome sequencing Illumina raw data for 23 specimens are available on NCBI under project code PRJNA1043593. Mitochondrial genomes for 21 specimens are accessible in NCBI under accession codes PP886534–PP886554. Annotation files for the reference genome assembly were stored in a Mendeley Data project accessible at <https://doi.org/10.17632/bwtznysk9v.2>.

## ORCID

Bernat Burriel-Carranza  <https://orcid.org/0000-0001-9832-1782>

Gabriel Mochales-Riaño  <https://orcid.org/0000-0002-9130-2308>

Adrián Talavera  <https://orcid.org/0000-0002-4007-0965>

Johannes Els  <https://orcid.org/0000-0002-5614-1887>

Maria Estarellas  <https://orcid.org/0000-0001-7360-6458>

Juan Diego Urriago Suarez  <https://orcid.org/0000-0002-2294-9335>

<https://orcid.org/0000-0002-2294-9335>

Michael Matschiner  <https://orcid.org/0000-0003-4741-3884>

Salvador Carranza  <https://orcid.org/0000-0002-5378-3008>

## REFERENCES

- Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12(1), 1–6. <https://doi.org/10.1186/1471-2105-12-246>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/GR.094052.109>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697–709. <https://doi.org/10.1038/nrg2844>
- Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdoci, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*, 20(4), 892–905. <https://doi.org/10.1111/1755-0998.13160>
- Arnold, E. N., & Gardner, A. S. (1994). A review of the middle eastern leaf-toed geckoes (Gekkonidae: *Asaccus*) with descriptions of two new species from Oman. *Fauna of Saudi Arabia*, 14, 424–441.
- Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & van Oosterhout, C. (2022). Genetic load: Genomic estimates and applications in non-model animals. *Nature Reviews Genetics*, 23(8), 492–503. <https://doi.org/10.1038/s41576-022-00448-x>
- Böhme, M., Spassov, N., Majidifard, M. R., Gärtner, A., Kirscher, U., Marks, M., Dietzel, C., Uhlig, G., El Atfy, H., Begun, D. R., & Winkhofer, M. (2021). Neogene hyperaridity in Arabia drove the directions of mammalian dispersal between Africa and Eurasia. *Communications Earth & Environment*, 2(1), 1–13. <https://doi.org/10.1038/s43247-021-00158-y>
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/JOURNAL.PCBI.1006650>
- Brúna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012). Inferring species trees directly from Biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917–1932. <https://doi.org/10.1093/MOLBEV/MSS086>
- Burriel-Carranza, B., Els, J., & Carranza, S. (2022). Reptiles & Amphibians of the Hajar Mountains. Editorial CSIC. [http://libros.csic.es/producto\\_info.php?products\\_id=1605&language=en](http://libros.csic.es/producto_info.php?products_id=1605&language=en)
- Burriel-Carranza, B., Tejero-Cicuéndez, H., Carné, A., Riaño, G., Talavera, A., Saadi, S. A., Els, J., Šmíd, J., Tamar, K., Tarroso, P., & Carranza, S. (2023). The origin of a mountain biota: Hyper-aridity shaped reptile diversity in an Arabian biodiversity hotspot. *bioRxiv*, p. 2023.04.07.536010. <https://doi.org/10.1101/2023.04.07.536010>
- Calderón, P. L., & Pigozzi, M. I. (2006). MLH1-focus mapping in birds shows equal recombination between sexes and diversity of cross-over patterns. *Chromosome Research*, 14(6), 605–612. <https://doi.org/10.1007/s10577-006-1059-0>
- Card, D. C., Jennings, W. B., & Edwards, S. V. (2023). Genome evolution and the future of Phylogenomics of non-avian reptiles. *Animals*, 13(3), 3. <https://doi.org/10.3390/ani13030471>
- Carranza, S., Arnold, E. N., Mateo, J. A., & López-Jurado, L. F. (2000). Long-distance colonization and radiation in gekkonid lizards, *Tarentola* (Reptilia: Gekkonidae), revealed by mitochondrial DNA sequences. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1444), 637–649. <https://doi.org/10.1098/rspb.2000.1050>
- Carranza, S., & Els, J. (2021). *Asaccus caudivolvulus*. The IUCN red list of threatened species 2021. *IUCN*, e.T125056608A125056621.
- Carranza, S., Els, J., & Burriel-Carranza, B. (2021). A field guide to the reptiles of Oman. Digital CSIC. [http://libros.csic.es/producto\\_info.php?products\\_id=1558](http://libros.csic.es/producto_info.php?products_id=1558)
- Carranza, S., Simó-Riudalbas, M., Jayasinghe, S., Wilms, T., & Els, J. (2016). Microendemism in the northern Hajar Mountains of Oman



- and The United Arab Emirates with the description of two new species of geckos of the genus *Asaccus* (Squamata: Phyllodactylidae). *PeerJ*, 4, e2371. <https://doi.org/10.7717/peerj.2371>
- Ceballos, G., Ehrlich, P. R., & Raven, P. H. (2020). Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 117(24), 13596–13602. <https://doi.org/10.1073/pnas.1922686117>
- Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit—Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, 10(4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. <https://doi.org/10.1186/S13742-015-0047-8/2707533>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dodge, T. O., Farquharson, K. A., Ford, C., Cavanagh, L., Schubert, K., Schumacher, M., Belov, K., & Hogg, C. J. (2023). Genomes of two extinct-in-the-wild reptiles from Christmas Island reveal distinct evolutionary histories and conservation insights. *Molecular Ecology Resources*, 1–17. <https://doi.org/10.1111/1755-0998.13780>
- 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 Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., ... Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics*, 1(1), 100002. <https://doi.org/10.1016/j.xgen.2021.100002>
- Escoda, L., & Castresana, J. (2021). The genome of the Pyrenean desman and the effects of bottlenecks and inbreeding on the genomic landscape of an endangered species. *Evolutionary Applications*, 14(7), 1898–1913. <https://doi.org/10.1111/eva.13249>
- Farkas, B., Buzás, B., & Farkas, V. (2018). Additions to the herpetofauna of the Wadi Wurayah National Park, Fujairah. *Tribulus*, 26, 42–45.
- Fattahi, A., Rastegar-Pouyani, N., Rastegar-Pouyani, E., Karamiani, R., Yousefkhani, S. S. H., & Fathinia, B. (2020). Molecular phylogeny and taxonomic evaluation of the genus *Asaccus* Dixon and Anderson, 1973 (Reptilia: Phyllodactylidae) in Iran. *Herpetological Journal*, 30(4), 207–214. <https://doi.org/10.33256/hj30.4.207214>
- Formenti, G., Abueg, L., Brajuka, A., Brajuka, N., Gallardo-Alba, C., Giani, A., Fedrigo, O., & Jarvis, E. D. (2022). Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*, 38(17), 4214–4216. <https://doi.org/10.1093/bioinformatics/btac460>
- Frankham, R. (1998). Inbreeding and extinction: Island populations. *Conservation Biology*, 12(3), 665–675. <https://doi.org/10.1111/j.1523-1739.1998.96456.x>
- Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126(2), 131–140. <https://doi.org/10.1016/j.biocon.2005.05.002>
- Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., & Stanke, M. (2021). TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics*, 22(1), 566. <https://doi.org/10.1186/s12859-021-04482-0>
- Gabrielli, M., Benazzo, A., Biello, R., Ancona, L., Fuselli, S., Iannucci, A., Balacco, J., Mountcastle, J., Tracey, A., Ficetola, G. F., Salvi, D., Solitto, M., Fedrigo, O., Formenti, G., Jarvis, E. D., Gerdel, M., Ciofi, C., Trucchi, E., & Bertorelle, G. (2023). A high-quality reference genome for the critically endangered Aeolian wall lizard, *Podarcis raifonei*. *Journal of Heredity*, 114(3), 279–285. <https://doi.org/10.1093/jhered/esad014>
- Gamble, T. (2010). A review of sex determining mechanisms in geckos (Gekkota: Squamata). *Sexual Development*, 4(1–2), 88–103. <https://doi.org/10.1159/000289578>
- Gaughran, S. J., Quinzin, M. C., Miller, J. M., Garrick, R. C., Edwards, D. L., Russello, M. A., Poulakakis, N., Ciofi, C., Beheregaray, L. B., & Caccone, A. (2018). Theory, practice, and conservation in the age of genomics: The Galápagos giant tortoise as a case study. *Evolutionary Applications*, 11(7), 1084–1093. <https://doi.org/10.1111/eva.12551>
- Gemmell, N. J., Rutherford, K., Prost, S., Tollis, M., Winter, D., Macey, J. R., Adelson, D. L., Suh, A., Bertozzi, T., Grau, J. H., Organ, C., Gardner, P. P., Muffato, M., Patricio, M., Billis, K., Martin, F. J., Flicek, P., Petersen, B., Kang, L., ... Stone, C. (2020). The tuatara genome reveals ancient features of amniote evolution. *Nature*, 584(7821), 403–409. <https://doi.org/10.1038/s41586-020-2561-9>
- Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A. M., & Koren, S. (2019). Integrating hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology*, 15(8), e1007273. <https://doi.org/10.1371/journal.pcbi.1007273>
- Girgis, H. Z. (2015). Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics*, 16(1), 227. <https://doi.org/10.1186/s12859-015-0654-5>
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9), 2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Hung, C.-M., Shaner, 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 of the United States of America*, 111(29), 10636–10641. <https://doi.org/10.1073/pnas.1401526111>
- Iannucci, A., Benazzo, A., Natali, C., Arida, E. A., Zein, M. S. A., Jessop, T. S., Bertorelle, G., & Ciofi, C. (2021). Population structure, genomic diversity and demographic history of komodo dragons inferred from whole-genome sequencing. *Molecular Ecology*, 30(23), 6309–6324. <https://doi.org/10.1111/mec.16121>
- Irizarry, K. J. L., Bryant, D., Kalish, J., Eng, C., Schmidt, P. L., Barrett, G., & Barr, M. C. (2016). Integrating genomic data sets for knowledge discovery: An informed approach to Management of Captive Endangered Species. *International Journal of Genomics*, 2016, e2374610. <https://doi.org/10.1155/2016/2374610>
- Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., & Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, 2(1), 124–131. <https://doi.org/10.1038/s41559-017-0375-4>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., & Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, 19(1), 189. <https://doi.org/10.1186/s12859-018-2203-5>
- Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, 44(9), e89. <https://doi.org/10.1093/nar/gkw092>

- King, M. (1987). Monophyleticism and polyphyleticism in the Gekkonidae: A chromosomal perspective. *Australian Journal of Zoology*, 35, 641–654.
- Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernández, J., Saveljev, A. P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., & Godoy, J. A. (2022). Purging of deleterious burden in the endangered Iberian lynx. *Proceedings of the National Academy of Sciences of the United States of America*, 119(11), e2110614119. <https://doi.org/10.1073/pnas.2110614119>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAXML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., & Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, 102, 3–11. <https://doi.org/10.1016/j.ymeth.2016.02.020>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM* (arXiv:1303.3997). arXiv. <https://doi.org/10.48550/arXiv.1303.3997>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lisachov, A. P., Tishakova, K. V., Tsepilov, Y. A., & Borodin, P. M. (2019). Male meiotic recombination in the steppe agama, *Trapelus sanguinolentus* (Agamidae, Iguania, Reptilia). *Cytogenetic and Genome Research*, 157(1–2), 107–114. <https://doi.org/10.1159/000496078>
- Lisachov, A. P., Trifonov, V. A., Giovannotti, M., Ferguson-Smith, M. A., & Borodin, P. M. (2017). Immunocytological analysis of meiotic recombination in two anole lizards (Squamata, Dactyloidae). *Comparative Cytogenetics*, 11(1), 129–141. <https://doi.org/10.3897/CompCytogen.v11i1.10916>
- Liu, Y., Zhou, Q., Wang, Y., Luo, L., Yang, J., Yang, L., Liu, M., Li, Y., Qian, T., Zheng, Y., Li, M., Li, J., Gu, Y., Han, Z., Xu, M., Wang, Y., Zhu, C., Yu, B., Yang, Y., ... Gu, X. (2015). *Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneration. *Nature Communications*, 6(1), 1. <https://doi.org/10.1038/ncomms10033>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, 7(9), 999–1007. <https://doi.org/10.1111/eva.12193>
- Mochales-Riaño, G., Fontseré, C., de Manuel, M., Talavera, A., Burriel-Carranza, B., Tejedo-Cicuéndez, H., Algethami, R. H. M., Shobrak, M., Marques-Bonet, T., & Carranza, S. (2023). Genomics reveals introgression and purging of deleterious mutations in the Arabian leopard (*Panthera pardus nimr*). *iScience*, 26(9), 107481. <https://doi.org/10.1016/j.isci.2023.107481>
- Morin, P. A., Archer, F. I., Avila, C. D., Balacco, J. R., Bukhman, Y. V., Chow, W., Fedrigo, O., Formenti, G., Fronczek, J. A., Functammasan, A., Gulland, F. M. D., Haase, B., Peter Heide-Jorgensen, M., Houck, M. L., Howe, K., Misuraca, A. C., Mountcastle, J., Musser, W., Paez, S., ... Jarvis, E. D. (2021). Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Molecular Ecology Resources*, 21(4), 1008–1020. <https://doi.org/10.1111/1755-0998.13284>
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/MOLBEV/MSU300>
- Nielsen, S. V., Daza, J. D., Pinto, B. J., & Gamble, T. (2019). ZZ/ZW sex chromosomes in the endemic Puerto Rican leaf-toed gecko (*Phyllodactylus wirshingi*). *Cytogenetic and Genome Research*, 157(1–2), 89–97. <https://doi.org/10.1159/000496379>
- Peel, E., Silver, L., Brandies, P., Zhu, Y., Cheng, Y., Hogg, C. J., & Belov, K. (2022). Best genome sequencing strategies for annotation of complex immune gene families in wildlife. *GigaScience*, 11, giac100. <https://doi.org/10.1093/gigascience/giac100>
- Peterson, D. G., Stack, S. M., Healy, J. L., Donohoe, B. S., & Anderson, L. K. (1994). The relationship between synaptonemal complex length and genome size in four vertebrate classes (Osteichthyes, Reptilia, Aves, Mammalia). *Chromosome Research*, 2(2), 153–162. <https://doi.org/10.1007/BF01553494>
- Petraglia, M. D., Breeze, P. S., & Groucutt, H. S. (2018). Blue Arabia, green Arabia: Examining human colonisation and dispersal models. In *Geological Setting, Palaeoenvironment and Archaeology of the Red Sea* (pp. 675–683). [https://doi.org/10.1007/978-3-319-99408-6\\_30/COVER](https://doi.org/10.1007/978-3-319-99408-6_30/COVER)
- 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(6187), 1246752. <https://doi.org/10.1126/science.1246752>
- Pinto, B. J., Gamble, T., Smith, C. H., Keating, S. E., Havird, J. C., & Chiari, Y. (2023). The revised reference genome of the leopard gecko (*Eublepharis macularius*) provides insight into the considerations of genome phasing and assembly. *bioRxiv*, p. 2023.01.20.523807. <https://doi.org/10.1101/2023.01.20.523807>
- Pinto, B. J., Keating, S. E., Nielsen, S. V., Scantlebury, D. P., Daza, J. D., & Gamble, T. (2022). Chromosome-level genome assembly reveals dynamic sex chromosomes in Neotropical leaf-litter geckos (Sphaerodactylidae: *Sphaerodactylus*). *Journal of Heredity*, 113(3), 272–287. <https://doi.org/10.1093/jhered/esac016>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian Phylogenetics using tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/SYSBIO/SYY032>
- Roberts, P., Stewart, M., Alagaili, A. N., Breeze, P., Candy, I., Drake, N., Groucutt, H. S., Scerri, E. M. L., Lee-Thorp, J., Louys, J., Zalmout, I. S., Al-Mufarrej, Y. S. A., Zech, J., Alsharekh, A. M., al Omari, A., Boivin, N., & Petraglia, M. (2018). Fossil herbivore stable isotopes reveal middle Pleistocene hominin palaeoenvironment in 'Green Arabia'. *Nature Ecology & Evolution*, 2(12), 1871–1878. <https://doi.org/10.1038/s41559-018-0698-9>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Simó-Riudalbas, M., Tarroso, P., Papenfuss, T., Al-Sariri, T., & Carranza, S. (2018). Systematics, biogeography and evolution of *Asaccus galagheri* (Squamata, Phyllodactylidae) with the description of a new endemic species from Oman. *Systematics and Biodiversity*, 16(4), 323–339. <https://doi.org/10.1080/14772000.2017.1403496>
- Stange, M., Sánchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian isthmus.

- Systematic Biology*, 67(4), 681–699. <https://doi.org/10.1093/SYSBIO/SYY006>
- Tamar, K., Mitsi, P., & Carranza, S. (2019). Cryptic diversity revealed in the leaf-toed gecko *Asaccus montanus* (Squamata, Phyllodactylidae) from the Hajar Mountains of Arabia. *Journal of Zoological Systematics and Evolutionary Research*, 57(2), 369–382. <https://doi.org/10.1111/jzs.12258>
- Tang, S., Lomsadze, A., & Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research*, 43(12), e78. <https://doi.org/10.1093/nar/gkv227>
- Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2), 301–326. <https://doi.org/10.1534/genetics.112.148825>
- Uetz, P., Freed, P., Aguilar, R., Reyes, F., & Hošek, J. (Eds.). (2023). *The reptile database*. <https://www.reptile-database.org>
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, 27(5), 737–746. <https://doi.org/10.1101/gr.214270.116>
- Vinogradov, A. E. (1998). Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. *Cytometry*, 31(2), 100–109. [https://doi.org/10.1002/\(SICI\)1097-0320\(19980201\)31:2<100::AID-CYTO5>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0320(19980201)31:2<100::AID-CYTO5>3.0.CO;2-Q)
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Xie, H.-X., Liang, X.-X., Chen, Z.-Q., Li, W.-M., Mi, C.-R., Li, M., Wu, Z.-J., Zhou, X.-M., & Du, W.-G. (2022). Ancient demographics determine the effectiveness of genetic purging in endangered lizards. *Molecular Biology and Evolution*, 39(1), msab359. <https://doi.org/10.1093/molbev/msab359>
- Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D. N., de Manuel, M., Hernandez-Rodriguez, J., Lobon, I., Siegmund, H. R., Pagani, L., Quail, M. A., Hvilsom, C., Mudakikwa, A., Eichler, E. E., ... Scally, A. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, 348(6231), 242–245. <https://doi.org/10.1126/science.aaa3952>
- Yan, J., Tian, C., Lv, L., Bauer, A. M., & Zhou, K. (2014). Complete mitochondrial genome of the san Lucan gecko, *Phyllodactylus unctus* (Sauria, Gekkota, Phyllodactylidae), in comparison with *Tarentola mauritanica*. *Mitochondrial DNA*, 25(3), 202–203. <https://doi.org/10.3109/19401736.2013.796464>
- Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35(10), 1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Burriel-Carranza, B., Mochales-Riaño, G., Talavera, A., Els, J., Estarellas, M., Al Saadi, S., Urriago Suarez, J. D., Olsson, P. O., Matschiner, M., & Carranza, S. (2024). Clinging on the brink: Whole genomes reveal human-induced population declines and severe inbreeding in the Critically Endangered Emirati Leaf-toed Gecko (*Asaccus caudivolvulus*). *Molecular Ecology*, 00, e17451. <https://doi.org/10.1111/mec.17451>