Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

High-quality de novo genome assembly reveals historic population decline in the critically endangered Pternistis bicalcaratus ayesha

Harshita Rana
Ashoka University
Abdellah Ichen
Mohammed V University in Rabat
Giovanni Forcina
Universidad de Alcalá
Kirnalee Patel
Ashoka University
Jose Antonio Dávila
Instituto de Investigación en Recursos Cinegéticos IREC (CSIC-UCLM-JCCM)
Balaji Chattopadhyay
Ashoka University
Kritika M. Garg
kritikamgarg@iisermohali.ac.in

Indian Institute of Science Education and Research Mohali

Research Article

Keywords: Gamebirds, double-spurred spurfowl, Pternistis bicalcaratus ayesha, birds of Morocco, demographic history, PSMC, gene family evolution, keratinisation, Galliformes

Posted Date: June 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9588231/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Galliform birds represent one of the most threatened avian groups globally, yet genomic resources are limited for many species. The double-spurred spurfowl (*Pternistis bicalcaratus)* is widely distributed in West Africa, but its northern subspecies, *P. b. ayesha*, is restricted to northwestern Morocco and is classified as critically endangered. In this study, we describe the first high-quality reference genome for *P. b. ayesha* and investigate its evolutionary and demographic history.

Results

We generated Illumina short-read and PacBio continuous long-read and high-fidelity sequencing data from an adult male individual. Multiple genome assembly strategies, employing both hybrid and long-read approaches, were employed to obtain the most contiguous genome. The final genome assembly was 1.01 gigabases in size with 599 scaffolds and an N50 of 16.36 megabases, recovering 99 percent of conserved avian single-copy genes. Gene family evolution analyses revealed expansion in families related to cytoskeletal organisation, motor activity, muscle function, and keratinisation, consistent with adaptations observed in ground-dwelling game birds. The historical effective population size was reconstructed using the Pairwise Sequentially Markovian Coalescent model. Demographic reconstruction revealed that the effective population size peaked during MIS-7 and started declining in the Penultimate Glacial Period approximately 190 thousand years ago (kya), followed by a sustained long-term decline through the Late Pleistocene to the Holocene, with no evidence of recovery.

Conclusion

These results suggest that prolonged demographic decline is in concordance with the present low population size of *P. b. ayesha*. Together with its restricted distribution, it highlights the vulnerability of the subspecies to environmental and anthropogenic pressures. The genomic resource generated here provides a foundation for future population genomic studies and will be valuable for informing conservation strategies for this geographically isolated and threatened subspecies.

INTRODUCTION

Game birds are a diverse group of ground-dwelling wild birds that are widely hunted for food and sport. Most game birds belong to the order Galliformes and are characterised by heavy bodies, strong legs for running and foraging, and short, rounded wings for explosive short-distance flights, making them a popular choice for hunting [1, 2]. Beyond their value to humans, game birds play a vital role in terrestrial ecosystems as important seed dispersers and serve as a primary prey for many carnivores [2, 3]. Furthermore, they hold immense economic, cultural, and subsistence value globally as several species

are domesticated for food, while others are valued for hunting, traditional practices and ornamental keeping [4, 5, 6].

Despite this importance, Galliformes are one of the most rapidly declining avian orders globally, with more than one-quarter of extant species listed as 'threatened' [1, 7, 8]. Habitat loss, overhunting, agricultural intensification, and increasing climatic variability have been identified as major drivers of these declines [1, 5, 8, 9]. Their predominantly ground-dwelling nature, limited dispersal capabilities and specific habitat requirements [8] make them particularly vulnerable to climate change and environmental disturbances [1, 7, 8, 10].

Within this broader pattern of declining populations of many Galliformes species, the Double-spurred Spurfowl (*Pternistis bicalcaratus*) highlights how species-level conservation assessments can overlook declines in regionally restricted subspecies. Although the species is widely distributed across tropical West Africa and classified as 'Least Concern' by the IUCN [7], its northern subspecies, *P. b. ayesha* (Hartert, 1917), is endemic to Morocco [11, 12] and is recognised as Critically Endangered [11, 13, 14, 15]. This species was historically placed within the genus *Francolinus* but is currently assigned to *Pternistis* following taxonomic revisions based on morphological and molecular evidence [16, 17, 18]. *Pternistis bicalcaratus* consists of three recognised subspecies that are generally similar in overall appearance with slight variations in plumage colouration and patterning. *P. b. ayesha* is similar (1.0% cytochrome *b* divergence; CYTB – 1143 base pairs) but distinguished as the more heavily patterned form, with faint vermiculation, small arrow-shaped buff marks on its belly feathers and slightly more rufous lower neck [17]. This subspecies represents an isolated population separated from other populations by the Sahara Desert, making *P. bicalcaratus* one of the few bird taxa occurring naturally in both north and south of the Sahara [17, 19]. *P. bicalcaratus* occurs in almost the rest of the sub-Saharan region with the exception of *P. b. ayesha* that expanded its range to Morocco in North Africa. It occupies a highly restricted and fragmented range in north-western Morocco and has experienced a marked contraction from its historical distribution [12] (Fig. 1a). This decline is largely attributed to hunting pressure and habitat degradation [11, 14, 20], with increasing climatic variability in North Africa likely imposing additional demographic stress. Similar distribution patterns are also seen in a few other Afrotropical species, including Marsh Owl (*Asio capensis tingitanus*), Black-crowned Tchagra (*Tchagra senegalus cucullatus*), and Brown-throated Martin (*Riparia paludicola mauritanica*), along with formerly occurring taxa such as Dark-chanting Goshawk (*Melierax metabates theresae*), Helmeted Guineafowl (*Numida meleagris sabyi*), and Arabian Bustard (*Ardeotis arabs lynesi*) [21].

This study presents the first high-quality genome assembly of the critically endangered Double-spurred spurfowl subspecies, *P. b. ayesha*, generated using a hybrid approach that combines Illumina short reads, PacBio long reads (continuous long reads, CLR), and PacBio HiFi reads. The assembled genome is approximately 1 Gb in size, which is consistent with other game bird genomes. Assembly metrics, including the number of scaffolds, scaffold N50, and Benchmarking Universal Single-Copy Orthologs (BUSCO) scores, indicate high genome contiguity and completeness. This genome provides a valuable resource for the genomic characterisation of this endangered species. Comparative analyses with other

bird genomes were carried out to explore genomic features of *P. b. ayesha* in an evolutionary framework. Additionally, demographic history analysis was conducted to assess the impact of past climate on this species. The findings of this study establish a foundation for future research on population genetic structure, conservation strategies, and the identification of adaptive genetic mechanisms in response to environmental challenges faced by the sub-species *ayesha*.

MATERIALS AND METHODS

Sampling and sequencing

Liver tissue from an adult male *P. b. ayesha* was obtained opportunistically from licensed hunters through collaborators in central Morocco. The ethical clearance for the project was obtained from the Faculté des Sciences de Rabat, Université Mohammed V (clearance number: PR002-0426). We extracted high-molecular-weight DNA with the Nanobind PanDNA kit (PacBio) following the manufacturer's instructions at the Laboratory of Genetics of the Institute for Game and Wildlife Research (Ciudad Real, Spain). Library preparation and sequencing were performed by Novogene Europe (Cambridge, UK). We generated short-read data using the Illumina NovaSeq platform and long-read and HiFi data using the Pacific Biosciences Sequel II platform based on single-molecule real-time (SMRT) sequencing technology.

Data processing

We obtained ~ 427 million Illumina short reads (150 bp paired-end), 8.7 million PacBio CLR reads, and 1.4 million PacBio HiFi reads, achieving ~ 64× Illumina and ~ 99× PacBio sequencing depth.

We assessed the quality of Illumina raw sequencing reads using FastQC ver. 0.11.9 [22]. We trimmed Illumina short reads to remove adapter sequences, low-quality bases (Phred score < 20), and reads shorter than 36 bp using Trimmomatic ver. 0.39 [23]. We converted PacBio CLR and HiFi reads from BAM to FASTQ format using samtools bam2fq ver. 1.10 [24, 25] and concatenated the resulting files using the cat command for downstream analyses.

Genome assembly

To generate a high-quality *de novo* genome assembly, we evaluated a total of seven genome assemblers. We first implemented a hybrid assembly strategy that combined Illumina short reads with PacBio long reads (CLR and HiFi) to leverage both high base accuracy and long-range contiguity. For this approach, we used five hybrid assemblers: MaSuRCA ver. 4.1.0 [26, 27], WenganA ver. 0.2 [28], WenganD ver. 0.2 [28], HASLR ver. 0.8a1 [29], and DBG2OLC ver. 20200723 [30]. Given the high accuracy and contiguity of PacBio HiFi reads, we also assessed whether assemblies generated solely from long-read data could

match or outperform hybrid assemblies. For this purpose, we used two long-read assemblers: Flye ver. 2.9.2 [31] and Hifiasm ver. 0.23.0-r [32].

For all hybrid assembly approaches, we concatenated PacBio HiFi reads and PacBio continuous long reads (CLR) and used the combined dataset as the long-read input together with Illumina short reads. Illumina paired-end short reads were used in their raw form for MaSuRCA, HASLR, and DBG2OLC whereas adapter-trimmed and quality-filtered reads were used for WenganA and WenganD hybrid assemblies. All assemblers were run using default or software-recommended parameters unless stated otherwise.

MaSuRCA uses a combination of de Bruijn graph and overlap-layout-consensus approaches and performs all steps from data preprocessing to final assembly [26, 27]. We performed the assembly in Flye mode, which is optimised for incorporating long-read data. Wengan is a hybrid genome assembler that uses long reads to scaffold a short-read assembly backbone without performing all-vs-all long-read alignments, thereby reducing computational requirements while maintaining assembly accuracy [28]. We ran Wengan in two modes: WenganD and WenganA. In WenganD mode, we first generated contigs from cleaned Illumina short reads using DiscovarDenovo [33] and then performed hybrid assembly. In WenganA mode, short-read assembly was performed using Abyss2 [34] instead.

DBG2OLC genome assembler first constructs accurate contigs from short reads using a de Bruijn graph approach and then anchors long, error-prone reads to these contigs to generate the final assembly, without requiring all-vs-all long-read alignments, thereby improving computational efficiency [30]. We followed the recommended three-step DBG2OLC pipeline. We first generated contigs from Illumina short reads using SparseAssembler [35]. We improved the quality of the short-read contigs by fine-tuning SparseAssembler parameters, including NodeCovTh, EdgeCovTh, and k, and by reloading previous computations using the LD = 1 flag to refine the de Bruijn graph, following recommendations provided in the DBG2OLC user manual. We then combined the resulting contigs with concatenated PacBio long reads and used them as input for DBG2OLC during the overlap and layout step. Finally, we polished the draft assembly using Sparc [36] to generate the final consensus.

HASLR genome assembler constructs a backbone graph from long reads and anchors accurate short-read contigs to guide assembly [29]. We ran HASLR by specifying the input read type as PacBio (-x pacbio), setting the estimated genome size to 1.1 Gb, and keeping all other parameters at their default values.

Flye is a *de novo* long-read assembler that constructs repeat graphs from raw reads and iteratively refines contigs through repeat resolution and polishing [31]. As Flye requires different PacBio read types to be processed separately, we performed two independent assemblies: one using only PacBio HiFi reads in --pacbio-hifi mode, and another using the concatenated PacBio CLR and HiFi dataset in --pacbio-raw mode, with all parameters set to default.

We also used Hifiasm, a genome assembler specifically designed for PacBio HiFi data that constructs phased assembly graphs directly from highly accurate HiFi reads to generate haplotype-resolved assemblies without prior error correction [32]. PacBio HiFi reads were first converted from BAM to FASTQ format using samtools bam2fq ver. 1.10 [24, 25], and the genome was assembled in primary mode using default parameters.

Quality assessment

We assessed the quality of all genome assemblies using QUAST v5.0.2 [37] under default settings. We also evaluated genome completeness using BUSCO v6.0.0 [38] with the Aves ortholog database ver. 12 [39], which comprises 6,251 conserved single-copy genes. We selected the assembly generated with MaSuRCA hybrid assembler (IBDC submission ID: INGCA000001241; NCBI BioProject ID PRJNA1459051; assembly accession no. JBXUZW000000000) for all downstream analyses because of its highest contiguity, lowest scaffold count, and high completeness.

Repeat masking and genome annotation

We soft-masked the MaSuRCA genome using RepeatMasker v4.1.2-p1 [40]. Repeats from the Chicken (*Gallus gallus*) genome was used for initial identification and masking. To account for phylogenetic proximity, we also performed repeat masking using the Quail (*Coturnix coturnix*) repeat library. The masked genome generated with the Chicken repeat library was subsequently used for genome annotation. We annotated the genome using the BRAKER2 pipeline [41], which integrates GeneMark-EP+ [42] and AUGUSTUS [41, 43, 44] in protein-supported mode [41, 42, 45, 46, 47, 48]. We provided protein evidence from the vertebrate dataset OrthoDB protein database, version 11 [49] to generate hints for annotation using only protein sequence data. The predicted proteome was functionally annotated using both Pannzer2 [50] web-based server and eggNOGMapper ver. 4.5 [51, 52]. We also compared the predicted proteome to other bird proteomes using OrthoVenn3 [53]. We used proteomes of seven avian species Collared Flycatcher (*Ficedula albicollis*), Australian Zebra Finch (*Taeniopygia guttata*), Mallard Duck (*Anas platyrhynchos*), Turkey (*Meleagris gallopavo*), Indian Peafowl (*Pavo cristatus*), Chicken and Quail from the available database in Orthovenn3. The *E* value was set to 1e − 5 and the inflation value to 1.5 to identify orthologous genes using the Ortho Finder algorithm. We also performed gene family fluctuations using CAFE5 [54] as a part of OthoVenn3. CAFE uses a statistical framework to identify the gene family expansion and contraction on a phylogenetic tree [54]. The divergence time between *G. gallus* and *A. platyrhynchos*, *G. gallus* and *C. japonica*, and *F. albicollis* and *T. guttata* were obtained from TimeTree 5 [55] and set to 83.37, 37.90 and 25.22 million years ago respectively. GO enrichment analyses were performed within OrthoVenn3, and significance was determined using Benjamini-Hochberg false discovery rate correction (FDR < 0.05) [56].

Demographic History

We reconstructed the demographic history of *P. b. ayesha* using Pairwise Sequentially Markovian Coalescent (PSMC) model [57]. We mapped the genome to sex chromosomes of Chicken (GCF_016700215.2) using Satsuma ver. 2.0 [58] to identify scaffolds corresponding to Z Chromosome. We aligned the cleaned Illumina short reads to the whole genome using BWA-MEM ver. 0.7.17 [59, 60]. We used samtools ver. 1.10 [24, 25, 61] to sort the BAM file, remove duplicates, and estimate the depth of coverage. We then used mpileup from bcftools ver. 1.10.2 [24, 25, 61] to remove the scaffolds mapping to Z chromosome and identify variable sites using autosomes with the following parameters -Q 30 -q 30 -C 50 -A -Ou. We obtained consensus sequences using vcfutils.pl [24, 25, 61], with the minimum and maximum read depth set to one-third and twice the average genome coverage, respectively [62, 63] and generated a diploid.psmcfa file using the fq2psmcfa tool from PSMC package [57] for downstream processing. PSMCs were performed with the following sets of parameters: *-N*30 *-t*5 *-r*5 *-p* '4 + 30*2 + 4 + 6 + 10', *-N*30 *-t*5 *-r*5 *-p* '2 + 2 + 30*2 + 4 + 6 + 10' and *-N*30 *-t*5 *-r*5 *-p* '1 + 1+1 + 1 + 30*2 + 4 + 6 + 10' (*N* = number of iterations, *t* = the upper limit of time to most recent common ancestor, *r* = ratio of the scaled mutation rate and the recombination rate, and *p* = number of free atomic time intervals) following previous studies on birds and to avoid spurious peaks in effective population size [57, 62, 63, 64, 65]. We ran 30 iterations for optimisation of parameters and ran 100 bootstrap replicates for determining uncertainty in our estimates. We used a generation time of 2 years and a mutation rate of 1.9×10^{-9} per base per year [66] to plot the results from PSMC analyses.

RESULTS

Genome assembly and quality assessment

The estimated genome size of *P. b. ayesha* based on k-mer analysis was ~ 1.01 Gb, consistent with other galliform genomes (Fig. 1b; Table 1). Among all assemblies, the MaSuRCA assembly demonstrated the highest contiguity (N50 = 16.36 Mb), the lowest number of scaffolds (599), and high completeness as measured by BUSCO (99% of conserved single-copy orthologs), and was therefore selected for downstream analyses (Fig. 1b and 1c; Table 1). The high contiguity and completeness of the Flye assembly (N50 = 12.87 Mb, 1,067 scaffolds, 99.1% BUSCO) further demonstrate that long-read sequencing alone can produce nearly complete genomes (Table 1; Fig. 1c). Other assemblies exhibited lower contiguity, higher scaffold counts, and/or reduced completeness (Table 1; Fig. 1c).

Table 1
Comparison of genome assemblies for Pternistis bicalcaratus ayesha

Assembler	Genome size (Gb)	No. of Scaffolds	N50 (Mb)	BUSCO (%)
MaSuRCA	1.01	599	16.36	99
Hifiasm	1.09	923	4.96	98.9
Flye	1.03	1067	12.87	99.1
Flye (HiFi only)	1.6	8833	0.33	49.3
WenganD	1.01	1193	3.67	98.8
WenganA	0.95	2285	1.85	93.4
DBG2OLC	1.05	1008	8.1	94.4
HASLR	0.03	20474	0.026	25.6

Repeat masking

The MaSuRCA assembly was soft-masked using RepeatMasker 4.1.2. We tested repeat libraries from both the Chicken and Quail genomes and found identical results. Only 5.64% of the genome consisted of repeat elements, with a majority of them being Long interspersed nuclear elements (Table 2).

Table 2
Percentage of repeat elements in the MaSuRCA assembly of Pternistis bicalcaratus ayesha

Repeat element	Percentage
Short interspersed nuclear elements (SINEs)	0.05
Long interspersed nuclear elements (LINEs)	4.01
DNA transposons	0.11
Unclassified	0.03
Simple repeats	1.16
Low complexity	0.28

Genome annotation and gene family fluctuations

We identified 23,388 protein-coding genes, out of which 9,401 could be functionally annotated using Pannzer2 and 11,473 using eggNOGMapper (Table S1 and Table S2). The phylogenetic tree generated within OrthoVenn3 using orthologous gene families was consistent with the established phylogeny of the Phasianidae (Fig. 2) [67]. Further we observed expansion in 75 gene families and contraction within 567 gene families in *P. b. ayesha* relative to *F. albicollis*, *T. guttata*, *A. platyrhynchos*, *M. gallopavo*, *G. gallus*, *P. cristatus*, and *C. japonica* (Fig. 2). Gene Ontology (GO) enrichment analysis of these families identified

79 significantly enriched terms in the expanding gene families and 274 significantly enriched terms in the contracting gene families after false discovery rate correction (FDR < 0.05) (Table S3 and Table S4). We observed significant expansion in families related to cytoskeletal organisation and motor activity (e.g., structural constituent of cytoskeleton, motor activity, actin filament organisation), immune processes (e.g., immune response and adaptive immune response), sensory perception (e.g., olfactory receptor activity and visual perception), muscle-related processes, and keratinisation (Fig. 2; Table S3). We also observed significant contraction in families associated with immune responses, sensory perception of smell and vision (Fig. 2; Table S4).

Demographic history

Our PSMC reconstruction of the *P. b. ayesha* from Morocco showed a peak in effective population size (Ne) during the Marine Isotope Stage (MIS) 7 (~ 200 kya), where N_e reached approximately 2.3 million individuals. Following this peak, N_e declined steadily, throughout MIS 6, decreasing to ~ 1.3 million during the Last Interglacial (~ 130 kya) and further to only ~ 28,000 individuals during the Late Pleistocene (Fig. 3; Table 3). By the end of Pleistocene (~ 10 kya), the effective population size reached its lowest inferred value (~ 11,000 individuals) and showed no signs of recovery (Fig. 3; Table 3). We do not see any evidence of population expansion at Last Glacial Maxima (LGM). This declining trend in N_e values was consistent across the three PSMC parameter settings used for reconstructing demographic history (Fig. 3).

Table 3
Estimated effective population size (N_e) at key time points inferred using PSMC for *P. b. ayesha*.

Time (kya)	Time period	Estimated N_e
~ 190	Beginning of MIS-6	2,307,259
~ 130	Last Interglacial (LIG)	1,376,358
~ 20–30	Late Pleistocene	28,277
~ 10	End of the Pleistocene	11,686

DISCUSSION

Game birds are among the most heavily exploited birds in the world, yet genomic resources remain limited for many threatened species [1, 9]. The availability of a high-quality reference genome for *P. b. ayesha* addresses a significant gap in genomic data for North African galliform birds. Although *P. bicalcaratus* is listed as Least Concern at the species level, its Moroccan subspecies (*ayesha*) is geographically isolated making it one of the few sub-Saharan bird species with natural populations both north and south of the Sahara (Fig. 1a). This isolation renders the subspecies both evolutionarily distinct and at an elevated risk of extinction. A reference genome enables studies on population structure,

genetic diversity, inbreeding and adaptive variation. For small and fragmented populations, it is particularly important to understand the hidden genetic erosion before developing conservation strategies. The genomic resource generated here provides a foundational baseline for future population genomic research and conservation efforts tailored to this subspecies.

We report the first high-quality reference genome for the critically endangered subspecies *P. b. ayesha*. Out of seven assemblers tested, the hybrid MaSuRCA assembly generated the most complete and contiguous genome, with an N50 of 16.36 Mb and 99% BUSCO completeness (Fig. 1a and b; Table 1). While the Flye long-read-only assembly also performed well, with an N50 of 12.87 Mb and 99.1% BUSCO completeness (Fig. 1c; Table 1), the hybrid approach generated a substantially lesser number of scaffolds and thus led to better scaffold continuity and overall quality. Using a combination of short and long reads along with benchmarking with multiple assemblers allowed for a robust genome assembly. Benchmarking studies in eukaryotic genomes indicate that long-read assemblers such as Flye perform effectively, particularly with error-prone long-read data [68]. However, assembly performance depends on sequencing strategy and evaluation criteria and no single approach consistently performs best across all datasets [68, 69]. In vertebrates, large-scale initiatives such as the Vertebrate Genomes Project highlight the importance of long-read sequencing for generating high-quality assemblies [70]. Additionally, highly accurate reads such as PacBio HiFi further enhance contiguity and the resolution of complex genomic regions [71]. In alignment with these findings, the Flye assembly in this study achieved high contiguity, while the hybrid MaSuRCA assembly produced a higher N50 and fewer scaffolds (Table 1; Fig. 1c), indicating improved scaffold continuity and overall assembly quality. These results suggest that integrating multiple sequencing strategies may be particularly advantageous for assembling complex vertebrate genomes.

The final assembly size, about 1.01 Gb is consistent with the size of other Galliform genomes, which are usually around 1Gb. The high BUSCO completeness also confirms that the assembly covers most of the conserved avian genes (Fig. 1b and 1c). Relatively low repeat content (5.64%; Table 2) is also in line with other published bird genomes, which show that birds have a compact genome with fewer repeat elements [72]. Overall, this assembly provides a strong genomic foundation for future evolutionary and conservation studies.

Gene family evolution in *Pternistis bicalcaratus* (*P. b. ayesha*) revealed expansion of gene families associated with cytoskeletal organisation (Table S3), which are involved in muscle contraction and force generation. This is important for *P. b. ayesha*, a ground dwelling game bird that relies on muscle strength and endurance for rapid running and short bursts of flight [73]. This is further supported by expansion in gene families associated with motor activity and muscle-related processes (Fig. 2; Table S3). Interestingly, we also observed expansion in gene families associated with keratinisation. This is particularly relevant for *P. b. ayesha*, commonly known as the double-spurred spurfowl, which is characterised by keratinised leg spurs. The spurs are a distinctive secondary sexual character in males and may play a role in mate-choice [74, 75, 76, 77] (Table S3). In contrast, several categories related to immune response and sensory perception were observed in both expanding and contracting gene

families, although they were more frequently represented among contracted families (Table S3 and S4). These observations suggest lineage-specific evolution of several functional gene families in *Pternistis bicalcaratus* (*P. b. ayesha*), highlighting the need for further investigation.

Demographic reconstruction using PSMC under all three parameter settings revealed a consistent pattern in the historical population dynamics of *P. b. ayesha*, with effective population size (N_e) reaching a peak during MIS 7 (~ 200 kya), followed by a steady decline (Fig. 3). This pattern aligns with major Pleistocene climatic transitions, particularly attributed to shifts in environmental conditions from a relatively warm and moist period in MIS-7 (~ 243 – 190 kya) to severe glaciation in MIS-6 (~ 190 – 130 kya), also known as the Penultimate Glacial Period [78, 79]. These climatic shifts during the the Middle to Late Pleistocene were characterised by alternating humid and arid phases, which strongly influenced vegetation cover and habitat structure in Africa. *P. b. ayesha*, is a ground-dwelling galliform associated with dense vegetation structure, including cork oak woodlands, thick shrub cover, and areas near cultivated land and water sources, as documented for extant populations in Morocco [14, 20].

Paleoenvironmental records indicate that North African hydroclimate over the last ~ 200 kyr was characterised by alternating humid and arid phases driven by monsoon variability. The humid periods during MIS 7 and early MIS 6 were associated with reduced dust input, increased vegetation cover, and expansion of habitable landscapes across the Sahara [80, 81]. Long-term vegetation records also suggest that much of the last ~ 200 kyr was predominantly arid with widespread C4 vegetation, and there were only brief wetter phases of C3 vegetation expansions in the Sahara/Sahel region [82]. Under these conditions, structurally suitable habitats would likely have expanded northward into North-West Africa during the MIS 7, increasing landscape connectivity rather than forming strict barriers. Similar dynamics have been inferred from human occupation patterns in North Africa, where humid phases are associated with increased dispersal opportunities across the Sahara, while arid phases correspond to reduced connectivity and more restricted distributions [82, 83]. This supports the interpretation that these humid intervals corresponded to increased ecological connectivity. The peak in N_e around ~ 200 kya (Fig. 3) is therefore consistent with maximum habitat availability and connectivity in North Africa. In contrast, the subsequent decline in N_e coincides with increasing aridity during MIS 6 (Fig. 3), a period characterised by high C4 vegetation dominance indicating arid conditions. This time period is associated with rising dust flux and cooler sea surface temperatures in the tropical Atlantic, indicating more severe arid conditions than those observed during the Last Glacial Maximum [80, 81, 82]. The re-establishment of these strong arid conditions would have fragmented dense woodland-shrub habitats and reduced connectivity, particularly at the northern range margin. For a species dependent on dense cover and access to water, this habitat fragmentation likely contributed to the observed decline in effective population size.

The increase in the population size around 200 kya (Fig. 3) could also possibly be due to subspecies divergence as population split can often result in increase in effective population size estimates in PSMC analysis [57]. However, based on the standard avian mitochondrial substitution rate of ~ 2% sequence divergence per million years [84] the observed 1% sequence divergence between *P. b. ayesha* and *P.*

bicalcaratus pushes the divergence time to ~ 500 kya [17]. In contrast, demographic inference based on the nuclear genome in this study indicates a more recent population decline beginning around ~ 190 kya (Fig. 3).

Notably, *P. b. ayesha* shows a long-term decline in N_e, reaching its lowest inferred level (~ 11,000 inferred individuals by ~ 10 kya) by the end of the Pleistocene (Fig. 3; Table 3). Small effective population sizes can reduce genetic diversity and adaptive potential, making a species more susceptible to environmental change. Similar long-term declines in N_e have been reported in several threatened avian species, particularly during the Last Glacial Period [62]. For example, currently endangered species such as the Crested Ibis (*Nipponia nippon*) and Gray Crowned Crane (*Balearica regulorum)* show substantial reductions in N_e from large ancestral populations to markedly lower N_e values during Late Pleistocene, with the Crested Ibis declining to ~ 2000 inferred individuals by ~ 10 kya [62]. Likewise, the vulnerable Brown Mesite (*Mesitornis unicolor*) shows a relatively low but stable historical N_e followed by a sharp decline to ~ 23000 inferred individuals during the Late Pleistocene [62]. These findings suggest that many species now listed threatened have undergone prolonged demographic declines rather than only recent reductions. Comparable trends have also been observed in mammals, such as the currently critically endangered Sumatran Rhinoceros (*Dicerorhinus sumatrensis*), which also shows a sustained decline in N_e from the Middle to Late Pleistocene, with no clear evidence of population recovery [85].

Overall, *P. b. ayesha* also shows a sustained long-term decline in N_e with no clear evidence of population expansion or recovery (Fig. 3), highlighting its potential vulnerability, which may have led to its present low population size. Given the restricted distribution of this subspecies in northwestern Africa, these results highlight the need for continued monitoring and conservation efforts for *P. b. ayesha*. Further population genomic studies incorporating multiple individuals across the range of this subspecies would help to better resolve recent demographic history and inform conservation strategies

CONCLUSION

This study presents the first high-quality reference genome for the critically endangered subspecies *Pternistis bicalcaratus ayesha*, providing an essential genomic resource for a geographically isolated and understudied galliform lineage. By evaluating multiple assembly strategies, we demonstrate that a hybrid approach yields a highly contiguous and complete genome, establishing a robust foundation for downstream evolutionary and conservation analyses.

Comparative genomic analyses reveal lineage-specific patterns of gene family evolution, particularly in functions associated with locomotion, structural integrity, and adaptation to a ground-dwelling lifestyle. Demographic reconstruction further indicates a long-term population decline beginning in the middle to late Pleistocene, likely driven by climatic shifts that altered habitat availability and connectivity in North Africa. Consistent with patterns observed in other threatened taxa, such prolonged demographic decline may have contributed to reduced population size in this subspecies, further emphasizing its conservation concern. Together, these findings highlight both the evolutionary distinctiveness and the

vulnerability of *P. b. ayesha*. The genomic resources generated here will facilitate future studies on population structure, genetic diversity, and adaptive potential, and will be critical for informing conservation strategies aimed at preserving this isolated and threatened subspecies.

Declarations

Ethics approval and consent to participate:

The study was reviewed and approved by the Animal Ethics Committee of the Faculty of Sciences of Rabat, Mohammed V University (Approval No. PR002-0426).

Availability of data and materials:

The genome assembly is available on the Indian Biological Data Centre (IBDC submission ID: INGCA000001241) and NCBI BioProject ID PRJNA1459051 under the accession JBXUZW000000000.

Competing interests:

The authors declare no conflicts of interest.

Funding:

This work was supported by the Department of Biotechnology grant to KMG (BT/PR42830/BRB/10/1995/2021). KMG acknowledges the support DBT-Ramalingaswami Fellowship (BT/HRD/35/02/2006). BC acknowledges the support of Trivedi School of Biosciences. GF was partially supported by a fellowship of the autonomous community of Madrid (“César Nombela” Research Talent Attraction, Ref. 2024-T1/SAL-GL-31424).

Authors' contributions:

H.R. performed genome assembly, genome annotation, gene family evolution analyses and demographic inference, and wrote the manuscript; AI provided the sample for sequencing and reviewed the manuscript; GF conceived the study, sample processing, data generation, and reviewed the manuscript; KP contributed to the data analysis for demographic inference and reviewed the manuscript; JAD provided the sample for sequencing and reviewed the manuscript; BC co-supervised the research, contributed to data interpretation and reviewed the manuscript; K.M.G. conceived and designed the study, supervised the research, contributed to data analysis and interpretation, and reviewed the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

The authors thank Mr. Durbadal Chatterjee for help with PSMC analysis.

References

1. McGowan PJ, Owens LL, Grainger MJ. Galliformes science and species extinctions: what we know and what we need to know. Anim Biodivers Conserv. 2012;35(2):321–31. https://doi.org/10.32800/abc.2012.35.0321.
2. Crespo R, França MS, Fenton H, Shivaprasad HL, Galliformes. and columbiformes. InPathology of wildlife and zoo animals 2018 Jan 1 (pp. 747–773). Academic Press. https://doi.org/10.1016/B978-0-12-805306-5.00031-6
3. Anderson G, Holmes C. Galliformes: species, varieties, and behavior. Gamebird Med Manage. 2022;11:13–35. https://doi.org/10.1002/9781119712244.ch3.
4. Geldenhuys G, Hoffman LC, Muller N, Gamebirds. A sustainable food source in Southern Africa? Food Secur. 2013;5(2):235–49. https://doi.org/10.1007/s12571-013-0245-0.
5. Tian S, Xu J, Li J, Zhang Z, Wang Y. Research advances of Galliformes since 1990 and future prospects. Avian Res. 2018;9(1):32. https://doi.org/10.1186/s40657-018-0124-7.
6. Wang MS, Thakur M, Peng MS, Jiang YU, Frantz LA, Li M, Zhang JJ, Wang S, Peters J, Otecko NO, Suwannapoom C. 863 genomes reveal the origin and domestication of chicken. Cell Res. 2020;30(8):693–701. https://doi.org/10.1038/s41422-020-0349-y.
7. IUCN. The IUCN Red List of Threatened Species. Version 2025-2 [Internet]. 2025 [cited 2025 Dec 14]. Available from: https://www.iucnredlist.org
8. Wang B, Ye W, Xu Y, Zhong X, Zhang J, Yang N, Yang B, Zhou C. Climate change affects Galliformes taxonomic, phylogenetic and functional diversity indexes, shifting conservation priority areas in China. Divers Distrib. 2023;29(3):409–22. https://doi.org/10.1111/ddi.13667.
9. Keane A, Brooke MD, McGowan PJ. Correlates of extinction risk and hunting pressure in gamebirds (Galliformes). Biol Conserv. 2005;126(2):216–33. https://doi.org/10.1016/j.biocon.2005.05.011.
10. Fuller RA, Carroll JP, McGowan PJ, editors. Partridges, quails, francolins, snowcocks, guineafowl, and turkeys: status survey and conservation action plan 2000–2004. IUCN; 2000.
11. Thévenot M, Vernon R, Bergier P. The birds of Morocco: an annotated checklist. British Ornithologists' Union; 2003.
12. Bergier P, Thevenot M, Qninba A, Houllier JR. Oiseaux du Maroc. SEOF; 2022.
13. McGowan PJ. Partridges, quails, francolins, snowcocks, and guineafowl: status survey and conservation action plan, 1995–1999. IUCN; 1995.
14. Hanane S, Alahyane N, Magri N, El Agbani MA, Qninba A. Habitat preference of the sole wild population of *Francolinus bicalcaratus ayesha* in the Palearctic: implications for conservation and management. Revue d'Écologie. 2016;71(3):288–97.
15. Iallaten H, Saadi K, Cherkaoui SI, Bazairi H, Qninba A. New record of *Pternistis bicalcaratus ayesha* at Korifla valley (Zaër, North-West Morocco). Éléments d'Ornithologie Marocaine. 2025:2.
16. Crowe TM, Bowie RC, Bloomer P, Mandiwana TG, Hedderson TA, Randi E, Pereira SL, Wakeling J. Phylogenetics, biogeography and classification of, and character evolution in, gamebirds (Aves:

Galliformes): effects of character exclusion, data partitioning and missing data. Cladistics. 2006;22(6):495–532. https://doi.org/10.1111/j.1096-0031.2006.00120.x.

17. Mandiwana-Neudani TG, Little RM, Crowe TM, Bowie RC. Taxonomy, phylogeny and biogeography of African spurfowls Galliformes, Phasianidae, Phasianinae, Coturnicini: *Pternistis spp*. Ostrich. 2019;90(2):145–72. https://doi.org/10.2989/00306525.2019.1584925.
18. Mandiwana-Neudani TG, Little RM, Crowe TM, Bowie RC. Taxonomy, phylogeny and biogeography of 'true'francolins: Galliformes, Phasianidae, Phasianinae, Gallini; Francolinus, Ortygornis, *Afrocolinus gen. nov., Peliperdix* and *Scleroptila spp*. Ostrich. 2019;90(3):191–221. https://doi.org/10.2989/00306525.2019.1632954
19. Moreau RE. The bird faunas of Africa and its islands. 1966.
20. Hanane S, Magri N. Towards understanding habitat use patterns of an endangered galliform subspecies, *Pternistis bicalcaratus ayesha*: implications for conservation. Ostrich. 2017;88(3):217–24. https://doi.org/10.2989/00306525.2017.1287134.
21. Isenmann P, Thevenot M. Endemism in terrestrial bird species of continental North Africa. Go-South Bull. 2020;17:161–84.
22. Andrews S, FastQC:. A Quality Control Tool for High Throughput Sequence Data. 2010.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.
24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.
25. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008. https://doi.org/10.1093/gigascience/giab008.
26. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–77. https://doi.org/10.1093/bioinformatics/btt476.
27. Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome research. 2017;27(5):787 – 92. https://doi.org/10.1101/gr.213405.116
28. Di Genova A, Buena-Atienza E, Ossowski S, Sagot MF. Efficient hybrid de novo assembly of human genomes with WENGAN. Nat Biotechnol. 2021;39(4):422–30. https://doi.org/10.1038/s41587-020-00747-w.
29. Haghshenas E, Asghari H, Stoye J, Chauve C, Hach F. HASLR: fast hybrid assembly of long reads. Iscience. 2020;23(8). https://doi.org/10.1016/j.isci.2020.101389.
30. Ye C, Hill CM, Wu S, Ruan J, Ma Z. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep. 2016;6(1):31900. https://doi.org/10.1038/srep31900.

31. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. https://doi.org/10.1038/s41587-019-0072-8.
32. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5. https://doi.org/10.1038/s41592-020-01056-5.
33. Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, Sogoloff B, Tabbaa D, Williams L, Russ C, Nusbaum C. Comprehensive variation discovery in single human genomes. Nat Genet. 2014;46(12):1350–5. https://doi.org/10.1038/ng.3121.
34. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 2017;27(5):768–77. https://doi.org/10.1101/gr.214346.116.
35. Ye C, Ma ZS, Cannon CH, Pop M, Yu DW. SparseAssembler: de novo Assembly with the Sparse de Bruijn Graph. arXiv preprint arXiv:1106.2603. 2011. https://doi.org/10.48550/arXiv.1106.2603
36. Ye C, Ma ZS. Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads. PeerJ. 2016. https://doi.org/10.7717/peerj.2016. 4:e2016.
37. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142–50. https://doi.org/10.1093/bioinformatics/bty266.
38. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54. https://doi.org/10.1093/molbev/msab199.
39. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2019;47(D1):D807–11. https://doi.org/10.1093/nar/gky1053.
40. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. http://www.repeatmasker.org. 2013–2015.
41. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. NAR Genomics Bioinf. 2021;3(1):lqaa108. https://doi.org/10.1093/nargab/lqaa108.
42. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinf. 2020;2(2):lqaa026. https://doi.org/10.1093/nargab/lqaa026.
43. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7(1):62. https://doi.org/10.1186/1471-2105-7-62.
44. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008;24(5):637–44.

https://doi.org/10.1186/1471-2105-7-62.

45. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60. https://doi.org/10.1038/nmeth.3176.
46. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33(20):6494–506. https://doi.org/10.1093/nar/gki937.
47. Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 2012;40(20):e161. https://doi.org/10.1093/nar/gks708.
48. Gotoh O, Morita M, Nelson DR. Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment. BMC Bioinformatics. 2014;15(1):189. https://doi.org/10.1186/1471-2105-15-189.
49. Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023;51(D1):D445–51. https://doi.org/10.1093/nar/gkac998.
50. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 2018;46(W1):W84–8. https://doi.org/10.1093/nar/gky350.
51. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44(D1):D286–93. https://doi.org/10.1093/nar/gkv1248.
52. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825–9. https://doi.org/10.1093/molbev/msab293.
53. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 2023;51(W1):W397–403. https://doi.org/10.1093/nar/gkad313.
54. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020;36(22–23):5516–8. https://doi.org/10.1093/bioinformatics/btaa1022.
55. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 2022;39(8):msac174. https://doi.org/10.1093/molbev/msac174.
56. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol). 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
57. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011 8;475(7357):493-6. https://doi.org/10.1038/nature10231

58. Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics. 2010;26(9):1145–51. https://doi.org/10.1093/bioinformatics/btq102.
59. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–95. https://doi.org/10.1093/bioinformatics/btp698.
60. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013. https://doi.org/10.48550/arXiv.1303.3997
61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93. https://doi.org/10.1093/bioinformatics/btr509.
62. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr Biol. 2015;25(10):1375–80. 10.1016/j.cub.2015.03.047.
63. Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 2016;25(5):1058–72. https://doi.org/10.1111/mec.13540.
64. Hilgers L, Liu S, Jensen A, Brown T, Cousins T, Schweiger R, Guschanski K, Hiller M. Avoidable false PSMC population size peaks occur across numerous studies. Curr Biol. 2025;35(4):927–30. 10.1016/j.cub.2024.09.028.
65. Karjee R, Iyer V, Chatterjee D, Ray R, Garg KM, Chattopadhyay B. Species biology and demographic history determine species vulnerability to climate change in tropical island endemic birds. eLife. 2025. DOI: https://elifesciences.org/reviewed-preprints/106369
66. Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backström N, Künstner A, Balakrishnan CN, Heger A, Ponting CP, Clayton DF. Molecular evolution of genes in avian genomes. Genome Biol. 2010;11(6):R68. https://doi.org/10.1186/gb-2010-11-6-r68.
67. Chen D, Hosner PA, Dittmann DL, O'Neill JP, Birks SM, Braun EL, Kimball RT. Divergence time estimation of Galliformes based on the best gene shopping scheme of ultraconserved elements. BMC Ecol Evol. 2021;21(1):209. https://doi.org/10.1186/s12862-021-01935-1.
68. Cosma BM, Shirali Hossein Zade R, Jordan EN, van Lent P, Peng C, Pillay S, Abeel T. Evaluating long-read de novo assembly tools for eukaryotic genomes: insights and considerations. GigaScience. 2023;12:giad100. https://doi.org/10.1093/gigascience/giad100.
69. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, Chitsaz H. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience. 2013;2(1):2047–X17. https://doi.org/10.1186/2047-217X-2-10.
70. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021;592(7856):737–46. https://doi.org/10.1038/s41586-021-03451-0.

71. Hotaling S, Wilcox ER, Heckenhauer J, Stewart RJ, Frandsen PB. Highly accurate long reads are crucial for realizing the potential of biodiversity genomics. BMC Genomics. 2023;24(1):117. https://doi.org/10.1186/s12864-023-09193-9.
72. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. Proceedings of the National Academy of Sciences. 2017;114(8):E1460-9. https://doi.org/10.1073/pnas.1616702114
73. Zhao S, Hu Y, Bai H, Chang G, Guo Q. Cytoskeletal signaling as a shared pathway for convergent genetic mechanisms underlying loss of flight in birds. Poult Sci. 2025;104(9):105503. https://doi.org/10.1016/j.psj.2025.105503.
74. Davison GW. Avian spurs. J Zool. 1985;206(3):353–66. https://doi.org/10.1111/j.1469-7998.1985.tb05664.x.
75. Göransson G, von Schantz T, Fröberg I, Helgee A, Wittzell H. Male characteristics, viability and harem size in the pheasant, *Phasianus colchicus*. Anim Behav. 1990;40(1):89–104. https://doi.org/10.1016/S0003-3472(05)80668-2.
76. Mateos C, Carranza J. Female choice for morphological features of male ring-necked pheasants. Anim Behav. 1995;49(3):737–48. https://doi.org/10.1016/0003-3472(95)80206-1.
77. Mateos C, Carranza J. Effects of male dominance and courtship display on female choice in the ring-necked pheasant. Behav Ecol Sociobiol. 1999;45(3):235–44. https://doi.org/10.1007/s002650050558.
78. Legrain E, Stevenard N, Capron E, Parrenin F, Vazquez Riveiros N. The Marine Isotopic Stage 7: a relic of the\41-ka world\? Perspectives from a global-scale sea-surface temperature synthesis. EGUsphere [preprint]. 2025. https://doi.org/10.5194/egusphere-2025-5840.
79. Wendt KA, Li X, Edwards RL, Cheng H, Spötl C. Precise timing of MIS 7 substages from the Austrian Alps. Clim Past. 2021;17(4):1443–54. https://doi.org/10.5194/cp-17-1443-2021.
80. Ehrmann W, Schmiedl G. Nature and dynamics of North African humid and dry periods during the last 200,000 years documented in the clay fraction of Eastern Mediterranean deep-sea sediments. Q Sci Rev. 2021;260:106925. https://doi.org/10.1016/j.quascirev.2021.106925.
81. Camuera J, Ramos-Román MJ, Jiménez-Moreno G, García-Alix A, Ilvonen L, Ruha L, Gil-Romera G, González-Sampériz P, Seppä H. Past 200 kyr hydroclimate variability in the western Mediterranean and its connection to the African Humid Periods. Sci Rep. 2022;12(1):9050. https://doi.org/10.1038/s41598-022-12047-1.
82. Castañeda IS, Mulitza S, Schefuß E, Lopes dos Santos RA, Sinninghe Damsté JS, Schouten S. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. Proceedings of the National Academy of Sciences. 2009;106(48):20159-63. https://doi.org/10.1073/pnas.0905771106
83. Drake N, Breeze P. Climate change and modern human occupation of the Sahara from MIS 6 – 2. InAfrica from MIS 6 – 2: Population Dynamics and, Paleoenvironments. 2016; (pp. 103–122). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-7520-5_6

84. Shields GF, Wilson AC. Calibration of mitochondrial DNA evolution in geese. J Mol Evol. 1987;24(3):212–7. https://doi.org/10.1007/BF02111234.

85. Mays HL, Hung CM, Shaner PJ, Denvir J, Justice M, Yang SF, Roth TL, Oehler DA, Fan J, Rekulapally S, Primerano DA. Genomic analysis of demographic history and ecological niche modeling in the endangered Sumatran *rhinoceros Dicerorhinus sumatrensis*. Curr Biol. 2018;28(1):70–6. 10.1016/j.cub.2017.11.021.

Figures

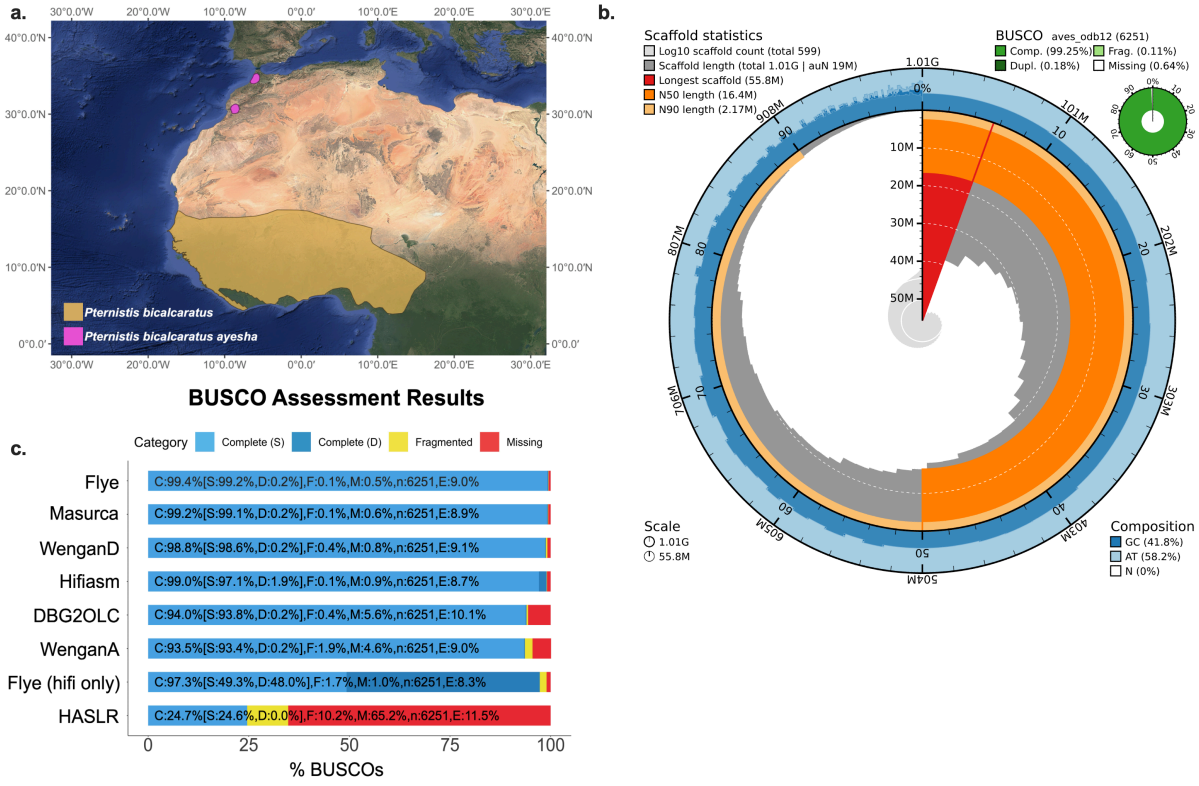


Figure 1

(a) Map showing the geographical distribution of *Pternistis bicalcaratus* and *Pternistis bicalcaratus ayesha*, South and North of Sahara, respectively (map adapted from IUCN). (b) Snail plot describing and representing the assembly stats of the final genome assembly using MaSuRCA. (c) BUSCO assessment results for genome assemblers compared using aves_odb12.

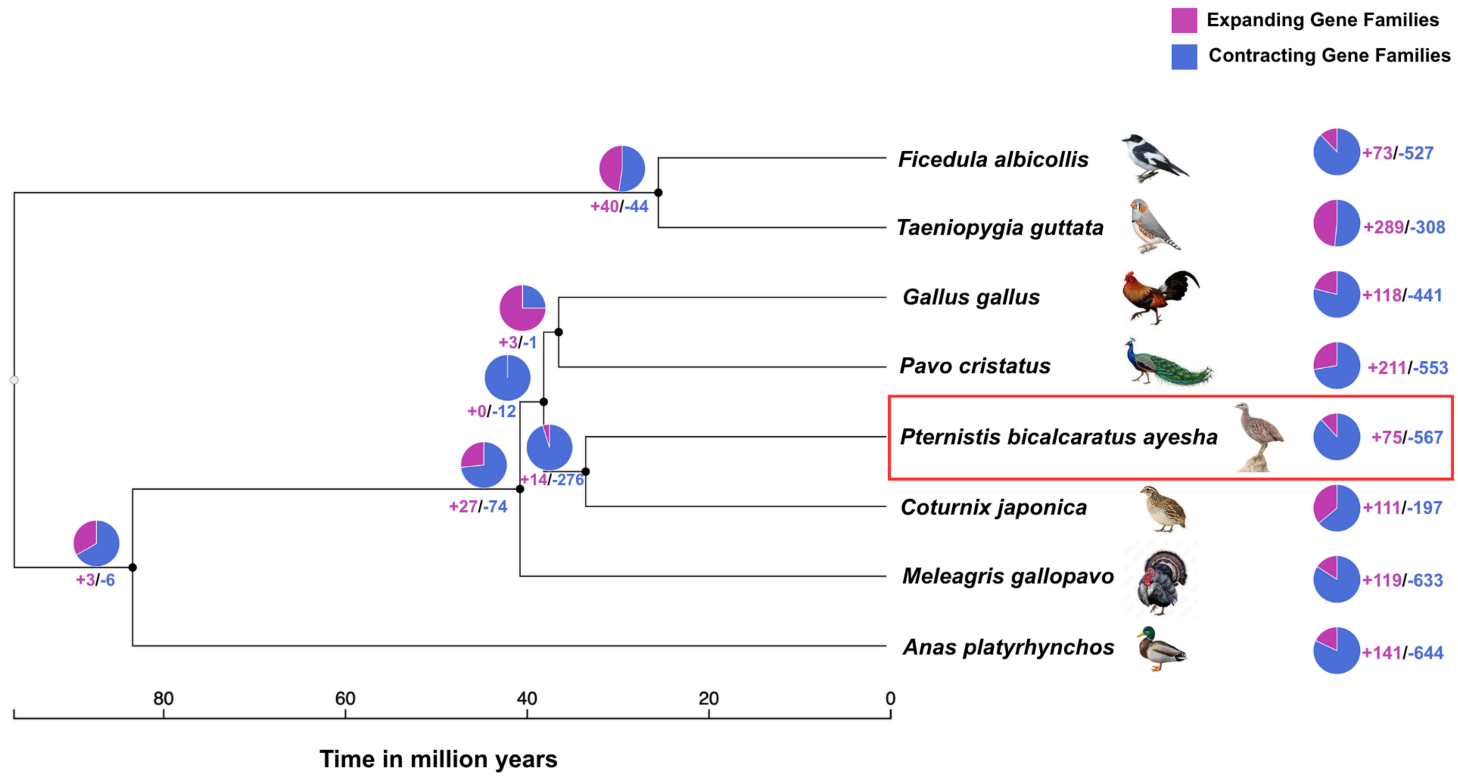


Figure 2

Ultrametric tree depicting the phylogenetic relationship and gene family fluctuations in select bird species.

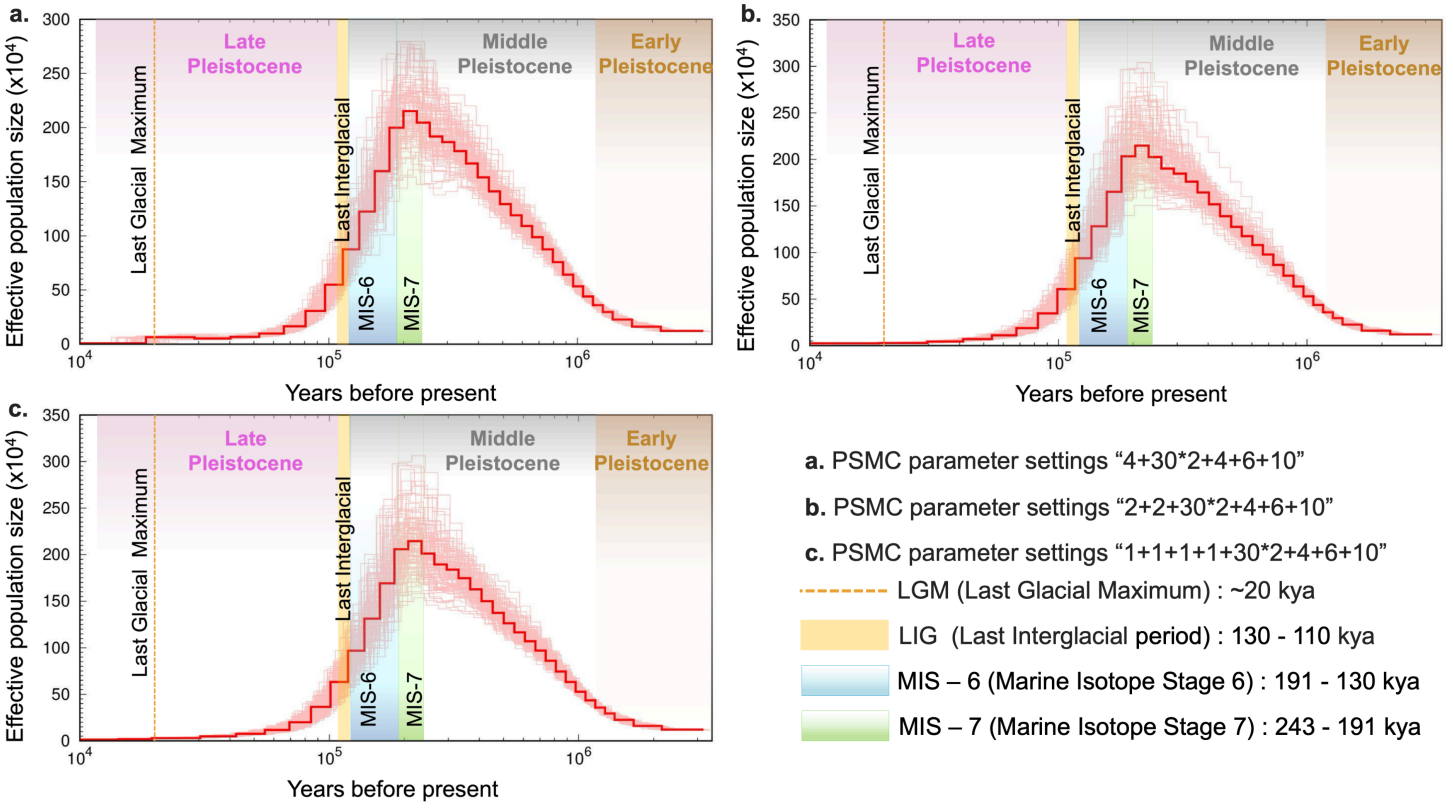


Figure 3

PSMC plots showing fluctuations in the historical demography of *Pternistis bicalcaratus ayesha* using different parameter settings (a) 4+30*2+4+6+10 (b) 2+2+30*2+4+6+10 (c) 1+1+1+1+30*2+4+6+10. Thick red lines represent estimates of effective population size (N_e), and thin red lines are bootstrap replicates. Peak in N_e was observed during the MIS-7 followed by a gradual decline up to the end of the Late Pleistocene.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.xlsx