Supplementary Materials for

Rapid speciation along the world's longest landbird migration

Merondun *et al.*

Corresponding authors: Jochen Wolf, and Kasper Thorup

Supplementary Materials

The PDF file includes:

Table Captions S1 to S6

Figs. S1 to S13

Other Supplementary Materials for this manuscript include the following:

Supplementary Tables (Supplementary_Tables.xlsx)

Data and materials availability

All associated genomic sequence data are uploaded to the NCBI SRA with biosample and run accessions indicated in **TABLE S1** within Bioproject PRJNA1170948. All bioinformatic code to reproduce the genomic resequencing aspects of the manuscript is available on GitHub (https://github.com/EvoBioWolf/CUCKOO_migration) and Zenodo (https://doi.org/10.5281/zenodo.14873435). The migratory track data used in this study are uploaded to the Movebank Data Repository.

Supplementary Tables

Table S1. Metadata for whole-genome resequencing samples included in the study.

Summary of biosamples and sequencing data. Samples subset for specific analyses are indicated in the 'Analysis_' columns. All sequencing libraries have been archived in the SRA with corresponding accession numbers. The total reads and bases represent the combined sequencing output for each biosample. Geographic clusters were assigned using k-means clustering.

Table S2. Model fit and parameter estimates for demographic history.

Model fit and estimates, including effective population size (N_e) and time to divergence in years between CC_W and CC_E (T_DIV1) and the ancestor of CC and CO (T_DIV2).

Table S3. Demographic model bootstrap correction and confidence interval for estimated parameters.

Bootstrap correction of each parameter of the best fit model and the corresponding confidence interval. NPOP1-3: Effective population size (2Ne) of CC_W, CC_E and CO, respectively. NANC1-2: Effective population size (2Ne) of the ancestral population of CC and CC_CO, respectively. TMIG1S-2S: Time when the first and second pulse of migration started, respectively, in generations. TMIG1E-2E: Time when the first and second pulse of migration ended, respectively, in generations. MIG01: Migration rate of CCE to CCW. MIG10: Migration rate of CC_W to CC_E. MIG12: Migration rate of CO to CC_E. MIG21: Migration rate of CC_E to CO. MIGA0A2: Migration rate of CO to ancestral population CC. MIGA2A0: Migration rate of ancestral population CC to CO. TDIV1-2: Time of divergence of CC and CC_CO, respectively, in generations.

Table S4. Metadata for migratory route and residency for ancestral reconstruction.

Metadata and maps (which probably can't be included for copyright) for the *Cuculidae* (*n* = 127) used for deep-time ancestral reconstruction of route and residency. Abbreviations residency: Africa = A, Indo-Malay/Australia = I, Americas = SA. Abbreviations route: intra-African (AA), Africa-Palearctic eastern (APe), Africa-Palearctic western (APw), Africa-Palearctic polymorphic (APwe), intra-Indo-Malay/Australian (II), North-South American (NS), sedentary (S), and intra-South American (SS).

Table S5. Bioacoustics metadata

The complete dataset of song parameters. The table also includes the country where the recording was made (RF represents Russia) and the start time of each syllable within the given recording. Record ID refers to unique catalogue numbers from the xeno-canto.org database.

Table S6. Individual track details

Details of individual tracks, including tag type (PTT, PinPoint GPS, ICARUS Basic GPS and population western (CC_W) and eastern (CC_E) common cuckoos (*C. canorus*) and southerly-migrating oriental cuckoos (CO, *C. optatus*).

Supplementary Figures

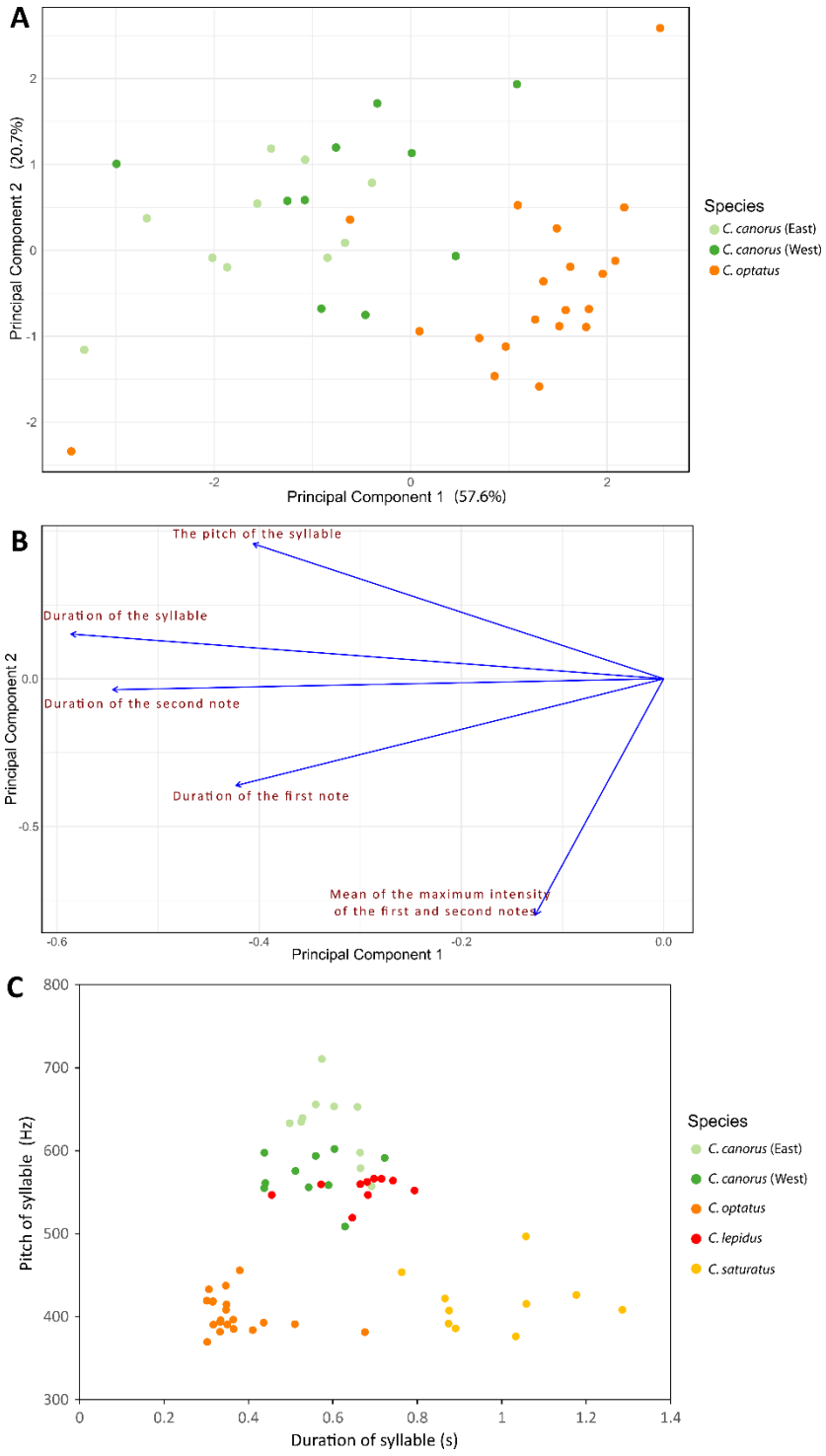

Fig S1 Analysis of song parameters in cuckoos.

(**A**) Principal Component Analysis (PCA) scatter plot shows individuals' distribution along the first two principal components, with points colored by species. **(B)** The loading plot illustrates the contribution of each variable to PC1 and PC2, with the direction and length of the arrows indicating the influence of each parameter. **(C)** Pitch and duration of song syllables for 60 individual cuckoos of closely related *Cuculus* taxa.

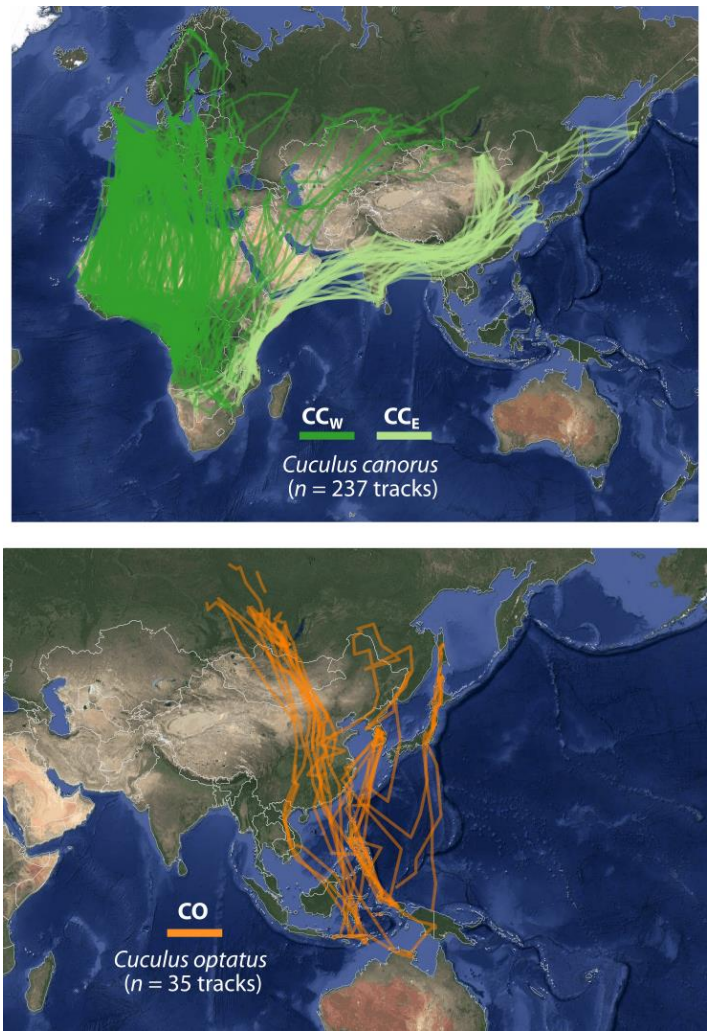

Fig S2. Individual-level tracking data depicted in satellite imagery.

Tracks showing satellite tracks for individual common (*C. canorus*) and oriental (*C. optatus*) cuckoos. Colors correspond to broad-scale migratory routes categorized by breeding locations of western (CC_W) and eastern (CC_E) common cuckoos (*C. canorus*) and southerly-migrating oriental cuckoos (CO, *C. optatus*).

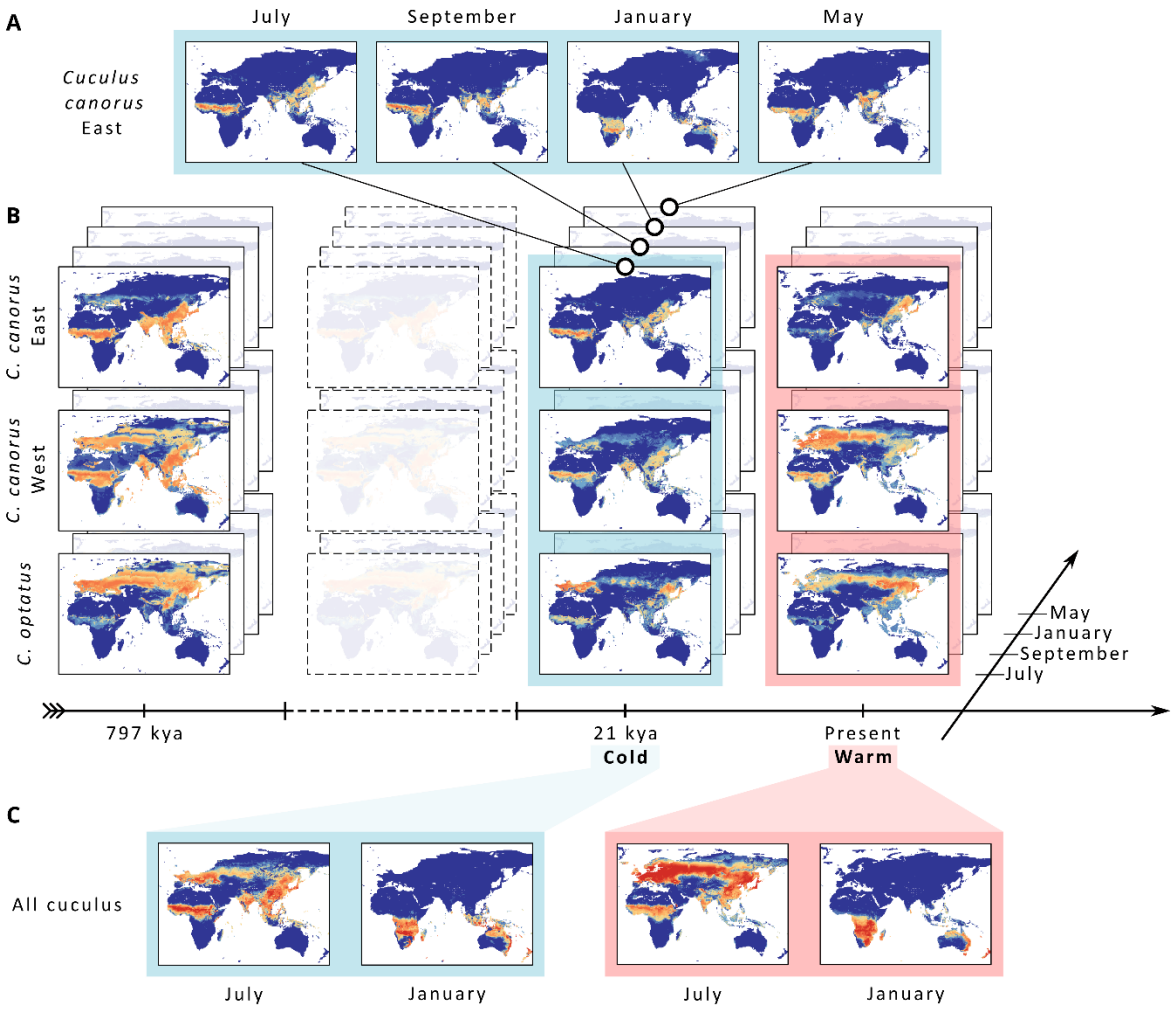

Fig S3. Overview of hindcasted habitat suitability models and illustration of fundamental and realised niche.

Suitabilities were modelled month by month every two thousand years back to 797 Ka including both cold (e.g. the Last Glacial Maximum, LGM, 21 Ka) and warm (e.g. current, 0 Ka) periods. Suitability is indicated from 0 (blue) to 1 (dark red). (**A**) Distribution of suitable habitat for eastern *Cuculus canorus* during major sedentary periods during the annual cycle at the LGM. (**B**) Realised niche habitat suitability modelled back in time during breeding for the each of the three populations considered (western and eastern common cuckoos *C. canorus* and oriental cuckoo *C. optatus*). (**C**) Fundamental niche habitat suitability during breeding and wintering for all populations combined at present and 21 Ka. Because of the morphological similarity, we assume that the combined distributions reflect the fundamental niche of the populations combined.

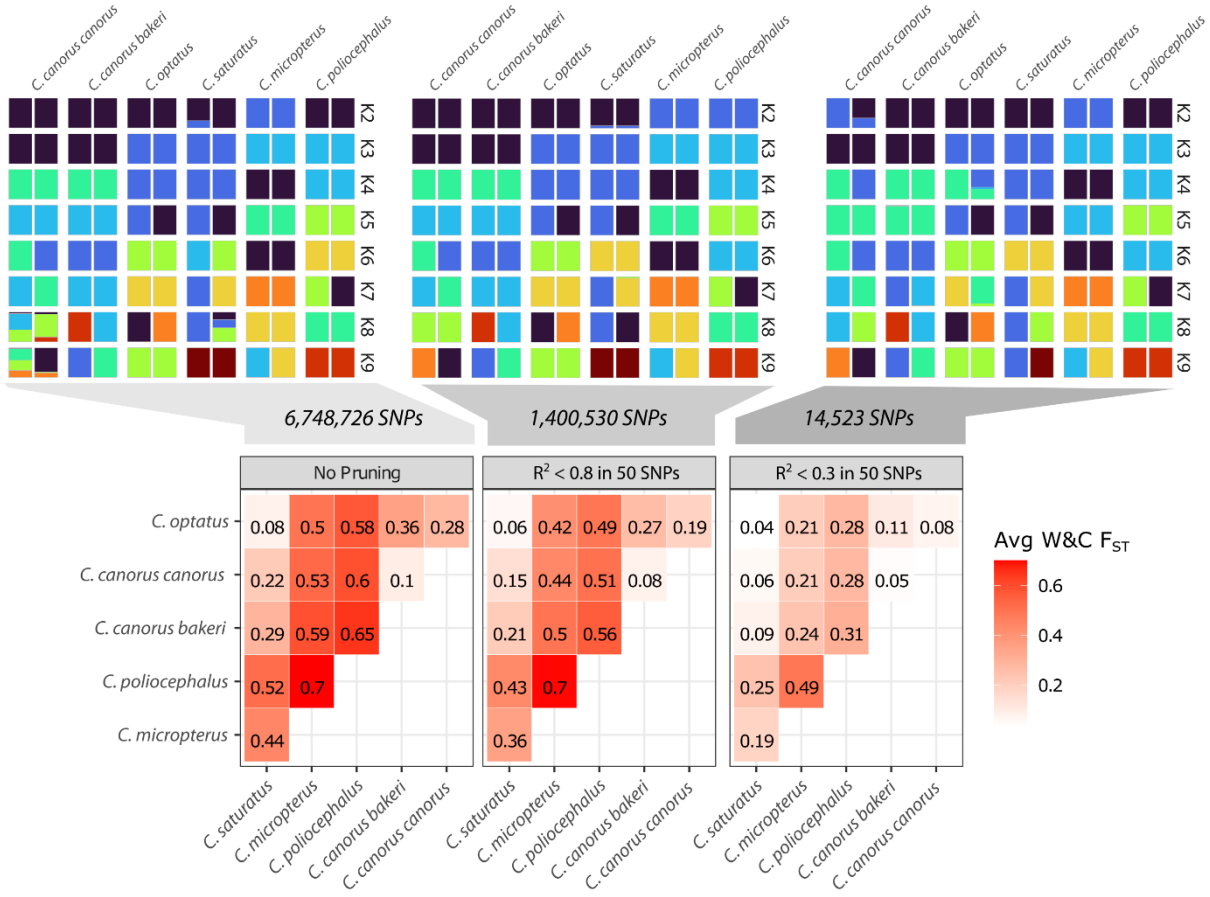

Fig S4. Genetic landscape of differentiation among *Cuculus* species.

Genetic differentiation (F_{ST}) estimated on chromosome 1 among *Cuculus canorus* (including specimens labeled as *C. c. canorus* and *C. c. bakeri*), *C. optatus*, *C. saturatus*, *C. micropterus*, and *C. poliocephalus*. Two randomly selected individuals per species were analyzed, with SNPs filtered to exclude sites with more than 20% missing genotypes. Sensitivity analyses assessed the impact of linkage disequilibrium (LD) pruning: no pruning (left), moderate pruning (middle), and strict pruning (right), leading to a variable number of SNPs used in ADMIXTURE and F_{ST} calculations.

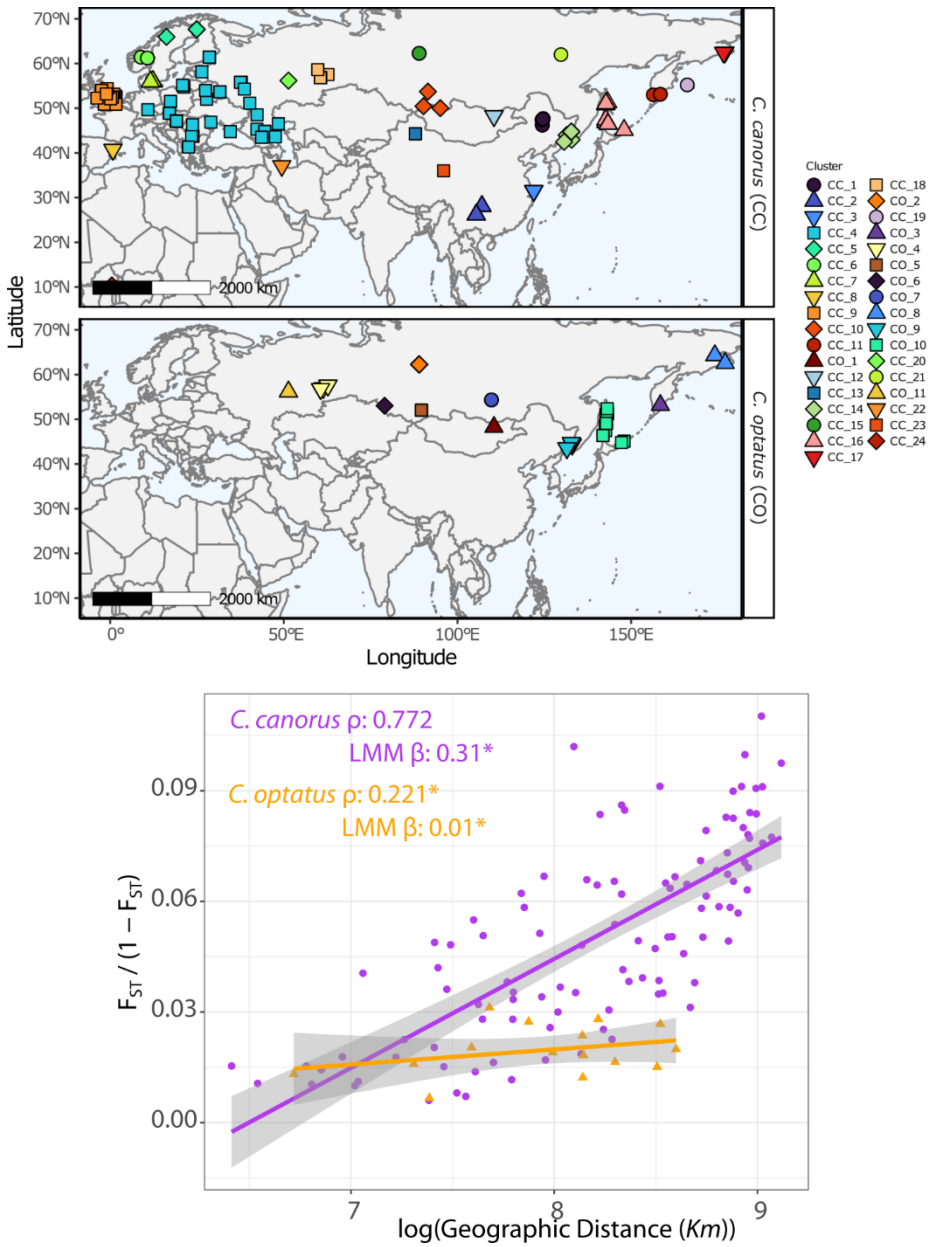

Fig S5. Association between genetic and geographic distance.

Relationship between geographic distance (*Km*) and genetic distance in *Cuculus canorus* and *Cuculus optatus*, estimated as F_{ST} / (1 - F_{ST}) for autosomal chromosomes. Each point represents a pairwise comparison between geographic groups (e.g., CC_11 vs. CC_10, CO_9 vs. CO_8 see **TABLE S1**). Spearman rank correlations between geographic and genetic distances are shown above for each species after Bonferroni correction. Shaded regions indicate 95% confidence intervals from linear regression smoothing (geom_smooth, method = '*lm*').

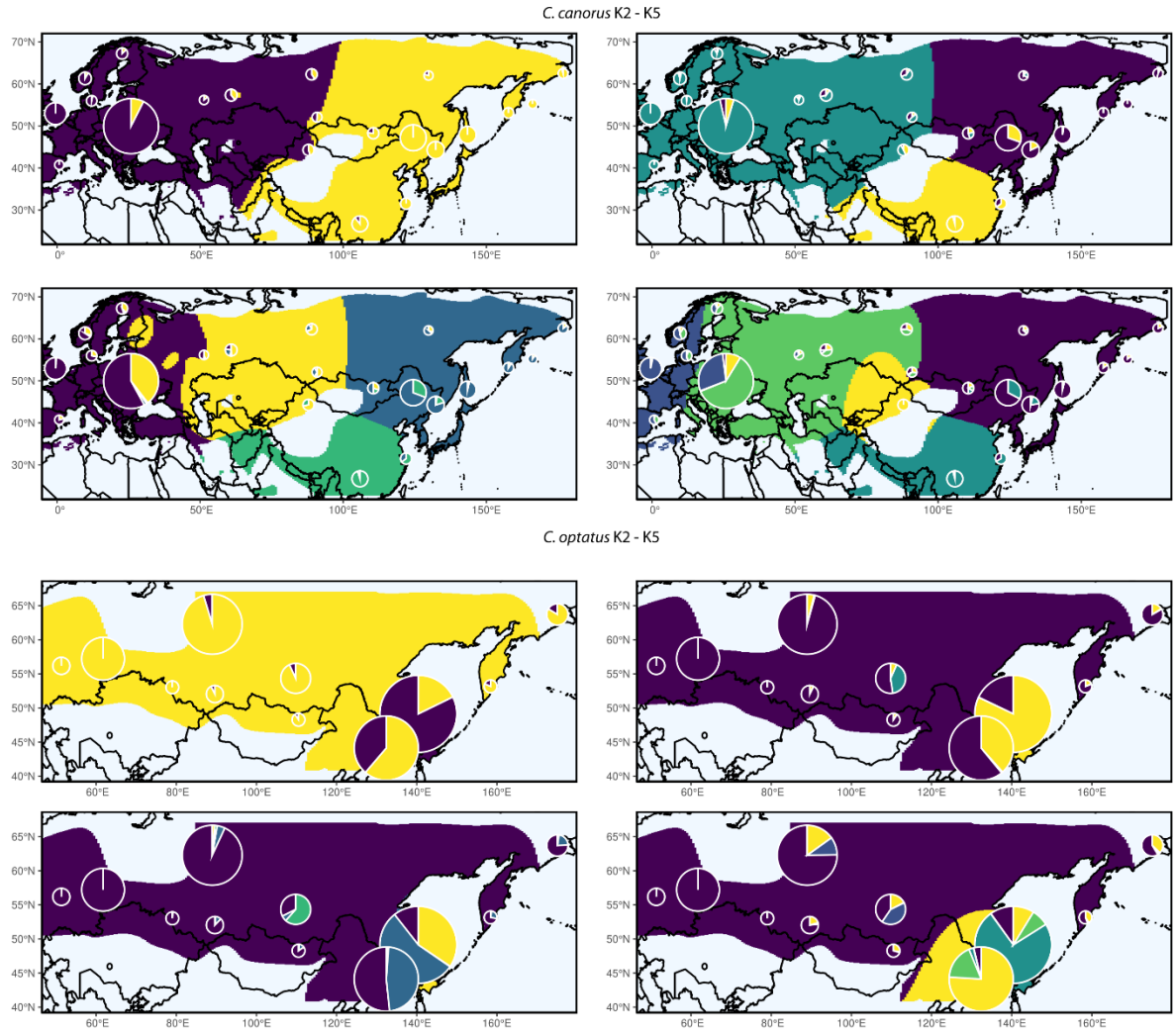

Fig S6. Inferred genetic ancestry across the landscape

Tesselation of genetic ancestry inferred across Eurasia using ADMIXTURE ancestry coefficients from K2 – K5 clusters, clipped to known breeding areas for each species. Pie charts represent the average ancestry coefficients of individuals within each geographic region (FIG S5), with the size of the pie scaled to the sample size.

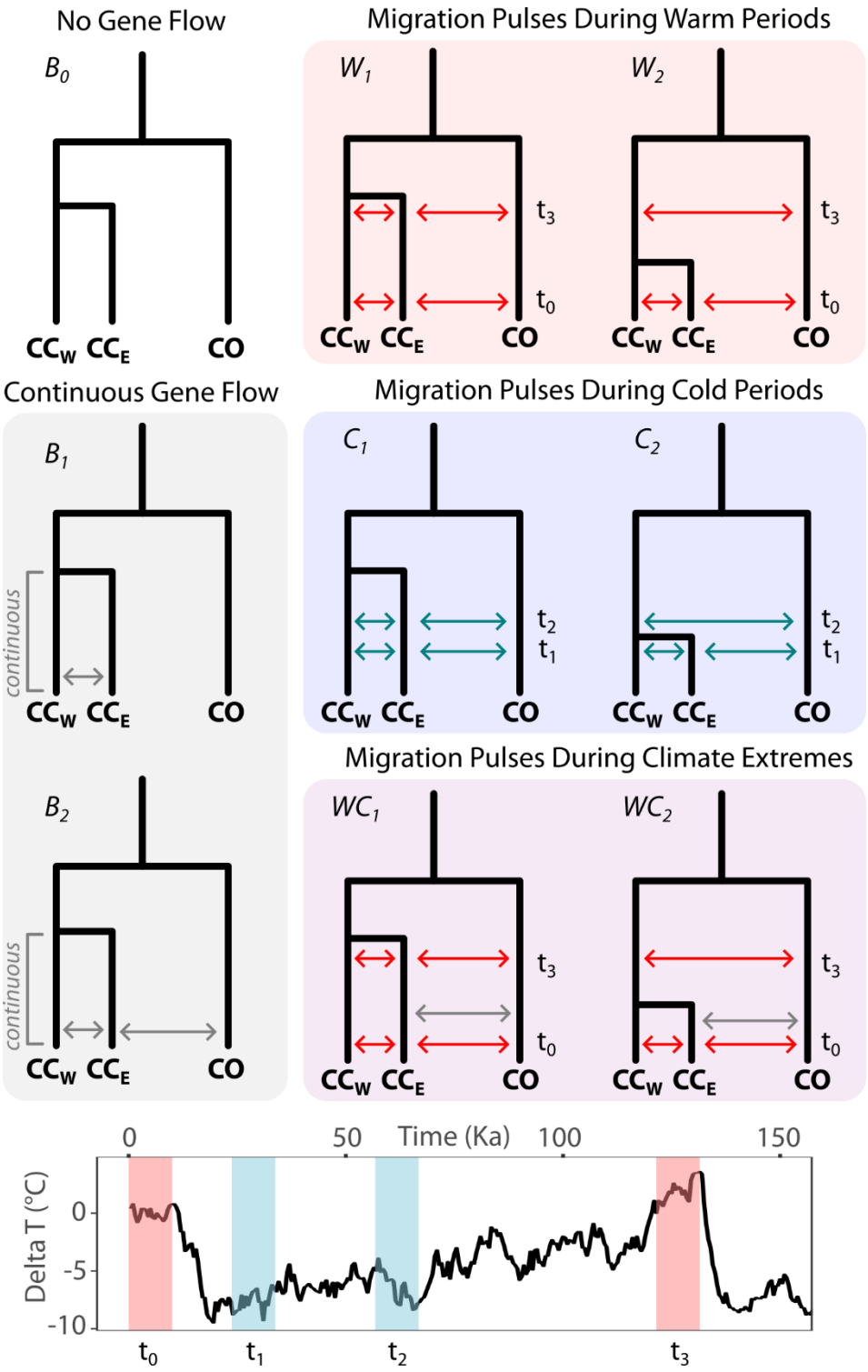

Fig S7. Demographic models simulated using Fastsimcoal2

Schematics representing the demographic models assessed in this study. Specifically, models include: (i) isolation without gene flow (base model B_0), (ii) isolation with continuous gene flow between ancestral lineages (base models B_1–B_2), and (iii) discrete pulses of gene flow occurring during two warm periods (W_1–W_2), two cold periods (C_1–C_2), or a combination of both (WC_1–WC_2). Temperature change corresponding to climatic periods indicated below, from (*36*).

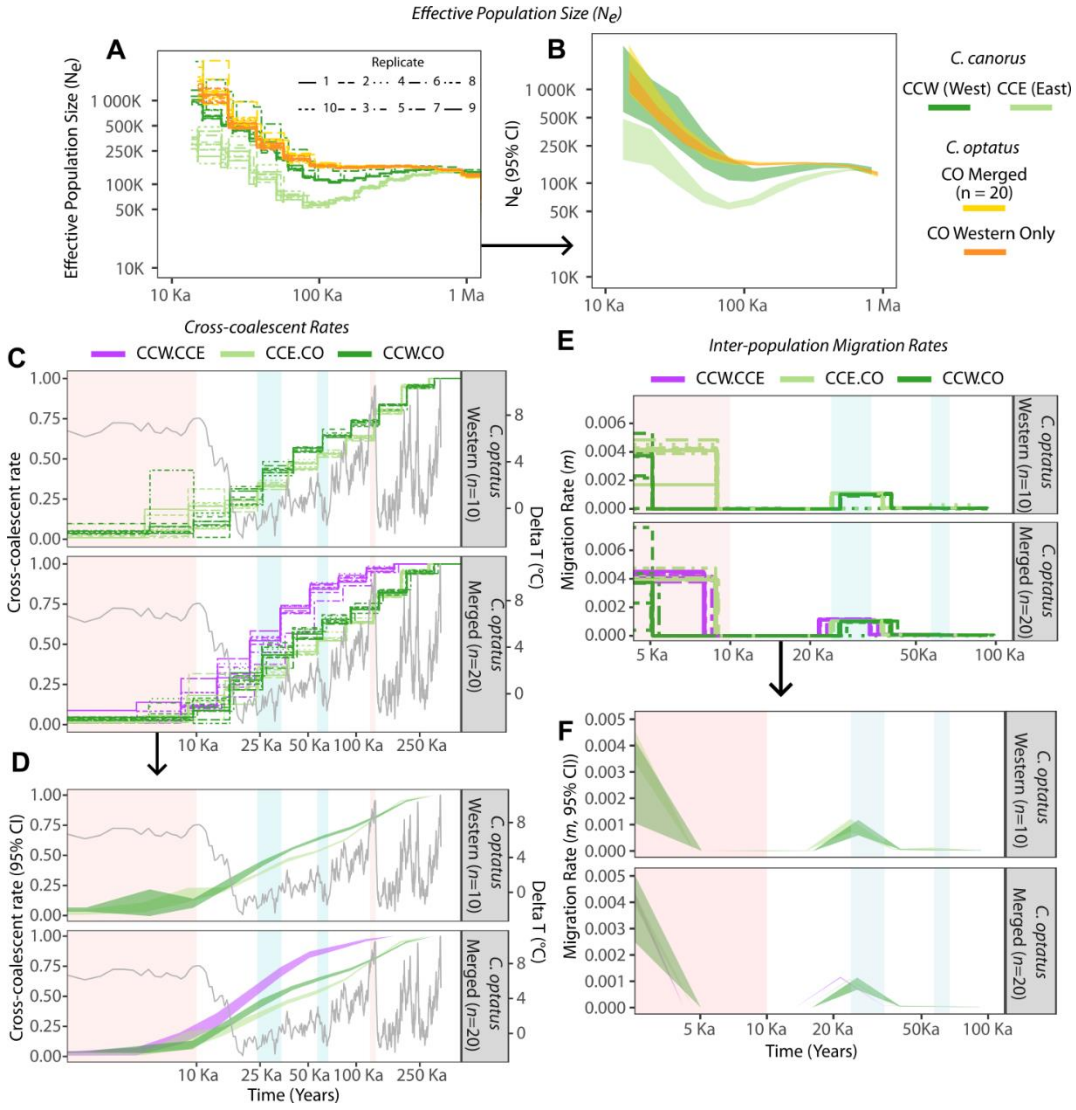

Fig S8. Demographic history inferred with the sequential Markovian coalescent

Demographic history inferred using MSMC2 for the three target populations (CC_W, CC_E, CO). Sensitivity to including the entire *C. optatus* species distribution is also shown in gold. (**A-B**) Effective population size (N_e) estimated by MSMC2, showing results from 10 independent replicates with two different randomly sampled representatives for each population, totalling 8 haplotypes for each replicate, (thin lines) and the smoothed 95% confidence intervals (ribbons). (**C-D**) Cross-coalescent rates between groups, illustrating higher cross-coalescent rate with more genetic mixing between populations, while a lower rate suggests the populations are more isolated from each other. (**E-F**) MSMC-IM (Multiple Sequentially Markovian Coalescent with Isolation-with-Migration) migration rate estimates, showing changes in bi-directional gene flow between populations over time. Shaded background regions highlight 10 Ka climate periods used for demographic modelling. Time is plotted on a logarithmic scale.

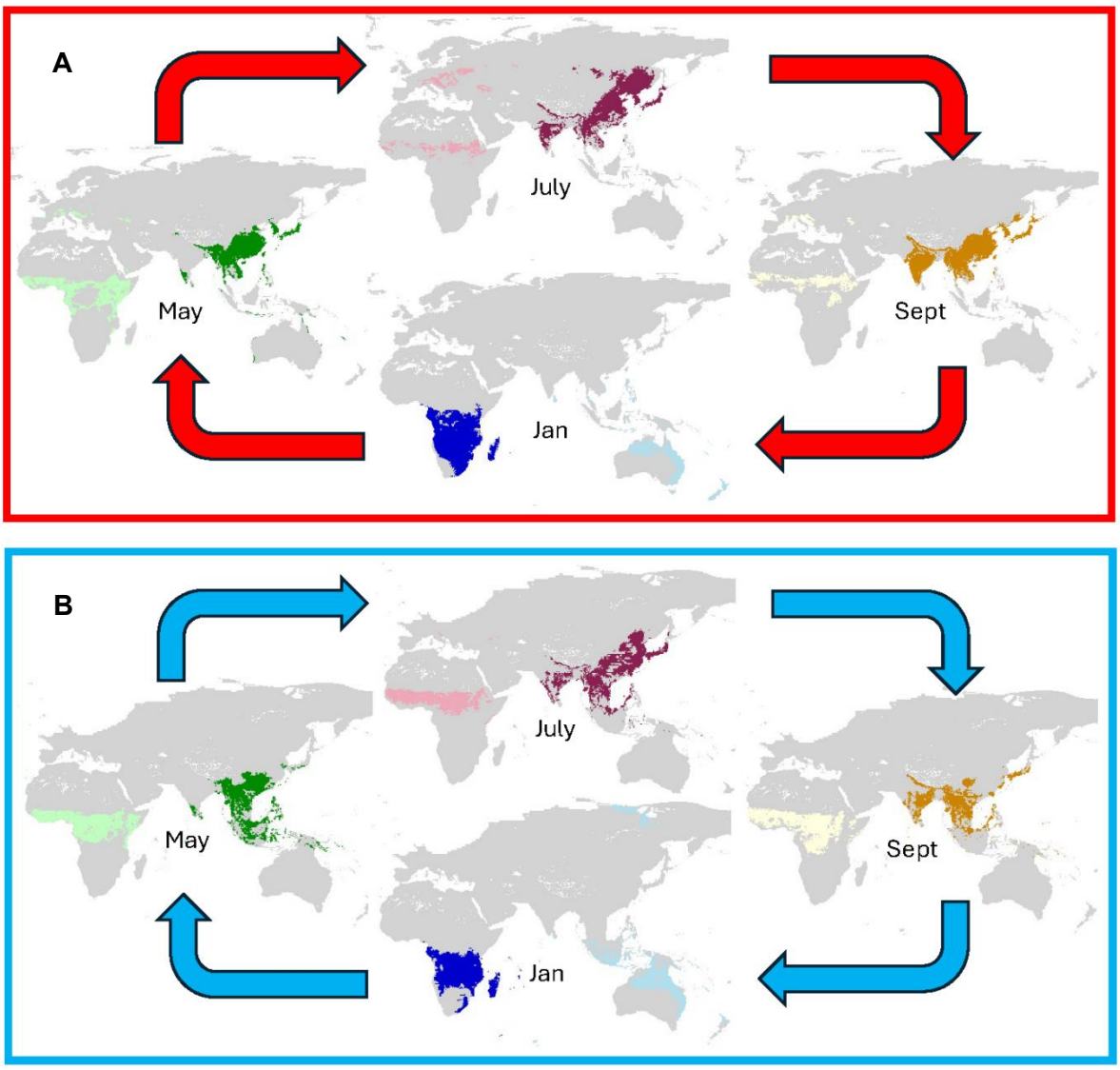

Fig S9. Estimated seasonal distributions of eastern *Cuculus canorus* during warm (current) and cold (Last Glacial Maximum, LGM) periods

Distribution of estimated habitat suitabilities during major sedentary stays in July (breeding), September (fall stopover), January (wintering) and May (spring stopover). (**A**) The current strategy involves breeding and stopping over in the East and wintering in the West (Africa) and the area currently used is colored darker. (**B**) This strategy remains possible during cold periods even with a shrinking total area.

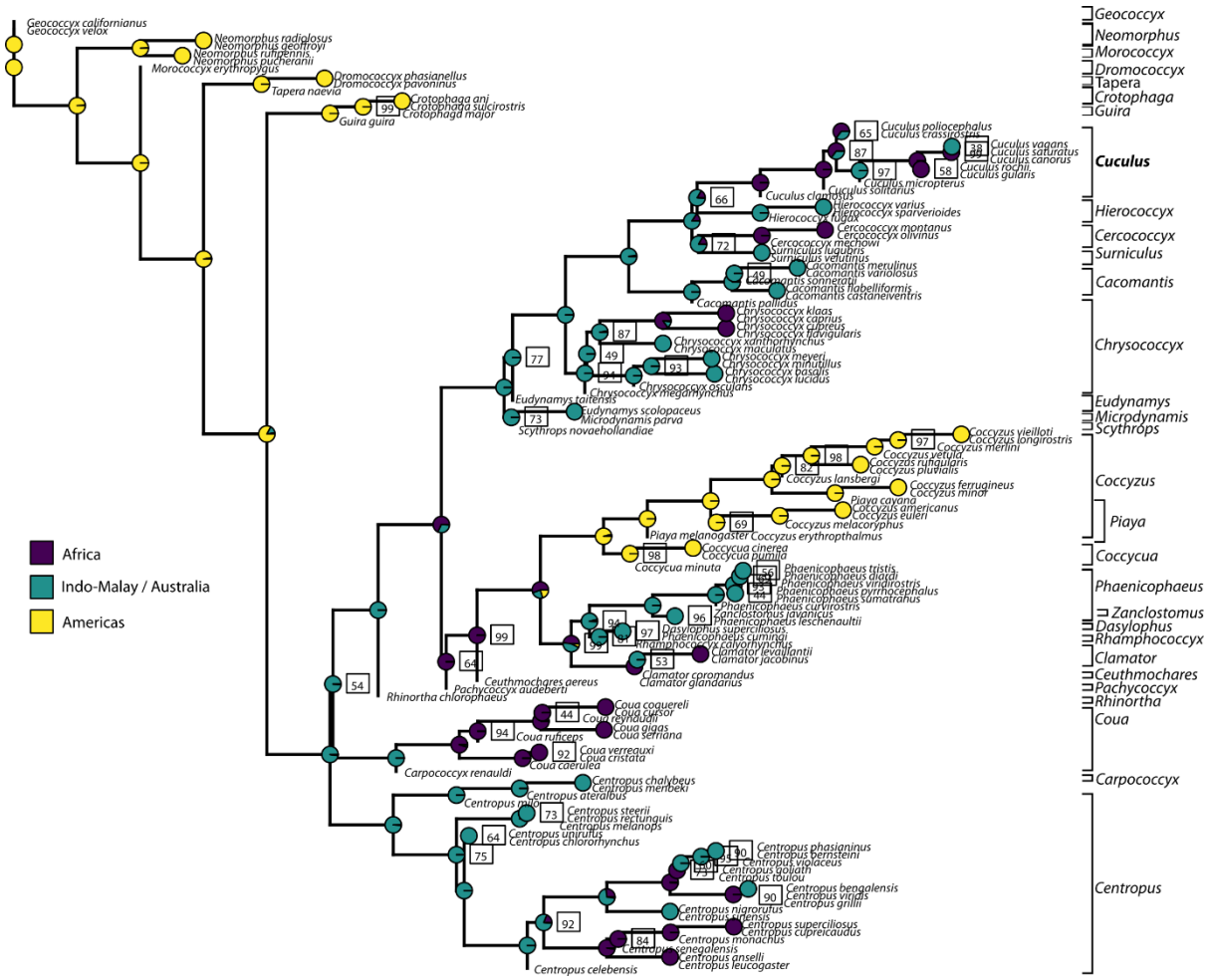

Fig S10. Ancestral reconstruction of overwintering locations across *Cuculidae*

Phylogenetic ancestral state reconstruction of *Cuculidae* showing likelihoods of overwintering location, or residency for non-migrants, across three major biogeographic regions (Africa, Indo-Malay/Australia, and Americas). Nodes with less than 100% bootstrap support are indicated with bootstrap support values, with pie charts at nodes representing the likelihood of ancestral state distributions. The tree is rooted with *Geococcyx californianus*, and major cuckoo genera are labelled on the right.

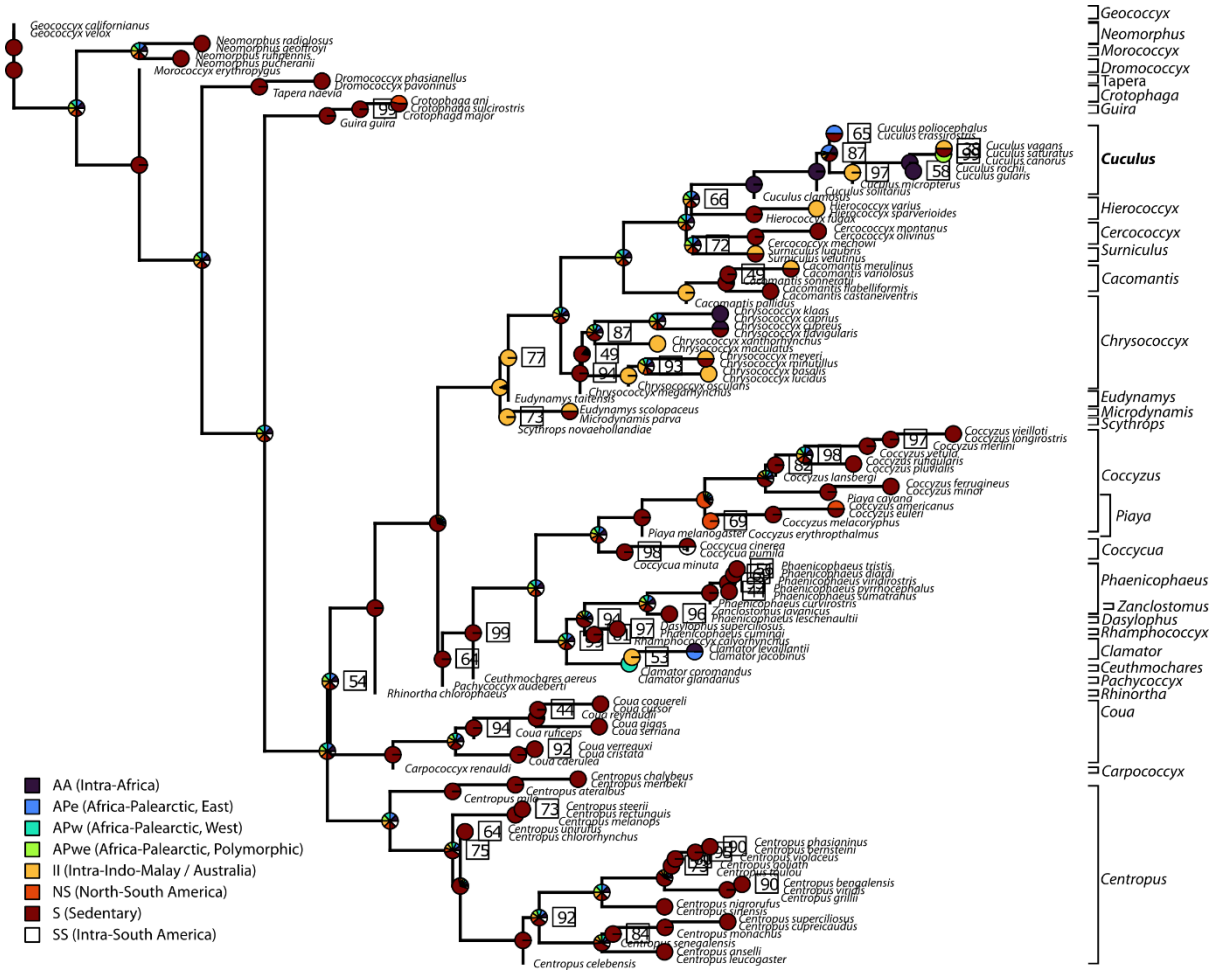

Fig S11. Ancestral reconstruction of migratory behavior across *Cuculidae*

Phylogenetic ancestral state reconstruction of *Cuculidae* showing migratory routes and residency patterns. Node values indicate bootstrap support (values <100%), with pie charts representing the likelihood of ancestral migratory behaviors: intra-African (AA), Africa-Palearctic eastern (APe), Africa-Palearctic western (APw), Africa-Palearctic polymorphic (APwe), intra-Indo-Malay/Australian (II), North-South American (NS), sedentary (S), and intra-South American (SS) movements. The tree is rooted with *Geococcyx californianus*, with major cuckoo genera labelled on the right.

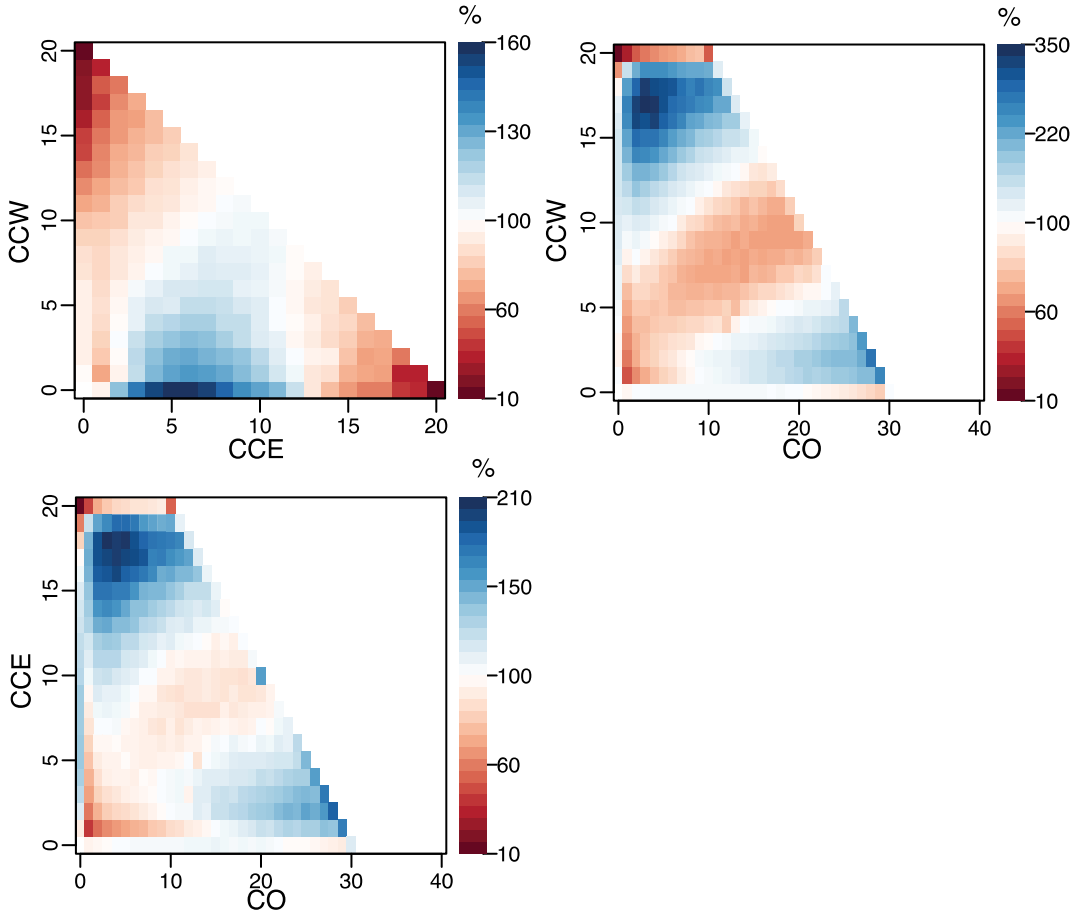

Fig S12. Demographic model site frequency spectra.

Comparison of observed and expected joint-site frequency spectra (SFS) simulated by fastsimcoal for the best fit model. Blue and red regions reflect overestimated and underestimated sites, respectively.

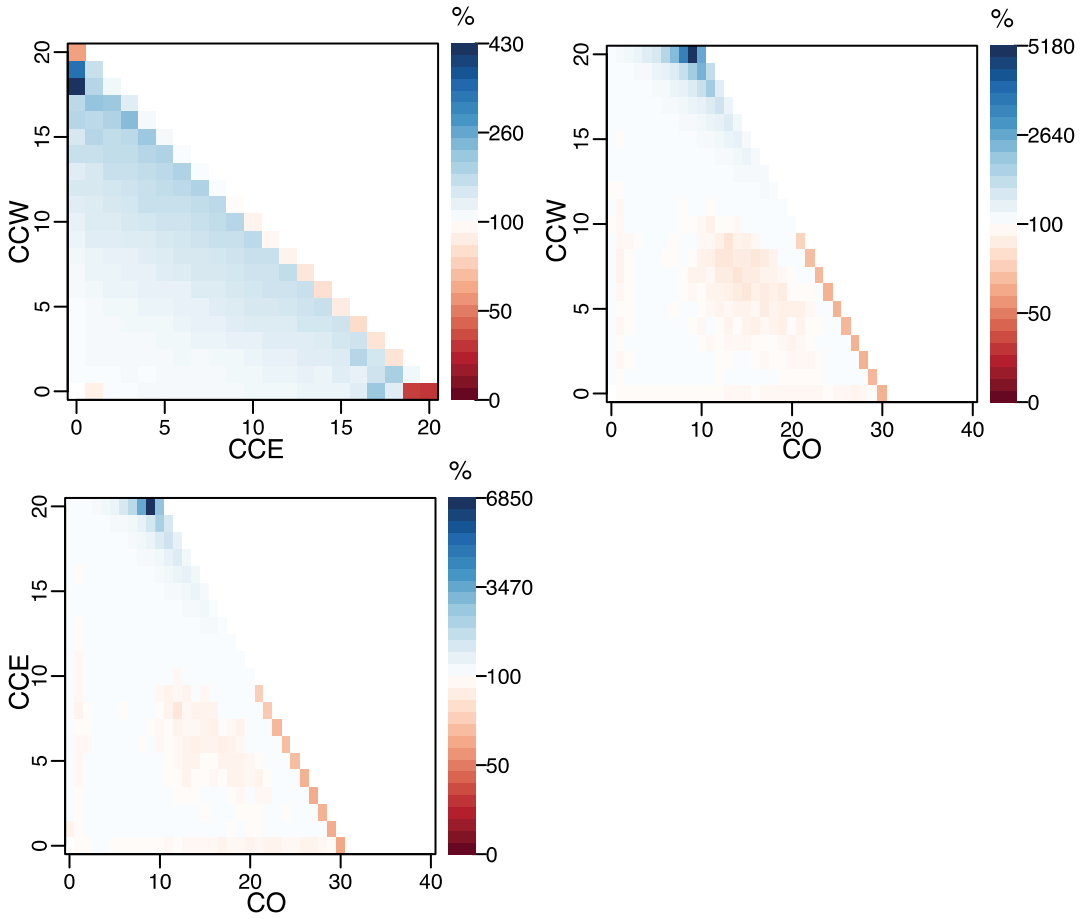

Fig S13. Bootstrap site frequency spectra

Expected j-SFS simulated from the parameter estimates of the best fit model using finite-sites model.