



## Phylogeny and historical biogeography of the swallow family (Hirundinidae) inferred from comparisons of thousands of UCE loci

Drew R. Schield<sup>a,b,\*</sup>, Clare E. Brown<sup>c,1</sup>, Subir B. Shakya<sup>c,d</sup>, Gina M. Calabrese<sup>a</sup>, Rebecca J. Safran<sup>a</sup>, Frederick H. Sheldon<sup>c</sup>

<sup>a</sup> Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

<sup>b</sup> Department of Biology, University of Virginia, Charlottesville, VA, USA

<sup>c</sup> Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

<sup>d</sup> Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA

### ARTICLE INFO

#### Keywords:

Ancestral reconstruction  
Biogeography  
Passerines  
Species tree  
Trait evolution  
Ultraconserved elements

### ABSTRACT

Swallows (Hirundinidae) are a globally distributed family of passerine birds that exhibit remarkable similarity in body shape but tremendous variation in plumage, sociality, nesting behavior, and migratory strategies. As a result, swallow species have become models for empirical behavioral ecology and evolutionary studies, and variation across the Hirundinidae presents an excellent opportunity for comparative analyses of trait evolution. Exploiting this potential requires a comprehensive and well-resolved phylogenetic tree of the family. To address this need, we estimated swallow phylogeny using genetic data from thousands of ultraconserved element (UCE) loci sampled from nearly all recognized swallow species. Maximum likelihood, coalescent-based, and Bayesian approaches yielded a well-resolved phylogenetic tree to the generic level, with minor disagreement among inferences at the species level, which likely reflect ongoing population genetic processes. The UCE data were particularly useful in helping to resolve deep nodes, which previously confounded phylogenetic reconstruction efforts. Divergence time estimates from the improved swallow tree support a Miocene origin of the family, roughly 13 million years ago, with subsequent diversification of major groups in the late Miocene and Pliocene. Our estimates of historical biogeography support the hypothesis that swallows originated in the Afrotropics and have subsequently expanded across the globe, with major *in situ* diversification in Africa and a secondary major radiation following colonization of the Neotropics. Initial examination of nesting and sociality indicates that the origin of mud nesting – a relatively rare nest construction phenotype in birds – was a major innovation coincident with the origin of a clade giving rise to over 40% of extant swallow diversity. In contrast, transitions between social and solitary nesting appear less important for explaining patterns of diversification among swallows.

### 1. Introduction

Swallows (Hirundinidae) are a globally distributed passerine family containing roughly 88 species. All swallows are aerial insectivores that share a common morphology with aerodynamic bodies, long pointed wings, and short broad bills (see illustrations in Fig. 2). Beyond this basic shape, however, swallows vary in traits such as plumage, sociality and territoriality, migratory strategy, and nesting behavior (Turner and Rose, 1989; Turner, 2018; Billerman et al., 2020). This variation, against the background of conserved morphology and basic life history, makes swallows a valuable system for comparative analyses of trait evolution.

However, the effectiveness of such studies requires a comprehensive, well-resolved phylogenetic tree with which to study trait change over time, and no such resource is available for swallows despite a substantial body of phylogenetic inquiry.

Although swallows are nearly globally distributed, the majority of taxa have breeding distributions that occur in the Afrotropics or Neotropics (32 and 19 species, respectively). In contrast, few swallows breed in Australia (four species, three of which are endemic) or Oceania (two species, neither of which is endemic). The Nearctic, Panamanian, Oriental, and Palearctic realms each contain between 10 and 13 breeding species. Only one species, the barn swallow (*Hirundo rustica*), occupies a

\* Corresponding author at: Department of Biology, University of Virginia, Charlottesville, VA 22903, USA.

E-mail address: [drew.schild@virginia.edu](mailto:drew.schild@virginia.edu) (D.R. Schield).

<sup>1</sup> Authors contributed equally.

<https://doi.org/10.1016/j.ympev.2024.108111>

Received 9 February 2024; Received in revised form 13 May 2024; Accepted 18 May 2024

Available online 25 May 2024

1055-7903/© 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

breeding distribution that includes more than three geographic realms (Palearctic, Nearctic, and Neotropical), and only two species, *H. rustica* and the bank swallow (*Riparia riparia*), occur in both the Old and New World. Understanding the development of this geographic distribution will illuminate not only the family's geographic origins, but also the evolutionary innovations that have facilitated swallow diversification.

With respect to nesting, swallows can be grouped into three broad categories (Supplementary Fig. S1; Winkler and Sheldon, 1993): species that excavate burrows (e.g., *Riparia*), species that adopt burrows, holes, or tree cavities (e.g., *Tachycineta*), and species that build mud-nests. Mud-nests can be subdivided further into open mud cups (e.g., *Hirundo*), enclosed mud cups (e.g., *Delichon*), and retort-shaped mud-nests, which have an enclosed cup with an entrance tunnel (e.g., *Cecropis*). Mud nesting and cavity adoption generally correspond to geographic distribution; most African swallows are mud-nesters and most Neotropical swallows adopt cavities. Swallows also vary in breeding sociality, with some species exhibiting solitary breeding (i.e., only a single breeding pair at a nesting site), some breeding in small groups, and others being highly colonial, forming large aggregations of hundreds to thousands of breeding pairs (Turner, 2018). As with historical biogeography, understanding the evolution of sociality and nesting behavior will provide substantial insight into the forces responsible for swallow diversification.

Before now, the phylogeny of Hirundinidae has been reconstructed three times using molecular data. Sheldon and Winkler (1993) compared single-copy DNA of 21 species in 19 putative genera using DNA-DNA hybridization. This was followed by comparisons of cytochrome *b* sequences from the same taxa (Sheldon et al., 1999). Finally, using sequences at three genetic loci, Sheldon et al. (2005) then inferred relationships among 75 swallow species. These studies provided significant insight into swallow phylogeny but left several deeper relationships, and relationships of species that were not sampled, unresolved. Beyond these intergeneric comparisons, multiple phylogenetic and phylogeographic studies have examined more recent patterns of divergence within various swallow subgroups, including: cave swallows (*Petrochelidon fulva*; Kirchman et al., 2000); tree swallows (*Tachycineta*; Whittingham et al., 2002; Cerasale et al., 2012; Dor et al., 2012); New World martins (*Progne*; Moyle et al., 2008); sand martins (*Riparia riparia* and *R. diluta*; Pavlova et al., 2008); the barn swallow subspecies complex (*Hirundo rustica*) and *Hirundo sensu stricto* (Zink et al., 2006; Dor et al., 2010); saw-wings (*Psalidoprocne*; Barrow et al., 2016); rough-winged swallows (*Stelgidopteryx*; Babin, 2005); and forest swallows (*Atronanus fuliginosus* [previously *Petrochelidon fuliginosa*]; de Silva et al., 2018). These studies have addressed inter- and intraspecific relationships but have not provided insight on deeper relationships within Hirundinidae.

Our objectives in this study were to reconstruct a well-supported, comprehensive phylogenetic hypothesis for Hirundinidae, and to then use this tree to estimate historical biogeography and examine character trait evolution. We addressed the first objective by producing a phylogenomic dataset of thousands of ultraconserved element (UCE) sequences from at least one individual of the majority of swallow species (82 species and numerous subspecies) and all genera. Next, we reconstructed the tree using a combination of maximum likelihood inference from a concatenated data matrix, as well as coalescent-based gene-tree summary methods and Bayesian inference. We then used this phylogenetic hypothesis to infer the evolution of sociality, nest type, and geographic breeding distribution across the swallow family.

## 2. Materials and methods

### 2.1. Sampling

We obtained samples from 82 of the 88 swallow species recognized by (Clements et al., 2023), including representatives of all 20 genera (Supplementary Table S1). We were unable to sample the Red Sea swallow (*Petrochelidon perditia*), which is known from only one specimen

(Fry and Smith, 1985), or the rufous-bellied swallow (*Cecropis badia*), which has a very restricted range. In most cases we sampled a single individual per species, but included multiple individuals for species when prior work (e.g. Sheldon et al., 2005; Moyle et al., 2008; Pavlova et al., 2008; Dor et al., 2010; Barrow et al., 2016) suggested significant intraspecific divergence, or when authorities disagree on classification (e.g., the red-rumped swallow [*Cecropis daurica*] complex). We obtained DNA from frozen or ethanol preserved tissue samples when possible. For taxa for which preserved tissues were not available, we sampled from museum skins (i.e., 'historic DNA'). Because swallows are small birds with tiny toe pads, from which historic DNA samples are typically sampled, we took pieces of skin from the apteria of the breast to avoid excessive damage to specimen feet. Such samples have been shown to produce usable DNA for UCE sequencing (Tsai et al., 2020). In total, we acquired samples from 122 individuals (99 tissues and 23 skins) representing 113 distinct species, subspecies, or populations (Supplementary Table S1). We included other sylvioid species, lanceolated warbler (*Locustella lanceolata*) and chestnut-capped flycatcher (*Erythrocerus mccallii*), as outgroups, using previously generated sequence data from Moyle et al. (2016).

### 2.2. DNA extraction, target enrichment, and sequencing

We extracted genomic DNA from frozen and ethanol preserved tissue samples using QIAGEN DNeasy Blood and Tissue kits following the manufacturer's protocol, and from skin samples using a phenol-chloroform purification and ethanol precipitation protocol optimized for historic DNA extraction (Tsai et al., 2020). Prior to purification, we washed skin samples in 100 % ethanol and 1X STE buffer, and digested them in a solution of buffer ATL, IM DTT, and proteinase K. We were able to obtain usable DNA from all skin samples with the exception of the pied-winged swallow (*Hirundo leucosoma*). We quantified DNA extracts using a Qubit 2.0 fluorometer (Life Technologies, Inc.). We mechanically sheared tissue samples using an Episonic Multi-Functional Bioprocessor to approximately 500 bp on average, which we assessed using gel electrophoresis. Historic DNA typically suffers some degree of degradation and is sufficiently fragmented to not require further shearing. This was the case for all but the two Peruvian martin (*Progne murphyi*) samples, which we mechanically sheared. We used a KAPA HyperPrep Kit (Roche Sequencing Systems) to prepare a barcoded library for Illumina sequencing from each sample by ligating dual indexed iTru adapters (Glenn et al. 2016) to the fragmented DNA. We followed the kit protocol but used half reagent volumes for the end-repair and A-tailing, adapter ligation, and post-ligation cleanup steps. Tissue libraries were amplified with nine PCR cycles, while historic DNA libraries were amplified with twelve cycles. We combined the amplified, barcoded libraries into equimolar pools of between seven and nine samples each, comprising 11 tissue pools and three historic DNA pools.

We used hybrid capture to enrich each pool for 5,060 UCE loci. We first denatured the libraries and hybridized a set of 5,472 biotinylated RNA probes (Tetrapods-UCE-5Kv1 sold by MYcroarray; Faircloth et al., 2012) to their target sequences on the single-stranded DNA and captured the resulting probe-target hybrids on streptavidin coated magnetic beads. This allowed unhybridized, non-target DNA to be washed away. We then released and amplified the target DNA, following the protocol in the MYbaits manual v3.0.1 through bait-target hybridization, hybrid binding, and bead washing steps. Tissue libraries were hybridized at 65 °C, and historic DNA libraries at 55 °C. After washing the beads to remove non-target DNA, we resuspended the enriched libraries along with the beads in 10 mM Tris-Cl with 0.05 % TWEEN and performed 15 cycles of limited-cycle PCR on the suspension with KAPA HiFi HotStart polymerase. We purified the PCR product with a 1.2 × or 3.0 × AMPure bead cleanup for tissue libraries and historic DNA libraries, respectively, and verified the absence of adapter dimers using an Agilent Bioanalyzer. We quantified libraries with qPCR using a KAPA Library Quantification Kit and sequenced the final libraries using 150 bp paired-end reads on an

Illumina NovaSeq 6000.

### 2.3. Processing of sequence data and generation of datasets for phylogenetic inference

We evaluated the quality of demultiplexed sequences using FastQC v0.11.7 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and removed adapter sequences and low quality bases from the raw reads using Illumiprocessor v2.0.9 (Faircloth et al., 2013). Illumiprocessor is a wrapper program for Trimmomatic v0.36 (Bolger et al., 2014), which we implemented with default settings. This included trimmed leading and trailing bases with Phred33 scores below 5 and 15, respectively, and scanning reads with a four bp sliding window, trimming the remainder of a read when the average Phred33 score was below 15. We also removed reads below 40 bp after trimming.

We then generated UCE datasets for phylogenetic inference from the trimmed read data using PHYLUCe v1.6.3 (Faircloth, 2016). We first assembled reads for each sample into contigs with SPAdes v3.12.0 (Bankevich et al., 2012), implemented with `phyluce_assembly_assemble_spades`, running SPAdes in single-cell mode with the default k-mer sizes of 21, 33, and 55 and a coverage threshold of 5. Prior to assembly, we used the `insilico_read_normalization.pl` script from Trinity v2.6.6 (Grabherr et al., 2011) to normalize reads of four historic DNA samples which had more than 10 million read pairs. Preliminary analysis revealed long terminal branch lengths for the majority of skin samples. To mitigate the effect of low quality data on terminal branch lengths, we used the mapping and correction steps implemented using `phyluce-workflow` to map read data to the assembled contigs of each skin sample and filter low depth, low quality base calls, using default settings. We then extracted assembled contigs that matched the UCE loci in the Tetrapods-UCE-5Kv1 probe set (available at <https://ultraconserved.org/>). We aligned loci matching the probes using Mafft v7.130 (Katoh et al., 2002; Katoh and Standley, 2013), implemented using `phyluce-align_seqcap_align`, then trimmed the resulting alignments with Gblocks v0.91b (Castresana, 2000), using `phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed`. We used default settings, with the exception of relaxing the minimum number of sequences required for a flank position to 65 %. We then generated datasets for phylogenetic analysis by excluding loci that were not recovered in  $\geq 95$  % of samples. Four historic DNA samples (Congo martin [*Riparia congica*], one Galápagos martin [*Progne modesta*], and two Cuban martins [*Progne cryptoleuca*]) had many fewer recovered UCE loci than the remaining samples, resulting in spurious placement in preliminary phylogenetic analyses. The remaining *P. modesta* sample fell within the clade containing the outgroup taxa, a potential result of contamination. These five samples were removed from downstream analyses, yielding a ‘full’ dataset including 118 samples (Supplementary Table S1). For comparison, we also generated a ‘reduced’ dataset consisting of only tissue samples and *Pseudochelidon* species (103 samples). Here, we explicitly included *Pseudochelidon* species because a previous analysis inferred *Pseudochelidon* to be sister to all other extant swallows (Sheldon et al., 2005), and it was critical to assess support for this ancestral split using the current dataset. Removing low quality samples from the ‘full’ dataset and further pruning in the ‘reduced’ dataset did not affect the number of genera represented in analyses.

### 2.4. Phylogenetic analysis

We performed phylogenetic inference using concatenated maximum likelihood (ML) and coalescent-based approaches. For ML, we performed 40 independent tree searches (20 random and 20 parsimony starting trees) using RAXML-NG v0.7.0 (Kozlov et al., 2019) on the concatenated alignment and selected the best scoring tree. We analyzed the alignment as a single partition and specified a GTR + GAMMA substitution model. We assessed support for the topology using 100 bootstrap replicate trees, also generated using RAXML-NG, and verified

bootstrap convergence with the autoMRE criterion with a cutoff of 0.3. The bootstrapped trees were used to calculate support values for the nodes of the best scoring tree from the initial tree searches. We then analyzed the reduced dataset using the same procedure for comparison to the inferred phylogeny from the full dataset.

We used two coalescent-based approaches to estimate the Hirundinidae species tree. We first implemented SVDquartets (Chifman and Kubatko, 2014, 2015) to analyze the full and reduced datasets for direct comparison with concatenated ML analyses. SVDquartets uses a coalescent framework to analyze quartets of taxa from site patterns in the alignment, then employs a supertree approach to estimate a species tree from these quartets. We ran SVDquartets in PAUP\* v4.0 (Swofford, 2003) to perform species tree inference, specifying *L. lanceolata* and *E. mccallii* as a monophyletic outgroup and 100 bootstrap replicates in the analysis. We then used ASTRAL v5.6.1 (Mirarab et al., 2014; Zhang et al., 2017) to perform coalescent-based species tree inference on a further reduced dataset consisting of only tissue samples (i.e., without *Pseudochelidon*) to enable broad comparison of deeper topological relationships within Hirundinidae across analyses. ASTRAL estimates a species tree from a set of gene trees under the multispecies coalescent model. We used PHYLUCe together with RAXML v8.2.11 (Stamatakis, 2014) on individual UCE alignments to conduct multi-locus bootstrapping, by which loci, and then sites within loci, are sampled with replacement (Seo, 2008). We generated 500 multi-locus bootstrapped alignments for each UCE locus and estimated bootstrapped gene trees from those alignments. We then ran ASTRAL to estimate species trees from each set of 500 bootstrapped gene trees and constructed a 50 % majority-rule consensus tree from the species trees using SumTrees v4.4.0 (DendroPy v4.4.0; Sukumaran and Holder, 2010). We performed plotting and annotation of resulting trees in R v4.3.1 (R Core Team, 2023) using the packages `ape` v5.7.1 (Paradis and Schliep, 2019), `phytools` v1.9.16 (Revell, 2012), and `treeio` v1.25.4 (Wang et al., 2020).

### 2.5. Divergence time estimation

We used Bayesian inference to estimate divergence times within Hirundinidae and to provide a time-calibrated phylogeny for analyses of ancestral character evolution and historical biogeography. Bayesian analysis of large multi-locus datasets is computationally expensive, so to balance robust inference with efficiency we sampled five sets of 50 UCE loci at random for analysis. We performed analyses using BEAST2 v2.7.4 (Bouckaert et al., 2014), specifying the HKY substitution model with empirical site frequencies and the optimized relaxed clock model with the default mean clock rate of 1. We used default priors for additional site and clock model parameters, and further specified two normally distributed most recent common ancestor (MRCA) prior constraints based upon estimates from Oliveros et al. (2019). We first constrained the split between outgroup and ingroup taxa to 22 million years ago (MYA) with a standard deviation = 2.5, based on the previously inferred divergence time between *Locustella* and swallow taxa (*Hirundo* and *Progne*). Second, we set a prior MRCA constraint on the clade containing all ingroup swallow taxa excluding the saw-wings (*Psalidoprocne*) and river martins (*Pseudochelidon*) to 9 MYA with a standard deviation = 1.5, based on the previously estimated divergence between *Hirundo* and *Progne*; this prior distribution also encompasses late Miocene fossils for putative *Hirundo*, *Delichon*, and *Riparia* species described by Kessler (2013). We applied these priors to each locus set and performed two independent runs for each using 10 million MCMC generations. After assessing chain mixing and convergence among independent runs using Tracer (Rambaut et al., 2018), we combined posterior samples using LogCombiner (Bouckaert et al., 2014) and used TreeAnnotator to generate a maximum clade credibility tree, discarding the first 20 % of samples as burn in. Finally, to focus downstream ancestral reconstruction analyses, we pruned the time-calibrated tree to 88 representative ingroup taxa for which data on geographic range, nest type, and sociality were available. The majority of remaining tips correspond to species,

although in a few cases we either grouped clades into a single tip (e.g., non-African *Cecropis daurica/striolata*), or included multiple subspecies or geographic variants (e.g., subspecies of the plain martin [*Riparia paludicola*]), again based on the availability of trait data.

## 2.6. Historical biogeography and analyses of trait evolution

### 2.6.1. Ancestral geographic range estimation

We reconstructed the evolution of geographic breeding range across Hirundinidae, which includes species that breed on every continent other than Antarctica. We included eight biogeographic realms in the analysis, modified from Holt et al. (2013). The eight realms are: (A) Afrotropical (including Madagascar); (B) Palearctic, a combination of Holt et al.'s Palearctic, Sino-Arabian, and Saharo-Arabian realms, excluding northern North America and Greenland; (C) Oriental; (D) Australian, including New Zealand; (E) Oceanian, including New Guinea; (F) Nearctic, including northern North America and Greenland; (G) Panamanian, including the Caribbean; and (H) Neotropical. We downloaded a digitized map of these biogeographic realms (available from <https://macroecology.ku.dk/resources/wallace>), and applied modifications as described above. We then built a presence/absence matrix describing the breeding range overlap of each swallow taxon with the eight realms, using digitized distribution maps obtained from BirdLife International (<https://www.birdlife.org>) and del Hoyo et al. (2004). These procedures were conducted in R v4.3.1 (R Core Team, 2023) using the packages sp v1.3-1 (Pebesma and Bivand, 2005; Bivand et al., 2013), rgdal v1.3.6 (Bivand et al., 2018), rgeos v0.4-2 (Bivand and Rundel, 2018), maptools v0.9-4 (Bivand and Lewin-Koh, 2018), raster v2.8-4 (Hijmans, 2018), rworldmap v1.3-6 (South, 2011), and rnaturlaearth v0.1.0 (South, 2017).

To infer the biogeographic history of Hirundinidae, we fit a series of models of geographic range evolution to the pruned time-calibrated swallow tree and geographic range data using the R package BioGeoBEARS (Matzke, 2018). For a global-scale system such as the Hirundinidae, the assumptions made by the dispersal, local extinction, and cladogenesis model (DEC; Ree and Smith, 2008) are likely the most realistic. The DEC model allows two anagenetic processes, dispersal and extinction, and two types of cladogenetic processes, sympatry and vicariance. In both cladogenetic processes, one daughter lineage is always assumed to have a range of only one area, meaning that in vicariant cladogenesis the ancestral range cannot be split evenly between daughter lineages. This is a reasonable assumption for Hirundinidae, as the occupied biogeographic realms are large enough for divergence to occur within a single area. The dispersal-vicariance analysis model (DIVA; Ronquist, 1997; Yu et al., 2010) is broadly similar to the DEC model, but allows equal division of the ancestral range under vicariant cladogenesis, and requires that daughter lineages occupy completely overlapping ranges under sympatric cladogenesis. The BAYAREA model (Landis et al., 2013) allows the same anagenetic processes of dispersal and extinction as the DIVA and DEC models, but does not allow range evolution to occur during cladogenesis, meaning that daughter lineages must inherit a range identical to their ancestor. This assumption renders the BAYAREA model less plausible for Hirundinidae than the DEC and DIVA models. Both DIVA and BAYAREA are implemented in BioGeoBEARS as likelihood interpretations of the models (i.e., DIVALIKE and BAYAREALIKE). BioGeoBEARS also expands each model to allow founder event speciation, in which range expansion into a new geographic area accompanies cladogenesis (Matzke, 2014). This process is incorporated by adding a parameter (J) that allows one daughter lineage to inherit the ancestral range, while the other disperses to occupy a new area outside of the ancestral range. We compared the fit of the three models to the data, with and without founder event speciation. No taxa in the dataset have breeding ranges including more than four realms, so we set this as the maximum range size. We used dispersal multipliers to constrain dispersal among regions; we assume that swallows do not disperse directly across vast ocean distances, disallowing

dispersal between the Afrotropical and the Australian or Oceanian realms, and between the Old World and the New World (other than between the Nearctic and Palearctic). We used AIC to select the best-fitting model to the data and to produce the best reconstruction of swallow geographic range evolution.

### 2.6.2. Ancestral character estimation of sociality and nest type

We performed ancestral character estimation of sociality and nest type (1) to examine shifts in solitary versus colonial nesting, (2) to determine the most likely ancestral nest type and the sequence of mud-nest evolution, and (3) to relate shifts in nest type to swallow diversification and geographic range evolution. We reviewed species accounts in Turner (2018), categorizing sociality and nest type as discrete traits with the following categories for sociality: strictly solitary breeders, small-group breeders (which includes species that are facultatively solitary or breed in small assemblages), and colonial breeders; and the following categories for nest type: burrow-excavating, cavity- or burrow-adopting, constructing open-cup mud nests (e.g., *Hirundo*), constructing closed-cup mud-nests (e.g., *Delichon*), and constructing retort mud-nests (enclosed mud-nests with entrance tunnels: e.g., *Cecropis*). The Mascarene martin (*Phedina borbonica*) and mountain sawwing (*Psalidoprocne fuliginosa*) place a nest made of plant material on a ledge, in a crevice, or in a similar fashion, and were classified as cavity-adopters. We excluded eight taxa from the analysis that are of unknown sociality, one of which (white-eyed river martin, *Pseudochelidon sirintarae*) was also the only species of unknown nest type. We reconstructed character evolution of sociality and nest type simultaneously on the Bayesian time-calibrated tree, pruned to remove the eight taxa with missing trait data, using hidden Markov models in the package corHMM (Beaulieu et al., 2013) in R v4.3.1 (R Core Team, 2023). We allowed all transition rate parameters in the model to differ, and fit models with 0, 1 and 2 hidden traits, and used AICc scores to select the best model among these. Such hidden traits allow rate heterogeneity across the phylogeny to be modeled without necessarily ascribing all observed heterogeneity to the traits of interest (i.e., sociality and nest type), thereby relaxing the assumption that there is a single transition rate between two states across the entire tree (Beaulieu et al., 2013). We then visualized the ancestral reconstruction using a stochastic character map of the best model, with posterior probabilities of each character state plotted on internal nodes.

## 3. Results

### 3.1. Sequencing results, phylogenetic inference, and divergence dating

Characteristics of sequences produced from skin and preserved tissue samples differed (Supplementary Table S1). After trimming adapters and low-quality bases from raw reads, we obtained 2.71–43.82 (mean  $\pm$  SD = 11.86  $\pm$  11.18) million reads per skin sample and 2.0–19.57 (8.53  $\pm$  3.33) million reads per tissue sample, with mean trimmed read lengths of 115 bp and 146 bp among skin and tissue samples, respectively. The higher read counts for skin samples were driven by an extremely high number of reads from four samples (Supplementary Table S1). The proportion of reads removed during trimming was higher and more variable among skin samples (0.26 %–28.09 %) than tissue samples, which had at most 0.16 % reads removed. Trimmed reads were assembled into 1352–512,718 (mean  $\pm$  SD = 86,392  $\pm$  136,204) contigs per skin sample and 8323–223,554 (26,579  $\pm$  35,283) contigs per tissue sample. The number of UCE loci successfully enriched in tissue samples, including outgroups, ranged from 4503–4830 (mean  $\pm$  SD = 4767  $\pm$  44) with mean lengths from 705–1521 bp (1,072  $\pm$  132 bp). We recovered 544–4851 (3713  $\pm$  1480) UCE loci from the 16 skin samples retained after preliminary analysis.

Following the mapping and correction steps, mean locus lengths from the skin samples range from 161–1087 bp (385  $\pm$  273 bp). The full 118 sample dataset with 95 % complete data per locus includes 4,009

aligned loci ranging 416–1591 bp in length (mean = 976 bp) and containing 0–414 (mean = 106) informative sites, an average of 11 % missing characters, and 35 % GC content. The reduced dataset consisting of only tissue samples and *Pseudochelidon* (103 samples) with 95 % complete data per locus includes 4,518 aligned loci ranging 394–1973 bp in length (mean = 946 bp) and containing 0–402 (mean = 99) informative sites, an average of 5 % missing characters, and 37 % GC content. The further reduced dataset of only tissue samples (101 samples; used for species tree inference within Hirundininae) with 95 % complete data includes 4565 loci ranging 394–1973 bp in length (mean = 943 bp) with 0–400 informative sites (mean = 99), an average of 4 % missing characters, and 37 % GC content. The total aligned sequence lengths for the three datasets are 3.9 Mbp, 4.3 Mbp, and 4.3 Mbp, respectively. UCE locus dataset summaries are also provided in Table 1.

ML phylogenetic reconstruction using RAxML on the concatenated alignment yielded a well resolved tree, with strong support for most relationships to the species level (Fig. 1). Further, there is complete support (i.e., bootstrap support = 100) for the monophyly of each genus, except for the sister relationship between Brazza's martin (*Phedina brazzae*) and banded martin (*Neophedina cincta*), rendering *P. brazzae* and *P. borbonica* paraphyletic. Most historic DNA samples are characterized by long terminal branch lengths, even after correction for low data quality during UCE locus processing. This is an expected result for decades-old skin samples with comparatively degraded DNA (McCor-mack et al., 2016). However, the branch lengths for these samples do not appear to have strongly influenced the overall estimate of the tree topology, and there is strong support for their placement within respective genera, with the only exceptions being *Progne murphyi* and the Sinaloa martin (*P. sinaloae*; bootstrap support = 42 and 35, respectively). The coalescent-based estimate using SVDquartets is similarly well resolved, albeit with lower support at several deeper nodes, including the split between the river martins (Pseudochelidoninae) and Hirundininae and the split between *Psalidoprocne* and remaining Hirundininae (Supplementary Fig. S2). The inferred topology is also largely congruent with the concatenation estimate at the species level, with conflicting relationships concentrated at shallower nodes (14 %), the majority of which had low support (i.e., bootstrap support < 70) estimated from one or both approaches (Supplementary Fig. S3). We performed an enrichment test to determine if these conflicting relationships are influenced by the inclusion of historic DNA samples but find no evidence of enrichment for conflicting nodes relative to tissue samples (Fisher's Exact Test,  $P$ -value = 0.314). The concatenation and coalescent estimates are nearly completely concordant at the genus level, with the only exception being the placement of the red-throated swallow (*Petrochelidon rufigula*; a historic DNA sample) as sister to the *Petrochelidon* + *Cecropis* clade in the coalescent analysis; *P. rufigula* is strongly supported as sister to Preuss's swallow (*P. preussi*) in the concatenated analysis. The concatenation and coalescent topologies inferred from the reduced dataset are nearly identical to those from the full dataset (Supplementary Figs. S4-S5), except for minor disagreement at shallow nodes within *Hirundo* and *Riparia* and *Psalidoprocne* and *Hirundo* in the concatenated and coalescent analyses, respectively (all of which have relatively low support values in both full and reduced analyses). The Hirundininae species tree topology inferred using ASTRAL is further congruent with the genus-level relationships recovered from the other approaches (Supplementary Fig. S6), with strong local branch posterior probability support at the majority of deeper nodes.

Consensus from the concatenation and coalescent approaches (as

well as Bayesian inference; Supplementary Fig. S7) provides clear support for the ancestral split between the subfamilies Pseudochelidoninae and Hirundininae (Fig. 2). Further, there is strong support for four major clades within Hirundininae, including the saw-wing clade (*Psalidoprocne*), the mud-nesting clade (*Ptyonoprogne*, *Hirundo*, *Atrionanus*, *Delichon*, *Petrochelidon*, and *Cecropis*), the clade containing the monotypic genera *Pseudhirundo* and *Cheramoeca*, and finally the clade containing Old World excavators (*Phedina*, *Neophedina*, and *Riparia*) and New World endemics (*Tachycineta*, *Progne*, *Stelgidopteryx*, *Atticora*, *Pygochelidon*, *Orochelidon*, and *Alopochelidon*). Here, *Psalidoprocne* is supported as sister to the remaining clades and the *Pseudhirundo*/*Cheramoeca* clade is sister to the Old World excavators/New World endemics. The mud-nesting clade comprises three subclades, with an initial divergence between *Ptyonoprogne* + *Hirundo* and the other genera followed by a split between *Atrionanus* + *Delichon* and *Petrochelidon* + *Cecropis*. At shallower timescales, the consensus among approaches suggests additional lineage diversity within some clades that is not captured by the present classification. Examples include the nesting of *Psalidoprocne fuliginosa* and the white-headed saw-wing (*P. albiceps*) within the black saw-wing (*P. pristopectera*) clade, rendering *P. pristopectera* subspecies paraphyletic, the non-sister relationship between house martin (*Delichon urbicum*) subspecies, and the nesting of the striolated swallow (*Cecropis striolata*) within *C. daurica* (Fig. 1; Supplementary Figs. S2, S6).

The Bayesian time-calibrated tree supports a mid-late Miocene origin of Hirundinidae, with a median posterior divergence time estimate between Pseudochelidoninae and Hirundininae of 12.83 MYA (Supplementary Fig. S7). Other major splits are estimated to have occurred during the late Miocene, including between the saw-wings and other ingroup Hirundininae (median posterior estimate = 11.07 MYA), between mud-nesters and the clade containing *Pseudhirundo*/*Cheramoeca* + Old World excavators/New World endemics (10.31 MYA), and between *Pseudhirundo*/*Cheramoeca* and Old World excavator/New World endemic clades (9.7 MYA). Subsequent diversification of most genera occurred during the Pliocene–Pleistocene, though median estimates within *Pseudochelidon* and *Phedina*/*Neophedina* are comparatively ancient (6.06 and 7.19 MYA, respectively), during the very late Miocene.

### 3.2. Estimation of historical biogeography and trait evolution

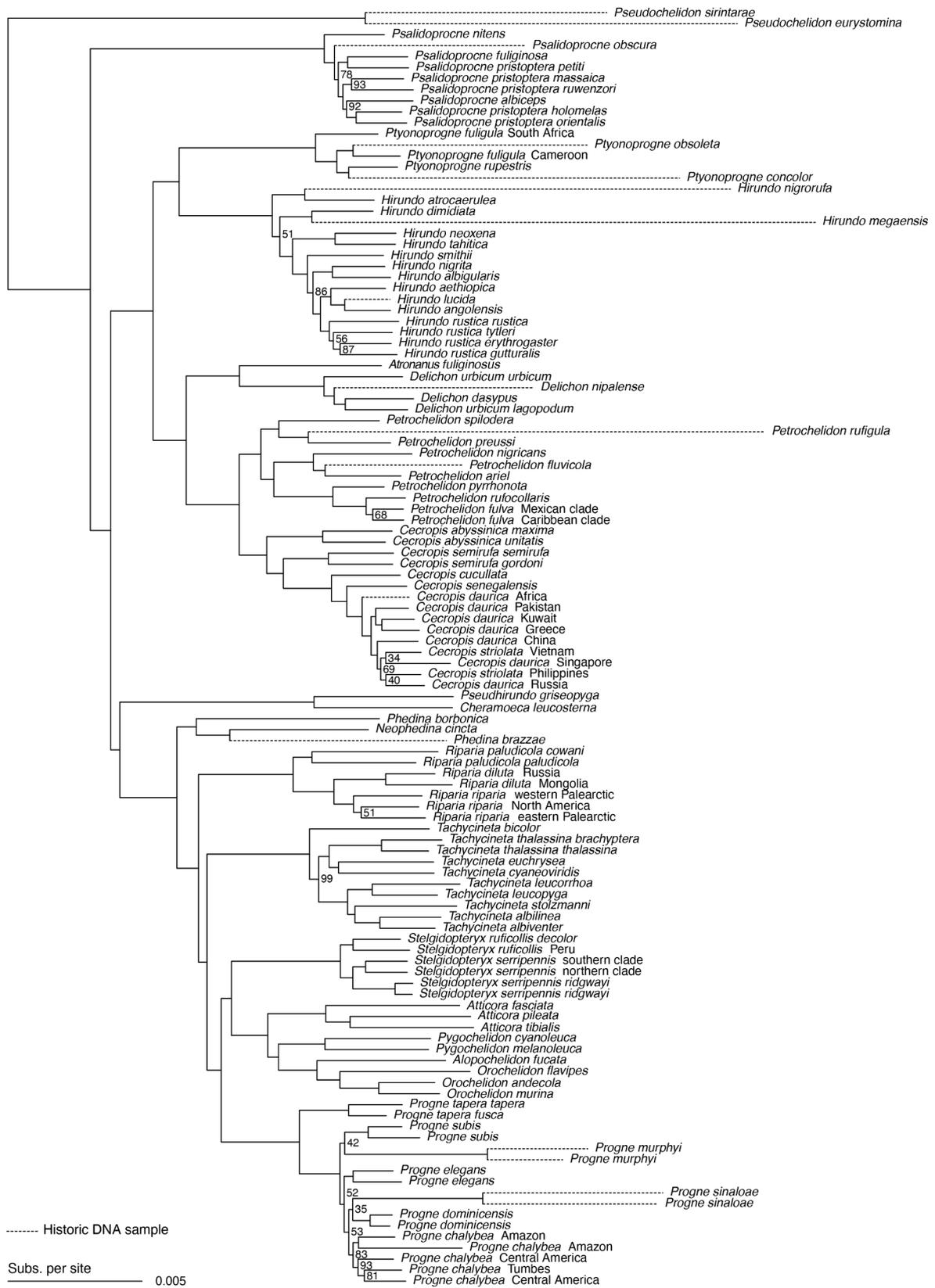
We investigated the evolution of geographic breeding ranges across Hirundinidae by evaluating support for three alternative biogeographic models, with and without an additional founder event speciation parameter (J). The inclusion of this parameter significantly increases the likelihood of the data under all three models (Table 2). The DEC + J and DIVALIKE + J models have similar support on the basis of AIC and Akaike Weights, and are each better fitting than the BAYAREALIKE + J model. The DEC model is appropriate for analysis of range evolution in swallows due to the global scale of the system and large geographic regions contained within the eight biogeographic realms considered in our analysis. Because the DEC + J is a similarly good fit to the data as the DIVALIKE + J model and has fewer underlying assumptions, we focus on the ancestral range reconstructions under the DEC + J model, even though results of both models are highly concordant (Supplementary Fig. S8).

Ancestral range estimation under the DEC + J model supports the hypothesis that swallows originated in Africa and reached the Palearctic through multiple independent colonization events (e.g., by the common

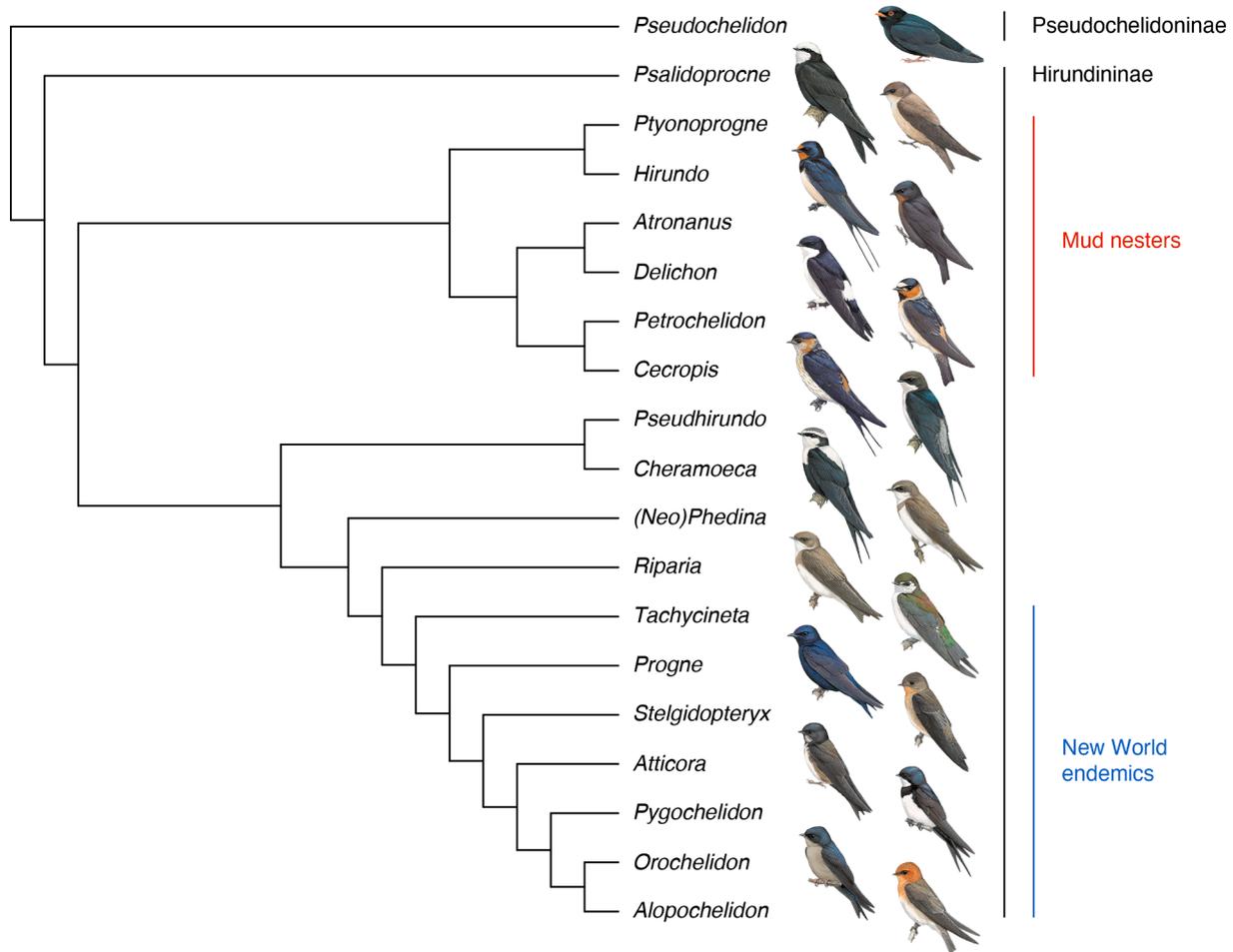
**Table 1**

Summary of UCE locus datasets used in this study under various sampling schemes. Mean values are shown for locus length, informative sites, and missing characters.

Dataset	$n$ taxa	$n$ loci	Locus length	Informative sites	Missing bp	Total Mbp
Full	118	4009	976	106	11 %	3.9
Reduced	103	4518	946	99	5 %	4.3
Reduced (no <i>Pseudochelidon</i> )	101	4565	943	99	4 %	4.3



**Fig. 1.** Phylogenetic relationships among Hirundinidae estimated using maximum likelihood on a concatenated matrix of 4009 UCE loci with 95 % complete data among 118 taxa. Nodal values indicate bootstrap support, and only values below 100 % are shown. Outgroup taxa are not shown. Historic DNA samples are denoted with dashed terminal branches.



**Fig. 2.** Consensus phylogenetic relationships among swallow genera based on concatenated maximum likelihood, coalescent species tree, and Bayesian approaches. The relationships shown are strongly supported across most, if not all approaches. Genera belonging to the two subfamilies Pseudochelidoninae and Hirundininae are labeled. *Phedina* and *Neophedina* form a clade, yet species-level relationships render *Phedina* paraphyletic, so this branch is labeled *(Neo)Phedina*. Swallow illustrations by Hilary Burn © Lynx Edicions.

**Table 2**

Support for models of geographic range evolution fitted to the time-calibrated Hirundinidae tree, based on classification of presence or absence of taxa across eight biogeographic realms. Three basic models are compared: 1) dispersal, local extinction, and cladogenesis (DEC), 2) likelihood interpretations of dispersal-vicariance analysis (DIVALIKE), and 3) BAYAREA (BAYAREALIKE). All models were fitted with and without an additional founder event speciation parameter (*J*). The number of free parameters, parameter estimates (i.e., rates of range expansion, *d*, range contraction, *e*, and jump dispersal weight, *j*), Akaike information criterion score (AIC), and Akaike weights are given for each model.

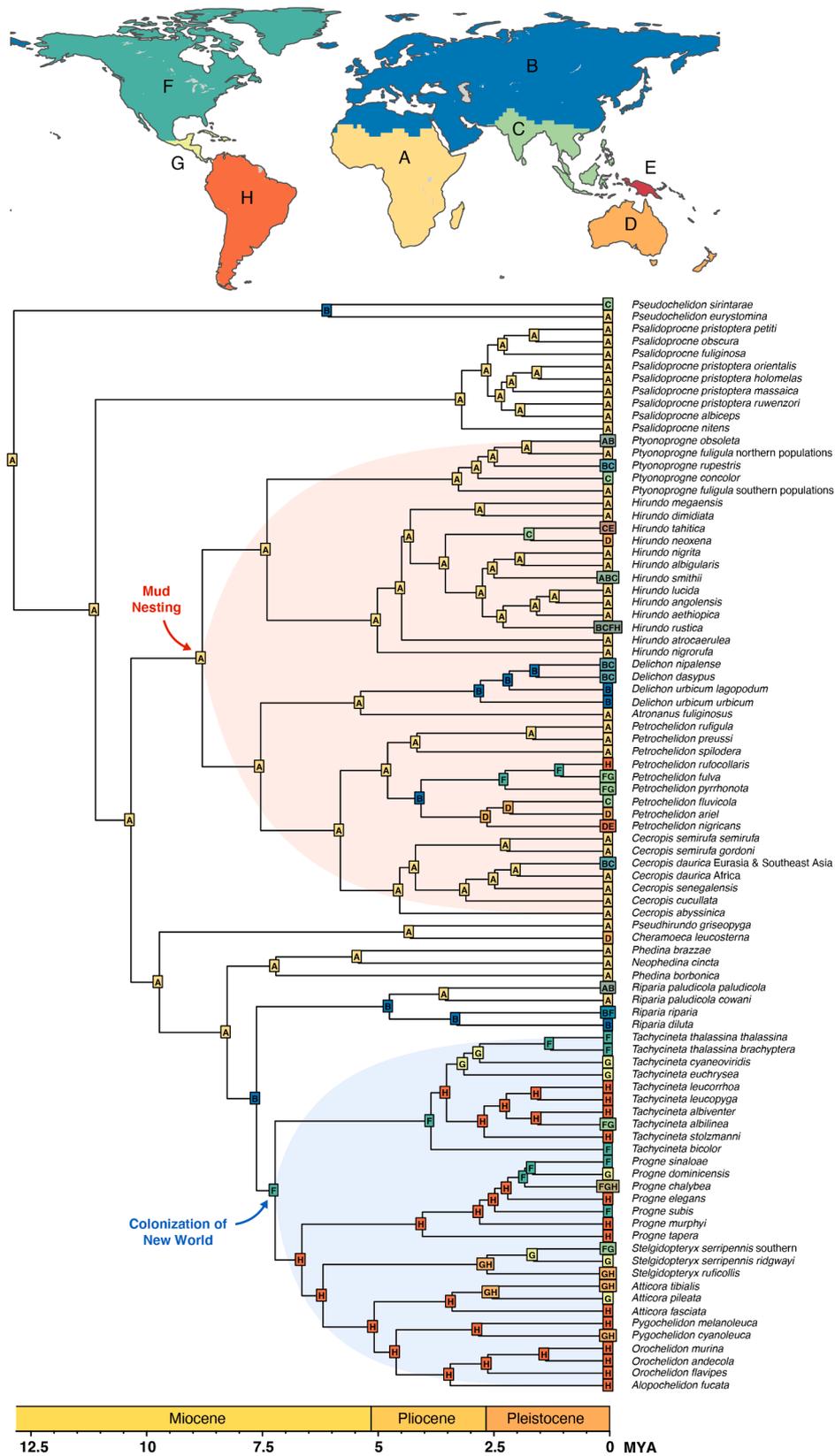
Model	Parameters	LnL	<i>d</i>	<i>e</i>	<i>j</i>	AIC	Akaike weight
DIVALIKE + <i>J</i>	3	-175.6	0.037	0.0069	0.078	357.2	0.5
DEC + <i>J</i>	3	-175.9	0.032	0.0067	0.089	357.8	0.39
BAYAREALIKE + <i>J</i>	3	-177.2	0.026	0.0072	0.1	360.4	0.11
DIVALIKE	2	-190.8	0.06	0.01	0	385.6	$3.6 \times 10^{-07}$
DEC	2	-193.9	0.05	0.015	0	391.8	$1.6 \times 10^{-08}$
BAYAREALIKE	2	-223.3	0.049	0.17	0	450.7	$2.6 \times 10^{-21}$

ancestors of *Delichon*, *Petrochelidon*, and *Riparia* + New World endemics, respectively, and multiple times within *Hirundo*; Fig. 3; Supplementary Fig. S8). Dispersal into the Australian and Oceanian realms also occurred multiple times (e.g., within *Hirundo*, *Petrochelidon*, and *Cheramoeca*), although there has not been a substantial degree of subsequent diversification within these regions. In contrast, much of extant swallow diversity resulted from *in situ* diversification within the Afrotropics, including the African-endemic *Psalidoprocne* and within *Hirundo* and *Cecropis*. Our reconstructions support an African origin of mud nesting. Major colonization of the New World from the Palearctic occurred during the late Miocene (~7.5 MYA), with subsequent (primarily Neotropical) diversification giving rise to *Tachycineta*, *Progne*,

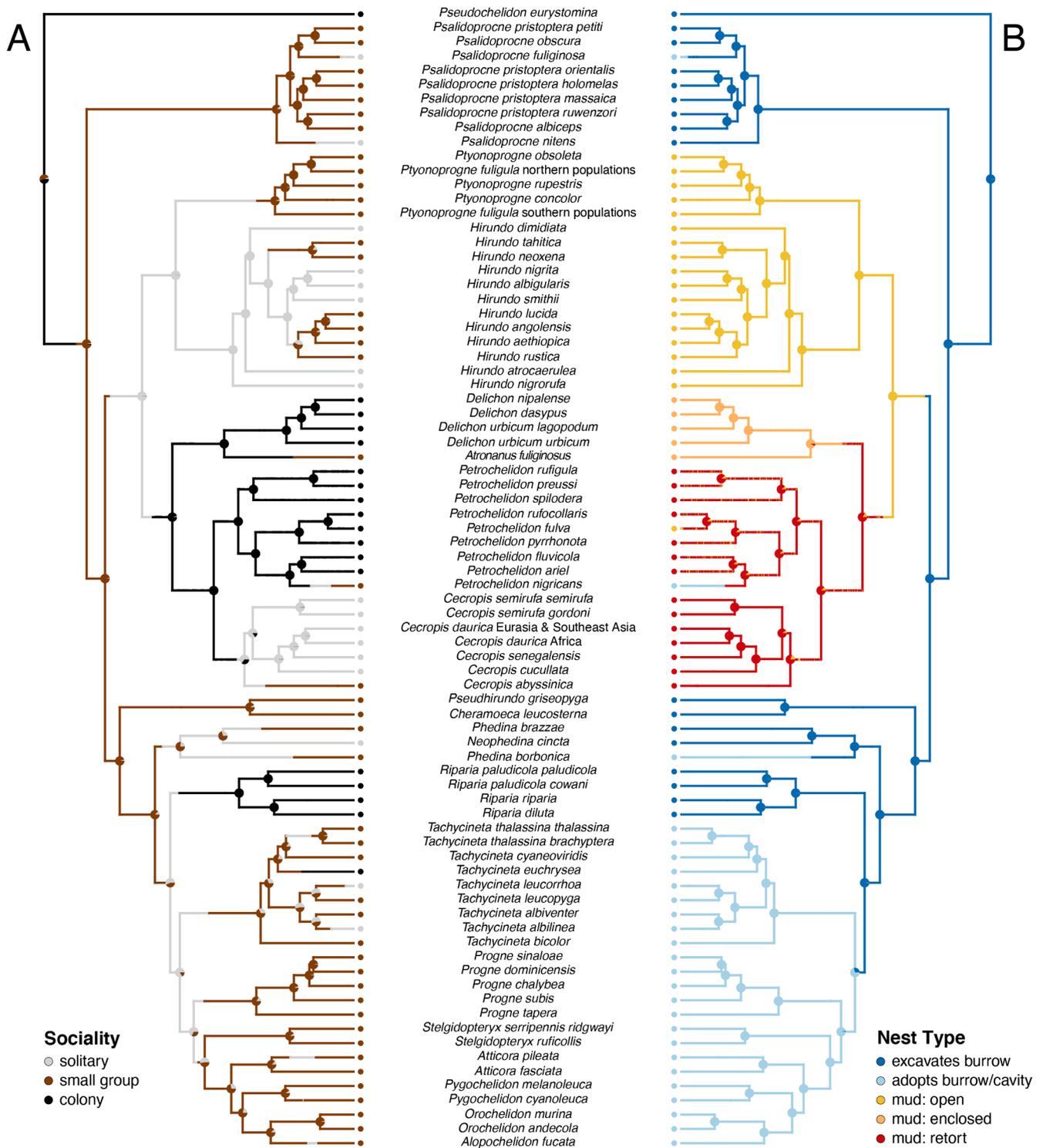
*Stelgidopteryx*, *Atticora*, *Pygochelidon*, *Orochelidon*, and *Alopochelidon*. Later colonizations of the New World from the Palearctic occurred within *Hirundo rustica*, *Riparia riparia*, and *Petrochelidon* species during the Pliocene and Pleistocene.

We explored the evolution of swallow sociality and nest type by testing models of trait evolution and reconstructing trait histories across the Hirundinidae phylogeny (Fig. 4). The model including one hidden trait (AICc = -180.1) was a better fit than the model with no hidden traits (AICc = 12619.6) but adding a second hidden trait (AICc -108.3) did not result in an improved fit over one hidden trait. We therefore report the results from the model with one hidden trait.

Sociality appears to be a much more evolutionarily labile trait than



**Fig. 3.** Historical biogeography of Hirundinidae estimated under a DEC + J model based on the time-calibrated Bayesian inference topology. In the model, dispersal is constrained to not occur between the Afrotropical and either Australian or Oceanian realms or between the Old and New World, except for between the Palearctic and Nearctic realms. The observed ranges of extant taxa are shown at the tips of the topology, with estimated ancestral ranges shown at the internal nodes. The red and blue arrows indicate the origin of mud nesting and the colonization of the New World, respectively. The eight biogeographic realms in the analysis are abbreviated: A, Afrotropical; B, Palearctic; C, Oriental; D, Australian; E, Oceanian; F, Nearctic; G, Panamanian; H, Neotropical. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Ancestral reconstruction of A) sociality and B) nest type in the Hirundinidae for 80 taxa for which trait data are available. Branches are color-coded by trait to show a stochastic character map based on the ancestral reconstruction model. Pies at internal nodes show posterior probabilities of each trait; pies at the tips show the trait data for extant taxa.

nest type (Fig. 4). The root ancestral sociality phenotype is uncertain, with high probabilities of both coloniality and small group nesting in the swallow common ancestor. Following the split from *Pseudochelidoninae*, the common ancestor of Hirundininae was likely small-group nesting (Fig. 4A). Our results support that solitary nesting has arisen at least six times independently, and coloniality at least three times. However, we note that ancestral node probabilities are equivocal between two states in some areas of the tree (e.g., solitary and small group

nesting in the *Tachycineta* clade). There was overall more uncertainty in the ancestral estimation of sociality than nest type, with the lowest maximum posterior probability at a node being 49.5 %, and 16 of the 79 nodes having < 80 % maximum posterior probability.

The ancestral character estimation of nest type recovers burrow excavation as the ancestral state (Fig. 4B). Overall, there are a few major transitions in nest type, and the trait is otherwise relatively evolutionarily stable. We recover a single transition to building mud-nests, with

open cup mud-nests being the likely ancestral state. Retort mud-nests likely first occurred in the ancestor of *Delichon* + (*Petrochelidon* + *Cecropis*), while enclosed mud-nests originated once in the ancestor of *Atrionanus* + *Delichon*. *Petrochelidon fulva* reverted to open cup nesting from a mud-retort ancestor, this being the only other origin of open cup mud nesting. Burrow and cavity adoption appears to be comparatively labile, having originated four times independently. We infer that the first origin of cavity adoption was from a burrow-excavating lineage following the divergence between *Riparia* and New World endemics. Cavity adoption also originated in the *Phedina borbonica* lineage (from an excavating ancestor), the Tree Martin (*Petrochelidon nigricans*) lineage (from a mud retort ancestor), and the *Psalidoprocne fuliginosa* lineage (from an excavating ancestor). Although *P. nigricans*, as a member of the mud nesting clade, is unique in adopting natural cavities for nesting sites, it nevertheless uses mud to dam the entrance to its nest hole. In general, our reconstructions estimate high certainty for nest type in ancestral nodes, with maximum posterior probabilities at each node >70 % in all cases, and >95 % in most cases.

## 4. Discussion

### 4.1. Phylogeny

We reconstructed the phylogeny of the swallow family, Hirundinidae, using the most comprehensive taxonomic sampling to date and a genome-scale dataset including thousands of UCE loci. We found broad consensus among concatenated ML, coalescent-based, and Bayesian trees in branching to the genus level (Figs. 1-2; Supplementary Figs. S2-S7). These trees agree both when historic DNA samples are included and excluded. Perhaps most importantly, the consensus phylogenetic hypothesis clarifies swallow relationships at deeper nodes, which have proved especially problematic in previous analyses (Sheldon and Winkler, 1993; Sheldon et al., 2005).

The structure of the swallow phylogeny is as follows. Pseudochelidoninae, comprising the two river martin species (*Pseudochelidon*), is sister to all other swallows (Hirundininae). These two species have a relictual distribution, one in west-central Africa (*P. eurystomina*) and the other in Indochina (*P. sirintarae*). *Pseudochelidon eurystomina* nests in excavated burrows; nesting of *P. sirintarae* is unknown. Hirundininae is divided into four main clades. (1) The saw-wings (*Psalidoprocne*), of which there are 5 species and 16 subspecies, are the sister group to the rest of Hirundininae. They are restricted to sub-Saharan Africa and nest in excavated burrows. (2) A clade of largely Old World species that construct mud-nests. This group includes 39 species in four genera (*Hirundo*, *Delichon*, *Cecropis*, and *Petrochelidon*). (3) A clade consisting of the white-backed swallow (*Cheramoeca leucosterna*) of Australia and the gray-rumped swallow (*Pseudhirundo griseopyga*) of sub-Saharan Africa. These two species are nest-hole excavators with a presumably relictual distribution (i.e., not the result of dispersal across the Indian Ocean). And (4) an eclectic group comprising the Mascarene species *Phedina borbonica*, the African species *Neophedina cinta* and *Phedina brazzae*, the largely Old World genus *Riparia* (comprising five species), and seven genera endemic to the Nearctic and Neotropics (*Tachycineta*, *Atticora*, *Pygochelidon*, *Alopochelidon*, *Orochelidon*, *Stelgidopteryx*, and *Progne*). The New World endemics build their nests in existing holes or crevices. The *Riparia* species, *Phedina brazzae*, and *Neophedina* excavate nest burrows, and *P. borbonica* builds its nest in crevices or on ledges.

Resolution of all generic relationships within the Hirundininae is notable because the branching pattern among some groups was unclear in previous molecular studies. In particular, while earlier studies discovered the remarkable long-distance relationship between *Cheramoeca* and *Pseudhirundo*, the position of this clade in the swallow tree was equivocal, with alternative data and analyses yielding sister relationships between it and saw-wings, or mud-nesters, or the New World endemics (Sheldon and Winkler, 1993; Sheldon et al., 2005). In most of our analyses, the latter relationship was strongly supported (e.g., ML

bootstraps of 100; Bayesian posterior probabilities of 100 %). The only exception was ASTRAL analysis using preserved tissue samples only (posterior probability = 0.52; Supplementary Fig. S6). Broadly, the results from our UCE dataset provide a major advance in resolving relationships at deep nodes that have been previously problematic.

Within the major clades of Hirundininae, all genera (except *Phedina*) are supported as monophyletic. Moreover, we confirm that the African mud-nesting forest swallow (*Atrionanus fuliginosus*) is sister to *Delichon* rather than the cliff swallows (*Petrochelidon*). The generic assignment of this taxon was in doubt until the recent acquisition of modern specimens, and subsequent comparisons of three genes by de Silva et al. (2018) led to the designation of a new genus, *Atrionanus*, and discovery of its sister relationship to *Delichon*. At the intrageneric level, relationships among some taxa can be complicated by population genetic processes, such as ongoing or recent gene flow, and poor geographic sampling, and we experienced mixed results in resolving some of these problems. The saw-wings (*Psalidoprocne*) are paramount in this respect. Our branching pattern disagrees with established classifications (Fig. 1; Supplementary Figs. S2-S7), as does that of Barrow et al. (2016), which is based on better taxonomic (if not genomic) sampling. Although the square-tailed saw-wing (*P. nitens*) is clearly sister to other saw-wing species, based on both morphology and molecules (Barrow et al., 2016; our study), relationships among many other taxa, among them the 12 subspecies of black saw-wing (*P. pristoptera*), remain uncertain. Another problematic genus is *Cecropis*, the group that makes enclosed mud-nests with long entrance tunnels. Our findings, for example, contradict the division of *C. daurica* into red-rumped swallow (*C. daurica*) and striated swallow (*C. striolata*), as suggested by Dickinson and Christidis (2014). Although our sampling of these taxa is not extensive, all analyses indicate that *C. striolata* is nested within *C. daurica* (Fig. 1). With respect to some other intrageneric splits, our tree supports the division of *Delichon urbicum* into western house martin (*D. urbicum*) and Siberian house martin (*D. lagopodum*) (del Hoyo et al., 2023), and the separation of Ridgway's rough-winged swallow (*Stelgidopteryx ridgwayi*) from the northern rough-winged swallow (*S. serripennis*), as done originally by Ridgway (1904) and subsequently by Phillips (1986).

### 4.2. Origins and diversification of the swallow family

Of 88 total taxa included in our biogeographic analyses, 39 (44 %) breed in the Afrotropics and 36 breed exclusively in Africa. Our comparisons indicate that this concentration of diversity results from substantial *in situ* diversification, with major pulses during the Miocene and Pliocene. These events include diversification within the saw-wings, as well as the origin of the common ancestor of *Cheramoeca* + *Pseudhirundo* and the clade containing *Phedina*, *Neophedina*, *Riparia* and New World endemics. Perhaps the most notable case of *in situ* diversification in Africa is the origin and evolution of mud-nesting species, with mud nesting itself representing a major innovation in swallow evolution (Winkler and Sheldon, 1993; Sheldon et al., 2005) which we discuss in more detail below. Nearly 60 % of the species breeding in the Afrotropics are mud-nesters, and there are particularly high rates of African endemism within *Hirundo* and *Cecropis*. Several mud-nesting lineages subsequently dispersed independently to Eurasia, then into South and Southeast Asia (Fig. 3). A few of these lineages reached Australia and Oceania, which are currently occupied by two species each of *Petrochelidon* and *Hirundo*; the only other species within the Australian realm is *Cheramoeca leucosterna*, resulting from what was likely a much earlier independent expansion into this region during the Pliocene.

The second most species-rich biogeographic realm is the Neotropics, in which 18 swallow species have breeding distributions, roughly half that of the Afrotropics. Most of these species constitute the clade of New World endemics, the exceptions being *Petrochelidon ruficollaris* and *Hirundo rustica*. We infer that swallow diversity in the Neotropics arose following the colonization of the Nearctic from the Palearctic. We note

that in our biogeographic models we constrained dispersal between the Old and New World to occur only between the Palearctic and Nearctic realms. While mid-oceanic dispersal of avian species is possible, as observed famously in Cattle Egrets (Crosby, 1972), this is a rare phenomenon. In contrast, a land connection or near-connection between North America and Siberia across the Bering Strait was present during much of the Miocene (Hopkins, 1959), and the colonization of the Nearctic by multiple bird families from the Palearctic has long been recognized (Mayr, 1946) and supported in subsequent analyses (e.g., Moyle et al., 2016; Sangster et al., 2022). Moreover, because modern swallows rarely migrate or disperse over long over-water distances (Turner, 2018), the Palearctic-Nearctic dispersal constraint is a reasonable assumption. With this constraint in place, we estimate that the major dispersal event to the Nearctic occurred during the late Miocene roughly 7.5 MYA, giving rise to *Tachycineta*, *Progne*, and other lineages largely endemic to the Neotropics. Dispersal events from the Palearctic within *Petrochelidon* and in *Hirundo rustica* and *Riparia riparia* occurred much later (Zink et al., 2006; Pavlova et al., 2008).

Colonization of the Neotropics within the cavity adopting clade appears to have been by Nearctic rather than Panamanian ancestors. The direct colonization of the Neotropics from the Nearctic realm is plausible, based both on seasonal migration of numerous Nearctic-breeding taxa to Neotropical non-breeding regions, and on the recent establishment of a breeding population of *Hirundo rustica* in Argentina (Martínez, 1983; Winkler et al., 2017). We infer that dispersal into the Panamanian region by swallows represents secondary independent colonization events from the Neotropics (e.g., by *Tachycineta* and *Stelgidopteryx*) or direct dispersal by Nearctic ancestors, as in the colonization of the Caribbean by the Caribbean Martin (*Progne dominicensis*) (Fig. 3). The beginning of the Pliocene, at 5.3 MYA, is roughly coincident with the gradual uplift of the Isthmus of Panama, which completed the land connection between North and South America by about 3 MYA (Smith and Klicka, 2010; O’Dea et al., 2016). This timing approximately matches the inferred dispersal to the Panamanian region by cavity-adopting lineages.

The location and timing of mud-nester and cavity-adopter radiations are straightforward to detect because they are consolidated geographically and a large number of species are involved. Identifying the center of origin of the swallow family, however, is more difficult. No closely related sylviid lineages exist for comparison (Oliveros et al., 2019), and there are only two species of river martins, which have disjunct, relictual distributions in Africa and Indochina. Recognizing these limitations, our estimates of divergence time and historical biogeography support swallow origination in the Afrotropics during the Miocene, roughly 13 MYA (Fig. 3; Supplementary Fig. S7). Alternative hypotheses have considered the possibility that the geographic range of the swallow ancestor may have been more expansive, covering Eurasia as well as Africa, with the center of origin at any point in this range. Sheldon et al. (2005), for example, hypothesized a broad range for ancestral swallow populations based on the distribution of the two river-martins in Africa and Indochina and the apparently relictual distribution of the relatively old, sister genera *Cheramoeca* and *Pseudhirundo* in Australia versus Africa. This ancestral distribution would have encompassed Africa and southern Asia, leading to the invasion of Australia. Moyle et al. (2016) also inferred a broad ancestral distribution for the swallow common ancestor, encompassing much of Africa and Asia. Their inference, however, was based on sampling of only two swallow species, one of which was *Hirundo rustica*, which has the largest breeding distribution of all swallows.

The distribution of *Pseudochelidon* species presents an especially intriguing alternative to an Afrotropical origin. With its extant species in central Africa and southern Asia, a common ancestor in the Palearctic is a possible inference. Yet, it seems equally plausible that there was an independent expansion out of the Afrotropics in the ancestor of *Pseudochelidon* and extinction of geographically intermediate populations. We also note that *P. sirintarae* was very rarely observed and only at a

single locality in Thailand during the putative non-breeding season (Thonglongya, 1968); thus, there is some uncertainty about the extent of its range and a reasonable possibility that this species is now extinct (Turner, 2018). The *Cheramoeca* + *Pseudhirundo* clade, as a member of the Hirundininae, is younger than *Pseudochelidon*, and thus unlikely to have played a role in the origin of swallows. However, it holds an interesting ancestral position in the hirundinine tree, having branched soon after the saw-wings as sister to the large clade of older African lineages (*Phedina Neophedina*, and *Riparia*) and all New World groups. This hints at a broad early distribution. Given that our current analysis dissects the historical biogeography of swallows using data from the majority of swallow species, we consider it to be the most robust assessment in support of an African origin. Nevertheless, we recognize that African and south Asian passerines moved back and forth across that part of the world (Fuchs et al., 2012; Fjeldså 2013), and even into the New World (Sangster et al., 2022), essentially instantaneously in geological time, and that our analysis is strongly influenced by the preponderance of swallow species currently residing in Africa. This modern distribution may not have been the situation in the past.

#### 4.3. Major innovations and the evolutionary lability of traits

Our reconstruction of nesting behavior is largely concordant with Winkler and Sheldon (1993): nest-building behavior is generally highly conserved; the ancestral state is burrowing; there was one major origin each of mud nesting (probably in Africa) and cavity adopting (in the New World endemics); and three independent shifts to cavity adopting in single-species lineages. Evolutionary conservation of broad nest-type categories (e.g., cavity, cup, or burrow) is typical at the family level in all birds (Sheldon and Winkler, 1999; Turner, 2018), with some notable exceptions, as in the ovenbirds (Furnariidae; Zyskowski and Prum, 1999). Our findings suggest that the origin of mud nesting – a relatively rare nest-construction phenotype – was a key innovation (Heard and Hauser, 1995; Sheldon and Whittingham, 1997) coincident with the founding of the clade comprising over 40 % of extant swallow species. Both the ancestral swallow nest type (burrow or cavity excavation) and mud nesting are unusual in passerines, which typically build a vegetative cup nest or, less commonly, adopt cavities. While other passerines incorporate mud into their nests to varying degrees, pure mud pellet nests of the type constructed by swallows are unique (Winkler and Sheldon, 1993).

The mud nesting clade appears to have originated during the mid-Miocene in the Afrotropics. Appropriate nesting sites for burrow-excavating swallows, which require sandy soil, are inherently limited (Silver and Griffin, 2009), and the evolutionary innovation of mud-nest construction likely allowed swallows to escape that limitation and expand into areas and habitats where appropriate burrow sites were scarce. The availability of mud, which is often collected from puddles, depends on sufficient rainfall during the nest building period, but excessively wet and humid conditions make nest construction more difficult, and may even cause completed nests to crumble (Emlen, 1954). The mid- to late Miocene saw a decrease in rainfall in much of Africa (Feakins and Demenocal, 2010), and this drying trend may have facilitated the radiation of the mud-nesting clade. This period was also characterized by a worldwide expansion of open, C4-dominated habitats, such as grasslands (Cerling et al., 1997), in which swallows thrive.

The origins and transitions between different nesting types also presents the opportunity to consider selective pressures and behaviors associated with nest morphology in a phylogenetic context. When nest construction behavior is viewed as a genetic process, nest morphology is a complex phenotype characterized by some combination of shape, material, and site or substrate. As such, categorization tends to differ across studies (Hall et al., 2013; Price and Griffith, 2017; Nagy et al., 2019) and becomes complicated at large taxonomic scales that necessarily include more diverse nest phenotypes (Fang et al., 2018). Nest type is also subject to a variety of evolutionary pressures, including: egg

and clutch traits (Nagy et al., 2019); predators, parasites, and brood parasites; sexual selection; availability and competition for sites and materials; and the effects of climate on nest materials and nest microclimate (Perez et al., 2020). Little is known about the genetic mechanisms or evolutionary development of nest building and morphology (reviewed in Healy et al., 2023) to help provide context for hypotheses about the frequency or direction of evolutionary transitions in nest type. However, these trajectories exist (Winkler and Sheldon, 1993; Sheldon and Winkler, 1999). Studies have found a genetic basis for the complex burrowing behavior of mice during nest construction (Weber et al., 2013), and this may provide a starting point for genetic studies of nest construction behavior in other animals.

We find that the original mud-nest was likely the open-cup form (Fig. 4). Our results differ from Winkler and Sheldon (1993) in that we found the retort nest likely evolved from the open-cup and preceded the enclosed-cup nest. However, this difference relies upon character estimation at a single node (the common ancestor of retort and enclosed cup-nesting clades). Enclosed mud-nests differ from retort nests in that they lack an entrance tunnel; thus, the transition to enclosed nesting in *Delichon* + *Atrionanus* could be viewed as a loss of tunnel building. Such an evolutionary trajectory of nest ‘simplification’ may not be unusual; for example, open cup nesting repeatedly evolved from enclosed dome nesting in early passerines (Price and Griffith, 2017). More research on the genetic and developmental mechanisms of nest construction is needed to understand whether these transitions can be understood as gains or losses of behavioral modules.

The major dispersal event to the New World took place in the latter part of the Miocene by the common ancestor of *Tachycineta*, *Progne*, and the other Neotropical endemics, coincident with the inferred transition of nest type from burrow excavation to secondary cavity adoption. As with mud nesting in the Afrotropics, the transition from burrow excavation to cavity adoption possibly facilitated the radiation of swallows in the Nearctic and Neotropics. Winkler and Sheldon (1993) suggested that secondary cavity adopters were able to exploit the region’s widespread forests and the cavities of their rich primary cavity-excavating avifauna, and that the formation of major mountain ranges and resulting disjunct areas during the Miocene provided conditions conducive to diversification. Also, some taxa (e.g., *Orochelidon*, *Alopochelidon*, and *Pygochelidon*) build their nests in crevices and niches in rock faces, rather than adopting tree holes. These would have further benefited from mountain building. Extant richness of primary cavity excavators (relative to secondary cavity adopters) is high in the Nearctic, but less so in the Neotropics (van der Hoek et al., 2020), where most of the New World swallow diversity is currently found. It is possible that ecological opportunity for cavity adopters in the Neotropics was higher in the late Miocene than today due to subsequent niche filling. Further, the ratio of primary:secondary cavity nesters is also very high in central Africa, yet cavity adoption occurs in only one swallow species there (*Psalidoprocne fuliginosa*). The circumstances surrounding the Neotropical swallow radiation are intriguing and merit further study.

In contrast to nest construction, transitions in sociality were common within major swallow clades. This is consistent with high evolutionary lability of coloniality found across birds (Rolland et al., 1998). In particular, solitary nesting frequently evolved from small-group nesting, and once from colonial nesting at the origin of *Cecropis*. Small-group nesting also evolved multiple times from solitary ancestors (twice within *Hirundo*). Interestingly, coloniality evolved directly from solitary nesters, evidently without transitioning through small group nesting in the ancestor of *Delichon* and *Petrochelidon* + *Cecropis* and possibly also in the ancestor of *Riparia* (although certainty of the solitary ancestor is lower in this case). Colonial nesting is most common in the mud-nesters, although burrow excavators (*Riparia* and *Pseudochelidon eurystomina*) and one cavity-adopter (*Tachycineta euchrysea*) also nest colonially. Burrowing species may have also tended to nest in groups due to scarcity of suitable vertical nesting sites, limited by both landscape and soil types. Coloniality is rare among secondary cavity adopters in many bird

species (Eberhard, 2002), presumably because this would require the availability of a highly clumped distribution of cavities provided by primary excavators.

#### 4.4. Future directions

Although we find strong support for generic relationships, we also observe several instances of discordance between concatenated ML and coalescent-based species tree approaches at the species level (Supplementary Fig. S3). In some cases, conflicting relationships were highly supported in both approaches. These discrepancies point to interesting avenues of future inquiry, especially into ongoing population genetic processes (e.g., incomplete lineage sorting and introgression) in such groups as saw-wings, several mud-nesting groups, rough-wings (*Stelgidopteryx*), and tree swallows (*Tachycineta*). An additional consideration is that data from UCEs may be less helpful at resolving relationships at very shallow evolutionary timescales (Carter et al., 2023).

The highly supported, time-calibrated species level phylogenetic tree presented here provides the opportunity for future comparative studies of trait evolution. Swallows are an excellent system in which to explore trait innovation and its relationship to diversification, due to their global distribution and radiations in both Africa and the Neotropics. The evolution of migratory strategy is of particular interest in this group; there are multiple independent origins of long-distance migration in the Hirundinidae, as well as intraspecific variation in migration behavior (e.g., migratory divides on three continents in the *Hirundo rustica* species complex [Hobson et al., 2012,2015; Scordato et al., 2020; Turbek et al., 2022], and the migratory falloff and seasonal breeding reversal in Argentine *H. rustica erythrogaster* [Garcia-Perez et al., 2013]). It is plausible that variation in migratory traits (e.g., timing, route, and distance) has had a substantial influence on swallow diversification, with long-distance migration presenting a means for dispersal and colonization of new geographic regions followed by *in situ* speciation events. Hirundinidae also includes numerous non-migratory species or species that exhibit short-distance dispersal, and natural variation among swallows in sedentary versus migratory life histories presents a useful framework for testing whether migration is a catalyst or deterrent to diversification. The detailed biogeographic hypotheses provided here lend important context for the evolution of migratory behavior that can be leveraged in future studies.

## 5. Data Accessibility

Data supporting the conclusions of this study are available from the NCBI short-read archive (accession PRJNA1117501). The computational workflow and associated analysis scripts used in this work are available at [https://github.com/drewschield/hirundinidae\\_phylogeny](https://github.com/drewschield/hirundinidae_phylogeny).

### CRediT authorship contribution statement

**Drew R. Schield:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Clare E. Brown:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Subir B. Shakya:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Gina M. Calabrese:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Rebecca J. Safran:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Frederick H. Sheldon:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data

curation, Conceptualization.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data and code used in this study are freely available via public repositories (NCBI and GitHub).

## Acknowledgments

The specimens and tissues used in this study were collected by dedicated field workers from the Academy of Natural Sciences of Philadelphia, American Museum of Natural History, Cornell University Museum of Vertebrates, Field Museum of Natural History, Louisiana State University Museum of Natural Science, Museum of Vertebrate Zoology, Natural History Museum of Los Angeles County, Smithsonian Institution, University of Kansas Natural History Museum, University of Michigan Museum of Zoology, University of Washington Burke Museum, and Yale Peabody Museum, with essential help from colleagues in the Americas, Europe, Africa, Asia, and Australia. Data collection for this work was supported by the OpenWings project (<https://www.openwings.org/>) and associated National Science Foundation (NSF) grants (DEB-1655559, DEB-1655624, DEB-1655683, and DEB-1655736). The project was also supported by NSF grants DBI-1906188 to DRS, DBI-2209166 to GMC, IOS 1856266 to RJS, DEB 1241059 to FHS, and by OpenWings, Louisiana Board of Regents, American Ornithologists Union, LSU Biograds, and LSU Museum of Natural Science grants to CEB. We are grateful to David Winkler for providing the early inspiration for this work and helping to collect many of the swallows used in the study. We are also grateful to Anna Hiller, Brant Faircloth, Jessie Salter, and Carl Oliveros for assistance with data generation and suggestions for analyses. We thank Daniel Field and Daniel Ksepka for information about swallow fossils and suggestions for calibrations in divergence time estimation.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2024.108111>.

## References

- Babin, M.J., 2005. Geographic Variation and Speciation in Roughwinged Swallows (Hirundinidae: Stelgidopteryx). Louisiana State University and Agricultural & Mechanical College.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Pribelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477.
- Barrow, L.N., Dalton, D.L., Kotze, A., Evans, S.W., 2016. Geographically widespread mitochondrial lineages of the African saw-wings inconsistent with species boundaries. *Ostrich* 87, 271–275.
- Beaulieu, J.M., O'Meara, B.C., Donoghue, M.J., 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in Campanulid Angiosperms. *Syst. Biol.* 62, 725–737.
- Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S., 2020. *Swallows* (Hirundinidae). BirdLife International. 2021. IUCN Red List for birds. <http://www.birdlife.org>.
- Bivand, R., Lewin-Koh, N., 2018. mapproj: Tools for handling spatial objects. R package.
- Bivand, R., Rundel, C., 2018. rgeos: Interface to geometry engine - Open source ('GEOS'). R package.
- Bivand, R., Keitt, T.H., Rowlingson, B., 2018. rgdal: Bindings for the Geospatial Data Abstraction Library. R package.
- Bivand, R.S., Pebesma, E., Gómez-Rubio, V., Bivand, R.S., Pebesma, E., Gómez-Rubio, V., 2013. Spatial data import and export. *Appl. Spat. Data Anal.* with R 83–125.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537.
- Carter, J.K., Kimball, R.T., Funk, E.R., Kane, N.C., Schield, D.R., Spellman, G.M., Safran, R.J., 2023. Estimating phylogenies from genomes: a beginners review of commonly used genomic data in vertebrate phylogenomics. *J. Hered.* 114, 1–13.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Cerasale, D.J., Dor, R., Winkler, D.W., Lovette, I.J., 2012. Phylogeny of the *Tachycineta* genus of New World swallows: Insights from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 63, 64–71.
- Cerling, T.E., Harris, J.M., MacFadden, B.J., Leakey, M.G., Quade, J., Eisenmann, V., Ehleringer, J.R., 1997. Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389, 153–158.
- Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324.
- Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J. Theor. Biol.* 374, 35–47.
- Clements, J.F., Rasmussen, P.C., Schulenberg, T.S., Iliff, M.J., Fredericks, T.A., Gerbracht, J.A., Lepage, D., Spencer, A., Billerman, S.A., Sullivan, B.L., Wood, C.L., 2023. The eBird/Clements checklist of Birds of the World v2023.
- Crosby, G.T., 1972. Spread of the cattle egret in the Western Hemisphere. *Bird-Banding* 205–212.
- de Silva, T.N., Fernando, S.W., Robbins, M.B., Cooper, J.C., Fokam, E.B., Peterson, A.T., 2018. Recognition of a new generic-level swallow taxon from central Africa. *J. Avian Biol.* 49, e01698.
- del Hoyo, J., Elliott, A., Christie, D.A. (Eds.), 2004. Handbook of the Birds of the World. Vol. 9. Cotingas to Pipits and Wagtails. Lynx Edicions, Barcelona.
- del Hoyo, J., Turner, A., Kirwan, G.M., Collar, N., Boesman, P.F.D., 2023. Western House-Martin (*Delichon urbicum*), version 1.0.
- Dickinson, E.C., Christidis, L., 2014. The Howard and Moore complete checklist of the birds of the world, 4th ed., vol. 2. Eastbourne, UK Aves Press. Sch.
- Dor, R., Safran, R.J., Sheldon, F.H., Winkler, D.W., Lovette, I.J., 2010. Phylogeny of the genus *Hirundo* and the Barn Swallow subspecies complex. *Mol. Phylogenet. Evol.* 56, 409–418.
- Dor, R., Carling, M.D., Lovette, I.J., Sheldon, F.H., Winkler, D.W., 2012. Species trees for the tree swallows (Genus *Tachycineta*): an alternative phylogenetic hypothesis to the mitochondrial gene tree. *Mol. Phylogenet. Evol.* 65, 317–322.
- Eberhard, J.R., 2002. Cavity adoption and the evolution of coloniality in cavity-nesting birds. *Condor* 104, 240–247.
- Emlen, J.T., 1954. Territory, nest building, and pair formation in the cliff swallow. *Auk* 71, 16–35.
- Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32, 786–788.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61, 717–726.
- Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS One* 8, e65923.
- Fang, Y.-T., Tuanmu, M.-N., Hung, C.-M., 2018. Asynchronous evolution of interdependent nest characters across the avian phylogeny. *Nat. Commun.* 9, 1863.
- Feakins, S.J., Demeo, P.B., 2010. Global and African regional climate during the Cenozoic. *Cenozoic Mamm. Africa* 45–56.
- Fjeldså, J., 2013. The global diversification of songbirds (Oscines) and the build-up of the Sino-Himalayan diversity hotspot. *Avian Res.* 4, 132–143.
- Fry, C.H., Smith, D.A., 1985. A new swallow from the Red Sea. *Ibis* 127, 1–6.
- Fuchs, J., Irestedt, M., Fjeldså, J., Couloux, A., Pasquet, E., Bowie, R.C.K., 2012. Molecular phylogeny of African bush-shrikes and allies: tracing the biogeographic history of an explosive radiation of corvid birds. *Mol. Phylogenet. Evol.* 64, 93–105.
- García-Pérez, B., Hobson, K.A., Powell, R.L., Still, C.J., Huber, G.H., 2013. Switching hemispheres: a new migration strategy for the Disjunct Argentinean breeding population of Barn Swallow (*Hirundo rustica*). *PLoS One* 8, e55654.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644.
- Hall, Z.J., Street, S.E., Healy, S.D., 2013. The evolution of cerebellum structure correlates with nest complexity. *Biol. Lett.* 9, 20130687.
- Healy, S.D., Tello-Ramos, M.C., Hébert, M., 2023. Bird nest building: visions for the future. *Philos. Trans. R. Soc. B* 378, 20220157.
- Heard, S.B., Hauser, D.L., 1995. Key evolutionary innovations and their ecological mechanisms. *Hist. Biol.* 10, 151–173.
- Hijmans, R.J., 2018. raster: Geographic data analysis and modeling. R package.
- Hobson, K.A., Møller, A.P., Van Wilgenburg, S.L., 2012. A multi-isotope ( $\delta^{13}\text{C}$ ,  $\delta^{15}\text{N}$ ,  $\delta^2\text{H}$ ) approach to connecting European breeding and African wintering populations of barn swallow (*Hirundo rustica*). *Animal Migration* 1, 8–22.
- Hobson, K.A., Kardynal, K.J., Van Wilgenburg, S.L., Albrecht, G., Salvadori, A., Cadman, M.D., Liechti, F., Fox, J.W., 2015. A Continent-Wide Migratory Divide in North American Breeding Barn Swallows (*Hirundo rustica*). *PLoS One* 10, e0129340.
- Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Aratijo, M.B., Dimitrov, D., Fabre, P.-H., Graham, C.H., Graves, G.R., Jönsson, K.A., 2013. An update of Wallace's zoogeographic regions of the world. *Science* 339, 74–78.
- Hopkins, D.M., 1959. Cenozoic History of the Bering Land Bridge: the seaway between the Pacific and Arctic basins has often been a land route between Siberia and Alaska. *Science* 129, 1519–1528.

- Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kessler, E., 2013. Neogene songbirds (Aves, Passeriformes) from Hungary. *Hantkeniana* 8, 37–149.
- Kirchman, J.J., Whittingham, L.A., Sheldon, F.H., 2000. Relationships among cave swallow populations (*Petrochelidon fulva*) determined by comparisons of microsatellite and cytochrome b data. *Mol. Phylogenet. Evol.* 14, 107–121.
- 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, 4453–4455.
- Landis, M.J., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P., 2013. Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* 62, 789–804.
- Martínez, M.M., 1983. Nidificación de *Hirundo rustica erythrogaster* (Boddaert) en la Argentina (Aves, Hirundinidae). *Neotropica* 29, 83–86.
- Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Syst. Biol.* 63, 951–970.
- Matzke, N.J., 2018. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis with R scripts.
- Mayr, E., 1946. History of the North American bird fauna. *Wilson Bull.* 58, 3–41.
- McCormack, J.E., Tsai, W.L.E., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. *Mol. Ecol. Resour.* 16, 1189–1203.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30, i541–i548.
- Moyle, R.G., Slikas, B., Whittingham, L.A., Winkler, D.W., Sheldon, F.H., 2008. DNA sequence assessment of phylogenetic relationships among New World martins (Hirundinidae: *Progne*). *Wilson J. Ornithol.* 120, 683–691.
- Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M., Faircloth, B.C., 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nat. Commun.* 7, 12709.
- Nagy, J., Hauber, M.E., Hartley, I.R., Mainwaring, M.C., 2019. Correlated evolution of nest and egg characteristics in birds. *Anim. Behav.* 158, 211–225.
- O’Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., de Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.-P., Berggren, W.A., Budd, A.F., Cozzuol, M.A., Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the Isthmus of Panama. *Sci. Adv.* 2, e1600883.
- Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J., 2019. Earth history and the passerine superradiation. *Proc. Natl. Acad. Sci. USA* 116, 7916–7925.
- Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528.
- Pavlova, A., Zink, R.M., Drovetski, S.V., Rohwer, S., 2008. Pleistocene evolution of closely related sand martins *Riparia riparia* and *R. diluta*. *Mol. Phylogenet. Evol.* 48, 61–73.
- Pebesma, E., Bivand, R.S., 2005. S classes and methods for spatial data: the sp package. *R News* 5, 9–13.
- Perez, D.M., Gardner, J.L., Medina, I., 2020. Climate as an evolutionary driver of nest morphology in birds: a review. *Front. Ecol. Evol.* 8, 566018.
- Phillips, A.R., 1986. The known birds of North and Middle America: Distributions and Variation, Migration, Changes, Hybrids, Etc. Part 1, Hirundinidae to Mimidae; Certhiidae.
- Price, J.J., Griffith, S.C., 2017. Open cup nests evolved from roofed nests in the early passerines. *Proc. r. Soc. B Biol. Sci.* 284, 20162708.
- R core Team. 2023. R: A language and environment for statistical computing.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904.
- Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* 57, 4–14.
- Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223.
- Ridgway, R., 1904. The birds of North and Middle America, Part III. *Bull. U.S. Natl. Mus.* 50, 1–801.
- Rolland, C., Danchin, E., de Fraipont, M., 1998. The evolution of coloniality in birds in relation to food, habitat, predation, and life-history traits: a comparative analysis. *Am. Nat.* 151, 514–529.
- Ronquist, F., 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol.* 46, 195–203.
- Sangster, G., Cibois, A., Reddy, S., 2022. Pteruthiidae and Erpornithidae (Aves: Corvides): two new family-group names for babbler-like outgroups of the vireos (Vireonidae). *Bull. Br. Ornithol. Club* 142, 239–243.
- Scordato, E.S.C., Smith, C.C.R., Semenov, G.A., Liu, Y., Wilkins, M.R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., Turbek, S.P., Wunder, M.B., Stricker, C.A., Safran, R.J., 2020. Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecol. Lett.* 23, 231–241.
- Seo, T.-K., 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. Evol.* 25, 960–971.
- Sheldon, F.H., Whittingham, L.A., 1997. Phylogeny in studies of bird ecology, behavior, and morphology. *Avian Mol. Evol. Syst.* 279–299.
- Sheldon, F.H., Whittingham, L.A., Winkler, D.W., 1999. A comparison of cytochrome b and DNA hybridization data bearing on the phylogeny of swallows (Aves: Hirundinidae). *Mol. Phylogenet. Evol.* 11, 320–331.
- Sheldon, F.H., Winkler, D.W., 1999. Nest Architecture and Avian Systematics. *Auk* 116, 875–877.
- Sheldon, F.H., Whittingham, L.A., Moyle, R.G., Slikas, B., Winkler, D.W., 2005. Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 35, 254–270.
- Sheldon, F.H., Winkler, D.W., 1993. Intergeneric phylogenetic relationships of swallows estimated by DNA-DNA hybridization. *Auk* 110, 798–824.
- Silver, M., Griffin, C.R., 2009. Nesting habitat characteristics of bank swallows and belted kingfishers on the Connecticut River. *Northeast. Nat.* 16, 519–534.
- Smith, B.T., Klicka, J., 2010. The profound influence of the Late Pliocene Panamanian uplift on the exchange, diversification, and distribution of New World birds. *Ecography* 33, 333–342.
- South, A., 2011. rworldmap: A new R package for mapping global data. *R J.* 3, R package.
- South, A., 2017. naturalearth: world map data from natural earth. R package.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571.
- Swofford, D.L., 2003. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4.
- Thonglongya, K., 1968. A new martin of the genus *Pseudochelidon* from Thailand. *Applied Scientific Research Corporation of Thailand*.
- Tsai, W.L.E., Schedl, M.E., Maley, J.M., McCormack, J.E., 2020. More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Mol. Ecol. Resour.* 20, 1220–1227.
- Turbek, S.P., Schield, D.R., Scordato, E.S.C., Contina, A., Da, X.-W., Liu, Y., Liu, Y., Pagani-Núñez, E., Ren, Q.-M., Smith, C.C.R., Stricker, C.A., Wunder, M., Zonana, D.M., Safran, R.J., 2022. A migratory divide spanning two continents is associated with genomic and ecological divergence. *Evolution* 76, 722–736.
- Turner, A., 2018. Swallows and Martins (Hirundinidae). In: del Hoyo, J., Elliott, A., Christie, D., de Juana, E. (Eds.), *Handbook of the Birds of the World. Lynx Edicions, Barcelona*.
- Turner, A.K., Rose, C., 1989. Swallows & martins: an identification guide and handbook. Houghton Mifflin, Boston.
- van der Hoek, Y., Gaona, G.V., Ciach, M., Martin, K., 2020. Global relationships between tree-cavity excavators and forest bird richness. *R. Soc. Open Sci.* 7, 192177.
- Wang, L.-G., Lam, T.-T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., 2020. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* 37, 599–603.
- Weber, J.N., Peterson, B.K., Hoekstra, H.E., 2013. Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature* 493, 402–405.
- Whittingham, L.A., Slikas, B., Winkler, D.W., Sheldon, F.H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 22, 430–441.
- Winkler, D.W., Gandoy, F.A., Areta, J.I., Iliff, M.J., Rakhimberdiev, E., Kardynal, K.J., Hobson, K.A., 2017. Long-distance range expansion and rapid adjustment of migration in a newly established population of barn swallows breeding in Argentina. *Curr. Biol.* 27, 1080–1084.
- Winkler, D.W., Sheldon, F.H., 1993. Evolution of nest construction in swallows (Hirundinidae): a molecular phylogenetic perspective. *Proc. Natl. Acad. Sci. USA* 90, 5705–5707.
- Yu, Y., Harris, A.J., He, X., 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Mol. Phylogenet. Evol.* 56, 848–850.
- Zhang, C., Sayyari, E., Mirarab, S., 2017. ASTRAL-III: increased scalability and impacts of contracting low support branches. In: RECOMB International Workshop on Comparative Genomics. Springer International Publishing, Cham, pp. 53–75.
- Zink, R.M., Pavlova, A., Rohwer, S., Drovetski, S.V., 2006. Barn swallows before barns: population histories and intercontinental colonization. *Proc. R. Soc. B Biol. Sci.* 273, 1245–1251.
- Zyskowski, K., Prum, R.O., 1999. Phylogenetic analysis of the nest architecture of neotropical ovenbirds (Furnariidae). *Auk* 116, 891–911.