



Genomic Regulatory Networks Underlying the Evolution of Avian Beak Shape

Nicolas Alexandre^{1*}, Juan Zanin^{1*}, **Lucas Moreira^{1*}**, Olivia Plateau², Jeffrey N. Stout¹, Irene M. Kaplow^{3,4}, Savannah J. Hoyt⁵, Nicole M. Tillquist⁶, Rachel J. O'Neill^{5,6,7}, Olivier Fedrigo¹, Richard A. Schneider^{6†}, Anna L. Keyte^{1†}

¹Colossal Biosciences, 1401 Lavaca Street, Austin, TX, USA; ²Evolutionary Functional Morphology Research Group, University of Bern, Bern Switzerland; ³Biological Sciences Department, Carnegie Mellon University, Pittsburgh, PA, USA; ⁴Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA; ⁵Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA; ⁶Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA; ⁷Department of Genetics and Genome Sciences, UConn Health, Farmington, CT, USA;

*These authors contributed equally to this work

✉ lucas@colossa.com | nicolas@colossal.com

Introduction

Avian beaks display extraordinary phenotypic diversity, arguably the most conspicuous adaptive radiation in vertebrates. Yet the genomic architecture underlying craniofacial variation in birds remains largely unresolved.

- We generated paired chromatin accessibility (ATAC-seq) and transcriptomic profiles (RNA-seq) across the three primary beak primordia (frontonasal process, mandible, and maxilla) at six developmental stages (HH25–HH40) in four species representing divergent beak morphologies. Integrating these data with geometric morphometric shape measurements from 363 bird species enabled us to map cis-regulatory elements to quantitative axes of beak shape variation across the avian tree.
- To extend regulatory inference beyond our focal species, we applied a machine learning framework¹ that predicts chromatin accessibility from DNA sequence alone, linking sequence variation in putative enhancers to continuous beak shape phenotypes without requiring chromatin data from every species.

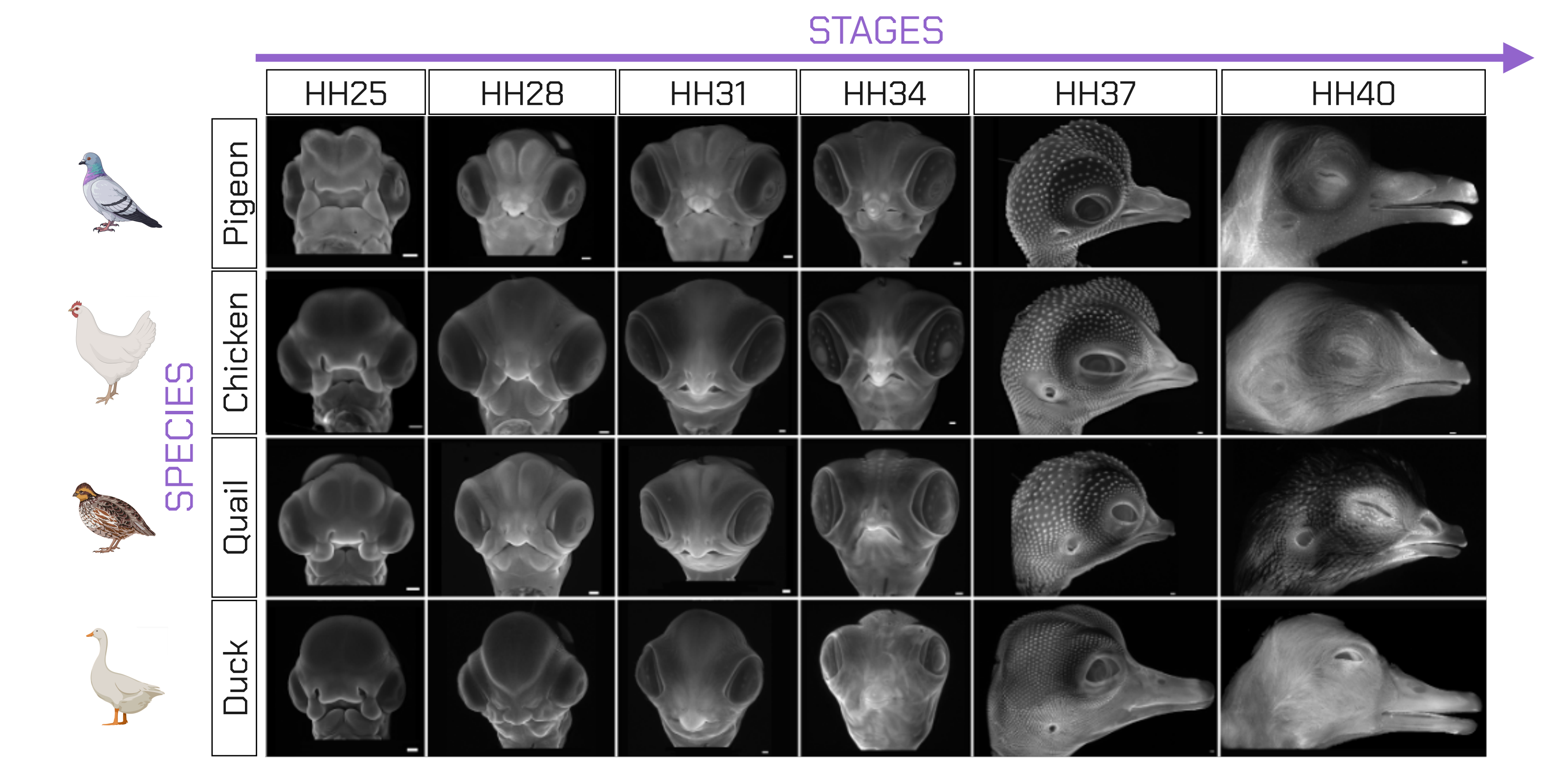


Fig 1. Craniofacial morphology across four avian species and six developmental stages. Epifluorescent images of the craniofacial outgrowth in pigeon (*Columba livia*), chicken (*Gallus gallus*), quail (*Coturnix japonica*), and duck (*Anas platyrhynchos*) at Hamburger–Hamilton stages HH25–HH40 using *sybr safe*. Scale bars indicated in each panel.

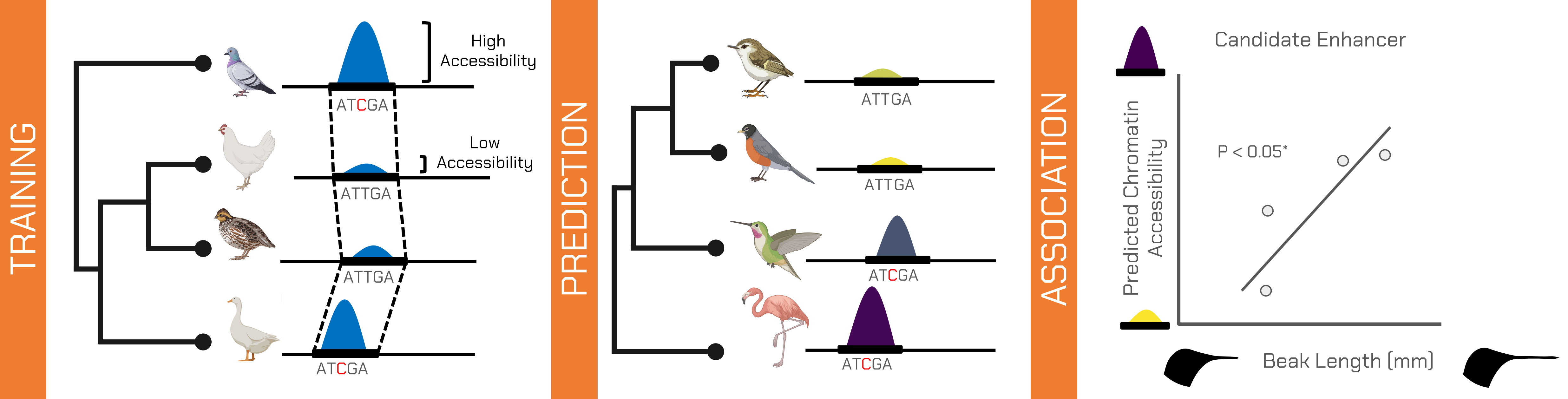


Fig 2. CNN workflow for linking regulatory sequence to beak morphology. (Training) A convolutional neural network learns to predict chromatin accessibility from DNA sequence using ATAC-seq data from four model species [pigeon, chicken, quail, duck]. (Prediction) The trained model projects chromatin accessibility across hundreds of non-model bird species using whole-genome alignments. (Association) Predicted accessibility at each candidate enhancer locus is tested for correlation with beak length across 363 species via phylogenetic regression, identifying regulatory elements whose activity correlates with beak shape evolution.

Linking regulatory activity to phenotype

We used a convolutional neural network method¹, to learn the tissue-specific regulatory code connecting genomic sequence to enhancer activity. We trained on open chromatin data from craniofacial dissections of four model species (chicken, quail, pigeon, and duck) and predicted across the full B10K² 363-way bird alignment.

- We selected distal open chromatin regions (OCRs) as candidate enhancers and extended their summits 250 bp on either flank.
- Reproducible peaks served as the positive set. Within species models were trained on 60% of chromosomes, tuned on 20%, and evaluated on the remaining 20%, achieving an average AUROC of 0.863. Cross-species models (train on one species, predict on others) achieved an average AUROC of 0.848.
- We ran phylogenetic regression between predicted chromatin accessibility and log10 beak length, obtained from VertNet³, across all 363 species in our Cactus⁴ alignment.
- Of 293,686 OCRs tested, 210 were significant at FDR < 0.2.

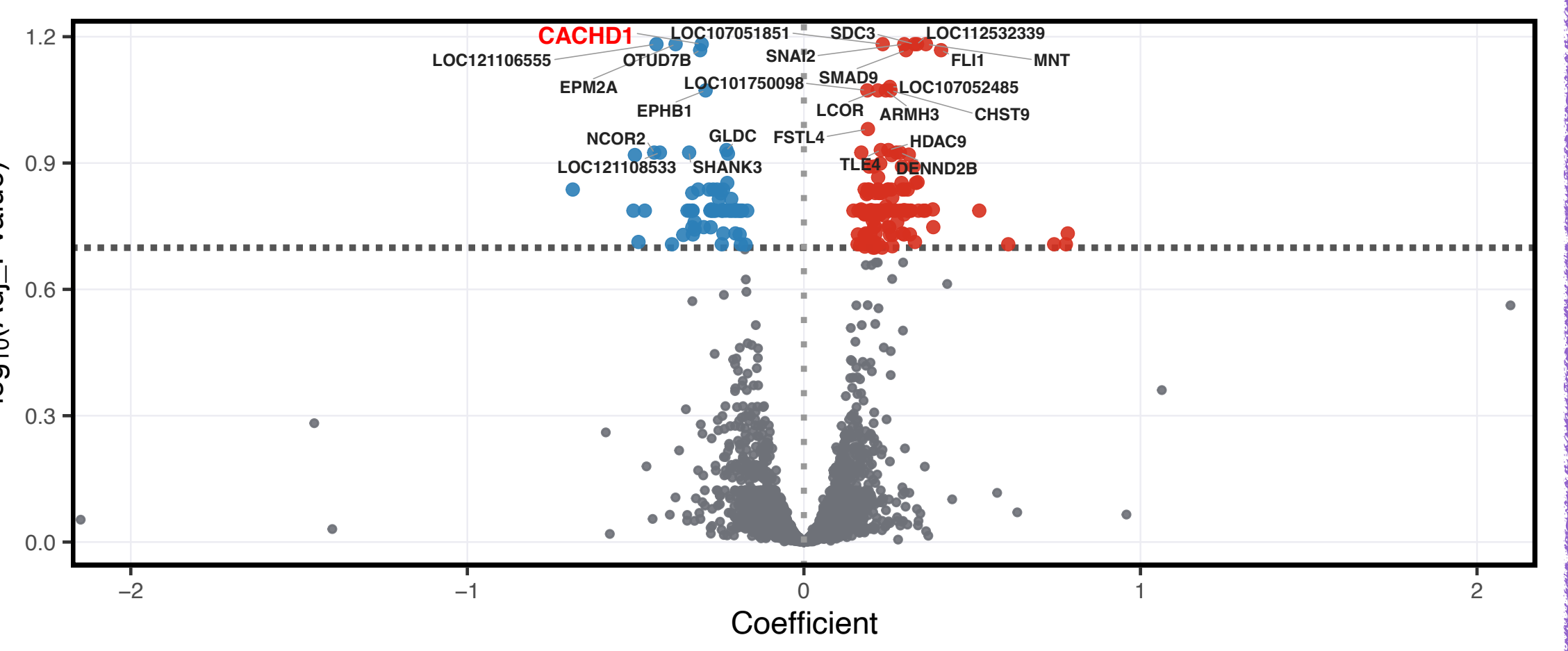


Fig 3. Phylogenetic regression identifies candidate enhancers associated with beak length. Volcano plot showing the association between predicted chromatin accessibility and beak length across 363 bird species. Significant loci are colored by direction of effect (red: positive association, blue: negative association), with the nearest gene labeled. CACHD1, a top candidate, is highlighted in red.

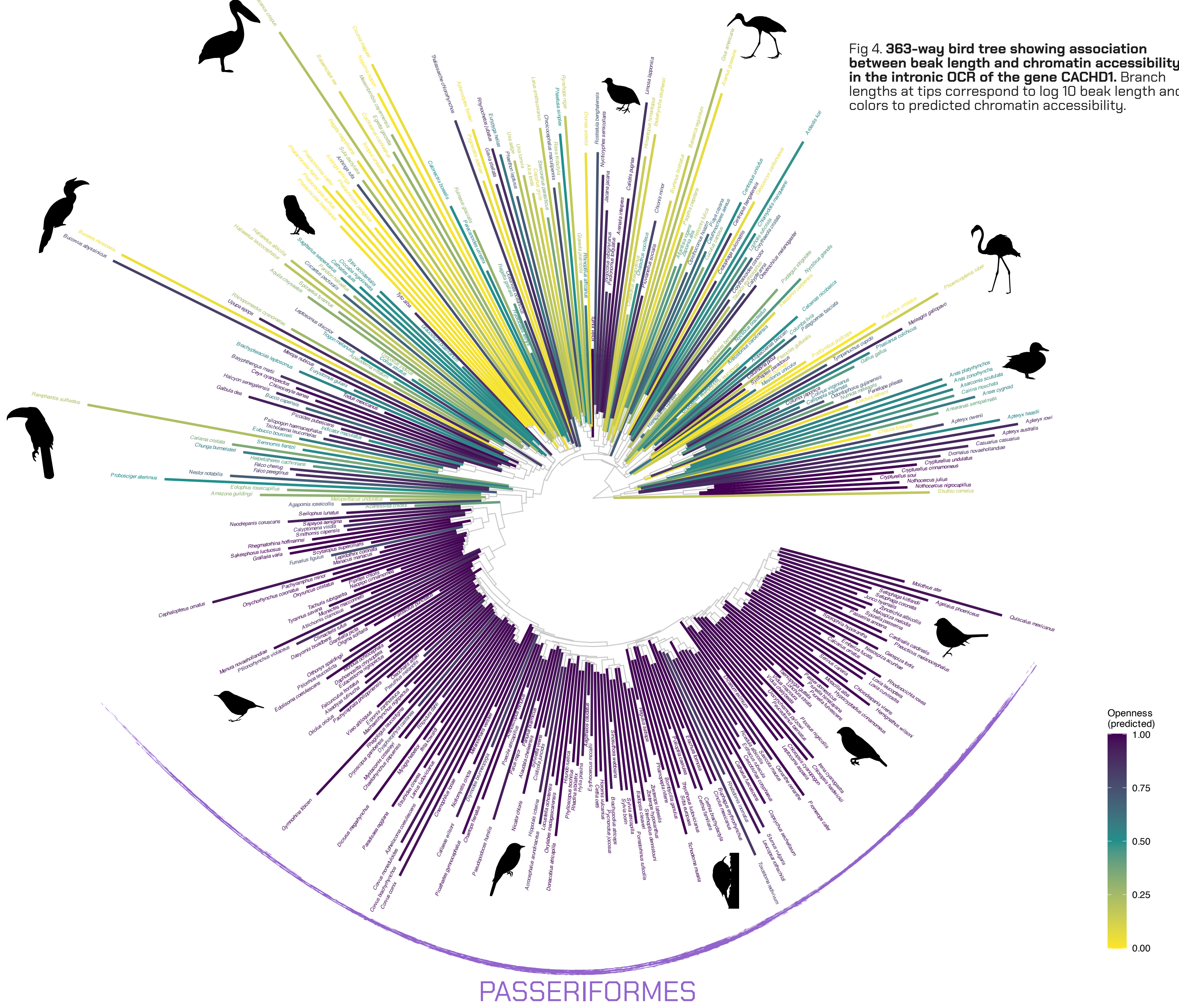


Fig 4. 363-way bird tree showing association between beak length and chromatin accessibility in the intronic OCR of the gene CACHD1. Branch lengths at tips correspond to log10 beak length and colors to predicted chromatin accessibility.

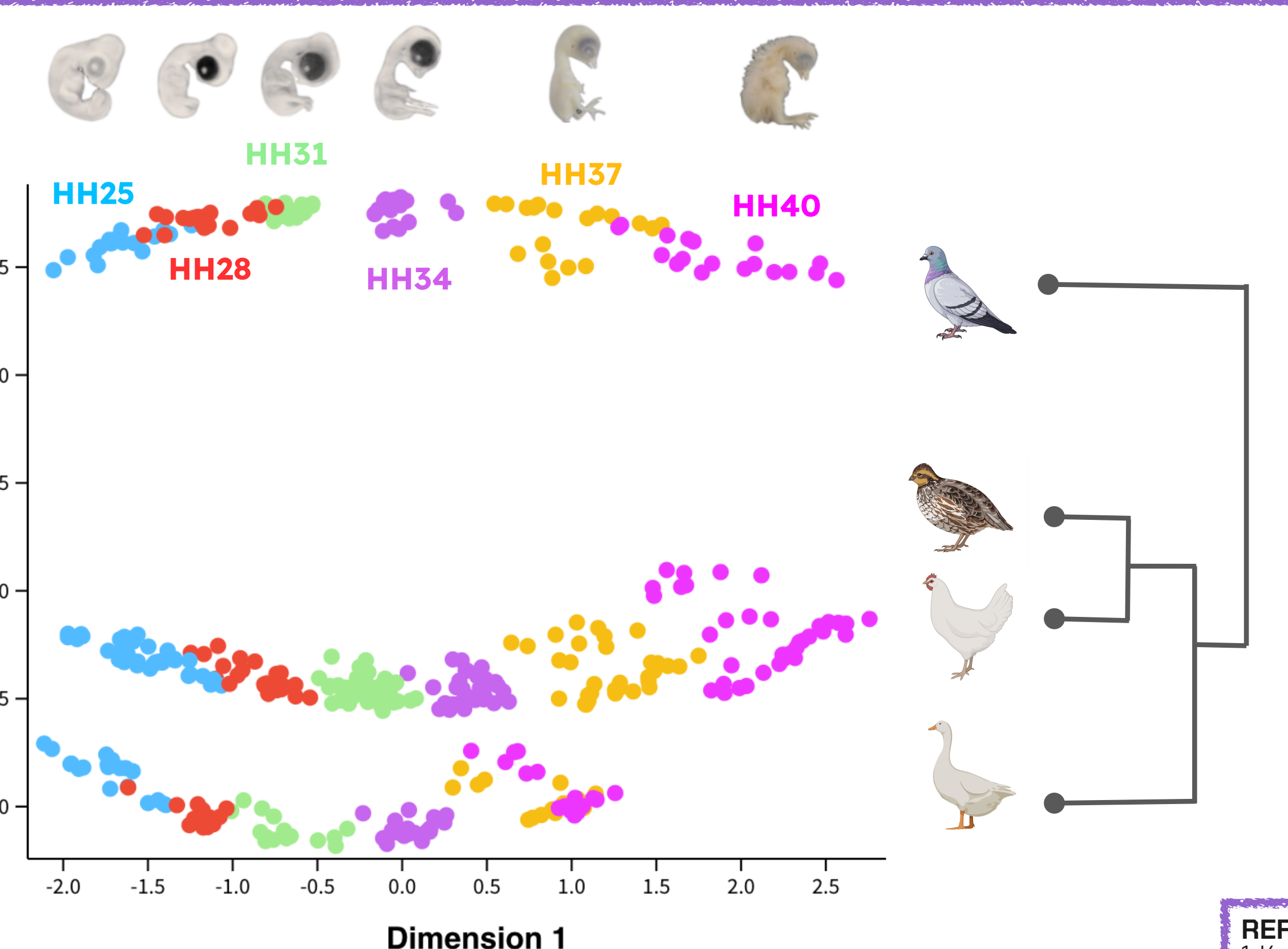


Fig 5. Transcriptional trajectories are driven by species and developmental stage. MDS plot of bulk RNA-seq samples from four avian species (pigeon, quail, chicken, duck) across six Hamburger–Hamilton stages (HH25–HH40). Each point represents one biological sample, colored by developmental stage. Samples separate primarily along Dimension 1 by stage, recapitulating the known temporal progression of craniofacial development (embryo images shown above), while species separate along Dimension 2, consistent with a shared developmental program overlaid with species-specific divergence in gene expression trajectories.

Detecting differences in gene expression across species

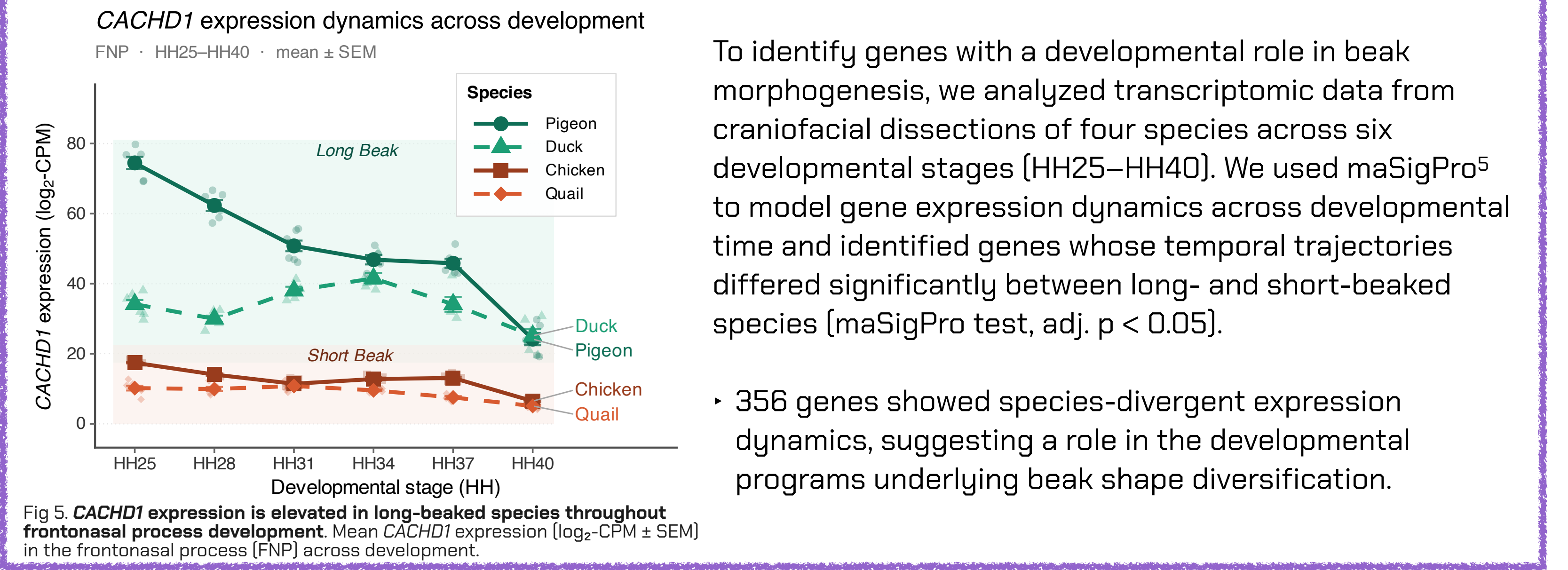


Fig 6. CACHD1 expression is elevated in long-beaked species throughout frontonasal process development. Mean CACHD1 expression (log₂-CPM ± SEM) in the frontonasal process (FNP) across development.

To identify genes with a developmental role in beak morphogenesis, we analyzed transcriptomic data from craniofacial dissections of four species across six developmental stages (HH25–HH40). We used maSigPro⁵ to model gene expression dynamics across developmental time and identified genes whose temporal trajectories differed significantly between long- and short-beaked species [maSigPro test, adj. p < 0.05].

- 356 genes showed species-divergent expression dynamics, suggesting a role in the developmental programs underlying beak shape diversification.

REFERENCES

1. Kaplow IM, Lawler AJ, Schäffer DE, et al. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science*. 2023;380(6643):eabm7993. doi:10.1126/science.abm7993
2. Feng S, Stiller J, Deng Y, et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature*. 2020;587(7833):252–257. doi:10.1038/s41586-020-2873-9
3. Constable H, Guralnick R, Wiczorek J, Spencer C, Peterson AT. VertNet: A New Model for Biodiversity Data Sharing. *PLoS Biology*. 2010;8(12):e1000309. doi:10.1371/journal.pbio.1000309
4. Armstrong J, Hickey G, Diekhans M, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*. 2020;587(7833):246–251. doi:10.1038/s41586-020-2871-y
5. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics*. 2014;30(18):2598–2602.