

RESEARCH ARTICLE SUMMARY

EVOLUTIONARY GENOMICS

A genetic signature of the evolution of loss of flight in the Galapagos cormorant

Alejandro Burga,* Weiguang Wang, Eyal Ben-David, Paul C. Wolf, Andrew M. Ramey, Claudio Verdugo, Karen Lyons, Patricia G. Parker, Leonid Kruglyak*

INTRODUCTION: Changes in the size and proportion of limbs and other structures have played a key role in the evolution of species. One common class of limb modification is recurrent wing reduction and loss of flight in birds. Indeed, Darwin used the occurrence of flightless birds as an argument in favor of his theory of natural selection. Loss of flight has evolved repeatedly and is found among 26 families of birds in 17 different orders. Despite the frequency of these modifications, we have a limited understanding of their underpinnings at the genetic and molecular levels.

RATIONALE: To better understand the evolution of changes in limb size, we studied a classic case of recent loss of flight in the Galapagos cormorant (*Phalacrocorax harrisi*). Cormorants are large water birds that live in coastal areas or near lakes, and *P. harrisi* is the only flightless cormorant among approximately 40 extant species. The entire population is distributed along the coastlines of Isabela and Fernandina islands in the Galapagos archipelago. *P. harrisi* has a pair of short wings, which are smaller

than those of any other cormorant. The extreme reduction of the wings and pectoral skeleton observed in *P. harrisi* is an attractive model for studying the evolution of loss of flight because it occurred very recently; phylogenetic evidence suggests that *P. harrisi* diverged from its flighted relatives within the past 2 million years. We developed a comparative and predictive genomics approach that uses the genome sequences of *P. harrisi* and its flighted relatives to find candidate genetic variants that likely contributed to the evolution of loss of flight.

RESULTS: We sequenced and de novo assembled the whole genomes of *P. harrisi* and three closely related flighted cormorant species. We identified thousands of coding variants exclusive to *P. harrisi* and classified them according to their probability of altering protein function based on conservation. Variants most likely to alter protein function were significantly enriched in genes mutated in human skeletal ciliopathies, including *Odf1*, *Evc*, *Wdr34*, and *Ift122*. We carried out experiments in *Caenorhabditis elegans* to confirm that a missense variant present in

the Galapagos cormorant IFT122 protein is sufficient to affect ciliary function. The primary cilium is essential for Hedgehog (Hh) signaling in vertebrates, and individuals affected by ciliopathies have small limbs and ribcages, mirroring the phenotype of *P. harrisi*. We also identified a 4-amino acid deletion in the regulatory domain of *Cux1*, a highly con-

ON OUR WEBSITE

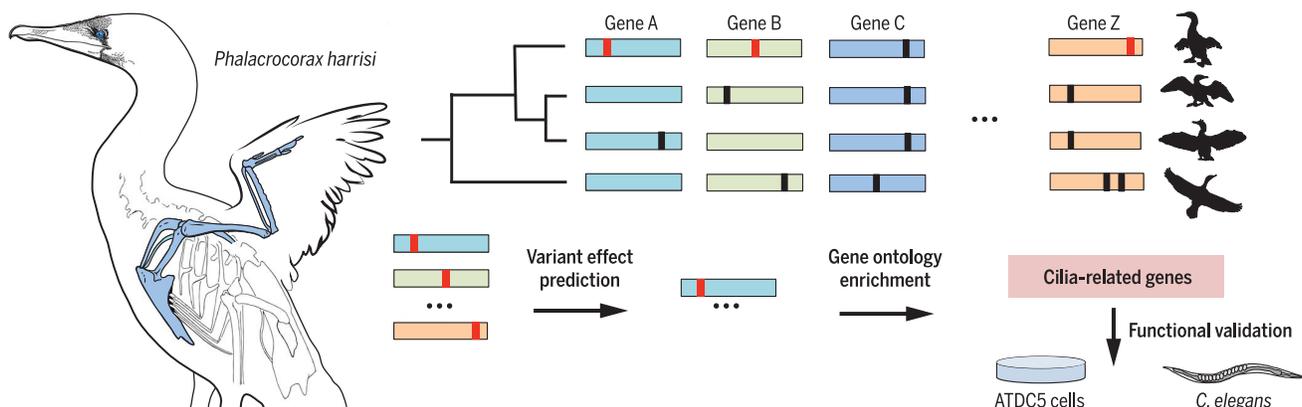
Read the full article at <http://dx.doi.org/10.1126/science.aal3345>

served transcription factor that has been experimentally shown to regulate limb growth in chicken. The four missing amino acids are perfectly conserved in all birds and mammals sequenced to date. We tested the consequences of this deletion in a chondrogenic cell line and showed that it impairs the ability of CUX1 to transcriptionally up-regulate cilia-related genes (some of which contain function-altering variants in *P. harrisi*) and to promote chondrogenic differentiation. Finally, we show that positive selection may have played a role in the fixation of the variants associated with loss of flight in *P. harrisi*.

CONCLUSION: Our results indicate that the combined effect of variants in genes necessary for the correct transcriptional regulation and function of the primary cilium likely contributed to the evolution of highly reduced wings and other skeletal adaptations associated with loss of flight in *P. harrisi*. Our approach may be generally useful for identification of variants underlying evolutionary novelty from genomes of closely related species. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: aburga@mednet.ucla.edu (A.B.); lkruglyak@mednet.ucla.edu (L.K.)
Cite this article as A. Burga et al., *Science* 356, eaal3345 (2017). DOI: 10.1126/science.aal3345

Comparative and predictive genomics of loss of flight. Comparison of the genomes of four closely related cormorant species allowed us to predict function-altering variants exclusively affecting the Galapagos cormorant and to test their functional consequences. Our results implicate ciliary dysfunction as a likely contributor to the evolution of loss of flight.



RESEARCH ARTICLE

EVOLUTIONARY GENOMICS

A genetic signature of the evolution of loss of flight in the Galapagos cormorant

Alejandro Burga,^{1*} Weiguang Wang,² Eyal Ben-David,¹ Paul C. Wolf,³ Andrew M. Ramey,⁴ Claudio Verdugo,⁵ Karen Lyons,² Patricia G. Parker,^{6,7} Leonid Kruglyak^{1*}

We have a limited understanding of the genetic and molecular basis of evolutionary changes in the size and proportion of limbs. We studied wing and pectoral skeleton reduction leading to flightlessness in the Galapagos cormorant (*Phalacrocorax harrisi*). We sequenced and de novo assembled the genomes of four cormorant species and applied a predictive and comparative genomics approach to find candidate variants that may have contributed to the evolution of flightlessness. These analyses and cross-species experiments in *Caenorhabditis elegans* and in chondrogenic cell lines implicated variants in genes necessary for transcriptional regulation and function of the primary cilium. Cilia are essential for Hedgehog signaling, and humans affected by skeletal ciliopathies suffer from premature bone growth arrest, mirroring skeletal features associated with loss of flight.

The evolution of loss of flight is one of the most recurrent limb modifications encountered in nature (1). Indeed, Darwin used the occurrence of flightless birds as an argument in favor of his theory of natural selection (2). He proposed that loss of flight could evolve as a result of selection in favor of larger bodies and relaxed selection due to the absence of predators. Loss of flight has evolved repeatedly and is found among 26 families of birds in 17 different orders (3). Moreover, recent studies strongly suggest that the ratites (ostriches, emus, rheas, cassowaries, and kiwis), long thought to derive from a single flightless ancestor, may constitute a polyphyletic group characterized by multiple independent instances of loss of flight and convergent evolution (3–5). However, despite the ubiquity and evolutionary importance of loss of flight (6), the underlying genetic and molecular mechanisms remain unknown.

The Galapagos cormorant (*Phalacrocorax harrisi*) is the only flightless cormorant among approximately 40 extant species (7). The entire population

is distributed along the coastlines of Isabela and Fernandina islands in the Galapagos archipelago. *P. harrisi* has a pair of short wings, which are smaller than those of any other cormorant (Fig. 1A)—a deviation from the allometric relationship between wing length and body mass (7). The radius and ulna are disproportionately small relative to the humerus, but no digits have been fused or lost, unlike in some ratites (8). In addition, the Galapagos cormorant differs from its flighted relatives in a delay in the onset of several developmental landmarks after hatching (9), shortened remiges (flight feathers), underdeveloped pectoral muscles, a long and narrow skull and pelvis, a disproportionately long tibiotarsus, a factor of 1.6 increase in body mass, and a highly reduced keel (7). The keel is an extension of the sternum that runs along its midline and provides an attachment surface for the flight muscles, the largest muscles in birds. Flightless taxa, such as ratites and Cretaceous *Hesperornis*, have evolved flat sternums in which the keel has been largely reduced or lost (10).

In contrast to ratites and penguins, which became flightless more than 50 million years ago (Ma) (5, 11), the Galapagos cormorant and its flighted relatives are estimated to share a common ancestor at ~2 Ma (12). This recent and extreme modification of wing size and pectoral skeleton makes *P. harrisi* an attractive model for studying loss of flight.

High-quality genome sequences of four cormorant species

To identify variants associated with loss of flight, we sequenced and de novo assembled the 1.2-Gb genomes of the Galapagos cormorant (Galapagos Islands, Ecuador) and three flighted cormorant spe-

cies: the double-crested cormorant (*Phalacrocorax auritus*; Minnesota, USA), the neotropical cormorant (*Phalacrocorax brasilianus*; Valdivia, Chile), and the pelagic cormorant (*Phalacrocorax pelagicus*; Alaska, USA). *P. auritus* and *P. brasilianus* are the closest relatives of *P. harrisi* (12–14), and *P. pelagicus* is part of a sister clade and served as an outgroup. Genomes were assembled from a combination of short insert and mate-pair Illumina libraries with SOAPdenovo2 (15) (table S1). Among these four genomes, the Galapagos cormorant's assembly had the longest contig and scaffold N50 metrics (contig N50, 103 kb; scaffold N50, 4.6 Mb; table S1B). We evaluated the completeness of the cormorants' genomes by estimating the total number of uniquely annotated proteins in each assembly and by using the CEGMA pipeline (16, 17). Overall, we found agreement between these two independent metrics in a data set including the four cormorant genomes and 17 recently published bird genomes ($r^2 = 0.75$, $P = 4.3 \times 10^{-7}$; Fig. 1B and table S2).

Commonly used metrics of assembly quality, such as contig and scaffold N50, were very poor predictors of the total number of proteins present in each assembly ($r^2 = 0.13$, $P = 0.15$, and $r^2 = 0.06$, $P = 0.81$; fig. S1 and table S2). Three of the four cormorant genomes (*P. harrisi*, *P. auritus*, and *P. pelagicus*) obtained the highest CEGMA scores and numbers of uniquely annotated genes among all bird genomes (red triangles, Fig. 1B). The following CEGMA scores, a means to estimate genome completeness, were obtained for the cormorants: *P. harrisi*, 90.3%; *P. auritus*, 91.3%; *P. brasilianus*, 72.6%; *P. pelagicus*, 87.1%. In contrast, Sanger and PacBio genomes had lower scores for other birds: *Gallus gallus* (Sanger assembly) (18), 80.7%; *Taeniopygia guttata* (Sanger assembly) (19), 71.4%; and *Melopsittacus undulatus* (PacBio assembly) (20), 79.0%. Thus, our cormorant genomes perform even better than genomes assembled from Sanger sequences and PacBio long reads (complete statistics in table S2).

Phylogeny and genetic diversity

We reconstructed the cormorant phylogeny using a Bayesian framework (17) and confirmed the phylogenetic relationship among the four sequenced species (Fig. 1C). Moreover, our results indicate that *P. harrisi* last shared a common ancestor with *P. auritus* and *P. brasilianus* at ~2.37 Ma, in agreement with an estimate from mitochondrial DNA (12) (Fig. 1C). Española, the oldest extant island in the Galapagos archipelago, emerged no earlier than 4 Ma, and proto-Galapagos islands existed at 9 Ma or earlier (21). Our results are consistent with the view that *P. harrisi* lost the ability to fly while inhabiting the archipelago.

We calculated the proportion of single-nucleotide polymorphism (SNP) heterozygous sites for each sequenced individual to estimate the levels of intraspecific genetic diversity (Fig. 1D). *P. harrisi* showed the lowest proportion of heterozygous SNPs among the sequenced cormorants (0.00685%; Fig. 1D). The heterozygosity of *P. harrisi* is even lower than that of the crested ibis, *Nipponia nippon*, a highly endangered bird with a small effective

¹Department of Human Genetics, Department of Biological Chemistry, and Howard Hughes Medical Institute, University of California, Los Angeles, CA, USA. ²Departments of Molecular, Cell and Developmental Biology and Orthopaedic Surgery, University of California and Orthopaedic Institute for Children, Los Angeles, CA, USA. ³Wildlife Services, U.S. Department of Agriculture, Roseburg, OR, USA. ⁴U.S. Geological Survey Alaska Science Center, Anchorage, AK, USA. ⁵Instituto de Patología Animal, Facultad de Ciencias Veterinarias, Universidad Austral de Chile, Valdivia, Chile. ⁶Department of Biology and Whitney Harris World Ecology Center, University of Missouri, St. Louis, MO, USA. ⁷WildCare Institute, Saint Louis Zoo, St. Louis, MO, USA. *Corresponding author. Email: aburga@mednet.ucla.edu (A.B.); lkruglyak@mednet.ucla.edu (L.K.)

population size and a known recent population bottleneck (22) (0.0172%; Fig. 1D). The low level of heterozygosity found in the Galapagos cormorant is most likely due to its small population size (~1500 individuals) and multiple population bottlenecks (23).

Discovery and characterization of function-altering variants in *P. harrisi*

To investigate the genetics of flightlessness evolution, we developed a comparative and predictive genomics approach (24, 25) that uses the genome sequences of *P. harrisi* and its flighted relatives to identify genetic variants that likely contributed to the evolution of loss of flight. Both coding and cis-regulatory variants have been implicated in the evolution of morphological traits (26, 27). However, determining the impact of regulatory variants is not straightforward. To identify the contribution of regulatory variants to the evolution of loss of flight in *P. harrisi*, we searched for ultraconserved noncoding sequences showing accelerated molecular evolution (28–30). We identified 11 ultraconserved noncoding regions in tetrapods that show accelerated evolution in *P. harrisi* but not in the other cormorants [false discovery rate (FDR) < 5%]. One of these regions was located in an intron of the gene *FTO* (fig. S2), which has been associated with obesity in humans (31); however, none of these regions overlapped with experimentally validated or putative mouse limb enhancers (17, 32, 33) (table S7).

We thus focused on characterizing coding variants because we are better able to predict their molecular consequences. For our variant discovery approach to be comprehensive, it was imperative to interrogate most of the Galapagos cormorant's genes. To increase our power to do so, we annotated genes using homology-based and transcriptome-based gene annotations. The latter were derived with mRNA expression data from the developing wing of a double-crested cormorant embryo (fig. S3C) (17). We then predicted all missense, deletion, and insertion variants in ortholog pairs between *P. harrisi* and each of its three flighted relatives (fig. S4) (17).

We used PROVEAN (34), a phylogeny-corrected variant effect predictor, to evaluate the impact on protein function of each of the Galapagos cormorant's variants on a genome-wide scale. PROVEAN predictions have been validated in experimental evolution studies that mimic the process of gradual accumulation of mutations in nature (35). A PROVEAN score is calculated for each variant; the more negative the score, the more likely a given variant is to alter protein function. We examined the distribution of PROVEAN scores obtained when comparing 12,442 ortholog pairs between *P. harrisi* and *P. auritus* (Fig. 2A). Of these, 4959 pairs (40%) did not contain coding variants; the remaining 7483 pairs contained a total of 23,402 coding variants: 22,643 single amino acid substitutions, 456 deletions, and 303 insertions (Fig. 2B). Most variants were predicted to be neutral (the distribution is centered around zero). As expected, deletions and insertions were enriched in the tails of the distribution (Fig. 2B).

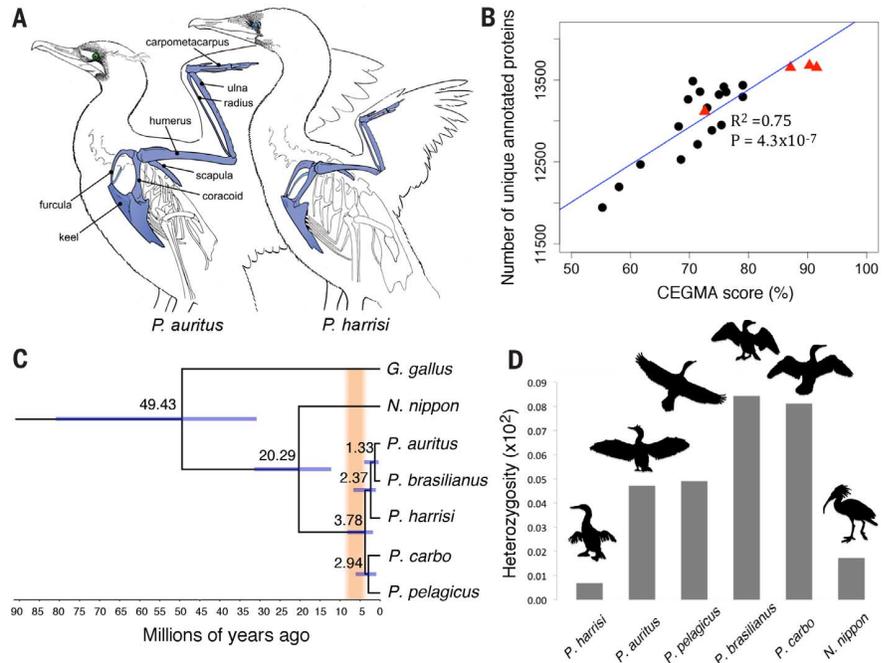


Fig. 1. The Galapagos cormorant, a model for studying flightlessness evolution. (A) The average wing length of an adult Galapagos cormorant male is 19 cm (3.6 kg body mass), whereas the wing length of its closest relative, the double-crested cormorant, is 31.5 cm (2.2 kg body mass). [Illustration by Katie Bertsche from specimens 134079 and 151575, Museum of Vertebrate Zoology at Berkeley] (B) The CEGMA score is a good predictor of genome completeness from a gene-centric perspective. The blue line is the linear regression model ($r^2 = 0.75$, $P = 4.3 \times 10^{-7}$). Genomes reported in this study are red triangles; other published avian genomes are black circles (table S2). (C) Bayesian phylogram reconstructed with fourfold degenerate sites from whole genome sequences. The orange bar illustrates the time span between the approximate origin of the proto-Galapagos archipelago (9 Ma) and the origin of the oldest extant island, San Cristobal (4 Ma). Nodes represent median divergence ages. Blue bars indicate the 95% highest posterior density interval. (D) Heterozygosity levels inferred from whole genome sequences. Birds are not drawn to scale.

Table 1. Function-altering variants in *P. harrisi* are enriched for genes that cause skeletal ciliopathies in humans. Sanger-validated examples of function-altering variants (PROVEAN score < -5) in *P. harrisi*. Cilia/Hh-related genes were found on the basis of functional enrichment for human syndromes. PCP (planar cell polarity) genes were selected according to literature evidence linking cilia and PCP. These variants are fixed in the population.

Gene	Pathway	Variant	PROVEAN score	Human syndrome
<i>Ofd1</i>	Cilia/Hh	R325C	-6.913	Orofaciodigital and Joubert
		K517T	-5.673	
		E889G	-5.068	
<i>Talpid3</i>	Cilia/Hh	D759V	-7.805	Joubert and Jeune
<i>Evc</i>	Cilia/Hh	T341I	-5.546	Ellis-van Creveld
<i>Dync2h1</i>	Cilia/Hh	P2733S	-7.431	Short-rib thoracic dysplasia
<i>Ift122</i>	Cilia/Hh	Q691L	-5.491	Cranioectodermal dysplasia
<i>Wdr34</i>	Cilia/Hh	P188R	-6.337	Short-rib thoracic dysplasia
<i>Kif7</i>	Cilia/Hh	R833W	-6.827	Joubert and acrocallosal
<i>Gli2</i>	Hh	P1086T	-5.117	Culler-Jones
		Fat1	PCP	
		Y2462C		-8.592
<i>Dchs1</i>	PCP	G2063D	-6.45	Van Maldergem
<i>Dvl1</i>	PCP	P103L	-8.23	Robinow

*Based on phenotypic similarity to mutant mouse model.

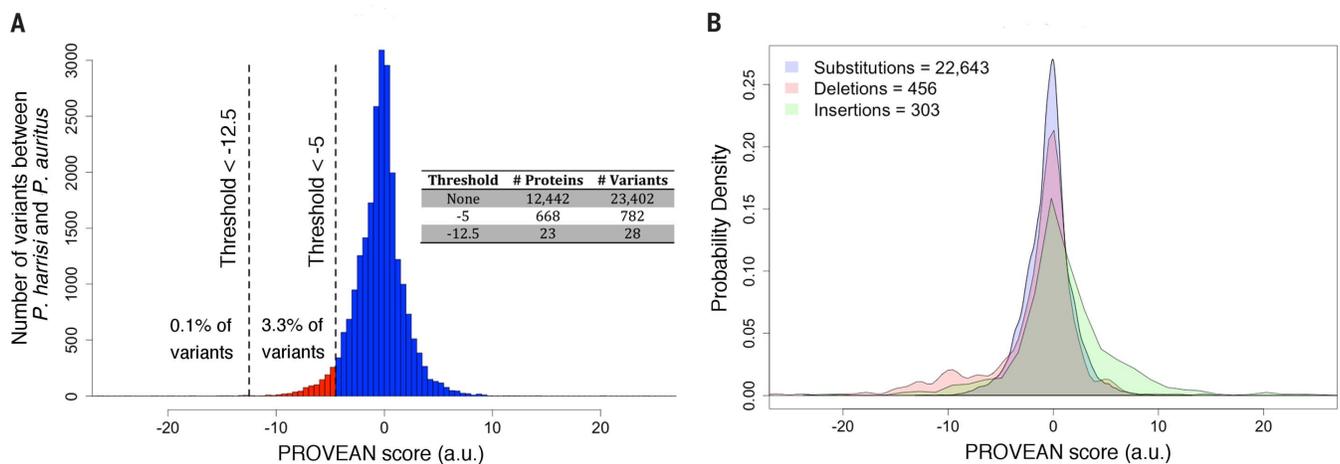


Fig. 2. Distribution of the effect of variants between *P. auritus* and *P. harrisi*. (A) We used PROVEAN to predict the effect on protein function of 23,402 variants contained in 12,442 orthologous pairs between *P. auritus* and *P. harrisi*; 4966 pairs contained no variants. The more negative the score, the more likely the variant affects protein function. PROVEAN score thresholds used in this study are drawn as vertical dashed lines. Numbers of proteins and variants found for each threshold are shown in the inset. (B) Density of PROVEAN scores for each class of variant. The same variants shown in (A) were classified as single amino acid substitutions, deletions, and insertions. Numbers of variants in each class are indicated. a.u., arbitrary units.

Very similar numbers of variants and PROVEAN score distributions were obtained for the other homology-based (fig. S5) and transcriptome-based annotations (fig. S3).

Enrichment for genes mutated in skeletal ciliopathies

To identify proteins carrying function-altering variants in the Galapagos cormorant, we applied a stringent threshold to our four prediction data sets: PROVEAN score less than -5 , twice the threshold for discovery of human disease variants (17, 34) (Fig. 2A). In our data set, variants with a PROVEAN score less than -5 typically occur at residues that have been perfectly conserved at least since mammals and birds last shared a common ancestor (~ 300 Ma; fig. S6). Consequently, changes in these residues are likely to alter protein function or stability.

On the basis of theoretical and experimental considerations (36), we hypothesized that flightlessness is likely to have a polygenic basis and that the underlying variants would be enriched in certain biological pathways. Consistent with this hypothesis, gene enrichment analysis of function-altering variants in the Galapagos cormorant revealed that genes implicated in human developmental disorders were significantly overrepresented (17) (table S3A). Strikingly, 8 of the 19 significantly enriched categories (Human Phenotype Ontology) consisted of genes that affect limb development when mutated, leading to disorders such as polydactyly, syndactyly, and duplication of limb bones. Control analyses showed no enrichment of these categories in the flighted cormorants (17) (table S3, B and C).

Many of the genes underlying the enrichment for limb syndromes are those mutated in a family of human disorders known as ciliopathies. For instance, 17 of 25 genes (65%) in the “duplication of hand bones” category and all 12 genes (100%) in the “preaxial hand polydactyly” category are

mutated in human ciliopathies (table S4). Moreover, ciliopathy-associated genes were present in all of the enriched categories (table S4). Ciliopathies comprise a phenotypically diverse group of rare genetic disorders that result from defects in the formation or function of cilia (37).

Cilia are hairlike microtubule-based structures that are nucleated by the basal body (centriole and associated proteins) and project from the surface of cells. Primary cilia are essential for mediating Hedgehog (Hh) signaling in vertebrates, serving as antennae for morphogens during development (38). We confirmed by Sanger sequencing the presence of predicted function-altering variants in *Ofd1*, *Evc*, *Talpid3*, *Dync2h1*, *Ift122*, *Wdr34*, and *Kif7*, all of which are necessary for the assembly or functioning of the primary cilium and are mutated in human ciliopathies, particularly those affecting the skeleton (Table 1). We also found a likely function-altering variant in *Gli2*, a transcription factor necessary for Hh signaling (39) (Table 1). Humans affected by diverse skeletal ciliopathies have small limbs and ribcages (37), suggesting a parallel with the main features of the Galapagos cormorant: small wings and a flattened sternum. However, the consequences of ciliopathies in humans are often more severe and pleiotropic, likely as a consequence of the overrepresentation of loss-of-function alleles in patients (40). Interestingly, although ciliopathies do not exclusively affect the forelimb in humans, differences between forelimbs and hindlimbs are commonly found among patients. For example, digital abnormalities affect the hands more often than the feet among oral-facial-digital syndrome type I (40, 41) and Ellis-van Creveld (42) patients, which suggests that forelimbs and hindlimbs differ in their intrinsic sensitivity to ciliary dysfunction.

The Galapagos cormorant ortholog of human *OFDI* (mutated in oral-facial-digital syndrome 1) contains three predicted function-altering var-

iants with a PROVEAN score less than -5 : Arg³²⁵ \rightarrow Cys (R325C), -6.913 ; Lys⁵¹⁷ \rightarrow Thr (K517T), -5.673 ; and Glu⁸⁸⁹ \rightarrow Gly (E889G), -5.068 (fig. S6, A and B). *Ofd1* knockout mice display polydactyly and shortened long bones (43). Also, a function-altering missense variant [Gln⁶⁹¹ \rightarrow Leu (Q691L), score -5.491] was found in IFT122, a component of the IFT complex that controls ciliogenesis and the ciliary localization of Shh pathway regulators (44). Null *Ift122* mutants show severe limb and skeletal phenotypes in mice (45), and mutations in *Ift122* have been associated with Sensenbrenner syndrome in humans, which is characterized by craniofacial, ectodermal, and skeletal abnormalities, including limb shortening (46). Strikingly, the mutated Gln in IFT122 is virtually invariant among eukaryotes as different as green algae, *C. elegans*, *Drosophila*, and vertebrates (fig. S6C). To directly test whether the nonsynonymous substitution in IFT122 affects protein function, we generated a *C. elegans* knock-in strain using CRISPR/Cas9 homology-directed genome editing (Fig. 3A). The edited strain carries the Galapagos cormorant missense variant at the corresponding orthologous position in *daf-10* [Gln⁸⁶² \rightarrow Leu (Q862L)], the ortholog of IFT122 in *C. elegans* (47).

In invertebrates, cilia do not mediate Shh signaling but are necessary for detecting external sensory inputs (48). The only ciliated cells in *C. elegans* are sensory neurons, and mutations in cilia components affect dispersal behavior, chemotaxis, and dauer formation (49). We tested the bordering behavior of worms (accumulation of animals on the thickest part of a bacterial lawn), which is known to be mediated by ciliated neurons (50). We found that *daf-10(e1387)* mutants carrying a premature stop (Q892X) displayed an increase in bordering behavior in a dispersal assay relative to the wild type [73% for *daf-10(e1387)* versus 30% for N2; $P = 0.019$, t test; Fig. 3, B, C, and E]. Two independently generated *daf-10* Q862L knock-in lines

phenocopied the effect of the premature stop allele (70% for line 1 and 69% for line 2 versus 30% for N2; $P = 0.018$ and 0.016 , respectively; Fig. 3, B, D, and E; see movie S1). *Daf-10(e1387)* ciliary neurons fail to incorporate fluorescent dyes, like many other loss-of-function mutants in cilia components (49). In contrast, the *daf-10* Q862L knock-in worms incorporated the DiO dye in the same manner as wild-type worms, which suggests that this allele is hypomorphic (fig. S7). Overall, these results indicate that the IFT122 Q691L missense variant present in the Galapagos cormorant can affect protein function in vivo in *C. elegans*.

The planar cell polarity (PCP) pathway exhibits a genetic link to cilia (38, 51–53). We found function-altering variants in members of the PCP pathway in *P. harrisi*: Fat1 atypical cadherin (*Fat1*), Dachous cadherin-related 1 (*Dchs1*), and Disheveled-1 (*Dvl1*) (Table 1). The Galapagos cormorant FAT1 contains two function-altering variants, Ser¹⁷¹⁷ → Leu and Tyr²⁴⁶² → Cys (S1717L and Y2462C; Table 1). The mutated Ser and Tyr are conserved from zebrafish to humans (fig. S5D). *Fat1* knockout mice show very selective defects in muscles of the upper body but not in posterior muscles (54). In addition, *Dvl1* is mutated in humans with Robinow syndrome, characterized by limb shortening (55, 56).

Sanger sequencing of 20 Galapagos cormorant individuals from two different populations (Cabo Hammond and Cañones Sur) (57) revealed only homozygous carriers for all of the variants in Table 1, indicating that these variants are most likely fixed in the Galapagos cormorant. In summary, we found an overrepresentation of predicted function-altering variants in genes that, when mutated in humans and mice, cause skeletal ciliopathies and bone growth defects.

CUX1 is mutated in *P. harrisi*

To identify the most likely function-altering variants in *P. harrisi*, we applied a more stringent PROVEAN score threshold: -12.5 delta alignment score, five times the threshold used for discovery of human disease variants (34). This strategy narrowed our search to 23 proteins (0.16% of annotated proteins in *P. harrisi*) (table S5). We manually curated these 23 proteins and performed additional Sanger sequencing, reducing the list of proteins with confirmed or putative variants to 12 (table S5) (17). These variants were exclusively small deletions. Among these 12 proteins, two stood out from their known role in development: LGALS-3 and CUX1. LGALS-3 is affected by a 7–amino acid deletion in *P. harrisi* (PROVEAN score, -26.319). LGALS-3 (Galectin-3) is localized at the base of the primary cilium and is necessary for correct ciliogenesis in mice (58), but it has not been implicated in human ciliopathies. Moreover, LGALS-3 physically interacts with SUFU, an important regulator of mammalian Hh signaling (59), and knockout mice show pleiotropic defects in chondrocyte differentiation (60).

In addition, we found a 4–amino acid deletion (PROVEAN score, -15.704) in CUX1. CUX1 (cut-like homeobox 1), also known as CDP, is a highly conserved transcription factor with diverse roles

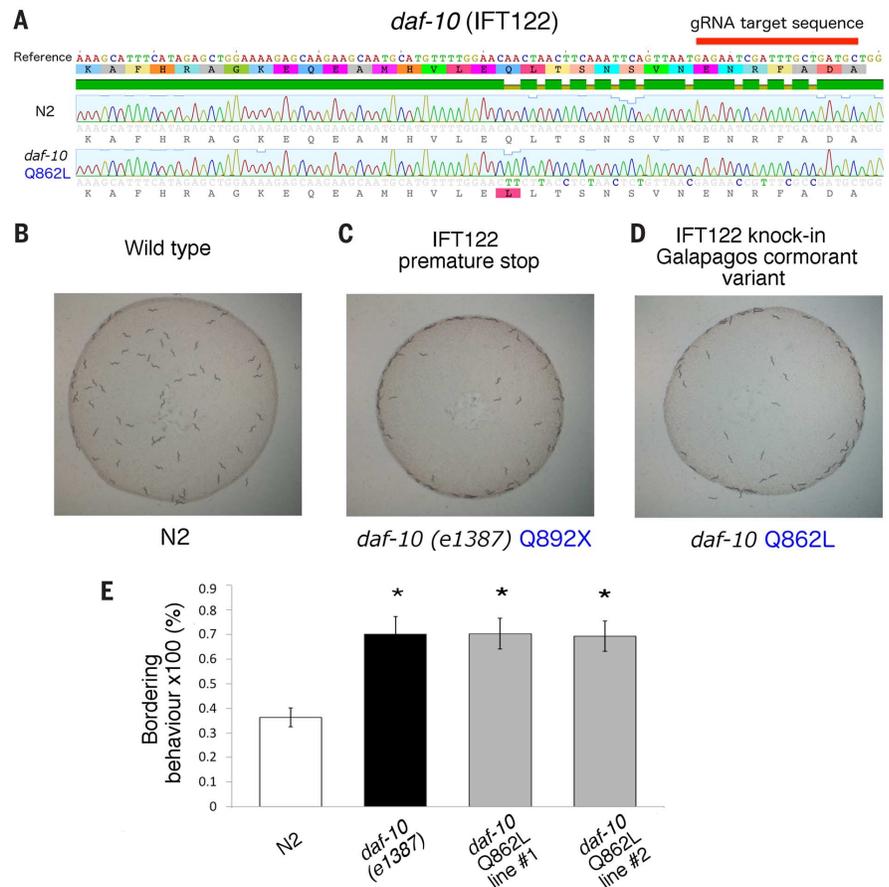


Fig. 3. The Galapagos cormorant variant IFT122 Q691L affects ciliary function in vivo. (A) The *daf-10* gene (IFT122 ortholog) was targeted with CRISPR/Cas9 homology-mediated repair in *C. elegans* to introduce a nonsynonymous substitution present exclusively in the Galapagos cormorant (IFT122 Q691L). The resulting edited knock-in strain contains the *daf-10* Q862L substitution and 10 synonymous substitutions (17). Edited strains were sequenced with Sanger sequencing to confirm genotypes. gRNA, guide RNA. (B to D) Representative bordering behavior of N2 wild-type worms (B), *daf-10(e1387)* containing a premature stop codon Q892X (C), and *daf-10* Q862L knock-in strain (D). (E) Quantification of bordering behavior in N2, *daf-10(e1387)*, and two independently generated knock-in *daf-10* Q862L strains ($n = 3$, t test). * $P < 0.05$; error bars denote SE.

in development. CUX1 contains four DNA binding domains: three CUT domains (CR1 to CR3) and one homeodomain (HD) (Fig. 4A) (61). The full-length isoform, which contains four DNA binding domains (CR1-3HD), acts exclusively as a transcriptional repressor and has rapid and unstable DNA binding dynamics. In contrast, smaller isoforms such as CR2-3HD and CR3HD can act as both repressors and activators of gene expression, and show slow and stable DNA binding dynamics in vitro (61). Although insect and bird wings evolved independently, it is noteworthy that *cut*, the *Drosophila* ortholog of Cux1, is necessary for the proper development of wings and flight muscles in flies (62). In chicken, *Cux1* mRNA expression in the limb at embryonic stage 23 is restricted to the ectoderm bordering the apical ectodermal ridge (AER) (63). The AER is one of the key signaling centers that drive limb development. At later stages, *Cux1* is expressed in the developing joints of both chicken (64) and mouse (fig. S8) and is also de-

tected in chondrocytes in developing bones of mice (fig. S9). A function-altering variant in CUX1 is a strong candidate to contribute to loss of flight in *P. harrisi*, because adenovirus-mediated overexpression in the developing chicken wing of a form of CUX1 missing the Cut2 DNA binding domain results in severe wing truncation (63, 65). These truncations most strongly affect distal skeletal elements (digits, radius, and ulna). As already noted, in *P. harrisi* the radius and ulna are disproportionately small relative to the humerus (7).

We Sanger-sequenced and confirmed the predicted *Cux1* 12–base pair (bp) deletion in *P. harrisi*. We also confirmed that this variant is fixed in the population and absent in the other cormorant species (fig. S10A). The 12–bp deletion in Cux1 removes four amino acids, Ala-Gly-Ser-Gln (AGSQ), immediately adjacent to the C-terminal end of the homeodomain (Fig. 4B). We refer to this variant as CUX1- Δ 4aa. Alignment of CUX1 orthologs from available vertebrate genomes revealed that

the four missing residues are extremely conserved among tetrapods (Fig. 4B). The deleted Ser is phosphorylated in human cells (66), but the consequences of this modification are unknown.

The *Cux1* deletion does not include any of the predicted residues responsible for DNA contact and recognition (67), but given its close proximity to the homeodomain, we decided to test whether the DNA binding activity of CUX1 was affected. We chose to express the CR3HD isoform because Western blot analysis revealed that this was the most abundant CUX1 isoform expressed in the developing wing of mallard embryos (~50 kDa; Fig. 4A). We performed electrophoretic mobility shift assay (EMSA) with purified CR3HD CUX1-Ancestral and CUX1-Δ4aa protein variants (fig. S10B) as described (68) and found that DNA binding was not abolished in the deletion variant (fig. S10C). CUX1 is able to both directly repress and activate gene expression through its C-terminal tail (69, 70). We performed a luciferase reporter assay (69, 71) and found that both variants were equally capable of repressing the expression of a UAS/tk luciferase reporter (Fig. 4D). Thus, the Galapagos cormorant CUX1-Δ4aa variant appears to not affect DNA binding in vitro or the C-terminal repression activity in COS-7 cells.

CUX1 regulates the expression of cilia and PCP genes

We hypothesized that the *Cux1* deletion variant is mechanistically related to the enrichment of function-altering variants in ciliopathy-related genes. This inference came from the fact that transgenic mice overexpressing the CR3HD-CUX1 isoform develop polycystic kidneys. Cilia in cystic epithelial cells from these animals were twice as long as the ones in control epithelial cells (72). Furthermore, the CR2CR3HD-CUX1 isoform has been shown to directly up-regulate the expression of *RPGRIPL1*, also known as *FTM*, a component of the cilia basal body that is involved in Shh signaling and mutated in human ciliopathies (73). Also, *Cux1* knockout mice show deregulation of SHH expression in hair follicles (74).

To test whether *Cux1* could globally regulate expression of cilia genes, we analyzed expression array data from human-derived Hs578t cells stably expressing a short hairpin RNA against *Cux1*, as well as cells overexpressing the human CR2CR3HD-CUX1 isoform (75). In concordance with the role of *Cux1* as a regulator of cell growth and proliferation (76), genes significantly up- or down-regulated in both conditions ($P < 0.05$ and $>10\%$ change) were enriched for pathways such as “cell cycle” and “mitotic G₁-G₁/S phases” ($P = 3.99 \times 10^{-5}$ and 0.016, respectively; table S6). We also found enrichment for cilia-related categories such as “assembly of the primary cilium” and “intraflagellar transport” ($P = 0.00012$ and 0.0057, respectively; table S6). These results suggest that cilia-related genes are enriched among *Cux1* targets.

To further test whether *Cux1* can regulate ciliary genes in an appropriate cellular context, we generated ATDC5 stable lines expressing N-terminal His-tagged versions of CR3HD CUX1-Ancestral and CUX1-Δ4aa variants. ATDC5 is a

well-characterized mouse chondrogenic cell line that largely recapitulates in vitro the differentiation landmarks of chondrocytes (77). We performed quantitative reverse transcription polymerase chain reaction (RT-qPCR) on a selected number of genes containing predicted strong function-altering variants in *P. harrisi* (Table 1) and showing detectable levels of expression in ATDC5 cells. In addition, we measured the expression of *Ptch1*, the receptor of the Hh pathway. Our experiments indicate that the CUX1-Ancestral variant transcriptionally up-regulated the expression of *Ofd1* (factor of 1.7, $P = 1.2 \times 10^{-6}$; Fig. 4C) and *Fat1* (factor of 1.8, $P = 0.029$; Fig. 4C) and down-regulated the expression of *Ift122* (factor of 0.77, $P = 0.0025$; Fig. 4C) and *Ptch1* (factor of 0.53, $P = 0.014$; Fig. 4C) relative to the control line. In contrast, neither *Dync2h1*

nor *Wdr34* expression levels were changed by CUX1-Ancestral overexpression (Fig. 4C). These results suggest that cilia- and Hh-related genes are likely transcriptional targets of CUX1 in chondrocytes.

Impaired transcriptional activity of the Galapagos cormorant CUX1

The Galapagos cormorant CUX1 showed impaired transcriptional activity relative to the ancestral variant. *Ofd1* was significantly up-regulated in CUX1-Δ4aa cells relative to control cells [factor of 1.2, $P = 0.021$, analysis of variance (ANOVA) and Tukey honest significant difference (HSD) test, Fig. 4C]; however, *Ofd1* up-regulation was significantly reduced in CUX1-Δ4aa cells relative to CUX1-Ancestral cells (factor of 1.2 versus 1.7, $P = 6 \times 10^{-5}$;

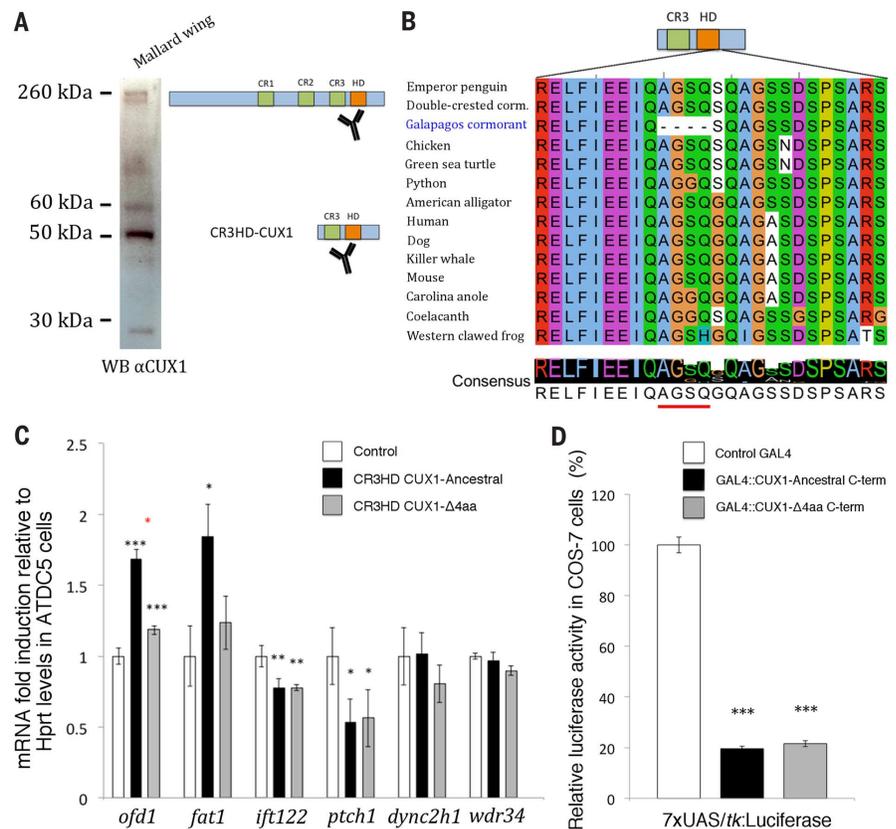


Fig. 4. The Galapagos cormorant *Cux1* is a transcriptional activation hypomorph. (A) Western blot showing the expression of CUX1 isoforms in the developing wing of a mallard embryo (22 days). The most abundant band corresponds to the predicted size of the CR3HD CUX1 isoform. (B) Protein alignment showing the deleted AGSQ residues in the Galapagos cormorant CUX1 and their high degree of conservation among vertebrates. (C) Differential up-regulation of genes by CUX1-Ancestral and CUX1-Δ4aa variants in ATDC5 cells. Pooled stable lines carrying CR3HD CUX1-Ancestral or CUX1-Δ4aa variants were generated by lentiviral transduction and puromycin selection. Control cells were transduced with an empty vector. Gene expression levels were measured by RT-qPCR ($n = 5$ biological replicates, each comprising three technical replicates). (D) Luciferase-based assay to test the repression activity of CUX1 C-terminal domain lacking CR3 and HD domains. GAL4 DNA binding domain was fused to CUX1-Ancestral or CUX1-Δ4aa variants. Both constructs equally repressed a promoter containing UAS binding sites in COS-7 cells ($n = 3$ biological replicates, each comprising three technical replicates). In (C) and (D), error bars denote SE. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (ANOVA and Tukey HSD); black and red asterisks respectively denote significant difference of CUX1 variant versus control and CUX1-Ancestral variant versus CUX1-Δ4aa variant.

Fig. 4C). Similarly, although *Fat1* was significantly up-regulated in CUX1-Ancestral cells relative to control cells (factor of 1.8, $P = 0.029$), *Fat1* expression levels in CUX1-Δ4aa cells were not significantly different from control lines (factor of 1.2, $P = 0.57$; Fig. 4C). The difference between *Fat1*

up-regulation in CUX1-Δ4aa and CUX1-Δ4aa cells was not significant (factor of 1.8 versus 1.3, $P = 0.17$; Fig. 4C). In contrast, CUX1-Δ4aa cells significantly repressed both *Ift122* (factor of 0.78, $P = 0.0026$) and *Ptch1* (factor of 0.56, $P = 0.024$), and there were no significant differences between

CUX1-Ancestral and CUX1-Δ4aa cells (factor of 0.78 versus 0.78 for *Ift122*, $P = 0.99$; factor of 0.53 versus 0.56 for *Ptch1*, $P = 0.96$; Fig. 4C). These results suggest that the 4-amino acid deletion in the Galapagos cormorant CUX1 affects its ability to activate but not to repress gene expression; they are also consistent with our luciferase reporter assays, which showed no effect on repression (Fig. 4D). It is notable that both the transcriptional activator (*Cux1*) and its target genes (*Odf1* and *Fat1*) exhibit function-altering variants in the Galapagos cormorant.

CR3HD-CUX1 promotes chondrogenesis

Chondrocytes are the main engine of bone growth. The growth of skeletal elements depends on the precise regulation of chondrocyte proliferation and hypertrophy. Mutations that affect cilia result in premature arrest of bone growth due to defects in Indian Hedgehog (IHH) signaling in chondrocytes (78). To test the role of CR3HD-CUX1 in chondrogenesis, we differentiated control, CUX1-Ancestral, and CUX1-Δ4aa ATDC5 cell lines and quantified the expression of *Ihh* and *Sox9*, two well-established markers of chondrocyte differentiation in vitro and in vivo (78). Overexpression of both CR3HD CUX1-Ancestral and CUX1-Δ4aa variants promoted chondrogenic differentiation of ATDC5 cells after 7 and 12 days of differentiation (Fig. 5A). However, the CUX1-Δ4aa variant was not as efficient as the ancestral variant, showing significant differences from CUX1-Ancestral in *Ihh* expression after 7 days of differentiation (~50% decrease, $P = 5.9 \times 10^{-4}$, ANOVA and Tukey HSD test; Fig. 5A) and in *Sox9* expression after 12 days (~15% decrease, $P = 1.6 \times 10^{-2}$; Fig. 5A). These results suggest that the Galapagos cormorant CUX1 is probably not as effective as the ortholog from its flighted relatives in promoting chondrogenic differentiation, and that mutations in *Cux1* may affect the dynamics of chondrogenesis. This observation is further supported by findings that CUX1 is expressed in the hypertrophic chondrocytes of developing bones in mice, and that the bones of *Cux1* mutant mice are thin and flaky (79).

Possible evolutionary scenarios

Loss of flight has traditionally been attributed to relaxed selection. In this scenario, the first cormorants that inhabited the Galapagos Islands found a unique environment that lacked predators and provided food year-round, drastically reducing the need to migrate. However, we found no evidence for pseudogenization of developmental genes in *P. harrisi* (17) (tables S10 and S11). On the other hand, loss of flight in the Galapagos cormorant is thought to confer an advantage for diving by decreasing buoyancy via shorter wings and by indirectly allowing increased oxygen storage via larger body size (80). This advantage could make flightlessness a target of positive selection.

We evaluated whether any of our candidate genes (Table 1) showed signatures of positive selection in the Galapagos cormorant lineage by estimating the ratio of nonsynonymous to synonymous substitutions ($\omega = dN/dS$). This is a very stringent test of selection because it assumes that

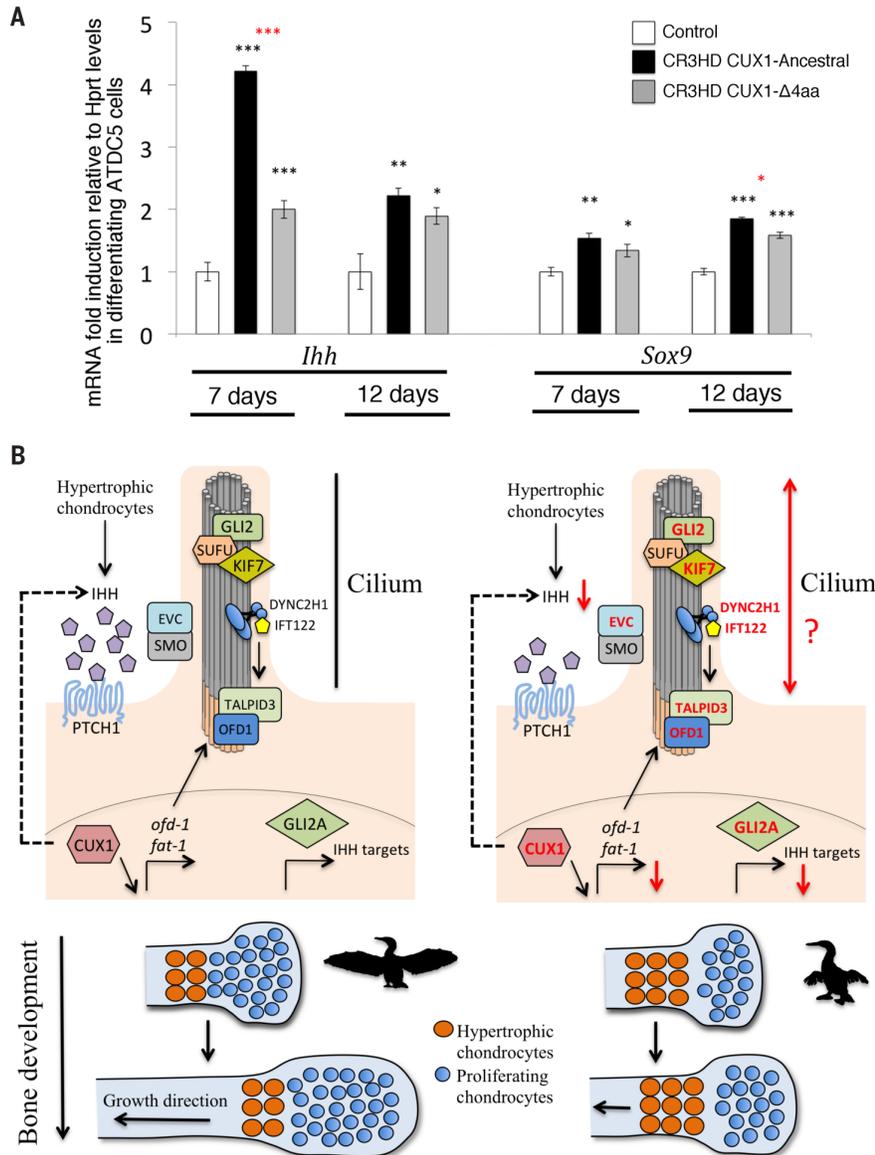


Fig. 5. CR3HD CUX1 promotes chondrogenesis. (A) ATDC5 control cells and cells carrying CR3HD CUX1-Ancestral or CUX1-Δ4aa variants were differentiated into chondrocytes. Gene expression levels were measured by RT-qPCR ($n = 4$ biological replicates, each comprising three technical replicates) after 7 and 12 days. Error bars denote SE. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (ANOVA and Tukey HSD); black and red asterisks respectively denote significant difference of CUX1 variant versus control and CUX1-Ancestral variant versus CUX1-Δ4aa variant. (B) Proposed mechanism for the reduction of wing size in *P. harrisi*. Left: Normal functioning of IHH signaling pathway in vertebrates. CUX1 regulates the expression of cilia-related genes such as *Odf1* and promotes chondrogenesis. Right: State of the IHH pathway in *P. harrisi*. Proteins in red have predicted function-altering variants in *P. harrisi*. We propose that these variants would affect both cilia formation and functioning, leading to a reduction in IHH pathway activity. As a result, the pool of proliferating chondrocytes would decrease in wing bones and the number of hypertrophic chondrocytes would increase, resulting in impaired bone growth.

all sites in a protein are evolving under the same selective pressure, a condition rarely met in highly conserved regulatory genes (81). We found that 3 of 11 tested genes showed signs of positive selection ($\omega > 1$) in the Galapagos cormorant lineage compared to a background phylogeny of 35 taxa (*Ofd1* $\omega = 1.92$, *Evc* $\omega = 1.93$, *Gli2* $\omega = 1.10$; table S8).

One of these three genes, *Gli2*, showed a statistically significant difference [$\omega = 1.10$ (Galapagos branch) versus $\omega = 0.11$ (background branch), $P = 0.0024$; table S8]. In contrast, *Gli2* showed no sign of selection in the sister group of *P. harrisi* (*P. auritus* and *P. brasiliensis*) [$\omega = 0.0001$ (sister branch) versus $\omega = 0.11$ (background branch), $P = 0.46$]. As a control, we also analyzed *Gli3*, the partially redundant paralog of *Gli2*, which also mediates Hh signaling but has no predicted function-altering variants in *P. harrisi*, and found no evidence for positive selection [$\omega = 0.04$ (Galapagos branch) versus $\omega = 0.15$ (background branch), $P = 0.11$; table S8]. These results suggest that selection toward flightlessness may be partially responsible for the phenotype of *P. harrisi*.

Discussion

The study of evolution of flightlessness in the Rallidae family led to the hypothesis that flightlessness could be a fast-evolving heterochronic condition (10, 82). Heterochrony, the relative change in the rate or timing of developmental events among species, is thought to be an important factor contributing to macroevolutionary change (83). Yet virtually nothing is known about its genetic and molecular mechanisms.

Diverse myological, osteological, and developmental observations suggest that flightlessness in the Galapagos cormorant is caused by the retention into adulthood of juvenile characteristics affecting pectoral and forelimb development (a class of heterochrony known as paedomorphosis) (7). Here, we propose a genetic and molecular model that may explain this heterochronic condition, where the perturbations of cilia/Ihh signaling may be responsible for the reduction in growth of both keel and wings in the Galapagos cormorant (Fig. 5B). However, we cannot rule out a role of *Cux1* in the AER. Of special interest is the gene *Fat1*, a target of *Cux1* (Fig. 4D), which contains two putative function-altering variants (Table 1). *Fat1*^{-/-} mouse mutants are viable and show selective defects in facial, pectoral, and shoulder muscles but not in hindlimb muscles (54). Thus, variants in *Fat1* could explain the underdeveloped pectoral muscles of *P. harrisi*.

Although we have identified multiple variants that likely contribute to the flightless phenotype of *P. harrisi*, we cannot exclude the possibility that other genes and pathways may contribute to the phenotype, nor the contribution of noncoding regulatory variants (17). Further characterization of the individual and joint contributions of the variants found in this study will help us to reconstruct the chain of events leading to flightlessness and to genetically dissect macroevolutionary change. We hypothesize that mutations in cilia or functionally related genes could be responsible for limb and other skeletal heterochronic trans-

formations in birds and diverse organisms, including humans.

REFERENCES AND NOTES

1. D. A. Roff, The evolution of flightlessness: Is history important? *Evol. Ecol.* **8**, 639–657 (1994). doi: [10.1007/BF01237847](https://doi.org/10.1007/BF01237847)
2. C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
3. J. Harshman et al., Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13462–13467 (2008). doi: [10.1073/pnas.0803242105](https://doi.org/10.1073/pnas.0803242105); pmid: [18765814](https://pubmed.ncbi.nlm.nih.gov/18765814/)
4. A. J. Baker, O. Haddrath, J. D. McPherson, A. Cloutier, Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Mol. Biol. Evol.* **31**, 1686–1696 (2014). doi: [10.1093/molbev/msu153](https://doi.org/10.1093/molbev/msu153); pmid: [24825849](https://pubmed.ncbi.nlm.nih.gov/24825849/)
5. K. J. Mitchell et al., Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**, 898–900 (2014). doi: [10.1126/science.1251981](https://doi.org/10.1126/science.1251981); pmid: [24855267](https://pubmed.ncbi.nlm.nih.gov/24855267/)
6. N. A. Wright, D. W. Steadman, C. C. Witt, Predictable evolution toward flightlessness in volant island birds. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4765–4770 (2016). doi: [10.1073/pnas.1522931113](https://doi.org/10.1073/pnas.1522931113); pmid: [27071105](https://pubmed.ncbi.nlm.nih.gov/27071105/)
7. B. C. Livezey, Flightlessness in the Galapagos cormorant (*Compsohalieu* [Nannopterum] *harrisi*): Heterochrony, gigantism, and specialization. *Zool. J. Linn. Soc.* **105**, 155–224 (1992). doi: [10.1111/j.1096-3642.1992.tb01229.x](https://doi.org/10.1111/j.1096-3642.1992.tb01229.x)
8. M. A. G. de Bakker et al., Digit loss in archosaur evolution and the interplay between selection and constraints. *Nature* **500**, 445–448 (2013). doi: [10.1073/pnas.1522931113](https://doi.org/10.1073/pnas.1522931113); pmid: [27071105](https://pubmed.ncbi.nlm.nih.gov/27071105/)
9. The onset times of several homologous developmental events, such as the appearance of immature feathers, dropping of egg tooth, and independence from parental care, occur progressively later in the Galapagos cormorant chick than in *P. auritus*.
10. A. Feduccia, *The Origin and Evolution of Birds* (Yale Univ. Press, ed. 2, 1999).
11. K. E. Slack et al., Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. *Mol. Biol. Evol.* **23**, 1144–1155 (2006). doi: [10.1093/molbev/msj124](https://doi.org/10.1093/molbev/msj124); pmid: [16533822](https://pubmed.ncbi.nlm.nih.gov/16533822/)
12. M. Kennedy, C. A. Valle, H. G. Spencer, The phylogenetic position of the Galapagos cormorant. *Mol. Phylogenet. Evol.* **53**, 94–98 (2009). doi: [10.1016/j.ympev.2009.06.002](https://doi.org/10.1016/j.ympev.2009.06.002); pmid: [19523526](https://pubmed.ncbi.nlm.nih.gov/19523526/)
13. M. Kennedy, H. G. Spencer, Classification of the cormorants of the world. *Mol. Phylogenet. Evol.* **79**, 249–257 (2014). doi: [10.1016/j.ympev.2014.06.020](https://doi.org/10.1016/j.ympev.2014.06.020); pmid: [24994028](https://pubmed.ncbi.nlm.nih.gov/24994028/)
14. A recent phylogeny of the Phalacrocoracidae suggested classifying *P. harrisi*, *P. auritus*, and *P. brasiliensis* as members of the monophyletic genera Nannopterum.
15. R. Luo et al., SOAPdenovo: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012). doi: [10.1186/2047-217X-1-18](https://doi.org/10.1186/2047-217X-1-18); pmid: [23587118](https://pubmed.ncbi.nlm.nih.gov/23587118/)
16. G. Parra, K. Bradnam, Z. Ning, T. Keane, I. Korf, Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289–297 (2009). doi: [10.1093/nar/gkn916](https://doi.org/10.1093/nar/gkn916); pmid: [19042974](https://pubmed.ncbi.nlm.nih.gov/19042974/)
17. See supplementary materials.
18. International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004). doi: [10.1038/nature03154](https://doi.org/10.1038/nature03154); pmid: [15592404](https://pubmed.ncbi.nlm.nih.gov/15592404/)
19. W. C. Warren et al., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: [10.1038/nature08819](https://doi.org/10.1038/nature08819); pmid: [20360741](https://pubmed.ncbi.nlm.nih.gov/20360741/)
20. G. Ganapathy et al., High-coverage sequencing and annotated assemblies of the budgerigar genome. *Gigascience* **3**, 11 (2014). doi: [10.1186/2047-217X-3-11](https://doi.org/10.1186/2047-217X-3-11); pmid: [25061512](https://pubmed.ncbi.nlm.nih.gov/25061512/)
21. D. J. Geist, H. Snell, H. Snell, C. Goddard, M. D. Kurz, in *The Galapagos: A Natural Laboratory for the Earth Sciences*, K. S. Harpp, E. Mittelstaedt, N. D'Ozouville, D. W. Graham, Eds. (Wiley, 2014), pp. 145–166.
22. G. Wang, X. Li, Population dynamics and recovery of endangered crested ibis (*Nipponia nippon*) in central China. *Waterbirds* **31**, 489–494 (2008). doi: [10.1675/1524-4695-31.3.489](https://doi.org/10.1675/1524-4695-31.3.489)
23. C. A. Valle, M. C. M. Coulter, Present status of the flightless cormorant, Galapagos penguin and greater flamingo populations in the Galapagos Islands, Ecuador, after the 1982–83 El Niño. *Condor* **89**, 276–281 (1987). doi: [10.2307/1368480](https://doi.org/10.2307/1368480)

24. J. Alföldi, K. Lindblad-Toh, Comparative genomics as a tool to understand evolution and disease. *Genome Res.* **23**, 1063–1068 (2013). doi: [10.1101/gr.157503.113](https://doi.org/10.1101/gr.157503.113); pmid: [23817047](https://pubmed.ncbi.nlm.nih.gov/23817047/)
25. R. Jelier, J. I. Semple, R. Garcia-Verdugo, B. Lehner, Predicting phenotypic variation in yeast from individual genome sequences. *Nat. Genet.* **43**, 1270–1274 (2011). doi: [10.1038/ng.1007](https://doi.org/10.1038/ng.1007); pmid: [22081227](https://pubmed.ncbi.nlm.nih.gov/22081227/)
26. M. D. Shapiro et al., Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004). doi: [10.1038/nature02415](https://doi.org/10.1038/nature02415); pmid: [15085123](https://pubmed.ncbi.nlm.nih.gov/15085123/)
27. D. L. Stern, V. Orgogozo, The loci of evolution: How predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008). doi: [10.1111/j.1558-5646.2008.00450.x](https://doi.org/10.1111/j.1558-5646.2008.00450.x); pmid: [18616572](https://pubmed.ncbi.nlm.nih.gov/18616572/)
28. K. S. Pollard et al., An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006). doi: [10.1038/nature05113](https://doi.org/10.1038/nature05113); pmid: [16915236](https://pubmed.ncbi.nlm.nih.gov/16915236/)
29. B. M. Booker et al., Bat accelerated regions identify a bat forelimb specific enhancer in the *HoxD* locus. *PLoS Genet.* **12**, e1005738 (2016). doi: [10.1371/journal.pgen.1005738](https://doi.org/10.1371/journal.pgen.1005738); pmid: [27019019](https://pubmed.ncbi.nlm.nih.gov/27019019/)
30. S. Prabhakar et al., Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008). doi: [10.1126/science.1159974](https://doi.org/10.1126/science.1159974); pmid: [18772437](https://pubmed.ncbi.nlm.nih.gov/18772437/)
31. T. M. Frayling et al., A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007). doi: [10.1126/science.1141634](https://doi.org/10.1126/science.1141634); pmid: [17434869](https://pubmed.ncbi.nlm.nih.gov/17434869/)
32. A. Visel et al., ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009). doi: [10.1038/nature07730](https://doi.org/10.1038/nature07730); pmid: [19212405](https://pubmed.ncbi.nlm.nih.gov/19212405/)
33. J. Cotney et al., Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* **22**, 1069–1080 (2012). doi: [10.1101/gr.129817.111](https://doi.org/10.1101/gr.129817.111); pmid: [22421546](https://pubmed.ncbi.nlm.nih.gov/22421546/)
34. Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, A. P. Chan, Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688 (2012). doi: [10.1371/journal.pone.0046688](https://doi.org/10.1371/journal.pone.0046688); pmid: [23056405](https://pubmed.ncbi.nlm.nih.gov/23056405/)
35. L. Rockah-Shmuel, Á. Tóth-Petróczy, D. S. Tawfik, Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. *PLoS Comput. Biol.* **11**, e1004421 (2015). doi: [10.1371/journal.pcbi.1004421](https://doi.org/10.1371/journal.pcbi.1004421); pmid: [26274323](https://pubmed.ncbi.nlm.nih.gov/26274323/)
36. D. L. Stern, Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000). doi: [10.1111/j.0014-3820.2000.tb00544.x](https://doi.org/10.1111/j.0014-3820.2000.tb00544.x)
37. A. M. Waters, P. L. Beales, Ciliopathies: An expanding disease spectrum. *Pediatr. Nephrol.* **26**, 1039–1056 (2011). doi: [10.1007/s00467-010-1731-7](https://doi.org/10.1007/s00467-010-1731-7); pmid: [21210154](https://pubmed.ncbi.nlm.nih.gov/21210154/)
38. S. C. Goetz, K. V. Anderson, The primary cilium: A signalling centre during vertebrate development. *Nat. Rev. Genet.* **11**, 331–344 (2010). doi: [10.1038/nrg2774](https://doi.org/10.1038/nrg2774); pmid: [20395968](https://pubmed.ncbi.nlm.nih.gov/20395968/)
39. C. J. Haycraft et al., *Gli2* and *Gli3* localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS Genet.* **1**, e53 (2005). doi: [10.1371/journal.pgen.0010053](https://doi.org/10.1371/journal.pgen.0010053); pmid: [16254602](https://pubmed.ncbi.nlm.nih.gov/16254602/)
40. I. J. Bisschoff et al., Novel mutations including deletions of the entire *OFD1* gene in 30 families with type 1 orofaciodigital syndrome: A study of the extensive clinical variability. *Hum. Mutat.* **34**, 237–247 (2013). doi: [10.1002/humu.22224](https://doi.org/10.1002/humu.22224); pmid: [23033313](https://pubmed.ncbi.nlm.nih.gov/23033313/)
41. K. L. M. Coene et al., *OFD1* is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am. J. Hum. Genet.* **85**, 465–481 (2009). doi: [10.1016/j.ajhg.2009.09.002](https://doi.org/10.1016/j.ajhg.2009.09.002); pmid: [19800048](https://pubmed.ncbi.nlm.nih.gov/19800048/)
42. G. Baujat, M. Le Merrer, Ellis-van Creveld syndrome. *Orphanet J. Rare Dis.* **2**, 27 (2007). doi: [10.1186/1750-1172-2-27](https://doi.org/10.1186/1750-1172-2-27); pmid: [17547743](https://pubmed.ncbi.nlm.nih.gov/17547743/)
43. S. Bimonte et al., *Ofd1* is required in limb bud patterning and endochondral bone development. *Dev. Biol.* **349**, 179–191 (2011). doi: [10.1016/j.ydbio.2010.09.020](https://doi.org/10.1016/j.ydbio.2010.09.020); pmid: [20920500](https://pubmed.ncbi.nlm.nih.gov/20920500/)
44. P. J. R. Ocbina, J. T. Eggenschwiler, I. Moskowitz, K. V. Anderson, Complex interactions between genes controlling trafficking in primary cilia. *Nat. Genet.* **43**, 547–553 (2011). doi: [10.1038/ng.832](https://doi.org/10.1038/ng.832); pmid: [21552265](https://pubmed.ncbi.nlm.nih.gov/21552265/)
45. J. Qin, Y. Lin, R. X. Norman, H. W. Ko, J. T. Eggenschwiler, Intraflagellar transport protein 122 antagonizes Sonic Hedgehog signaling and controls ciliary localization of pathway components. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1456–1461 (2011). doi: [10.1073/pnas.101140108](https://doi.org/10.1073/pnas.101140108); pmid: [21209331](https://pubmed.ncbi.nlm.nih.gov/21209331/)

46. J. Walczak-Szulpa *et al.*, Cranioectodermal dysplasia, Sensenbrenner syndrome, is a ciliopathy caused by mutations in the IFT122 gene. *Am. J. Hum. Genet.* **86**, 949–956 (2010). doi: [10.1016/j.ajhg.2010.04.012](https://doi.org/10.1016/j.ajhg.2010.04.012); pmid: 20493458
47. L. R. Bell, S. Stone, J. Yochem, J. E. Shaw, R. K. Herman, The molecular identities of the *Caenorhabditis elegans* intraflagellar transport genes *dyf-6*, *daf-10* and *osm-1*. *Genetics* **173**, 1275–1286 (2006). doi: [10.1534/genetics.106.056721](https://doi.org/10.1534/genetics.106.056721); pmid: 16648645
48. V. Singla, J. F. Reiter, The primary cilium as the cell's antenna: Signaling at a sensory organelle. *Science* **313**, 629–633 (2006). doi: [10.1126/science.1124534](https://doi.org/10.1126/science.1124534); pmid: 16888132
49. L. A. Perkins, E. M. Hedgecock, J. N. Thomson, J. G. Culotti, Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986). doi: [10.1016/0012-1606\(86\)90314-3](https://doi.org/10.1016/0012-1606(86)90314-3); pmid: 2428682
50. E. Z. Macosko *et al.*, A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009). doi: [10.1038/nature07886](https://doi.org/10.1038/nature07886); pmid: 19349961
51. T. J. Park, S. L. Haigo, J. B. Wallingford, Ciliogenesis defects in embryos lacking inturnd or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. *Nat. Genet.* **38**, 303–311 (2006). doi: [10.1038/ng1753](https://doi.org/10.1038/ng1753); pmid: 16493421
52. R. S. Gray *et al.*, The planar cell polarity effector Fuz is essential for targeted membrane trafficking, ciliogenesis and mouse embryonic development. *Nat. Cell Biol.* **11**, 1225–1232 (2009). doi: [10.1038/ncb1966](https://doi.org/10.1038/ncb1966); pmid: 19767740
53. H. Zeng, A. N. Hoover, A. Liu, PCP effector gene Inturnd is an important regulator of cilia formation and embryonic development in mammals. *Dev. Biol.* **339**, 418–428 (2010). doi: [10.1016/j.ydbio.2010.01.003](https://doi.org/10.1016/j.ydbio.2010.01.003); pmid: 20067783
54. N. Caruso *et al.*, Deregulation of the protocadherin gene *FAT1* alters muscle shapes: Implications for the pathogenesis of facioscapulohumeral dystrophy. *PLoS Genet.* **9**, e1003550 (2013). doi: [10.1371/journal.pgen.1003550](https://doi.org/10.1371/journal.pgen.1003550); pmid: 23785297
55. J. White *et al.*, DVL1 frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. *Am. J. Hum. Genet.* **96**, 612–622 (2015). doi: [10.1016/j.ajhg.2015.02.015](https://doi.org/10.1016/j.ajhg.2015.02.015); pmid: 25817016
56. K. J. Bunn *et al.*, Mutations in *DVL1* cause an osteosclerotic form of Robinow syndrome. *Am. J. Hum. Genet.* **96**, 623–630 (2015). doi: [10.1016/j.ajhg.2015.02.010](https://doi.org/10.1016/j.ajhg.2015.02.010); pmid: 25817014
57. C. V. Duffie, T. C. Glenn, F. H. Vargas, P. G. Parker, Genetic structure within and between island populations of the flightless cormorant (*Phalacrocorax harrisi*). *Mol. Ecol.* **18**, 2103–2111 (2009). doi: [10.1111/j.1365-294X.2009.04179.x](https://doi.org/10.1111/j.1365-294X.2009.04179.x); pmid: 19635072
58. A. Koch, F. Poirier, R. Jacob, D. Delacour, Galectin-3, a novel centrosome-associated protein, required for epithelial morphogenesis. *Mol. Biol. Cell* **21**, 219–231 (2010). doi: [10.1091/mbc.E09-03-0193](https://doi.org/10.1091/mbc.E09-03-0193); pmid: 19923323
59. M. Paces-Fessy, D. Boucher, E. Petit, S. Paute-Briand, M.-F. Blanchet-Tournier, The negative regulator of Gli, Suppressor of fused (Sufu), interacts with SAP18, Galectin3 and other nuclear proteins. *Biochem. J.* **378**, 353–362 (2004). doi: [10.1042/bj20030786](https://doi.org/10.1042/bj20030786); pmid: 14611647
60. C. Colnot, S. S. Sidhu, N. Balmain, F. Poirier, Uncoupling of chondrocyte death and vascular invasion in mouse galectin 3 null mutant bones. *Dev. Biol.* **229**, 203–214 (2001). doi: [10.1006/dbio.2000.9933](https://doi.org/10.1006/dbio.2000.9933); pmid: 11133164
61. L. Sansregret, A. Nepveu, The multiple roles of CUX1: Insights from mouse models and cell-based assays. *Gene* **412**, 84–94 (2008). doi: [10.1016/j.gene.2008.01.017](https://doi.org/10.1016/j.gene.2008.01.017); pmid: 18313863
62. V. Sudarsan, S. Anant, P. Gupta, K. VijayRaghavan, H. Skaer, Myoblast diversification and ectodermal signaling in *Drosophila*. *Dev. Cell* **1**, 829–839 (2001). doi: [10.1016/S1534-5807\(01\)00089-2](https://doi.org/10.1016/S1534-5807(01)00089-2); pmid: 11740944
63. A. T. Tavares, T. Tsukui, J. C. Izipisúa Belmonte, Evidence that members of the Cut/Cux/CDP family may be involved in AER positioning and polarizing activity during chick limb development. *Development* **127**, 5133–5144 (2000). pmid: 11060239
64. G. Lizarraga, A. Lichtler, W. B. Upholt, R. A. Koshier, Studies on the role of Cux1 in regulation of the onset of joint formation in the developing limb. *Dev. Biol.* **243**, 44–54 (2002). doi: [10.1006/dbio.2001.0559](https://doi.org/10.1006/dbio.2001.0559); pmid: 11846476
65. The chicken CUX1-WT variant studied by Tavares *et al.* was inadvertently missing the Cut2 DNA binding domain, as revealed by the chicken genome generated subsequently to their study (Ensembl ID: ENSGALP00000002788). This isoform is likely hypomorphic, and its overexpression may exert a dominant negative effect.
66. S. Matsuoka *et al.*, ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007). doi: [10.1126/science.1140321](https://doi.org/10.1126/science.1140321); pmid: 17525332
67. D. Iyaguchi, M. Yao, N. Watanabe, J. Nishihira, I. Tanaka, DNA recognition mechanism of the ONECUT homeodomain of transcription factor HNF-6. *Structure* **15**, 75–83 (2007). doi: [10.1016/j.str.2006.11.004](https://doi.org/10.1016/j.str.2006.11.004); pmid: 17223534
68. N. S. Moon, G. Bérubé, A. Nepveu, CCAAT displacement activity involves CUT repeats 1 and 2, not the CUT homeodomain. *J. Biol. Chem.* **275**, 31325–31334 (2000). doi: [10.1074/jbc.M002912200](https://doi.org/10.1074/jbc.M002912200); pmid: 10864926
69. F. Mailly *et al.*, The human cut homeodomain protein can repress gene expression by two distinct mechanisms: Active repression and competition for binding site occupancy. *Mol. Cell. Biol.* **16**, 5346–5357 (1996). doi: [10.1128/MCB.16.10.5346](https://doi.org/10.1128/MCB.16.10.5346); pmid: 8816446
70. M. Truscott *et al.*, Carboxyl-terminal proteolytic processing of CUX1 by a caspase enables transcriptional activation in proliferating cells. *J. Biol. Chem.* **282**, 30216–30226 (2007). doi: [10.1074/jbc.M702328200](https://doi.org/10.1074/jbc.M702328200); pmid: 17681953
71. H. Nishio, M. J. Walsh, CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11257–11262 (2004). doi: [10.1073/pnas.0401343101](https://doi.org/10.1073/pnas.0401343101); pmid: 15269344
72. C. Cadieux *et al.*, Polycystic kidneys caused by sustained expression of Cux1 isoform p75. *J. Biol. Chem.* **283**, 13817–13824 (2008). doi: [10.1074/jbc.M709332200](https://doi.org/10.1074/jbc.M709332200); pmid: 18356167
73. G. Stratigopoulos, C. A. LeDuc, M. L. Cremona, W. K. Chung, R. L. Leibell, Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator-interacting protein-1-like (RPGRIPL) genes and coordinates leptin receptor signaling. *J. Biol. Chem.* **286**, 2155–2170 (2011). doi: [10.1074/jbc.M110.188482](https://doi.org/10.1074/jbc.M110.188482); pmid: 21037323
74. T. Ellis *et al.*, The transcriptional repressor CDP (Cut1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001). doi: [10.1101/gad.200101](https://doi.org/10.1101/gad.200101); pmid: 11544187
75. C. Vadnais *et al.*, Long-range transcriptional regulation by the p110 CUX1 homeodomain protein on the ENCODE array. *BMC Genomics* **14**, 258 (2013). doi: [10.1186/1471-2164-14-258](https://doi.org/10.1186/1471-2164-14-258); pmid: 23590133
76. Z. M. Ramdzan, A. Nepveu, CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nat. Rev. Cancer* **14**, 673–682 (2014). doi: [10.1038/nrc.3805](https://doi.org/10.1038/nrc.3805); pmid: 25190083
77. C. Shukunami *et al.*, Chondrogenic differentiation of clonal mouse embryonic cell line ATDC5 in vitro: Differentiation-dependent gene expression of parathyroid hormone (PTH)/PTH-related peptide receptor. *J. Cell Biol.* **133**, 457–468 (1996). doi: [10.1083/jcb.133.2.457](https://doi.org/10.1083/jcb.133.2.457); pmid: 8609176
78. H. M. Kronenberg, Developmental regulation of the growth plate. *Nature* **423**, 332–336 (2003). doi: [10.1038/nature01657](https://doi.org/10.1038/nature01657); pmid: 12748651
79. A. M. Sinclair *et al.*, Lymphoid apoptosis and myeloid hyperplasia in CCAAT displacement protein mutant mice. *Blood* **98**, 3658–3667 (2001). doi: [10.1182/blood.V98.13.3658](https://doi.org/10.1182/blood.V98.13.3658); pmid: 11739170
80. Y. Y. Watanabe, A. Takahashi, K. Sato, M. Viviant, C.-A. Bost, Poor flight performance in deep-diving cormorants. *J. Exp. Biol.* **214**, 412–421 (2011). doi: [10.1242/jeb.050161](https://doi.org/10.1242/jeb.050161); pmid: 21228200
81. Z. Yang, Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573 (1998). doi: [10.1093/oxfordjournals.molbev.a025957](https://doi.org/10.1093/oxfordjournals.molbev.a025957)
82. S. L. Olson, Evolution of the rails of the South Atlantic islands (Aves: Rallidae). *Smithson. Contrib. Zool.* **152**, 1–53 (1973). doi: [10.5479/si.00810282.152](https://doi.org/10.5479/si.00810282.152)
83. S. J. Gould, *Ontogeny and Phylogeny* (Belknap/Harvard Univ. Press, 1977).

ACKNOWLEDGMENTS

Supported by the Jane Coffin Childs Memorial Fund for Medical Research (A.B.), the Howard Hughes Medical Institute (L.K.), the U.S. Geological Survey through the Wildlife Program of the Ecosystems Mission Area (A.M.R.), a Gruss-Lipper postdoctoral fellowship from the EGL Charitable Foundation (E.B.-D.), FONDECYT grant 11130305 (C.V.), and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (K.L.). Field collection permits and export permits were by the Galapagos National Park, facilitated by the Charles Darwin Foundation. Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. government. We thank C. Koo and M. Arbing at the UCLA-DOE Protein Expression Laboratory Core Facility for protein purification; S. Feng at the Broad Stem Cell Research Center High Throughput Sequencing Core for assistance; C. Cicero for granting access to cormorant specimens (134079 and 151575) at the Museum of Vertebrate Zoology at Berkeley; C. Valle (Universidad San Francisco de Quito) for advice; and K. Garrett (Natural History Museum of Los Angeles), K. Burns (San Diego Natural History Museum and San Diego State University Museum of Biodiversity), and R. Duerr (International Bird Rescue) for providing samples used in preliminary stages of this study. Author contributions: A.B. and L.K. conceived the study; A.B. coordinated the collection of samples, prepared libraries, assembled and annotated genomes, and performed analyses and experiments; E.B.-D. and A.B. performed the accelerated evolution analysis; P.C.W., A.M.R., C.V., and P.G.P. provided DNA or tissue samples; W.W. and A.B. carried out ATDC5 cell line experiments supervised by K.L.; A.B. and L.K. wrote the manuscript; and all authors discussed and agreed on the final version of the manuscript. All sequencing data from this study is available through the NCBI Sequence Read Archive under Bioproject accession number PRJNA321273. Alignments used for phylogenetic analysis and selection test are available at DRYAD doi:10.5061/dryad.8m2t5. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/356/6341/eaal3345/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S13
Movie S1
References (84–109)

4 November 2016; accepted 13 April 2017
10.1126/science.aal3345

A genetic signature of the evolution of loss of flight in the Galapagos cormorant

Alejandro Burga, Weiguang Wang, Eyal Ben-David, Paul C. Wolf, Andrew M. Ramey, Claudio Verdugo, Karen Lyons, Patricia G. Parker and Leonid Kruglyak

Science **356** (6341), eaal3345.
DOI: 10.1126/science.aal3345

Loss of flight in the Galapagos cormorant

Although rare among existing birds, the loss of flight appears to have occurred multiple times in evolutionary history. However, the genetic changes that ground avian species are not well understood. Burga *et al.* sequenced genomes from three cormorant species and compared them with that of the flightless Galapagos cormorant (see the Perspective by Cooper). They identified variants in genes involved in primary ciliogenesis. Functional analyses of these variants suggest that the impaired function of the genes may be responsible for skeletal changes associated with the Galapagos cormorant's loss of flight.

Science, this issue p. eaal3345; see also p. 904

ARTICLE TOOLS

<http://science.sciencemag.org/content/356/6341/eaal3345>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2017/05/31/356.6341.eaal3345.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/356/6341/904.full>

REFERENCES

This article cites 101 articles, 29 of which you can access for free
<http://science.sciencemag.org/content/356/6341/eaal3345#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2017, American Association for the Advancement of Science