EVOLUTIONARY BIOLOGY

A nuclear genome assembly of an extinct flightless bird, the little bush moa

Scott V. Edwards[1,2]*, Alison Cloutier[1], Glenn Cockburn[3], Robert Driver[4], Phil Grayson[1,2], Kazutaka Katoh[5], Maude W. Baldwin[3], Timothy B. Sackton[6], Allan J. Baker[7,8]†

We present a draft genome of the little bush moa (*Anomalopteryx didiformis*)—one of approximately nine species of extinct flightless birds from Aotearoa, New Zealand—using ancient DNA recovered from a fossil bone from the South Island. We recover a complete mitochondrial genome at 249.9× depth of coverage and almost 900 megabases of a male moa nuclear genome at ~4 to 5× coverage, with sequence contiguity sufficient to identify more than 85% of avian universal single-copy orthologs. We describe a diverse landscape of transposable elements and satellite repeats, estimate a long-term effective population size of ~240,000, identify a diverse suite of olfactory receptor genes and an opsin repertoire with sensitivity in the ultraviolet range, show that the wingless moa phenotype is likely not attributable to gene loss or pseudogenization, and identify potential function-altering coding sequence variants in moa that could be synthesized for future functional assays. This genomic resource should support further studies of avian evolution and morphological divergence.

INTRODUCTION

The extinct moa of New Zealand (Aves: Dinornithiformes) comprise nine currently recognized species (*1*) and belong to the Palaeognathae, which encompasses the flightless ratites (ostrich, emu, cassowary, kiwi, rheas, moa, and elephant birds) and the volant, or flying, tinamous. Extinction of all moa species is thought to have closely followed the Polynesian settlement of New Zealand in the late 13th century as the result of direct human exploitation compounded by anthropogenic land-use changes and impacts associated with introduced species (*2*).

In addition to a rich history of paleontological study [reviewed in (*3*)], ancient DNA (aDNA) has yielded multiple insights into moa biology (*4*, *5*). Cooper *et al*. (*6*) used aDNA amplified by the polymerase chain reaction (PCR) to show that moa are not most closely related to kiwi, indicating independent arrivals of these two lineages to New Zealand. Instead, aDNA places moa as sister to the volant tinamous, consistent with multiple independent losses of flight in ratites (*7*–*11*). aDNA has also helped clarify moa taxonomy (*1*, *12*) and was instrumental in identifying extreme reversed sexual size dimorphism that misled some morphological taxonomic designations (*3*, *13*). Contributions from aDNA have "clothed" moa by assigning feathers to their species of origin (*14*), identified males as the likely incubating sex from eggshell aDNA (*15*), and investigated moa feeding ecology and parasites using coprolites (*16*, *17*).

This diversity of aDNA research testifies to the wealth of relatively well-preserved moa remains (*4*). Yet, most molecular studies of moa have relied heavily upon mitochondrial DNA (mtDNA) since mtDNA occurs in high copy number per cell and is, therefore, more readily recovered than nuclear DNA from subfossil substrates, where aDNA is often highly degraded (*4*, *5*, *18*). High-throughput sequencing (HTS) has revolutionized the field of aDNA by allowing recovery of these short segments of nuclear DNA. Unlike mtDNA, which is uniparentally inherited and represents only a tiny fraction of the total genomic "blueprint" in an individual, nuclear DNA can provide much more detail about the evolutionary history and unique adaptations of extinct species (*5*, *18*). It is therefore likely that we have only just begun to access the available genetic information for moa.

We isolated aDNA from the little bush moa (*Anomalopteryx didiformis*) and used HTS to characterize its genome. Little bush moa (Fig. 1A) were distributed in lowland forests across the North and South Islands of New Zealand and were among the smallest of moa species, reaching heights of 50 to 90 cm and body weights of around 30 kg (*1*, *3*). In addition to a complete mitochondrial genome, we report the nuclear genome of a little bush moa, assembled by mapping little bush moa reads to a high-quality draft genome for the emu (*Dromaius novaehollandiae*), which previously had one of the most completely assembled genomes of any paleognath (*11*), and which now has been augmented by extensive long-read data (*19*). Moa are estimated to have diverged from their closest living relatives, the tinamous, about 53 million years ago (Ma ago), and from rheas and ostrich about 70 Ma and 79 Ma ago, respectively (*20*), making comparative mapping challenging. Nonetheless, we use this moa nuclear genome to explore diverse aspects of moa genome biology, including sensory biology, and to present genomic resources, including polymorphic microsatellites for future population-level studies and coding sequences for a suite of candidate genes for limb development, to investigate their possible association with flightlessness in moa and other ratites.

RESULTS

Library characterization and endogenous DNA content

Illumina sequencing yielded 143.4 Gb of raw data (SRA accession SRP132423; fig. S1 and table S1). Most data incorporated into the mitochondrial and nuclear genomes described below originated from TruSeq library A_didi_CTTGTA (fig. S1A) due, in part, to greater

[1]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. [2]Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. [3]Evolution of Sensory Systems Research Group, Max Planck Institute for Biological Intelligence, 82319 Seewiesen, Germany. [4]Department of Biology, East Carolina University, E 5th Street, Greenville, NC 27605, USA. [5]Department of Genome Informatics, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita 565-0871, Japan. [6]Informatics Group, Harvard University, 38 Oxford Street, Cambridge, MA 02138, USA. [7]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcox Street, Toronto, ON M5S 3B2, Canada. [8]Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON M5S 2C6, Canada.
*Corresponding author. Email: sedwards@fas.harvard.edu
†Deceased.

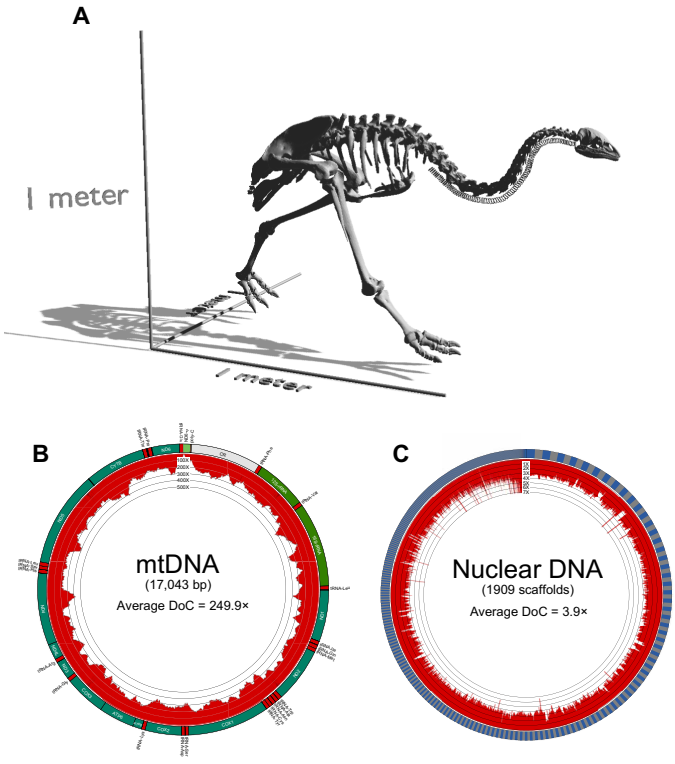

Fig. 1. Draft nuclear and mitochondrial genome assemblies of the little bush moa. (**A**) 3D depiction of a little bush moa skeleton. (**B**) De novo assembled mitochondrial genome, with locations of annotated genes and RNAs indicated. The inward-facing plot shows the per-base depth of coverage (DoC). (**C**) Reference-based nuclear genome assembly (illustrated for the original moa assembly). Alternating gray and blue sections along the outer circle indicate individual scaffolds in order of decreasing size. The inward-facing plot shows DoC calculated in 10 nonoverlapping windows per scaffold. Image in (A) by nzfauna (https://skfb.ly/FQAU) is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

sequencing effort for this library. TruSeq library A_didi_GCCAAT produced fewer reads than expected and had a high level of sequence duplication due to suboptimal cluster density (fig. S1 and table S1). Recovery of moa DNA from the three Nextera libraries was also limited, a result that could reflect the smaller amount of input DNA used in the Nextera protocol and/or a decreased amount of endogenous DNA in the size fractions assayed for these preparations.

Taxonomic profiling of reads, which represent a mixture of endogenous moa aDNA and environmental DNA, assigned taxonomy to 26 to 35% of reads across libraries (fig. S1B). The TruSeq libraries contained much higher proportions of reads assigned to Aves (e.g., all birds, 13 and 10% respectively for libraries CTTGTA and GCCAAT; fig. S1B) than the Nextera libraries, with most of these reads further assigned to Palaeognathae (fig. S1C). Total mapping rates before duplicate removal mirror estimated amounts of endogenous DNA in each library (fig. S1C and table S1), suggesting that the use of a reference genome, emu, which was relatively divergent from moa for mapping, nevertheless recovered most of the recognizable moa DNA within library extracts. Levels of read duplication (table S1) further indicate that sequencing saturation was reached to recover the maximum possible amount of endogenous DNA.

aDNA is typically degraded to fragments smaller than 500 base pairs (bp) and displays characteristic postmortem modifications leading to an excess of purines immediately preceding strand breaks and increasing cytosine (C) to thymine (T) substitutions toward fragment ends (*21*, *22*). We cannot fully assess the extent of DNA damage because library construction involves DNA shearing, meaning that fragment ends represent a mixture of naturally occurring DNA breakage as well as strand breaks induced during library preparation. However, mean lengths of mapped reads and estimated insert sizes, especially for the two TruSeq libraries, are consistent with well-preserved DNA (table S1). Consequently, although we do observe signatures of aDNA damage, the amount of damage appears minimal (fig. S2). These observations are not unprecedented for well-preserved moa specimens. Haddrath and Baker (*23*), Cooper *et al.* (*24*), and Baker *et al.* (*12*), successfully amplified moa PCR products 250 to 600 bp in length, and Cooper *et al.* (*24*) reported high endogenous DNA content and little DNA damage for samples used to sequence complete mitochondrial genomes. In addition, both the mitochondrial genome described below and phylogenetic analysis of genome-wide datasets of nuclear markers for this specimen corroborate its aDNA sequence authenticity (*7*, *8*, *11*).

Assembly of mitochondrial and nuclear genomes

We recovered a complete 17,043-bp mitochondrial genome at 249.9× average depth of coverage (DoC) following duplicate removal (GenBank accession no. MK778441; Fig. 1B). mapDamage correction yielded an identical mtDNA genome sequence. This little bush moa assembly spans the entire 1478-bp control region (D loop), which was not fully represented in the published mtDNA genome assembled from PCR-based sequencing of the same specimen (*23*). We confirm a previously identified (*25*), moa-specific mitochondrial gene order upstream of the control region and consisting of cytb:tRNA-thr:tRNA-pro:ND6:tRNA-glu:ND6-pseudogene-fragment:control-region (Fig. 1B). The assembly is nearly identical to the existing reference, with only two single-nucleotide polymorphisms (SNPs) across 777 bp of alignable control region sequence, and five SNPs and three single–base pair indels across 15,565 bp lying outside the control region (99.9% identity), with all differences supported by >50× DoC in the new HTS assembly.

The mitochondrial genome allows more refined evidence of the provenance of the sample. Haddrath (*26*) noted that the sample used here was "uncovered from a cave site on the South Island of New Zealand" (p. 27) without further detail. A 30-bp hypervariable control region "snippet" diagnostic for moa lineages (*27*) confirms the taxonomic assignment of the sequenced specimen, and a longer (382 bp) segment spanning this region is identical to a haplotype from little bush moa sampled at multiple sites across the South Island of New Zealand (fig. S3) (*1*). We combined the full-length CR1 sequence from the sample with 263 additional CR1 samples from Bunce *et al.* (*1*) and conducted a maximum likelihood phylogenetic analysis using default settings of IQ-TREE (fig. S4) (*28*). The moa sample used here appeared identical to samples from Tākaka on the north coast of the South Island, and Punakaiki, also from the northern South Island. Hopefully, this more detailed determination of the locality of the sample will aid efforts to identify the appropriate cultural context for the assembly (Supplementary Text).

Iterative mapping to a high-quality emu reference (*11*) recovered almost 900 Mb of the little bush moa nuclear genome (GenBank accession no. GCA_006937325.1 for the original assembly and

GCA_006938045.1 for the mapDamage corrected version; see also Aotearoa Genomic Data Repository project ID TAONGA-AGDR00049) or approximately 75% of the 1.2-Gb emu reference in 1844 moa scaffolds (Fig. 1C and Table 1). Average DoC across called bases was 3.9× for the original, uncorrected assembly, with 87% of bases having DoC ≥2 (corresponding values for the mapDamage-corrected assembly are 4.0× average DoC, with 87% of called bases having DoC ≥2; Fig. 1C). Although a large number of moa contigs are relatively short (maximum contig length = 12.2 kb for the original assembly and 12.0 kb for mapDamage-corrected; Table 1), and the fraction of ambiguous bases (Ns) for moa scaffolds averages ~26% (fig. S5), the average break between contigs is small (mean contig break = 218 bp for the original assembly and 226 bp for mapDamage-corrected; Table 1). As a result, more than 85% of BUSCO single-copy orthologs for birds were identified in moa, with more than 72% of BUSCOs represented by complete coding sequence (Table 1). The use of the mapDamage-corrected nuclear assembly for retrieval of loci for phylogenetic analysis made little difference to the phylogenetic conclusions of paleognath species of Sackton *et al*. (*11*) (Supplementary Text).

Genome size and GC content

The nuclear genome size of any moa (Dinornithiformes) is currently unknown. By measuring osteoblast cell volume and extrapolating from a Bayesian phylogenetic regression of nuclear genome size on cell volume among extant tetrapods, Organ *et al*. (*29*) estimated genome size of an unknown species of dinornithiform at 1.7 ± 0.0024 Gb, which we now know is considerably larger than the genome sizes of other paleognaths as measured by Fuelgen densitometry, biochemical methods, or their assembly sizes (*11*, *30*, *31*). We cannot use the moa assembly as scaffolded on a reference emu assembly as a proxy for genome size, nor is the unscaffolded moa assembly useful here because it is incomplete and has low coverage. Even if we had high-coverage sequence data for moa, assembly sizes are often smaller than genome sizes determined by other means (*32*). We therefore used a *k*-mer counting method tailored to low-coverage genomes, RESPECT, to estimate genome size (*33*). RESPECT conducts an optimization search for the repeat spectrum vector whose elements are assumed to fall within a set of constrained conditions. We applied RESPECT to raw moa sequencing reads after stringent filtering of potentially contaminating reads (*34*). A variety of filtering methods and proportion of mapped reads yielded a range of estimated genome sizes of 1.07 to 1.12 Gb, suggesting that this moa genome may be on the small side among avian genomes (fig. S6). Analysis of downsampled sequence reads from five birds and a crocodilian suggests that RESPECT appears to estimate assembly size well (slope = 1.049), although the correlation between RESPECT estimates of genome size and assembly size was not significant (fig. S7). We compared these estimates of moa genome size to assembly sizes of 11 species of the sister group of moas, the tinamous (Tinamidae), as a rough proxy. We found that genome size in tinamous varied from 0.958 Gb in *Eudromia* to 1.195 in *Crypturellus*, with evidence for some effect of the sequencing method. Still, tinamou genomes, particularly those in *Nothoprocta* and *Eudromia*, seem particularly small, on par with those of hummingbirds (*35*). Genome size in paleognaths generally seems to covary strongly with phylogeny, and our estimate of genome size in *Anomalopteryx* is consistent with the small genome size found in some tinamous.

We calculated a GC content of the total assembly of *Anomalopteryx*, corrected for ambiguous base calls, of 0.4277 (the value of 0.32 reported on NCBI is not corrected for Ns). This value falls broadly within the variation of GC content across paleognaths and is slightly higher than the global GC content of tinamous (0.420 to 0.422), despite some effect of the sequencing method on estimated GC content (fig. S8). As expected for Illumina data (*36*, *37*), short scaffolds had a higher GC content than longer scaffolds (fig. S9). To investigate details of the GC landscape across long scaffolds and facilitate comparisons between species, we first mapped moa scaffolds to a close relative of the moa with a good assembly, the Chilean Tinamou (*Nothoprocta perdicaria*) (fig. S10). We then developed an algorithm within the suite of methods within the MAFFT alignment package (*38*) to align long (>10 Mb) scaffolds with base-pair precision (Supplementary Text). We found that the GC landscape of unambiguous regions of moa scaffolds closely mirrored that of scaffolds of the Chilean tinamou (fig. S11). Despite removing ambiguous bases in these comparisons, sliding windows of moa GC content are systematically higher than those of comparable regions of the *N. perdicaria* genome, suggesting

Table 1. Assembly statistics for emu reference (GenBank accession no. GCA_006938045.1) and the reference-based little bush moa nuclear genomes (Aotearoa Genomic Data Repository, Project ID TAONGA-AGDR00049; and GenBank).

	Emu	Little bush moa	
		Original assembly	MapDamage-corrected
Number of scaffolds	2,777	1,909	1,843
Total scaffold length (bp, gapped)	1,192,237,364	1,190,759,065	1,189,933,016
Total ACGT bases (bp)	1,179,039,690	889,719,452	868,145,842
Number of contigs*	18,796	1,335,633	1,368,477
Contig N50 (bp)	187,464	1,127	1,068
Mean contig size (bp)	62,805	673	643
Average break between contigs (bp)	733	218	226
Total BUSCOs	4,797/4,915 (97.6%)	4,299/4,915 (87.5%)	4,190/4,915 (85.2%)
Complete BUSCOs	4,629/4,915 (94.2%)	3,701/4,915 (75.3%)	3,550/4,915 (72.2%)

*Contig breaks are defined by strings of ≥25 consecutive Ns.

the possibility of differences in nucleotide substitution process between the two lineages, which are implicit given the strong differences in nuclear DNA substitution rate between the two lineages (*10*, *11*, *39*). As expected, GC content also increased in scaffolds composed of a higher proportion of exons (fig. S12).

Heterozygosity and effective population size

We considered quantifying the demographic history of moa by analyzing the divergence between the two moa haplotypes via the PSMC technique (*40*), but the quality of the assembly was not high enough to have confidence in that approach. We therefore calculated nucleotide diversity (π) in 10-kb windows across moa scaffolds by mapping the filtered, mapDamage-corrected moa reads back to the moa assembly. We found that π per window (on scaffolds at least 100 kb and corrected for coverage by at least 2500 bases) ranged from 0.0001 to 0.024, with a mean of 0.00102 for autosomes. Calculating as above across whole scaffolds (rather than windows) yields a mean π of 0.00097 [95% confidence interval (CI): 0.00096 to 0.00098] (fig. S13). Under neutrality, we expect $\pi = 4N_e m$, where N_e is the long-term effective population size and m is the mutation/substitution rate. Consistent with estimates of female N_e based on mtDNA (*1*, *20*), we estimate that the moa lineage experienced a substitution rate for nuclear introns of 1.0245×10^{-08} (95% CI: 1.011796×10^{-08} to 1.037205×10^{-08}) substitutions per site per year (Supplementary Text and fig. S14). With a generation time of ~10 years for some moa species (*2*), we therefore estimate an effective population size of 237,555 (95% CI: 231,897 to 243,355).

We mapped moa scaffolds to a long-read female assembly of an emu (table S2) (*19*) and identified 69 scaffolds totaling 81,993,503 bp (52,587,049 non-*N* bases) mapping to the emu Z (82,723,677 bp). Median SNP density and π per site were significantly lower for the Z chromosome than for autosomes (fig. S15); we attribute this trend to the expected lower N_e for Z chromosomes than for autosomes due to the effects of drift (*41*). Overall, the mean π for the Z (0.0015) was slightly higher than for autosomes (0.00097; $t = -2.3629$, df = 43.06, $P = 0.02272$, Cohen's $d = 1.26$). and we found no strong and systematic deficiency of SNPs on longer (>100 kb) putatively Z-linked scaffolds of the moa assembly compared to autosomal scaffolds (fig. S15), suggesting that there were no regions of the Z with very low or zero polymorphism and that the reads likely derived from a male bird.

We also searched for similarities between the sex-linked probe of Huynen *et al.* (*42*) in the moa assembly. This 676-bp sequence (GenBank accession no. AF308932) had three blast hits to three different moa scaffolds, indicating that the moa assembly contained sequences similar to those of male ratites and did not have the diagnostic ratite deletion indicating a female bird (fig. S16) (*42*). However, the scaffolds with sequences matching this sequence all mapped to either chromosome 1 or 2 of the emu. We also found that the marker developed by Huynen *et al.* (*42*) had 72 blast hits to different sites in the long-read emu assembly (*19*), including one site each on the Z and W chromosomes. These blast hits to the Z and W chromosomes covered 90 and 100% of the query sequence, respectively. It seems likely that the marker by Huynen *et al.* (*42*) is a moderately repetitive DNA that may not be well covered by the moa assembly but may be present in some unannotated repetitive sequences identified in our search for transposable elements (TEs; see below). Overall, we find this analysis broadly consistent with the evidence from coverage of individual chromosomes suggesting that the sample is a male bird.

Identification of polymorphic microsatellite markers

Microsatellites offer an appealing option for aDNA studies since these nuclear markers are often highly polymorphic, are spread throughout the genome, and are sufficiently short to allow amplification in degraded samples (*43*). However, wetlab approaches for microsatellite isolation are not amenable to degraded aDNA samples, and cross-species amplification of markers from extant taxa is often unsuccessful (*43*–*45*). HTS can circumvent these difficulties by identifying microsatellites directly from sequencing reads in the target species. This approach was used in moa where Allentoft *et al.* (*44*, *45*) developed six polymorphic microsatellites from GS FLX 454 pyrosequencing reads and demonstrated their utility for studies of moa kinship (*46*) and population demography (*2*).

We used a complementary approach to isolating polymorphic microsatellites from the little bush moa nuclear assembly. We identified 27,127 dinucleotide and 25,170 trinucleotide repeats, approximately half of which met our criteria for inclusion based on flanking sequence contiguity (retaining 14,902 dinucleotides and 13,951 trinucleotides). From these, we identified 40 microsatellites (28 dinucleotides and 12 trinucleotides; table S3) that are heterozygous in the sequenced individual and hence at least minimally polymorphic in the species as a whole. We offer the realigned BAM files for each locus and alignments to other ratites as a community resource for future studies (available in Dryad, an example of each data type is given in fig. S17).

Analysis of transposable elements

TEs and other repeated sequences form an important component of avian genomes, and with more sensitive methods for the detection of TEs, the estimated fraction of each avian genome composed of TEs may be increasing (*47*). Previous estimates of paleognath genome sizes and repeat content have indicated that paleognath genomes are larger than those of neognaths, by up to 27% in some estimates (*48*). Yet, estimates of the fraction of paleognath genomes composed of TEs remain unexpectedly low, usually less than 10% (*30*, *49*, *50*). To quantify the TE landscape of moa and relatives, we used a sensitive pipeline for TE quantification involving the production of an annotated clade-specific repeat library augmented by deep learning approaches for the annotation of TEs (*51*, *52*). We then used this library to estimate TE distribution and abundance with both reference-based and reference-free approaches.

Using reference-based approaches such as RepeatModeler2 and RepeatMasker (v. 4.1.4) (*52*), our estimates of TE content for non-moa paleognaths are typically about double previous estimates from short-read genomes and, where available, consistent with recent estimates from long-read genomes (table S4) (*48*, *53*), with annotated TE fractions ranging from 5.6% in the elegant crested tinamou to 7.99% in Okarito kiwi (table S4). For long-read genomes, we annotate 14.46% of the chicken genome and 8.01% for emu, as TEs. As expected with this pipeline, estimates of total TE abundance based on the little bush moa assembly (3.3%) were somewhat lower and likely biased downward due to the fragmented assembly. Using unassembled sequence read–based approaches such as dnaPipeTE (*54*), which is not sensitive to the quality of the reference genome, our estimates for TE fractions in paleognaths range from 7.06% in the elegant crested tinamou to 15.51% in emu (Fig. 2). At 5.5% for annotated repeats, read-based estimates of TE abundances from the little bush moa genome are lower than those from emu and kiwi and similar to the tinamou. The low TE abundances in moa estimated by the read-based approach are

less likely to be an artifact of the DNA source and quality and instead may be part of a lineage-specific trend toward more simplified TE landscapes in the tinamou-moa clade (Fig. 3 and fig. S15). Paleognath genomes appear particularly rich in DNA elements, and the abundance of retrotransposons with long terminal repeats (LTRs), such as CR1, varies considerably across the clade.

The moa assembly was distinctive among paleognath genomes in appearing to have a high fraction of unannotated repeats (fig. S18). The genomic percentage of unannotated repeats using dnaPipeTE was as high as 18.55%, even when using a custom annotated TE library made from moa and tinamou. However, such estimates can be sensitive to genome-wide heterozygosity as well as uneven read depths across the target genome (*47*, *54*). dnaPipeTE constructed more than 70,000 contigs of unannotated, putatively repetitive sequence reads in the little bush moa genome, all but 200 of which mapped uniquely or nearly so to the moa genome and to other paleognath, neognath, and crocodilian genomes (fig. S19). We, therefore, conclude that most of these putative unannotated repeats represent unique sequences that were sequenced to high depth from the moa genome. However, we found five motifs common to clusters of repeated moa sequences that mapped to the long-read emu genome thousands of times, comprising up to 37.3 kb (fig. S19). These motifs—likely satellite DNAs and intersecting to varying degrees with annotated repeats in emu—were also found in large numbers in the kiwi genome but, unexpectedly, four of them were completely absent from the genome of the elegant crested tinamou, despite the sister relationship of moas and tinamous (fig. S19). We explored the phylogenetic diversity of additional repeats unannotated in moa and found these repeats in varying abundances in other paleognath genomes, often in proportion to reference genome quality and phylogenetic proximity to moa. Examples of phylogenetically widespread repeats vary in abundance across paleognaths and exhibit diverse patterns of diversification throughout the clade (fig. S19). Together, these results suggest that analysis of unassembled sequence reads identified diverse classes of repetitive elements in the little bush moa genome, some of which contain previously unknown but phylogenetically widespread motifs.

Genes informing moa sensory biology

To better understand moa sensory biology, we investigated olfactory receptor (OR) genes, bitter and umami taste receptor genes, and visual opsins in the moa genome. On the basis of the branching pattern of olfactory nerves, Johnston and Mitchell (*55*) concluded that the upland moa (*Megalapteryx didinus*) and other moa were olfactory specialists; despite the small olfactory bulb that has led most authors to suggest poor olfactory capabilities (*56*), they point out that some moa have a large olfactory chamber, and the presence of bill pits suggest high sensitivity of the bill tip, potentially in combination with olfaction while foraging. Others, however, have asserted that the olfactory bulb is not small but instead has been buried in an unprecedentedly large expansion of the cerebellum rostrally (*3*). We detected a total of 43 partial or complete OR genes in the moa genome. To place the moa OR genes in a broader phylogenetic context, we identified 1767 OR genes from the high-quality genomes of saltwater crocodile, emu, little spotted kiwi, elegant crested tinamou, and chicken (Fig. 3). To these, we also aligned 23 "index" OR genes from Vandewege *et al.* (*57*), 13 from crocodilians, and 10 from chicken that helped identify the lineages of specific moa ORs. Together, these index OR genes encompassed 13 different OR lineages as defined by Vandewege *et al.* (*57*), including OR51 and OR52. Although the number of OR genes in moa was small and likely biased downward due to poor assembly quality, we were able to find OR genes closely related to all 13 index OR lineages, suggesting that moa have a complete repertoire of reptilian and avian OR genes.

Vision is one of the most important senses for birds, yet moa had relatively small eyes, small optic lobes in the brain, and a limited visual field (*55*). Previous morphological and anatomical analyses suggest that moa had a visual acuity on par with extant diurnal ratites (*56*). We found complete coverage of rhodopsin and evidence for three of the four major cone opsins found in extant birds, including middle-wavelength sensitive (*MWS*) and short-wavelength sensitive (*SWS1*–violet sensitivity and *SWS2*–blue sensitivity; Fig. 4, A to C, and fig. S21). All opsins except *LWS* are recovered from the moa assembly, with at least 50% of the sequence present relative to the orthologous emu sequence. A cysteine at residue 90 is sufficient to shift the tuning of the *SWS1* visual pigment from a violet-sensitive (VS) to an ultraviolet-sensitive (UVS) state in many species (*58*). We found clear evidence of cysteine in this position (Fig. 4B), consistent with previous work on other moa species (*59*).

Little is known of the sense of taste in paleognaths, including moa, although the presence of taste buds on the tongue has been confirmed in several extant ratites (*60*, *61*). We found that the little

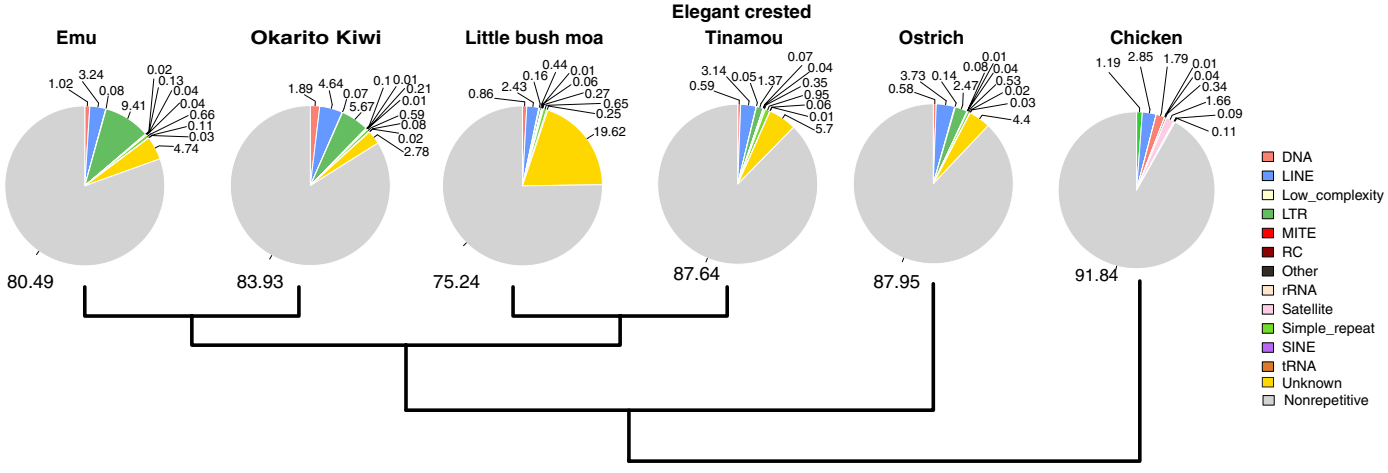

Fig. 2. Read-based quantification of the TE landscape in moa and other birds. Percentages of TEs were quantified using dnaPipeTE (*54*).

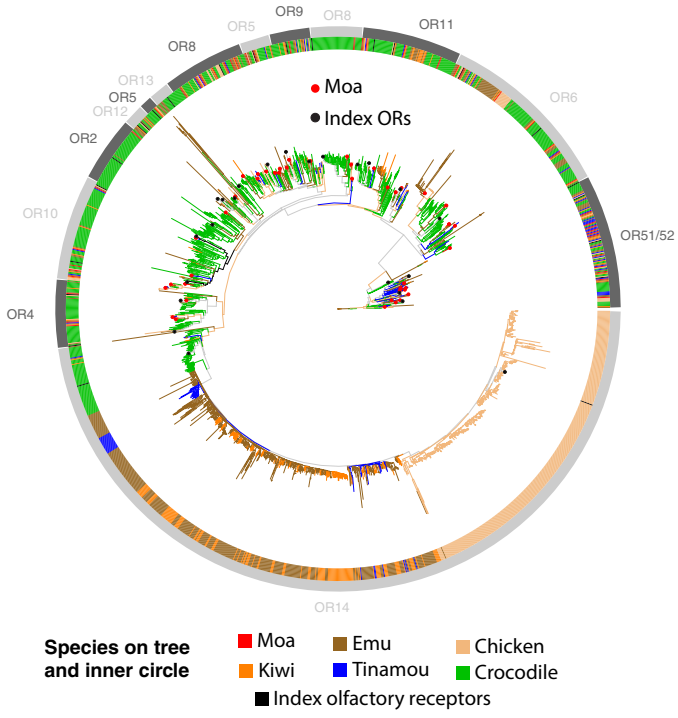

Fig. 3. Evolution of moa olfactory receptor genes. Using sequence from 14 protein queries from Vandewege *et al.* (*57*), we retrieved 1833 moa OR genes, indicated with a red ball. The 14 OR index sequences are indicated with a black ball. Branches of the tree are colored by species as indicated at the bottom and along the inner ring. Approximate phylogenetic limits of each index OR clade are indicated in the outer circle in gray tones. Further details are found in Supplementary Text.

bush moa possessed a member of each of the three well-supported avian clades of bitter taste receptors (*T2Rs*), suggesting a sensitivity of bitter taste equivalent to that of other birds (Fig. 4, D and E), although one *T2R* (part of the clade including chicken *T2R40*) may be a pseudogene. In each clade, the moa sequence was phylogenetically close to sequences from other ratites, as expected. Comparison with 13 additional bird *T2R* sequences suggests a bout of positive selection on the branch leading to one of the neoavian *T2R* clades, including moa, orthologous to the chicken the *T2R40L* gene.

The umami taste receptor consists of a heterodimer (T1R1 and T1R3) that has undergone marked evolution at the amino acid level to enable new sensitivities in nectarivorous birds, including hummingbirds and oscine passerine birds (*62*, *63*). We found evidence for exons 2 to 5 from *T1R1*, but exons 1 and 6 were missing (Fig. 4F). No exons from *T1R3* were recovered, possibly due to assembly gaps: genes expected to flank *T1R3*, including *CPTP* and *DVL1* were found in the expected position and orientation on the same scaffold, but large gaps were present between these two loci.

Genes of the major histocompatibility complex and Toll-like receptors

A recent report detected major histocompatibility complex (*MHC*) class I and II genes in the moa genome (*64*) but used the version of the genome uncorrected for base substitutions incurred in aDNA and did not provide a phylogenetic context for moa immunodiversity. We reassessed the presence of *MHC* class I and II genes in the moa genome with a tblastn approach using chicken class I and II protein exons as probes. Overall, we detected fragments of two *MHC* class I genes and three *MHC* class II genes (Fig. 5). We found evidence for three copies of class I exon 3, which partially encodes the peptide-binding region (PBR), in moa, each on a separate scaffold. We also found fragments of three copies of the conserved exon 3 of *MHC* class II (*BLB1* in chicken) on moa scaffold SZQD01000727.1 (scaffold 726), located approximately 25.5 and 2.25 kb away from each other. One of these copies had evidence for a near full-length copy of class II exon 2, encoding the PBR.

Alignments of both exon 3 of moa class I and exon 2 of moa class II to other paleognaths showed that moa sequences had several unique substitutions, although it was unclear whether any of the retrieved sequences were functional. Analysis of adaptive evolution by a Bayesian method (*65*) revealed five (class I) and two (class II) codons experiencing positive selection across alignments that included a moa sequence (fig. S22).

We searched for Toll-like receptors (TLRs) in the moa assembly using blast. Out of 13 *TLR* query exons or full-length coding regions comprising nine genes, we found strong evidence for all nine *TLR* genes as judged by low *e* values (table S5). Overall, these results suggest that the moa assembly contains the majority if not all of TLRs thus far known in birds (*66*–*68*).

Tests of selection in candidate genes for limb development

Arguably, the most remarkable moa trait is the complete absence of forelimb skeletal elements comprising the wings. All ratites exhibit some degree of forelimb reduction; however, moa are unique in retaining only a fused scapulocoracoid within the pectoral girdle (*3*, *69*). Huynen *et al.* (*69*) recovered the moa coding sequence for the T-box transcription factor (*TBX5*), which plays a key role in forelimb specification and outgrowth (*70*–*72*), and demonstrated that the moa *TBX5* sequence activates promoters of downstream genes in developing chicken embryos. Therefore, alterations to this coding region alone appear unlikely to underlie the wingless moa phenotype (*69*). We build upon this work by reporting moa sequence for a more comprehensive suite of candidate genes with established involvement in vertebrate limb development (Table 2), as well as candidates with putative function-altering variants in the flightless cormorant (*Phalacrocorax harrisi*) hypothesized to accompany forelimb reduction in this flightless species (Table 2) (*73*).

We recovered moa sequence for all investigated genes, with an average of 88% of coding sequence per gene recovered from the original moa assembly (Table 2; 87% in the mapDamage-corrected version, table S3). We found no frameshift mutations and only a single in-frame stop codon in *HOXD4* which, however, occurred at 1× coverage in both assembly versions and could represent a sequence artifact (this codon was intentionally masked by Ns for further tests). There was no evidence for lineage-specific diversifying selection in moa, with $P > 0.05$ in aBSREL tests for each gene (*74*). RELAX tests also found no significant difference in the strength of selection in moa relative to other ratites for candidate genes with established roles in limb development (Table 2 and table S6). However, RELAX did identify a significant intensification of selection in the *FAT1* gene in moa and relaxation in *GLI2* relative to other ratites among the 11 candidates originating from the study of the flightless cormorant (Table 2 and table S6). Neither of these results remains significant under a more conservative genome-wide correction for an estimated 16,255 genes in birds as opposed to correcting only for the set of 37 candidates tested here (both $P > 0.05$).

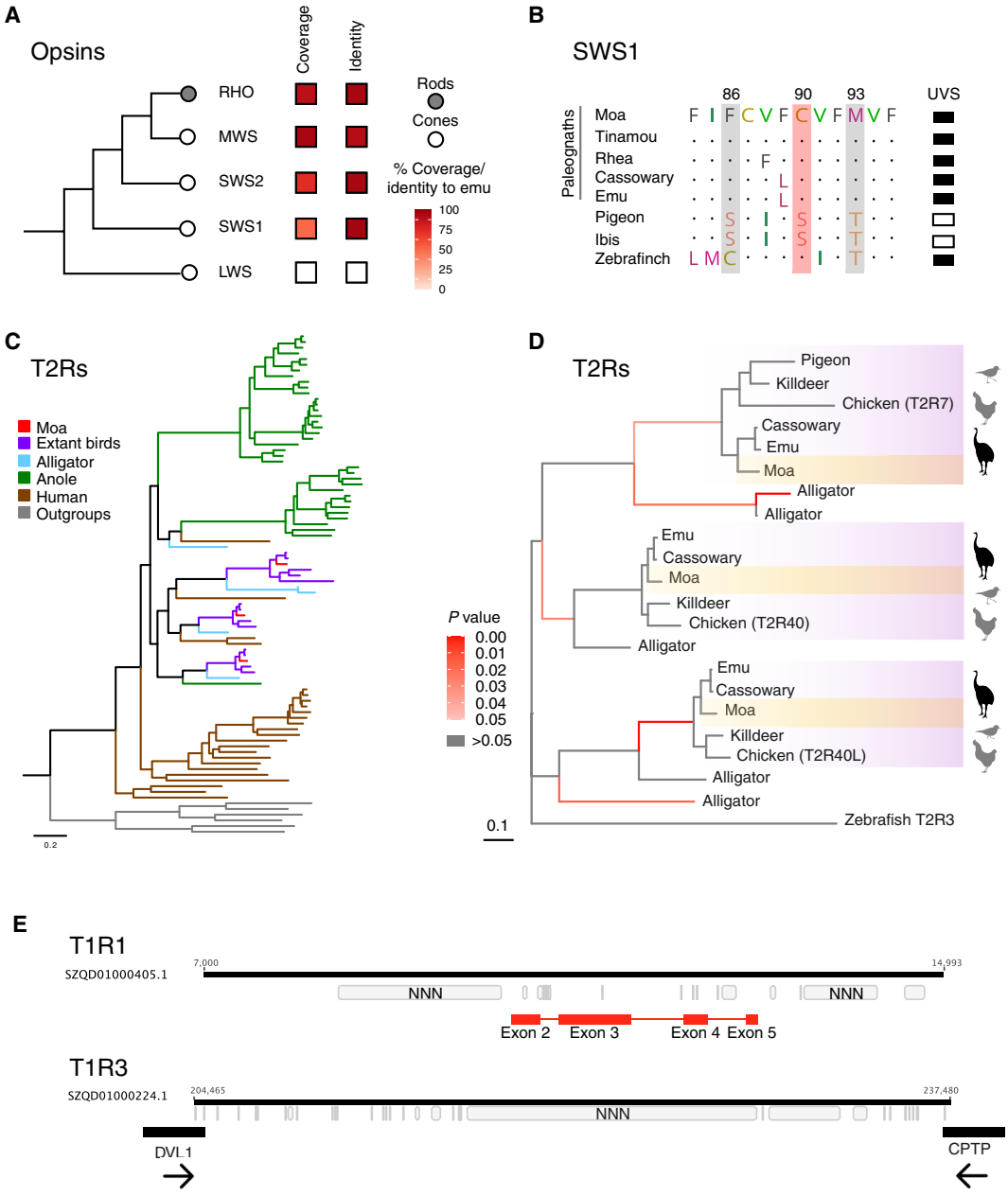

Fig. 4. Overview of visual opsins present in the moa. (**A**) Cone-expressed opsins encoding four visual pigments [long-wavelength sensitive, LWS; short-wavelength sensitive (SWS1, SWS2); middle-wavelength sensitive (MWS), as well as the rod-expressed rhodopsin (RHO)] were examined. The percentage of coverage and identity relative to emu opsins is depicted: All opsins except *LWS* are recovered from the moa assembly, with at least 50% of the sequence present relative to the orthologous emu sequence. See fig. S21 for phylogenetic analysis and selection tests. (**B**). Alignment of tuning sites (gray and red shading) responsible for shifts in the optimal wavelength absorption of SWS1. A cysteine at residue 90, shaded red, is sufficient to shift the tuning of the SWS1 pigment from VS (violet-sensitive) to UVS (ultraviolet-sensitive) state in many species (*58*); alignment numbers correspond to positions in bovine rhodopsin; predicted or tested UVS sensitivity depicted (black boxes). (**C**) Maximum likelihood tree of bitter taste receptors (*T2Rs*). The little bush moa has a member of each of the three well-supported avian clades (moa, red; extant birds, purple; Chinese alligator, blue; anole, green, human, brown; see fig. S21). Anole and human receptors are from Dong *et al*. (*99*). (**D**) Patterns of positive selection within bird *T2Rs* (red, branches with $P < 0.05$; uncorrected *P* value, aBSREL branch models (*74*); avian clades shaded purple and moa *T2Rs* are shaded red. (**E**). Exons 2 to 5 from *T1R1* (top) were recovered from the moa assembly (exons 1 and 6 were missing); shaded gray rectangles represent gaps in the assembly. No exons from *T1R3* were recovered, likely due to assembly gaps; flanking loci *CPTP* and *DVL1* (which flank *T1R3* in chicken) are found on the same scaffold. Silhouettes in (D) are from phylopic.org.

RELAX tests also identified seven candidates with significant selective shifts in flightless lineages relative to other birds (Table 2; 10 candidates, including the previous 7, when the mapDamage-corrected sequence is used, table S3). Of these, we find evidence for intensified selection in *HOXA2*, *HOXA4*, and *HOXD4* (additionally, *HOXA10*, *SHH*, and *WNT2B* using the corrected moa sequence), and relaxation in four genes (*GLI3*, *EVC*, *FAT1*, and *TALPID3*; note that *FAT1* shows not only intensified selection in moa relative to other ratites but also relaxed selection in flightless birds generally). However, only the intensification for *HOXA2* remains significant under the more stringent genome-wide false discovery rate correction ($P = 0.021$ for both datasets).

PROVEAN analysis (*75*) identified 24 moa sequence variants of possible functional relevance compared to the emu reference (table S7). Identified sites were identical for both moa assemblies, except *DVL1* where only the alternate moa allele with codon AAC was recovered after base quality recalibration. However, half of these variants (12 of 24) are either shared with other species or are polymorphic in moa, with the emu residue present as an alternative moa allele, indicating that this subset of sites is unlikely to underlie the wingless moa phenotype. In addition, 16 of the 24 sites display alternative residues in other birds that are often accompanied by PROVEAN scores comparable to moa (table S7). Comparison to an inferred reconstruction of the common ancestor of moa and tinamous yielded broadly similar results, with 17 of 19 potentially functionally relevant moa variants identical to those identified from comparison to the emu reference (table S8).

Putative function-altering variants in the flightless cormorant are not shared with other flightless lineages, indicating that any commonality in the genetic basis for independent losses of flight involving these genes is likely not attributable to convergent or parallel amino acid changes (table S9). Burga *et al.* (*73*) also identified a deletion in *CUX1* of the flightless cormorant, with experimental assays indicating this gene acts as a transcriptional activator of targets *FAT1* and *OFD1*. As with the other reported flightless cormorant variants, the *CUX1* deletion is not shared by moa or other ratites (identical sequence occurs in both moa assemblies; fig. S23). Together, we conclude that the loss of wings in moa is not attributable to gene loss or pseudogenization within this candidate gene set, although the functional relevance of variants unique to moa requires further experimental work.

DISCUSSION

Here, we have described an assembly of a nuclear genome from an extinct species of moa, the little bush moa (*Anomalopteryx didiformes*). This genomic resource has already proven useful for assembling genome-wide datasets of nuclear markers for phylogenetic inference (*7*, *8*) and for analyses of mitogenomes (*25*) and specific nuclear genes. Here, we have used a variety of additional ratite genomes, including a long-read genome of an emu (*19*), to elucidate diverse aspects of moa biology. Some of our conclusions and ability to detect specific genes are likely contingent on the quality of the short-read reference genome to which we mapped the moa reads for this particular assembly; mapping moa reads to long-read paleognath genomes (*19*) could reveal additional genes and assembly contiguity. Nonetheless, the current mapDamage-corrected moa assembly contains an unexpectedly high proportion of genes and other markers for analysis. We further provide estimates of the nuclear genome size and sex of the sample and explore moa sensory biology through its genome.

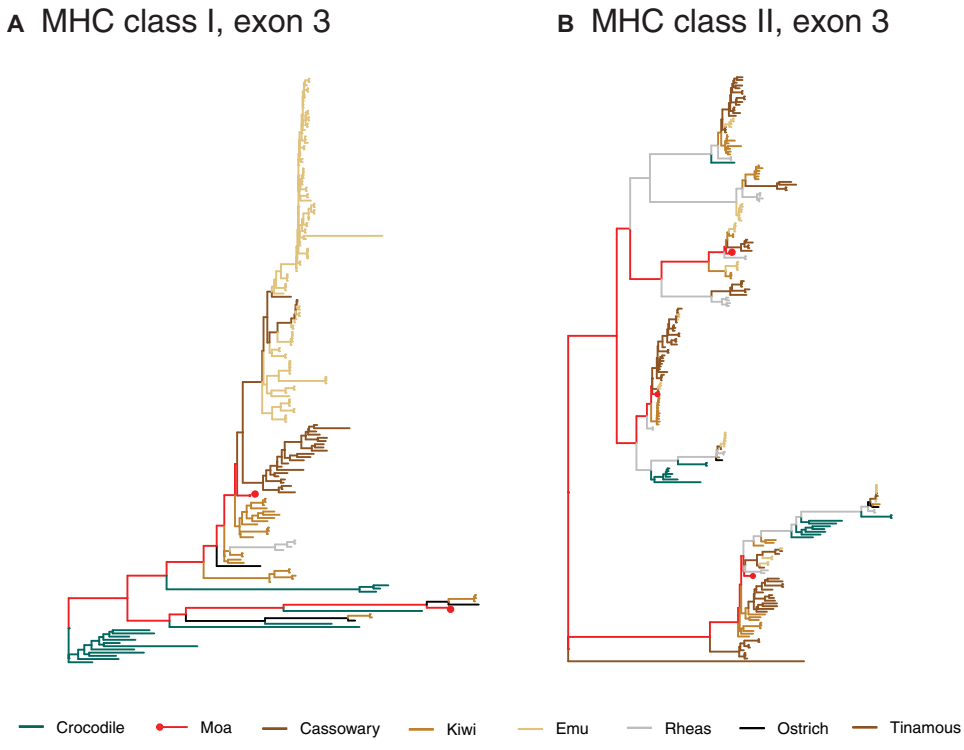

Fig. 5. Evolution of major histocompatibility complex (MHC) class I and class II genes of the little bush moa. Trees of (**A**) exon 3 of class I peptide-binding region (PBR) and (**B**) exon 2 of class II PBR were made using maximum likelihood and optimal nucleotide substitution models as described in Supplementary Text.

Table 2. Tests of selection for candidate limb development genes using moa sequence from the original genome assembly. *K*, relaxation parameter (values <1 indicate relaxed selection on foreground branches, and values >1 denote intensified selection); P_{adj}, adjusted *P* value (*Q* value) controlling for the false discovery rate at a significance level of 0.05 based on $N = 37$ genes tested. CDS, coding sequence. AA, amino acid.

Gene	Description	CDS length (AA, % of total)			RELAX tests			
					Moa		All flightless	
		Chicken	Emu	Moa	*K*	P_{adj}	*K*	P_{adj}
Candidate limb development genes								
FGF8	Fibroblast growth factor 8	214	214	189 (88%)	3.912	0.991	49.994	0.081
FGF10	Fibroblast growth factor 10	212	212	204 (96%)	0.494	0.419	1.052	0.542
GLI3	GLI family zinc finger 3	1576	1575	1570 (99%)	1.491	0.389	0.359	0.006
HOXA1	Homeobox A1	320	319	308 (96%)	0.999	0.991	0.931	0.459
HOXA2	Homeobox A2	375	374	357 (95%)	0.178	0.074	3.424	< 0.001
HOXA3	Homeobox A3	413	413	414 (100%)	1.491	0.389	1.033	0.525
HOXA4	Homeobox A4	309	145*	155 (50%)	0.365	0.389	2.564	0.006
HOXA5	Homeobox A5	270	270	251 (93%)	29.348	0.289	0.633	0.215
HOXA6	Homeobox A6	231	231	231 (100%)	0.828	0.808	0.272	0.365
HOXA7	Homeobox A7	219	219	219 (100%)	28.556	0.389	1.156	0.425
HOXA9	Homeobox A9	260	261	249 (95%)	1.985	0.389	2.040	0.084
HOXA10	Homeobox A10	364	317*	289 (79%)	1.239	0.934	2.914	0.062
HOXA11	Homeobox A11	297	297	255 (86%)	0.303	0.389	1.453	0.110
HOXA13	Homeobox A13	290	290	269 (93%)	1.179	0.934	0.858	0.459
HOXD3	Homeobox D3	413	248*	247 (60%)	1.211	0.808	1.326	0.425
HOXD4	Homeobox D4	237	237	202 (85%)	1.125	0.966	9.667	0.006
HOXD8	Homeobox D8	268	147*	146 (54%)	0.607	0.497	< 0.001	0.379
HOXD9	Homeobox D9	302	299	283 (94%)	0.295	0.339	0.816	0.217
HOXD10	Homeobox D10	339	339	339 (100%)	0.923	0.991	49.998	0.152
HOXD11	Homeobox D11	280	282	272 (96%)	0.967	0.976	0.642	0.081
HOXD12	Homeobox D12	266	266	266 (100%)	1.452	0.397	1.006	0.569
HOXD13	Homeobox D13	301	82*	74 (25%)	0.926	0.991	0.939	0.525
SALL4	Spalt-like transcription factor 4	1108	1111	1023 (92%)	0.793	0.389	0.860	0.146

(Continued)

(Continued)

Gene	Description	CDS length (AA, % of total)			RELAX tests			
					Moa		All flightless	
		Chicken	Emu	Moa	*K*	P_{adj}	*K*	P_{adj}
SHH	Sonic hedgehog	425	422	398 (94%)	1.979	0.389	0.974	0.542
TBX5	T-box 5	521	538	426 (79%)	0.240	0.389	1.021	0.545
WNT2B	Wnt family member 2B	385	330*	260 (68%)	0.636	0.339	1.703	0.081
				Candidate genes from the flightless cormorant				
DCHS1	Dachsous cadherin-related 1	3266	3267	3086 (94%)	1.094	0.389	1.006	0.554
DVL1	Disheveled segment polarity protein 1	712	655*	633 (89%)	1.326	0.397	0.814	0.127
DYNC2H1	Dynein cytoplasmic 2 heavy chain 1	4301	4295	3991 (93%)	1.152	0.389	0.763	0.089
EVC	EvC ciliary complex subunit 1	984	927*	871 (89%)	0.169	0.389	0.654	0.023
FAT1	FAT atypical cadherin 1	4645	4644	4489 (97%)	1.493	0.013	0.880	0.024
GLI2	GLI family zinc finger 2	1528	1528	1528 (100%)	0.408	0.013	0.927	0.262
IFT122	Intraflagellar transport 122	1245	1239	1202 (97%)	0.035	0.808	1.311	0.146
KIF7	Kinesin family member 7	1412	1279*	1226 (87%)	0.683	0.389	1.204	0.146
OFD1	OFD1, centriole, and centriolar satellite protein	1012	1014	1002 (99%)	0.978	0.991	0.006	0.102
TALPID3	KIAA0586	1523	1527	1474 (97%)	0.748	0.389	0.035	0.023
WDR34	WD repeat domain 34	500	502	459 (91%)	18.909	0.251	0.776	0.102

*Partial CDS recovered in the emu reference sequence.

We found approximately equal levels of diversity on the Z chromosome and autosomes. Our estimates of π for the autosomes and Z chromosome suggest a ratio of nearly one for median π (π_Z/π_A = 0.00097/0.00095 = 0.979) and mean π is slightly higher for the Z than for the autosomes (π_Z/π_A = 0.00150/0.00098 = 1.53). The slightly higher mean π on the Z is consistent with the effects of reverse sexual dimorphism, but a broader sampling of autosomes and sex chromosomes is necessary for further insight. Given that female-driven sexual selection can drive Z chromosome diversity down (*76*–*78*), it is tempting to speculate that the reverse sexual dimorphism found in some moa species (*79*, *80*) may not drive diversity on the Z chromosome down. However, demographic effects can also alter the ratio of diversity on sex chromosomes and autosomes; in particular, population expansions can result in transiently higher diversity on the X (=Z) chromosome relative to autosomes (*81*). In general, a variety of selective and demographic effects can shift the expected ratio of Z/autosome diversity from the expected ¾ that strictly reflects differences in population size when the sex ratio is 1:1 (*76*, *77*).

Our estimate of moa effective population size is rough and does not incorporate all sources of variance, such as mutational variance, or accurate generation time, appropriately. With their smaller body size, little bush moa likely had a shorter generation time than other, larger-bodied moa. Our estimate of ~240,000 for *A. didiformes* falls in the middle of multiple estimates ranging from ~9200 from microsatellite variation to more than 1 million with mtDNA (*2*, *82*). Allentoft and Rawlence (*4*) discuss the challenges of estimating population sizes of extinct moa from ecological or genomic data and suggest that the highest estimates from genetic data are too high. Our estimate of ~240,000 for this one species is also larger than estimates based on morphological and ecological considerations, which estimate up to 100,000 birds total for the entire moa clade across all moa species (*83*). Our estimate of N_e for *A. didiformes* is uncertain due to uncertainties of mutation rate and is further compromised by our not having forced the divergence rate of introns through the origin, which could lead to an upwardly biased estimate of the mutation rate. Further study of the substitution rate along the moa lineage, which is known to be lower than in the sister tinamou lineage, as well as more detailed demographic estimates from nuclear DNA, are warranted (*10*, *11*, *39*).

The relatively complete array of OR genes suggests that the little bush moa was not deficient in olfactory capabilities, lending support to an anatomical interpretation of the endocranium favoring a normal-sized olfactory bulb in moa and other giant flightless birds (*84*, *85*). We also demonstrate its utility in isolating markers for population-level studies and in investigating sequence evolution in candidate protein-coding genes. The relative contributions of coding sequence variation and mutations in noncoding regulatory elements to phenotypic variation constitute an area of active research (*11*, *70*, *86*), and we anticipate that the availability of a moa nuclear genome will further contribute to experimental studies of regulatory changes associated with flightless phenotypes. Additional nuclear genome assemblies from extinct moa, as well as better assemblies of tinamous and other paleognaths, will no doubt further enable exploration of the genetic basis of phenotypic traits of these extraordinary birds.

MATERIALS AND METHODS

DNA extraction and sequencing

DNA was extracted from a single toe bone of a poorly provenanced moa specimen held in the collections of the Royal Ontario Museum (ROM; Toronto, Canada). Some of the HTS reads reported here were previously used for phylogenetic analysis of paleognath relationships using 1448 ultraconserved elements and protein-coding loci (*7*), and PCR-based sequences obtained from this same specimen have been reported by Haddrath and Baker (*9*), Baker *et al.* [(*12*), sample A. did. OH], and Haddrath and Baker [(*9*), sample TW95].

The sample analyzed here was also presented in (*87*); for transparency and clarification, we here present important details of extraction and analysis. DNA extraction followed by Baker *et al.* (*12*). Briefly, the outer 1 to 2 mm was removed from the bone surface by microblasting with an Airbrasive System (MicroBlaster; Comco, Burbank CA, USA), and 0.2 g of the remaining material was ground into fine powder. Enzymatic digestion proceeded overnight at 56°C in a buffer containing final concentrations of 0.5 M EDTA, proteinase K (200 μg/ml), and 0.5% *N*-lauroylsarcosine at pH 8.0 (*88*), and DNA was purified using commercially available silica spin columns (DNeasy Blood & Tissue Kit; Qiagen, Germantown, MD, USA). Sample preparation occurred in a dedicated aDNA workspace in the ROM following established best practices (*89*, *90*).

Library preparation and sequencing were performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. Library A_didi_CTTGTA was constructed from 200- to 400-bp size-selected DNA sheared to 200-bp insert size followed by library preparation with the Illumina TruSeq DNA v3 DNA Prep Kit. Paired-end sequencing (2 × 101 bp) was carried out on three lanes of a HiSeq 2500 platform using Illumina v3 chemistry. A second TruSeq library (A_didi_GCCAAT) was prepared from the same input DNA and sequenced on two partial lanes of a HiSeq 2500. Three additional libraries were constructed with the Illumina Nextera XT Sample Preparation Kit. The A_didi_CAGAGA and A_didi_CTCTCT libraries used input DNA < 500 bp with no additional shearing, while the A_didi_AGGCAG library used DNA 500 bp to 2 kb in size subsequently sheared to <700 bp. These Nextera libraries were pooled for sequencing on a single HiSeq 2500 lane.

Read processing and genome assembly

Trimmomatic v. 0.32 (*91*) was run in paired-end mode for adapter removal and quality trimming and reads with post-trimming length below 25 bp were discarded (options ILLUMINACLIP:[adapter_file]:2:30:10:1:true SLIDINGWINDOW:10:13 MINLEN:25). A de novo mitochondrial genome assembly was built with MITObim v. 1.8 (*92*) using the published little bush moa mtDNA genome as a starting seed [GenBank accession no. NC_002779, (*23*)]

To obtain a nuclear sequence assembly, we first mapped reads to a draft genome for emu [*D. novaehollandiae*; (*11*), GenBank accession no. GCA_003342905.1], and then remapped reads to the initial moa consensus for improved recovery of short and/or variant reads. First, to estimate an appropriate substitution parameter between the emu and moa, a random subset of reads was mapped to the emu reference with Stampy v. 1.0.28 (*93*) using default settings. The full data were then mapped to emu with Stampy and this user-specified substitution parameter (estimated at 0.0839). Samtools v. 1.3.1 (*94*) was used to filter reads with mapping quality score below 30, and duplicates within each library were marked and removed with Picard Tools v. 2.6.0 (https://broadinstitute.github.io/picard/) before merging mapped reads across libraries. Samtools 'mpileup' was used to output variant call format files with minimum mapping quality 30 and base quality 20, and a consensus sequence was called with BCFTools v. 1.2. Reads were remapped to this initial consensus with Bowtie2 v. 2.2.9 (*95*), with subsequent post-processing as above. In addition to the nuclear assembly detailed above (hereafter referred to as the "original assembly"), an error-corrected version of the genome was also generated to account for nucleotide misincorporations due to postmortem DNA damage characteristic of aDNA, described in Supplementary Text. Taxonomic read profiling is described in Supplementary Text.

Heterozygosity and effective population size

We mapped moa reads to the emu-scaffolded moa assembly and called variants using the Genome Analysis Toolkit (GATK) following standard practices as implemented in the snpArcher Snakemake pipeline (*96*). TEs were estimated from raw sequence reads using dnaPipeTE v.1.3.1 (*54*) as described in Supplemenatary Text.

Sensory receptor genes

Genes for ORs, opsins, and taste receptor types 1 and 2 (T1Rs and T2Rs) were obtained by blast (*97*) using diverse query sequences

from birds and crocodilians (see Supplementary Text for further details and tests for positive selection). OR genes from moa and other paleognaths were isolated by blast using 23 index OR sequences from Vandewege *et al.* (*57*). Further details are in Supplementary Text.

Tests of selection for candidate limb development genes

Multiple sequence alignments were compiled for 26 candidate genes with established roles in vertebrate limb development [listed in Table 2, reviewed in (*70*, *71*, *98*)] and for 11 genes with potential function-altering variants in the flightless cormorant (*P. harrisi*) hypothesized to accompany phenotypic modifications typical of flightless birds [listed in Table 2, reviewed in (*73*)]. Gene models were manually curated for 10 new draft genome assemblies for paleognaths (*11*). The moa coding sequence was obtained from pairwise whole-scaffold alignments of moa to emu using reference emu coordinates (alignments and an accessory Perl script are available in Dryad). Sequences from draft paleognath genomes were combined with available avian sequences from GenBank and flightless cormorant sequences from Burga *et al.* (*73*). Amino acid translations were aligned with MAFFT v. 7.245 (*38*). Partial (<70% of total alignment length) and poorly aligning (<60% mean pairwise amino acid identity) sequences were removed, and the resulting alignment was used to guide gap insertion in the corresponding nucleotide sequences. GenBank source information, curated gene models, and sequence alignments are available in Dryad. Tests of selection and functional effects of moa sequence variants in candidate limb development genes are described in Supplementary Text. Identification of polymorphic microsatellite repeats are described in Supplementary Text.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S23
Tables S1 to S9
References

REFERENCES AND NOTES

1. M. Bunce, T. H. Worthy, M. J. Phillips, R. N. Holdaway, E. Willerslev, J. Haile, B. Shapiro, R. P. Scofield, A. Drummond, P. J. J. Kamp, A. Cooper, The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20646–20651 (2009).
2. M. E. Allentoft, R. Heller, C. L. Oskam, E. D. Lorenzen, M. L. Hale, M. T. P. Gilbert, C. Jacomb, R. N. Holdaway, M. Bunce, Extinct New Zealand megafauna were not in decline before human colonization. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4922–4927 (2014).
3. M. Bunce, A. Cooper, The lost world of the moa: Prehistoric life of New Zealand. Life of the Past. By Trevor H Worthy and Richard N Holdaway; principal photography by Rod Morris. Bloomington (Indiana): Indiana University Press. $89.95. xxxv + 718 p. ill; index. ISBN: 0–253–34034–9. 2002. Quart. Rev. Biol. 78, 469 (2003).
4. M. E. Allentoft, N. J. Rawlence, Moa's Ark or volant ghosts of Gondwana? Insights from nineteen years of ancient DNA research on the extinct moa (Aves: Dinornithiformes) of New Zealand. *Anat. Anz.* **194**, 36–51 (2012).
5. A. Grealy, N. J. Rawlence, M. Bunce, Time to spread your wings: A review of the avian ancient DNA field. *Genes* **8**, 184 (2017).
6. A. Cooper, C. Mourer-Chauviré, G. K. Chambers, A. von Haeseler, A. C. Wilson, S. Pääbo, Independent origins of New Zealand moas and kiwis. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8741–8744 (1992).
7. 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).
8. A. Cloutier, T. B. Sackton, P. Grayson, M. Clamp, A. J. Baker, S. V. Edwards, Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. *Syst. Biol.* **68**, 937–955 (2019).
9. O. Haddrath, A. J. Baker, Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* **279**, 4617–4625 (2012).
10. M. J. Phillips, G. C. Gibb, E. A. Crimp, D. Penny, Tinamous and moa flock together: Mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst. Biol.* **59**, 90–107 (2009).
11. T. B. Sackton, P. Grayson, A. Cloutier, Z. Hu, J. S. Liu, N. E. Wheeler, P. P. Gardner, J. A. Clarke, A. J. Baker, M. Clamp, S. V. Edwards, Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
12. A. J. Baker, L. J. Huynen, O. Haddrath, C. D. Millar, D. M. Lambert, Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8257–8262 (2005).
13. L. Huynen, C. D. Millar, R. P. Scofield, D. M. Lambert, Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**, 175–178 (2003).
14. N. J. Rawlence, J. R. Wood, K. N. Armstrong, A. Cooper, DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* **276**, 3395–3402 (2009).
15. L. Huynen, B. J. Gill, C. D. Millar, D. M. Lambert, Ancient DNA reveals extreme egg morphology and nesting behavior in New Zealand's extinct moa. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16201–16206 (2010).
16. J. R. Wood, J. M. Wilmshurst, N. J. Rawlence, K. I. Bonner, T. H. Worthy, J. M. Kinsella, A. Cooper, A megafauna's microfauna: Gastrointestinal parasites of New Zealand's extinct moa (Aves: Dinornithiformes). *PLOS ONE* **8**, e57315 (2013).
17. J. R. Wood, J. M. Wilmshurst, S. J. Richardson, N. J. Rawlence, S. J. Wagstaff, T. H. Worthy, A. Cooper, Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes). *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16910–16915 (2013).
18. M. Hofreiter, J. L. A. Paijmans, H. Goodchild, C. F. Speller, A. Barlow, G. G. Fortes, J. A. Thomas, A. Ludwig, M. J. Collins, The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays* **37**, 284–293 (2015).
19. J. Liu, Z. Wang, J. Li, L. Xu, J. Liu, S. Feng, C. Guo, S. Chen, Z. Ren, J. Rao, K. Wei, Y. Chen, E. D. Jarvis, G. Zhang, Q. Zhou, A new emu genome illuminates the evolution of genome configuration and nuclear architecture of avian chromosomes. *Genome Res.* **31**, 497–511 (2021).
20. T. Yonezawa, T. Segawa, H. Mori, P. F. Campos, Y. Hongoh, H. Endo, A. Akiyoshi, N. Kohno, S. Nishida, J. Wu, H. Jin, J. Adachi, H. Kishino, K. Kurokawa, Y. Nogi, H. Tanabe, H. Mukoyama, K. Yoshida, A. Rasoamiaramanana, S. Yamagishi, Y. Hayashi, A. Yoshida, H. Koike, F. Akishinonomiya, E. Willerslev, M. Hasegawa, Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of the ratites. *Curr. Biol.* **27**, 68–77 (2017).
21. J. Dabney, M. Meyer, S. Pääbo, Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
22. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7**, e34131 (2012).
23. O. Haddrath, A. J. Baker, Complete mitochondrial DNA genome sequences of extinct birds: Ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. Roy. Soc. London, Ser. B, Biol. Sci.* **268**, 939–945 (2001).
24. A. Cooper, C. Lalueza-Fox, S. Anderson, A. Rambaut, J. Austin, R. Ward, Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704–707 (2001).
25. A. D. Urantówka, A. Kroczak, P. Mackiewicz, New view on the organization and evolution of Palaeognathae mitogenomes poses the question on the ancestral gene rearrangement in Aves. *BMC Genomics* **21**, 874 (2020).
26. O. Haddrath, "Ratite molecular evolution, phylogeny and biogeography inferred from complete mitochondrial genomes," thesis, University of Toronto, Toronto, Canada (2000).
27. J. McCallum, S. Hall, I. Lissone, J. Anderson, L. Huynen, D. M. Lambert, Highly informative ancient DNA 'snippets' for New Zealand moa. *PLOS ONE* **8**, e50732 (2013).
28. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
29. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007).
30. Z.-J. Wang, G.-J. Chen, G.-J. Zhang, Q. Zhou, Dynamic evolution of transposable elements, demographic history, and gene content of paleognathous birds. *Zool. Res.* **42**, 51–61 (2021).
31. E. D. Jarvis, S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, A. Suh, C. C. Weber, R. R. da Fonseca, J. W. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, M. S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldon, S. Capella-Gutierrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. Schierup, B. Lindow, W. C. Warren, D. Ray, R. E. Green, M. W. Bruford, X. J. Zhan, A. Dixon, S. B. Li, N. Li, Y. H. Huang, E. P. Derryberry, M. F. Bertelsen,

F. H. Sheldon, R. T. Brumfield, C. V. Mello, P. V. Lovell, M. Wirthlin, M. P. C. Schneider, F. Prosdocimi, J. A. Samaniego, A. M. V. Velazquez, A. Alfaro-Nunez, P. F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D. M. Lambert, Q. Zhou, P. Perelman, A. C. Driskell, B. Shapiro, Z. J. Xiong, Y. L. Zeng, S. P. Liu, Z. Y. Li, B. H. Liu, K. Wu, J. Xiao, X. Yinqi, Q. M. Zheng, Y. Zhang, H. M. Yang, J. Wang, L. Smeds, F. E. Rheindt, M. Braun, J. Fjeldsa, L. Orlando, F. K. Barker, K. A. Jonsson, W. Johnson, K. P. Koepfli, S. O'Brien, D. Haussler, O. A. Ryder, C. Rahbek, E. Willerslev, G. R. Graves, T. C. Glenn, J. McCormack, D. Burt, H. Ellegren, P. Alstrom, S. V. Edwards, A. Stamatakis, D. P. Mindell, J. Cracraft, E. L. Braun, T. Warnow, W. Jun, M. T. P. Gilbert, G. J. Zhang, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).

32. A. N. Black, K. J. Bondo, A. Mularo, A. Hernandez, Y. Yu, C. M. Stein, A. Gregory, K. A. Fricke, J. Prendergast, D. Sullins, D. Haukos, M. Whitson, B. Grisham, Z. Lowe, J. A. DeWoody, A highly contiguous and annotated genome assembly of the lesser prairie-chicken (*Tympanuchus pallidicinctus*). *Genome Biol. Evol.* **15**, evad043 (2023).
33. S. Sarmashghi, M. Balaban, E. Rachtman, B. Touri, S. Mirarab, V. Bafna, Estimating repeat spectra and genome length from low-coverage genome skims with RESPECT. *PLOS Comput. Biol.* **17**, e1009449 (2021).
34. E. Rachtman, V. Bafna, S. Mirarab, CONSULT: Accurate contamination removal using locality-sensitive hashing. *NAR Genom. Bioinform.* **3**, lqab071 (2021).
35. T. R. Gregory, C. B. Andrews, J. A. McGuire, C. C. Witt, The smallest avian genomes are found in hummingbirds. *Proc. R. Soc B Biol. Sci.* **276**, 3753–3757 (2009).
36. Y. C. Chen, T. Liu, C. H. Yu, T. Y. Chiang, C. C. Hwang, Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLOS ONE* **8**, e62856 (2013).
37. Y. Benjamini, T. P. Speed, Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72 (2012).
38. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
39. J. Harshman, E. L. Braun, M. J. Braun, C. J. Huddleston, R. C. K. Bowie, J. L. Chojnowski, S. J. Hackett, K. L. Han, R. T. Kimball, B. D. Marks, K. J. Miglia, W. S. Moore, S. Reddy, F. H. Sheldon, D. W. Steadman, S. J. Steppan, C. C. Witt, T. Yuri, Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13462–13467 (2008).
40. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
41. D. E. Irwin, Sex chromosomes and speciation in birds and other ZW systems. *Mol. Ecol.* **27**, 3831–3851 (2018).
42. L. Huynen, C. D. Millar, D. M. Lambert, A DNA test to sex ratite birds. *Mol. Ecol.* **11**, 851–856 (2002).
43. K. A. Selkoe, R. J. Toonen, Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* **9**, 615–629 (2006).
44. M. E. Allentoft, C. Oskam, J. Houston, M. L. Hale, M. T. P. Gilbert, M. Rasmussen, P. Spencer, C. Jacomb, E. Willerslev, R. N. Holdaway, M. Bunce, Profiling the dead: Generating microsatellite data from fossil bones of extinct megafauna—protocols, problems, and prospects. *PLOS ONE* **6**, e16670 (2011).
45. M. E. Allentoft, S. C. Schuster, R. N. Holdaway, M. L. Hale, E. McLay, C. Oskam, M. T. P. Gilbert, P. Spencer, E. Willerslev, M. Bunce, Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**, 195–200 (2009).
46. M. E. Allentoft, R. Heller, R. N. Holdaway, M. Bunce, Ancient DNA microsatellite analyses of the extinct New Zealand giant moa (Dinornis robustus) identify relatives within a single fossil site. *Heredity* **115**, 481–487 (2015).
47. F. Termignoni-García, J. J. Kirchman, J. Clark, V. S., Edwards, Comparative population genomics of cryptic speciation and adaptive divergence in Bicknell's and gray-cheeked thrushes (Aves: Catharus bicknelli and Catharus minimus). *Genome Biol. Evol.* **14**, evab255 (2022).
48. K. Wang, H. Hu, Y. Tian, J. Li, A. Scheben, C. Zhang, Y. Li, J. Wu, L. Yang, X. Fan, G. Sun, D. Li, Y. Zhang, R. Han, R. Jiang, H. Huang, F. Yan, Y. Wang, Z. Li, G. Li, X. Liu, W. Li, D. Edwards, X. Kang, The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. *Mol. Biol. Evol.* **38**, 5066–5081 (2021).
49. A. Kapusta, A. Suh, Evolution of bird genomes—A transposon's-eye view. *Ann. N. Y. Acad. Sci.* **1389**, 164–185 (2017).
50. A. Kapusta, A. Suh, C. Feschotte, Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1460–E1469 (2017).
51. H. Yan, A. Bombarely, S. Li, DeepTE: A computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* **36**, 4269–4275 (2020).
52. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
53. Z. Huang, Z. Xu, H. Bai, Y. Huang, N. Kang, X. Ding, J. Liu, H. Luo, C. Yang, W. Chen, Q. Guo, L. Xue, X. Zhang, L. Xu, M. Chen, H. Fu, Y. Chen, Z. Yue, T. Fukagawa, S. Liu, G. Chang, L. Xu, Evolutionary analysis of a complete chicken genome. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2216641120 (2023).
54. C. Goubert, L. Modolo, C. Vieira, C. ValienteMoro, P. Mavingui, M. Boulesteix, De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **7**, 1192–1205 (2015).
55. P. Johnston, K. J. Mitchell, Contrasting patterns of sensory adaptation in living and extinct flightless birds. *Diversity* **13**, 538 (2021).
56. K. W. Ashwell, R. P. Scofield, Big birds and their brains: Paleoneurology of the New Zealand moa. *Brain Behav. Evol.* **71**, 151–166 (2008).
57. M. W. Vandewege, S. F. Mangum, T. Gabaldon, T. A. Castoe, D. A. Ray, F. G. Hoffmann, Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biol. Evol.* **8**, 470–480 (2016).
58. S. Yokoyama, F. B. Radlwimmer, N. S. Blow, Ultraviolet pigments in birds evolved from violet pigments by a single amino acid change. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7366–7371 (2000).
59. Z. Aidala, L. Huynen, P. L. Brennan, J. Musser, A. Fidler, N. Chong, G. E. Machovsky Capuska, M. G. Anderson, A. Talaba, D. Lambert, M. E. Hauber, Ultraviolet visual sensitivity in three avian lineages: Paleognaths, parrots, and passerines. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **198**, 495–510 (2012).
60. M. R. Crole, J. T. Soley, Morphology of the tongue of the emu (Dromaius novaehollandiae). II. Histological features. *Onderstepoort J. Vet. Res.* **76**, 347–361 (2009).
61. M. R. Crole, J. T. Soley, Contrasting morphological evidence for the presence of taste buds in *Dromaius novaehollandiae* and *Struthio camelus* (Palaeognathae, Aves). *Zoomorphology* **134**, 499–507 (2015).
62. Y. Toda, M.-C. Ko, Q. Liang, T. Miller Eliot, A. Rico-Guevara, T. Nakagita, A. Sakakibara, K. Uemura, T. Sackton, T. Hayakawa, S. Y. W. Sin, Y. Ishimaru, T. Misaka, P. Oteiza, J. Crall, S. V. Edwards, W. Buttemer, S. Matsumura, M. W. Baldwin, Early origin of sweet perception in the songbird radiation. *Science* **373**, 226–231 (2021).
63. M. W. Baldwin, Y. Toda, T. Nakagita, M. J. O'Connell, K. C. Klasing, T. Misaka, S. V. Edwards, S. D. Liberles, Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* **345**, 929–933 (2014).
64. K. He, C.-H. Liang, Y. Zhu, P. Dunn, A. Zhao, P. Minias, Reconstructing macroevolutionary patterns in avian MHC architecture with genomic data. *Front. Genet.* **13**, 823686 (2022).
65. B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S. L. Kosakovsky Pond, K. Scheffler, FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196–1205 (2013).
66. M. Alcaide, S. V. Edwards, Molecular evolution of the toll-like receptor multigene family in birds. *Mol. Biol. Evol.* **28**, 1703–1715 (2011).
67. H. Velová, M. W. Gutowska-Ding, D. W. Burt, M. Vinkler, Toll-like receptor evolution in birds: Gene duplication, pseudogenization, and diversifying selection. *Mol. Biol. Evol.* **35**, 2170–2184 (2018).
68. J. D. Yang, M. Zhou, Y. Zhong, L. Q. Xu, C. J. Zeng, X. L. Zhao, M. Zhang, Gene duplication and adaptive evolution of Toll-like receptor genes in birds. *Dev. Comp. Immunol.* **119**, 103990 (2021).
69. L. Huynen, T. Suzuki, T. Ogura, Y. Watanabe, C. D. Millar, M. Hofreiter, C. Smith, S. Mirmoeini, D. M. Lambert, Reconstruction and in vivo analysis of the extinct *tbx5* gene from ancient wingless moa (Aves: Dinornithiformes). *BMC Evol. Biol.* **14**, 75–75 (2014).
70. F. Petit, K. E. Sears, N. Ahituv, Limb development: A paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245–258 (2017).
71. M. Tanaka, Molecular and evolutionary basis of limb field specification and limb initiation. *Dev. Growth Differ.* **55**, 149–163 (2013).
72. C. Tickle, How the embryo makes a limb: Determination, polarity and identity. *J. Anat.* **227**, 418–430 (2015).
73. A. Burga, W. Wang, E. Ben-David, P. C. Wolf, A. M. Ramey, C. Verdugo, K. Lyons, P. G. Parker, L. Kruglyak, A genetic signature of the evolution of loss of flight in the Galapagos cormorant. *Science* **356**, doi/10.1126/science.aal3345 (2017).
74. M. D. Smith, J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, S. L. Kosakovsky Pond, Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
75. Y. Choi, A. P. Chan, PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
76. H. Ellegren, The different levels of genetic diversity in sex chromosomes and autosomes. *Trends Genet.* **25**, 278–284 (2009).
77. J. E. Mank, B. Vicoso, S. Berlin, B. Charlesworth, Effective population size and the faster-x effect: Empirical results and their interpretation. *Evolution* **64**, 663–674 (2010).
78. S. J. Oyler-McCance, R. S. Cornman, K. L. Jones, J. A. Fike, Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity* **115**, 452–459 (2015).

79. V. A. Olson, S. T. Turvey, The evolution of sexual dimorphism in New Zealand giant moa (Dinornis) and other ratites. *Proc. Biol. Sci.* **280**, 20130401 (2013).
80. M. Bunce, T. H. Worthy, T. Ford, W. Hoppitt, E. Willerslev, A. Drummond, A. Cooper, Extreme reversed sexual size dimorphism in the extinct New Zealand moa Dinornis. *Nature* **425**, 172–175 (2003).
81. J. E. Pool, R. Nielsen, Population size changes reshape genomic patterns of diversity. *Evolution* **61**, 3001–3006 (2007).
82. N. J. Gemmell, M. K. Schwartz, B. C. Robertson, Moa were many. *Biol. Lett.* **271** (Suppl 6), S430–S432 (2004).
83. G. L. W. Perry, A. B. Wheeler, J. R. Wood, J. M. Wilmshurst, A high-precision chronology for the rapid extinction of New Zealand moa (Aves, Dinornithiformes). *Quat. Sci. Rev.* **105**, 126–135 (2014).
84. T. H. Worthy, R. P. Scofield, Twenty-first century advances in knowledge of the biology of moa (Aves: Dinornithiformes): A new morphological analysis and moa diagnoses revised. *N. Z. J. Zool.* **39**, 87–153 (2012).
85. W. D. Handley, T. H. Worthy, Endocranial anatomy of the giant extinct australian mihirung birds (Aves, Dromornithidae). *Diversity* **13**, 10.3390/d13030124 (2021).
86. S. Lamichhaney, D. C. Card, P. Grayson, J. F. R. Tonini, G. A. Bravo, K. Näpflin, F. Termignoni-Garcia, C. Torres, F. Burbrink, J. A. Clarke, T. B. Sackton, S. V. Edwards, Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180248 (2019).
87. T. B. Sackton, N. Clark, Convergent evolution in the genomics era: New insights and directions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190102 (2019).
88. E. Hagelberg, Mitochondrial DNA from ancient bones, in *Ancient DNA: Recovery and Genetic Analysis of Genetic Material From Paleontological, Archaeological, Museum, Medical, and Forensic Specimens*, B. Hermann, S. Hummell, Eds. (Springer-Verlag, 1994), pp. 195–204.
89. A. Cooper, Ancient DNA: Do it right or not at all. *Science* **289**, 1139 (2000).
90. M. Knapp, A. C. Clarke, K. A. Horsburgh, E. A. Matisoo-Smith, Setting the stage–Building and working in an ancient DNA laboratory. *Anat. Anz.* **194**, 3–6 (2012).
91. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
92. C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
93. G. Lunter, M. Goodson, Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
94. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
95. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
96. C. D. Mirchandani, A. J. Shultz, G. W. C. Thomas, S. J. Smith, M. Baylis, B. Arnold, R. Corbett-Detig, E. Enbody, T. B. Sackton, A fast, reproducible, high-throughput variant calling workflow for evolutionary, ecological, and conservation genomics. *Mol. Biol. Evol.* **41**, msad270 (2024).
97. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
98. J. Zakany, D. Duboule, The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**, 359–366 (2007).
99. D. Dong, G. Jones, S. Zhang, Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol. Biol.* **9**, 12 (2009).
100. S. Koot, J. Grant, M. Khumûb, K. Fernando, T. Mushavanga, T. Dommerholt, C. Gressier, D. Pienaar, S. Ui Kunta, R. F. Puckett, A. Paksi, S. Moeti, L. O. Tsamkgao, L. Steenkamp, R. Hitchcock, J. Maruyama, R. Gordon, D. Mushavanga, Research codes and contracts do not guarantee equitable research with Indigenous communities. *Nat. Ecol. Evol.* **7**, 1543–1546 (2023).
101. L. Jennings, T. Anderson, A. Martinez, R. Sterling, D. D. Chavez, I. Garba, M. Hudson, N. A. Garrison, S. R. Carroll, Applying the 'CARE principles for indigenous data governance' to ecology and biodiversity research. *Nat. Ecol. Evol.* **7**, 1547–1551 (2023).
102. N. C. Ban, A. Frid, M. Reid, B. Edgar, D. Shaw, P. Siwallace, Incorporate Indigenous perspectives for impactful research and effective management. *Nat. Ecol. Evol.* **2**, 1680–1683 (2018).
103. A. D. Cortez, D. A. Bolnick, G. Nicholas, J. Bardill, C. Colwell, An ethical crisis in ancient DNA research: Insights from the Chaco Canyon controversy as a case study. *J. Soc. Archaeol.* **21**, 157–178 (2021).
104. M. C. Avila-Arcos, C. D. Castro, M. A. Nieves-Colon, M. Raghavan, Recommendations for sustainable ancient DNA research in the Global South: Voices from a new generation of paleogenomicists. *Front. Genet.* **13**, 880170 (2022).
105. E. Kowal, L. S. Weyrich, J. M. Arguelles, A. C. Bader, C. Colwell, A. D. Cortez, J. L. Davis, G. Figueiro, K. Fox, R. S. Malhi, E. Matisoo-Smith, A. Nayak, E. A. Nelson, G. Nicholas, M. A. Nieves-Colón, L. Russell, S. Ulm, F. Vergara-Silva, F. A. Villanea, J. K. Wagner, J. M. Yracheta, K. S. Tsosie, Community partnerships are fundamental to ethical ancient DNA research. *H. G. G. Adv.* **4**, 100161 (2023).
106. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: A fast and accurate illumina paired-end reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
107. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
108. F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
109. B. S. Pedersen, A. R. Quinlan, Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
110. International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
111. D. Le Duc, G. Renaud, A. Krishnan, M. S. Almén, L. Huynen, S. J. Prohaska, M. Ongyerth, B. D. Bitarello, H. B. Schiöth, M. Hofreiter, P. F. Stadler, K. Prüfer, D. Lambert, J. Kelso, T. Schöneberg, Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biol.* **16**, 147–147 (2015).
112. G. J. Zhang, C. Li, Q. Y. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Odeen, J. Cui, Q. Zhou, L. H. Xu, H. L. Pan, Z. J. Wang, L. J. Jin, P. Zhang, H. F. Hu, W. Yang, J. Hu, J. Xiao, Z. K. Yang, Y. Liu, Q. L. Xie, H. Yu, J. M. Lian, P. Wen, F. Zhang, H. Li, Y. L. Zeng, Z. J. Xiong, S. P. Liu, L. Zhou, Z. Y. Huang, N. An, J. Wang, Q. M. Zheng, Y. Q. Xiong, G. B. Wang, B. Wang, J. J. Wang, Y. Fan, R. R. da Fonseca, A. Alfaro-Nunez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farre, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Hastad, R. H. Sawyer, H. Kim, K. W. Kim, H. J. Kim, S. Cho, N. Li, Y. H. Huang, M. W. Bruford, X. J. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. B. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alstrom, J. Fjeldsa, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. M. Yang, J. Wang; Avian Genome Consortium, E. D. Jarvis, M. T. P. Gilbert, J. Wang, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
113. D. H. Huson, S. Beier, I. Flade, A. Górska, M. El-Hadidi, S. Mitra, H.-J. Ruscheweyh, R. Tappu, MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput. Biol.* **12**, e1004957 (2016).
114. C. The Galaxy, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351 (2022).
115. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
116. H. Li, in https://github.com/lh3/seqtk. (2023).
117. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
118. H. Li, New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
119. F. Cabanettes, C. Klopp, D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
120. S. M. Kielbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
121. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
122. S. V. Edwards, P. Beerli, Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839–1854 (2000).
123. S. V. Edwards, A. Cloutier, A. J. Baker, Conserved non-exonic elements: A novel class of marker for phylogenomics. *Syst. Biol.* **66**, 1028–1044 (2017).
124. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
125. E. W. Sayers, E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, S. T. Sherry, Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–d26 (2022).
126. J. L. Steenwyk, T. J. Buida III, Y. Li, X. X. Shen, A. Rokas, ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biol.* **18**, e3001007 (2020).

127. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
128. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, GGTREE: Anrpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
129. K. Katoh, K. Kuma, H. Toh, T. Miyata, MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
130. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
131. Y. Niimura, M. Nei, Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6039–6044 (2005).
132. W. C. Warren, D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. S. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y. C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Voelker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. J. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S. P. Yang, X. C. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. F. Hou, E. R. Mardis, R. K. Wilson, The genome of a songbird. *Nature* **464**, 757–762 (2010).
133. F. Abascal, R. Zardoya, M. J. Telford, TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13 (2010).
134. S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
135. V. Lefort, J. E. Longueville, O. Gascuel, SMS: Smart model selection in PhyML. *Mol. Biol. Evol.* **34**, 2422–2424 (2017).
136. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
137. S. Weaver, S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, S. L. Kosakovsky Pond, Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **35**, 773–777 (2018).
138. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
139. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
140. L. Liu, L. Yu, S. V. Edwards, A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302–302 (2010).
141. A. A. Digiacomo, A. Cloutier, P. Grayson, T. B. Sackton, S. V. Edwards, The unfinished synthesis of comparative genomics and phylogenetics: examples from flightless birds, in *Species Tree Inference: A Guide to Methods and Applications*. L. S. Kubatko, L. L. Knowles, Eds. (Princeton Univ. Press, 2023), pp. 215–231.
142. S. L. Pond, S. D. Frost, S. V. Muse, HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
143. J. O. Wertheim, B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, K. Scheffler, RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
144. J. D. Storey, A. J. Bass, A. Dabney, D. Robinson, qvalue: Q-value estimation for false discovery rate control. doi:10.18129/B9.bioc.qvalue, R package version 2.34.0, https://bioconductor.org/packages/qvalue (2023).
145. 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).
146. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
147. B. C. Faircloth, MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* **8**, 92–94 (2008).
148. K. Kojima, Y. Kawai, K. Misawa, T. Mimori, M. Nagasaki, STR-realigner: A realignment method for short tandem repeat regions. *BMC Genomics* **17**, 991–991 (2016).
149. G. Highnam, C. Franck, A. Martin, C. Stephens, A. Puthige, D. Mittelman, Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. *Nucleic Acids Res.* **41**, e32–e32 (2013).
150. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
151. S. Louca, M. Doebeli, Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**, 1053–1055 (2018).
152. A. Ghosh, M. G. Johnson, A. B. Osmanski, S. Louha, N. J. Bayona-Vásquez, T. C. Glenn, J. Gongora, R. E. Green, S. Isberg, R. D. Stevens, D. A. Ray, A high-quality reference genome assembly of the saltwater crocodile, crocodylus porosus, reveals patterns of selection in crocodylidae. *Genome Biol. Evol.* **12**, 3635–3646 (2020).

Acknowledgments: We thank O. Haddrath for extracting moa aDNA, and T. Worthy and P. Scofield for helpful discussion concerning the provenance of the sequenced moa specimen. We thank S. Pereira of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada, for overseeing library construction and sequencing and M.-C. Ko from the Evolution of Sensory Systems Research Group for assistance with the analysis of sensory receptors. T. Worthy provided helpful comments on the manuscript. We thank P. Deardon, M. Bunce, J. Huh, and C. Rye of New Zealand eScience Infrastructure and Genomics Aotearoa for advice on best practices for scientific and cultural engagement and collaboration. Computations were performed on the GPC supercomputer at the SciNet HPC Consortium funded by Compute Canada, the Government of Ontario, and the University of Toronto, as well as the Cannon cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. During the analysis and writing of this paper, S.V.E. was supported by a King XVI Carl Gustaf Professorship in Environmental Science from the King of Sweden. The costs of publishing this paper were covered in part by a grant from the Wetmore Colles Fund of the Museum of Comparative Zoology. This work is dedicated to the memory of coauthor A.J.B., who was the driving force behind this project but passed away before its completion. **Funding:** This work was supported by the Natural Science and Engineering Research Council of Canada (to A.J.B.), Royal Ontario Museum Governors Fund (to A.J.B.), National Science Foundation [NSF grant DEB 1355343 (EAR 1355292) to A.J.B. and S.V.E.], and Japan Society for the Promotion of Science (KAKENHI grant number JP20K06767 to K.K.). **Author contributions:** Conceptualization: A.J.B. and S.V.E. Methodology: A.C., T.B.S., P.G., and K.K. Resources: P.G. and K.K. Investigation: A.C., S.V.E., T.B.S., M.W.B., and G.C. Data curation: S.V.E. Validation: S.V.E. and T.B.S. Supervision: S.V.E. and M.W.B. Formal analysis: A.C., S.V.E., T.B.S., P.G., K.K., M.W.B., and R.D. Software: T.B.S., S.V.E., and K.K. Visualization: A.C., S.V.E., and M.W.B. Project administration: A.J.B. and S.V.E. Funding acquisition: A.J.B. and S.V.E. Writing—original draft: A.C. and S.V.E. Writing—review and editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The nuclear and mitochondrial assemblies have been deposited in the Aotearoa Genomic Data Repository (https://doi.org/10.57748/M42Z-SW23). The moa assemblies and sequence reads have also been deposited on NCBI under BioProject PRJNA534317, under accession numbers GCA_006937325.1 (original assembly), GCA_006938045.1 (map-damage-corrected assembly), and MK778441 (mitochondrial genome) and Short Read Archive accession SRP132423. Additional datasets and alignments for sensory genes, MHC, transposable elements, and population genetics are available on the Dryad database at https://doi.org/10.5061/dryad.d51c59zxp.

Submitted 10 July 2023
Accepted 17 April 2024
Published 23 May 2024
10.1126/sciadv.adj6823