Ancient proteins resolve controversy over the identity of *Genyornis* eggshell

Beatrice Demarchi[a,1], Josefin Stiller[b], Alicia Grealy[c,d], Meaghan Mackie[e,f], Yuan Deng[b,g], Tom Gilbert[h], Julia Clarke[i], Lucas J. Legendre[i], Rosa Boano[a], Thomas Sicheritz-Pontén[e,h,j], John Magee[k], Guojie Zhang[b,g,l,m], Michael Bunce[c,n], Matthew James Collins[e,o], and Gifford Miller[p,q]

[a]Department of Life Sciences and Systems Biology, Palaeoproteomics Laboratory, University of Turin, 10123 Turin, Italy; [b]Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark; [c]School of Molecular and Life Science, Trace and Environmental DNA Laboratory, Curtin University, Bentley WA 6102, Australia; [d]National Research Collections Australia, Australian National Herbarium, Commonwealth Scientific and Industrial Research Organization, Canberra ACT 2601, Australia; [e]Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1350 Copenhagen, Denmark; [f]The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark; [g]China National GeneBank, Beijing Genomics Institute Shenzhen, 518083 Shenzhen, China; [h]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, 2100 Copenhagen, Denmark; [i]Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712; [j]Center of Excellence for Omics-Driven Computational Biodiscovery, Faculty of Applied Sciences, Asian Institute of Medicine, Science, and Technology University, 08100 Bedong, Malaysia; [k]Research School of Earth Sciences, The Australian National University, Acton ACT 0200, Australia; [l]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 65000 Kunming, China; [m]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 65000 Kunming, China; [n]Institute of Environmental Science and Research (ESR), Porirua 5022, New Zealand; [o]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom; [p]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO 80309; and [q]Department of Geological Sciences, University of Colorado, Boulder, CO 80309

Edited by Paul Szpak, Anthropology, Trent University, Peterborough, ON, Canada; received August 4, 2021; accepted December 20, 2021 by Editorial Board Member Dolores R. Piperno

The realization that ancient biomolecules are preserved in "fossil" samples has revolutionized archaeological science. Protein sequences survive longer than DNA, but their phylogenetic resolution is inferior; therefore, careful assessment of the research questions is required. Here, we show the potential of ancient proteins preserved in Pleistocene eggshell in addressing a longstanding controversy in human and animal evolution: the identity of the extinct bird that laid large eggs which were exploited by Australia's indigenous people. The eggs had been originally attributed to the iconic extinct flightless bird *Genyornis newtoni* (†Dromornithidae, Galloanseres) and were subsequently dated to before 50 ± 5 ka by Miller et al. [*Nat. Commun.* 7, 10496 (2016)]. This was taken to represent the likely extinction date for this endemic megafaunal species and thus implied a role of humans in its demise. A contrasting hypothesis, according to which the eggs were laid by a large mound-builder megapode (Megapodiidae, Galliformes), would therefore acquit humans of their responsibility in the extinction of *Genyornis*. Ancient protein sequences were reconstructed and used to assess the evolutionary proximity of the undetermined eggshell to extant birds, rejecting the megapode hypothesis. Authentic ancient DNA could not be confirmed from these highly degraded samples, but morphometric data also support the attribution of the eggshell to *Genyornis*. When used in triangulation to address well-defined hypotheses, paleoproteomics is a powerful tool for reconstructing the evolutionary history in ancient samples. In addition to the clarification of phylogenetic placement, these data provide a more nuanced understanding of the modes of interactions between humans and their environment.

Genyornis eggshell | Australia | paleoproteomics | ancient DNA | megafaunal extinction

One long-standing controversy in animal and human evolution rests on the taxonomic identity of the eggshell of an extinct giant bird. Thousands of Pleistocene sites in Australia yield vast amounts of eggshell fragments, which are typically found eroding out of sandhills (Fig. 1*C*) or, more rarely, locally reworked into beach sediments, and have provided geochemists and archaeologists with excellent material for geochronological, geochemical, and paleoclimatic studies (1–3). Some of the eggshells exhibit evidence of having been cooked and then discarded in and around a hearth. This is revealed not only by visible charring of the eggshell but also by a specific signature of amino acid decomposition: Miller et al. (4) measured amino acid concentrations along a transect in partially charred eggshell, starting from the blackened end and moving away from it (see figure 2 in ref. 4). They found that amino acids were fully decomposed in the burnt end but that concentrations increased along the transect. This is consistent with a high-temperature gradient typical of contact with hot embers and cannot be attributed to bush fires. Strikingly, burnt *Dromaius* (emu, part of ratites and tinamous Palaeognathae) eggshell appears around 55 ka B.P. and remains frequent through to near-modern time. A second type of eggshell, which we dub here "undetermined ootaxon" (UO), bears signs of cooking but only during a narrow temporal window [50 ± 5 ka B.P. (4)]. This

Significance

The controversy over the taxonomic identity of the eggs exploited by Australia's first people around 50,000 y ago is resolved. The birds that laid these eggs are extinct, and distinguishing between two main candidates, a giant flightless "mihirung" *Genyornis* and a large megapode *Progura*, had proven impossible using morphological and geochemical methods. Ancient DNA sequencing remains inconclusive because of the age and burial temperature of the eggshell. In contrast, ancient protein sequences recovered from the eggshell enabled estimation of the evolutionary affinity between the egg and a range of extant taxa. The eggs are those of a Galloanseres (a group that includes extinct Dromornithidae, as well as extant landfowl and waterfowl), *Genyornis*, and not of the megapode (Megapodiidae, crown Galliformes).

Author contributions: B.D., J.S., M.B., M.J.C., J.M. and G.M. designed research; B.D., A.G., M.M., J.C., L.J.L., T.S.-P., and M.J.C. performed research; B.D., T.G., R.B., G.Z., M.B., and M.J.C. contributed new reagents/analytic tools; B.D., J.S., A.G., Y.D., J.C., L.J.L., and M.J.C. analyzed data; and B.D., J.S., A.G., J.C., L.J.L., M.J.C., and G.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. P.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: beatrice.demarchi@unito.it.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2109326119/-/DCSupplemental.

Published May 24, 2022.

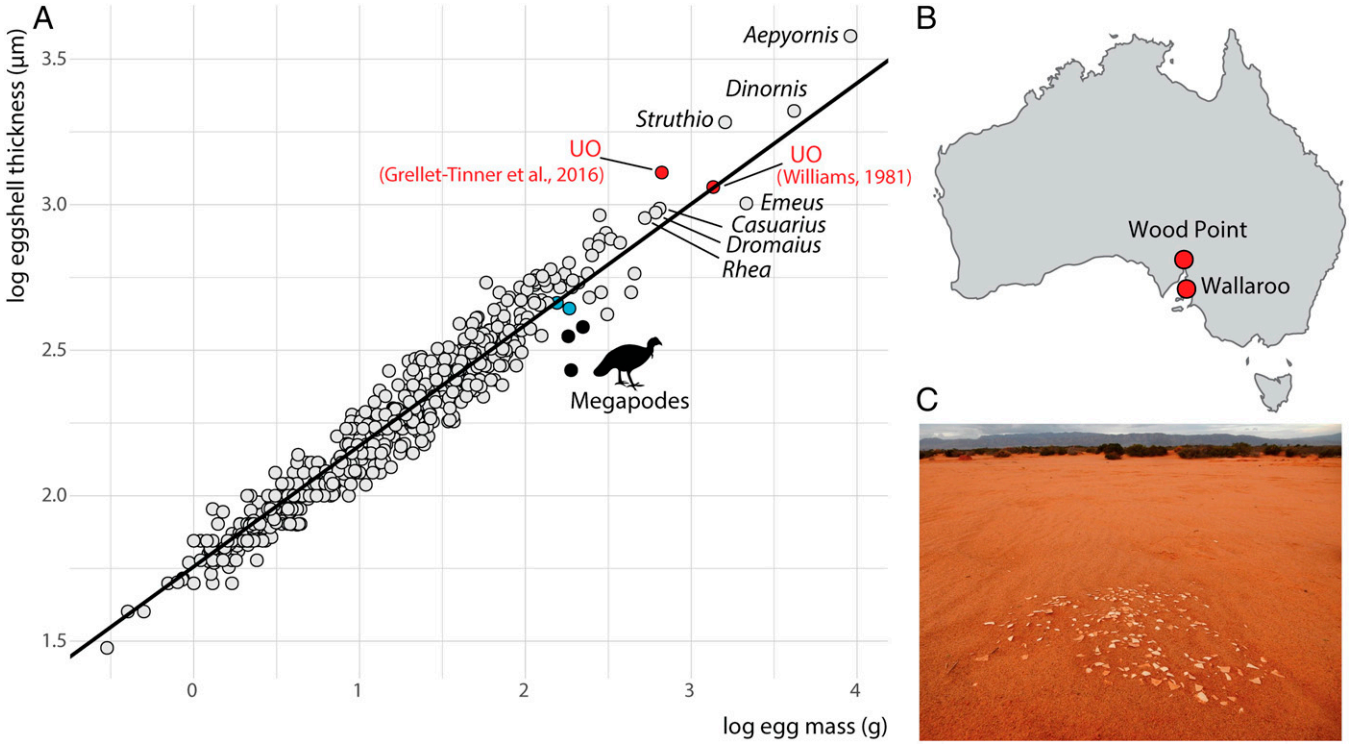

Fig. 1. (*A*) Scaling relationship of eggshell thickness with egg mass for 538 species of extant birds with two estimates for UO (red) included for comparison. Megapodes (Megapodiidae) and screamers (Anhimidae) are colored in black and blue, respectively. Fitted line corresponds to phylogenetic generalized least-squares regression of eggshell thickness on egg mass (pseudo R-squared: 0.97). Silhouette of *Aepypodius bruijnii* taken from PhyloPic (Public Domain). (*B*) Location of the archaeological site of Wood Point and of the Wallaroo sand dunes. (*C*) Concentration of surface and near-surface UO eggshell recently exposed in the South Australia sandhills. Based on the mass of the Spooner Egg (17), the recovered eggshell represents a concentration of 10 to 13 whole eggs, which we interpret to be a nest. Several such clusters have appeared in the same degrading sandhills revisited over 15 y. All have the same level of aa epimerization; the optically stimulated luminescence age of dune sand at one such cluster is 55.5 ± 2.3 ka (122).

interval is broadly contemporaneous with the spread of people across Australia, with the earliest robust date of arrival (currently) estimated at ∼65 ka B.P. (5). The simple repeated action of egg cooking, carried out many thousand years ago, now represents a unique insight into human–fauna interactions and early anthropic impacts on virgin landscapes. Whether this behavior also contributed to the extinction of some megafaunal species is hotly debated and hinges on the identity of the UO eggshell remains. In one scenario, the UO eggshell belong to *Genyornis newtoni*, the last surviving member of the giant (∼200 kg) flightless "mihirungs" (i.e., Dromornithidae), an extinct clade within Galloanseres (6–10). Modern Galloanseres comprise Galliformes, which are land fowl such as chickens, pheasants, quails, and megapodes, and Anseriformes, which are waterfowl such as ducks, geese, and screamers. In a contrasting hypothesis, these eggshells belong to a large (∼5 kg) volant extinct megapode (i.e., Megapodiidae) within the crown Galliformes.

The attribution of UO eggshell [from the site of Coopers Dune, 50 ± 5 ka B.P. (4)] to *Genyornis* was first proposed by Williams (11). Despite the lack of direct association between skeletal remains and UO eggshells, they were in close proximity and approximately the right size for the egg layer. If the Williams identification is correct, *Genyornis* was among the megafaunal species that disappeared a few thousand years after the arrival of people, implying a role of humans in its extinction (1, 4, 5, 7). In this debate, the closest parallel would appear to be the demise of the New Zealand megaherbivore, moa (Palaeognathae), which occurred within 150 y of the arrival of a small group of people (∼2,000 individuals), around 1300 to 1500 CE, as the result of habitat removal and hunting (12–16). However, the Australian Pleistocene archaeological record reveals the near absence of evidence for kill sites or human modification of *Genyornis* skeletal remains. This could be due to taphonomic biases (e.g., sea level rise), but if we accept the classification of the UO as *Genyornis*, it most likely implies that mihirung egg exploitation was routinely carried out by Australia's first people, while hunting was not (4). Could egg exploitation have become unsustainable and driven *Genyornis* to extinction?

A different scenario is evoked by the recent challenge to the identity of the UO eggshell layer and its attribution to an extinct large megapode such as *Progura* (17). The hotly debated (17–19) megapode hypothesis is primarily based on estimates of a thin eggshell, the eggs being rather small for the estimated size of *Genyornis*, and the widespread presence of remains from the extinct large megapode *Progura* in Australia. If this hypothesis is proven correct, it would imply that humans may have had an impact on the extinction of a volant, mound-building, relatively small (∼5 kg) bird around 50 ka B.P. People would thus have preferentially targeted eggs in hidden mounds and burrows rather than collecting them from the ground nests of large flightless birds, the latter being a typical behavior of modern humans, who had been exploiting ostrich eggs in Africa and Asia since at least 120 ka B.P. (20). More importantly, there would be no direct proof of human–*Genyornis* interaction, and it is indeed possible that *Genyornis* had already gone extinct by the time humans arrived in Australia. A scenario of human–megapode interaction therefore throws a completely different light on the debate over human role(s) in Australian megafauna extinctions, supporting environmental-driven hypotheses. This would be in line with recent studies, which have argued that there is no link between the arrival of humans and the extinction of megafaunas for the global Pleistocene

record (21) and that Sahul extinctions were chronologically coupled with hydroclimate deterioration and not with anthropogenic effects (22).

In summary, the debate hinges on two mutually exclusive hypotheses: either the eggshell belongs to *Genyornis*, a flightless member of Galloanseres, or to a *Progura*-like volant megapode in crown Galliformes. The key piece of missing evidence is therefore a morphometric character or molecular marker (or combination of both) able to unambiguously falsify either hypothesis. Here, we used three independent approaches in order to obtain information on the identity of UO eggshell: 1) comparison of eggshell thickness scaling in UO with that of over 500 species of extant birds using phylogenetic regressions, 2) hybridization capture and next-generation sequencing of ancient DNA extracted from the eggshell, and 3) ancient protein sequencing using high-resolution tandem mass spectrometry followed by phylogenetic analyses of "fossil" sequences. The results are internally consistent, allowing us to resolve the *Genyornis*/megapode controversy and bringing to the fore important implications for the use of biomolecular methods in archaeology.

Results

Eggshell Morphology. The taxonomy of UO has been debated by examining a range of morphometrical, microstructural, and geochemical parameters and datasets (17–19), but the initial argument was based on a proposed unusual small size and thin eggshell. The eggshell fragments are 0.95 to 1.36 mm (17), that is, slightly thicker than that of midsized Palaeognathae [e.g., emu, rheas, and cassowaries (23–25)]. Using the eggshell fragment curvature and a range of equations relating body mass (estimated from skeletal remains of *Genyornis*) to egg size, the eggs' dimensions can be estimated as ∼125 × 155 mm (11, 19) or ∼97 × 125 mm (17, 18). Our phylogenetic regression of eggshell thickness on egg mass for extant birds (Fig. 1*A*) shows that for either estimate of egg dimensions, UO eggs fall very close to the general relationship of these two traits among birds. All three megapodes in our sample are clear outliers to that relationship, far below the regression line due to their very thin eggshells and large eggs (Fig. 1*A*). If the UO eggshells had a thickness/mass scaling relationship similar to that of extant megapodes, the eggs would be expected to be further away from the regression line (i.e., closer in size to those of large moa [*Dinornis*, Palaeognathae]). Conversely, the screamers (Anhimidae)—that is, early diverging Anseriformes—fall close to the regression line (Fig. 1*A*), making the thickness/mass scaling of UO more consistent with a *Genyornis* identity.

Other aspects of UO eggs are either not diagnostic or inconsistent with the hypothesis of their layer being a megapode. Their Y-shaped pore structure, for example, is present in megapodes but also in other large dinosaur eggs [e.g., flightless Palaeognathae, sauropods (23, 26, 27)]. UO eggs present a typical ovoid shape (19), very different from that of megapode eggs, which are much larger and elongated than those of other similar-sized Galliformes and have a very large yolk [over 50% of the egg content (28, 29)]. The thin eggshell of megapodes also shows a dense horizontal pore network, undocumented in any other extant birds (30). In extant birds, egg elongation is correlated with an enlarged yolk (31) and powered flight (32), and both enlarged yolk and high eggshell conductance are correlated with precociality (31, 33, 34). Megapodes incubate their eggs in mounds of rotting vegetation—or, for some species, in sand heated by the sun or by geothermal activity (35, 36). These highly controlled temperature and humidity conditions (36, 37) contribute to their hyperprecociality and associated egg traits (35). Therefore, if UO eggs belonged to a giant megapode, that species would have had a growth pattern and incubation strategy highly distinct from those of its extant relatives.

It is also worth noting that the concentrations of near-surface and surface eggshell ("nests," Fig. 1*C*) where UO eggshells were discovered are located in sand dunes devoid of significant organic matter that could have been used for nest building, rendering the presence of vegetation mounds unlikely (19). Alternatively, a giant megapode nesting on that site could have burrowed its eggs in the sand; this nesting strategy, however, is only found in seven species (four of which can also use mound nesting and none of which live in Australia) out of 22 (28). The ancestral megapode is estimated to be a mound nester, and burrow-nesting megapodes are inferred to have acquired this strategy only after their Pleistocene dispersal to islands north of Australia (e.g., New Guinea, Philippines), where vegetative material was then scarcely available (36). However, since burrowing species cannot readily adjust the structure of the burrow like that of a mound, using the sun as the sole heat source for incubation requires a very loose substrate structure (e.g., volcanic or siliceous sand), and this can only happen during the dry season when solar radiation is maximal (38). For this reason, all seven burrowing megapode species tend to either bury their eggs with organic matter that also contributes to incubation (sometimes near rotting tree roots) or dig their burrows in volcanic sand so that incubation is in part ensured by geothermal activity (38, 39). Thus, since solar heating is almost always used in addition to other heat sources for incubation (39), the lack of vegetation in the nonvolcanic sand dunes where UO eggshells were recovered would always be a constraint on them being incubated in a mound or a burrow as built by a megapode, especially a giant one. While the acquisition of a novel nesting strategy associated with body size increase cannot be excluded, this makes the existence of a giant Australian burrow-nesting megapode highly unlikely.

Paleogenomics. Guided by amino acid racemization (isoleucine epimerization, A/I) data, previously shown to be correlated with biomolecular preservation in ostrich eggshell (40), 14 of the most promising samples were selected for ancient DNA (aDNA) analysis, a strategy previously successful in Madagascar (41). While aDNA has been retrieved from million-year-old tissues (42), the hot climate of Australia is not conducive to the survival of aDNA in 50-ka UO eggshell (∼0.3 Ma thermal years). An avian *12S* ribosomal DNA (rDNA) mini-barcode amplified was 100% identical to chicken (*Gallus gallus*), a common contaminant, and any avian reads that mapped to various reference avian mitochondrial genomes (*Materials and Methods*) cannot be distinguished from nonendogenous contamination, as most were identical to domestic Galloanseres (*SI Appendix*, *Ancient DNA supplementary results*). However, no reads from the laboratory controls mapped to any of the avian mitochondrial reference genomes used, and reads from the controls were largely human or bacterial in origin. The phylogenies that were reconstructed from avian mitochondrial reads show all three samples placed sister to *Gallus gallus*, indicating that the chicken is a likely contaminant: if the samples were truly Galliformes, we would expect them to fall within Galliformes but not immediately sister to the chicken (*SI Appendix*, Fig. S11 *A*, *C*, and *E*). When reads assigned to *Gallus* are excluded from mapping, we find one specimen is placed within Galliformes but not sister to *Gallus* or Megapodiidae (*SI Appendix*, Fig. S11*B*), while two others are both placed within Anseriformes (*SI Appendix*, Fig. S11 *D* and *F*). A higher proportion of C to T nucleotide misincorporations at both the 5 and 3′ ends of mapped reads was not observed (*SI Appendix*, Fig. S12); however, this was expected because the coverage was so low. As such, we cannot be confident the mapped reads are truly ancient in origin, and aDNA data remains inconclusive.

Paleoproteomics. Ancient proteins were successfully extracted from three fragments of UO eggshell—one from the archaeological site of Wood Point, where partially charred eggshell fragments were found amongst un-burnt fragments (43), and two from sand dunes near Wallaroo (Fig. 1*B*). The samples were selected based upon A/I values (Table 1) because we have previously shown a negative correlation between the number of identified peptides and increasing racemization in ostrich eggshells (40). The proteins were extracted from eggshell powders that had been extensively bleached in order to isolate the intracrystalline organic fraction and remove external contamination (40, 44). The tandem mass spectrometry raw data were searched against all proteins downloaded from the National Center for Biotechnology Information (NCBI) repository, restricting the taxonomy to Aves and including common contaminants. These preliminary searches highlighted that the proteome composition was similar across the three samples and that the top-scoring proteins were neuronal pentraxin, lactadherin, and C-type lectin (1- and 2-). Neuronal pentraxin is a Ca-binding protein involved in central nervous system development. Lactadherin/milk fat globule-EGF factor 8 (MFGE8) is consistently detected in eggshell proteomes and is involved in mineralization, particularly in the vesicle-mediated transport of amorphous calcium carbonate (45, 46). MFGE8 has calcium-binding EGF-like domains, and it is overabundant in the proteome of mechanically stronger shells (47). The C-type lectins are highly specific for biomineralization and are present as two forms (XCA-1 and XCA-2, following the nomenclature of 46) in ratites, while Neognathae can possess both or either paralog forms.

Preliminary analyses showed that XCA-1 and lactadherin were the most suitable sequences, as they yielded higher protein sequence coverages and the appropriate phylogenetic resolution. Neuronal pentraxin was not considered for phylogenetic analyses because it had less informative sites and was not available for the reference Megapodiidae, *Alectura lathami*. Raw tandem mass spectrometry data were then searched against custom-made databases containing XCA-1, lactadherin, and common laboratory contaminant sequences. The top XCA-1 match for the UO was clearly with *Anseranas semipalmata* and *Chauna torquata* (~70% coverage), with other taxa yielding lower identities (~50%). The match between the data from the UO and the lactadherin references was around 40 to 60% identity with a range of avian taxa. This is a common issue arising from the difficulty of automatically ranking conserved peptide sequences, which, in this case, required manual checking of the quality of individual spectra and peptide spectrum matches. The "fossil" UO proteins were reconstructed by aligning all peptide hits from each reference sequence and constructing consensus sequences (a selection of annotated spectra supporting the reconstruction of the UO XCA-1 and lactadherin are reported in *SI Appendix*, *Selection of annotated tandem mass spectra supporting the reconstruction of UO protein sequences*). *SI Appendix*, Figs. S1 and S2, respectively, display the alignment of UO XCA-1 and lactadherin with reference species; the coverage obtained was sufficient for obtaining a prediction of UO XCA-1 structure using AlphaFold (48) by inclusion of a short inferred sequence (*SI Appendix*, Fig. S3 and *Supplementary Figures*).

Phylogenetic Placement of UO. Consensus UO sequences for XCA-1 and lactadherin, together with data for 364 bird species, were used to assess the evolutionary placement of UO relative to extant birds, particularly to Palaeognathae, Galliformes, and Anseriformes. Concatenated maximum likelihood phylogenetic inference (756 amino acids, aa) resulted in a tree topology with major groups of birds—that is, ratites and tinamous (Palaeognathae), land fowl and waterfowl (Galloanseres), and all other modern birds (Neoaves)—being monophyletic (Fig. 2 and *SI Appendix*, Fig. S4). The average bootstrap support (bs) was notably lower within Neoaves (bs = 32) than in Palaeognathae (bs = 66) and Galloanseres (bs = 72), which is consistent with the long-standing difficulties in resolving the relationships among the deep branches of Neoaves (49–51). Because protein and DNA sequences can be informative at different time scales, we compared the aa dataset to the corresponding DNA dataset for the extant species only, because of the unavailability of DNA data for UO. We found the same relationships but higher bs in the DNA (bs = 55) compared to the protein dataset (bs = 36) (*SI Appendix*, Fig. S5). The results demonstrate that eggshell XCA-1 and lactadherin protein sequences can recover deep time divergences and are therefore adequate to our scope.

In all phylogenetic models, whether unconstrained or constrained to published topologies, UO fell firmly outside Palaeognathae and within Galloanseres (Fig. 2 and *SI Appendix*, Fig. S6). While the first nodes of Galliformes are poorly supported and characterized by short branch lengths, the inclusion of UO within Galloanseres has good support (bs = 83). UO was supported as the sister to a clade of Galliformes and of Anatidae plus Anseranatidae (bs = 52, Fig. 2). The exact placement within Galloanseres is complicated by the paraphyly of Anseriformes, with *Chauna torquata* (Anhimidae) being the sister group to the remaining Galloanseres and UO. Anseriformes are monophyletic in both morphological and DNA-based phylogenies (9, 50, 52). When constraining the topology to established phylogenetic relationships that include a monophyletic Anseriformes, we find that UO is placed as the sister group to Anatidae and Anseranatidae with good support (bs = 91, *SI Appendix*, Fig. S6). However, given that support for an enforced monophyletic Anseriformes was low (bs = 19), we consider it more conservative to conclude that the phylogenetic placement of UO is certainly within the more inclusive Galloanseres clade, with the possibility of a placement as the sister of the group including Anatidae and Anseranatidae.

UO was never recovered within the crown Galliformes (bs = 91) and was several nodes distant from the reference megapode *Alectura lathami*. Although the exact relative position of UO, Anhimidae, and Anatidae + Anseranatidae was not always identical, subsets of this analysis resulted in the same placement of UO in Galloanseres, when only analyzing taxa with both genes were represented (*SI Appendix*, Fig. S7), when only species of Palaeognathae and Galloanseres were included (*SI Appendix*, Fig. S8), or when only Australasian taxa were included (*SI Appendix*, Fig. S9). A genetic affinity between UO and Megapodiidae is therefore rejected.

Discussion

Ancient proteins have recently been in the spotlight as the most promising source of information for clarifying the Pleistocene evolutionary history of extinct organisms in the absence of aDNA, having been retrieved from a 3.8-Ma eggshell (40) as well as

Table 1. Details of samples analyzed in this study

Turin Sample ID	Miller Lab ID	aDNA Lab ID	Site	Coordinates	MAT (°C)	Avg A/I	OSL ± 1σ (ka)
PALTO 215	M14-A031	AD2078	Wood Point	−33.3475°, 137.8988°	17.9	0.291 ± 0.001	55.0 ± 5.0
PALTO 216	M14-A035	AD2079* (CSS3622)	Wallaroo	−33.8842°, 137.6110°	17.0	0.371 ± 0.001	
PALTO 217	M14-A036	AD2080		−33.8841°, 137.6109°		0.354 ± 0.004	

Mean annual temperature (MAT) and geochronological data from ref. 4. Asterisked samples had their DNA sequenced albeit unsuccessfully.

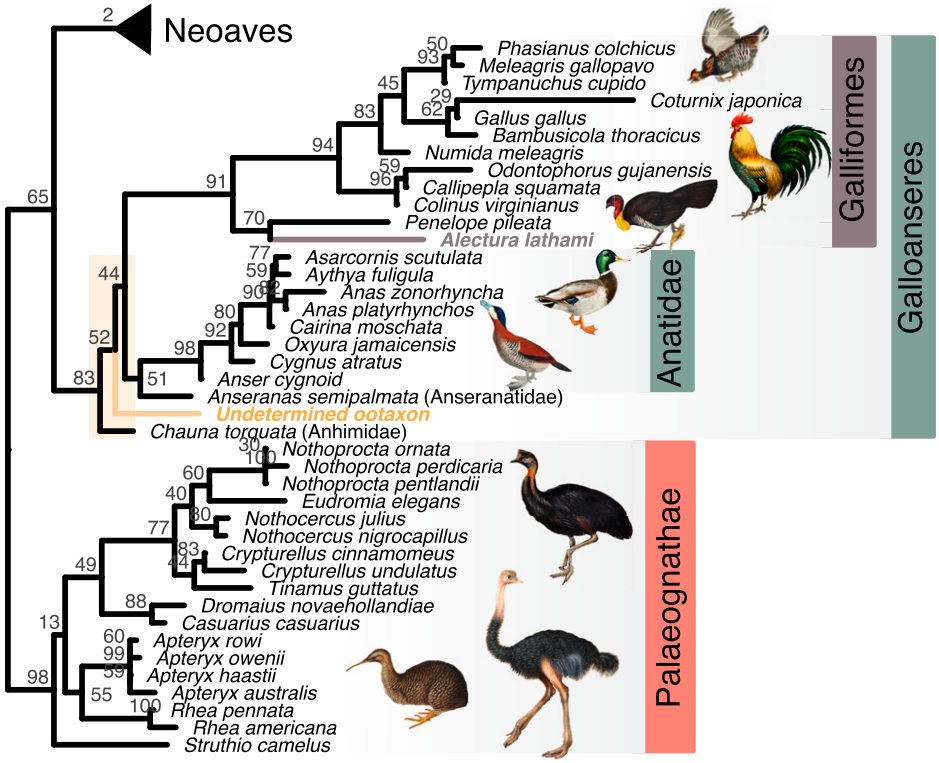

Fig. 2. Maximum likelihood phylogenetic tree based on the concatenation of two eggshell protein sequences. The UO is highlighted in orange, and the orange box indicates the area of phylogenetic uncertainty in which the taxon fell in different analyses. Node numbers are bootstrap support.

million-year-old dental enamel from both animal and hominin taxa (53–55). Their preservation in 50,000-y-old eggshells from warm Australian environments (mean annual temperature = 17 to 17.9 °C, thermal age ∼300 ka) is highly significant: eggshell proteins are firmly established as a reliable source of information for assessing the evolutionary proximity between extant and extinct organisms. In the case of the UO, we can confidently exclude a close affinity with Megapodiidae and support its original determination as *Genyornis* eggs, thus contributing to our understanding of human–megafauna interactions in the past.

In our study, protein sequences could recover clades of Palaeognathae, Galloanseres, Galliformes, Anatidae, and Neoaves, which diverged more than 50 Mya. The topology had shortcomings within Neoaves, the paraphyly of Anseriformes, and often low bs, which are likely a consequence of the relatively low number of aa sites analyzed here. We cannot unambiguously place the ootaxon as a relative of any extant species, but we can confidently place it among the first branches of Galloanseres and certainly outside of crown Galliformes. Protein sequences are often advocated for phylogenies because they generally suffer less from saturation, while DNA sequences have an advantage of having more sites (56). While we cannot directly test the placement of the ootaxon based on aDNA data, we found that the relationships among the main clades were similar when comparing the aa and the corresponding nucleotide data. It is possible that noncoding regions would be more phylogenetically informative than the protein-coding sequences as has been found in genome-wide comparisons (57, 58), but noncoding data may never become available for such ancient samples. This suggests a role for these ancient protein sequences in addressing evolutionary placement in other systems.

The limited number of modifications that can occur within protein sequences is dependent on chemical and steric factors. As shown in *SI Appendix*, Figs. S1 and S2, common substitutions are between aa with similar characteristics, for example, Asp ↔ Asn, Glu ↔ Gln, Asp ↔ Glu, Ile ↔ Leu or Ile ↔ Val, Tyr ↔ Ser, and Arg ↔ Lys. The protein structure also plays a fundamental role; for example, Cys residues are highly conserved in XCA-1 sequences, as they ensure the formation of disulfide bridges and, thus, the correct folding. These chemical and structural constraints imply that aa substitutions are neither wholly random nor occur at a constant rate (see also ref. 59), which is a key requirement for an accurate molecular clock. The proteins can therefore be considered somewhat more akin to morphological characters (i.e., conserved and subjected to environmental pressures). Issues of resolution can be further compounded by diagenesis and differential preservation. We have recently described (40) a mechanism of molecular preservation which involves mineral-surface stabilization, ensuring the survival of Asp-rich peptides from ostrich XCA-1 sequences (residues 92 to 99 in *SI Appendix*, Fig. S1). Unfortunately, this region is among the most highly conserved and therefore of limited use for taxonomic determinations. On the contrary, a taxonomically variable region (residues 76 to 83 in *SI Appendix*, Fig. S1) is predicted to be a surface-exposed loop between two β-sheets and is not preserved in the UO sequence (*SI Appendix*, Figs. S1 and S3). aa substitutions can therefore be driven by functional constraints. This also leaves protein-coding sequences susceptible to homoplasy, as is the case with skeletons, where convergent morphological adaptations are mirrored by the same or similar molecular changes. This phenomenon appears widespread in birds, as waterbird and predatory morphs have evolved multiple times and are associated with convergent molecular changes (49), while flight loss was associated with convergent changes in regulatory DNA (60).

These caveats do not imply that ancient protein sequences cannot give accurate taxonomic information. On the contrary, in the presence of well-defined hypotheses and a good set of reference sequences, we show here that species-level resolution could be achieved by targeting specific peptides. For example, positions 12, 25, 36, and 80 of the XCA-1 alignment in *SI Appendix*, Fig. S1 could be used to

separate *Nothoprocta perdicaria* and *Nothoprocta ornata* archaeological eggshells. While the suitability of peptide mass fingerprinting for the rapid order-level identification of archaeological eggshell is well established (61–65), these earlier studies consistently recognized that a major limitation in taxonomic resolution was the lack of reference sequences and stressed the importance of integrating morphological and molecular evidence. Two methodological breakthroughs have occurred recently, which have the potential to revolutionize the way eggshell is studied archaeologically: thanks to the Bird 10,000 Genomes (B10K) project (http://b10k.genomics.cn), we can now for the first time rely on a growing dataset of well-annotated tissue-specific protein sequences, and tools such as AlphaFold (48) can provide additional structural insights. Well-supported shared derived sequence characteristics can now be identified that are apomorphic for key taxa (e.g., Tyr → Met in Galliformes at position 332 in *SI Appendix*, Fig. S2). The presence or absence of these in fossil targets allow key tests for specific hypotheses of identity but will also allow for the reconstruction of the evolution of protein traits and the investigation of potential taphonomic effects. Combined with zooarchaeological, environmental, and historical information, this is certainly the beginning of a new era in the study of human–avifauna interactions.

The *Genyornis* study is a clear example of the potential of this approach: we defined a testable hypothesis and gathered three independent lines of evidence based on morphology, paleogenomics, and paleoproteomics. aDNA data (while suggestive) were inconclusive, but ancient protein data rejected the megapode hypothesis on the basis of evolutionary distance between UO and extant Megapodiidae and were supported by morphological considerations. The UO eggshell was therefore attributed to *Genyornis*, a conclusion which is pivotal for understanding how Australia's first people interacted with their new environment and reaffirms the value of the *Genyornis* eggshell as a tool to study paleoclimate and extinction processes. While these data cannot directly assess the extent of the interaction between humans and *Genyornis*, we note that, given the geographical extent of burnt *Genyornis* eggshell (4), this interaction was likely nonsporadic.

As a minimum, at least one human exploited at least one *Genyornis* egg at the site of Wood Point around 50,000 y ago. In doing so, they were likely reproducing the same mode of interaction that they had established with another large, flightless bird: ostrich, albeit on other continents. Ostrich bones are rarely found in archaeological sites (66), but ostrich eggshell (OES) has been ubiquitous in human occupation sites in southern and eastern Africa since the Middle Stone Age (20, 67–69). OES is also common in northern Africa, the Arabian Peninsula, southwestern and northern Asia, and present-day India and China throughout the Pleistocene and Holocene (70–74). In the past, and today, ostrich eggs were widely used for a variety of purposes: the egg itself is an important source of nourishment, while the shell can be worked to make beads, which reify social relationships and identities, or it can be used as a water container. Ostriches and humans therefore have coexisted for at least 100,000 y, and this interaction was seemingly a very specific one, mainly involving egg collection. The secure biomolecular identification of *Genyornis* eggshell therefore hints at the antiquity, continuity, and persistence of specific patterns of human–megafaunal-birds interaction over large temporal and spatial scales.

Materials and Methods

Phylogenetic Comparative Methods for Eggshell Thickness Scaling. A dataset for eggshell thickness (μm) and egg mass (g) was compiled for 602 species of extant and subfossil birds from Maurer et al. (75), Juang et al. (76), and Legendre and Clarke (25). Two additional megapode species (*Alectura lathami* and *Leipoa ocellata*) were sampled from Grellet-Tinner et al. (30). Egg mass was estimated from egg length and width using the equation of Hoyt [ref. 77; see also ref. 25]. Both traits were log converted [common logarithm (78)] prior to analysis. All subsequent analyses were performed in R 4.1.0 (79). Following Rubolini et al. (80), we sampled 100 phylogenetic trees from BirdTree.org using the topology of Jetz et al. (81) with the backbone of Hackett et al. (82) and generated a consensus calibrated tree using "consensus.edges" in phytools (83). The dataset was reduced to the 538 species sampled in this tree for subsequent analyses. We performed phylogenetic generalized least-squares [PGLS (84)] regressions of eggshell thickness on egg mass using nlme (85) and ape (86). The goodness of fit was assessed using pseudo R-squared (R_{pred}) in rr2 (87, 88). Normality and homoscedasticity of the residuals were assessed using a Shapiro–Wilk test and a Q–Q plot and a residuals versus fits plot, respectively (89). The regression plot includes two values of eggshell thickness and egg mass for the *Genyornis* eggs, taken respectively from Williams (11) and Grellet-Tinner et al. (17).

Paleogenomics. Four UO (putative *Genyornis*) eggshell specimens were collected north of Spencer Gulf, South Australia, along with an additional 10 from Arcoona Station, Woomera, South Australia, by G.M. (*SI Appendix*, Table S1). In a designated cleanroom facility at Curtin University, DNA was extracted from 200 mg eggshell powder following the method described by Dabney et al. (90) with minor changes outlined in Grealy et al. (41) (*SI Appendix, Ancient DNA supplementary methods*). A DNA-free extraction control was included for every 11 samples. We attempted to amplify a 53-bp (base-pair) barcoding region of mitochondrial *12S rDNA* using universal bird-specific primers (91) as described elsewhere (SI 1.4 of ref. 92; see also *SI Appendix, Amplification of the 12S rDNA mini-barcode*). DNA-free PCR and extraction controls were included. One barcode only was able to be sequenced on a MiSeq Nano flowcell single-end 150-cycle V2. Five samples in which the previously described 53-bp barcoding region could be amplified were prioritized for single-stranded shotgun library preparation (*SI Appendix*, Table S1) following the methods of Gansauge et al. (93) with minor changes (*SI Appendix, Shotgun library preparation*). Negative control libraries and extraction controls were also prepared and carried through capture to sequencing. A total of 100 mer mitochondrial and nuclear avian baits with 50-bp tiling were previously designed using the sequences in table S5.9.4 of Grealy et al. (39) and manufactured by MYBaits (MYcroarray). The hybridization capture was performed on three libraries (*SI Appendix*, Table S1) according to the MYcroarray MYBaits Sequence Enrichment for Targeted Sequencing v.3 manual with minor changes (*SI Appendix, Hybridisation capture*). The final sequencing library was sequenced on a MiSeq standard flowcell paired-end 300-cycle V2 (*SI Appendix, Hybridisation capture*). The reads were demultiplexed, trimmed, filtered, and de-replicated according to Grealy et al. (94) and *SI Appendix, Bioinformatics*. Any reads that mapped to the reads within the control libraries were discarded. All libraries were individually mapped to several reference mitochondrial genomes (*SI Appendix, Bioinformatics*), and mapped reads that were avian in origin were used to reconstruct a consensus mitochondrial genome (*SI Appendix, Bioinformatics*). Mapping was also repeated, excluding reads that were assigned to *Gallus gallus* as potential contamination. Consensus sequences were then separately aligned with 40 other avian mitochondrial genomes (*SI Appendix*, Fig. S10), and protein-coding genes were partitioned in codon positions, while RNA-coding genes were partitioned into stems and loops (*SI Appendix, Bioinformatics*). IQ-Tree 1.6.12 (95, 96) was used to find the best nucleotide substitution model for each partition (*SI Appendix*, Table S6) and generate a consensus maximum likelihood tree from 500 bootstrap replicates (*SI Appendix, Bioinformatics*). MapDamage 2.2.1 (97) was used to examine C to T misincorporations at the 5 and 3′ terminus of reads (*SI Appendix, Bioinformatics* and Fig. S12). The raw data, alignments, and phylogenetic analyses are available on FigShare (10.6084/m9.figshare.15084879)

Paleoproteomics. The eggshell fragments targeted for ancient protein analysis were collected at the site of Wood Point (one sample) and sand dunes near Wallaroo (two samples). Wood Point is about 10 m above sea level, along Spencer Gulf, south of Port Augusta. This is a human occupation site with strong evidence of human exploitation of emu and UO eggs for alimentary purposes (2, 43). It has been dated by Miller et al. (4) to 44,094 to 47,419 cal B.P. The Wallaroo sand dunes are 65 km south of Wood Point. The three eggshell fragments come from three separate eggs. It is worth noting that the UO eggshells found throughout the continent are all consistent in terms of thickness, pore pattern and degree of curvature, and, importantly, taphonomy—which is similar to that of the emu's eggshell found in the same sites.

Sample preparation. The preparation of samples was carried out in a dedicated laboratory at the University of Turin, handling the samples under a laminar flow cabinet in accordance with international guidelines (98). Eggshell is an excellent substrate for paleoproteomics, as it retains a tight closed system of proteins, which can be isolated by extensive bleaching (40, 44). Here, we bleached powdered eggshell samples (~40 mg) for 72 h and then extracted the intracrystalline fraction by demineralizing in 0.6 M hydrochloridric acid

(fresh solution, cooled to 4 °C) followed by ultrafiltration (Nanosep Centrifugal Devices, 3-kDa MWCO, Pall Laboratory). Following the reduction and alkylation of disulphide bonds with dithiothreitol and iodoacetamide according to usual eggshell protocols (40, 70), digestion was carried out overnight at 37 °C on two subsamples for each sample, adding 4 μL trypsin (0.5 μg/μL; Promega) for "T" subsamples and 4 μL elastase (1 μg/μL) for "E" subsamples. Digestion was stopped by adding trifluoroacetic acid (TFA) to a final concentration of ∼0.1% (vol/vol), and peptide digests were purified using C18 solid-phase extraction (Pierce zip-tip; Thermo Fisher) according to the manufacturer's instructions and evaporated to dryness. A procedural blank was included during sample preparation and was analyzed alongside the eggshell samples.

LC-MS/MS analysis. Trypsin and elastase digests for each sample were resuspended with 25 μL 80% acetonitrile (ACN) and combined. A total of 40 μL was transferred to a 96-well plate prior to liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis. To remove the ACN, the plate was vacuum centrifuged until ∼5 μL remained. Samples were then resuspended with 8 μL 0.1% TFA 5% can, and 5 μL was then separated by an EASY-nLC 1200 (Proxeon) attached to a Q-Exactive HF-X mass spectrometer (Thermo Scientific) on a 15-cm column (75 μm inner diameter; made in-house laser pulled and packed with 1.9-μm C18 beads [Dr. Maisch] over a 77-min gradient). The parameters were the same as those already published for historical samples (99). In short, MS1: 120k resolution, maximum injection time (IT) 25 ms, scan target 3E6. MS2: 60k resolution, top 10 mode, maximum IT 118 ms, minimum scan target 3E3, normalized collision energy of 28, dynamic exclusion 20 s, and isolation window of 1.2 *m/z*. A wash-blank consisting of 0.1% TFA 5% ACN was run before and after each sample to hinder cross-contamination.

Data analysis. Bioinformatic analysis was carried out using PEAKS Studio 8.5 [Bioinformatics Solutions, Inc (100)]. After conducting preliminary screening in order to ensure uniformity across the results, the raw tandem mass spectrometry data from the three samples were combined in a single search. The thresholds for peptide and protein identification were set as follows: peptide score $-10\lg P \geq 30$, protein score $-10\lg P \geq 40$, de novo sequences scores (ALC%) ≥ 80, unique peptides ≥ 2 (threshold lowered to unique peptides ≥ 0 for sequence reconstruction so that all reference sequences could be taken into account). The NCBI database (taxonomy restricted to Aves) was used for carrying out preliminary searches, and a database including common contaminants was included (Common Repository of Adventitious Proteins: https://www.thegpm.org/crap/). Further searches were performed against the annotated lactadherin and XCA-1 sequences as described in *Annotations using B10K genomes*. The proteomics datasets have been deposited to the ProteomeXchange Consortium via the Proteomics Identifications Database (PRIDE) partner repository with the dataset identifier PXD027713. The protein structure of XCA-1 was inferred using the ColabFold AlphaFold2 notebook (101, 102) (*SI Appendix*, Fig. S3, data available in FigShare [10.6084/m9.figshare.15084879]).

Quality controls and authentication. The results of the eggshell analysis were evaluated against established checks for quality control and authenticity of ancient proteins. 1) Contamination: bleaching removes exogenous contaminants which may have been incorporated into the samples during burial and post-excavation. However, contamination can occur during protein extraction in the laboratory or during analysis. No sign of contamination was detected in the procedural blank, showing that the precautions we used during preparation were effective. 2) Carryover: blanks were analyzed between each eggshell sample and showed some evidence of carryover (46 XCA-1 peptide-spectrum matches [PSMs] in the water wash analyzed before sample PALTO 215; 100 XCA-1 and 2 lactadherin PSMs between PALTO 215 and PALTO 216; 9 XCA-1 and 12 lactadherin PSMs between PALTO 216 and PALTO 217). Given that signal reduction is at least 100-fold (more often 1,000- or 10,000-fold) between each injection (40), it is unlikely that the eggshell samples were affected by carryover. However, in order to ensure the reproducibility of the results, we repeated the injection and LC-MS/MS analysis using two leftover aliquots of samples PALTO 215 and PALTO 216, 18 mo after the first analysis. The results obtained were equivalent, that is, the top sequences identified in the two separate analyses were the same, and therefore cannot be due to carryover (all files are available within the PRIDE dataset PXD027713). 3) Extent of degradation: closed-system eggshell proteins degraded in situ exhibit a clear signal of increasing levels of degradation with increasing thermal age—including aa racemization, oxydation, water loss, and deamidation (40). UO eggshell sequences accordingly yielded clear evidence of diagenesis-induced modifications (Asn and Gln deamidation, dehydration, and pyroglutamic acid formation; oxidation of Trp and Met; and ornithine formation from Arg). For example, 45% Asn and 80% Gln were deamidated in UO XCA-1 (the low %Asn deamidation likely being due to decomposition), while for UO lactadherin, the extent of deamidation was found to be 65% for both Asn and Gln. Deamidation processes can be affected by many factors in complex, open systems (e.g., calculus), and their use for authentication is not always straightforward (103), but high extents of Gln/Asn deamidation in a closed system further support the authenticity of UO ancient sequences.

Annotations using B10K genomes. The B10K dataset comprised 363 bird species from 92% of bird families (104). After an initial analysis indicated that the ootaxon fell within Galloanseres, we added sequences for five additional Galloanseres species from NCBI using BLAST (Basic Local Alignment Search Tool). More specifically, 1) for lactadherin-MFGE8, a total of 353 species from the B10K dataset were annotated and supplemented with four additional Galloanseres genomes (*Bambusicola thoracicus* POI30273.1, *Cygnus atratus* XP_035410506.1, *Oxyura jamaicensis* XP_035191991.1, and *Aythya fuligula* XP_032051139.1); 2) for XCA, a total of 108 species from the B10K dataset were annotated, which were supplemented with data from four additional Galloanseres genomes (*Oxyura_jamaicensis* XM_035313502.1, *Aythya_fuligula* XM_032185620.1, *Cygnus_atratus* XM_035569317.1, and *Numida_meleagris* XM_021383313.1). Unfiltered sequence files are available on FigShare (10.6084/m9.figshare.15084879).

We used GeneWise v.2.4.1 (105), Exonerate v.2.2.0 (106), and SPALN v.2.3.3 (107) to align reference protein sequences of XCA-1 and lactadherin-MFGE8 to 363 bird genomes and to predict gene models. The predicted coding regions were translated into protein sequences and aligned to the reference protein by MUSCLE v.3.8.31 (108). Predicted proteins with an identity of <30% to the reference protein sequences were filtered out. A second filtering step discarded those annotations that were only supported by one method and also had identity <40% to the reference protein.

Phylogenetic analysis. The computation was performed on the National Life Science Supercomputing Center—Computerome 2.0. Commands, alignment files, and output files from model selection and phylogenetic analysis are available from FigShare (10.6084/m9.figshare.15084879). The initial analyses were done without UO to assess the phylogenetic resolution of the eggshell protein sequences. We also analyzed the corresponding DNA sequences to test whether the tree topologies from the same dataset analyzed under different substitution models were congruent. Subsequent analyses included the consensus aa sequence of UO.

Prior to alignment, we used PREQUAL v.1.02 (109) to mask stretches of sequences that did not have simple homology with other sequences. PREQUAL masked 0.3% of residues of lactadherin-MFGE8 (157,532/158,071 residues) and 2.6% of XCA-1 (13,507/13,863 residues). Masked sequences were aligned using MAFFT l-ins-i v7.453 (110, 111), and trailing ends at the beginning and the end of the alignment were trimmed manually (alignments are available on FigShare [10.6084/m9.figshare.15084879]). This resulted in an alignment of 514 sites for lactadherin-MFGE8 and 242 sites for XCA-1. DNA sequences were aligned using the codon-aware aligner MACSE v.2.05 (112), resulting in 1,479 sites for lactadherin-MFGE8 and 684 sites for XCA-1. We selected the appropriate aa or DNA substitution model for each alignment using ModelTest-NG v.0.1.3 (113). Maximum likelihood trees were inferred using RAxML-NG v.1.0.3 (114) using 10 parsimony and 10 random starting trees, and the number of bootstrap replicates were determined using bootstopping (115). The trees were inferred for both genes individually and for the two loci concatenated with Alignment Manipulation and Summary (AMAS) (116) in a partitioned analysis.

We investigated the impacts of restricting the analysis to include only 1) taxa that had sequences for both genes (107 species), 2) Palaeognathae and Galloanseres (42 species), and 3) taxa occurring in Australia and adjacent islands (93 species). In all analyses, the sequences of the target taxa were realigned, filtered, and analyzed as described in the previous paragraph of this section. The tree files are available on FigShare (10.6084/m9.figshare.15084879).

We also placed UO in a tree topology that was constrained to match accepted relationships among birds. A constraint tree was built according to the phylogenetic relationships of Galloanseres (50, 117, 118) and Palaeognathae (119). In Neoaves, we constrained the relationships between different orders as obtained from genome-wide analyses (49) but left relationships within the orders unresolved because some of these have not been sufficiently resolved. The placement of UO as a member of Neoaves was not supported in the unconstrained analyses, and constraints within Neoaves should therefore not impact the placement of UO. We used this constrained tree topology in RAxML-NG (available on FigShare [10.6084/m9.figshare.15084879]) to resolve the polytomies and to freely place UO maximizing likelihood.

Data Availability. Palaeoproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD027713 (120). Ancient DNA, phylogenetic trees, and AlphaFold models data have been deposited in FigShare (DOI: 10.6084/m9.figshare.15084879) (121). All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank two anonymous reviewers for their comments, which have helped us improve the manuscript. We also thank Dr. Jorune Sakalauskaite and Prof. Kirsty Penkman for support and useful discussion. B.D. was supported by the Italian "Ministero dell'Università e della Ricerca" (Young Researchers Program—"Rita Levi Montalcini") and M.J.C. by the Carlsberg Foundation Semper Ardens grant CF18-1110. G.M. was funded by US-NSF grant 0914821. M.B. received support from an Australian Research Council Discovery grant. J.M. was funded by the Australian Research Council and the Australian National University. M.M. is supported by Danish National Research Foundation Award PROTEIOS ("Understanding and exploiting the survival of ancient proteins") (DNRF128). We are grateful to the Science Faculty at the University of Copenhagen for free access to Computerome 2.0. We thank Prof. Jesper Velgaard Olsen at the Novo Nordisk Center for Protein Research for providing access and resources, which were also funded in part by a donation from the Novo Nordisk Foundation (Grant No. NNF14CC0001). The Pearcey Supercomputing Cluster at Commonwealth Scientific and Industrial Research Organisation was used to carry out aDNA phylogenetic analysis.

1. G. H. Miller *et al.*, Pleistocene extinction of *Genyornis newtoni*: Human impact on Australian megafauna. *Science* **283**, 205–208 (1999).
2. M. I. Bird *et al.*, Radiocarbon dating of organic- and carbonate-carbon in *Genyornis* and *Dromaius* eggshell using stepped combustion and stepped acidification. *Quat. Sci. Rev.* **22**, 1805–1812 (2003).
3. G. H. Miller, M. L. Fogel, Calibrating $\delta^{18}O$ in *Dromaius novaehollandiae* (emu) eggshell calcite as a paleo-aridity proxy for the Quaternary of Australia. *Geochim. Cosmochim. Acta* **193**, 1–13 (2016).
4. G. Miller *et al.*, Human predation contributed to the extinction of the Australian megafaunal bird *Genyornis newtoni* ~47 ka. *Nat. Commun.* **7**, 10496 (2016).
5. C. Clarkson *et al.*, Human occupation of northern Australia by 65,000 years ago. *Nature* **547**, 306–310 (2017).
6. P. F. Murray, D. Megirian, The skull of dromornithid birds: Anatomical evidence for their relationship to Anseriformes. *Rec. South Aust. Mus.* **31**, 51–97 (1998).
7. P. F. Murray, P. Vickers-Rich, *Magnificent Mihirungs: The Colossal Flightless Birds of the Australian Dreamtime* (Indiana University Press, 2004).
8. G. Mayr, Cenozoic mystery birds—On the phylogenetic affinities of bony-toothed birds (Pelagornithidae). *Zool. Scr.* **40**, 448–467 (2011).
9. T. H. Worthy, F. J. Degrange, W. D. Handley, M. S. Y. Lee, The evolution of giant flightless birds and novel phylogenetic relationships for extinct fowl (Aves, Galloanseres). *R. Soc. Open Sci.* **4**, 170975 (2017).
10. W. D. Handley, T. H. Worthy, Endocranial anatomy of the giant extinct Australian Mihirung Birds (Aves, Dromornithidae). *Diversity (Basel)* **13**, 124 (2021).
11. D. L. G. Williams, *Genyornis* eggshell (Dromornithidae; Aves) from the Late Pleistocene of South Australia. *Alcheringa* **5**, 133–140 (1981).
12. M. E. Allentoft *et al.*, Extinct New Zealand megafauna were not in decline before human colonization. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4922–4927 (2014).
13. N. J. Rawlence, A. Cooper, Youngest reported radiocarbon age of a moa (Aves: Dinornithiformes) dated from a natural site in New Zealand. *J. R. Soc. N. Z.* **43**, 100–107 (2013).
14. R. N. Holdaway *et al.*, An extremely low-density human population exterminated New Zealand moa. *Nat. Commun.* **5**, 5436 (2014).
15. R. N. Holdaway, C. Jacomb, Rapid extinction of the moas (Aves: Dinornithiformes): Model, test, and implications. *Science* **287**, 2250–2254 (2000).
16. C. L. Oskam *et al.*, Ancient DNA analyses of early archaeological sites in New Zealand reveal extreme exploitation of moa (Aves: Dinornithiformes) at all life stages. *Quat. Sci. Rev.* **52**, 41–48 (2012).
17. G. Grellet-Tinner, N. A. Spooner, T. H. Worthy, Is the "*Genyornis*" egg of a mihirung or another extinct bird from the Australian dreamtime? *Quat. Sci. Rev.* **133**, 147–164 (2016).
18. G. Grellet-Tinner, N. A. Spooner, W. D. Handley, T. H. Worthy, The *Genyornis* egg: Response to Miller et al.'s commentary on Grellet-Tinner et al., 2016. *Quat. Sci. Rev.* **161**, 128 (2017).
19. G. H. Miller, M. L. Fogel, J. W. Magee, S. J. Clarke, The *Genyornis* egg: A commentary on Grellet-Tinner et al., 2016. *Quat. Sci. Rev.* **161**, 123–127 (2017).
20. E. M. Niespolo, W. D. Sharp, G. Avery, T. E. Dawson, Early, intensive marine resource exploitation by Middle Stone Age humans at Ysterfontein 1 rockshelter, South Africa. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020042118 (2021).
21. J. Louys *et al.*, No evidence for widespread island extinctions after Pleistocene hominin arrival. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023005118 (2021).
22. S. A. Hocknull *et al.*, Extinction of eastern Sahul megafauna coincides with sustained environmental deterioration. *Nat. Commun.* **11**, 2250 (2020).
23. C. Tyler, K. Simkiss, A study of the egg shells of ratite birds. *Proc. Zool. Soc. Lond.* **133**, 201–243 (2009).
24. G. Maurer, S. J. Portugal, I. Boomer, P. Cassey, Avian embryonic development does not change the stable isotope composition of the calcite eggshell. *Reprod. Fertil. Dev.* **23**, 339–345 (2011).
25. L. J. Legendre, J. A. Clarke, Shifts in eggshell thickness are related to changes in locomotor ecology in dinosaurs. *Evolution* **75**, 1415–1430 (2021).
26. A. Mikhailov, Fossil and recent eggshell in amniotic vertebrates: Fine structure comparative morphology and classification. *Special Papers in Palaeontology* **56**, 1–76 (1997).
27. E. M. Hechenleitner, G. Grellet-Tinner, L. E. Fiorelli, What do giant titanosaur dinosaurs and modern Australasian megapodes have in common? *PeerJ* **3**, e1341 (2015).
28. R. W. R. J. Dekker, T. G. Brom, Megapode phylogeny and the interpretation of incubation strategies. *Zool. Verh.* **278**, 19–31 (1992).
29. R. Dekker, T. G. Brom, Maleo eggs and the amount of yolk in relation to different incubation strategies in megapodes. *Aust. J. Zool.* **38**, 19–24 (1990).
30. G. Grellet-Tinner, S. Lindsay, M. B. Thompson, The biomechanical, chemical and physiological adaptations of the eggs of two Australian megapodes to their nesting strategies and their implications for extinct titanosaur dinosaurs. *J. Microsc.* **267**, 237–249 (2017).
31. D. C. Deeming, Effect of composition on shape of bird eggs. *J. Avian Biol.* **49**, jav–01528 (2018).
32. M. C. Stoddard *et al.*, Avian egg shape: Form, function, and evolution. *Science* **356**, 1249–1254 (2017).
33. A. Ar, Y. Yom-Tov, The evolution of parental care in birds. *Evolution* **32**, 655–669 (1978).
34. A. Ar, H. Rahn, Pores in avian eggshells: Gas conductance, gas exchange and embryonic growth rate. *Respir. Physiol.* **61**, 1–20 (1985).
35. J. M. Starck, E. Sutter, Patterns of growth and heterochrony in moundbuilders (Megapodiidae) and fowl (Phasianidae). *J. Avian Biol.* **31**, 527–547 (2000).
36. R. B. Harris, S. M. Birks, A. D. Leaché, Incubator birds: Biogeographical origins and evolution of underground nesting in megapodes (Galliformes: Megapodiidae). *J. Biogeogr.* **41**, 2045–2056 (2014).
37. R. S. Seymour, D. Vleck, C. M. Vleck, Gas exchange in the incubation mounds of megapode birds. *J. Comp. Physiol. B* **156**, 773–782 (1986).
38. D. N. Jones, W. R. J. Dekker, C. S. Roselaar, *The Megapodes: Megapodiidae* (Oxford University Press, 1995).
39. D. T. Booth, D. N. Jones, "Underground nesting in the megapodes" in *Avian Incubation: Behaviour, Environment, and Evolution*, D. C. Deeming, Ed. (Oxford University Press, 2002), pp. 192–206.
40. B. Demarchi *et al.*, Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**, e17092 (2016).
41. A. Grealy *et al.*, Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Mol. Phylogenet. Evol.* **109**, 151–163 (2017).
42. T. van der Valk *et al.*, Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).
43. M. A. Smith, G. Miller, G. F. Van Tets, Burnt ratite eggshell from Pleistocene aeolian sediments. *CAVEPS 1993. Rec. South Aust. Mus.* **27**, 228 (1994).
44. M. Crisp *et al.*, Isolation of the intra-crystalline proteins and kinetic studies in *Struthio camelus* (ostrich) eggshell for amino acid geochronology. *Quat. Geochronol.* **16**, 110–128 (2013).
45. L. Stapane, N. Le Roy, M. T. Hincke, J. Gautron, The glycoproteins EDIL3 and MFGE8 regulate vesicle-mediated eggshell calcification in a new model for avian biomineralization. *J. Biol. Chem.* **294**, 14526–14545 (2019).
46. N. Le Roy, L. Stapane, J. Gautron, M. T. Hincke, Evolution of the avian eggshell biomineralization protein toolkit—New insights from multi-omics. *Front. Genet.* **12**, 672433 (2021).
47. C. Sun, G. Xu, N. Yang, Differential label-free quantitative proteomic analysis of avian eggshell matrix and uterine fluid proteins associated with eggshell mechanical property. *Proteomics* **13**, 3523–3536 (2013).
48. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. E. D. Jarvis *et al.*, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
50. R. O. Prum *et al.*, A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573 (2015).
51. A. Suh, The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool. Scr.* **45**, 50–62 (2016).
52. H. Kuhl *et al.*, An unbiased molecular approach using 3′-UTRs resolves the avian family-level tree of life. *Mol. Biol. Evol.* **38**, 108–127 (2021).
53. E. Cappellini *et al.*, Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature* **574**, 103–107 (2019).
54. F. Welker *et al.*, The dental proteome of *Homo antecessor*. *Nature* **580**, 235–238 (2020).
55. F. Welker *et al.*, Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature* **576**, 262–265 (2019).
56. J. P. Townsend, F. López-Giráldez, R. Friedman, The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. *J. Mol. Evol.* **67**, 437–447 (2008).
57. S. Reddy *et al.*, Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst. Biol.* **66**, 857–879 (2017).
58. E. L. Braun, R. T. Kimball, Data types and the phylogeny of neoaves. *Birds North America* **2**, 1–22 (2021).

59. V. L. Harvey, J. N. Keating, M. Buckley, Phylogenetic analyses of ray-finned fishes (Actinopterygii) using collagen type I protein sequences. *R. Soc. Open Sci.* **8**, 201955 (2021).
60. T. B. Sackton *et al.*, Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
61. J. Stewart *et al.*, Walking on eggshells: A study of egg use in Anglo-Scandinavian york based on eggshell identification Using ZooMS. *Int. J. Osteoarchaeol.* **24**, 247–255 (2014).
62. S. Presslee *et al.*, The identification of archaeological eggshell using peptide markers. *Sci. Technol. Archaeol. Res.* **3**, 89–99 (2017).
63. B. Demarchi *et al.*, The role of birds at Çatalhöyük revealed by the analysis of eggshell. *Quat. Int.* **543**, 50–60 (2020).
64. B. Demarchi *et al.*, Birds of prey and humans in prehistoric Europe: A view from El Mirón Cave, Cantabria (Spain). *J. Archaeol. Sci. Rep.* **24**, 244–252 (2019).
65. T. Jonuks *et al.*, Multi-method analysis of avian eggs as grave goods: Revealing symbolism in conversion period burials at Kukruse, NE Estonia. *Environ. Archaeol.* **23**, 109–122 (2018).
66. B. Collins, T. E. Steele, An often overlooked resource: Ostrich (*Struthio* spp.) eggshell in the archaeological record. *J. Archaeol. Sci. Rep.* **13**, 121–131 (2017).
67. F. d'Errico *et al.*, Trajectories of cultural innovation from the Middle to Later Stone Age in Eastern Africa: Personal ornaments, bone artifacts, and ocher from Panga ya Saidi, Kenya. *J. Hum. Evol.* **141**, 102737 (2020).
68. B. A. Stewart *et al.*, Ostrich eggshell bead strontium isotopes reveal persistent macroscale social networking across late Quaternary southern Africa. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6453–6462 (2020).
69. P.-J. Texier *et al.*, From the cover: A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6180–6185 (2010).
70. L. Janz, R. G. Elston, G. S. Burr, Dating North Asian surface assemblages with ostrich eggshell: Implications for palaeoecology and extirpation. *J. Archaeol. Sci.* **36**, 1982–1989 (2009).
71. D. T. Potts, Ostrich distribution and exploitation in the Arabian peninsula. *Antiquity* **75**, 182–190 (2001).
72. A. Pitarch Martí, Y. Wei, X. Gao, F. Chen, F. d'Errico, The earliest evidence of coloured ornaments in China: The ochred ostrich eggshell beads from Shuidonggou Locality 2. *J. Anthropol. Archaeol.* **48**, 102–113 (2017).
73. G. L. Badam, A note on the ostrich in India since the Miocene. *Man and Environment* **30**, 97–104 (2005).
74. J. Blinkhorn, H. Achyuthan, M. D. Petraglia, Ostrich expansion into India during the Late Pleistocene: Implications for continental dispersal corridors. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **417**, 80–90 (2015).
75. G. Maurer, S. J. Portugal, P. Cassey, A comparison of indices and measured values of eggshell thickness of different shell regions using museum eggs of 230 European bird species. *Ibis* **154**, 714–724 (2012).
76. J.-Y. Juang *et al.*, The avian egg exhibits general allometric invariances in mechanical design. *Sci. Rep.* **7**, 14205 (2017).
77. D. F. Hoyt, Practical methods of estimating volume and fresh weight of bird eggs. *Auk* **96**, 73–77 (1979).
78. R. R. Sokal, F. James Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research* (WH Freeman and Company, New York, 1995).
79. R core team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, 2021).
80. D. Rubolini, A. Liker, L. Z. Garamszegi, A. P. Møller, N. Saino, Using the BirdTree.org website to obtain robust phylogenies for avian comparative studies: A primer. *Curr. Zool.* **61**, 959–965 (2015).
81. W. Jetz *et al.*, Global distribution and conservation of evolutionary distinctness in birds. *Curr. Biol.* **24**, 919–930 (2014).
82. S. J. Hackett *et al.*, A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008).
83. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
84. M. R. E. Symonds, S. P. Blomberg, "A primer on phylogenetic generalised least squares" in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, L. Z. Garamszegi, Ed. (Springer Berlin Heidelberg, 2014), pp. 105–130.
85. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar; R Core Team, nlme: Linear and nonlinear mixed effects models. R package version 3.1-152. https://cran.r-project.org/web/packages/nlme/nlme.pdf. Accessed 18 January 2022.
86. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
87. A. Ives, D. Li, rr2: An R package to calculate R^2s for regression models. *J. Open Source Softw.* **3**, 1028 (2018).
88. A. R. Ives, R^{2}s for correlated data: Phylogenetic models, LMMs, and GLMMs. *Syst. Biol.* **68**, 234–251 (2019).
89. R. Mundry, "Statistical issues and assumptions of phylogenetic generalized least squares" in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, L. Z. Garamszegi, Ed. (Springer Berlin Heidelberg, 2014), pp. 131–153.
90. J. Dabney *et al.*, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
91. A. Cooper, Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthisittidae). *Experientia* **50**, 558–563 (1994).
92. A. Grealy, N. E. Langmore, L. Joseph, C. E. Holleley, Genetic barcoding of museum eggshell improves data integrity of avian biological collections. *Sci. Rep.* **11**, 1605 (2021).
93. M.-T. Gansauge *et al.*, Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).
94. A. Grealy, M. Bunce, C. E. Holleley, Avian mitochondrial genomes retrieved from museum eggshell. *Mol. Ecol. Resour.* **19**, 1052–1062 (2019).
95. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
96. 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).
97. 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).
98. J. Hendy *et al.*, A guide to ancient protein studies. *Nat. Ecol. Evol.* **2**, 791–799 (2018).
99. M. Mackie *et al.*, Palaeoproteomic profiling of conservation layers on a 14th Century Italian Wall Painting. *Angew. Chem. Int. Ed. Engl.* **57**, 7369–7374 (2018).
100. J. Zhang *et al.*, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).
101. M. Steinegger, S. Ovchinnikov, M. Mirdita, Data from "AlphaFold2 w/MMseqs2." Google Colaboratory. https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb. Accessed 28 July 2021.
102. R. R. Ruiz-Arellano, F. J. Medrano, A. Moreno, A. Romero, Structure of struthiocalcin-1, an intramineral protein from *Struthio camelus* eggshell, in two crystal forms. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 809–818 (2015).
103. A. Ramsøe *et al.*, Assessing the degradation of ancient milk proteins through site-specific deamidation patterns. *Sci. Rep.* **11**, 7795 (2021).
104. S. Feng *et al.*, Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
105. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
106. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
107. H. Iwata, O. Gotoh, Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* **40**, e161 (2012).
108. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
109. S. Whelan, I. Irisarri, F. Burki, PREQUAL: Detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).
110. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
111. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
112. V. Ranwez, S. Harispe, F. Delsuc, E. J. P. Douzery, MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One* **6**, e22594 (2011).
113. D. Darriba *et al.*, ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
114. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
115. N. D. Pattengale, M. Alipour, O. R. P. Bininda-Emonds, B. M. E. Moret, A. Stamatakis, "How many bootstrap replicates are necessary?" in *Research in Computational Molecular Biology* (Springer Berlin Heidelberg, 2009), pp. 184–200.
116. M. L. Borowiec, AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660 (2016).
117. Z. Sun *et al.*, Rapid and recent diversification patterns in Anseriformes birds: Inferred from molecular phylogeny and diversification analyses. *PLoS One* **12**, e0184529 (2017).
118. N. Wang, R. T. Kimball, E. L. Braun, B. Liang, Z. Zhang, Assessing phylogenetic relationships among galliformes: A multigene phylogeny with expanded taxon sampling in Phasianidae. *PLoS One* **8**, e64312 (2013).
119. A. Cloutier *et al.*, Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. *Syst. Biol.* **68**, 937–955 (2019).
120. B. Demarchi *et al.*, Ancient proteins resolve controversy over the identity of Genyornis eggshell. PRIDE. https://www.ebi.ac.uk/pride/archive/projects/PXD027713. Deposited 5 August 2021.
121. B. Demarchi *et al.*, Ancient proteins resolve controversy over the identity of Genyornis eggshell. FigShare. 10.6084/m9.figshare.15084879. Deposited 30 July 2021.
122. G. H. Miller *et al.*, Ecosystem collapse in Pleistocene Australia and a human role in megafaunal extinction. *Science* **309**, 287–290 (2005).