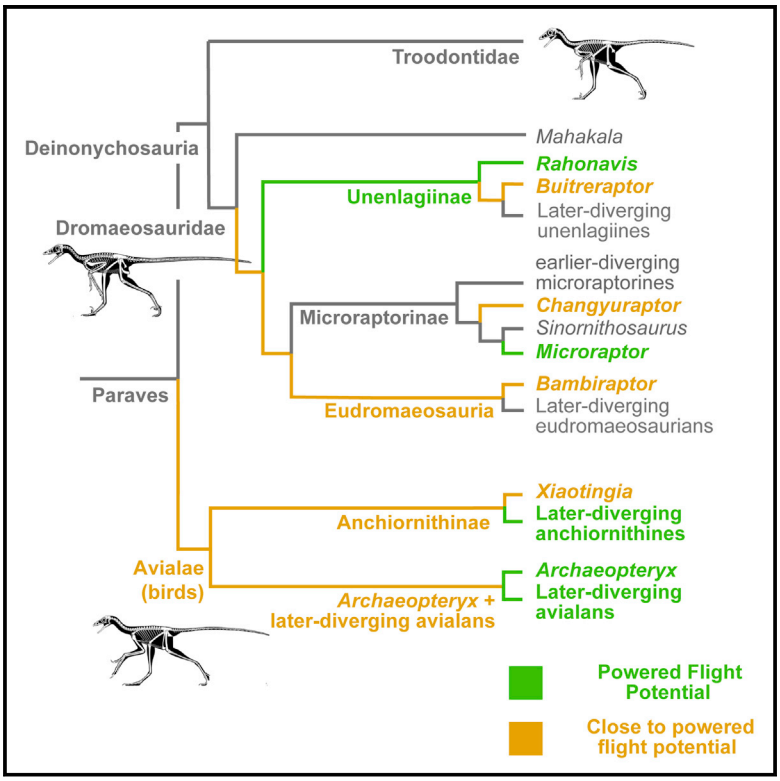
Article

Current Biology

Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds

Graphical Abstract



Highlights

- Support Deinonychosauria as sister taxon to birds and Anchiornithinae as early birds
- Powered flight potential evolved $\geq$3 times: once in birds and twice in dromaeosaurids
- Many ancestors of bird relatives neared thresholds of powered flight potential
- Broad experimentation with wing-assisted locomotion before theropod flight evolved

Authors

Rui Pei, Michael Pittman, Pablo A. Goloboff, ..., Mark A. Norell, Stephen L. Brusatte, Xing Xu

Correspondence

mpittman@hku.hk

In Brief

Pei et al. use an updated phylogeny of early birds and their closest relatives to reconstruct powered flight potential, showing it evolved at least three times. Many ancestors of the closest bird relatives neared thresholds of powered flight potential, suggesting broad experimentation with wing-assisted locomotion before theropod flight evolved.

Pei et al., 2020, Current Biology *30*, 1–14
October 5, 2020
https://doi.org/10.1016/j.cub.2020.06.105

CellPress

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

Current Biology

CellPress
OPEN ACCESS

Article

Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds

Rui Pei,[1,10] Michael Pittman,[2,10,11,*] Pablo A. Goloboff,[3] T. Alexander Dececchi,[4] Michael B. Habib,[5] Thomas G. Kaye,[6] Hans C.E. Larsson,[7] Mark A. Norell,[8] Stephen L. Brusatte,[9] and Xing Xu[1]

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology & Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China
[2]Vertebrate Palaeontology Laboratory, Division of Earth and Planetary Science, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[3]Unidad Ejecutora Lillo, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Tucumán, Tucumán, Argentina
[4]Division of Natural Sciences, Mount Marty College, Yankton, SD 57078, USA
[5]Dinosaur Institute, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, USA
[6]Foundation for Scientific Advancement, Sierra Vista, AZ 85650, USA
[7]Redpath Museum, McGill University, Montréal, QC H3A 0C4, Canada
[8]Division of Paleontology, American Museum of Natural History, New York City, NY 10024, USA
[9]School of Geosciences, University of Edinburgh, Edinburgh EH9 3FE, UK
[10]These authors contributed equally
[11]Lead Contact
*Correspondence: mpittman@hku.hk
https://doi.org/10.1016/j.cub.2020.06.105

SUMMARY

Uncertainties in the phylogeny of birds (Avialae) and their closest relatives have impeded deeper understanding of early theropod flight. To help address this, we produced an updated evolutionary hypothesis through an automated analysis of the Theropod Working Group (TWiG) coelurosaurian phylogenetic data matrix. Our larger, more resolved, and better-evaluated TWiG-based hypothesis supports the grouping of dromaeosaurids + troodontids (Deinonychosauria) as the sister taxon to birds (Paraves) and the recovery of Anchiornithinae as the earliest diverging birds. Although the phylogeny will continue developing, our current results provide a pertinent opportunity to evaluate what we know about early theropod flight. With our results and available data for vaned feathered pennaraptorans, we estimate the potential for powered flight among early birds and their closest relatives. We did this by using an ancestral state reconstruction analysis calculating maximum and minimum estimates of two proxies of powered flight potential—wing loading and specific lift. These results confirm powered flight potential in early birds but its rarity among the ancestors of the closest avialan relatives (select unenlagiine and microraptorine dromaeosaurids). For the first time, we find a broad range of these ancestors neared the wing loading and specific lift thresholds indicative of powered flight potential. This suggests there was greater experimentation with wing-assisted locomotion before theropod flight evolved than previously appreciated. This study adds invaluable support for multiple origins of powered flight potential in theropods ($\geq$3 times), which we now know was from ancestors already nearing associated thresholds, and provides a framework for its further study.

INTRODUCTION

The origin of birds (Avialae) and modern powered flapping flight were iconic events in the history of life. Recent studies of early birds and their closest dinosaurian relatives (non-avialan paravian theropods) have provided key insights into this major evolutionary transition. It is now clear that anatomies and behaviors traditionally associated with birds were first acquired by non-avialan dinosaurs before the origin of birds and modern powered flapping flight. These include smaller body size, accelerated evolutionary rates [1–3], early feathers of "modern" aspect [4–17], complex plumage coloration, flapping-based locomotion, non-powered flight capabilities (among some non-avialan paravians; M. Habib, et al., 2012, Geol. Assoc. Can., abstract) [18–22, 23], and even an avian-like sleeping posture [24]. As a result of these advances, the origin of birds has emerged as one of the best documented examples of a major macroevolutionary transition. Despite the extensive array of new specimens and data, the phylogenetic relationships within and between paravian clades have been challenging to reconstruct, and uncertainties continue to present obstacles toward reaching consensus and improving resolution at important nodes.

Traditionally, dromaeosaurids and troodontids were united together as the Deinonychosauria by the "sickle claw" of their second toe and other characters [2, 25–28]. They were considered the sister group of birds and were altogether known as the Paraves [2, 25–28]. The rapid discovery of paravian species over the last decade [13, 14, 17, 28–32], especially from East

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

Asia, has called this into question. Most notably, many of the historically diagnostic features of Dromaeosauridae, Troodontidae, and Deinonychosauria are now recognized as synapomorphies of more inclusive theropod groups (e.g., Maniraptora) or in some cases appear to have been acquired convergently in different taxa. The number of evolutionary hypotheses has grown with these new fossil discoveries [1, 2, 4, 13, 14, 28, 30, 32–37]. They now encompass a range of possible interrelationships between birds and other paravians, even challenging the monophyly of Deinonychosauria and the composition of stem avialans [4, 13, 33–35, 38]. A primary issue concerns troodontids, which have been grouped with either dromaeosaurids [2, 4, 28, 32, 33] as the traditional Deinonychosauria or with Avialae [13, 14, 34, 35] exclusive of dromaeosaurids. Each phylogenetic hypothesis has different implications on the origin of birds and the morphological, biomechanical, and ecological states of their transitional antecedents.

We contribute toward addressing these prevailing phylogenetic issues by presenting an updated parsimony-based reconstruction of paravian interrelationships by using the large coelurosaurian theropod phylogenetic dataset of the Theropod Working Group (TWiG), produced by a long-standing consortium of international theropod experts [1] (and references therein). Updated with recently discovered taxa, our expanded version of the TWiG dataset includes nine new dromaeosaurid terminal taxa (*Acheroraptor*, *Changyuraptor*, *Dakotaraptor*, IVPP V22530, *Linheraptor*, *Luanchuanraptor*, *Velociraptor osmolskae*, *Yurgovuchia*, and *Zhenyuanlong*); among the largest number of dromaeosaurids (31) included in a phylogenetic analysis so far. It also incorporates a wealth of new data from existing paravians that have been recently described in more detail [15–17, 28, 30–32, 35, 39] and studied first-hand, including the key early-diverging paravians *Anchiornis* and *Archaeopteryx* [40]. For additional details, see "Phylogenetic dataset" in the STAR Methods section.

Paleontological datasets often pose challenges to phylogenetic analysis, especially in leading to taxa that can be placed equally well in distant parts of the tree ("wildcards"), typically as a consequence of missing entries from incomplete preservation. Further challenges might emerge from high degrees of morphological variation, manifesting as homoplasies, in densely sampled phylogenetic regions. Both of these confounding issues are expected in bird origin studies. One of the key goals of this study is to provide accurate phylogenetic placement for as many paravians as possible, but several are missing over 90% of their scoring entries (e.g., the dromaeosaurids *Atrociraptor* and *Shanag*). We have therefore placed emphasis in developing a pipeline of analysis that can automatically deal with the numerous wildcards commonly seen in paleontological datasets, including some newly developed techniques. The steps in a phylogenetic analysis subsequent to finding the optimal trees are prone to human error, particularly when wildcards are involved (which, with many of the steps being sequential, could easily carry over to subsequent steps). Thus, to minimize the risk of human error, we used scripts to automate all analytical procedures, in a way that was reproducible and appropriate for other paleontological datasets. An important benefit of this automated analysis is that it encourages the dataset to be rechecked and corrected for scoring errors/problems, as many more analytical iterations can be made in the same time as a handful of manual analyses would take. We hope this automated analytical pipeline can increase our community's access to more in-depth parsimony-based analyses.

We use our more resolved and better evaluated TWiG-based phylogeny to infer when and how the potential for powered flight developed in early birds and their closest relatives. Previous work has proposed that powered theropod flight evolved once or maybe even multiple times [18, 41, 42]. It has even been suggested that birds should be defined by the possession of flight alone, as an apomorphic feature [43]. With the phylogenetic placement of the iconic early bird *Archaeopteryx* with Deinonychosauria, a single origin of powered flight has been proposed at Paraves, polarizing the evolution of proportionally longer and more robust arms at that node [4]. However, the wing and body dimensions of many early-diverging paravians do not surpass the minimal thresholds for flight ability [18] as defined in modern birds and other taxa [44–46]. A quantitative study found that non-volant flapping-based locomotion was confined to Paraves: flap running, wing-assisted incline running (WAIR), and wing-assisted leaping [18]. In showing that this was optimized at Paraves and that significant capabilities were derived independently in microraptorine dromaeosaurids and avialans [18], that study supported the potential for multiple origins of powered flight in theropods. However, that study was restricted by a problematic phylogeny and small taxon sample, but more importantly, it did not focus on testing taxa and lineages against known minimal thresholds for flight ability in modern birds [44–46].

We have overcome these restrictions by using our larger, more-resolved, and better evaluated TWiG-based tree topology across 43 taxa sharing lift-compatible vaned feathers (Pennaraptora; but see [47]) to provide maximum and minimum estimates of wing loading and specific lift in the ancestors of our study taxa by using ancestral state reconstruction analysis. These provide a proxy of the potential for powered flight through the transition from non-flying to flying theropods. These parameters are estimated from morphological features measurable from the fossils and are commonly used to evaluate flight capability in extant avians [44–46].

Wing loading is a major determinant of minimum flight speed, required launch and landing speeds, and maneuverability in powered flyers [48]. Wing loading is also a major determinant of required flapping frequency in powered flyers (wings must be flapped faster if they are smaller and/or body size is greater) [49].

In powered flyers, specific lift is critical to weight support and generation of thrust (thrust is primarily a component of lift in vertebrate flapping flyers) [50]. Although small powered flyers with sophisticated wing kinematics (particularly during the upstroke phase) [51] can generate some drag-based thrust, we consider this to have been unlikely in early-diverging birds because of their less-refined aerofoils and motion control. Three major types of biomechanical competency are assessed by wing loading and specific lift that allows us to determine whether early birds and their closest relatives had the potential for powered flight.

Anatomical Requirements for Flight

Takeoff in flying animals is initiated by leaping [52], so the primary anatomical requirements to initiate launch in paravians are

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

related to hind-limb characteristics [18]. Because of the terrestrial ancestry of theropod dinosaurs, large hind-limb muscle mass and robust hind-limb skeletal elements were plesiomorphic for paravians, so all of our study taxa inherited sufficient hind-limb strength for leaping [18]. For powered flapping flight (i.e., after takeoff), the primary anatomical requirements are summarized by wing loading, which simultaneously includes potential lift-producing surface area and body weight in one single variable. A key assumption we make is that our fossil taxa had body densities within the range known for living birds. We constrained the estimated body mass range of taxa potentially capable of powered flight by assuming that they were roughly similar in mass to living birds with similar wingspans and body volumes. This gave us a narrower set of body mass estimates within the relative large confidence intervals around the regressions used to estimate body mass [53].

Aerodynamic Force Production Requirements for Flight

Early in theropod flight evolution, the lift:drag ratios of wings were not necessarily equivalent to those of modern birds [20, 48, 54]. However, morphospace comparisons of wing shape show significant overlap between early taxa and modern ones [55]. There is good evidence that the wings of non-avialan taxa were capable of maximum lift coefficients broadly similar to those of living birds, and so we model them as such. In particular, fossil taxa also possessed similar leading edge shapes with well-developed propatagia [56–58], and the range of wing shapes lies within the overall aspect ratio morphospace of living birds. Given the heavy feather layering and aerodynamically symmetric feather vanes in the wings of some non-avialan taxa (e.g., *Anchiornis* [59]), we consider it unlikely that these taxa utilized slotted wingtips. To account for this, we model the fossil taxa as having non-slotted wingtips (i.e., anatomical aspect ratio and aerodynamic aspect ratio taken as equivalent). The long bone cross-sections in the forelimbs of early birds and microraptorine dromaeosaurids have similar shapes and comparable bending strengths to those of living birds [60]. Analysis of feather stiffness [61] and vane asymmetry ratios [62] demonstrates that the feathers of early paravians might have been less competent as individual airfoils than the primary feathers of living birds (but see [63]). This might have limited earlier taxa to the use of unslotted wings. Furthermore, some questions remain regarding the upstroke kinematics available to early paravians [18, 64, 65]. Taken together, these data indicate that early paravians were capable of similar aerodynamic force production to that seen during steady-state conditions in living birds, excluding the use of slotted wing tips. In quantitative terms, these data suggest that lift coefficients up to 1.6 (typical steady-state maximum for living birds) were possible, but the larger lift coefficients sometimes achieved by living taxa by using dynamic stall and similar unsteady mechanisms (as high as 5.3—see Norberg [66]) might not have been possible. Wing loading and specific lift estimates of fossil theropods that pass value thresholds characterizing all volant modern birds therefore indicate a potential for powered flight.

Physiological Requirements for Flight

Our estimates of specific lift utilize a range of potential muscle power available to our theropod taxa to reflect prevailing uncertainty in this parameter. We assume that at least some anaerobic power was available for climb out after takeoff, and we have included this in our estimates, but we have kept the estimates of this anaerobic fraction conservative (see Supplemental Information). The specific lift estimates also take into account the likely limitations on the maximum coefficient of lift in early taxa mentioned above.

Our Approach

Estimates of wing loading and specific lift were calculated from reconstructed ancestral morphologies by using our own direct measurements of specimens as well as parameters reported in the literature. This allowed us to identify ancestors that fall within the range seen in extant volant birds, which we consider more accurate than just mapping diagnostic features or single metrics of flight capability in isolation. To consider parametric differences in past studies and differences from ongoing uncertainties in paravian anatomy (see STAR Methods), we calculated a maximum and minimum estimate for wing loading and specific lift. These estimates bracket the range of calculation permutations currently available, producing the most conservative results currently possible (for additional information, see STAR Methods and Methods S1G–S1J). Our approach contrasts with the concept of body weight support examined in [18], as we wanted to avoid using poorly known behavioral capabilities (e.g., flapping speed, flap angle, and running speed) in our calculations in order to maximize the precision of our model. We interpret our results in the context of the osteological and feather anatomy changes recovered by our updated phylogenetic analysis to provide a detailed account of when and how powered flight evolved.

RESULTS AND DISCUSSION

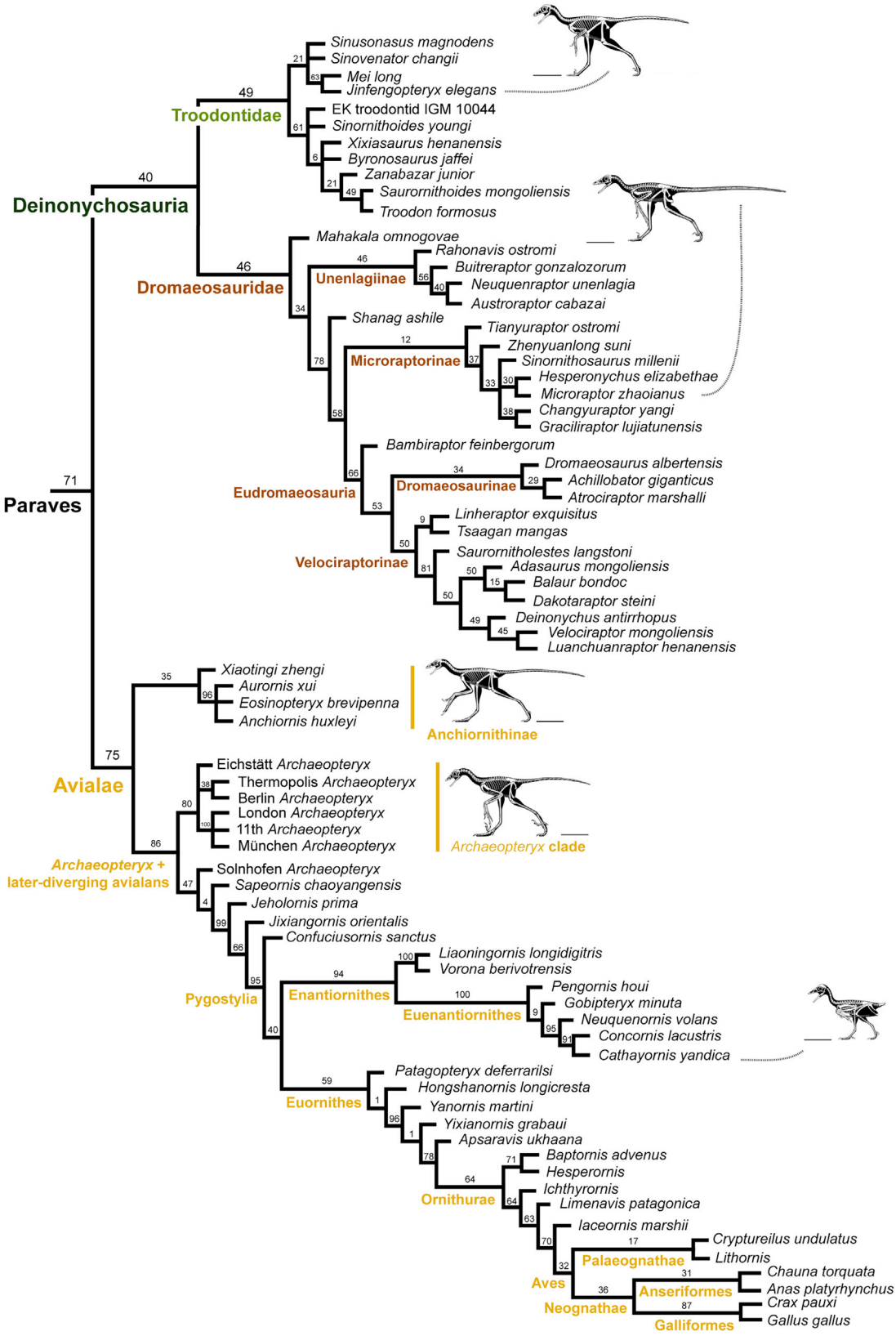
Paravian Phylogeny

Our TWiG-based phylogenetic results help to confirm important details recovered in previous non-TWiG and TWiG-based studies and show interrelationships for taxa that have never been evaluated in a phylogenetic analysis. All extended implied weighting (XIW) and equal weighting (EW) topologies support the monophyly of each of the traditionally recognized paravian clades: Paraves comprises Deinonychosauria and Avialae as in [2, 25–28] and Deinonychosauria comprises Dromaeosauridae and Troodontidae as in [2, 4, 28, 32, 33] (Figure 1; Methods S1A–S1D; see "Character weighting" in STAR Methods).

Dromaeosaurid interrelationships are significantly improved relative to previous TWiG-based studies with better supported internal resolution. Four anchiornithine taxa scattered throughout Paraves in previous TWiG analyses [1, 2] are gathered into a distinct clade of early-diverging avialans (Anchiornithinae) as in [34, 40, 67] and, in part, [4] (Figure 1; Methods S1A–S1D). Our results generate a revised list of evolutionary synapomorphies for the major paravian clades, including a refined sequence of evolutionary changes at the base of Avialae (Figure 1). A succinct description of the results is given here with additional details provided in Methods S1 (see "Additional description of results, including previously unreported synapomorphies of select paravian clades").

Despite there being some differences in how paravian interrelationships are recovered between this study and previous ones

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105



(legend on next page)

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

[1, 2, 4, 13, 14, 28, 30, 32–37], some of these past studies share the same synapomorphies for Paraves, making them useful traits for identifying members of the clade. A laterally facing glenoid fossa is an especially useful trait for identification (character [char.] 136.1 in this study), as it was recovered at the equivalent node in two recent TWiG studies (char. 138.1 in Turner et al. [2] and char. 136.1 in Brusatte et al. [1]) and is related to the extension of the glenoid floor onto the external surface of the scapula, which is a paravian synapomorphy in the non-TWiG-based studies of Agnolin and Novas [67] (char. 138.1 in [67]), Senter et al. [28] (char. 216.1 in [28]), and Xu et al. [28] (char. 122.1 in [33]) as well as a (*Jinfengopteryx* + Paraves) synapomorphy of Foth et al. [34] (char. 133.1 in [34]).

The basic dromaeosaurid topology is maintained in this study, with unenlagiines and microraptorines at earlier diverging positions (*contra* [67]) and eudromaeosaurians at later diverging ones [1, 2, 15, 17, 28, 30, 32]. Regardless of the character weighting employed, *Mahakala* is recovered as the earliest diverging dromaeosaurid followed by the Unenlagiinae (*Rahonavis* sister to *Buitreraptor* and (*Neuquenraptor* + *Austroraptor*)) and then *Shanag*; the latter taxon is sister to the clade (Microraptorinae + Eudromaeosauria). Several recent phylogenetic analyses have had difficulties resolving microraptorine interrelationships with the TWiG-based studies of Brusatte et al. [1], Turner et al. [2], and Lü and Brusatte [17] recovering polytomies in their reduced strict consensus trees (Figure 64 of [2]; Figure S2 of [1]; Figure 5 of [17]). Only the non-TWiG-based topologies of Senter et al. [28] and Xu et al. [33] were resolved. As originally suggested by Senter et al. [28], we recover the largest microraptorine, *Tianyuraptor*, as the earliest diverging member of Microraptorinae under both XIW and EW results but contrary to its previously suggested eudromaeosaurian affinity [32]. For the first time, we recover another relatively large-bodied microraptorine, *Zhenyuanlong*, as the next earliest diverging microraptorine, which previously had an unresolved phylogenetic position within a polytomy of Liaoning dromaeosaurids [17]. *Tianyuraptor* and *Zhenyuanlong* lack both a characteristic tubercle along the lateral edge of the mid-shaft of the posteriorly curved pubis and the presence of a subarctometatarsalian foot, unlike other microraptorines (chars. 228.0 and 200.1). However, *Tianyuraptor* is unique among microraptorines in having a straight pubis (char. 177.0), resembling other non-unenlagiine and non-velociraptorine dromaeosaurids. Other smaller, later-diverging microraptorines (char. 200.1) are recovered as a polytomy in both the XIW and EW strict consensus trees. As expected, the later-diverging microraptorines include the Early Cretaceous Chinese dromaeosaurids *Changyuraptor* and IVPP V22530 [15, 31].

Bambiraptor was recovered as the earliest diverging eudromaeosaurian in all XIW results, whereas, in all EW results, *Bambiraptor* is the second earliest diverging eudromaeosaurian after *Saurornitholestes.* A relatively early-diverging position for *Bambiraptor* within Eudromaeosauria was recovered by the non-TWiG studies of Agnolín and Novas [67], Senter et al. [28], and DePalma et al. [32], and both *Bambiraptor* and *Saurornitholestes* were nested within Velociraptorinae in the TWiG studies of Turner et al. [2] and Brusatte et al. [1]. Under both equal and differential weighting, the remaining traditionally identified eudromaeosaurians do not have resolved interrelationships in the strict consensus tree except for *Linheraptor* and *Tsaagan*, which are recovered in a sister relationship as expected [39]. Pruning *Yurgovuchia*, *Acheroraptor*, *V. osmolskae*, and *Utahraptor* from the eudromaeosaurian polytomy in both the XIW and EW strict consensus tree reveals a much more resolved topology, with a monophyletic Dromaeosaurinae and Velociraptorinae (see "Additional description of results, including previously unreported synapomorphies of select paravian clades" in Methods S1).

Our results under XIW and EW fail to recover Anchiornithinae as part of Troodontidae (*contra* [1, 2]). The remaining troodontids were recovered in at least two clades in recent TWiG studies (Figure S2 of Brusatte et al. [1] and Figure 57 of [2], but their composition differs; see "Additional description of results, including previously unreported synapomorphies of select paravian clades" in Methods S1).

In the XIW and EW topologies, the anchiornithine paravians from northeastern China—*Anchiornis*, *Aurornis*, *Eosinopteryx*, and *Xiaotingia*—were recovered as the earliest diverging avialan clade (earlier diverging than *Archaeopteryx* and later-diverging avialans) as in [34, 40, 67] and, in part, [4]. The avialan node recovered in this study is shared by the anchiornithines and traditional avialans (we define Avialae as a stem-based taxon after [2]).

Group support calculated by using symmetric resampling [68] was generally very low within Paraves, indicating that nodes are supported with relatively high levels of character conflict (Methods S1E and S1F). Disregarding the most unstable taxa, reasonable support values between 76 and 100 were found for the rest of the tree, including Paraves, Troodontidae, Dromaeosauridae, Unenlagiinae, and Eudromaeosauria (Methods S1E and S1F). In other datasets [1, 2, 28, 33, 34], the nodal supports of paravian nodes are generally low, and Bremer supports are typically between 1 and 2 for most paravian nodes (but see Xu et al. [33]). See "Group supports and conflict" in the STAR Methods for more information. See Pol and Goloboff [69] for a more extended discussion of coelurosaurian group support.

Major Types of Biomechanical Competency for Flight

Our expanded, more resolved, and better evaluated TWiG-based phylogeny provides the opportunity to assess the three major types of biomechanical competency for theropod flight—anatomical, aerodynamic force production, and physiological. This allows us to identify which paravian ancestors had

Figure 1. Revised Paravian Phylogeny

Reduced strict consensus tree showing topology common to analyses using extended implied character weighting and equal character weighting. Group support values are the lowest ones recovered under either character weighting scheme and under any of the combinations of pruned taxa (see Methods S1E for the complete set of support values and "Group supports and conflict" in the STAR Methods for more information). See Results and Discussion section for synapomorphies shared with previous studies. For the details about previously unreported synapomorphies, see "Additional description of results, including previously unreported synapomorphies of select paravian clades" in Methods S1. See STAR Methods for the details of the phylogenetic analysis. Skeletal reconstructions (scale bar represents 10 cm) are used with the permission of Scott A. Hartman.

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

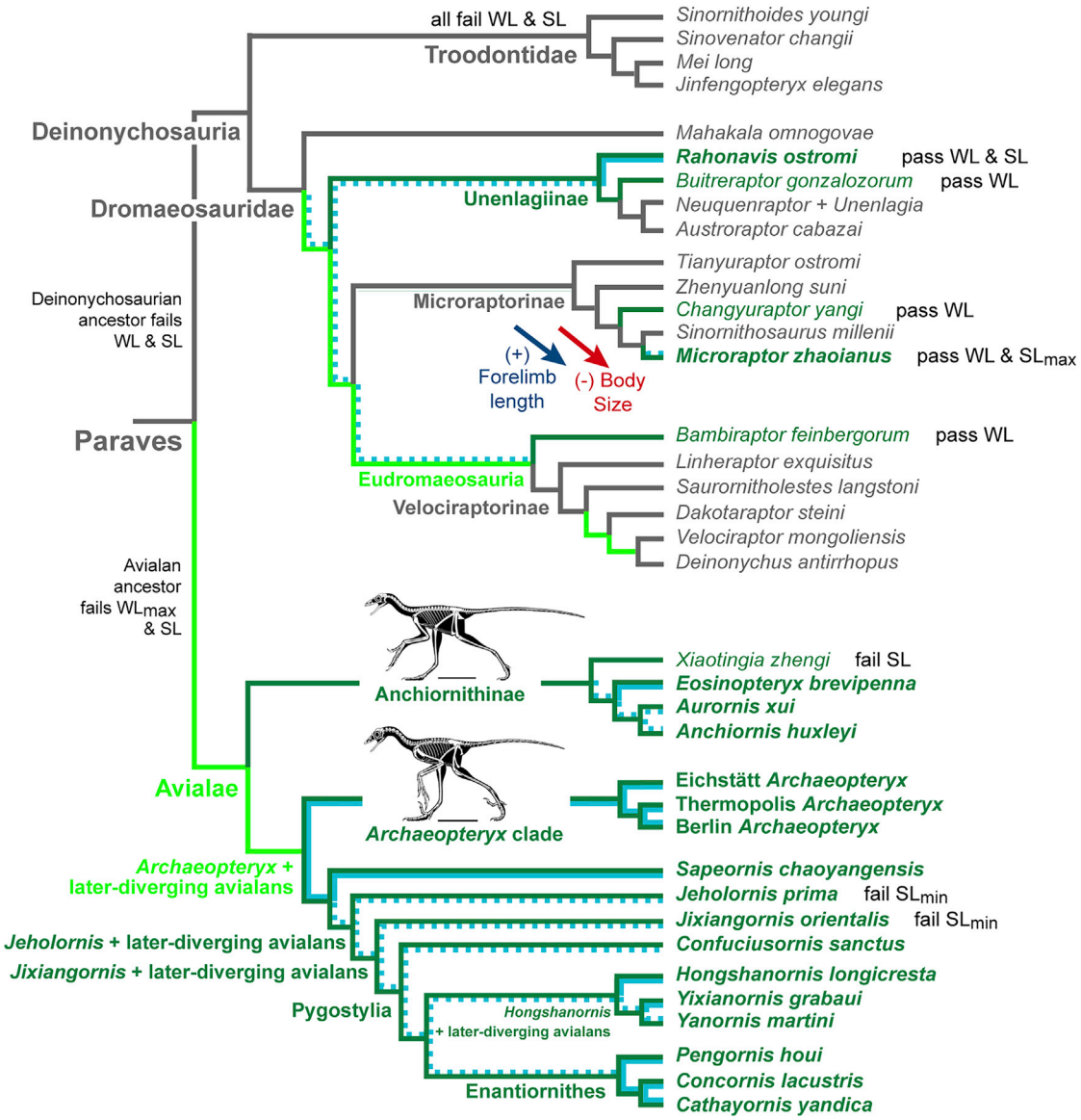


Figure 2. Parsimony-Based Ancestral State Reconstruction Analysis of Paravian Wing Loading and Specific Lift Using the Updated TWiG-Based Phylogeny

Powered flight potential in an ancestor or terminal taxon as suggested by wing loading estimates below the 2.5 gcm^{-2} threshold (Methods S1G and S1H): marked in dark green shading if relating to the maximum reconstructed ancestral value (i.e., minimum powered flight potential; minimum wing area and maximum body mass); marked in light green shading if only relating to the minimum reconstructed ancestral value (i.e., maximum powered flight potential; maximum wing area and minimum body mass). Gray shading denotes wing loading estimates above the 2.5 gcm^{-2} threshold. Powered flight potential in an ancestor or terminal taxon as suggested by specific lift values above the 9.8 Nkg^{-1} threshold (Methods S1I and S1J): marked in stippled light blue shading if relating to a maximum estimate (SL_{max} uses maximum muscle-mass-specific power output [Po,m] of 287 Wkg^{-1}); marked in light blue shading if relating to minimum estimate (SL_{min} uses Po,m of 225 Wkg^{-1}, respectively). Specific lift estimates below 9.8 Nkg^{-1} threshold are not marked. The figure includes the trends of increased forelimb length (dark blue arrow) and decreased body size (red arrow) along the microraptorine lineage. See main text for flight-related synapomorphies. See also STAR Methods, Methods S1, Tables S1 and S2, and the "Parsimony analysis" folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk. Skeletal reconstructions are used with the permission of Scott A. Hartman (scale bar represents 10 cm).

the potential for powered flight and to determine the extent to which they meet the requirements for flight we see in modern birds.

Anatomical Requirements for Flight

The primary anatomical requirements for flapping flight are summarized by wing loading. All of the non-paravian vane-feathered theropods sampled have estimated wing loading values well above 2.5 gcm^{-2}, with the lowest values of ~6.0 gcm^{-2} (Figure 2; Methods S1G and S1H) being larger than values previously measured in extant flightless birds, such as flightless ducks, cormorants, or Kakapos [70–72]. Using ancestral morphologies calculated by using ancestral state reconstruction analysis, we

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

determined all avialans (including anchiornithines) and five dromaeosaurids (*Bambiraptor*, *Buitreraptor*, *Changyuraptor*, *Microraptor*, and *Rahonavis*) among the vane-feathered paravians sampled have wing loading estimates at or below the 2.5 gcm^{-2} threshold for modern flapping flyers (Figure 2; Methods S1G and S1H). The dromaeosaurid *Mahakala* and the troodontid *Jinfengopteryx* are just above this threshold (Methods S1G and S1H). The early-diverging positions of the larger microraptorines *Tianyuraptor* and *Zhenyuanlong* imply a decrease in body size and an increase in relative forelimb length across Microraptorinae. This was confirmed quantitatively by using parsimony-based ancestral state reconstruction of body mass (from femoral length) and relative forelimb length (from [humerus + ulna length]/femur length). In other dromaeosaurid lineages, there appear to be multiple trends of body size and forelimb change (decreases as well as increases) [2, 73]. See Figure 2, STAR Methods, and Methods S1K–S1N.

Using ancestral morphologies calculated with ancestral state reconstruction analysis as well as direct calculations of wing loading and specific lift reveals that all fossil avialans sampled met the thresholds for powered flight seen in modern birds. This suggests that they had at least the potential for powered flight: wing loading values at or below 2.5 gcm^{-2} and specific lift estimates that exceed 9.8 Nkg^{-1} (except for minimum estimates of specific lift where Po,m = 225 Wkg^{-1}) (Figure 2; Methods S1J). Interestingly, the closest common ancestor of anchiornithines and later-diverging avialans fails the specific lift thresholds of powered flight potential, supporting a lack of powered flight in this ancestor (Figure 2). The potential for powered flight we found in anchiornithines (*contra* [18]) is supported by a reduced capacity for terrestrial running and greater emphasis on wing-based locomotion implied by the more proximal attachment of tail musculature, elongation of the acromion process, and more slender distal tibia found at the node shared between anchiornithines and traditional avialans [41]. The exception among anchiornithines is *Xiaotingia*, which was reasonably close to these thresholds for powered flight potential (Figures 2, 3, and 4). The potential for powered flight in anchiornithines should be treated cautiously though. This is because aspects of their anatomy might have affected their flight-relevant forelimb capabilities negatively, e.g., lack of functionally asymmetrical feathers (char. 1.0), relatively short ulnae and humerii compared with those of later-diverging avialans (char. 233.0; char. 262.2), limited pectoral musculature indicated by a weakly developed deltopectoral crest (char. 138.2) with an apex located closer to its proximal end (char. 684.2), and the lack of a bony sternum (at least in *Anchiornis*) [56, 74]. Paradoxically, *Xiaotingia* has a bowed rather than straight ulna, a feature linked with better takeoff potential in modern birds [75]. It also has a narrower radius (char. 438.1) like the aerodynamically capable *Microraptor*, suggesting some potential benefits related to flapping-based locomotion. Alternatively, these features might yield mechanical advantages in contexts other than powered flight that deserve further investigation.

The more active muscle-based shoulder stabilization expected in early-diverging birds is an anatomical limitation of powered flight potential that also needs to be considered, because it would have used energy that could have otherwise been used for lift generation (their acrocoracoid process and/or its homologs is at or below the level of the glenoid: char. 342.0 of [77]). A more passive and efficient intermediate condition of shoulder stabilization did not appear until at least the node uniting *Jeholornis* and later-diverging birds, although the earliest members of this clade still lacked a bony sternum and modern arm-flapping capabilities [41] (the acrocoracoid process became elevated above the glenoid: char. 342.1; the strut-like coracoid appeared: char. 134.3; a more passive ligament system enables compressive forces to be transmitted from the wing to the sternum [77]). The stabilizing role of a bony sternum is a synapomorphy of the more inclusive clade of *Jixiangornis* (possibly a synonym of *Jeholornis*) and more modern birds (char. 126.1). Shoulder stabilization becomes even more passive and efficient at the node uniting *Hongshanornis* and more modern birds, when the humeral head becomes enlarged through the development of a proximal convexity (char. 352.1). However, modern-style arm-flapping capabilities did not appear until later ornithurans [41]. Thus, fossil paravians that we suggest have the potential for powered flight likely did so less efficiently and with greater energy costs than do modern birds.

Aerodynamic Force Production Requirements for Flight

At the nodes Pennaraptora and Paraves, our wing loading estimates decrease, and to a lesser extent, our specific lift estimates increase (Figure 2). This coincides with a notable reduction in body size [3, 41, 73, 78, 79], the appearance of pennaceous feathers [34, 41, 80] (symmetrical at Pennaraptora [81] and char. 456.1; asymmetrical at Paraves) [41, 81], and a respiratory system more suited to higher intensity aerobic activity (advanced costosternal ventilator pump appearing among pennaraptorans [41]). Taken together, these findings support the suggested arm-flapping capabilities of pennaraptorans [41], as well as the potential for wing-assisted locomotion among paravians [41]. In other words, the data suggest that the wings of these taxa might have been used to assist in locomotion, such as running speed, turning, braking, and jumping [18]. However, it is only at the node Paraves that either of our ranges of wing loading and specific lift estimates approach the minimal thresholds of powered flight potential (cf. initial aerial locomotion [41]) and only in Avialae and a few independent lineages within Paraves (Unenlagiinae and Microraptorinae), where both thresholds are reached, thus indicating high probabilities of powered flight potential (Figures 2, 3, and 4). This supports the disconnect between the origins of pennaceous feathers and their incorporation into a flight-capable regime in non-avialan theropods [18, 34]. Pennaceous feathers (char. 456.1) appear at Pennaraptora [81] (but see [47]), becoming asymmetrical in Paraves [41, 81]. Asymmetrical forelimb feathers are found in *Microraptor* and are widespread among birds diverging later than Anchiornithinae (char. 1.1). This, in turn, lends credence to the hypothesis that pennaceous feathers and wings first evolved for non-flight purposes, e.g., other wing-assisted locomotion [18], display, or egg brooding [82, 83].

Physiological Requirements for Flight

Accounting for these constraints, among non-avialan paravians, only *Microraptor* and *Rahonavis* have specific lift estimates above 9.8 Nkg^{-1}; no other non-avialan taxa were possibly volant (Figures 2, 3, and 4). Direct calculation of specific lift in *Microraptor* (Figure 3) shows that it passes the specific lift threshold for powered flight potential, with our maximum estimate of specific

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

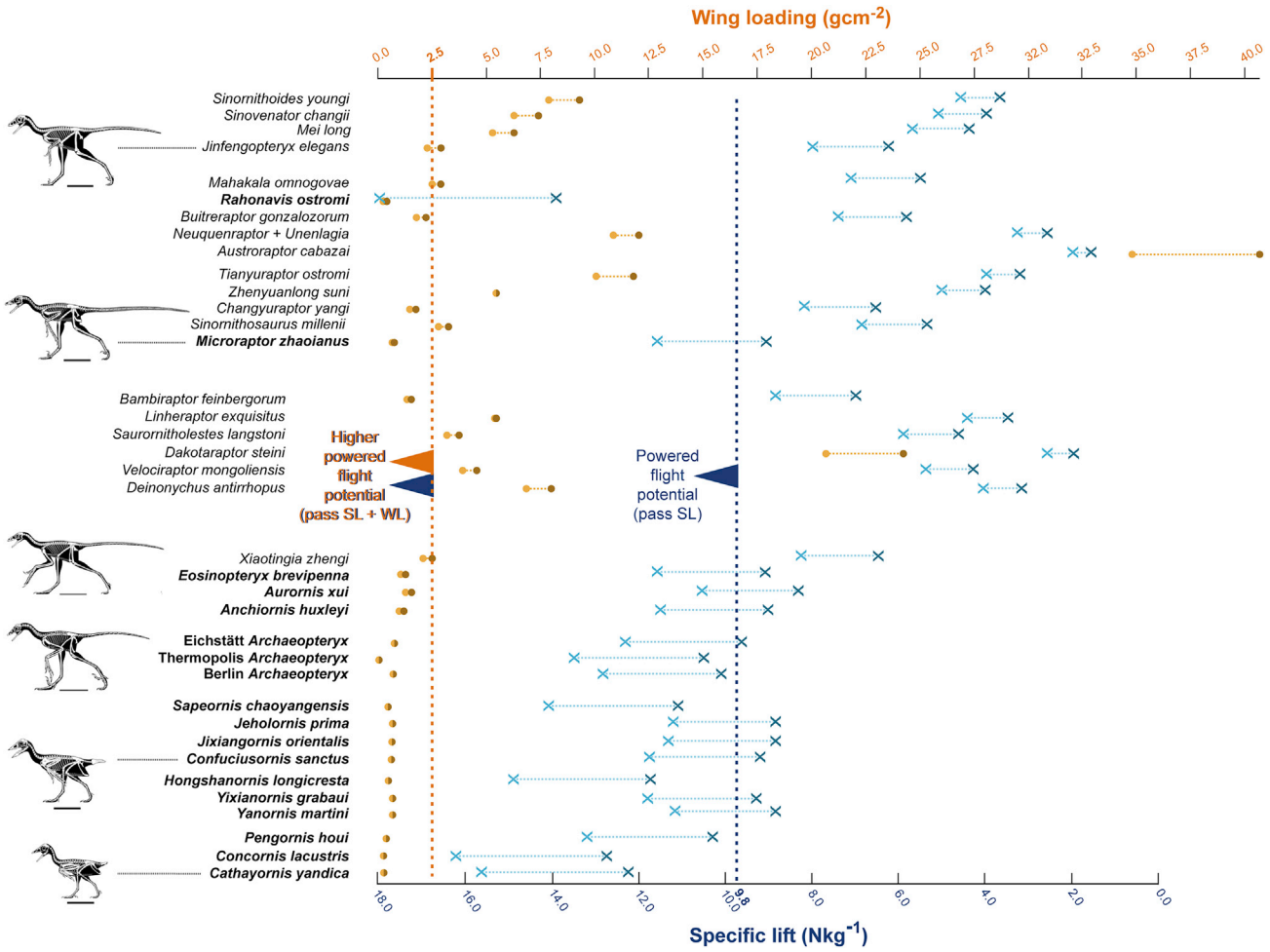


Figure 3. Maximum and Minimum Estimates of Wing Loading and Specific Lift for Each Terminal Taxon

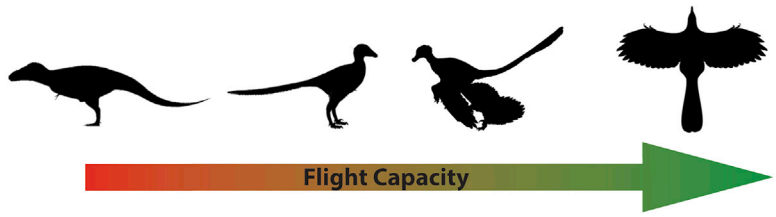
Maximum and minimum estimates of wing loading calculated using conservative and ultra-conservative wing areas (light orange and brown dots, respectively). Maximum and minimum estimates of specific lift were calculated using a broad range of Po,m values (287 Wkg^{-1} and 225 Wkg^{-1}). These terminal taxon values were not calculated from ancestral morphologies using ancestral state reconstruction analysis. See plotted values and their derivation in the "Parsimony analysis" folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk and in Tables S1 and S2. Skeletal reconstructions are used with the permission of Scott A. Hartman (scale bar represents 10 cm). See also Methods S1.

lift using a Po,m of 287 Wkg^{-1}. A wide range of plausible body mass estimates for *Microraptor* and *Rahonavis* derived from fossil measurements and 3D volumetric methods recovered the potential for powered flight: permutations, including body masses up to double what would be expected for a living bird of similar span, still retrieve a powered flight potential. For example, using the regression equations for wingspan versus mass calculated by Witton [84], using a larger dataset of extant birds (n = 96) or bats (n = 102), we generate mass estimates of 0.445 and 0.323 kg, respectively. Also, if we compare our estimate to a commonly used analog, the Common raven *Corvus corax*, we find that similar-sized individuals have wing spans well in excess of one meter (data from [55, 85]). Larger or earlier diverging relatives of *Microraptor* and *Rahonavis*, such as *Mahakala*, *Buitreraptor*, *Changyuraptor*, and *Bambiraptor*, as well as all troodontids, show lower values of lift (Figure 2; Methods S1I and S1J). Ancestral wing loading and specific lift can be calculated by using the wing loading and specific lift calculated for terminal taxa. However, we consider this less accurate than calculating ancestral morphologies and then using the nodal values to calculate ancestral wing loading and specific lift, because it does not independently assess values for the individual variables that determine wing loading and specific lift (which can change independently). Despite that, ancestral values of wing loading and specific lift calculated from wing loading and specific lift in terminals still produced similar results (Figure 3).

A wide range of deinonychosaurs showed wing loading values below the 2.5 gcm^{-2} threshold (Figures 2, 3, and 4). What is of particular interest in our results is the subset of taxa below the wing loading threshold but near or above the specific lift threshold (Figures 2, 3, and 4). Ancestral paravians shared several traits that presumably benefited the development of flapping-based locomotion, including smaller body size [41, 73], asymmetrically vaned feathers [41, 81], elongated and robust forelimbs [4, 41], a laterally orientated glenoid fossa articulation surface (char. 136.1; at equivalent node in [2]: char. 138.1;

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

CellPress
OPEN ACCESS



Fail All		Pass WL Only	Pass WL and Some SL	Pass All
Zuolong	*Incisvosaurus*	*Xiaotingia*	*Anchiornis* (2)	*Archaeopteryx* Berlin
Tanycolagreus	*Caudipteyrx*	*Buitreraptor*	*Aurornis* (1)	*Archaeopteryx* Thermopolis
Coelurus	*Avimius*	*Changyuraptor*	*Eosinopteryx* (2)	*Sapeornis*
Guanlong	*Microvenator*	*Bambiraptor*	*Archaeopteryx* Eichstätt (2)	*Pengornis*
Dilong	*Citipati*		*Jeholornis* (1)	*Cathyornis*
Eotyrannus	*Conchoraptor*		*Jixiangornis* (2)	*Concornis*
Tyrannosaurus	*Ingenia*		*Confuciusornis* (2)	*Hongshanornis*
Tarbosaurus	*Oviraptor*		*Yixianornis* (2)	*Asparavis*
Daspletosaurus	*Mei*		*Yanornis* (2)	*Rahonavis*
Albertosaurus	*Sinovenator*		*Microraptor* this study(2)	*Microraptor* Dyke 2013
Gorgosaurus	*Sinornithoides*		*Microraptor* C&T 2007 (2)	
Juravenator	*Jinfengopteryx*		*Microraptor* Allen 2013 (2)	
Sinocalliopteryx	*Mahakala*			
Huaxiagnathus	*Austroraptor*			
Sinosauropteryx	*Unenlagia*			
Compsognathus	*Tianyuraptor*			
Nqwebasaurus	*Zhenyuanlong*			
Beishanlong	*Sinornithosaurus*			
Harpymimus	*Saurornitholestes*			
Sinornithomimus	*Linheraptor*			
Gallimimus	*Deinonychus*			
Anserimimus	*Dakotaraptor*			
Struthiomimus	*Velociraptor*			
Ornithomimus				
Haplocheirus				
Mononykus				
Falcarius				
Beipiaosaurus				
Alxasaurus				
Erliansaurus				
Segnosaurus				
Nothronychus				

Figure 4. Overview of Flight Capabilities in Avialan and Non-avialan Theropods Tested in This Study

We grade taxa as more capable of some degree of flight capacity if they surpass minimum thresholds in several or all testing regimes administered here. For taxa that only pass the wing loading criteria, we deem them least likely up to the specimens and taxa that pass both tests at all three power level permutations that have strong potential for powered flight. Because the taxon *Microraptor* is crucial to our analyses, we included several mass and wing area estimate permutations (see text) labeled as follows: Allen 2013, Allen et al., 2013 [76]; C& T 2007, Chatterjee and Templin, 2007 [19]; and Dyke 2013, Dyke et al., 2013 [20]. The numbers in brackets beside a taxon indicate the number of specific lift permutations, out of 3, that said taxon succeeded in meeting or surpassing the minimum threshold. SL, specific lift tests; WL, wing loading. See also STAR Methods, Methods S1, Tables S1 and S2, and the "Parsimony analysis" folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk.

related to the extension of the glenoid floor onto the external surface of the scapula: char. 138.1 in [67], char. 216.1 in [28], char. 122.1 in [33], and char. 133.1 in [34] for node (*Jinfengopteryx* + Paraves)), a symmetrical furcula (char. 469.1), a laterally everted anterior edge of the acromion margin (char. 131.1), and elaborated brain regions associated with vision [41]. Although the origin of powered flight has been proposed at Paraves [41], our data do not support this but suggest the possibility of powered flight originating independently *outside* avialans. The unenlagiine *Rahonavis* is our strongest deinonychosaur candidate for flight potential, passing all wing loading and specific lift requirements (Figures 2, 3, and 4). This relates to its extremely elongated forearms of *Rahonavis* (ulna is longer than the femur as well as the tibia: Table 1 of [86]; see additional details in the spreadsheet in the "Parsimony analysis" folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk), which suggest very large wings.

The microraptorine *Microraptor* is another strong non-avialan candidate for flight, being below 9.8 Nkg^{-1} only for the strictest calculations (and even then displaying values approaching cutoff) (Figures 2, 3, and 4). Its robust, asymmetrically feathered forelimbs controlled by muscles attached to a fused, ossified sternum [6] support this. The vane asymmetry of its feathers though are less than the 3:1 vane ratio required for aeroelastic stability [62], which might have limited *Microraptor* to relatively short flights (aeroelastic stability requires at least a 3:1 vane ratio). However, this is the case for many early avialan taxa that otherwise seem flight capable. This is complementary to reconstructed aerodynamic prowess by several independent studies using traditional functional morphology [5], physical [20–22, 87], and theoretical modeling [19, 20]. Dececchi et al. [18] modeled launching in *Microraptor* (and other paravian taxa) similar to living birds. To further evaluate *Microraptor*'s candidacy for powered flight, we modeled its thrust-assisted launch potential under the alternative approach of [88], which used wing-generated thrust to supplement running takeoff in *Archaeopteryx.* We used their original parameters and calculated permutations that incorporated a larger flap angle and considered the effects of drag with both our model of *Microraptor* and the models of existing published studies [19, 20, 22]. In all cases, we found that *Microraptor* was capable of generating sufficient speed and flight forces for a ground-based takeoff and were within the range of values estimated for an arboreal launch [19, 20, 22]. Modeling approaches suggest that the 10% flight muscle ratio is probably underestimated for microraptorines (and *Archaeopteryx*) [89, 76], although this low ratio is found in some living volant birds, such as particular owl species [90]. If we increase this ratio slightly to 13%–15%, which is well within the range of living flying birds and is supported by volumetric modeling of these taxa, values well in excess of 9.8 Nkg^{-1} are achieved. For comparison, the average flight muscle fraction for living volant birds is around 18.4%, and this average is somewhat elevated by later-diverging forms, such as passerines that often have high flight muscle fractions [90]. Such promising flight potential provides a compelling context for interpreting unusual, potentially flight-related anatomies in more detail (e.g., the

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

elliptical fenestra of the deltopectoral crest found in *Microraptor* and the volant early birds *Confuciusornis* and *Sapeornis*) [16, 91, 92]. See "Additional description of results, including previously unreported synapomorphies of select paravian clades" in Methods S1 for additional information about microraptorines.

Although other paravian taxa, such as the troodontid *Jinfengopteryx* and the dromaeosaurids *Bambiraptor*, *Buitreraptor*, *Changyuraptor*, and *Mahakala*, are close to these thresholds, they never surpass them, despite the generous wing and flight muscle ratio reconstructions adopted (Figures 2, 3, and 4). Even though they did not pass the specific lift threshold, their high scores as well as low wing loading values make these taxa—particularly the microraptorine *Changyuraptor*—deserving of further study from the flight potential perspective, using more fine-grained techniques and modeling. This will distill the extant nature of the changes that are necessary to transition from non-volant, flapping-based locomotion to active flight. The recent suggestion of a short-armed clade at the base of Dromaeosauridae [93] supports the idea that flight capability is not ancestral to paravians.

Model Confidence and Hypothesis Testing

Phylogenetic distributions of wing loading and specific lift (see STAR Methods and Methods S1G–S1J) combined with osteological, integumentary, and body size changes reconstructed from our phylogeny contribute to a more holistic and integrated view of the origin of powered theropod flight. The robust phylogenetic context allows us to examine the evolutionary transitions of powered flight requirements from the perspective of anatomy, aerodynamic force production, and muscle physiology. Of these three categories of flight requirements, we are most confident that some small non-avialan paravians had the required anatomical competency for flight. We are highly confident that aerodynamic force production was sufficient for flight in *Microraptor*, *Rahonavis*, and early birds. Because muscle physiology is not known for fossil taxa, and because our specific lift estimations must necessarily make more assumptions than the other aspects of the analysis, we are less confident regarding the precise patterns of specific lift evolution recovered in the analysis. However, our results do show that, for conservative muscle power outputs, some of the non-avialan paravian taxa could likely fly, even if only briefly at lower power outputs. All permutations for *Rahonavis* suggest powered flight potential, as do 9 of our 12 permutations of the *Microraptor gui* model using 10% muscle mass and all 12 using 13% muscle mass, as estimated by Allen et al. [76]. From these results, we suggest that muscle physiology might have been the limiting constraint for flight in early paravians. Under this model, any time muscle physiology crossed the critical power output boundary, flight could have originated—and this could have happened multiple times. Our results also allow us to test a number of other hypotheses relating to four areas.

Muscle Size and Physiology

We reject the hypothesis that flight muscle fractions above 10% would be required for large-winged, non-avian pennaraptorans to engage in powered flight. We further reject the hypothesis that flight muscle physiologies outside those seen in modern birds would be required for large-winged, non-avian pennaraptorans to engage in powered flight.

Wing Area

We reject the hypothesis that large-winged, non-avialan pennaraptorans would have been prevented from flight on account of insufficient wing area in relation to body mass. Only under the most extreme body mass estimates for large-winged non-avialan pennaraptorans do we retrieve wing loading results above the powered flight thresholds observed in living birds. For example, even using the heaviest mass estimate per Allen et al. [76] of 1.59 kg for *Microraptor gui* and the lowest wing area estimate of Chatterjee and Templin [19], on the basis of the incomplete estimate of feather length, we still obtain wing loading values of 1.69 gcm^{-2}, well below the 2.5 gcm^{-2} maximum and similar to values seen in adult chukar partridges [85] and turkeys [94].

Duration of Aerial Behaviors

We cannot reject the hypothesis that powered aerial behaviors in large-winged, non-avialan pennaraptorans were typically brief in duration. Although the gross wing structure in large-winged, non-avian pennaraptorans appears to be very similar to that of living birds, the structure of individual feathers suggests that aeroelastic instability in early taxa might have reduced wing performance. In some permutations of our estimated parameter set, recovery of power flight potential in large-winged, non-avialan pennaraptorans was dependent on a portion of the flight muscle mass being anaerobic. For example, all anchiornithines at a 10% muscle fraction require muscle outputs of minimally 250 Wkg^{-1} to achieve sufficient lift for takeoff (except *Xiaotingia*, which never achieves sufficient lift). In these cases, low muscle endurance would necessitate short-ranged aerial behaviors.

Powered Flight Potential across Paraves

We cannot reject the hypothesis that large-winged, non-avialan pennaraptorans potentially had powered flight and that some kind of potential evolved multiple times among paravians. The majority of our parameter permutations recover some level of powered flight potential in large-winged, non-avialan paravians. On the basis these results, we feel that it is likely that powered flight evolved multiple times from a range of paravian ancestors that were already nearing powered flight potential: twice in Dromaeosauridae; once or twice in Avialae (depending on character optimizations for *Xiaotingia*); and potentially once in Troodontidae (if more capable examples than *Jinfengopteryx* are found; Figures 2, 3, and 4) but originated neither with Avialae nor at a single node within Paraves. We consider this more likely than a scenario in which the body densities or muscle physiologies of non-avialan pennaraptorans were far outside those measured in living birds. These potential powered flyers are all associated with size reductions as well as forelimb elongation (in this study and in [18, 73]). Notable anatomical differences between subclades (e.g., the absence of ossified sterna in troodontids and their presence in dromaeosaurids) [74, 95] suggest these independent origins of flight are not entirely parallel—i.e., they do not share the same anatomical starting points and might have achieved functional flight in different ways. This also appears to be the case for non-paravian pennaraptorans, as suggested by the bizarre membranous wings of the scansoriopterygids [33, 96]. The relative paucity of preserved troodontid forelimbs compared with those of dromaeosaurids hinders some of these reconstructions, but known forelimb differences have intriguing

implications for the evolution and ecology of paravian powered flight.

Revised Evolutionary Scenario and New Frontiers

Our analysis suggested multiple origins of powered flight from differing initial conditions; some members exhibited some capacity for wing-based locomotory assistance that, although not flight capable, might have assisted non-volant behaviors [18]. This implies that Paraves, in general, might have been experimenting with wing-assisted, non-volant behaviors to expand into locomotory repertoires otherwise unexplored by Late Jurassic and Early Cretaceous vertebrates. These include high-speed running and starts, leaping, rapid braking and turning, and dynamic balance. Emphasis on some of these behaviors in different paravian, and even pennaraptoran, clades might have presented opportunities for diverse ecological niches for these agile taxa. Only when some clades evolved smaller body sizes did these independent biomechanical repertoires, adapted for high-speed terrestrial or scansorial locomotion, become capable of powered flight. This evolutionary scenario emphasizes further examination of the non-volant, large-bodied paravians, with the goal of estimating their differing anatomical and biomechanical specializations. The results presented here suggest paravians were exploring a wider range of locomotory niches than previously appreciated and might have set the stage for the origin of birds and powered flight from a rapidly evolving, highly diverse suite of locomotory and ecological experimentations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Phylogenetic Dataset
- METHOD DETAILS
  - Automated phylogenetic analysis
  - Reconstruction of wing loading and specific lift values
  - Trends in body mass change and forelimb length
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Uncertainty quantification and estimation confidence

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.06.105.

A video abstract is available at https://doi.org/10.1016/j.cub.2020.06.105#mmc4.

ACKNOWLEDGMENTS

This study was supported by the Research Grant Council of Hong Kong's General Research Fund(17103315) to M.P. It was also supported by The University of Hong Kong (RAE Improvement Fund of the Faculty of Science to M.P. and T.G.K., the University Research Committee Postdoctoral Fellow Scheme to M.P. for R.P., Conference Support to M.P., and the Seed Fund for Basic Research [two awards] to M.P.), the National Science Foundation of China (41688103, 41120124002, 91514302 and 41972025) to X.X. and R.P., and a Proyecto de Unidad Ejecutora from CONICET (PUE0070) to P.A.G. This study benefited from discussions at the International Pennaraptoran Symposium, held at the University of Hong Kong and supported by Kenneth HC Fung and First Initiative Foundation. Scott A. Hartman is thanked for the use of his skeletal reconstructions in Figures 1, 2, and 3 and in the graphical abstract.

AUTHOR CONTRIBUTIONS

Conceptualization, M.P., P.A.G., and X.X. with input from T.A.D. and R.P.; Methodology, M.P., P.A.G., R.P., and T.A.D.; Software, P.A.G.; Validation, M.P., P.A.G., T.A.D., and R.P.; Formal Analysis, M.P., R.P., P.A.G., and T.A.D.; Investigation, M.P., R.P., P.A.G., and T.A.D.; Resources, M.P., P.A.G., and T.G.K.; Data Curation, M.P., P.A.G., R.P., T.A.D., and T.G.K.; Writing – Original Draft, M.P., R.P., P.A.G., T.A.D., M.B.H., T.G.K., H.C.E.L., M.A.N., S.L.B., and X.X.; Writing – Review and Editing, M.P., P.A.G., and T.A.D.; Visualization, M.P. and P.A.G.; Supervision, M.P. and P.A.G.; Project Administration, M.P.; Funding Acquisition, M.P., P.A.G., T.G.K., and X.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 2, 2019
Revised: March 19, 2020
Accepted: June 30, 2020
Published: August 6, 2020

REFERENCES

1. Brusatte, S.L., Lloyd, G.T., Wang, S.C., and Norell, M.A. (2014). Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. Curr. Biol. *24*, 2386–2392.
2. Turner, A.H., Makovicky, P.J., and Norell, M.A. (2012). A review of dromaeosaurid systematics and paravian phylogeny. Bull. Am. Mus. Nat. Hist. *371*, 1–206.
3. Benson, R.B.J., Campione, N.E., Carrano, M.T., Mannion, P.D., Sullivan, C., Upchurch, P., and Evans, D.C. (2014). Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. PLoS Biol. *12*, e1001853.
4. Xu, X., You, H., Du, K., and Han, F. (2011). An *Archaeopteryx*-like theropod from China and the origin of Avialae. Nature *475*, 465–470.
5. Xu, X., Zhou, Z., and Wang, X. (2000). The smallest known non-avian theropod dinosaur. Nature *408*, 705–708.
6. Xu, X., Zhou, Z., Wang, X., Kuang, X., Zhang, F., and Du, X. (2003). Four-winged dinosaurs from China. Nature *421*, 335–340.
7. Xu, X., Zhao, Q., Norell, M., Sullivan, C., Hone, D., Erickson, G., Wang, X., Han, F., and Guo, Y. (2009). A new feathered maniraptoran dinosaur fossil that fills a morphological gap in avian origin. Chin. Sci. Bull. *54*, 430–435.
8. Xu, X., Wang, X.L., and Wu, X.C. (1999). A dromaeosaurid dinosaur with a filamentous integument from the Yixian Formation of China. Nature *401*, 262–266.
9. Ji, Q., Norell, M.A., Gao, K.Q., Ji, S.A., and Ren, D. (2001). The distribution of integumentary structures in a feathered dinosaur. Nature *410*, 1084–1088.
10. Ji, Q., Ji, S.A., Lu, J.C., You, H.L., Chen, W., Liu, Y.Q., and Liu, Y.X. (2005). First avialian bird from China (*Jinfengopteryx elegans* gen. et sp. nov.). Geol. Bull. China *24*, 197–210.
11. Norell, M., Ji, Q., Gao, K., Yuan, C., Zhao, Y., and Wang, L. (2002). Palaeontology: 'modern' feathers on a non-avian dinosaur. Nature *416*, 36–37.
12. Norell, M.A., and Xu, X. (2005). Feathered dinosaurs. Annu. Rev. Earth Planet. Sci. *33*, 277–299.

13. Godefroit, P., Cau, A., Hu, D.Y., Escuillié, F., Wu, W.H., and Dyke, G. (2013). A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. Nature *498*, 359–362.

14. Godefroit, P., Demuynck, H., Dyke, G., Hu, D., Escuillié, F., and Claeys, P. (2013). Reduced plumage and flight ability of a new Jurassic paravian theropod from China. Nat. Commun. *4*, 1394.

15. Han, G., Chiappe, L.M., Ji, S.A., Habib, M., Turner, A.H., Chinsamy, A., Liu, X., and Han, L. (2014). A new raptorial dinosaur with exceptionally long feathering provides insights into dromaeosaurid flight performance. Nat. Commun. *5*, 4382.

16. Pei, R., Li, Q., Meng, Q., Gao, K.Q., and Norell, M.A. (2014). A new specimen of *Microraptor* (Theropoda: Dromaeosauridae) from the Lower Cretaceous of western Liaoning, China. Am. Mus. Novit. *3821*, 1–28.

17. Lü, J., and Brusatte, S.L. (2015). A large, short-armed, winged dromaeosaurid (Dinosauria: Theropoda) from the Early Cretaceous of China and its implications for feather evolution. Sci. Rep. *5*, 11775.

18. Dececchi, T.A., Larsson, H.C.E., and Habib, M.B. (2016). The wings before the bird: an evaluation of flapping-based locomotory hypotheses in bird antecedents. PeerJ *4*, e2159.

19. Chatterjee, S., and Templin, R.J. (2007). Biplane wing planform and flight performance of the feathered dinosaur *Microraptor gui*. Proc. Natl. Acad. Sci. USA *104*, 1576–1580.

20. Dyke, G., de Kat, R., Palmer, C., van der Kindere, J., Naish, D., and Ganapathisubramani, B. (2013). Aerodynamic performance of the feathered dinosaur *Microraptor* and the evolution of feathered flight. Nat. Commun. *4*, 2489.

21. Koehl, M.A.R., Evangelista, D., and Yang, K. (2011). Using physical models to study the gliding performance of extinct animals. Integr. Comp. Biol. *51*, 1002–1018.

22. Alexander, D.E., Gong, E., Martin, L.D., Burnham, D.A., and Falk, A.R. (2010). Model tests of gliding with different hindwing configurations in the four-winged dromaeosaurid *Microraptor gui*. Proc. Natl. Acad. Sci. USA *107*, 2972–2976.

23. Palmer, C. (2014). The aerodynamics of gliding flight and its application to the arboreal flight of the Chinese feathered dinosaur *Microraptor*. Biol. J. Linn. Soc. Lond. *113*, 828–835.

24. Xu, X., and Norell, M.A. (2004). A new troodontid dinosaur from China with avian-like sleeping posture. Nature *431*, 838–841.

25. Gauthier, J.A. (1986). Saurischian monophyly and the origin of birds. In The origin of birds and the evolution of flight, K. Padian, ed. (California Academy of Science), pp. 1–55.

26. Sereno, P.C. (1997). The origin and evolution of dinosaurs. Annu. Rev. Earth Planet. Sci. *25*, 435–489.

27. Norell, M.A., Clark, J.M., and Makovicky, P.J. (2001). Phylogenetic relationships among coelurosaurian theropods. In New perspectives on the origin and early evolution of birds, J. Gauthier, and L.F. Gall, eds. (Peabody Museum of Natural History), pp. 49–68.

28. Senter, P., Kirkland, J.I., DeBlieux, D.D., Madsen, S., and Toth, N. (2012). New Dromaeosaurids (Dinosauria: Theropoda) from the lower cretaceous of Utah, and the evolution of the Dromaeosaurid tail. PLoS ONE *7*, e36790.

29. Lü, J.C., Xu, L., Zhang, X.L., Ji, Q., Jia, S.H., Hu, W.Y., et al. (2007). New dromaeosaurid dinosaur from the Late Cretaceous Qiupa Formation of Luanchuan area, western Henan, China. Geol. Bull. China *26*, 777–786.

30. Evans, D.C., Larson, D.W., and Currie, P.J. (2013). A new dromaeosaurid (Dinosauria: Theropoda) with Asian affinities from the latest Cretaceous of North America. Naturwissenschaften *100*, 1041–1049.

31. Pittman, M., Pei, R., Tan, Q., and Xu, X. (2015). The first dromaeosaurid (Dinosauria: Theropoda) from the Lower Cretaceous Bayan Gobi Formation of Nei Mongol, China. PeerJ *3*, e1480.

32. DePalma, R.A., Burnham, D.A., Martin, L.D., Larson, P.L., and Bakker, R.T. (2015). The first giant raptor (Theropoda: Dromaeosauridae) from the Hell Creek Formation. Paleontol. Contrib. *2015*, 1–16.

33. Xu, X., Zheng, X., Sullivan, C., Wang, X., Xing, L., Wang, Y., Zhang, X., O'Connor, J.K., Zhang, F., and Pan, Y. (2015). A bizarre Jurassic maniraptoran theropod with preserved evidence of membranous wings. Nature *521*, 70–73.

34. Foth, C., Tischlinger, H., and Rauhut, O.W.M. (2014). New specimen of *Archaeopteryx* provides insights into the evolution of pennaceous feathers. Nature *511*, 79–82.

35. Cau, A., Brougham, T., and Naish, D. (2015). The phylogenetic affinities of the bizarre Late Cretaceous Romanian theropod *Balaur bondoc* (Dinosauria, Maniraptora): dromaeosaurid or flightless bird? PeerJ *3*, e1032.

36. Zheng, X.T., You, H.L., Xu, X., and Dong, Z.M. (2009). An Early Cretaceous heterodontosaurid dinosaur with filamentous integumentary structures. Nature *458*, 333–336.

37. Xu, X., Choiniere, J.N., Pittman, M., Tan, Q., Xiao, D., Li, Z., Tan, L., Clark, J.M., Norell, M.A., Hone, D.W.E., et al. (2010). A new dromaeosaurid (Dinosauria: Theropoda) from the Upper Cretaceous Wulansuhai Formation of Inner Mongolia, China. Zootaxa *2403*, 1–9.

38. Norell, M.A., Clark, J.M., Turner, A.H., Makovicky, P.J., Barsbold, R., and Rowe, T. (2006). A new dromaeosaurid theropod from Ukhaa Tolgod (Ömnögov, Mongolia). Am. Mus. Novit. *3545*, 1–51.

39. Xu, X., Pittman, M., Sullivan, C., Choiniere, J.N., Tan, Q.W., Clark, J.M., Norell, M.A., and Wang, S. (2015). The taxonomic status of the Late Cretaceous dromaeosaurid *Linheraptor exquisitus* and its implications for dromaeosaurid systematics. Vertebrata Pal Asiatica *53*, 29–62.

40. Pei, R., Li, Q., Meng, Q., Norell, M.A., and Gao, K.Q. (2017). New specimens of *Anchiornis huxleyi* (Theropoda: Paraves) from the Late Jurassic of northeastern China. Bull. Am. Mus. Nat. Hist. *411*, 1–67.

41. Xu, X., Zhou, Z., Dudley, R., Mackem, S., Chuong, C.M., Erickson, G.M., and Varricchio, D.J. (2014). An integrative approach to understanding bird origins. Science *346*, 1253293.

42. Makovicky, P.J., Apesteguía, S., and Agnolín, F.L. (2005). The earliest dromaeosaurid theropod from South America. Nature *437*, 1007–1011.

43. Gauthier, J., and de Queiroz, K. (2001). Feathered dinosaurs, flying dinosaurs, crown dinosaurs, and the name 'Aves'. In New perspectives on the origin and early evolution of birds, J. Gauthier, and L.F. Gall, eds. (Peabody Museum of Natural History).

44. Meunier, K. (1951). Korrelation und umkonstruktion in den größenbeziehungen zwischen vogelflügel und vogelkörper (Correlation and restructuring in the size relationship between avian wing and avian body). Biol. Gen. *19*, 403–443.

45. Marden, J.H. (1987). Maximum lift production during takeoff in flying animals. J. Exp. Biol. *130*, 235–258.

46. Guillemette, M., and Ouellet, J.F. (2005). Temporary flightlessness in prelaying Common Eiders *Somateria mollissima*: are females constrained by excessive wing-loading or by minimal flight muscle ratio? Ibis *147*, 293–300.

47. Zelenitsky, D.K., Therrien, F., Erickson, G.M., DeBuhr, C.L., Kobayashi, Y., Eberth, D.A., and Hadfield, F. (2012). Feathered non-avian dinosaurs from North America provide insight into wing origins. Science *338*, 510–514.

48. Pennycuick, C.J. (2008). Modelling the Flying Bird (Academic).

49. Pennycuick, C.J. (1990). Predicting wingbeat frequency and wavelength of birds. J. Exp. Biol. *150*, 171–185.

50. Alexander, D.E. (2004). Nature's Flyers: Birds, Insects, and the Biomechanics of Flight (Johns Hopkins University).

51. Crandell, K.E., and Tobalske, B.W. (2011). Aerodynamics of tip-reversal upstroke in a revolving pigeon wing. J. Exp. Biol. *214*, 1867–1873.

52. Earls, K.D. (2000). Kinematics and mechanics of ground take-off in the starling *Sturnis vulgaris* and the quail *Coturnix coturnix*. J. Exp. Biol. *203*, 725–739.

53. Campione, N.E., Evans, D.C., Brown, C.M., and Carrano, M.T. (2014). Body mass estimation in non-avian bipeds using a theoretical conversion to quadruped stylopodial proportions. Methods Ecol. Evol. *5*, 913–923.

54. Thomas, A.L.R. (1993). The aerodynamic costs of asymmetry in the wings and tail of birds: asymmetric birds can't fly round tight corners. Proc. R. Soc. Lond. B Biol. Sci. *254*, 181–189.

55. Serrano, F.J., and Chiappe, L.M. (2017). Aerodynamic modelling of a Cretaceous bird reveals thermal soaring capabilities during early avian evolution. J. R. Soc. Interface *14*, 20170182.

56. Wang, X., Pittman, M., Zheng, X., Kaye, T.G., Falk, A.R., Hartman, S.A., and Xu, X. (2017). Basal paravian functional anatomy illuminated by high-detail body outline. Nat. Commun. *8*, 14576.

57. Falk, A.R., Kaye, T.G., Zhou, Z., and Burnham, D.A. (2016). Laser fluorescence illuminates the soft tissue and life habits of the Early Cretaceous bird *Confuciusornis*. PLoS ONE *11*, e0167284.

58. Zheng, X., O'Connor, J.K., Wang, X., Pan, Y., Wang, Y., Wang, M., and Zhou, Z. (2017). Exceptional preservation of soft tissue in a new specimen of *Eoconfuciusornis* and its biological implications. Natl. Sci. Rev. *4*, 441–452.

59. Longrich, N.R., Vinther, J., Meng, Q., Li, Q., and Russell, A.P. (2012). Primitive wing feather arrangement in *Archaeopteryx lithographica* and *Anchiornis huxleyi*. Curr. Biol. *22*, 2262–2267.

60. Voeten, D.F.A.E., Cubo, J., de Margerie, E., Röper, M., Beyrand, V., Bureš, S., Tafforeau, P., and Sanchez, S. (2018). Wing bone geometry reveals active flight in *Archaeopteryx*. Nat. Commun. *9*, 923.

61. Wang, X., Nudds, R.L., Palmer, C., and Dyke, G.J. (2012). Size scaling and stiffness of avian primary feathers: implications for the flight of Mesozoic birds. J. Evol. Biol. *25*, 547–555.

62. Feo, T.J., Field, D.J., and Prum, R.O. (2015). Barb geometry of asymmetrical feathers reveals a transitional morphology in the evolution of avian flight. Proc. Biol. Sci. *282*, 20142864.

63. Lees, J., Garner, T., Cooper, G., and Nudds, R. (2017). Rachis morphology cannot accurately predict the mechanical performance of primary feathers in extant (and therefore fossil) feathered flyers. R. Soc. Open Sci. *4*, 160927.

64. Senter, P. (2006). Comparison of forelimb function between *Deinonychus* and *Bambiraptor* (Theropoda: Dromaeosauridae). J. Vertebr. Paleontol. *26*, 897–906.

65. Senter, P., and Robins, J.H. (2015). Resting orientations of dinosaur scapulae and forelimbs: a numerical analysis, with implications for reconstructions and museum mounts. PLoS ONE *10*, e0144036.

66. Norberg, U.M. (1975). Hovering flight in the pied flycatcher (*Ficedula hypoleuca*). In Swimming and Flying in Nature, T.T. Wu, C.J. Brokaw, and C. Brennen, eds. (Plenum), pp. 869–881.

67. Agnolín, F.L., and Novas, F.E. (2013). Avian ancestors: a review of the phylogenetic relationships of the theropods Unenlagiidae, Microraptoria, *Anchiornis* and Scansoriopterygidae (Springer).

68. Goloboff, P.A., Farris, J.S., Källersjö, M., Oxelman, B., Ramírez, M., and Szumik, C.A. (2003). Improvements to resampling measures of group support. Cladistics *19*, 324–332.

69. Pol, D., and Goloboff, P.A. (2020). The impact of unstable taxa in coelurosaurian phylogeny and resampling support measures for parsimony analyses. In Pennaraptoran theropod dinosaurs: past progress and new frontiers, M. Pittman, and X. Xu, eds. (American Museum of Natural History).

70. Livezey, B.C. (1992). Flightlessness in the Galápagos cormorant (*Compsohalieus* [*Nannopterum*] *harrisi*): heterochrony, giantism and specialization. Zool. J. Linn. Soc. *105*, 155–224.

71. Livezey, B.C., and Humphrey, P.S. (1986). Flightlessness in steamer-ducks (Anatidae: Tachyeres): its morphological bases and probable evolution. Evolution *40*, 540–558.

72. Livezey, B.C. (1992). Morphological corollaries and ecological implications of flightlessness in the kakapo (Psittaciformes: *Strigops habroptilus*). J. Morphol. *213*, 105–145.

73. Turner, A.H., Pol, D., Clarke, J.A., Erickson, G.M., and Norell, M.A. (2007). A basal dromaeosaurid and size evolution preceding avian flight. Science *317*, 1378–1381.

74. Zheng, X., O'Connor, J., Wang, X., Wang, M., Zhang, X., and Zhou, Z. (2014). On the absence of sternal elements in *Anchiornis* (Paraves) and *Sapeornis* (Aves) and the complex early evolution of the avian sternum. Proc. Natl. Acad. Sci. USA *111*, 13900–13905.

75. Dial, K.P. (1992). Avian forelimb muscles and nonsteady flight: can birds fly without using the muscles in their wings? Auk *109*, 874–885.

76. Allen, V., Bates, K.T., Li, Z., and Hutchinson, J.R. (2013). Linking the evolution of body shape and locomotor biomechanics in bird-line archosaurs. Nature *497*, 104–107.

77. Baier, D.B., Gatesy, S.M., and Jenkins, F.A. (2007). A critical ligamentous mechanism in the evolution of avian flight. Nature *445*, 307–310.

78. Lee, M.S.Y., Cau, A., Naish, D., and Dyke, G.J. (2014). Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. Science *345*, 562–566.

79. Dececchi, T.A., and Larsson, H.C.E. (2013). Body and limb size dissociation at the origin of birds: uncoupling allometric constraints across a macroevolutionary transition. Evolution *67*, 2741–2752.

80. Brusatte, S.L., O'Connor, J.K., and Jarvis, E.D. (2015). The origin and diversification of birds. Curr. Biol. *25*, R888–R898.

81. Xu, X., Currie, P., Pittman, M., Xing, L., Meng, Q., Lü, J., Hu, D., and Yu, C. (2017). Mosaic evolution in an asymmetrically feathered troodontid dinosaur with transitional features. Nat. Commun. *8*, 14972.

82. Brusatte, S.L. (2017). A Mesozoic aviary. Science *355*, 792–794.

83. Koschowitz, M.C., Fischer, C., and Sander, M. (2014). Evolution. Beyond the rainbow. Science *346*, 416–418.

84. Witton, M. (2008). A new approach to determining pterosaur body mass and its implications for pterosaur flight. Zitteliana *B28*, 143–159.

85. Jackson, B.E. (2009). The allometry of bird flight performance. PhD thesis (University of Montana).

86. Forster, C.A., Sampson, S.D., Chiappe, L.M., and Krause, D.W. (1998). The theropod ancestry of birds: new evidence from the late cretaceous of madagascar. Science *279*, 1915–1919.

87. Davis, M. (2008). The Four-Winged Dinosaur. In NOVA (series), February 26, 2008. https://www.pbs.org/wgbh/nova/evolution/four-winged-dinosaur.html.

88. Burgers, P., and Chiappe, L.M. (1999). The wing of *Archaeopteryx* as a primary thrust generator. Nature *399*, 60–62.

89. Heers, A.M., Baier, D.B., Jackson, B.E., and Dial, K.P. (2016). Flapping before flight: high resolution, three-dimensional skeletal kinematics of wings and legs during avian development. PLoS ONE *11*, e0153446.

90. Greenewalt, C.H. (1962). Dimensional relationships for flying animals. Smithson. Misc. Collect. *144*, 1–46.

91. Chiappe, L.M., Ji, S., Ji, Q., and Norell, M.A. (1999). Anatomy and systematics of the Confuciusornithidae (Theropoda: Aves) from the late Mesozoic of northeastern China. Bull. Am. Mus. Nat. Hist. *242*, 1–89.

92. Zhou, Z., and Zhang, F. (2003). Anatomy of the primitive bird *Sapeornis chaoyangensis* from the Early Cretaceous of Liaoning, China. Can. J. Earth Sci. *40*, 731–747.

93. Cau, A., Beyrand, V., Voeten, D.F.A.E., Fernandez, V., Tafforeau, P., Stein, K., Barsbold, R., Tsogtbaatar, K., Currie, P.J., and Godefroit, P. (2017). Synchrotron scanning reveals amphibious ecomorphology in a new clade of bird-like dinosaurs. Nature *552*, 395–399.

94. Tobalske, B.W., and Dial, K.P. (2000). Effects of body size on take-off flight performance in the Phasianidae (Aves). J. Exp. Biol. *203*, 3319–3332.

95. Lambertz, M., and Perry, S.F. (2015). Remarks on the evolution of the avian sternum, dinosaur gastralia, and their functional significance for the respiratory apparatus. Zool. Anz. *255*, 80–84.

96. Wang, M., O'Connor, J.K., Xu, X., and Zhou, Z. (2019). A new Jurassic scansoriopterygid and the loss of membranous wings in theropod dinosaurs. Nature *569*, 256–259.

97. Kaye, T.G., Falk, A.R., Pittman, M., Sereno, P.C., Martin, L.D., Burnham, D.A., Gong, E., Xu, X., and Wang, Y. (2015). Laser-stimulated fluorescence in paleontology. PLoS ONE *10*, e0125923.

98. Goloboff, P.A., Farris, J.S., and Nixon, K.C. (2008). TNT, a free program for phylogenetic analysis. Cladistics *24*, 774–786.

99. Goloboff, P.A., and Catalano, S.A. (2016). TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics *32*, 221–238.

100. Goloboff, P.A. (1999). Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics *15*, 415–428.

101. Coddington, J., and Scharff, N. (1994). Problems with zero-length branches. Cladistics *10*, 415–423.

102. Goloboff, P.A., and Farris, J.S. (2001). Methods for quick consensus estimation. Cladistics *17*, S26–S34.

103. Goloboff, P.A. (1993). Estimating character weights during tree search. Cladistics *9*, 83–91.

104. Goloboff, P.A. (2014). Extended implied weighting. Cladistics *30*, 260–272.

105. Goloboff, P.A., and Szumik, C.A. (2015). Identifying unstable taxa: Efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. Mol. Phylogenet. Evol. *88*, 93–104.

106. Goloboff, P.A., and Catalano, S.A. (2012). GB-to-TNT: facilitating creation of matrices from GenBank and diagnosis of results in TNT. Cladistics *28*, 503–513.

107. Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. *55*, 539–552.

108. Pinegar, R.T., Loewen, M.A., Cloward, K.C., Hunter, R.J., and Weege, C.J. (2003). A juvenile allosaur with preserved integument from the basal Morrison Formation of central Wyoming. J. Vertebr. Paleontol. Prog. Abstr. *23*, 87A–88A.

109. Campione, N.E., Barrett, P.M., and Evans, D.C. (2020). On the ancestry of feathers in Mesozoic dinosaurs. In The Evolution of Feathers: From Their Origin to the Present, C. Foth, and O.W.M. Rauhut, eds. (Springer), pp. 213–243.

110. Li, Q., Gao, K.Q., Meng, Q., Clarke, J.A., Shawkey, M.D., D'Alba, L., Pei, R., Ellison, M., Norell, M.A., and Vinther, J. (2012). Reconstruction of *Microraptor* and the evolution of iridescent plumage. Science *335*, 1215–1219.

111. Yalden, D.W., Bramwell, C.D., and Heptonstall, W.B. (1971). Flying ability of *Archaeopteryx*. Nature *231*, 127–128.

112. Serrano, F.J., Palmqvist, P., and Sanz, J.L. (2015). Multivariate analysis of neognath skeletal measurements: implications for body mass estimation in Mesozoic birds. Zool. J. Linn. Soc. *173*, 929–955.

113. Zhang, F., Zhou, Z., Xu, X., Wang, X., and Sullivan, C. (2008). A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. Nature *455*, 1105–1108.

114. Marden, J.H. (1994). From damselflies to pterosaurs: how burst and sustainable flight performance scale with size. Am. J. Physiol. *266*, R1077–R1084.

115. Askew, G.N., Marsh, R.L., and Ellington, C.P. (2001). The mechanical power output of the flight muscles of blue-breasted quail (*Coturnix chinensis*) during take-off. J. Exp. Biol. *204*, 3601–3619.

116. Lewis, P.O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. *50*, 913–925.

117. Goloboff, P.A., and Arias, J.S. (2019). Likelihood approximations of implied weights parsimony can be selected over the Mk model by the Akaike information criterion. Cladistics *35*, 695–716.

118. Tuffley, C., and Steel, M. (1997). Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. *59*, 581–607.

119. Goloboff, P.A., Pittman, M., Pol, D., and Xu, X. (2019). Morphological data sets fit a Common Mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. *68*, 494–504.

120. Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M. (1996). Phylogenetic inference. In Molecular Systematics, D.M. Hillis, C. Moritz, and B.K. Marble, eds. (Sinauer), pp. 407–514.

121. Felsenstein, J. (2004). Inferring Phylogenies (Sinauer).

122. Maddison, W.P., and Maddison, D.R. (2019). Mesquite: a modular system for evolutionary analysis. Version 3.61. http://mesquiteproject.org.

123. Humphrey, P.S., and Livezey, B.C. (1982). Flightlessness in flying steamer-ducks. Auk *99*, 368–372.

124. Livezey, B.C. (1989). Flightlessness in grebes (Aves, Podicipedidae): its independent evolution in three genera. Evolution *43*, 29–54.

125. Ouellet, J.F., Guillemette, M., and Blier, P.U. (2008). Morphological and physiological aspects of takeoff aptitudes of female common eiders (*Somateria mollissima*) during the pre-laying period. Can. J. Zool. *86*, 462–469.

126. Livezey, B.C. (1988). Morphometrics of flightlessness in the Alcidae. Auk *105*, 681–698.

127. Christiansen, P., and Fariña, R.A. (2004). Mass prediction in theropod dinosaurs. Hist. Biol. *16*, 85–92.

128. Campione, N.E., and Evans, D.C. (2012). A universal scaling relationship between body mass and proximal limb bone dimensions in quadrupedal terrestrial tetrapods. BMC Biol. *10*, 60.

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Phylogenetic data matrix	This paper	.tnt file in 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk
Ancestral state reconstruction analysis and direct calculation of wing loading and specific lift	This paper	.xlsx file in 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk; Methods S1
Parsimony-based phylogenetic analysis and mapping	This paper	'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk; Methods S1
Likelihood mapping of wing loading and specific lift	This paper	'Tests with Mk model' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk; Methods S1
Software and Algorithms		
New phylogenetic analysis scripts	This paper	'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk
New parsimony mapping scripts	This paper	'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk
New likelihood mapping scripts	This paper	'Tests with Mk model' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk
Tree analysis using New Technology (TNT) V 1.5	http://www.lillo.org.ar/phylogeny/tnt/	N/A

RESOURCE AVAILABILITY

Lead Contact

Michael Pittman (mpittman@hku.hk).

Materials Availability

Further information and requests for resources and regents should be directed to and will be fulfilled by the Lead Contact.

Data and Code Availability

The data and code generated or analyzed during this study (including new custom computer code) are available from the Lead Contact Michael Pittman (mpittman@hku.hk) and on the public repository Dryad: https://doi.org/10.5061/dryad.866t1g1nk.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental models are coelurosaurian theropod dinosaur fossils deposited in international public repositories. The authors obtained institutional permission to study these specimens and collect data from them. Other scientists can freely access these fossil specimens for scientific study. For additional information, see .xlsx file in 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk.

Phylogenetic Dataset

The coelurosaurian theropod dataset of Brusatte et al. [1] - the most recent version of the Theropod Working Group dataset (TWiG dataset) - was significantly expanded with data pertinent to paravian phylogeny, especially data concerning dromaeosaurids. Nine dromaeosaurid terminals were added to this expanded version of the TWiG dataset for the first time, including the Late Cretaceous microraptorine IVPP V22530, *Changyuraptor*, *Zhenyuanlong*, *Luanchuanraptor*, *Acheroraptor*, *Linheraptor*, *Yurgovuchia*, *Dakotaraptor* and *Velociraptor osmolskae*. The current dataset has thirty-one dromaeosaurid taxa, including all valid genera that have been included in previous phylogenetic analyses, except for *Pyroraptor* (represented by a fragmentary specimen lacking recognizable synapomorphies [2]). Codings of many other dromaeosaurids, troodontids, and early-diverging avialans have been revised, including *Anchiornis*, *Aurornis*, *Eosinopteryx* and individual specimens of *Archaeopteryx*. Codings for several non-paravian maniraptorans have also been revised or added for the first time e.g., the scansoriopteryid *Yi*. Phylogenetic coding changes in this study were relative to Brusatte et al. [1] and other past studies. Terminals added were the following: IVPP V22530, *Changyuraptor*, *Zhenyuanlong*,

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

Luanchuanraptor, *Acheroraptor*, *Linheraptor*, *Yurgovuchia*, *Dakotaraptor*, *Velociraptor osmolskae*, *Sinusonasus*, *Archaeopteryx* London specimen, *Archaeopteryx* Berlin specimen, *Archaeopteryx* Munich specimen, *Archaeopteryx* Eichstätt specimen, *Archaeopteryx* Thermopolis specimen, *Archaeopteryx* Haarlem specimen, *Archaeopteryx* Solnhofen specimen and *Archaeopteryx* 11th specimen. Terminals with revised codings were as follows: *Dromaeosaurus*, *Deinonychus*, *Velociraptor mongoliensis*, *Balaur*, *Tsaagan*, *Bambiraptor*, *Tianyuraptor*, *Sinornithosaurus*, *Microraptor*, *Graciliraptor*, *Hesperonychus*, *Rahonavis*, *Buitreraptor*, *Neuquenraptor* + *Unenlagia*, *Austroraptor*, *Shanag*, *Mahakala*, *Atrociraptor*, *Utahraptor*, *Adasaurus*, *Saurornitholestes*, *Troodon*, *Jinfengopteryx*, *Mei*, *Sinovenator*, *Sinornithoides*, *Byronosaurus*, *Xixiasaurus*, *Saurornithoides*, *Zanabazar*, *Epidexipteryx*, *Sapeornis*, *Jeholornis*, *Incisivosaurus* and *Caudipteryx*.

All taxa were coded based on first-hand examinations, relevant literature and photographs. Some codings for the newly included taxa e.g., *Acheroraptor*, *Yurgovuchia*, *Dakotaraptor* and *Yi* were also adopted from non-TWiG datasets [28, 30, 32, 33]. Four *Archaeopteryx* specimens (Eichstätt, Berlin, Haarlem and Munich) were re-examined first-hand using Laser-Stimulated Fluorescence (LSF) imaging [97], revealing additional anatomical details.

The character list of Brusatte et al. [1] consists of 853 characters compiled from multiple sources. A new character state was added to character 229 to reflect a potential synapomorphy of *Archaeopteryx* specimens on the ischium (marked with *). A new character state was also added to character 744 to reflect the variations of pedal phalanx II-2 in deinonychosaurians (marked with *).

Character 229 is as follows: Ischium, distally placed dorsal process along posterior edge of shaft

0: absent; 1: present; 2*present, and forming a narrow notch distally

Character 743 is as follows: Pes, phalanx II-2, length

0: more than 60% of length of pedal phalanx II-1; 1: less than 60% of length of pedal phalanx II-1; 2*: more than 60% of length of pedal phalanx II-1, and pedal phalanx with distinctly enlarged condyles and ventral heel

METHOD DETAILS

Automated phylogenetic analysis

The phylogenetic analysis was carried out with TNT version 1.5 [98, 99]. In order to make the analysis fully reproducible and less time-consuming to run, a master script was used to automate thorough searches, as well as the subsequent diagnosis and characterization of results. All the scripts and batch files for initial analysis, diagnosis, and other tasks, are available on Dryad: https://doi.org/10.5061/dryad.866t1g1nk READ ME FIRST file provides step-by-step instructions.

Tree searches

Tree searches used the extended search algorithms of TNT (initially using 5 random addition sequence Wagner trees followed by TBR, sectorial searches [CSS, RSS, and XSS], and 5 cycles of tree-drifting, followed by tree-fusing (see [100] for details). The search calculated the consensus as trees of optimal score were repeatedly found (eliminating branches of minimum length zero [101]; subsequent consensus calculation after pruning rogue taxa more conservatively collapsed trees with TBR branch-swapping [102]), stopping the search only when the consensus becomes stable to new hits, so validating the accuracy of the consensus for the corresponding dataset and optimality criterion. Thus, optimal score was found independently as many times as needed to obtain a stable consensus; this validation procedure was performed three times for greater reliability. A common problem in palaeontological studies is the extremely large numbers of equally parsimonious trees (due to missing entries), but the consensus stabilization procedure can produce an accurate consensus without finding all equally possible most parsimonious trees (thus greatly saving in computational effort [100]). The trees found by the routines used, therefore, are a representative sample of all possible most parsimonious trees (instead of all the trees), and their consensus is expected to be identical to the consensus of all possible most parsimonious trees for the dataset. For all the analyses performed here, trees were rooted on *Allosaurus* (following previous TWiG analyses [1, 2]).

Parsimony and character weighting

All of our analyses included implied weighting (XIW) [103], with modifications proposed by [104] that prevent improper inflation of weights due to missing entries (downweighting the weight of characters with missing entries faster). The missing entries were assumed to have 0.8 of the homoplasy in observed entries, but not extrapolating beyond 5 times the observed homoplasy. The analyses used a concavity of 10, which weights mildly against characters with homoplasy (see 'Parsimony analysis' folder on Dryad (https://doi.org/10.5061/dryad.866t1g1nk) for the cost of adding a step to each of the characters in the dataset, for the trees optimal under XIW; for ca. 80% of the characters, the cost of adding a step is within a 5:1 ratio, relative to the cost for no homoplasy; for only 8% of the characters is the cost of adding an extra step above a 10:1 ratio, relative to no homoplasy).

Identification of rogue taxa

The rogue taxa were identified automatically with TNT, using two types of routines combined. First, the taxa were identified heuristically saving to a list those (combinations of) prunes that improve resolution, with the commands *prunnelsen* (which tries combinations of taxa pruned for each polytomy), in the case of strict consensus trees, and the command *chkmoves* (which swaps taxa with TBR and records taxa that can be moved within a certain distance from the original position with a specified difference in score), for the case of measuring group supports. The initial heuristic list is then refined with a new technique (implemented in the *prupdn* command of TNT), which takes possible combinations of taxa from the list produced heuristically, measuring whether the resolution or the overall degree of support is increased when the taxon combination is added to, or removed from the consensus. The refinement procedure is potentially more time consuming, but the number of combinations to try is reduced by taking candidates from the initial heuristic

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

list. This two-step procedure allowed us to produce reduced consensus trees with a good degree of resolution or support, fully automatically. The two-step procedure was first applied to increasing the general resolution of the tree, and then applied (with more intensive parameters) to resolving reduced trees comprising only the Dromaeosauridae (which in being a focus of the present paper includes a significant proportion of wildcards, so that stable taxa are harder to identify). See Methods S1 for details on specific strategies and parameters used in the scripts, and a description of the optimality criterion used to evaluate a given set of taxon prunings in the refinement step. The rogue taxa were excluded from the trees when calculating the consensus; they were never excluded from the matrix in any search for optimal trees (thus, the character information in the rogues can influence the relationships of the other taxa; see [105] for a recent discussion of this problem).

The identification of rogue taxa (both for improving the resolution of the strict consensus, and for improving the group support values) was carried out in two steps, first identifying a list of possible rogues with a heuristic procedure, then selecting from that list with an optimality-based method. The identification of rogues for the strict consensus was carried out with a routine implemented in a script called *improvecombin*, and for the supports (i.e., frequency differences) with a routine implemented in a script called *bothprunes*.

To improve the strict consensus tree (for EW, XIW, and EW-XIW combined), the initial heuristic list for the whole tree was created (with *prunnelsen =3/>heuristic_prunset*) by exhaustively trying all triple combinations of nodes connected to polytomies, saving to the list all cases of prunings that increase the resolution of the tree. If alternative sets of taxa provide different ways to improve the consensus trees, both will be saved, so that the final list may contain more taxa than the minimum needed for optimal resolution. This is why the second stage, using an optimality criterion, helps refine the identification of rogues. This is a command implemented in TNT as part of the research for this paper, *prupdn* (for prune-up-and-down). The command takes the list of heuristic candidates, and (in the "up" mode, the "<" option) reinserts into the consensus combinations of up to 2 taxa (the "=2" option), evaluating the result of the reinsertion with $E = (P + \sum v) / (T - 2)$, where $\sum v$ = sum of support values across all branches of the resulting tree, T = number of taxa for the complete tree (i.e., with the least possible prunings), and P = penalty for pruned taxa. The penalty P is calculated as $P = R \times 100 \times (1 - F^2)$, where R = number of taxa removed (relative to the full taxon set), and F = factor to penalise removal varying between 0 and 1 (with larger values providing a stronger penalty and thus accepting only those prunings that improve the tree more; for the entire tree, a factor of 0.5 was used, with the option ":0.5"). When evaluating E for the pruning for strict consensus (i.e., with the "&100" option), the value v for each group is either 0 or 1, and the program in this case saves time by not pruning combinations of taxa that attach to different polytomies in the full consensus. Thus, to summarize, for the entire tree, the options used were *prupdn heuristic_prunset final_prunset :0.5 =2 < &100*.

Since the Dromaeosauridae included several very fragmentary taxa, that jointly float in the tree, their identification required more intensive parameters, with more taxa pruned in the initial heuristic stage (5 taxa, with *prunnelsen =5/>heuristic_prunset*), and combinations of up to 3 taxa (instead of 2) reinserted into the tree, evaluating prunes with a factor of 0.9 (instead of 0.5, so that pruning a taxon is more costly), with the command *prupdn heuristic_prunset final_prunset :0.9 =3 < &100*.

The same procedure was used for improving the consensus of EW, XIW, and EW combined with XIW (in the *bothprunes* script).

For improving the group supports, it was necessary to use more trees than just the optimal trees, since a taxon with a unique position in the optimal trees may nonetheless move around with minimal differences in score, thus decreasing group supports. To detect which taxa may be pruned to further improve group supports, then, the initial heuristic list was obtained by subjecting 11 arbitrarily chosen optimal trees to a round of TBR swapping, accepting moves suboptimal by a score difference of 2 and a relative score difference [102] of 0.25 (i.e., equivalent to the conflict of 3 versus 4 characters), and recording the list of taxa that can be moved to up to 5 nodes away from the original position, in the case of EW, and up to 3 nodes away, in the case of XIW (i.e., *chkmoves [0.10 /5 > heuristic_prunset*). The initial heuristic prunset included both the taxa that move with small score differences under either EW or XIW. The taxa identified as rogues for the individual sets of EW, and of XIW trees, were added to this heuristic set of prunes. Then, this set of possible prunes was submitted to the iterative improver of the *prupdn* command, in the default "down" mode (i.e., pruning combinations of up to 4 taxa from the heuristic set), to work on the set of trees obtained by resampling. The evaluation in this case considered the frequency differences (the "*" option), for the groups in the reference tree (the tree obtained with the prunings found by the previous routine, for the EW results, or for the XIW results, or *base* prunings), with a penalty factor for taxon exclusion of 0.5. This is accomplished with the command *prupdn heuristic_prunset final_prunset =4 :0.5 *[0] 1. / (base)*.

Taxonomic correspondence

The taxon names incorporated the full taxonomic hierarchy, which TNT can subsequently use for automatic diagnosis of results [106]. Tree branches were color-labeled with the taxonomic group that most closely matched the corresponding set of taxa after pruning rogues (the number of taxa in the reference taxonomic group that are removed from or added to the group in the tree are preceded with a minus [−] and plus [+] sign respectively; groups with no indication are fully monophyletic). The taxonomic correspondence of the groups in the tree was evaluated after pruning rogue taxa.

Synapomorphy lists

Lists of synapomorphies cannot be meaningfully produced by mapping consensus trees (as they are less parsimonious than the original optimal trees used to produce the consensus, and can imply synapomorphies not implied by any of the optimal trees). Once the rogues had been identified, lists of synapomorphies were produced by pruning the rogues from all trees, and then finding the list of synapomorphies in each of the trees. The changes common to all the pruned trees were then plotted on the strict consensus. As the

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

lists of synapomorphies must be plotted on the pruned consensus (i.e., for better resolution), the rogue taxa had to be pruned from the trees for character optimization (thus, there may be minor differences in the clades corresponding to the possible alternative placements of the rogues).

Group supports and conflict

Group supports are typically assessed by means of resampling. A problem in datasets with many missing entries is that standard measures of support may produce very low values for groups characterized by few synapomorphies, even if contradicted by no (or few) characters. Resampling may be intended to assess several aspects of the support (see review in [68, 107]); the goal of evaluating whether there is conflicting information regarding the monophyly of a given clade is best served if the prior weights of the characters are altered (as in bootstrapping or jackknifing) but no character is eliminated. Also, since some of the analyses performed here use implied character weighting, the resampling should increase or decrease the weight of the characters with the same probability (i.e., a no-zero weight symmetric resampling) to avoid distortions produced by different implied weights [68]. Thus, the prior weights were doubled or halved, with the same probability of 25%. The results (for 100 pseudoreplicates) were summarized by means of frequency-differences (GC; [68]), for the groups in the (pruned) strict consensus; frequency differences avoid problems associated with using raw frequencies to evaluate poorly supported groups, and measure support on a scale between 0 and 100% (unlike raw frequencies, which can only be meaningfully used to measure supports for groups with frequencies above 50%). By virtue of using this method, groups that are contradicted by almost as many characters as support the group will receive values approaching 0. Note that the values obtained by using the no-zero weight symmetric resampling cannot be interpreted in the same way as standard bootstrap values (typically much lower, often seen as measuring statistical significance [107]), but are best seen instead as a measure of the degree to which the characters in the dataset present conflict in regard to the monophyly of the group. For simplicity, the term "group support" is used throughout the paper, but the specific meaning of the evaluation carried out should be borne in mind.

Another problem that had to be taken into account for producing the final evaluation of groups is that superficial tree-searches (often used to speed up support calculations) may easily fail to find trees of optimal score in some of the resampled matrices in such complex datasets, thus decreasing even more the degree to which supported groups are consistently recovered. Thus, a more accurate evaluation was obtained by using a search strategy similar to that used for the observed dataset, but hitting the best score (for the resampled dataset) 4 times. In a first step, a single uncollapsed tree was saved from each pseudoreplicate. These uncollapsed trees can be used to identify more easily the rogue taxa that decrease group supports. After identifying the rogues, for the final calculation of supports, all trees of equal score in the TBR-neighborhood were considered for each pseudoreplicate. For doing this, instead of searching again, the datasets were created by resampling with the same sequence of random seeds, and using the optimal trees already found by the first step (this avoids having to repeat the computationally costly searches of the first step).

In the *bothprunes* script, the group supports combined were evaluated (as explained in the main text) by reading into TNT the trees saved in a previous resampling to a file, both for EW and XIW, using after rogue identification those trees as starting point (for TBR collapsing) in a resampling with the same sequence of random seeds. The results that use XIW as base considered the supports for the topology with the rogue taxa identified for the strict consensus tree, showing the values above branches; if additional prunes are identified as improving the group supports for those groups (with the *chkmoves* and *prupdn* combination described in the previous section), then the values for the improved groups are shown preceded by an asterisk (only if better than those prior to the additional pruning). These asterisk-preceded values are those obtained when additional taxa are pruned from the trees, and these additional taxa are marked on the tree with a triple slash. Note that in some cases, a group with the additional prunings may correspond to two or more nodes of the tree with a larger taxon set [e.g., the node common to DE in the tree (A(B(DE))) may correspond to either the node DE or the node CDE in the tree with the larger taxon set (A(B(C(DE))))]. In that case, the value for the pruned tree is displayed in the smallest corresponding group (i.e., DE, instead of CDE).

When using XIW as base, then the values of support under EW are indicated below the branches. To obtain those, the resampled trees resulting from analyzing under EW were read, and reused for collapsing (under EW). This is the proper evaluation of the support under EW, of a group found by XIW. When the additional set of prunings improves the support under equal weights, then that value is shown below the branch, preceded by a star.

When using the EW results as base for the supports, the sets of trees are switched. These results are not identical to those using XIW as base because the initial set of taxon prunings is the one corresponding to the EW strict consensus, not to the XIW strict consensus. Thus, the values may be slightly different, and some of the additional prunings may be different.

Evaluation of character conflict

In addition to automatically identifying the rogue taxa, the script used also facilitates discovering the characters (if any) responsible for the alternative placements of the rogues. All the trees produced by the XIW analysis had the same numbers of steps for each characters (so that the alternative placements of all rogues in the XIW results must correspond only to missing entries or ambiguities in optimization). But in the case of EW, there were differences in steps, for many characters, between trees with different topologies. Then, identifying these characters helps identify the cause of the conflicting resolutions.

To better evaluate conflict, for a given location of the rogue taxon, the steps for each of the characters in the matrix were stored in memory (with the *xcomp* command of TNT). Then, the number of steps for each of the characters was compared on trees where the rogue was located in alternative optimal placements. These trees with alternative placements were found by (a) taking one of the distinct topologies for the full taxon set, (b) creating a set of skeleton constraints for each of the groups in the tree, with the rogue taxon in question left floating (i.e., constraints allowing the rogue taxon to be either inside or outside of each of the constrained groups), and (c) performing a TBR search under the optimality criterion and settings in effect. The resulting trees will then display

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

the rogue taxon in alternative placements, but all the rest of taxa in fixed positions, so as to filter out the differences in character fits that have to do with the relationships of non-rogue taxa. When no differences are found in the fit of the characters in the resulting multiple trees, then the taxon is inferred to be a rogue only as a result of ambiguity in optimisations, or missing entries, instead of actual conflict between characters (and thus not amenable to resolution by a different method of analysis). This method is a simplification, as some of the alternative placements of rogue taxa may possibly occur only given some specific resolution for the rest of the tree, and those placements would be missed with the present approach (which considers only one resolution for the rest of the tree). Thus, the method helps to identify some of the characters responsible for the alternative locations of rogues, but not necessarily all of them (an improved method would use different topologies; it would be relatively easy to modify the scripts to use multiple topologies).

A similar procedure was used to identify the characters responsible for the different resolutions of the trees that do not correspond to rogue taxa. In that case, all the rogue taxa were pruned from the full set of optimal trees, and the remaining polytomies in the consensus tree were identified. For each of these polytomies, a set of constraints was created from one of the trees with pruned rogues, excluding from the constraints all the groups that correspond to the polytomy in question. A subsequent search under those constraints will thus leave the rest of the tree fixed (filtering out step differences in characters that have to do with conflict in *other* parts of the tree), and find alternative resolutions of the polytomy in question. Then, a character-by-character comparison of the step differences between the trees found by such constrained search will help identify conflicting characters (if none, different resolutions will be the result of ambiguity due to alternative character optimisations or lack of information due to missing entries). As for individual rogue taxa, this is a simplification (given that the rest of the tree is fixed at one resolution for the analysis of each of the polytomies in the consensus); some of the characters responsible for the different resolutions will be identified by this method, but not necessarily all of them. A more thorough analysis would require using different trees from the original set as basis for constraints.

Reconstruction of wing loading and specific lift values

Two criteria – wing loading and specific lift – were taken from theoretical and *in vivo* work on extant avialans and applied to fossils; they present easily testable benchmarks to discern volant from flightless taxa [44–46, 70]. For taxa without preserved complete primary feathers (e.g., *Rahonavis* and *Mahakala*), feather length was modeled on closely related taxa (for additional information see 'ancestral_state_analysis_and_direct calc_of_WL_SL.xls' in the 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk). We did not reconstruct pennaceous feathers on our outgroup taxon (*Allosaurus*) as there are reported to be absent [108, 109]. Wing area was calculated based on the methods presented in Dececchi et al. [18] (for additional information see 'ancestral_state_analysis_and_direct calc_of_WL_SL.xls' in the 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk). Wing span was taken as 2.1 times the summation of the lengths of the humerus, ulna and metacarpal II and the longest distal primary [18]. Wing chord was taken as 55% of the longest distal primary length, a modification of the methodology used in [18]. This is because it better reflects the differences between primary and secondary lengths seen in *Microraptor* [110] and produces wing area estimates that are within less than 1% of those measured by Yalden [111] for *Archaeopteryx* and by Lü and Brusatte for *Zhenyuanlong*. For avian theropods, wing length, wingspan and wing area were based on the reconstructions of Serrano et al. [112]. This was adopted to better reflect the potential differences in wing size and shape in later-diverging birds. Scansoriopterygians are included in the phylogenetic analysis but are excluded from the flight parameters because *Yi*'s wing is skin-based rather than feather-based like the other winged taxa in this dataset [33], while *Epidexipteryx* does not possess pennaceous feathers [113]. For *Rahonavis*, given only the radius and ulna are known, we reconstructed its wing with similar intralimb proportions to *Microraptor* where the ulna is ~37% of the forelimb length. A maximum muscle mass-specific power output (Po,m) value of 225 Wkg^{-1} was suggested by Marden [114] as the mean value for burst flight in birds. Work by Guillemette and Ouellet [46] suggested a range between 225-250 Wkg^{-1} more accurately mimics values seen in a Common Eider, a bird with short wings that displays some of the highest wing loading values seen in extant birds, two features that resemble the condition seen in the extinct taxa examined here. 287 Wkg^{-1} was based on the values calculated for Chukar partridges [115], a short burst flight taxon previously used as a model for early flight in theropods. To improve optimization of the data we screened these coelurosaurians from 77 to 43 taxa based on their presence of vaned feathers which are integral to the production of aerodynamic forces; terminals for which feather condition is unknown were considered to have the same state as their ancestor, which is the condition predicted by our phylogenetic hypothesis (if absent marked as '—'). Our primary mapping results used parsimony (for data and scripts, see 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk).

The reviewers requested that we include an ML analysis. Many phylogenetic studies use approaches based on maximum likelihood and specific models; for morphological analyses, the most commonly used model is Mk [116]. To apply that model, the estimates of flight capability were converted into a discrete variable, indicating whether the wing loading and specific lift thresholds were crossed in each of the terminal taxa. This discretized (0/1) character was then mapped by means of likelihood onto the tree obtained. Perhaps the most important assumption of the Mk model is that character changes are more likely to occur on longer than on shorter branches (branch length being a composite of duration and mutation rate); the choice of branch lengths may critically affect the mapping, but which branch lengths should be used to map this discretized character is far from obvious, and there are four main ways to do this.

The first alternative is to use lengths equal for all the branches of the tree, estimated from the data (as in [117]), thus leading changes in the mapped character to occur equiprobably in any branch of the tree. Uniform branch lengths were tested, and they – expectedly – produced results (Methods S1O and S1P) very similar to the parsimony mappings (Methods S1G–S1J; see [117]), with several

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

independent originations of flight capability in *Microraptor*, *Rahonavis*, and Avialae. Uniform branch lengths seem appropriate when we have no knowledge of whether changes are more likely along some branches (and which branches).

A second alternative is estimating branch lengths solely from the character to be mapped. This is mathematically guaranteed to produce the same results as a parsimony mapping [118], so it would add nothing to the parsimony-based analysis.

A third alternative is using actual dates for branch lengths. Actual dates for branch lengths are often (and perhaps appropriately) used in some studies (e.g., biogeography), but they are difficult to obtain and seem misplaced for studying alternate theories of morphological evolution – doing so deprives the analysis of one of the big advantages of standard maximum likelihood methods, namely not having to assume that mutation rate is constant through time. If branch lengths depended only on time, a morphological "clock" would be in effect, and all extant sister terminal taxa would have the same amount of morphological divergence, which is a strong assumption rejected by many empirical datasets.

The fourth alternative is to estimate branch lengths from the whole dataset, which is the most common course of action in comparative studies, but it is far from ideal, as it assumes that the rates of evolution along each branch of the tree is the same for all characters, including the character being mapped; empirical analyses strongly reject such uniformity in morphological datasets (see Goloboff et al.'s [119] analysis of 86 datasets).

Despite our reservations, at the specific request of the reviewers, the branch lengths were calculated on (one of) the trees found with XIW, using standard methods (see [116, 120, 121]), excluding the discretized flight characters (the tree with optimized branch lengths is included in the 'Tests with Mk model' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk). For every one of the alternative variables used to measure flight capability (i.e., wing loading and specific lift with Po,m = 225, 250 and 287 Wkg^{-1}), the likelihood for both states was calculated at each node; as is standard (e.g., see documentation for *Mesquite* v3.61 [122]), the proportion to the total likelihood contributed by each state was used for mapping the character. Although based on principles rather different from those of parsimony, such mappings (Methods S1O and S1P) were very similar to the parsimony ones (Methods S1G–S1J), indicating separate increases of flight capability, relative to their ancestors, for *Microraptor* (except when Po,m = 225 Wkg^{-1} for specific lift), *Rahonavis* and Avialae. The scripts, data and results used in the ML analyses are included in the 'Tests with Mk model' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk.

In addition to the uncertainty in calculating wing loading and specific lift in fossil taxa near the threshold of flight capability, there is of course uncertainty in the phylogenetic conclusions. The trees recovered by our EW and XIW analyses imply that flight capability evolved at least three independent times (with no, or at most one, secondary loss; this depends on which variable is mapped). A typical way to assess the confidence on such a conclusion is to generate the best possible trees where flight originates a single time, then comparing the number of additional steps required by those trees over the best possible ones. That is easily accomplished by giving a high cost to the 0→1 transformation (i.e., acquiring flight) relative to 1→0 (i.e., becoming non-volant again); this requires that the root of the tree is forced to be non-volant (with the *ancstates* command of TNT). The results of finding the optimal trees under such a cost regime, however, are in this particular case of no help in testing the support for the hypothesis of multiple flight origins: under EW, the trees obtained have the exact same length for the rest of the characters, with flight originating only once at the Paraves, and becoming secondarily lost between 6 to 11 times (depending on the variable being mapped; for wing loading it is lost the fewest, and for specific lift with Po,m = 225 Wkg^{-1} it is lost 11 times). Under XIW, the trees with a single 0→1 transformation have minor differences relative to the unconstrained trees, but the same result as with EW is obtained: the single origination does not result from placing volant taxa together, but instead from keeping them separate and mapping non-volancy as lost numerous times (the minor differences in tree topology under XIW result from uniting some secondarily non-volant taxa on the basis of this reversal). Comparing the trees, therefore, cannot provide clues as to the degree of support for the hypothesis of multiple originations of flight capabilities – the hypothesis depends on how the *character* is mapped, not on which *phylogeny* is preferred. There has been a lot of work in statistical tests for comparing trees, but almost none for comparing alternative reconstructions on fixed trees. The only course of action, in such a situation, is comparing individual reconstructions on the same tree, where (a) flight originates multiple times and is lost never or only once, and where (b) flight originates a single time and is lost multiple times. To provide a number that is easily interpretable as a ratio between probabilities, we calculated the likelihood (with branch lengths optimized under the Mk model for the whole dataset, as done for the plots in Methods S1O and S1P) for individual reconstructions (a) and (b). Note that both reconstructions (a) and (b) were obtained under parsimony, with reconstruction (a) also optimal under the Mk model, but reconstruction (b) strongly suboptimal under likelihood; the comparison is precisely based on calculating how suboptimal that second reconstruction is, given the assumptions of the Mk model. This depends on the variable being mapped, and the trees on which it is mapped, but the reconstruction with a single origination of flight capability is generally much less probable than the alternative. The two reconstructions that come closest in probability are the ones obtained for wing loading on the XIW tree, where the reconstruction (a), with multiple originations, is 2.5 times more probable than the reconstruction (b), with multiple secondary losses. For the 7 remaining combinations (wing loading on the EW tree, and all the alternatives of Po,m values for specific lift on either the EW or XIW trees), the reconstruction with multiple originations is at least 1.16×10^5 times more probable than the reconstruction with multiple losses. These numbers suggest a very strong statistical support for the hypothesis of multiple flight originations, although it should be borne in mind that they result from numerous simplifying assumptions, and that the uncertainty in estimating wing loading and specific lift themselves is large (and even harder to quantify).

Wing loading

Meunier and others demonstrated that volant extant birds always have wing loading values below 2.5 gcm^{-2} [44, 71, 123–126], and so the present study deems a fossil taxon with values above 2.5 gcm^{-2} as certainly flightless. Fossil taxa with values above 2.5 gcm^{-2}

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

are seen to have had the potential for powered flight. Wing loading is based on body mass estimated as per above (kg; see 'Trends in body mass change and forelimb length' in STAR Methods) over wing area (cm^2).

Specific lift

In the case of specific lift, the cut-off used to identify fossil taxa with the potential for powered flight is 9.8 Nkg^{-1} (gravity), as used by studies involving volant extant birds [45, 46]. In practice the value is slightly *greater* than 9.8 Nkg^{-1} since some lift is oriented as thrust in powered flyers [88]. Specific lift is based on Marden's model [45]:

$$\text{Specific lift} = \text{FMR x Po, m x}(L / P)$$

Where FMR is the flight muscle ratio which was assigned at a constant value of 10% of body weight across all taxa examined here. This is at the lower range of the values seen in volant birds and is likely a significant overestimation for all non-paravian taxa, though lower than those for *Archaeopteryx* and *Microraptor* based on recent 3D modeling work [76]. Po,m is the maximum muscle mass-specific power output based on values from extant birds. As Po,m is unknown for non-avialan theropods, two separate calculations were made that bracket the range of Po,m values that could have reasonably been expected (225 and 287 Wkg^{-1}; see Methods S1I and S1J as well as TNT scripts and script results in the 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk) to reconstruct minimum and maximum powered flight potential. L/P (lift/power) is calculated from:

$$\log_{10}(L / P) = -0.440 \log_{10}\text{muscle mass} + 0.845 \log_{10}(\text{wingspan} / 2) - 2.239$$

Trends in body mass change and forelimb length

Paravian body masses were calculated from femoral length measurements using the empirical equation of Christiansen & Fariña [127]:

$$\log_{10}\text{body mass} = -6.288 \pm 0.500 + 3.222 \pm 0.181 \times \log_{10}\text{femur length}$$

This methodology is a widely used estimator for body size across Theropoda [73]. While limb bone circumference has been shown to be a more accurate proxy of theropod body size [3, 53], this measurement was not available in many important Chinese paravian taxa because their long bones are crushed or flattened on mudrock slabs (a survey of ~1000 specimens covering dozens of species failed to recover reliable circumferences). Thus, the femoral length proxy was adopted because the measurement itself is available across our sample and because it has been widely used in previous theropod literature. As *Microraptor* is critical for our analysis we compared our mass value to one generated from an estimate of femoral circumference using the empirical equation of Campione et al. [128] as well as comparisons to mass estimates generated through 3-D computer and displacement methods [19, 22, 76]. All of these produce similar estimates to the one obtained using Christiansen & Fariña [127] (see Tables S1 and S2). The mapping of body mass change considered all possible reconstructions, and used the proportional change of size instead of the absolute differences. Thus, an increase of 0.5 units is considered as more significant if occurring from an ancestor with size 0.5 (i.e., increasing in 100%), than if occurring from an ancestor with size 2 (i.e., increasing in 25%). Size differences were normalized always relative to the smaller value (i.e., relative to the descendant instead of the ancestral value, in the case of decreases). Among all possible reconstructions, the maximum possible increases in size (or the minimum decrease) were calculated.

For completeness, the maximum possible decrease at each branch (or the minimum increase) was calculated as well. This was accomplished with a TNT script, *ptrends.run* (for "proportional trends"), available in the 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk. Trends in forelimb length were reconstructed in the same way as body mass. In consideration of the need for accurate geometric measurements across our microraptorine sample, relative changes in the length of the forelimbs and hindlimbs were estimated using a combined length of the humerus and the ulna for the forelimb (normalized with femoral length) and femoral length for the hindlimb. This is because the manus and the rest of the leg were not sufficiently preserved across our sample to estimate complete forelimb and hindlimb lengths meaningfully. For raw data, see script data and spreadsheet in the 'Parsimony analysis' folder on Dryad: https://doi.org/10.5061/dryad.866t1g1nk.

QUANTIFICATION AND STATISTICAL ANALYSIS

Uncertainty quantification and estimation confidence

One potential weakness of our modeling approach is the sensitivity to scaling assumptions in the assessment of locomotor performance. This sensitivity does not affect taxa recovered as far below thresholds for volancy, but it could potentially affect conclusions for those taxa recovered as performing near thresholds for volant behavior i.e., near powered flight potential. To address this, we used an iterative resampling method in which we varied the starting parameters and reran the analyses for taxa recovered as having performance estimates near threshold values. We found that our model was most sensitive to assumptions regarding specific lift, and so we focused resampling on varying FMR. As noted above, Po,m was automatically varied for all taxa by performing calculations at two values that encompass the range of maximum power outputs measured by prior teams (see section 'Reconstruction of wing loading and specific lift values'). Error in mass estimation was found to be less critical, with marginal performance taxa typically requiring both a significant deviation in wing area and a significant deviation in body mass from the expected values to change our expectations of volancy. However, varying body mass by applying standard errors from the mean as scalars is arguably not

Please cite this article in press as: Pei et al., Potential for Powered Flight Neared by Most Close Avialan Relatives, but Few Crossed Its Thresholds, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.06.105

the most robust method, since this ignores the underlying frequency distribution. To further validate masses for the most critical taxa (particularly *Microraptor*), we validated our estimates against wholly independent methods of mass estimation, including those derived through 3-D computer and displacement methods (see 'Trends in body mass change and forelimb length' section above). Validating our mass estimates against volumetric-based estimates is a particularly robust option because it allows us to eliminate extraneous potential minima and maxima that would result in unrealistic body densities (i.e., those well above or below those measured for living birds).