ABSTRACT

ALGER-MEYER, EVAN LAUREL. Assessment of Surface Shape of Ornament-like Cranial Structures in Birds: Are Bony Crests Convergent? (Under the direction of Dr. Terry A. Gates).

Prominent enlargements on the skull (bony ornament-like structures) are widespread in modern animals, particularly birds, and also occur in many fossil animals. These structures frequently appear independently in different lineages, and so are often described as convergent. However, few studies have examined whether these structures truly are convergent in their shape as assumed. This study uses multiple qualitative and quantitative approaches to assess: 1) whether birds with bony ornament-like structures are convergent with one another, 2) whether the region of the skull containing the bony ornament-like structure affects convergence metrics, and 3) whether any ecological factors correlate with the presence of the bony ornament-like structure. The study found that in most cases, these structures are not actually convergent, and may be divergent morphologically due to their diverse forms. For bony ornament-like structures with similar forms, specifically those found on the premaxilla, RRphylo *search.conv* and C-value analyses report evidence for convergence. This convergence is actually more likely to be an example of parallel evolution, as these taxa only group together loosely in their morphology but appear to have followed similar evolutionary trajectories. The ornamental region of the skull does appear to contribute positively to the strength of the signal for parallel evolution. However, even though these structures are often presumed to serve a function as sexual ornaments, the present study indicates that under most definitions, the taxa that have these structures are ecologically differentiable from the taxa that lack these structures, suggesting ecological factors likely influence the evolution of bony ornament-like structures as well. Sexual selection alone is not likely the sole influence on their evolution, though this study does not preclude the

possibility that bony ornament-like structures are affected by both sexual and non-sexual selection.

Assessment of Surface Shape of Ornament-like Cranial Structures in Birds: Are Bony Crests Convergent?

by
Evan Laurel Alger-Meyer

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Biology

Raleigh, North Carolina
2023

APPROVED BY:

_____	_____
Dr. Terry A. Gates Committee Chair	Dr. Brian Langerhans
_____	_____
Dr. Maria Servedio	Dr. Barbara Ballentine

DEDICATION

To my family, my lab group, and my cat, Allaidh, who all helped me through this journey. Thank you.

BIOGRAPHY

Evan Alger-Meyer grew up in Westminster, Colorado and cultivated an early interest in living and extinct animals. They carried this interest with them through high school when, in 2010, they joined the paleontology laboratory at the Denver Museum of Nature and Science, collecting paleoecology data for local Pleistocene fossil sites. They attended Washington University in St. Louis as an undergraduate in 2013, where they aided the work of four distinct biology and archaeology laboratories, completed a study abroad semester in Australia, and spent six weeks mapping geological formations in the Rocky Mountains and Black Hills. Evan earned their bachelor's degree in Earth and Planetary Sciences from Washington University in St. Louis in 2017, returning to Denver to work until 2020, when they were admitted into to a master's program studying evolutionary biology at North Carolina State University. Moving to Raleigh in the midst of the COVID-19 pandemic, Evan molded their prospective master's research project into the work contained in this manuscript. They aim to continue their research through a PhD program, using this work to address questions about how life has evolved in the past and will continue to evolve into the future.

ACKNOWLEDGMENTS

I would like to acknowledge the people who supported me at various points in my graduate career. A huge thanks to my advisor, Terry Gates, who took me on as his first graduate student and gave me valuable feedback throughout the planning, research, and writing process of this project. You helped me get here, and I can't thank you enough for that.

Thank you also to the members of the Zanno-Gates lab group who have offered insight into the complexities of evolutionary science, project feedback, and ample moral support throughout my program here. I'm lucky to have worked with this group of people. Thank you to Kyla Beguesse, Josh Hedge, Haviv Avrahami, Emi Bender, Willie Freimuth, Ty Harris, Evan Jevnikar, John David Kanipe, Thomas LaMaster, Aubrey Knowles, Luka Rolleri, Lindsay Zanno, Terry Gates, Lisa Herzog, Eric Lund, Chinzorig Tsogtbaatar, and Aurore Canoville. I would particularly like to thank Kyla Beguesse, my Paleosciences Peer Mentoring Program mentor, whose ceaseless enthusiasm and dedicated support helped carry me through my thesis defense. I would also like to thank Katie Berry, my Paleosciences Peer Mentoring Program mentee, whose hard work and sincerity inspire me to push myself. I hope I've done you all proud.

I would like to thank my committee members, Brian Langerhans, Barbara Ballentine, and Maria Servedio, whose expertise and guidance helped improve this project tremendously. Your kind words during my defense mean more than I can express.

Thank you to Theresa Johansson and Terry Gates, my teaching assistantship supervisors, for helping me become a better instructor to the undergraduates I taught, and for making my working livelihood during this time rewarding on many levels. Thank you also to my fellow TA, Josh Hedge, for the same.

Thank you to Freha Legoas for guiding me through the logistical part of graduate school and fielding my many questions over the years. Your kindness in this regard has been invaluable.

Thank you also to Brian O'Shea at the North Carolina Museum of Natural Sciences for facilitating my use of the ornithology collection specimens, and Justin Gladman at the Duke University SMiF lab for training me in use of the microCT scanner.

I would like to thank my previous mentors, Joe Sertich, Jennifer Smith, and Jill Pasteris, whose instruction and kind words helped me get into this program.

Lastly, I would like to thank my first paleontology mentor, Richard Stucky, who sadly passed away in 2022. His tutelage was formative for me in high school, setting me on my current path as an evolutionary scientist. I would likely not be where I am today if not for him.

TABLE OF CONTENTS

LIST OF TABLES viii
LIST OF FIGURES x

Chapter 1: Introduction 1
1.1: On Convergence 3
1.2: On Selection 10
1.3: On Bills and Bony Ornament-Like Structures 14
1.4: Defining and Testing Convergence in Avian Bony Ornament-Like Structures 20
Chapter 2: Materials and Methods 28
2.1: Data Collection and Landmarking 28
2.2: Phylogeny 35
2.3: Analyses 39
2.3.1: Principle Components Analysis and Hierarchical Cluster Analysis 39
2.3.2: Convergence Analyses 40
2.3.2.1 Stayton's C-values 40
2.3.2.2: RRphylo *search.conv* Function 44
2.3.2.3: Wheatsheaf Index 46
2.3.3: Ecological Analyses 49
2.3.3.1: ANCOVA 50
2.3.3.2: MANCOVA 50
2.3.3.3: Pearson's Chi-Square 51
Chapter 3: Results Section 1 53
3.1: Phylomorphoplot and PCAs 53
3.2: Cluster Analysis Dendrogram 57
3.3: C-Values 63
3.3.1: Test Definition 1.0–1.1 73
3.3.2: Test Definition 2.0–2.2 74
3.3.3: Test Definition 3.0–3.3 75
3.4: *search.conv* Time-Adjusted θ 78
3.5: Wheatsheaf Index 80
Chapter 4: Results Section 2 82
4.1: *search.conv* Time-Adjusted θ 84
4.2: Wheatsheaf Index 87
Chapter 5: Results Section 3 90
5.1: Visualizations of Ecological Data 92
5.2: Allometry 97
5.3: Phylogenetic ANCOVA 100
5.4: Phylogenetic MANCOVA 105
5.5: Pearson's Chi-Square Test 108
Chapter 6: Conclusions and Discussion 110
6.1: Conclusions 110
6.1.1: Question 1 Conclusions 110
6.1.2: Question 2 Conclusions 112
6.1.3: Question 3 Conclusions 113
6.2: Discussion 114

6.2.1: On Test Definitions for Bony Ornament-Like Structures 114
6.2.2: Notes on Analyses 121
6.2.2.1: pANCOVA, pMANCOVA, and Pearson's Chi-Square Tests 121
6.2.2.2: C-Values 123
6.2.3: Interpreting Convergence Using the Bolnick et al. (2018) Definition 130
6.2.4: Interpretating Convergence and Parallelism in Bony Ornament-Like Structures 133
6.3: Future Work 135

LIST OF TABLES

Table 1 Different definitions of convergence and parallelism reported by authors over the last two decades. Arbuckle et al. 2014, Stayton 2015, and Castiglione et al. 2019 all propose means of quantifying convergence that I have used for this study. 5

Table 2 Which methods will be used to address the three main questions in the present study. Light gray cells represent qualitative methods, while dark gray cells represent quantitative methods. 21

Table 3 The nine test definitions for ornamentation, organized by three categories (literature, personal, and premaxilla definitions) used in this study. Taxa that were tested as convergent and did have ornament-like structures are described as "ornamented" taxa, while taxa lacking these structures are described as "unornamented" for each of the test definitions. Please note that "ornamented" is a useful shorthand for discussion of the test results, and does not necessarily mean these taxa are using their bony ornament-like structure as sexually-selected biological ornaments. These definitions are purely morphological. n = the number of taxa designated as ornamented. The full 82 sample taxa were included in each test. 22

Table 4 Percent of variance accounted for by each principal component. 54

Table 5 The *search.conv* mean θ-values between species defined as ornamented, divided by the time distance between them (the ang.state.time output for the *search.conv* function), along with the associated p-values for the time-adjusted angles. The results are given for both the full skull and each of the bones in the skull that were landmarked for this study. Blue boxes indicate test-region subsamples with low p-values ($p<0.05$). For light blue boxes, $p<0.1$. 78

Table 6 Wheatsheaf index values and associated p-values for each of the different ornamentation definition tests. As with the θ-values from the *search.conv* analyses, the results here are given for both the full sampled skull and each of the individual cranial bones. Blue boxes indicate test-region subsamples with low p-values ($p<0.05$). P-values listed as 0 are not actually 0 but simply low enough that they exceed the reporting limit for R 4.1.2 and are rounded down. They may be accepted as $p<<0.05$. 80

Table 7 The top table shows the *search.conv* θ values for the "non-ornamental" region, and the bottom table shows the same for the ornamental region defined in the Materials and Methods. Orange values in the top table are those with $p<0.5$. Blue values in the bottom table are those where $p<0.05$ and the θ -value was greater (indicating greater convergence) compared to the non-ornamental region analyses. Orange values in the bottom table indicate the θ-value was either the same or lower in the ornamental region analysis for $p<0.05$ (for light orange, $p<0.1$). 84

Table 8 The top table shows the Wheatsheaf index values for the non-ornamental region, and the bottom table shows the same for the ornamental region defined in the Materials and Methods. Orange values in the top table are those with $p<0.5$. Blue values in the bottom table are those where $p<0.05$

(for light blue, p<0.1) and the w-value was greater (indicating greater convergence) compared to the non-ornamental region analyses. Orange values in the bottom table indicate the w-value was either the same or lower in the ornamental region analysis for p<0.05. P-values listed as 0 are not actually 0 but simply low enough that they exceed the reporting limit for R 4.1.2 and are rounded down. They may be accepted as p<<0.05. 87

Table 9 pANCOVA for weighted average PC score for bill surface landmarks compared to the absolute value of the centroid latitude of species range, weighted average PC compared to the log of species range size, weighted average PC compared to the log of mass, log of mass compared to absolute value of centroid latitude, log of mass compared to log of average range size, and absolute value of centroid latitude compared to log of species range size. P-values reported as 0 are rounded down by R, and should be taken as p<<0.05. 100

Table 10 pMANCOVA for weighted average PC score for bill surface landmarks compared to the log of average species mass, absolute value of the centroid latitude of species range, log of species range size, and interaction factors between mass and latitude, mass and range size, and latitude and range size. P-values reported as 0 are rounded down by R, and should be taken as p<<0.05. 106

Table 11 Chi-square results for habitat, trophic niche, trophic level, and primary lifestyle for the ornamented versus unornamented taxa defined by the different test definitions..... 108

LIST OF FIGURES

Figure 1 Illustration of definitions used by Bolnick et al. (2018) that show the five possible combinations of parallelism/nonparallelism and convergence/divergence. Parallel evolution by their definition is measured independently of convergence, and is assessed by the angle θ between vectors (θ is at or near 0 for parallel evolution and >0 for nonparallel evolution). Bolnick et al. (2018) define convergence as a smaller distance between the descendant taxa (D1, D2) than the distance between the ancestral taxa (A1, A2), with divergence being a greater distance between descendant taxa than ancestral taxa. According to this definition, both convergence and divergence can occur at the same θ values. 7

Figure 2 Visualizations for how the definitions for convergence described by the methods used in this manuscript compare to the definitions of Bolnick et al. (2018). Dots represent taxa in trait space, with red dots representing purported convergent taxa and black dots representing the rest of the taxa in the dataset. Phylomorphoplots connect taxa for the search.conv and c-values methods. A convex hull is drawn around the convergent taxa and full dataset for the Wheatsheaf index method. The search.conv method calculates adjusted θ values between the vectors of the focal lineages from ancestor to tip taxa, comparable to the θ defined in Bolnick et al. 2018. The c-values method compares distances between the end tip taxa (Dtip) to the greatest distance between taxa from the common ancestor to the end tip (Dmax). The Wheatsheaf index compares the distances between convergent taxa in trait space to the distances between all of the taxa in the dataset. 8

Figure 3 Examples of different forms of ornamentation used for the different test definitions in this study (shown in right lateral and anterior views; *Crotophaga* and *Fratercula* are also shown in dorsal view). The left side of the figure illustrates different levels of bony ornament-like structure prominence, with the last two categories lacking bony ornament-like structures, but the taxa with non-bony ornament-like structures having 3D ornament-like structures in life. The right side of the figure illustrates ornamentation groupings based on ornament-like structure location and general characteristics. The premaxillary bony ornament-like structures are highly varied but they can be broadly differentiated by the width of the ornament-like structure posterior base and the angle at which the ornament-like structure meets the frontal. Test definition 2.0 only defines taxa with prominent bony ornament-like structures as "ornamented." Test definition 2.1 defines taxa with either prominent or ambiguous bony ornament-like structures as "ornamented." Test definition 2.2 defines all taxa but those clearly lacking an ornament-like structure as "ornamented." Test definition 3.0 defines taxa with any premaxillary bony ornament-like structure as "ornamented." Test definition 3.1 defines taxa with premaxillary-wide (not backswept) bony ornament-like structures as "ornamented." Test definition 3.2 defines taxa with premaxillary-wide

(backswept) bony ornament-like structures as “ornamented.” Test definition 3.3 defines taxa with either type of premaxillary-wide ornament-like structures as “ornamented.” *Porphyrio poliocephalus* photograph adapted from Charles J. Sharp. 24
Figure 4 Approximation of sampled skull area for two taxa in the dataset: *Casuarius bennetti* (left) and *Uria lomvia* (right). The sampled area is highlighted in blue and yellow and includes the external surface of the frontal, nasal, and premaxillary bones. The blue region represents the region of the sampled area designated as “ornamental,” meant to represent parts of the skull where bony ornament-like structures occur in certain species. The yellow region represents the portion designated as “non-ornamental,” where bony ornament-like structures never occur in any of our sampled taxa. *Casuarius bennetti* has a well-defined ornament-like structure, much of which is formed from the nasal bone. *Uria lomvia* is defined as “unornamented” in all definitions. 33
Figure 5 The phylogeny used in the present study. See supplementary materials for details about branch lengths and the original study sources for those branch lengths..... 37
Figure 6 Phylogeny with nodes and groups depicted in the phylomorphoplot of Figure 7. *Paleognathae* (orange); *Galloanserae* (yellow); *Cuculiformes*, *Columbiformes*, *Musophagiformes*, and *Strisores* (teal); *Charadriformes* and *Gruiformes* (blue); *Pelecaniformes*, *Suliformes*, *Sphenisciformes*, *Procellariiformes*, *Eurypygiformes*, and *Phaethontiformes* (green); *Australaves* (lavender); *Piciformes*, *Coraciiformes*, *Bucerotiformes*, *Accipitriformes*, *Strigiformes*, and *Trogoniformes* (red); *Phoenicopteriformes* (pink)..... 38
Figure 7 PCA with superimposed phylogeny depicting the phylomorphospace of this study’s dataset. Different colors correspond to major clades defined by specific nodes on the phylogeny in Figure 6: *Paleognathae* (orange); *Galloanserae* (yellow); *Cuculiformes*, *Columbiformes*, *Musophagiformes*, and *Strisores* (teal); *Charadriformes* and *Gruiformes* (blue); *Pelecaniformes*, *Suliformes*, *Sphenisciformes*, *Procellariiformes*, *Eurypygiformes*, and *Phaethontiformes* (green); *Australaves* (lavender); *Piciformes*, *Coraciiformes*, *Bucerotiformes*, *Accipitriformes*, *Strigiformes*, and *Trogoniformes* (red); *Phoenicopteriformes* (pink). See Figure 6 and supplementary materials for precise node locations on the phylogeny. 54
Figure 8 PCA of bird taxa in this study, comparing PC1 and PC2. PC1 was also plotted against PC3, which contributed to 12.33% of the variance in the morphological data. The PC1 by PC3 plot showed similar overall clustering, so the slightly higher PC2 score is favored for all subsequent visualizations. However, the PC1 by PC3 plot is available in the supplementary materials. Note that even though *Balaeniceps rex* has a broad bill, it is unusual among the birds of this dataset in that a large portion of its bill (mainly the sides) is made up of the maxilla, not the premaxilla. Because the maxilla was not sampled in this study, the *Balaeniceps* bill sampled area appears to be long and thin, which is likely

why that taxon plots out with *Fregata* and *Limosa*, which are more traditionally thin-billed taxa. 55
Figure 9 PCAs for PCs 1 and 2 indicating spread and overlap between ornamented and unornamented taxa. Designations for presence of bony ornament-like structures are defined by test definition 1.1 (left) and 3.3 (right). Overall, there is very high spread and overlap. Test definition 3.3 shows the combined lowest overlap and tightest clustering of ornamented taxa out of any of the test definitions. 56
Figure 10 Orange dots correspond to ornamented taxa defined in test definition 1.1, blue dots correspond to ornamented taxa defined in test definition 3.3. The two most numerically distinct clusters of taxa are marked in orange and blue branches. Point A indicates the most closely clustered node containing Test 3.3 taxa. Note that these relationships are purely morphological and on their own do not account for similarity due to homology. 60
Figure 11 Scatterplot of branch lengths on the phylogeny compared to the dendrogram based on the VCV chart of all taxon pairings. The slope of the trendline is 0.307. A slope of 0 or infinity would mean no association between phylogeny and morphology while a slope of 1 would mean that morphology is perfectly predicted by phylogeny (as predicted by the Brownian Motion model). A slope greater than 1 would indicate that morphology changes dramatically with small increases in time (more time influence), and a slope of less than 1 indicates morphology changes only a small amount with larger increases in time (less time influence). The slope of ~0.3 suggests that time has a large influence on morphology, but it is not a one-to-one ratio. The high spread also indicates that the relationship between time and morphology is not especially strong, as many taxa deviate from the average and skew the trendline. Generally, taxa that share more time (left side of the graph) also have a higher morphological similarity (bottom of the graph). However, the shape spreads out over time, with several taxa deviating considerably in morphology from one another after only about 5-20 million years of separation. One would expect taxa above the trendline to be those that have diverged more rapidly than expected, while one would expect taxa below the trendline to be those that have either experienced morphological stasis or convergence to some extent. 61
Figure 12 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 1.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 63
Figure 13 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 1.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 64

Figure 14 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 65
Figure 15 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 66
Figure 16 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.2 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 67
Figure 17 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 68
Figure 18 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 69
Figure 19 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.2 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. 70
Figure 20 ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.3 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing. Because the TD 3.3 graphs had an extremely high proportion of C-values with values of 0, I removed the values of 0 in the bottom two density plot graphs to provide a more legible comparison between the different points. The bottom two graphs are not meant to offer a direct point of comparison to the density plots for the other tests aside test definition Test 3.2. 71
Figure 21 Comparison of non-ornamental to ornamental region *search.conv* θ-values for subtests where at least one θ-value had an associated p-value of <0.05. Lower θ-values indicate more convergence among the ornamented taxa.
Figure 22. Comparison of non-ornamental to ornamental region values for subtests where at least one w-value (Wheatsheaf index value) had an

associated p-value of <0.05. This chart is read differently than Figure 21; higher w-values indicate more convergence among the ornamented taxa. 85
Figure 22 Comparison of non-ornamental to ornamental region values for subtests where at least one w-value (Wheatsheaf index value) had an associated p-value of <0.05. This chart is read differently than Figure 21; higher w-values indicate more convergence among the ornamented taxa. 88
Figure 23 Bill surface landmark PCA with points colored by the absolute value of centroid latitude for species range. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 93
Figure 24 Bill surface landmark PCA with points colored by the log of species average mass. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 93
Figure 25 Bill surface landmark PCA with points colored by the log of geographic species range size. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 94
Figure 26 Bill surface landmark PCA with points colored by the primary lifestyle of the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 94
Figure 27 Bill surface landmark PCA with points colored by the typical habitat of the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 95
Figure 28 Bill surface landmark PCA with points colored by the general trophic level for the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 95
Figure 29 Bill surface landmark PCA with points colored by the general trophic niche for the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3. 96
Figure 30 Display of PC1 against PC2, with taxa colored by the log of the ratio of bill volume to mass. The blue convex hull denotes the TD 3.3 taxa while the orange hull denotes the TD 1.1 taxa. In general, the values increase across PC1, with birds on the right of the chart having larger bills relative to their body and birds on the left having smaller bills. There doesn't appear to be any association for this metric along PC2. Curiously, while several ornamented taxa occupy the upper right portion of the chart, the cassowaries, noted for having very large casques, are among the furthest left taxa and have a very low bill to body mass ratio. This is likely because although their casques are large, as flightless birds, their bodies are also much heavier than other birds. It is possible that the relationship between bony ornament-like structures and their cooling function could differ for

flighted birds compared to flightless birds, as flight is a major source of body heat production. 97
Figure 31 Plots comparing different measures of mass and volume against each other and PC score values. Ornamented taxa show different trends from the full dataset except for the log volume over mass compared to PC score. Spread is very high for the log mass and fairly high for the ornamented mass by volume, and the ornamented and full dataset volume by PC score trends. However, there is a tighter association between log mass by log volume for the full dataset and the log volume over mass compared to PC score for the ornamented taxa, as well as a very strong trend with very low spread for the ornamented taxa log volume over mass compared to the PC score. Taxa highlighted as "ornamented" are those listed in the Conclusions section for this study, as those are the consensus heavily ornamented taxa. 99
Figure 32 Pairwise comparison table of C1 values for all 82 taxa compared with one another. Darker green colors represent higher C-values, lighter values represent lower values. Values are expected to be low when morphology is dissimilar or phylogenetic distance is low (hence a white diagonal of values representing 0 when taxa are compared with themselves). 123
Figure 33 This chart shows a regression of the C1 values compared against a scaled ratio of phylogenetic time since separation represented by phylogenetic branch lengths. Overall, there is a small but significant decline indicating slightly higher C1 values on average for pairs of taxa that split more recently within the dataset. This effect is especially pronounced for taxa that diverged within ~20 million years (0-0.2 on the x-axis), indicating C-value measurements may have a tendency to report recently diverged taxa as more convergent than expected due to the influence of phylogenetic constraints. Datasets that include more closely related taxa for comparison could be less subject to this bias, but it merits future consideration for those who wish to use C-values. The spread of the data overall is high, indicating the potential bias toward higher C1 values in more recently-diverged taxa is an average trend; there are both very high and very low C-values across the range of divergence times. The C1 values for greater genetic divergence times (right side of the graph) likely reflect a realistic relative comparison of convergence values for the pairings. The vertical columns are a consequence of the way the bird phylogeny splits, with the column at 1 representing comparisons between paleognath and neognath species, and the column near 0.85 representing comparisons between galloanserine neognaths and all other neognaths. 127
Figure 34 This chart shows a regression of the C2 values compared against a scaled ratio of phylogenetic time since separation represented by phylogenetic branch lengths. Overall, there is a very small decline indicating slightly higher C2 values on average for pairs of taxa that split more recently within the dataset, similar to the linear regression for the C1 values. Here, the data is more evenly spread, and the bias toward higher C-values in more recently-diverged taxa is minimal, though may still be worth considering for certain studies depending on their focus. A regression

slope of 0 would indicate that the phylogenetic time separation has no influence on the reported C2 values, which is expected if the C2 values are providing an accurate calculation of morphological similarity adjusted by phylogenetic relatedness, as intended. Of the four C-values provided, the C2 values appear to be least influenced by the magnitude of time shared between taxa and might therefore be interpreted as providing the most suitable measurement for convergence out of the possible C-values..... 128

Figure 35 The above chart shows a regression of the C3 values compared against a scaled ratio of phylogenetic time shared for pairs of taxa. As with the other C-values, there is a slight but potentially important bias where higher C-values are reported for more closely-related taxa. The trend for the C3 values closely resembles that of the C1 values. It differs largely in that its values are overall much smaller than the C1 and C2 values, as expected by the C3 calculation, and the maximum C3 value for any given shared phylogenetic time (x-axis) tends to increase with increasing time shared (more recent divergence, to the right side of the graph). The reason for this is that C3 varies with the total variation within a lineage, which is generally greater for more distantly-related taxa. 129

CHAPTER 1

1: Introduction

Convergence is a biological phenomenon where separate lineages develop similar traits independent of one another. Trait similarity that is not based in homology offers independent examples of evolutionary processes and has attracted considerable attention from biologists throughout the history of evolutionary biology. Darwin himself warned about how similar, independently-acquired traits obscure historical relationships, and this consideration remains important for fields such as phylogenetics (Darwin 1872). Convergence can interact with many other processes and greatly complicate interpretation of an ecological context, so understanding its nuances is important for the study of evolutionary biology.

Although there are plentiful exceptions, convergence is generally expected when different lineages of organisms are affected by similar selection (Losos 2011). When convergent evolution occurs as a result of similar selection acting upon distinct lineages, the convergent lineages offer independent examples of how selection affects the evolution of particular traits (Losos 2011, Langerhans 2018). Non-sexual forms of selection are a common cause of convergence, such as the independent evolution of spiny hairs in mammals to protect against predators, or the evolution of long, thin bills among piscivorous birds. However, traits subject to sexual selection, such as ornaments, may also appear to be convergent in some species. Identifying traits that are convergent in different groups of organisms allows these traits to be studied in different contexts, including in the fossil record, which in turn means these traits can offer insight into selection on macroevolutionary scales. If traits in different groups of organisms can evolve convergently due to sexual selection specifically, convergence can potentially be used to study sexual selection on a macroevolutionary scale.

Numerous methods have been used to detect and quantify selection, and the basic premise that some individuals will be more likely to survive or successfully reproduce in a given situation can be observed directly, but most current methods of studying sexual selection in particular rely on behavioral observations in living populations (for example, Jones and Montgomerie 1992). As behaviors cannot be observed on long evolutionary timescales, detection of sexual selection in extinct populations has relied on ancestral state reconstructions, phylogenetic comparative methods involving extant species, and the presence of structures associated with sexual selection in modern animals, such as body size dimorphism and presence of presumed weapons or ornaments (Krishtalka et al. 1990, Regh & Leigh 1999, Peters & Peters 2009, Padian & Horner 2011, Knell et al. 2013, Mallon 2017).

Comparisons to modern analogues are often suitable to infer general trends in sexual selection. For example, Krishtalka et al. (1990) propose enlarged canines in an early primate likely evolved in response to sexual selection, and based on the consistency of canine dimorphism across various modern primates, this is a reasonable proposition. However, such comparisons become more tenuous for organisms without close modern relatives, necessitating studies that are capable of incorporating other analogues, such as taxa believed to have evolved convergently to extant organisms.

Exaggerated structures, especially when these appear to be dimorphic, are often assumed to be the result of sexual selection in fossil species (Padian & Horner 2011, Hone et al. 2012, Knell et al. 2013). Bony structures on the skull, like casques, crests, domes, shields, and humps (henceforth called “bony ornament-like structures” in this manuscript), are especially useful to study in fossil vertebrates because they can preserve well, are associated with elements that tend to be diagnostic for fossil species, and are found in a diverse number of modern taxa.

Gates et al. (2016) used bony ornament-like structures to investigate macroevolutionary trends in non-avian theropod dinosaurs, finding a positive association between the evolution of gigantism and the evolution of these structures. Animals in the dataset only showed bony ornament-like structures if above a certain body mass threshold, and adorned lineages evolved gigantism on macroevolutionary scales at a more sustained pace than unadorned lineages. Evolution of gigantism in predatory theropods is likely associated with interspecific interactions and ecology, but the association between bony ornament-like structures and body size in these species suggests that, if these structures did function as sexual display structures, body size evolution in these species may have additionally been driven by sexual selection and a mix of socio-sexual and ecological interactions (Gates et al. 2011).

Studies like this offer valuable insight into sexual selection in the fossil record over immense time scales. However, the use of bony ornament-like structures to study sexual selection in fossils relies on two important assumptions: 1) that different bony ornament-like structures are comparable to one another, and 2) that they are shaped by sexual selection in the first place. Superficially, both of these assumptions seem to generally be true, but they have rarely been tested. Quantifying the convergence of bony ornament-like structures is a means of testing that first assumption.

1.1: On Convergence

Despite the general idea of convergence as phenotypic similarity in two or more lineages that arises independently being well-accepted, there is not currently consensus on a precise, testable definition for convergence.

Two additional concepts are important to clarify when discussing convergent evolution: divergent evolution and parallel evolution.

Divergent evolution is a widely observed phenomenon in which taxa within lineages become more different from one another across evolutionary time as they speciate and diversify. Divergent evolution itself is not always defined in discussions of convergence, but it is essentially the opposite of convergence; in convergence, species independently evolve to become more similar, whereas in divergence, species become more dissimilar (Stayton 2015, Bolnick et al. 2018). Divergent evolution as a term should be distinguished from divergent selection, which is characterized by selection that acts in different directions across environments (Rundle & Nosil 2005, Langerhans 2010). Divergent selection is likely a substantial cause of divergent evolution in most cases, so the two concepts are connected, but divergent evolution, when defined in relation to convergent evolution, is the end product of selection rather than the process itself.

Parallel evolution, meanwhile, is a concept similar to convergence with multiple disparate definitions that muddle consensus about what it is and how it fully works. Generally, parallel evolution is described as a process by which organisms end up with similar morphologies due to some phylogenetic constraints in how those organisms can evolve. Like with convergent evolution, authors who distinguish parallel evolution as a unique concept describe parallel traits to be distinct from homologies, but the traits are often characterized as being less independent than those evolving through convergent evolution, with the evolutionary pathway perhaps following a similar route because of a separate shared trait that is homologous (Gould 2002, Zakon 2002). Some authors reject parallelism or place it under convergence

because common definitions suggest a continuum between convergence and parallelism (Ingraham & Mahler 2013, Stern 2013).

Definitions for convergence often characterize it in relation to parallel evolution. Table 1 provides a sample of previous definitions for convergence relevant to the present study.

Table 1. Different definitions of convergence and parallelism reported by authors over the last two decades. Arbuckle et al. 2014, Stayton 2015, and Castiglione et al. 2019 all propose means of quantifying convergence that I have used for this study.

Gould 2002	Gould defines parallelism as "a positive deep constraint of homology in underlying generators" and convergence as "the opposite sign of domination for external natural selection upon a yielding internal substrate that imposes no constraint."
Zakon 2002	"Convergent evolution occurs when a similar-appearing trait evolves independently in two or more lineages from different ancestral states. Convergence is often distinguished from parallelism in which a similar trait evolves independently in two lineages from the same ancestral state."
Serb and Ernisse 2008	"Convergence is the independent evolution of similar structural or functional components in two or more unrelated or distantly related lineages. In other words, two lineages have evolved convergently if they started from different ancestral morphologies and evolved towards the same (adaptive) phenotype."
Losos 2011	Losos defines convergence as "the independent evolution of similar features in different evolutionary lineages."
Stern 2013	Stern 2013 states, "Here, I reserve the term convergence to mean that different populations or species evolved similar phenotypic solutions. I use the term parallel genetic evolution only for convergent evolution at the level of the mechanisms that generate phenotypes."
Ingraham & Mahler 2013	Ingraham and Mahler convergence as "repeated evolution of similar phenotypes in similar environments." They also clarify, "we do not distinguish between convergence and parallelism, as either represents evidence for nonrandom evolutionary change."
Arbuckle et al. 2014	Arbuckle et al. describe convergence as "the independent occurrence of phenotypic similarity."

Table 1 (continued).

Stayton 2015	Nearly all definitions of convergent evolution agree that it manifests as a pattern in which similar phenotypes evolve independently in multiple lineages" and defines two general approaches, pattern-based and process-based, which seem to characterize most uses of the term. He describes that, "The methods proposed in this article were developed to be pattern based, process neutral, and to provide a quantitative measure of convergent for any data that allow distances to be calculated between taxa. Phenotypic similarity is taken into account, but a certain level of similarity is not required for convergence."
Bolnick et al. 2018	Bolnick et al.'s leading definition for convergent evolution is "the evolution of similar phenotypes or genotypes in multiple independent populations, in response to similar selection pressures, from different initial conditions. A geometric definition is when the endpoints of two evolutionary vectors are closer together than the vectors origins."
Castiglione et al. 2019	Castiglione et al. define convergence as "the morphological resemblance between phylogenetically distant taxa."

Many definitions are somewhat vague, contradictory, or difficult to test. Distinguishing between parallel and convergent evolution is a common problem, and the researcher's approach differs depending on whether the focus is on phenotypic traits or the genome itself, as genes can develop convergent sequences independent of phenotypic convergence.

The definition proposed by Bolnick et al. (2018) is useful because it defines convergence and parallelism on a geometric basis in trait space (see Figure 1). According to this definition, the ancestors of two or more lineages will be further apart in trait space than the end point taxa; in other words, the sum trajectory of the lineage vectors in trait space from ancestor to descendant will converge with one another. In Bolnick et al.'s definition, convergence is when two or more species evolve to become more similar to each other than their ancestors were, either due to the distance between two vectors or the angle of their trajectories. They define divergence as two or more species evolving to become less similar to each other than their ancestors were – an opposite trend to convergence. They also define parallel evolution as a

distinct phenomenon from convergence, with parallel evolution being a situation where the trajectories of two lineages are the same, and nonparallel evolution being any situation where there is a non-zero angle between the vectors of those lineages over time.

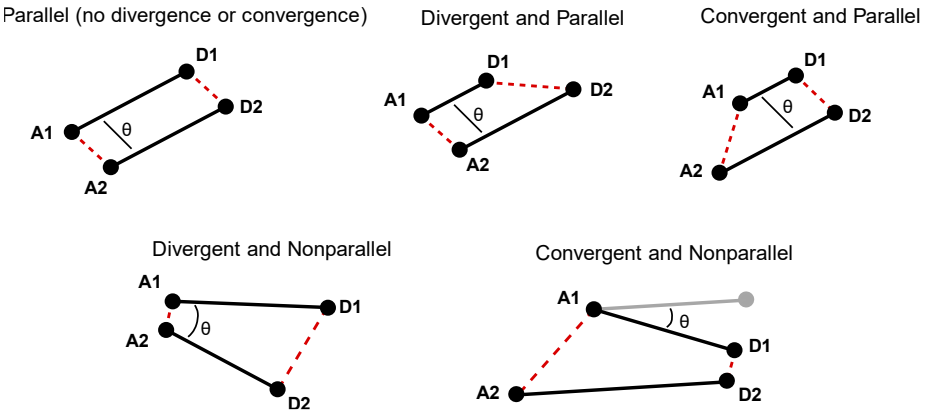

Figure 1. Illustration of definitions used by Bolnick et al. (2018) that show the five possible combinations of parallelism/nonparallelism and convergence/divergence. Parallel evolution by their definition is measured independently of convergence, and is assessed by the angle θ between vectors (θ is at or near 0 for parallel evolution and >0 for nonparallel evolution). Bolnick et al. (2018) define convergence as a smaller distance between the descendant taxa (D1, D2) than the distance between the ancestral taxa (A1, A2), with divergence being a greater distance between descendant taxa than ancestral taxa. According to this definition, both convergence and divergence can occur at the same θ values.

This set of definitions is similar to that of two existing methods for quantifying convergence: Stayton's C-values and the Castiglione et al. (2019) vector angle. I favor the Bolnick et al. (2018) definition for creation of future convergence metrics, but as the established methods are not defined with respect to Bolnick et al. (2018), I will describe here how the three main quantitative methods used for this study differ from Bolnick et al. in their definition of convergence.

C-values are described in Stayton (2015) and determine convergence based on differences in maximum distances between lineage vectors rather than the angle between the vectors directly. The maximum distance can be defined at any point along the lineage, however, not just at a given set of ancestors (see Figure 2).

The Castiglione et al. (2019) vector angle (henceforth called “the search.conv method” in this manuscript) uses phylogenetic ridge regression to create a weighted phenotypic vector between the ancestor and descendant, which is compared against two or more lineages to produce an angle comparing the vectors of these lineages. The angle between vectors defined in Bolnick et al. (2018) is comparable to the angle θ produced in the search.conv method, though Castiglione et al. (2019) interpret this angle differently than Bolnick et al. (2018), defining convergence as a smaller θ angle regardless of original placement of end tip taxa in trait space relative to their ancestors. The search.conv method does not distinguish between non-convergent parallel evolution and convergent parallel evolution defined by Bolnick et al. (2018).

The Wheatsheaf index (Arbuckle et al. 2014) is not a test for the presence of convergence, but rather the strength of convergence. It compares how tightly and separately purported convergent taxa cluster in trait space compared to a random sample from the dataset, with adjustments made by phylogenetic distance.

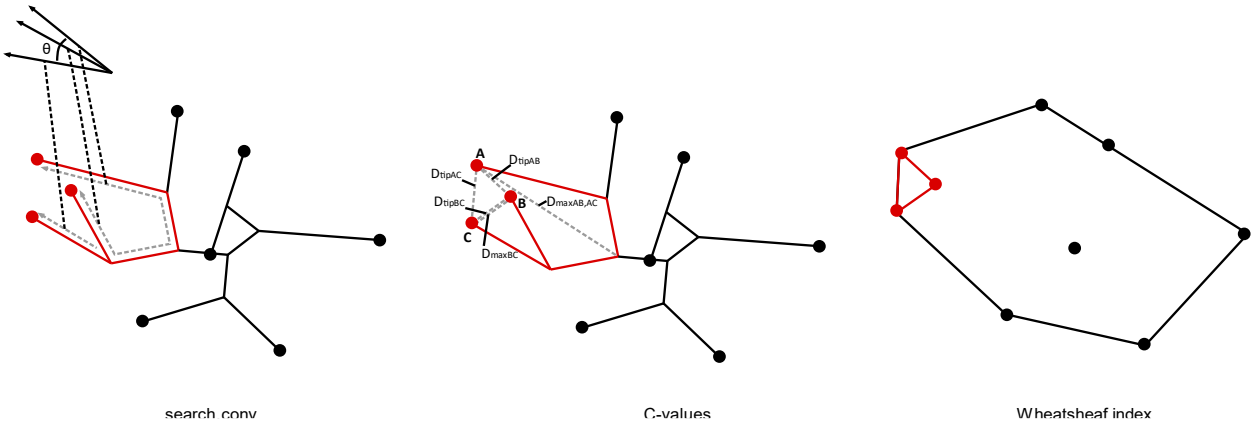

Figure 2 Visualizations for how the definitions for convergence described by the methods used in this manuscript compare to the definitions of Bolnick et al. (2018). Dots represent taxa in trait space, with red dots representing purported convergent taxa and black dots representing the rest of the taxa in the dataset. Phylomorphoplots connect taxa for the search.conv and c-values methods. A convex hull is drawn around the convergent taxa and full dataset for the Wheatsheaf index method. The search.conv method calculates adjusted θ values between the vectors of the focal lineages from ancestor to tip taxa, comparable to the θ defined in Bolnick et al. 2018.

Figure 2 (continued).

The c-values method compares distances between the end tip taxa (D_{tip}) to the greatest distance between taxa from the common ancestor to the end tip (D_{max}). The Wheatsheaf index compares the distances between convergent taxa in trait space to the distances between all of the taxa in the dataset.

Convergence is often the result of natural selection for similar adaptations, though it can arise through other means such as coincidence, exaptation, and in correlation to another character (Losos 2011). Convergence often implies similar selection is at play in the convergent lineages, though exceptions are common (Losos 2011); Stayton (2008), for instance, found that convergence in trait space can occur in different lineages solely through genetic drift. Stayton (2015) goes on to explain that convergence can be defined as process-based or pattern-based, where process-based convergence is similarity in morphology driven by a common selection or other evolutionary process in both lineages, and pattern-based convergence is the observation of similar morphologies in separate lineages that are presumed to have had a common ancestor without this morphology.

It is not necessary to identify the cause for pattern-based convergence in order to recognize pattern-based convergence. Where possible, though, it is important to attempt to determine the selection and other evolutionary processes that have led to pattern-based convergence, as the causes for convergence can be complex (Stayton 2008, Losos 2011, Ord et al. 2013, Stayton 2015). Pattern-based convergence is often assumed to represent a common evolutionary process, but that assumption usually goes untested (Stayton 2015). In some cases, organisms may evolve convergent traits by different means (Stayton 2015), or through exaptation of traits evolved for a separate function rather than direct adaptation (Losos 2011).

1.2: On Selection

Selection is one of the most fundamental concepts in evolutionary biology. Darwin is largely credited with devising the concept, though he himself clarifies that elements of the idea have much older roots and the concept as defined in his work is the result of work by both himself and Alfred Russel Wallace (Darwin 1872). In *On the Tendency of Species to Form Varieties*, Darwin describes what he would later distinguish as natural and sexual selection: "Besides this natural means of selection, by which those individuals are preserved, whether in their egg, or larval, or mature state, which are best adapted to the place they fill in nature, there is a second agency at work in most unisexual animals, tending to produce the same effect, namely, the struggle of the males for the females.... This kind of selection, however, is less rigorous than the other: it does not require the death of the less successful, but gives to them fewer descendants" (Darwin & Wallace 1858, p. 15). Similar views are expressed by Wallace in his portion of the presentation (Darwin & Wallace 1858). Darwin would later go on to clarify natural and sexual selection in more detail in *On the Origin of Species* and *The Descent of Man, and Selection in Relation to Sex* (Darwin 1871, 1872).

The details about how selection works have been clarified through countless studies by subsequent authors, in some cases refuting assertions made by Darwin and Wallace about the circumstances, mechanism, and effects of selection. Different types of selection aside from natural and sexual selection have also been defined in subsequent work, such as artificial selection. These other forms of selection are usually described as subcategories that fit under a broad umbrella of "selection."

"Natural selection" is sometimes described as encompassing all other forms of selection, but some authors prefer to distinguish between sexual selection and natural selection as separate

entities due to the unique association between sexual selection and reproductive success that is not found in other forms of selection (Alonzo & Servedio 2019, Shuker & Kvarnemo 2021). The former use of the term "natural selection" may be referred to as "broad sense natural selection," whereas the latter definition, which excludes sexual selection, may be described as "narrow sense natural selection" (Shuker & Kvarnemo 2021). Distinguishing between the two and reaching a clear definition is not always straightforward due to examples in nature where established definitions become incomplete (Alonzo & Servedio 2019). Shuker & Kvarnemo (2021) offer a definition that describes sexual selection as "any selection that arises from fitness differences associated with nonrandom success in the competition for access to gametes for fertilization." This definition is useful because it is widely-applicable, clear, and testable, but it is important to realize that any one definition is likely not going to be capable of encompassing the full complexity of sexual selection; indeed, many responses to Shuker & Kvarnemo (2021) caution against universal adoption of this definition to the exclusion of other, broader definitions (Alonzo & Servedio 2021, Andersson 2021, Clutton-Brock 2021, Kokko 2021, Shuster & Wade 2021, Simmons & Parker 2021, Zuk 2021).

The differentiation between sexual and non-sexual forms of selection, such as narrow sense natural selection, is not always straightforward. There are many cases where elaborate display structures or weapons evolved primarily in relation to sexual selection also impact survival. This is particularly prominent in cases where ornamentation increases vulnerability to predators, or when individuals risk injury and infection during combat over mating opportunities. The opposite is also true; Heinen-Kay et al. (2015) demonstrate a "selection trade-off hypothesis" experimentally in *Gambusia hubbsi* populations, which show female preference for brighter fin coloration in males. In environments containing predators, the reproductive benefit

of bright coloration is insufficient to overcome the survival risk of highly visible ornaments, and as a consequence, more muted and camouflaged males are found in blue hole environments containing predators. A similar trend has been observed in guppies (*Poecilia reticulata*) (Abrahams 1993, Ghalambor et al. 2004).

Sexual selection primarily involves intraspecific interactions by influencing reproductive success of individuals within a species. However, the consequences of sexual selection also have broader effects on biodiversity and speciation that can be observed on long time scales among multiple species; these larger-scale recurring patterns related to sexual selection are often useful for macroevolutionary studies (Sampson 1999, Padian & Horner 2011, Hone et al. 2012, Knell et al. 2013, Gates et al. 2016, Mallon 2017). Distinction between the microevolutionary and macroevolutionary study of sexual selection is important because the former relies on direct observation of sexual interaction or recent genetic data, and macroevolutionary studies must incorporate more inference, relying instead on indirect methods for detecting sexual selection, particularly when fossil taxa are involved.

Because of these practical research constraints, recurring patterns in morphology of extant species that are associated with sexual selection are often referenced as indicators of sexual selection in the fossil record, such as exaggerated cranial ornament-like structures (Sampson 1999, Padian & Horner 2011, Hone et al. 2012, Knell et al. 2013, Gates et al. 2016). The hypothesis that these structures evolved in response to sexual selection is partly based on models by Lande (1981) and Kirkpatrick (1982), who establish that choice in one sex (usually females) for ornamentation in the other (usually males) can lead to maintenance within the population and experience Fischerian runaway if the genes involved in preference become connected to those involved in the ornamentation itself. The rapid evolution and diversity of

bony ornament-like structures in fossils fits the broad macroevolutionary expectations of the Lande-Kirkpatrick model (sensu Prum 2010). Some authors (e.g., Padian & Horner 2011) question whether these structures in dinosaur species are the result of sexual selection, as many do not appear to be dimorphic and may have served non-sexual roles such as species recognition and non-sexual communication. We currently have limited understanding for how modern animals use similar structures, which limits assessments of how plausible it is that bony ornament-like structures in the fossil record were subject to sexual selection.

Two additional considerations are also relevant to discussion of casques and similar structures: mutual sexual selection and the socio-sexual selection gradient.

Sexual dimorphism is common in vertebrates, with many species have much more elaborate sexual ornamentation in one sex than the other. Yet, this is not the case for all traits influenced by sexual selection. In some species, both males and females may evolve the same form of sexually-selected trait, influenced in similar ways by mutual sexual selection (i.e., when both males and females are the choosy sex; Kraaijeveld et al. 2007). This is especially common in the feather ornaments of monogamous birds in which both males and females actively choose their mates. Mutual sexual selection has been proposed as an explanation for why elaborate bony ornament-like structures have evolved in dinosaurs and pterosaurs despite many species not appearing to be sexually dimorphic (Hone et al. 2012, Knell et al. 2013).

Further complicating discussion of sexual versus natural selection is the concept of social selection, a type of selection primarily defined by interactions between individuals, mainly individuals of the same species (West-Eberhard 1984). Lyon & Montgomerie (2012) present the example of juvenile ornamentation in American coots (*Fulica americana*), which functions like sexual ornamentation in serving a signaling purpose, but is exclusively survival-based, with more

brilliantly-ornamented chicks receiving more food from their parents. Similarly, for ornaments serving a role, sexual selection may in some cases serve non-sexual signaling roles as well, allowing competition for territory that has resource advantages beyond reproductive success (West-Eberhard 1984, Lyon & Montgomerie 2012, Tobias et al. 2012). In species where both sexes have exaggerated ornaments, the ornaments can have a direct impact on reproductive success, but they likely also affect social standing, particularly in colonial-nesting species like seabirds, for which mutual ornaments are common (Kraaijeveld et al. 2007). Some authors suggest that traits subject to social selection can be described on a spectrum from sexual to non-sexual in their function (Tobias et al. 2012).

1.3: On Bills and Bony Ornament-Like Structures

Keratinous bills are found in all extant birds and are frequently subject to both sexual and non-sexual selection. Bill structure is closely linked to feeding, which is not usually tied directly to mating success, though feeding strategies in the parent may impact offspring survival. The recently-extinct huia (*Heteralocha acutirostris*) of New Zealand displayed an unusually dimorphic bill that impacted feeding behaviors of males and females (Buller 1871). The bill dimorphism in huias was not necessarily the result of mate preference, but was still likely tied to sexual selection, if indirectly, by necessitating foraging differences in males and females.

Bill size and shape can impact song characteristics as well as feeding habit. In some North American sparrows such as swamp sparrows (*Melospiza georgiana*), species and subspecies have diverged in response to habitat requirements, with birds occupying salt marshes evolving larger bills (Greenberg & Olsen 2010, Ballentine et al. 2013). Bill shape influences birdsong, and song plays an important role in mate choice, so environmental impacts on bill

characteristics may in turn influence non-random mating patterns and sexual selection (Greenberg & Olsen 2010, Ballentine et al. 2013).

Bird bills are also often highly vascularized, and in species with large bill surface areas, the bill can serve as a thermal conductor for temperature regulation (Tattersall et al. 2009, van de Ven et al. 2016). This is well-documented in tropical birds such as the toco toucan (*Ramphastos toco*) and southern yellow-billed hornbill (*Tockus leucomelas*) (Tattersall et al. 2009, van de Ven et al. 2016), and has even been proposed as a reason behind the morphological and ecological similarities in the two groups (Hughes 2014, van de Ven et al. 2016, van Vuuren et al. 2020). Phillips & Sanborn (1994) and Eastick et al. (2009) found cassowary casques can also serve as thermal windows, despite being much smaller relative to the body than the large bills of hornbills, toucans, or puffins. This effect is not limited to tropical birds; Greenberg et al. (2012) found a similar trend in hot-climate subspecies of song sparrows (*Melospiza melodia*), and Schraft et al. (2019) found tufted puffins (*Fratercula arctica*) can use bill surface area to dissipate heat from high-energy flight.

Enlarged surface areas increase conductivity, but they also provide increased display space. In toucans, many hornbills, and puffins, the bill is brightly pigmented, often by carotenoids, which in birds can be an honest indicator of fitness and are associated with mate preference (Delhey & Peters 2017). In puffins, portions of the bill are also fluorescent, suggesting porphyrins or similar fluorescent molecules contribute to pigment as well (Dunning et al. 2018). In puffins and other fraterculines, the rhamphotheca only displays its characteristic vibrant colors during the summer breeding months for mature adults (Harris 2014). Outside of the breeding season, portions of the rhamphotheca are shed, exposing a softer skin underneath

the inner bill along with a decrease in color intensity on the red outer bill that resembles the juvenile form except for the physical surface area of the bill (Harris 2014).

Given the complexity of bills, discussion of bony ornament-like structures on the skulls of birds requires clarity. I use the term "bony ornament-like structure" because prior terminology is useful but inconsistent when discussing trends across different clades.

Mayr (2018) surveys many of the forms that these structures take in crown Aves, describing them as extravagant bony cranial protuberances, and more specifically as casques or frontal humps. "Casque" tends to be the most common and widely-used term for these structures, being applied to the cranial enlargements of hornbills, cassowaries, oropendolas, and some non-avian dinosaurs such as oviraptorosaurs. However, the term "casque," derived from a type of helmet, often implies a relatively bulbous structure with a keratinous covering. Bony enlargements do not need to be bulbous, nor are they always covered in keratin, even though "casque" is an adequate descriptor in many cases, I favor a more neutral term when referring to any type of enlarged feature on the skull that appears to be ornamental. "Shield" is the favored term for the type of structure found on the heads of musophagids and rails (the latter of which do not have bony structures but instead a soft-tissue enlargement covered in keratin), whereas "dome" is the descriptor often used for certain ducks and *Balearica* species with bony cranial enlargements. Other terms applied to birds and non-avian dinosaurs include "crest," "knob," and "hump."

Protrusions toward the back of the skull can include "horns" and "frills," found in lizards, ceratopsid dinosaurs, and meiolaniid turtles, though horns in particular may be defensive or related to combat, and should be distinguished from presumed ornamental structures where relevant. Some domes and casques may also serve combat purposes, however, so it may still be

useful to label these structures as "ornament-like," distinguishing between ornament-like weapons as a type of ornament-like structure.

I use "ornament-like structure" as opposed to "ornament" in my preferred terminology for these collective structures because of their unknown status as biological ornaments, especially when referring to fossil taxa. An "ornament-like structure" is defined by its morphology, not its evolution or function, so ornament-like structures may be found to be true ornaments or serve non-ornamental functions by later studies, but their description as ornament-like structures would remain.

Ornament-like structures found on the heads of birds are diverse in their presentation. Many of these structures, particularly bony ornament-like structures, are well-described anatomically, but their evolution remains poorly understood. The association between casques, bill size, and thermal regulation offers a non-sexual explanation for the evolution of some bony ornament-like structures. Thermoregulatory functions have also been found in enlarged fleshy ornament-like structures such as the wattles and throat sacs of *Meleagris gallopavo* and *Numida meleagris* (Buchholz et al. 1996, Van Niekirk et al. 2022). *Meleagris* wattles and body size are sexually dimorphic whereas some *N. meleagris* subspecies show sexual dimorphism in throat sac and body size. Van Niekirk et al. (2022) suggest that thermoregulatory capabilities of these structures may facilitate size dimorphism, but that thermoregulation is not directly associated with dimorphism and a more complex interaction appears to be happening between the influences of sexual and non-sexual selection in *N. meleagris* throat sacs.

In some species with bony ornament-like structures, there is dimorphism between males and females, but this pattern is highly variable and not always clear-cut. In some hornbills, the female may have little to no casque alongside the males with well-developed casques (Kemp

1995). *Numida meleagris* crests tend to be more well-developed and larger in males than females, but both sexes bear some form of crest and there is considerable overlap between individuals (Angst et al. 2020); this is also seen in many hornbills and cracids. Cassowaries, on the other hand, have a great deal of individual variation in casque shape, but male and female shape variation is largely indistinguishable (Green et al. 2022).

Horrocks et al. (2009) note that the fleshy knob of the mute swan (Cygnus olor) increases in size during key parts of the breeding season for both males and females. This structure is generally larger in males, but is present in both sexes and appears to be used in mutual mate choice for this species (Horrocks et al. 2009). The fleshy knob overlies a bony core.

In some cases, bill protrusions associated with ornamentation may not be directly selected by mate choice. Jones and Montgomerie (1992) report that least auklets (*Aethia pusilla*) bear a small keratinous knob on the bill in both sexes, respond favorably to feather ornaments and bill coloration, but show no preference for increased size in the bill knob.

It is also important to note that even though casques and similar structures are often described as "ornaments" with presumed ornamental function in sexual selection, enlarged bony structures may also be used as weapons. Weapons are similar to ornaments, and may serve a signaling purpose similar to ornaments in some cases, but are a distinct form of structure subject to slightly different sexually selective forces (McCollough et al. 2016). Casques in at least two species of hornbill, the Indian grey hornbill (*Ocyceros birostris*) and the helmeted hornbill (*Buceros vigil*) have additional structural reinforcement in their casques that aid intraspecific combat suspected to be related to combat for mating opportunities (Kinnarid et al. 2003, Kasambe et al. 2011). This behavior has only been reported among males, but whereas *O.*

birostris is sexual dimorphic in the casque, *B. vigil* females have large reinforced casques similar to the males (Kinnarid et al. 2003, Kasambe et al. 2011).

Similarity between all or part of the bony ornament-like structure of different vertebrates has been reported as convergence in several previous studies (Brown & Henderson 2015, Gates et al. 2016, O'Brien et al. 2016), but most of these claims are qualitative. Goméz & Milevicich (2021) offers one of the few quantitative studies of convergence in casque structure using birds from separate clades within the family Icterydae. Goméz & Milevicich report convergence in the casque between the giant cowbird (*Molothrus oryzivorus*) and cacicines such as *Cacicus cela*. The casque in icterids is usually small, though some oropendolas develop well-formed casques similar to those of small-casqued hornbills; *M. oryzivorus* and *C. cela* both bear casque structures that has a dorsoventrally short but mediolaterally wide posterior boss. Goméz & Milevicich (2021) suggest that because the casque formation in these species does not appear to be associated with size or feeding ecology, unlike the rest of the skull, a possible explanation for the convergence between *M. oryzivorus* and *C. cela* could be brood parasitism. Giant cowbirds parasitize *C. cela* and other cacicines, so similarity in casque appearance could have evolved in response to selection for better mimicry of host chicks among young giant cowbirds. Because of the close relationship between *Molothrus* and *Cacius*, this example could reflect both convergence and parallel evolution according to the Bolnick et al. (2018) definition. *Cacius* and other cacicines appear to have evolved their casques several million years before *M. oryzivorus* (Goméz & Milevicich 2021), so the selection for initial casque evolution in cacicines was likely not a response to brood parasitism, but instead driven by other factors such as thermoregulation or mate preference.

1.4: Defining and Testing Convergence in Avian Bony Ornament-Like Structures

The goal of the present study is to use surface morphometrics to assess convergence in bony ornament-like structures across a broad sample of extant birds and determine if these structures are comparable to one another as is often presumed. A secondary goal of this study is to begin to determine what selective forces influence the evolution of bony ornament-like structures. Are these structures true ornaments evolving in response to sexual selection, or are they predominantly influenced by narrow-sense natural selection as described by Shuker & Kvarnemo (2021)? Our study is not capable of answering this latter question directly, but we can use phylogenetic comparative methods to explore trends among taxa with these structures and begin characterizing how bony ornament-like structures correlate with other aspects of animal biology such as size, diet, and habitat. Laying the groundwork for study of these bony ornament-like structures has wide-reaching implications for the evolution of extinct organisms, and it will also inform our understanding of how these structures have evolved and continue to evolve in modern birds.

To pursue these goals, this study has three main questions (Table 2):

1. To what extent does the morphological similarity of bird skulls from species with ornament-like structures surpass the similarity expected by their genetic relationships?
2. How does the exclusion of the skull region where bony ornament-like structures occur affect the convergence estimates between taxa?
3. What, if any, ecological conditions correlate with the presence and form of bony ornament-like structures?

The first question identifies and quantifies convergence between the bills of birds in my dataset. This study uses several established methods for quantifying convergence to address this question, allowing comparison between slightly different definitions for convergence. The

second question assesses the influence of the bony ornament-like structure itself on the initial convergence values. In some cases, birds may have similar bills but differ in their bony ornament-like structure; we need to be able to detect this difference if we are to attribute the conclusions in this study to the ornament-like structure itself. Finally, the third question offers a simple examination of possible associations between bony ornament-like structures and traits with known ecological associations. Behavioral and experimental studies are necessary to clarify how bony ornament-like structures are influenced by different types of selection; however, they are widely believed to be sexual ornaments. By comparing the presence and morphology of these structures to ecological traits, I can generally assess of whether bony ornament-like structures might be influenced by similar selection.

This study is limited in how completely it can answer these questions, but the tests (Table 2) are designed to provide a starting point for how our data can inform our understanding of these structures and how they relate to convergence and sexual selection, if at all.

Table 2. Which methods will be used to address the three main questions in the present study. Light gray cells represent qualitative methods, while dark gray cells represent quantitative methods.

	PCA	Cluster Analysis	Basic C-values	C-value ANOVA and t-test comparisons	Wheatsheaf index	RRphylo search.conv	ANCOVA	MANCOVA	Pearson's chi-square test
Question 1									
Question 2									
Question 3									

In order to study these bony ornament-like structures, we need a consistent definition for bony ornament-like structures that allows us to differentiate between taxa that have these structures and taxa that do not have them.

To account for the variety of ways one could define or small bony ornament-like structures, I have tested nine different ways (henceforth called “test definitions") of categorizing taxa as ornamented and unornamented. Table 3 describes how these test definitions are organized. Taxon inclusion can be found in Supplementary Table 2.

Table 3. The nine test definitions for ornamentation, organized by three categories (literature, personal, and premaxilla definitions) used in this study. Taxa that were tested as convergent and did have ornament-like structures are described as “ornamented” taxa, while taxa lacking these structures are described as “unornamented” for each of the test definitions. Please note that “ornamented” is a useful shorthand for discussion of the test results, and does not necessarily mean these taxa are using their bony ornament-like structure as sexually-selected biological ornaments. These definitions are purely morphological. n = the number of taxa designated as ornamented. The full 82 sample taxa were included in each test.

Test Definition 1	Literature definitions	1.0	n=15	Consensus designations for bony ornament-like structures in previous studies (literature definition conservative)
		1.1	n=17	Any designations for bony ornament-like structures in previous studies (literature definition broad)
Test Definition 2	Personal definitions	2.0	n=15	Personal designations for prominent bony ornament-like structures (personal definition conservative)
		2.1	n=40	Personal designations for prominent and ambiguous bony ornament-like structures (personal definition broad)
		2.2	n=49	Personal designations for bony or soft-tissue ornament-like structures, prominent and ambiguous (personal definition broad + soft tissue)
Test Definition 3	Premaxilla definitions	3.0	n=9	Combined Test 1.1 and 2.0 designations for bony ornament-like structure primarily made up of premaxillary bone (premaxilla definition broad)

Table 3 (continued).

		3.1	n=3	Combined Test 1.1 and 2.0 designations for bony ornament-like structure primarily made up of premaxillary bone in the non-backswept crest/casque morphology found in species like *Crotophaga ani* (*Crotophaga*-type premaxilla definition)
		3.2	n=4	Combined Test 1.1 and 2.0 designations for bony ornament-like structure primarily made up of premaxillary bone in the backswept shield/casque morphology found in species like *Musophaga violacea* (*Musophaga*-type premaxilla definition)
		3.3	n=6	Same definition as for Test 3.0 but excluding *Fratercula* species and *Tauraco persa* (premaxilla broad, *Fratercula* excluded)

Test definitions (TD) 1.0-1.1 are based on work by Livezey (1996), Webster (2003), and Mayr (2018), which describe extant bird taxa which bear casques or similar bony protrusions on the skull. Not all taxa in my dataset have been specifically assessed in the literature with regard to the presence or absence of casques, so if casques are not mentioned in descriptions of the animal, they are categorized as not having bony ornament-like structures on the skull.

Test Definitions 2.0-2.2 are modified from the TD 1.0-1.1 with changes made based on my personal preference in comparing these structures across a broad range of taxa. In definition 2.0, I designate only those taxa with prominent bony protrusions to be ornamented. A broader definition for bony ornament-like structures is provided for TD 2.1, in which slight enlargements are categorized as bony ornament-like structures (see Figure 3). TD 2.2 determines all taxa with any non-feather ornament-like structure to be ornamented, even if the structure does not appear to have a bony core. The 2.2 test definition is included largely as a point of comparison for the tests that focus exclusively on visible bony ornament-like structures, and accounts for similarities in the skulls of animals which may have accessory features related to the presence of ornament-

like structures, whether those structures are composed of bone or soft tissues like collagen and keratin alone.

In addition to the definitions marking all taxa with prominent bony structures as ornamented, I also wished to explore the possibility that bony ornament-like structures with similar regional origins or general shapes might be more convergent with one another than they were to other types of bony ornament-like structures. Test definitions 3.0-3.3, used to test subcategories of ornamentation, are demonstrated in Figure 3.

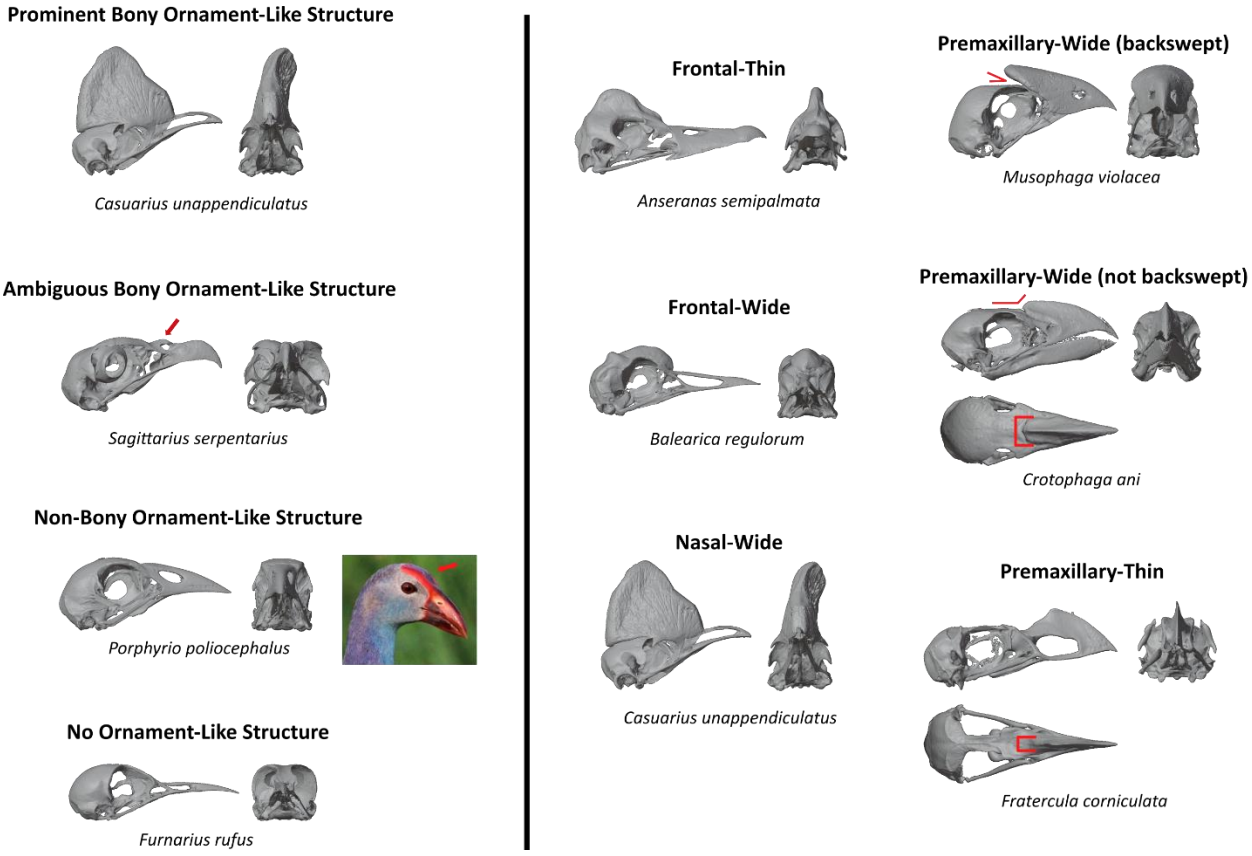

Figure 3. Examples of different forms of ornamentation used for the different test definitions in this study (shown in right lateral and anterior views; Crotophaga *and* Fratercula *are also shown in dorsal view). The left side of the figure illustrates different levels of bony ornament-like structure prominence, with the last two categories lacking bony ornament-like structures, but the taxa with non-bony ornament-like structures having 3D ornament-like structures in life. The right side of the figure illustrates ornamentation groupings based on ornament-like structure location and general characteristics.*

Figure 3 (continued).

The premaxillary bony ornament-like structures are highly varied but they can be broadly differentiated by the width of the ornament-like structure posterior base and the angle at which the ornament-like structure meets the frontal. Test definition 2.0 only defines taxa with prominent bony ornament-like structures as "ornamented." Test definition 2.1 defines taxa with either prominent or ambiguous bony ornament-like structures as "ornamented." Test definition 2.2 defines all taxa but those clearly lacking an ornament-like structure as "ornamented." Test definition 3.0 defines taxa with any premaxillary bony ornament-like structure as "ornamented." Test definition 3.1 defines taxa with premaxillary-wide (not backswept) bony ornament-like structures as "ornamented." Test definition 3.2 defines taxa with premaxillary-wide (backswept) bony ornament-like structures as "ornamented." Test definition 3.3 defines taxa with either type of premaxillary-wide ornament-like structures as "ornamented." Porphyrio poliocephalus *photograph adapted from Charles J. Sharp.*

Additional test definitions were explored but ultimately rejected due to small sample size or inconsistency. Evan though additional bony ornament-like structure forms are differentiable in taxa within our dataset (see Figure 3), only the premaxillary forms were found in non-sister taxa and sufficiently numerous to allow convergence tests. These additional definitions could possibly be tested in the future as more taxa with ornament-like structures are sampled.

Phylogenetic Flexible Discriminant Analysis (PFDA) was also explored as a means of categorizing taxa in a more quantitative manner after the work of Motani & Schmitz (2011) and Schmitz & Motani (2011). In an FDA, well-defined datapoints are categorized by a human researcher and these are used to train an algorithm to recognize these categories, then new datapoints of unknown categorization are fed into the algorithm to categorize them according to its learned rules; PFDA applies this same method, but incorporates a phylogenetic adjustment to account for non-independence of taxonomic data (Motani & Schmitz 2011). While this method holds promise for sorting bony ornament-like structures into binary categories with less subjectivity than a human-based categorization, the PFDA tests for my dataset were unable to consistently identify parameters for gauging the presence of bony ornament-like structures. The

code produced by Motani & Schmitz (2011) allows one to feed the original data back into the trained algorithm as unknowns to test the accuracy of the sorting. For my dataset, all PFDA tests trained on the established TDs accurately categorized less than 25% of the original taxa, including some of the least ambiguously ornamented taxa in the dataset such as Casuarius unappendiculatus. The PFDA is limited in the amount of input data it can use for its definitions, so the large number of landmark points and the general disparity in morphology across the dataset may be part of why the PFDA struggles to accurately categorize the taxa by ornamentation, especially as the shape and location of different bony ornament-like structures may differ substantially.

Human categorization can be biased by the researcher's perception, this bias may be meaningful for identifying casques. Human experience with birds and other animals means that a person can filter and weight different regions of the skull based on expectation for what type of variation defines a "typical" bird without necessarily realizing they are doing so. Extreme structures, especially prominent casques growing out of the bill or the top of the skull, are immediately recognizable to a human as unusual based on prior experience. A computer will weight variation in all areas of the skull equally unless instructed otherwise, and while it can develop the same filtration and weighting system used by the human brain if given enough practice, humans tend to acquire this skillset as part of our development. Some casques have evolved specifically to be eye-catching, as they are brightly colored and centrally-located near or between the eyes. Even though humans are not the intended audience for attention-grabbing ornaments in birds and other animals, I tend to gravitate toward unexpected, flamboyant, or exaggerated structures regardless.

Human bias in defining casques is important, but it is not necessarily biologically meaningless for the animals sporting these structures. The physiological consequences of large casques on cassowaries and hornbills are reflected in their thermal conductivity (Phillips & Sanborn 1994, Eastick et al. 2009, Schraft et al. 2019, Van Niekirk et al. 2022), and even if the enlargement is relatively small (such as the small but distinct bony ornament-like structure on certain species of *Philemon*), the increased surface area of the bill may still have thermoregulatory and structural effects.

Ambiguous or small ornaments are a likely place where human bias can affect the definition for casques. It is easy to find small regions near the nasal or the premaxilla-frontal joint where the bill appears to get thicker, similar to the early stages of casque growth for cassowaries and hornbills. Sometimes, these thicker areas might actually reflect the small base of an ornament-like structure, but other times, these could be an artifact of how the boundaries between different bones look or similar things that do not actually indicate ornamentation. For this study, the use of different test definitions that include or exclude ambiguous bony ornament-like structures should allow us to see whether these taxa impact the results, but future studies may use additional quantitative methods such as a modified PFDA to determine how ambiguous bony ornament-like structures should be categorized.

CHAPTER 2

2: Materials and Methods

2.1: Data Collection and Landmarking

I assembled a list of publicly available CT-scan files and meshes (as .ply or .stl file formats) available through the digital repository databases Morphosource and Phenome10k. Additional CT-scan files were collected from the Duke University Shared Materials Instrumentation Facility Nikon XTH 225 ST High Resolution X-ray Computed Tomography Scanner (see supplementary materials for details about sourcing). This list contained 305 taxa from all major branches of extant Aves. In order to obtain the least biased data possible, the list of available bird species was randomized using the *sample* function in R v.4.1.2, then I removed specimens that were found to be unsuitable due to cropping or damage in the area of interest (e.g., broken nasal bone), disruptive scanning artifacts, or other issues that would hinder landmarking and comparison between specimens.

Data was collected in order of the randomized list, with additional specimens possessing prominent or ambiguous bony ornament-like structures prioritized non-randomly due to the scarcity of prominently-ornamented taxa within the list of available specimens. In total, 82 species were sampled (see supplementary information for details). Members of the Alcidae are more heavily represented in the sample than other groups relative to their species diversity, because the clade includes species with all four grades of ornamentation (e.g., *Fratercula arctica* has a prominent bony ornament-like structure, *Alca torda* has an ambiguous bony ornament-like structure, *Aethia cristatella* has non-bony ornament-like structures, and *Alle alle* has no ornament-like structure). *Uria aalge* was selected as a taxon with no ornament-like structure to compare to alcids with some form of bony ornament-like structure.

Bird skulls were collected as 3D mesh files for landmarking. Some specimens were downloaded as pre-made meshes formatted from CT-scans and photogrammetric scans. CT-scanning uses composite x-ray photographs of a physical object moved by small increments to capture the three-dimensional form of the internal and external form of an object. The CT images (saved in DICOM, TIFF, or other file formats) can then be uploaded to a program such as Dragonfly or Avizo and combined together to form a 3D form that can be subdivided and saved as a mesh file. Dragonfly was used to process CT files in this project. Photogrammetry functions in a similar manner, using visible light rather than x-rays. Photogrammetry can only capture the external surface of skulls, but for the purposes of this project, a high-resolution photogrammetric scan is sufficient to produce a mesh of the external skull surface for sampling.

All specimens were saved as .ply files and oriented in Blender v.2.93 LTS with the centroid of the mesh volume at the origin and the tip of the bill positioned four units along the Y axis. The purpose of this orientation is primarily to facilitate data collection and preliminary comparisons in shape data. Where necessary, specimens were scaled, rotated, and flipped in Blender to allow data collection of points on the scan surface of the right side of the skull. Small gaps in the mesh surface could disrupt placement of semilandmarks, so when gaps were substantial, portions of the mesh surface were covered with an additional mesh plane that was then shrink-wrapped to the skull surface in Blender to provide a base that would prevent semilandmarks from falling through gaps in the mesh surface.

Specimens with asymmetry (e.g., *Picus viridis* has a naturally asymmetrical bony ridge to allow connection of the hyoid bone and associated tissues to the right frontal and nasal) were kept in the dataset. The right side of these asymmetrical specimens was measured unless damaged (with the exception of *Casuarius unappendiculatus*, which was digitally flipped to

measure its left side to facilitate data collection on its crest – see supplementary materials). Bony ornament-like structures have been noted to vary between individuals in a species based on population, age, sex, and other factors (Kemp 1995, Mayr 2018, Angst et al. 2020, Green et al. 2022, personal observation). Specimens in this study were taken to be average representatives of each species, because all skulls were from adults and none of the sampled species shows considerable sexual dimorphism in the cranium relative to the overall variation in the dataset. However, this assumption should be scrutinized during interpretation. Assessment of the individual variation in bony ornament-like structures in different species is outside of the purview of this report, but we recognize it is important for future studies.

Some specimens were subsampled in ORS Dragonfly v.2021.1.0.977 and subdivided in Blender to reduce file size (see supplementary materials for details). Note that different programs use different subdivision algorithms, some of which are not suitable for surface data collection. Large meshes subdivided in Dragonfly, for instance, render the surface as prismoidal structures with small gaps between them due to the program subdividing internal structures alongside the external surface. This essentially breaks up the surface into a grainy mass that does not work well for curve- or patch-based landmarking methods. Generally, subdividing structures by large amounts (90% or greater) results in a grainy mesh, so subdivision was only used where necessary and only to the amount that would reduce the .ply file to a sufficiently small size that it could be loaded into a landmark program.

Landmarks were placed on the surface of each mesh using Stratovan Checkpoint v.2020.10.13.0659 (see supplementary materials for details). The program allows different types of landmarks and allows semilandmark placement relative to established landmarks, with semilandmarks placed along the mesh to minimize bending energy. For this project, a large

number of semilandmarks were desired to characterize the details of the surface of the bird skulls, as is established practice for assessing 3D skull morphometric variation (Wiley 2005, Bardua et al. 2019). Nine landmarks were placed at recognizable anatomical points around the edge of three bones on the skull – the frontal, nasal, and premaxilla – with other bones excluded from the analysis. Square patches of semilandmarks were mapped to the anatomical landmarks, distorting to capture the shape of the element. The frontal merges with the parietal and other posterior bones in most mature birds, rendering the boundary between individual elements difficult to see on actual skulls and scans, so only the anterior portion of the frontal was sampled, with the posterior limit of the frontal defined in this study as a straight line starting at the dorsal apex of the orbit and intersecting the sagittal plane at a right angle (see Supplementary Figure 6).

Some birds have bony ornament-like structures further back on the skull, and because only the anterior half of the frontal was sampled, the posterior-most portion of the bony ornament-like structure in *Balearica* and *Anseranas* was not sampled. The posterior portion of the crest in *Numida* was sampled, with the posterior apex of the crest defined as Landmark F1 (see Supplementary Figure 6) due to the distortion of the skull as the animal grows (Mayr 2018, Angst et al. 2020).

The number of semilandmarks used for each specimen was identical throughout the dataset, but the number of semilandmarks differed between the three bones sampled. This was to allow approximately even semilandmark densities across the sampled area so that elements generally contributed proportionally to the overall skull shape. The number of semilandmark points was selected based on visual estimation for the number of points necessary to capture the surface shape of each element on the meshes in as much detail as possible without becoming computationally burdensome or introducing semilandmark placement errors. In many birds, the

nasal is a thin, L- or V-shaped bone with limited space for landmark placement, and semilandmark placement tends to become distorted when dense patches are used, so a smaller number of semilandmarks were used for the nasal than the other two elements. The premaxilla is usually the largest bone of the bill in terms of external surface area, so in order to adequately capture its shape, a larger number of semilandmarks were placed for the premaxillary patch. In total, 289 landmark and semilandmark points were used for the nasal (17 on each side of the patch), 441 points were used for the frontal (21 on each side of the patch), and 1089 points were used for the premaxilla (33 on each side of the patch), for 1,819 total 3D coordinate points used to characterized each specimen.

The 1,819 total bill landmarks were then subcategorized according to their contribution to the bony ornament-like structure (see Figure 4). A simple designation was created for "ornamental" and "non-ornamental" regions of the sampled area to allow comparison for how the bony ornament-like structure may contribute to the convergence metrics, as per Question 2 for in this study. The ornamental region was any portion of the skull that contributed to a bony ornament-like structure in any taxon, as the regions needed to be comparable for all specimens. Because of the variation in location for bony ornament-like structures, the ornamental region contained a much larger number of landmark points than the non-ornamental region. In total, 263 landmarks were classified as part of the non-ornamental region (35 on the nasal, 40 on the frontal, and 188 on the premaxilla), while the remaining 1,556 were classified as part of the ornamental region (254 on the nasal, 401 on the frontal, and 901 on the premaxilla).

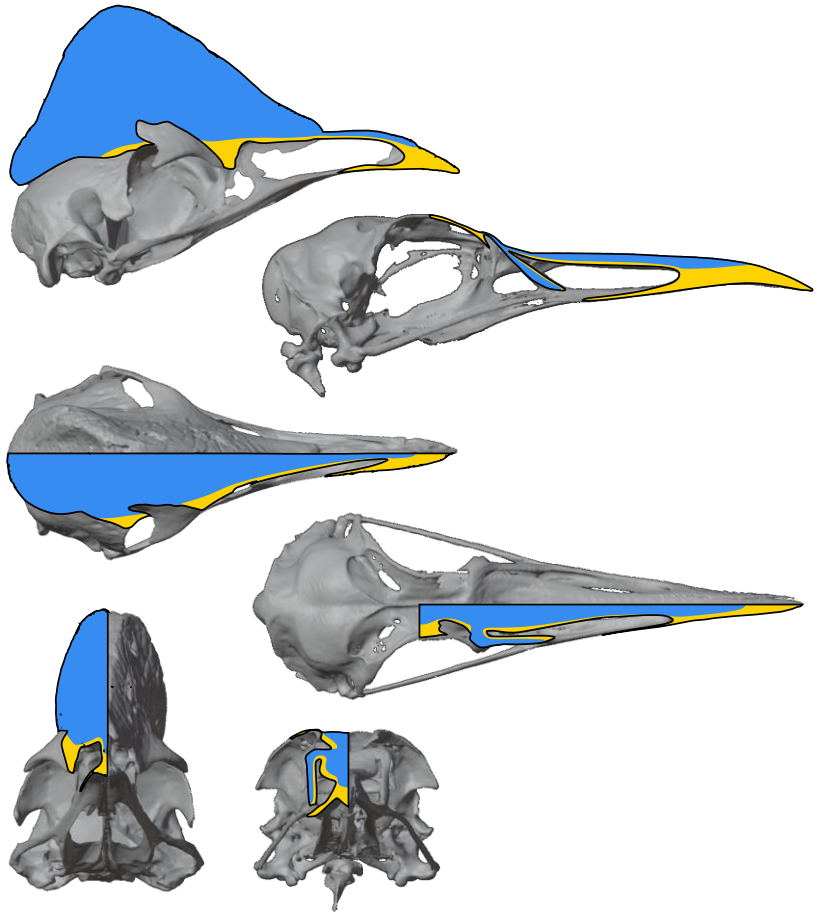

Figure 4. Approximation of sampled skull area for two taxa in the dataset: Casuarius bennetti *(left) and* Uria lomvia *(right). The sampled area is highlighted in blue and yellow and includes the external surface of the frontal, nasal, and premaxillary bones. The blue region represents the region of the sampled area designated as "ornamental," meant to represent parts of the skull where bony ornament-like structures occur in certain species. The yellow region represents the portion designated as "non-ornamental," where bony ornament-like structures never occur in any of our sampled taxa.* Casuarius bennetti *has a well-defined ornament-like structure, much of which is formed from the nasal bone.* Uria lomvia *is defined as "unornamented" in all definitions.*

Procrustes superimposition was performed on the raw coordinate value data using the *ProcSym* function in the R package Morpho v.2.10.

According to Cardini (2016), Procrustes superimposition of single-sided 3D cranial landmarks can differ significantly from adjustments made when landmarks account for both

sides of the skull. Even though these differences may be less disruptive for macroevolutionary studies, we found that the influence on specimen correction did affect our results, so in the interest of maintaining Procrustes superimposition corrections similar to what we would get from measuring landmarks on both the right and left sides of the skull, we opted to mirror the right-side landmarks across the sagittal plane prior to completing the superimposition. Analyses were then based only on the adjusted right-side landmarks. This adjustment remains subject to minor inaccuracies relative to both-sides landmarking of anatomical points for some species because of cranial asymmetry.

Procrustes superimposition is, by definition, sensitive to shape differences between specimens, which are a concern in this dataset due to the great range of morphological variation in the birds sampled. Because the study is interested in assessing shape variation of these skulls relative to each other, the purpose of performing Procrustes superimposition is to reduce noise from variation that cannot be explained by morphological differences, but some amount of uncertainty will always be present depending on what landmarks are selected for the adjustment. We decided to include the "ornamental region" of the skull in the superimposition. We have defined the ornamental region as the set of landmarks that occur in areas of the skull which, on at least some taxa in the sample, contribute to the bony ornament-like structure. The reason for including this region in the Procrustes superimposition is that the region itself occupies a large portion of the landmarks in certain regions, but relatively few of the taxa actually have a bony ornament-like structure in these regions. The interest of adjusting the samples to a true average of the sample distances supersedes the desire to align only the regions of the skull that never have visible bony ornament-like structures, especially because the taxa with these structures are not always clearly defined. Mirroring the landmarks across the sagittal plane slightly mitigates

the influence of the bony ornament-like structures on the total adjustment by preventing medio-lateral rotation.

The result is that the bird bills and frontal elements are approximately aligned, and distances between enlarged bony ornament-like structures are minimized as much as possible while still regarding these structures as part of the larger skull complex. The difference in bony ornament-like structure landmark differences are conservative estimates for this study.

2.2: Phylogeny

This study incorporated a phylogeny for analyses to establish cladistic relationships between taxa and estimate time since divergence. Phylogenetic relationships and time since divergence between branches was compiled in Mesquite v.3.70 from node ages given in molecular phylogenies from existing literature (Wright 2008, Gonzales et al. 2009, Krajewski et al. 2010, Viseshakul et al. 2011, Gibb et al. 2013, Mattos-Reano 2013, Subramanian et al. 2013, Price et al. 2014, Prum et al. 2015, Smith & Clarke 2015, Andersen 2017, Hooper & Price 2017, Gargia-R et al. 2020, Khul et al. 2021; see Table 1 in the supplementary materials for details). Molecular information offers a more direct and high-resolution dataset for determining taxon relationships than morphological data alone. The phylogeny used in this study generally agrees with recent fossil studies and offers reasonable consensus for the complex relationships between major bird orders (Prum et al. 2015, Smith & Clarke 2015, Hooper and Price 2017, Garcia-R et al. 2020, Khul et al. 2021). However, it is important to note that molecular phylogenies require bounding by fossils for accurate time calibration, and some bird clades, particularly those found in the tropics, do not have the extensive fossil records necessary for high-precision and high-accuracy time calibration. The general divergence time estimates appear to hold true, but there

remains disagreement on the exact timing of splits between major clades within Aves. Because all three of the convergence metrics used in this study incorporate phylogenetic branch length as a correction factor, error in the phylogeny will likely affect the convergence estimates.

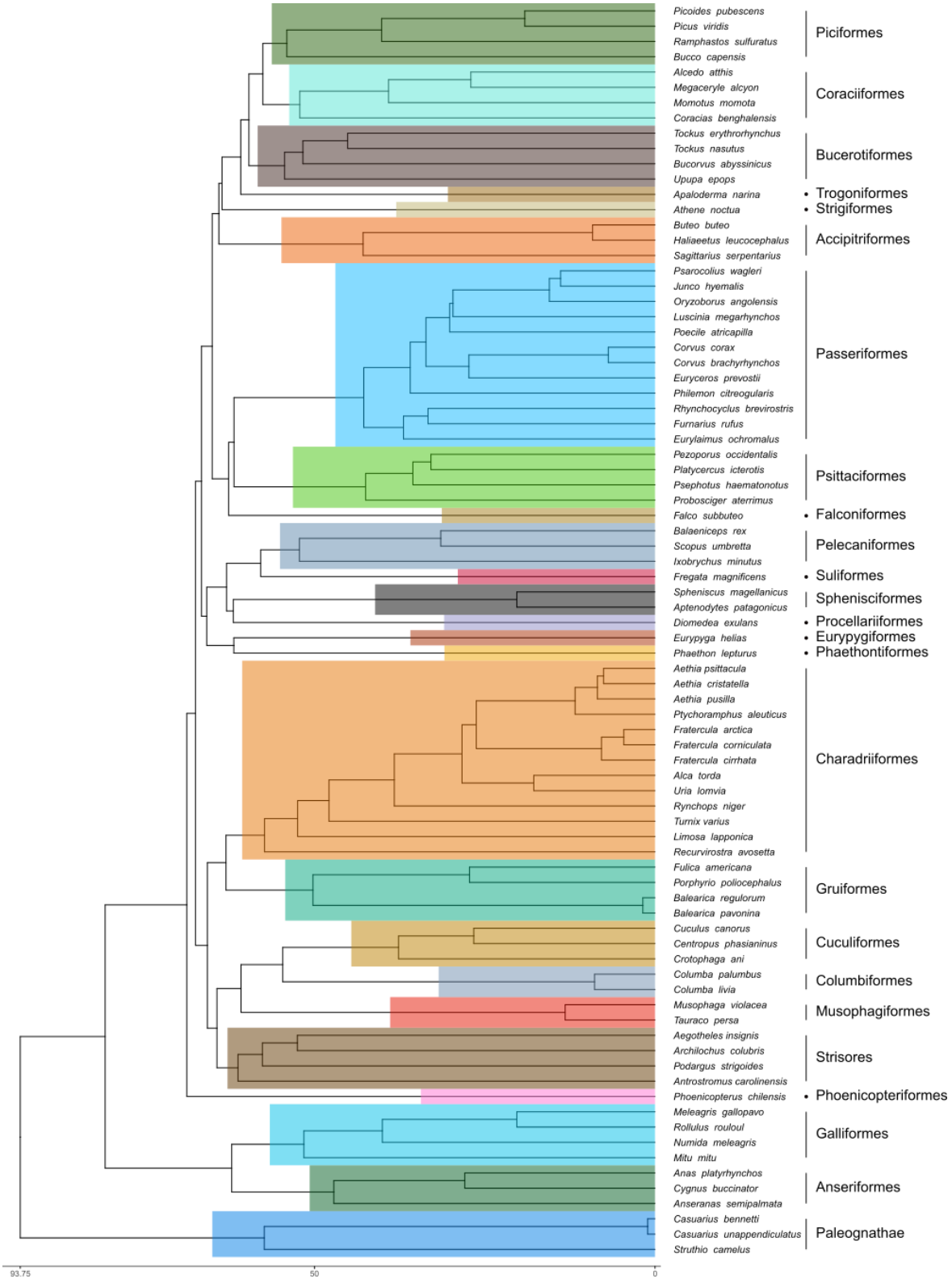

Figure 5. The phylogeny used in the present study. See supplementary materials for details about branch lengths and the original study sources for those branch lengths.

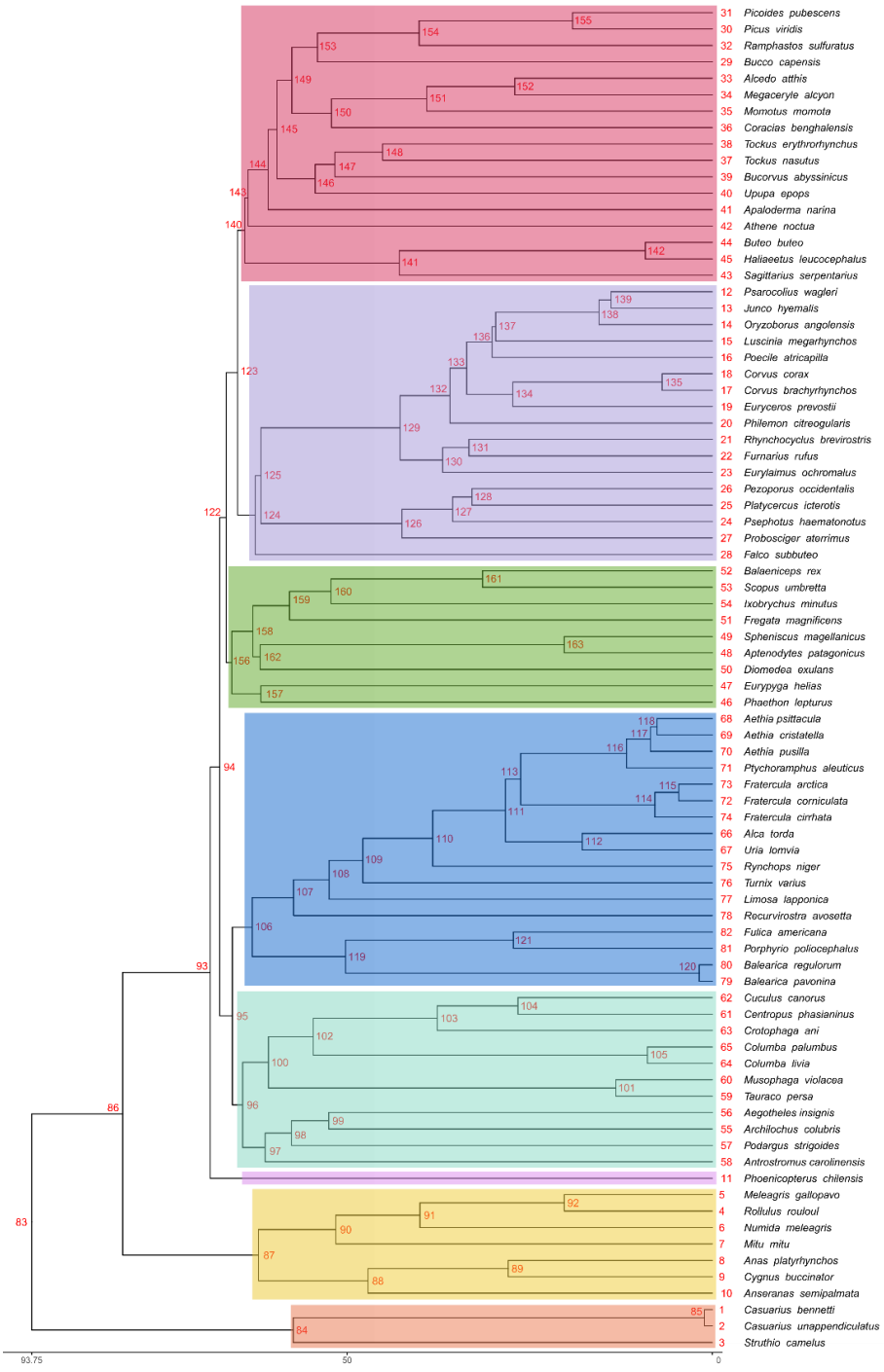

Figure 6. Phylogeny with nodes and groups depicted in the phylomorphoplot of Figure 7. Paleognathae *(orange);* Galloanserae *(yellow);* Cuculiformes, Columbiformes, Musophagiformes, *and* Strisores *(teal);* Charadriformes *and* Gruiformes *(blue);* Pelecaniformes, Suliformes, Sphenisciformes, Procellariiformes, Eurypygiformes, *and* Phaethontiformes *(green);* Australaves *(lavender);* Piciformes, Coraciiformes, Bucerotiformes, Accipitriformes, Strigiformes, *and* Trogoniformes *(red);* Phoenicopteriformes *(pink).*

2.3: Analyses

2.3.1: Principle Components Analysis and Hierarchical Cluster Analysis

In response to the disagreement in convergence definitions and measurements, the analyses for convergence in this study incorporate several different methods.

Principle Components Analyses (PCA) were calculated to characterize the shape variation between specimens. A phylomorphoplot superimposed on a scatterplot of PC1 measured against PC2 was established to illustrate the phylogenetic clustering of taxa in trait space. The C-values and *search.conv* convergence analyses both incorporate phylomorphoplot information in their metrics, though these are not visualized. PCA points on some plots are color-coded according to ecological and mass data to visualize trends in shape related to other factors.

I also ran a hierarchical cluster analysis to characterize the similarities in the morphology of the taxa in this dataset. The cluster analysis, like a PCA, does not incorporate any phylogenetic information, but groupings from the cluster analysis can be compared to phylogenetic groupings to determine whether certain taxa cluster more closely than expected by their phylogenetic relationship, as would be the case for most definitions of pattern-based convergence.

This study tests convergence within a specific region of the bird skull associated with the bill using three quantitative methods: Stayton's C-values (Stayton 2015), the RRphylo search.conv function (Castiglione et al. 2019), and the Wheatsheaf index (Arbuckle et al. 2014). C-values greater than 0 indicate the descendent taxa have evolved to become more similar than at least some of their ancestors, indicating convergence, with higher c-values reflecting greater similarity among the birds. Convergence within the search.conv metric would mean something similar, though with an emphasis on the direction of the vectors rather than the distance between

them. The Wheatsheaf index provides a phylogenetically-counterweighted estimate of how similar the convergent taxa are relative to the rest of the sample, and offers an indication of how strongly convergent those taxa are to one another. In total, these results offer an incomplete but useful picture of how our selected bird taxa compare to one another in the bill region. Detection of convergence (or in the case of the Wheatsheaf index, significantly substantial convergence) for any of these metrics would suggest similarity among our focal taxa. A single test suggesting convergence should not be accepted on its own without scrutiny, especially if the other tests fail to determine the focal taxa for that test are convergent; however, any such result would warrant further scrutiny and could suggest convergence is indeed present. As our various definitions for focal taxa are based on the presence or absence of a bony ornament-like structure, similarity in the selected region among taxa bearing these structures would suggest that these structures may have evolved due to similar selection acting on these species independently.

2.3.2: Convergence Analyses

2.3.2.1 Stayton's C-values

The C-value metrics defined in Stayton (2015) are a set of four (sometimes five) convergence values (C1-C4) calculated from euclidean distances on a phylomorphoplot:

$$C_1 = 1 - \left(\frac{D_{tip}}{D_{max}}\right)$$

$$C_2 = D_{max} - D_{tip}$$

$$C_3 = C_2 / L_{tot.lineage}$$

$$C_4 = C_2 / L_{tot.clade}$$

D_{tip} = distance between convergent tips in phenotypic space

D_{max} = maximum distance between any pair of taxa in those lineages

$L_{tot.lineage}$ = sum of all phenotypic distances along morphological branch length from common ancestor of taxa of interest to branch tips for those taxa

$L_{tot.clade}$ = sum of all phenotypic distances along morphological branch length in entire clade defined by common ancestor of taxa of interest.

C-values were calculated using the convevol v.1.3 R package.

C1 is the scaled difference of phenotypic similarity, C2 is the total difference in phenotypic similarity, C3 is a ration of tip taxa difference to total lineage differece, and C4 is a ration of tip taxa difference to whole clade difference. C1 is reported on a scale of 0-1, with 0 reported as exceedingly low or no convergence between the taxa since their most recent common ancestor, and 1 reported as exceedingly high or complete convergence since the most recent common ancestor. C2, C3, and C4 each establish a value with similar significance, but without an upper limit.

C4 is not used in this study because it scales logarithmically and is not particularly comparable to the other C-values. C4 may be relevant for some studies, but it is of limited value in a broad-scale morphometric study such as this one.

C-values are usually 0 or for sister taxa because as species diverge from one another, they separate in trait space from one another but usually tend to remain close in trait space to the common ancestor shared between the sister taxa. 0 represents taxa that show no more similarity to one another than any of their ancestors did to one another. However, sister taxa can end up with non-zero C-values if they are more phenotypically similar to one another than to their most recent common ancestor, as would be the case if they evolved according to the Bolnick et al. (2018) definition for parallel evolution. According to Bolnick et al. (2018), sister taxa are always

divergent. C-values are not able to differentiate between parallel and convergent evolution according to the Bolnick et al. (2018) definition.

C2 and C3 are bounded at 0, because if the tips are more morphologically dissimilar than any of their ancestors, D_{max} becomes 0. For C2, as D_{tip} becomes smaller (i.e., the taxa become more morphologically similar), the C2 value increases for any given D_{max} value. C3 tends to increase between more distantly related taxa, largely because there are more possible datapoints to add to the sum, as well as more opportunity for divergence within the lineage.

A large D_{max} value coupled with a moderate D_{tip} value will show similar results to a moderate D_{max} value coupled with a small D_{tip} value, which is possibly why the C-values show a bias toward larger C-values for more closely-related taxa. This trend seems to diminish in taxa that are separated by more than 10 million years (see Results Section 1), but is an important consideration for studies among closely-related taxa.

Because the C-values are calculated based on ancestral state reconstructions, their accuracy is heavily tied to these reconstructions, and in modern datasets that are not tied to fossil morphological data, the ancestral state reconstructions may be swayed by the presence of unusual synapomorphies in the taxa. Similarly, missing data tends to make ancestral state reconstructions assume that the existing lineages have traits found in the ancestor REF, but in cases where one taxon has evolved a unique morphology whereas other members of the lineage have retained many of the ancestral traits, a phylomorphospace character reconstruction will usually weight all taxa equally for the reconstruction REF.

C-values have a tendency to over-estimate convergence when parallel evolution occurs. If two taxa split from a common ancestor and do not get closer to each other in morphospace, but both of them do get further away from the ancestral state, the D_{max} will likely be calculated from

one of the tips to one of the points closer to the ancestral state, increasing the C-values even though the tip taxa are not evolving closer to one another at any time since their split.

C-values calculate divergent taxa as 0 because at that point, $D_{tip} = D_{max}$. This is true regardless of how divergent the taxa are – taxa that have diverged by an extremely large amount and taxa that have only diverged a small amount, but have diverged more than any of their ancestors, will still be calculated as 0.

C-values appear to be robust against homology but are dependent upon the phylomorphospace assumptions. The C-values determine convergence following Stayton (2008, 2015) based on differences in maximum distances between points along each lineage in trait space, rather than the distance between ancestors and descendants. As a result, taxa will appear convergent according to the C-values either if two of the ancestral points on the separate lineages are more distant in trait space than the end tip taxa, or if the distance between an end tip taxon and an ancestral point on another lineage is greater than the distance between the end tip taxa. D_{max} can be calculated between one of the end-tip taxa and an ancestral taxon, which may skew convergence values to make them appear higher than they actually are when compared to the Bolnick et al. (2018) definition.

C-values are computationally intensive, but are flexible in their ability to assess convergence between any two lineages. For this study, I chose to calculate C-values between each of the 82 taxa in the dataset, producing 3321 unique taxon pairings. Ideally, a complete inquiry into which taxa pairings produce the highest and lowest C-values would be expanded across all possible combinations of taxa grouped by increasing number, but this is computationally impractical, so instead, pair-based comparisons will offer a general assessment from which C-value tests grouping more than two taxa could be derived if desired.

Once C-values were obtained, I conducted a one-way ANOVA and two-tail pairwise t-tests between the pairings that compared ornamented taxa to ornamented taxa, ornamented taxa to unornamented taxa, and unornamented taxa to unornamented taxa, all based on the different test definitions for what constituted ornamented taxa. These calculations offered a means for comparing whether ornamented taxa pairings for C-values differed from those that were not between ornamented taxa, and if so, by how much.

2.3.2.2: RRphylo *search.conv* Function

The search.conv function in the R package RRphylo v. v.2.7.0 calculates an angle (θ) representing mean phenotypic difference between lineages, adjusted by phylogenetic time since separation and compared against random values in the dataset to obtain a p-value (Castiglione et al. 2019). This method allows quantitative comparison of lineage similarity based on phylogenetic ridge regression REF. The phenotypic change between a node and a daughter tip is defined by the equation:

$$\Delta y = \vec{B}_1 l_1 + \vec{B}_2 l_2 \ldots + \vec{B}_n l_n$$

n = # of branches between node and tip

$\vec{B}_{1\ldots n}$ = vectors of regression coefficients at each branch

$l_{1\ldots n}$ = branch lengths

Regression coefficients are calculated for each branch and also independently calculated for each variable using a normalization factor (lambda) to prevent multicollinearity (Castiglione et al. 2019).

Each tip is represented by a phenotypic vector. For phenotypic vectors A and B (representing two tips in the tree), the angle θ is calculated by the equation:

$$\theta = arccos \frac{A \bullet B}{|A||B|}$$

This equation is essentially calculating the dot product of two vectors and then extracting the angle between them. That angle, θ, represents the correlation coefficient between the vectors and is used as a measure of morphological similarity. θ goes from 0° to 180°, where 0° represents identical morphological evolution along the two lineages and 180° represents a completely contrasting morphological evolution for each variable. Values of 90° or greater indicate that the two lineages are dissimilar in their evolution.

The phenotypic dissimilarity between any two taxa under the Brownian Motion model is expected to be proportional to the time elapsed since their most recent common ancestor. The *search.conv* calculation samples the phylogenetic tree to calculate a mean observed θ of taxa designated as convergent and compares this to a random θ generated from 1000 random samples of tips within the tree to calculate its statistical significance. If the sample designated convergent produces values outside of the BM model, the p-value is low. Low p-values associated with small angles are interpreted by the authors as convergence, whereas low p-values associated with high angles are interpreted by the authors as non-convergence or divergence.

The search.conv method can perform its computations in a short span of time, but it cannot calculate large numbers of individual trait inputs, necessitating the use of PC scores as a proxy for large trait datasets like the one used in the present study.

The *search.conv* method appears to be robust against homology and offers quantification of evolutionary trajectory and lineage similarity. Actual p-values are only given for the definition of convergence used by Castiglione et al. (2019), and further details about trajectories greater than expected under the Brownian Motion model are not currently provided by this function. As mentioned in the introduction, the θ-value used by Castiglione et al. (2019) is comparable to the

θ defined between lineages by Bolnick et al. (2018). However, Bolnick et al. define θ as an indication of parallel evolution and distinguish this from convergence. In this study, I will use the terminology of Castiglione et al. (2019) when discussing this method as a means of measuring convergence, but I agree with the Bolnick et al. (2018) definition and believe the *search.conv* method is actually an assessment of parallel evolution, and I will clarify this use when interpreting my results in the Conclusions section of this manuscript.

Like the C-values, the *search.conv* method is dependent upon accuracy of ancestral state reconstructions. Without fossil material adjusting phenotypic expectations for the ancestral state reconstruction, the accuracy of ancestral state reconstructions for certain clades (for example, the ratites) are likely to be skewed.

2.3.2.3: Wheatsheaf Index

The Wheatsheaf index (Arbuckle et al. 2014) takes phenotypic distance in trait space between taxa, weights these distances by phylogenetic distance between the lineages, and compares weighted distances of all taxa to weighted distances of focal taxa by creating a ratio (the titular "index"). This method measures convergence as the consequence of convergent evolution; it cannot distinguish between parallel evolution and variations on convergent evolution because it is not a process-based calculation method. Within Arbuckle et al.'s definition, similar organisms are measured by their morphological similarity with the presumption that niche occupation combined with phylogenetic evolutionary history defines an organism's similarity.

The Wheatsheaf index compares a focal group (the purported convergent taxa) to a non-focal group. The index measures how similar the focal species are to one another and how

distinct they are from the non-focal species in phenotypic space, adjusting for phylogenetic relatedness. The process involves creating a phenotypic distance matrix using the following equation:

$$\dot{d_{ij}} = \frac{d_{ij}}{1 - \log\left(p_{ij} + 0.01\right)}$$

The index itself is then calculated with this equation:

$$w = \frac{\overline{\dot{d_a}}}{\overline{\dot{d_f}}}$$

d_{ij} = phenotypic (Euclidean) distance between species i and j

p_{ij} = shared proportional distance between species i and j obtained from the phylogeny

$\overline{\dot{d_a}}$ = mean $\dot{d_{ij}}$ for pairwise comparisons between all species

$\overline{\dot{d_f}}$ = mean $\dot{d_{ij}}$ for pairwise comparisons between focal species

Wheatsheaf index values were calculated using the *windex* v.2.0.3 R package. The index itself is a ratio of the phenotypic distance matrix for all species compared to the phenotypic distance matrix for the focal species specifically. The more similar the focal species are relative to the rest of the taxa, the higher the index will be. If the focal taxa are all closely-related, their phenotypic similarity will be artificially inflated to compensate, meaning the ratio will have a lower value than it would otherwise (taxa are "penalized" proportional to their relatedness).

The p-value is created by bootstrapping the tree tips and the trait values for each taxon and comparing the subsamples simulated by bootstrapping to the calculated Wheatsheaf index value. If the sampled w-values (Wheatsheaf index values) are above the random subsample, the p-value will be low.

Notably, the p-values for the Wheatsheaf tests in this study were usually so high or so low that they were reported within R as 1 or 0. This was not always the case, but it was the case very frequently. Based on the unusually binary p-values, I suspect this test may be sensitive to large morphological datasets. It is also possible that the method it uses for its bootstrapping does not sample broadly enough, hence the very narrow range of randomized subsamples.

There were also many instances where the focal Wheatsheaf value was clearly isolated from the overall distribution based on the histogram output, but was still reported as a high p-value because any Wheatsheaf index value lower than the overall sample average is regarded as evidence of low- or non-convergence. A Wheatsheaf index lower than the sample average indicates the focal species have a wide spread and are concentrated near the outskirts of the overall trait space occupied by the data, so most taxa cluster more closely than they do. According to the Wheatsheaf index, these taxa would be considered non-convergent, though the degree to which they are divergent relative to the rest of the taxa is not part of the Wheatsheaf index output.

This method is not robust against homology, which is why the authors specify that the focal taxa must be determined to be convergent through a separate metric beforehand. The authors describe the Wheatsheaf index as a measure of the strength of convergence, with higher Wheatsheaf values corresponding to higher convergence, and low p-values indicating that the focal taxa have a significantly higher Wheatsheaf index value than would be expected from a random sample. However, if the focal taxa share many homologous traits, the phylogenetic adjustment may not be sufficient to compensate and may suggest that these taxa are highly convergent when in fact they have merely retained their ancestral characters. Use of other

methods such as Stayton's C-values to establish convergence among the focal taxa reduces the issue of homology yielding a false positive.

For the purposes of this study, the limitations of the Wheatsheaf index in relation to homology suggest that this method is still useful for gauging the significance of convergence if detected through the other two methods (with preference given to the C-values due to the search.conv output providing a closer estimate of parallel evolution than convergence). However, if the Wheatsheaf values indicate convergence but the other methods do not, homology is the likely cause for the discrepancy.

2.3.3: Ecological Analyses

Question 3 of this study concerns potential differences between the ornamented and unornamented taxa that are not a direct result of the presence or absence of bony ornament-like structures. In particular, I was interested in seeing whether there could be ecological correlates associated with the presence of bony ornament-like structures. To assess this, ecological trait data was extracted from AVONET for the sampled taxa. Some of these traits are quantitative while others are qualitative. For quantitative traits (mass, centroid latitude, and species range size), data for each species was plotted against an average of the PC scores weighted by contribution of each score to total morphological variance for ANCOVA and MANCOVA tests. The weighted average PC score comparisons assess how the overall bill shape changes in relation to ecological variables and mass across the taxa. A difference between the weighted average PC score comparisons between the ornamented and unornamented groups would indicate that the predictability for how a given ecological trait trends with the bill shape is different between the two groups. Differences in trends between the groups imply that members

of the two groups follow different evolutionary pathways, while overlap between them would indicate they are subject to similar selection and evolutionary responses.

Qualitative traits (lifestyle, trophic niche, trophic level, and habitat) were compared in their distribution between the ornamented and unornamented taxa using Pearson's chi-squared tests. These tests were not corrected for phylogenetic relationship, so are subject to phylogenetic biases in the datasets. However, the broad distribution of both ornamented and unornamented taxa across the dataset for certain definitions still allows a meaningful comparison in the qualitative traits that occur in both groups.

2.3.3.1: ANCOVA

Six two-way phylogenetic analysis of covariance (pANCOVA) tests were performed on the ecological, mass, and shape data for the different test definitions using modifications to the pANCOVA R code developed by Smaers & Rohlf (2016). This code performs a standard ANCOVA while incorporating a phylogenetic tree to weight species in the dataset using generalized least squares by adjusting the contribution of taxa to the dataset according to their covariance with other taxa in order to better fit the assumption of independence in an ANCOVA. The use of ecological and mass data is a means of ruling out (or fail to rule out) explanations for the presence of bony ornament-like structures as a consequence of non-sexual selection.

2.3.3.2: MANCOVA

Two-way phylogenetic multivariate analysis of covariance (pMANCOVA) tests were performed on the ecological, mass, and shape data for the different test definitions using modifications to the pANCOVA R code developed by Smaers & Rohlf (2016). The

pMANCOVA tests measured shape, characterized by a weighted PC score for each taxon, against mass with covariates of latitude, range size, and the interaction factors of mass with latitude, mass with range size, and latitude with range size.

The dependent variables in a MANCOVA are assembled into a multi-dimensional distribution, and when this multi-dimensional distribution is plotted against the independent variable, the relationship between the independent variable and the collection of dependent variables can be summarized as a straight line. The MANCOVA then compares the slopes of the summary lines between two different groups. This line represents the estimated predictability of the dependent variable at given values for the independent variables, with the estimate adjusted to all independent variables simultaneously, as well as how those independent variables covary.

2.3.3.3: Pearson's Chi-Square

Pearson's chi-square was performed using the *chisq.test* function in R v.4.1.2 for the ornamented and unornamented taxa defined by each test designation. The Pearson's chi-square test adjusts for differences in sample size assuming that the expected frequency (calculated as (row sum x column sum)/table sum)) within each category is 5 or greater in at least 80% of cells and the data within the groups are independent of one another. The first assumption works for TDs 1.0-2.2, but for TDs 3.0-3.3, the small number of taxa in the ornamented group becomes a problem and leads the chi-square results to become less reliable.

The assumption of independence is an issue with taxonomic data because of phylogenetic relatedness between clades that means some taxa are more closely related to one another than other taxa in the dataset. For phylogenies that approximate a polytomy, this is less of an issue, but phylogenetic corrections would provide more robust conclusions for any taxonomic

comparison involving more than two taxa. For our purposes, the scattered distribution of ornamented and unornamented taxa for the 1.0-2.2 TDs mitigates the effect of non-independence somewhat, though it is still a concern given the close relatedness of some of the species in our dataset, such as the genera *Balearica* and *Casuarius*, which are very likely to have similar ecological traits to other members of their respective genera. The smaller number of ornamented taxa for TDs 3.0-3.3 also makes those chi-square results dubious because of the low likelihood that these taxa are ecologically representative of all possible bird taxa with these types of bony ornament-like structures.

Nonetheless, we wished to report these data as a starting comparison to illustrate the relative distributions of taxa in these different qualitative categories. Small samples sizes with the chi-square test are prone to Type 2 errors, meaning the chi-square test is unlikely to pick up differences in ecology between the ornamented and unornamented distributions. However, rejection of the null hypothesis is likely to be accurate, so these tests still provide useful data in that regard.

CHAPTER 3

3: Results Section 1

3.1: Phylomorphoplot and PCAs

The following visualizations are the preliminary attempts to answer study Question 1:

To what extent does the morphological similarity of bird skulls surpass the similarity expected by their genetic relationships?

The phylomorphoplot depicted in Figure 7 shows that most taxa in PC1 and PC2 cluster near the left side of the graph—the presumed ancestral region of the phylomorphoplot—but several taxa from at least six of the eight major clades are separated from the main cluster because they possess more disparate morphologies. Many of the outlying branches show taxa that cluster more closely with taxa from other clades rather than their own relatives, which would constitute convergence under some of the aforementioned definitions. Some amount of convergence does appear present, but it is worth noting that few of the outlying taxa cluster tightly together. Instead, extreme divergence away from the central cluster more accurately describes the phylomorphoplot pattern, with extremely long branches from various clades occasionally crossing paths.

The phylomorphoplot and PCA diagrams reflect the variation within the taxa along the first two principal components of the morphometric dataset representing a sum of 54.43% of the total variance within the dataset. Table 4 breaks down the contribution of each principal component to the overall variance in the morphometric data.

Table 4. Percent of variance accounted for by each principal component.

PC1	38.76% of variance
PC2	15.67% of variance
PC3	12.33% of variance
PC4	8.12% of variance
PC5	6.97% of variance
PC6-PC81	<5% of variance each

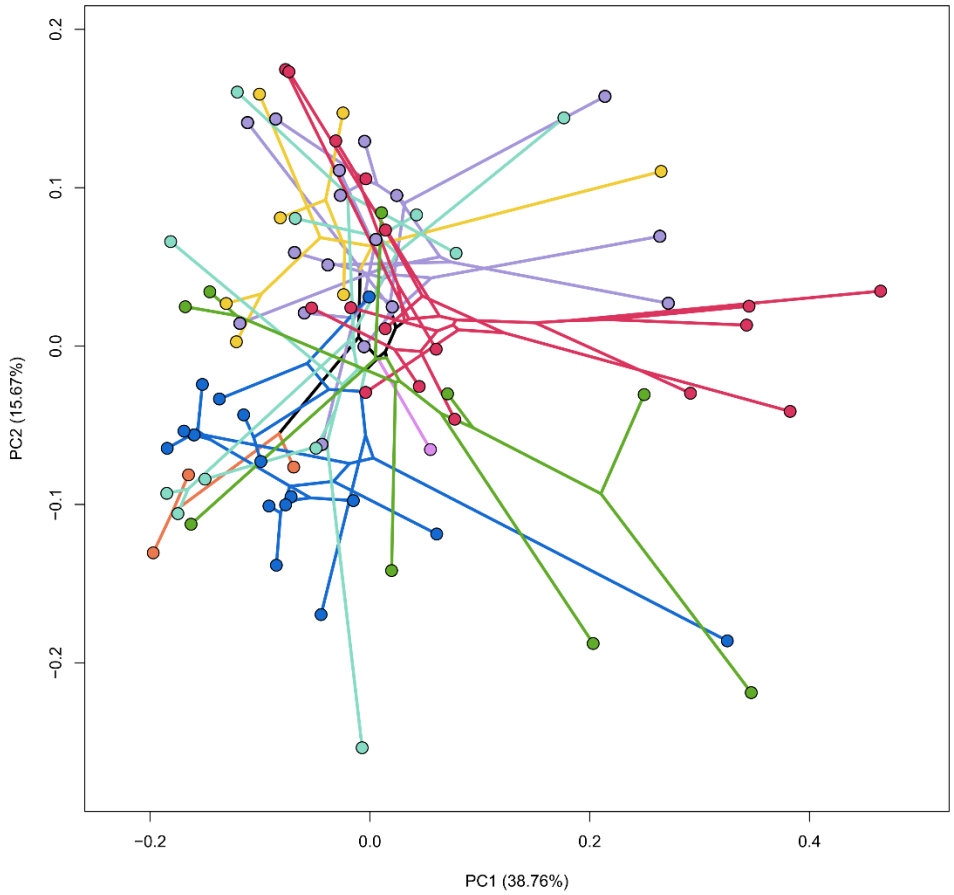

Figure 7. PCA with superimposed phylogeny depicting the phylomorphospace of this study's dataset. Different colors correspond to major clades defined by specific nodes on the phylogeny in Figure 6: Paleognathae *(orange);* Galloanserae *(yellow);* Cuculiformes, Columbiformes, Musophagiformes, *and* Strisores *(teal);* Charadriformes *and* Gruiformes *(blue);* Pelecaniformes, Suliformes, Sphenisciformes, Procellariiformes, Eurypygiformes, *and* Phaethontiformes *(green);* Australaves *(lavender);* Piciformes, Coraciiformes, Bucerotiformes, Accipitriformes, Strigiformes, *and* Trogoniformes *(red);* Phoenicopteriformes *(pink). See Figure 6 and supplementary materials for precise node locations on the phylogeny.*

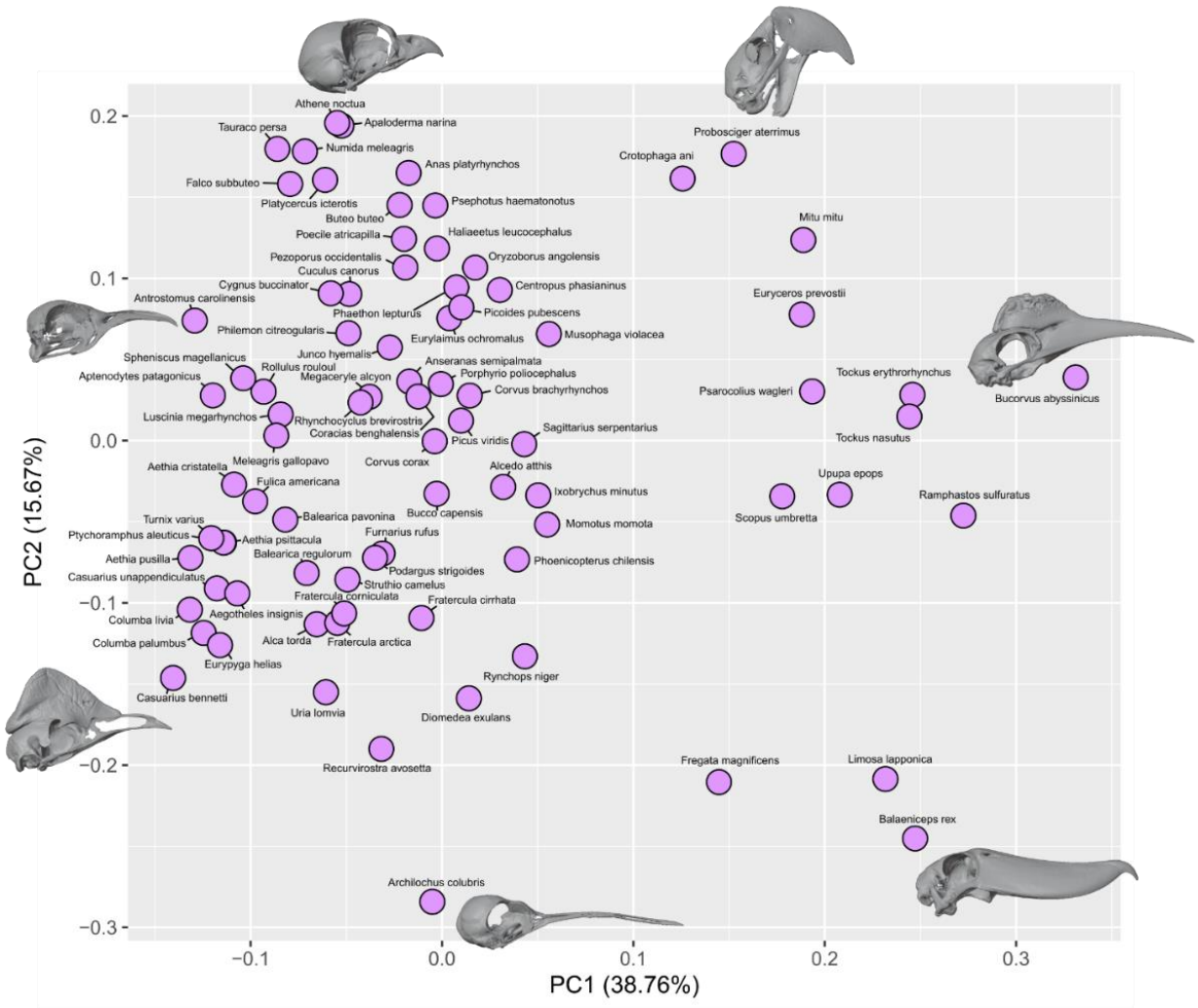

Figure 8*. PCA of bird taxa in this study, comparing PC1 and PC2. PC1 was also plotted against PC3, which contributed to 12.33% of the variance in the morphological data. The PC1 by PC3 plot showed similar overall clustering, so the slightly higher PC2 score is favored for all subsequent visualizations. However, the PC1 by PC3 plot is available in the supplementary materials. Note that even though* Balaeniceps rex *has a broad bill, it is unusual among the birds of this dataset in that a large portion of its bill (mainly the sides) is made up of the maxilla, not the premaxilla. Because the maxilla was not sampled in this study, the* Balaeniceps *bill sampled area appears to be long and thin, which is likely why that taxon plots out with* Frigata *and* Limosa*, which are more traditionally thin-billed taxa.*

Figure 8 shows what components of the bill are likely contributing to the taxa that branch away from the more concentrated cluster. Bill length relative to the rest of the skull appears to be the primary cause for outlying taxa. *Probosciger* (a cockatoo) and *Casuarius bennettii* (a cassowary) at first seem to indicate that extreme bill curvature and expansion is contributing to a

high PC2 score and low PC1 score, respectively, which may be partly true, but this doesn't explain why the other parrots and *C. unappendiculatus* plot closer to the central cluster, or why *Phoenicopterus* (a flamingo, also with a highly curved bill like *Probosciger*) plots in the center of the chart.

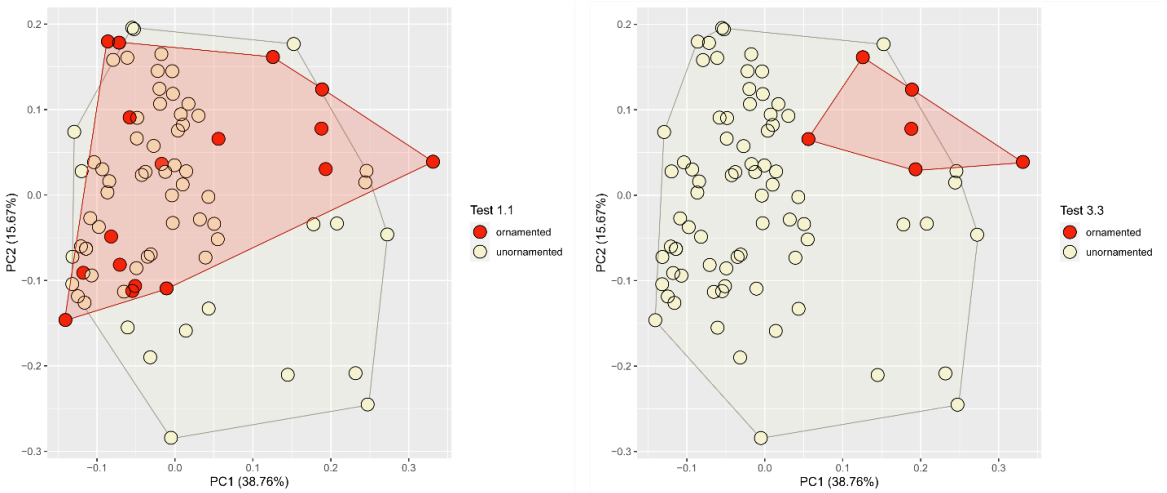

Figure 9. PCAs for PCs 1 and 2 indicating spread and overlap between ornamented and unornamented taxa. Designations for presence of bony ornament-like structures are defined by test definition 1.1 (left) and 3.3 (right). Overall, there is very high spread and overlap. Test definition 3.3 shows the combined lowest overlap and tightest clustering of ornamented taxa out of any of the test definitions.

Figure 9 indicates which taxa were designated as ornamented according to the different test definitions (TDs). Only the PCAs for Test 1.1 and Test 3.3 definitions for indicated ornamented taxa were included in Figure 9 to illustrate some of the variation possible within the convex hull shape for the "ornamented" taxa. Refer to Supplementary Materials Figure 2 for PCAs that highlight the other test definitions. Overall, there is considerable spread among the ornamented taxa for most of the definitions, with test definition 3.3 showing the most tightly-grouped cluster, though even this grouping showed considerable spread.

The taxa designated as ornamented tend to be located in the upper portion of the PCA, though TD 2.2 shows almost complete occupation of the available trait space by taxa designated as ornamented.

There appears to be a general relationship between bill length and/or size relative to the rest of the bill along the axes. This is not entirely surprising, as PCAs tend to be sensitive to size differences, and even though Procrustes superimposition can reduce scaling effects, length proportions tend to affect overall shape considerably.

3.2: Cluster Analysis Dendrogram

The dendrogram in Figure 10 is entirely based on morphology without accounting for phylogeny, therefore, closely related taxa will tend to group together, especially among taxa that retain similar bill shapes from a common ancestor (for instance, both species of *Columba* group together as do the woodpeckers, *Picus* and *Picoides*). However, there are some unexpected groupings visible. In some cases, closely related taxa, like the two different species of *Corvus*, may show one member morphologically grouping with a different taxon. *Corvus brachyrhynchos* (a passerine) groups with *Coracias benghalensis* (a coraciiform), whereas *Corvus corax* (a passerine) groups with *Porphyrio poliocephalus* (a gruiform). This may be related to morphological differences in these taxa related to their ecology; *Corvus corax* is a larger bird with a heavier bill suited to feeding on larger prey items, *Corvus brachyrhynchos* is more slender-billed, and the same is true of *Porphyrio* and *Coracias*, respectively. *Porphyrio* species have more plant material in their omnivorous diets than *Corvus* species and *Coracias* species feed on less plant material than *Corvus*.

For ornamented taxa, closely related species group together as expected (*Casuarius bennettii* with *C. unappendiculatus*, etc.), but there are noted exceptions. The two musophagids, *Musophaga violacea* and *Tauraco persa* fall out in completely different regions of the dendrogram, and *Crotophaga ani* and *Euryceros prevostii* group next to one another despite *Crotophaga* being more closely related to *Centropus* and *Cuculus*, and *Euryceros* being more closely related to the *Corvus* species. This suggests that the overall skull shape of *Crotophaga ani* and *Euryceros prevostii* could potentially be convergent from a pattern-based perspective.

In general, there is little clustering of ornamented taxa with one another, with a few exceptions. As expected, closely related ornamented taxa group together, such as *Balearica* and *Casuarius*, but there is also an exceptional cluster of eight ornamented taxa (marked by the "A" node in Figure 10) that includes a disproportionate number of ornamented taxa with few close phylogenetic relationships to one another, and which generally do not group with their closed unornamented relatives. These taxa all share a similar type of ornament, formed from an extension of the dorsal process of the premaxilla, though not all taxa (notably the *Fratercula* species and *Tauraco persa*) with this type of ornament are included in this cluster. Non-ornamented members of this cluster, as well as members of another cluster that includes *Tockus nasutus* (a bucerotid) and *Ramphastos sulfuratus* (a piciform), are generally large-billed species. Members of Bucerotidae and Ramphastidae are suspected to be convergent in bill morphology and likely ecological function of their bill enlargement, so the grouping of these taxa together is consistent with those observations (Van Den Ven 2016). The taxa marked by the "A" node are similar, but not identical, to the taxa identified as ornamented in TD 3.3 (see Figure 10).

The Figure 10 dendrogram includes TD 1.1 and TD 3.3 definitions for ornamented taxa to illustrate how differences in definitions alter whether the ornamented taxa cluster together. Most

definitions produced ornamented taxa distributed widely across the dendrogram, indicating that either the morphology of the bony ornament-like structure was not similar between the ornamented taxa, or did not contribute enough to the overall morphology to affect the morphological clustering of the taxa. (Refer to Supplementary Materials Figure 1 for dendrograms showing the distributions of other test definitions.) The dendrogram branch lengths show a generally positive correlation with phylogenetic branch lengths (Figure 11), but there is a large amount of scatter in this trend.

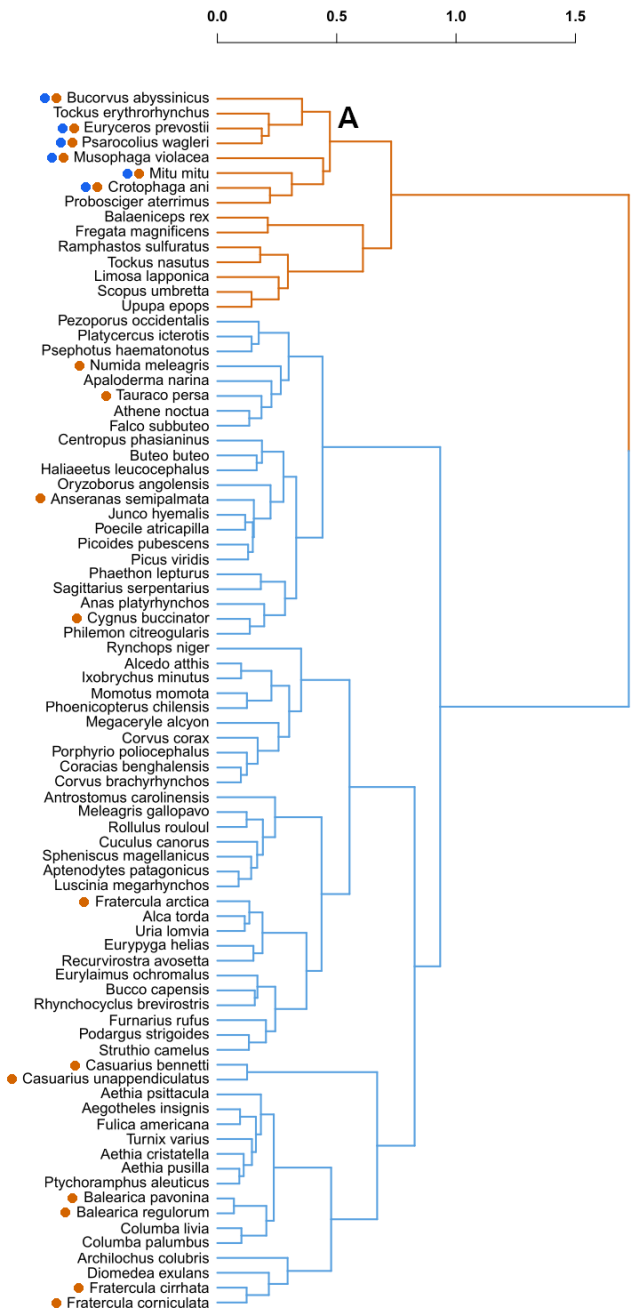

Figure 10. Orange dots correspond to ornamented taxa defined in test definition 1.1, blue dots correspond to ornamented taxa defined in test definition 3.3. The two most numerically distinct clusters of taxa are marked in orange and blue branches. Point A indicates the most closely clustered node containing Test 3.3 taxa. Note that these relationships are purely morphological and on their own do not account for similarity due to homology.

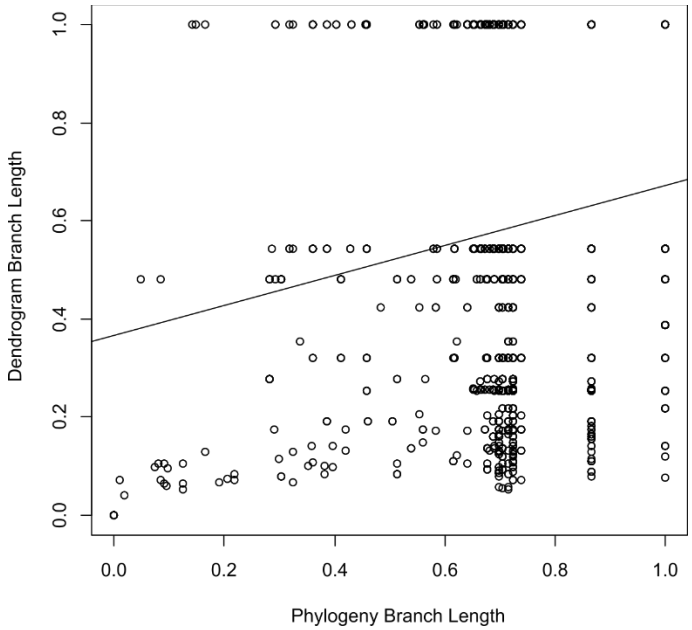

Figure 11. Scatterplot of branch lengths on the phylogeny compared to the dendrogram based on the VCV chart of all taxon pairings. The slope of the trendline is 0.307. A slope of 0 or infinity would mean no association between phylogeny and morphology while a slope of 1 would mean that morphology is perfectly predicted by phylogeny (as predicted by the Brownian Motion model). A slope greater than 1 would indicate that morphology changes dramatically with small increases in time (more time influence), and a slope of less than 1 indicates morphology changes only a small amount with larger increases in time (less time influence). The slope of ~0.3 suggests that time has a large influence on morphology, but it is not a one-to-one ratio. The high spread also indicates that the relationship between time and morphology is not especially strong, as many taxa deviate from the average and skew the trendline. Generally, taxa that share more time (left side of the graph) also have a higher morphological similarity (bottom of the graph). However, the shape spreads out over time, with several taxa deviating considerably in morphology from one another after only about 5-20 million years of separation. One would expect taxa above the trendline to be those that have diverged more rapidly than expected, while one would expect taxa below the trendline to be those that have either experienced morphological stasis or convergence to some extent.

ANOVA tests were run to determine if there are differences between the three possible groupings: ornamented taxon compared to ornamented taxon (OO), ornamented taxon compared to unornamented taxon (OU), and unornamented taxon compared to unornamented taxon (UU) (Figures 23–29). All analyses were repeated for each TD. If taxa with bony ornament-like structures have some level of morphological convergence with one another, the OO pairings will

have non-zero average C-values, and if these birds are exceptionally convergent, the OO pairing C-values will be greater than the OU pairings.

The phylomorphoplot, PCAs, and hierarchical cluster analysis dendrograms all suggest that there is some amount of convergence among certain taxa in this dataset, including some potential convergence among ornamented taxa, depending on the definition used. The C-values, search.conv θ-values, and Wheatsheaf index provide quantitative assessment of this apparent convergence.

3.3: C-Values

Test 1.0	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	9.585	**7.07e-05**	5.313	**0.00497**	14.2	**7.22e-07**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	0.622	0.723	0.932			
OO-UU	**0.0395**	0.128	**0.0413**			
OU-UU	**0.0000459**	**0.00236**	**0.000000305**			

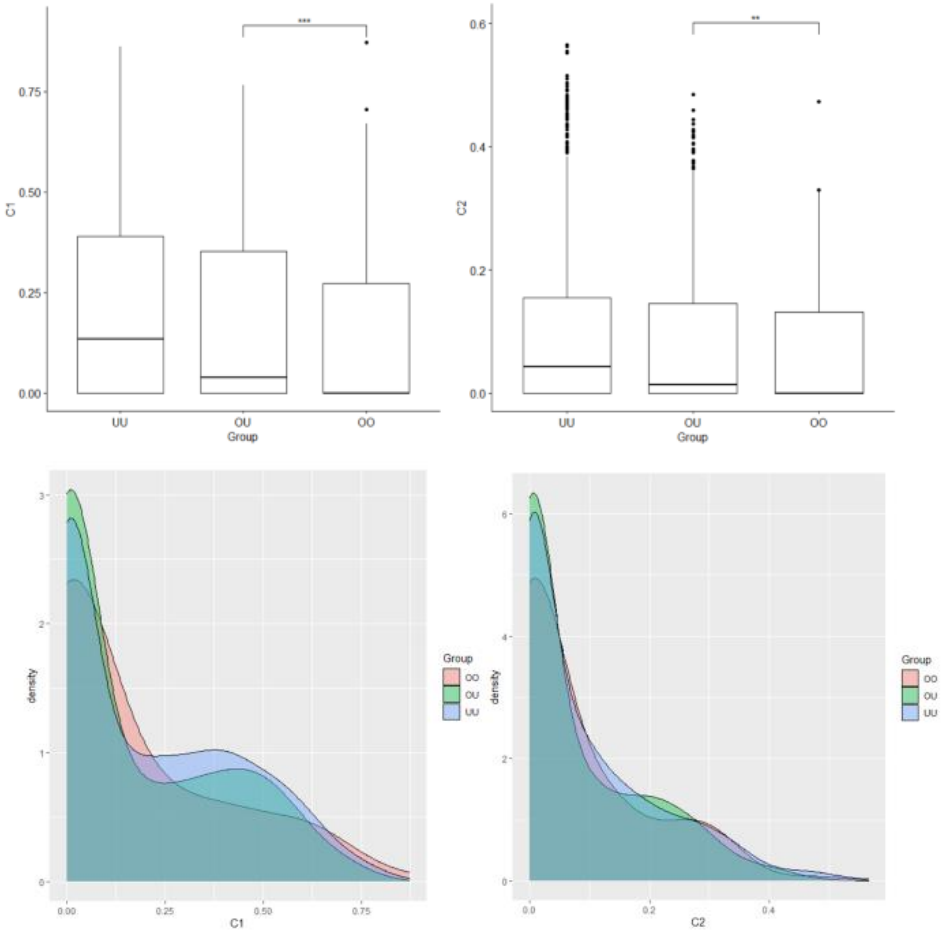

Figure 12. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 1.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 1.1	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	18.02	**1.65e-08**	10.41	**3.11e-05**	23.26	**9.27e-11**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	0.609	0.745	8.89e- 1			
OO-UU	**0.00386**	**0.0322**	**3.85e- 3**			
OU-UU	**0.0000000199**	**0.0000177**	**7.4 e-11**			

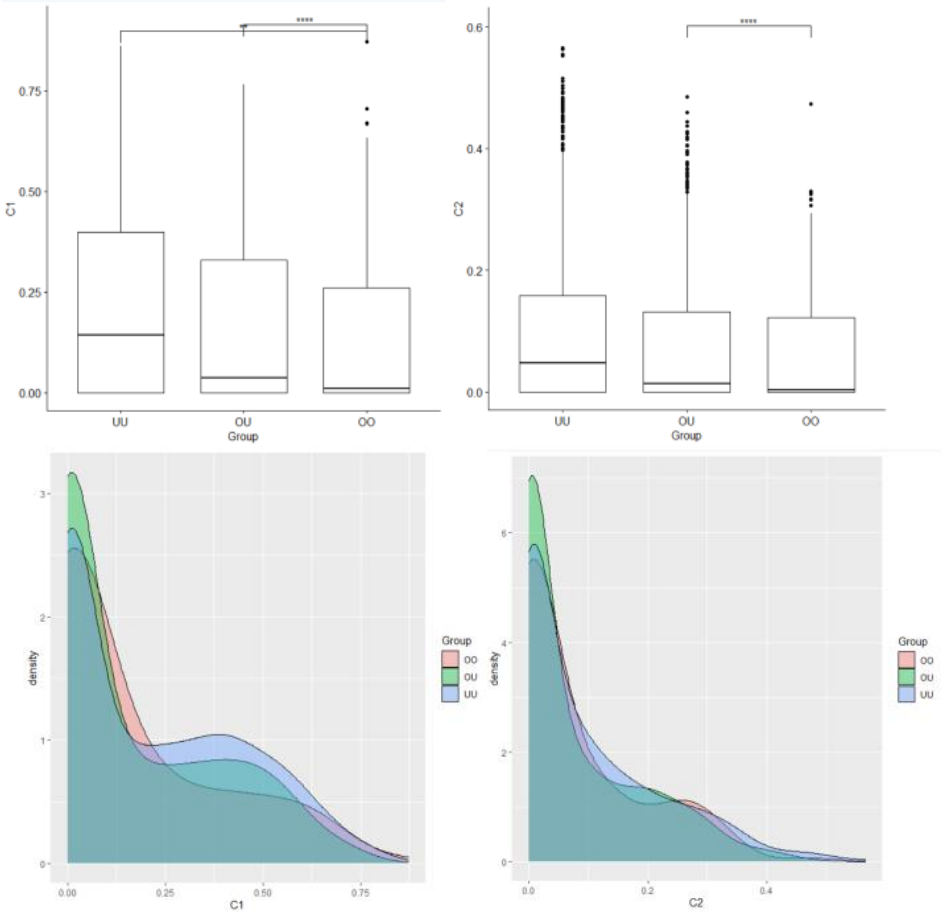

Figure 13. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 1.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 2.0	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	20.59	**1.29e-09**	5.784	**0.00311**	19.93	**2.49e-09**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	0.952	0.628	8.34e- 1			
OO-UU	**2.11e- 2**	0.43	**3.33e- 2**			
OU-UU	**5.57e-10**	**0.000736**	**8.43e-10**			

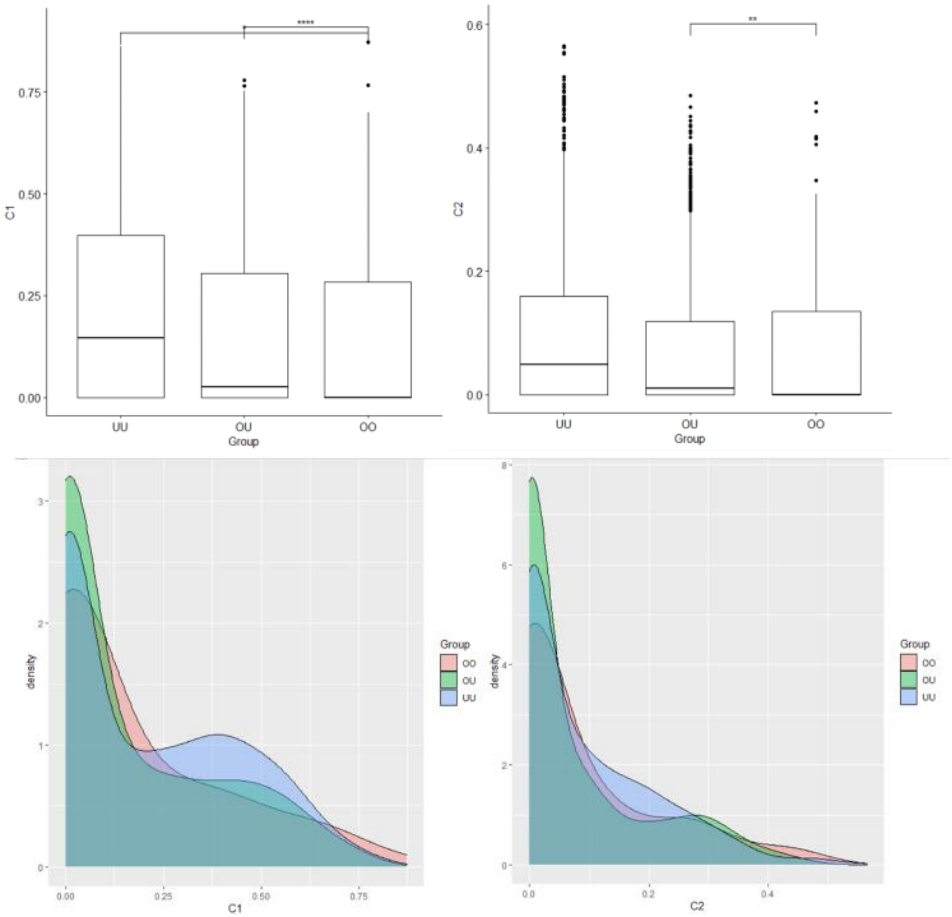

Figure 14. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 2.1	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	52.76	**<2e-16**	29.08	**3.02e-13**	57.36	**<2e-16**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	**3.85e- 3**	**3.33e- 2**	**5.78e- 3**			
OO-UU	**1.17e-21**	**1.14e-12**	**5.64e-23**			
OU-UU	**8.8 e-17**	**5.57e-10**	**1.15e-18**			

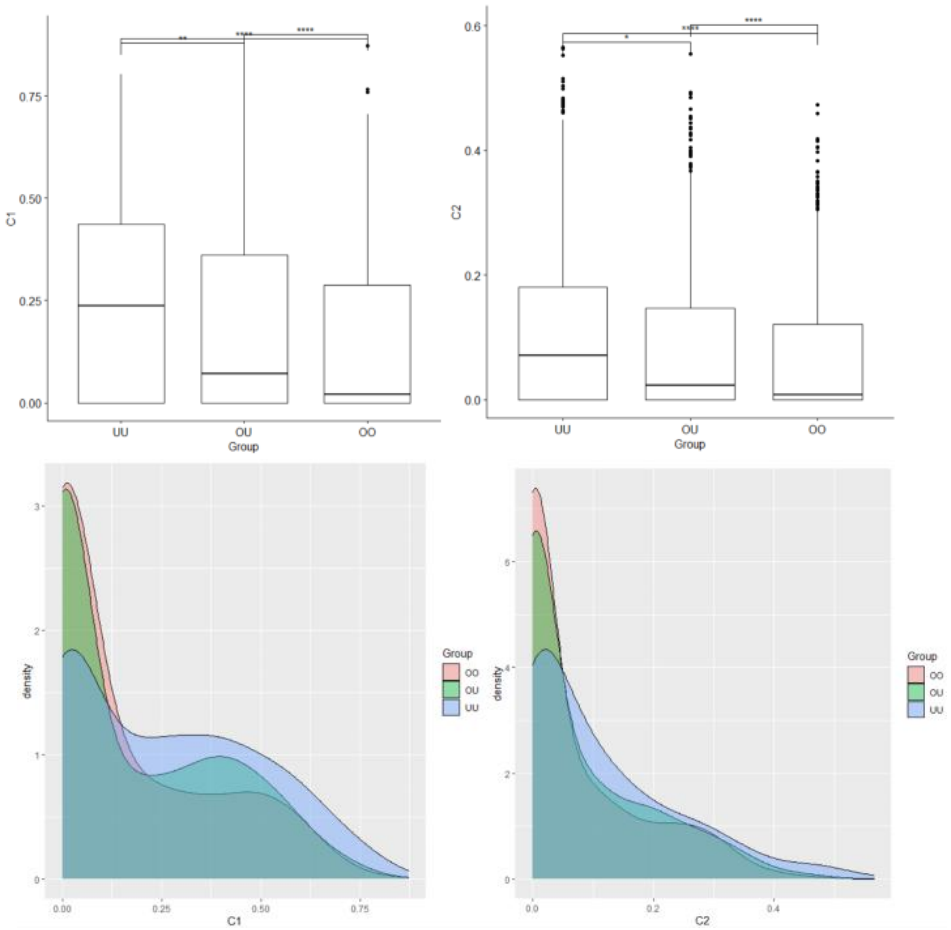

Figure 15. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 2.2	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	5.463	**0.00428**	2.331	0.0973	15.35	**2.31e-07**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	0.775	0.634	0.426			
OO-UU	**0.00424**	0.088	**0.000000121**			
OU-UU	**0.00134**	**0.0318**	**0.000000838**			

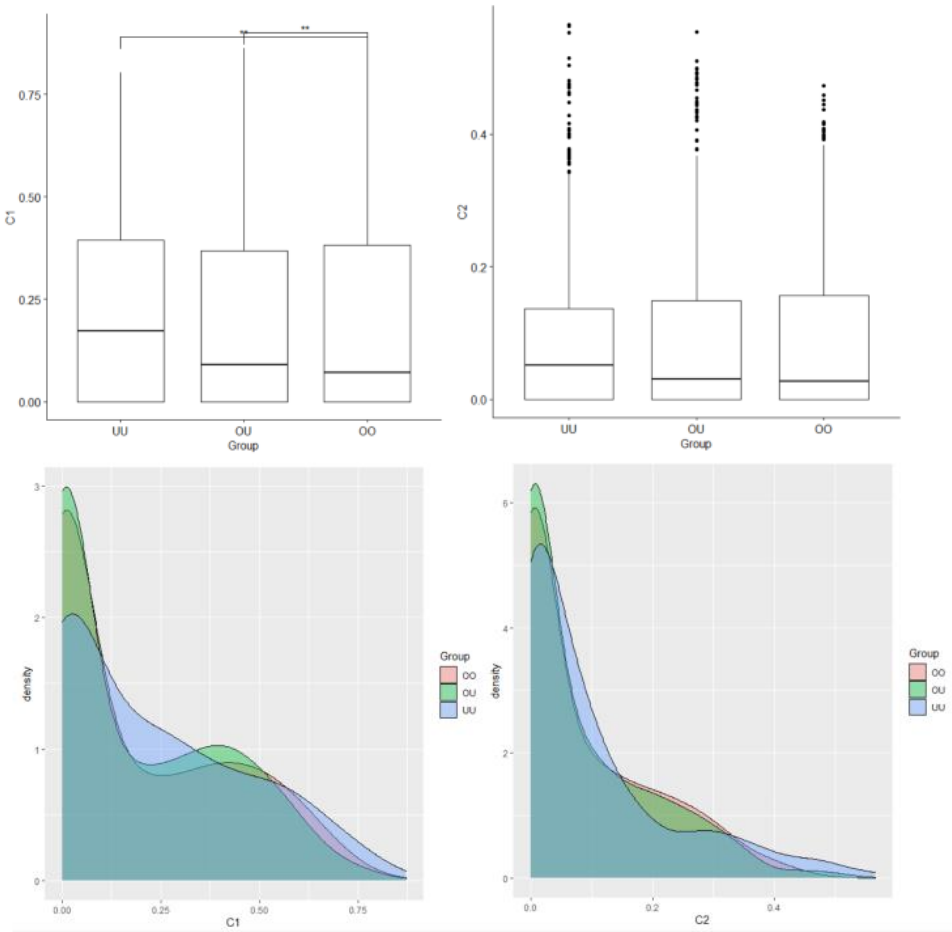

Figure 16. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 2.2 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 3.0	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	94.25	**<2e-16**	74.26	**<2e-16**	89.07	**<2e-16**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	9.15e- 1	9.48e- 1	8.49e- 1			
OO-UU	**3.26e- 4**	**1.57e- 3**	**1.33e- 3**			
OU-UU	**5.08e-40**	**4.28e-32**	**2.73e-38**			

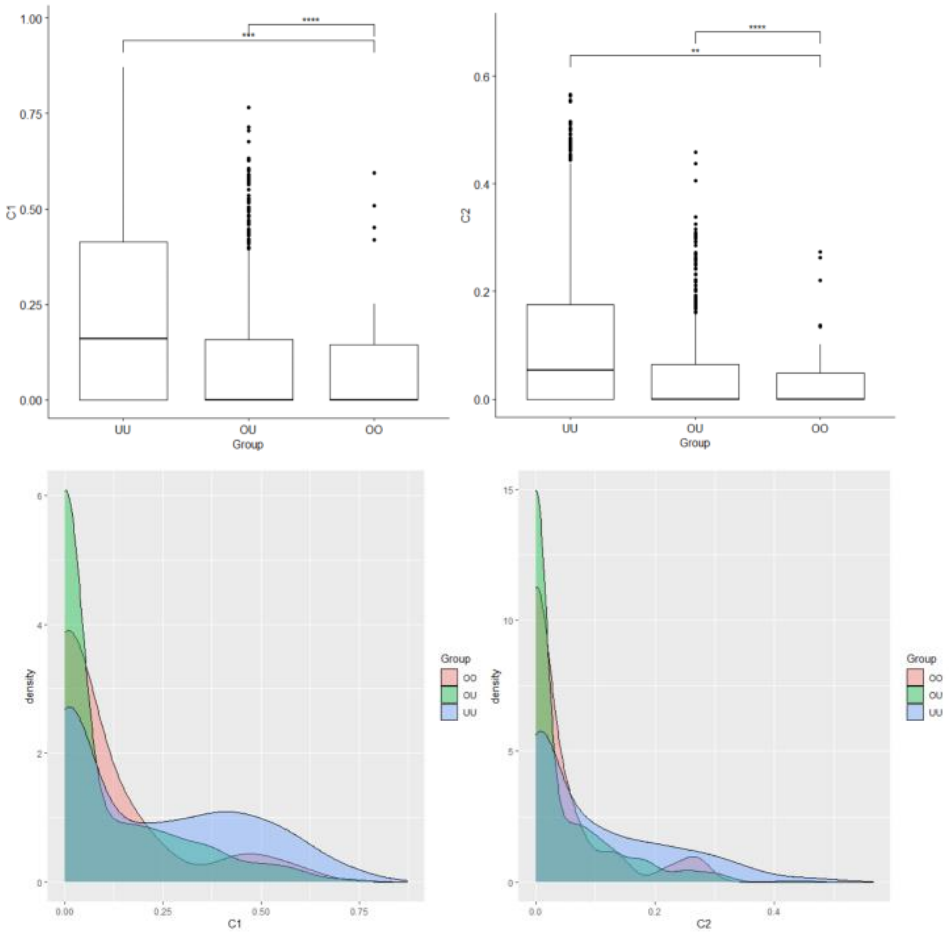

Figure 17. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.0 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 3.1	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	48.57	**<2e-16**	37.66	**<2e-16**	42.3	**<2e-16**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	6.21e- 1	6.73e- 1	6.71e- 1			
OO-UU	1.02e- 1	1.54e- 1	1.37e- 1			
OU-UU	**4.31e-22**	**1.53e-17**	**1.68e-19**			

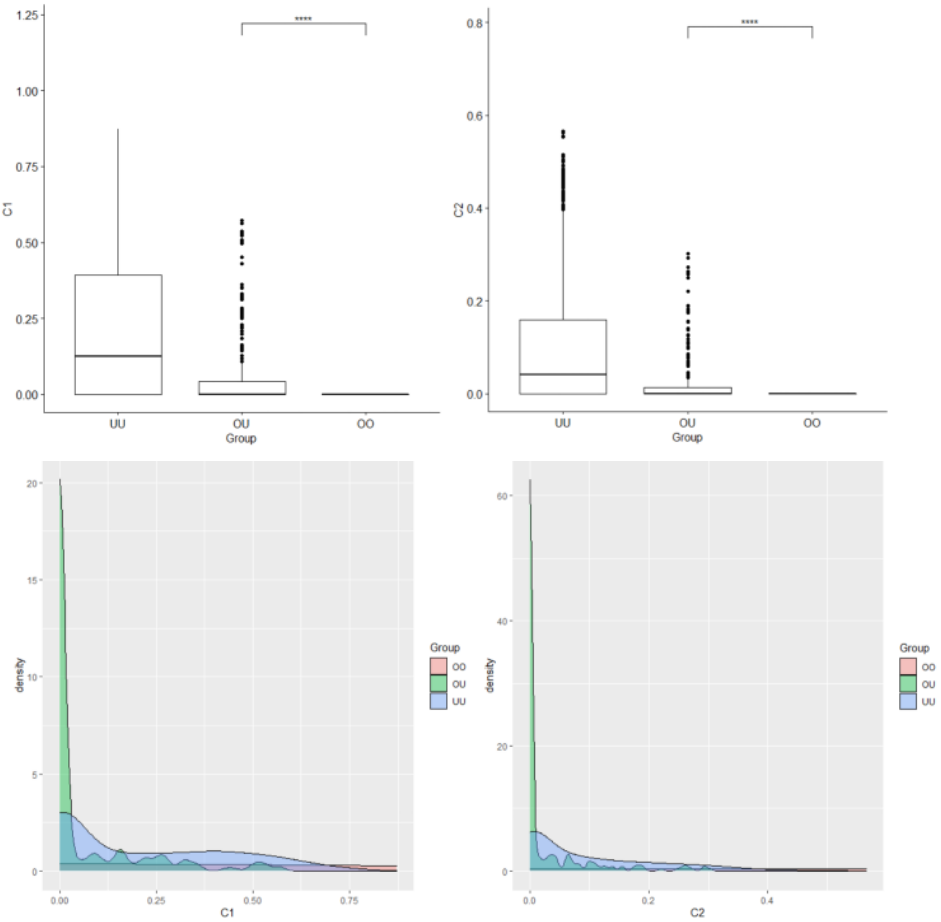

Figure 18. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.1 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Test 3.2	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	70.89	**<2e-16**	46.7	**<2e-16**	55.19	**<2e-16**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	2.39e- 1	3.47e- 1	2.55e- 1			
OO-UU	5.86e- 1	6.47e- 1	7.03e- 1			
OU-UU	**5.17e-32**	**8.74e-22**	**2.08e-25**			

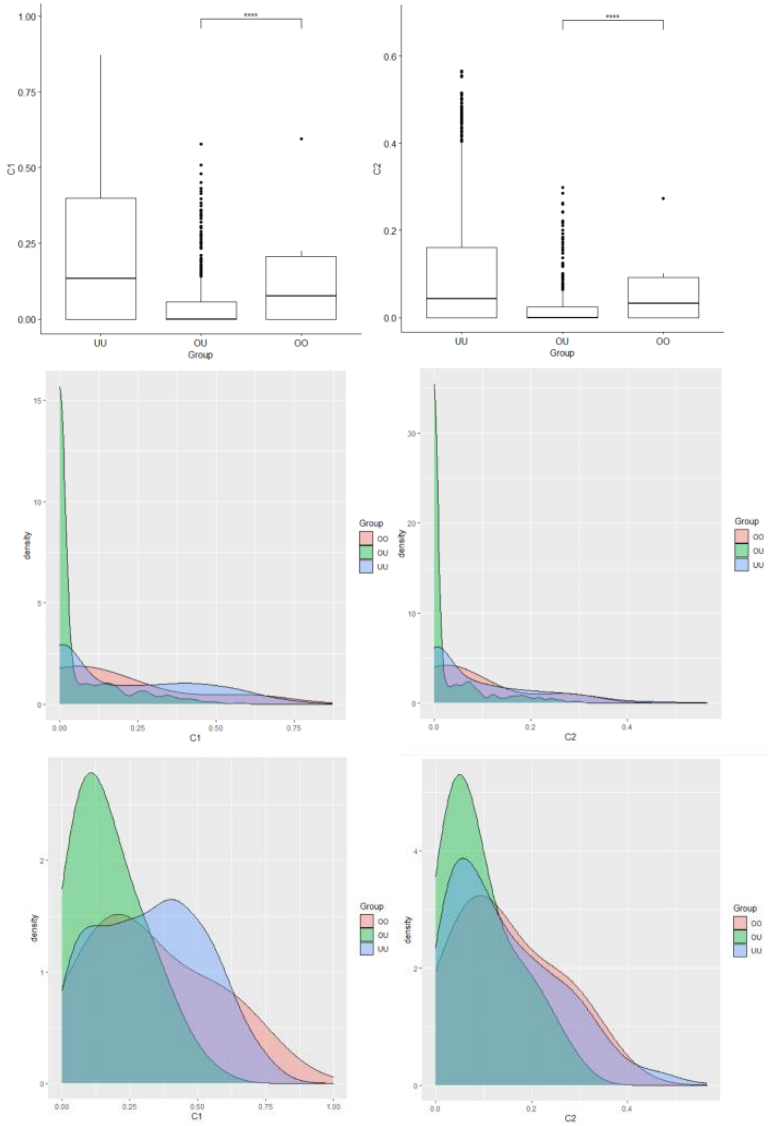

Figure 19. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.2 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Because the TD 3.2 graphs had an extremely high proportion of C-values with values of 0, I removed the values of 0 in the bottom two density plot graphs to provide a more legible comparison between the different points. The bottom two graphs are not meant to offer a direct point of comparison to the density plots for the other tests aside from test definition 3.3.

Test 3.3	C1 F-value	C1 p-value	C2 F-value	C2 p-value	C3 F-value	C3 p-value
ANOVA	130.7	**<2e-16**	89.5	**<2e-16**	102.3	**<2e-16**
T-test Pairing	C1 p-value	C2 p-value	C3 p-value			
OO-OU	**2.56e- 2**	5.29e- 2	**3.44e- 2**			
OO-UU	3.77e- 1	5.2 e- 1	5.21e- 1			
OU-UU	**1.32e-56**	**8.78e-40**	**4.39e-45**			

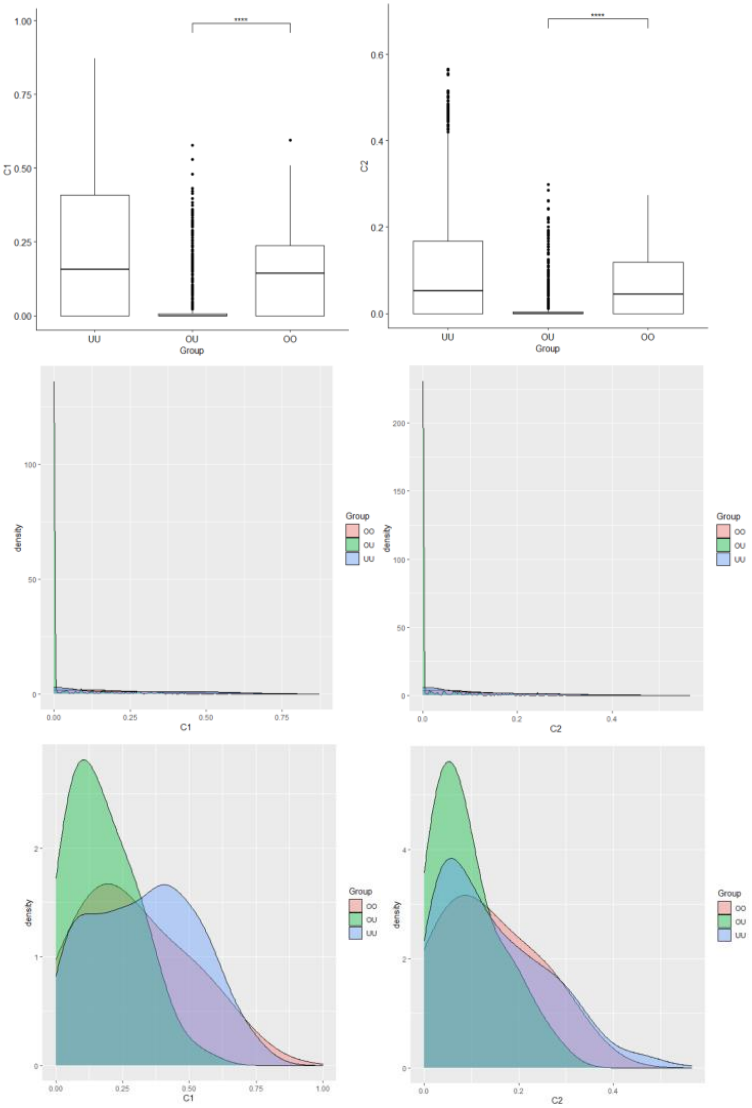

Figure 20. ANOVA, t-tests, box-and-whisker plots, and density plots using test definition 3.3 for ornamented taxa. Only C1 and C2 values are shown for histogram and density plots. OO = ornamented-ornamented pairing, OU = ornamented-unornamented pairing, UU = unornamented-unornamented pairing.

Because the TD 3.3 graphs had an extremely high proportion of C-values with values of 0, I removed the values of 0 in the bottom two density plot graphs to provide a more legible comparison between the different points. The bottom two graphs are not meant to offer a direct point of comparison to the density plots for the other tests aside test definition Test 3.2.

From these tests, we can see that different TDs offer different C-value interpretations, emphasizing the importance of clearly defining what taxa are being considered to be ornamented within the study. ANOVA and t-tests also differed between the C-value types for many of the given tests, so the interpretation of whether the groups differ or not depends on the metric used. For the present study, I prioritize the C1 and C2 values over the C3 values. Agreement between two or more tests offers stronger support for a given interpretation, so for instance, in cases where the C2 and C3 values both show low p-values for the ANOVA, but the C1 value does not, the C2 and C3 values will be taken as the more consistent reflection of how the different groupings compare.

All of the ANOVA tests found a difference between the OO, OU, and UU groups. Test definition 2.2 has slightly higher p-values, so the possibility of falsely rejecting the null hypothesis that all of the groups are the same should be considered for that test definition, but the other test definitions all have very low p-values for the ANOVA.

Test definition 1.0, 2.0, 3.0, and 3.1 had a median value of 0 for their C-values in the OO comparison, indicating that despite a few taxa in these definitions for ornament-like structures may be convergent, the groupings do not indicate convergence according to the C-value tests. Test definitions 1.1 and 2.1 offered slightly higher but still low C-values. Test definition 2.2 had a distinctly non-zero median C-value score among its pairings, suggesting the designated ornamented taxa are potentially convergent with one another. However, these values were comparable to the UU pairings and almost identical to the OU pairings, which when combined with the large number of taxa designated as ornamented for this test definition suggests that there is not necessarily anything unique to the OO pairings that suggests they are more convergent with one another than they are to the unornamented taxa according to the test definition 2.2. Test

definitions 3.2 and 3.3 showed some of the most distinct results, with a distinct separation between the ornamented and unornamented taxa, and ornamented taxa having relatively high C-values within the dataset. The 3.2 and 3.3 test definitions OO grouping had a smaller upper limit and 2nd quartile distribution of C-values than the UU grouping. Overall, the C-values suggest that the TDs 1.0-2.1 and 3.0-3.1 definitions for ornamentation do not indicate convergence. The TD 2.2 might indicate convergence, and the 3.2-3.3 test definitions offer the strongest support for convergence in the dataset. However, none of the ornamental taxa groupings show greater convergence to one another than can be observed in the rest of the dataset.

The spread of possible C-values within each pairing category and TD was very high, except for the 3.1 OO group, which was exclusively 0 for all C-values. Note that because C-values can be skewed by parallel evolution (Bolnick et al. 2018 definition) in some cases, TDs 5.2 and 5.3 could be detecting parallel evolution instead of convergence.

3.3.1: Test Definition 1.0–1.1

The TD 1.0 and 1.1 results are similar, as might be expected given these definitions only differ from one another by a few taxa. The C1 values for both these tests show $p>0.05$ for the OO-OU comparison, $p<0.05$ for the OO-UU comparison, and $p<<0.05$ for the OU-UU comparison. C3 values follow the same trend as the C1 values, but C2 values differ slightly, with the C2 comparisons for TD 1.0 showing $p>0.05$ for the OO-OU and OO-UU comparisons, but $p<0.05$ for the OU-UU comparison. For C2 values in TD 1.1, $p>0.05$ for OO-OU, $p<0.05$ for OO-UU, and $p<<0.05$ for OU-UU, as with the C1 values. C4 values for these tests show an inverted trend (OO-OU and OO-UU $p<<0.05$, OU-UU $p<0.05$), but will largely be ignored for these analyses in favor of agreement amongst the other values, as mentioned earlier.

The box-and-whisker plots for the C-value distributions grouped by comparison type reveal subtle but potentially meaningful differences among these groups. For the C1 and C2 values in both TD 1.0 and 1.1, UU comparisons have generally greater values than OU comparisons, and OU comparisons have greater C-values than OO comparisons, with the median OO value hovering at or near a value of 0. The minimum and 2nd quartile values are bounded at 0, but the maximum and 3rd quartile values follow the same trend.

The density plots reveal that all of the grouping have a considerable number of C-values at or near 0, but for higher C-values, there are a large amount of OU and UU C1 values between 0.25 and 0.5 compared to the OO C1 values in both TD 1.0 and 1.1. The C2 density plots are less clear, but seem to show an even trend of fewer UU points at higher C2 values, as well as a larger than expected cluster of OU C2 values around 0.2, and a similar unexpectedly large cluster of OO C2 values that are slightly higher, around 0.3.

3.3.2: Test Definition 2.0–2.2

For C1, TD 2.0 has $p>0.05$ for the OO-OU comparison, $p<0.05$ for the OO-UU comparison, and $p<<0.05$ for the OU-UU comparison. This trend is slightly different for the C2 values, where OO-OU has $p>0.05$, OO-UU has $p>0.05$, and OU-UU has $p<0.05$. For TD 2.1, all of the C1 and C2 comparisons have $p<0.05$, with the OO-UU and OU-UU comparisons having exceptionally low p-values. Test 2.2 shows a similar trend to TD 2.0, with OO-OU $p>0.05$, OO-UU $p<0.05$, and OU-UU $p<0.05$. For all of these tests, the C3 comparisons agree with the C1 comparisons.

The box-and-whisker plots show similar trends to the TD 1 subtests, where the UU values are highest for all of these values and OO values are generally very low. The median for

OO is at or around 0 for TD 2.0 and 2.1, but higher for TD 2.2. In TD 2.2, the median for the OO and OU groups is very similar, as is their overall distribution. Distributions are generally more similar for C2 test values than the C1 values between the different groups, perhaps related to the slightly higher phylogenetic bias in the C1 tests.

The C1 density distribution plots for TD 2.0 and 2.1 are similar to those of the TD 1 subtests. For the TD 2.2 C1 density plot, the OO and OU groupings show a peak near the value of 0 and a smaller peak between a C1 value of 0.4 to 0.5, but the UU grouping shows an almost linear relationship between C-value size and density for C1. The C2 density plot for TD 2.0 has a small peak in OU values around 0.3, but no clear peak in OO and UU values aside from the initial peak at 0 seen in almost all of the tests. The TD 2.0 C2 values show slightly more OO values above 0.4 than in the other groups. The TD 2.1 C2 density plot looks similar to that of the TD 1 subtests, though with less pronounced peaks. The TD 2.2 C2 density plot shows almost identical trends between the OO and OU groups, but a sharp drop in UU values above 0.2 relative to the other groups. The OO and OU groups decline in frequency at higher C2 values for TD 2.2 around a value of 0.4, so there are more UU values above 0.4 than in the other groups.

3.3.3: Test Definition 3.0–3.3

The TD 3 subtests showed considerable variation in their results compared to the other groupings, despite many of these groups including many of the same taxa. That itself could indicate that at least one or more taxa from the 3.0 definition (from which the TD 3.1, 3.2, and 3.3 were selected) has a disproportionate effect on the C-value results compared to other taxa. The TD 3.0 t-tests produced $p>0.05$ for the OO-OU comparison, $p<0.05$ for the OO-UU comparison, and $p<<0.05$ for the OU-UU comparison for all of the C1, C2 and C3 values. For

Test 3.1, $p>0.05$ for the OO-OU and OO-UU comparisons, but $p<<0.05$ for the OU-UU comparisons for all of the C1, C2 and C3 values. The same was true for TD 3.2. Test definition 3.3 produced unique results compared to the other tests, with $p<0.05$ for the OO-OU comparison but $p>0.05$ for the OO-UU comparison and $p<<0.05$ for the OU-UU comparison for the C1 and C3 values; the C2 results were similar for TD 3.3, except that the p-value for the OO-OU comparison was slightly higher than 0.05.

Test definitions 3.0 and 3.1 show box-and-whisker plots with low OO and OU median values (near or at 0) and much higher UU values. The maximum values of the UU comparisons are greater than those of the OU comparisons, which are greater than the maximum values of the OO comparisons (though note that all TD 5.1 OO comparison values are 0). The TD 3.2 and 3.3 box-and-whisker plots, meanwhile, showed unique trends compared to the other tests. For each of these, the median C1 and C2 values for the OU comparisons was near 0, though there were several higher OU outliers. The OO and UU values were much higher, showing very similar median values to each other, while the general spread and 3rd quartile values were higher in the UU than OO groups.

The density plots for the TD 3 subtests were generally more complex than in the other tests, with a larger number of peaks in C-value densities across the range of possible values, especially for the OU pairingts. This could have a variety of interpretations. All of the TD 3 subtest density plots showed far more 0 values in the OU group than the OO or UU groups. In TD 3.0, there was a greater propensity for C1 values between about 0.1 and 0.2 in the OO group compared to the others, and larger numbers of high UU values than in the other groups. Both the OO and UU groups see a secondary peak in C1 values around 0.4. UU C2 values are also

generally greater for TD 3.0, though the OO C2 values show relatively high abundance from about 0.03-0.07 and around 0.36.

The TD 3.1 density plot has a massive concentration of OU values at 0, with some higher OU values. The UU values are more evenly spread across the possible C1 and C2 values by comparison, decreasing in frequency at higher values. The OO values appear to be evenly distributed across the C-values for both C1 and C2, but this is an artifact of the small OO sample size and the way the kernel density is calculated; in reality, all of the OO values are 0.

The TD 3.2 density plot is similar to that of test definition 3.1, but with more substantial OO values. The OO values show a wide peak around 0.1 for the C1 values and 0.05 for the C2 values, while the UU values show a peak at 0 and then a relatively even distribution of higher C1 and C2 values generally exceeding that of the other categories. There is a very broad peak in C1 values between around 0.37 and 0.5 for the UU group.

Test definition 3.3 has such a substantial number of 0-value OU comparisons relative to the OO and UU groups that the legibility of the plot is compromised, but when the 0-values are removed, as seen in Figures 19 and 20, the plots for test definition 3.2 and 3.3 are very similar to one another. The C1 OO values show a slight peak near 0.15 and the UU values show a slight peak near 0.40, but the C2 values show little to no differentiation in their distribution between the OO and UU values.

It should be noted that there were only three taxa in TD 3.1, producing a total of three unique comparisons between different taxa, and there were only four taxa in TD 3.2, producing a total of six unique comparisons, both of which are low sample sizes that could easily skew the t-values. This dataset is vulnerable to small sample size bias among its ornamentation selections. Test definition 3.3, which has many of the same ornamented taxa as test definitions 3.1 and 3.2

but includes more individuals for a total of 15 unique OO comparisons, and TD 3.0, which has 36 unique OO comparisons, are likely less prone to small sample size effects, though perhaps not as much as in some of the other tests.

3.4: *search.conv* Time-Adjusted θ

*Table 5. The search.conv mean θ-values between species defined as ornamented, divided by the time distance between them (the ang.state.time output for the search.*conv *function), along with the associated p-values for the time-adjusted angles. The results are given for both the full skull and each of the bones in the skull that were landmarked for this study. Blue boxes indicate test-region subsamples with low p-values (p<0.05). For light blue boxes, p<0.1.*

search.conv Full Bill Region

		Test 1.0	Test 1.1	Test 2	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	**θ**	0.7620	0.7441	0.7277	0.6689	0.6689	0.7850	0.6383	0.3256	0.3039
	θ p-value	0.989	0.972	0.927	0.553	0.557	0.965	0.445	**0.006**	**0.001**
Nasal Bone	**θ**	0.8620	0.8051	0.8415	0.6646	0.6646	0.7679	0.2221	0.3041	0.2682
	θ p-value	1	0.999	0.998	0.133	0.131	0.914	**0.013**	**0.016**	**0.001**
Frontal Bone	**θ**	0.7487	0.7548	0.7404	0.6823	0.6823	0.8256	0.7406	0.2072	0.2783
	θ p-value	0.963	0.987	0.944	0.784	0.781	0.986	0.623	**0.002**	**0.002**
Premaxillary Bone	**θ**	0.7302	0.7255	0.6826	0.6763	0.6763	0.8070	0.7299	0.4567	0.3562
	θ p-value	0.921	0.937	0.617	0.602	0.595	0.977	0.623	**0.06**	**0.002**

In general, few of the TDs for ornamented focal taxa yielded evidence of convergence. However, three TDs–3.1, 3.2, and 3.3–gave values suggesting the ornamented taxa grouped more tightly with other ornamented taxa than a random assortment of taxa within the dataset ($p \leq 0.1$), suggesting these taxa could be considered convergent (Table 5). All elements of TDs 3.2 and 3.3 grouped as convergent according to the search.conv analyses, providing decent evidence that the taxa designated in these tests could be considered convergent across the skull. The 3.0-3.3 TDs defined ornamented taxa as those with bony ornament-like structures primarily formed from the premaxillary bone, and superficially, there is a lot of visual similarity among many of these taxa based on the placement of the ornamental structure (enlargement of the bone is largely toward

the dorsoposterior portion of the premaxilla) as well as a reduced narial fenestra in many taxa, especially in Test 3.3.

Convergence among the TD 3.3 taxa is supported by the hierarchical cluster analysis grouping, which places these taxa close together relative to most other taxa, including related unornamented taxa. Convergence for the TD 3.2 taxa grouping is less well-supported by the cluster analysis, though convergence among the TD 3.2 taxa is still supported.

It is worth noting that the TD 3.2 and TD 3.3 taxa both have only a small number of taxa in their ornamented taxa groupings (n = 4 and n=6, respectively). While it is not unreasonable to conduct convergence tests with small numbers of taxa, even as low as two, provided they are not sister taxa within the dataset, smaller groupings of focal taxa may be more subject to bias from these tests. Where possible, multiple specimens from each species in the focal and non-focal groupings can provide a more complete picture and mitigate the sample size biases. This was not done in these tests due to practical limitations, but could be conducted in the future and would clarify whether the apparently strong results from our *search.conv* tests hold up to scrutiny.

3.5: Wheatsheaf Index

Table 6. Wheatsheaf index values and associated p-values for each of the different ornamentation definition tests. As with the θ-values from the search.conv analyses, the results here are given for both the full sampled skull and each of the individual cranial bones. Blue boxes indicate test-region subsamples with low p-values (p<0.05). P-values listed as 0 are not actually 0 but simply low enough that they exceed the reporting limit for R 4.1.2 and are rounded down. They may be accepted as p<<0.05.

Wheatsheaf Full Bill Region

		Test 1.0	Test 1.1	Test 2.0	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	w	0.9301	0.9158	0.9171	0.9680	1.0015	0.9672	1.2260	1.1012	1.0484
	w p-value	1	1	1	1	1	1	1	1	1
Nasal Bone	w	1.0363	1.0328	1.0123	0.9911	0.9891	1.1688	1.4477	1.4791	1.3367
	w p-value	1	0	1	0	1	0	0	1	0
Frontal Bone	w	0.9713	0.9548	0.9302	0.9597	1.0109	0.9855	1.0628	1.1483	1.0038
	w p-value	1	1	1	1	1	1	1	1	1
Premaxillary Bone	w	0.8763	0.8569	0.8526	0.9439	0.9906	0.8923	1.255701	0.9076	0.8976
	w p-value	1	1	1	1	1	1	0.83	1	1

As with the search.conv analyses, many tests yielded no evidence for convergence among the ornamented taxa, though some test-sample combinations did produce results with very low p-values ($p << 0.05$), suggesting the ornamented taxa group together more closely than expected (Table 6). Curiously, the nasal bone showed the only indication of convergence among any the test definitions. All elements taken together did not signal convergence among the ornamented taxa for any of the Wheatsheaf index analyses, which is a good indication that this test is easily swayed by non-convergent regions of the skull and breaking areas of interest up into modules, such as individual bones, would be useful for future studies.

The nasal bone occupies a small overall portion of the physical area of the skull and is highly variable in morphology across bird taxa. In our study, the nasal also had a smaller number of landmarks compared to the other regions (n = 289), so the apparent convergence in the nasal bone for the Wheatsheaf index analyses should be accepted with caution. Similarly, many of the tests that showed positive indications for convergence in the Wheatsheaf index values (TDs 1.1,

2.1, 3.0, and 3.1) showed little or no evidence of convergence in either the C-values or search.conv analyses, meaning these indicators of "convergence" may in fact be retention of homologous characters.

However, support for some of the test-sample combinations comes from agreement between the Wheatsheaf and search.conv analyses. In both tests, the nasal bone landmarks were convergent among the ornamented taxa for TDs 3.1 and 3.3.

It is curious that the nasal should show as convergent for the TD 3.1 and 3.3 taxa whereas the premaxilla does not show consensus between the two tests. The total strength of the convergence signal is stronger for the nasal than the premaxilla in both tests, which might partly explain the difference. It is possible that the convergence in the nasal bone landmarks is greater because of the element's generally smaller size compared to the premaxilla, or its more central location on the skull where it tends to form internal supports to lateral structures rather than directly contributing to lateral structures itself. If landmarks that are distributed on the fringes of the bony ornament-like structure are less similar between taxa, but the shape of the combined premaxilla and nasal bones are generally similar between those same taxa, then the more central region of the premaxilla and most of the nasal bone would be expected to have stronger convergence values in the analyses than more lateral regions such as the bill edge and culmen.

For both the search.conv and Wheatsheaf analyses, many of the definitions produced different results, and almost none of the traditionally accepted groupings for skull ornamentation as a whole indicated any convergence. Clear definitions of these structures and the decision to group these structures based on specific morphological characteristics are important elements for assessing whether these structures are convergent or not.

CHAPTER 4

4: Results Section 2

The Section 2 analyses concern Question 2 of this study:

How does the exclusion of the skull region where bony ornament-like structures occur affect the convergence estimates between taxa?

If convergence among the ornamented taxa determined in Results Section 1 is related to the presence of the bony ornament-like structure, one would expect that convergence tests for the region containing the ornament-like structure would be greater than convergence tests for the rest of the skull among the designated ornamented taxa.

The elements selected for landmarking in this study (nasal, frontal, and premaxillary bones) primarily make up the anterior and dorsal portions of the skull, where the majority of bony ornament-like structures are found in birds. For each skull bone and test definition compared between the ornamental and non-ornamental skull regions, there are four possible outcomes:

Outcome 1: Neither the ornamental nor non-ornamental regions is convergent.

Outcome 2: There is convergence in the non-ornamental region, but either lower or no convergence in the ornamental region.

Outcome 3: There is convergence in the ornamental region, but either lower or no convergence in the non-ornamental region.

Outcome 4: There is convergence in both regions, with a negligible difference between the two regions.

Outcome 1 is the expected outcome if the methods in Results Section 1 recovered no convergence for the test definition ornamented taxa and the lack of convergence is unaffected by

any particular contribution from the ornamental region. Based on the results from the previous section which showed convergence in at least some of the groupings, we would expect at last test definition 3.2 and 3.3 to show convergence the ornamental and/or non-ornamental regions. Outcome 2 would suggest that the ornament-like structure negatively contributes to overall skull convergence among ornamented taxa. Outcome 3 would fit the expectation that the convergence among the ornamented taxa is due to shared morphological similarity in their ornament-like structures. Outcome 4 would suggest that even though the ornamented taxa are convergent, either the ornament-like structure does not contribute to this or the means of differentiating the ornamental from the non-ornamental regions is inadequate to detect the regional contributions to convergence.

4.1: *search.conv* Time-Adjusted θ

Table 7. The top table shows the search.conv *θ values for the "non-ornamental" region, and the bottom table shows the same for the ornamental region defined in the Materials and Methods. Orange values in the top table are those with p<0.5. Blue values in the bottom table are those where p<0.05 and the θ -value was greater (indicating greater convergence) compared to the non-ornamental region analyses. Orange values in the bottom table indicate the θ-value was either the same or lower in the ornamental region analysis for p<0.05 (for light orange, p<0.1).*

search.conv Non-Ornamental Region

		Test 1.0	Test 1.1	Test 2.0	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	**θ**	0.7245	0.7135	0.6856	0.6691	0.6691	0.7774	0.6857	0.5193	0.4591
	θ p-value	0.905	0.886	0.622	0.358	0.378	0.966	0.539	0.111	**0.014**
Nasal Bone	**θ**	0.9111	0.8303	0.8756	0.6783	0.6783	0.7894	0.2935	0.2338	0.2304
	θ p-value	0.997	0.994	0.998	0.366	0.367	0.945	0.945	**0.007**	**0.001**
Frontal Bone	**θ**	0.7429	0.7457	0.7410	0.6789	0.6789	0.8020	0.6708	0.3489	0.3963
	θ p-value	0.956	0.974	0.944	0.604	0.588	0.973	0.485	**0.011**	**0.002**
Premaxillary Bone	**θ**	0.7250	0.7188	0.6763	0.6772	0.6772	0.8189	0.8437	0.7804	0.6519
	θ p-value	0.908	0.897	0.503	0.572	0.544	0.984	0.823	0.809	0.385

search.conv Ornamental Region

		Test 1.0	Test 1.1	Test 2.0	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	**θ**	0.7693	0.7503	0.7435	0.6701	0.6701	0.7999	0.6285	0.3048	0.2879
	θ p-value	0.992	0.988	0.959	0.631	0.615	0.981	0.391	**0.008**	**0.001**
Nasal Bone	**θ**	0.8621	0.8068	0.8418	0.6629	0.6629	0.7692	0.2137	0.3281	0.2796
	θ p-value	1	0.999	1	0.114	0.11	0.933	**0.017**	**0.027**	**0.002**
Frontal Bone	**θ**	0.7491	0.7568	0.7389	0.6829	0.6829	0.8312	0.7597	0.1934	0.2638
	θ p-value	0.978	0.985	0.954	0.807	0.806	0.986	0.68	**0.002**	**0.001**
Premaxillary Bone	**θ**	0.7369	0.7369	0.7049	0.6806	0.6806	0.8473	0.7580	0.4335	0.3385
	θ p-value	0.939	0.947	0.787	0.702	0.716	0.989	0.699	**0.043**	**0.004**

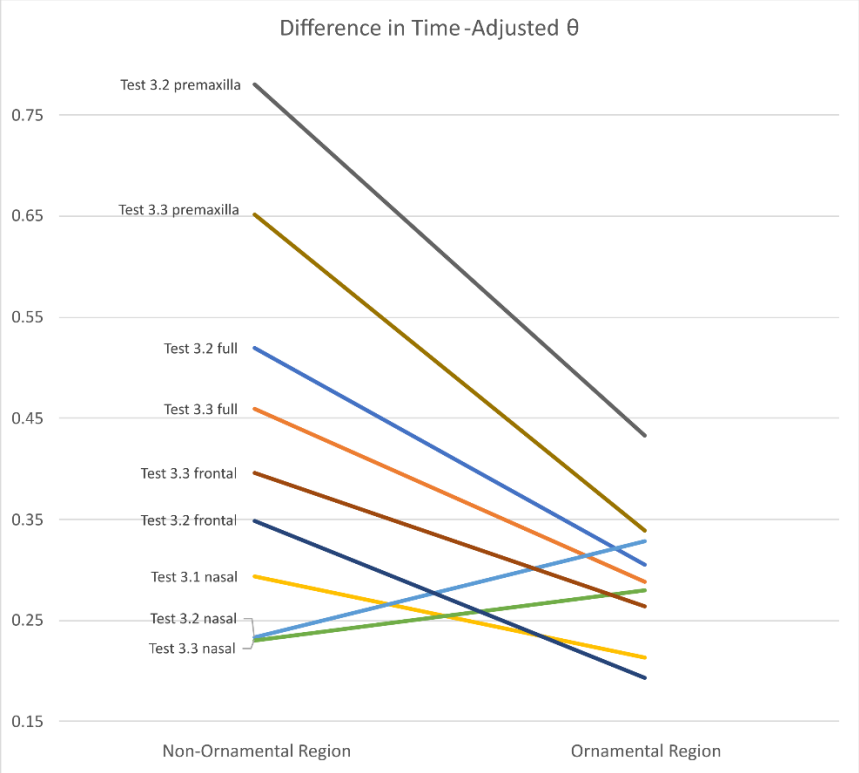

Figure 21. Comparison of non-ornamental to ornamental region search.conv *θ-values for subtests where at least one θ-value had an associated p-value of <0.05. Lower θ-values indicate more convergence among the ornamented taxa.*

Table 7 and Figure 21 show the results of the *search.conv* regional tests. Only subsamples from TDs 3.1, 3.2, and 3.3 showed any significant convergence according to this method.

Five subsamples (test 3.2 nasal, 3.2 frontal, 3.3 full, 3.3 nasal, and 3.3 frontal) had p-values lower than 0.01 when the non-ornamental region was included, indicating convergence is occurring even outside of the ornamental region. The test 3.2 frontal, 3.3 full, and 3.3 frontal subsamples all showed lower θ-values in the ornamental region than the non-ornamental region, suggesting that even though these subsamples were convergent according to the search.conv definition in both the ornamental and non-ornamental regions, the ornamented taxa are more convergent in the ornamental region for those three subsamples. This trend did not hold true for all subsamples showing convergence in the non-ornamental region; test 3.2 nasal and 3.3 nasal

had lower θ-values for the non-ornamental region, suggesting the nasal bone is less convergent among ornamented taxa in the ornamental region even when the other elements appear to be more convergent.

There were four additional subsamples (test 3.2 full, 3.1 frontal, 3.2 premaxilla, and 3.3 premaxilla) that had high p-values in the non-ornamental region but low p-values in the ornamental region, indicating these ornamental region samples were distinct from the θ-values randomized across the sample taxa. These four subsamples also had substantially lower associated θ-values than their equivalents in the non-ornamental region (Figure 21). One would expect the premaxillary bone to be more convergent among the 3.2 and 3.3 test definitions, as these tests defined ornamented taxa on the basis of the presence of premaxilla bony ornament-like structures specifically.

Overall, the search.conv results generally showed greater evidence of convergence in the ornamental than the non-ornamental region (Outcome 3) for the 3.2 and 3.3 TDs, though the nasal subsamples showed greater evidence for convergence in the non-ornamental region (Outcome 2) instead. Test definition 3.1 showed a slightly greater convergence in the ornamental region compared to the non-ornamental, concentrated in the nasal. All other TDs recovered no evidence of convergence (Outcome 1), which indicates no convergence and no unique contribution to convergence from the region where the ornament-like structure occurs.

4.2: Wheatsheaf Index

Table 8. The top table shows the Wheatsheaf index values for the non-ornamental region, and the bottom table shows the same for the ornamental region defined in the Materials and Methods. Orange values in the top table are those with p<0.5. Blue values in the bottom table are those where p<0.05 (for light blue, p<0.1) and the w-value was greater (indicating greater convergence) compared to the non-ornamental region analyses. Orange values in the bottom table indicate the w-value was either the same or lower in the ornamental region analysis for p<0.05. P-values listed as 0 are not actually 0 but simply low enough that they exceed the reporting limit for R 4.1.2 and are rounded down. They may be accepted as p<<0.05.

Wheatsheaf Non-Ornamental Region

		Test 1.0	Test 1.1	Test 2.0	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	**w**	0.9437	0.8955	0.9464	0.9393	0.9783	0.9044	1.0789	1.0416	1.0100
	w p-value	1	1	1	1	1	1	1	1	1
Nasal Bone	**w**	1.0680	1.0212	0.9993	0.9869	1.0150	1.0865	1.2283	1.2280	1.1503
	w p-value	**0**	**0**	1	0.54	0.98	**0**	0.49	1	**0**
Frontal Bone	**w**	1.0750	1.0683	1.0440	1.0156	1.0574	1.0642	1.2452	1.2498	1.0978
	w p-value	**0**	**0**	**0**	**0**	**0**	**0.01**	0.49	0.68	0.77
Premaxillary Bone	**w**	0.9301	0.8883	0.9434	0.9383	0.9853	0.9107	1.1188	1.0364	0.9850
	w p-value	1	1	1	1	1	1	0.95	1	1

Wheatsheaf Ornamental Region

		Test 1.0	Test 1.1	Test 2.0	Test 2.1	Test 2.2	Test 3.0	Test 3.1	Test 3.2	Test 3.3
Full Skull	**w**	0.9547	0.9412	0.9302	0.9693	1.0015	1.0045	1.2616	1.1011	1.0375
	w p-value	1	1	1	1	1	1	0.47	1	1
Nasal Bone	**w**	1.0352	1.0379	1.0233	0.9906	0.9853	1.1903	1.4758	1.5463	1.3718
	w p-value	0.93	**0**	0.98	**0**	1	**0**	**0**	0.63	**0**
Frontal Bone	**w**	0.9659	0.9496	0.9250	0.9522	1.0030	0.9857	1.0747	1.1599	1.0156
	w p-value	1	1	1	1	1	1	1	1	1
Premaxillary Bone	**w**	0.9038	0.8858	0.8950	0.9510	0.9837	0.9290	1.2760	0.9869	0.9859
	w p-value	1	1	1	1	1	1	**0.06**	1	1

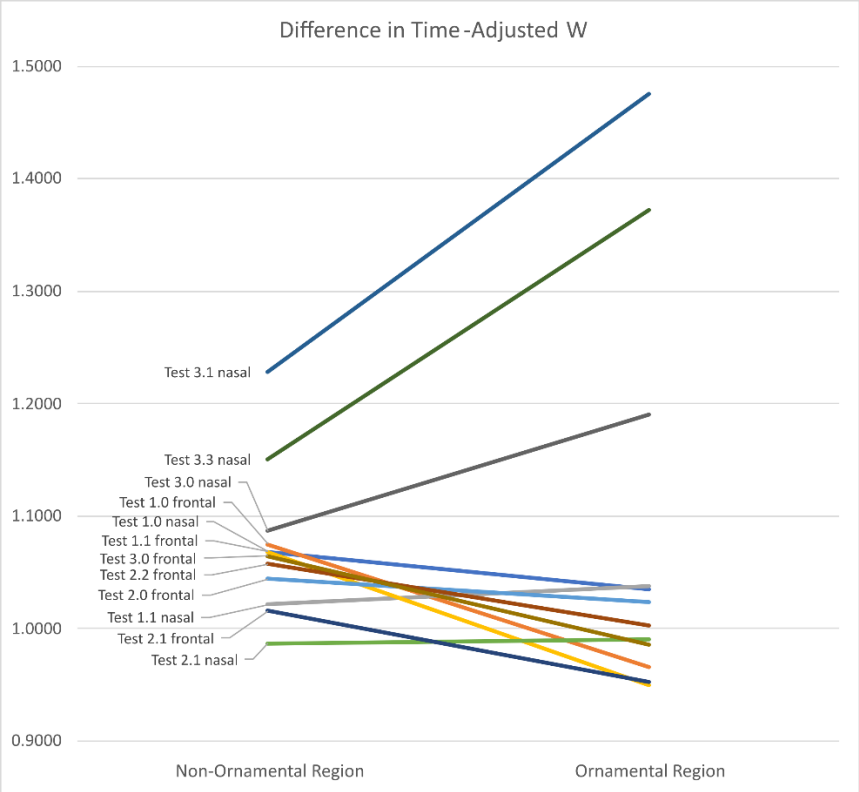

Figure 22. Comparison of non-ornamental to ornamental region values for subtests where at least one w-value (Wheatsheaf index value) had an associated p-value of <0.05. This chart is read differently than Figure 21; higher w-values indicate more convergence among the ornamented taxa.

As with the full sample region reported above, the Wheatsheaf index results showed little consensus with the *search.conv* method for the regional analyses.

Of the data samples that had p-values lower than 0.01 when the ornamental region was included, five test-sample combinations showed increased convergence metrics (higher w-values with correspondingly low p-values) for the ornamental region relative to the non-ornamental region. These were test 1.1 frontal, 2.1 nasal, 3.0 nasal, 5.1 nasal, and 5.3 nasal. Several frontal bone subsamples (TDs 1.0-3.0) and a few nasal bone subsamples (tests 2.1 and 3.1) showed the opposite results, with the non-ornamental region appearing to be more convergent (Table 7, Figure 22).

As mentioned previously, the Wheatsheaf index is not sensitive to homology, therefore TDs 1.0-2.1 should likely be disregarded for the Wheatsheaf index tests. Test definitions 2.2, 3.0, and 3.1 also show little evidence for convergence, though 3.1 does appear to show ornament-related convergence in the nasal according to the *search.conv* definition for convergence. Test definition 3.1 does appear to indicate increased evidence for convergence in the ornamental compared to the non-ornamental region according to the Wheatsheaf index (Outcome 3). The results for TDs 2.2 and 3.0 are mixed.

None of the Wheatsheaf index tests found convergence stronger than a random sample within the dataset for TD 3.2 (Outcome 1). There was similarly little evidence for substantial convergence for TD 3.3 across the skull according to the Wheatsheaf index, though the ornamental region of the nasal bone does appear to be more convergent than the non-ornamental region among the ornamented taxa (Outcome 3).

CHAPTER 5

5: Results Section 3

The Section 3 analyses concern Question 3 of this study:

> **Question 3: What, if any, ecological conditions correlate with the presence and form of bony ornament-like structures?**

According to Results Section 1, there does appear to be a statistically significant and potentially biologically meaningful difference between the ornamented and unornamented groups for at least some of the TDs. Some of this appears to be attributable to the bill ornament structure. Whether these structures are sexually-selected remains unclear, and this study is not able to determine the selection affecting their morphology. However, the available data can begin to clarify the relationship between the taxa with these structures and common factors known to influence bill shape such as body mass, diet, and habitat (e.g., Darwin 1872, Greenberg et al. 2012, Luther & Greenberg 2013, Tatersall et al. 2017, Chávez-Hoffmeister 2020, Pigot et al. 2020, Crouch & Tobias 2022, Freeman et al. 2022, Tobias 2022).

Correlations, or lack thereof, among ecological factors can help rule out alternative explanations to sexual selection. If there are correlations between traits such as diet, latitude, range, and mass (a physiological trait with important ecological consequences) among taxa with bony ornament-like structures, and these principally ecological traits differ between the ornamented and unornamented taxa, then differences in ecology might explain the presence of these structures rather than sexual selection.

Correlation does not clarify causation; it is possible that bony ornament-like structures selected primarily through sexual interactions may have impacts on an organism's ecology. For instance, if these structures tend to be co-opted for thermoregulation, then covariation between

bill morphology, size, and latitude could be a consequence of interactions between sexual selection and non-sexual selection, with sexual selection for ornament-like structures reflecting honest signals about thermoregulatory capability.

Nonetheless, correlation remains a useful means of assessing how likely it is that the evolution of these structures is driven by sexual selection alone. Similar ecological distributions for ornamented and unornamented taxa would strengthen the hypothesis that bony ornament-like structures are best explained by non-ecological means, such as sexual selection, whereas differences in the ecologies of ornamented taxa compared to unornamented taxa would weaken the hypothesis that sexual selection is the primary factor driving the evolution of these structures.

The traits of interest for this portion of the study are average mass, centroid latitude, species range size, primary lifestyle, habitat, trophic level, and trophic niche. Mass is a physiological trait that influences many ecological characters, including feeding strategy, metabolism, and lifestyle, but the rest of the selected traits of interest pertain directly to the ecological interactions between the organism and its environment. Species averages were used because species traits like mass, latitude, and geographical range size were not available for all of the individual specimens for which shape data was collected. Latitude was adjusted by absolute value, and mass and range were log-corrected to make more meaningful comparisons between taxa at the extremes. Range is complicated somewhat by the inclusion of critically endangered species like *Mitu mitu*, which is extinct in the wild. Rare and endangered birds threatened by habitat loss tend to have much smaller range sizes than they did historically, so these range values are likely to be an underestimate for birds dependent upon dwindling habitats.

5.1: Visualizations of Ecological Data

Figures 33–39 illustrate the mass and ecological data associated with PCs 1 and 2. Mass and range show little relationship to the PC axes, though latitude does seem to vary slightly across PC1, with generally more high-latitude species found toward the left of the graph, though there is overlap at both extremes. Primary lifestyle sees more swimming, aerial, and insessorial birds toward the left side of the plot, and more generalists in the top half, with terrestrial birds throughout. Habitat is fairly well-mixed across the plot, but with more birds in human-modified environments found in the upper portion of the plot. For trophic level, herbivores seem to be restricted to the upper left-hand side of the plot, with some spread lower and further right.

Trophic niche, like habitat, appears to be largely mixed across the upper portion of the plot, but the lower half of the plot is unique in that almost all taxa below -0.15 on the PC2 axis are aquatic predators, either diving or wading birds. The only exception is *Archilochus*, a hummingbird, and the only nectarivore in our dataset. This may be because long, thin bills (common in birds in the lower portion of the graph) often evolve in piscivorous species such as herons due to the spear-fishing advantages of the form, and the nectivorous lifestyle of some hummingbirds happens to overlap in this morphology based on the flowers with which they co-evolve.

PC1 appears to be largely linked with overall bill volume. Large bill volumes tend to be more common in tropical species, as they can provide increased thermal conductivity. Large bill volumes also tend to interfere with aerodynamic efficiency, so are more common in terrestrial, wading, and climbing birds than in gliding, diving, and perching birds.

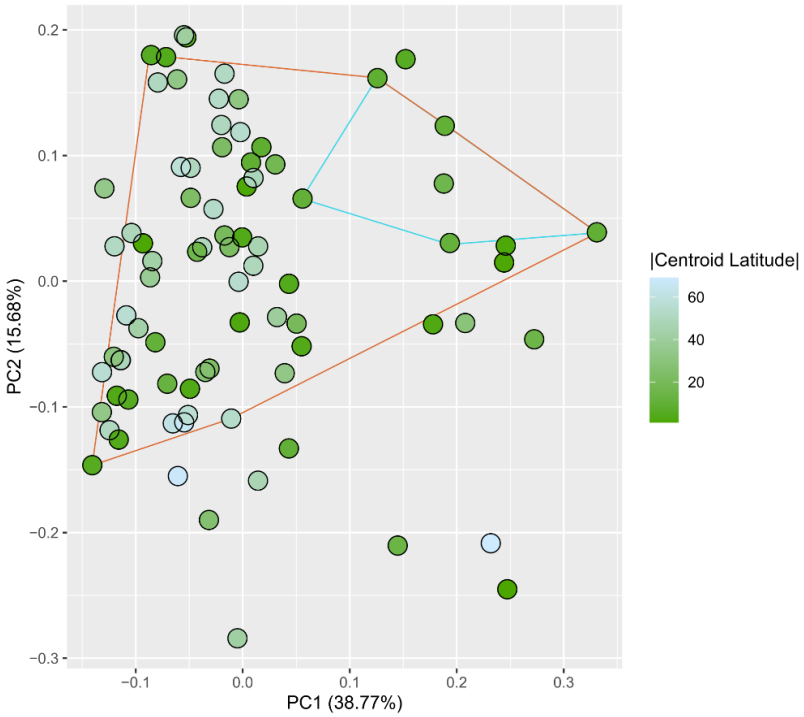

Figure 23. Bill surface landmark PCA with points colored by the absolute value of centroid latitude for species range. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

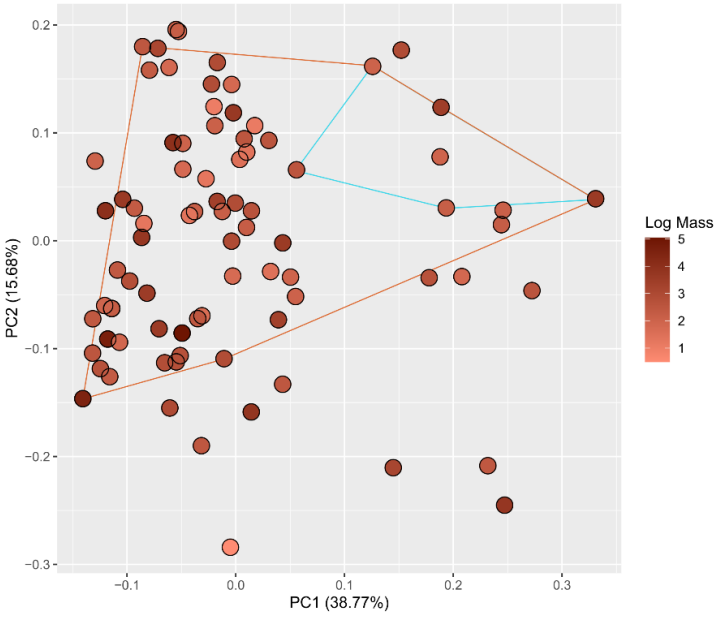

Figure 24. Bill surface landmark PCA with points colored by the log of species average mass. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

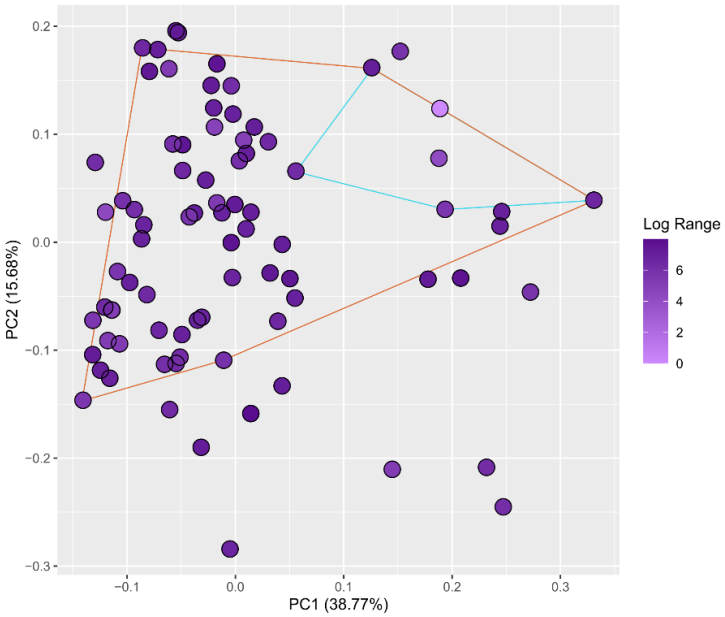

Figure 25. Bill surface landmark PCA with points colored by the log of geographic species range size. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

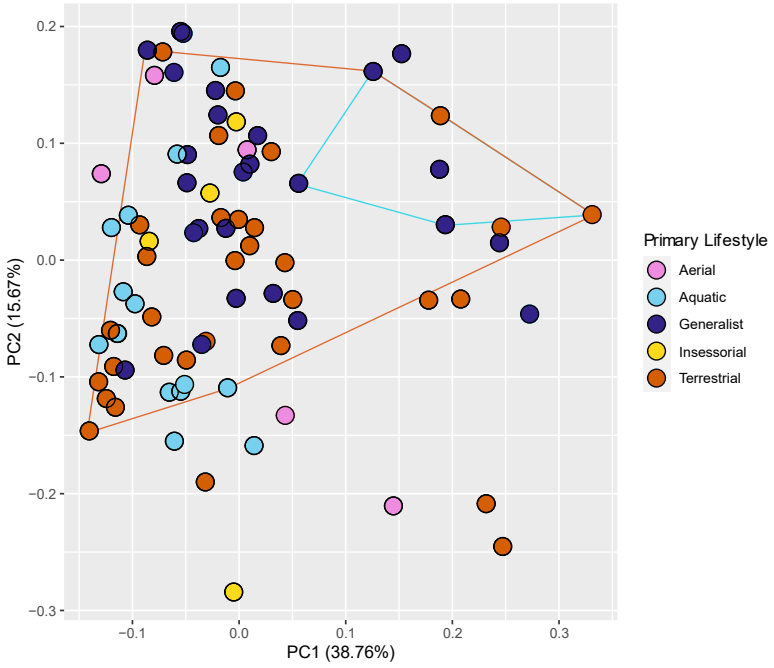

Figure 26. Bill surface landmark PCA with points colored by the primary lifestyle of the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

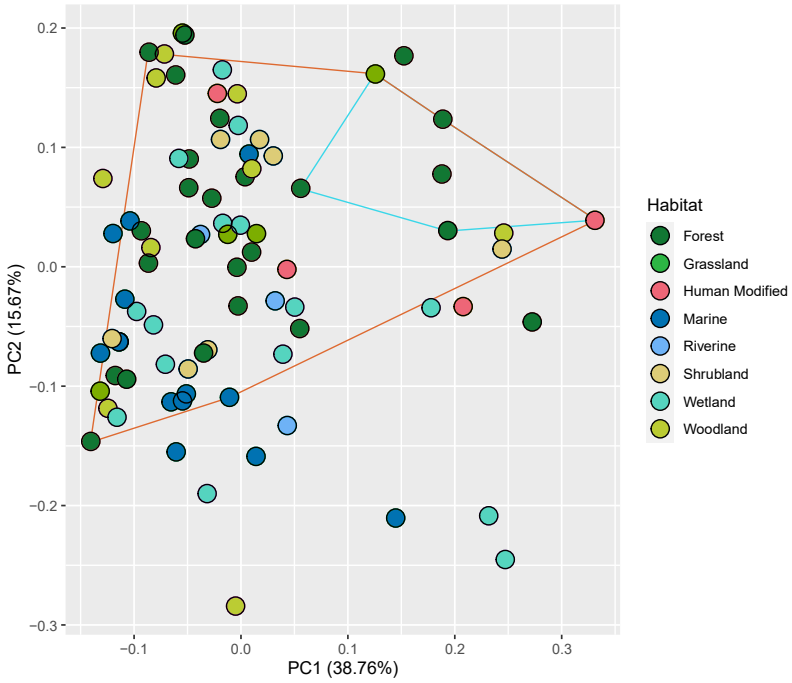

Figure 27. Bill surface landmark PCA with points colored by the typical habitat of the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

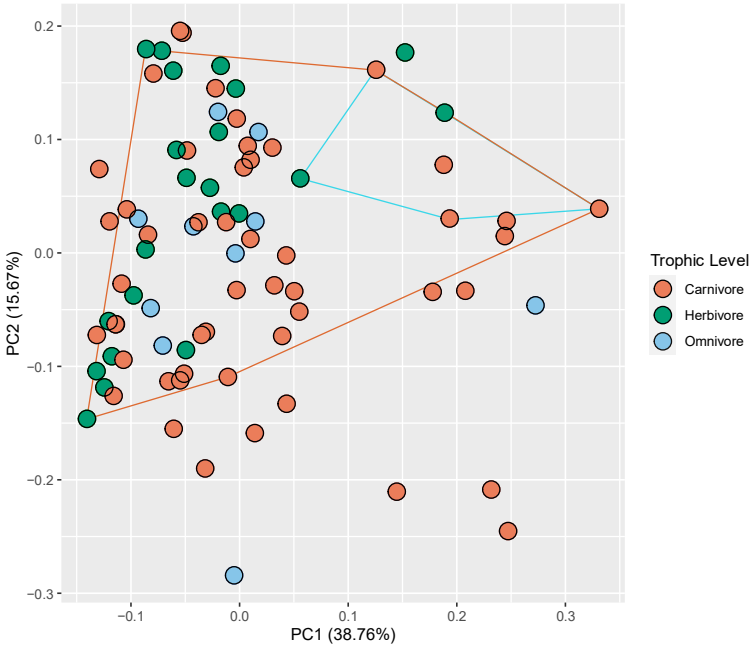

Figure 28. Bill surface landmark PCA with points colored by the general trophic level for the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

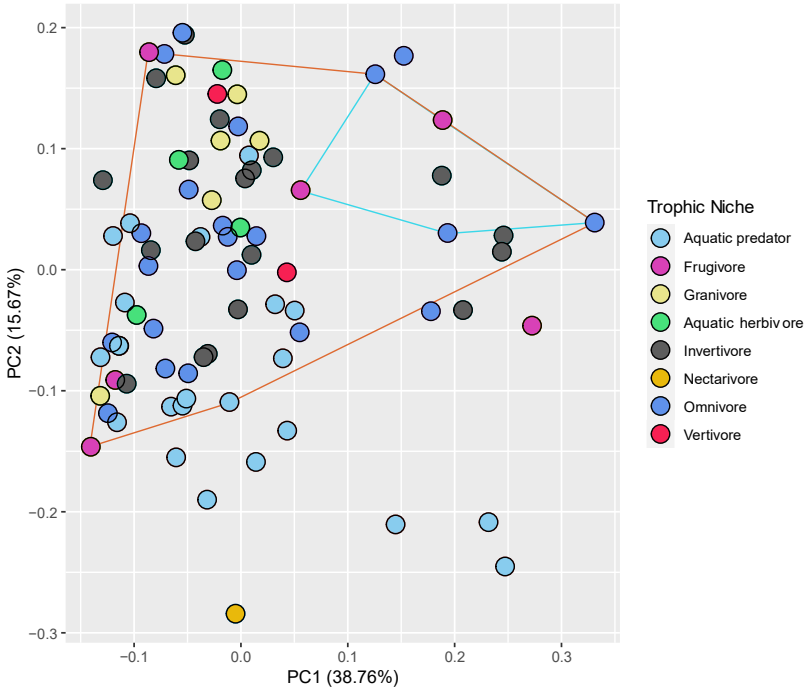

Figure 29. Bill surface landmark PCA with points colored by the general trophic niche for the species. The orange convex hull surrounds ornamented taxa defined by test definition 1.1, and the blue convex hull surrounds ornamented taxa defined by test definition 3.3.

In examining the PCAs, I also wished to point out that there is an interesting association between a basic approximation of bill volume and the average mass of the bird species that plots in association to morphology along PC1 (Figure 30). When the logarithm of the ration of bill volume to mass is taken, a gradient forms where birds with a lower bill volume to mass ratio plot on the left side of PC1 while birds with proportionally larger bills compared to their mass plot on the right with a gradient between. Bill volume is thought to be related to thermoregulation, though, curiously, two of the groups for which this trends has been recorded plot on either extreme of PC1 for our data, indicating that relative body mass may be less important of a control on the presence of enlarged bony structures than other physiological characteristics.

Note that mass and bill volume for these calculations were made using species-wide averages reported in AVONET (Tobias et al. 2022), rather than data for these particular

individuals. The reason is that mass was not recorded for most individuals in the dataset, as most were museum specimens and many were skeletonized or catalogued without mass data.

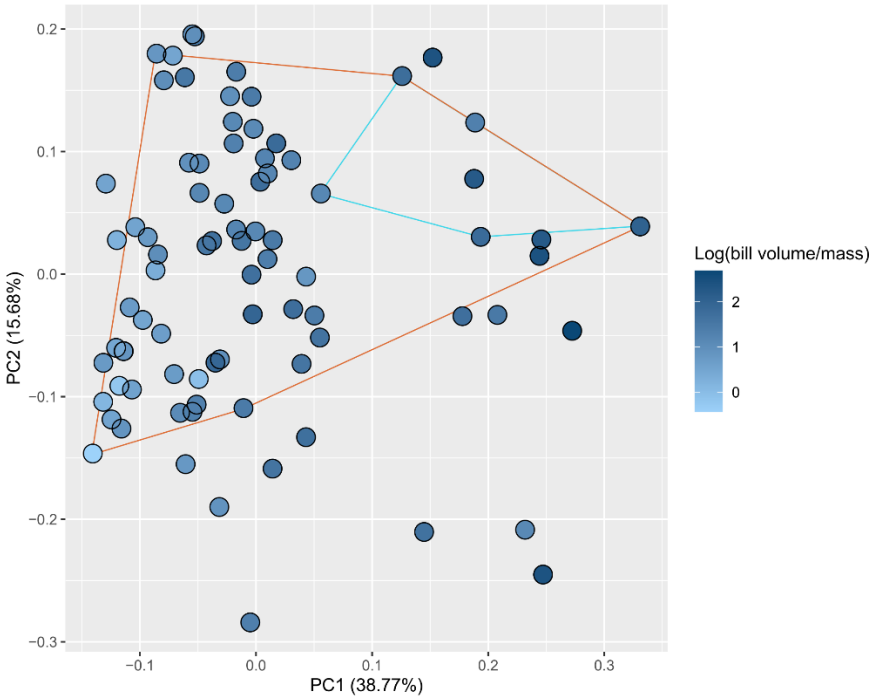

Figure 30. Display of PC1 against PC2, with taxa colored by the log of the ratio of bill volume to mass. The blue convex hull denotes the TD 3.3 taxa while the orange hull denotes the TD 1.1 taxa. In general, the values increase across PC1, with birds on the right of the chart having larger bills relative to their body and birds on the left having smaller bills. There doesn't appear to be any association for this metric along PC2. Curiously, while several ornamented taxa occupy the upper right portion of the chart, the cassowaries, noted for having very large casques, are among the furthest left taxa and have a very low bill to body mass ratio. This is likely because although their casques are large, as flightless birds, their bodies are also much heavier than other birds. It is possible that the relationship between bony ornament-like structures and their cooling function could differ for flighted birds compared to flightless birds, as flight is a major source of body heat production.

5.2: Allometry

Many species exhibit allometric trends in the size of various body structures, and this trends is frequently observed in ornamentation and weapons (Kodric-Brown et al. 2006, Eberhard et al. 2018, O'Brien et al. 2018). Larger individuals within a species often bear larger ornaments and weapons, and similarly, larger species tend to be able to support proportionally

larger weapons and ornaments than smaller species. I wanted to determine whether allometric trends could be contributing to the morphometric trends, so I compared the weighted average of the bird bill PC scores to the species average mass as an approximation of size, and approximate species average bill volume as an estimate of bill size. Species vary in mass and bill volumes, especially between sexes, but species average mass and volume were used in place of individual specimen mass and volume because these data were unavailable for most specimens in this study.

Figure 31 shows the results of allometric comparisons. Mass, curiously, does not show a significant relationship to PC score across the sample set, but there is potentially a strong association between mass and PC score for the taxa most consistently defined as having prominent bony ornament-like structures. There is a moderate association between bill volume and PC score, with the prominently ornamented taxa showing a slightly different slope than the set of all taxa, likely because of skew from the two cassowaries. There are strong associations between the log of bill volume over mass compared against PC score, with similar trends for ornamented taxa and the full dataset. Mass generally tracks positively with volume, as might be expected; larger birds are capable of supporting heavier bills and may need more robust bills for processing food, but if bill volume is also associated with thermal regulation, larger birds may also require larger bills in general to cool down. Previous studies (Bright et al. 2016) have found associations between the size of a bird species and the shape of its bill.

Based on the Figure 31 plots, it appears that there is a positive allometric relationship between mass and volume, but not between mass and PC score, although volume and PC score and the ratio of volume to mass and PC score do show a relationship. It seems likely that these relationships differ for the ornamented taxa compared to the overall group, so I suspect allometry

may contribute to interspecific variation between these taxa, though to what extent is presently unclear.

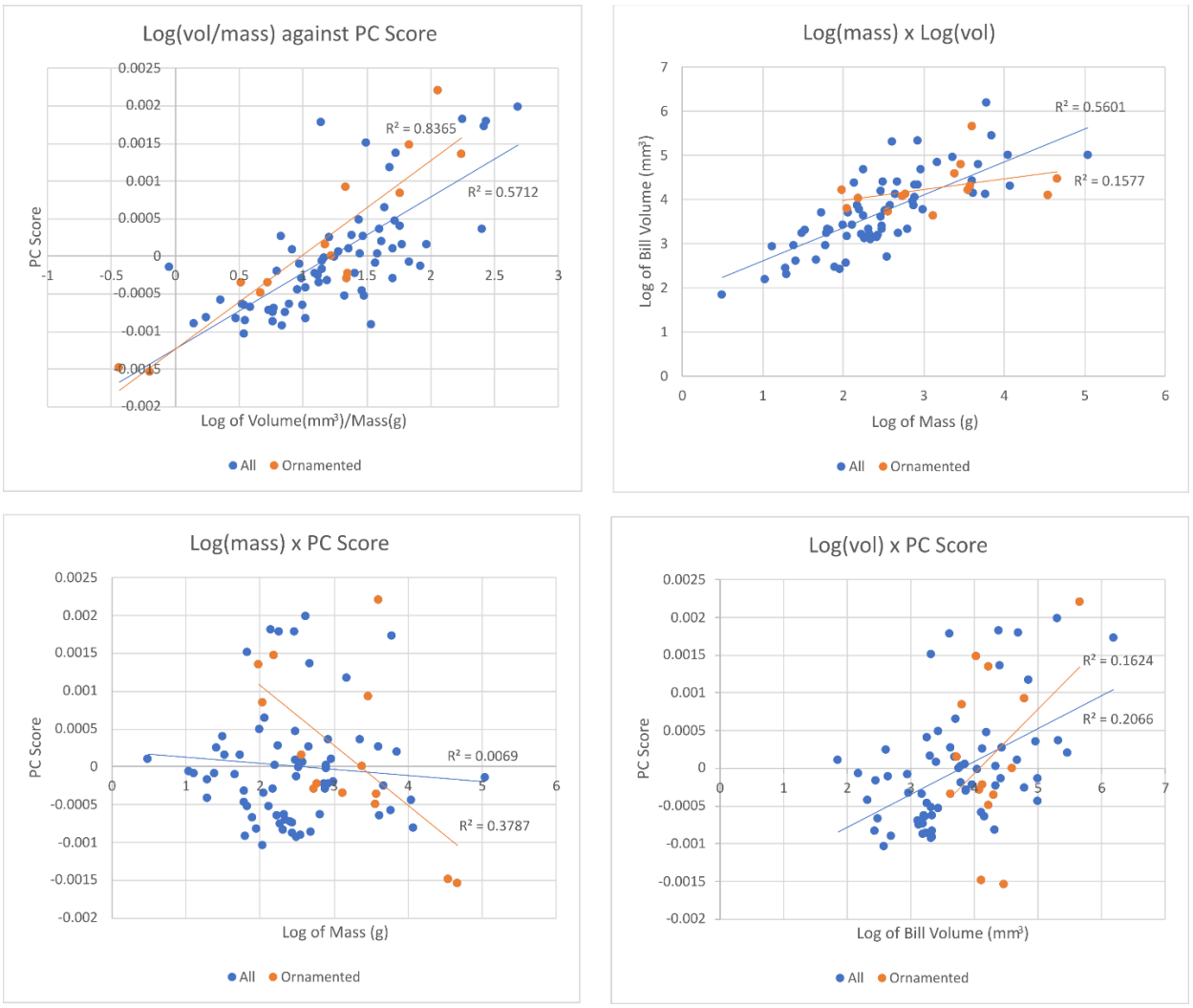

Figure 31. Plots comparing different measures of mass and volume against each other and PC score values. Ornamented taxa show different trends from the full dataset except for the log volume over mass compared to PC score. Spread is very high for the log mass and fairly high for the ornamented mass by volume, and the ornamented and full dataset volume by PC score trends. However, there is a tighter association between log mass by log volume for the full dataset and the log volume over mass compared to PC score for the ornamented taxa, as well as a very strong trend with very low spread for the ornamented taxa log volume over mass compared to the PC score. Taxa highlighted as "ornamented" are those listed in the Conclusions section for this study, as those are the consensus heavily ornamented taxa.

5.3: Phylogenetic ANCOVA

Table 9 details the output of various phylogenetic ANCOVA for the individual quantitative variables of weighted average PC score of bill landmarks, log range size, log mass, and the absolute value of the centroid latitude of the bird species in the dataset according to the database AVONET (Tobias et al. 2022). Slopes and intercepts of regression lines were compared between different combinations of ecological and summary morphological traits to assess differences between ornamented and unornamented groupings. Low slope p-values indicate dissimilarity in the relationship trends of selected traits between the two groups, while low intercept p-values indicate dissimilarity at low values for the trait designated as the independent variable within the ANCOVA (treated here as the "starting point" for reference between the different tests).

Table 9. pANCOVA for weighted average PC score for bill surface landmarks compared to the absolute value of the centroid latitude of species range, weighted average PC compared to the log of species range size, weighted average PC compared to the log of mass, log of mass compared to absolute value of centroid latitude, log of mass compared to log of average range size, and absolute value of centroid latitude compared to log of species range size. P-values reported as 0 are rounded down by R, and should be taken as p<<0.05.

Test 1.0	Weighted Avg. PC by Latitude	Weighted Avg. PC by Range Size	Weighted Avg. PC by Mass	Mass by Latitude	Mass by Range Size	Latitude by Range Size
F-value slope	5.6952	8.758	4.2436	10.8928	6.1354	0.3306
p-value slope	**0.0194**	**0.0041**	**0.0427**	**0.0015**	**0.0154**	0.5669
F-value intercept	11.9499	9.1744	9.2736	8.0388	6.5335	0.8379
p-value intercept	**9e-04**	**0.0033**	**0.0032**	**0.0058**	**0.0125**	0.3628

Table 9 (continued).

Test 1.1						
F-value slope	7.6574	5.7908	3.3655	9.8851	5.7483	1.0509
p-value slope	**0.007**	**0.0184**	**0.0703**	**0.0023**	**0.0189**	0.3084
F-value intercept	10.5045	7.322	8.4115	7.3554	5.7334	2.3944
p-value intercept	**0.0017**	**0.0083**	**0.0048**	**0.0082**	**0.019**	0.1258
Test 2.0						
F-value slope	6.5385	5.7304	4.1216	3.0653	1.5103	1.0884
p-value slope	**0.0125**	**0.019**	**0.0457**	0.0839	0.2227	0.3
F-value intercept	7.9733	6.0714	7.2474	1.896	1.2927	1.9234
p-value intercept	**0.006**	**0.0159**	**0.0087**	0.1724	0.259	0.1694
Test 2.1						
F-value slope	5.6362	6.4285	4.1392	9.9738	6.4408	3.9318
p-value slope	**0.02**	**0.0132**	**0.0453**	**0.0022**	**0.0131**	0.**0509**
F-value intercept	7.5379	6.5958	6.318	7.4627	7.1047	5.2731
p-value intercept	**0.0075**	**0.0121**	**0.014**	**0.0078**	**0.0093**	**0.0243**
Test 2.2						
F-value slope	1.4873	2.7986	1.125	7.3551	7.1687	3.7732
p-value slope	0.2263	0.0983	0.2921	**0.0082**	**0.009**	0.0556
F-value intercept	3.1395	2.8812	2.3671	7.5208	7.5385	5.7548
p-value intercept	0.0803	0.0936	0.1279	**0.0075**	**0.0075**	**0.0188**

Table 9 (continued).

Test 3.0						
F-value slope	20.4118	6.7715	8.9981	2.5089	1.0262	2.882
p-value slope	**0**	**0.0111**	**0.0036**	0.1172	0.3141	0.0935
F-value intercept	10.8156	8.453	10.7984	1.2295	0.7258	5.2707
p-value intercept	**0.0015**	**0.0047**	**0.0015**	0.2709	0.3968	**0.0243**
Test 3.1						
F-value slope	1.8959	0.0444	0.017	0.0939	0.2734	0.8259
p-value slope	0.1724	0.8337	0.8965	0.7601	0.6025	0.3662
F-value intercept	0.0078	0.0147	0.0016	0.1325	0.298	0.807
p-value intercept	0.9298	0.9037	0.9679	0.7168	0.5867	0.3717
Test 3.2						
F-value slope	24.8277	19.8095	25.2616	3.8561	1.2159	1.2489
p-value slope	**0**	**0**	**0**	0.0531	0.2735	0.2672
F-value intercept	29.1012	24.2382	27.6748	1.9509	0.7317	3.822
p-value intercept	**0**	**0**	**0**	0.1664	0.3949	0.0541
Test 3.3						
F-value slope	30.271	30.0604	35.3548	2.9045	1.9428	2.2854
p-value slope	**0**	**0**	**0**	0.0923	0.1673	0.1346
F-value intercept	40.2953	35.7323	38.7961	2.542	1.4523	4.9698
p-value intercept	**0**	**0**	**0**	0.1148	0.2318	0.0286

Test definitions 1.0 and 1.1 showed differences in both the slope and intercept between the ornamented and unornamented taxa for all of the pANCOVA except for the one that measured latitude against range size. This means that the relationship between latitude and range is indistinguishable between the different groups, but the influence of latitude, range size, and mass on the generalized morphology of the taxa, as well as the influence of latitude and range size on mass, differs considerably for the two groups.

For TD 2.0, there were differences in the slope and intercept for the comparison of generalized morphology against latitude, range size, and mass, but no difference in the relationships between mass and latitude, mass and range size, or latitude and range size. This suggests that the bill morphology has a relationship to mass and certain ecological traits that differs between the ornamented and unornamented groups, but that the relationships among those other traits are indistinguishable between the two groups.

In test definition 2.1, the intercept differed between the ornamented and unornamented taxa for all of the comparisons and the slope differed for all of the comparisons except for latitude compared to range size. This means that the ornamented and unornamented taxa in test definition 2.1 appear to follow different evolutionary trends from one another with the exception of how their range size varies with latitude. The overall trend in range size covariance with latitude is functionally the same for the two groups, but for taxa at low range sizes, the latitude of their centroid latitude is statistically different.

Test definition 2.2 shows similar relationships between the ornamented and unornamented taxa for the mass covariance with range, mass covariance with latitude, and latitude covariance with range size, but shows no difference in the relationship between generalized morphology and mass, range size, or latitude. These results suggest that mass, range

size, and latitude have the same predictive effect on generalized morphology between the ornamented and unornamented groups.

Test definition 3.0 showed a difference between the morphology by latitude, morphology by mass, morphology by range, and latitude by range comparisons for each of the groups. Note that the p-value of the slope for the generalized morphology compared to latitude is especially low compared to most of the other p-values for this dataset, indicating a very high certainty that the trend for the two groups differs for this comparison.

Test definition 3.1 showed no difference between the ornamented and unornamented taxa for any of the trait comparisons.

Tests definitions 3.2 and 3.3 showed similar results. The slopes and intercepts of generalized morphology with latitude, range size, and mass differed between the ornamented and unornamented groups for both Test definitions 3.2 and 3.3. The p-values for all of the comparisons that differed were extremely low, indicating high certainty that these values differ significantly between the two groups.

Please note that for test definitions 3.0-3.3, the number of ornamented taxa is less than 10. For test definition 3.1, N=3, so although these datasets can produce a trend line, the accuracy of that trendline in representing all possible taxa that could fit those ornamentation definitions is likely fairly low.

Although certain values produced in this pANCOVA analysis are below the traditional 0.05 α level threshold for p-values, I suggest caution in fully rejecting the null hypothesis that there is no difference in the groups. The values showing low p-values (intercept of weighted average PC scores against range size, slope of log mass against range size, and intercept of log mass against range size), all of these p-values are still on the same order of magnitude for the

0.05 cutoff. The 0.05 cutoff itself was defined by Ronald Fischer to approximate about two standard deviations on a normal curve, or a 1 in 20 chance of procuring a type-1 error, which may be sufficiently low for certain studies, but should not be applied universally or taken as clear evidence against the null hypothesis on their own (Wasserstein & Lazar 2016). For this study's data, the 0.05 cutoff appears to be sufficient, but tests with values nearer this cutoff may be less biologically meaningful than those with much lower p-values.

5.4: Phylogenetic MANCOVA

Table 10 provides the output for the phylogenetic MANCOVA tests. The independent variable for the pMANCOVA tests was the weighted average PC score, measured against the following dependent variables: the log of mass, the absolute value of latitude, the log of range size, the interaction between log mass and absolute latitude, the interaction between log mass and log range size, and the interaction between absolute latitude and log range size.

The intercept for a pMANCOVA is established in relation to a common reference point between the two groups. Because the dependent variables are multi-dimensional, the intercept does not pass through an axis per se, but rather assesses what the independent variable's predicted value is at that reference point according to the trend line established for the dataset. Similarity in the intercept but differences in the trend would likely be interpreted in much the same way as differences in both intercept and trend between two groups, as the specific reference point used to calculate the intercept may not necessarily be biologically meaningful, depending on the data in question. However, differences in the intercept that are coupled with similarities in the slope for a pMANCOVA indicate that the two groups are likely subject to similar selection, but at differing magnitudes.

Table 10. pMANCOVA for weighted average PC score for bill surface landmarks compared to the log of average species mass, absolute value of the centroid latitude of species range, log of species range size, and interaction factors between mass and latitude, mass and range size, and latitude and range size. P-values reported as 0 are rounded down by R, and should be taken as p<<0.05.

Test 1.0	Weighted Avg. PC by Mass + Latitude + Range Size + Interaction Factors
F-value slope	4.0009
p-value slope	**0.0491**
F-value intercept	7.2924
p-value intercept	**0.0086**
Test 1.1	
F-value slope	1.6714
p-value slope	0.2001
F-value intercept	3.7567
p-value intercept	0.0564
Test 2.0	
F-value slope	2.181
p-value slope	0.144
F-value intercept	3.5626
p-value intercept	0.063
Test 2.1	
F-value slope	2.6854
p-value slope	0.1055
F-value intercept	3.6726
p-value intercept	0.0592
Test 2.2	
F-value slope	0.2753
p-value slope	0.6014
F-value intercept	0.5755
p-value intercept	0.4505
Test 3.0	
F-value slope	3.3068
p-value slope	0.073
F-value intercept	4.0264
p-value intercept	**0.0484**

Table 10 (continued).

Test 3.1	
F-value slope	0.0036
p-value slope	0.9521
F-value intercept	0.0105
p-value intercept	0.9188
Test 3.2	
F-value slope	14.4792
p-value slope	**3e-04**
F-value intercept	15.5612
p-value intercept	**2e-04**
Test 3.3	
F-value slope	22.5876
p-value slope	**0**
F-value intercept	24.2811
p-value intercept	**0**

When all variables were summed in a pMANCOVA along with their interaction factors, most of the different test definitions showed no statistically substantial difference between the ornamented and unornamented groups. However, there were a few exceptions.

For TD 3.0, the slopes were the functionally the same, but the intercepts differed. In an pMANCOVA, this would indicate that the association between the correlated traits is the same for both groups, but the magnitude of that association differs between them.

Tests 1.0, 3.2, and 3.3 differed in their pMANCOVA intercepts and slopes, indicating the collective influence of range, mass, and latitude on weighted average PC sore differed between the ornamented and unornamented groups for each of these definitions. The associated p-values were especially low for Test 3.3, indicating a low chance of producing a type one error and accordingly a high likelihood that the groups are statistically different in their evolutionary trends.

5.5: Pearson's Chi-Square Test

Table 11 shows the results of the chi-square tests comparing the distribution of taxa falling into different habitat, trophic niche, trophic level, and lifestyle qualitative categories between the ornamented and unornamented groups.

Table 11. Chi-square results for habitat, trophic niche, trophic level, and primary lifestyle for the ornamented versus unornamented taxa defined by the different test definitions.

Test 1.0	Habitat	Trophic Niche	Trophic Level	Primary Lifestyle
χ-square	3.5241	19.385	4.0369	4.4754
Df	7	7	2	4
p-value	0.8327	**0.007063**	0.1329	0.3455
Test 1.1				
χ-square	4.2975	22.329	4.6533	3.0657
Df	7	7	2	4
p-value	0.7449	**0.002229**	0.09762	0.5469
Test 2.0				
χ-square	3.3457	18.065	1.8229	3.1545
df	7	7	2	4
p-value	0.8513	**0.01168**	0.4019	0.5323
Test 2.1				
χ-square	7.2012	12.629	0.35143	2.0313
df	7	7	2	4
p-value	0.4082	0.08169	0.8389	0.73
Test 2.2				
χ-square	7.149	14.941	3.8917	7.0243
df	7	7	2	4
p-value	0.4135	**0.03676**	0.1429	0.1346
Test 3.0				
χ-square	7.6544	11.947	1.446	3.7397
df	7	7	2	4
p-value	0.3641	0.1023	0.4853	0.4424

Table 11 (continued).

Test 3.1				
χ-square	10.788	7.0187	0.44539	2.1375
Df	7	7	2	4
p-value	0.1481	0.4269	0.8004	0.7105
Test 3.2				
χ-square	9.5877	12.323	1.4374	3.9284
Df	7	7	2	4
p-value	0.2132	0.09043	0.4874	0.4158
Test 3.3				
χ-square	9.599	10.452	0.92593	4.4437
df	7	7	2	4
p-value	0.2125	0.1644	0.6294	0.3493

Test definitions 1.0 and 1.1 differed in their trophic niche between the ornamented and unornamented groups.

Test definition 2.1 showed no difference between the two groups for any of the tested traits, but the test definition 2.0 and 2.2 definitions showed differences in trophic niche between the ornamented and unornamented groups.

There was no difference between the two groups for test definitions 3.0-3.3.

Note again that the number of ornamented taxa in the test definitions 3.0-3.3 is less than 10, meaning the taxa within the subsample may not be a good representation of the total trait variability for all possible taxa categorized as ornamented by these subtest definitions. Test definition 3.0 is more likely to be influenced by phylogenetic effects because it contains two closely-related species of *Fratercula* as well as two musophagids. It is therefore more likely that the chi-squared tests would recover differences between the ornamented and unornamented groups for TD 3.0.

CHAPTER 6

6: Conclusions and Discussion

6.1: Conclusions

6.1.1: Question 1 Conclusions

The ornamented taxa categorized in TDs 1.0, 1.1, 2.0, 2.1, 3.0, and 3.1 showed distinct C-value distributions compared to the unornamented taxa, but these ornamented taxa had median C-values at or near 0 when compared with one another, suggesting little to no convergence among the ornament-like structures for those test definitions. Test definition 2.2 showed higher C-values, but the pairings among ornamented taxa were similar to those between ornamented and unornamented taxa, suggesting the convergence values were likely not due to the ornament-like structures themselves and may merely be a consequence of the large number of taxa designated as ornamented by this definition. Test definitions 3.2 and 3.3 had ornamented taxa that were distinct from the unornamented taxa based on much lower C-values for comparisons between ornamented and unornamented taxa than for either the ornamented or unornamented groups alone. The median C-values for the ornamented group were slightly less in test definitions 3.2 and 3.3 than those for the unornamented group, so even though these TDs provide evidence for convergence among ornamented taxa according to the Stayton (2015) definition, the amount of convergence among the ornamented taxa is not unique among birds in the sample.

The *search.conv* results seem to broadly agree with the convergence assessment of the C-values. The full skull tests found no evidence for convergence according to the Castiglione et al (2019) convergence definition for TDs 1.0-3.0. There was some evidence for convergence in TD 3.2, but this was localized to the nasal bone and did not appear when all elements were sampled together. Test definitions 3.2 and 3.3, on the other hand, showed very strong evidence for convergence across all or most skull elements.

The Wheatsheaf index showed evidence for convergence only within the nasal bone under TDs 1.1, 2.1, 3.0, 3.1, and 3.3. Based on the C-value and search.conv results, the Wheatsheaf positive results for TDs 1.1 and 3.0 should probably be disregarded, because the Wheatsheaf index is prone to false positives and groupings used to calculate the index must be checked against methods that detect for the presence or absence of convergence. The Wheatsheaf index suggests that TD 2.1 may be reflecting actual convergence, at least in the nasal bone. Test definition 3.1 is congruent between the Wheatsheaf value and the *search.conv* results, so despite the evidence for convergence across the whole skull appearing weak overall, there may be strong convergence in the nasal bone for this definition's ornamented taxa. The Wheatsheaf index did not find that the convergence among the TD 3.2 taxa was exceptionally strong, but it provides support for TD 3.3 having strong convergence among the ornamented taxa in the nasal bone.

Most of the ornamented taxa for the TDs do not cluster tightly on the PCA or the hierarchical cluster analysis, but the 3.2 and 3.3 taxa do show some clustering on both and are not closely related to one another. Combined with the quantitative results from the C-values, search.conv, and Wheatsheaf, it appears that most ornamented taxa are not necessarily convergent with one another in their bill shape, but there is evidence that the 3.2 TD ornamented taxa are convergent, and stronger evidence for convergence in the 3.3 TD ornamented taxa. These two test definitions share four of the same taxa.

In response to Question 1 of this study, the similarity among most definitions for birds with bony ornament-like structures is lower than expected by their genetic relationships, implying that they may be more divergent than most birds rather than more convergent. However, birds with wide bony ornament-like structures on the premaxilla appear to be an exception, as they appear to be convergent.

6.1.2: Question 2 Conclusions

The search.conv regional tests found that for most parts of the skull, there was either greater convergence or increased evidence for convergence in the ornamental region compared to the non-ornamental region. This trend was not consistent across all regions, as the nasal bones for TDs 3.2 and 3.3 were less convergent in the ornamental region than the non-ornamental region, contradicting the Wheatsheaf results for the full skull.

The Wheatsheaf index regional tests showed complex relationships between the frontal and nasal bones for several of the tests. However, because of the limitations of the Wheatsheaf methodology and absence of evidence for convergence in the C-values and search.conv tests for TDs 1.0-2.1 and 3.0, I am inclined to regard most of the Wheatsheaf regional test results as a consequence of factors other than convergence. The meaningful Wheatsheaf results were for TDs 2.1, 3.0, and 3.3. Test definition 2.1 and 3.0 showed greater convergence for the nasal in the ornamental region but lower convergence for the frontal in the ornamental region, suggesting, as in the *search.conv* test, that convergence values differ regionally in different parts of the bill. The 3.3 TD showed convergence in the nasal for both the ornamental and non-ornamental regions, but the proportional value of the Wheatsheaf index was greater in the ornamental region.

In response to Question 2, it appears that the inclusion or exclusion of the ornamental region where bony and other ornament-like structures occur can significantly affect the convergence results for the sampled area. This trend is also not always straightforward; sometimes certain bone elements may be more convergent for the ornamental region than the non-ornamental region among the designated taxa, so not all of the convergence contribution is due to the ornamental region. For TDs 3.2 and 3.3, however, the ornamental region appears to be

the main reason for this contribution, with the non-ornamental region indicating little convergence among the TD 3.2 and 3.3 ornamented taxa.

6.1.3: Question 3 Conclusions

The pANCOVA tests found differences between the ornamented and unornamented taxa for most of the test definitions. This is surprising given the Results Section 1 conclusions about non-convergence in most TDs 1.0-3.0, but differences between the ornamented and unornamented taxa could indicate that although these taxa are not necessarily convergent with one another, they may still differ as a group in other ways from the unornamented taxa. Test definitions 1.0, 1.1, 2.1, 2.2, and 3.0 had differences in covariation between various ecological traits, and of these test definitions, all but TD 2.2 found associations between the birds' ecological traits and bill morphology. Test definitions 2.0, 3.2, and 3.3 had correlations between bill morphology and ecology, but their ecological values did not differ noticeably from the rest of the sample. The pMANCOVA tests only found differences between the sum trajectories of the ornamented taxa compared to the unornamented taxa for TDs 1.0, 3.2, and 3.3.

The chi-square tests found differences between the ornamented and unornamented groups in the distribution of trophic niche among the taxa. Ornamented taxa for TD 1.0, 1.1, 2.0, and 2.2 all differed in their trophic niche relative to the unornamented taxa. The chi-square tends to give false negatives at low sample sizes, so similarity between ornamented and unornamented taxa is not conclusive, especially for the 3.0-3.3 TDs. However, the differences between ornamented and unornamented taxa for the four test definitions that did show differences appears to reflect the expectation from the PCA, which shows a disproportionate number of piscivores among the unornamented taxa relative to the ornamented taxa.

Collectively, these results indicate that even for the test definitions that did not indicate convergence, there does appear to be a difference in ecological traits among the ornamented taxa that might explain why they developed ornament-like structures. Test definitions 3.2 and 3.3 showed very strong correlations between bill morphology and ecological traits for both the pANCOVA and pMANCOVA. It is possible that small sample size among ornamented taxa may be affecting these results. However, given that there is a difference between the TD 3.1 results and those of 3.2 and 3.3, and also given than the pMANCOVA p-values for TD 3.3 are much lower than those for 3.2, despite 3.3 having slightly more taxa, I believe the pANCOVA and pMANCOVA patterns for these tests are reasonably accurate. If true, these results suggest that the strong convergence among the bills for the test 3.2 and 3.3 definitions is likely tied in some way to ecology, and likely other ecological traits simultaneously.

In response to Question 3, there are ecological traits that correlate with the presence of bony ornament-like structures, and it appears that these ecological traits include latitude, species range size, mass, and trophic niche. This means that sexual selection alone is not likely the sole reason for why bony ornament-like structures develop, though sexual selection should not be discounted as an additional, potentially important, factor in the evolution of the surface shape of these structures.

6.2: Discussion

6.2.1: On Test Definitions for Bony Ornament-Like Structures

Test definition 1.0 represents a well-established definition for bony ornament-like structures based on the work of Livezey (1996), Webster (2003), and Mayr (2018). This definition is moderately conservative in the types of taxa it includes, potentially limiting its

scope, and is similarly limited by the small number of studies that compare casques and similar structures across different bird orders. Nonetheless, TD 1.0 offers a good baseline for which taxa are generally considered to possess bony ornament-like structures.

Test definitions 1.0 and 1.1 differ by the inclusion or exclusion of two specific taxa: *Euryceros prevostii* and *Tauraco persa*. *Euryceros* was excluded in Test 1.0 because of a comment by Mayr (2018) that the enlargement of its bill might be influenced by structural properties, and *Tauraco persa* was included because the Musophagidae are considered by Mayr (2018) to all have bony enlargements. I chose to include *Euryceros* on the basis that bright blue color on the bill extending into the bony ornament-like structure could be an indicator that the bill is involved in sexual selection, as pigments are well-known sexual signals in other bird bills; I also chose to exclude *T. persa* for TD 1.1 because this species was not specifically noted as having a bony ornament-like structure in any texts, even though its family is often characterized by these structures. Given that these two taxa altered the results between TDs 1.0 and 1.1, they and their bill structure merit close attention in future studies.

Test definition 2.0, which was modified from definition 1.0, did not show much difference from test 1.0 in its results across the board, except for a lack of correlations among the ecological traits for the ornamented taxa relative to the unornamented in the pANCOVA.

Test definition 2.1 used a broader definition for bony ornament-like structure and showed slight differences from TDs 1.0-2.0 in the results, but these differences were not substantial. This definition groups structures like casques alongside small bony enlargements, some of which are known to have other purposes, such as the structural reinforcement in *Oryzoborus* or the tongue attachment for *Picus viridis*. Future studies would do well to clarify what types of enlarged structures actually serve an ornamental purpose in life and which are exclusively functional, if

there is a distinction. The 2.1 TD characterizes a large number of taxa as ornamented, and accordingly, those taxa occupy a large amount of morphospace. Because this definition is so broad, it is likely of limited value in studies of bony ornament-like structures directly, but it could be useful for establishing which taxa are consistently defined as unornamented. It is noteworthy that this group did show slight convergence for the C-values, even though it was very low, and the TD 2.1 taxa were distinct within the pANCOVA tests from the unornamented taxa. There may still be some signal unifying these taxa, albeit loosely, based on their bill morphology.

Definition 2.2 is similarly broad like Test 2.1, but is unique in that it includes some taxa with small to prominent soft-tissue structures that typically do not show enlargements on the underlying bone. Part of the purpose of this definition was to search for associations between taxa with bony ornament-like structures and soft-tissue ornament-like structures that may not be visually apparent. The Test 2.2 definition is not meant to be a good definition for bony ornament-like structures, but is rather a test to see if taxa with any sort of ornament-like structure would show convergence with one another even if only the skull is included. The TD 2.2 ornamented taxa did have distinctly non-zero median C-values when compared to one another, even if that signal was not necessarily unique within the dataset. These taxa also grouped differently from the unornamented taxa in the pANCOVA tests ecologically, so whereas their similarities may not be related to their bill morphology, the results suggest that there may be different ecological factors associated with the presence or absence of any type of ornament-like structure, whether it is made out of soft tissue or out of bone. Among the taxa classified as unornamented for Test 2.2, there is a disproportionate number of small aerial specialists, aquatic predators, and birds with especially small or long, thin bills. For these taxa, feather ornaments may be advantageous over

fleshy and bony ornaments, as some of the unornamented taxa still display elaborate feather ornaments or color-based displays. Several species of hummingbird and nightjar, taxa for which large three-dimensional structures might be expected to severely hinder flight efficiency, are known to have extreme feather ornaments on their wings or tails (for example, *Caprimulgus longipennis* and *Loddigesia mirabilis*). It could be that feeding strategy specifically restricts the shape of the bill and front of the head in these smaller taxa, and although our dataset only has one nectarivore, we would expect this to be especially true for nectar-feeding birds that have co-evolved bills to fit inside specific flowers.

Test definition 3.0, based on the presence of premaxillary ornamentation, showed similar results to TDs 1.0-2.0 and as such, the 3.0 ornamented taxa do not appear to be convergent. They are, however, ecologically distinct from the unornamented taxa, including in how their bills covary

Test definition 3.1 suffers from a small sample size, and it is likely its ecological test results are inconclusive because of this. Ornamented taxa from this test definition did appear slightly convergent in the nasal bone specifically, and that convergence does appear to be related to the bony ornament-like structure, or some part of the ornamental region. I suspect that because two of the TD 3.1 taxa overlap with those of TD 3.3, morphological similarities between those two taxa are the source for the weak convergence signal in this definition, with the third taxon (*Tauraco persa*) not contributing substantially to the convergence signal.

Test definitions 3.2 and 3.3 were the ones with greatest support for convergence according to the analytical methods used. Test 3.3 is a slightly broader definition that encompasses TD 3.2 and two additional taxa (*Bucorvus abyssinicus* and *Crotophaga ani*). The inclusion of these taxa strengthened the convergence results for the ornamented taxa in the TD

3.3 analyses, so it appears that TD 3.3 provides the more complete morphological definition for convergent bony ornament-like structures in this dataset.

Species included in convergence tests alter the results considerably, so choice of ornamented taxa is important and should be carefully considered in relation to the question you wish to study.

Taxa with prominent bony ornament-like structures seem to be the most consistently used in the literature, and I would agree that these are easy to identify as ornamented. I would remove some of the ambiguous taxa, mainly most *Cygnus* species aside from *C. olor* (not included in this dataset), *Tauraco persa* and other turacos with smaller bony crests, oropendolas with smaller casques (not included in this dataset), and *Fratercula cirrhata*, all of which have distinct textural differences and small swellings, but don't fit quite as well into the same category as animals like cassowaries and hornbills in terms of their bony ornament-like structures.

I would also include *Euryceros* as having a prominent bony ornament-like structure. The curvature of the bill does likely impact its structural properties, but that doesn't exclude the role of these structures as ornamentation on its own. Further studies should be done to differentiate between the display and survival functions of these ornamental structures, if there is a way to distinguish them (I suspect doing so might be difficult). Also, according to my TD 1.1 results, this species and *Tauraco persa* can sway the results of the tests on their own, and based on the prominence of their ornamentation, I am inclined to put *Euryceros* in the ornamented categorization and *T. persa* in an ambiguously-ornamented or semi-ornamented category that requires more study. Birds with a similar bill morphology to *Euryceros*, specifically *Corvus crassirostris*, should be assessed for whether they better fit with the taxa that have prominent bony ornament-like structures or ambiguous ones.

Taxa that have no type of ornament-like structure, ambiguous or otherwise (see Supplementary Table 2) seem to be somewhat restricted in their ecology, but I think these could be a useful comparison to the prominently ornamented taxa in some studies. Comparisons that cut out taxa with ambiguous ornament-like structures or compare taxa with ambiguous ornament-like structures to those with prominent ornament-like structures, could offer insight into the constraints on these structures as well as their early evolutionary morphologies, given that ambiguous ornament-like structures are presumed to be a necessary intermediary step for the ontogenetic development of prominent ornament-like structures.

Based on the combined test results and visual prominence, I would offer the following list for which extant bird taxa from this study's dataset should be described as having prominent bony ornament-like structures:

Anseranas semipalmata

Balearica pavonina

Balearica regulorum

Bucorvus abyssinicus

Casuarius bennetti

Casuarius unappendiculatus

Crotophaga ani

Euryceros prevostii

Fratercula arctica

Fratercula corniculata

Mitu mitu

Musophaga violacea

Numida meleagris

Psarocolius wagleri

Four of the aforementioned taxa have bony ornament-like structures on the frontal in one of at least two styles (wide and thin):

Anseranas semipalmata

Balearica pavonina

Balearica regulorum

Numida meleagris

Eight of the listed taxa have bony ornament-like structures on the premaxilla in one of two styles (wide or thin):

Bucorvus abyssinicus

Crotophaga ani

Euryceros prevostii

Fratercula arctica

Fratercula corniculata

Mitu mitu

Musophaga violacea

Psarocolius wagleri

Numida may have a separate form of bony ornament-like structure from *Anseranas*, given that the former's ornamentation is tall and recurved, whereas *Anseranas* has a simpler crest of bone.

The two *Casuarius* species have bony ornament-like structures on the nasal that appear to be a single common style (wide), presumably inherited from a common ancestor given the close evolutionary history of those species and the presence of bony ornament-like structures in all extant *Casuarius*.

The taxa with wide bony ornament-like structures on the premaxilla showed unique evolutionary patterns within the dataset, marked by higher C-values for within-group comparisons than comparisons between ornamented and unornamented taxa, as well as low angle differences for the *search.conv* method. Other subcategories of ornamentation were not tested due to sample size, so at present, it is not clear whether the signature of the wide premaxillary bony ornament-like structures is unique among birds with or if other forms of bony ornament-like structures would show similar patterns.

6.2.2: Notes on Analyses

6.2.2.1: pANCOVA, pMANCOVA, and Pearson's Chi-Square Tests

The pANCOVA and pMANCOVA used in this study are not meant to provide a clear cause for convergence or parallelism among ornamented taxa within the dataset. These methods offer a simple assessment of how and whether the ornamented and unornamented groups differ in the correlations between their traits, and as a simple method, the pANCOVA/pMANCOVA results are necessarily limited. Further study would be required to assess the rigor of the results these methods provide, particularly given the small sample size for some of the test definitions.

However, the initial assessment by the pANCOVA and pMANCOVA can point to the likelihood of bony ornament-like structures having evolved exclusively in response to sexual interaction. If sexual selection along affects the evolution of bony ornament-like structures, one would expect these structures to have little to no effect on the non-sexual ecology of the organism. Because of the interactions between sexual- and non-sexual selection, the ecology of ornamented taxa is likely not tied exclusively to non-sexual selection, even if the bony ornament-like structures evolved directly in response to sexual selection. The pANCOVA/pMANCOVA tests are not designed to clarify the nuanced influences that sexual and non-sexual selection can have on ecology. However, non-sexual selection remains the primary contributor to an organism's ecology in most species, so the use of ecological factors as a proxy for non-sexual selection remains useful. This reasoning is also applied to the Pearson's chi-square tests.

None of the methods used in Results Section 3 are meant to rule out sexual selection as a contributor to the evolution of bony ornament-like structures. However, dissimilarity in the ecological traits of the ornamented taxa compared to the unornamented taxa points to the possibility that non-sexual selection may contribute to these structures. More detailed study of the selection influencing bony ornament-like structures in populations of living birds from ornamented species would be necessary to clarify what factors directly influence the evolution of these structures.

6.2.2.2: C-Values

Figure 32. Pairwise comparison table of C1 values for all 82 taxa compared with one another. Darker green colors represent higher C-values, lighter values represent lower values. Values are expected to be low when morphology is dissimilar or phylogenetic distance is low (hence a white diagonal of values representing 0 when taxa are compared with themselves).

Figure 32 shows a pairwise comparison table of C1 (standardized convergence) values for all of the sampled taxa. *Haliaeetus leucocephalus* (bald eagle), *Luscinia megarhynchos* (rufous nightingale), and *Bucorvus abyssinicus* (Abyssinian ground hornbill), are indicated as examples. *Haliaeetus* and *Luscinia* both have very median bill morphologies compared to other birds within the dataset, hence why their rows and columns form nearly solid green lines.

Bucorvus has a fairly unusual morphology within the dataset, its bill being exceptionally large in volume with a slight curvature and distortion of the nasal compared with many other taxa.

Of note, there are a few taxa with which *Bucorvus* does show higher C-values (0.2+), notably *Euryceros*, *Probosciger*, *Tockus erythrorhynchus*, and *Mitu*. The latter taxon and *Euryceros* are notably classified as ornamented under many of our test definitions, along with *Bucorvus*. *Tockus* (another genus of hornbill) and *Probosciger* (a cockatoo) are generally classified as unornamented except in a few tests, as they have small bony protrusions on their skills with slightly enlarged soft tissue coverings (and feather attachments in the case of *Probosciger*) that may serve similar functions to ornament-like structures on birds with more obvious enlargements. The non-0 C-values for *Bucorvus* include several taxa often classified as ornamented, as well as taxa that are not necessarily ornamented but have especially long or large bills. The *Bucorvus* C-values generally agree with the grouping identified in the cluster analysis that placed *Bucorvus* morphologically together with other hornbills, several taxa with well-defined bony ornament-like structures on the premaxilla, and some long-billed bird taxa.

There are several instances where some species or groups of closely-related species show unusually low or high values for all comparisons to other taxa. Low C-values across the board, such as for *Bucorvus abyssinicus*, likely represent taxa that are so morphologically distinct that they have little physical similarity to the other taxa within the dataset. However, taxa that are convergent with almost every other taxon should be viewed with scrutiny because they likely represent a bias with this type of calculation. C-values are based on PCA distances, which are a useful summary of morphology, but also cluster taxa in multi-dimensional space in a way that requires certain taxa be more central to the cluster than others. *Haliaeetus leucocephalus*, a typical accipitrine, and *Luscinia megarhynchos*, a typical passerine, have a similar morphology

to related taxa, but they also happen to have very median skull shapes within the dataset, meaning the Euclidean distance between where these samples plot along the PCA axes is going to be inherently closer to many taxa, regardless of where those other taxa are located. To some extent, a median morphology can represent true morphological convergence, but because of their central location on the chart, their apparent convergence to other taxa may be somewhat exaggerated relative to that of more outlying taxa, and the chance of producing false positives within the dataset is likely higher for taxa with a median morphology like these.

Comparisons of C-values within a particular row or column may be a way to mitigate some of these biases, as it would set the C-values as comparisons of convergence relative to a specific taxon. This could be useful for studies that wish to assess which taxa another taxon is convergent with and how distinct that convergence is. The Wheatsheaf and search.conv tests can also provide this information once desired convergence groupings are determined, such as from the C-value results.

As seen in Figure 32, many of the total comparison C-values (~50%) were reported by R as 0, which indicates either a rounded value of extremely low convergence or that the end tip taxa have a greater distance between each other in trait space than any of their ancestors. Values of 0 are taken to be true measures of no convergence, and the high quantity of 0 values in this dataset indicates that many taxa are likely divergent from one another rather than convergent.

The highest values in the C1 dataset (0.87 and 0.86) represent comparisons between *Balearica regulorum* and *Balearica pavonina*, and between *Aptenodytes patagonicus* and *Luscinia megarhynchos*, respectively.

The *Balearica* pairing represents two sister taxa with a very recent divergence and high C-values may be related to inaccuracy in the C-value calculations for taxa with a very similar

morphology due to phylogenetic constraints. The *Balearica* comparison may also be a case of parallel evolution, or related to inaccuracy in the ancestral state reconstruction. It is worth mentioning that one of the major differences in the skulls of the two *Balearica* species, which are considered "ornamented" within many of our test definitions, is the presence of two posterolateral protrusions on the skull. The posterolateral protrusions allow the skulls of these *Balearica* species to be readily distinguished from one another, as the postero-lateral protrusions are spine-like in *B. regulorum* but more rounded and greatly reduced in *B. pavonina*. These features are not captured within the sampled skull region for our study due to them being found on the parietal bones. I would expect the C-values for the whole skull might show less similarity between these two taxa.

The *Aptenodytes-Luscinia* pairing is more likely a true reflection of convergence, though surprising considering that *Aptenodytes* is a species of large penguin while *Luscinia* is a species of nightingale. Even though both taxa often hunt for small mobile prey and likely require aerodynamic bill properties for efficient locomotion through fluids, they are several magnitudes different in size and share almost no other obvious ecological similarities with one another. However, convergence between *Luscinia* and the other penguin in the dataset, *Spheniscus*, was also high, around 0.78 for the C1 value, and the cluster analysis also recovered *Luscinia* with the two penguins, suggesting the surprisingly high convergence value could be a true reflection of the similarity in geometry between the sampled area of the skulls of these birds, even considering the generally high C-values for *Luscinia* in general as mentioned earlier.

To analyze the general trends in the C-value data and identify potential biases, I plotted scaled regressions of the three C-values used in this study against the time since divergence for each pair of species tested.

In Figures 33–35, a slope of 0 for the regression line would represent a C-value that completely accounts for phylogenetic distance, meaning the only influence on the C-value at time x should be the morphological similarity or dissimilarity of the taxa. A slope of 1 would indicate that the C-values correlate directly with evolutionary time shared, indicating that time is the only control on the C-values. A slope of 1 between morphology and time shared is the expectation of the Brownian motion model and would indicate the C-values have inadequate correction for time shared to accurately assess convergence. Similarly, a slope of -1 or similar values would indicate too much overcorrection for the phylogenetic time separation to provide an accurate estimate of convergence.

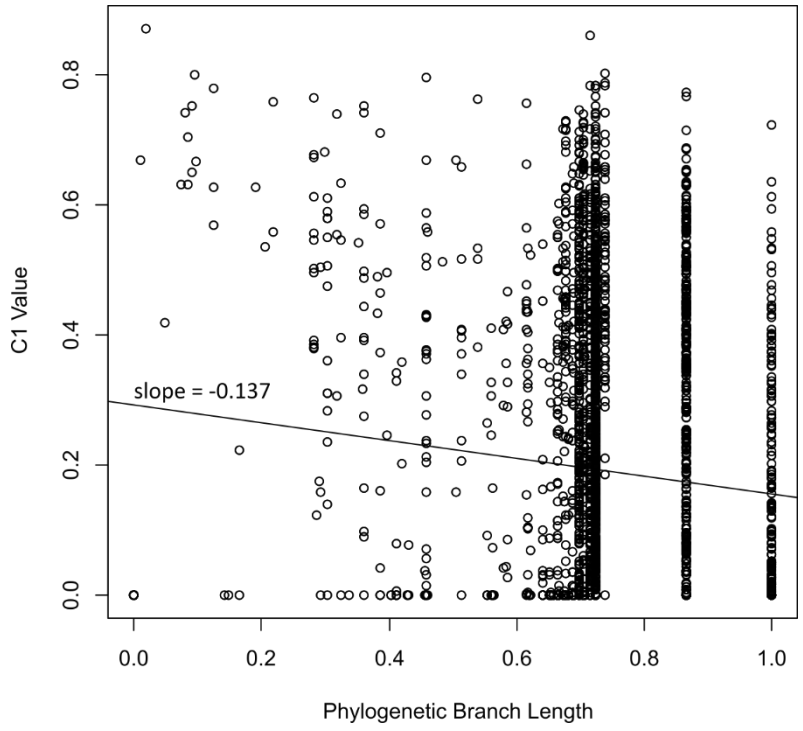

Figure 33. This chart shows a regression of the C1 values compared against a scaled ratio of phylogenetic time since separation represented by phylogenetic branch lengths. Overall, there is a small but significant decline indicating slightly higher C1 values on average for pairs of taxa that split more recently within the dataset. This effect is especially pronounced for taxa that diverged within ~20 million years (0-0.2 on the x-axis), indicating C-value measurements may have a tendency to report recently diverged taxa as more convergent than expected due to the influence of phylogenetic constraints.

Figure 33 (continued).

Datasets that include more closely related taxa for comparison could be less subject to this bias, but it merits future consideration for those who wish to use C-values. The spread of the data overall is high, indicating the potential bias toward higher C1 values in more recently-diverged taxa is an average trend; there are both very high and very low C-values across the range of divergence times. The C1 values for greater genetic divergence times (right side of the graph) likely reflect a realistic relative comparison of convergence values for the pairings. The vertical columns are a consequence of the way the bird phylogeny splits, with the column at 1 representing comparisons between paleognath and neognath species, and the column near 0.85 representing comparisons between galloanserine neognaths and all other neognaths.

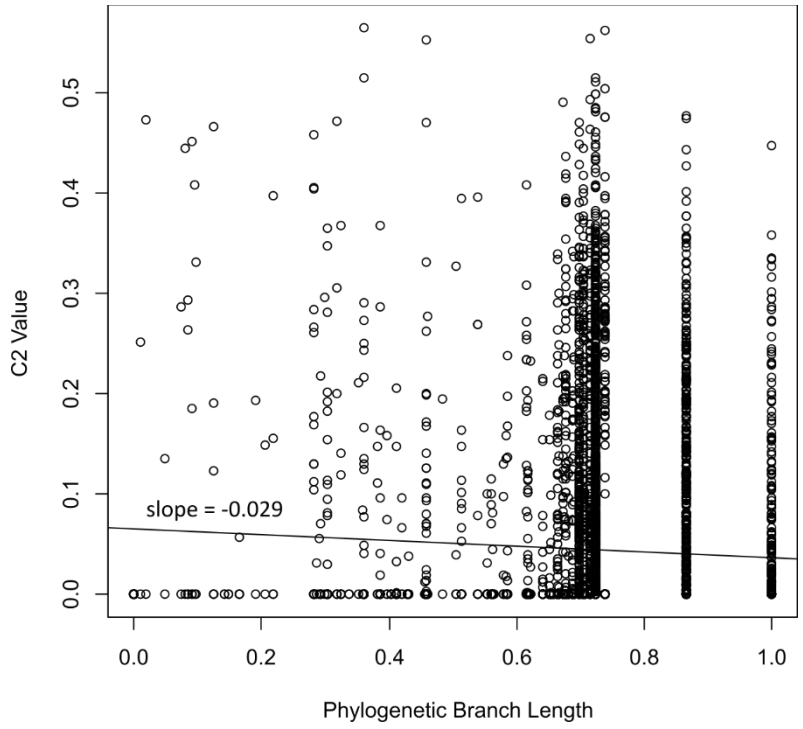

Figure 34. This chart shows a regression of the C2 values compared against a scaled ratio of phylogenetic time since separation represented by phylogenetic branch lengths. Overall, there is a very small decline indicating slightly higher C2 values on average for pairs of taxa that split more recently within the dataset, similar to the linear regression for the C1 values. Here, the data is more evenly spread, and the bias toward higher C-values in more recently-diverged taxa is minimal, though may still be worth considering for certain studies depending on their focus. A regression slope of 0 would indicate that the phylogenetic time separation has no influence on the reported C2 values, which is expected if the C2 values are providing an accurate calculation of morphological similarity adjusted by phylogenetic relatedness, as intended. Of the four C-values provided, the C2 values appear to be least influenced by the magnitude of time shared between taxa and might therefore be interpreted as providing the most suitable measurement for convergence out of the possible C-values.

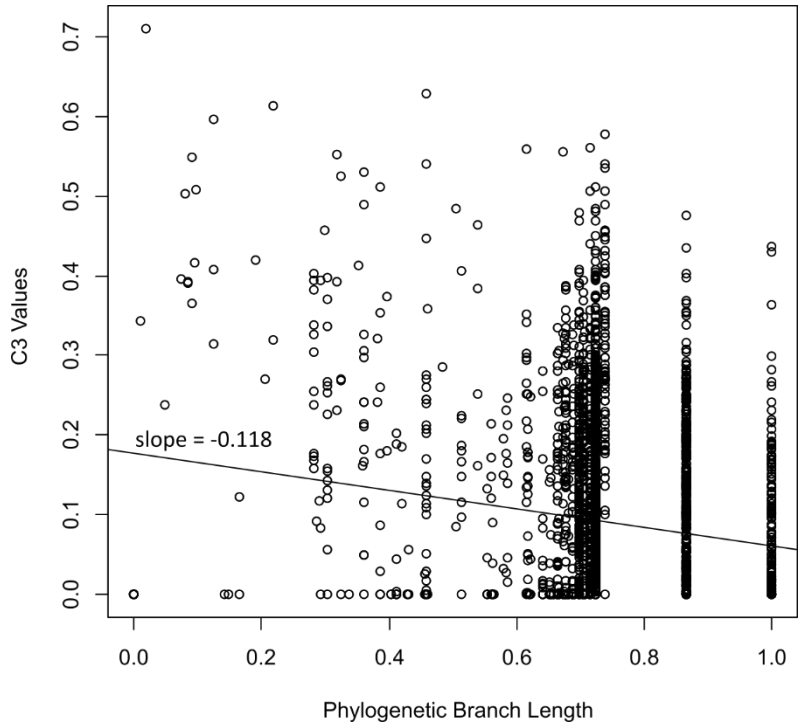

Figure 35. The above chart shows a regression of the C3 values compared against a scaled ratio of phylogenetic time shared for pairs of taxa. As with the other C-values, there is a slight but potentially important bias where higher C-values are reported for more closely-related taxa. The trend for the C3 values closely resembles that of the C1 values. It differs largely in that its values are overall much smaller than the C1 and C2 values, as expected by the C3 calculation, and the maximum C3 value for any given shared phylogenetic time (x-axis) tends to increase with increasing time shared (more recent divergence, to the right side of the graph). The reason for this is that C3 varies with the total variation within a lineage, which is generally greater for more distantly-related taxa.

Figures 33–35 show that there is some bias related to phylogenetic distance to take into account when interpreting C-value results for our dataset, but the influence is generally low and should not affect our interpretations greatly. Given that the upper bounds of the C3 values can vary considerably by phylogenetic distance whereas the C1 and C2 values tend to have a more even spread, the subsequent analyses for the C-values method will focus on the C1 and C2 values. The C2 values given that they more completely account for phylogenetic relatedness than the C1 values based on the smaller regression slope.

Based on the spread of C-value scores and relatively low phylogenetic bias in the C-values, especially for the C2 and to a lesser extent the C1 values, it appears that there are indeed several convergent taxa within the dataset. Some of these high C-values include pairings of taxa designated as having bony ornament-like structures in at least some of the TDs. However, there is not a clear delineation between ornamented and unornamented taxa based on the C-values alone, as convergence in bill shape could be due to a variety of factors, particularly in relation to diet.

6.2.3: Interpreting Convergence Using the Bolnick et al. (2018) Definition

The discussion of convergence or non-convergence in this study's results has so far assumed that the methods used are capable of discerning convergence according to a consistent definition. However, as addressed in the Introduction, the C-values, *search.conv* method, and Wheatsheaf index all use slightly different definitions for convergence, and each of these differs from my preferred definition provided by Bolnick et al. (2018). Disagreement between these results can be partially attributed to each of these metrics providing different information.

Using the Bolnick et al. (2018) definition, we can see that the C-values provide a general estimate for Bolnick et al.'s convergence, but the C-value definition sometimes struggles to differentiate between Bolnick et al. convergence and parallel evolution; sister taxa, expected to only be able to experience divergence by the Bolnick et al. definition, may produce non-zero C-values if they have evolved along a Bolnick et al. parallel evolution trajectory for long enough that at least one of the end-tip taxa is less similar to their common ancestor than they are to one another. This is demonstrated in the unexpectedly high C-values for the *Balearica* species in our dataset. It is possible that the higher C-values in our dataset are due to a mixture of Bolnick et al.

convergence, parallelism and convergence, parallelism and divergence, and parallelism with neither divergence nor convergence (see Figure 1). Non-zero C-values are capable of ruling out Bolnick et al. nonparallel divergence, and all of the other geometric possibilities defined by Bolnick et al. have been described as convergent evolution by previous authors, so the C-values are not inherently invalidated as a means of estimating morphological convergence. However, Bolnick et al. (2018) differentiates convergence from parallelism because, when defined geometrically, these terms reflect related but distinct biological concepts. Parallel and nonparallel evolution reflect evolutionary processes, and both can be predicted in real-world scenarios that inform our understanding of adaptation and speciation (Langerhans 2018). Convergence is a phenomenon dependent upon morphological distance, making it important for studies in which net similarity between two or more taxa is important, and convergence can arise from different forms of parallel and nonparallel evolution (Bolnick et al. 2018).

I would like to suggest that the *search.conv* method does not measure convergence at all, but instead, because it measures a weighted angle of evolutionary trajectories within a lineage, the *search.conv* method is a useful assessment of the strength of parallel evolution between two lineages leading back to their common ancestor. Under this interpretation, the strong evidence for convergence in the search.conv results for TDs 3.2 and 3.3 is actually evidence for parallel evolution in the ornamented taxa relative to the general parallelism of the overall dataset. The angles of the total trajectories are still fairly large (around 32.6° and 30.4° respectively, as seen in Table 7), but because they are measured in relation to a common ancestor, the angle between the lineages for the search.conv method should rarely, if ever, be 0°. The ornamented taxa in definitions 3.2 and 3.3 appear to be evolving along trajectories that are more parallel than expected within the dataset, in which case, bony ornament-like structures may or may not be

convergent by the Bolnick et al. (2018) definition. The results of this study provide strong evidence that they are the product of parallel evolution. The other definitions for ornamentation do not show parallel trends.

If the *search.conv* method only provides information about parallelism, and the C-values can be skewed by parallel evolution if the lineages evolve in parallel for long enough (which is likely the case for the 3.2 and 3.3 TDs according to Figure 7), then neither of these methods appears to be able to offer a clear assessment of whether Bolnick et al. (2018) convergence is actually occurring in the ornamented taxa. The Wheatsheaf index, being based on clustering distances, can offer a more direct assessment of tip distances, but because it does not estimate ancestral states, it remains vulnerable to falsely reporting convergence when there is none if the C-values are predominantly influenced by parallel evolution in the lineages rather than Bolnick et al. convergence. Based on the phylomorphoplot branch lengths compared to the clustering for the taxa in Figure 7, I suspect that the signal of parallel evolution is likely greater than any true convergence signal for these taxa. Even if the Wheatsheaf values do reflect true convergence, the Wheatsheaf index results found no significant convergence among the TD 3.2 taxa and only a small amount of convergence among the TD 3.3 taxa. *Euryceros precostii* and *Mitu mitu* have the smallest distances between the tips in trait space for PCs 1 and 2 relative to their phylogenetic distances to one another, so if true convergence is reflected in the 3.2 or 3.3 datasets, I would expect these two taxa to be the most convergent to one another.

Analysis of all ornamented taxa simultaneously often shows a distinction between these taxa and unornamented taxa. However, the skulls of these species are not consistently convergent; the test 3.2 and 3.3 ornamented taxa appear to be evolving in parallel to one another, but both these and the rest of the taxa with bony ornament-like structures are more likely

divergent in relation to one another rather than convergent. Taxa with prominent bony ornament-like structures could be exceptionally divergent in bill surface shape, but none of the methods used in this study was designed to measure degrees of divergence, so this claim remains untested.

6.2.4: Interpretating Convergence and Parallelism in Bony Ornament-Like Structures

It is likely that species that independently evolve bony ornamentation follow similar evolutionary pathways for the development of these structures. If the specific location and form of the enlargement is the focus of the research, enlarged structures on different regions may cause the morphology of these organisms to diverge from one another considerably more than even in the unornamented relatives of those taxa. In more distantly related taxa, especially those with ranges that do not overlap, this apparent divergence would likely be a consequence of random occupation of available morphospace rather than specific selection for divergent characters. If the process of enlargement itself is the focus of research, regardless of location, it is important to realize that these structures may develop through similar mechanisms, so parallel evolution may be occurring even in cases where the taxa themselves are not becoming more similar over time. Constraints on ornament-like structures developed in similar regions of the skull, or subject to similar physical forces, could limit the possible morphospace these structures can occupy, resulting in some structures in unrelated groups of organisms appearing to converge on a similar morphology. In the latter case, both convergent and divergent patterns could be produced through the same mechanism.

Among closely related organisms, part of the function of casques and similar ornament-like structures is suspected to be species identification and socio-sexual display. If this is true,

these structures may be subject to more direct selection for morphological divergence from related taxa, particularly if mate preference for a particular morphology drives selection against hybridization (Servedio 2001, 2004). Ornamentation in closely related species is often subject to phylogenetic constraints that limit where and how the structure can grow, and for bony ornament-like structures, evolutionary divergence between the structures of related species may alter the height, curvature, color, or outline of the ornament, but not its early growth or location. This pattern is seen in many species of cracids, hornbills, and all extant species of cassowary, and individual variation within species also varies among the more plastic components of the structure. My study did not look at individual variation, but previous studies (Green et al. 2022) have recognized that casques in species, such as cassowaries, can vary considerably among individuals, and casque variation also appears to reflect subspecies differences in the rhinoceros hornbill, *Buceros rhinoceros* (Kemp 1995).

Within a given possible morphospace, bony ornament-like structures could be divergent among closely related taxa and simultaneously appear convergent among more distantly related taxa. The divergence in this case would likely be the consequence of direct selection, where the convergence would be coincidental, driven by genetic drift or a similar random process (Stayton 2008). Selection against similar ornament morphologies is less likely to occur among species that do not share a range or are unlikely to hybridize due to behavioral or other morphological differences.

As a final consideration, modularity of the skull means that, in particular cases, some regions of the skull may show divergence whereas others show convergence within a selected group of organisms. For example, nasals may appear convergent while premaxillae appear divergent, and the same is true for different degrees of parallel and non-parallel evolution. For

more independent modules, these patterns could result from different selection, but the co-evolution of skull regions should also be considered, especially for structures connected directly to the bill.

6.3: Future Work

The work in this study highlights new information about bony protrusions associated with the bills of birds. This study also emphasizes that there are substantial gaps in our knowledge about these structures in modern animals. Taxa bearing bony ornament-like structures can be recovered as convergent, parallel, or neither, depending on the taxa selected and the methods used. If we wish to be able to compare casques and similar structures across clades and use them to glean behavioral and ecological information about extinct species, we need consistent, testable definitions for bony ornament-like structures, as well as consistent use of convergent evolution and parallel evolution. We need to be able to communicate why we consider bony ornament-like structures to be similar to one another, especially because they often arise independently and take on a variety of morphological shapes.

Based on the work conducted here, I do think that it is reasonable to categorize bony ornament-like structures as a collective type of trait in vertebrates, as more prominent bony ornament-like structures are readily distinguished and at least some of these structures appear to follow similar evolutionary patterns. However, the evolution of these structures can be complex, particularly across broad evolutionary clades. Even though we can recognize the independent evolution of bony enlargements in different clades, that does not mean that these structures are necessarily convergent in the details of their shape, size, or anatomical structure.

Birds with bony ornament-like structures are rarer than those without these structures, but these structures still appear to be quite common. The diversity of bony ornament-like structures across birds has yet to be fully characterized with computed tomography and morphometrics, but efforts to quantify these structures are worthwhile. Hornbills and cracids are among the most well-studied taxa with bony ornament-like structures, but despite detailed records of the biology of many of these species collected by authors like Kemp (1995), anatomical detail about their casques is limited to a small handful of studies (e.g., Alexander et al. 1994, Gamble 2007). The diversity of bony ornament-like structures within these groups alone implies an association between speciation and bony ornament-like structure evolution. Almost all of my tests found differences in the covariation between bill morphology and ecological variables between ornamented and unornamented taxa, but the nature of these associations is unclear. It seems unlikely that sexual selection should be discarded as an influence on the form and function of bony ornament-like structures, given many of them bear bright colors and are highly variable between closely related species (Kemp 1995). However, associations between bony ornament-like structures and thermoregulation, as well as other possible ecologically significant traits, suggest a complex relationship between ornament-like structures, sexual selection, and non-sexual selection that offers interesting opportunities for future research.

Methodology for describing the skull shape should be carefully scrutinized. This study used the standard practice of marking anatomical points with patches and filling those patches with semilandmarks placed along the surface of the bone (Wiley 2005, Bardua et al. 2019), which is useful for quantifying shape changes among the bones of the skull, but may not be the best method for studying convergence in overall skull shape. The cassowary, for example, forms its casque from an expansion of the nasal and a unique central bone element (Green & Gignac

2021), which is not the case for similar-looking birds like cracids. If the landmarks on these regions are different for cassowaries than they are for cracids, then the convergence metrics used in this study will likely have a hard time detecting similarity that may be obvious to the human eye. In convergent evolution, similar structures can originate from different anatomical regions. Volumetric studies, or surface studies that treat the skull as a single bone could give different results than the element-based approach to describing the morphology of these birds.

Future work built upon this study would ideally incorporate a broader range of specimens from unsampled species, as well as assess how intraspecific variation, such as sexual dimorphism, influences convergence metrics for taxa with bony ornament-like structures. The goal of this research is to provide information about ornament-like structures within birds, and eventually use this group to study similar structures in broader groups of animals, including extinct taxa.

My sample set is fairly broad and covers most of the major bird groups with bony ornament-like structures, but there are several important taxa I did not have access to that I would like to explore in future studies. Some pigeons, including members of the *Ducula* genus and *Ptilinopus insolitus*, have protrusions that appear to contain a bony core. The recently-extinct Rodrigues solitaire (*Pezophaps solitaria*) is also reported to have a bony frontal growth, similar to that of *Balearica*, though slightly smaller (Mayr 2018). The extinct bird *Sylviornis* would be a good fossil comparison to the modern bird dataset, as it is still within crown Aves. *Cygnus olor*, *Melanitta*, or other Anseriformes with prominent bony ornament-like structures would offer a useful comparison to the related *Anseranas* and *Anas* within the present dataset. While I was able to sample one member of the genus *Philemon*, this was an unornamented taxon. Three species of *Philemon* (*P. corniculatus*, *P. buceroides*, and *P. argenticeps*) are reported by Mayr (2018) to

have prominent bony ornament-like structures. *P. corniculatus* in particular has a tall, isolated, horn-like ornament that would likely constitute its own category of bony ornament-like structure. Finally, while I sampled three hornbill species available on Morphosource and Phenome10k, the range of hornbill casque morphologies is far more extensive than my sample set, and these birds merit considerably more research into their casque variation.

Among non-avian taxa, there are several beaked tetrapods with somewhat similar cranial structures that might be suitable to incorporate into a broader comparative dataset of bony ornament-like structures. Oviraptorosaurs are the closest toothless relatives of birds which also bear casque-like structures in some species. Hadrosaurs, a more distantly-related group of non-avian dinosaurs, also have cranial ornament-like structures that appear to follow a similar developmental pattern to that seen in some birds (Evans et al. 2007). Pterosaurs also have an extensive record of bony ornament-like structures and soft tissue ornament-like structures as well (Bennett 2002). Some fossil ornament-like structures appear to take on forms that are not represented in modern birds, which suggests these structures merit special attention. *Citipati*, *Guanlong*, *Saurolophus*, and *Dilophosaurus*, for instance, all have external concavities and vacuities in the bone of their ornament-like structures. External cavities like these do not occur in modern birds.

Limitations should be considered when comparing modern species to fossil taxa. Sometimes, there is not a good analogue for a trait seen in a fossil species, at which point concrete comparisons must focus on what shared elements can be studied. Models can also help assess the point at which comparisons to modern taxa break down – for instance, whether size or lifestyle might affect the way non-avian taxa develop bony ornament-like structures. Macroevolutionary trends in prominent bony ornament-like structures are still observable across

broad groups and time scales (Gates et al. 2016), and my data support the general idea that taxa with ornament-like structures are distinguishable from taxa lacking these structures, so the patterns in the evolution of these structures merit further exploration on both long and short time scales.

Comparisons between the skulls of birds and those of other groups with bony ornament-like structures can provide information for how consistent this study's results are among non-avian taxa. Broader taxonomic comparisons can also clarify the limitations of existing convergence metrics, and what bony ornament-like structures can tell us about the macroevolution of the lineages that have acquired them. The prospect that these structures may be simultaneously parallel and divergent in their evolution offers possible guidance for how to approach broad comparisons of ornament-like structures in the future.

REFERENCES

Abrahams, M. V. (1993). The trade-off between foraging and courting in male guppies. *Animal Behaviour*, 45(4), 673–681.

Alexander, G. D., Houston, D. C., & Campbell, M. (1994). A possible acoustic function for the casque structure in hornbills (Aves: Bucerotidae). *Journal of Zoology*, 233(1), 57–67.

Alonzo, S. H. & Servedio, M. R. (2019). Grey zones of sexual selection: why is finding a modern definition so hard? *Proceedings of the Royal Society B*, 286(1909), 1-8.

Alonzo, S. H. & Servedio, M. R. (2021). Embracing complexity might help us find clarity: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 800–800.

Andersen, M. J., McCullough, J. M., Mauck III, W. M., Smith, B. T., & Moyle, R. G. (2018). A phylogeny of kingfishers reveals an Indomalayan origin and elevated rates of diversification on oceanic islands. *Journal of Biogeography*, 45(2), 269–281.

Andersson, M. (2021). Amending Darwin on "sexual selection": a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 797.

Arbuckle, K., Bennett, C. M., & Speed, M. P. (2014). A simple measure of the strength of convergent evolution. *Methods in Ecology and Evolution*, 5(7), 685–693.

Ballentine, B., Horton, B., Brown, E. T., & Greenberg, R. (2013). Divergent selection on bill morphology contributes to nonrandom mating between swamp sparrow subspecies. *Animal Behaviour*, 86(2), 467–473.

Bardua, C., Felice, R. N., Watanabe, A., Fabre, A. C., & Goswami, A. (2019). A practical guide to sliding and surface semilandmarks in morphometric analyses. *Integrative Organismal Biology*, 1(1), 1–34.

Bennett, S. C. (2002). Soft tissue preservation of the cranial crest of the pterosaur *Germanodactylus* from Solnhofen. *Journal of Vertebrate Paleontology*, 22(1), 43–48.

Bolnick, D.I., Barrett, K.B.O., Rennison, D.J., & Stuart, Y.E. (2018). (Non)Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics*, 49, 303–330.

Borm, G. F., Fransen, J., & Lemmens, W. A. (2007). A simple sample size formula for analysis of covariance in randomized clinical trials. *Journal of clinical epidemiology*, 60(12), 1234–1238.

Bright, J. A., Marugán-Lobón, J., Cobb, S. N., & Rayfield, E. J. (2016). The shapes of bird beaks are highly controlled by nondietary factors. *Proceedings of the National Academy of Sciences*, 113(19), 5352–5357.

Brown, C. M. & Henderson, D. M. (2015). A new horned dinosaur reveals convergent evolution in cranial ornamentation in Ceratopsidae. *Current Biology*, 25(12), 1641–1648.

Buchholz, R. (1996). Thermoregulatory role of the unfeathered head and neck in male wild turkeys. *The Auk*, 113(2), 310–318.

W.L. Buller (1871). On the structure and habits of the Huia (*Heteralocha gouldii*). *Transactions and Proceedings of the New Zealand Institute*, 3(1871), 24–29.

Cardini, A. (2017). Left, right or both? Estimating and improving accuracy of one-side-only geometric morphometric analyses of cranial variation. *Journal of Zoological Systematics and Evolutionary Research*, 55(1), 1–10.

Castiglione, S., Serio, C., Tamagnini, D., Melchionna, M., Mondanaro, A., Di Febbraro, M., Profico, A., Piras, P., Barattolo, F. & Raia, P. (2019). A new, fast method to search for morphological convergence with shape data. *PLOS ONE*, 14(12), 1–20.

Chávez-Hoffmeister, M. (2020). Bill disparity and feeding strategies among fossil and modern penguins. *Paleobiology*, 46(2), 176–192.

Clutton-Brock, T. (2021). The costs and benefits of re-defining sexual selection: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 796–797.

Craig, A. J. F. K. (1996). The annual cycle of wing-moult and breeding in the Wattled Starling *Creatophora cinerea*. *Ibis*, 138(3), 448–454.

Crouch, N. M. & Tobias, J. A. (2022). The causes and ecological context of rapid morphological evolution in birds. *Ecology Letters*, 25(3), 611–623.

Darwin, C. (1872). *The Origin of Species (6th ed.)*. John Murray.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex (1st ed.)*. John Murray.

Darwin, C. & Wallace, A (1858). On the Tendency of Species to Form Varieties; and on the perpetuation of Varieties and Species by Natural Means of Selection. *Zoological Journal of the Linnean Society*, 3(9), 46–62.

Delhey, K. & Peters, A. (2017). The effect of colour-producing mechanisms on plumage sexual dichromatism in passerines and parrots. *Functional Ecology*, 31(4), 903–914.

Dunning, J., Diamond, A. W., Christmas, S. E., Cole, E. L., Holberton, R. L., Jackson, H. J., Kelly, K. G., Brown, D., Rivera, I. J., & Hanley, D. (2018). Photoluminescence in the bill of the Atlantic Puffin *Fratercula arctica*. *Bird Study*, 65(4), 570–573.

Eastick, D. L., Tattersall, G. J., Watson, S. J., Lesku, J. A., & Robert, K. A. (2019). Cassowary casques act as thermal windows. *Scientific reports*, 9(1), 1–7.

Eberhard, W. G., Rodríguez, R. L., Huber, B. A., Speck, B., Miller, H., Buzatto, B. A., & Machado, G. (2018). Sexual selection and static allometry: the importance of function. *The Quarterly Review of Biology*, 93(3), 207–250.

Evans, D. C., Reisz, R. R., & Dupuis, K. (2007). A juvenile *Parasaurolophus* (Ornithischia: Hadrosauridae) braincase from Dinosaur Provincial Park, Alberta, with comments on crest ontogeny in the genus. *Journal of Vertebrate Paleontology*, 27(3), 642–650.

Freeman, B. G., Weeks, T., Schluter, D., & Tobias, J. A. (2022). The latitudinal gradient in rates of evolution for bird beaks, a species interaction trait. *Ecology Letters*, 25(3), 635–646.

Gamble, K. C. (2007). Internal anatomy of the hornbill casque described by radiography, contrast radiography, and computed tomography. *Journal of Avian Medicine and Surgery*, 21(1), 38–49.

Garcia-R, J. C., Lemmon, E. M., Lemmon, A. R., & French, N. (2020). Phylogenomic reconstruction sheds light on new relationships and timescale of rails (Aves: Rallidae) evolution. *Diversity*, 12(2), 1–10.

Ghalambor, C. K., Reznick, D. N., & Walker, J. A. (2004). Constraints on adaptive evolution: the functional trade-off between reproduction and fast-start swimming performance in the Trinidadian guppy (*Poecilia reticulata*). *The American Naturalist*, 164(1), 38–50.

Gibb, G. C., Kennedy, M., & Penny, D. (2013). Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. *Molecular Phylogenetics and Evolution*, 68(2), 229–238.

Gómez, R. O. & Lois-Milevicich, J. (2021). Why the long beak? Phylogeny, convergence, feeding ecology, and evolutionary allometry shaped the skull of the Giant Cowbird *Molothrus oryzivorus* (Icteridae). *Journal of Morphology*, 282(11), 1587–1603.

Gonzalez, J., Düttmann, H., & Wink, M. (2009). Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *Journal of Zoology*, 279(3), 310–318.

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard university press.

Green, T. L., & Gignac, P. M. (2021). Osteological description of casque ontogeny in the southern cassowary (*Casuarius casuarius*) using micro-CT imaging. *The Anatomical Record*, 304(3), 461–479.

Green, T. L., Kay, D. I., & Gignac, P. M. (2022). Intraspecific variation and directional casque asymmetry in adult southern cassowaries (*Casuarius casuarius*). *Journal of Anatomy*, 241(4), 951–965.

Greenberg, R. & Olsen, B. (2010). Bill size and dimorphism in tidal-marsh sparrows: island-like processes in a continental habitat. *Ecology*, 91(8), 2428–2436.

Greenberg, R., Cadena, V., Danner, R. M., & Tattersall, G. (2012). Heat loss may explain bill size differences between birds occupying different habitats. *PLOS ONE*, 7(7), 1–9.

Heinen-Kay, J. L., Morris, K. E., Ryan, N. A., Byerley, S. L., Venezia, R. E., Peterson, M. N., & Langerhans, R. B. (2015). A trade-off between natural and sexual selection underlies diversification of a sexual signal. *Behavioral Ecology*, 26(2), 533–542.

Hone, D. W., Naish, D., & Cuthill, I. C. (2012). Does mutual sexual selection explain the evolution of head crests in pterosaurs and dinosaurs? *Lethaia*, 45(2), 139–156.

Hooper, D. M. & Price, T. D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology & Evolution*, 1(10), 1526–1534.

Horrocks, N., Perrins, C., & Charmantier, A. (2009). Seasonal changes in male and female bill knob size in the mute swan *Cygnus olor*. *Journal of Avian Biology*, 40(5), 511–519.

Hughes, A. L. (2014). Evolution of bill size in relation to body size in toucans and hornbills (Aves: Piciformes and Bucerotiformes). *Zoologia* (Curitiba), 31, 256–263.

Ingram, T., & Mahler, D. L. (2013). SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5), 416–425.

Janse van Vuuren, A. K., Kemp, L. V., & McKechnie, A. E. (2020). The beak and unfeathered skin as heat radiators in the Southern Ground-hornbill. *Journal of Avian Biology*, 51(6), 1–7.

Jones, I. L., & Montgomerie, R. (1992). Least Auklet ornaments: Do they function as quality indicators?. *Behavioral Ecology and Sociobiology*, 30(1), 43–52.

Kasambe, R., Charde, P., & Yosef, R. (2011). Aerial jousting and bill grappling in Indian Grey Hornbill (*Ocyceros birostris*). *acta ethologica*, 14(1), 13–15.

Kemp, A. C. (1995). *The hornbills: bucerotiformes* (Vol. 1). Oxford University Press.

Kinnaird, M. F., Hadiprakarsa, Y. Y., & Thiensongrusamee, P. (2003). Aerial jousting by Helmeted Hornbills *Rhinoplax vigil*: observations from Indonesia and Thailand. *Ibis*, 145(3), 506–508.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, 36(1), 1–12.

Knell, R. J., Naish, D., Tomkins, J. L., & Hone, D. W. (2013). Sexual selection in prehistoric animals: detection and implications. *Trends in Ecology & Evolution*, 28(1), 38–47.

Kodric-Brown, A., Sibly, R. M., & Brown, J. H. (2006). The allometry of ornaments and weapons. *Proceedings of the National Academy of Sciences*, 103(23), 8733–8738.

Kokko, H. (2021). The tired copepod and the definition of sexual selection: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 795–796.

Kraaijeveld, K., Kraaijeveld-Smit, F. J., & Komdeur, J. (2007). The evolution of mutual ornamentation. *Animal Behaviour*, 74(4), 657–677.

Krajewski, C., Sipiorski, J. T., & Anderson, F. E. (2010). Complete mitochondrial genome sequences and the phylogeny of cranes (Gruiformes: Gruidae). *The Auk*, 127(2), 440–452.

Krishtalka, L., Stucky, R. K., & Beard, K. C. (1990). The earliest fossil evidence for sexual dimorphism in primates. *Proceedings of the National Academy of Sciences*, 87(13), 5223–5226.

Kuhl, H., Frankl-Vilches, C., Bakker, A., Mayr, G., Nikolaus, G., Boerno, S. T., Klages, S., Timmermann, B., & Gahr, M. (2021). An unbiased molecular approach using 3′-UTRs resolves the avian family-level tree of life. *Molecular Biology and Evolution*, 38(1), 108–127.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6), 3721–3725.

Langerhans, R. B. (2010). Predicting evolution with generalized models of divergent selection: a case study with poeciliid fish. *Integrative and Comparative Biology*, 50(6), 1167–1184.

Langerhans, R. B. (2018). Predictability and parallelism of multitrait adaptation. *Journal of Heredity*, 109(1), 59–70.

Luther, D., & Greenberg, R. (2014). Habitat type and ambient temperature contribute to bill morphology. *Ecology and Evolution*, 4(6), 699–705.

Lyon, B. E., & Montgomerie, R. (2012). Sexual selection is a form of social selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1600), 2266–2273.

Mallon, J. C. (2017). Recognizing sexual dimorphism in the fossil record: lessons from nonavian dinosaurs. *Paleobiology*, 43(3), 495–507.

Mattos-Reano, J. C. (2013). Phylogenetic analysis and phylogeography of the family Casuariidae (Aves: Casuariiformes). [Master's thesis, The California State University]. ScholarWorks.

McCullough, E. L., Miller, C. W., & Emlen, D. J. (2016). Why sexually selected weapons are not ornaments. *Trends in Ecology & Evolution*, 31(10), 742–751.

O'Brien, H. D., Faith, J. T., Jenkins, K. E., Peppe, D. J., Plummer, T. W., Jacobs, Z. L., Li, B., Joannes-Boyau, R., Price, G., Feng, Y., & Tryon, C. A. (2016). Unexpected convergent evolution of nasal domes between Pleistocene bovids and Cretaceous hadrosaur dinosaurs. *Current Biology*, 26(4), 503–508.

O'Brien, D. M., Allen, C. E., Van Kleeck, M. J., Hone, D., Knell, R., Knapp, A., Christiansen, S., & Emlen, D. J. (2018). On the evolution of extreme structures: static scaling and the function of sexually selected signals. *Animal Behaviour*, 144, 95–108.

Ord, T. J., Stamps, J. A., & Losos, J. B. (2013). Convergent evolution in the territorial communication of a classic adaptive radiation: Caribbean *Anolis* lizards. *Animal Behaviour*, 85(6), 1415–1426.

Padian, K. & Horner, J. R. (2011). The evolution of 'bizarre structures' in dinosaurs: biomechanics, sexual selection, social selection or species recognition? *Journal of Zoology*, 283(1), 3–17.

Peters, W. S. & Peters, D. S. (2009). Life history, sexual dimorphism and 'ornamental' feathers in the Mesozoic bird *Confuciusornis sanctus*. *Biology Letters*, 5(6), 817–820.

Phillips, P. K. & Sanborn, A. F. (1994). An infrared, thermographic study of surface temperature in three ratites: ostrich, emu and double-wattled cassowary. *Journal of Thermal Biology*, 19(6), 423–430.

Pigot, A. L., Sheard, C., Miller, E. T., Bregman, T. P., Freeman, B. G., Roll, U., Seddon, N., Trisos, C. H., Weeks, B. C., & Tobias, J. A. (2020). Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology & Evolution*, 4(2), 230-239.

Price, T. D., Hooper, D. M., Buchanan, C. D., Johansson, U. S., Tietze, D. T., Alström, P., Olsson, U., Ghosh-Harihar, M., Ishtiaq, F., Gupta, S. K., Martens, J., Harr, B., Singh, P., & Mohan, D. (2014). Niche filling slows the diversification of Himalayan songbirds. *Nature*, 509(7499), 222–225.

Prum, R. O. (2010). The Lande–Kirkpatrick mechanism is the null model of evolution by intersexual selection: implications for meaning, honesty, and design in intersexual signals. *Evolution: International Journal of Organic Evolution*, 64(11), 3085–3100.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526(7574), 569–573.

Rehg, J. A. & Leigh, S. R. (1999). Estimating sexual dimorphism and size differences in the fossil record: a test of methods. *American Journal of Physical Anthropology*, 110(1), 95–104.

Rundle, H. D. & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352.

Servedio, M. R. (2001). Beyond reinforcement: the evolution of premating isolation by direct selection on preferences and postmating, prezygotic incompatibilities. *Evolution*, 55(10), 1909–1920.

Servedio, M. R. (2004). The evolution of premating isolation: local adaptation and natural and sexual selection against hybrids. *Evolution*, 58(5), 913–924.

Shuker, D. M. & Kvarnemo, C. (2021). The definition of sexual selection. *Behavioral Ecology*, 32(5), 781–794.

Shuster, S. M., Wade, M. J., & Simmons, L. (2021). A grander view of life that includes gametic processes: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 798–798.

Simmons, L. W. & Parker, G. A. (2021). The devil is in the details: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, 32(5), 798–799.

Smaers, J. B. & Rohlf, F. J. (2016). Testing species' deviation from allometric predictions using the phylogenetic regression. *Evolution*, 70(5), 1145–1149.

Smith, N. A. & Clarke, J. A. (2015). Systematics and evolution of the Pan-Alcidae (Aves, Charadriiformes). *Journal of Avian Biology*, 46(2), 125–140.

Soghigian, J., Andreadis, T. G., & Livdahl, T. P. (2017). From ground pools to treeholes: convergent evolution of habitat and phenotype in *Aedes* mosquitoes. *BMC evolutionary biology*, 17(1), 1–13.

Speed, M. P. & Arbuckle, K. (2017). Quantification provides a conceptual basis for convergent evolution. *Biological Reviews*, 92(2), 815–829.

Stayton, C. T. (2015). The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution*, 69(8), 2140–2153.

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14(11), 751–764.

Subramanian, S., Beans-Picón, G., Swaminathan, S. K., Millar, C. D., & Lambert, D. M. (2013). Evidence for a recent origin of penguins. *Biology Letters*, 9(6), 1–4.

Tattersall, G. J., Andrade, D. V., & Abe, A. S. (2009). Heat exchange from the toucan bill reveals a controllable vascular thermal radiator. *Science*, 325(5939), 468–470.

Tattersall, G. J., Arnaout, B., & Symonds, M. R. (2017). The evolution of the avian bill as a thermoregulatory organ. *Biological Reviews*, 92(3), 1630–1656.

Tobias, J. A., Montgomerie, R., & Lyon, B. E. (2012). The evolution of female ornaments and weaponry: social selection, sexual selection and ecological competition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1600), 2274–2293.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., Darski, B., Freeman, B. G., Bregman, T. P., Cooney, C. R., Hughes, E. C., Capp, E. J. R., Varley, Z. K., Friedman, N. R., Korntheuer, H., Corrales-Vargas, A., Trisos, C. H., Weeks, B. C., Hanz, D. M., Töpfer, T., Bravo, G. A., Remeš, V., Nowak, L., Carneiro, L. S., Moncada R., A. J., Matysioková, B., Baldassarre, D. T., Martínez-Salinas, A., Wolfe, J. D., Chapman, P. M., Daly, B. G., Sorensen, M. C., Neu, A., Ford, M. A., Mayhew, R. J., Silveira, L. F., Kelly, D. J., Annorbah, N. N. D., Pollock, H. S., Grabowska-Zhang, A. M., McEntee, J. P., Gonzalez, J. C. T., Meneses, C. G., Muñoz, M. C., Powell, L. L., Jamie, G. A., Matthews, T. J., Johnson, O., Brito, G. R. R., Zyskowski, K., Crates, R., Harvey, M. G., Zevallos, M. J., Hosner, P. A., Bradfer-Lawrence, T., Maley, J. M., Stiles, F. G., Lima, H. S., Provost, K. L., Chibesa, M., Mashao, M., Howard, J. T., Mlamba, E., Chua, M. A. H., Li, B., Gómez, B. I., García, N. C., Päckert, M., Fuchs, J., Ali, J. R., Derryberry, E. P., Carlson, M. L., Urriza, R. C., Brzeski, K. E., Prawiradilaga, D. M., Rayner, M. J., Miller, E. T., Bowie, R. C. K., Lafontaine, R., Scofield, R. P., Lou, Y., Somarathna, L., Lepage, D., Illif, M., Neuschulz, E. L., Templin, M., Dehling, D. M., Cooper, J. C., Pauwels, O. S. G., Analuddin, K., Fjeldså, J., Seddon, N., Sweet, P. R., DeClerck, F. A. J., Naka, L. N., Brawn, J. D., Aleixo, A., Böhning-Gaese, K., Rahbek, C., Fritz, S. A., Thomas, G. H., & Schleuning, M. (2022). AVONET: morphological, ecological and geographical data for all birds. *Ecology Letters*, 25(3), 581–597.

Tobias, J. A. (2022). A bird in the hand: Global-scale morphological trait datasets open new frontiers of ecology, evolution and ecosystem science. *Ecology Letters*, 25(3), 573–580.

Van De Ven, T. M. F. N., Martin, R. O., Vink, T. J. F., McKechnie, A. E., & Cunningham, S. J. (2016). Regulation of heat exchange across the hornbill beak: functional similarities with toucans? *PLOS ONE*, 11(5), 1–14.

Van Niekerk, J. H., Megía-Palma, R., & Forcina, G. (2022). Thermoregulatory function and sexual dimorphism of the throat sack in Helmeted Guineafowl (*Numida meleagris*) across Africa. *Avian Research*, 13(2022), 1–10.

Viseshakul, N., Charoennitikul, W., Kitamura, S., Kemp, A., THONG-AREE, S., Surapunpitak, Y., Poonswad, P., & Ponglikitmongkol, M. (2011). A phylogeny of frugivorous hornbills linked to the evolution of Indian plants within Asian rainforests. *Journal of Evolutionary Biology*, 24(7), 1533–1545.

Wasserstein, R. L. & Lazar, N. A. (2016). ASA statement on statistical significance and P-values. *The American Statistician*, 70(2), 129–133.

West-Eberhard, M. J. (1984). Sexual selection, competitive communication and species-specific signals in insects. In *Insect communication (Lewis, T., ed.)*. Academic Press.

Wiley, D. F., Amenta, N., Alcantara, D. A., Ghosh, D., Kil, Y. J., Delson, E., Harcourt-Smith, W., Rohlf, F. J., John, K. S., & Hamann, B. (2005). Evolutionary morphing. *Proceedings of the IEEE Visualization 2005*, 431–438.

Wright, T. F., Schirtzinger, E. E., Matsumoto, T., Eberhard, J. R., Graves, G. R., Sanchez, J. J., Capelli, S., Müller, H., Scharpegge, J., Chambers, G. K., & Fleischer, R. C. (2008). A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a Gondwanan origin during the Cretaceous. *Molecular Biology and Evolution*, 25(10), 2141–2156.

Ye, C. Y., Tang, W., Wu, D., Jia, L., Qiu, J., Chen, M., Mao, L., Lin, F., Xu, H., Yu, X., Lu, Y., Wang, Y., Olsen, K. M., Timko, M. P., & Fan, L. (2019). Genomic evidence of human selection on Vavilovian mimicry. *Nature Ecology & Evolution*, 3(10), 1474–1482.

Zuk, M. (2021). Sex differences, sexual selection, and gamete size: a comment on Shuker and Kvarnemo. *Behavioral Ecology*, XX(XX), 1–2.