Title

Sex estimation from morphology in living animals and dinosaurs

Journal

Zoological Journal of the Linnean Society, 192(4)

ISSN

0024-4082

Author

Motani, Ryosuke

Publication Date

2021-07-30

DOI

10.1093/zoolinnean/zlaa181

Peer reviewed

Sex estimation from morphology in living animals and dinosaurs

Ryosuke Motani

Department of Earth and Planetary Sciences, University of California, Davis, One Shields Avenue, Davis, California, 95616 USA; rmotani@ucdavis.edu

ABSTRACT

Sexual dimorphism is a prevalent feature of sexually reproducing organisms yet its presence in dinosaurs has recently been questioned. The inferred absence of sexual dimorphism, however, may be a methodological artefact, rooted in the lack of systematic knowledge concerning how sexual dimorphism of living animals behaves statistically. To start building such knowledge, I reanalyzed published data of 139 species of living animals that are sexually dimorphic. The previous method used for dinosaurs recognized only 5% of the living species correctly as dimorphic. This low rate is largely caused by tilting of ordinated multivariate space due to interactions between size and shape dimorphisms, low signal/noise ratios, and inclusion of outliers. The rate can be improved to 50% by modifying the method but not further, unless the information on the sex of individual specimens is used—such information is unavailable in dinosaurs, so sexual dimorphism probably cannot be established for a large proportion of sexually dimorphic dinosaurs. At the same time, about 32% of the 139 are strongly sexually dimorphic, and can be re-sexed from shape with misclassification rates below 0.05. A reassessment of dinosaurian data suggests that sexual dimorphism likely existed at least in some species, such as *Allosaurus fragilis*.

Keywords: dinosaurs, sex estimation, sexual dimorphism, shape dimorphism, size dimorphism

INTRODUCTION

Sexual dimorphism and its evolution have fascinated biologists over two centuries. The notion of sexual dimorphism was already noted by John Hunter [1728-1793], who stated that males of many species differed from their respective females not only in “organs of generation” but also in “secondary properties” (Hunter, 1837). Darwin later called these traits primary and secondary sexual characters (Darwin, 1859, 1871), with the primary sexual characters being “organs of reproduction” (Darwin, 1871) whereas secondary sexual characters being “ attached to one sex, but are not directly connected with the act of reproduction” (Darwin, 1859). Darwin also noted that dimorphism between sexes may also be seen outside of the primary and secondary sexual characters, in “structures related to habits of life” (Darwin, 1871), e.g., if males and females have different dietary preferences, their morphologies may differ although the difference would have limited impacts on reproduction per se. This third category is rarely mentioned in the recent literature, in which it is considered a part of secondary sexual characters sensu lato. Darwin separated this third category from the secondary sexual characters probably because he did not expect it to be sexually selected—he noted that sexually selected characters were seen when females and males “have the same general habits of life” (Darwin, 1859). In the modern context, sexual dimorphism encompasses not only morphology per se but also “morphs” in other aspects of biology, such as physiology and genetics. The present study focuses on sexual dimorphism in the morphology of primary and secondary sexual characters in the broad sense.

Sexual dimorphism is typically seen in two components of morphology—size and shape. Sexual size dimorphism, in which the mean sizes of adult females and males differ, is a feature seen in many animals including humans (Fairbairn, 1997). It can be quantified uniformly across

taxa by simply measuring the body mass or a representative length, such as the snout-vent length of reptiles, of many individuals. Probably thanks to this simplicity of quantification, it has been studied extensively, especially in relation to sexual selection and Rensch's Rule (Fairbairn, Blanckenhorn, & Székely, 2007). Sexual shape dimorphism, on the other hand, is difficult to quantify uniformly across taxa because it is expressed in different body parts depending on the taxon. It has been hypothesized that sexual shape dimorphism is detectable as bimodality along the principal axes from a Principal Component Analysis (PCA) of morphological variables, especially PC2 (Chapman *et al.*, 1997; Mallon, 2017)—size is not removed a priori in these analyses so PC1 is expected to represent size. If true, the assumption would provide a uniform method to quantify sexual shape dimorphism across taxa.

There are, however, a few reasons why this assumption may not hold. First is the signal/noise ratio in the total data—given that sexual shape dimorphism may only be seen in a limited number of characters, it is questionable if the signal remains clearly detectable when mixed with noise from variations in many other characters that are not sexually dimorphic. Second, the presence of two distinctive groups, females and males, in the data is expected to bias the principal component axes, as with the famous examples of the mismatch between total and groupwise regression lines (Sokal & Rohlf, 2012)(Fig. 1A). In that case, principal components, which are theoretically orthogonal to each other, may be correlated when viewed sex-wise (Fig. 1B). Third, how the co-existence of sexual shape and size dimorphisms affects the ordinated multivariate space is poorly understood. In the simplest case of sexual size dimorphism without shape dimorphism, the morphospace may appear as in Fig. 1C, whereas it may resemble Fig. 1D in the simple case of sexual shape dimorphism without size dimorphism, as postulated by Chapman et al. (1997). When the shape and size dimorphisms coexist, however, there is no

guarantee that the morphospace is shaped simply as suggested by Chapman et al. (1997) (Fig. 1E)—the coexistence of the two may accentuate the second bias to make the morphospace resemble Fig. 1F. PC2 may be bimodal in Fig. 1D and E but bimodality would be obscure in Fig. 1F, especially if the sample size is small. Therefore, the assumption needs to be tested with empirical data from extant species with sexual dimorphism.

The vast majority of the studies on sexual dimorphism has been made on extant organisms but the subject has also been studied in fossil organisms, for which the sex of the individual specimens is usually unknown a priori. Without the knowledge of sex, statistical methods that are usually used to test sexual dimorphism, such as Analysis of Variance (ANOVA) and Multiple Analysis of Variance (MANOVA), cannot be used. Also, some fossil samples contain multiple species with sexual dimorphism that are similar to each other, complicating the problem (Motani *et al.*, 2018). Despite these handicaps, there have been successful case studies of sexual dimorphism in fossil animals. Among Mesozoic reptiles, for example, sexual size dimorphism has been suggested in the pterosaur *Pteranodon* based on the presence of two size classes that also differ in pelvic morphology (Bennett, 1992), whereas sexual shape dimorphism has been suggested for pachypleurosaurs, a group of Mesozoic marine reptiles, based on the size and robustness of limb bones relative to the body (Rieppel, 1989; Sander, 1989; Cheng *et al.*, 2009; Xue *et al.*, 2015)—the observed dimorphism is comparable to sexual dimorphism in extant salamanders with amplexus (Motani *et al.*, 2015). For one species of pachypleurosaurs, sexual selection has been suggested based on male-biased sexual size dimorphism and female-biased adult sex ratio (Motani *et al.*, 2015). These studies involved sex estimation of individual specimens based on morphology, either by eyeballing the threshold between sexes (Sander, 1989; Cheng *et al.*, 2009; Xue *et al.*, 2015) or by an explicitly statistical

procedure (Motani *et al.*, 2015). The accuracy of the sex estimation, however, has never been tested.

In contrast to these aerial and aquatic Mesozoic reptiles, sexual dimorphism in dinosaurs remains murky despite the long-standing research interest since the first half of the 20th Century (Chapman *et al.*, 1997; Mallon, 2017; Saitta *et al.*, 2020). Many cases of sexual dimorphisms were suggested in the 1970s to 90s, especially for forms with cranial ornamentations, such as ceratopsians and some ornithopods, but also for others, such as theropods, as summarized by previous workers (Chapman *et al.*, 1997; Mallon, 2017; Saitta *et al.*, 2020). Probably the most famous case was made by Peter Dodson, who quantified the cranial morphology of *Protoceratops andrewsi* and detected sexual dimorphism with due considerations of allometry (Dodson, 1976). However, all of these studies have recently been rejected based on statistical analyses of new and previously published data (Maiorino *et al.*, 2015; Mallon, 2017), leaving no valid case of sexual dimorphism in dinosaurs. The inferred absence of sexual dimorphism in dinosaurs is puzzling given that sexual dimorphism is not a rare phenomenon in reptiles, as well as other vertebrates. A reassessment is due, especially given the questions raised earlier in this paper concerning methodology. There was a recent paper that tried to establish sexual size dimorphism in some dinosaurs (Saitta *et al.*, 2020) but it did not address the methodological issue.

Thanks to the recent movement to make research data freely available online, it is now possible to assess the questions raised in the previous paragraphs by using morphological data from extant species that are known to be sexually dimorphic. The purpose of the present paper is to test the following three hypotheses using the data: 1) Sexual dimorphism is not always detectable in principal components of all measurements because of the biases suggested above;

2) Individuals can be sexed with high accuracy based on sexual dimorphism under specific conditions; and 3) At least some dinosaurs were sexually dimorphic.

MATERIALS AND METHODS

I will first describe the outline of the data analysis procedure, which was designed to maximize the applicability to fossil species, for which the true sex of individual specimens is unknown. The data for each extant species were first screened for sample size, dimensional compatibility, and outliers, and then tested for dimorphism without using the information of the true sex of individual specimens. If dimorphism is detected, then the specimens were divided into two morphotypes based on the dimorphic traits discovered, again without using the knowledge of their true sex. Then, the resulting morphotypes were labeled with most likely sex identities and misclassification rate was calculated by comparing this inferred sex with the true sex. Finally, common statistical tests of sexual dimorphism, such as MANOVA based on all traits and ANOVA of size, were conducted based on both the true and inferred sexes to find the proportions of discrepancy in the outcome. Fossil species were treated by the same procedure but without the part that involves the true sex of the specimens. Details of each step of this procedure are given below.

Data Source

Data for this study were all derived from the literature. Published data of sexual dimorphism in living animals, containing morphological measurements for individuals, were located using online search tools, namely Dryad, Mendeley Data, and Google Data Search. After examining more than 100 published datasets, 16 publications were found to contain suitable data, encompassing 153 species. The data from these publications were first screened for sample

size—species with less than 10 individuals in total, as well as those with less than 4 individuals per sex, were removed. The remaining species were then tested for sexual dimorphisms using Multivariate Analysis of Variance (MANOVA) based on all morphological measurements, and species with at least weakly significant differences between the sexes ($p < 0.1$) were retained. This left 139 species in the data (Table 1), containing 53 arthropods (38 arachnids (Buzatto *et al.*, 2014), 4 crustacean (Fernandes Martins *et al.*, 2017; Sørdalen *et al.*, 2020), 7 insects (Punzalan & Rowe, 2015)), 1 cnidarian (González-Espinosa *et al.*, 2018), and 89 vertebrates (1 amphibian (De Lisle, Paiva, & Rowe, 2018), 2 birds (Hsu *et al.*, 2014; Poissant *et al.*, 2016), 6 mammals (Christiansen & Harris, 2012; Roseman *et al.*, 2020), 8 osteichthyes (Ronco, Roesti, & Salzburger, 2019; Garcia & Zuanon, 2019), and 72 reptiles (Sanger *et al.*, 2013; Massetti *et al.*, 2017; Burbrink, 2019)).

Data for four dinosaurs were found to have suitable sizes and completeness, namely *Allosaurus fragilis* (Smith, 1998), *Hesperosaurus mjosi* (Saitta, 2015), *Protoceratops andrewsi* (Maiorino *et al.*, 2015), and *Plateosaurus* sp. The data for *A. fragilis* and *Pl.* sp. were originally unpublished but were later published with reanalyses (Mallon, 2017). The femoral data for *A. fragilis* was kindly provided by David Smith, revealing mostly minor typos in the published version of the data. They have been corrected and included herein (Supplementary Table S1). Additionally, data for the sauropterygian *Keichousaurus hui* (Xue *et al.*, 2015) were reanalyzed.

Data Screening

All morphological measurements in the data were $\log_{10}$ transformed to address scaling effects. Specimens with missing values were removed—imputation was not used because such would add extra assumptions. Areas, if any, were converted to their square-roots. Most of the data contained only length measurements but lengths, areas, and angles were mixed in the data

for *Hesperosaurus* (Saitta, 2015). Angles were replaced by their length equivalents, i.e., their sin values multiplied by the length of the plate (i.e., approximately the plate height in situ)—this allows $\log_{10}$ transformation of the data while maintaining the information content. Also, when raw angles are mixed in with length, it tends to skew the principal component space (e.g., PC 1 is no longer a simple addition of all variables). The same dataset contained misplaced data—there is a typo to make one of the plates look as if it is 23 m long yet only 56 cm^2 in area but it is most likely 56 cm long and about 2300 cm^2 in area.

After the procedures above, outliers were identified and removed. A Q-Q plot of the square of the Mahalanobis distance (D^2) against expected Chi-squared value quantile was produced for each dataset, using the outlier function of the psych package of R (Revelle, 2019). Specimens with unusually large D^2 values, i.e., those outside of the third quartile plus 1.5 times the interquartile space, were removed. Similarly, specimens with unusually large difference between D^2 and the expected Chi-squared quartile were removed, based on the same quartile-based criterion. This process was iterated until there was no more outlier, although no iteration was needed in most species.

Test of Normality

Morphological data from sexually dimorphic species are probably not normally distributed, especially if dimorphism is strongly expressed. Nevertheless, we tested for the normality of the data to see if it has any ramification on the results of the analyses herein. Two tests were run—Shapiro Wilk test (Shapiro & Wilk, 1965) of univariate normality for each of the traits in the data, and Mardia's test of multivariate normality (Mardia, 1974) for each species, as implemented in the MVN package of R (Korkmaz, Goksuluk, & Zararsiz, 2014).

Test of Dimorphism

Dimorphism in each species was tested in four ways—three statistical tests and a visual inspection. The data were initially inspected visually. For each species, all combinations of variables were plotted in biplots to see if two separate clusters of data points, suggesting dimorphism, may be visible in any pair of variables. Species with two clusters, or morphotypes, were recorded, and notes were taken on whether the two clusters are completely separated from each other (i.e., there are two distinctive morphotypes) or partially overlapped (i.e., there are two morphotypes that are partly overlapped). If there is only one cluster in all biplots, the species was coded to have continuous female and male morphotypes. If two concentrations are largely different in sample sizes, the species was not recorded as dimorphic. The variable pair with the most distinctive separation of two clusters was recorded as the dimorphic variable pair (DVP). Ideally, the independent variable of DVP represents the size of individual animals, such as the body mass or total length.

The data were then statistically tested for dimorphism in three ways, using ACR tests of unimodality and bimodality, as well as the gap test. The gap test uses the difference between observed and expected log Residual Orthogonal Sum of Squared Distances to test for the number of groups in bivariate or multivariate data (Dunbar, 2018), whereas ACR test combines excess mass and smoothing approaches to tests for the number of modes in a univariate data with fewer errors than the commonly-used dip-test (Ameijeiras-Alonso, Crujeiras, & Rodríguez-Casal, 2019). ACR test was used to test two null hypotheses, of unimodality and bimodality of a metric in question, respectively. A metric was considered bimodal only when its unimodality is rejected while its bimodality is not rejected. Double testing was necessary because only one of them cannot establish bimodality for sure: when the number of mode(s) is ambiguous, it often happens

that both hypotheses are rejected together or neither hypothesis is rejected. The three statistical tests were run on data ordinated with Principal Component Analysis (PCA), as well as some raw metrics: PC1 and 2 from the entire data (PC_{All} 1 and PC_{All} 2), PC1 and 2 from DVP (PC_{DVP} 1 and PC_{DVP} 2), and the raw metrics of DVP, referred to as X_{DVP} and Y_{DVP} hereafter, with X_{DVP} being larger than Y_{DVP}. These six metrics can be obtained without the knowledge of sex identities of the specimens. In addition, two metrics that require the prior knowledge of sex were also tested, namely regression residuals from PC_{All} 1-PC_{All} 2 and X_{DVP}-Y_{DVP}—they will be referred to as RES_{PCS} and RES_{DVP}, respectively. The residuals were calculated relative to the common Standardized Major Axis (SMA) regression line between males and females. ACR test was run on each of these eight metrics, whereas the gap test was run on three pairs, PC_{All} 1-PC_{All} 2, PC_{DVP}1-PC_{DVP} 2, and X_{DVP}-Y_{DVP}.

Morphotype Establishment

For each species, samples were divided into two morphotypes based on the following statistical procedure. First, the metric that is most characteristically bimodal was selected based on the results of ACR tests. Then the samples were divided into two groups using k-means clustering of the metric identified—the two groups are recognized as two morphotypes. For comparisons, Linear Grouping Algorithm (LGA) (Van Aelst *et al.*, 2006) based on DVP was also used as an alternative method to divide the specimens into two groups, in case LGA may perform better than k-means clustering.

Sex Estimation

Once the specimens were divided into morphotypes, their sexes were inferred. For fossil species, the two morphotypes were labelled with sex based on analogies with the known sexual dimorphism in living species. For example, if one morphotype has a more expanded pelvic inlet

than the other, then it is most likely that it is female. The labelling process is unfortunately not quantitative but there is no alternative method at this point. For extant species, sexual dimorphism in each species is already known, so the information was used to label the morphotypes with sex identities.

Species were screened for inferred sex ratios. That is, if the inferred sex ratio is too skewed for one sex, it is possible that the statistical procedure was misled by discontinuity in data that does not reflect sex, e.g., uneven sampling across size due to a small sample size. An arbitrary ratio of 1:3 was used as the threshold to remove species with strongly skewed inferred sex ratios. Whereas sex ratios further skewed than 1:3 may exist in nature, experimentations with the data at hand revealed that the proportion of misleading data is sufficiently high to render this screening worthwhile.

True versus Inferred Sex

For living species, it is possible to quantify how accurately the true sexes of the specimens were re-classified from morphology. This was done by making a contingency table between the true and inferred sexes for each species, and calculating the proportion of misclassification.

The effects of using inferred instead of the true sex in tests of sexual dimorphism were also examined, so as to illuminate the possibilities of using inferred sex in calculations of some statistics related to sexual dimorphism in fossil species. For example, MANOVA of all variables, which would test for overall sexual dimorphism, was performed with the true as well as inferred sex and the results were compared to see if any false positives or negatives were produced at $p < 0.1$ and $p < 0.05$. Similarly, the bias from the use of inferred sex was examined with Analysis of Variance (ANOVA) of the geometric means of all variables, testing for sexual size dimorphism,

as well as Analysis of Covariance (ANCOVA) of DVP, capturing some cases of sexual shape dimorphism. Lastly, ACR tests of RES_{DVP} and RES_{PCS}, respectively, were also performed with the true and inferred sex to find the degree of discrepancy.

RESULTS

All statistics appearing below are calculated after excluding 11 species whose inferred sex ratios are strongly skewed, as defined in Methods. Also, the numbers reflect the results at $p < 0.1$, unless otherwise noted. A complete set of test statistics, including those for fossil species, are given in Supplementary Table S2.

Test of Normality

Only 40 out of the 139 species are considered normal by both univariate and multivariate normality tests. Most of them are only weakly dimorphic—only 3 of the 40 are dimorphic according to the four tests of dimorphism. In contrast, strongly dimorphic species tend to perform poorly in one or both of the tests while, as reported below, performing best in the tests of bimodality and sex estimation from morphology. Therefore, the present statistical procedures seem to be robust against violation of normality. The outcome of the tests for each species can be found in Supplementary Table S2.

Test of Dimorphism

Inspection of biplots suggests that 46 species have at least one biplot in which two morphotypes are visible. The morphotypes are completely distinctive in 33 of them whereas two morphotypes are largely separated but with some overlaps in the remaining 13. ACR tests reveal that 103 species have at least one metric that is bimodal out of the six metrics examined, while gap test suggests that 74 species have two or more morphotypes in at least one of the three pairs of

metrics examined. When combining the results of ACR and gap tests, 69 species have at least two morphotypes in both tests. Therefore, ACR and gap tests are more forgiving overall in recognizing dimorphism than visual inspection. See Table 2 for counts at $p < 0.05$.

Morphotype Establishment

There are many ways in which dimorphism is expressed through the metrics examined, while simple morphospace as in Fig. 1C-E is never found. The various types of dimorphism expressions can be summarized into five types based on which metric is most useful in establishing morphotypes (Table 3). Types 1 and 3 reflect shape dimorphism whereas Types 4 and 5 are for cases in which size dimorphism dominates. Type 2 is known in both shape (Fig. 2E-F) and size (Fig. 2G-H) dimorphisms. Type 1, in which PC_{DVP} 2 defines morphotypes (Fig. 2A-D), is the most common type for shape dimorphism whereas Type 4, in which PC_{All} 1 defines morphotypes (Fig. 2K-N), is the most common expression of size dimorphism. There is a rare variant of Type 1, called Type 1' in Table 3, in which PC_{DVP} 2 does not appear bimodal (Fig. 2Q-T), most likely because of the limitation of ACR test. In such cases, bimodality can be established only through RES_{PCS} and RES_{DVP} while PC_{DVP} 2 may still be useful in morphotype establishment. Type 2, in which PC_{DVP} 1 defines morphotypes (Fig. 2E-H), is seen when two morphotypes are seen along the major axis regression line. Type 3, in which PC_{All} 2 defines morphotypes (Fig. 2I-J), is seen when signals from shape dimorphism rules the variation in PC_{All} 2. One example is when there is less noise to obscure the signals because the number of variables is small. Type 5, in which X_{DVP} 1 defines morphotypes (Fig. 2O), is a rare extension of Type 4, in which PC_{All} 1 is not bimodal despite the presence of size dimorphism because the principal component space is tilted—see Discussion. Also see Discussion for the sixth type of dimorphism expression.

True versus Inferred Sex

When comparing the true versus inferred sex, the misclassification rate varies depending on how clearly separated the morphotypes are. The average misclassification rates is lowest at 0.027 for the 33 species for which two separate morphotypes are visible in a biplot. The rate becomes slightly higher at 0.038 when including species for which two morphotypes are visible in a biplot but with a partial overlap (n=46). In contrast, 69 species that are considered bimodal by both ACR and gap tests have a much higher average misclassification rate of 0.11, so the forgivingness of these methods over visual inspection comes at a cost. Misclassification rates under various scenarios are listed in Table 2. When using an alternative method (LGA) to divide the specimens into morphotypes, the average misclassification rate worsens greatly on average (0.23). However, LGA outperforms the present method in three species with large sample sizes (> ~500), dimorphism type 1 or 2, and two slightly overlapping morphotypes in a biplot, namely *Homarus gammarus* (Fig. 2C-D), *Mus musculus* (Fig. 2E-F) and *Serracutisoma proximum*.

When using inferred sex instead of the true sex in statistical tests, most of the test outcomes do not change drastically at $p < 0.05$ except in ANCOVA of DVP (Supplementary Tables S3). There are no false positives or negative for MANOVA of all variables in any of the 46 species that visually appear dimorphic, or among the 69 that are considered dimorphic by both ACR and gap tests. For ANOVA of the geometric means of all variables, no false positives or negatives are found among the 46 species but 7 false negatives are found among the 69 species, without any false positives. For ACR test of RES_{DVP}, 4 false positive and 0 false negatives result from the 46 species whereas the numbers are 8 and 5, respectively, for the 69 species. Therefore, as long as dimorphism is well defined (i.e., 46 out of 139 species), inferred sex provides the same results as the true sex in MANOVA of all variables, ANOVA of size, and

ACR test of RES_{DVP}, although a small proportion of false positives may be expected in the last statistic

Dimorphism in Fossil Reptiles

The results of statistical analyses for fossil species are summarized in Table 4. In four of the five species examined, namely *Allosaurus fragilis*, *Hesperosaurus mjosi*, *Protoceratops andrewsi*, and *Keichousaurus hui*, two morphotypes are visible in biplots (Fig. 3), although the two may overlap in parts. These species also pass both ACR and gap tests. Therefore, they are dimorphic. Dimorphism is of Type 1 for the first two and Type 6 for the rest (see Discussion for Type 6). See below for discussions of whether these dimorphisms are of sexual nature. The remaining species, *Plateosaurus* sp., exhibits weakest signs of dimorphism—two groups are not clearly seen in biplots, whereas bimodality is not recognized in any of the metrics, although the gap test finds two groups.

DISCUSSION

The results of the present analyses do not reject any of the three hypotheses proposed earlier in this paper. Thus: individual principal components of all measurements often fail to clearly exhibit sexual dimorphism; high-accuracy sex estimation from morphology alone is possible under specific conditions; and at least some dinosaurs were sexually dimorphic. At the same time, the results reiterate the overall difficulty of identifying sexual dimorphism from morphology alone (Mallon, 2017; Hone *et al.*, 2020). Only up to about 50% of the 139 species could be recognized as being dimorphic when sex identities of the specimens are unknown, whereas all 139 species are sexually dimorphic according to MANOVA, which benefits from the knowledge of the sex of individual specimens. This in turn suggests that a species may still be sexually

dimorphic even if statistical tests of bimodality fail to establish it based on morphological data. Sex estimation from morphology alone is also difficult—only about 33% of the 139 species examined could be sexed from morphology with the high accuracies, i.e., e misclassification rates below 0.05, and the ratio increases to 37% if slightly less accurate sexing results are combined (i.e., misclassification rates below 0.1). At the same time, sex may be inferred with high accuracies as long as sexual dimorphism in a species is well-defined, i.e., the mean misclassification rare was 0.03 for the 46 species in which dimorphism is visible in a biplot of traits.

The assumption that sexual dimorphism is detected as bimodality of the individual principal components, especially PC2, from the total data was revealed to be false most of the time. Only 13 out of the 139 species examined in this study showed multimodal PC_{All} 2 according to ACR test at $p<0.05$, so over 90% of the sexually dimorphic species would appear unimodal by examining only PC_{All} 2. When using the dip-test instead, as done by a previous worker (Mallon, 2017), only 7 out of 139 would be recognized as having multimodal PC_{All} 2. The main reason why this assumption did not hold is that the actual principal components do not resemble Fig. 1C but are tilted groupwise as in Fig. 1D and Fig. 4. This tilting has multiple causes. A low signal/noise ratio is one of the causes, as seen by comparing Fig. 4A-B to C-D, E-F to G-H, and I-J to K-L. In all three cases, the tilting of the groupwise trend becomes less by removing characters that are not dimorphic (i.e., noise), and the bimodality of PC_{All} 2 increases accordingly. However, the signal/noise ratio is not the only controlling factor as seen in Fig. 4H—females and males are still left with trends after removing excess data. This is most likely because of the degree of size dimorphism that co-exists with shape dimorphism—when comparing the three cases for the degree of sexual size dimorphism along PC_{All} 1, Fig. 4A and 4I

exhibit limited sexual size dimorphism unlike Fig. 4E, so the mixture of size and shape dimorphism likely led to the observed tilting. Additionally, outliers, if present, would add groupwise trends in principal components space, although they are not relevant to Fig. 4 because they had been removed in this study.

The best practice of sex estimation from morphology would follow the steps below. First, biplots of all combinations of variables are made and DVP is identified. If there are many variables that are not dimorphic, consider removing a majority of them to increase the signal/noise ratio but leave at least one size variable in the data, if any. Then, ACR tests of unimodality and bimodality, as well as the gap test, are run on six metrics, PC_{All} 1, PC_{All} 2, PC_{DVP} 1, PC_{DVP} 2, X_{DVP}, and Y_{DVP}. Compare the outcome of these tests to Table 3 to identify the type of dimorphism seen in the species. Once the type is known, use k-means clustering of the metric suggested in Table 3 to separate specimens into two morphotypes. However, when the sample size is large (> ~500) and the dimorphism is of Type 1 or 2, LGA of DVP should be used instead of the k-means approach to separate specimens into morphotypes. The resulting morphotypes are then labeled with sex identities based on the analogy from the knowledge of female and male morphologies in living species. If labeling is impossible, then the observed dimorphism may not be of sexual nature, and cause of the observed dimorphism needs to be sought outside of sexual dimorphism. Also, if the resulting sex ratio is highly skewed (e.g., more extreme than, say, 1:3), then the observed dimorphism may also not be of sexual origin, unless there is an independent reason to suggest a highly skewed sex ratio. A highly skewed sex ratios may exist in nature but are more often derived from insufficient sampling, so a careful consideration is necessary. Sex estimation is expected to be of high accuracy if two morphotypes are largely separated in any of the biplots (average misclassification rate < 0.05 in the species

examined in this study). Also, MANOVA based on the sexes inferred this way is probably trustworthy given that it always gave the same results as the one based on the true sex in this study. ANOVA of sexual size dimorphism based on inferred sex is trustworthy as long as dimorphism is visible in a biplot of DVP.

Sample size is an important factor. There is a tendency for both ACR and dip-tests of unimodality to result in excessive false positives (i.e., unimodality is accepted even when the sample was drawn from a bimodal distribution) at smaller sample sizes. For example, when randomly drawing samples from arbitrary bimodal distributions as in Fig. 5A-C, respectively, these samples are considered unimodal by ACR test half of the time when the size of each sample is at n=20 even if the population distribution is symmetrical (i.e., sample size and standard deviations are the same between the two modes)(Fig. 5G). A larger sample size of about n=50 is required to reduce the proportion of false positives to about 25%. The numbers are even larger if the population distribution is asymmetrical (Fig. 5H-I), and dip-test performs further poorly at small samples sizes (Fig. 5D-F). When the sample size is greater than about 100, ACR test seems to perform well against weak asymmetry of population distribution as in Fig. 5B-C but dip-test would require about twice the sample size. Therefore, it is best to have a large sample size. At the same time, some of the extant species with samples size as small as about 20 were considered dimorphic by ACR and gap tests while dimorphism is clearly visible in a biplot of traits. Therefore, small sample sizes are tolerated as long as there is a strong signal of dimorphism in the data.

It is necessary to discuss two exceptions from the best practice outlined above. First, apart from the five types of dimorphism expression listed earlier, there is a sixth type that was not present in the 139 species examined, yet is known in fossil reptiles, i.e., Type 6 of Table 3. In

Type 6, many variables are dimorphic when plotted against size, so it is unfair to pick just one pair of DVP on which to base morphotype identification. In that case, it is best to use all principal components from the entire data except PC_{All} 1 to divide the sample into two morphotypes—this procedure was previously proposed for *Keichousaurus hui* (Motani *et al.*, 2015). Second, it is sometimes possible to inferre sex with a reasonably high accuracy even when two morphotypes are not visually distinguishable in biplots. Many of these examples are from cases with large sample sizes (>200). For example, in *Parus major* (n=2372 after cleaning; Fig. 2G-H) (Poissant *et al.*, 2016) and *Panthera leo* (n=230; Fig. 2O-P) (Christiansen & Harris, 2012), it is difficult to recognize two morphotypes in biplots without prior knowledge of sex but ACR tests suggested that they were dimorphic.

The previous study by Mallon reached a more pessimistic conclusion about the possibility of identifying sexual dimorphism from morphology alone (Mallon, 2017). The discrepancy arises from the methodological differences. As discussed earlier, it was previously assumed that sexual dimorphism should be visible as bimodality of some principal components from all data, especially PC_{All} 2 (Mallon, 2017), but sexual dimorphism may indeed be seen in any of PC_{All} 1, PC_{All} 2, PC_{DVP} 1, PC_{DVP} 2, X_{DVP}, Y_{DVP}, RES_{All}, or RES_{DVP}. Another methodological difference is in data treatment. Mallon used the published data on dinosaurs almost as is, while augmenting missing data through imputation. In contrast, outliers were removed through a statistical procedure, all data were converted to length-equivalents and log transformed to account for scaling effects, and specimens with missing data were removed without imputation in this study.

The observed dimorphism in *Allosaurus fragilis* is most strongly expressed when plotting the width of the femoral head against the femoral length (Fig. 3A-B). The presence of

dimorphism was previously noticed based on a plot of the femoral head width against the femoral width at the lesser trochanter (Smith, 1998), but the separation is less clear at larger sizes. Sexual dimorphism of the femoral head is not well investigated in extant reptiles but it is at least known in humans, in which wide heads belong to males (Papaloucas, Fiska, & Demetriou, 2008; Clavero, Salicrú, & Turbón, 2015). Given that humans and *A. fragilis* are both bipedal, the femoral heads of the two probably share at least a part of their mechanical roles. Therefore, it is reasonable to consider the dimorphism in the femur of *A. fragilis* an example of sexual dimorphism, as originally suggested (Smith, 1998), pending future examination of femoral head dimorphism in extant reptiles, including birds. Specimen counts also support the interpretation—there are 17 suspected females and 14 suspected males, so the inferred sex ratio is not unusually skewed. Mallon did not find sexual dimorphism in this species (Mallon, 2017) but that is partly because he did not clean the data for outliers—experimentation showed that, with outliers, the principal component space is trended groupwise and $PC2_{All}$ appears unimodal but the trend lessened greatly when removing the outliers.

A possible caveat to the argument above is the age of the specimens—at least about half of the specimens in the data of *Allosaurus fragilis* are expected to be sexually immature (Lee & Werning, 2008). The presence of sexual dimorphism outside of primary sexual characters in juveniles may appear contradictory to the commonly held notion of secondary sexual characters, that these characters are not present at birth but developed later in life. This notion goes back at least to John Hunter, who stated "…in animals just born, or very young, there are no peculiarities to distinguish one sex from the other, exclusive of what relates to the organ of generation" and "…toward the age of maturity the discriminating changes before mentioned begin to appear" (Hunter, 1837). However, it is known today that some secondary sexual characters are already

present at birth, or in childhood. For example, male chickens have more muscular mass than female chickens even at hatching (Rose, Nudds, & Codd, 2016), whereas *Keichousaurus* males already have wider humeral end than females at newborn size (Motani *et al.*, 2015). Even in humans, the vertebral cross-sectional area is greater in males at birth (Ponrartana *et al.*, 2015) while school-aged children (6.2 years old on average) already have bone sexual dimorphism whereby male bones are stronger (Medina-Gomez *et al.*, 2016). Therefore, it is not strange to detect sexual dimorphism of secondary sexual nature in immature individuals of *A. fragilis*, or in other dinosaurs, e.g., *Protoceratops andrewsi* whose data also contained juveniles (Maiorino *et al.*, 2015).

The case for the stegosaurid *Hesperosaurus mjosi* is more complicated. Dimorphism in this species was first examined by Saitta et al. (Saitta, 2015), who recognized round and elongated morphotypes in the plates on the back They interpreted it as sexual dimorphism related to display purposes, with the round plates belonging to males. The present study largely supports the original suggestions by Saitta et al. (Saitta, 2015)—two morphotypes are most clearly present in the width of the plate relative to the distance from the tip to the center of the base, i.e., whether the plate is broad or narrow (Fig. 3C-D). There are 21 narrow and 15 broad plates in the data after outliers were removed. The morphotypes are mostly specimen dependent—seven specimens have slender plates whereas six have round plates, with one plate each of three of the specimens being of the other type. The ratio of 7:6 is not unusual for a sex ratio. Therefore, the probability of this dimorphism being linked to sex, as suggested bySaitta (2015), cannot be rejected.

A previous reexamination the plate dimorphism in *Hesperosaurus mjosi* based on the data of Saitta (2015) did not find sexual dimorphism unlike the present study (Mallon, 2017).

The conclusion, however, was likely compromised by a methodological issue. The analysis was based on a mixture of angles, areas, and lengths without log transformation, and consequently had a skewed principal component space where PC_{All} 1 does not represent size while angles dominated PC_{All} 2. This led to a conclusion that dimorphism of the plates was because of the variations in the angle between the apex and base (Mallon, 2017). It had already been shown that this angle depended on the part of the body rather than individuals (Saitta, 2015), and that led Mallon to conclude that the plate dimorphism in *H. mjosi* was not of a sexual nature. When the measurements are converted to their length-equivalents when necessary and log transformed to account for scaling effects, and statistical outliers are removed, the angle between the apex and base is no longer the only ruling factor behind plate dimorphism because the ratio between the major and minor axes becomes slightly more dominant in PC_{All} 2. Thus, plate dimorphism in *H. mjosi* is in the aspect ratios as characterized by the original authors (Saitta, 2015) and therefore largely depends on individuals.

The observed dimorphism in *Protoceratops andrewsi* is most clearly seen in the width of the external naris relative to the nasal height of the skull, but also in many other metrics, such as the length and height of the cranial frill. One morphotype has narrower external nares than the other morphotype, while also having longer and higher frills for size (Fig. 3E-F), while width measurements did not covary greatly with this dimorphism. These combinations of traits approximately match the morphotypes originally proposed (Dodson, 1976), whereby the morphotype with large frills and narrow external nares was considered male. Whereas it is difficult to find the reasons why the narrowness of the nares may be a sexual character, enlarged frills would serve a display purpose as Dodson pointed out. It was previously suggested that ceratopsian frills facilitated species recognition rather than sexual appeals (Padian & Horner,

2011). However, whereas the role of the frill in species recognition is plausible, the inferred dimorphism in frill shape cannot be explained by species recognition alone. Then, there is insufficient reasons to reject Dodson's suggestion that the observed dimorphs are of sexual nature. At the same time, the case for sexual dimorphism for *P. andrewsi* is the weakest of the three cases of dinosaurian sexual dimorphism suggested in this paper. For example, the sex ratio inferred in the present study is strongly male-biased, with 6 females and 12 males. However, this could be due to sampling biases.

The possibility that the observed dimorphism is an artefact of taphonomic deformation of fossils deserves a discussion. It is unlikely that the presumed male morph is an artefact of deformation because deformation that narrows the nose while also heightening the nose horn and frill and lengthening the frill would be non-linear while involving extension, and it is unlikely that many specimens experienced such complicated deformation in the same manner. Some of the features of the postulated female morph could be the results of dorsoventral compaction of the skull—a simple compaction in that orientation would lessen the height of the nose horn and frill. However, such deformation would not widen the nares or shorten the frill in dorsal view. Therefore, it would again take non-linear deformation to turn a male morph into a female morph, although the deformation would not involve much extension this time. Then, it is unlikely that such complicated deformation was shared among many specimens. Therefore, it is unlikely at this point that the observed dimorphism is an artifact of taphonomic deformation, although further studies would be necessary to scrutinize this point.

Sexual dimorphism in *Plateosaurus* sp. remains unclear. Given that the gap test detected two morphotypesand that ACR tests are often misleading at small sample sizes, the species was probably sexually dimorphic. However, there is currently insufficient data to firmly establish

sexual dimorphism in this species. If the sample size increases in the future (e.g., tripled), a more robust conclusion may be reached.

The presence of sexual dimorphism in *Keichousaurus hui* and other pachypleurosaurs was never tested in terms of p-values, and the accuracies of sex estimations in these species were never explored, despite the three decades of research (Rieppel, 1989; Sander, 1989; Cheng *et al.*, 2009; Motani *et al.*, 2015; Xue *et al.*, 2015). The present study has shown that at least *K. hui* was dimorphic based on the gap and ACR tests, showing for the first time that the observed dimorphism is statistically significant without the knowledge of the sex of individual specimens. Given that dimorphism is seen in the results of these tests as well as from visual inspections of biplots, it is likely that sex estimation from dimorphic traits of this species is of high accuracy. The observed dimorphs exhibit features similar to what is seen in sexually dimorphic salamanders in which males use elongated limbs to hold females during copulation (Motani *et al.*, 2015), so it is likely that the observed dimorphism is of sexual nature—*K. hui* was a small, amphibious reptile. The same is expected in other pachypleurosaurs but the lack of published data prevents further examination of these other species.

As previous studies pointed out (Mallon, 2017), sample size is usually the most restricting factor when studying sexual dimorphisms in fossil reptiles. Of the fossil species examined, only *Keichousaurus hui* has a sample size greater than 50 (Supplementary Table S2). Smaller sample sizes may be tolerated if the signals for dimorphism is strong, as discussed above. Therefore, interpretations given here for dinosaurs are plausible at least tentatively, although they should be retested if large sample sizes become available in the future. Samples of fossil reptiles also suffer from inclusion of temporal variations that may be mistaken as sexual dimorphism (Evans & Reisz, 2007), unless all specimens in a sample are coeval. Large, coeval, and

approximately sympatric samples may become available in the future from bone beds with limited taxonomic diversities, although deformed or fragmentary specimens need to be carefully removed from the sample.

CONCLUSIONS

1. Sexual dimorphism could be established without knowledge of the sex of individual specimens in about 50% of the 139 sexually dimorphic species examined.
2. Specimens could be sexed from morphology alone in about a third of the 139 sexually dimorphic species with accuracies higher than 95%, and these species are recognizable prior to sex estimation from a combination of statistical metrics (i.e., results from the gap and ACR tests) and visual recognition of dimorphism in biplots of traits.
3. Sexual dimorphism is expressed in various combinations of six statistical metrics, PC_{All} 1, PC_{All} 2, PC_{DVP} 1, PC_{DVP} 2, X_{DVP}, and Y_{DVP}, depending on mixtures of size and shape dimorphism as well as signal/noise ratio in the data. A previous study of dinosaurian sexual dimorphism examined only one of the six and consequently found no case for sexual dimorphism.
4. Test of sexual dimorphism in fossil species is difficult because the sex of individual specimens is unknown, preventing the use of common statistical methods such as MANOVA and ANOVA. Yet it is still possible to establish sexual dimorphism in fossil species if they pass all the tests that are stated in item 2 above.
5. Dimorphism can be established in at least three species of dinosaurs and a sauropterygian, namely *Allosaurus fragilis*, *Hesperosaurus mjosi*, *Protoceratops andrewsi*, and *Keichousaurus hui*, based on the six statistical metrics listed in 3. For these species, it is expected that specimens can be sexed from morphology with high accuracies.

ACKNOWLEDGMENTS

David Smith kindly provided a printout version of the data for *Allosaurus fragilis* from his 1998 publication. Geerat J. Vermeij read an earlier version of the manuscript and provided suggestions. Colin Boisvert obtained the data from David Smith. James Farlow and an anonymous reviewer provided useful suggestions.

REFERENCES

Van Aelst S, Wang X, Zamar RH, Zhu R. 2006. Linear grouping using orthogonal regression. *Computational Statistics and Data Analysis* 50: 1287–1312.

Ameijeiras-Alonso J, Crujeiras RM, Rodríguez-Casal A. 2019. Mode testing, critical bandwidth and excess mass. *TEST* 28: 900–919.

Bennett SC. 1992. Sexual dimorphism of *Pteranodon* and other pterosaurs, with comments on cranial crests. *Journal of Vertebrate Paleontology* 12: 422–434.

Burbrink FT. 2019. Female - biased gape and body - size dimorphism in the New World watersnakes (tribe : Thamnophiini) oppose predictions from Rensch ' s rule. : 9624–9633.

Buzatto BA, Tomkins JL, Simmons LW, Machado G. 2014. Correlated evolution of sexual dimorphism and male dimorphism in a clade of neotropical harvestmen. *Evolution* 68: 1671–1686.

Chapman RE, Weishampel DB, Hunt G, Rasskin-Gutman D. 1997. Sexual dimorphism in dinosaurs. *Dinofest International Proceedings* 1: 83–93.

Cheng YNN, Holmes R, Wu XCC, Alfonso N. 2009. Sexual dimorphism and life history of *Keichousaurus hui* (Reptilia: Sauropterygia). *Journal of Vertebrate Paleontology* 29: 401–408.

Christiansen P, Harris JM. 2012. Variation in craniomandibular morphology and sexual

dimorphism in pantherines and the sabercat *Smilodon fatalis* (AA Farke, Ed.). *PLoS ONE* 7: e48352.

Clavero A, Salicrú M, Turbón D. 2015. Sex prediction from the femur and hip bone using a sample of CT images from a Spanish population. *International Journal of Legal Medicine* 129: 373–383.

Darwin C. 1859. *On the Origin Of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: J. Murray.

Darwin C. 1871. *The Descent of Man, and Selection in Relation to Sex*. London: J. Murray.

Dodson P. 1976. Quantitative aspects of relative growth and sexual dimorphism in *Protoceratops*. *Journal of Paleontology* 50: 929–940.

Dunbar RIM. 2018. The Anatomy of. *Trends in Cognitive Sciences* 22: 32–51.

Evans DC, Reisz RR. 2007. Anatomy and relationships of *Lambeosaurus magnicristatus*, a crested hadrosaurid dinosaur (Ornithischia) from the Dinosaur Park Formation, Alberta. *Journal of Vertebrate Paleontology* 27: 373–393.

Fairbairn DJ. 1997. Allometry for sexual size dimorphism: Pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics* 28: 659–687.

Fairbairn DJ, Blanckenhorn WU, Székely T. 2007. *Sex, Size and Gender Roles*. Oxford University Press.

Fernandes Martins MJ, Hunt G, Lockwood R, Swaddle JP, Horne DJ. 2017. Correlation between investment in sexual traits and valve sexual dimorphism in *Cyprideis* species (Ostracoda) (P Raia, Ed.). *PLOS ONE* 12: e0177791.

Garcia EQ, Zuanon J. 2019. Sexual dimorphism in the electric knifefish, *Gymnorhamphichthys*

rondoni (Rhamphichthyidae: Gymnotiformes). *Acta Amazonica* 49: 213–220.

González-Espinosa PC, Paz-García DA, Reyes-Bonilla H, Cabral-Tena RA, Balart EF. 2018. Evidence of sexual dimorphism in skeletal morphology of a gonochoric reef coral. *Royal Society Open Science* 5: 171843.

Hone D, Mallon JC, Hennessey P, Witmer LM. 2020. Ontogeny of a sexually selected structure in an extant archosaur *Gavialis gangeticus* (Pseudosuchia: Crocodylia) with implications for sexual dimorphism in dinosaurs. *PeerJ* 2020.

Hsu Y cheng, Shaner P jen, Chang C i, Ke L, Kao S ji. 2014. Trophic niche width increases with bill-size variation in a generalist passerine: a test of niche variation hypothesis (S Meiri, Ed.). *Journal of Animal Ecology* 83: 450–459.

Hunter J. 1837. An account of an extraordinary phesant. In: Palmer JF, ed. *The Works of John Hunter, F. R. S. Vol. IV*. London: Longman, Rees, Orme, Brown, Green, and Longman, 44–49.

Korkmaz S, Goksuluk D, Zararsiz G. 2014. MVN: An R package for assessing multivariate normality. *R Journal* 6: 151–162.

Lee AH, Werning S. 2008. Sexual maturity in growing dinosaurs does not fit reptilian growth models. *Proceedings of the National Academy of Sciences of the United States of America* 105: 582–587.

De Lisle SP, Paiva S, Rowe L. 2018. Habitat partitioning during character displacement between the sexes. *Biology Letters* 14: 20180124.

Maiorino L, Farke AA, Kotsakis T, Piras P. 2015. Males resemble females: re-evaluating sexual dimorphism in *Protoceratops andrewsi* (Neoceratopsia, Protoceratopsidae) (MC Mihlbachler, Ed.). *PLOS ONE* 10: e0126464.

Mallon JC. 2017. Recognizing sexual dimorphism in the fossil record: lessons from nonavian

dinosaurs. *Paleobiology* 43: 495–507.

Mardia K V. 1974. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics* 36: 115–128.

Massetti F, Gomes V, Perera ANA, Rato C, Kaliontzopoulou A. 2017. Morphological and functional implications of sexual size dimorphism in the Moorish gecko, *Tarentola mauritanica*. *Biological Journal of the Linnean Society* 122: 197–209.

Medina-Gomez C, Heppe DHM, Yin JL, Trajanoska K, Uitterlinden AG, Beck TJ, Jaddoe VWV, Rivadeneira F. 2016. Bone mass and strength in school-age children exhibit sexual dimorphism related to differences in lean mass: the generation R study. *Journal of Bone and Mineral Research* 31: 1099–1106.

Motani R, Huang JJD, Jiang DY yong, Tintori A, Rieppel O, You HLH, Hu YCYC chao, Zhang R. 2018. Separating sexual dimorphism from other morphological variation in a specimen complex of fossil marine reptiles (Reptilia, Ichthyosauriformes, *Chaohusaurus*). *Scientific Reports* 8: 1–14.

Motani R, Jiang DY, Rieppel OC, Xue YF, Tintori A. 2015. Adult sex ratio, sexual dimorphism and sexual selection in a Mesozoic reptile. *Proceedings of the Royal Society B-Biological Sciences* 282: 20151658.

Padian K, Horner JR. 2011. The evolution of ‘bizarre structures’ in dinosaurs: Biomechanics, sexual selection, social selection or species recognition? *Journal of Zoology* 283: 3–17.

Papaloucas C, Fiska A, Demetriou T. 2008. Sexual dimorphism of the hip joint in Greeks. *Forensic Science International* 179: 83.e1-83.e3.

Poissant J, Morrissey MB, Gosler AG, Slate J, Sheldon BC. 2016. Multivariate selection and intersexual genetic constraints in a wild bird population. 29: 2022–2035.

Ponrartana S, Aggabao PC, Dharmavaram NL, Fisher CL, Friedlich P, Devaskar SU, Gilsanz V. 2015. Sexual dimorphism in newborn vertebrae and its potential implications. *The Journal of Pediatrics* 167: 416–421.

Punzalan D, Rowe L. 2015. Evolution of sexual dimorphism in phenotypic covariance structure in Phymata. *Evolution* 69: 1597–1609.

Revelle W. 2019. psych: Procedures for Psychological, Psychometric, and Personality Research. R Package.

Rieppel O. 1989. A new pachypleurosaur (Reptilia: Sauropterygia) from the Middle Triassic of Monte San Giorgio, Switzerland. *Philosophical Transactions - Royal Society of London, Series B* 323: 1–73.

Ronco F, Roesti M, Salzburger W. 2019. A functional trade-off between trophic adaptation and parental care predicts sexual dimorphism in cichlid fish. *Proceedings of the Royal Society B: Biological Sciences* 286: 20191050.

Rose KA, Nudds RL, Codd JR. 2016. Variety, sex and ontogenetic differences in the pelvic limb muscle architectural properties of leghorn chickens (*Gallus gallus domesticus*) and their links with locomotor performance. *Journal of Anatomy* 228: 952–964.

Roseman CC, Capellini TD, Jagoda E, Williams SA, Grabowski M, O'Connor C, Polk JD, Cheverud JM. 2020. Variation in mouse pelvic morphology maps to locations enriched in Sox9 Class II and Pitx1 regulatory features. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 334: 100–112.

Saitta ET. 2015. Evidence for sexual dimorphism in the plated dinosaur *Stegosaurus mjosi* (Ornithischia, Stegosauria) from the Morrison Formation (Upper Jurassic) of western USA. *PLoS ONE* 10: 1–20.

Saitta ET, Stockdale MT, Longrich NR, Bonhomme V, Benton MJ, Cuthill IC, Makovicky PJ. 2020. An effect size statistical framework for investigating sexual dimorphism in non-avian dinosaurs and other extinct taxa. : 1–43.

Sander PM. 1989. The pachypleurosaurids (Reptilia, Nothosauria) from the Middle Triassic of Monte-San-Giorgio (Switzerland) with the description of a new species. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 325: 563–666.

Sanger TJ, Sherratt E, McGlothlin JW, Brodie ED, Losos JB, Abzhanov A. 2013. Convergent evolution of sexual dimorphism in skull shape using distinct developmental strategies. *Evolution* 67: 2180–2193.

Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality. *Biometrika* 52: 591–611.

Smith DK. 1998. A morphometric analysis of *Allosaurus*. *Journal of Vertebrate Paleontology* 18: 126–142.

Sokal RR, Rohlf FJ. 2012. *Biometry : the principles and practice of statistics in biological research*. New York: W.H. Freeman.

Sørdalen TK, Halvorsen KT, Vøllestad LA, Moland E, Olsen EM. 2020. Marine protected areas rescue a sexually selected trait in European lobster. *Evolutionary Applications*: eva.12992.

Xue YF, Jiang DY, Motani R, Rieppel O, Sun YL, Sun ZY, Ji C, Yang PF. 2015. New information on sexual dimorphism and allometric growth in *Keichousaurus hui*, a pachypleurosaur from the middle triassic of Guizhou, South China. *Acta Palaeontologica Polonica* 60: 681–687.

Figure 1. Schematic plots a sexually dimorphic sample to explain basic notions. A, a biplot showing the difference between the total and sex-wise regression lines. B, a biplot of a case in which PC1 and 2 are correlated when viewed sex-wise. C, an idealized case of sexual size dimorphism without shape dimorphism. D, an idealized case of sexual shape dimorphism without size dimorphism. E, an idealized mixture of sexual size and shape dimorphism. F, a typically observed mixture of sexual size and shape dimorphism. Blue and red indicate two sexes (specific sex is irrelevant in these theoretical plots).

Figure 2. Five types of dimorphism expression observed, with exemplar species. For each species, a biplot of DVP and a histogram of the most bimodal metric is given, except for *Cyperideis torosa*, for which two additional plots, of PC_{All} 1-2 and PC_{DVP} 1-2, are also provided to demonstrate sex-wise correlations in these variable pairs. A-D, type 1; E-H, type 2; I-J, type 3; K-N, type 4; O-P, type 5, Q-T, type 1'. Species are: A-B, *Seminatrix pygae*; C-D, *Homarus gammarus*; E-F, *Mus musculus*; G-H, *Parus major*; I-J, *Gymnorhamphichthys rondoni*; K-L, *Anolis hendersoni*; M-N, *Longiperna concolor*; O-P, *Panthera leo*; Q-T, *Cyperideis torosa*. Blue is male and red is female (true sex).

Figure 3. Biplot of DVP and histogram of the most bimodal metric in dinosaurs. A-B, *Allosaurus fragilis*; C-D, *Hesperosaurus mjosi*, E-F, *Protoceratops andrewsi*; G-H, *Plateosaurus* sp. Blue is male and red is female (inferred sex).

Figure 4. Biases in the total versus dimorphic part of the data in three exemplar species. Biases, seen as sex-wise correlation of principal component space, is less when analyzing the dimorphic

part of the data only as opposed to the total data, likely because the signal/noise ratio is higher in the former A-D, *Seminatrix pygae*; E-H, *Cyperideis torosa*; I-L, *Mus musculus*. Blue is male and red is female (true sex).

Figure 5. Effects of sample size on ACR and dip-tests of unimodality. A, population distribution for D and G, with equal sample sizes and standard deviations between the modes. B, population distribution for E and H, with equal standard deviations between the modes but the sample sizes are unequal, with a ratio of 2:3. C, population distribution for F and I, with equal samples sizes between the modes but the standard deviations are unequal, with a ratio of 2:1. D-F, boxplots of p-values from dip-test of unimodality based on 1000 random samples per sample size form populations distributions in A-C, respectively. G-I, boxplots of p-values from ACR test of unimodality based on 1000 random samples per sample size form populations distributions in A-C, respectively. Samples drawn from the bimodal distributions tend to be falsely judged unimodal (i.e., p-values above 0.05) by the tests at small sample sizes. Red lines are for a p-value of 0.05.

Table 1. Taxonomic composition of the living species examined.

	pre-screening	post-screening
Vertebrates	99	89
Osteichthyes	18	8
Amphibia	1	1
Reptilia	72	72
Aves	2	2
Mammalia	6	6
Invertebrates	54	50
Crustacea	4	4
Arachnida	42	38
Insecta	7	7
Cnidaria	1	1

Table 2. Numbers of species that passed ACR, Gap and visual tests of dimorphism at two probability levels, out of the 139 sexually dimorphic extant species examined.

Positive results in…	p<0.1		p<0.05	
	n	misclassification	n	misclassification
Visual Test*	46	0.038	33	0.027
ACR Test	103	0.146	84	0.129
Gap Test	74	0.114	72	0.114
V+A	45	0.038	33	0.027
V+G	46	0.038	33	0.027
A+G	69	0.109	59	0.098
V+A+G	45	0.038	33	0.027

Table 3. Characterization of the five types of dimorphism expression found in extant species plus one type (Type 6) that is known in fossil species.

Type	Bimodality observed in…	Cluster using…	Dominant Dimorphism	Note
1	PC_{DVP} 2, (Y_{DVP}, PC_{All} 2)	PC_{DVP} 2	Shape	Common
1'	RES_{DVP}, RES_{PCS}	PC_{DV}P 2	Shape	Rare
2	PC_{DVP} 1, (Y_{DVP}, PC_{All} 2)	PC_{DVP} 1	Shape/Size	Rare
3	PC_{All} 2, (Y_{DVP})	PC_{All} 2	Shape	PC_{DVP} 1 and 2 are not bimodal
4	PC_{All} 1, (X_{DVP})	PC_{All} 1	Size	Common
5	X_{DVP}	XDVP	Size	PC_{All} 1 is not bimodal
6	PC_{All} 2, (Y_{DVP}, PC_{DVP} 2)	PC_{All} 2-N*	Shape	Many traits with shape dimorphism

*N is the number of principal components

Table 4. Test results for fossil reptiles.

Species	Visually Dimorphic	Gap Test	ACR Unimodality	ACR Bimodality	Bimodal Metric	MANOVA
Allosaurus fragilis	Partly	2 groups	rejected	not rejected	PCDVR 2	p<0.01
Hesperosaurus mjosi	Partly	2 groups	rejected	not rejected	PCDVR 2	p<0.01
Protoceratops andrewsi	Partly	2 groups	rejected	not rejected	PCDVR 2	p<0.01
Plateosaurus sp.	No	2 groups	not rejected	not rejected	—	—
Keichousaurus hui	Partly	2 groups	rejected	not rejected	PCAll2, PCDVR 2	p<0.01

Figure 1. Schematic plots a sexually dimorphic sample to explain basic notions. A, a biplot showing the difference between the total and sex-wise regression lines. B, a biplot of a case in which PC1 and 2 are correlated when viewed sex-wise. C, an idealized case of sexual size dimorphism without shape dimorphism. D, an idealized case of sexual shape dimorphism without size dimorphism. E, an idealized mixture of sexual size and shape dimorphism. F, a typically observed mixture of sexual size and shape dimorphism. Blue and red indicate two sexes (specific sex is irrelevant in these theoretical plots).

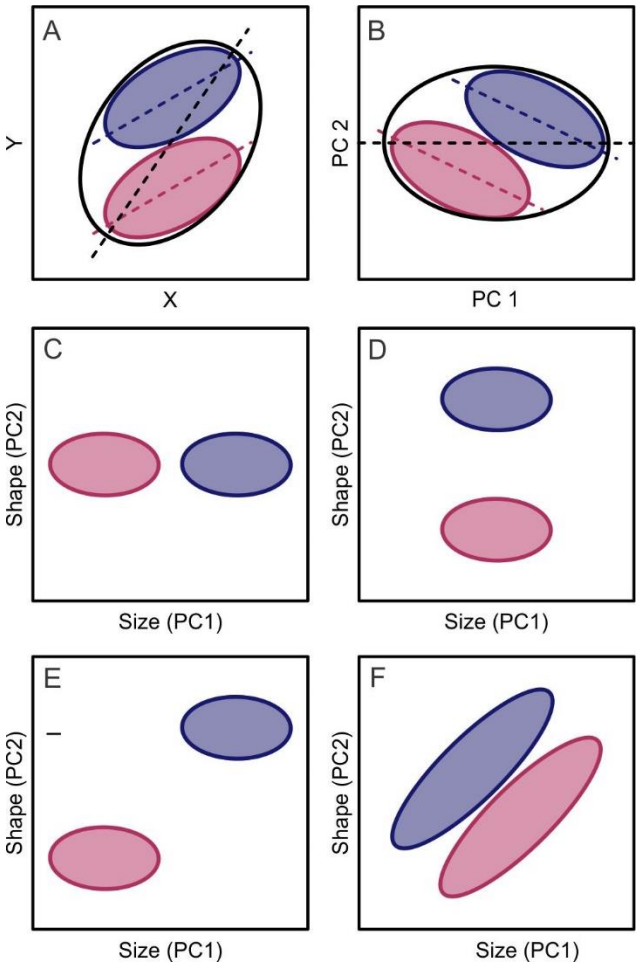


Figure 2. Five types of dimorphism expression observed, with exemplar species. For each species, a biplot of DVP and a histogram of the most bimodal metric is given, except for *Cyperideis torosa*, for which two additional plots, of PC_{All} 1-2 and PC_{DVP} 1-2, are also provided to demonstrate sex-wise correlations in these variable pairs. A-D, type 1; E-H, type 2; I-J, type 3; K-N, type 4; O-P, type 5, Q-T, type 1'. Species are: A-B, *Seminatrix pygae*; C-D, *Homarus gammarus*; E-F, *Mus musculus*; G-H, *Parus major*; I-J, *Gymnorhamphichthys rondoni*; K-L, *Anolis hendersoni*; M-N, *Longiperna concolor*; O-P, *Panthera leo*; Q-T, *Cyperideis torosa*. Blue is male and red is female (true sex).

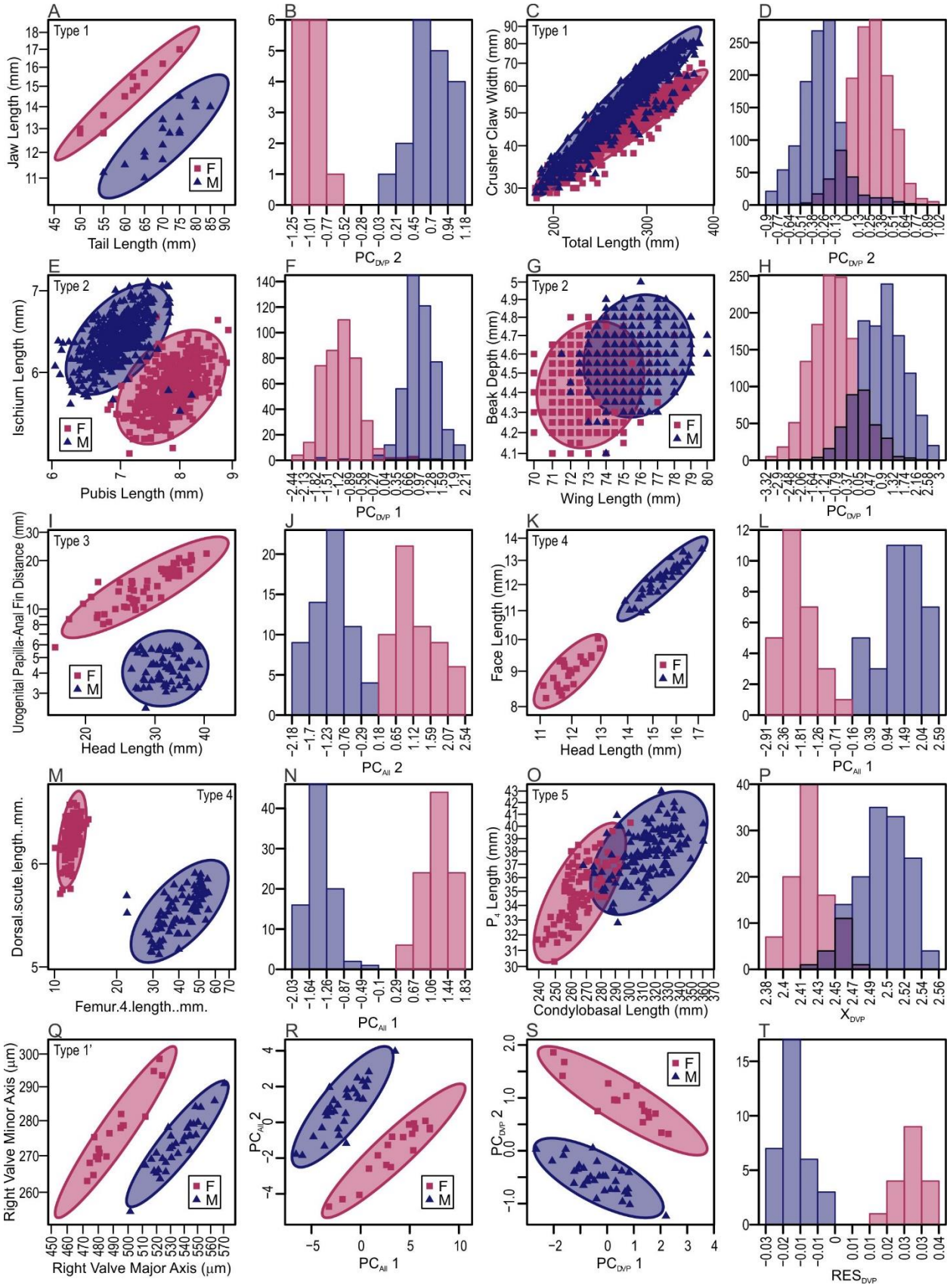

A
Type 1
Jaw Length (mm)
Tail Length (mm)
F
M
B
PC_DVP 2
C
Type 1
Crusher Claw Width (mm)
Total Length (mm)
D
PC_DVP 2
E
Type 2
Ischium Length (mm)
Pubis Length (mm)
F
M
F
PC_DVP 1
G
Type 2
Beak Depth (mm)
Wing Length (mm)
F
M
H
PC_DVP 1
I
Type 3
Urogenital Papilla-Anal Fin Distance (mm)
Head Length (mm)
F
M
J
PC_All 2
K
Type 4
Face Length (mm)
Head Length (mm)
F
M
L
PC_All 1
M
Type 4
Dorsal.scute.length..mm.
Femur.4.length..mm.
N
PC_All 1
O
Type 5
P_4 Length (mm)
Condylobasal Length (mm)
P
X_DVP
Q
Type 1'
Right Valve Minor Axis (µm)
Right Valve Major Axis (µm)
F
M
R
PC_All 2
PC_All 1
F
M
S
PC_DVP 2
PC_DVP 1
F
M
T
RES_DVP

Figure 3. Biplot of DVP and histogram of the most bimodal metric in dinosaurs. A-B, *Allosaurus fragilis*; C-D, *Hesperosaurus mjosi*, E-F, *Protoceratops andrewsi*; G-H, *Plateosaurus* sp. Blue is male and red is female (inferred sex).

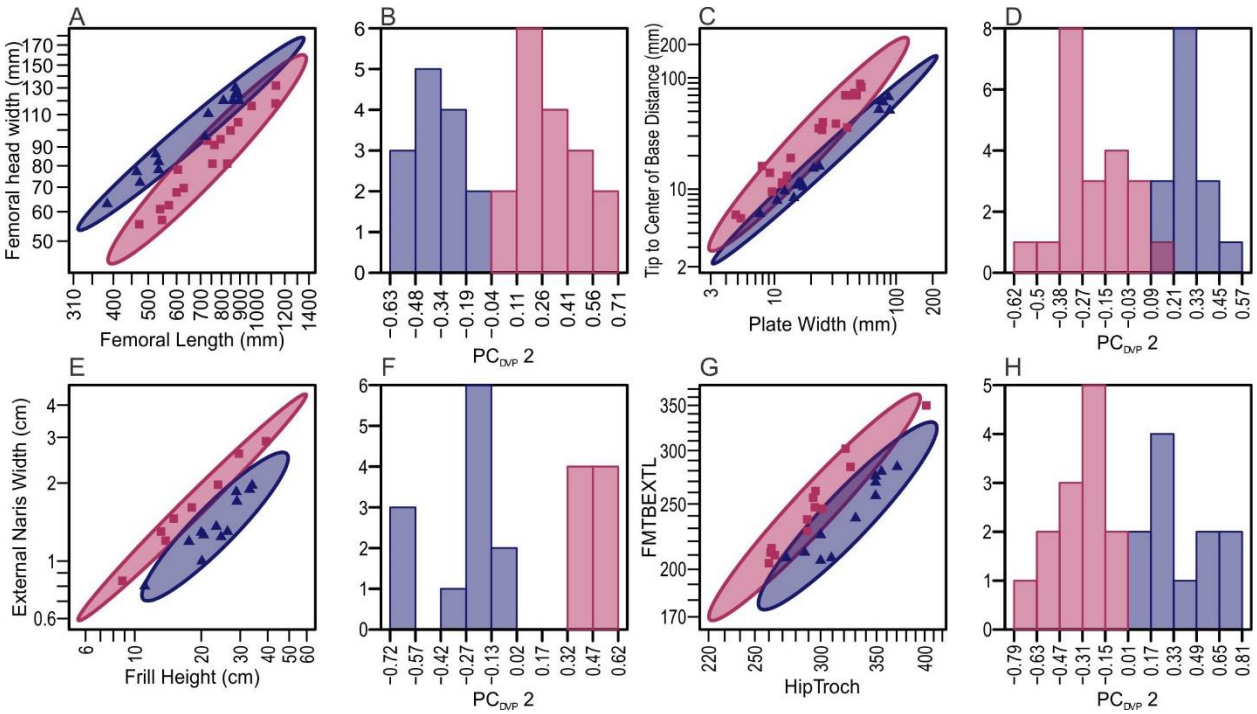


Figure 4. Biases in the total versus dimorphic part of the data in three exemplar species. Biases, seen as sex-wise correlation of principal component space, is less when analyzing the dimorphic part of the data only as opposed to the total data, likely because the signal/noise ratio is higher in the former A-D, *Seminatrix pygae*; E-H, *Cyperideis torosa*; I-L, *Mus musculus*. Blue is male and red is female (true sex).

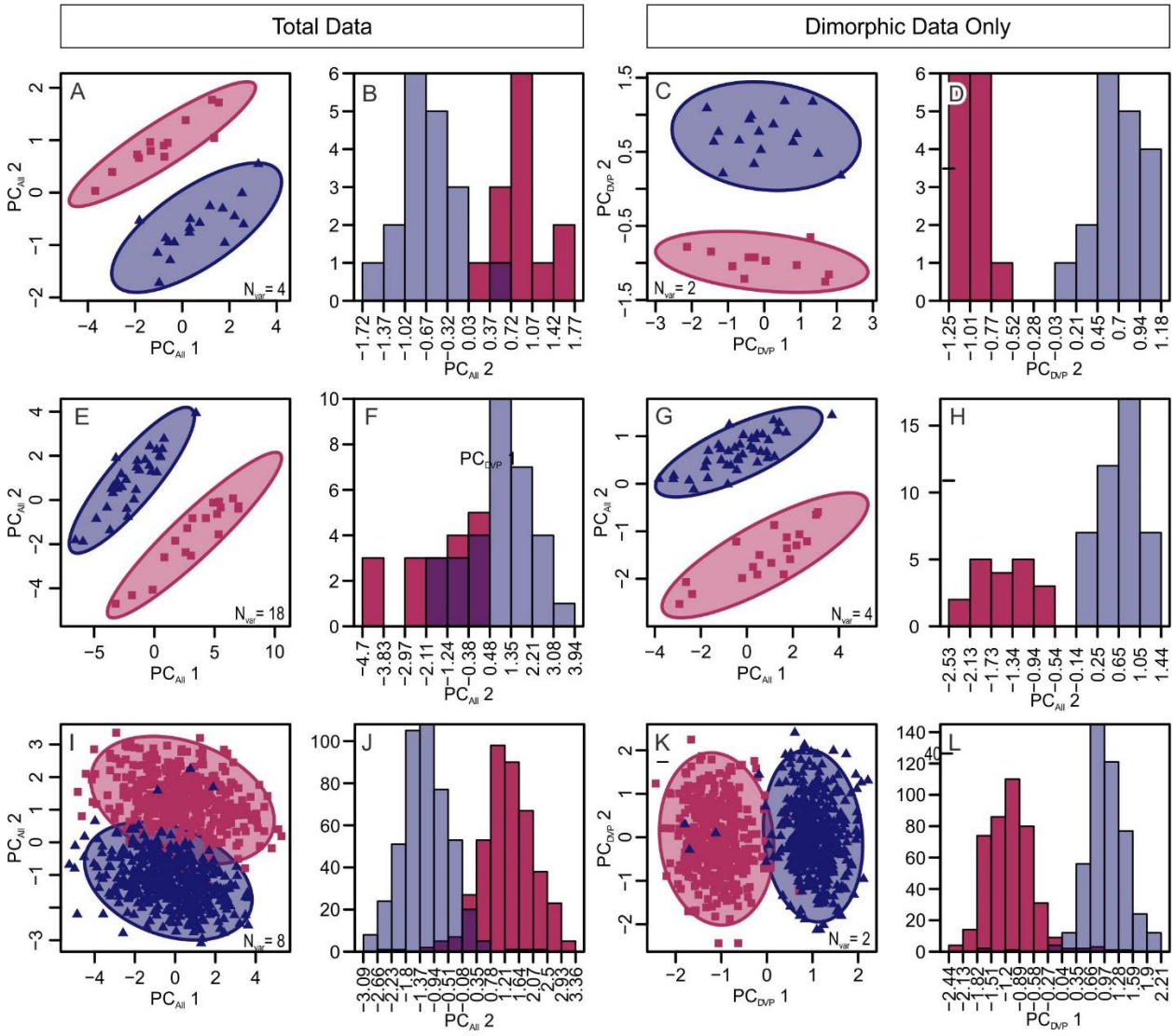
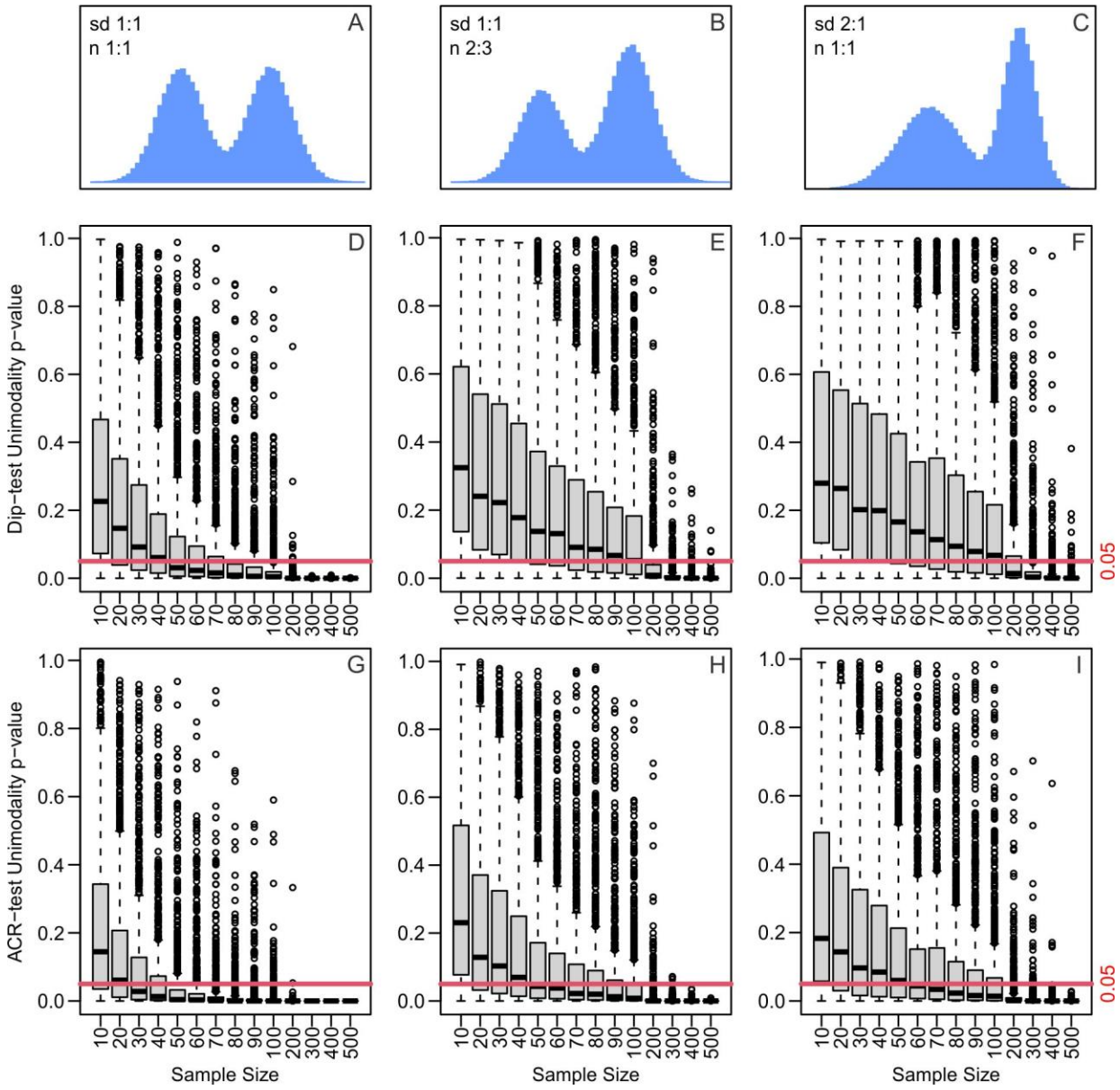


Figure 5. Effects of sample size on ACR and dip-tests of unimodality. A, population distribution for D and G, with equal sample sizes and standard deviations between the modes. B, population distribution for E and H, with equal standard deviations between the modes but the sample sizes are unequal, with a ratio of 2:3. C, population distribution for F and I, with equal samples sizes between the modes but the standard deviations are unequal, with a ratio of 2:1. D-F, boxplots of p-values from dip-test of unimodality based on 1000 random samples per sample size form populations distributions in A-C, respectively. G-I, boxplots of p-values from ACR test of unimodality based on 1000 random samples per sample size form populations distributions in A-C, respectively. Samples drawn from the bimodal distributions tend to be falsely judged unimodal (i.e., p-values above 0.05) by the tests at small sample sizes. Red lines are for a p-value of 0.05.


sd 1:1
n 1:1
A
sd 1:1
n 2:3
B
sd 2:1
n 1:1
C
D
E
F
G
H
I
Dip−test Unimodality p−value
ACR−test Unimodality p−value
0.05
Sample Size
Sample Size
Sample Size