EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION
SSE
SOCIETY for the STUDY of EVOLUTION

Evolution of large males is associated with female-skewed adult sex ratios in amniotes

Journal:	*Evolution*
Manuscript ID	Draft
Manuscript Type:	Original Article
Keywords:	sexual selection, mating competition, mating opportunity, sex-biased mortality, comparative method

Evolution of large males is associated with female-skewed adult sex ratios in amniotes

Abstract

Body size often differs between the sexes (leading to sexual size dimorphism, SSD), as a consequence of differential responses by males and females to selection pressures. Adult sex ratio (the proportion of males in the adult population, ASR) should influence SSD because ASR relates to both the number of competitors and available mates, which shape the intensity of mating competition and thereby promotes SSD evolution. However, whether ASR correlates with SSD variation among species has not been yet tested across a broad range of taxa. Using phylogenetic comparative analyses of 462 amniotes (i.e. reptiles, birds and mammals), we fill this knowledge gap by showing that male bias in SSD increases with increasingly female-biased ASRs in both mammals and birds. This relationship is not explained by the higher mortality of the larger sex because SSD is not associated with sex differences in either juvenile or adult mortality. Phylogenetic path analysis indicates that higher mortality in one sex leads to skewed ASR, which in turn may generate selection for SSD biased towards the rare sex. Taken together, our findings provide evidence that skewed ASRs in amniote populations can result in the rarer sex evolving large size to capitalise on enhanced mating opportunities.

Keywords: sexual selection, mating competition, mating opportunity, sex-biased mortality, comparative method

INTRODUCTION

Sexual size dimorphism (SSD, measured as the size of males relative to females) is widespread in nature and is one of the most conspicuous phenotypic difference between the sexes (Darwin 1871; Andersson 1994; Fairbairn et al. 2007). It is the consequence of different optimal body size for the sexes resulting from opposing selection forces (some of which may influence only one of the sexes) that equilibrate differently in males and females (Blanckenhorn 2005).

A large volume of research has focused on how sex-specific behaviour (e.g. mating system, parental care), ecological processes (e.g. abundance and quality of resources), and life history (e.g. fecundity in indeterminate growers) can generate size differences between the sexes (Andersson 1994; Blanckenhorn 2005). These studies have concluded that sexual selection is often a major driver of SSD evolution by either intra-sexual competition for access to mates or inter-sexual mate choice, although other evolutionary mechanisms (e.g. fertility selection and competition for resources) may also be important (Jehl and Murray 1986; Andersson 1994; Blanckenhorn 2005; Fairbairn et al. 2007; Clutton-Brock 2016). Strong sexual selection for large body size in one sex is particularly likely in species where that sex competes for mates by physical contests or endurance rivalry, as observed in several vertebrate taxa (e.g. reptiles, birds, and mammals; Jehl and Murray 1986; Andersson 1994; Cox *et al.* 2007; Székely *et al.* 2007; Clutton-Brock 2016).

Adult sex ratio (ASR), best measured as the proportion of males in the adult population (Ancona et al. 2017) is a key demographic property of populations that influences both the number of competitors for mates and the number of mates available to an individual (Murray 1984; Székely et al. 2014b; Jennions and Fromhage 2017; Schacht et al. 2017). For example, a male-skewed ASR means potentially more competitors and fewer available partners for males than for females. An increasing number of studies show that ASR covaries

with several reproductive traits such as mating system, parental sex roles, divorce rate, extra-pair mating and cooperative breeding both in non-human animals and humans (Liker et al. 2013, 2014; Schacht et al. 2014; Kappeler 2017; Komdeur et al. 2017; Eberhart-Phillips et al. 2018; Grant and Grant 2019). However, whether and how ASR is related to the evolution of SSD is still poorly understood.

Theories suggest that ASR can drive the evolution of SSD in at least two ways. First, the intensity of sexual competition may increase with the number of competitors. As Darwin wrote (1871, p. 217): *"That some relation exists between polygamy and development of secondary sexual characters, appears nearly certain; and this supports the view that a numerical preponderance of males would be eminently favourable to the action of sexual selection"*. According to his idea, highly skewed ASRs may intensify selection for competitive traits such as weapons and large body size in the more numerous sex. Thus this 'mating competition hypothesis' predicts that the extent of male-bias in SSD should increase with the degree of male skew in the ASR. Later work refined Darwin's (1871) original idea by suggesting that the operational sex ratio (OSR, the number of sexually active males per receptive female at a given time) rather than the ASR determines the intensity of mating competition in a population (Emlen and Oring 1977). Thus, according to this latter theory ASR would predict SSD if ASR covaries with OSR, for example because OSR is in part determined by ASR (together with sex differences in behaviour like parental care; Kokko et al. 2012). Although the relationship between ASR and OSR is yet to be fully explored, their positive association has been demonstrated both by theoretical models (Kokko and Jennions 2008: Fig. 4a; Fromhage and Jennions 2016: Fig. 3c,d) and comparative analyses (Mitani et al. 1996, correlation between ASR and OSR in 18 primates: $r = 0.4$, $P = 0.002$; unpublished result using data from their Table 1). Empirical studies commonly use ASR and OSR

74 interchangeably in testing their relationship with SSD (Poulin 1997) and other proxies of
75 sexual selection (Janicke and Morrow 2018).
76 Second, models of reproductive sex roles predict that ASR should influence the
77 evolution of SSD because individuals of a given sex may allocate less to parental care when
78 the sex ratio is skewed towards the opposite sex than when it is skewed towards their own sex
79 (Queller 1997; McNamara et al. 2000). According to these models, males in female-skewed
80 populations display a higher reproductive success due to increased probability of breeding
81 with multiple partners and therefore may evolve to reduce parental care (Queller 1997:
82 section 3., McNamara et al. 2000: section 'Sex ratio'). This association between ASR and
83 parental sex roles can drive the evolution of SSD because more elaborate trait expression in
84 males is evolutionarily linked to female-biased care and stronger sexual selection on males
85 (the so called 'sex-role syndrome', Janicke et al. 2016: Fig 3.). Thus, this 'mating
86 opportunity hypothesis' predicts that the extent of male-bias in mating competition, and hence
87 in SSD, should decrease with increasing male skew in the ASR. A demographic analysis of
88 mating systems by Murray (1984) also predicts that female-skewed ASRs should be
89 associated with both polygyny and male-biased SSD, whereas male-skewed ASRs should be
90 associated with polyandry and female-biased SSD.
91 Alternatively, SSD may drive changes in sex ratios through sex differences in
92 mortality resulting from sexual competition. According to this 'mortality cost hypothesis', the
93 skewed ASR is a consequence rather than a cause of intense sexual selection, because when
94 males allocate a lot to mating competition they may suffer increased mortality, which in turn
95 leads to female-skewed ASR (Trivers 1972; Clutton-Brock et al. 1985; Liker and Székely
96 2005; Kalmbach and Benito 2007). This hypothesis predicts that in species exhibiting SSD
97 (1) the larger sex should have higher mortality due to the costs of being large, including the

direct costs associated with competition (e.g. fights, displays); which leads to (2) decreasing male skew in the ASR with increasing degree of male bias in the SSD.

Studies that have investigated the relationships between sex ratios, SSD and sex-specific mortality have so far yielded inconsistent results. While some studies found a positive link between SSD and ASR or OSR (i.e. an increasing male bias in SSD with increasing male skew in the sex ratios; Mitani *et al.* 1996; Poulin 1997), others reported negative associations (Clutton-Brock et al. 1977; Wittenberger 1978; Georgiadis 1985; Haro et al. 1994; Johansson et al. 2005; Lovich et al. 2014), or found no consistent relationships (Owen-Smith 1993; Hirst and Kiørboe 2014; Muralidhar and Johnson 2017). Similarly, mortality costs paid by the larger sex in dimorphic species were reported in some studies (Clutton-Brock et al. 1985; Promislow 1992; Promislow et al. 1992; Moore and Wilson 2002; Benito and González-Solís 2007; Kalmbach and Benito 2007), whereas no consistent relationship between SSD and sex differences in mortality was found by others (Owens and Bennett 1994; Toïgo and Gaillard 2003; Lemaître and Gaillard 2013; Székely *et al.* 2014a; Tidière *et al.* 2015). Many of these studies focused on a narrow range of taxonomic groups and were based on a relatively small number of species (typically fewer than 50) in comparative analyses. Furthermore, none of the studies tested explicitly whether statistical models assuming that ASR drives variation in SSD (as proposed by the mating competition and mating opportunity hypotheses) or alternative models (like the mortality costs hypothesis) fit better to the data.

Here we investigate the strength and direction of the relationship between ASR and SSD in populations of wild amniotes, using the largest existing comparative dataset on ASR compiled to date (462 species). First, we investigate whether SSD increases or decreases with ASR across species, as predicted by the mating competition and mating opportunity hypotheses, respectively. We also test whether the relationship is consistent among three

major amniote taxa (reptiles, birds, and mammals) because these taxa differ in multiple ecological, behavioural and life-history traits. Since the extent and direction of SSD can be influenced by ecological, life-history and behavioural factors besides mating competition, we also control for several potential confounding variables in the analyses. Second, we study whether SSD drives ASR variation by generating sex-biased mortality as proposed by the mortality cost hypothesis. We test this latter hypothesis by investigating whether SSD is related to sex differences in juvenile or adult mortality, and by comparing path models representing different structural relationships between SSD, ASR and sex-specific mortality.

METHODS

Data collection

Data were extracted from published sources (see Appendix S1 in Supporting Information). The initial dataset was based on Pipoly *et al.* (2015) that contains ASR and SSD for 344 amniote species. We excluded amphibians included in Pipoly *et al.* (2015) because sex-specific mortality data (see below) are very scarce for this taxon, especially in juveniles. The initial dataset was augmented with additional reptile and mammal species, and with information on sex-specific mortality. These additional data were taken from existing comparative datasets (Berger and Gompper 1999 and Bókony *et al.* 2019 for ASR in mammals and reptiles, respectively, and Székely *et al.* 2014a for mortality in birds) or from primary publications. In the latter case we searched the literature through the search engines Web of Science and Google Scholar, using the search terms ‘sex ratio’, ‘sex-specific mortality OR survival’ or ‘male female mortality OR survival’ together with taxonomic names. Data for different variables for the same species were often available only from different populations or studies. The final dataset includes 462 species with both ASR and SSD available (155 reptiles, 185 birds, 122 mammals).

148

149 **Body mass and SSD**

150 Sex-specific body mass (g) was available for all birds and mammals in our dataset. Since
151 body mass data were missing for many reptiles, we also collected body length data (mm) for
152 this taxon in the form of snout-vent length for squamates and crocodilians and plastron or
153 carapace length for turtles. We estimated body mass from body length using published
154 allometric equations (Appendix S2). We used estimated body mass for reptiles instead of
155 body length in the combined analyses of all species because (1) data on mass are more readily
156 available than data on body length in birds and mammals, which provided the majority of
157 species, and (2) body mass is measured in a standardized way in all taxa, whereas the
158 measurement of body length varies because different parts of the body are recorded as a proxy
159 for length in different taxa. If multiple mass or length data were available for a species, we
160 used the mean value. Average adult body mass was calculated as $\log_{10}$-transformed mean
161 mass of the sexes.

162 SSD was calculated as $\log_{10}$(male mass / female mass), which has been recommended
163 as a statistically appropriate and preferable measure of dimorphism (Smith 1999; Fairbairn
164 2007). To test whether the results are sensitive to conversion of length to mass in reptiles, we
165 repeated the main analyses (1) with SSD calculated from body length ($\log_{10}$(male length /
166 female length)) of reptiles, and (2) with SSD calculated from body mass for a subset (31
167 species) of reptiles that has sex-specific mass data available from Myhrvold et al. (2015).
168 Whatever approach was used to assess the degree of sexual size dimorphism the results were
169 qualitatively unchanged (see Results). In the main text we thus report results based on body
170 mass estimated from body length for reptiles.

171

172 **Sex ratio**

We followed Wilson and Hardy (2002) and Ancona et al. (2017) in expressing ASR as the proportion of males in the adult population. We defined the adult population here broadly as adult individuals living in the study area during ASR sampling. Wilson and Hardy (2002) showed that analysing sex ratios as a proportion variable is appropriate when sex ratios are estimated from samples of ≥ 10 individuals and the dataset has ≥ 50 sex ratio estimates. These conditions are fully met in our analyses because sample sizes for ASR estimates were always larger than 10 individuals per species (and typically much larger), and more than 50 species were included in most analyses.

ASR data from Pipoly *et al.* (2015) were augmented with new species and updated with more recent and/or better quality information (e.g. based on a more reliable method or a larger sample size) for some reptiles. ASR estimates were collected by different observers for the different taxa: reptiles by V.B. and I.P. (Pipoly et al. 2015; Bókony et al. 2019), birds by A.L. (Liker *et al.* 2014), and mammals by Berger and Gompper (1999), Donald (2007) and Anile and Devillard (2018). Details of data selection criteria are given in the original publications (see also Ancona *et al.* 2017). Mean values were calculated for species with multiple ASR data. ASR estimates are repeatable between populations of the same species as measured by the intraclass correlation coefficient (ICC), although the magnitude of repeatability varies among taxa: reptiles with genetic and environmental sex determination: ICC= 0.55 and 0.14, respectively (Bókony et al. 2019), birds: ICC= 0.64 (Ancona et al. 2017), mammals: ICC= 0.60 (Valentine Federico, J-F.L., J-M.G., A.L., I.P., T.S. unpublished results). ASR estimates are not influenced by the sample size of the ASR studies (Székely et al. 2014a; Bókony et al. 2019).

Sex-specific mortality

Annual mortality rates were collected from studies in which mortality (or survival) was estimated for each of both sexes. Juvenile and adult mortality refer to age classes before and after the age of first reproduction, respectively. For reptiles, data were collected by V.B. (Bókony et al. 2019). Most adult mortality data on birds are taken from Székely *et al.* (2014a) with the addition of new data for juvenile mortality by A.L. Reptile and bird mortality includes estimates by various methods (capture-recapture and demographic analyses, return rates), although we used better quality estimates (e.g. those from capture-recapture analyses) whenever we had a choice (Székely *et al.* 2014a; Bókony *et al.* 2019). For mammals, all sex-specific estimates were collected by J-M.G. and J-F.L. (Lemaître et al. 2020). Sex differences in juvenile and adult mortality rates were calculated as the magnitude of male-biased mortality (i.e. $\log_{10}$(juvenile or adult male mortality / juvenile or adult female mortality)), also referred to as 'mortality bias'. These measures of mortality bias are not related to the overall mortality rate of the species, as estimated by the average mortality rates of the sexes (phylogenetic generalised least squares models, juvenile mortality bias: slope ± *SE* = - 0.068 ± 0.101, $t = 0.7$, $P = 0.497$, $n = 100$; adult mortality bias: slope ± *SE* = - 0.05 ± 0.08, $t = 0.7$, $P = 0.513$, $n = 230$).

Other predictors

We controlled for the potential effects of ecological and life-history variables related to either ASR or SSD (or both) that may confound the assessment of their relationship. First, we collected data on the type of sex determination system because it is associated with both ASR (Pipoly et al. 2015) and SSD (Adkins-Regan and Reeve 2014). We divided the species into three categories according to the Tree of Sex database (Ashman et al. 2014): male-heterogametic (XY) or female-heterogametic (ZW) genetic sex determination, or temperature-dependent sex determination (TSD). For species that were not included in the Tree of Sex

database we assumed the same type of sex determination as reported for the genus (or family, respectively; Bókony *et al.* 2019) when the genus (or family) to which it belongs had invariable sex determination system. All birds were assigned to ZW, and all mammals to XY sex-determination (Ashman et al. 2014).

Second, we controlled for the potential effects of environmental variation among species by using two proxies. Breeding latitude correlates with life-history traits in many organisms and may also influence the potential for polygamy, hence also sexual selection (Fischer 1960; Isaac 2005; Balasubramaniam and Rotenberry 2016). We used absolute values of the geographic latitude of the ASR studies included in our dataset (i.e. average values for species with multiple ASR estimates) to represent the distance from the Equator. When the authors did not report latitude, we used Google Earth to estimate it as the center of the study sites based on the site descriptions. For 30 birds and 10 mammals, accurate population locations were not reported, hence, we used the latitudinal midpoint of the breeding ranges of these species (birds: V. Remeš, A. Liker, R. Freckleton and T. Székely unpublished data, mammals: PanTHERIA database).

In addition to latitude, we investigated environmental harshness as a second environmental variable, which also has been hypothesized to influence SSD (Isaac 2005). We quantified the harshness of the breeding environment using a proxy proposed by Botero *et al.* (2014). This is the PC1 score extracted from Principal Component Analysis (PCA) performed on a set of climatic and ecological variables (e.g. temperature and precipitation, net primary productivity, habitat heterogeneity; see Botero *et al.* 2014 for a detailed description of the variables and the analysis). The PC1 scores have higher values for a higher level of exposure to drier, less productive environments, with colder, less predictable and more variable annual temperatures (see Table 1 in Botero *et al.* 2014). In birds and mammals, we used the data

published in Botero *et al.* (2014), whereas for reptiles we calculated PC1 scores by performing a PCA with the same set of variables.

Third, we characterised courtship displays in birds because earlier studies showed that birds with aerial displays have less male-biased SSD compared to species with ground displays, probably because selection favours male agility in aerially displaying species constraining male body size (Jehl and Murray 1986; Székely et al. 2007). We followed Székely *et al.* (2007) and divided species into two display groups: (1) mating displays that may favour male agility, including species that mainly have aerial displays (both non-acrobatic and acrobatic, categories 4 and 5 in Székely *et al.* 2007), and (2) displays that may not favour male agility, including all other display types, typically performed on ground (categories 1-3 in Székely *et al.* 2007). Although SSD can also be influenced by display type and display habitat in reptiles and mammals (e.g. see Agha *et al.* 2018), we were not able to collect reliable data for these taxa, therefore we analysed the effect of display type only in birds.

Fourth, we tested for the potential effect of social mating system, because the scope for mating competition may be more limited in monogamous than in polygamous species (Andersson 1994). Thus, although there is ASR variation among monogamous species that can generate some variation in mating competition and/or opportunity, the relationship between ASR and SSD is expected to be weaker in monogamous than in polygamous species. To test this idea, we characterised social mating system for birds and mammals, because we found reliable information in these taxa for most species (Liker *et al.* 2014; Lukas and Clutton-Brock 2013). We categorized species as either socially monogamous or polygamous (most often polygynous) according to the sources. In birds, mating system was originally scored on a five point scale (Liker et al. 2014), and here we considered a species monogamous if it had score 0 or 1 (polygamy frequency <1%) for both sexes.

Finally, in reptiles, the evolution of viviparity and reduced reproductive frequency are generally correlated with shifts toward female-biased SSD due to fecundity selection for large female size (Pincheira-Donoso and Hunt 2017). To control for its potential effect on SSD, we categorized the reproductive mode of reptiles as either viviparous or oviparous (Uetz et al. 2019).

Statistical analyses

Phylogenetic generalised least squares (PGLS) models were built to conduct bivariate and multi-predictor analyses. To control for phylogenetic relationships among taxa, we used the composite phylogeny applied in Pipoly *et al.* (2015) with the addition of new species according to a family-level (Sarre et al. 2011) and other recent phylogenies (Squamata: Nicholson *et al.* 2012, Pyron *et al.* 2013, Gamble *et al.* 2014; Testudines: Barley *et al.* 2010, Guillon *et al.* 2012, Spinks *et al.* 2014; Crocodylia: Oaks 2011; mammals: Fritz *et al.* 2009, Meredith *et al.* 2011). Since composite phylogenies do not have true branch lengths, we used three methods to generate branch lengths (Nee's method, Pagel's method, and unit branch lengths, using the PDAP:PDTREE module of Mesquite; Midford *et al.* 2011), and repeated key analyses with these alternative trees. We present results with Nee's branch lengths in the paper, except for the sensitivity analyses (see Results). Freckleton et al. (2002) showed that PGLS is relatively insensitive to branch length assumptions. In each model we used the maximum-likelihood estimate of phylogenetic dependence (Pagel's λ). PGLS models were run using the 'caper' R package (Orme et al. 2013).

First, using all species, we applied bivariate PGLS models to test interspecific associations between ASR, SSD and sex differences in juvenile and adult mortality rates. Then we built two multi-predictor models. In Multi-predictor model 1, we tested the relationship between ASR and SSD while controlling for potential confounding effects of

mean mass, sex determination system, and breeding latitude. In Multi-predictor model 2, we tested the ASR - SSD relationships while controlling for the effects of sex differences in juvenile and adult mortality rates, and mean mass. We built these two separate multi-predictor models because we have much lower sample sizes for sex-specific mortalities than for the other predictors, thus the statistical power would be reduced for variables of Multi-predictor model 1 if all predictors were combined in a single model. We ran the models in two alternative versions in which either SSD or ASR was the dependent variable, respectively, since we had no *a priori* knowledge about the cause-effect direction of these relationships and results may differ between these analyses if the two models have different values for Pagel's λ (see Appendix S3).

We investigated whether the ASR – SSD relationship, which is the main focus of our study, differed among taxa by testing the interaction between ASR and the taxonomic class. To explore differences among taxa in the multivariate relationships, we repeated all analyses separately for reptiles, birds and mammals. In taxon-specific Multi-predictor models 1, we included reproductive mode for reptiles and display type for birds as further predictors. In reptiles, we also tested whether the relationship between ASR and SSD is sensitive (1) to the inclusion of species that have environmental sex determination, because ASR shows low repeatability in such reptiles (Bókony et al. 2019), and (2) to the inclusion of species in which the type of sex determination was inferred from data on related species in the genus or family. Finally, we ran two additional separate analyses to test whether social mating system and environmental harshness confounded the ASR - SSD relationship. We did not include these latter two predictors in the multi-predictor models because these data were available only for subsets of species that would have resulted in considerably reduced sample sizes. All numeric variables were standardized before analyses to make parameter estimates comparable, and model assumptions were also checked and met. We report two-tailed statistics. Sample sizes

differed between models because not all variables were available for all species (see Appendix S1).

In addition to PGLS models, we used phylogenetic path analyses (Santos 2012; Gonzalez-Voyer and von Hardenberg 2014) to compare two sets of path models corresponding to different hypotheses for the relationships linking ASR, SSD and sex differences in mortality. Although path analyses – unlike experiments – cannot infer causality, it is a suitable method to compare alternative scenarios representing different causal relationships between variables (Shipley 2016). Model 1 assumes that sex-biased mortality influences ASR, which in turn influences SSD through its effects on mating competition (as proposed by the mating opportunity hypothesis; Fig. 1). Three variants of this model were tested: Model 1a assumes that sex differences in both juvenile and adult mortality rates influence ASR, while Models 1b-c include only one of these mortality effects. Model 2 assumes that SSD has sex-specific effects on juvenile and/or adult mortality, which then drives ASR variation (representing the mortality cost hypothesis; Fig. 1). We tested all the three variants of this latter scenario, assuming SSD effects on both juvenile and adult mortality (Model 2a) or only on one mortality component (Models 2b-c).

We followed the approach proposed by Santos (2012) for phylogenetic path analyses. In the first step, we conducted phylogenetic transformation on the data to control for effects of phylogenetic relatedness among species. For this purpose, we (1) determined λ separately for each variable by maximum likelihood, (2) used this variable-specific λ value to re-scale the phylogenetic tree to a unit tree, and (3) used the transformed tree to calculate phylogenetically independent contrasts for the variable (using 'pic' function of the R package 'ape'; Paradis 2012). We repeated this process for each variable, and the resulting phylogenetically transformed values were used for fitting path models. In the second step of the analyses, we evaluated model fit using d-separation method (Shipley 2016) as implemented in the R

package 'piecewiseSEM' (Lefcheck 2016). In this method, Fisher's *C* statistic is used to test the goodness of fit of the whole path model, and the model is rejected (i.e. it does not provide a good fit to the data) if the result of this *C* statistic is statistically significant (and conversely a statistically non-significant result means acceptable fit; Lefcheck 2016). We compared model fit between the six path models by their AICc values. Note that this approach ensures that the same variables (i.e. the contrasts) are used in each path model and that correlations are non-directional (i.e. for a pair of variables *X* and *Y*, $r_{XY} = r_{YX}$ as assumed in path analysis).

To test the robustness of the results, we repeated the path analyses using two other methods. First, we repeated the above procedure (i.e. followed Santos 2012) except that we used the covariance matrix comparison method for model fit instead of d-separation, as implemented in the R package 'lavaan' (Rosseel 2012). Second, we repeated the analyses using the method developed by von Hardenberg and Gonzalez-Voyer (2013). Unlike Santos' (2012) method, in this latter approach a single value of Pagel's λ is estimated for each pair of traits in a directional statistical model, rather than a value of λ for each variable (see the Discussion and Appendix S3 where we explain why this approach may be problematic). We used the R package 'phylopath' (van der Bijl 2018) for this latter analysis, which relies on the d-separation method for model fitting (similarly to 'piecewiseSEM', see above). We provide further justifications for our approach and additional analyses to test the robustness of the path analysis' results in Appendix S3.

RESULTS

Mating competition versus mating opportunity hypotheses

Consistent with the mating opportunity hypothesis, and in contrast to the mating competition hypothesis, we found a negative relationship between our measures of ASR and SSD: the size of males relative to females increases when ASR becomes more female-skewed (Fig. 2, Table

1). This correlation was statistically significant when all species were analysed together and did not differ among the three amniote classes (ASR × class interaction on SSD: $F_{2,456}$= 2.177, P= 0.115). The increase of SSD with increasingly female-skewed ASR was statistically significant within birds and mammals but was not in reptiles when the three taxa were analysed separately (Fig. S1, Tables S1-4). These results remained consistent when we used SSD estimates based on length instead of estimated mass in reptiles (Tables S1, S2 and S5), and also when SSD for reptiles were estimated from published body mass data (Table S5).

These results are robust because the direction of the ASR - SSD relationship and its statistical significance were not sensitive to branch length assumptions (Table S6), and to the inclusion of other predictors (Table 1). In multi-predictor models (Table 1), mean body mass was positively related to SSD, supporting the Rensch rule (Abouheif and Fairbairn 1997), and the type of sex determination influenced ASR variation as previously reported by Pipoly *et al.* (2015). Nevertheless, ASR remained negatively associated with SSD when the effects of mass and sex determination systems were accounted for (Table 1). This result also did not change when environmental variation was included in the models using either breeding latitude (Table 1) or environmental harshness (Table S5). Finally, excluding reptiles with TSD (that have the lowest consistency in ASR; Bókony *et al.* 2019) or with assumed sex determination also did not influence the relationship (Table S5).

The multi-predictor model for birds showed that species with aerial courtship displays have lowered SSD as found in earlier studies (Jehl and Murray 1986; Székely et al. 2007); however, the relationship between ASR and SSD remained statistically significant and negative when this effect was included in the model (Table S3). Furthermore, data in birds and mammals showed that, as expected, the relationship was weaker in monogamous than in polygamous species, although the same trend occurred in both mating systems (Table S7).

Finally, reproductive mode was not associated with SSD or ASR in reptiles in our dataset (Tables S1-2).

Mating opportunity versus mortality costs hypotheses

Both the mating opportunity hypothesis and the mortality cost hypothesis predict female-skewed ASRs in species with male-biased SSD. However, our results are more consistent with the mating opportunity hypothesis for two reasons. First, ASR but not SSD was associated with the extent of sex differences in juvenile or adult mortality, and ASR remained strongly and negatively correlated with SSD when sex differences in juvenile and adult mortality were statistically controlled for (Table 1). Second, phylogenetic path analyses showed that models of the mating opportunity hypothesis provided better fit to the data (Models 1a-c, Fisher' *C* statistic: $P = 0.07 - 0.97$) than models corresponding to the mortality cost hypothesis (Models 2a-c, $P < 0.001$; Table 2). The strongest support was for Model 1a because it had the lowest AICc (ΔAICc = 4.1 - 43.2; Table 2). This model proposes that sex-biased mortality in both juveniles and adults generates skewed ASR, which in turn leads to SSD biased towards the rarer sex (Fig. 3). These results are robust because we obtained the same results when the analyses were repeated using two other implementations of the path analysis (see Table S8 for the results obtained using 'phylopath', and Appendix S3 for the results obtained using 'lavaan'). Finally, path analyses that excluded reptiles (for which the ASR - SSD relationship was not statistically significant, see above) also yielded results qualitatively consistent with the full dataset (Table S9).

DISCUSSION

Our analyses provided three major findings: (1) adult sex ratio is related to sexual size dimorphism among amniote species, although the association is the opposite of the one

proposed by Darwin; (2) sex-biased mortality is unrelated to the extent of SSD in amniotes; and (3) confirmatory path analyses indicate that sex-biased mortality influences ASR, which in turn induces changes in SSD. Collectively, these findings support the mating opportunity hypothesis, indicating that selection is likely to favour an increased resource allocation toward mating competition (by growing and maintaining a large body mass) in the rarer sex, which has a higher chance of getting mates than the other sex.

Theoretical models show that skewed ASRs can promote evolutionary changes that may generate this association between ASR and SSD. First, models of sex role evolution showed that skewed ASR can result in divergences in reproductive roles between the sexes leading to less parental care and more frequent desertion and remating in the rarer sex and opposite changes (i.e. more parental care and less frequent remating) in the more abundant sex (Queller 1997; McNamara et al. 2000). Similarly, a demographic analysis based on the relationships between mating systems and sex ratio, sex-specific patterns of survivorship, age of first reproduction, and annual fecundity predicts that skewed ASRs promote the evolution of polygamy (i.e. polygyny and polyandry in female-biased and male-biased populations, respectively; Murray 1984). Since both frequent remating and polygamy can intensify sexual selection, the above effects of skewed ASR can promote the evolution of SSD by favouring increased body size in the rare sex. In line with the predictions of these models, an increasing number of recent studies in birds and humans show that polygyny is more frequent and parental care by males is reduced in female-skewed populations (Liker et al. 2013, 2014, 2015; Remeš et al. 2015; Schacht and Borgerhoff Mulder 2015; Eberhart-Phillips et al. 2018; Grant and Grant 2019). Our results are also concordant with experimental studies in voles and lizards, which reported that female-skewed ASRs exert directional selection for large body size in males (Klemme et al. 2007; Fitze and Le Galliard 2008), and increase variance in male reproductive success (Dreiss et al. 2010).

Theoretical models predict that the effects of ASR may depend on other life-history and behavioural traits of the populations. For example, Fromhage and Jennions (2016) highlighted the importance of the specific processes generating ASR skews for the outcomes of sex role evolution, and that a coevolutionary feedback between parental care and sexually selected traits can greatly amplify sex role divergence. In addition, sexual competition for mates may favour different traits in species with distinct ecology and behaviour, leading to inconsistent relationships between sex differences in mating competition and sexual dimorphisms in behavioural or morphological trait across species (Clutton-Brock 2017). Collectively, these factors may account for the relatively low amount of variation in SSD explained by ASR in some of our analyses.

The association between intense sexual selection in males and female-skewed ASRs was proposed decades ago by avian evolutionary ecologists (e.g. Mayr 1939), although it was usually explained by the mortality cost hypothesis (Wittenberger 1976). Our analyses do not support this hypothesis because sex-biased SSD is not associated with sex-biased juvenile or adult mortality in the studied amniote species, and the results of the confirmatory path analyses are also inconsistent with the mortality cost hypothesis. We propose that the lack of relationship between SSD and sex differences in mortality may be explained by variation in the environmental context (Lemaître et al. 2020). Studies in birds and mammals showed that having a large body size may only be costly in terms of mortality in populations subjected to harsh environmental conditions (Toïgo and Gaillard 2003; Kalmbach and Benito 2007; Jones et al. 2009; Clutton-Brock 2017). The effect of SSD may thus be reduced or absent when the sex-specific mortality estimates correspond to average conditions, that may often be the case in wild populations.

The ASR - SSD relationship may also be influenced by sex differences in the time of maturation because longer maturation time in the larger sex can result in a shortage of that sex

in the adult population (Lovich et al. 2014) because immature life stages are generally characterized by higher mortality (e.g. Gaillard et al. 2000). Furthermore, Fromhage & Jennions (2016) showed that female-skewed sex ratios at maturation (MSR) can result in the evolution of increased female care and male allocation to traits facilitating mating success. Thus, if variation in ASR is determined at least in part by MSR, then the effects of sex-biased MSR on sex roles can contribute to the observed association of ASR with the intensity of mating competition, and, hence, SSD. This latter mechanism would deserve further investigations.

Although the relationship between ASR and SSD is not statistically significant in reptiles, it is qualitatively consistent with our findings in birds and mammals. Other selective processes (e.g. fertility selection for large female size in indeterminate growers, Cox *et al.* 2007) might have masked the influence of sexual selection on SSD in reptiles. Consistent with this explanation, selection often favours delayed maturation in female reptiles, which enables them to produce larger clutches, which in turn also influences their body size and the extent of SSD (Shine 2005; Agha et al. 2018). Follow-up studies using different proxies of sexual selection are needed to investigate further how sexual selection is related to ASR in reptiles.

Biased estimates of ASR may generate spurious relationship with SSD, which may potentially affect our results. For example, the larger sex may have lower detectability in polygamous species if some members of that sex are excluded from breeding sites (Ancona et al. 2017). However, highly polygamous species in which populations have been thoroughly surveyed showed skewed ASR even when all individuals in the population were accurately counted (Granjon et al. 2017), and fairly consistent ASR estimates were obtained when both breeding and non-breeding individuals were included (Emlen and Wrege 2004). In general, ASR estimates show a moderate but statistically significant repeatability across populations in

most of the studied taxa, except reptiles with temperature-dependent sex determination (Ancona et al. 2017; Bókony et al. 2019; Valentine Federico, J-F.L., J-M.G., A.L., I.P., T.S. unpublished result), and in 80% of bird species the direction of ASR skew is the same for all repeated estimates (Székely et al. 2014a).

The paths of causality in comparative data are difficult to untangle. Path analysis is a valuable tool for contrasting different causal models, although it cannot reveal causality (Shipley 2016). Path analysis assumes that each variable includes independent variations or 'errors' and that these errors are independent among variables. This is not true for comparative data, because the errors will be correlated across species. Our approach follows Santos (2012), an innovative but overlooked method that satisfies the assumptions of path analysis better than an alternative method based on phylogenetic regressions proposed by von Hardenberg and Gonzalez-Voyer (2013). As outlined in more detail in Appendix S3, this latter approach is problematic because it is not robust to changes in the specification of the model: if variable Y is regressed on X and λ estimated, then the estimates of the partial correlations and λ may be different from those obtained if Y is regressed on X with λ estimated. The approach we have taken avoids this problem. However, there is still room for methodological improvement. For instance, our approach has the drawback to be a 'subtractive' comparative method (sensu Harvey and Pagel 1991). Multivariate approaches such as path analyses for complex data have to be developed further.

Concluding remarks

Our findings indicate that sex-specific selection for large body size is associated with skewed ASRs across amniotes, and this process appears to produce SSD biased towards the rare sex in birds and mammals. Although this conclusion contrasts with Darwin's initial suggestion that intense sexual selection among males occurs when there is a surplus of males in the

521 population (Darwin 1871), theoretical and empirical work have suggests mechanisms that can
522 favour large size in the rare sex (Murray 1984; Klemme et al. 2007; Fitze and Le Galliard
523 2008; Dreiss et al. 2010). Further analyses of these processes and their application to species
524 with differing mating systems offer exciting opportunities for future investigations of the
525 interplay among sexual selection, SSD and ASR across the tree of life.
526

REFERENCES

Abouheif, E., and D. J. Fairbairn. 1997. A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. Am. Nat. 149:540–562.

Adkins-Regan, E., and H. K. Reeve. 2014. Sexual dimorphism in body size and the origin of sex-determination systems. Am. Nat. 183:519–536.

Agha, M., J. R. Ennen, A. J. Nowakowski, J. E. Lovich, S. C. Sweat, and B. D. Todd. 2018. Macroecological patterns of sexual size dimorphism in turtles of the world. J. Evol. Biol. 31:336–345.

Ancona, S., F. V. Dénes, O. Krüger, T. Székely, and S. R. Beissinger. 2017. Estimating adult sex ratios in nature. Philos. Trans. R. Soc. B Biol. Sci. 372:20160313.

Andersson, M. B. 1994. Sexual Selection. Princeton University Press, Princeton, New Jersey.

Anile, S., and S. Devillard. 2018. Camera-trapping provides insights into adult sex ratio variability in felids. Mamm. Rev. 48:168–179..

Ashman, T.-L., D. Bachtrog, H. Blackmon, E. E. Goldberg, M. W. Hahn, M. Kirkpatrick, J. Kitano, J. E. Mank, I. Mayrose, R. Ming, S. P. Otto, C. L. Peichel, M. W. Pennell, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, and J. C. Vamosi. 2014. Tree of Sex: A database of sexual systems. Sci. Data 1:140015.

Balasubramaniam, P., and J. T. Rotenberry. 2016. Elevation and latitude interact to drive life-history variation in precocial birds: a comparative analysis using galliformes. J. Anim. Ecol. 85:1528–1539.

Barley, A. J., P. Q. Spinks, R. C. Thomson, and H. B. Shaffer. 2010. Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. Mol. Phylogenet. Evol. 55:1189–1194.

Benito, M. M., and J. González-Solís. 2007. Sex ratio, sex-specific chick mortality and sexual size dimorphism in birds. J. Evol. Biol. 20:1522–1530.

Berger, J., and M. E. Gompper. 1999. Sex ratios in extant ungulates: products of contemporary predation or past life histories? J. Mammal. 80:1084–1113.

Blanckenhorn, W. U. 2005. Behavioral causes and consequences of sexual size dimorphism. Ethology 1016:977–1016.

Bókony, V., G. Milne, I. Pipoly, T. Székely, and A. Liker. 2019. Sex ratios and bimaturism differ between temperature-dependent and genetic sex-determination systems in reptiles. BMC Evol. Biol. 19:57.

Botero, C. A., R. Dor, C. M. McCain, and R. J. Safran. 2014. Environmental harshness is

positively correlated with intraspecific divergence in mammals and birds. Mol. Ecol. 23:259–268.

Clutton-Brock, T. 2017. Reproductive competition and sexual selection. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160310.

Clutton-Brock, T. H. 2016. Mammal Societies. Wiley-Blackwell.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness. 1985. Parental investment and sex differences in juvenile mortality in birds and mammals. Nature 313:131–133.

Clutton-Brock, T. H., P. H. Harvey, and B. Rudder. 1977. Sexual dimorphism, socionomic sex ratio and body weight in primates. Nature 269:797–800.

Cox, R. M., M. A. Butler, and H. B. John-Alder. 2007. The evolution of sexual size dimorphism in reptiles. Pp. 38–49 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles. Oxford University Press, Oxford.

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

Donald, P. F. 2007. Adult sex ratios in wild bird populations. Ibis 149:671–692.

Dreiss, A. N., J. Cote, M. Richard, P. Federici, and J. Clobert. 2010. Age-and sex-specific response to population density and sex ratio. Behav. Ecol. 21:356–364.

Eberhart-Phillips, L. J., C. Küpper, M. C. Carmona-Isunza, O. Vincze, S. Zefania, M. Cruz-López, A. Kosztolányi, T. E. X. Miller, Z. Barta, I. C. Cuthill, T. Burke, T. Székely, J. I. Hoffman, and O. Krüger. 2018. Demographic causes of adult sex ratio variation and their consequences for parental cooperation. Nat. Commun. 9:1651.

Emlen, S. T., and L. W. Oring. 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197:215–23.

Emlen, S. T., and P. H. Wrege. 2004. Size dimorphism, intrasexual competition, and sexual selection in Wattled jacana (Jacana jacana), a sex-role-reversed shorebird in Panama. Auk 121:391–403.

Fairbairn, D. J. 2007. Introduction: The enigma of sexual size dimorphism. Pp. 1–10 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism. Oxford University Press.

Fairbairn, D. J., W. U. Blanckenhorn, and T. Székely. 2007. Sex, Size and Gender Roles. Oxford University Press, Oxford.

Fischer, A. G. 1960. Latitudinal variations in organic diversity. Evolution 14:64–81. John

Fitze, P. S., and J. F. Le Galliard. 2008. Operational sex ratio, sexual conflict and the intensity

of sexual selection. Ecol. Lett. 11:432–439.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Fritz, S. A., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecol. Lett. 12:538–549.

Fromhage, L., and M. D. Jennions. 2016. Coevolution of parental investment and sexually selected traits drives sex-role divergence. Nat. Commun. 7:12517.

Gaillard, J.-M., M. Festa-Bianchet, N. G. Yoccoz, A. Loison, and C. Toïgo. 2000. Temporal Variation in Fitness Components and Population Dynamics of Large Herbivores. Annu. Rev. Ecol. Syst. 31:367–393.

Gamble, T., A. J. Geneva, R. E. Glor, and D. Zarkower. 2014. Anolis sex chromosomes are derived from a single ancestral pair. Evolution 68:1027–1041.

Georgiadis, N. 1985. Growth patterns, sexual dimorphism and reproduction in African ruminants. Afr. J. Ecol. 23:75–87.

Gonzalez-Voyer, A., and A. von Hardenberg. 2014. An introduction to phylogenetic path analysis. Pp. 201–229 *in* L. Z. Garamszegi, ed. Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer Berlin Heidelberg.

Granjon, A.-C., C. Rowney, L. Vigilant, and K. E. Langergraber. 2017. Evaluating genetic capture-recapture using a chimpanzee population of known size. J. Wildl. Manage. 81:279–288.

Grant, P. R., and B. R. Grant. 2019. Adult sex ratio influences mate choice in Darwin's finches. Proc. Natl. Acad. Sci. U. S. A. 116:12373–12382.

Guillon, J. M., L. Guéry, V. Hulin, and M. Girondot. 2012. A large phylogeny of turtles (Testudines) using molecular data. Contrib. to Zool. 81:147–158.

Haro, R. J., K. Edley, and M. J. Wiley. 1994. Body size and sex ratio in emergent stonefly nymphs (Isogenoides olivaceus: Perlodidae): variation between cohorts and populations. Can. J. Zool. 72:1371–1375.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press.

Hirst, A. G., and T. Kiørboe. 2014. Macroevolutionary patterns of sexual size dimorphism in copepods. Proc. R. Soc. B Biol. Sci. 281.

Isaac, J. L. 2005. Potential causes and life-history consequences of sexual size dimorphism in

626 mammals. Mamm. Rev. 35:101–115.
627 Janicke, T., I. K. Haderer, M. J. Lajeunesse, and N. Anthes. 2016. Darwinian sex roles
628 confirmed across the animal kingdom. Sci. Adv. 2:e1500983.
629 Janicke, T., and E. H. Morrow. 2018. Operational sex ratio predicts the opportunity and
630 direction of sexual selection across animals. Ecol. Lett. 21:384–391.
631 Jehl, J. R., and B. G. Murray. 1986. The evolution of normal and reverse sexual size
632 dimorphism in shorebirds and other birds. Pp. 1–86 *in* R. F. Johnston, ed. Current
633 Ornithology, vol. 3. Springer US, Boston, MA.
634 Jennions, M. D., and L. Fromhage. 2017. Not all sex ratios are equal: The Fisher condition,
635 parental care and sexual selection. Philos. Trans. R. Soc. B Biol. Sci. 372.
636 Johansson, F., P. H. Crowley, and T. Brodin. 2005. Sexual size dimorphism and sex ratios in
637 dragonflies (Odonata). Biol. J. Linn. Soc. 86:507–513.
638 Jones, K. S., S. Nakagawa, and B. C. Sheldon. 2009. Environmental sensitivity in relation to
639 size and sex in birds: meta-regression analysis. Am. Nat. 174:122–133.
640 Kalmbach, E., and M. M. Benito. 2007. Sexual size dimorphism and offspring vulnerability in
641 birds. Pp. 133–142 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex,
642 Size and Gender Roles. Oxford University Press.
643 Kappeler, P. M. 2017. Sex roles and adult sex ratios: insights from mammalian biology and
644 consequences for primate behaviour. Philos. Trans. R. Soc. B Biol. Sci. 372:20160321.
645 Klemme, I., H. Ylönen, and J. A. Eccard. 2007. Reproductive success of male bank voles
646 (Clethrionomys glareolus): the effect of operational sex ratio and body size. Behav. Ecol.
647 Sociobiol. 61:1911–1918.
648 Kokko, H., and M. D. Jennions. 2008. Parental investment, sexual selection and sex ratios. J.
649 Evol. Biol. 21:919–948.
650 Kokko, H., H. Klug, and M. D. Jennions. 2012. Unifying cornerstones of sexual selection:
651 operational sex ratio, Bateman gradient and the scope for competitive investment. Ecol.
652 Lett. 15:1340–1351.
653 Komdeur, J., T. Székely, X. Long, and S. A. Kingma. 2017. Adult sex ratios and their
654 implications for cooperative breeding in birds. Philos. Trans. R. Soc. B Biol. Sci. 372:5–
655 9.
656 Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for
657 ecology, evolution, and systematics. Methods Ecol. Evol. 7:573–579.
658 Lemaître, J. F., and J. M. Gaillard. 2013. Male survival patterns do not depend on male

allocation to sexual competition in large herbivores. Behav. Ecol. 24:421–428.

Lemaître, J. F., V. Ronget, M. Tidière, D. Allainé, V. Berger, A. Cohas, F. Colchero, D. A. Conde, M. Garratt, A. Liker, G. A. B. Marais, A. Scheuerlein, T. Székely, and J. M. Gaillard. 2020. Sex differences in adult lifespan and aging rates of mortality across wild mammals. Proc. Natl. Acad. Sci. U. S. A. 117:8546–8553.

Liker, A., R. P. Freckleton, V. Remeš, and T. Székely. 2015. Sex differences in parental care: Gametic investment, sexual selection, and social environment. Evolution 69:2862–2875.

Liker, A., R. P. Freckleton, and T. Székely. 2014. Divorce and infidelity are associated with skewed adult sex ratios in birds. Curr. Biol. 24:880–884.

Liker, A., R. P. Freckleton, and T. Székely. 2013. The evolution of sex roles in birds is related to adult sex ratio. Nat. Commun. 4:1587.

Liker, A., and T. Székely. 2005. Mortality costs of sexual selection and parental care in natural populations of birds. Evolution 59:890–897.

Lovich, J. E., J. W. Gibbons, and M. Agha. 2014. Does the timing of attainment of maturity influence sexual size dimorphism and adult sex ratio in turtles? Biol. J. Linn. Soc. 112:142–149.

Lukas, D., and T. H. Clutton-Brock. 2013. The evolution of social monogamy in mammals. Science 341:526–530.

Mayr, E. 1939. The Sex Ratio in Wild Birds. Am. Nat. 73:156–179.

McNamara, J. M., T. Székely, J. N. Webb, and A. I. Houston. 2000. A dynamic game-theoretic model of parental care. J. Theor. Biol. 205:605–623.

Meredith, R. W., J. E. Janecka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simao, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, and W. J. Murphy. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524.

Midford, P. E., T. J. Garland, and W. P. Maddison. 2011. PDAP:PDTREE module of Mesquite.

Mitani, J. C., J. Gros-Louis, and A. F. Richards. 1996. Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. Am. Nat. 147:966–980.

Moore, S. L., and K. Wilson. 2002. Parasites as a viability cost of sexual selection in natural

692 populations of mammals. Science 297:2015–2018.
693 Muralidhar, P., and M. A. Johnson. 2017. Sexual selection and sex ratios in Anolis lizards. J.
694 Zool. 302:178–183.
695 Murray, B. G. 1984. A demographic theory on the evolution of mating systems as
696 exemplified by birds. Pp. 71–140 *in* Hecht M.K., Wallace B., and Prance G.T., eds.
697 Evolutionary Biology. Springer US, Boston, MA.
698 Myhrvold, N. P., E. Baldridge, B. Chan, D. Sivam, D. L. Freeman, and S. K. M. Ernest. 2015.
699 An amniote life-history database to perform comparative analyses with birds, mammals,
700 and reptiles. Ecology 96:3109.
701 Nicholson, K. E., B. I. Crother, C. Guyer, and J. M. Savage. 2012. It is time for a new
702 classification of anoles (Squamata: Dactyloidae). Zootaxa 3477:1–108.
703 Oaks, J. R. 2011. A time-calibrated species tree of crocodylia reveals a recent radiation of the
704 true crocodiles. Evolution (N. Y). 65:3285–3297.
705 Orme, D., R. P. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2013.
706 caper: Comparative Analyses of Phylogenetics and Evolution in R. Available at:
707 https://cran.r-project.org/web/packa.
708 Owen-Smith, N. 1993. Comparative mortality rates of male and female kudus: the costs of
709 sexual size dimorphism. J. Anim. Ecol. 62:428.
710 Owens, I. P. F., and P. M. Bennett. 1994. Mortality costs of parental care and sexual
711 dimorphism in birds. Proc. R. Soc. B Biol. Sci. 257:1–8.
712 Paradis, E. 2012. Analysis of Phylogenetics and Evolution with R. Springer.
713 Pincheira-Donoso, D., and J. Hunt. 2017. Fecundity selection theory: concepts and evidence.
714 Biol. Rev. 92:341–356.
715 Pipoly, I., V. Bókony, M. Kirkpatrick, P. F. Donald, T. Székely, and A. Liker. 2015. The
716 genetic sex-determination system predicts adult sex ratios in tetrapods. Nature 527:91–
717 94.
718 Poulin, R. 1997. Covariation of sexual size dimorphism and adult sex ratio in parasitic
719 nematodes. Biol. J. Linn. Soc. 62:567–580.
720 Promislow, D. E. L. 1992. Costs of sexual selection in natural populations of mammals. Proc.
721 B Biol. Sci. 247:203–210.
722 Promislow, D. E. L., R. Montgomerie, and T. E. Martin. 1992. Mortality costs of sexual
723 dimorphism in birds. Proc. R. Soc. B Biol. Sci. 250:143–150.
724 Pyron, R., F. T. Burbrink, and J. J. Wiens. 2013. A phylogeny and revised classification of

Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13:93.

Queller, D. C. 1997. Why do females care more than males? Proc. R. Soc. London. Ser. B Biol. Sci. 264:1555–1557.

Remeš, V., R. P. Freckleton, J. Tökölyi, A. Liker, and T. Székely. 2015. The evolution of parental cooperation in birds. Proc. Natl. Acad. Sci. U. S. A. 112:13603–13608.

Rosseel, Y. 2012. Lavaan: An R package for structural equation modelling. J. Stat. Softw. 48:1–36.

Santos, J. C. 2012. Fast molecular evolution associated with high active metabolic rates in poison frogs. Mol. Biol. Evol. 29:2001–2018.

Sarre, S. D., T. Ezaz, and A. Georges. 2011. Transitions between sex-determining systems in reptiles and amphibians. Annu. Rev. Genomics Hum. Genet. 12:391–406.

Schacht, R., and M. Borgerhoff Mulder. 2015. Sex ratio effects on reproductive strategies in humans. R. Soc. Open Sci. 2:140402.

Schacht, R., K. L. Kramer, T. Székely, and P. M. Kappeler. 2017. Adult sex ratios and reproductive strategies: A critical re-examination of sex differences in human and animal societies. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160309.

Schacht, R., K. L. Rauch, and M. Borgerhoff Mulder. 2014. Too many men: the violence problem? Trends Ecol. Evol. 29:214–222.

Shine, R. 2005. Life-history evolution in reptiles. Annu. Rev. Ecol. Evol. Syst. 36:23–46.

Shipley, B. 2016. Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference with R. 2nd editio. Cambridge University Press.

Smith, R. J. 1999. Statistics of sexual size dimorphism. J. Hum. Evol. 36:423–458.

Spinks, P. Q., R. C. Thomson, M. Gidiş, and H. Bradley Shaffer. 2014. Multilocus phylogeny of the New-World mud turtles (Kinosternidae) supports the traditional classification of the group. Mol. Phylogenet. Evol. 76:254–260.

Székely, T., A. Liker, R. P. Freckleton, C. Fichtel, and P. M. Kappeler. 2014a. Sex-biased survival predicts adult sex ratio variation in wild birds. Proc. R. Soc. B Biol. Sci. 281:20140342.

Székely, T., T. Lislevand, and J. Figuerola. 2007. Sexual size dimorphism in birds. Pp. 27–37 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles. Oxford University Press, Oxford.

Székely, T., F. J. Weissing, and J. Komdeur. 2014b. Adult sex ratio variation: Implications for breeding system evolution. J. Evol. Biol. 27:1500–1512.

Tidière, M., J. M. Gaillard, D. W. H. Müller, L. B. Lackey, O. Gimenez, M. Clauss, and J. F. Lemaître. 2015. Does sexual selection shape sex differences in longevity and senescence patterns across vertebrates? A review and new insights from captive ruminants. Evolution 69:3123–3140.

Toïgo, C., and J. M. Gaillard. 2003. Causes of sex-biased adult survival in ungulates: Sexual size dimorphism, mating tactic or environment harshness? Oikos 101:376–384.

Trivers, R. L. 1972. Parental investment and sexual selection. Pp. 136–179 *in* In: Sexual Selection and the Descent of Man (ed. Campbell B), London: Heinemann.

Uetz, P., P. Freed, and J. (eds) Hošek. 2019. The Reptile Database.

van der Bijl, W. 2018. phylopath: Easy phylogenetic path analysis in R. PeerJ 2018:e4718.

von Hardenberg, A., and A. Gonzalez-Voyer. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution 67:378–387.

Wilson, K., and I. C. W. Hardy. 2002. Statistical analysis of sex ratios: an introduction. Pp. 48–92 *in* I. C. W. Hardy, ed. Sex Ratios. Cambridge University Press.

Wittenberger, J. F. 1976. The ecological factors selecting for polygyny in altrical birds. Am. Nat. 110:779–799.

Wittenberger, J. F. 1978. The evolution of mating systems in grouse. Condor 80:126–137.

Table 1. Phylogenetically corrected analyses of sexual size dimorphism (SSD) and adult sex ratio (ASR) in amniotes (reptiles, birds and mammals).

Predictors	$b \pm SE$	t	P	R^2	λ	n
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.183 ± 0. 036	5.037	**< 0.0001**	0.050	0.828 *#	462
Juvenile mortality bias	0.035 ± 0.070	0.493	0.623	< 0.001	0.737 *#	100
Adult mortality bias	0.014 ± 0.050	0.272	0.786	< 0.001	0.887 *#	230
Multi-predictor model 1:				*0.126*	*0.841 **	*457*
ASR	- 0.160 ± 0.035	4.555	**< 0.0001**			
Mean body mass	0.515 ± 0.087	5.950	**< 0.0001**			
Latitude	0.004 ± 0.038	0.103	0.918			
Sex determination, TSD [1]	- 0.297 ± 0.251	1.184	0.237			
Sex determination, ZW [1]	- 0.685 ± 0.264	2.592	**0.009**			
Multi-predictor model 2:				*0.273*	*0.814 *#*	*97*
ASR	- 0.271 ± 0.061	4.452	**< 0.0001**			
Mean body mass	0.377 ± 0.134	2.824	**0.006**			
Juvenile mortality bias	0.001 ± 0.060	0.011	0.992			
Adult mortality bias	- 0.019 ± 0.067	0.277	0.783			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.234 ± 0.051	4.593	**< 0.0001**	0.042	0.359 *#	462
Juvenile mortality bias	- 0.214 ± 0.099	2.151	**0.034**	0.035	0.281 *#	100
Adult mortality bias	- 0.257 ± 0.060	4.313	**< 0.0001**	0.071	0.288 *#	230
Multi-predictor model 1:				*0.071*	*0.247 *#*	*457*
SSD	- 0.188 ± 0.050	3.727	**0.0002**			
Mean body mass	- 0.106 ± 0.080	1.330	0.184			
Latitude	- 0.095 ± 0.045	2.135	**0.033**			
Sex determination, TSD [1]	0.481 ± 0.221	2.178	**0.030**			
Sex determination, ZW [1]	0.712 ± 0.205	3.471	**0.0006**			
Multi-predictor model 2:				*0.402*	*0.030 #*	*97*
SSD	- 0.457 ± 0.120	3.794	**0.0003**			
Mean body mass	- 0.249 ± 0.108	2.316	**0.023**			
Juvenile mortality bias	- 0.146 ± 0.086	1.702	0.092			
Adult mortality bias	- 0.259 ± 0.100	2.591	**0.011**			

Results of bivariate and multi-predictor phylogenetic generalized least-squares (PGLS) models with either (A) SSD ($\log_{10}$(male mass/female mass)) or (B) ASR (proportion of males in the adult population) as dependent variable. Mortality biases were calculated as $\log_{10}$(male mortality/female mortality) for juveniles and adults, respectively. $b \pm SE$ is the model's parameter estimate with its standard error (intercepts are not shown), t and P are the associated test statistic and its significance, λ is Pagel's lambda, n is number of species.
* λ statistically different from 0, # λ statistically different from 1.
[1] Differences from species with XY sex determination; overall effect of sex determination on SSD: $F_{2,451}$= 3.411, P= 0.034; on ASR: $F_{2,451}$= 6.135, P= 0.002.

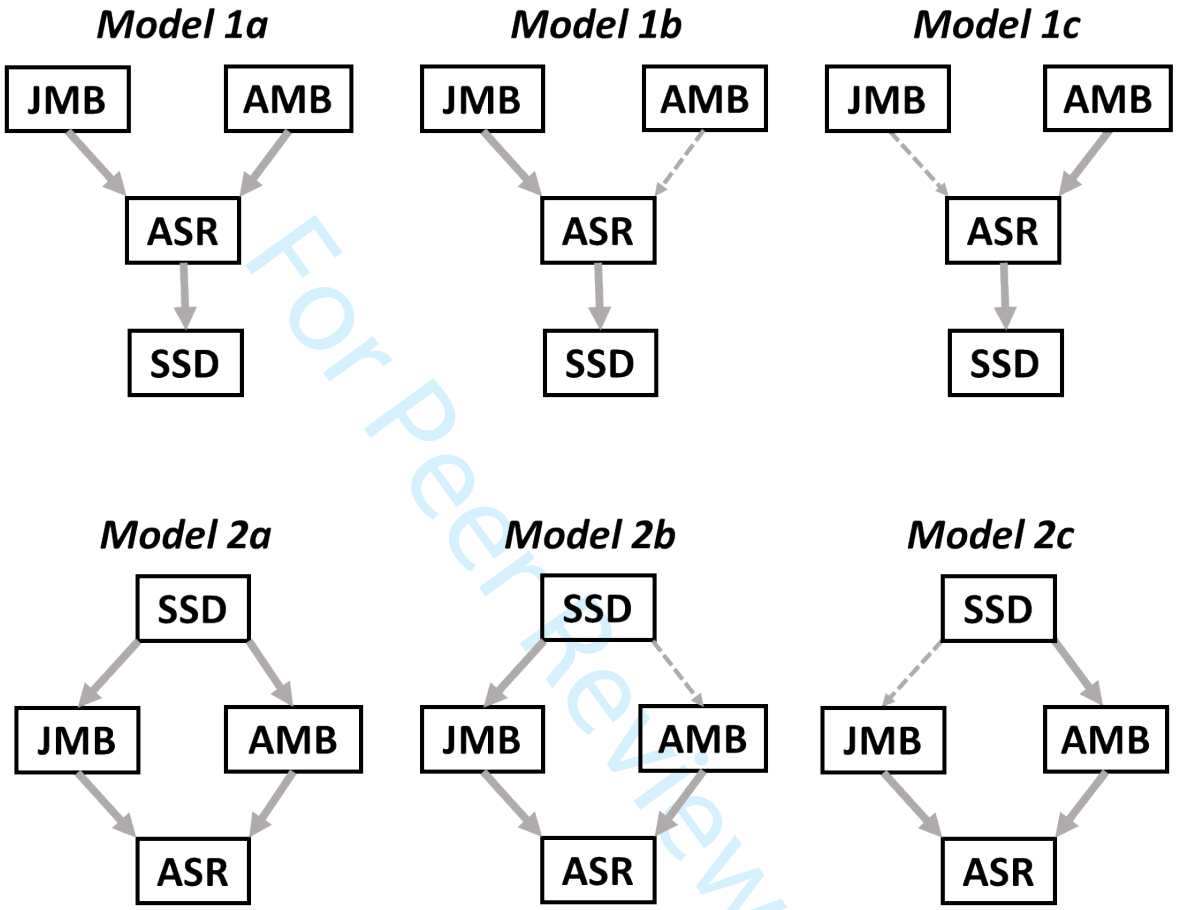
Table 2. Phylogenetic path models of the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c) in amniotes (reptiles, birds and mammals).

Model/Path	Path coefficient ± *SE*	*Z*	*P*
Model 1a	***P_C= 0.972, df= 4, AICc= 15.8, ΔAICc= 0.0***		
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 1b	***P_C= 0.065, df= 6, AICc= 25.7, ΔAICc= 9.9***		
(AMB → ASR)[1]	0	-	-
JMB → ASR	- 0.258 ± 0.107	- 2.417	**0.018**
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 1c	***P_C= 0.376, df= 6, AICc= 19.9, ΔAICc= 4.1***		
AMB → ASR	- 0.378 ± 0.113	- 3.334	**0.001**
(JMB → ASR)[1]	0	-	-
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 2a	*P_C= 0.0, df= 4, AICc= 59.0, ΔAICc= 43.2*		
SSD → AMB	0.171 ± 0.105	1.631	0.106
SSD → JMB	0.111 ± 0.115	0.958	0.341
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
Model 2b	*P_C= 0.0, df= 4, AICc= 50.4, ΔAICc= 34.6*		
SSD → JMB	0.111 ± 0.115	0.958	0.341
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
Model 2c	*P_C= 0.0, AICc= 50.4, ΔAICc= 34.6*		
SSD → AMB	0.171 ± 0.105	1.631	0.106
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052

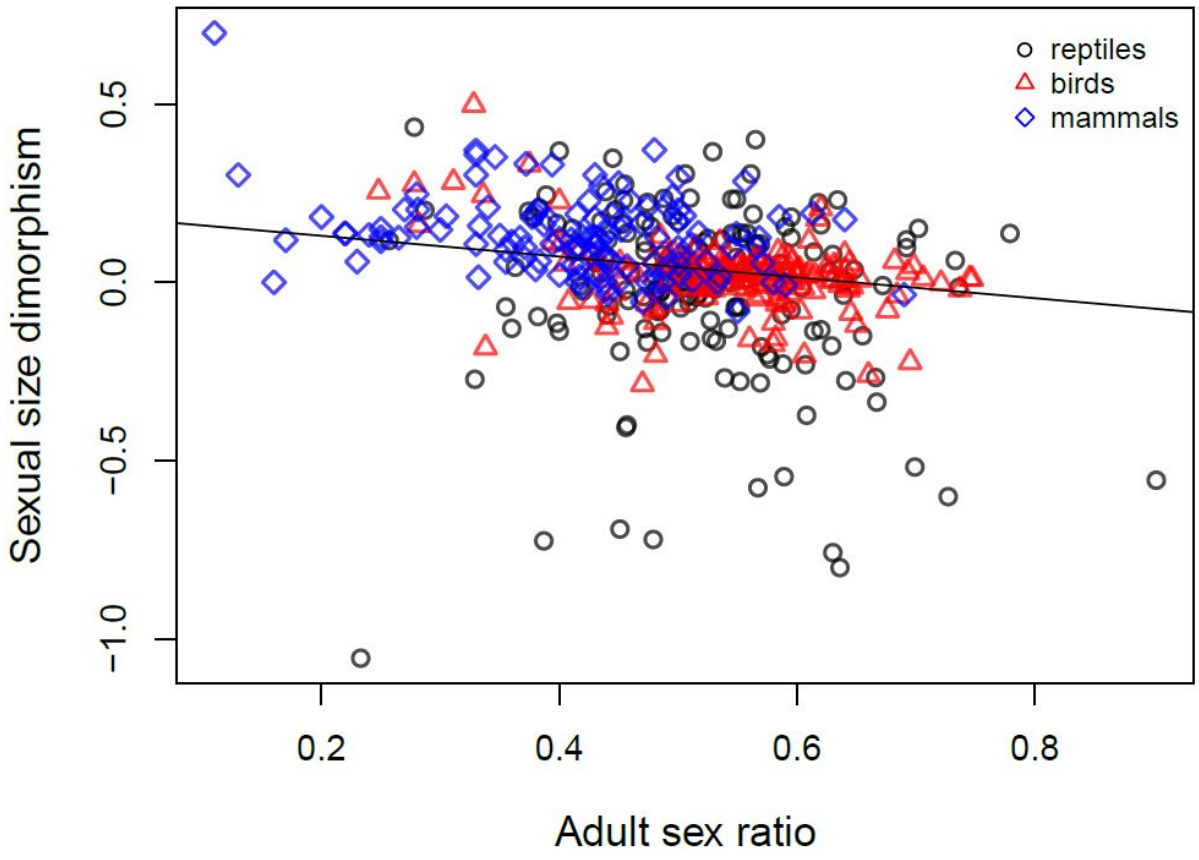
Model structures are shown in Figure S1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively (variables are explained in footnotes of Table 1). P_C is *P*-value for Fisher's *C* statistic for model fit, with non-significant values (> 0.05) indicating an acceptable fit. ΔAICc indicates difference in AICc values between the most supported model (lowest AICc, Model 1a) and the focal models. ΔAICc > 2 indicates substantially higher support for the best model than for the other models. The analyses include 97 species of reptiles, birds and mammals with data for all for variables.

[1] Path coefficient set to zero to keep the variable in the model.

Figure 1. Path models tested in the phylogenetic path analyses. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB: juvenile mortality bias, AMB: adult mortality bias. Dashed arrows indicate paths with coefficients set to zero to keep the variable in the model. Models 1a-c and 2a-c represent relationships as predicted by the mating opportunity hypothesis and the mortality cost hypothesis, respectively.



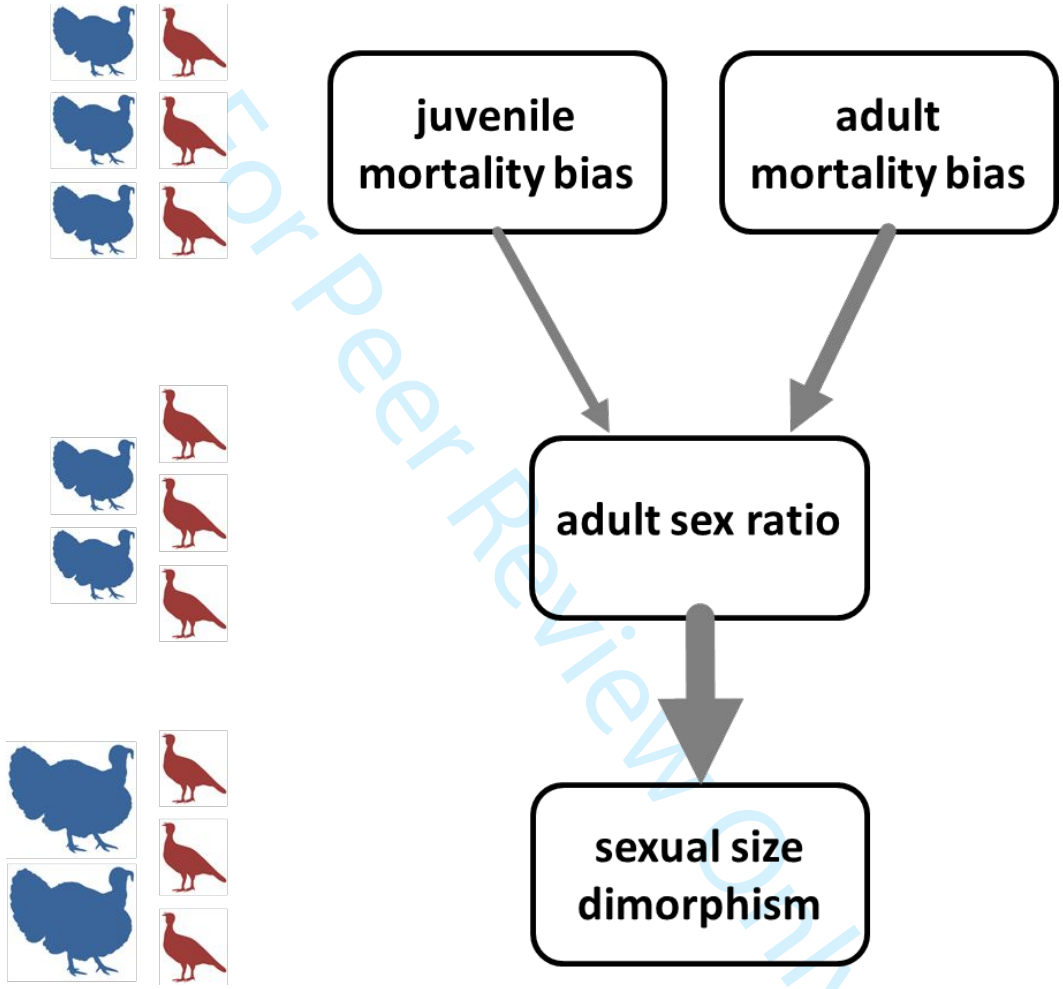
843 **Figure 2**. Sexual size dimorphism (SSD) in relation to adult sex ratio (ASR) in amniotes.
844 SSD was calculated as $\log_{10}$(male mass/female mass); ASR is the proportion of males in the
845 adult population. Each data point represents a species; the regression line is fitted by
846 phylogenetic generalized least-squares (PGLS) model (see Table 1 for statistics).
847
848



849

850

Figure 3. Path diagram of the best-fitting phylogenetic path model (Model 1a in Table 2, *n* = 97 species of reptiles, birds and mammals). The model supports the scenario that sex-biased juvenile and adult mortalities lead to skewed adult sex ratio, which in turn results in increased size dimorphism by sexual selection. Width of the arrows is proportional to path coefficients (see Table 2 for statistical details of the model). Bird pictures on the left illustrate the case when differential mortality generates female-skewed ASR, which then leads to a more male-biased SSD (i.e. larger body size in males relative to females). The path analyses were based on the approach proposed by Santos (2012), see Appendix S3 for details.



Electronic Supporting Information: tables and figures

Table S1. Relationship between SSD, ASR and sex-biased mortalities in reptiles, using estimated body mass data for SSD calculation.

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.060 ± 0.052	1.143	0.253	0.002	0.931 *#	155
Juvenile mortality bias	- 0.109 ± 0.102	1.065	0.304	0.008	0.352	17
Adult mortality bias	- 0.142 ± 0.088	1.610	0.113	0.030	1.0 *	62
Multi-predictor model 1:				*0.116*	*0.956 **	*153*
ASR	- 0.090 ± 0.075	1.203	0.231			
Mean body mass	0.715 ± 0.178	4.019	< **0.0001**			
Latitude	- 0.175 ± 0.126	1.389	0.167			
Reproductive mode [1]	0.348 ± 0.313	1.112	0.268			
Sex determination, TSD [2]	- 0.463 ± 0.384	1.206	0.230			
Sex determination, ZW [2]	- 1.003 ± 0.313	2.344	**0.020**			
Multi-predictor model 2:				*< 0.001*	*0.0*	*17*
ASR	- 0.022 ± 0.252	0.086	0.933			
Mean body mass	- 0.452 ± 0.523	0.865	0.404			
Juvenile mortality bias	- 0.500 ± 0.374	1.339	0.205			
Adult mortality bias	0.284 ± 0.429	0.662	0.520			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.145 ± 0.080	1.819	0.071	0.015	0.0 #	155
Juvenile mortality bias	- 0.251 ± 0.217	1.156	0.266	0.021	0.0#	17
Adult mortality bias	- 0.179 ± 0.103	1.732	0.088	0.032	0.155 #	62
Multi-predictor model 1:				*0.078*	*0.0 #*	*153*
SSD	- 0.049 ± 0.055	0.891	0.374			
Mean body mass	0.173 ± 0.108	1.599	0.112			
Latitude	- 0.001 ± 0.109	0.013	0.990			
Reproductive mode [1]	- 0.140 ± 0.216	0.650	0.517			
Sex determination, TSD [2]	0.209 ± 0.224	0.934	0.352			
Sex determination, ZW [2]	0.667 ± 0.216	3.091	**0.002**			
Multi-predictor model 2:				*0.165*	*0.0 #*	*17*
SSD	- 0.028 ± 0.331	0.086	0.933			
Mean body mass	0.929 ± 0.556	1.671	0.121			
Juvenile mortality bias	- 0.044 ± 0.459	0.095	0.926			
Adult mortality bias	- 0.641 ± 0.465	1.377	0.194			

* Pagel's lambda statistically different from 0, # lambda statistically different from 1.

[1] Differences from oviparous species.

[2] Differences from XY species; overall effect of sex determination on SSD: $F_{2,146}$= 2.5, P= 0.083; on ASR: $F_{2,146}$= 5.3, P= 0.006.

For further explanation, see the footnotes of Table 1 in the main text.

872 **Table S2.** Relationship between SSD, ASR and sex-biased mortalities in reptiles, using body
873 length data for SSD calculation.
874

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.060 ± 0.053	1.132	0.259	0.002	0.924 *#	155
Juvenile mortality bias	- 0.141 ± 0.115	1.221	0.241	0.030	0.531	17
Adult mortality bias	- 0.154 ± 0.090	1.707	0.093	0.030	1.0 *	62
Multi-predictor model 1:				*0.122*	*0.52 **	*153*
ASR	- 0.048 ± 0.042	1.126	0.262			
Mean body mass	0.391 ± 0.100	3.891	**< 0.001**			
Latitude	- 0.103 ± 0.071	1.459	0.147			
Reproductive mode [1]	0.179 ± 0.177	1.015	0.312			
Sex determination, TSD [2]	- 0.223 ± 0.216	1.032	0.304			
Sex determination, ZW [2]	- 0.633 ± 0.241	2.628	**0.009**			
Multi-predictor model 2:				*< 0.001*	*0.0*	*17*
ASR	- 0.059 ± 0.161	0.368	0.719			
Mean body mass	- 0.347 ± 0.334	1.038	0.320			
Juvenile mortality bias	- 0.354 ± 0.239	1.484	0.164			
Adult mortality bias	0.092 ± 0.274	0.337	0.742			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.147 ± 0.080	1.841	0.068	0.015	0.0 #	155
Juvenile mortality bias	- 0.251 ± 0.217	1.156	0.266	0.021	0.0#	17
Adult mortality bias	- 0.179 ± 0.103	1.732	0.088	0.032	0.155 #	62
Multi-predictor model 1:				*0.078*	*0.0 #*	*153*
SSD	- 0.085 ± 0.098	0.870	0.386			
Mean body mass	0.172 ± 0.109	1.588	0.114			
Latitude	- 0.001 ± 0.108	0.007	0.994			
Reproductive mode [1]	- 0.141 ± 0.216	0.654	0.514			
Sex determination, TSD [2]	0.214 ± 0.223	0.958	0.340			
Sex determination, ZW [2]	0.667 ± 0.216	3.089	**0.002**			
Multi-predictor model 2:				*0.174*	*0.0 #*	*17*
SSD	- 0.188 ± 0.512	0.368	0.720			
Mean body mass	0.867 ± 0.570	1.522	0.154			
Juvenile mortality bias	- 0.096 ± 0.463	0.208	0.839			
Adult mortality bias	- 0.624 ± 0.457	1.366	0.197			

875
876 * Pagel's lambda statistically different from 0, # lambda statistically different from 1.
877 [1] Differences from oviparous species.
878 [2] Differences from XY species; overall effect of sex determination on SSD: $F_{2,146}$= 3.4, *P*=
879 0.035; on ASR: $F_{2,146}$= 5.2, *P*= 0.006.
880 For further explanation, see the footnotes of Table 1 in the main text.
881

882 **Table S3.** Relationship between SSD, ASR and sex-biased mortalities in birds.
883

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.393 ± 0.055	7.186	**< 0.001**	0.216	0.770 *#	185
Juvenile mortality bias	0.185 ± 0.167	1.108	0.274	0.005	0.861 *	47
Adult mortality bias	0.131 ± 0.082	1.608	0.110	0.013	0.625 *#	123
Multi-predictor model 1:				*0.272*	*0.720 *#*	*178*
ASR	- 0.242 ± 0.038	6.390	**< 0.0001**			
Mean body mass	0.261 ± 0.100	2.599	**0.010**			
Latitude	- 0.020 ± 0.034	0.584	0.560			
Display type, agile [1]	- 0.338 ± 0.090	3.748	**< 0.001**			
Multi-predictor model 2:				*0.386*	*1.0 **	*47*
ASR	- 0.346 ± 0.080	4.318	**< 0.0001**			
Mean body mass	0.424 ± 0.246	1.719	0.093			
Juvenile mortality bias	- 0.032 ± 0.065	0.489	0.627			
Adult mortality bias	- 0.139 ± 0.141	0.992	0.327			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.499 ± 0.077	6.520	**< 0.001**	0.184	0.480 *#	185
Juvenile mortality bias	- 0.494 ± 0.160	3.084	**0.003**	0.156	0.0 #	47
Adult mortality bias	- 0.401 ± 0.082	4.866	**< 0.001**	0.157	0.0 #	123
Multi-predictor model 1:				*0.239*	*0.244 #*	*178*
SSD	- 0.717 ± 0.116	6.183	**< 0.0001**			
Mean body mass	- 0.191 ± 0.136	1.406	0.161			
Latitude	- 0.127 ± 0.058	2.201	**0.029**			
Display type, agile [1]	- 0.589 ± 0.161	3.667	**< 0.001**			
Multi-predictor model 2:				*0.397*	*0.0 #*	*47*
SSD	- 0.382 ± 0.153	2.499	**0.016**			
Mean body mass	- 0.128 ± 0.198	0.646	0.522			
Juvenile mortality bias	- 0.199 ± 0.109	1.831	0.074			
Adult mortality bias	- 0.468 ± 0.139	3.368	**0.002**			

884
885 * Pagel's lambda statistically different from 0, # lambda statistically different from 1.
886 [1] Difference from non-agile species.
887 For further explanation, see the footnotes of Table 1 in the main text.
888

889

890 **Table S4.** Relationship between SSD, ASR and sex-biased mortalities in mammals.
891

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.314 ± 0.088	3.561	**< 0.001**	0.088	0.190 ǂ	122
Juvenile mortality bias	0.089 ± 0.216	0.414	0.682	< 0.001	0.096 ǂ	36
Adult mortality bias	0.071 ± 0.171	0.414	0.681	< 0.001	0.039 ǂ	45
Multi-predictor model 1:				*0.166*	*0.342* *ǂ	*120*
ASR	- 0.153 ± 0.058	2.646	**0.009**			
Mean body mass	0.418 ± 0.131	3.191	**0.002**			
Latitude	0.106 ± 0.053	2. 016	**0.046**			
Multi-predictor model 2:				*0.250*	*0.0* ǂ	*33*
ASR	- 0.374 ± 0.129	2.900	**0.007**			
Mean body mass	0.209 ± 0.237	0.880	0.386			
Juvenile mortality bias	0.088 ± 0.116	0.767	0.452			
Adult mortality bias	- 0.016 ± 0.119	0.134	0.894			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.302 ± 0.085	3.539	**< 0.001**	0.087	0.252 *ǂ	122
Juvenile mortality bias	- 0.036 ± 0.184	0.195	0.847	< 0.001	0.0 ǂ	36
Adult mortality bias	- 0.080 ± 0.163	0.493	0.624	< 0.001	0.0 ǂ	45
Multi-predictor model 1:				*0.008*	*0.367* *ǂ	*114*
SSD	- 0.375 ± 0.140	2.670	**0.009**			
Mean body mass	- 0.314 ± 0.209	1.500	0.136			
Latitude	- 0.075 ± 0.083	0.907	0.366			
Multi-predictor model 2:				*0.293*	*0.0* ǂ	*33*
SSD	- 0.617 ± 0.213	2.900	**0.007**			
Mean body mass	- 0.494 ± 0.294	1.678	0.104			
Juvenile mortality bias	- 0.043 ± 0.150	0.285	0.778			
Adult mortality bias	0.022 ± 0.153	0.142	0.888			

892
893 * Pagel's lambda statistically different from 0, ǂ lambda statistically different from 1.
894 For further explanation, see the footnotes of Table 1 in the main text.
895
896

Table S5. Sensitivity analyses of the relationship between sexual size dimorphism (SSD, dependent variable in all models) and adult sex ratio (ASR). Table shows results when (A) reptiles are included with SSD based on body length, (B) reptiles are included with SSD calculated from sex-specific body mass, (C) reptiles with temperature-dependent sex determination (TSD) are excluded, (D) reptiles with assumed sex determination, based on related species, are excluded, and (E) environmental harshness is included in the model.

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) Bivariate model, reptiles included with SSD calculated from body length:						
ASR (all species)	- 0.252 ± 0.039	6.492	**< 0.0001**	0.082	0.664 *#	462
(B) Bivariate model, reptiles included with SSD calculated from body mass[1]:						
ASR (all species)	- 0.306 ± 0.051	6.037	**< 0.0001**	0.095	0.745 *#	338
(C) Bivariate models, TSD reptiles excluded:						
ASR (all species)	- 0.271 ± 0.045	6.077	**< 0.0001**	0.082	0.756 *#	402
ASR (reptiles only)	- 0.119 ± 0.131	0.909	0.366	< 0.001	0.915 *	95
(D) Bivariate model, assumed sex determination excluded[2]:						
ASR (all species)	- 0.186 ± 0.038	4.968	**< 0.0001**	0.055	0.817 *#	409
(E) Multi-predictor model with environmental harshness[3]:						
birds and mammals:				*0.110*	*0.757 *#*	*219*
ASR	- 0.192 ± 0.036	5.525	**< 0.0001**			
Environmental harshness	0.028 ± 0.046	0.618	0.538			
reptiles:				*0.026*	*0.944 **	*58*
ASR	- 0.033 ± 0.133	0.246	0.807			
Environmental harshness	0.117 ± 0.063	1.865	0.066			
all species:				*0.078*	*0.849 *#*	*277*
ASR	- 0.167 ± 0.039	4.301	**< 0.0001**			
Environmental harshness	0.082 ± 0.034	2.439	**0.015**			

* Pagel's lambda statistically different from 0, # lambda statistically different from 1

[1] Sex-specific body mass data from Myhrvold et al. (2015).

[2] Sex determination mechanism assumed to be the same type as reported for the genus or family (see Methods).

[3] The influence of environmental harshness was tested in birds and mammals using data from Botero et al. (2014), in reptiles using data calculated in this study (following the method of Botero et al 2014), and in all species by pooling the harshness scores from the two studies.

Table S6. Analyses of the relationship between SSD (dependent variable) and ASR with branch lengths calculated by three different methods for the phylogeny used in the PGLS models. The analyses included reptiles, birds, and mammals.

Branch length method[1]	$b \pm SE$	t	P	R^2	λ	n
Nee's method	- 0.183 ± 0.036	5.037	**< 0.0001**	0.050	0.828 *#	462
Pagel's method	- 0.180 ± 0.036	4.966	**< 0.0001**	0.049	0.839 *#	462
unit branch length	- 0.195 ± 0.033	5.833	**< 0.0001**	0.067	1.0 *	462

* Pagel's lambda statistically different from 0, # lambda statistically different from 1
[1] See Methods for details of branch length calculations

Table S7. Analyses of the relationship between SSD (dependent variable) and ASR in socially monogamous and socially polygamous species, respectively. The analyses included birds and mammals.

Mating system[1]	$b \pm SE$	t	P	R^2	λ	n
Monogamy	- 0.070 ± 0.038	1.829	0.070	0.021	1.0 *	109
Polygamy	- 0.275 ± 0.049	5.599	**< 0.0001**	0.159	0.385 *#	162

* Pagel's lambda statistically different from 0, # lambda statistically different from 1
[1] When monogamous and polygynous species are analysed together, there is a statistically significant interaction between the effects of mating system and ASR ($b \pm SE = 0.218 \pm 0.087$, $t = 2.510$, $P = 0.013$).

Table S8. Results of the phylogenetic path analyses using the R package 'phylopath'. Models represent the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c). Analyses based on data of all species (birds, mammals, and reptiles; *n*= 97 species).

Model	*k*	*q*	*C*	*P*	*CICc*	*ΔCICc*
Model 1a	**3**	**7**	**6.4**	**0.383**	**21.6**	**0.0**
Model 1b	4	6	18.7	0.017	31.6	10.0
Model 1c	**4**	**6**	**11.2**	**0.188**	**24.2**	**2.6**
Model 2a	2	8	32.4	<0.001	50.0	28.4
Model 2b	3	7	34.8	<0.001	50.0	28.4
Model 2c	3	7	36.6	<0.001	51.9	30.3

Model structures are shown in Figure 1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively. The table shows the number of independence claims (*k*), the number of parameters (*q*), Fisher's C statistic (*C*) and its accompanying probability (*P*), C-statistic information criterion corrected for small sample sizes (CICc), and the difference in CICc from the top model (ΔCICc). A *P*-value less than 0.05 indicates a poor model fit (i.e. rejection of the model), whereas a ΔCICc > 2 indicates substantial support for the top path model over the alternative models.

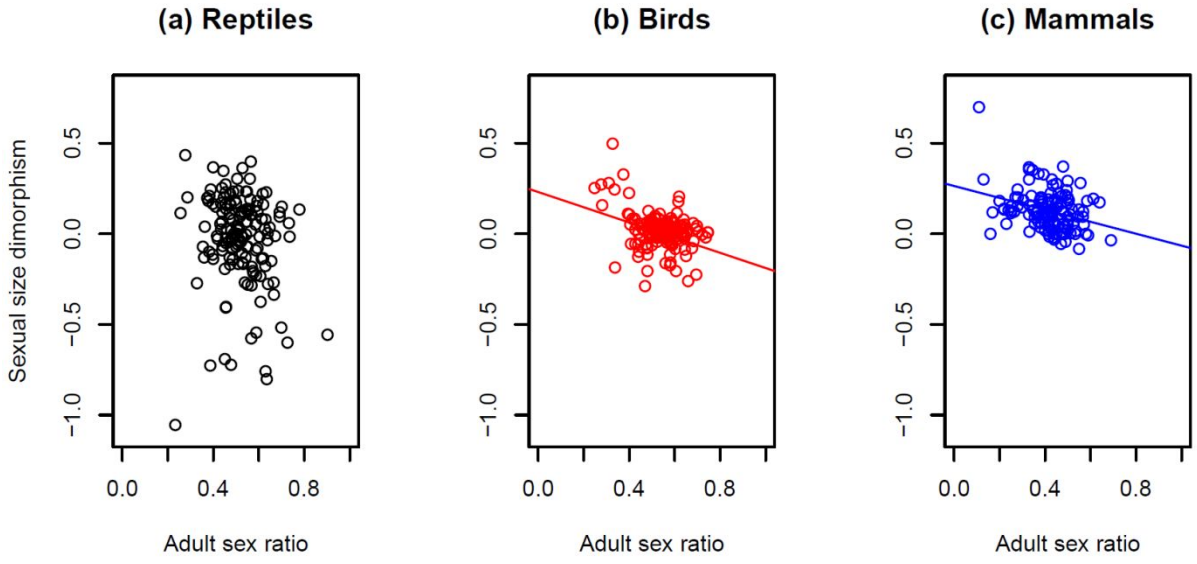
Table S9. Phylogenetic path models representing the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c). Analyses with data of birds and mammals (i.e. excluding reptiles; *n*= 81 species).

Model/Path	**Path coefficient ± *SE***	***Z***	***P***
Model 1a	***P_C= 0.991, df= 4, AICc= 15.9, ΔAICc= 0.0***		
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 1b	***P_C= 0.142, df= 6, AICc= 23.7, ΔAICc= 7.8***		
(AMB → ASR)[1]	0	-	-
JMB → ASR	- 0.257 ± 0.112	- 2.289	**0.025**
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 1c	***P_C= 0.428, df= 6, AICc= 19.7, ΔAICc= 3.8***		
AMB → ASR	- 0.358 ± 0.121	- 2.958	**0.004**
(JMB → ASR)[1]	0	-	-
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 2a	*P_C= 0.0, df= 4, AICc= 59.7, ΔAICc= 43.8*		
SSD → AMB	0.117 ± 0.077	1.503	0.137
SSD → JMB	0.102 ± 0.086	1.187	0.239
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
Model 2b	*P_C= 0.0, df= 4, AICc= 50.5, ΔAICc= 34.6*		
SSD → JMB	0.102 ± 0.086	1.187	0.239
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
Model 2c	*P_C= 0.0, AICc= 50.5, ΔAICc= 34.6*		
SSD → AMB	0.117 ± 0.077	1.503	0.137
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059

Model structures are shown in Figure 1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively (variables are explained in footnotes of Table 1). P_C is *P*-value for Fisher's *C* statistic for model fit, with non-significant values (> 0.05) indicating an acceptable fit. ΔAICc indicates difference in AICc values between the most supported model (lowest AICc, Model 1a) and the focal models. ΔAICc > 2 indicates substantially higher support for the best model than for the other model.

[1] Path coefficient set to zero to keep the variable in the model.

Figure S1. Sexual size dimorphism in relation to adult sex ratio in (a) reptiles (PGLS, b ± *SE* = -0.060 ± 0.052, *P*= 0.253, *n*= 155 species), (b) birds (*b* ± *SE* = -0.392 ± 0.055, *P*< 0.001, *n*= 185), and (c) mammals (*b* ± *SE* = -0.314 ± 0.088, *P* < 0.001, *n*= 122). Each data point represents a species, and lines show statistically significant regressions fitted by PGLS (see Tables S1-4 for further statistical details).



Electronic Supporting Information: Appendix S1

References for data sources are provided in a separate excel file. The full dataset will be published together with references after the manuscript is accepted for publication.

Electronic Supporting Information: Appendix S2

Parameters of the allometric equations between body length and body mass (log10(mass in g) = a + b*log10(length in mm)) used for the calculation of mass estimates in reptiles. n is the number of species included in the analyses.

Taxon	**Intercept (*a*)**	**Slope (*b*)**	***n***	**Reference**
Snakes	-5.773	2.786	336	Feldman and Meiri (2013)
Squamates and crocodilians	-4.52	2.923	600	Meiri (2010)
Turtles	-3.535	2.887	199	Regis and Meik (2017)

References:

Feldman, A., and S. Meiri. 2013. Length-mass allometry in snakes. Biol. J. Linn. Soc. 108:161–172.

Meiri, S. 2010. Length-weight allometries in lizards. J. Zool. 281:218–226.

Regis, K. W., and J. M. Meik. 2017. Allometry of sexual size dimorphism in turtles: a comparison of mass and length data. PeerJ 5:e2914.

Electronic Supporting Information: Appendix S3

Methodological notes on path analyses applied to comparative data

Several approaches have been proposed for applying path analysis in phylogenetic comparative studies (von Hardenberg and Gonzalez-Voyer 2013). However, some methodological aspects of this method still pose challenges for its phylogenetic applications. Here we discuss two of such aspects: (1) the problem of using bivariate phylogenetic generalized least squares (PGLS) models to estimate correlations between variables in the path models, and (2) the reliability of AIC statistics to compare non-nested path models. We suggest ways to avoid these problems which may help further applications of path analysis to phylogenetic data.

1. The problem of using PGLS in path analyses

The analysis of multivariate dependent data is a notoriously thorny problem. In this dataset we, as in common with many similar analyses, had to deal with the problem of phylogenetic non-independence (Harvey and Pagel 1991). Techniques for analysing cause-effect relationships using linear models are well developed (Felsenstein 1985; Harvey and Pagel 1991; Hansen and Martins 1996; Pagel 1997; Freckleton et al. 2002). These techniques are designed for analysis of data in which there is a dependent variable of interest, which is modelled as a function of independent predictors. In these models the effect of phylogeny is accounted for by modelling phylogenetic dependence in the residual term, and we have used these for several analyses (e.g. Table 1 in the main text). However, analyses of data in which variables are treated as multivariate responses are much less common. Phylogenetic principal components analysis is one exception, although recent research has stressed that this may be more complex than previously realised (Uyeda et al. 2015).

Path analysis is a method of multivariate trait analysis that allows complex dependencies among variables to be modelled. von Hardenberg & Gonzalez-Voyer (2013) presented a new method for performing path analysis on phylogenetically dependent data. This approach accounts for phylogenetic non-independence through constructing a series of bivariate PGLS models in which one variable is treated as the dependent variable, and the other is the independent variable. Importantly, this approach permits variable levels of

dependence to be modelled through estimating Pagel's λ, which accounts for varying contributions of phylogeny to trait variation.

However the analysis of such data is complex. Consider a simple example, in which we have three variables x, y, and z. If we model $x \rightarrow y$ (i.e. where x is the predictor and y is the response variable) using the PGLS-λ approach, we might well estimate a different correlation than if we model $y \rightarrow x$. This is because in PGLS the estimate of λ depends on the direction of the relationship. In path analysis correlations between pairs of variables are the input and the directionality of the statistical model should not be an issue. Specifically, the net correlation r_{xz} should then be the product of the component pairwise correlations, i.e. $r_{xz} = r_{xy} \times r_{yz}$. However, if we use PGLS and $\lambda \neq 0$ and $\lambda \neq 1$ (like in most cases in our analyses, see Table 1 in the main text), then in general $r_{xz} = r_{xy} \times r_{yz} \neq r_{yx} \times r_{yz} = r_{xz}$.

A further issue is that when we extend to several variables, the approach can produce correlations that are difficult to compare because they have different evolutionary interpretations. For instance, if the value of λ is 1 (strong phylogenetic signal) for the path $x \rightarrow y$, but the value is 0 (no phylogenetic signal) for $y \rightarrow z$, then the overall model relating x to z is not consistent in an evolutionary sense. This is because the correlation between x and y is an evolutionary one, but the relationship between y and z is not. This is equivalent to saying that the two correlations r_{xy} and r_{yz} are measured on two different scales. To see this, in this specific example we can estimate r_{xy} from phylogenetic contrasts, by transforming x to $p(x|tr)$ and y to $p(y|tr)$, where the transformation $p()$is the contrasts given tree tr. The correlation estimated for x and y is thus, effectively, the correlation between p(x) and p(y), i.e. is effectively $r_{p(x)\,p(y)}$. In contrast the correlation between y and z is estimated on the untransformed scale, i.e. r_{yz}. Thus, the variable y is not the same in the two cases, which calls into question the validity of calculating a net correlation from r_{xy} and r_{yz} estimated using different error models.

In our analyses we overcame the above difficulties by avoiding the use of bivariate PGLS models to estimate correlations between the variables in the path models. We followed the approach proposed by Santos (2012), in which first we calculated λ for each individual variable and conducted a phylogenetic transformation on the variable using that estimate of λ. Then we fitted path models to these already transformed data using ordinary fitting methods developed for non-phylogenetic data. Details of the analyses are described in the Methods section of the main text.

2. The problem of comparing non-nested path models by AIC statistics

In phylogenetic comparative studies the direction of causality between variables is often unknown, and different evolutionary hypotheses may propose opposing cause - effect relationships (like the mating competition and the mortality cost hypotheses in our study, see Fig. S1). These hypotheses may be represented by different path models, and then their fit to the data can be compared by some comparative fit indices, most commonly by AIC (West et al. 2012). However, simulations suggest that conclusions of path model comparisons based on information theory approach (like AIC) can be unreliable (Preacher and Merkle 2012). In addition the competing models can be non-nested (non-hierarchical) (e.g. Models 1a versus 2a in Fig. S1), for which AIC-based comparison should be applied with caution (Kline 2015).

To explore the problem of model comparison in the context of our study, first we fitted our path models to the real dataset by two alternative methods: (1) by covariance matrix comparison, as implemented in the R package *lavaan* (Rosseell 2012), and (2) by piecewise structural equation modelling (or d-separation) method, as implemented in the *piecewiseSEM* (Lefcheck 2016) package. We compared path coefficient estimates and various model fit indices between these two methods to evaluate whether they produce consistent conclusions. Second, we used the same two methods and R implementations to fit the models to simulated datasets, and tested which of the methods produces more reliable (less biased) model comparisons.

2.1. Fitting path models to real data

The general steps of model fitting procedure we followed in this study are described in the Methods section of the main text. We performed model fitting with the two R packages *piecewiseSEM* and *lavaan*. In *piecewiseSEM* and *lavaan* the global model fit for each individual path model is evaluated by Fisher's C and χ^2 statistics, respectively, where a statistically non-significant result means acceptable fit. In *lavaan*, several other measures for model fit of individual models are also available, and here we report four of the most widely used indices (TLI, CFI, RMSEA, SRMR). It has been proposed that that the values of TLI and CFI > 0.95, RMSEA < 0.06, and SRMR < 0.08 indicate acceptable/good fit of models to the data (West et al. 2012).

We found that the two methods produced highly consistent estimates for the standardised path coefficients in all path models (*piecewiseSEM*: Table 1 in the main text, *lavaan*: Table S9 below). The effect of juvenile mortality on ASR was marginally not

significant in most *piecewiseSEM* models whereas it was significant with all *lavaan* models. For all other relationships the two methods produced consistent results.

Table S9. Estimates of standardised path coefficients for the six path models representing various relationships between SSD, ASR, and sex biases in adult (AMB) and juvenile (JMB) mortality, obtained by the R package *lavaan* (see Fig. S1 for model details). Significant relationships are highlighted in bold.

Model/Path	Path coefficient ± SE	Z	P
Model 1a			
AMB → ASR	**- 0.340 ± 0.112**	**- 3.048**	**0.002**
JMB → ASR	**- 0.205 ± 0.102**	**- 2.002**	**0.045**
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 1b			
(AMB → ASR)[1]	0	-	-
JMB → ASR	**- 0.258 ± 0.105**	**- 2.443**	**0.015**
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 1c			
AMB → ASR	**- 0.378 ± 0.112**	**- 3.370**	**0.001**
(JMB → ASR)[1]	0	-	-
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 2a			
SSD → AMB	0.117 ± 0.070	1.680	0.093
SSD → JMB	0.089 ± 0.077	1.157	0.247
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**
Model 2b			
SSD → JMB	0.089 ± 0.077	1.157	0.247
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**
Model 2c			
SSD → AMB	0.117 ± 0.070	1.680	0.093
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**

[1] Path coefficient set to zero

The two methods also produced highly consistent results for model fit as evaluated by global fit indices (i.e. C and χ^2 statistics, respectively, see Table S10). The only difference was that for Model 1b *piecewiseSEM* indicated 'marginally acceptable' model fit whereas *lavaan* indicated poor model fit for this path model. The other fit indices (TLI, CFI, RMSEA, and SRMR) suggest conclusions that are fully consistent with C statistics and χ^2 tests, i.e. acceptable fit for Models 1a and 1c by all of these indices and unacceptable fit for all other models (Table S10).

Table S10. Fit indices for the six path models, obtained by *piecewiseSEM* and *lavaan*. Values indicating acceptable fit are highlighted in bold.

Model	piecewiseSEM				lavaan					
	C	df	P_c		χ^2	df	P_{χ^2}	TLI	CFI	RMSEA	SRMR
1a	**0.29**	**4**	**0.972**		**0.02**	**2**	**0.991**	**1.119**	**1.000**	**0.000**	**0.004**
1b	**11.6**	**6**	**0.065**		8.9	3	0.031	0.764	0.858	0.143	0.101
1c	**6.2**	**6**	**0.376**		**3.9**	**3**	**0.267**	**0.962**	**0.977**	**0.057**	**0.065**
2a	34.7	4	0.000		30.6	2	0.000	-0.978	0.341	0.386	0.154
2b	34.7	4	0.000		30.6	2	0.000	-0.719	0.313	0.386	0.154
2c	34.7	4	0.000		30.6	2	0.000	-0.661	0.336	0.386	0.154

2.2. AIC-based model comparisons using real and simulated data

To assess which of these models provides the best account of the data, first we calculated the AIC value for each model (in *piecewiseSEM* this is corrected for small sample size, i.e. AIC_c) using the real dataset. Second, we used simulated data to test which of the two methods produces less biased conclusions. For this latter purpose, we generated simulated datasets using the R function 'rnorm'. The simulated datasets have the same number of variables and sample size as the phylogenetically transformed real dataset. We fitted path models with both *piecewiseSEM* and *lavaan* to obtain the AIC (or AIC_c) values. Then we compared Model 1a (the model that got the highest support for model fit by the global fit indices, see Table S10) to the other five models (Models 1b,1c, 2a, 2b, and 2c), thus conducted five pairwise comparisons, repeated with the two methods. These paired comparisons between models mimic the comparison we conducted with the real dataset in our study (Table 2 in the main

text). We calculated ΔAIC for each comparison as the difference between AIC values of the two models (i.e. AIC of compared model - AIC of Model 1a, thus a positive ΔAIC value indicates better fit for Model 1a). We repeated this procedure with 1000 simulated datasets that resulted in 1000 ΔAIC values for each pairwise comparison. To assess whether the comparison of two particular models produces biased results with simulated data we calculated (1) the mean ΔAIC value of the 1000 runs ($\Delta AIC_{simulation}$), and (2) the probability that the simulated ΔAIC was larger than the ΔAIC value we got with the real dataset ($P_{\geq \Delta AIC\_sim}$).

Using real data, *piecewiseSEM* gave the lowest AIC_c for Model 1a (Table S11), a result consistent with global model fit evaluation (see Table S10). ΔAICc values suggested strong support for this model in all comparisons (ΔAICc ≥ 4.1, Table S11). In contrast, *lavaan* results were inconsistent with global model fit evaluation because it gave very strong support for Model 2c (Table S11), a model that had an unacceptable fit by all fit indices (see Table S10).

Table S11. AIC-based model comparison using real and simulated data by the two methods. AIC_c (*piecsewiseSEM*) and AIC (*lavaan*) values provided for all models are based on analyses of our real data. ΔAIC_{data} and $\Delta AIC_{simulation}$ show differences from Model 1a in pairwise comparisons, based on analyses of real or simulated data, respectively. $P_{\geq \Delta AIC\_sim}$ indicates the probability that analyses of random data result in as large or larger AIC differences in support for Model 1a than the ΔAIC values obtained with real data.

Model	**piecewiseSEM**					**lavaan**			
	AIC_c	**ΔAIC_{data}**	**$\Delta AIC_{simulation}$**	**$P_{\geq \Delta AIC\_sim}$**		**AIC**	**ΔAIC_{data}**	**$\Delta AIC_{simulation}$**	**$P_{\geq \Delta AIC\_sim}$**
1a	**15.8**	**0.0**	-	-		382.4	0.0	-	-
1b	25.7	9.9	-0.2	0.003		389.3	6.9	-1.0	0.004
1c	19.9	4.1	-0.9	0.042		384.3	1.9	-1.1	0.041
2a	59.0	43.2	7.4	0.0		521.3	138.9	274.3	1.0
2b	50.4	34.6	-0.1	0.0		360.0	-22.4	273.4	1.0
2c	50.4	34.6	0.4	0.0		**341.7**	**-40.7**	273.4	1.0

Using simulated data, we found that *piecewiseSEM* produced less biased results than *lavaan*. First, in most cases mean simulated ΔAIC values were small and there was no strong bias in favour of one specific model (see $\Delta AIC_{simulation}$ in Table S11), as one would expect with

random data. The only exception was the comparison between Model 1a and Model 2a in which simulated ΔAIC produced by *piecewiseSEM* was 7.4, favouring Model 1a. Importantly, however, these simulations indicated only a low probability for random data resulting in as large or larger AIC differences (43.2) in support for Model 1a than the ΔAIC values we obtained with real data (see low $P_{\geq \Delta AIC\_sim}$ values in Table S11), suggesting that support for Model 1a was unlikely the result of biased AIC estimates.

In contrast, simulations showed that *lavaan* produced highly biased ΔAIC values in all non-nested comparisons (see the high $\Delta AIC_{simulation}$ and $P_{\geq \Delta AIC\_sim}$ values for Models 2a, 2b and 2c in Table S9). On the other hand, for nested model comparisons (i.e. with Models 1b and 1c) *lavaan* produced unbiased results similarly to those we got with *piecsewiseSEM* (Table S11).

These analyses suggest that the two methods gave consistent results for (1) path coefficients estimates and for (2) evaluating model fit of individual path models by global fit indices (using *C* statistics in *piecewiseSEM*, and χ^2, TLI, CFI, RMSEA, and SRMR in *lavaan*). On the other hand, simulation results indicate that AIC-based model comparisons are less biased when performed by the piecewise structural equation modelling method, at least for comparisons between non-nested models.

References

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press.

Kline, R. B. 2015. Principles and practice of structural equation modeling. Guilford.

Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol. Evol. 7:573–579.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

Preacher, K. J., and E. C. Merkle. 2012. The problem of model selection uncertainty in structural equation modeling. Psychol. Methods 17:1–14.

1218 Rosseel, Y. 2012. Lavaan: An R package for structural equation modelling. J. Stat. Softw.
1219 48:1–36.
1220 Santos, J. C. 2012. Fast molecular evolution associated with high active metabolic rates in
1221 poison frogs. Mol. Biol. Evol. 29:2001–2018.
1222 Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal
1223 components can be misleading. Syst. Biol. 64:677–689.
1224 von Hardenberg, A., and A. Gonzalez-Voyer. 2013. Disentangling evolutionary cause-effect
1225 relationships with phylogenetic confirmatory path analysis. Evolution 67:378–387.
1226 West, S. G., A. B. Taylor, and W. Wu. 2012. Model fit and model selection in structural
1227 equation modeling. Pp. 209–231 *in* R. Hoyle, ed. Handbook of structural equation
1228 modeling. Guilford.
1229

Evolution of large males is associated with female-skewed adult sex ratios in amniotes

Abstract

Body size often differs between the sexes (leading to sexual size dimorphism, SSD), as a consequence of differential responses by males and females to selection pressures. Adult sex ratio (the proportion of males in the adult population, ASR) should influence SSD because ASR relates to both the number of competitors and available mates, which shape the intensity of mating competition and thereby promotes SSD evolution. However, whether ASR correlates with SSD variation among species has not been yet tested across a broad range of taxa. Using phylogenetic comparative analyses of 462 amniotes (i.e. reptiles, birds and mammals), we fill this knowledge gap by showing that male bias in SSD increases with increasingly female-biased ASRs in both mammals and birds. This relationship is not explained by the higher mortality of the larger sex because SSD is not associated with sex differences in either juvenile or adult mortality. Phylogenetic path analysis indicates that higher mortality in one sex leads to skewed ASR, which in turn may generate selection for SSD biased towards the rare sex. Taken together, our findings provide evidence that skewed ASRs in amniote populations can result in the rarer sex evolving large size to capitalise on enhanced mating opportunities.

Keywords: sexual selection, mating competition, mating opportunity, sex-biased mortality, comparative method

INTRODUCTION

Sexual size dimorphism (SSD, measured as the size of males relative to females) is widespread in nature and is one of the most conspicuous phenotypic difference between the sexes (Darwin 1871; Andersson 1994; Fairbairn et al. 2007). It is the consequence of different optimal body size for the sexes resulting from opposing selection forces (some of which may influence only one of the sexes) that equilibrate differently in males and females (Blanckenhorn 2005).

A large volume of research has focused on how sex-specific behaviour (e.g. mating system, parental care), ecological processes (e.g. abundance and quality of resources), and life history (e.g. fecundity in indeterminate growers) can generate size differences between the sexes (Andersson 1994; Blanckenhorn 2005). These studies have concluded that sexual selection is often a major driver of SSD evolution by either intra-sexual competition for access to mates or inter-sexual mate choice, although other evolutionary mechanisms (e.g. fertility selection and competition for resources) may also be important (Jehl and Murray 1986; Andersson 1994; Blanckenhorn 2005; Fairbairn et al. 2007; Clutton-Brock 2016). Strong sexual selection for large body size in one sex is particularly likely in species where that sex competes for mates by physical contests or endurance rivalry, as observed in several vertebrate taxa (e.g. reptiles, birds, and mammals; Jehl and Murray 1986; Andersson 1994; Cox *et al.* 2007; Székely *et al.* 2007; Clutton-Brock 2016).

Adult sex ratio (ASR), best measured as the proportion of males in the adult population (Ancona et al. 2017) is a key demographic property of populations that influences both the number of competitors for mates and the number of mates available to an individual (Murray 1984; Székely et al. 2014b; Jennions and Fromhage 2017; Schacht et al. 2017). For example, a male-skewed ASR means potentially more competitors and fewer available partners for males than for females. An increasing number of studies show that ASR covaries

with several reproductive traits such as mating system, parental sex roles, divorce rate, extra-pair mating and cooperative breeding both in non-human animals and humans (Liker et al. 2013, 2014; Schacht et al. 2014; Kappeler 2017; Komdeur et al. 2017; Eberhart-Phillips et al. 2018; Grant and Grant 2019). However, whether and how ASR is related to the evolution of SSD is still poorly understood.

Theories suggest that ASR can drive the evolution of SSD in at least two ways. First, the intensity of sexual competition may increase with the number of competitors. As Darwin wrote (1871, p. 217): *"That some relation exists between polygamy and development of secondary sexual characters, appears nearly certain; and this supports the view that a numerical preponderance of males would be eminently favourable to the action of sexual selection"*. According to his idea, highly skewed ASRs may intensify selection for competitive traits such as weapons and large body size in the more numerous sex. Thus this 'mating competition hypothesis' predicts that the extent of male-bias in SSD should increase with the degree of male skew in the ASR. Later work refined Darwin's (1871) original idea by suggesting that the operational sex ratio (OSR, the number of sexually active males per receptive female at a given time) rather than the ASR determines the intensity of mating competition in a population (Emlen and Oring 1977). Thus, according to this latter theory ASR would predict SSD if ASR covaries with OSR, for example because OSR is in part determined by ASR (together with sex differences in behaviour like parental care; Kokko et al. 2012). Although the relationship between ASR and OSR is yet to be fully explored, their positive association has been demonstrated both by theoretical models (Kokko and Jennions 2008: Fig. 4a; Fromhage and Jennions 2016: Fig. 3c,d) and comparative analyses (Mitani et al. 1996, correlation between ASR and OSR in 18 primates: $r = 0.4$, $P = 0.002$; unpublished result using data from their Table 1). Empirical studies commonly use ASR and OSR

74 interchangeably in testing their relationship with SSD (Poulin 1997) and other proxies of
75 sexual selection (Janicke and Morrow 2018).
76 Second, models of reproductive sex roles predict that ASR should influence the
77 evolution of SSD because individuals of a given sex may allocate less to parental care when
78 the sex ratio is skewed towards the opposite sex than when it is skewed towards their own sex
79 (Queller 1997; McNamara et al. 2000). According to these models, males in female-skewed
80 populations display a higher reproductive success due to increased probability of breeding
81 with multiple partners and therefore may evolve to reduce parental care (Queller 1997:
82 section 3., McNamara et al. 2000: section 'Sex ratio'). This association between ASR and
83 parental sex roles can drive the evolution of SSD because more elaborate trait expression in
84 males is evolutionarily linked to female-biased care and stronger sexual selection on males
85 (the so called 'sex-role syndrome', Janicke et al. 2016: Fig 3.). Thus, this 'mating
86 opportunity hypothesis' predicts that the extent of male-bias in mating competition, and hence
87 in SSD, should decrease with increasing male skew in the ASR. A demographic analysis of
88 mating systems by Murray (1984) also predicts that female-skewed ASRs should be
89 associated with both polygyny and male-biased SSD, whereas male-skewed ASRs should be
90 associated with polyandry and female-biased SSD.
91 Alternatively, SSD may drive changes in sex ratios through sex differences in
92 mortality resulting from sexual competition. According to this 'mortality cost hypothesis', the
93 skewed ASR is a consequence rather than a cause of intense sexual selection, because when
94 males allocate a lot to mating competition they may suffer increased mortality, which in turn
95 leads to female-skewed ASR (Trivers 1972; Clutton-Brock et al. 1985; Liker and Székely
96 2005; Kalmbach and Benito 2007). This hypothesis predicts that in species exhibiting SSD
97 (1) the larger sex should have higher mortality due to the costs of being large, including the

direct costs associated with competition (e.g. fights, displays); which leads to (2) decreasing male skew in the ASR with increasing degree of male bias in the SSD.

Studies that have investigated the relationships between sex ratios, SSD and sex-specific mortality have so far yielded inconsistent results. While some studies found a positive link between SSD and ASR or OSR (i.e. an increasing male bias in SSD with increasing male skew in the sex ratios; Mitani *et al.* 1996; Poulin 1997), others reported negative associations (Clutton-Brock et al. 1977; Wittenberger 1978; Georgiadis 1985; Haro et al. 1994; Johansson et al. 2005; Lovich et al. 2014), or found no consistent relationships (Owen-Smith 1993; Hirst and Kiørboe 2014; Muralidhar and Johnson 2017). Similarly, mortality costs paid by the larger sex in dimorphic species were reported in some studies (Clutton-Brock et al. 1985; Promislow 1992; Promislow et al. 1992; Moore and Wilson 2002; Benito and González-Solís 2007; Kalmbach and Benito 2007), whereas no consistent relationship between SSD and sex differences in mortality was found by others (Owens and Bennett 1994; Toïgo and Gaillard 2003; Lemaître and Gaillard 2013; Székely *et al.* 2014a; Tidière *et al.* 2015). Many of these studies focused on a narrow range of taxonomic groups and were based on a relatively small number of species (typically fewer than 50) in comparative analyses. Furthermore, none of the studies tested explicitly whether statistical models assuming that ASR drives variation in SSD (as proposed by the mating competition and mating opportunity hypotheses) or alternative models (like the mortality costs hypothesis) fit better to the data.

Here we investigate the strength and direction of the relationship between ASR and SSD in populations of wild amniotes, using the largest existing comparative dataset on ASR compiled to date (462 species). First, we investigate whether SSD increases or decreases with ASR across species, as predicted by the mating competition and mating opportunity hypotheses, respectively. We also test whether the relationship is consistent among three

major amniote taxa (reptiles, birds, and mammals) because these taxa differ in multiple ecological, behavioural and life-history traits. Since the extent and direction of SSD can be influenced by ecological, life-history and behavioural factors besides mating competition, we also control for several potential confounding variables in the analyses. Second, we study whether SSD drives ASR variation by generating sex-biased mortality as proposed by the mortality cost hypothesis. We test this latter hypothesis by investigating whether SSD is related to sex differences in juvenile or adult mortality, and by comparing path models representing different structural relationships between SSD, ASR and sex-specific mortality.

METHODS

Data collection

Data were extracted from published sources (see Appendix S1 in Supporting Information). The initial dataset was based on Pipoly *et al.* (2015) that contains ASR and SSD for 344 amniote species. We excluded amphibians included in Pipoly *et al.* (2015) because sex-specific mortality data (see below) are very scarce for this taxon, especially in juveniles. The initial dataset was augmented with additional reptile and mammal species, and with information on sex-specific mortality. These additional data were taken from existing comparative datasets (Berger and Gompper 1999 and Bókony *et al.* 2019 for ASR in mammals and reptiles, respectively, and Székely *et al.* 2014a for mortality in birds) or from primary publications. In the latter case we searched the literature through the search engines Web of Science and Google Scholar, using the search terms ‘sex ratio’, ‘sex-specific mortality OR survival’ or ‘male female mortality OR survival’ together with taxonomic names. Data for different variables for the same species were often available only from different populations or studies. The final dataset includes 462 species with both ASR and SSD available (155 reptiles, 185 birds, 122 mammals).

148

149 **Body mass and SSD**

150 Sex-specific body mass (g) was available for all birds and mammals in our dataset. Since
151 body mass data were missing for many reptiles, we also collected body length data (mm) for
152 this taxon in the form of snout-vent length for squamates and crocodilians and plastron or
153 carapace length for turtles. We estimated body mass from body length using published
154 allometric equations (Appendix S2). We used estimated body mass for reptiles instead of
155 body length in the combined analyses of all species because (1) data on mass are more readily
156 available than data on body length in birds and mammals, which provided the majority of
157 species, and (2) body mass is measured in a standardized way in all taxa, whereas the
158 measurement of body length varies because different parts of the body are recorded as a proxy
159 for length in different taxa. If multiple mass or length data were available for a species, we
160 used the mean value. Average adult body mass was calculated as $\log_{10}$-transformed mean
161 mass of the sexes.

162 SSD was calculated as $\log_{10}$(male mass / female mass), which has been recommended
163 as a statistically appropriate and preferable measure of dimorphism (Smith 1999; Fairbairn
164 2007). To test whether the results are sensitive to conversion of length to mass in reptiles, we
165 repeated the main analyses (1) with SSD calculated from body length ($\log_{10}$(male length /
166 female length)) of reptiles, and (2) with SSD calculated from body mass for a subset (31
167 species) of reptiles that has sex-specific mass data available from Myhrvold et al. (2015).
168 Whatever approach was used to assess the degree of sexual size dimorphism the results were
169 qualitatively unchanged (see Results). In the main text we thus report results based on body
170 mass estimated from body length for reptiles.

171

172 **Sex ratio**

We followed Wilson and Hardy (2002) and Ancona et al. (2017) in expressing ASR as the proportion of males in the adult population. We defined the adult population here broadly as adult individuals living in the study area during ASR sampling. Wilson and Hardy (2002) showed that analysing sex ratios as a proportion variable is appropriate when sex ratios are estimated from samples of ≥ 10 individuals and the dataset has ≥ 50 sex ratio estimates. These conditions are fully met in our analyses because sample sizes for ASR estimates were always larger than 10 individuals per species (and typically much larger), and more than 50 species were included in most analyses.

ASR data from Pipoly *et al.* (2015) were augmented with new species and updated with more recent and/or better quality information (e.g. based on a more reliable method or a larger sample size) for some reptiles. ASR estimates were collected by different observers for the different taxa: reptiles by V.B. and I.P. (Pipoly et al. 2015; Bókony et al. 2019), birds by A.L. (Liker *et al.* 2014), and mammals by Berger and Gompper (1999), Donald (2007) and Anile and Devillard (2018). Details of data selection criteria are given in the original publications (see also Ancona *et al.* 2017). Mean values were calculated for species with multiple ASR data. ASR estimates are repeatable between populations of the same species as measured by the intraclass correlation coefficient (ICC), although the magnitude of repeatability varies among taxa: reptiles with genetic and environmental sex determination: ICC= 0.55 and 0.14, respectively (Bókony et al. 2019), birds: ICC= 0.64 (Ancona et al. 2017), mammals: ICC= 0.60 (Valentine Federico, J-F.L., J-M.G., A.L., I.P., T.S. unpublished results). ASR estimates are not influenced by the sample size of the ASR studies (Székely et al. 2014a; Bókony et al. 2019).

Sex-specific mortality

Annual mortality rates were collected from studies in which mortality (or survival) was estimated for each of both sexes. Juvenile and adult mortality refer to age classes before and after the age of first reproduction, respectively. For reptiles, data were collected by V.B. (Bókony et al. 2019). Most adult mortality data on birds are taken from Székely *et al.* (2014a) with the addition of new data for juvenile mortality by A.L. Reptile and bird mortality includes estimates by various methods (capture-recapture and demographic analyses, return rates), although we used better quality estimates (e.g. those from capture-recapture analyses) whenever we had a choice (Székely *et al.* 2014a; Bókony *et al.* 2019). For mammals, all sex-specific estimates were collected by J-M.G. and J-F.L. (Lemaître et al. 2020). Sex differences in juvenile and adult mortality rates were calculated as the magnitude of male-biased mortality (i.e. $\log_{10}$(juvenile or adult male mortality / juvenile or adult female mortality)), also referred to as 'mortality bias'. These measures of mortality bias are not related to the overall mortality rate of the species, as estimated by the average mortality rates of the sexes (phylogenetic generalised least squares models, juvenile mortality bias: slope ± *SE* = - 0.068 ± 0.101, $t = 0.7$, $P = 0.497$, $n = 100$; adult mortality bias: slope ± *SE* = - 0.05 ± 0.08, $t = 0.7$, $P = 0.513$, $n = 230$).

Other predictors

We controlled for the potential effects of ecological and life-history variables related to either ASR or SSD (or both) that may confound the assessment of their relationship. First, we collected data on the type of sex determination system because it is associated with both ASR (Pipoly et al. 2015) and SSD (Adkins-Regan and Reeve 2014). We divided the species into three categories according to the Tree of Sex database (Ashman et al. 2014): male-heterogametic (XY) or female-heterogametic (ZW) genetic sex determination, or temperature-dependent sex determination (TSD). For species that were not included in the Tree of Sex

database we assumed the same type of sex determination as reported for the genus (or family, respectively; Bókony *et al.* 2019) when the genus (or family) to which it belongs had invariable sex determination system. All birds were assigned to ZW, and all mammals to XY sex-determination (Ashman et al. 2014).

Second, we controlled for the potential effects of environmental variation among species by using two proxies. Breeding latitude correlates with life-history traits in many organisms and may also influence the potential for polygamy, hence also sexual selection (Fischer 1960; Isaac 2005; Balasubramaniam and Rotenberry 2016). We used absolute values of the geographic latitude of the ASR studies included in our dataset (i.e. average values for species with multiple ASR estimates) to represent the distance from the Equator. When the authors did not report latitude, we used Google Earth to estimate it as the center of the study sites based on the site descriptions. For 30 birds and 10 mammals, accurate population locations were not reported, hence, we used the latitudinal midpoint of the breeding ranges of these species (birds: V. Remeš, A. Liker, R. Freckleton and T. Székely unpublished data, mammals: PanTHERIA database).

In addition to latitude, we investigated environmental harshness as a second environmental variable, which also has been hypothesized to influence SSD (Isaac 2005). We quantified the harshness of the breeding environment using a proxy proposed by Botero *et al.* (2014). This is the PC1 score extracted from Principal Component Analysis (PCA) performed on a set of climatic and ecological variables (e.g. temperature and precipitation, net primary productivity, habitat heterogeneity; see Botero *et al.* 2014 for a detailed description of the variables and the analysis). The PC1 scores have higher values for a higher level of exposure to drier, less productive environments, with colder, less predictable and more variable annual temperatures (see Table 1 in Botero *et al.* 2014). In birds and mammals, we used the data

246 published in Botero *et al.* (2014), whereas for reptiles we calculated PC1 scores by
247 performing a PCA with the same set of variables.
248 Third, we characterised courtship displays in birds because earlier studies showed that
249 birds with aerial displays have less male-biased SSD compared to species with ground
250 displays, probably because selection favours male agility in aerially displaying species
251 constraining male body size (Jehl and Murray 1986; Székely et al. 2007). We followed
252 Székely *et al.* (2007) and divided species into two display groups: (1) mating displays that
253 may favour male agility, including species that mainly have aerial displays (both non-
254 acrobatic and acrobatic, categories 4 and 5 in Székely *et al.* 2007), and (2) displays that may
255 not favour male agility, including all other display types, typically performed on ground
256 (categories 1-3 in Székely *et al.* 2007). Although SSD can also be influenced by display type
257 and display habitat in reptiles and mammals (e.g. see Agha *et al.* 2018), we were not able to
258 collect reliable data for these taxa, therefore we analysed the effect of display type only in
259 birds.
260 Fourth, we tested for the potential effect of social mating system, because the scope
261 for mating competition may be more limited in monogamous than in polygamous species
262 (Andersson 1994). Thus, although there is ASR variation among monogamous species that
263 can generate some variation in mating competition and/or opportunity, the relationship
264 between ASR and SSD is expected to be weaker in monogamous than in polygamous species.
265 To test this idea, we characterised social mating system for birds and mammals, because we
266 found reliable information in these taxa for most species (Liker *et al.* 2014; Lukas and
267 Clutton-Brock 2013). We categorized species as either socially monogamous or polygamous
268 (most often polygynous) according to the sources. In birds, mating system was originally
269 scored on a five point scale (Liker et al. 2014), and here we considered a species
270 monogamous if it had score 0 or 1 (polygamy frequency <1%) for both sexes.

Finally, in reptiles, the evolution of viviparity and reduced reproductive frequency are generally correlated with shifts toward female-biased SSD due to fecundity selection for large female size (Pincheira-Donoso and Hunt 2017). To control for its potential effect on SSD, we categorized the reproductive mode of reptiles as either viviparous or oviparous (Uetz et al. 2019).

Statistical analyses

Phylogenetic generalised least squares (PGLS) models were built to conduct bivariate and multi-predictor analyses. To control for phylogenetic relationships among taxa, we used the composite phylogeny applied in Pipoly *et al.* (2015) with the addition of new species according to a family-level (Sarre et al. 2011) and other recent phylogenies (Squamata: Nicholson *et al.* 2012, Pyron *et al.* 2013, Gamble *et al.* 2014; Testudines: Barley *et al.* 2010, Guillon *et al.* 2012, Spinks *et al.* 2014; Crocodylia: Oaks 2011; mammals: Fritz *et al.* 2009, Meredith *et al.* 2011). Since composite phylogenies do not have true branch lengths, we used three methods to generate branch lengths (Nee's method, Pagel's method, and unit branch lengths, using the PDAP:PDTREE module of Mesquite; Midford *et al.* 2011), and repeated key analyses with these alternative trees. We present results with Nee's branch lengths in the paper, except for the sensitivity analyses (see Results). Freckleton et al. (2002) showed that PGLS is relatively insensitive to branch length assumptions. In each model we used the maximum-likelihood estimate of phylogenetic dependence (Pagel's λ). PGLS models were run using the 'caper' R package (Orme et al. 2013).

First, using all species, we applied bivariate PGLS models to test interspecific associations between ASR, SSD and sex differences in juvenile and adult mortality rates. Then we built two multi-predictor models. In Multi-predictor model 1, we tested the relationship between ASR and SSD while controlling for potential confounding effects of

mean mass, sex determination system, and breeding latitude. In Multi-predictor model 2, we tested the ASR - SSD relationships while controlling for the effects of sex differences in juvenile and adult mortality rates, and mean mass. We built these two separate multi-predictor models because we have much lower sample sizes for sex-specific mortalities than for the other predictors, thus the statistical power would be reduced for variables of Multi-predictor model 1 if all predictors were combined in a single model. We ran the models in two alternative versions in which either SSD or ASR was the dependent variable, respectively, since we had no *a priori* knowledge about the cause-effect direction of these relationships and results may differ between these analyses if the two models have different values for Pagel's λ (see Appendix S3).

We investigated whether the ASR – SSD relationship, which is the main focus of our study, differed among taxa by testing the interaction between ASR and the taxonomic class. To explore differences among taxa in the multivariate relationships, we repeated all analyses separately for reptiles, birds and mammals. In taxon-specific Multi-predictor models 1, we included reproductive mode for reptiles and display type for birds as further predictors. In reptiles, we also tested whether the relationship between ASR and SSD is sensitive (1) to the inclusion of species that have environmental sex determination, because ASR shows low repeatability in such reptiles (Bókony et al. 2019), and (2) to the inclusion of species in which the type of sex determination was inferred from data on related species in the genus or family. Finally, we ran two additional separate analyses to test whether social mating system and environmental harshness confounded the ASR - SSD relationship. We did not include these latter two predictors in the multi-predictor models because these data were available only for subsets of species that would have resulted in considerably reduced sample sizes. All numeric variables were standardized before analyses to make parameter estimates comparable, and model assumptions were also checked and met. We report two-tailed statistics. Sample sizes

differed between models because not all variables were available for all species (see Appendix S1).

In addition to PGLS models, we used phylogenetic path analyses (Santos 2012; Gonzalez-Voyer and von Hardenberg 2014) to compare two sets of path models corresponding to different hypotheses for the relationships linking ASR, SSD and sex differences in mortality. Although path analyses – unlike experiments – cannot infer causality, it is a suitable method to compare alternative scenarios representing different causal relationships between variables (Shipley 2016). Model 1 assumes that sex-biased mortality influences ASR, which in turn influences SSD through its effects on mating competition (as proposed by the mating opportunity hypothesis; Fig. 1). Three variants of this model were tested: Model 1a assumes that sex differences in both juvenile and adult mortality rates influence ASR, while Models 1b-c include only one of these mortality effects. Model 2 assumes that SSD has sex-specific effects on juvenile and/or adult mortality, which then drives ASR variation (representing the mortality cost hypothesis; Fig. 1). We tested all the three variants of this latter scenario, assuming SSD effects on both juvenile and adult mortality (Model 2a) or only on one mortality component (Models 2b-c).

We followed the approach proposed by Santos (2012) for phylogenetic path analyses. In the first step, we conducted phylogenetic transformation on the data to control for effects of phylogenetic relatedness among species. For this purpose, we (1) determined λ separately for each variable by maximum likelihood, (2) used this variable-specific λ value to re-scale the phylogenetic tree to a unit tree, and (3) used the transformed tree to calculate phylogenetically independent contrasts for the variable (using 'pic' function of the R package 'ape'; Paradis 2012). We repeated this process for each variable, and the resulting phylogenetically transformed values were used for fitting path models. In the second step of the analyses, we evaluated model fit using d-separation method (Shipley 2016) as implemented in the R

package 'piecewiseSEM' (Lefcheck 2016). In this method, Fisher's *C* statistic is used to test the goodness of fit of the whole path model, and the model is rejected (i.e. it does not provide a good fit to the data) if the result of this *C* statistic is statistically significant (and conversely a statistically non-significant result means acceptable fit; Lefcheck 2016). We compared model fit between the six path models by their AICc values. Note that this approach ensures that the same variables (i.e. the contrasts) are used in each path model and that correlations are non-directional (i.e. for a pair of variables *X* and *Y*, $r_{XY} = r_{YX}$ as assumed in path analysis).

To test the robustness of the results, we repeated the path analyses using two other methods. First, we repeated the above procedure (i.e. followed Santos 2012) except that we used the covariance matrix comparison method for model fit instead of d-separation, as implemented in the R package 'lavaan' (Rosseel 2012). Second, we repeated the analyses using the method developed by von Hardenberg and Gonzalez-Voyer (2013). Unlike Santos' (2012) method, in this latter approach a single value of Pagel's λ is estimated for each pair of traits in a directional statistical model, rather than a value of λ for each variable (see the Discussion and Appendix S3 where we explain why this approach may be problematic). We used the R package 'phylopath' (van der Bijl 2018) for this latter analysis, which relies on the d-separation method for model fitting (similarly to 'piecewiseSEM', see above). We provide further justifications for our approach and additional analyses to test the robustness of the path analysis' results in Appendix S3.

RESULTS

Mating competition versus mating opportunity hypotheses

Consistent with the mating opportunity hypothesis, and in contrast to the mating competition hypothesis, we found a negative relationship between our measures of ASR and SSD: the size of males relative to females increases when ASR becomes more female-skewed (Fig. 2, Table

1). This correlation was statistically significant when all species were analysed together and did not differ among the three amniote classes (ASR × class interaction on SSD: $F_{2,456}$= 2.177, P= 0.115). The increase of SSD with increasingly female-skewed ASR was statistically significant within birds and mammals but was not in reptiles when the three taxa were analysed separately (Fig. S1, Tables S1-4). These results remained consistent when we used SSD estimates based on length instead of estimated mass in reptiles (Tables S1, S2 and S5), and also when SSD for reptiles were estimated from published body mass data (Table S5).

These results are robust because the direction of the ASR - SSD relationship and its statistical significance were not sensitive to branch length assumptions (Table S6), and to the inclusion of other predictors (Table 1). In multi-predictor models (Table 1), mean body mass was positively related to SSD, supporting the Rensch rule (Abouheif and Fairbairn 1997), and the type of sex determination influenced ASR variation as previously reported by Pipoly *et al.* (2015). Nevertheless, ASR remained negatively associated with SSD when the effects of mass and sex determination systems were accounted for (Table 1). This result also did not change when environmental variation was included in the models using either breeding latitude (Table 1) or environmental harshness (Table S5). Finally, excluding reptiles with TSD (that have the lowest consistency in ASR; Bókony *et al.* 2019) or with assumed sex determination also did not influence the relationship (Table S5).

The multi-predictor model for birds showed that species with aerial courtship displays have lowered SSD as found in earlier studies (Jehl and Murray 1986; Székely et al. 2007); however, the relationship between ASR and SSD remained statistically significant and negative when this effect was included in the model (Table S3). Furthermore, data in birds and mammals showed that, as expected, the relationship was weaker in monogamous than in polygamous species, although the same trend occurred in both mating systems (Table S7).

Finally, reproductive mode was not associated with SSD or ASR in reptiles in our dataset (Tables S1-2).

Mating opportunity versus mortality costs hypotheses

Both the mating opportunity hypothesis and the mortality cost hypothesis predict female-skewed ASRs in species with male-biased SSD. However, our results are more consistent with the mating opportunity hypothesis for two reasons. First, ASR but not SSD was associated with the extent of sex differences in juvenile or adult mortality, and ASR remained strongly and negatively correlated with SSD when sex differences in juvenile and adult mortality were statistically controlled for (Table 1). Second, phylogenetic path analyses showed that models of the mating opportunity hypothesis provided better fit to the data (Models 1a-c, Fisher' *C* statistic: $P = 0.07$ - 0.97) than models corresponding to the mortality cost hypothesis (Models 2a-c, $P < 0.001$; Table 2). The strongest support was for Model 1a because it had the lowest AICc (ΔAICc = 4.1 - 43.2; Table 2). This model proposes that sex-biased mortality in both juveniles and adults generates skewed ASR, which in turn leads to SSD biased towards the rarer sex (Fig. 3). These results are robust because we obtained the same results when the analyses were repeated using two other implementations of the path analysis (see Table S8 for the results obtained using 'phylopath', and Appendix S3 for the results obtained using 'lavaan'). Finally, path analyses that excluded reptiles (for which the ASR - SSD relationship was not statistically significant, see above) also yielded results qualitatively consistent with the full dataset (Table S9).

DISCUSSION

Our analyses provided three major findings: (1) adult sex ratio is related to sexual size dimorphism among amniote species, although the association is the opposite of the one

proposed by Darwin; (2) sex-biased mortality is unrelated to the extent of SSD in amniotes; and (3) confirmatory path analyses indicate that sex-biased mortality influences ASR, which in turn induces changes in SSD. Collectively, these findings support the mating opportunity hypothesis, indicating that selection is likely to favour an increased resource allocation toward mating competition (by growing and maintaining a large body mass) in the rarer sex, which has a higher chance of getting mates than the other sex.

Theoretical models show that skewed ASRs can promote evolutionary changes that may generate this association between ASR and SSD. First, models of sex role evolution showed that skewed ASR can result in divergences in reproductive roles between the sexes leading to less parental care and more frequent desertion and remating in the rarer sex and opposite changes (i.e. more parental care and less frequent remating) in the more abundant sex (Queller 1997; McNamara et al. 2000). Similarly, a demographic analysis based on the relationships between mating systems and sex ratio, sex-specific patterns of survivorship, age of first reproduction, and annual fecundity predicts that skewed ASRs promote the evolution of polygamy (i.e. polygyny and polyandry in female-biased and male-biased populations, respectively; Murray 1984). Since both frequent remating and polygamy can intensify sexual selection, the above effects of skewed ASR can promote the evolution of SSD by favouring increased body size in the rare sex. In line with the predictions of these models, an increasing number of recent studies in birds and humans show that polygyny is more frequent and parental care by males is reduced in female-skewed populations (Liker et al. 2013, 2014, 2015; Remeš et al. 2015; Schacht and Borgerhoff Mulder 2015; Eberhart-Phillips et al. 2018; Grant and Grant 2019). Our results are also concordant with experimental studies in voles and lizards, which reported that female-skewed ASRs exert directional selection for large body size in males (Klemme et al. 2007; Fitze and Le Galliard 2008), and increase variance in male reproductive success (Dreiss et al. 2010).

Theoretical models predict that the effects of ASR may depend on other life-history and behavioural traits of the populations. For example, Fromhage and Jennions (2016) highlighted the importance of the specific processes generating ASR skews for the outcomes of sex role evolution, and that a coevolutionary feedback between parental care and sexually selected traits can greatly amplify sex role divergence. In addition, sexual competition for mates may favour different traits in species with distinct ecology and behaviour, leading to inconsistent relationships between sex differences in mating competition and sexual dimorphisms in behavioural or morphological trait across species (Clutton-Brock 2017). Collectively, these factors may account for the relatively low amount of variation in SSD explained by ASR in some of our analyses.

The association between intense sexual selection in males and female-skewed ASRs was proposed decades ago by avian evolutionary ecologists (e.g. Mayr 1939), although it was usually explained by the mortality cost hypothesis (Wittenberger 1976). Our analyses do not support this hypothesis because sex-biased SSD is not associated with sex-biased juvenile or adult mortality in the studied amniote species, and the results of the confirmatory path analyses are also inconsistent with the mortality cost hypothesis. We propose that the lack of relationship between SSD and sex differences in mortality may be explained by variation in the environmental context (Lemaître et al. 2020). Studies in birds and mammals showed that having a large body size may only be costly in terms of mortality in populations subjected to harsh environmental conditions (Toïgo and Gaillard 2003; Kalmbach and Benito 2007; Jones et al. 2009; Clutton-Brock 2017). The effect of SSD may thus be reduced or absent when the sex-specific mortality estimates correspond to average conditions, that may often be the case in wild populations.

The ASR - SSD relationship may also be influenced by sex differences in the time of maturation because longer maturation time in the larger sex can result in a shortage of that sex

in the adult population (Lovich et al. 2014) because immature life stages are generally characterized by higher mortality (e.g. Gaillard et al. 2000). Furthermore, Fromhage & Jennions (2016) showed that female-skewed sex ratios at maturation (MSR) can result in the evolution of increased female care and male allocation to traits facilitating mating success. Thus, if variation in ASR is determined at least in part by MSR, then the effects of sex-biased MSR on sex roles can contribute to the observed association of ASR with the intensity of mating competition, and, hence, SSD. This latter mechanism would deserve further investigations.

Although the relationship between ASR and SSD is not statistically significant in reptiles, it is qualitatively consistent with our findings in birds and mammals. Other selective processes (e.g. fertility selection for large female size in indeterminate growers, Cox *et al.* 2007) might have masked the influence of sexual selection on SSD in reptiles. Consistent with this explanation, selection often favours delayed maturation in female reptiles, which enables them to produce larger clutches, which in turn also influences their body size and the extent of SSD (Shine 2005; Agha et al. 2018). Follow-up studies using different proxies of sexual selection are needed to investigate further how sexual selection is related to ASR in reptiles.

Biased estimates of ASR may generate spurious relationship with SSD, which may potentially affect our results. For example, the larger sex may have lower detectability in polygamous species if some members of that sex are excluded from breeding sites (Ancona et al. 2017). However, highly polygamous species in which populations have been thoroughly surveyed showed skewed ASR even when all individuals in the population were accurately counted (Granjon et al. 2017), and fairly consistent ASR estimates were obtained when both breeding and non-breeding individuals were included (Emlen and Wrege 2004). In general, ASR estimates show a moderate but statistically significant repeatability across populations in

most of the studied taxa, except reptiles with temperature-dependent sex determination (Ancona et al. 2017; Bókony et al. 2019; Valentine Federico, J-F.L., J-M.G., A.L., I.P., T.S. unpublished result), and in 80% of bird species the direction of ASR skew is the same for all repeated estimates (Székely et al. 2014a).

The paths of causality in comparative data are difficult to untangle. Path analysis is a valuable tool for contrasting different causal models, although it cannot reveal causality (Shipley 2016). Path analysis assumes that each variable includes independent variations or 'errors' and that these errors are independent among variables. This is not true for comparative data, because the errors will be correlated across species. Our approach follows Santos (2012), an innovative but overlooked method that satisfies the assumptions of path analysis better than an alternative method based on phylogenetic regressions proposed by von Hardenberg and Gonzalez-Voyer (2013). As outlined in more detail in Appendix S3, this latter approach is problematic because it is not robust to changes in the specification of the model: if variable *Y* is regressed on *X* and λ estimated, then the estimates of the partial correlations and λ may be different from those obtained if *Y* is regressed on *X* with λ estimated. The approach we have taken avoids this problem. However, there is still room for methodological improvement. For instance, our approach has the drawback to be a 'subtractive' comparative method (sensu Harvey and Pagel 1991). Multivariate approaches such as path analyses for complex data have to be developed further.

Concluding remarks

Our findings indicate that sex-specific selection for large body size is associated with skewed ASRs across amniotes, and this process appears to produce SSD biased towards the rare sex in birds and mammals. Although this conclusion contrasts with Darwin's initial suggestion that intense sexual selection among males occurs when there is a surplus of males in the

521 population (Darwin 1871), theoretical and empirical work have suggests mechanisms that can
522 favour large size in the rare sex (Murray 1984; Klemme et al. 2007; Fitze and Le Galliard
523 2008; Dreiss et al. 2010). Further analyses of these processes and their application to species
524 with differing mating systems offer exciting opportunities for future investigations of the
525 interplay among sexual selection, SSD and ASR across the tree of life.
526

REFERENCES

Abouheif, E., and D. J. Fairbairn. 1997. A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. Am. Nat. 149:540–562.

Adkins-Regan, E., and H. K. Reeve. 2014. Sexual dimorphism in body size and the origin of sex-determination systems. Am. Nat. 183:519–536.

Agha, M., J. R. Ennen, A. J. Nowakowski, J. E. Lovich, S. C. Sweat, and B. D. Todd. 2018. Macroecological patterns of sexual size dimorphism in turtles of the world. J. Evol. Biol. 31:336–345.

Ancona, S., F. V. Dénes, O. Krüger, T. Székely, and S. R. Beissinger. 2017. Estimating adult sex ratios in nature. Philos. Trans. R. Soc. B Biol. Sci. 372:20160313.

Andersson, M. B. 1994. Sexual Selection. Princeton University Press, Princeton, New Jersey.

Anile, S., and S. Devillard. 2018. Camera-trapping provides insights into adult sex ratio variability in felids. Mamm. Rev. 48:168–179..

Ashman, T.-L., D. Bachtrog, H. Blackmon, E. E. Goldberg, M. W. Hahn, M. Kirkpatrick, J. Kitano, J. E. Mank, I. Mayrose, R. Ming, S. P. Otto, C. L. Peichel, M. W. Pennell, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, and J. C. Vamosi. 2014. Tree of Sex: A database of sexual systems. Sci. Data 1:140015.

Balasubramaniam, P., and J. T. Rotenberry. 2016. Elevation and latitude interact to drive life-history variation in precocial birds: a comparative analysis using galliformes. J. Anim. Ecol. 85:1528–1539.

Barley, A. J., P. Q. Spinks, R. C. Thomson, and H. B. Shaffer. 2010. Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. Mol. Phylogenet. Evol. 55:1189–1194.

Benito, M. M., and J. González-Solís. 2007. Sex ratio, sex-specific chick mortality and sexual size dimorphism in birds. J. Evol. Biol. 20:1522–1530.

Berger, J., and M. E. Gompper. 1999. Sex ratios in extant ungulates: products of contemporary predation or past life histories? J. Mammal. 80:1084–1113.

Blanckenhorn, W. U. 2005. Behavioral causes and consequences of sexual size dimorphism. Ethology 1016:977–1016.

Bókony, V., G. Milne, I. Pipoly, T. Székely, and A. Liker. 2019. Sex ratios and bimaturism differ between temperature-dependent and genetic sex-determination systems in reptiles. BMC Evol. Biol. 19:57.

Botero, C. A., R. Dor, C. M. McCain, and R. J. Safran. 2014. Environmental harshness is

positively correlated with intraspecific divergence in mammals and birds. Mol. Ecol. 23:259–268.

Clutton-Brock, T. 2017. Reproductive competition and sexual selection. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160310.

Clutton-Brock, T. H. 2016. Mammal Societies. Wiley-Blackwell.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness. 1985. Parental investment and sex differences in juvenile mortality in birds and mammals. Nature 313:131–133.

Clutton-Brock, T. H., P. H. Harvey, and B. Rudder. 1977. Sexual dimorphism, socionomic sex ratio and body weight in primates. Nature 269:797–800.

Cox, R. M., M. A. Butler, and H. B. John-Alder. 2007. The evolution of sexual size dimorphism in reptiles. Pp. 38–49 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles. Oxford University Press, Oxford.

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

Donald, P. F. 2007. Adult sex ratios in wild bird populations. Ibis 149:671–692.

Dreiss, A. N., J. Cote, M. Richard, P. Federici, and J. Clobert. 2010. Age-and sex-specific response to population density and sex ratio. Behav. Ecol. 21:356–364.

Eberhart-Phillips, L. J., C. Küpper, M. C. Carmona-Isunza, O. Vincze, S. Zefania, M. Cruz-López, A. Kosztolányi, T. E. X. Miller, Z. Barta, I. C. Cuthill, T. Burke, T. Székely, J. I. Hoffman, and O. Krüger. 2018. Demographic causes of adult sex ratio variation and their consequences for parental cooperation. Nat. Commun. 9:1651.

Emlen, S. T., and L. W. Oring. 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197:215–23.

Emlen, S. T., and P. H. Wrege. 2004. Size dimorphism, intrasexual competition, and sexual selection in Wattled jacana (Jacana jacana), a sex-role-reversed shorebird in Panama. Auk 121:391–403.

Fairbairn, D. J. 2007. Introduction: The enigma of sexual size dimorphism. Pp. 1–10 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism. Oxford University Press.

Fairbairn, D. J., W. U. Blanckenhorn, and T. Székely. 2007. Sex, Size and Gender Roles. Oxford University Press, Oxford.

Fischer, A. G. 1960. Latitudinal variations in organic diversity. Evolution 14:64–81. John

Fitze, P. S., and J. F. Le Galliard. 2008. Operational sex ratio, sexual conflict and the intensity

of sexual selection. Ecol. Lett. 11:432–439.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Fritz, S. A., O. R. P. Bininda-Emonds, and A. Purvis. 2009. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecol. Lett. 12:538–549.

Fromhage, L., and M. D. Jennions. 2016. Coevolution of parental investment and sexually selected traits drives sex-role divergence. Nat. Commun. 7:12517.

Gaillard, J.-M., M. Festa-Bianchet, N. G. Yoccoz, A. Loison, and C. Toïgo. 2000. Temporal Variation in Fitness Components and Population Dynamics of Large Herbivores. Annu. Rev. Ecol. Syst. 31:367–393.

Gamble, T., A. J. Geneva, R. E. Glor, and D. Zarkower. 2014. Anolis sex chromosomes are derived from a single ancestral pair. Evolution 68:1027–1041.

Georgiadis, N. 1985. Growth patterns, sexual dimorphism and reproduction in African ruminants. Afr. J. Ecol. 23:75–87.

Gonzalez-Voyer, A., and A. von Hardenberg. 2014. An introduction to phylogenetic path analysis. Pp. 201–229 *in* L. Z. Garamszegi, ed. Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology. Springer Berlin Heidelberg.

Granjon, A.-C., C. Rowney, L. Vigilant, and K. E. Langergraber. 2017. Evaluating genetic capture-recapture using a chimpanzee population of known size. J. Wildl. Manage. 81:279–288.

Grant, P. R., and B. R. Grant. 2019. Adult sex ratio influences mate choice in Darwin's finches. Proc. Natl. Acad. Sci. U. S. A. 116:12373–12382.

Guillon, J. M., L. Guéry, V. Hulin, and M. Girondot. 2012. A large phylogeny of turtles (Testudines) using molecular data. Contrib. to Zool. 81:147–158.

Haro, R. J., K. Edley, and M. J. Wiley. 1994. Body size and sex ratio in emergent stonefly nymphs (Isogenoides olivaceus: Perlodidae): variation between cohorts and populations. Can. J. Zool. 72:1371–1375.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press.

Hirst, A. G., and T. Kiørboe. 2014. Macroevolutionary patterns of sexual size dimorphism in copepods. Proc. R. Soc. B Biol. Sci. 281.

Isaac, J. L. 2005. Potential causes and life-history consequences of sexual size dimorphism in

626 mammals. Mamm. Rev. 35:101–115.
627 Janicke, T., I. K. Haderer, M. J. Lajeunesse, and N. Anthes. 2016. Darwinian sex roles
628 confirmed across the animal kingdom. Sci. Adv. 2:e1500983.
629 Janicke, T., and E. H. Morrow. 2018. Operational sex ratio predicts the opportunity and
630 direction of sexual selection across animals. Ecol. Lett. 21:384–391.
631 Jehl, J. R., and B. G. Murray. 1986. The evolution of normal and reverse sexual size
632 dimorphism in shorebirds and other birds. Pp. 1–86 *in* R. F. Johnston, ed. Current
633 Ornithology, vol. 3. Springer US, Boston, MA.
634 Jennions, M. D., and L. Fromhage. 2017. Not all sex ratios are equal: The Fisher condition,
635 parental care and sexual selection. Philos. Trans. R. Soc. B Biol. Sci. 372.
636 Johansson, F., P. H. Crowley, and T. Brodin. 2005. Sexual size dimorphism and sex ratios in
637 dragonflies (Odonata). Biol. J. Linn. Soc. 86:507–513.
638 Jones, K. S., S. Nakagawa, and B. C. Sheldon. 2009. Environmental sensitivity in relation to
639 size and sex in birds: meta-regression analysis. Am. Nat. 174:122–133.
640 Kalmbach, E., and M. M. Benito. 2007. Sexual size dimorphism and offspring vulnerability in
641 birds. Pp. 133–142 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex,
642 Size and Gender Roles. Oxford University Press.
643 Kappeler, P. M. 2017. Sex roles and adult sex ratios: insights from mammalian biology and
644 consequences for primate behaviour. Philos. Trans. R. Soc. B Biol. Sci. 372:20160321.
645 Klemme, I., H. Ylönen, and J. A. Eccard. 2007. Reproductive success of male bank voles
646 (Clethrionomys glareolus): the effect of operational sex ratio and body size. Behav. Ecol.
647 Sociobiol. 61:1911–1918.
648 Kokko, H., and M. D. Jennions. 2008. Parental investment, sexual selection and sex ratios. J.
649 Evol. Biol. 21:919–948.
650 Kokko, H., H. Klug, and M. D. Jennions. 2012. Unifying cornerstones of sexual selection:
651 operational sex ratio, Bateman gradient and the scope for competitive investment. Ecol.
652 Lett. 15:1340–1351.
653 Komdeur, J., T. Székely, X. Long, and S. A. Kingma. 2017. Adult sex ratios and their
654 implications for cooperative breeding in birds. Philos. Trans. R. Soc. B Biol. Sci. 372:5–
655 9.
656 Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for
657 ecology, evolution, and systematics. Methods Ecol. Evol. 7:573–579.
658 Lemaître, J. F., and J. M. Gaillard. 2013. Male survival patterns do not depend on male

allocation to sexual competition in large herbivores. Behav. Ecol. 24:421–428.

Lemaître, J. F., V. Ronget, M. Tidière, D. Allainé, V. Berger, A. Cohas, F. Colchero, D. A. Conde, M. Garratt, A. Liker, G. A. B. Marais, A. Scheuerlein, T. Székely, and J. M. Gaillard. 2020. Sex differences in adult lifespan and aging rates of mortality across wild mammals. Proc. Natl. Acad. Sci. U. S. A. 117:8546–8553.

Liker, A., R. P. Freckleton, V. Remeš, and T. Székely. 2015. Sex differences in parental care: Gametic investment, sexual selection, and social environment. Evolution 69:2862–2875.

Liker, A., R. P. Freckleton, and T. Székely. 2014. Divorce and infidelity are associated with skewed adult sex ratios in birds. Curr. Biol. 24:880–884.

Liker, A., R. P. Freckleton, and T. Székely. 2013. The evolution of sex roles in birds is related to adult sex ratio. Nat. Commun. 4:1587.

Liker, A., and T. Székely. 2005. Mortality costs of sexual selection and parental care in natural populations of birds. Evolution 59:890–897.

Lovich, J. E., J. W. Gibbons, and M. Agha. 2014. Does the timing of attainment of maturity influence sexual size dimorphism and adult sex ratio in turtles? Biol. J. Linn. Soc. 112:142–149.

Lukas, D., and T. H. Clutton-Brock. 2013. The evolution of social monogamy in mammals. Science 341:526–530.

Mayr, E. 1939. The Sex Ratio in Wild Birds. Am. Nat. 73:156–179.

McNamara, J. M., T. Székely, J. N. Webb, and A. I. Houston. 2000. A dynamic game-theoretic model of parental care. J. Theor. Biol. 205:605–623.

Meredith, R. W., J. E. Janecka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simao, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, and W. J. Murphy. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524.

Midford, P. E., T. J. Garland, and W. P. Maddison. 2011. PDAP:PDTREE module of Mesquite.

Mitani, J. C., J. Gros-Louis, and A. F. Richards. 1996. Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. Am. Nat. 147:966–980.

Moore, S. L., and K. Wilson. 2002. Parasites as a viability cost of sexual selection in natural

populations of mammals. Science 297:2015–2018.

Muralidhar, P., and M. A. Johnson. 2017. Sexual selection and sex ratios in Anolis lizards. J. Zool. 302:178–183.

Murray, B. G. 1984. A demographic theory on the evolution of mating systems as exemplified by birds. Pp. 71–140 *in* Hecht M.K., Wallace B., and Prance G.T., eds. Evolutionary Biology. Springer US, Boston, MA.

Myhrvold, N. P., E. Baldridge, B. Chan, D. Sivam, D. L. Freeman, and S. K. M. Ernest. 2015. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. Ecology 96:3109.

Nicholson, K. E., B. I. Crother, C. Guyer, and J. M. Savage. 2012. It is time for a new classification of anoles (Squamata: Dactyloidae). Zootaxa 3477:1–108.

Oaks, J. R. 2011. A time-calibrated species tree of crocodylia reveals a recent radiation of the true crocodiles. Evolution (N. Y). 65:3285–3297.

Orme, D., R. P. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2013. caper: Comparative Analyses of Phylogenetics and Evolution in R. Available at: https://cran.r-project.org/web/packa.

Owen-Smith, N. 1993. Comparative mortality rates of male and female kudus: the costs of sexual size dimorphism. J. Anim. Ecol. 62:428.

Owens, I. P. F., and P. M. Bennett. 1994. Mortality costs of parental care and sexual dimorphism in birds. Proc. R. Soc. B Biol. Sci. 257:1–8.

Paradis, E. 2012. Analysis of Phylogenetics and Evolution with R. Springer.

Pincheira-Donoso, D., and J. Hunt. 2017. Fecundity selection theory: concepts and evidence. Biol. Rev. 92:341–356.

Pipoly, I., V. Bókony, M. Kirkpatrick, P. F. Donald, T. Székely, and A. Liker. 2015. The genetic sex-determination system predicts adult sex ratios in tetrapods. Nature 527:91–94.

Poulin, R. 1997. Covariation of sexual size dimorphism and adult sex ratio in parasitic nematodes. Biol. J. Linn. Soc. 62:567–580.

Promislow, D. E. L. 1992. Costs of sexual selection in natural populations of mammals. Proc. B Biol. Sci. 247:203–210.

Promislow, D. E. L., R. Montgomerie, and T. E. Martin. 1992. Mortality costs of sexual dimorphism in birds. Proc. R. Soc. B Biol. Sci. 250:143–150.

Pyron, R., F. T. Burbrink, and J. J. Wiens. 2013. A phylogeny and revised classification of

Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13:93.

Queller, D. C. 1997. Why do females care more than males? Proc. R. Soc. London. Ser. B Biol. Sci. 264:1555–1557.

Remeš, V., R. P. Freckleton, J. Tökölyi, A. Liker, and T. Székely. 2015. The evolution of parental cooperation in birds. Proc. Natl. Acad. Sci. U. S. A. 112:13603–13608.

Rosseel, Y. 2012. Lavaan: An R package for structural equation modelling. J. Stat. Softw. 48:1–36.

Santos, J. C. 2012. Fast molecular evolution associated with high active metabolic rates in poison frogs. Mol. Biol. Evol. 29:2001–2018.

Sarre, S. D., T. Ezaz, and A. Georges. 2011. Transitions between sex-determining systems in reptiles and amphibians. Annu. Rev. Genomics Hum. Genet. 12:391–406.

Schacht, R., and M. Borgerhoff Mulder. 2015. Sex ratio effects on reproductive strategies in humans. R. Soc. Open Sci. 2:140402.

Schacht, R., K. L. Kramer, T. Székely, and P. M. Kappeler. 2017. Adult sex ratios and reproductive strategies: A critical re-examination of sex differences in human and animal societies. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160309.

Schacht, R., K. L. Rauch, and M. Borgerhoff Mulder. 2014. Too many men: the violence problem? Trends Ecol. Evol. 29:214–222.

Shine, R. 2005. Life-history evolution in reptiles. Annu. Rev. Ecol. Evol. Syst. 36:23–46.

Shipley, B. 2016. Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference with R. 2nd editio. Cambridge University Press.

Smith, R. J. 1999. Statistics of sexual size dimorphism. J. Hum. Evol. 36:423–458.

Spinks, P. Q., R. C. Thomson, M. Gidiş, and H. Bradley Shaffer. 2014. Multilocus phylogeny of the New-World mud turtles (Kinosternidae) supports the traditional classification of the group. Mol. Phylogenet. Evol. 76:254–260.

Székely, T., A. Liker, R. P. Freckleton, C. Fichtel, and P. M. Kappeler. 2014a. Sex-biased survival predicts adult sex ratio variation in wild birds. Proc. R. Soc. B Biol. Sci. 281:20140342.

Székely, T., T. Lislevand, and J. Figuerola. 2007. Sexual size dimorphism in birds. Pp. 27–37 *in* D. J. Fairbairn, W. U. Blanckenhorn, and T. Székely, eds. Sex, Size and Gender Roles. Oxford University Press, Oxford.

Székely, T., F. J. Weissing, and J. Komdeur. 2014b. Adult sex ratio variation: Implications for breeding system evolution. J. Evol. Biol. 27:1500–1512.

Tidière, M., J. M. Gaillard, D. W. H. Müller, L. B. Lackey, O. Gimenez, M. Clauss, and J. F. Lemaître. 2015. Does sexual selection shape sex differences in longevity and senescence patterns across vertebrates? A review and new insights from captive ruminants. Evolution 69:3123–3140.

Toïgo, C., and J. M. Gaillard. 2003. Causes of sex-biased adult survival in ungulates: Sexual size dimorphism, mating tactic or environment harshness? Oikos 101:376–384.

Trivers, R. L. 1972. Parental investment and sexual selection. Pp. 136–179 *in* In: Sexual Selection and the Descent of Man (ed. Campbell B), London: Heinemann.

Uetz, P., P. Freed, and J. (eds) Hošek. 2019. The Reptile Database.

van der Bijl, W. 2018. phylopath: Easy phylogenetic path analysis in R. PeerJ 2018:e4718.

von Hardenberg, A., and A. Gonzalez-Voyer. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution 67:378–387.

Wilson, K., and I. C. W. Hardy. 2002. Statistical analysis of sex ratios: an introduction. Pp. 48–92 *in* I. C. W. Hardy, ed. Sex Ratios. Cambridge University Press.

Wittenberger, J. F. 1976. The ecological factors selecting for polygyny in altrical birds. Am. Nat. 110:779–799.

Wittenberger, J. F. 1978. The evolution of mating systems in grouse. Condor 80:126–137.

Table 1. Phylogenetically corrected analyses of sexual size dimorphism (SSD) and adult sex ratio (ASR) in amniotes (reptiles, birds and mammals).

Predictors	$b \pm SE$	t	P	R^2	λ	n
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.183 ± 0. 036	5.037	**< 0.0001**	0.050	0.828 *#	462
Juvenile mortality bias	0.035 ± 0.070	0.493	0.623	< 0.001	0.737 *#	100
Adult mortality bias	0.014 ± 0.050	0.272	0.786	< 0.001	0.887 *#	230
Multi-predictor model 1:				*0.126*	*0.841 **	*457*
ASR	- 0.160 ± 0.035	4.555	**< 0.0001**			
Mean body mass	0.515 ± 0.087	5.950	**< 0.0001**			
Latitude	0.004 ± 0.038	0.103	0.918			
Sex determination, TSD [1]	- 0.297 ± 0.251	1.184	0.237			
Sex determination, ZW [1]	- 0.685 ± 0.264	2.592	**0.009**			
Multi-predictor model 2:				*0.273*	*0.814 *#*	*97*
ASR	- 0.271 ± 0.061	4.452	**< 0.0001**			
Mean body mass	0.377 ± 0.134	2.824	**0.006**			
Juvenile mortality bias	0.001 ± 0.060	0.011	0.992			
Adult mortality bias	- 0.019 ± 0.067	0.277	0.783			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.234 ± 0.051	4.593	**< 0.0001**	0.042	0.359 *#	462
Juvenile mortality bias	- 0.214 ± 0.099	2.151	**0.034**	0.035	0.281 *#	100
Adult mortality bias	- 0.257 ± 0.060	4.313	**< 0.0001**	0.071	0.288 *#	230
Multi-predictor model 1:				*0.071*	*0.247 *#*	*457*
SSD	- 0.188 ± 0.050	3.727	**0.0002**			
Mean body mass	- 0.106 ± 0.080	1.330	0.184			
Latitude	- 0.095 ± 0.045	2.135	**0.033**			
Sex determination, TSD [1]	0.481 ± 0.221	2.178	**0.030**			
Sex determination, ZW [1]	0.712 ± 0.205	3.471	**0.0006**			
Multi-predictor model 2:				*0.402*	*0.030 #*	*97*
SSD	- 0.457 ± 0.120	3.794	**0.0003**			
Mean body mass	- 0.249 ± 0.108	2.316	**0.023**			
Juvenile mortality bias	- 0.146 ± 0.086	1.702	0.092			
Adult mortality bias	- 0.259 ± 0.100	2.591	**0.011**			

Results of bivariate and multi-predictor phylogenetic generalized least-squares (PGLS) models with either (A) SSD ($\log_{10}$(male mass/female mass)) or (B) ASR (proportion of males in the adult population) as dependent variable. Mortality biases were calculated as $\log_{10}$(male mortality/female mortality) for juveniles and adults, respectively. $b \pm SE$ is the model's parameter estimate with its standard error (intercepts are not shown), t and P are the associated test statistic and its significance, λ is Pagel's lambda, n is number of species.
* λ statistically different from 0, # λ statistically different from 1.
[1] Differences from species with XY sex determination; overall effect of sex determination on SSD: $F_{2,451}$= 3.411, P= 0.034; on ASR: $F_{2,451}$= 6.135, P= 0.002.

Table 2. Phylogenetic path models of the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c) in amniotes (reptiles, birds and mammals).

Model/Path	Path coefficient ± *SE*	*Z*	*P*
Model 1a	***P_C= 0.972, df= 4, AICc= 15.8, ΔAICc= 0.0***		
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 1b	***P_C= 0.065, df= 6, AICc= 25.7, ΔAICc= 9.9***		
(AMB → ASR)[1]	0	-	-
JMB → ASR	- 0.258 ± 0.107	- 2.417	**0.018**
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 1c	***P_C= 0.376, df= 6, AICc= 19.9, ΔAICc= 4.1***		
AMB → ASR	- 0.378 ± 0.113	- 3.334	**0.001**
(JMB → ASR)[1]	0	-	-
ASR → SSD	- 0.425 ± 0.074	- 5.723	**0.000**
Model 2a	*P_C= 0.0, df= 4, AICc= 59.0, ΔAICc= 43.2*		
SSD → AMB	0.171 ± 0.105	1.631	0.106
SSD → JMB	0.111 ± 0.115	0.958	0.341
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
Model 2b	*P_C= 0.0, df= 4, AICc= 50.4, ΔAICc= 34.6*		
SSD → JMB	0.111 ± 0.115	0.958	0.341
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052
Model 2c	*P_C= 0.0, AICc= 50.4, ΔAICc= 34.6*		
SSD → AMB	0.171 ± 0.105	1.631	0.106
AMB → ASR	- 0.340 ± 0.113	- 3.000	**0.004**
JMB → ASR	- 0.205 ± 0.104	- 1.970	0.052

Model structures are shown in Figure S1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively (variables are explained in footnotes of Table 1). P_C is *P*-value for Fisher's *C* statistic for model fit, with non-significant values (> 0.05) indicating an acceptable fit. ΔAICc indicates difference in AICc values between the most supported model (lowest AICc, Model 1a) and the focal models. ΔAICc > 2 indicates substantially higher support for the best model than for the other models. The analyses include 97 species of reptiles, birds and mammals with data for all for variables.

[1] Path coefficient set to zero to keep the variable in the model.

Figure 1. Path models tested in the phylogenetic path analyses. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB: juvenile mortality bias, AMB: adult mortality bias. Dashed arrows indicate paths with coefficients set to zero to keep the variable in the model. Models 1a-c and 2a-c represent relationships as predicted by the mating opportunity hypothesis and the mortality cost hypothesis, respectively.

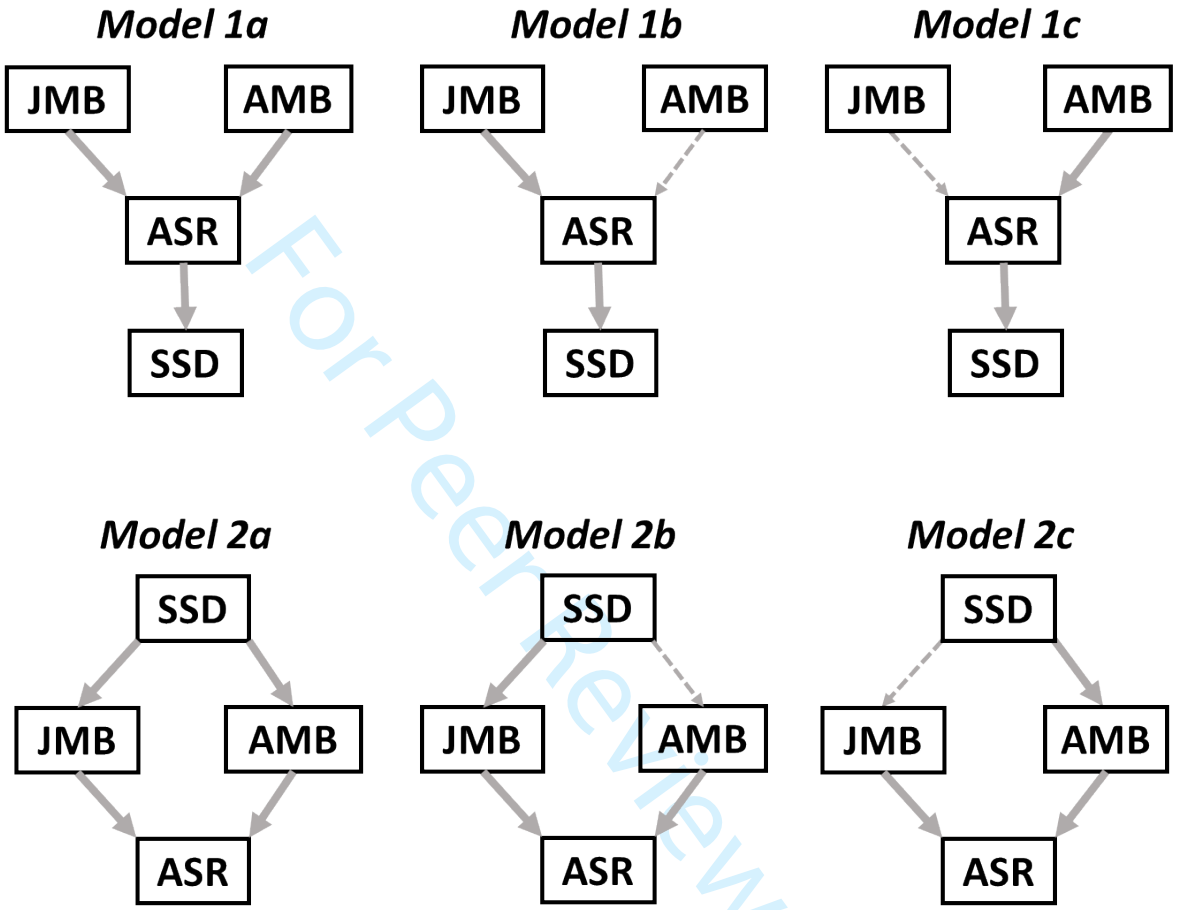


Figure 2. Sexual size dimorphism (SSD) in relation to adult sex ratio (ASR) in amniotes. SSD was calculated as $\log_{10}$(male mass/female mass); ASR is the proportion of males in the adult population. Each data point represents a species; the regression line is fitted by phylogenetic generalized least-squares (PGLS) model (see Table 1 for statistics).

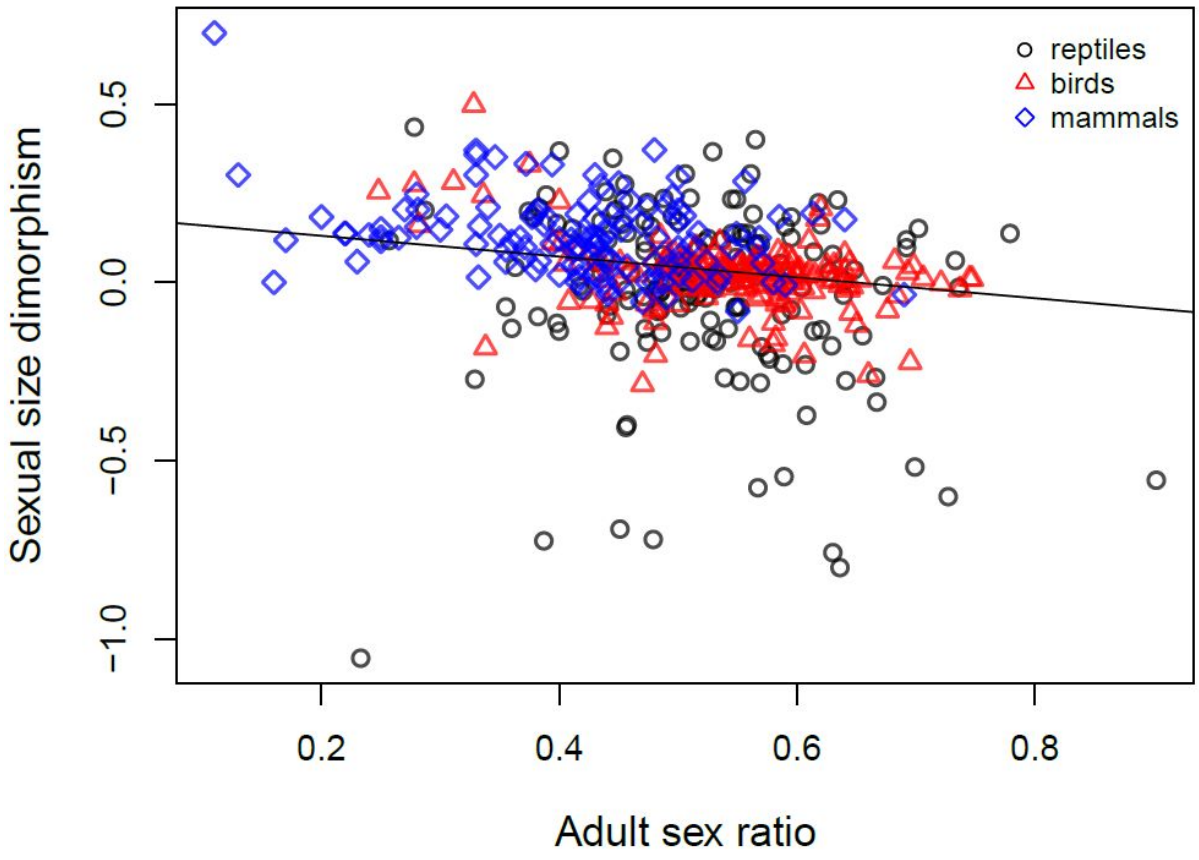
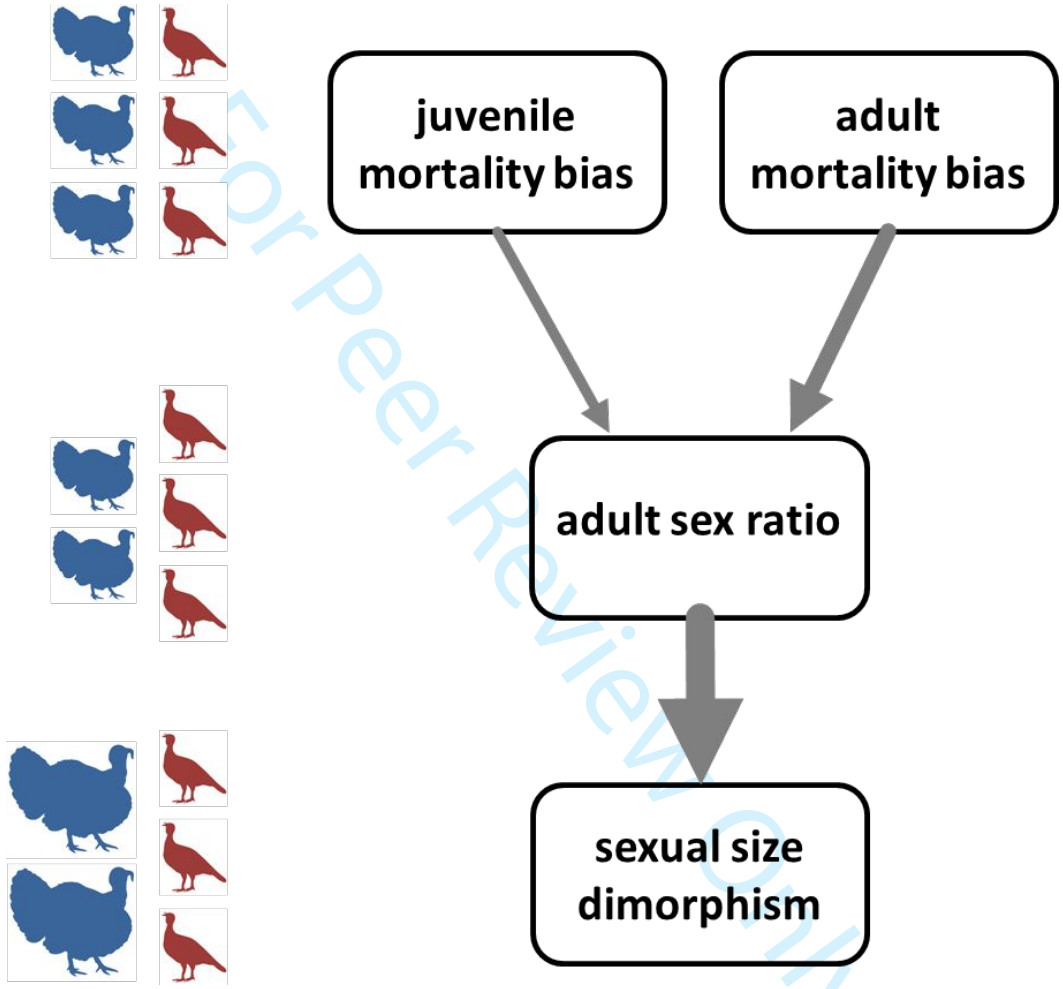


Figure 3. Path diagram of the best-fitting phylogenetic path model (Model 1a in Table 2, *n* = 97 species of reptiles, birds and mammals). The model supports the scenario that sex-biased juvenile and adult mortalities lead to skewed adult sex ratio, which in turn results in increased size dimorphism by sexual selection. Width of the arrows is proportional to path coefficients (see Table 2 for statistical details of the model). Bird pictures on the left illustrate the case when differential mortality generates female-skewed ASR, which then leads to a more male-biased SSD (i.e. larger body size in males relative to females). The path analyses were based on the approach proposed by Santos (2012), see Appendix S3 for details.



Electronic Supporting Information: tables and figures

Table S1. Relationship between SSD, ASR and sex-biased mortalities in reptiles, using estimated body mass data for SSD calculation.

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.060 ± 0.052	1.143	0.253	0.002	0.931 *#	155
Juvenile mortality bias	- 0.109 ± 0.102	1.065	0.304	0.008	0.352	17
Adult mortality bias	- 0.142 ± 0.088	1.610	0.113	0.030	1.0 *	62
Multi-predictor model 1:				*0.116*	*0.956 **	*153*
ASR	- 0.090 ± 0.075	1.203	0.231			
Mean body mass	0.715 ± 0.178	4.019	< **0.0001**			
Latitude	- 0.175 ± 0.126	1.389	0.167			
Reproductive mode [1]	0.348 ± 0.313	1.112	0.268			
Sex determination, TSD [2]	- 0.463 ± 0.384	1.206	0.230			
Sex determination, ZW [2]	- 1.003 ± 0.313	2.344	**0.020**			
Multi-predictor model 2:				*< 0.001*	*0.0*	*17*
ASR	- 0.022 ± 0.252	0.086	0.933			
Mean body mass	- 0.452 ± 0.523	0.865	0.404			
Juvenile mortality bias	- 0.500 ± 0.374	1.339	0.205			
Adult mortality bias	0.284 ± 0.429	0.662	0.520			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.145 ± 0.080	1.819	0.071	0.015	0.0 #	155
Juvenile mortality bias	- 0.251 ± 0.217	1.156	0.266	0.021	0.0#	17
Adult mortality bias	- 0.179 ± 0.103	1.732	0.088	0.032	0.155 #	62
Multi-predictor model 1:				*0.078*	*0.0 #*	*153*
SSD	- 0.049 ± 0.055	0.891	0.374			
Mean body mass	0.173 ± 0.108	1.599	0.112			
Latitude	- 0.001 ± 0.109	0.013	0.990			
Reproductive mode [1]	- 0.140 ± 0.216	0.650	0.517			
Sex determination, TSD [2]	0.209 ± 0.224	0.934	0.352			
Sex determination, ZW [2]	0.667 ± 0.216	3.091	**0.002**			
Multi-predictor model 2:				*0.165*	*0.0 #*	*17*
SSD	- 0.028 ± 0.331	0.086	0.933			
Mean body mass	0.929 ± 0.556	1.671	0.121			
Juvenile mortality bias	- 0.044 ± 0.459	0.095	0.926			
Adult mortality bias	- 0.641 ± 0.465	1.377	0.194			

* Pagel's lambda statistically different from 0, # lambda statistically different from 1.

[1] Differences from oviparous species.

[2] Differences from XY species; overall effect of sex determination on SSD: $F_{2,146}$= 2.5, *P*= 0.083; on ASR: $F_{2,146}$= 5.3, *P*= 0.006.

For further explanation, see the footnotes of Table 1 in the main text.

872 **Table S2.** Relationship between SSD, ASR and sex-biased mortalities in reptiles, using body
873 length data for SSD calculation.
874

Predictors	*b* ± *SE*	*t*	*P*	R^2	λ	*n*
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.060 ± 0.053	1.132	0.259	0.002	0.924 *#	155
Juvenile mortality bias	- 0.141 ± 0.115	1.221	0.241	0.030	0.531	17
Adult mortality bias	- 0.154 ± 0.090	1.707	0.093	0.030	1.0 *	62
Multi-predictor model 1:				*0.122*	*0.52 **	*153*
ASR	- 0.048 ± 0.042	1.126	0.262			
Mean body mass	0.391 ± 0.100	3.891	**< 0.001**			
Latitude	- 0.103 ± 0.071	1.459	0.147			
Reproductive mode [1]	0.179 ± 0.177	1.015	0.312			
Sex determination, TSD [2]	- 0.223 ± 0.216	1.032	0.304			
Sex determination, ZW [2]	- 0.633 ± 0.241	2.628	**0.009**			
Multi-predictor model 2:				*< 0.001*	*0.0*	*17*
ASR	- 0.059 ± 0.161	0.368	0.719			
Mean body mass	- 0.347 ± 0.334	1.038	0.320			
Juvenile mortality bias	- 0.354 ± 0.239	1.484	0.164			
Adult mortality bias	0.092 ± 0.274	0.337	0.742			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.147 ± 0.080	1.841	0.068	0.015	0.0 #	155
Juvenile mortality bias	- 0.251 ± 0.217	1.156	0.266	0.021	0.0#	17
Adult mortality bias	- 0.179 ± 0.103	1.732	0.088	0.032	0.155 #	62
Multi-predictor model 1:				*0.078*	*0.0 #*	*153*
SSD	- 0.085 ± 0.098	0.870	0.386			
Mean body mass	0.172 ± 0.109	1.588	0.114			
Latitude	- 0.001 ± 0.108	0.007	0.994			
Reproductive mode [1]	- 0.141 ± 0.216	0.654	0.514			
Sex determination, TSD [2]	0.214 ± 0.223	0.958	0.340			
Sex determination, ZW [2]	0.667 ± 0.216	3.089	**0.002**			
Multi-predictor model 2:				*0.174*	*0.0 #*	*17*
SSD	- 0.188 ± 0.512	0.368	0.720			
Mean body mass	0.867 ± 0.570	1.522	0.154			
Juvenile mortality bias	- 0.096 ± 0.463	0.208	0.839			
Adult mortality bias	- 0.624 ± 0.457	1.366	0.197			

875
876 * Pagel's lambda statistically different from 0, # lambda statistically different from 1.
877 [1] Differences from oviparous species.
878 [2] Differences from XY species; overall effect of sex determination on SSD: $F_{2,146}$= 3.4, *P*=
879 0.035; on ASR: $F_{2,146}$= 5.2, *P*= 0.006.
880 For further explanation, see the footnotes of Table 1 in the main text.
881

882 **Table S3.** Relationship between SSD, ASR and sex-biased mortalities in birds.
883

Predictors	$b \pm SE$	t	P	R^2	λ	n
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.393 ± 0.055	7.186	**< 0.001**	0.216	0.770 *#	185
Juvenile mortality bias	0.185 ± 0.167	1.108	0.274	0.005	0.861 *	47
Adult mortality bias	0.131 ± 0.082	1.608	0.110	0.013	0.625 *#	123
Multi-predictor model 1:				*0.272*	*0.720 *#*	*178*
ASR	- 0.242 ± 0.038	6.390	**< 0.0001**			
Mean body mass	0.261 ± 0.100	2.599	**0.010**			
Latitude	- 0.020 ± 0.034	0.584	0.560			
Display type, agile [1]	- 0.338 ± 0.090	3.748	**< 0.001**			
Multi-predictor model 2:				*0.386*	*1.0 **	*47*
ASR	- 0.346 ± 0.080	4.318	**< 0.0001**			
Mean body mass	0.424 ± 0.246	1.719	0.093			
Juvenile mortality bias	- 0.032 ± 0.065	0.489	0.627			
Adult mortality bias	- 0.139 ± 0.141	0.992	0.327			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.499 ± 0.077	6.520	**< 0.001**	0.184	0.480 *#	185
Juvenile mortality bias	- 0.494 ± 0.160	3.084	**0.003**	0.156	0.0 #	47
Adult mortality bias	- 0.401 ± 0.082	4.866	**< 0.001**	0.157	0.0 #	123
Multi-predictor model 1:				*0.239*	*0.244 #*	*178*
SSD	- 0.717 ± 0.116	6.183	**< 0.0001**			
Mean body mass	- 0.191 ± 0.136	1.406	0.161			
Latitude	- 0.127 ± 0.058	2.201	**0.029**			
Display type, agile [1]	- 0.589 ± 0.161	3.667	**< 0.001**			
Multi-predictor model 2:				*0.397*	*0.0 #*	*47*
SSD	- 0.382 ± 0.153	2.499	**0.016**			
Mean body mass	- 0.128 ± 0.198	0.646	0.522			
Juvenile mortality bias	- 0.199 ± 0.109	1.831	0.074			
Adult mortality bias	- 0.468 ± 0.139	3.368	**0.002**			

884
885 * Pagel's lambda statistically different from 0, # lambda statistically different from 1.
886 [1] Difference from non-agile species.
887 For further explanation, see the footnotes of Table 1 in the main text.
888

889

890 **Table S4.** Relationship between SSD, ASR and sex-biased mortalities in mammals.
891

Predictors	$b \pm SE$	t	P	R^2	λ	n
(A) response: sexual size dimorphism						
Bivariate models:						
ASR	- 0.314 ± 0.088	3.561	**< 0.001**	0.088	0.190 ǂ	122
Juvenile mortality bias	0.089 ± 0.216	0.414	0.682	< 0.001	0.096 ǂ	36
Adult mortality bias	0.071 ± 0.171	0.414	0.681	< 0.001	0.039 ǂ	45
Multi-predictor model 1:				*0.166*	*0.342 *ǂ*	*120*
ASR	- 0.153 ± 0.058	2.646	**0.009**			
Mean body mass	0.418 ± 0.131	3.191	**0.002**			
Latitude	0.106 ± 0.053	2. 016	**0.046**			
Multi-predictor model 2:				*0.250*	*0.0 ǂ*	*33*
ASR	- 0.374 ± 0.129	2.900	**0.007**			
Mean body mass	0.209 ± 0.237	0.880	0.386			
Juvenile mortality bias	0.088 ± 0.116	0.767	0.452			
Adult mortality bias	- 0.016 ± 0.119	0.134	0.894			
(B) response: adult sex ratio						
Bivariate models:						
SSD	- 0.302 ± 0.085	3.539	**< 0.001**	0.087	0.252 *ǂ	122
Juvenile mortality bias	- 0.036 ± 0.184	0.195	0.847	< 0.001	0.0 ǂ	36
Adult mortality bias	- 0.080 ± 0.163	0.493	0.624	< 0.001	0.0 ǂ	45
Multi-predictor model 1:				*0.008*	*0.367 *ǂ*	*114*
SSD	- 0.375 ± 0.140	2.670	**0.009**			
Mean body mass	- 0.314 ± 0.209	1.500	0.136			
Latitude	- 0.075 ± 0.083	0.907	0.366			
Multi-predictor model 2:				*0.293*	*0.0 ǂ*	*33*
SSD	- 0.617 ± 0.213	2.900	**0.007**			
Mean body mass	- 0.494 ± 0.294	1.678	0.104			
Juvenile mortality bias	- 0.043 ± 0.150	0.285	0.778			
Adult mortality bias	0.022 ± 0.153	0.142	0.888			

892
893 * Pagel's lambda statistically different from 0, ǂ lambda statistically different from 1.
894 For further explanation, see the footnotes of Table 1 in the main text.
895
896

Table S5. Sensitivity analyses of the relationship between sexual size dimorphism (SSD, dependent variable in all models) and adult sex ratio (ASR). Table shows results when (A) reptiles are included with SSD based on body length, (B) reptiles are included with SSD calculated from sex-specific body mass, (C) reptiles with temperature-dependent sex determination (TSD) are excluded, (D) reptiles with assumed sex determination, based on related species, are excluded, and (E) environmental harshness is included in the model.

Predictors	$b \pm SE$	t	P	R^2	λ	n
(A) Bivariate model, reptiles included with SSD calculated from body length:						
ASR (all species)	- 0.252 ± 0.039	6.492	**< 0.0001**	0.082	0.664 *#	462
(B) Bivariate model, reptiles included with SSD calculated from body mass[1]:						
ASR (all species)	- 0.306 ± 0.051	6.037	**< 0.0001**	0.095	0.745 *#	338
(C) Bivariate models, TSD reptiles excluded:						
ASR (all species)	- 0.271 ± 0.045	6.077	**< 0.0001**	0.082	0.756 *#	402
ASR (reptiles only)	- 0.119 ± 0.131	0.909	0.366	< 0.001	0.915 *	95
(D) Bivariate model, assumed sex determination excluded[2]:						
ASR (all species)	- 0.186 ± 0.038	4.968	**< 0.0001**	0.055	0.817 *#	409
(E) Multi-predictor model with environmental harshness[3]:						
birds and mammals:				*0.110*	*0.757 *#*	*219*
ASR	- 0.192 ± 0.036	5.525	**< 0.0001**			
Environmental harshness	0.028 ± 0.046	0.618	0.538			
reptiles:				*0.026*	*0.944 **	*58*
ASR	- 0.033 ± 0.133	0.246	0.807			
Environmental harshness	0.117 ± 0.063	1.865	0.066			
all species:				*0.078*	*0.849 *#*	*277*
ASR	- 0.167 ± 0.039	4.301	**< 0.0001**			
Environmental harshness	0.082 ± 0.034	2.439	**0.015**			

* Pagel's lambda statistically different from 0, # lambda statistically different from 1

[1] Sex-specific body mass data from Myhrvold et al. (2015).

[2] Sex determination mechanism assumed to be the same type as reported for the genus or family (see Methods).

[3] The influence of environmental harshness was tested in birds and mammals using data from Botero et al. (2014), in reptiles using data calculated in this study (following the method of Botero et al 2014), and in all species by pooling the harshness scores from the two studies.

Table S6. Analyses of the relationship between SSD (dependent variable) and ASR with branch lengths calculated by three different methods for the phylogeny used in the PGLS models. The analyses included reptiles, birds, and mammals.

Branch length method[1]	$b \pm SE$	t	P	R^2	λ	n
Nee's method	- 0.183 ± 0.036	5.037	**< 0.0001**	0.050	0.828 *#	462
Pagel's method	- 0.180 ± 0.036	4.966	**< 0.0001**	0.049	0.839 *#	462
unit branch length	- 0.195 ± 0.033	5.833	**< 0.0001**	0.067	1.0 *	462

* Pagel's lambda statistically different from 0, # lambda statistically different from 1
[1] See Methods for details of branch length calculations

Table S7. Analyses of the relationship between SSD (dependent variable) and ASR in socially monogamous and socially polygamous species, respectively. The analyses included birds and mammals.

Mating system[1]	$b \pm SE$	t	P	R^2	λ	n
Monogamy	- 0.070 ± 0.038	1.829	0.070	0.021	1.0 *	109
Polygamy	- 0.275 ± 0.049	5.599	**< 0.0001**	0.159	0.385 *#	162

* Pagel's lambda statistically different from 0, # lambda statistically different from 1
[1] When monogamous and polygynous species are analysed together, there is a statistically significant interaction between the effects of mating system and ASR ($b \pm SE = 0.218 \pm 0.087$, $t = 2.510$, $P = 0.013$).

Table S8. Results of the phylogenetic path analyses using the R package 'phylopath'. Models represent the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c). Analyses based on data of all species (birds, mammals, and reptiles; *n*= 97 species).

Model	*k*	*q*	*C*	*P*	*CICc*	*ΔCICc*
Model 1a	**3**	**7**	**6.4**	**0.383**	**21.6**	**0.0**
Model 1b	4	6	18.7	0.017	31.6	10.0
Model 1c	**4**	**6**	**11.2**	**0.188**	**24.2**	**2.6**
Model 2a	2	8	32.4	<0.001	50.0	28.4
Model 2b	3	7	34.8	<0.001	50.0	28.4
Model 2c	3	7	36.6	<0.001	51.9	30.3

Model structures are shown in Figure 1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively. The table shows the number of independence claims (*k*), the number of parameters (*q*), Fisher's C statistic (*C*) and its accompanying probability (*P*), C-statistic information criterion corrected for small sample sizes (CICc), and the difference in CICc from the top model (ΔCICc). A *P*-value less than 0.05 indicates a poor model fit (i.e. rejection of the model), whereas a ΔCICc > 2 indicates substantial support for the top path model over the alternative models.

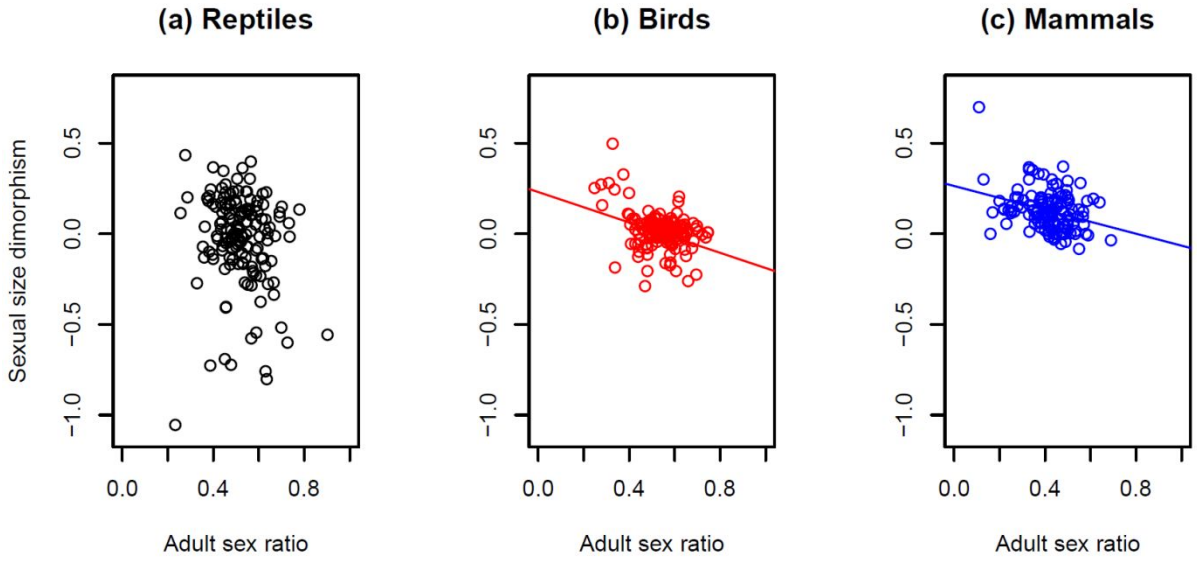
Table S9. Phylogenetic path models representing the mating opportunity hypothesis (Models 1a-c) and the mortality cost hypothesis (Models 2a-c). Analyses with data of birds and mammals (i.e. excluding reptiles; *n*= 81 species).

Model/Path	Path coefficient ± *SE*	*Z*	*P*
Model 1a	***P_C= 0.991, df= 4, AICc= 15.9, ΔAICc= 0.0***		
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 1b	***P_C= 0.142, df= 6, AICc= 23.7, ΔAICc= 7.8***		
(AMB → ASR)[1]	0	-	-
JMB → ASR	- 0.257 ± 0.112	- 2.289	**0.025**
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 1c	***P_C= 0.428, df= 6, AICc= 19.7, ΔAICc= 3.8***		
AMB → ASR	- 0.358 ± 0.121	- 2.958	**0.004**
(JMB → ASR)[1]	0	-	-
ASR → SSD	- 0.719 ± 0.122	- 5.887	**0.000**
Model 2a	*P_C= 0.0, df= 4, AICc= 59.7, ΔAICc= 43.8*		
SSD → AMB	0.117 ± 0.077	1.503	0.137
SSD → JMB	0.102 ± 0.086	1.187	0.239
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
Model 2b	*P_C= 0.0, df= 4, AICc= 50.5, ΔAICc= 34.6*		
SSD → JMB	0.102 ± 0.086	1.187	0.239
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059
Model 2c	*P_C= 0.0, AICc= 50.5, ΔAICc= 34.6*		
SSD → AMB	0.117 ± 0.077	1.503	0.137
AMB → ASR	- 0.321 ± 0.121	- 2.662	**0.009**
JMB → ASR	- 0.210 ± 0.109	- 1.920	0.059

Model structures are shown in Figure 1. SSD: sexual size dimorphism, ASR: adult sex ratio, JMB and AMB: juvenile and adult mortality biases, respectively (variables are explained in footnotes of Table 1). P_C is *P*-value for Fisher's *C* statistic for model fit, with non-significant values (> 0.05) indicating an acceptable fit. ΔAICc indicates difference in AICc values between the most supported model (lowest AICc, Model 1a) and the focal models. ΔAICc > 2 indicates substantially higher support for the best model than for the other model.

[1] Path coefficient set to zero to keep the variable in the model.

Figure S1. Sexual size dimorphism in relation to adult sex ratio in (a) reptiles (PGLS, b ± *SE* = -0.060 ± 0.052, *P*= 0.253, *n*= 155 species), (b) birds (*b* ± *SE* = -0.392 ± 0.055, *P*< 0.001, *n*= 185), and (c) mammals (*b* ± *SE* = -0.314 ± 0.088, *P* < 0.001, *n*= 122). Each data point represents a species, and lines show statistically significant regressions fitted by PGLS (see Tables S1-4 for further statistical details).



Electronic Supporting Information: Appendix S1

References for data sources are provided in a separate excel file. The full dataset will be published together with references after the manuscript is accepted for publication.

Electronic Supporting Information: Appendix S2

Parameters of the allometric equations between body length and body mass (log10(mass in g) = a + b*log10(length in mm)) used for the calculation of mass estimates in reptiles. n is the number of species included in the analyses.

Taxon	**Intercept (*a*)**	**Slope (*b*)**	***n***	**Reference**
Snakes	-5.773	2.786	336	Feldman and Meiri (2013)
Squamates and crocodilians	-4.52	2.923	600	Meiri (2010)
Turtles	-3.535	2.887	199	Regis and Meik (2017)

References:

Feldman, A., and S. Meiri. 2013. Length-mass allometry in snakes. Biol. J. Linn. Soc. 108:161–172.

Meiri, S. 2010. Length-weight allometries in lizards. J. Zool. 281:218–226.

Regis, K. W., and J. M. Meik. 2017. Allometry of sexual size dimorphism in turtles: a comparison of mass and length data. PeerJ 5:e2914.

Electronic Supporting Information: Appendix S3

Methodological notes on path analyses applied to comparative data

Several approaches have been proposed for applying path analysis in phylogenetic comparative studies (von Hardenberg and Gonzalez-Voyer 2013). However, some methodological aspects of this method still pose challenges for its phylogenetic applications. Here we discuss two of such aspects: (1) the problem of using bivariate phylogenetic generalized least squares (PGLS) models to estimate correlations between variables in the path models, and (2) the reliability of AIC statistics to compare non-nested path models. We suggest ways to avoid these problems which may help further applications of path analysis to phylogenetic data.

1. The problem of using PGLS in path analyses

The analysis of multivariate dependent data is a notoriously thorny problem. In this dataset we, as in common with many similar analyses, had to deal with the problem of phylogenetic non-independence (Harvey and Pagel 1991). Techniques for analysing cause-effect relationships using linear models are well developed (Felsenstein 1985; Harvey and Pagel 1991; Hansen and Martins 1996; Pagel 1997; Freckleton et al. 2002). These techniques are designed for analysis of data in which there is a dependent variable of interest, which is modelled as a function of independent predictors. In these models the effect of phylogeny is accounted for by modelling phylogenetic dependence in the residual term, and we have used these for several analyses (e.g. Table 1 in the main text). However, analyses of data in which variables are treated as multivariate responses are much less common. Phylogenetic principal components analysis is one exception, although recent research has stressed that this may be more complex than previously realised (Uyeda et al. 2015).

Path analysis is a method of multivariate trait analysis that allows complex dependencies among variables to be modelled. von Hardenberg & Gonzalez-Voyer (2013) presented a new method for performing path analysis on phylogenetically dependent data. This approach accounts for phylogenetic non-independence through constructing a series of bivariate PGLS models in which one variable is treated as the dependent variable, and the other is the independent variable. Importantly, this approach permits variable levels of

dependence to be modelled through estimating Pagel's λ, which accounts for varying contributions of phylogeny to trait variation.

However the analysis of such data is complex. Consider a simple example, in which we have three variables x, y, and z. If we model $x \rightarrow y$ (i.e. where x is the predictor and y is the response variable) using the PGLS-λ approach, we might well estimate a different correlation than if we model $y \rightarrow x$. This is because in PGLS the estimate of λ depends on the direction of the relationship. In path analysis correlations between pairs of variables are the input and the directionality of the statistical model should not be an issue. Specifically, the net correlation r_{xz} should then be the product of the component pairwise correlations, i.e. $r_{xz} = r_{xy} \times r_{yz}$. However, if we use PGLS and $\lambda \neq 0$ and $\lambda \neq 1$ (like in most cases in our analyses, see Table 1 in the main text), then in general $r_{xz} = r_{xy} \times r_{yz} \neq r_{yx} \times r_{yz} = r_{xz}$.

A further issue is that when we extend to several variables, the approach can produce correlations that are difficult to compare because they have different evolutionary interpretations. For instance, if the value of λ is 1 (strong phylogenetic signal) for the path $x \rightarrow y$, but the value is 0 (no phylogenetic signal) for $y \rightarrow z$, then the overall model relating x to z is not consistent in an evolutionary sense. This is because the correlation between x and y is an evolutionary one, but the relationship between y and z is not. This is equivalent to saying that the two correlations r_{xy} and r_{yz} are measured on two different scales. To see this, in this specific example we can estimate r_{xy} from phylogenetic contrasts, by transforming x to $p(x|tr)$ and y to $p(y|tr)$, where the transformation $p()$is the contrasts given tree tr. The correlation estimated for x and y is thus, effectively, the correlation between p(x) and p(y), i.e. is effectively $r_{p(x)\,p(y)}$. In contrast the correlation between y and z is estimated on the untransformed scale, i.e. r_{yz}. Thus, the variable y is not the same in the two cases, which calls into question the validity of calculating a net correlation from r_{xy} and r_{yz} estimated using different error models.

In our analyses we overcame the above difficulties by avoiding the use of bivariate PGLS models to estimate correlations between the variables in the path models. We followed the approach proposed by Santos (2012), in which first we calculated λ for each individual variable and conducted a phylogenetic transformation on the variable using that estimate of λ. Then we fitted path models to these already transformed data using ordinary fitting methods developed for non-phylogenetic data. Details of the analyses are described in the Methods section of the main text.

2. The problem of comparing non-nested path models by AIC statistics

In phylogenetic comparative studies the direction of causality between variables is often unknown, and different evolutionary hypotheses may propose opposing cause - effect relationships (like the mating competition and the mortality cost hypotheses in our study, see Fig. S1). These hypotheses may be represented by different path models, and then their fit to the data can be compared by some comparative fit indices, most commonly by AIC (West et al. 2012). However, simulations suggest that conclusions of path model comparisons based on information theory approach (like AIC) can be unreliable (Preacher and Merkle 2012). In addition the competing models can be non-nested (non-hierarchical) (e.g. Models 1a versus 2a in Fig. S1), for which AIC-based comparison should be applied with caution (Kline 2015).

To explore the problem of model comparison in the context of our study, first we fitted our path models to the real dataset by two alternative methods: (1) by covariance matrix comparison, as implemented in the R package *lavaan* (Rosseell 2012), and (2) by piecewise structural equation modelling (or d-separation) method, as implemented in the *piecewiseSEM* (Lefcheck 2016) package. We compared path coefficient estimates and various model fit indices between these two methods to evaluate whether they produce consistent conclusions. Second, we used the same two methods and R implementations to fit the models to simulated datasets, and tested which of the methods produces more reliable (less biased) model comparisons.

2.1. Fitting path models to real data

The general steps of model fitting procedure we followed in this study are described in the Methods section of the main text. We performed model fitting with the two R packages *piecewiseSEM* and *lavaan*. In *piecewiseSEM* and *lavaan* the global model fit for each individual path model is evaluated by Fisher's *C* and χ^2 statistics, respectively, where a statistically non-significant result means acceptable fit. In *lavaan*, several other measures for model fit of individual models are also available, and here we report four of the most widely used indices (TLI, CFI, RMSEA, SRMR). It has been proposed that that the values of TLI and CFI > 0.95, RMSEA < 0.06, and SRMR < 0.08 indicate acceptable/good fit of models to the data (West et al. 2012).

We found that the two methods produced highly consistent estimates for the standardised path coefficients in all path models (*piecewiseSEM*: Table 1 in the main text, *lavaan*: Table S9 below). The effect of juvenile mortality on ASR was marginally not

significant in most *piecewiseSEM* models whereas it was significant with all *lavaan* models. For all other relationships the two methods produced consistent results.

Table S9. Estimates of standardised path coefficients for the six path models representing various relationships between SSD, ASR, and sex biases in adult (AMB) and juvenile (JMB) mortality, obtained by the R package *lavaan* (see Fig. S1 for model details). Significant relationships are highlighted in bold.

Model/Path	Path coefficient ± SE	Z	P
Model 1a			
AMB → ASR	**- 0.340 ± 0.112**	**- 3.048**	**0.002**
JMB → ASR	**- 0.205 ± 0.102**	**- 2.002**	**0.045**
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 1b			
(AMB → ASR)[1]	0	-	-
JMB → ASR	**- 0.258 ± 0.105**	**- 2.443**	**0.015**
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 1c			
AMB → ASR	**- 0.378 ± 0.112**	**- 3.370**	**0.001**
(JMB → ASR)[1]	0	-	-
ASR → SSD	**- 0.657 ± 0.107**	**- 6.144**	**0.000**
Model 2a			
SSD → AMB	0.117 ± 0.070	1.680	0.093
SSD → JMB	0.089 ± 0.077	1.157	0.247
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**
Model 2b			
SSD → JMB	0.089 ± 0.077	1.157	0.247
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**
Model 2c			
SSD → AMB	0.117 ± 0.070	1.680	0.093
AMB → ASR	**- 0.340 ± 0.110**	**- 3.092**	**0.002**
JMB → ASR	**- 0.205 ± 0.101**	**- 2.031**	**0.042**

[1] Path coefficient set to zero

The two methods also produced highly consistent results for model fit as evaluated by global fit indices (i.e. C and χ^2 statistics, respectively, see Table S10). The only difference was that for Model 1b *piecewiseSEM* indicated 'marginally acceptable' model fit whereas *lavaan* indicated poor model fit for this path model. The other fit indices (TLI, CFI, RMSEA, and SRMR) suggest conclusions that are fully consistent with C statistics and χ^2 tests, i.e. acceptable fit for Models 1a and 1c by all of these indices and unacceptable fit for all other models (Table S10).

Table S10. Fit indices for the six path models, obtained by *piecewiseSEM* and *lavaan*. Values indicating acceptable fit are highlighted in bold.

Model	piecewiseSEM				lavaan					
	C	df	P_c		χ^2	df	P_{χ^2}	TLI	CFI	RMSEA	SRMR
1a	**0.29**	**4**	**0.972**		**0.02**	**2**	**0.991**	**1.119**	**1.000**	**0.000**	**0.004**
1b	**11.6**	**6**	**0.065**		8.9	3	0.031	0.764	0.858	0.143	0.101
1c	**6.2**	**6**	**0.376**		**3.9**	**3**	**0.267**	**0.962**	**0.977**	**0.057**	**0.065**
2a	34.7	4	0.000		30.6	2	0.000	-0.978	0.341	0.386	0.154
2b	34.7	4	0.000		30.6	2	0.000	-0.719	0.313	0.386	0.154
2c	34.7	4	0.000		30.6	2	0.000	-0.661	0.336	0.386	0.154

2.2. AIC-based model comparisons using real and simulated data

To assess which of these models provides the best account of the data, first we calculated the AIC value for each model (in *piecewiseSEM* this is corrected for small sample size, i.e. AIC_c) using the real dataset. Second, we used simulated data to test which of the two methods produces less biased conclusions. For this latter purpose, we generated simulated datasets using the R function 'rnorm'. The simulated datasets have the same number of variables and sample size as the phylogenetically transformed real dataset. We fitted path models with both *piecewiseSEM* and *lavaan* to obtain the AIC (or AIC_c) values. Then we compared Model 1a (the model that got the highest support for model fit by the global fit indices, see Table S10) to the other five models (Models 1b,1c, 2a, 2b, and 2c), thus conducted five pairwise comparisons, repeated with the two methods. These paired comparisons between models mimic the comparison we conducted with the real dataset in our study (Table 2 in the main

text). We calculated ΔAIC for each comparison as the difference between AIC values of the two models (i.e. AIC of compared model - AIC of Model 1a, thus a positive ΔAIC value indicates better fit for Model 1a). We repeated this procedure with 1000 simulated datasets that resulted in 1000 ΔAIC values for each pairwise comparison. To assess whether the comparison of two particular models produces biased results with simulated data we calculated (1) the mean ΔAIC value of the 1000 runs ($\Delta AIC_{simulation}$), and (2) the probability that the simulated ΔAIC was larger than the ΔAIC value we got with the real dataset ($P_{\geq \Delta AIC\_sim}$).

Using real data, *piecewiseSEM* gave the lowest AIC_c for Model 1a (Table S11), a result consistent with global model fit evaluation (see Table S10). ΔAICc values suggested strong support for this model in all comparisons (ΔAICc ≥ 4.1, Table S11). In contrast, *lavaan* results were inconsistent with global model fit evaluation because it gave very strong support for Model 2c (Table S11), a model that had an unacceptable fit by all fit indices (see Table S10).

Table S11. AIC-based model comparison using real and simulated data by the two methods. AIC_c (*piecsewiseSEM*) and AIC (*lavaan*) values provided for all models are based on analyses of our real data. ΔAIC_{data} and $\Delta AIC_{simulation}$ show differences from Model 1a in pairwise comparisons, based on analyses of real or simulated data, respectively. $P_{\geq \Delta AIC\_sim}$ indicates the probability that analyses of random data result in as large or larger AIC differences in support for Model 1a than the ΔAIC values obtained with real data.

Model	**piecewiseSEM**					**lavaan**			
	AIC_c	**ΔAIC_{data}**	**$\Delta AIC_{simulation}$**	**$P_{\geq \Delta AIC\_sim}$**		**AIC**	**ΔAIC_{data}**	**$\Delta AIC_{simulation}$**	**$P_{\geq \Delta AIC\_sim}$**
1a	**15.8**	**0.0**	-	-		382.4	0.0	-	-
1b	25.7	9.9	-0.2	0.003		389.3	6.9	-1.0	0.004
1c	19.9	4.1	-0.9	0.042		384.3	1.9	-1.1	0.041
2a	59.0	43.2	7.4	0.0		521.3	138.9	274.3	1.0
2b	50.4	34.6	-0.1	0.0		360.0	-22.4	273.4	1.0
2c	50.4	34.6	0.4	0.0		**341.7**	**-40.7**	273.4	1.0

Using simulated data, we found that *piecewiseSEM* produced less biased results than *lavaan*. First, in most cases mean simulated ΔAIC values were small and there was no strong bias in favour of one specific model (see $\Delta AIC_{simulation}$ in Table S11), as one would expect with

random data. The only exception was the comparison between Model 1a and Model 2a in which simulated ΔAIC produced by *piecewiseSEM* was 7.4, favouring Model 1a. Importantly, however, these simulations indicated only a low probability for random data resulting in as large or larger AIC differences (43.2) in support for Model 1a than the ΔAIC values we obtained with real data (see low $P_{\geq \Delta AIC\_sim}$ values in Table S11), suggesting that support for Model 1a was unlikely the result of biased AIC estimates.

In contrast, simulations showed that *lavaan* produced highly biased ΔAIC values in all non-nested comparisons (see the high $\Delta AIC_{simulation}$ and $P_{\geq \Delta AIC\_sim}$ values for Models 2a, 2b and 2c in Table S9). On the other hand, for nested model comparisons (i.e. with Models 1b and 1c) *lavaan* produced unbiased results similarly to those we got with *piecsewiseSEM* (Table S11).

These analyses suggest that the two methods gave consistent results for (1) path coefficients estimates and for (2) evaluating model fit of individual path models by global fit indices (using *C* statistics in *piecewiseSEM*, and χ^2, TLI, CFI, RMSEA, and SRMR in *lavaan*). On the other hand, simulation results indicate that AIC-based model comparisons are less biased when performed by the piecewise structural equation modelling method, at least for comparisons between non-nested models.

References

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. 160:712–726.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press.

Kline, R. B. 2015. Principles and practice of structural equation modeling. Guilford.

Lefcheck, J. S. 2016. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol. Evol. 7:573–579.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

Preacher, K. J., and E. C. Merkle. 2012. The problem of model selection uncertainty in structural equation modeling. Psychol. Methods 17:1–14.

Rosseel, Y. 2012. Lavaan: An R package for structural equation modelling. J. Stat. Softw. 48:1–36.

Santos, J. C. 2012. Fast molecular evolution associated with high active metabolic rates in poison frogs. Mol. Biol. Evol. 29:2001–2018.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal components can be misleading. Syst. Biol. 64:677–689.

von Hardenberg, A., and A. Gonzalez-Voyer. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution 67:378–387.

West, S. G., A. B. Taylor, and W. Wu. 2012. Model fit and model selection in structural equation modeling. Pp. 209–231 *in* R. Hoyle, ed. Handbook of structural equation modeling. Guilford.