Supporting Information S1 for the paper:

Testing for adaptive radiation: A new approach applied to Madagascar frogs

by

Daniel S. Moen, Rojo N. Ravelojaona, Carl R. Hutter, and John J. Wiens

Supplementary Methods and Results

SAMPLING

We collected data on Malagasy frogs from the four families on the island: Mantellidae (25 species), Microhylidae (8 species), Ptychadenidae (1 species), and Hyperoliidae (2 species). We collected frogs near Andasibe, a site in the tropical wet forest zone in the Moramanga District in the Alaotra-Mangoro Region. The forests surrounding Andasibe are known for having exceptional frog species richness (Glaw and Vences 2007; Vieites et al. 2009; Brown et al. 2016). We collected frogs near their peak of activity at the beginning of the wet season (December 2017 to January 2018; Heinermann et al. 2015). Our sampling of the focal clade (Mantellidae) included all three subfamilies and eight of the 12 genera. We also focused on sampling species from different microhabitat categories (see microhabitat section below).

We note here that our sampling of one species is uncertain, but it does not affect our results. After we completed our fieldwork and analyses, Rancilhac et al. (2020) re-evaluated the taxonomic status of the closely related mantellid species *Mantidactylus grandidieri* and *M. guttulatus*. Our samples of these two species were from Andasibe

(performance and morphology) and Ranomafana (morphology only). Following from the descriptions of the ranges of these two species in Glaw and Vences (2007), we originally considered these to be *M. grandidieri* (Andasibe) and *M. guttulatus* (Ranomafana). However, the genetic data and range revisions of Rancilhac et al. (2020) suggest that our sample from Andasibe was also *M. guttulatus*, as only this species has been recorded where we sampled in Mitsinjo Forest Preserve (in Andasibe, adjacent to An'Ala Special Reserve). We have thus changed our designation of this species in our data to *M. guttulatus* for those individuals sampled for performance, but we note that this does not affect our main 80-species phenotypic analyses or our diversification analyses. Changing this species in our large dataset of morphology, however, would lead to us having two samples of one species (*M. guttulatus*) from two geographically disparate locations. If we lumped them, we would have one less species in our large-tree analyses and those analyses could potentially be affected. To address this possibility, we compared rates of size and shape evolution between Mantellidae and outgroup taxa after lumping individuals from Andasibe and Ranomafana into a single *M. guttulatus* (and thus reducing this tree from 217 species to 216). We found qualitatively identical results, with highly similar parameter estimates between the 216- and 217-species analyses (e.g., Mantellidae shape rate: 217-species σ^2 = 0.00404; 216-species σ^2 = 0.00396). Moreover, our finding that mantellid frogs have accelerated rates of shape evolution remained supported (P = 0.026 with 216 species versus P = 0.015 in Table 2d with 217 species). Thus, we retain our original results based on the 217-species tree due to the slight uncertainty about which species we actually sampled at Andasibe (e.g., both species were found at Moramanga and Vohidrazana, two localities

47 that flank Andasibe on the west and east, respectively). Furthermore, keeping these two
48 populations as separate units should be equivalent to sampling two species in this
49 genus of morphologically very similar species.
50 Our sampling of species across all four sites (one new site here; three from Moen
51 et al. [2013] in Australia, China, and Colombia) combined different aspects of anuran
52 evolutionary history. Within advanced frogs (Neobatrachia; containing ~95% of frog
53 species; AmphibiaWeb 2020), there are two major clades that include almost all frog
54 species (Hyloidea, Ranoidea). Most hyloid families occur in South America, and many
55 were represented in the sample from Colombia. In contrast, Asia is dominated by
56 Ranoidea. The Chinese taxa included two major ranoid families (Ranidae,
57 Rhacophoridae), one of which is sister to Mantellidae (Rhacophoridae; Feng et al. 2017;
58 Jetz and Pyron 2018). Australia includes a possible adaptive radiation of hylid frogs
59 (Pelodryadinae; Wiens et al. 2011, Moen et al. 2013), as does Madagascar
60 (Mantellidae; Wollenberg Valero et al. 2017).
61 As done by previous authors (Moen et al. 2013; Mendoza et al. 2020), we
62 collected mostly adult males to avoid potential differences between males and females
63 (e.g., sexual-size dimorphism is common in anurans; DeLisle and Rowe 2013; Nali et al.
64 2014). Males were sexed primarily based on calling activity, distended throat sacs, and
65 species-specific secondary-sexual characters (e.g., nuptial pads or spines; Glaw and
66 Vences 2007). In some cases, males and females were impossible to distinguish in live
67 frogs (87.5% of frogs in our dataset were males). However, we do not expect this to
68 affect our results, given the high phenotypic similarity between the sexes in these
69 cases.

To this performance and morphology dataset, we added an additional 137 species from Moen et al. (2016), for which only morphological data were available. Those data included 11 additional Malagasy species from different localities than where we sampled (primarily Ranomafana National Park).

PERFORMANCE DATA COLLECTION

All animals were tested for performance within 3–4 days of capture, and all testing was conducted over 7–10 days. We tested each animal for performance only once per day, and data collection for jumping and swimming performance was conducted on alternating days. Animals may perform differently at different times of the day (e.g., they may be more alert during their typical activity time, which was at night for nearly all species in this study). Thus, we collected performance data for each individual at least once in the morning (08h00–12h00), afternoon (13h00–17h00), and night (20h00–01h00). Regardless, maximum jumping efforts were evenly distributed across all time periods (*G*-test for heterogeneity across time periods; $P = 0.363$). Swimming showed fewer peak efforts in the morning ($P = 0.017$), with equal peak efforts in the afternoon and night. These results are consistent with previous studies that also showed time of day largely does not affect peak performance in anurans (Moen et al. 2013; Mendoza et al. 2020).

In Madagascar we tested animals at ambient temperatures (mean ± standard deviation = 25.2 ± 1.33 °C), which were their typical activity temperatures (i.e., those at which we collected the frogs) and were within the range of temperatures tested by Moen et al. (2013; 21.8–27.6 °C). We do not expect temperature to largely influence our

93 results. While anuran muscle physiology can be sensitive to temperature (Hirano and
94 Rome 1984; Marsh 1994; Navas et al. 1999), whole-organism jumping and swimming
95 performance is largely uniform within the range of temperatures we considered here
96 (Wilson 2001; Navas et al. 2008; Careau et al. 2014).

97 We recorded jumps in the lateral plane within a performance arena and swims
98 overhead within a plastic container filled with water. We recorded both behaviors with a
99 Fastec TS5 high-speed camera (Fastec Imaging, San Diego, CA, USA). We collected
100 jumping performance data at a frame rate of 500 Hz and 0.5 ms exposure time. Jumps
101 were elicited until three strong efforts were recorded or until the frog showed signs of
102 fatigue. Swimming performance trials followed jumping trials in nearly all aspects. We
103 placed frogs in a 62 x 35 cm plastic container with 11 cm deep water. We captured
104 videos overhead at 250 Hz and 2 ms exposure time. Frogs usually swam upon entering
105 the water, but occasionally we solicited swimming by lightly tapping on the frog's
106 posterior. Swimming behavior varied among individuals but was usually species-
107 specific. Only efforts where frogs swam parallel to the camera (on the surface or the
108 bottom) were retained for later analysis (see below). While swimming on the surface
109 can involve higher drag forces due to wake formation (Johansson and Lauder 2004;
110 Biewener and Patek 2018), most frogs swim faster than the wake can form and the
111 overall swimming mechanics are similar for surface-swimming and submerged frogs
112 (Richards 2010).

113 We used a non-stick (PTFE-coated) frying pan for adhesion trials, as such
114 surfaces have a similar coefficient of friction as waxy rainforest leaves (Emerson 1991).
115 We rotated the pan on a door hinge from 0 degrees (frog right-side up) up to 180

degrees (frog upside down). We then marked the angle at which each individual lost adhesion. Not all frogs in this study have adhesive pads on their fingers and toes, so our measure of adhesion integrates adhesion due to all surfaces (e.g., pressing the belly to the surface; Endlein et al. 2013). Such data represent an individual's ability to cling due to any morphological or behavioral trait, which may be the most ecologically relevant measure of adhesion (Endlein et al. 2013).

We initially screened jumping and swimming videos and excluded those from individuals that showed sub-optimal performance (e.g., much slower motion than in other videos of the same species; see Moen et al. 2013; Mendoza et al. 2020). This reduced our sample from 241 to 127 individuals but ensured that we analyzed data on peak performance (see below). Final sample size per species averaged 3.7 and ranged from 1–6 individuals (see Supporting Information S2 for full performance dataset), depending on availability of animals in the field and collecting permit limits. Sample-size limits in our collecting permit were determined by justification based on previous work and a desire to limit collection as much as possible based on conservation concerns.

To estimate peak jumping and swimming performance, we first digitized the tip of the snout during the take-off (jumping) or power stroke (swimming) in ImageJ (Rasband 1997; Schneider et al. 2012). We then used custom R code to convert the digitized coordinates into one-dimensional displacement-by-time profiles. We smoothed these profiles to reduce digitization error using a quintic spline (Walker 1998) in the package *fda* v. 2.4.8 (Ramsay et al. 2009; Ramsay et al. 2020) in R (R Core Team 2020). We used an objective criterion (generalized cross-validation; GCV) to determine an initial smoothing parameter. We then checked the resulting smoothed profile's first and

second derivatives, representing velocity and acceleration profiles, respectively, to ensure proper smoothing. Simulations have shown quintic splines with GCV to be an optimal smoothing technique for movement data of fast animals (Walker 1998), and we often found that it gave reasonable results (see also Moen et al. 2013). However, at times the GCV-chosen smoothing parameter produced profiles that had biologically unreasonable shapes (e.g., velocity increasing after take-off; acceleration monotonically declining instead of peaking midway through take-off; Marsh 1994; Marsh and John-Alder 1994). In such cases we manually adjusted the smoothing parameter until we obtained biologically reasonable results, as in previous studies (Moen et al. 2013). In the last step of performance data processing, we calculated velocity profiles as the first derivative of the smoothed displacement-by-time profiles. We used the peak velocity across videos for each individual as raw data, then calculated species means.

We focus on peak values for many reasons. First, and most importantly, studies of performance across a wide diversity of organisms focus on peak performance to capture the aspect of performance most likely to be molded by past selection (Bennett and Huey 1990; Losos et al. 2002). For example, while much of an animal's daily life is conducted at lower levels of performance, the failure to escape a predator due to poor performance likely imposes strong selection (Irschick and Losos 1998). Our experience in the field is that frogs perform maximally when attempting to escape potential predators (i.e., human collectors). Studies on field escape performance in frogs are rare, with most supporting peak performance (Rand 1952; Hayes 1990; Heinen and Hammond 1997) and some not (Gomes et al. 2002). Second, an animal may perform submaximally when measured because of many factors (e.g., insufficient motivation,

fatigue). These factors can be temporary and are thus unlikely related to the evolution of performance in the measured behavior (Losos et al. 2002). Performance may also be reduced by variation among experimental replicates (e.g., when an individual's foot slips on the substrate, or when it jumps at a slight angle away from the camera). These cases suggest it would be problematic to use mean values. In our experience, all individuals show variation in performance over a week of trials, but their peak values tend to be consistent (i.e., most individuals show 2–3 efforts that are similar and better than most others). In other organisms, peak performance has been found to be repeatable across years in natural populations (Huey and Dunham 1987). Third, while maximum endurance has been measured in anurans (Herrel and Bonneaud 2012; Careau et al. 2014; Rebelo and Measey 2019), its relevance to the ecology of most anuran species is unclear. Most anurans are relatively sedentary, using sit-and-wait predation and avoiding predators by hiding and later escaping through single bounds (Heinen and Hammond 1997; Wells 2007; Bulbert et al. 2015). Moreover, studies on anuran jumping have shown that many species of frogs become quickly fatigued and are reluctant to continue jumping after as few as 3–4 jumps (Rand 1952; Zug 1978). Thus, burst performance is most relevant to their biology. This is consistent with our own fieldwork and labwork.

In this study we focus on jumping and swimming velocity. While we could have followed previous studies by also calculating acceleration and power profiles (Toro et al. 2003, 2004; Moen et al. 2013; Moen 2019), we focus on velocity here in order to equally balance representation of the three behaviors we analyze in our multivariate rate analyses below (i.e., a single performance variable for jumping, swimming, and

clinging). Preliminary analyses showed that including acceleration and power increased the overall rate of multivariate performance evolution (i.e., including more traits produces higher rates; Denton and Adams 2015). However, our statistical comparisons of clades were not affected by this decision (results not shown).

MORPHOLOGICAL DATA COLLECTION

We generally collected morphological data from the same individuals from which we collected performance data, but sample sizes for morphology were higher in many cases due to the exclusion of sub-optimally performing individuals from the performance dataset. All sample sizes per species are provided in Supporting Information S3. In three cases (*Heterixalus betsileo, Gephyromantis ventrimaculatus,* and *Mantella crocea*), we measured museum specimens that we size-matched to those from which we collected performance data. We did this because of Malagasy government restrictions from removing all specimens from the country. Due to their small size, these three species were difficult to measure within the country.

We first obtained data on linear measurements with electronic Mitutoyo calipers (precision: 0.01mm), including: snout-to-vent length (SVL; tip of the snout to the posterior tip of the pelvis), leg length (the sum of four leg segment measurements from the tip of the pelvis to the tip of the longest toe), arm length (the sum of three measurements from the insertion point of the arm into the body wall to the tip of the longest finger), head length (tip of the snout to the corner of the jaw), and head width (distance between the two corners of the jaw).

Second, we photographed the hands and feet of each specimen. We then used ImageJ (Rasband 1997; Schneider et al. 2012) to calculate the area of finger and toe tips (which are often expanded disks in arboreal species), webbing between the toes of the foot (which is often expanded in aquatic species), and the inner metatarsal tubercle (which is often expanded and keratinized in burrowing species; Emerson 1976; Moen et al. 2016) .

Third, we dissected the upper- and lower-leg muscles of a single leg and calculated their mass, as these leg muscles are key drivers of most frog locomotion (Calow and Alexander 1973; Marsh 1994; Moen et al. 2013; Astley 2016; Moen 2019). We used preserved muscle mass here because we were unable to accurately measure fresh muscle mass in the field, and fresh and preserved muscle mass give similar results in comparative studies of anuran locomotion (Mendoza et al. 2020). We removed the entire set of muscles around the femur (upper leg) and tibiofibula (lower leg), gently patted them dry to remove surface ethanol, and obtained their mass on a Sartorius Entris mass balance precise to 0.1 mg. We assumed leg symmetry and doubled the measured value to estimate total leg muscle mass (Moen 2019; Mendoza et al. 2020).

MICROHABITAT DATA

Microhabitat states are described in the Methods of the main text. For the species whose phenotypic data had been previously published (Moen et al. 2013, 2016), we used the microhabitat states of Moen and Wiens (2017). For the additional Malagasy species from our fieldwork, we assigned microhabitat states using this same source

(based largely on Glaw and Vences [2007]). We also used species descriptions to assign microhabitat states to more recently described species (Vences et al. 2010; Glaw et al. 2015). In all cases, the states we assigned were consistent with our experience with these species in the field. Microhabitat data are given in Supporting Information S3.

We found that our sample of mantellid frogs did not include all eight initial states. This precluded comparing potential differences between mantellids and non-mantellids in ecomorphology (see below). Moreover, Moen (2019) showed that many frog ecomorphs are similar in jumping performance, and subsets are similar in swimming performance. Thus, to increase statistical power to detect potential differences between mantellids and other frogs (see below), we lumped three pairs of states that were similar, arbitrarily choosing the state with more species as the combined state: aquatic with semiaquatic, semiarboreal with arboreal, and semiburrowing with burrowing.

PHYLOGENY

We primarily based our phylogenetic comparative analyses on the posterior distribution of trees from Jetz and Pyron (2018). We also conducted some diversification-rate analyses using an alternative tree (Feng et al. 2017), but this tree had insufficient sampling at the species level to use for estimating phenotypic rates (or diversification rates with a species-level phylogeny). Other recent large-scale phylogenies (e.g., Hime et al. 2021) share this same limitation. We discuss analyses using the tree of Feng et al. (2017) in the section on diversification rates.

We first pruned the posterior distribution of Jetz and Pyron (2018) to include only the 3449 species of anurans for which they had genetic data, using tools on the VertLife webpage (www.vertlife.org/phylosubsets) to download 1000 trees. We then calculated the maximum-clade credibility (MCC) topology and summarized branch lengths in TreeAnnotator (Bouckaert et al. 2019). We used the option "Common Ancestor heights," which generally produces more accurate clade ages than mean branch lengths (Heled and Bouckaert 2013). For diversification analyses, we used this full tree. For phenotypic analyses, we first pruned the full tree to the 217 species for which we had morphological data (Supporting Information S4). We then further pruned the 217-species tree for analyses of 80 species (Fig. 3; Supporting Information S5). We note that all nodes in the 80-species tree and all but three nodes in the 217-species tree had posterior probabilities greater than 0.50, making these MCC trees very similar to a majority-rule consensus of the posterior distribution of trees.

Two measured species were not in the phylogeny. This was due to taxonomic ambiguity at the time of our measurements or in the tree of Jetz and Pyron (2018). In these cases, we made simple substitutions based on the taxonomic history in Frost (2019). We considered *Discoglossus jeanneae* (our data) as synonymous with *Discoglossus galganoi* (phylogeny). Moreover, we used the branch for *Stumpffia grandis* (phylogeny) for *Stumpffia kibomena* (our data). The former species is the one most closely related to *S. kibomena* (Glaw et al. 2015; Rakotoarison et al. 2017) among those in the phylogeny of Jetz and Pyron (2018).

SIZE STANDARDIZATION

Given that all morphological characters used here increase with body size, we wanted to reduce size-related redundancy, since ecomorphs do not differ in overall body size (Moen et al. 2013, 2016). Such standardization leads to a measure of body shape. We also size-standardized performance data to ensure that rates of evolution were analogous to those in morphology.

For size standardization we conducted phylogenetic regressions of each variable on a measure of body size (Revell 2009). We then used the residuals of these regressions as data for comparative analyses. We regressed most variables on SVL. However, we regressed preserved leg muscle mass on preserved body mass, given that these variables have the same dimensions. In all cases, we $\log_e$-transformed variables prior to regressions, given our desire to model evolutionary change in terms of proportions (O'Meara et al. 2006; Pélabon et al. 2014). We used Brownian motion as a model of covariance in our regressions because we later used this model for calculating rates of evolution (see below). We used the function 'phyl.resid' in the R package *phytools* v.0.6.99 (Revell 2012) for these size regressions. The resulting data used for comparative analyses are available as Supporting Information S6.

We recognize that many methods have been proposed for size standardization. Each has advantages and disadvantages, and the decision about which method to use can be controversial (Packard and Boardman 1988; García-Berthou 2001; Price et al. 2019). Ultimately, all such methods are quantifying the same underlying shape and should give relatively similar results (Klingenberg 2016). Previously, Moen (2019) found in a subset of the current taxa (191 species) and traits (leg length, leg muscle mass) that anuran morphology scales isometrically, regardless of analysis method (i.e.,

298 standard or phylogenetic regression). Thus, while using regression residuals could
299 eliminate allometric shape variation (Klingenberg 2016; Price et al. 2019), they do not
300 here, because such variation does not occur (at least in two key variables). Moreover,
301 results of downstream macroevolutionary analyses of these taxa are nearly
302 quantitatively identical using ratios and residuals (Moen 2019). Thus, we do not expect
303 our choice of standardization to affect our results.
304
305 **TESTING THE FIT OF BODY SHAPE AND PERFORMANCE TO ECOLOGY**
306 No studies have tested the fit of morphology and performance to ecology in Mantellidae,
307 our focal clade. Thus, we wanted to explicitly test whether mantellid frogs conformed to
308 ecomorphological relationships found in other frogs (Gomes et al. 2009; Moen et al.
309 2013, 2016; Moen 2019). Moreover, a previous study showed strong covariation
310 between morphology and performance (Moen et al. 2013), but such analyses have not
311 been done in Mantellidae, our candidate adaptive radiation.
312 In testing ecomorphology, we tested three factors: (1) microhabitat; (2) a factor of
313 Mantellidae versus other frogs (i.e., an overall difference in phenotype between these
314 groups); and (3) an interaction of microhabitat and Mantellidae. This last factor tested
315 whether the effect of being a mantellid on the phenotype differs among microhabitats.
316 We conducted these analyses on the 80-species dataset so that results for both
317 morphology and performance were comparable (i.e., derived from the same set of taxa).
318 Because not all eight microhabitats were represented in our sample of Mantellidae
319 (Supporting Information S3), we could not estimate an interaction term as we initially
320 assigned the eight microhabitats (Sokal and Rohlf 1995). Thus, for these analyses we

merged aquatic, semiarboreal, and semiburrowing with semiaquatic, arboreal, and burrowing, respectively.

The method of Adams (2014a) uses Brownian Motion (BM) as the model of evolutionary covariance among species. This model has been criticized as overly simplistic (Butler and King 2004; Hansen 2014). We agree that Ornstein-Uhlenbeck (OU) models are better models of macroevolutionary adaptation (Hansen and Orzack 2005; Hansen 2014) and OU models have been used in our previous analyses of similar data (Moen et al. 2016; Moen 2019). However, current implementations of these models have high error rates when analyzing many traits (Adams and Collyer 2018, 2019). Thus, we use the method of Adams (2014a) to overcome current limitations of multivariate OU methods. Given similar ecomorphological results here as in other anuran papers (Moen et al. 2013, 2016; Moen 2019), we do not expect this difference in methodology to have much consequence on our key results. Moreover, in the next section we test explicitly the fit of BM to our data and find that it largely fits the data well.

TESTING BROWNIAN MOTION AS A MODEL OF PHENOTYPIC EVOLUTION

When comparing rates of evolution between a focal clade and its outgroup, we estimated BM rates of evolution (Adams 2014b) because non-Brownian models of evolutionary change cannot be modeled accurately with multivariate data (Adams and Collyer 2018, 2019), although recent approaches are promising (Clavel et al. 2019). Moreover, we wanted to compare rates that do not assume a particular evolutionary process: BM fits a broad diversity of evolutionary scenarios (including some models of

adaptation; Felsenstein 1988; Revell et al. 2008) and can be considered a generalized model of evolutionary change (O'Meara et al. 2006).

We did two sets of analyses to ensure our results were not compromised by assuming BM. First, we used λ to evaluate whether and to what extent our data deviated from BM (Pagel 1999a; Freckleton et al. 2002). This parameter scales internal branch lengths to best describe the relationship between phenotypic covariation among taxa and their phylogeny, with $\lambda = 1$ equivalent to BM. Thus it can be used to model deviations from BM that are agnostic about macroevolutionary process (Freckleton et al. 2002). We calculated the maximum-likelihood estimate (MLE) of λ and a support region of values within 2 log-likelihood units of the MLE (Edwards 1972; Pagel 1999b). We did this for all traits (13 total: SVL, nine size-standardized morphological variables, and three size-standardized performance variables). We conducted these analyses for all six trees that we later analyzed when comparing multivariate rates of evolution: (i) the 80-species tree for performance data, (ii) the 25-species sample of Mantellidae within this 80-species tree, (iii) the 55-species outgroup of the 80-species tree, (iv) the full 217-species morphology-only tree, (v) the 36-species sample of Mantellidae in the 217-species tree, and (vi) the 181-species outgroup of the 217-species tree. We conducted these analyses in *phytools* (Revell 2012). We found that across all traits and all trees, λ was generally very close to 1.0, and nearly all support regions included 1.0 (Table S8). These results suggest that BM was an appropriate model of evolution for our data.

Second, we also considered other models of evolution. In particular, a pattern of early-burst evolution (higher rates of evolution early in a clade's history) has been considered a hallmark of adaptive radiation (Slater and Friscia 2019). This pattern has

366 largely been unsupported in many clades considered to be adaptive radiations (Harmon
367 et al. 2010; Pennell et al. 2015; Slater 2015; but see Slater and Friscia 2019).
368 Nonetheless, given our goal of detecting adaptive radiation and the role of the early-
369 burst model in some tests of it, we applied the model-comparison framework of Harmon
370 et al. (2010) to our trees and traits. We fit BM, a single-optimum OU model, an early-
371 burst model, and a model that estimates λ. We did this for each combination of trait and
372 tree, as above for our estimates of λ alone. We used the function "fitContinuous" in the
373 R package *geiger* v.2.0.7 (Harmon et al. 2008; Pennell et al. 2014) to estimate the
374 likelihood and AICc of each model for each trait-tree combination. We then calculated
375 the AICc weights of the four models within each trait-tree combination.
376 Most analyses returned similar support for the different models (Table S9). No
377 model consistently had the highest weight, and the best-supported model within a trait-
378 tree combination rarely had a weight so high as to strongly exclude all other models. Of
379 these four models, the early-burst model was clearly the worst: across all 69
380 comparisons, it was the most-supported model only once and it had low weight across
381 all datasets. Overall, few analyses strongly supported BM as a model of evolution, but
382 neither did they strongly support other models. As BM is the most general model
383 consistent with multiple evolutionary scenarios (Felsenstein 1988; Revell et al. 2008),
384 we therefore consider it reasonable to assume BM for our analyses. Furthermore, as
385 noted above, most traits showed values of λ very close to 1.0, consistent with a BM
386 model.
387
388 **ACCOUNTING FOR PHYLOGENETIC UNCERTAINTY IN RATE TESTS**

We accounted for phylogenetic uncertainty in all our multivariate rate comparisons by conducting analyses on a set of 1000 trees from the posterior distribution of Jetz and Pyron (2018). On each tree, we estimated rates in the two-rate model (e.g., Mantellidae vs. all others), then calculated the ratio of the focal rate (e.g., Mantellidae) to non-focal rate (e.g., other species). If statistically significant results on the consensus tree are robust to uncertainty in phylogeny, most of these rate ratios across the posterior should exceed 1.0 (i.e., indicating that Mantellidae has a higher rate than other frogs across most trees in the posterior). As an alternative measure of phylogenetic uncertainty, we also did a full hypothesis test on each tree (i.e., calculated a *P*-value via parametric simulation). We considered a statistically significant result robust to phylogenetic uncertainty if over 95% of these *P*-values were less than 0.05.

SIMULATIONS TO TEST EFFECTS OF TREE SIZE ON STATISTICAL POWER

We found that some of our statistical results were much stronger (i.e., significant or very close to significant) when conducting analyses on the 217-species tree than on the 80-species tree (see Results). Thus, we wanted to assess the statistical power of the analyses on trees that differed in size by almost a factor of three. To do so, we first compared Pelodryadinae to all other frogs, given that our sample of this clade was smaller (11 species) than our sample of Mantellidae (25 species). Significantly higher rates in Pelodryadinae would suggest that our sample size for Mantellidae was not too small to show significantly elevated rates.

Second, we conducted power simulations to assess how different sampling (i.e., 80- vs. 217-species trees) could affect our tests at the observed rate ratios. In other

words, if we found that Mantellidae showed a rate of body-shape evolution 1.25 times higher than other frogs, what would be the statistical power of testing this rate difference in the 80-species tree versus the 217-species tree? Because the rates of the different groups (focal and other) could differ in the different datasets and cause the differing results, these power simulations avoided confounding observed rate differences in the two trees with differences in results based on taxon sampling (i.e., power). Thus these power simulations dissected the importance of rate differences versus differences in tree size alone.

We conducted these simulations following Adams (2014b) and Adams and Collyer (2018). We set the BM rate of evolution of other frogs at $\sigma^2_{Other} = 1.0$ for each trait dimension (see below), then considered rate differences in mantellids from a rate (and thus ratio) of $\sigma^2_{Mant} = 1.1–2.5$, considered on an interval of 0.1 (i.e., 15 total ratios). We also considered the four observed rate ratios in body-shape evolution in both the 80- and 217-species trees, in comparing Mantellidae to other frogs, as well as Mantellidae versus other frogs without Pelodryadinae in the analysis (i.e., Table 2). We calculated power for these four observed rates for body shape because they were the key results that differed between tree sizes.

Our empirical tests differed in their dimensions: body size was a single variable, body shape had nine variables, and performance had three. Thus, we conducted simulations for one, three, and nine variables. Moreover, the correlation among variables differed. Using the method of Revell and Collar (2009), we found that the average BM correlation was 0.364 and 0.124 among performance and body-shape variables, respectively. Thus, we specified variable correlation as 0.364 in the

435 simulations of three variables and 0.124 in simulations of nine variables. These
436 conditions best replicated our empirical conditions to optimally estimate the statistical
437 power of our tests, given that power is sensitive to variable correlation (Adams and
438 Collyer 2018). We also note that the approach we used assumes isotropic error in
439 specifying error covariances. We did not consider non-isotropic error, given that power
440 is very similar under both isotropic and non-isotropic error (Adams 2014b).
441 To simulate differing rates of trait evolution for mantellids versus other frogs, we
442 first used the function 'transformPhylo' in the R package *motmot* v. 2.1.3 (Thomas and
443 Freckleton 2012) to stretch Mantellidae's branch lengths by the rate difference (i.e., 1.1–
444 2.5 times longer, depending on the simulation). We stretched trees because simulating
445 homogeneous character evolution on stretched branches is equivalent to directly
446 simulating different rates on unstretched branches (O'Meara 2012). We then specified
447 simulation rate matrices based on the number of trait dimensions and trait correlations
448 indicated above with a per-character rate of 1.0. We inputted these rate matrices and
449 stretched trees to the function 'sim.char' in *geiger* v. 2.0.6.1 (Harmon et al. 2008;
450 Pennell et al. 2014) to simulate 1000 datasets per simulation condition. Finally, we
451 statistically compared rates in Mantellidae to other frogs on the original trees. We did
452 this final step with 'compare.evol.rates' in *geomorph* v. 3.1.3 (Adams and Otárola-
453 Castillo 2013), as in our empirical rate analyses. For each simulation condition, the
454 proportion of those replicates that had *P*-values less than 0.05 was our estimate of
455 power of that test (i.e., combination of rate difference and tree size/sampling). In total,
456 we explored 114 simulation conditions (two trees, three trait dimensions, and 19 rate
457 ratios). In Supporting Information S7 we provide an R function that allows users to

easily replicate these phenotypic analyses, including power simulations and predicting rate differences between groups.

Power at the body-shape rate ratio of 1.171 (i.e., our observed ratio of the mantellid rate to the other-frog rate for the smaller tree without Pelodryadinae; Table 2) was 0.246 in the 80-species tree versus 0.405 in the 217-species tree. In other words, if the true rate ratio were 1.171 during the evolution of these groups and this evolutionary process were repeated multiple times, the resulting rate test would be significant only ~25% of the time in an 80-species comparison, whereas ~40% of the time it would be significant in a 217-species comparison. Power at the ratio of 1.294 (observed in the larger tree without Pelodryadinae; Table 2) was 0.523 and 0.774 in the 80- and 217-species trees, respectively. Thus, increasing the rate ratio increased power more than increasing tree size (Fig. S1): increasing the rate ratio by 25% (1.171 to 1.294) roughly doubled power at both tree sizes, but increasing tree size for a given rate ratio only increased power by a factor of 1.5. These results suggest that the different *P*-values for mantellid shape evolution at the two tree sizes was more a result of rate differences in these trees than tree size per se.

TESTING EFFECTS OF SAMPLING DENSITY ON RATE COMPARISONS

We recognize that our sampling of Mantellidae and the outgroup for rate comparisons was quite low when considered as a proportion of total anuran species richness. In our 80-species tree, the 25 species of Mantellidae were 13.2% of those found in the 3449-species tree of Jetz and Pyron (2018), and the 55 species in the outgroup represented just 1.7% of potential outgroup taxa in the full tree. In the 217-species tree, these

numbers were somewhat higher (36 mantellids and 181 outgroup species, or 19.0% and 5.6%, respectively). Thus, we considered the possibility that sampling so few species could mislead our comparisons of rates of phenotypic evolution between Mantellidae and other anurans. We see two potential concerns with such sampling.

First, microhabitat-based ecomorphs seem to drive the evolution of morphology and performance in frogs (see Results), so they may also drive rates of morphological and performance evolution. Therefore, biased sampling of microhabitat types among species could strongly bias phenotypic rate estimates (e.g., mantellids would be almost guaranteed to have relatively low phenotypic rates if we only sampled arboreal species). We therefore estimated and compared microhabitat proportions (the number of sampled species with each microhabitat state divided by the total number of species sampled) for three datasets: our 80-species tree, our 217-species tree, and a much larger sample of anurans. For latter, we used data from Moen and Wiens (2017), who compiled microhabitat data for 3394 species in 53 of 54 families of anurans, the largest sample to date. We hereafter refer to this as “all anurans” for brevity. Using these data, we counted the frequency of microhabitat states in both Mantellidae and all other frogs (the outgroup) for each of the three cases (i.e., 80 species, 217 species, and all anurans). We then used chi-square tests (Sokal and Rohlf 1995) to test whether proportions of microhabitat states in our samples differed from proportions in the much larger sample of Moen and Wiens (2017). In the main manuscript we show how our sampling of mantellid microhabitats is consistent across all three samples of the family. Here, we address the sampling of other frogs.

503 We tabulated the numbers of species in each microhabitat category for the
504 outgroup for all three samples (Supporting Information S3; Moen and Wiens 2017). We
505 then conducted chi-square tests of proportions between all anurans and our samples,
506 using the base R function “chisq.test”. We calculated *P*-values by Monte Carlo
507 simulation. A *P*-value less than 0.05 in this case would indicate statistically significant
508 differences in the proportions of microhabitat states between our sample and across all
509 anurans. We found that for both the 80- and 217-species trees, our outgroup sampling
510 significantly deviated from the proportions of states across all anurans (80 species: P =
511 0.017; 217 species: P = 0.002). The key deviation between our samples and those of all
512 anurans is that we had no semiarboreal species in our outgroup, whereas Moen and
513 Wiens (2017) found that 8.2% of their sampled anurans fit this state. Subsequent chi-
514 square tests without this category show this to be the case: when comparing only the
515 remaining seven microhabitat categories found in all samples, proportions were much
516 more similar (80 species: P = 0.367; 217 species: P = 0.069). This omission in our
517 outgroup samples is particularly relevant because it is a state frequently found in
518 Mantellidae (Supporting Information S3). All things equal, the fact that Mantellidae has
519 this state that is absent in our outgroup makes it more likely to find significantly elevated
520 rates of phenotypic evolution in Mantellidae (i.e., more ecomorph states are likely to
521 lead to higher rates of phenotypic evolution). Given that we did not find elevated rates in
522 Mantellidae across most analyses, we consider our results robust to potential problems
523 associated with biased microhabitat sampling.
524 A second potential consequence of taxon sampling on comparing rates of
525 evolution is that proportionally low sampling itself could influence estimates of rates,

526 reduce power, or increase Type I error rates. We thus simulated trait evolution along the
527 full, 3449-species tree, subsampled species to match our empirical sample sizes, and
528 examined the statistical properties of the resulting rate comparisons between our focal
529 group (Mantellidae) and an outgroup. We simulated trait evolution for one, three, and
530 nine traits, which matched our empirical tests of body-size evolution, performance
531 evolution, and morphological evolution, respectively. In terms of trait correlations, our
532 mean empirical PGLS correlations on the 80-species tree were 0.110 and 0.364 for
533 morphology and performance, respectively. For these simulations we wanted to bracket
534 these empirical values so as to examine the full effects of trait correlation possible for
535 our dataset. Thus, we considered both no trait correlation and a correlation of 0.5 in
536 simulations of multiple traits. As in previous similar simulations (Adams 2014b; Adams
537 and Collyer 2018), we scaled the full tree to a length of 1.0 and simulated Brownian-
538 motion trait evolution at a rate of 1.0 for the outgroup and 1.0, 1.25, 1.5, 1.75, 2.0, 2.25,
539 and 2.5 for Mantellidae. We simulated trait evolution 2000 times for each of our 35
540 conditions (i.e., the seven mantellid rates for one trait, three and nine traits with no
541 correlation, and three and nine traits with correlation of 0.5), as described above for
542 power simulations. We then randomly subsampled 25 mantellids and 55 outgroup taxa
543 for 1000 trees of each condition, as well as 36 mantellids and 181 outgroup taxa for the
544 other 1000 trees. This totaled 70 distinct simulation scenarios. We then assessed
545 statistical performance at each simulation scenario. We examined potential bias by
546 calculating mean and standard deviation of rate estimates for mantellids and the
547 outgroup. We also calculated Type I error as the proportion of significant rate
548 differences at α = 0.05 when rates were the same for Mantellidae and the outgroup. We

549 assessed power as the proportion of statistically significant results when mantellids
550 were simulated as having a higher rate of evolution.
551 We found that parameter estimation was unbiased, with mean estimates in all
552 conditions fitting the simulated rates closely (Fig. S2). Variability in parameter estimates
553 increased as trait numbers decreased and as trait correlations increased, as expected
554 based on power estimates from previous simulations (Fig. S1; Adams 2014b; Adams
555 and Collyer 2018). We found that Type I errors were not elevated in any condition (Fig.
556 S3 at rate ratio = 1.0), nor was power affected by subsampling. For example, these
557 curves closely match those of Fig. S1, in which trait evolution was simulated only on the
558 already-pruned trees used in empirical analyses (note that the nine-trait curves in Fig.
559 S1 were simulated with a trait correlation of 0.110, closest to “no correlation” here).
560 Thus, these subsampling simulations showed that comparison of multivariate rates of
561 evolution is unbiased, accurate, and powerful even under extremely low sampling of the
562 phylogeny along which the traits evolved. We conclude that our empirical rate tests
563 herein are most likely uncompromised by our limited sampling of Mantellidae and other
564 anurans. More generally, comparison of rates of phenotypic evolution may be accurate
565 for traits whose data-collection intensity is so high that species sampling must be
566 reduced, as in biomechanics, performance, and physiology.
567
568 **DIVERSIFICATION RATES**
569 Some studies have suggested that Mantellidae is an adaptive radiation because its net
570 diversification rate slows through time (Wollenberg Valero et al. 2017). Yet mantellid net
571 diversification rates are intermediate among frog families (Moen and Wiens 2017). On

572 the other hand, Madagascar's amphibian diversity is considered to be underestimated
573 (Vieites et al. 2009; Perl et al. 2014), so these previous rates may be underestimates.
574 Therefore, we estimated net diversification rates to statistically compare rates across
575 anuran families and to account for undescribed species diversity.

576 To calculate net diversification rates, we primarily used the method-of-moments
577 estimator of Magallón and Sanderson (2001), calculated with their equations coded
578 directly in R. This approach requires only the species richness and age of each clade
579 (i.e., not requiring extensive sampling within clades, as required by most other tests;
580 Morlon 2014). This method is relatively accurate, showing strong correlations between
581 true and estimated rates in simulations (Kozak and Wiens 2016; Meyer et al. 2018;
582 Meyer and Wiens 2018). Moreover, its accuracy is not strongly impacted by faster rates
583 in younger clades, changing rates within clades over time, or heterogeneity in rates
584 between subclades (Kozak and Wiens 2016; Meyer et al. 2018; Meyer and Wiens
585 2018). Therefore, it does not require constant rates within or between clades.
586 Furthermore, it does not require separately estimating speciation and extinction rates,
587 and the stem-age estimator is not impacted by incomplete taxon sampling within clades.

588 For thoroughness, we also calculated net diversification rates using the time-
589 constant birth-death estimators of Nee et al. (1994). These estimators use branch-
590 length distributions to estimate speciation and extinction rates. We calculated both pure-
591 birth and birth-death estimates of net diversification rate, with the latter simply the
592 estimated speciation rate minus the estimated extinction rate for each clade. Given that
593 these estimates do not use the stem branches for the families, we note that these net
594 diversification rates are more directly analogous to the crown estimates from our other

595 analyses. Because these estimates use branch lengths, we limited analyses to families
596 for which Jetz and Pyron (2018) had genetic data for 5 or more species. The minimum
597 number of taxa to obtain reliable estimates with these methods is unclear, so this
598 threshold is admittedly arbitrary. It is also low. Yet including some smaller families
599 should make it easier to find a statistically significant rate increase in Mantellidae:
600 anuran families with few species tend to have the lowest diversification rates (Moen and
601 Wiens 2017), so including small families decreases the mean rate among the outgroup
602 families. More importantly, our comparison of net diversification rates in Mantellidae
603 versus other families was invariant to the minimum size of family included in these
604 analyses (results not shown). We estimated these diversification rates in R with the
605 package *diversitree* v.09.14 (FitzJohn 2012). We accommodated incomplete sampling
606 of phylogenies by calculating, for each family, the proportion of species in the phylogeny
607 relative to the total species diversity of the family (AmphibiaWeb 2020).
608 Importantly, as we apply them here, both approaches (Nee et al. 1994; Magallón
609 and Sanderson 2001) allow for a different rate of diversification in each anuran family.
610 Thus, they assume at least 38 different rates across the anuran tree, depending on the
611 approach for calculating rates (Table S6). In contrast, simulations show that Bayesian
612 Analysis of Macroevolutionary Mixtures (BAMM; Rabosky 2014) grossly underestimates
613 rate variation across large trees. Thus, in simulations with 10 or more rates simulated
614 across a tree of thousands of species, BAMM will typically infer that only 2–3 distinct
615 rates are present, assigning many clades with distinct rates as having identical rates
616 (Meyer and Wiens 2018). This result is also consistent with simulations in the original
617 paper in which BAMM was proposed, which showed that BAMM increasingly

underestimated rate variation as rate variation increased (Rabosky 2014). Thus, arguing that BAMM should be used to avoid assuming constant rates is not justified. Similarly, state-dependent speciation-extinction (SSE) methods (e.g., BiSSE, HiSSE; Beaulieu and O'Meara 2016) assign all species in a tree to only a handful of different diversification rates, typically including two observed states (and two hidden states for HiSSE). Thus, these methods also assume a limited number of different diversification rates across large trees, in contrast to the methods that we used here.

We acknowledge that there are many other diversification rate estimators available (e.g., Morlon 2014), and our goal here is not to provide an exhaustive review of all of their pros and cons. We simply emphasize that the methods that we used are relatively accurate and are not particularly sensitive to the limitation (i.e., rate constancy) that is often ascribed to them. Moreover, current methods that attempt to extract much more information from trees—particularly those that estimate separate speciation and extinction rates that vary through time—may be highly misleading (Louca and Pennell 2020). Our approach here integrates over variation in diversification histories over time and focuses on the net rate of diversification, thus reducing the probability of incorrect inference.

In addition to data from AmphibiaWeb (2020), we used a recent estimate of maximum richness of Mantellidae that accounts for undescribed species. Perl et al. (2014) estimated the total number of mantellid species to be 409 (i.e., both described and undescribed candidate species), so we also used this highest recent estimate of the family's diversity to calculate its diversification rate. However, many other anuran families have described richness that underestimates their actual species diversity. For

example, in the last five years, nine anuran families (of 54) had more species described per year than Mantellidae (AmphibiaWeb 2020). Thus, we primarily focus on results from the current species diversity of Mantellidae, to avoid biasing the results by only considering undescribed richness of mantellids and not other families.

In addition to the stem ages of Jetz and Pyron (2018), we also tested crown ages from the same tree, as well as the ages of the families (44 of 54) sampled by Feng et al. (2017). Sampling in the latter tree was too limited to calculate crown ages of most clades, so we only included their stem ages. For each of these trees and age estimates, we assumed one of three relative extinction fractions (ε = 0.0, 0.5, 0.9). In total we conducted 18 comparisons (two estimates of mantellid richness, three age estimates, three extinction fractions) with the method-of-moments estimates and two additional comparisons with the birth-death estimates. Nonetheless, all approaches gave similar results (Table S6). Therefore, our main results focused on the most comprehensive tree (Jetz and Pyron 2018) and accurate data (stem ages; Sanderson 1996). We provide all diversification-rate estimates in Supporting Information S9.

LITERATURE CITED

Adams, D. C. 2014a. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

———. 2014b. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. Syst. Biol. 63:166–177.

Adams, D. C. and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst. Biol. 67:14–31.

———. 2019. Phylogenetic comparative methods and the evolution of multivariate phenotypes. Annu. Rev. Ecol. Evol. Syst. 50:405–425.

Adams, D. C. and E. Otárola-Castillo. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. 4:393–399.

AmphibiaWeb. 2020. AmphibiaWeb: Information on amphibian biology and conservation. Accessed 15 January 2020. http://amphibiaweb.org.

Astley, H. C. 2016. The diversity and evolution of locomotor muscle properties in anurans. J. Exp. Biol. 219:3163–3173.

Beaulieu, J. M. and B. C. O'Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Bennett, A. F. and R. B. Huey. 1990. Studying the evolution of physiological performance. Pp. 251–284 *in* D. J. Futuyma, and J. Antonovics, eds. Oxford surveys in evolutionary biology. Oxford University Press, Oxford, UK.

Biewener, A. A. and S. Patek. 2018. Animal Locomotion. Oxford University Press, Oxford, UK.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchene, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kuhnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Muller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C. H. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 15:e1006650.

686 Brown, J. L., N. Sillero, F. Glaw, P. Bora, D. R. Vieites, and M. Vences. 2016. Spatial
687 biodiversity patterns of Madagascar's amphibians and reptiles. PLoS One
688 11:e0144076.
689 Bulbert, M. W., R. A. Page, and X. E. Bernal. 2015. Danger comes from all fronts:
690 predator-dependent escape tactics of túngara frogs. PLoS One 10:e0120546.
691 Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling
692 approach for adaptive evolution. Am. Nat. 164:683–695.
693 Calow, L. J. and R. M. Alexander. 1973. A mechanical analysis of a hind leg of a frog
694 (*Rana temporaria*). J. Zool. 171:293–321.
695 Careau, V., P. A. Biro, C. Bonneaud, E. B. Fokam, and A. Herrel. 2014. Individual
696 variation in thermal performance curves: swimming burst speed and jumping
697 endurance in wild-caught tropical clawed frogs. Oecologia 175:471–480.
698 Clavel, J., L. Aristide, and H. Morlon. 2019. A penalized likelihood framework for high-
699 dimensional phylogenetic comparative methods and an application to New-World
700 monkeys brain evolution. Syst. Biol. 68:93–116.
701 De Lisle, S. P. and L. Rowe. 2013. Correlated evolution of allometry and sexual
702 dimorphism across higher taxa. Am. Nat. 182:630–639.
703 Denton, J. S. and D. C. Adams. 2015. A new phylogenetic test for comparing multiple
704 high-dimensional evolutionary rates suggests interplay of evolutionary rates and
705 modularity in lanternfishes (Myctophiformes; Myctophidae). Evolution 69:2425–
706 2440.
707 Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge, UK.
708 Emerson, S. B. 1976. Burrowing in frogs. J. Morph. 149:437–458.

———. 1991. The ecomorphology of Bornean tree frogs (family Rhacophoridae). Zool. J. Linn. Soc. 101:337–357.

Endlein, T., W. J. Barnes, D. S. Samuel, N. A. Crawford, A. B. Biaw, and U. Grafe. 2013. Sticking under wet conditions: the remarkable attachment abilities of the torrent frog, *Staurois guttatus*. PLoS One 8:e73810.

Felsenstein, J. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.

Feng, Y. J., D. C. Blackburn, D. Liang, D. M. Hillis, D. B. Wake, D. C. Cannatella, and P. Zhang. 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. Proc. Natl. Acad. Sci. USA 114:E5864–5870.

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

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.

Frost, D. R. 2019. Amphibian species of the world: an online reference. American Museum of Natural History, New York, NY. Accessed 15 January 2020. https://amphibiansoftheworld.amnh.org.

García-Berthou, E. 2001. On the misuse of residuals in ecology: testing regression residuals vs. the analysis of covariance. J. Anim. Ecol. 70:708–711.

Glaw, F., D. Vallan, F. Andreone, D. Edmonds, R. Dolch, and M. Vences. 2015. Beautiful bright belly: A distinctive new microhylid frog (Amphibia: *Stumpffia*) from eastern Madagascar. Zootaxa 3925:120–128.

Glaw, F., and M. Vences. 2007. A field guide to the amphibians and reptiles of Madagascar. 3rd ed. Verlag G.B, Köln.

Gomes, F. R., C. R. Bevier, C. A. Navas, and R. E. Gatten Jr. 2002. Environmental and physiological factors influence antipredator behavior in *Scinax hiemalis* (Anura: Hylidae). Copeia 2002:994–1005.

Gomes, F. R., E. L. Rezende, M. B. Grizante, and C. A. Navas. 2009. The evolution of jumping performance in anurans: morphological correlates and ecological implications. J. Evol. Biol. 22:1088–1097.

Hansen, T. F. 2014. Use and misuse of comparative methods in the study of adaptation. Pp. 351–379 *in* L. Z. Garamszegi, ed. Modern phylogenetic comparative methods and their application in evolutionary biology. Springer-Verlag, Berlin.

Hansen, T. F. and S. H. Orzack. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59:2063–2072.

Harmon, L. J. 2018. Phylogenetic comparative methods: learning from trees. CreateSpace Independent Publishing Platform, Moscow, ID.

Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeek, F. Moreno-Roark, T. J. Near, A. Purvis, R. E. Ricklefs, D. Schluter, J. A. I. Schulte, O. Seehausen, B. L. Sidlauskas, O. Torres-Carvajal, J. T. Weir, and A. O. Mooers. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

755 Hayes, F. E. 1990. Comparative escape behaviour of adult green frogs (*Rana*
756 *clamitans*) and northern leopard frogs (*R. pipiens*). Bull. Maryland Herpetol. Soc.
757 26:81–99.
758 Heinen, J. T. and G. Hammond. 1997. Antipredator behaviors of newly metamorphosed
759 green frogs (*Rana clamitans*) and leopard frogs (*R. pipiens*) in encounters with
760 eastern garter snakes (*Thamnophis s. sirtalis*). Am. Midland Nat. 137:136–144.
761 Heinermann, J., A. Rodríguez, O. Segev, D. Edmonds, R. Dolch, and M. Vences. 2015.
762 Year-round activity patterns in a hyperdiverse community of rainforest
763 amphibians in Madagascar. J. Nat. Hist. 49:2213–2231.
764 Heled, J. and R. R. Bouckaert. 2013. Looking for trees in the forest: summary tree from
765 posterior samples. BMC Evol. Biol. 13:221.
766 Herrel, A. and C. Bonneaud. 2012. Trade-offs between burst performance and maximal
767 exertion capacity in a wild amphibian, *Xenopus tropicalis*. J. Exp. Biol. 215:3106–
768 3111.
769 Hirano, M. and L. C. Rome. 1984. Jumping performance of frogs (*Rana pipiens*) as a
770 function of muscle temperature. J. Exp. Biol. 108:429–439.
771 Huey, R. B. and A. E. Dunham. 1987. Repeatability of locomotor performance in natural
772 populations of the lizard *Sceloporus merriami*. Evolution 41:1116–1120.
773 Irschick, D. J. and J. B. Losos. 1998. A comparative analysis of the ecological
774 significance of maximal locomotor performance in Caribbean *Anolis* lizards.
775 Evolution 52:219–226.

776 Jetz, W. and R. A. Pyron. 2018. The interplay of past diversification and evolutionary
777 isolation with present imperilment across the amphibian tree of life. Nat. Ecol.
778 Evol. 2:850–858.
779 Johansson, L. C. and G. V. Lauder. 2004. Hydrodynamics of surface swimming in
780 leopard frogs (*Rana pipiens*). J. Exp. Biol. 207:3945–3958.
781 Jolliffe, I. T. and J. Cadima. 2016. Principal component analysis: a review and recent
782 developments. Philos. Trans. A Math. Phys. Eng. Sci. 374:20150202.
783 Klingenberg, C. P. 2016. Size, shape, and form: concepts of allometry in geometric
784 morphometrics. Dev. Genes Evol. 226:113–137.
785 Kozak, K. H. and J. J. Wiens. 2016. Testing the relationships between diversification,
786 species richness, and trait evolution. Syst. Biol. 65:975–988.
787 Losos, J. B., D. A. Creer, and J. A. Schulte II. 2002. Cautionary comments on the
788 measurement of maximum locomotor capabilities. J. Zool. 258:57–61.
789 Louca, S. and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of
790 diversification histories. Nature 580:502–505.
791 Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm
792 clades. Evolution 55:1762–1780.
793 Marsh, R. L. 1994. Jumping ability of anuran amphibians. Pp. 51–111 *in* J. H. Jones, ed.
794 Advances in veterinary science and comparative medicine. Academic Press,
795 New York, NY.
796 Marsh, R. L. and H. B. John-Alder. 1994. Jumping performance of hylid frogs measured
797 with high-speed cine film. J. Exp. Biol. 188:131–141.

798 Mendoza, E., E. Azizi, and D. S. Moen. 2020. What explains vast differences in jumping
799 power within a clade? Diversity, ecology and evolution of anuran jumping power.
800 Funct. Ecol. 34:1053–1063.
801 Meyer, A. L. S., C. Roman-Palacios, and J. J. Wiens. 2018. BAMM gives misleading
802 rate estimates in simulated and empirical datasets. Evolution 72:2257–2266.
803 Meyer, A. L. S. and J. J. Wiens. 2018. Estimating diversification rates for higher taxa:
804 BAMM can give problematic estimates of rates and rate shifts. Evolution 72:39–
805 53.
806 Moen, D. S. 2019. What determines the distinct morphology of species with a particular
807 ecology? The roles of many-to-one mapping and trade-offs in the evolution of
808 frog ecomorphology and performance. Am. Nat. 194:E81–95.
809 Moen, D. S. and J. J. Wiens. 2017. Microhabitat and climatic niche change explain
810 patterns of diversification among frog families. Am. Nat. 190:29–44.
811 Moen, D. S., D. J. Irschick, and J. J. Wiens. 2013. Evolutionary conservatism and
812 convergence both lead to striking similarity in ecology, morphology and
813 performance across continents in frogs. Proc. R. Soc. B Biol. Sci. 280:20132156.
814 Moen, D. S., H. Morlon, and J. J. Wiens. 2016. Testing convergence versus history:
815 convergence dominates phenotypic evolution for over 150 million years in frogs.
816 Syst. Biol. 65:146–160.
817 Morlon, H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett.
818 17:508–525.

819 Nali, R. C., K. R. Zamudio, C. F. Haddad, and C. P. Prado. 2014. Size-dependent
820 selective mechanisms on males and females and the evolution of sexual size
821 dimorphism in frogs. Am. Nat. 184:727–740.
822 Navas, C. A., F. R. Gomes, and J. E. Carvalho. 2008. Thermal relationship and exercise
823 physiology in anuran amphibians: integration and evolutionary implications.
824 Comp. Biochem. Physiol. 151A:344–362.
825 Navas, C. A., R. S. James, J. M. Wakeling, K. M. Kemp, and I. A. Johnston. 1999. An
826 integrative study of the temperature dependence of whole animal and muscle
827 performance during jumping and swimming in the the frog *Rana temporaria*. J.
828 Comp. Physiol. B 169:588–596.
829 Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process.
830 Phil. Trans. R. Soc. B Biol. Sci. 344:305–311.
831 O'Meara, B. C. 2012. Evolutionary inferences from phylogenies: a review of methods.
832 Annu. Rev. Ecol. Evol. Syst. 43:267–285.
833 O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for
834 different rates of continuous trait evolution using likelihood. Evolution 60:922–
835 933.
836 Packard, G. C. and T. J. Boardman. 1988. The misuse of ratios, indices, and
837 percentages in ecophysiological research. Physiol. Zool. 61:1–9.
838 Pagel, M. 1999a. Inferring the historical patterns of biological evolution. Nature
839 401:877–884.
840 ———. 1999b. The maximum likelihood approach to reconstructing ancestral character
841 states of discrete characters on phylogenies. Syst. Biol. 48:612–622.

842 Pélabon, C., C. Firmat, G. H. Bolstad, K. L. Voje, D. Houle, J. Cassara, A. L. Rouzic,
843 and T. F. Hansen. 2014. Evolution of morphological allometry. Ann. NY Acad.
844 Sci. 1320:58–75.
845 Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn,
846 M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods
847 for fitting macroevolutionary models to phylogenetic trees. Bioinformatics
848 30:2216–2218.
849 Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model
850 adequacy and the macroevolution of angiosperm functional traits. Am. Nat.
851 186:E33–50.
852 Perl, R. G. B., Z. T. Nagy, G. Sonet, F. Glaw, K. C. Wollenberg, and M. Vences. 2014.
853 DNA barcoding Madagascar's amphibian fauna. Amphibia-Reptilia 35:197–206.
854 Price, S. A., S. T. Friedman, K. A. Corn, C. M. Martinez, O. Larouche, and P. C.
855 Wainwright. 2019. Building a body shape morphospace of teleostean fishes.
856 Integr. Comp. Biol. 59:716–730.
857 R Core Team. 2020. R: a language and environment for statistical computing, version
858 4.0.2. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
859 Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-
860 dependence on phylogenetic trees. PLoS One 9:e89543.
861 Rakotoarison, A., M. D. Scherz, F. Glaw, J. Köhler, F. Andreone, M. Franzen, J. Glos,
862 O. Hawlitschek, T. Jono, A. Mori, S. H. Ndriantsoa, N. R. Raminosoa, J. C.
863 Riemann, M.-O. Rödel, G. M. Rosa, D. R. Vieites, A. Crottini, and M. Vences.
864 2017. Describing the smaller majority: integrative taxonomy reveals twenty-six

new species of tiny microhylid frogs (genus *Stumpffia*) from Madagascar. Vertebr. Zool. 67:271–398.

Ramsay, J. O., S. Graves, and G. Hooker. 2020. “fda”: functional data analysis. R package version 2.4.8. https:// https://cran.r-project.org/web/packages/fda/index.html.

Ramsay, J. O., G. Hooker, and S. Graves. 2009. Functional data analysis with R and MATLAB. Springer, New York, NY.

Rancilhac, L., T. Bruy, M. D. Scherz, E. A. Pereira, M. Preick, N. Straube, M. L. Lyra, A. Ohler, J. W. Streicher, F. Andreone, A. Crottini, C. R. Hutter, J. C. Randrianantoandro, A. Rakotoarison, F. Glaw, M. Hofreiter, and M. Vences. 2020. Target-enriched DNA sequencing from historical type material enables a partial revision of the Madagascar giant stream frogs (genus *Mantidactylus*). J. Nat. Hist. 54:87–118.

Rand, A. S. 1952. Jumping ability of certain anurans, with notes on endurance. Copeia 1952:15–20.

Rasband, W. S. 1997. ImageJ. U. S. National Institutes of Health, Bethesda, Maryland.

Rebelo, A. D. and J. Measey. 2019. Locomotor performance constrained by morphology and habitat in a diverse clade of African frogs (Anura: Pyxicephalidae). Biol. J. Linn. Soc.127:310–323.

Revell, L. J. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258–3268.

———. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

888 Revell, L. J. and D. C. Collar. 2009. Phylogenetic analysis of the evoluitonary correlation
889 using likelihood. Evolution 63:1090–1100.
890 Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary
891 process, and rate. Syst. Biol. 57:591–601.
892 Richards, C. T. 2010. Kinematics and hydrodynamics analysis of swimming anurans
893 reveals striking inter-specific differences in the mechanism for producing thrust.
894 J. Exp. Biol. 213:621–634.
895 Sanderson, M. J. 1996. How many taxa must be sampled to identify the root node of a
896 large clade? Syst. Biol. 45:168–173.
897 Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25
898 years of image analysis. Nature Methods 9:671–675.
899 Slater, G. J. 2015. Iterative adaptive radiations of fossil canids show no evidence for
900 diversity-dependent trait evolution. Proc. Natl. Acad. Sci. USA 112:4897–4902.
901 Slater, G. J. and A. R. Friscia. 2019. Hierarchy in adaptive radiation: A case study using
902 the Carnivora (Mammalia). Evolution 73:524–539.
903 Sokal, R. R. and F. J. Rohlf. 1995. Biometry. W.H. Freeman, New York, NY.
904 Thomas, G. H. and R. P. Freckleton. 2012. MOTMOT: models of trait macroevolution on
905 trees. Methods Ecol. Evol. 3:145–151.
906 Toro, E., A. Herrel, and D. J. Irschick. 2004. The evolution of jumping performance in
907 Caribbean *Anolis* lizards: solutions to biomechanical trade-offs. Am. Nat.
908 163:844–856.

909 Toro, E., A. Herrel, B. Vanhooydonck, and D. J. Irschick. 2003. A biomechanical
910 analysis of intra- and interspecific scaling of jumping and morphology in
911 Caribbean *Anolis* lizards. J. Exp. Biol. 206:2641–2652.
912 Vences, M., F. Glaw, J. Köhler, and K. C. Wollenberg. 2010. Molecular phylogeny,
913 morphology and bioacoustics reveal five additional species of arboreal microhylid
914 frogs of the genus *Anodonthyla* from Madagascar. Contrib. Zool. 79:1–32.
915 Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Köhler, F. Glaw, and M. Vences. 2009.
916 Vast underestimation of Madagascar's biodiversity evidenced by an integrative
917 amphibian inventory. Proc. Natl. Acad. Sci. USA 106:8267–8272.
918 Walker, J. A. 1998. Estimating velocities and accelerations of animal locomotion: a
919 simulation experiment comparing numerical differentiation algorithms. J. Exp.
920 Biol. 201:981–995.
921 Wells, K. D. 2007. The ecology and behavior of amphibians. Univ. Chicago Press,
922 Chicago, IL.
923 Wiens, J. J., R. A. Pyron, and D. S. Moen. 2011. Phylogenetic origins of local-scale
924 diversity patterns and the causes of Amazonian megadiversity. Ecol. Lett.
925 14:643–652.
926 Wilson, R. S. 2001. Geographic variation in thermal sensitivity of jumping performance
927 in the frog *Limnodynastes peronii*. J. Exp. Biol. 204:4227–4236.
928 Wollenberg Valero, K. C., J. Garcia-Porta, A. Rodriguez, M. Arias, A. Shah, R. D.
929 Randrianiaina, J. L. Brown, F. Glaw, F. Amat, S. Kunzel, D. Metzler, R. D.
930 Isokpehi, and M. Vences. 2017. Transcriptomic and macroevolutionary evidence

for phenotypic uncoupling between frog life history phases. Nat. Commun. 8:15213.

Zug, G. R. 1978. Anuran locomotion – structure and function. 2. Jumping performance of semiaquatic, terrestrial, and arboreal frogs. Smithsonian Contrib. Zool. 276:1–31.

Table S1. Results of two-block partial least-squares (2B-PLS) analysis of the relationship between body shape and performance in Mantellidae. All input variables were phylogenetically transformed residual values from a phylogenetic regression to size-standardize data. We interpreted weights in bold to be those representing the primary relationship between body shape and performance along each dimension. Singular values represent the total covariance between morphology and performance in each dimension; correlations are those between the 2B-PLS latent variables representing morphology and performance in each dimension.

		Dimensions		
Matrix	Variable	1	2	3
$\mathbf{F}_1$	Relative leg length	**0.554**	**0.594**	0.080
	Relative leg mass	**0.440**	0.176	**0.253**
	Relative arm length	-0.120	**0.303**	**-0.488**
	Relative head length	0.051	0.127	**0.387**
	Relative head width	-0.063	0.029	0.022
	Relative tubercle area	-0.210	**0.520**	**-0.475**
	Relative foot webbing area	**0.559**	-0.200	**-0.346**
	Relative toe tip area	0.316	**-0.380**	**-0.279**
	Relative finger tip area	0.147	-0.229	**-0.344**
$\mathbf{F}_2$	Rel. maximum clinging angle	0.352	**-0.768**	**0.536**
	Relative peak jump velocity	**0.552**	**0.632**	**0.543**
	Relative peak swim velocity	**0.756**	-0.104	**-0.646**
Singular value		1.592	0.722	0.221
Correlation		0.710	0.444	0.392
P-value		0.001	0.047	0.475

Table S2. Results of two-block partial least-squares (2B-PLS) analysis of the overall relationship between body shape and performance in non-mantellid frogs (55 species). We multiplied weights of dimension 3 by -1 to make them more easily comparable to the results for Mantellidae (Table S1). Given that only the relative magnitudes and signs of the weights matter, this transformation did not change interpretation of the results (Jolliffe and Cadima 2016). Note that the coefficient for clinging angle in dimension 2 only rounds to -1, which is why jumping and swimming velocity have non-zero coefficients. All other features of the table follow Table S1.

		Dimensions		
Matrix	Variable	1	2	3
$\mathbf{F}_1$	Relative leg length	**0.617**	0.178	0.002
	Relative leg muscle mass	**0.649**	0.161	-0.030
	Relative arm length	0.077	0.170	0.281
	Relative head length	0.252	0.043	**0.584**
	Relative head width	-0.120	0.183	**0.457**
	Relative tubercle area	-0.199	0.305	0.054
	Relative foot webbing area	0.235	-0.204	**-0.509**
	Relative toe tip area	0.135	**-0.712**	0.210
	Relative finger tip area	0.023	**-0.485**	0.254
$\mathbf{F}_2$	Rel. maximum clinging angle	-0.011	**-1.000**	-0.023
	Relative peak jump velocity	**0.720**	-0.023	**0.694**
	Relative peak swim velocity	**0.694**	0.009	**-0.720**
Singular value		1.544	0.755	0.251
Correlation		0.696	0.490	0.410
P value		0.001	0.001	0.014

Table S3. Evolutionary rate comparisons of Mantellidae and other anurans across Jetz and Pyron's (2018) posterior distribution of trees. "Mean σ^2 ratio" refers to the average ratio of the mantellid rate to the non-mantellid rate. "Prop. σ^2 mant > σ^2 other" reflects the number of trees in which the mantellid rate was higher than the non-mantellid rate. P_{mean} is the average *P*-value of trees in which Mantellidae showed a higher rate. "Prop. significant *P*" refers to the number of trees in which Mantellidae showed a significantly higher rate, as assessed through 1000 simulations on each tree.

(A) Mantellidae vs. other anurans (with Pelodryadinae): 80-species tree

Trait type	Mean σ^2 ratio	Prop. σ^2 mant > σ^2 other	P_{mean}	Prop. significant *P*
Body size	1.605	0.996	0.235	0.092
Body shape	1.044	0.521	0.413	0.072
Performance	0.694	0.023	0.617	0.000

(B) Mantellidae vs. other anurans (with Pelodryadinae): 217-species tree

Trait type	Mean σ^2 ratio	Prop. σ^2 mant > σ^2 other	P_{mean}	Prop. significant *P*
Body size	0.958	0.378	0.642	0.003
Body shape	1.161	0.779	0.204	0.357

(C) Mantellidae vs. others (without Pelodryadinae): 69-species tree

Trait type	Mean σ^2 ratio	Prop. σ^2 mant > σ^2 other	P_{mean}	Prop. significant *P*
Body size	1.818	0.996	0.143	0.212
Body shape	1.183	0.826	0.327	0.165
Performance	0.866	0.155	0.699	0.002

(D) Mantellidae vs. others (without Pelodryadinae): 206-species tree

Trait type	Mean σ^2 ratio	Prop. σ^2 mant > σ^2 other	P_{mean}	Prop. significant *P*
Body size	0.959	0.382	0.641	0.002
Body shape	1.236	0.878	0.155	0.501

962 **Table S4**. Results of evolutionary rate comparisons when grouping Mantellidae and
963 Pelodryadinae together and comparing their shared rate to the background rate of other
964 anurans, as assessed on the 80-species complete-data phylogeny. "Overall σ^2"
965 represents the rate estimated for the entire tree (i.e., without specifying groups). For
966 each comparison, "σ^2 ratio" represents the ratio of the higher rate to the lower rate, and
967 *P* reflects the probability that both groups have the same rate. Posterior-distribution
968 results follow the explanation in Table S3.

Maximum clade credibility (MCC) tree results

Trait type	Overall σ^2	Mant-pelo σ^2	Other σ^2	σ^2 ratio	*P*
Body size	0.00201	0.00265	0.00149	1.779	0.062
Body shape	0.00395	0.00452	0.00349	1.295	0.061
Performance	0.00078	0.00089	0.00070	1.269	0.286

Posterior-distribution results

Trait type	Mean σ^2 ratio	Prop. σ^2 mant-pelo > σ^2 other	P_{mean}	Prop. significant *P*
Body size	1.779	0.996	0.107	0.300
Body shape	1.318	0.988	0.129	0.500
Performance	1.295	0.971	0.304	0.082

969

970 **Table S5**. Evolutionary rate comparisons between Pelodryadinae and other anurans
971 (including Mantellidae) on the 80-species phylogeny, showing that low sample size (11
972 species of Pelodryadinae) does not necessarily compromise statistical power. All
973 analyses were done on the maximum clade credibility (MCC) tree calculated from the
974 posterior distribution of Jetz and Pyron (2018). All other table details follow the MCC
975 results of Table 2.

Trait type	Overall σ^2	Pelodryadinae σ^2	Other σ^2	σ^2 ratio	*P*
Body size	0.00201	0.00250	0.00194	1.293	0.571
Body shape	0.00395	0.00555	0.00370	1.501	0.046
Performance	0.00078	0.00153	0.00066	2.310	0.014

976

977 **Table S6**. Results from all diversification-rate analyses. Method was either the method-
978 of-moments estimator ("MoM"; Magallón and Sanderson 2001) or birth-death estimates
979 (Nee et al. 1994); the latter either assumed no extinction ("pure-birth") or estimated
980 extinction rates ("birth-death"). Species diversity of Mantellidae follows two estimates:
981 currently described species (AmphibiaWeb 2020) and the estimate of Perl et al. (2014).
982 Clade ages came from Jetz and Pyron (2018; J & P) or Feng et al. (2017; FEA). Net
983 diversification rates estimated using the stem or crown ages are calculated differently
984 (Magallón and Sanderson 2001); only stem ages were considered from Feng et al.
985 (2017) due to more limited taxon sampling. ε is the relative extinction fraction assumed
986 for the MoM estimates. *r* is the net diversification rate for Mantellidae. Number of
987 families in each comparison differed depending on the availability of ages in the source
988 phylogenies, as well as whether diversification rates were greater than zero (crown
989 rates are undefined at very low diversities; Magallón and Sanderson 2001). *P* reflects
990 phylogenetic ANOVA tests of whether mantellid rates were elevated above those of
991 other anuran families.
992

Method	Mantellidae diversity	Source clade ages	Type	ε	r	Families	P
MoM	229	J & P	Stem	0.0	0.0587	53	0.328
MoM	229	J & P	Stem	0.5	0.0512	53	0.268
MoM	229	J & P	Stem	0.9	0.0342	53	0.244
MoM	229	J & P	Crown	0.0	0.0613	51	0.421
MoM	229	J & P	Crown	0.5	0.0576	51	0.417
MoM	229	J & P	Crown	0.9	0.0403	51	0.350
MoM	229	FEA	Stem	0.0	0.1012	43	0.222
MoM	229	FEA	Stem	0.5	0.0884	43	0.228
MoM	229	FEA	Stem	0.9	0.0590	43	0.253
MoM	409	J & P	Stem	0.0	0.0649	53	0.172
MoM	409	J & P	Stem	0.5	0.0575	53	0.160
MoM	409	J & P	Stem	0.9	0.0403	53	0.121
MoM	409	J & P	Crown	0.0	0.0688	51	0.264
MoM	409	J & P	Crown	0.5	0.0651	51	0.233
MoM	409	J & P	Crown	0.9	0.0476	51	0.155
MoM	409	FEA	Stem	0.0	0.1120	43	0.127
MoM	409	FEA	Stem	0.5	0.0991	43	0.119
MoM	409	FEA	Stem	0.9	0.0695	43	0.103
Pure-birth	229	J & P	Crown	0.0	0.0547	38	0.943
Birth-death	229	J & P	Crown	NA	0.0489	38	0.369

993

Table S7. Results of phylogenetic principal components analysis of body-shape residuals.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eigenvalues	0.0177	0.0085	0.0026	0.0013	0.0005	0.0003	0.0001	0.0001	0.0000
Prop. variance explained	0.5679	0.2715	0.0827	0.0427	0.0169	0.0108	0.0044	0.0020	0.0013
Cumulative variance	0.5679	0.8393	0.9220	0.9647	0.9816	0.9923	0.9967	0.9987	1.0000

Eigenvectors

Original variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Leg length	-0.029	-0.002	0.027	-0.185	-0.069	-0.185	-0.466	-0.520	0.662
Leg muscle mass	-0.067	-0.072	0.132	-0.910	-0.307	0.092	0.077	0.174	-0.077
Arm length	-0.013	0.031	0.002	0.023	-0.068	-0.208	-0.833	0.311	-0.401
Head length	0.002	0.005	-0.001	-0.070	-0.111	-0.716	0.224	-0.472	-0.444
Head width	0.013	-0.003	-0.072	0.029	-0.024	-0.624	0.175	0.615	0.441
Metatarsal tubercle area	0.005	-0.002	-0.986	-0.149	0.013	0.048	-0.019	-0.035	-0.028
Foot webbing area	-0.969	-0.231	-0.017	0.084	0.015	-0.009	0.014	0.004	-0.006
Toe tip area	-0.158	0.613	0.047	-0.276	0.718	-0.069	-0.007	0.006	-0.029
Finger tip area	-0.174	0.752	-0.033	0.163	-0.607	0.072	0.050	0.004	0.037

997 **Table S8**. Maximum-likelihood estimates (MLE) of λ and its support region of values within 2 log-likelihood units of the
998 MLE (in brackets). Most trait-tree combinations are consistent with Brownian motion, as indicated by MLEs close to 1.0
999 and support regions that include this value. Note that we only evaluated λ values between 0.0–1.0 (Freckleton et al. 2002;
1000 Harmon 2018), as we generally found the likelihood unstable at λ > 1.0. All variables except SVL were standardized to
1001 body size (i.e., residuals from a phylogenetic regression on SVL or body mass, depending on the trait).

(A) 80-species dataset

Trait	All taxa (80 sp.)	Mantellidae (25 sp.)	Other taxa (55 sp.)
Body length (SVL)	0.599 [0.082-0.997]	0.000 [0.000-1.000]	0.758 [0.211-1.000]
Leg length	0.826 [0.405-0.998]	1.000 [0.000-1.000]	0.735 [0.065-0.974]
Leg muscle mass	0.962 [0.748-1.000]	1.000 [0.320-1.000]	0.929 [0.565-1.000]
Arm length	0.602 [0.119-0.959]	0.000 [0.000-1.000]	0.438 [0.000-0.958]
Head length	0.813 [0.460-1.000]	0.668 [0.076-1.000]	0.584 [0.172-0.982]
Head width	0.113 [0.000-0.730]	0.725 [0.000-1.000]	0.000 [0.000-0.431]
Tubercle area	1.000 [0.646-1.000]	0.545 [0.000-1.000]	1.000 [0.375-1.000]
Foot webbing area	1.000 [0.818-1.000]	1.000 [0.143-1.000]	1.000 [0.743-1.000]
Toe tip area	0.993 [0.799-1.000]	0.658 [0.048-1.000]	1.000 [0.883-1.000]
Finger tip area	0.931 [0.681-1.000]	0.722 [0.029-1.000]	0.946 [0.663-1.000]
Max. clinging angle	0.980 [0.847-1.000]	1.000 [0.000-1.000]	0.975 [0.806-1.000]
Peak jump velocity	0.934 [0.697-1.000]	1.000 [0.000-1.000]	0.901 [0.564-1.000]
Peak swim velocity	0.937 [0.708-1.000]	0.000 [0.000-1.000]	0.923 [0.641-1.000]

(B) 217-species dataset

Trait	All taxa (217 sp.)	Mantellidae (36 sp.)	Other taxa (181 sp.)
Body length (SVL)	0.927 [0.763-0.992]	1.000 [0.800-1.000]	0.855 [0.632-0.968]

Leg length	0.978 [0.917-1.000]	1.000 [0.570-1.000]	0.972 [0.896-1.000]
Leg muscle mass	0.974 [0.900-1.000]	0.848 [0.343-1.000]	0.973 [0.884-1.000]
Arm length	0.932 [0.800-0.992]	0.000 [0.000-0.983]	0.954 [0.825-1.000]
Head length	0.957 [0.877-0.996]	0.492 [0.000-0.961]	0.967 [0.884-1.000]
Head width	0.972 [0.872-1.000]	1.000 [0.054-1.000]	0.961 [0.828-1.000]
Tubercle area	0.902 [0.724-0.975]	0.924 [0.277-1.000]	0.876 [0.632-0.971]
Foot webbing area	1.000 [0.977-1.000]	1.000 [0.692-1.000]	1.000 [0.970-1.000]
Toe tip area	1.000 [0.967-1.000]	0.720 [0.159-1.000]	1.000 [0.984-1.000]
Finger tip area	0.987 [0.934-1.000]	0.779 [0.260-1.000]	1.000 [0.965-1.000]

Table S9. Results of evolutionary model-fitting to the traits and trees of this study. Values are AICc weights for four models, assessed within each combination of trait and phylogeny: BM = Brownian motion; OU = single-optimum Ornstein-Uhlenbeck model; EB = early-burst model; λ = internal branch lengths scaled by lambda. The model in bold is the one with the highest weight in each comparison set. All variables except SVL were standardized to body size (i.e., residuals from a phylogenetic regression). Weights may not sum to 1.0 due to rounding error. Note that the "all taxa" and "other taxa" results are highly similar due to the much larger proportion of shared species than between "all taxa" and "Mantellidae".

(A) 80-species dataset

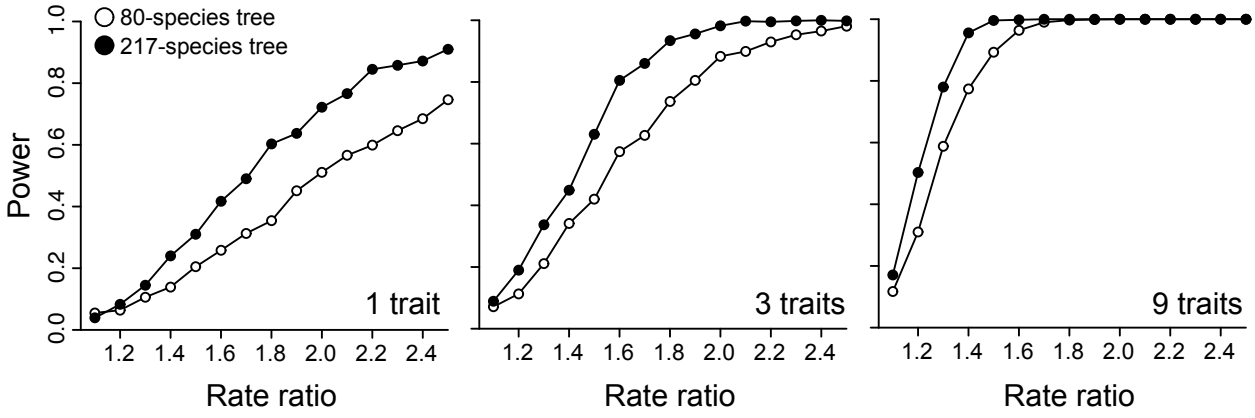
	All taxa (80 sp.)				Mantellidae (25 sp.)				Other taxa (55 sp.)			
Trait	BM	OU	EB	λ	BM	OU	EB	λ	BM	OU	EB	λ
Body length (SVL)	0.056	**0.776**	0.019	0.149	**0.377**	0.277	0.103	0.243	0.163	**0.534**	0.053	0.249
Leg length	0.086	**0.657**	0.029	0.228	**0.536**	0.146	0.171	0.146	0.031	**0.772**	0.010	0.187
Leg muscle mass	0.010	**0.822**	0.003	0.165	0.265	**0.381**	0.072	0.282	0.014	**0.802**	0.005	0.179
Arm length	0.150	**0.542**	0.051	0.256	**0.407**	0.178	0.111	0.304	0.064	**0.707**	0.021	0.209
Head length	0.000	**0.832**	0.000	0.168	**0.451**	0.243	0.123	0.183	0.000	**0.690**	0.000	0.310
Head width	0.338	**0.433**	0.115	0.115	**0.433**	0.224	0.118	0.225	0.361	**0.403**	0.118	0.118
Tubercle area	**0.494**	0.170	0.168	0.168	**0.547**	0.155	0.149	0.149	**0.504**	0.167	0.165	0.165
Foot webbing area	**0.440**	0.260	0.149	0.151	**0.429**	0.223	0.117	0.231	**0.502**	0.170	0.164	0.164
Toe tip area	0.249	**0.400**	0.085	0.266	**0.478**	0.208	0.130	0.184	**0.329**	0.311	0.107	0.252
Finger tip area	0.288	**0.471**	0.098	0.143	**0.546**	0.149	0.156	0.149	0.208	**0.586**	0.068	0.137
Max. clinging angle	**0.473**	0.168	0.161	0.198	**0.541**	0.164	0.148	0.148	**0.476**	0.167	0.155	0.201
Peak jump velocity	0.274	**0.355**	0.093	0.278	**0.536**	0.166	0.146	0.151	0.187	**0.502**	0.061	0.250
Peak swim velocity	**0.340**	0.318	0.115	0.227	**0.435**	0.261	0.119	0.185	0.328	**0.341**	0.107	0.225

(B) 217-species dataset

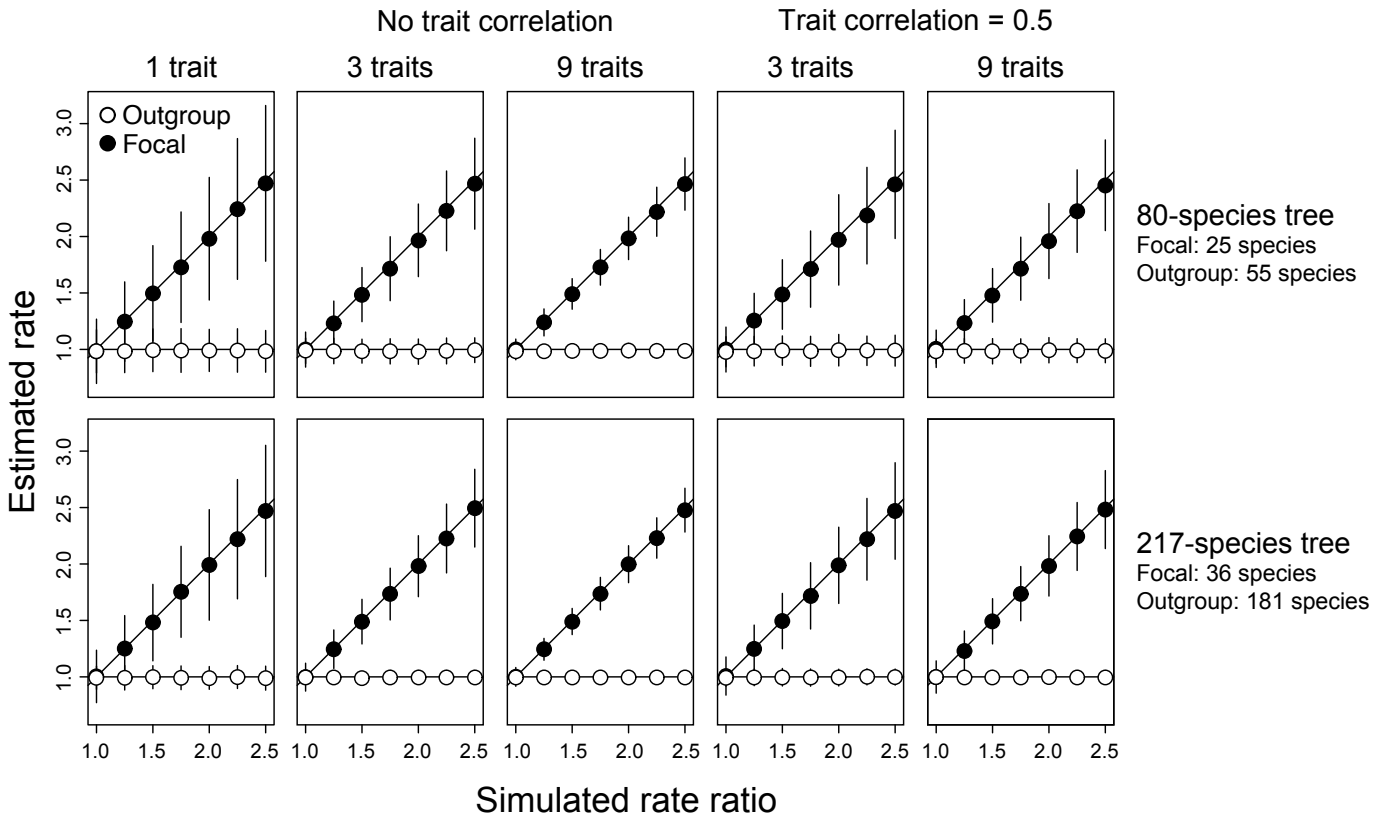
	All taxa (217 sp.)				Mantellidae (36 sp.)				Other taxa (181 sp.)			
Trait	BM	OU	EB	λ	BM	OU	EB	λ	BM	OU	EB	λ
Body length (SVL)	0.000	**0.997**	0.000	0.002	**0.518**	0.157	0.168	0.157	0.000	**0.985**	0.000	0.015
Leg length	0.300	0.165	0.107	**0.428**	**0.521**	0.163	0.158	0.158	0.261	0.149	0.093	**0.497**
Leg muscle mass	0.088	0.278	0.031	**0.603**	0.108	**0.569**	0.033	0.290	0.234	0.254	0.083	**0.428**
Arm length	0.157	0.076	0.056	**0.710**	0.121	0.370	0.037	**0.473**	0.282	0.117	0.100	**0.501**
Head length	0.244	**0.473**	0.087	0.196	**0.480**	0.229	0.146	0.146	0.193	**0.536**	0.069	0.202
Head width	0.001	**0.799**	0.000	0.200	**0.513**	0.159	0.156	0.172	0.000	**0.803**	0.000	0.197
Tubercle area	**0.482**	0.174	0.172	0.172	**0.523**	0.159	0.160	0.159	**0.484**	0.172	0.172	0.172
Foot webbing area	**0.473**	0.169	0.188	0.169	0.292	0.244	0.089	**0.376**	0.356	0.127	**0.391**	0.127
Toe tip area	**0.427**	0.212	0.153	0.208	**0.334**	0.236	0.101	0.328	**0.483**	0.172	0.173	0.172
Finger tip area	0.199	**0.530**	0.071	0.200	**0.452**	0.191	0.137	0.220	0.182	**0.599**	0.065	0.155

1011

Figure S1. Results of power simulations of multivariate evolutionary rate comparisons at our two different tree sizes. Power increases with rate ratio, number of traits, and tree size. The 80-species tree had 25 species of mantellids, while the 217-species tree had 36. Trait numbers corresponded to our phenotypic analyses (i.e., one trait for body size, three traits for performance, and nine traits for body shape). Trait correlations in simulations assumed the mean estimated correlation among traits in empirical analyses of the 80-species tree (i.e., 0.364 for three performance traits and 0.110 for nine morphological traits).



1021 **Figure S2**. The effect of subsampling clades on parameter estimation when comparing rates of continuous trait evolution.
1022 Evolution of one, three, and nine traits was simulated on a 3449-species phylogeny of Anura, with 189 species in the focal
1023 group Mantellidae. Multi-trait simulations had either no trait correlation or a correlation of 0.5. After simulation, trees were
1024 randomly pruned to include only 80 (25 focal and 55 outgroup) or 217 (36 focal and 181 outgroup) species, following tree
1025 sizes in our empirical analyses. Dots show mean estimated rates across 1000 simulations of trait evolution, with bars
1026 indicating standard deviations of estimates. Black dots show estimates for the focal group, which had the higher rate in
1027 simulations; diagonal lines represent a 1:1 correspondance between simulated values and estimates. White dots show
1028 estimates for the rest of the species in the outgroup, which were always simulated as 1.0; horizontal lines are drawn
1029 through 1.0. Overall, we found that rate estimates were unbiased under all simulation conditions, even under deep
1030 subsampling (e.g., 13% of the focal clade and 1.7% of other species in the 80-species tree). Moreover, increasing trait
1031 numbers and reducing trait correlation both increased precision in estimates, as expected (Adams 2014b; Adams and
1032 Collyer 2018) and consistent with statistical power estimates (Fig. S3). Increasing tree size only slightly increased
1033 precision (i.e., comparing the top vs. bottom rows of panels).


No trait correlation
Trait correlation = 0.5
1 trait
3 traits
9 traits
3 traits
9 traits
Outgroup
Focal
Estimated rate
Simulated rate ratio
80-species tree
Focal: 25 species
Outgroup: 55 species
217-species tree
Focal: 36 species
Outgroup: 181 species

1034

Figure S3. Type I error and power estimates for subsampling simulations. Type I error was assessed at a rate ratio of 1.0 (i.e., no difference in rate between focal group and outgroup) and estimates cluster around the dashed horizontal line, drawn at the expected Type I error rate of 0.05. Power was assessed at all rate ratios >1.0, where Mantellidae was simulated as having a higher rate of evolution. Values indicate the proportion of simulations in which the true difference in rates was detected at α = 0.05. White circles represent analyses conducted after subsampling 80 species from the full 3449-species tree of Jetz and Pyron (2018), whereas black circles indicated subsampling 217 species. Solid lines connect results when simulating trait evolution without correlation among traits, whereas dotted lines connect results with a correlation of 0.5. Overall, our results suggest that subsampling a small proportion of species from a large tree leads to neither increased Type I error rates nor decreased power due to subsampling per se (e.g., compare to Fig. S1).