SUPPORTING INFORMATION

A cautionary note on using quantitative measures of phenotypic convergence

David M. Grossnickle, William H. Brightly, Lucas N. Weaver, Kathryn E. Stanchak,
     Rachel A. Roston, Spencer K. Pevsner, C. Tristan Stayton, P. David Polly, Chris J. Law

SUPPLEMENTAL METHODS

Univariate model-fitting analyses

    One limitation of the *mvMORPH* multivariate models, which are used for our primary

model-fitting analyses, is that they do not permit the evolutionary rate (σ) or strength of

    attraction to optima (α) to vary between the two selective regimes ('gliders' and 'non-

gliders'). This likely results in poor model performance because the datasets were

    simulated such that 'gliders' and 'non-gliders' should have different rates and attraction

strengths. For example, the 'non-gliders' are evolved by BM, and thus they are not

    expected to exhibit attraction to a trait optimum, whereas the convergent 'glider'

lineages are expected to exhibit strong attraction due to being simulated by an OU

    process. Further, the phylogenetic half-life ($\ln(2)/\alpha$) of the 'glider' regime cannot be

calculated independent of the 'non-glider' regime if the α parameter is uniform across

    both regimes, which is the case with the multivariate models.

Thus, we also fit seven univariate evolutionary models to the subset of simulated

    datasets, including several multiple-regime OU models that permit σ and α to vary

between regimes. Using functions in the *OUwie R* package (Beaulieu et al. 2012), we fit

    these models to the first principal component (PC1) scores of a principal components

analysis of the six simulated traits. The univariate models include uniform (or single-

    regime) BM and OU models, as well as a suite of multiple-regime OU models (i.e.,

'OUM' models of Beaulieu et al. 2012). The OU2 model keeps α and σ constant for both

    regimes, the OU2A model allows α (but not σ) to vary between regimes, the OU2V

32   model allows σ^2 (but not α) to vary between regimes, and the OU2VA model allows both

        σ and α to vary between regimes. As with the multivariate analyses, all models were

34   fitted across 10 'simmaps' for each of the 30 datasets and relative support for models

        was measured using AICcW.

36   We recognize that fitting models to PC scores can lead to biased results (Uyeda

        et al. 2015), and thus our univariate results should be considered with caution.

38   However, we feel that using PC1 scores here is justified for two reasons. First, the

        alternative option is to fit models to each of the six simulated traits individually, but four

40   of the traits are evolved via a strong OU process and two traits are evolved via BM (in

        our subset of datasets used in model-fitting analyses; see Methods), and thus the

42   model-fitting results are expected to vary considerably between those two types of

        traits. PC1 provides a single value for which results can be more easily interpreted

44   compared to results for the six traits. Second, our conclusions concerning the use of

        model-fitting analyses for testing for convergence are based entirely on the multivariate

46   model-fitting analyses (see Results & Discussion), and thus the results of the univariate

        model-fitting analyses (which are congruent with the multivariate results; Tables 2 and

48   S1) do not influence the broad conclusions of this study. The univariate model-fitting

        analyses are simply a supplemental analysis that provide a fitted α value and

50   phylogenetic half-life for the 'glider' regime.

52   ***Ct*-measures**

        We used the *R* script from Zelditch et al. (2017) as a foundation for the updated

54   functions for calculating *Ct*1–*Ct*4 and simulation-based *p*-values because they are

        computationally faster than the original *R* functions in the *convevol R* package (Stayton

56   2015, Stayton 2018). Note that the relevant *R* functions are titled *calcConv* (*C*

        calculations) and *convSig* (significance testing) in the *R* code of Zelditch et al. (2017),

58   *convrat* and *convratsig* in the original *convevol R* package, and *convrat.t* and

        convratsig.t for our updated measures.

60   *$D_{max.t}$ measurement*. The primary change made by the *Ct*-measures in

        comparison to Stayton's (2015) original *C*-measures is the way in which D_{max} is defined.

62   *Ct*-measures were designed to ensure D_{max} (now referred to as $D_{max.t}$) was obtained

from comparisons of synchronous time points along the evolutionary paths leading to

64   the putatively convergent taxa of interest. In this way it prevents the inflation of $D_{max.t}$

that resulted from comparison of asynchronous nodes (e.g., tips and internal nodes)

66   which often occurred when using the original metrics on lineages with outlying

morphologies (Figs. 3C and 4). Several modifications to the source *R* code were made

68   to facilitate this change. Candidate $D_{max.t}$ measurements for putatively convergent

lineages are now measured at each internal node along the branch paths from the most

70   recent common ancestor (MRCA) of the lineages (e.g., see Figures 4 and 5B). At each

of these points we extracted the phenotypic distance between lineages as the euclidean

72   distance between the ancestral reconstruction at the focal node and the coincident

reconstruction along the branch path of the other lineage. Where this corresponds to a

74   point along a branch (which is most cases) the ancestral state is estimated using

formula [2] from Felsenstein (1985), which allows ancestral states to be interpolated at

76   any point along a given branch from reconstructions at the branch's ancestral and

descendant nodes. The code for this was largely repurposed from the *contMap* function

78   of the *phytools R* package (Revell 2012). If no contemporaneous point exists on the

opposite path for a given internal node (e.g., when comparing extinct and extant taxa),

80   then a measurement is not taken at that node. All distances measured between paths

are stored for each pair of user defined tips. $D_{max.t}$ is the maximum of these distance

82   values, but it is restricted to predate either focal tip (i.e., $D_{max.t}$ cannot equal D_{tip}).

Restriction of $D_{max.t}$ to predate the focal tips means the minimum $Ct1$ value is no

84   longer set to zero as in the original *C1*-measure. This allows for some degree of

divergence to be captured (i.e., relatively more negative $Ct1$ values may represent

86   greater divergence). However, users are cautioned from using this to test the magnitude

of divergence between clades. This is because in divergent clades $D_{max.t}$ will almost

88   always be the last time point before the oldest focal tip. The method will thus reflect only

a small portion of the period when lineages were undergoing divergent evolution.

90   Degree of divergence will then be a function of both phenotypic rates of evolution and of

subtending branch length. The latter will in many practical situations be a function of

92   sampling, with long subtending branches due to poor sampling likely to inflate

divergence measures substantially since they will provide the best scenario for a large

3

94 time difference between D_{max} and D_{tip} (and thus capture the greatest proportion of divergent evolution).

96   The changes to D_{max} were the most consequential of those made to modify the original *C*-measures. However, a number of other new options were also included.

98 These are briefly described below. Full documentation of these options will be available as part of the next update to the *convevol R* package (Stayton, 2018).

100   *User-defined groups*. The first new option is for users to provide grouping assignments to the tips being tested, thus allowing comparisons of clades with multiple

102 lineages, whereas the original *C*-measures are limited to comparisons of individual lineages. This option removes pairwise comparison between tips within the same group

104 (e.g., two flying squirrels would not be compared if all flying squirrels are defined as one group) and returns results for each unique comparison between groups in addition to

106 overall results. This option is useful if it is hypothesized that two (or more) clades converged, and relieves the user from needing to average tip values of a clade or

108 manually define all of the desired comparisons. When using this option, the overall (for all pairwise comparisons) and comparison-specific *Ct* and *p* values are returned. Overall

110 results are provided as both raw values (means of all pairwise comparisons, excluding within-group comparisons) and weighted values. The latter allows each inter-group

112 comparison to impact the overall average equally, so that larger within group sample sizes don't skew overall results. For instance, if there are three putatively convergent

114 groups (Group A, Group B, and Group C), and Groups A and B both include a single lineage and Group C includes 10 lineages, then there would be 21 total pairwise

116 comparisons among groups (one for A-B, 10 for A-C, and 10 for B-C). Although constituting one third of the unique inter-group comparisons, *Ct* measurements taken

118 from comparison of Groups A and B constitute less than 5% of those used to compute overall (average) *Ct* values. Thus, Groups A and B have a relatively smaller impact than

120 Group C on the overall *Ct* scores and *p*-values. The weighted output scales the *Ct* results (and associated *p*-values) so that each unique inter-group comparison

122 contributes equally to the overall results, whereas the raw overall result simply reports the mean value for all 21 pairwise comparisons. Both weighted and unweighted values

124 are reported in the default output printed by the updated *convSig* function, but we

recommend the weighted result be used by default when comparing groups.

126 Nevertheless, the raw result may be preferable in cases in which researchers believe that the more heavily sampled group(s) should have a larger impact on overall results.

128 Note that it is possible to define groups even when those consist of a single tip. While doing so will not change which pairwise comparisons the model considers, it will

130 provide the user with unique Ct scores and p-values for each comparison. This can be especially useful when the degree of convergence varies across the lineages of interest

132 (e.g., see the pairwise results for anole species in Figure S5 and Table S2).

Conservative $D_{max.t}$ option. When providing user-defined groups, a conservative

134 $D_{max.t}$ option is available that limits candidate $D_{max.t}$ measurements to a time point predating the origination of both focal groups (i.e., the nodes of the MRCAs of each

136 group). This is to prevent $D_{max.t}$ being skewed by an early transition of one lineage toward a shared adaptive optimum that is outlying in morphospace, which can result in

138 inflated Ct scores, especially when the origins of the clades are very different in age. This issue is discussed in the Supplemental Results and illustrated in Figure S3. Note

140 that this option is only meaningful when user defined groups are provided. When one of those groups consist of a single lineage the node immediately ancestral to the tip is

142 used. Using this method, long branches can substantially alter inferred $D_{max.t}$ values. We have provided the option to print relevant information about the restrictions put on $D_{max.t}$

144 when using this method (by setting VERBOSE = TRUE in *convrat.t*). We strongly suggest that users investigate the impact of using the conservative $D_{max.t}$ option before

146 committing to significance tests.

Updated Ct4 computation. In addition to changes to $D_{max.t}$, we also altered the

148 way in which the $C4$-measure is computed. The new version (called $Ct4$) redefines $L_{tot.clade}$, which is the value used to standardize the $C2$ value (D_{max} subtracted by D_{tip}) to

150 obtain $C4$. $L_{tot.clade}$ is described by Stayton (2015) as reflecting the total amount of morphological evolution which occurs in the clade originating with the MRCA of two

152 putatively convergent tips. In the original C-measures, $L_{tot.clade}$ values were obtained as a sum of the phenotypic distances from all pairwise comparisons between nodes in the

154 clade, but this does not fully account for phylogenetic structure and is heavily influenced by sampling intensity. We have updated this to now be the sum of the phenotypic

156 distances accumulated along each branch in the clade of interest. This change brings
*C*4 closer to the original description of the metric.

158 *Measuring convergence of single traits.* By default, the original *C*-measures do
not support investigation of convergence in a single trait (although see Spear and

160 Williams, 2020; Law, 2022). To circumvent this limitation we have added code to the
convrat.t function which appends an invariant trait (with value zero) to datasets

162 consisting of a single trait. This approach was taken due to ease of integration with
existing code, and although crude will provide the same phenotypic distances as would

164 be obtained from the single trait.

Model output. Additional changes were made to increase the amount of

166 information returned to the user and facilitate plotting of results. This includes the
addition of the novel *plot.C* function, which is described in the 'Measuring convergence

168 through time via Ct-measures' section of the main text (with example output in Figure
5B).

170

SUPPLEMENTAL RESULTS

172

Univariate model-fitting analyses

174 For univariate models fit to PC1 scores the OU2VA model, which allows varying rates
and attraction strengths between regimes, is the best fitting model at all trait optimum

176 values for both convergence and divergences datasets (Table S1). However for
convergence datasets, the null model (BM1) is the second best-fitting model when the

178 trait optimum is zero and 20, and the total AICcW values for all OU2 models increases
with greater optima values, indicating increased evidence of convergence in

180 morphological outliers. These results are consistent with the results of the multivariate
evolutionary models (Table 2).

182

184

186

188

190

Table S1. Tests of convergence among lineages of the simulated datasets using evolutionary models fit to univariate data (PC1 scores). Model-fitting results for each trait optimum are the mean AICcWs of 50 phylogenetic trees (five datasets with 10 'simmaps' each). Model support for the two-regime models (any variation of the OU2 model) could be interpreted as support for convergence because this model reflects evolution of the putatively convergent lineages toward a shared adaptive peak (but see the Results & Discussion). Abbreviations: AICcW, small-sample corrected Akaike weights; BM, Brownian motion; OU, Ornstein-Uhlenbeck.

|	Model	Trait optimum			
		0	20	50	100
Convergence simulations	BM1	0.157	0.043	0.000	0.000
	OU1	0.058	0.015	0.000	0.000
	OU2	0.041	0.020	0.001	0.000
	OU2A	0.016	0.014	0.103	0.338
	OU2V	0.058	0.022	0.023	0.000
	OU2VA	0.670	0.885	0.873	0.661
Divergence simulations	BM1	–	–	0.000	0.000
	OU1	–	–	0.000	0.000
	OU2	–	–	0.000	0.000
	OU2A	–	–	0.116	0.331
	OU2V	–	–	0.001	0.020
	OU2VA	–	–	0.882	0.649

200

202

In the main text, we discuss a few factors that likely explain why the two-regime OU models are unexpectedly the best-fitting models to divergent data. Namely, the two-regime OU may be the best-fitting of bad-fitting models, with the BM1 and OU1 models even worse fits to the data. An additional factor that may contribute to the relatively strong fits of two-regime OU models to divergence datasets is that we treated the datasets as we would with empirical datasets and used 'simmaps' for ancestral state reconstructions of regime states (gliding or non-gliding), rather than use the known node information (via the simulation data). For instance, the two marsupial glider groups in our dataset are closely related (but believed to have evolved gliding independently), and

212   thus the 'simmaps' might commonly (and mistakenly) reconstruct the MRCA of those

       lineages as having gliding behavior.

214

       C1–C4 and Ct1–Ct4 applied to simulated data

216   In the main text we only present results for C1 (Fig. 3C, Table 1) and Ct1 (Fig. 5A,

       Table 1), which were applied to both the simulated convergence datasets and the

218   simulated divergence datasets. However, Stayton (2015) developed four distance-

       based convergence measures (C1–C4) and one frequency-based measure (C5), with

220   C1 being the primary measure, and we altered C1-C4 to produce the Ct1–Ct4

       measures. Here, we provide full results for C1–C4 (Fig. S1) and Ct1–Ct4 (Fig. S2),

222   which are also applied to both the convergence and divergence datasets. See the

       Methods and Stayton (2015) for descriptions of the four convergence measures, and

224   see the Methods for information on the simulated datasets. Note that the Ct4 measure

       is calculated differently than the C4 measure (see Supplemental Methods). For C1–C4,

226   all results for divergence simulations are greater than zero (Fig. S1), incorrectly

       indicating convergence, whereas the Ct1–Ct4 scores for divergence datasets are

228   generally at or below zero (Fig. S2).

230
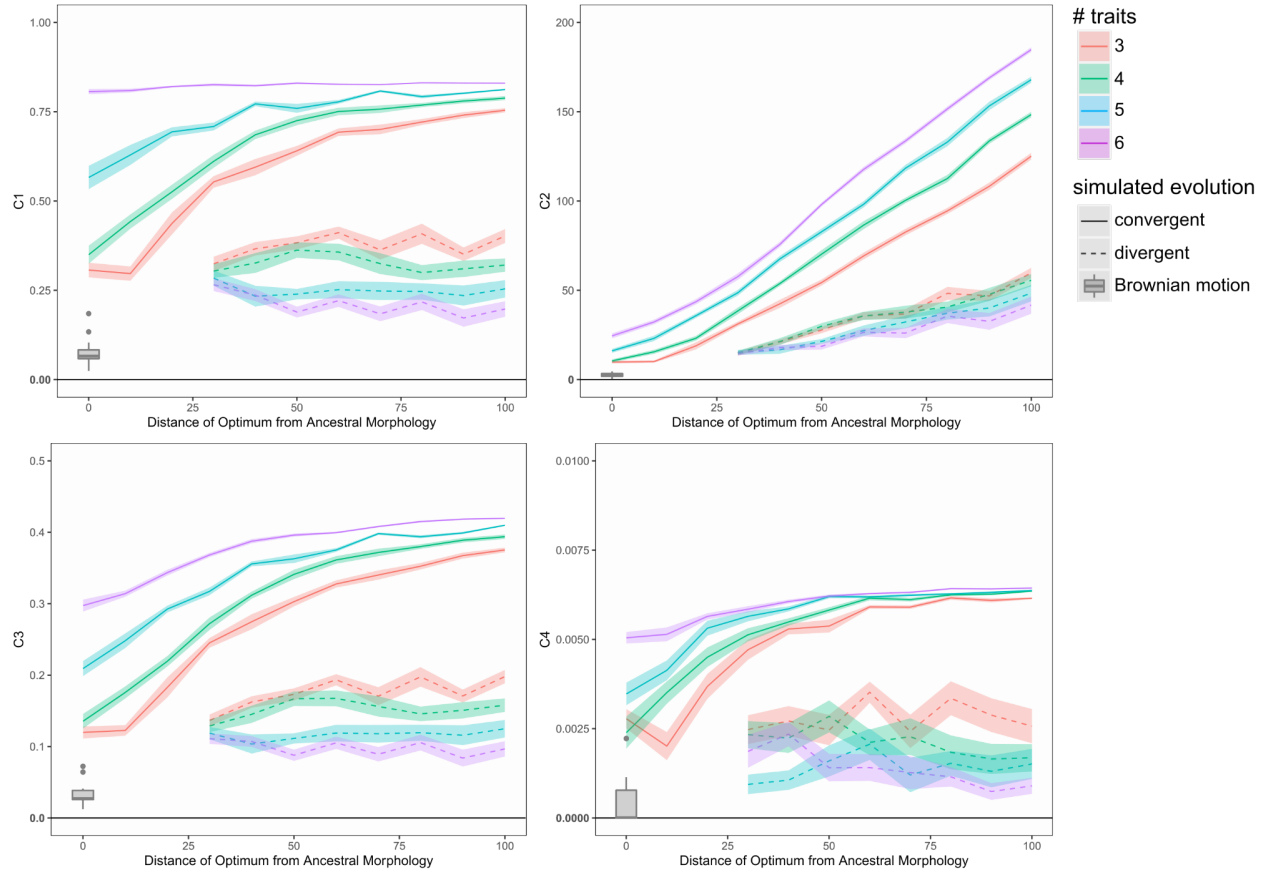
Figure S1. Plots of means and standard errors of *C*1–*C*4 scores for simulated convergent lineages (solid lines) and divergent lineages (dashed lines). Datasets varied in the number of convergent/divergent traits (represented by the different colored lines) and in the distance of trait optima from the ancestral morphology (approximated as the center of morphospace). Means and standard errors are computed from 15 simulated datasets. Greater *C*1–*C*4 values indicate greater convergence. We did not simulate divergence for trait optima of 0, 10, and 20 because at these optima our simulation methods may have inadvertently generated convergence patterns (see Methods and Figure 3). As a second means of simulating divergence, we allowed the lineages of interest ('gliders') to evolve via BM. These are provided as box-and-whisker plots, summarizing 15 simulated datasets of six traits (see Methods). Note that the divergence results are all greater than zero, incorrectly indicating convergence.

9
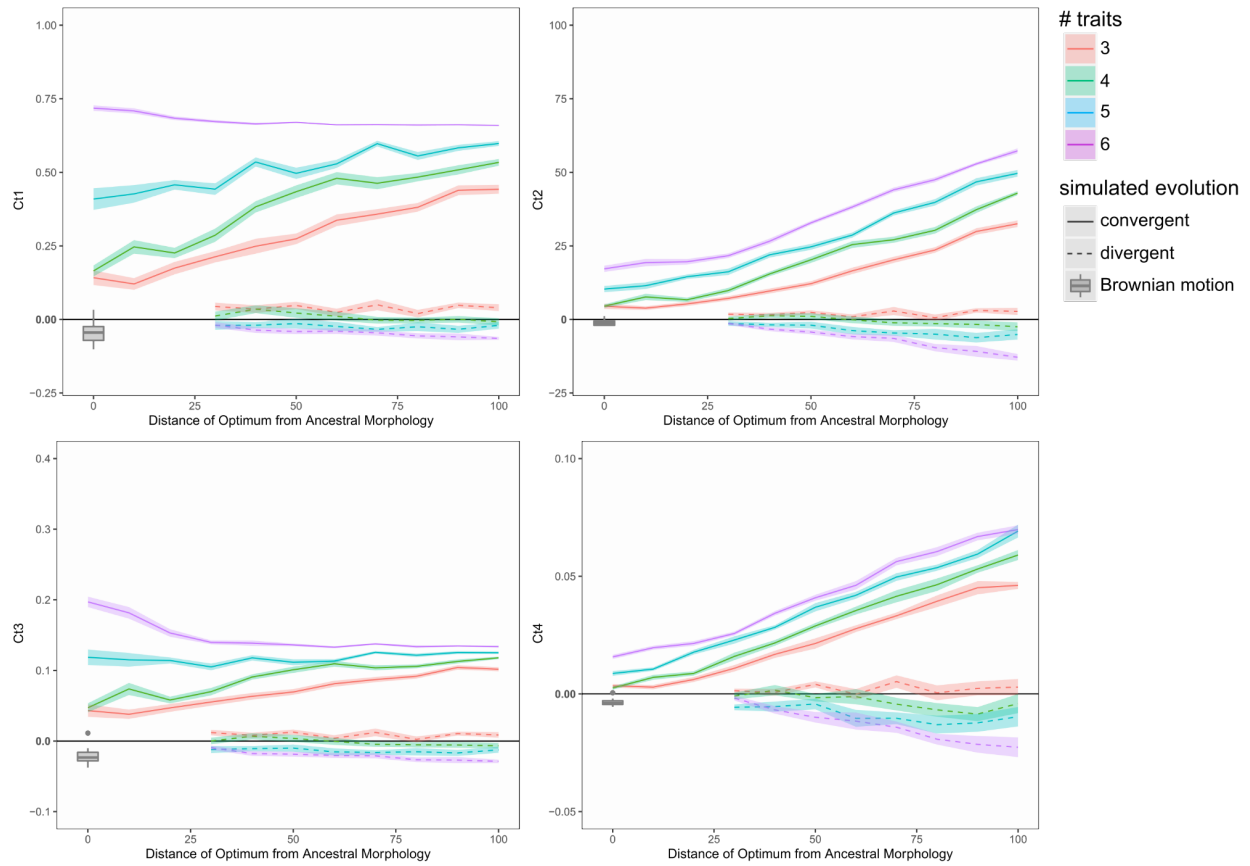
Figure S2. Plots of means and standard errors of $Ct1-Ct4$ scores for simulated convergent lineages (solid lines) and divergent lineages (dashed lines). Datasets varied in the number of convergent/divergent traits (represented by the different colored lines) and in the distance of trait optima from the ancestral morphology (approximated as the center of morphospace). Means and standard errors are each computed from 15 simulated datasets. Greater $Ct1-Ct4$ values indicate greater convergence. We did not simulate divergence for trait optima of 0, 10, and 20 because at these optima our simulation methods may have inadvertently generated convergence patterns (see Methods and Figure 3). As a second means of simulating divergence, we allowed the lineages of interest ('gliders') to evolve via BM. These are provided as box-and-whisker plots, summarizing 15 simulated datasets of six traits (see Methods). Note the differences in the scaling of the vertical axes of the $Ct2$ and $Ct3$ plots relative to the $C2$ and $C3$ plots (Fig. S1), respectively. (The scaling for $C4$ and $Ct4$ is different because these measures are calculated differently.) Also, note the different position of zero relative to results in the $Ct1-Ct4$ plots versus the position in $C1-C4$ plots (Fig. S1), as well as the overlap in the $Ct1-Ct4$ plots of divergence data simulated by both BM and OU processes.

10

270 ***Ct*-measures – the influence of origination times on results**

As discussed in the main text, the *Ct*-measures limit candidate $D_{max.t}$ measurements to

272 specific time slices at internal nodes, and thus the timing of evolutionary change among

putatively convergent lineages can influence the results of *Ct*-measures. For instance, if

274 different lineages of interest evolve toward (or away from) a specific morphology at

different points in time, then the $D_{max.t}$ measurement may not measure the

276 morphologically farthest distances between the lineages. This issue may be magnified

when convergence is expected to be linked to adaptive changes (e.g., adaptations for

278 gliding behavior) that evolved at specific times. For instance, if colugos (i.e.,

Dermoptera or 'flying lemurs') evolved traits associated with gliding behavior

280 approximately 60 Ma, and flying squirrels (Pteromyini) evolved traits associated with

gliding approximately 25 Ma (e.g., Grossnickle et al. 2020), then most of the candidate

282 $D_{max.t}$ measurements will be comparisons of dermopterans with gliding traits to stem

flying squirrels without gliding traits (from 60 to 25 Ma). If the older lineage (colugos)

284 has already undergone considerable evolutionary change by the time that the younger

lineage (flying squirrels) originated, then much of the convergent evolutionary change of

286 the older lineage is not captured by the morphological distances measured at 'time

slices,' which are limited to the time period in which the lineages overlap. Ideally, most

288 candidate $D_{max.t}$ measurements would be comparisons of non-gliding stem colugos and

non-gliding stem flying squirrels that lack the adaptive traits associated with gliding. This

290 issue might lead to candidate $D_{max.t}$ measurements being smaller than expected, or at

least smaller than those calculated by measures that ignore time (e.g., *C*-measures).

292 Conversely, if the putatively convergent taxa evolve toward outlying regions of

morphospace, then the asynchronous origins of the clades could inflate the *Ct*-

294 measures. We illustrate this in Figure S3. In the conceptual illustrations, the *Ct*1 score is

consistently 0.3 when convergent lineages originate at the same time and/or when

296 lineages evolve toward the ancestral morphology. However, when lineages originate at

different times and evolve toward an outlying region of morphospace, then the *Ct*1

298 score is 0.7. Thus, researchers should be cautious when applying Ct measures to

datasets with outlying taxa of various origination ages, and we offer some suggestions

300 in the main text for mitigating this issue. It is also worth noting that this latter scenario

11

assumes that the convergent lineages can reach adaptive zones; if the later-evolving

302   convergent lineage is still evolving toward outlying morphospace (i.e., it has yet to reach

an adaptive peak or zone) then the aforementioned issue may have less of an influence

304   on *Ct* results.

306

308
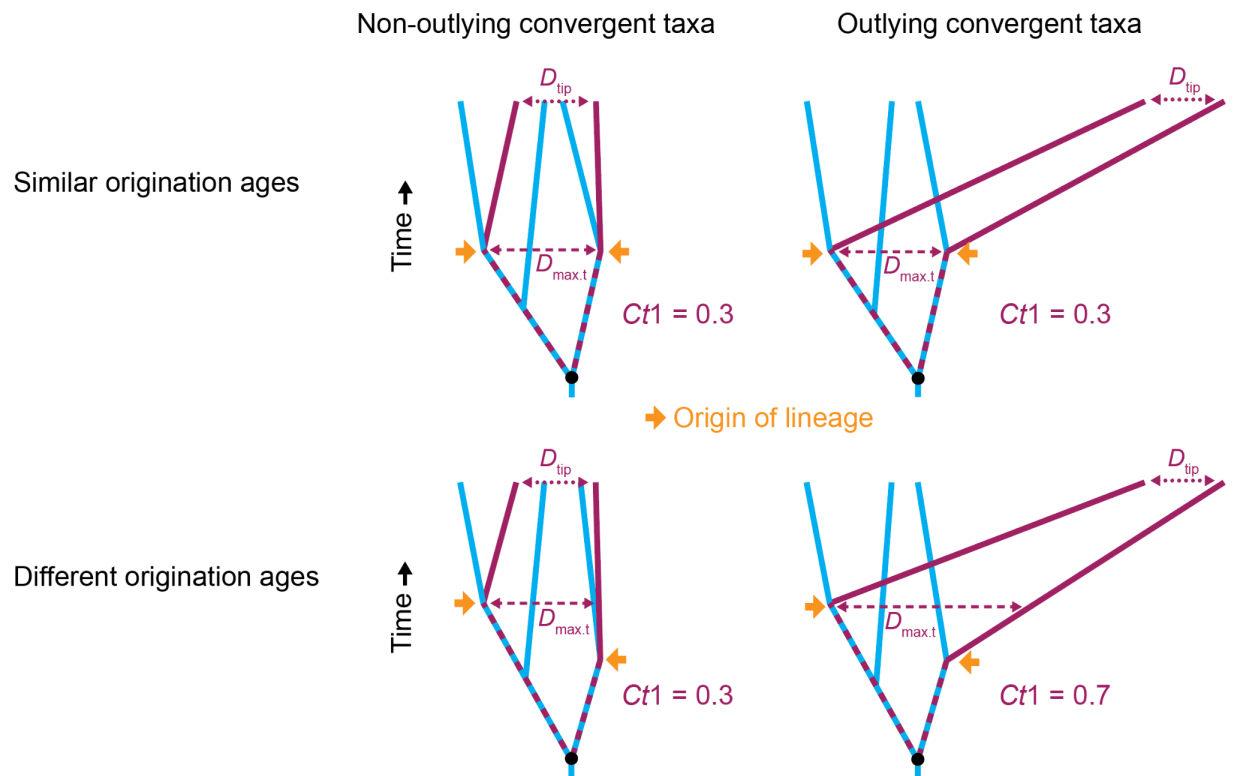
Figure S3. Conceptual illustrations demonstrating how *Ct*1 results can be influenced by a

310   combination of outlying morphologies and varying origination times among convergent

lineages. The *Ct*1 score is 0.3 in three of the scenarios but inflates to 0.7 when lineages

312   both originate at different times and are outliers in morphospace (bottom right). To help

mitigate this issue, we have included an option as part of the *convrat.t* function that allows

314   users to limit candidate $D_{max.t}$ measurement to the time period prior to the origination of the

focal lineages (see Supplemental Methods). See the main text for descriptions of *Ct*1, $D_{max.t}$,

316   and D_{tip}.

318

320

**Influence of the number of traits on *Ct* results**

As discussed in the main text (see Results & Discussion), the number of traits used in analyses (with all else equal) can bias the *Ct* scores. Inference of ancestral states via BM tends to average variation at internal nodes; thus, D_{tip} typically increases at a higher rate than $D_{max.t}$ for each non-convergent trait that is added to a dataset. (Here, we use "non-convergent traits" to refer to BM-evolved traits that are not selected to evolve toward a trait optimum via an OU process. These are often divergent, although it should be noted that BM-evolved traits could still be convergent by chance.) This is illustrated in Figure S4. The effect of this pattern is that an increased number of traits in analyses (with all else equal) could result in a relative decrease in *Ct* scores, unless those added traits are strongly convergent.

334
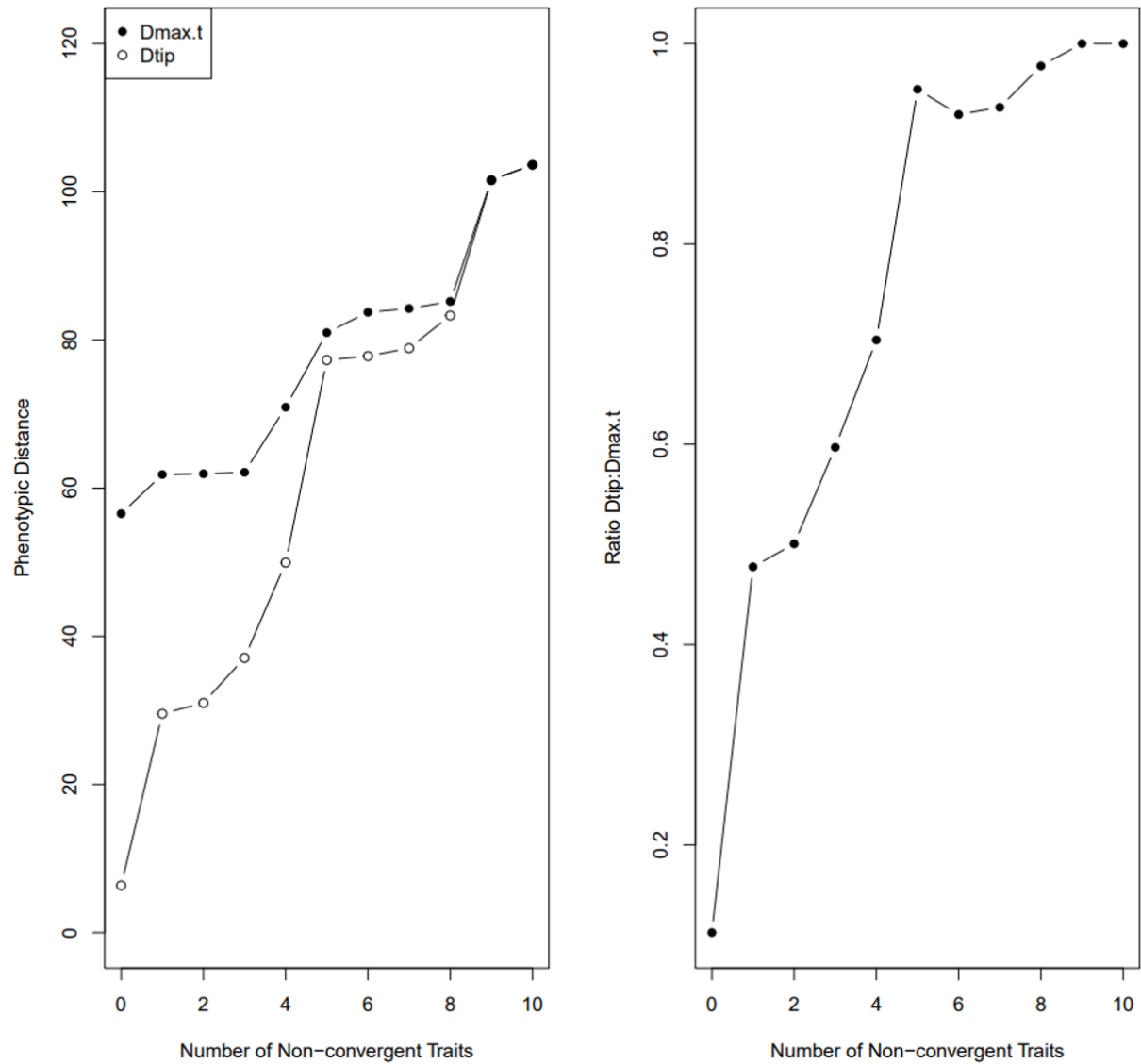
Figure S4. Illustration of how the number of traits used in analyses can influence *Ct*-
336   measures, demonstrating the increased rate at which D_{tip} values increase relative to $D_{max.t}$
as additional non-convergent traits are included in analyses. (Here, 'non-convergent traits'
338   refers to BM-evolved traits, which are expected to be divergent in most cases.) The left
panel shows D_{tip} and $D_{max.t}$ measured between two 'glider' lineages with two simulated
340   convergent traits (optimum = 100) and varying number of additional traits simulated via BM.
The right panel shows the ratio between the D_{tip} and $D_{max.t}$ values.

342

344

346

Empirical example - *Anolis* 'twig' ecomorphotype

348   To test the novel *Ct*-measures and compare *Ct* results to those of *C*-measures (see the *Empirical examples* subsection of the Results & Discussion), we re-analyzed a classic

350   example of convergence among *Anolis* lizards (Mahler et al. 2013), focusing specifically on five 'twig' ecomorphotype lineages. We chose this ecomorphotype because the taxa

352   are morphological outliers that occupy a unique region of *Anolis* morphospace (Huie et al. 2021), and they have especially strong *C*-measure scores (Stayton 2015, Huie et al.

354   2021), although we believe that this is due in part to the lineages being morphological outliers (see Results & Discussion). Following the methods of Mahler et al. (2013), we

356   size-corrected the traits via PGLS regression of each trait against the snout-to-vent length via PGLS. The *Ct*-measure results for this analysis are provided in Figure S5 and

358   Table S2. Whereas the *C*1 score is 0.36 (Stayton 2015), but we find the overall *Ct*1 score to be near zero for both the raw and weighted results (Table S2). This helps to

360   highlight the inflated *C*-measure results due to the issues highlighted in the Results & Discussion. However, note that there is considerable diversity in the results among the

362   ten pairwise comparisons; four are strongly statistically significant, whereas some (e.g., *Anolis occultus* and the *A. paternus* clade) show considerable divergence (*Ct*1 = -0.763;

364   Table S2). To highlight the differences between convergent and non-convergent (or not significant convergence) pairwise comparisons, we separate those comparisons in

366   Figure S5. Thus, we recommend that researchers examine and report results for pairwise comparisons whenever examining more than two putatively convergent

368   lineages.

370

372

374
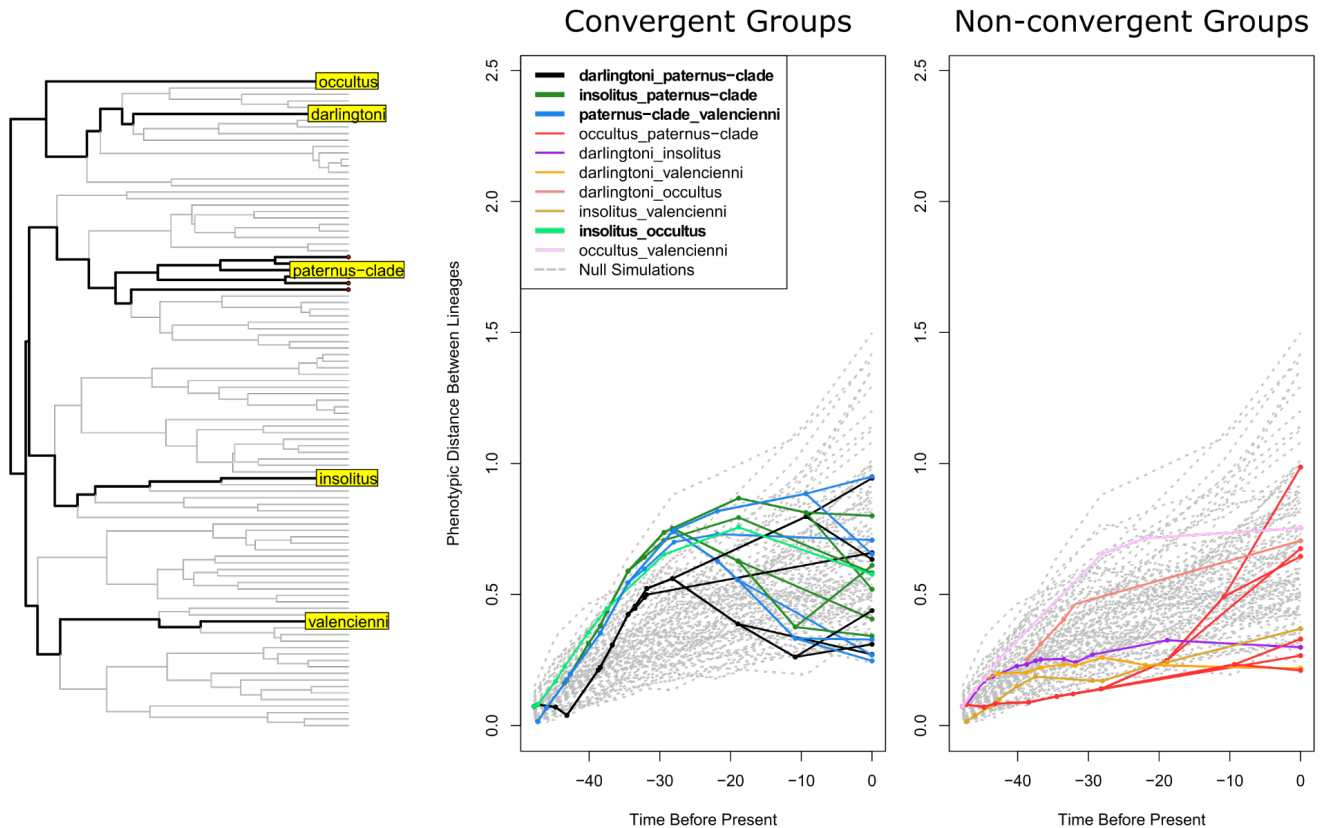
Figure S5. Summary of empirical tests of convergence in *Anolis* species belonging to the 'twig ecomorph' (Mahler et al. 2013). We size-corrected (via PGLS regression) and then analyzed the ten skeletal traits of the dataset of Mahler et al. (2013), with taxa assigned to groups based upon unique origins of the 'twig' ecomorphotype (see the *Ct-measures* section of the Supplemental Methods). The plots are the output of the *plot.C* function of the convevol R package, although the distance-through-time plot has been split to show statistically significant (left) and not significant (right) pairwise comparisons separately (see also Table S2). Significant pairwise comparisons are also indicated in bold in the key. Note that two of the 'non-convergent' comparisons in the right panel do have a positive *Ct*1 value, but they are statistically not significant (Table S2). There are 50 null simulations (light gray lines).

Table S2. *Ct*-measure values obtained for analyses run using the anole dataset of Mahler et al. (2013; ten standardized skeletal traits). Values are reported for overall comparison of ten'twig ecomorph' species in five groups (corresponding to each independent origin of the ecomorph; Fig. S5). Pairwise comparisons of groups are also illustrated in (Fig. S5). See the Supplemental Methods for an explanation of the difference between 'overall raw' and 'overall weighted' results. Note that 'pat' refers to a five-species clade that includes *Anolis paternus* and four closely related specie*s*, whereas all other 'twig' taxa include a single lineage (Fig. S5); see the Methods for updates to the *convevol R* package that allow for comparisons among taxa with more than one lineage. Asterisks denote values returned as significantly different from null simulations (. - p < 0.1, * - p < 0.05, ** - p < 0.01). Abbreviations: *dar, Anolis darlingtoni*; *ins, Anolis insolitus*; *occ, Anolis occultus*; *pat, Anolis paternus*; *val*, *Anolis valencienni*.

	Overall		Pairwise comparisons									
	Raw	**Weighted**	*dar - pat*	*ins - pat*	*pat - val*	*occ - pat*	*dar - ins*	*dar - val*	*dar - occ*	*ins - val*	*ins - occ*	*occ - val*
Ct 1	-0.01**	-0.057**	0.147**	0.323**	0.346**	-0.763	0.083 .	0.161 .	-0.521	-0.527	0.237**	-0.055
Ct 2	0.072**	0.022**	0.086**	0.254**	0.261**	-0.216	0.027 .	0.042 .	-0.241	-0.127	0.179**	-0.039
Ct 3	0.039**	0.012**	0.047**	0.111**	0.140**	-0.090	0.013 .	0.023 .	-0.117 .	-0.063	0.071*	-0.018
Ct 4	0.002**	-0.006 .	0.003**	0.019**	0.011**	-0.007	0.001 .	0.001 .	-0.089	-0.005	0.006**	-0.001

LITERATURE CITED (in the Supporting Information)

Beaulieu, J. M., D. C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

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

Huie, J. M., I. Prates, R. C. Bell, and K. de Queiroz. 2021. Convergent patterns of adaptive radiation between island and mainland *Anolis* lizards. Biol. J. Linn. Soc. Lond. 134:85–110.

Mahler, D. L., T. Ingram, L. J. Revell, and J. B. Losos. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science 341:292–295.

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

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

Stayton C. T. 2018. *Convevol: quantifies and assesses the significance of convergent evolution.*

432   R package version 1.3. https:cran.r-project.org/package=convevol.

Uyeda, J. C., D. S. Caetano, and M. W. Pennell. 2015. Comparative analysis of principal

434   components can be misleading. Syst. Biol. 64:677–689.