

ECOLOGICAL Society of America

Ecology/Ecological Monographs/Ecological Applications

PREPRINT

This preprint is a PDF of a manuscript that has been accepted for publication in an ESA journal. It is the final version that was uploaded and approved by the author(s). While the paper has been through the usual rigorous peer review process of ESA journals, it has not been copy-edited, nor have the graphics and tables been modified for final publication. Also note that the paper may refer to online Appendices and/or Supplements that are not yet available. We have posted this preliminary version of the manuscript online in the interest of making the scientific findings available for distribution and citation as quickly as possible following acceptance. However, readers should be aware that the final, published version will look different from this version and may also have some differences in content.

The doi for this manuscript and the correct format for citing the paper are given at the top of the online (html) abstract.

Once the final published version of this paper is posted online, it will replace the preliminary version at the specified doi.

ECOLOGICAL, SOCIETY OF AMERICA^{1/23}

- 1 Improving phylogenetic regression under complex evolutionary models
- 2 Florent Mazel, T. Jonathan Davies, Damien Georges, Sébastien Lavergne, Wilfried Thuiller and
- 3 Pedro R. Peres-Neto
- 4 <u>Running title:</u> PGLS under complex models of evolution

5 Author affiliations:

- 6 Florent Mazel. Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000
- 7 Grenoble, France ; CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble,
- 8 France, flo.mazel@gmail.com
- 9 Jonathan T. Davies. Department of Biology, McGill University, 1205, Avenue Docteur Penfield,
- 10 Montreal, Quebec, Canada. African Centre for DNA Barcoding, University of Johannesburg, APK
- 11 Campus, PO Box 524, Auckland Park 2006, Johannesburg, South Africa;
- 12 j.davies@mcgill.ca
- 13 Damien Georges. Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000
- 14 Grenoble, France ; CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France.
- 15 damien.georges2@gmail.com
- 16 <u>Sébastien Lavergne.</u> Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000
- 17 Grenoble, France ; CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France
- 18 sebastien.lavergne@ujf-grenoble.fr
- 19 Wilfried Thuiller. Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), F-38000
- 20 Grenoble, France ; CNRS, Laboratoire d'Écologie Alpine (LECA), F-38000 Grenoble, France
- 21 wilfried.thuiller@ujf-grenoble.fr
- 22 Pedro R. Peres-Neto. Canada Research Chair in Spatial Modelling and Biodiversity,
- 23 Département des sciences biologiques, Université du Québec à Montréal, Montréal, QC H3C 3P8,
- 24 Canada; peres-neto.pedro@uqam.ca
- 25 Abstract

ECOLOGICAL SOCIETY OF AMERICA^{2/23}

26 Phylogenetic Generalised Least Square (PGLS) is the tool of choice among phylogenetic 27 comparative methods to measure the correlation between species features such as morphological 28 and life-history traits or niche characteristics. In its usual form it assumes that the residual 29 variation follows a homogenous model of evolution across the branches of the phylogenetic tree. 30 Since a homogenous model of evolution is unlikely to be realistic in nature, we explored the 31 robustness of the phylogenetic regression when this assumption is violated. We did so by 32 simulating a set of traits under various heterogeneous models of evolution, and evaluating the 33 statistical performance (type I error -the % of tests based on samples that incorrectly rejected a 34 true null hypothesis- and power -the % of tests that correctly rejected a false null hypothesis) of 35 classical phylogenetic regression. We found that PGLS has good power but unacceptable type I 36 error rates. This finding is important since this method has been increasingly used in comparative 37 analyses over the last decade. To address this issue, we propose a simple solution based on 38 transforming the underlying variance-covariance matrix to adjust for model heterogeneity within 39 PGLS. We suggest that heterogeneous rates of evolution might be particularly prevalent in large 40 phylogenetic trees, whilst most current approaches assume a homogenous rate of evolution. Our 41 analysis demonstrates that overlooking rate heterogeneity can result in inflated type I errors, thus 42 misleading comparative analyses. We show that it is possible to correct for this bias even when the 43 underlying model of evolution is not known a priori.

44

45 <u>Keywords</u>. Comparative methods, PGLS, non-stationarity, statistical performance, Phylogenetic
46 Generalised Least Square.

47

48 Introduction

49 Comparative methods are among the key tools for understanding ecological and evolutionary 50 processes (Felsenstein 1985; Harvey & Pagel 1991) and are used to test hypotheses about the

ECOLOGICAL SOCIETY OF AMERICA^{3/23}

51 correlated evolution of traits (e.g. Pearman *et al.* 2014). Since species share common ancestry, 52 they should not be considered statistically independent units, thus traditional statistical methods 53 such as Ordinary Least Square (OLS) regression are not appropriate for analyzing comparative 54 data. When analyzed by OLS, the two major issues that arise from shared evolution are increased 55 type I error when traits are uncorrelated with each other and reduced precision in parameter 56 estimation when traits are correlated with each other (Revel 2010). Therefore, interspecific trait 57 data should be analysed within a phylogenetic framework (Harvey and Pagel 1991, Freckleton et 58 al. 2002, Revell 2010).

59 Multiple methods have been proposed to account for phylogenetic non-independence of species 60 when regressing two (or more) continuous or categorical traits (Felsenstein 1985, Grafen 1989, 61 Maddison 1990, Lynch 1991, Garland et al. 1992, Martins and Hansen 1997, Diniz-Filho et al. 62 1998, Freckleton et al. 2002, Paradis and Claude 2002, Ives and Garland 2010). The phylogenetic 63 regression based on a generalised least square, where the inverse of the phylogenetic covariance 64 matrix is used as weights, is perhaps now the most widely adopted (Grafen 1989; Martins & 65 Hansen 1997): it is a generalisation of phylogenetic independent contrasts (Rohlf 2001) originally 66 proposed by Felsenstein (1985) and a particular case of general linear models (Rencher & 67 Schaalje, 2007). Phylogenetic regression assumes that the model residual error ε is distributed 68 according to $\sigma_{\varepsilon}^2 C$ where and σ_{ε}^2 represents the residual variance and C is a n × n matrix (n is the 69 number of tips, i.e., species in most cases) describing the evolutionary relationships among species 70 (i.e., a phylogenetic covariance matrix with diagonal elements estimated as the total branch length 71 between each tip and the root, and off-diagonal elements estimated as the evolutionary time shared 72 by each species pair).

73 In its simplest form, PGLS assumes a Brownian Motion (BM, Edwards & Cavalli-Sforza 1964) 74 model of evolution with a single rate, σ^2 . Nevertheless, PGLS is highly flexible and it can be 75 extended to alternative evolutionary models (Martins and Hansen 1997). For example, recent

ECOLOGICAL SOCIETY OF AMERICA^{4/23}

76 PGLS implementations have incorporated tree transformation models that capture different modes
77 of evolution (e.g. early vs. late trait diversification, continuous vs. punctual evolution, Pagel 1997,
78 1999; Freckleton *et al.* 2002; Revell 2010) or selective regimes (e.g. Ornstein–Uhlenbeck [OU]
79 models, Hansen 1997; Butler & Kings 2004; Lavin *et al.* 2008). This model flexibility has been
80 key in reducing type I errors due to model misspecification, one of major issues in comparative
81 biology (Freckleton 2009).

82 Despite recent advances, however, current PGLS implementations still assume that the tempo and 83 mode of evolution remain constant across the phylogenetic tree (although they allow for rate 84 variation over time), whereas it is likely that both are highly heterogeneous (Simpson 1944, Gould 85 2002), particularly in the case of large phylogenetic trees (O'Meara 2012). The construction of 86 very large (Jetz et al. 2012) hand in hand with the increased availability of corresponding large 87 trait datasets (e.g. Wilman et al. 2014) has generated an increasing need to consider more complex 88 models of evolution within the regression framework. Heterogeneous trait evolution, where trait 89 evolution has been markedly different across multiple clades is a potential source of bias that has 90 been largely overlooked. If two traits evolved under complex models of evolution, standard PGLS 91 (i.e. assuming a single rate of evolution) may not be appropriate. Since it has been demonstrated 92 that an incorrectly defined VCV matrix in PGLS increases the type I error rate (Revell 2010) for 93 simple homogeneous models, it naturally follows that for large comparative phylogenetic datasets, 94 where evolutionary processes are likely heterogeneous, there will also be an increase in type I 95 error rates and/or reduced statistical power.

96 One potential solution arises from the development of heterogeneous models of evolution, which 97 allow the fit of highly complex VCV matrices. Variation in evolutionary rates across the 98 phylogenetic tree can be modeled, for instance, with a heterogeneous BM process where σ^2 varies 99 across the phylogenetic tree (O'Meara et al. 2006, Thomas et al. 2006). Similarly, heterogeneous 100 OU models with multiple optima, strength of selection and evolutionary rates have be proposed to

ECOLOGICAL SOCIETY OF AMERICA^{5/23}

101 model adaptive peaks (Beaulieu 2013; Ingram & Mahler, 2013). These models have been 102 successfully applied in the literature to investigate how a single trait has evolved across the 103 phylogeny (e.g. to study the evolution of a clade life form, Adams *et al.* 2009; Boucher *et al.* 104 2012) but they are not yet incorporated into the toolbox of comparative analyses of trait 105 correlation. Transforming the phylogenetic tree according to a heterogeneous model of trait 106 evolution fitted with the data and using this transformed tree to derive a new VCV matrix could 107 help in increasing the flexibility of the PGLS framework, but this potential has remained 108 unexplored.

109 Here using simulations, we first explore the performance (type I error rate and statistical power) of 110 PGLS under models of heterogeneous trait evolution. We simulate multiple models of trait 111 evolution and cross correlations among traits, and contrast power and type I error rates. We show 112 that complex models of evolution lead to inflated type I error rates, but PGLS is able to handle 113 such complexities when the correct VCV matrices is known. We then propose an implementation 114 of PGLS that has valid type I error rates even under models with large rate heterogeneity where 115 the underlying model of evolution is not known *a priori*

116

117 Methods & Results

118 Simulating traits under complex evolutionary models

119 Simulation model

120 Our simulations considered two traits (X and Y) generated by the following basic equation, setting 121 *a* to zero (following Revel 2010):

122 Y = $a + \beta X + \epsilon$ (Eqn. 1)

123 We defined evolutionary models by manipulating the phylogenetic covariance structure and rates 124 of evolution (see below) and simulate X and the residuals error ε assuming normally distributed 125 values N(0, σ^2); where σ^2 represents the instantaneous rate of evolution that was set depending on

ECOLOGICAL SOCIETY OF AMERICA^{6/23}

126 the specific scenarios of homogenous (one single evolutionary rate) and heterogeneous (multiple 127 changes in evolutionary rates) models of trait evolution (see below). Type I error for a given 128 method was assessed by simulating data with $\beta=0$ while statistical power was assessed by 129 simulating data with $\beta=1$ (Appendix A). In both cases, we tested the null hypothesis H₀: $\beta=0$ and 130 reported the percentage of simulations in which H₀ was rejected with an alpha level of 0.05 (i.e., 131 type I error rate as number of rejections when $\beta=0$ and statistical power as the number of 132 rejections when $\beta=1$). For $\beta=0$, Y= ϵ and the evolution of X and Y are simulated independently. 133 For $\beta=1$, the evolutionary model for Y was a function of the two evolutionary processes 134 generating X and ϵ . We consider three different scenarios representing different evolutionary 135 models for X and ϵ (Appendix A). We generated X and ϵ under (1) the same evolutionary model, 136 (2) different models or (3) assuming that only ϵ followed an evolutionary model while X was 137 drawn from a normal distribution (with mean of 0 and a standard deviation of one).

138 Phylogenetic trees

139 To make sure our results were not just representative of a particular phylogenetic topology, we ran 140 all analyses on two very different phylogenetic topologies of 128 species each (rescaled so that 141 their total depth equaled one). One tree was completely balanced, whereas the second one was 142 obtained using a pure birth process leading to a more realistic unbalanced tree (see Appendix B). 143 *Homogeneous models of trait evolution*

144 We considered three classic models (Brownian motion (BM, Edwards & Cavalli-Sforza 1964), 145 Ornstein–Uhlenbeck (OU, Hansen 1997) and the lambda (λ) tree transformation (Pagel 1999)). 146 For BM, the change in species traits over time was expressed as:

147 $dX(t) = \sigma dB(t)$ (Eq. 2)

148 where dX(t) is the change in trait X over time period dt. The parameter σ measures the rate of 149 evolution, while the term B(t) is random noise $\sim N(0, dt)$.

150 For an OU process, the change in species traits over time was expressed as:

ECOLOGICAL SOCIETY OF AMERICA^{1/23}

151 $dX(t) = \alpha[\theta - X(t)]dt + \sigma dB(t)$ (Eq. 3)

152 where θ represents the mean trait value (often interpreted as the trait optimum) and alpha measures 153 the rate of decay of trait similarity through time (often interpreted as the intensity of stabilizing 154 selection). When $\alpha = 0$, the OU model simplifies to a BM model (see Eq. 2).

155 The λ tree transformation model simply rescales the phylogenetic tree before applying the classic 156 BM model. In our case, λ is the multiplier of internal branches and we considered values between 157 1 (no transformation: the trait evolved under a classical BM) and 0 (the tree is a star phylogeny 158 and the trait has no 'phylogenetic signal', i.e. related species do not tend to share similar trait 159 value). Continuous characters (representing X and ε) under each model had a starting value of 0 at 160 the root of the tree and were evolved tipward according to each model.

161 Heterogeneous models of trait evolution

162 For heterogeneous BM models, we simulated traits with two different rates of evolution occurring 163 in different parts of the phylogenetic tree. In the simplest case, one sub-clade evolved with a rate 164 σ^2 =1 while the other sub-clade evolved with one of the following σ^2 : 1/1000, 1/100, 1/10, 1/4, 1/2 165 or 3/4. In this case we simulated a single rate shift near the root of the phylogenetic trees so that 166 the two major sub-clades of the phylogeny evolved under different σ^2 (see Appendix B). We also 167 generated traits evolving under multiple rate shifts (3, 5 and 9) within each of the two major 168 clades, but we kept the total number of rate values restricted to two (see Appendix B). 169 For heterogeneous OU models, we simulated optima shifts and kept σ^2 and α constant at 1 and 0.5, 170 respectively. The initial OU regime started with an optimal value of θ (either 1, 2, 3 or 4) and then 171 shifts to $-\theta$ at the same node as described for BM rate shifts above (see Appendix B). 172 For heterogeneous λ -transformed models, we simulated a single shift occurring near the root of the 173 tree (i.e. separating two major clades A and B, see Appendix C) by multiplying the internal branch 174 length of clade A by λ , keeping the internal branches lengths of clade B unchanged. The resulting

175 sub-clades therefore differed in their root-to-tip distances and tip-to-internal branch length ratios.

ECOLOGICAL, SOCIETY OF AMERICA^{8/23}

176 If we had used a single σ^2 for this transformed tree, the trait evolution of the two sub-clades would 177 thus have different rates and phylogenetic signal. Because we were interested here in the 178 differences in phylogenetic signal only, we rescaled the transformed branches of clade A so that 179 all species in the complete tree had the same root-to-tip distance (same overall evolutionary rate) 180 but differed in their ratio of tip-to-internal branch length (see Appendix C). For λ , we restricted 181 our analysis to the balanced tree (see above) because it is not straightforward to retain tree 182 ultrametricity for more complex tree topologies, potentially confounding comparisons between our 183 alternative models. As for the homogenously evolved traits, traits had a starting value of 0 at the 184 root of the tree and were evolved tipward according to each model and evolutionary rates σ^2 that 185 were clade dependent.

186

187 **Result 1:** assessing power and type 1 error rates of the phylogenetic regression. 188 We fitted two classical linear regression models to each of the different simulated datasets 189 (scenarios). We first fitted an OLS regression, for which we tested the significance of the slope 190 with a t-test using n-2=126 d.f. Note that a t-test was used given its common usage in the 191 comparative analysis literature, though a likelihood test contrasting the slope and the intercept-192 only models could had been equally applied. Implementation of the latter could be the object of 193 future studies. Second we fitted a PGLS (using the *pgls* function in the *caper* R package) that also 194 simultaneously optimized a single λ value for the residuals of the model (λ="ML" in the *pgls* 195 function, 'PGLS_{global_λ}' hereafter). The λ parameter (Pagel 1999) is a multiplier for the off-196 diagonal elements (i.e. the internal phylogenetic branches) of the VCV matrix and usually varies 197 between 0 and 1. If λ =1, the VCV is left unchanged and the PGLS assumes a BM while if λ =0, the 198 VCV is a diagonal matrix and the PGLS reduces to an OLS. Any value between 0 and 1 indicates 199 that the phylogenetic strength in the residuals is reduced in contrast to a BM. As such, the 200 optimization of λ in PGLS allows encompassing both classical OLS and PGLS. As for OLS, we

ECOLOGICAL SOCIETY OF AMERICA^{9/23}

201 then tested the significance of the slope with a t-test with n-2-1=125 d.f. (λ is an additional 202 parameter).

203 Power analysis

204 First, we simulated X and ε independently from each other to generate Y (Y = β X + ε with β =1). 205 Under this scenario, both OLS and PGLS_{global_ $\lambda}} had good statistical power (i.e. they both detected$ 206 the simulated correlation) for all models of evolution tested (Table 1 and Appendix D).</sub>

207 Type 1 error analysis

208 Second, we simulated Y and X independently from each other ($Y = \beta X + \varepsilon$ with $\beta = 0$, so $\varepsilon = Y$), and 209 evaluated the percentage of simulations where a correlation was (incorrectly) detected (i.e. false 210 positives). The type I error of a valid test should equal the alpha value selected when assessing test 211 significance (5% here). Results differed across the three simulated scenarios (Table 1 and see 212 Appendix E-F): (1) When X was simulated without phylogenetic signal and Y followed different 213 models of trait evolution (i.e. heterogeneous BM and OU models), both OLS and PGLS had 214 correct type I error rates (Fig. F1 in Appendix F); (2) When X and Y followed different models of 215 trait evolution (e.g. X followed a heterogeneous BM model and Y followed a heterogeneous OU 216 model), OLS showed inflated type I error while PGLS_{global λ} still performed well (i.e., it had 217 correct type I error rates, Fig. F2 in Appendix F); (3) When X and Y followed the same 218 heterogeneous model of trait evolution results were more mixed. When we simulated BM with 219 heterogeneous Pagel λ , all methods had an inflated type I error, which covaried with the 220 heterogeneity in the strength of the phylogenetic signal (see Figure 1). When we simulated BM 221 models with heterogeneous rates of evolution (different σ^2 across the tree), type I error rates were 222 also inflated, and varied with the strength of the evolutionary rate variation, but were only weakly 223 influenced by the number of rate shifts (Fig. 2). We did not evaluate type I error rates under a 224 heterogeneous OU model of evolution for reasons described in the Discussion section PGLS and 225 hidden selective trends. Unbalanced and balanced tree gave qualitatively similar results (see Fig. 2

226 and Fig. F3 in Appendix F, respectively), for brevity we report only values for the unbalanced tree 227 in the main text.

228 Providing the correct VCV to PGLS led to correct type I error rates in all cases (Fig. 2 and Fig. F3 229 in Appendix F). However, this is obviously not a viable solution for most empirical studies since 230 the true VCV is not known *a priori*, but it nonetheless shows that PGLS is able to deal with 231 complex models of evolution when they are correctly estimated.

232

233 Result 2: a solution to correct for inflated type 1 error

234 Our simulations showed that PGLS had inflated type I error rates under heterogeneous BM but
235 that it could theoretically handle such models when a correct VCV was provided.
236 To correct for inflated type I errors when the correct VCV is not known, we studied the statistical
237 performance of the following three step procedure: 1) fit heterogeneous BM models of trait
238 evolution to the raw OLS residuals; 2) use this fit to modify the VCV matrix used in the standard
239 PGLS; and 3) apply a significance criterion (see *Modified method for significance testing* below)
240 that allows for proper inference (i.e., correct type I error) associated to the two initial fits that
241 themselves involve one statistical test each. For clarity, note that the correction in step 3 is
242 completely independent from the issues of inflated type I errors in comparative analysis involving
243 multiple rates of evolution (the focus of this paper). However, because our proposed framework is
244 based on a two-step procedure, one would incur additional type I errors if a multiple testing
245 criterion was not applied.

246 We tested two approaches to detect rate shifts on the OLS residuals: the *auteur* bayesian approach 247 (Eastman *et al.* 2011, implemented in the R package *geiger*) and the *trait medusa* approach 248 (Thomas and Freckleton 2012, implemented in the R package *motmot*). As *auteur* and *trait* 249 *medusa* yielded similar results but *auteur* being much faster, we present in the main text the 250 detailed procedure and results for *auteur* (PGLS_{auteur} hereafter). The details and results of the

251 procedure involving *trait medusa* are provided in Appendix G. We provide R code and example 252 data of both procedures as a supplement (Supplements 1 & 2).

253 Detecting evolutionary rate shifts with auteur.

254 *auteur* (Eastman et al. 2011) is a recently developed Bayesian approach to model evolutionary rate 255 heterogeneity along a phylogeny. It uses a reversible jump Markov chain Monte Carlo procedure 256 to sample a distribution of multi-rate BM models in inverse proportion to their poorness of fit. In 257 the optimization procedure, the Markov chain jumps from models of differing complexities (i.e. 258 number and position of the shifts in the phylogenetic tree).

259 We sampled parameters every 500 generations for a total of 20,000 generations. We removed 25% 260 of the sampled parameters as burn-in to obtain marginalized distributions of relative rates for each 261 branch of the tree.

262 Using fitted multi-rate models in standard PGLS

263 We then used the parameter estimates from *auteur* to rescale the phylogenetic tree using the 264 function *rescale* in the *geiger* package. The rescaled tree was then used in a standard PGLS 265 procedure (PGLS_{auteur}, function *pgls* in the R package *caper*) as described above. We then tested 266 the significance of the slope with a t-test with n-2 d.f. (PGLS_{auteur}, see Fig. 3). As described above, 267 this stepwise procedure may be expected to have a slight inflated type I error rate because it 268 represents a two-step procedure (Fig. 3), and each step has independent errors (the same reasoning 269 applies to the PGLS_{trait medusa} procedure, Appendix G). Step 1 estimates the evolutionary model of 270 the OLS residuals, and step 2 fits PGLS using the estimated VCV matrix in step 1. As before, 271 because two statistical tests were involved here, we applied a modified significance testing 272 criterion.

273 Modified method for significance testing

274 Statistical testing when using a two-step procedure is likely to have inflated incorrect family wise 275 type I error. This has been previously recognized by ter Braak *et al.* (2012) when testing for the

ECOLOGICAL, SOCIETY OF AMERICA^{12/23}

276 links between trait to environmental variation. Their procedure was based on a two-step procedure 277 in which the links between species trait variation and species distributions, and the links between 278 environmental variation and species distributions were both tested. They adopted a rejection 279 criterion for establishing the significance of the links between environment and trait variation in 280 which the largest probability of the two tests involved needed to be below the pre-established 281 alpha-level (i.e., 0.05). Retaining the largest p-value is equivalent to conducting only one 282 statistical test instead of two, assuring correct type I errors. Following ter Braak *et al.* (2012), we 283 thus considered the relationship between Y and X (PGLS_{combination}) and retained the largest p-value 284 between the PGLS_{global_ λ} and PGLS_{auteur}. This approach produced correct type I error rate in all 285 cases (Fig. 3).

286

287 Discussion

288 With the increasingly availability of well resolved phylogenetic trees, PGLS has become routinely 289 employed in the analysis of interspecific data over the past decades (Felsenstein 1985, Grafen 290 1989, Martins and Hansen 1997, Freckleton et al. 2002). By assuming an explicit model of 291 evolution, PGLS contrasts with some other approaches, for example, non-parametric eigenvector 292 decomposition (Diniz-Filho et al. 1998, Freckleton et al. 2011). While it has been argued that the 293 inclusion of an explicit evolutionary model within PGLS allows for increased efficiency of 294 estimation, decreased variance in parameter estimates, and decreased Type I errors (Freckleton *et* 295 *al.* 2011), misspecifying the evolutionary model may have important consequences for hypotheses 296 testing (Revell 2010) (though not for parameter estimation, see e.g. Rohlf 2006). One solution is 297 to simultaneously estimate the parameters of the PGLS model and the model of evolution of the 298 residuals. For some simple models of trait evolution, in which both the tempo and mode of 299 evolution remain constant across the phylogenetic tree, it is possible to adjust the PGLS model 300 residuals using Pagel's (1999) lambda tree transformations (Freckleton 2002). However, such

301 simple models of evolution are likely rare, particularly for large trees (O'Meara 2012), which have 302 become widely used in the comparative literature with the increasingly availability of large scale 303 mega-phylogenies including several thousand species. Here we evaluated the performance (power 304 and type I errors) of PGLS methods, under more complex evolutionary scenarios. We show that 305 most developed methods around PGLS have good power, but unacceptably high type I errors 306 under some scenarios with heterogeneous evolutionary rates. Nonetheless, PGLS methods perform 307 well when the VCV matrix is estimated properly.

308 X and Y follow different models of evolution

309 When there is no phylogenetic signal in the independent variable (i.e. X was normally distributed 310 and independent from phylogeny) but the residuals (of Y) follow a heterogeneous model of 311 evolution (either BM or OU), all methods (OLS and PGLS) showed correct type I error rate. 312 Similar results have been reported for homogeneous BM models (see Revell 2010). However, 313 when we simulated a heterogeneous BM in the independent variable (X) and the dependent 314 variable (Y) followed a heterogeneous OU, OLS shows inflated type I error rates, but PGLS still 315 performs well.

316 *X* and *Y* follow the same model of evolution

317 When we simulated X and Y with the same heterogeneous model of evolution all classical
318 methods (OLS and PGLS) showed inflated type I error rates. Nevertheless PGLS is theoretically
319 able to handle such bias (Martins and Hansen 1997), and we demonstrated empirically that
320 providing the correct VCV leads to valid test of correlated evolution with appropriate type I error
321 rates.

322 Because knowing the correct VCV transformation *a priori* is difficult for traits evolving under 323 complex evolutionary models, we implemented a simple approach that allowed us to estimate the 324 model of evolution on the residuals of the OLS, and used this to transform the VCV matrix for 325 PGLS. Here we used two methods (a Bayesian approach and a bootstrap approach in conjunction

326 with an algorithm for estimating multiple rate shifts) to identify the appropriate model for 327 transforming the VCV matrix, though we acknowledge that alternative methods are available (see 328 Revell *et al.* 2012 for an example). We show that assuming the transformed VCV matrix, PGLS 329 had appropriate type I errors after correcting for the two-step model selection procedure (ter Braak 330 *et al.* 2012). However, we caution that one of the methods used is sensitive to overfitting (a known 331 problem with the medusa algorithm assuming the AICc stopping criterion) and that a wrongly 332 defined VCV matrix can lead to type I error rates as high as for raw OLS models. It is crucial, 333 therefore, to use a methodology that correctly assigns rate shifts, or to test overfitting by 334 parametric bootstraping (Boettiger et al. 2012).

335 It is possible that more efficient strategies will become soon available for fitting PGLS under 336 heterogeneous models of evolution, nonetheless we have shown that our approach works well, and 337 is reasonably transparent. Due to computational constrains, it was not feasible to explore the full 338 range of possible evolutionary models, tree shapes, and trees sizes (the combinations of which are 339 effectively infinite), but we see no reason why our approach should fail under different conditions 340 except when clades sizes for which rate changes have occurred are too small to accurately infer 341 the correct evolutionary model.

342 Last, we note that we do not provide a solution for the heterogeneous phylogenetic signal343 scenario; however, future advances that allow the fit of heterogeneous phylogenetic signal to the344 tree could be easily implemented within our framework.

345 PGLS and hidden selective trends

346 We did not evaluate type I error rates under a heterogeneous OU model of evolution. Under this 347 scenario, even for B=0 (i.e. no expected correlation between X and Y), the two traits might be 348 significantly correlated if both have followed the same selective trends (i.e. optima and selection 349 strength shifts). In this case, we are no longer considering type I error rates, but rather Power of 350 the method and it is not straightforward to derive expectations if the two traits follow slightly

351 different models. For example, one could imagine that the two traits show the same general trends 352 but with slightly different optima, strength of selection or location of shifts. One interesting idea 353 would be to simulate traits with different OU parameters and positions, and plot the percentage of 354 detected correlations against the amount of difference between the two models using a version of 355 PGLS that takes into account the drift part of the OU model. It might also be informative to 356 explore models where traits co-evolve, or the evolution of one is driven by the evolution of the 357 other, causing consistent time-lags between their evolutionary shifts. It would then be possible to 358 compare the likelihood of such a model with another in which traits evolve independently. This 359 would avoid the direct correlation of extant species traits. Finally, in the present study we have 360 assumed a parametric method for estimating changes in evolutionary rates in the sense that the 361 procedures here estimate variance-covariance phylogenetic matrices based on families of known 362 models of evolution. Another potential solution, not explored here, would be to use an iterative 363 weighted least-square implementation (e.g., Björck 1996) in order to estimate an appropriate 364 variance-covariance structure that would make model residuals independent.

365 Taken together, we have shown that currently implemented phylogenetic comparative methods 366 have unacceptable type I error rate when species' traits evolve under heterogeneous models of 367 evolution. We proposed a flexible solution based on PGLS and showed that it had correct type I 368 error. Our framework is potentially extendible to most complex evolutionary scenarios.

369 Acknowledgements:

370 FM would like to thank Rafael Wüest for stimulating discussions on the PGLS approach. The 371 research leading to these results had received funding from the European Research Council under 372 the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 373 281422 (TEEMBIO). FM, DG, SL and WT belong to the Laboratoire d'Écologie Alpine, which is 374 part of Labex OSUG@2020 (ANR10 LABX56). All computations performed using the CiGRI 375 platform of the CIMENT infrastructure (https://ciment.ujf-grenoble.fr), which is supported by the

376 Rhône-Alpes region (GRANT CPER07_13 CIRA) and France-Grille (<u>http://www.france-grilles.fr</u>).
377 TJD and PRPN were supported by a team grant from the Fonds de recherche du Québec – Nature et
378 technologies.

379 References

- 380 Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species
- diversification correlated with rates of morphological evolution? Proceedings of the
 Royal Society of London, Series B 276:2729–2738.
- Björck, Å. 1996. Numerical Methods for Least Squares Problems. SIAM. Philadeplphia, PA.
- Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? Measuring the

power of comparative methods. Evolution 66:2240–2251.

- Boucher, F. C., W. Thuiller, C. Roquet, R. Douzet, S. Aubert, N. Alvarez, and S. Lavergne.
- 2012. Reconstructing the origins of high-alpine niches and cushion life form in the genus
 Androsace sl (Primulaceae). Evolution 66:1255–1268.
- 389 Ter Braak, C., A. Cormont, and S. Dray. 2012. Improved testing of species traits-environment
- relationships in the fourth-corner problem. Ecology 93:1525–1526.
- 391 Butler, M. A., and A. A. Kings. 2004. Phylogenetic comparative analysis: A modeling

approach for adaptive evolution. The American naturalist 164:683–695.

- 393 Diniz-Filho, J. A. F., C. E. R. de Sant'Ana, and L. M. Bini. 1998. An Eigenvector Method for
- 394 Estimating Phylogenetic Inertia. Evolution 52:1247–1262.

ECOLOGICAL, SOCIETY OF AMERICA^{17/23}

395	Eastman.	J. M.,	M.E.	Alfaro	P. Joy	vce, A. l	L. Hipp.	and L. J.	Harmon.	2011. A	novel
						/					

396 comparative method for identifying shifts in the rate of character evolution on trees.

397 Evolution; international journal of organic evolution 65:3578–89.

- 398 Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees.
- Pages 67–76 *in* V. H. Heywood and J. McNeill, editors. Phenetic and phylogenetic
- 400 classification. London.
- Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist:1–
 15.
- Freckleton, R. 2009. The seven deadly sins of comparative analysis. Journal of evolutionary
 biology 22:1367–1375.
- Freckleton, R. P., N. Cooper, and W. Jetz. 2011. Comparative methods as a statistical fix: the
 dangers of ignoring an evolutionary model. The American naturalist 178:E10–E17.
- 407 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
- data: a test and review of evidence. The American Naturalist 160:712–726.
- Garland, T., P. Harvey, and A. Ives. 1992. Procedures for the analysis of comparative data
 using phylogenetically independent contrasts. Systematic Biology 41:18–32.
- 411 Gould, S. J. 2002. The Structure of Evolutionary Theory. Harvard University Press.
- Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal
 Society of London B:119–157.
- Hansen, T. F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation.
 Evolution 51:1341.

- 416 Harvey, P., and M. Pagel. 1991. the Comparative Method in Evolutionary Biology. Oxford
- 417 University Press, Oxford, UK.
- 418 Ingram, T., and D. Mahler. 2013. SURFACE: detecting convergent evolution from
- 419 comparative data by fitting Ornstein Uhlenbeck models with stepwise Akaike
- 420 Information Criterion. Methods in Ecology and Evolution 4:416–425.
- 421 Ives, A., and T. Garland. 2010. Phylogenetic logistic regression for binary dependent
 422 variables. Systematic Biology 59:9–26.
- 423 Jetz, W., G. Thomas, J. Joy, and A. Ø. Mooers. 2012. The global diversity of birds in space
- 424 and time. Nature 491:444–448.
- 425 Lavin, S. R., W. H. Karasov, A. R. Ives, K. M. Middleton, and T. Garland. 2008.
- 426 Morphometrics of the avian small intestine compared with that of nonflying mammals: a
- 427 phylogenetic approach. Physiological and biochemical zoology : PBZ 81:526–50.
- 428 Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology.
 429 Evolution 45:1065–1080.
- 430 Maddison, W. P. 1990. A Method for Testing the Correlated Evolution of Two Binary
- 431 Characters: Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic
 432 Tree? Evolution 44:223–539–557.
- Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general
 approach to incorporating phylogenetic information into the analysis of interspecific
- data. The American naturalist 149:646–667.

- 436 O'Meara, B. 2012. Evolutionary inferences from phylogenies: a review of methods. Annual
- 437 Review of Ecology, Evolution, and Systematics 43:267–285.
- 438 O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different
- rates of continuous trait evolution using likelihood. Evolution 60:922–933.
- 440 Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zoologica Scripta
- 441 26:331–348.
- 442 Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.
- 443 Paradis, E., and J. Claude. 2002. Analysis of Comparative Data Using Generalized Estimating

444 Equations. Journal of Theoretical Biology 218:175–185.

- 445 Pearman, P., S. Lavergne, C. Roquet, R. Wuest, N. E. Zimmermann, and W. Thuiller. 2014.
- 446 Phylogenetic patterns of climatic, habitat and trophic niches in a European avian

447 assemblage. Global Ecology and Biogeography 23:414–424.

- 448 Rencher, A. C., and G. B. Schaalje. 2007. Linear Models in Statistics. John Wiley & Sons,
- 449 Inc., Hoboken, NJ, USA.
- 450 Revell, L. 2010. Phylogenetic signal and linear regression on species data. Methods in
 451 Ecology and Evolution 1:319–329.
- 452 Revell, L., D. Mahler, P. Peres-Neto, and B. Redelings. 2012. A new phylogenetic method for
- 453 identifying exceptional phenotypic diversification. Evolution 66:135–146.
- 454 Rohlf, F. 2006. A comment on phylogenetic correction. Evolution 60:1509–1515.

- 455 Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric
- 456 interpretations. Evolution 55:2143–2160.
- 457 Simpson, G. G. 1944. Tempo and Mode in Evolution. Columbia University Press.
- 458 Thomas, G. H., and R. P. Freckleton. 2012. MOTMOT: models of trait macroevolution on
- trees. Methods in Ecology and Evolution 3:145–151.
- 460 Thomas, G. H., R. P. Freckleton, and T. Székely. 2006. Comparative analyses of the influence
- 461 of developmental mode on phenotypic diversification rates in shorebirds. Proceedings.
- Biological sciences / The Royal Society 273:1619–24.
- 463 Wilman, H., J. Belmaker, J. Simpson, C. de la Rosa, M. M. Rivadeneira, and W. Jetz. 2014.
- EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals.
- 465 Ecology 95:2027–2027.
- 466 List of Supplementary material
- 467 Appendix A. Simulation workflow
- 468 Appendix B. Phylogenetic trees and associated regime shifts
- 469 Appendix C. λ tree transformation
- 470 Appendix D. Power Analysis
- 471 Appendix E. Type 1 error Analysis workflow
- 472 Appendix F. Type 1 error Analysis results
- 473 Appendix G. Trait Medusa approach: methods & results
- 474 Supplement 1. R code Example
- 475 Supplement 2. Data Example
- 476

ECOLOGICAL SOCIETY OF AMERICA^{21/23}

477 **Table 1**. Table 1 summarizes the statistical performance (type I error -the % of tests based on 478 samples that incorrectly rejected a true null hypothesis- and power -the % of tests that correctly 479 rejected a false null hypothesis) of classical PGLS (optimized for a single λ) under different 480 simulation scenarios. Complex evolutionary models were used to simulate either X and the 481 residuals of Y (column 1), only X (column 2) or only the residuals (column 3). 'KO' indicates 482 reduced power or inflated type I error, 'OK' indicates good power and correct type I errors.

Complex Evolutionary	Residuals	Only X	Only residuals	
model in:	& X	(residuals are	(X is normally	
			distributed)	
		distributed)		
Statistical				
erformance:				
Type 1 error	КО	ОК	ОК	
Power	ОК	ОК	ОК	

484

ECOLOGICAL SOCIETY OF AMERICA^{22/23}

485 **Figure 1. Effect of heterogeneous phylogenetic signal on type I error.** Effect of variation in 486 phylogenetic signal heterogeneity on Type I error rates for the different comparative methods: 487 classical OLS, PGLS that jointly optimises a single λ value for the residuals together with 488 parameter estimates ('PGLS_{global, λ}') and a PGLS that uses the true VCV matrix ('PGLS_{TrueVCV}'). 489 The X-axis represents λ of clade B (λ of clade A is set to one). Plotted below the X-axis are the 490 corresponding transformed trees for a homogeneous signal (λ [Clade A] = λ [Clade B] =1) and a 491 heterogeneous signal (λ [Clade A] = 1; λ [Clade B] =0.01). The type I error represents the 492 percentage of simulation that detected a significant correlation at the 5% level (1000 simulations) 493 between the two traits which is expected to be 5% for a valid method (black horizontal line). 494 **Figure 2. Effect of heterogeneous rate of trait evolution on type I error.** Effect of variation in 495 evolutionary rate heterogeneity (i.e. the ratio of rate evolution between clades) on Type I error 496 rates for different comparative methods (see legend of Figure 1). Different models of rate 497 heterogeneity are presented (i.e. from 1 to 9 rate shifts). For the simplest case (one single rate 498 shift), we plotted below the X-axis the corresponding transformed trees for a homogeneous rate 499 (σ^2 [Clade 1] = σ^2 [Clade 2] =1) and a heterogeneous signal (σ^2 [Clade 1] = 1; σ^2 [Clade 2] =0.01).

500 The type I error represents the percentage of simulation that detected a significant correlation at 501 the 5% level (1000 simulations) between the two traits which is expected to be 5% for a valid 502 method (black horizontal line).

503 Figure 3. Type I error rates for modified PGLS. Comparison of type I error rate for classical 504 (OLS, PGLS_{global_ λ}, and PGLS_{TrueVCV}) and modified (PGLS_{auteur} and PGLS_{combination}) comparative 505 methods as a function of evolutionary rate heterogeneity between clades. We show here the result 506 for the simplest model of rate heterogeneity (i.e. one single rate shift) and plot below the X-axis 507 the corresponding transformed trees for a homogeneous rate (σ^2 [Clade 1] = σ^2 [Clade 2] =1) and a 508 heterogeneous rate (σ^2 [Clade 1] = 1; σ^2 [Clade 2] =0.01). The type I error represents the

ECOLOGICAL SOCIETY OF AMERICA^{23/23}

509 percentage of simulation that detected a significant correlation at the 5% level (1000 simulations)

510 between the two traits which is expected to be 5% for a valid method (black horizontal line).







