The need to include phylogeny in trait-based analyses of community composition

Daijiang Li[1*] and Anthony R. Ives[2]

[1]Department of Botany, University of Wisconsin-Madison, Madison, WI 53706, USA

[2]Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA

[*]Present address: Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL 32611, USA

Author contributions: DL and AI designed the study; DL conducted the simulations; DL and AI wrote the manuscript.

Emails: daijianglee@gmail.com arives@wisc.edu

Running title: Trait-environment phylogenetic mixed models

Word count:

Key words: Phylogeny, functional traits, fourth-corner problem, trait-environment relationship, phylogenetic linear mixed model

23   **Summary**

24   1. A growing number of studies incorporate functional trait information to analyse patterns and

25   processes of community assembly. These studies of trait-environment relationships generally

26   ignore phylogenetic relationships among species. When functional traits and the residual

27   variation in species distributions among communities have phylogenetic signal, however,

28   analyses ignoring phylogenetic relationships can decrease estimation accuracy and power, inflate

29   type I error rates, and lead to potentially false conclusions.

30   2. Using simulations, we compared estimation accuracy, statistical power, and type I error rates

31   of linear mixed models (LMM) and phylogenetic linear mixed models (PLMM) designed to test

32   for trait-environment interactions in the distribution of species abundances among sites. We

33   considered the consequences of both phylogenetic signal in traits and phylogenetic signal in the

34   residual variation of species distributions generated by an unmeasured (latent) trait with

35   phylogenetic signal.

36   3. When there was phylogenetic signal in the residual variation of species among sites, PLMM

37   provided better estimates (closer to the true value) and greater statistical power for testing

38   whether the trait-environment interaction regression coefficient differed from zero. LMM had

39   unacceptably high type I error rates when there was phylogenetic signal in both traits and the

40   residual variation in species distributions. When there was no phylogenetic signal in the residual

41   variation in species distributions, LMM and PLMM had similar performances.

42   4. LMMs that ignore phylogenetic relationships can lead to poor statistical tests of trait-

43   environment relationships when there is phylogenetic signal in the residual variation of species

44   distributions among sites, such as caused by unmeasured traits. Therefore, phylogenies and

45 PLMMs should be used when studying how functional traits affect species abundances among

46 communities in response to environmental gradients.

Introduction

48 Species composition and abundance in ecological communities depend in part on both the

49 environmental conditions at a site and the traits expressed by species that allow them to live

50 under these environmental conditions. Typically, environmental conditions at a site allow only a

51 subset of species from the regional species pool to reach high abundances, with different

52 functional traits favouring species in different sites. Therefore, both environmental conditions

53 and functional traits play an important role in explaining species abundances in communities. To

54 better understand community assembly, we need to study the statistical interaction between

55 environmental conditions at a site and the functional traits of species that live there (McGill et al.

56 2006; Westoby & Wright, 2006).

57

58 Common statistical approaches to analyse how traits mediate species responses to environmental

59 variables have used either ordination with permutation tests (the fourth-corner problem and RLQ

60 analysis, Legendre, Galzin & Harmelin-Vivien 1997; Dray & Legendre, 2008) or an indirect

61 two-step approach. The fourth-corner problem links three data matrix tables: a site × species

62 incidence/abundance matrix (L), a site × environmental variables matrix (R), and a species ×

63 traits matrix (Q). The traits × environmental variables matrix (R'LQ) is the fourth matrix (thus

64 explaining the etymology of the approach). While this approach provides a good qualitative

65 overview of how traits and environmental variables are associated, it does not give information

66 about species-specific variation in responses to environmental variables, and it is difficult to use

3

67   for prediction. The second, two-step approach first fits species-specific regressions of abundance

68   against environmental variables; the resulting regression coefficients are then regressed against

69   traits (e.g., Soudzilovskaia et al. 2013). This approach, while informative at the species level,

70   does not incorporate all community data in a single analysis and has low statistical power (Jamil

71   et al. 2013).

72

73   The interactions between traits and environmental variables can also be directly tested with

74   model-based methods (Bolker et al. 2009; Jamil et al. 2013; Brown et al. 2014; Warton et al.

75   2014, Ovaskaine, De Knegt & Delgado 2016). Statistically, the interaction between traits and

76   environmental variables can be estimated as the trait-environment interaction coefficient in

77   generalized linear models (GLMs, Brown et al. 2014), linear mixed models (LMMs, Ovaskaine,

78   De Knegt & Delgado 2016), or generalized linear mixed models (GLMMs, Pollock, Morris &

79   Vesk 2012; Jamil et al. 2013). These model-based methods allow model selection and prediction,

80   and are often more flexible, powerful, and informative than fourth-corner and two-step

81   approaches (Ives & Helmus 2011; Jackson et al. 2012; Brown et al. 2014; Warton et al. 2014).

82

83   Most analyses of trait-environment interactions ignore phylogenetic relationships among species,

84   despite the large literature on phylogenetic analyses in comparative studies (Felsenstein 1985;

85   Harvey & Pagel 1991; Paradis 2012; Garamszegi 2014) and the relevance of phylogeny to many

86   areas of ecology (Webb et al. 2002; Cavender-Bares et al. 2009). This can lead to statistical

87   problems because functional traits often exhibit a phylogenetic pattern in which closely related

88   species share similar trait values (i.e., phylogenetic signal, Blomberg, Garland & Ives 2003). If

89   there are multiple traits that affect species abundance or incidence (or other characteristic of

4

90   interest), then the unmeasured traits with phylogenetic signal may generate covariance in the

91   unexplained, residual variation after accounting for measured traits. This covariance in the

92   residual variation will reflect the phylogeny, and will affect model estimation and hypothesis

93   testing of regression coefficients (e.g., Felsenstein 1985, Martins & Hansen 1997; Garland,

94   Bennett & Rezende 2005; Revell 2010).

95

96   Here, we investigate the need to incorporate phylogenetic covariance among species in

97   regressions for trait-environment interactions. We considered a regression problem in which

98   there is a causal but unmeasured (latent) trait that introduces unexplained variability in species

99   abundance, and phylogenetic covariance in the unexplained variation if the unmeasured trait has

100   phylogenetic signal. This gives four possible cases (Revell 2010): the pairwise combinations of

101   whether or not there is phylogenetic signal in the measured trait in the regression, and whether or

102   not there is phylogenetic signal in the residual variation. We then compared the accuracy, type I

103   error rates, and statistical power of linear mixed models (LMMs) and phylogenetic linear mixed

104   models (PLMMs, Ives & Helmus 2011) in estimating the trait-environment interaction

105   coefficient. We show that when there is phylogenetic signal in the residual variation (latent trait),

106   PLMM outperformed LMM, with LMM performing particularly poorly when there is also

107   phylogenetic signal in the measured trait.

108

Materials and methods

110   We simulated data to test the importance of accounting for phylogenetic relationships when

111   studying how functional traits interact with environmental variables to affect species abundances.

112   All simulations and calculations were performed with R (R Core Team, 2015).

5

Simulations

113

114 We simulated the abundance Y of species j ($j = 1, \dots, n$) at site s ($s = 1, \dots, m$) that depends on

115 two site environmental variables (env1 and env2) and two species functional traits

116 (trait1 and trait2) using the model

117

118 $Y_i = \alpha + \beta_1 \, \text{env1}_{\text{site}[i]} + \beta_2 \, \text{env2}_{\text{site}[i]} + \beta_3 \, \text{trait1}_{\text{spp}[i]} + \beta_4 \, \text{trait2}_{\text{spp}[i]} +$

119 $\qquad \beta_5 \, (\text{env1}_{\text{site}[i]} \times \text{trait1}_{\text{spp}[i]}) + \beta_6 \, (\text{env1}_{\text{site}[i]} \times \text{trait2}_{\text{spp}[i]}) + e_i.$ (1)

120

121 Functions spp[i] and site[i] map the observation i to the identity of the species and site,

122 respectively (Gelman & Hill, 2007, p251-252), so i takes values from 1 to nm. We assume both

123 environmental variable env1 and functional trait trait1 are measured. Env1 (e.g., soil fertility,

124 canopy cover) affects the abundance of all species among sites ($\beta_1 \neq 0$), and trait1 (e.g., nutrient

125 absorption capacity, specific leaf area) determines in part the overall abundance of species ($\beta_3 \neq$

126 0). Furthermore, there is an interaction between env1 and trait1 ($\beta_5 \neq 0$) implying that trait1

127 affects the performance of species along the environmental gradient env1.

128

129 To introduce unexplained variation and phylogenetic signal, we treated env2 and trait2 as

130 unmeasured (latent) variables. Like env1, env2 has a direct effect on species abundances ($\beta_2 \neq 0$).

131 Like trait1, trait2 determines in part species abundances ($\beta_4 \neq 0$) and has an interactive effect

132 with env1 ($\beta_6 \neq 0$). As we are mainly interested in the trait × environment interactions for the

133 measured data (env1 and trait1), we did not include the interactions between env2 and trait1 or

134 trait2. Our goal is to investigate the interaction between env1 and trait1 which is given by β_5.

135 Consequently, we set all parameters in equation 1 other than β_5 to be 1. Finally, we simulated e_i

6

136   as a normal random variable that is independent among species and sites. In this way, we treated

137   the abundance of species *Y* as log-transformed values from count data. We did not simulate

138   abundance as raw count data because log-transformation of count data usually does not affect the

139   significance tests for regression coefficients when low count values (<5) are uncommon (Ives

140   2015; Warton et al. 2016).

141

142   We simulated the phylogeny as a uniform birth-death process with birth rate = 1 and death rate =

143   0 using the `sim.bdtree` function of the `geiger` R package (Harmon et al. 2008). The

144   phylogeny gives the expected phylogenetic covariances among species under Brownian motion

145   evolution (Grafen 1989; Martins & Hansen 1997) that can be used to construct a matrix **C**, and

146   when there is no phylogenetic signal the (zero) covariance structure is given by the identity

147   matrix **I**. Because functional traits may or may not have phylogenetic signal, we simulated four

148   scenarios for the two functional traits: trait1 with phylogenetic signal but not trait2 (trait1: **C**;

149   trait2: **I**); trait2 with phylogenetic signal but not trait1 (trait1: **I**; trait2: **C**); both traits with

150   phylogenetic signal (trait1: **C**; trait2: **C**); and neither trait with phylogenetic signal (trait1: **I**;

151   trait2: **I**). Functional traits without phylogenetic signal were simulated as $N(0, 1)$ normal random

152   variables; functional traits with phylogenetic signal were simulated using the `fastBM` function of

153   the `phytools` R package (Revell 2012). We simulated env1 as a uniform distribution ranging

154   from −1 and 1 to generate a strong environmental gradient. Variable env2 and residuals e_i were

155   simulated as $N(0, 1)$ normal random variables.

156

157   We conducted simulations with 30 sites. To study type I error rates (false positives that

158   incorrectly reject the true null hypothesis), we set $\beta_5 = 0$ and varied the number of species (20,

159   30, 40, 50, 60, 70, 80). To study statistical power, we varied the value of β_5 (0, 0.25, 0.5, 0.75, 1)

160   and fixed the number of species at 50. For each case we performed 1000 simulations.

161

Model fitting

163   We fit both LMM and PLMM to the simulated datasets with R package `pez` (Pearse et al. 2015).

164   The LMM has the form

165

166    $Y_i = \alpha + a_{\mathrm{spp}[i]} + b_{\mathrm{site}[i]} + (\beta_1 + c_{\mathrm{spp}[i]})\,\mathrm{env1}_{\mathrm{site}[i]} + \beta_3\,\mathrm{trait1}_{\mathrm{spp}[i]} + \beta_5\,\mathrm{env1}_{\mathrm{site}[i]} \times \mathrm{trait1}_{\mathrm{spp}[i]} + e_i$

167    $a \sim \mathrm{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$

168    $b \sim \mathrm{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{I}_m)$

169    $c \sim \mathrm{Gaussian}(\mathbf{0}, \sigma_c^2 \mathbf{I}_n)$ (2)

170

171   Here, we use the convention of multilevel models (Gelman & Hill, 2007), with fixed and random

172   effects given by Greek and Latin letters, respectively. The fixed effects β_1, β_3, and β_5 correspond

173   to the same coefficients in the simulation model (equation 1). Random effect $a_{\mathrm{spp}[i]}$ allows

174   different species to have different overall abundance to capture effects of the term $\beta_4\,\mathrm{trait2}_{\mathrm{spp}[i]}$ in

175   equation 1. Random effect $b_{\mathrm{site}[i]}$ allows different sites to have different overall abundance across

176   all species within that site to capture effects of the term $\beta_2\,\mathrm{env2}_{\mathrm{site}[i]}$ in equation 1. Finally,

177   random effect $c_{\mathrm{spp}[i]}$ allows different species to have different responses to env1 to capture effects

178   of the term $\beta_6\,\mathrm{env1}_{\mathrm{site}[i]} \times \mathrm{trait2}_{\mathrm{spp}[i]}$ in equation 1.

179

180   The PLMM includes all terms of equation 2, plus phylogenetic version of random terms $a_{\mathrm{spp}[i]}$

181   and $c_{\mathrm{spp}[i]}$:

8

182

183    $Y_i = \alpha + (a_{\text{spp}[i]} + a^p_{\text{spp}[i]}) + b_{\text{site}[i]} + (\beta_1 + c_{\text{spp}[i]} + c^p_{\text{spp}[i]})\ \text{env1}_{\text{site}[i]} + \beta_3\ \text{trait1}_{\text{spp}[i]} +$

184        $\beta_5\ \text{env1}_{\text{site}[i]} \times \text{trait1}_{\text{spp}[i]} + e_i$

185    $a \sim \text{Gaussian}(\mathbf{0}, \sigma^2_a \mathbf{I}_n)$

186    $a^p \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{ap} \mathbf{C})$

187    $b \sim \text{Gaussian}(\mathbf{0}, \sigma^2_b \mathbf{I}_m)$

188    $c \sim \text{Gaussian}(\mathbf{0}, \sigma^2_c \mathbf{I}_n)$

189    $c^p \sim \text{Gaussian}(\mathbf{0}, \sigma^2_{cp} \mathbf{C})$ (3)

190

191   Random effect $a^p_{\text{spp}[i]}$ implies closely related species to have similar overall abundance; this will

192   capture the main effects of traits in the simulations (equation 1) if trait2 has phylogenetic signal.

193   Similarly, random effect $c^p_{\text{spp}[i]}$ allows closely related species to have similar responses to env1,

194   thereby capturing the interactive effect of trait2 and env1 in the simulations if trait2 has

195   phylogenetic signal.

Results

196 **Results**

197   To compare LMMs and PLMMs, we focused on the regression coefficient β_5 for the interaction

198   between env1 and trait1. For each simulated dataset, we compared the accuracy of LMM and

199   PLMM by determining the frequency with which one gave a more accurate estimate of β_5 than

200   the other, and also by calculating the means and standard deviations of the estimates of β_5. We

201   also counted the number of estimates that were scored as significant at the $\alpha = 0.05$ level for both

202   models to determine their type I errors (when the true value of $\beta_5 = 0$) and statistical power

203   (when the true value of $\beta_5 > 0$).

204

205   *No phylogenetic signal in trait2*

206   When the unmeasured trait2 did not have phylogenetic signal (trait1: **I**; trait2: **I**, and trait1: **C**;

207   trait2: **I**), implying no phylogenetic signal in the unexplained variation in species abundances

208   among sites, LMM and PLMM had similar estimation accuracy (Fig. 1-2), type I error rates, and

209   power (Fig. 3). Averaged across all simulation scenarios, in roughly 50% of simulations LMM

210   produced better estimates (closer to the true value) of β_5 (Fig. 1). The estimators of β_5 from

211   LMM and PLMM had similar means and standard deviations (Fig. 2A, Fig. 2B, Fig. A1).

212   Furthermore, LMM and PLMM had almost identical type I error rates and power across all

213   simulation scenarios (Fig. 3). They also gave very similar estimates when $\beta_5 > 0$ (Fig. A2). These

214   results are explained, in part, by the fact that in about 65% of simulations across all scenarios we

215   investigated with no phylogenetic residual variation (trait2: **I**), the estimates of both σ^2_{ap} and σ^2_{cp}

216   in the PLMM were zero, so the PLMM collapsed to the LMM and estimates of β_5 were the same

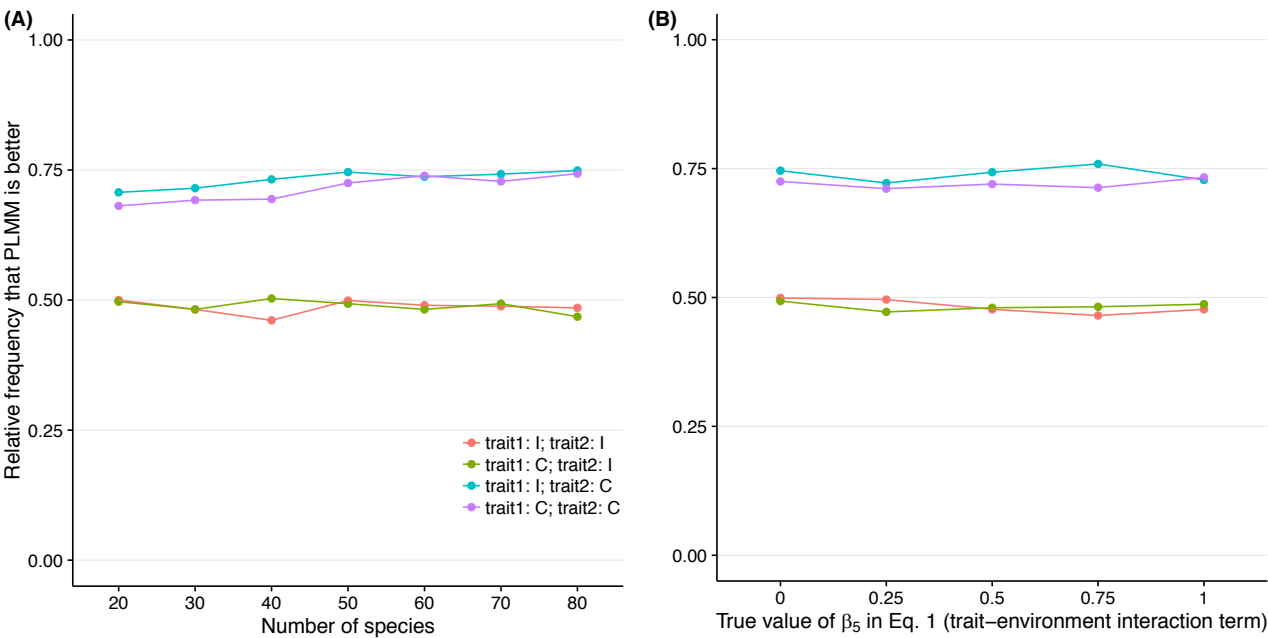217   (± numerical accuracy in the REML optimizations).

218

10

219

Figure 1. The fraction of simulations in which PLMM yielded a better estimate of β_5 (i.e., closer to its true value) than LMM versus (A) the number of simulated species and (B) the true value of β_5 (Eq. 1). The performance of PLMM was consistently better than LMM whenever there was phylogenetic signal in the residual variation (caused by unmeasured trait2). Abbreviations: trait1: **I** – measured trait1 does not have phylogenetic signal; trait1: **C** – measured trait1 has phylogenetic signal; trait2: **I** – unmeasured trait2 does not have phylogenetic signal; trait2: **C** – unmeasured trait2 has phylogenetic signal.
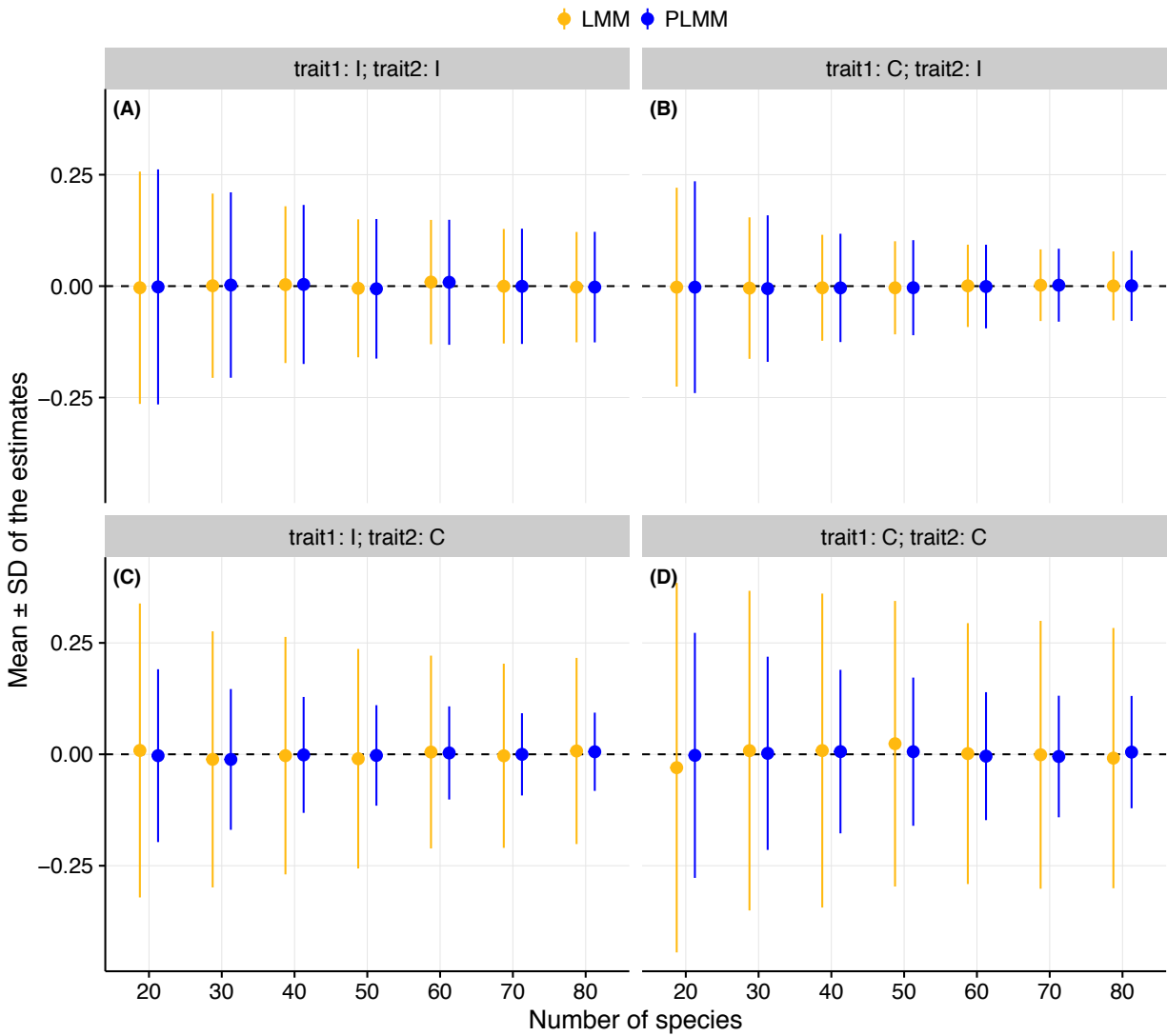
11

Figure 2: Mean (± standard deviation) of simulated estimates of β_5 (Eq. 1) using LMM and PLMM

versus the number of species in the simulations for cases (A) trait1: **I**; trait2: **I**, (B) trait1: **C**; trait2: **I**,

(C) trait1: **I**; trait2: **C**, and (D) trait1: **C**; trait2: **C**. Horizontal dash lines represent the true value of the

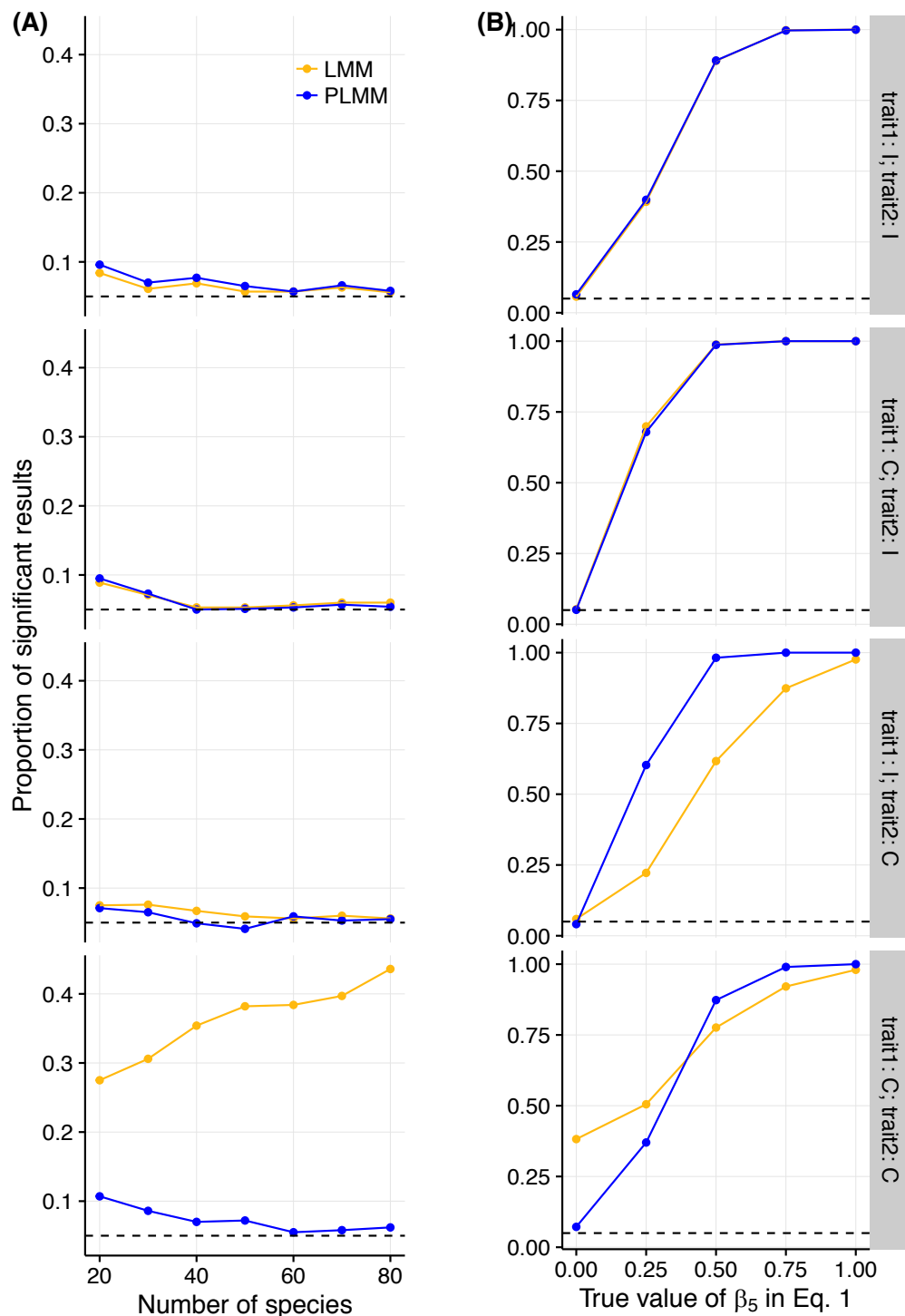parameter. Abbreviations are as in Fig. 1.

231

Figure 3. (A) Type I error rates and (B) statistical power of LMMs and PLMMs under four scenarios of

233 simulated functional traits (abbreviations as in Fig. 1). For all tests, a significance level of $\alpha = 0.05$ is used

234 (horizontal dashed lines).

235   *Phylogenetic signal in trait2*

236   When the unmeasured trait2 had phylogenetic signal (trait1: **I**; trait2: **C,** and trait1: **C**; trait2: **C**),

237   PLMM had substantially higher estimation accuracy (Fig. 1-2), better type I error control (Fig.

238   3A), and higher power (Fig. 3B) than LMM. Type I error control and power were particularly

239   poor for LMM when trait1 also had phylogenetic signal (i.e., trait1: **C**; trait2: **C**).

240

241   Averaged across all simulation conditions, in about 75% simulations PLMM produced more

242   accurate estimates of β_5 (Fig. 1), and the variance of the estimator of β_5 (Fig. 2 and A1) was

243   consistently lower than LMM. The was true regardless of the number of species, the true value of

244   β_5, and the status of the measured trait1 (with or without phylogenetic signal) used in

245   simulations. In addition, for type I error control and power, LMM had particularly poor

246   performance when the measured trait1 had phylogenetic signal (trait1: **C**; trait2: **C**). For

247   simulations with $\beta_5 = 0$ (Fig. 3), LMM rejected H_0: $\beta_5 = 0$ at the $\alpha = 0.05$ level in ~25% of the

248   datasets with 20 species, and type I error control became worse as the number of species

249   increased. When there was no phylogenetic signal in trait1 (trait1: **I**; trait2: **C**) and type I error

250   control was only slightly elevated, LMM had much lower power than PLMM (Fig. 3B)

251

252   We investigated further the particularly poor type I error control of LMM when there is

253   phylogenetic signal in both the measured trait and the unexplained residual variation (trait1: **C**;

254   trait2: **C**). Poor type I error control occurs when the estimate of the standard error of β_5 is smaller

255   than the true standard error. For cases both with phylogenetic signal in trait 1 (trait1: **C**; trait2: **C**)

256   and without (trait1: **I**; trait2: **C**), we plotted the estimate of the standard error of β_5 for each

257   simulated dataset against the estimate of β_5 using both LMM and PLMM (Fig. 4). For the case

14

258   (trait1: **I**; trait2: **C**), the decrease in accuracy of LMM relative to PLMM is seen in the greater

259   variance in the estimates of β_5 (variance in the horizontal direction). Despite this increase in the

260   variance in the estimates of β_5, false positives (given by values to the right of the dashed line of

261   Fig. 4) from the LMM are only slightly inflated, because the LMM estimates of the standard

262   error of β_5 are larger than those from PLMM. However, for the case (trait1: **C**; trait2: **C**), the

263   decrease in accuracy of LMM relative to PLMM is not accompanied by an appropriate increase

264   is the LMM estimates of the standard error, thereby leading to high type I error rates. In contrast

265   to LMM, even though the variance in the estimates of β_5 from PLMM increases when there is

266   phylogenetic signal in trait1 (Fig. 4A vs. 4B), the estimates of the standard error also increase,

267   leading to much better type I error control than LMM. In summary, the poor type I error control

268   for LMM when there is phylogenetic signal in trait1 occurs because, as phylogenetic signal in

269   trait1 increases the variance in the LMM estimates of β_5, phylogenetic signal in trait1 decreases

270   the LMM estimates of this variance. The decrease in power of LMM relative to PLMM for the

271   case without phylogenetic signal in trait1 (trait1: **I**; trait2: **C**) is caused by the increase in

272   variance in the estimator of β_5, that is, decreased accuracy. Given the very poor type I error

273   control for LMM for the case with phylogenetic signal in trait1 (trait1: **C**; trait2: **C**), it is

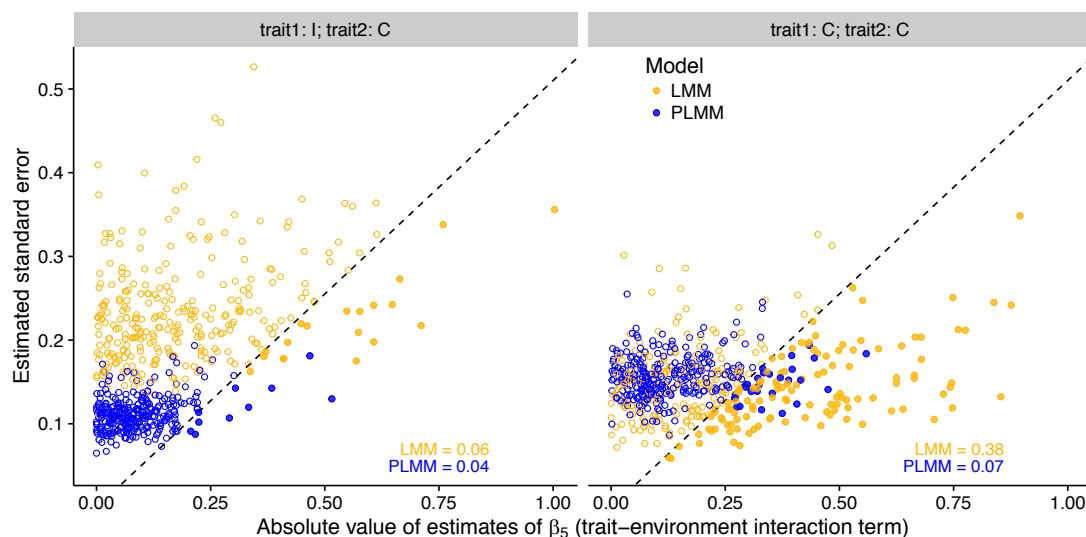274   inappropriate to assess power for this case.

275

15

Figure 4. The relationships between the absolute value of the estimates of β_5 and the estimates of the standard errors from LMM (yellow) and PLMM (blue). Solid points to the right of the dashed line would (approximately) reject the null hypothesis H_0: $\beta_5 = 0$ at the 0.05 α significance level. When unmeasured trait2 has phylogenetic signal but not measured trait1 (trait1: **I**; trait2: **C**), LMM estimates are more variable (horizontal axis) and have greater estimated standard errors (vertical axis) than PLMM, leading to only slightly inflated type I error control. When both trait1 and trait2 have phylogenetic signal (trait1: **C**; trait2: **C**), LMM estimates are more variable, but this is not correctly captured by increasing estimates of standard errors, leading to very high type I error rates. The dashed line has intercept of zero and slope of 1/1.96. Simulations were performed with 50 species; the fractions of simulations rejecting the null hypothesis (text in the panels) were calculated from 1000 simulations, of which only 300 are presented for clarity.

Discussion

Our simulations have demonstrated the importance of incorporating phylogeny into the study of how species functional traits interact with the environment to affect their abundance. In

16

292   simulations in which there was phylogenetic signal in the residual variation in abundances

293   caused by an unmeasured (latent) trait, we showed that LMMs have lower accuracy, poor type I

294   error control, and lower power than PLMMs in identifying the trait × environment interaction.

295   The performance of LMMs was particularly poor in terms of type I error control and power when

296   there was also phylogenetic signal in the measured trait. In contrast, PLMMs had better

297   accuracy, generally good type I error control (except when the number of species was small), and

298   good power.

299

300   Our results mirror the results of Revell (2010) who studied the performance of LMs and PLMs

301   applied to regression for phylogenetic comparative data. The model he considered that most

302   closely corresponds to our PLMM is a phylogenetic least-squares model in which Pagel's λ

303   branch-length transform is used. Pagel's λ transformation can be constructed by adding a

304   phylogenetic and a non-phylogenetic covariance matrix with λ scaling between them (i.e., $(1 -$

305    $\lambda)\mathbf{I} + \lambda\mathbf{C}$). In our PLMM (Eq. 3), covariance terms are similarly combined; for example, the

306   covariance for species-specific slopes across environmental variable 1 is $\sigma_c^2\mathbf{I}_n + \sigma_{cp}^2\mathbf{C}$. Revell

307   (2010) found that PLMs outperformed LMs when there was phylogenetic signal in the residual

308   variation, with the performance of LMs particularly poor when there was also phylogenetic

309   signal in the independent variable. Thus, we found similar results in the more-complex problem

310   of identifying trait × environment interactions in community data.

311

312   The better performance of PLMMs over LMMs is not surprising on theoretical grounds. For the

313   special, hypothetical case in which the variance parameters σ_a^2, σ_{ap}^2, σ_b^2, σ_c^2, and σ_{cp}^2 are known, the

314   PLMM in equation 3 will be the minimum variance estimator of the regression coefficients

17

315   (fixed effects), including the trait $\times$ environment interaction β_5; this is a consequence of the

316   Cramer-Rao Theorem applied to Generalized Least Squares (GLS) models (Judge et al. 1985).

317   This explains why PLMMs provide more accurate estimates of β_5 than LMMs, and the increase

318   in accuracy explains the increase in power of PLMMs relative to LMMs.

319

320   A particular warning derived from our simulations is the poor type I error control for LMMs

321   when there is phylogenetic signal in both the residual variation and in the independent variable.

322   When there is also phylogenetic signal in the measured trait1, the variance in the estimates of β_5

323   greatly increases. Nonetheless, the LMM estimates of the standard error of β_5 do not increase as

324   they should, leading to false rejections of the null hypothesis that $\beta_5 = 0$. Because PGLMMs are

325   close to the minimum variance estimators of β_5, the variance in its estimates of β_5 does not

326   increase as much as LMMs when there is phylogenetic signal in the independent variable, and

327   what increase occurs is correctly given by the estimates standard errors of β_5; thus, there is

328   generally good type I error control.

329

330   When the number of species is small (<60), however, PLMM had inflated type I error rates; for

331   simulations with 20 species and phylogenetic signal in both independent variable (measured

332   trait1) and residual variation (unmeasured trait2), the null hypothesis H_0: $\beta_5 = 0$ was rejected in

333   10% of the datasets at the α significance level of 0.05. In analyses with small numbers of species

334   and P-values computed from the data that are close to the significance level selected by the

335   researcher, we suggest using parametric bootstrapping. This can be performed by estimating

336   parameters from the data under H_0: $\beta_5 = 0$ (i.e., without the trait $\times$ environment interaction),

337   simulating a large number (e.g., 2000) datasets with these parameter values, fitting each dataset

18

338   with the full model (i.e., with the trait × environment interaction), and for each dataset recording

339   the Z-score of the estimate of β_5. The bootstrap approximate P-value of β_5 under an α

340   significance level of 0.05 is then given by the proportion of bootstrap Z-scores whose absolute

341   values exceed the absolute value of the Z-score from the observed data. Code for performing this

342   bootstrap is provided in Appendix S1.

343

344   Our analyses have been confined to abundance as a continuous dependent variable.

345   Presence/absence (incidence) community data can also be analysed with phylogenetic

346   information using PGLMM (Ives and Helmus 2011), and results will likely be similar. We did

347   not pursue this here, however, because the computational burden of PGLMMs with existing

348   software makes simulation studies difficult. Nonetheless, if tests of the existence of relationships

349   (i.e., testing H_0: $\beta_5 = 0$) are all that is needed, applying PLMMs to binary data generally provides

350   good type I error control, although at the expense of some power (Ives 2015; Warton et al. 2016).

351

352   Even when there was no phylogenetic signal in the residual variation, PLMMs performed as well

353   as LMMs. In part, this is because, when PLMMs detected no phylogenetic signal in the residual

354   variation, they give the same results as the corresponding LMMs (although their AIC values are

355   still penalized by the variance term that equals zero). The fact that PLMMs often collapse exactly

356   to LMMs as a special case suggests that PLMMs should be always used in analyses of trait ×

357   environment interactions, since there is no cost in the absence of phylogenetic signal and

358   considerable benefits when there is (which is likely).

359

19

Acknowledgements

Data accessibility:

No data were used in this article.

References:

Blomberg, S. P., Garland, T., & Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**:717-745.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**, 127–135.

Brown, A.M., Warton, D.I., Andrew, N.R., Binns, M., Cassis, G. & Gibb, H. (2014). The fourth-corner solution - using predictive models to understand how species traits interact with the environment. *Methods in Ecology and Evolution*, **5**, 344–352.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

Dray, S. & Legendre, P. (2008). Testing the species traits-environment relationships: The fourth-corner problem revisited. *Ecology*, **89**, 3400–3412.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, **125**, 1–15.

Garamszegi, L.Z. (2014). Modern phylogenetic comparative methods and their application in evolutionary biology. *Concepts and Practice. London, UK: Springer*.

Garland, T., Bennett, A.F. & Rezende, E.L. (2005). Phylogenetic approaches in comparative physiology. *Journal of experimental Biology*, **208**, 3015–3035.

383   Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*.

384   Cambridge University Press.

385   Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of*

386   *London. Series B, Biological Sciences*, **326**, 119–157.

387   Harmon, L., Weir, J., Brock, C., Glor, R. & Challenger, W. (2008). GEIGER: Investigating evolutionary

388   radiations. *Bioinformatics*, **24**, 129–131.

389   Harvey, P.H., Pagel, M.D. & others. (1991). *The comparative method in evolutionary biology*. Oxford

390   university press Oxford.

391   Ives, A.R. (2015). For testing the significance of regression coefficients, go ahead and log-transform

392   count data. *Methods in Ecology and Evolution*, **6**, 828–835.

393   Ives, A.R. & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses of

394   community structure. *Ecological Monographs*, **81**, 511–525.

395   Jackson, M.M., Turner, M.G., Pearson, S.M. & Ives, A.R. (2012). Seeing the forest and the trees:

396   Multilevel models reveal both species and community patterns. *Ecosphere*, **3**, art79.

397   Jamil, T., Ozinga, W.A., Kleyer, M. & ter Braak, C.J. (2013). Selecting traits that explain

398   speciesenvironment relationships: A generalized linear mixed model approach. *Journal of Vegetation*

399   *Science*, **24**, 988–1000.

400   Judge, G.G., Griffiths, W., Hill, R.C., Lutkepohl, H. & Lee, T.C. (1985). Introduction to the theory and

401   practice of econometrics.

402   Legendre P., M.L.H.-V., René Galzin. (1997). Relating behavior to habitat: Solutions to the fourth-corner

403   problem. *Ecology*, **78**, 547–562.

404   Martins, E.P. & Hansen, T.F. (1997). Phylogenies and the Comparative Method: A General Approach to

405   Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *The American Naturalist*,

406   **149**, 646–667.

407   McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from

408   functional traits. *Trends in Ecology & Evolution*, **21**, 178–185.

409   Ovaskainen, O., De Knegt, H.J. & Mar Delgado, M. del. (2016). *Quantitative ecology and evolutionary*

410   *biology: Integrating models with data*. Oxford University Press.

411   Paradis, E. (2011). *Analysis of phylogenetics and evolution with r*. Springer Science & Business Media.

412   Pearse, W.D., Cadotte, M.W., Cavender-Bares, J., Ives, A.R., Tucker, C.M., Walker, S.C., & Helmus,
413   M.R. (2015). pez: phylogenetics for the environmental sciences. *Bioinformatics*, **31**, 2888-2890.

414   Pollock, L.J., Morris, W.K. & Vesk, P.A. (2012). The role of functional traits in species distributions
415   revealed through a hierarchical model. *Ecography*, **35**, 716–725.

416   R Core Team. (2015). R: A Language and Environment for Statistical Computing.

417   Revell, L.J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and*
418   *Evolution*, **1**, 319–329.

419   Revell, L.J. (2012). Phytools: An r package for phylogenetic comparative biology (and other things).
420   *Methods in Ecology and Evolution*, **3**, 217–223.

421   Soudzilovskaia, N.A., Elumeeva, T.G., Onipchenko, V.G., Shidakov, I.I., Salpagarova, F.S., Khubiev,
422   A.B., Tekeev, D.K. & Cornelissen, J.H.C. (2013). Functional traits predict relationship between plant
423   abundance dynamic and long-term climate warming. *Proceedings of the National Academy of Sciences*,
424   **110**, 18180–18184.

425   Warton, D.I., Foster, S.D., De'ath, G., Stoklosa, J. & Dunstan, P.K. (2014). Model-based thinking for
426   community ecology. *Plant Ecology*, 1–14.

427   Warton, D.I., Lyons, M., Stoklosa, J. & Ives, A.R. (2016). Three points to consider when choosing a LM
428   or GLM test for count data. *Methods in Ecology and Evolution*, **7**, 882–890.

429   Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community
430   ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

431   Westoby, M. & Wright, I.J. (2006). Land-plant ecology on the basis of functional traits. *Trends in*
432   *Ecology & Evolution*, **21**, 261–268.

Supporting information:

434   Appendix S1: R scripts used for the simulations and analyses (in the `Rcode` folder). It is time

435   consuming to re-run all simulations and model fittings. Therefore, we also include all model

436   results for fixed effects (as `rds.zip` and `rds2.zip` files). (We cannot upload these two files

437   through manuscript central as they include >50 files; one can download them from here:
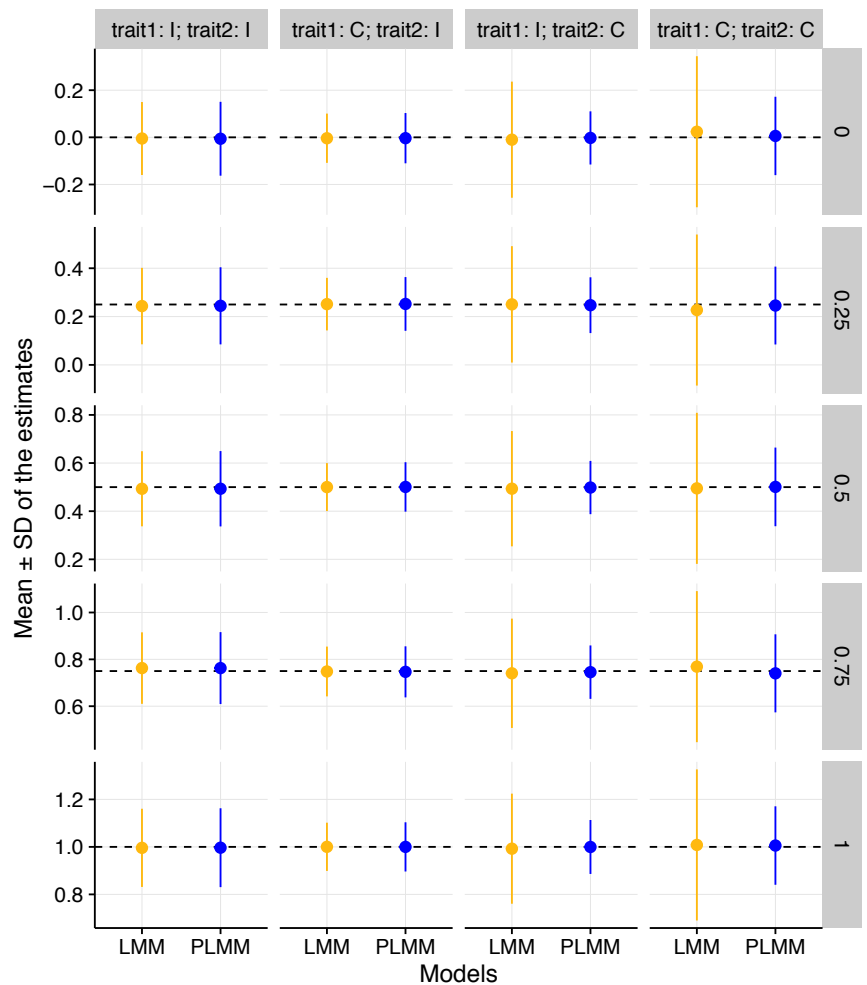
438   https://uwmadison.box.com/s/l1nnsmqlymg53xyxdarfksjrtegvuzci)

439

440

441   **Figure A1.** Mean (± standard deviation) of estimates of β_5 (Eq. 1) using LMM and PLMM varying the

442   true value of β_5 (0, 0.25, 0.5, 0.75, and 1). Horizontal dash lines represent the true value of the parameter.

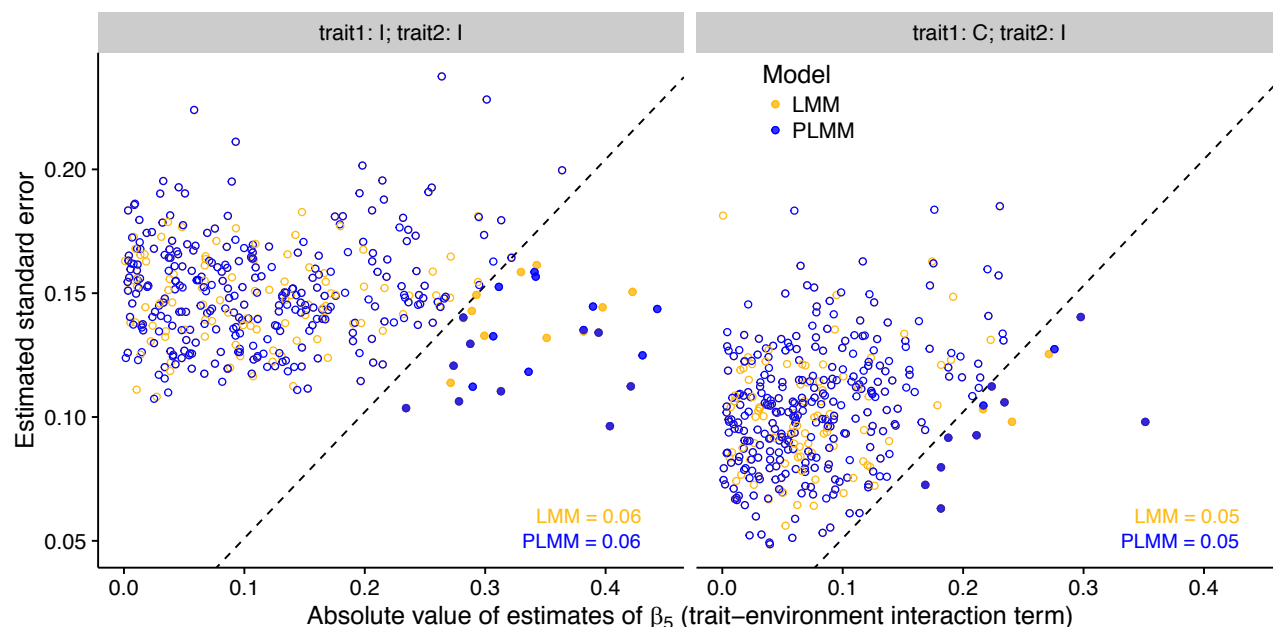443   Abbreviations are as in Fig. 1.

23

444

Figure A2. The relationships between the absolute value of the estimates of β_5 and the estimates of the standard errors for LMM (yellow) and PLMM (blue). Solid points to the right of the dashed line would reject the null hypothesis H_0: $\beta_5 = 0$. When neither trait1 nor trait2 has phylogenetic signal (trait1: **I**; trait2: **I**), LMM and PLMM estimates have similar variability (horizontal axis) and similar estimated standard errors (vertical axis). When trait1 has phylogenetic signal (trait1: **C**; trait2: **I**), both LMM and PLMM estimates become less variable, which is correctly captured by decreasing estimates of standard errors, leading to appropriate type I error rates. The dashed line has intercept of zero and slope of 1/1.96. Simulations were performed with 50 species; the fractions of simulations rejecting the null hypothesis (text in the panels) were calculated from 1000 simulations, of which only 300 are presented for clarity.

454