1     **Varying richness need not imply non-random species co-occurrence:**

2     **implications for specifying null models**

3     Short title: Species richness and co-occurrence

4     Author: Tom M. Fayle*

5     Address: Biology Centre of Czech Academy of Sciences, Institute of Entomology,

6     Branišovská 31, 370 05 České Budějovice, Czech Republic; Institute for Tropical

7     Biology and Conservation, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

8     *Corresponding author. Email: tmfayle@gmail.com

9

## Abstract

**Background.** Non-random species co-occurrence is of fundamental interest to ecologists. One approach to analysing non-random patterns is null modelling. This involves calculation of a metric for the observed dataset, and comparison to a distribution obtained by repeatedly randomising the data. Choice of randomisation algorithm, specifically whether null model species richness is fixed at that of the observed dataset, is likely to affect model results. This is particularly important in cases when there is high variation in species richness between sampling units in the observed data.

**Methods.** Here I demonstrate the effects of accounting for variation in species richness. I use the C-score, a metric measuring species segregation as "checkerboard units", applied to 289 datasets. First, I run null models in which sites are equally likely to be occupied (fixed-equiprobable algorithm). I do this both for the original datasets, and for the same datasets where occurrences are randomised with the species richness distribution fixed (pre-randomised datasets). Second, I run null models that fix site species richness to that observed (fixed-fixed algorithm).

**Results.** For real datasets, using the fixed-equiprobable algorithm (sites are equally likely to be colonised), C-score standardised effect size (SES) was positively related to variability in species richness between sites within a dataset. This effect was also found for pre-randomised datasets, indicating that variability in species richness can be exclusively responsible for detection of non-random species co-occurrence. When using the fixed-fixed algorithm (richness is constrained to that of real sites), there was no

32  relationship between SES and variability in species richness. There was also a reverse

33  in the effect direction, with 94% of significant tests indicating a lower C-score than

34  expected for the fixed-equiprobable algorithm, but 98% of significant tests indicating a

35  *higher* C-score than expected for the fixed-fixed algorithm.

36  **Discussion.** I speculate that when variation in species richness is high, fewer

37  checkerboard units are possible, regardless of segregation between species. Therefore,

38  use of fixed-equiprobable algorithms in situations where real species richness is highly

39  variable between sites within a dataset will yield significant results, even if species co-

40  occur randomly within the constraints of the species richness distribution. Consequently,

41  use of such tests makes the a priori assumption that high within-dataset variation in

42  species richness indicates non-random species co-occurrence. I recommend using

43  algorithms that explicitly take into account species richness distributions when one

44  wants to eliminate the effect of richness variation in terms of producing significant but

45  spurious positive co-occurrence results. Alternatively, non-null mechanistic models can

46  be created, in which hypothesised species assembly processes must be explicitly stated

47  and tested.

48  Keywords: competition, facilitation, species assembly, niche, Narcissus effect, C-score,

49  species segregation, species aggregation.

50  **Introduction**

51  Species commonly co-occur non-randomly with respect to each other. Determining the

52  particular deterministic patterns to which species conform lies at the heart of ecology.

53  Non-random co-occurrence patterns can be caused by a range of interactions between

3

54   species, including competition, predation, parasitism, facilitation, and mutualism (Bell et

55   al. 2010; Holdaway & Sparrow 2006). Drivers other than direct interactions can also

56   play a role, including similar or differing environmental preferences leading to positive or

57   negative co-occurrence respectively, and dispersal limitation over different scales (Holt

58   et al. 2013; Hubbell 2001). Often data for testing such patterns take the form of a

59   presence/absence species by sites matrix. One approach for analysing such matrices is

60   to use null models of species co-occurrence (Gotelli 2000). This suite of methods

61   involves the calculation of a metric measuring the way in which species co-occur.

62   However, solely from such a metric it is not possible to tell whether the structure of the

63   matrix differs from that which would be expected if species co-occurred independently.

64   For example, even in randomly structured matrices it is likely that some pairs of species

65   will not co-occur (Connor & Simberloff 1979). To overcome this problem, observed

66   species occurrences are randomised to break any associations between the species.

67   This is done repeatedly, and the metric calculated for each of these randomised

68   matrices, allowing estimation of the distribution of the metric expected under the null

69   hypothesis that species co-occur independently of each other (hence the term "null

70   modelling"). This then allows the calculation of a p-value, that is, the probability that the

71   observed metric, or one more extreme, would be observed under the null hypothesis of

72   random species co-occurrence.


73   A critical step in this null modelling process is defining what is "random" in a given

74   situation, and hence choosing exactly how to create randomised matrices from the

75   observed data (Gotelli & Ulrich 2012). However, this aspect of the field remains

76   contentious (e.g. Fayle & Manica 2011; Gotelli & Ulrich 2011). Specifically, the modeller

4

77    needs to select which aspects of the real dataset to maintain in the null model.

78    Algorithms range from those that do not constrain either species richness or species

79    occurrences across sites, keeping constant only the total number of species

80    occurrences across the entire matrix, to more restrictive methods that constrain both

81    richness and species occurrence distributions to be exactly the same as those from the

82    observed dataset (Connor & Simberloff 1979; Gotelli 2000; Miklós & Podani 2004). Most

83    analyses constrain the null occurrences of species to be the same as those observed in

84    the real population (Fayle & Manica 2010), thus assuming that the occurrences of

85    species in the observed sample correlate with the occurrences of species in the

86    unobserved species pool. However, deciding the degree to which null species richness

87    should reflect that of the real community is more problematic (Wilson 1995). Either null

88    species richness across sites can be constrained to exactly match those observed (or

89    some probabilistic version of this), or all sites can be defined to be equally likely to be

90    occupied by a particular species, although with the richness distribution being able to

91    vary to some degree (Gotelli 2000).

92    Constraining species richness values has the disadvantage that differences in species

93    richness might have been caused by non-random co-occurrence between species. For

94    example, presence of a keystone mutualist might lead to highly variable species

95    richness, dependent on the presence of that species. Since this is the very mechanism

96    that is under study it should therefore be excluded from the null model, potentially

97    making any test that includes such constraints overly conservative (i.e. risking type 2

98    errors). This has been called the "Narcissus effect" (Colwell & Winkler 1984), since the

5

99    signal of the processes under study is already "reflected" in the null model, and hence is

100   not detected.

101   However, allowing unconstrained occupation of sites means that factors causing non-

102   random differences in species richness between sites unrelated to species co-

103   occurrence patterns are also excluded from the null model (the "Jack Horner effect",

104   Wilson 1995). Thus, it is not clear whether significant results of analyses using this less

105   conservative method arise specifically from non-random species co-occurrence

106   patterns, or merely from the real species richness distribution being more variable

107   across sites within a dataset than is accounted for in the null model (in which case

108   significant results could be type 1 errors). An example of this is the need to account for

109   differences in island size when assessing co-occurrence between island dwelling

110   species, since larger islands will have more species, because they have more living

111   space and also present a larger target for incoming colonists (Connor & Simberloff

112   1979). Even if all islands represented identical habitats in terms of abilities of different

113   species to survive there, it would still be necessary to account for differences in island

114   size in a null model of species co-occurrence.

115   This problem was discussed during earlier development of null modelling of species co-

116   occurrence (Colwell & Winkler 1984; Wilson 1995). However, sometimes a range of

117   different methods for generating null matrices is used, and any non-random patterns

118   that emerge are discussed without reference to null model constraints (e.g. De los Ríos

119   et al. 2008). It is even possible to get opposite patterns (segregation vs aggregation) by

120   using different randomisation algorithms on the same dataset (Hétérier et al. 2008;

121   Rooney 2008), which is clearly a concern. The differences in results are expected to be

6

122   more severe the greater the variation in species richness between sites within a

123   dataset, since both randomisation methods should give similar results for datasets with

124   low variation in species richness, because the null models specified will be similar. Note

125   that co-occurrence metrics (even when standardised using null models) are correlated

126   with the richness of the entire assemblage across all sites (Ulrich et al. 2017; Ulrich et

127   al. 2018). Hence it is recommended that analyses including datasets with varying total

128   species richness (and number of sites), account for this, at a minimum by using these

129   measures of matrix size as a covariate during modelling (Ulrich et al. 2018). However,

130   the effects of variation in species richness between sites *within* an assemblage remain

131   unexplored. The purpose of this article is to demonstrate how variation in site species

132   richness (as opposed to total assemblage richness across all sites) impacts the results

133   of null models of species co-occurrence, with parallel analyses using randomisation

134   algorithms that either constrain or do not constrain site species richness. Note that here

135   I use "unconstrained species richness" to mean that null model species richness is not

136   dictated to be precisely that of the observed dataset; I do not mean that species

137   richness is completely free to vary.

138   I first quantify the degree to which within-dataset variability in species richness

139   correlates with the detection of significant patterns of species segregation in 289 real

140   datasets when that richness variability is not taken into account in the null model. To

141   assess the degree to which this pattern might relate solely to variability in species

142   richness itself, I then generate a series of datasets that match the real ones in terms of

143   species richness and species occurrence distributions, but for which species

144   occurrences are already randomised, and repeat this analysis. If detected patterns of

145    species segregation relate only to variability in species richness (and not to non-random

146    patterns of species co-occurrence), then we would expect the results from these two

147    analyses to be similar. Finally, I repeat both analyses using a randomisation algorithm

148    that constrains null site species richness to match the species richness distribution of

149    the real datasets. These models should allow detection of species segregation or

150    aggregation within the observed pattern of species richness.

151    **Materials and methods**

152    *Does variability in species richness between sites within a dataset correlate with*

153    *detection of deterministic species co-occurrence patterns?*

154    First, I tested the degree to which variability in species richness affects the probability of

155    finding significant patterns of co-occurrence between species in real datasets when this

156    observed variability is not included in the model (i.e. using a randomisation algorithm

157    that assumes uniform probability of species occurrence across sites). Note that for

158    these analyses there will still be variability in species richness between sites, since

159    species occupy sites randomly, but this variability will not necessarily reflect the

160    variability in the observed data. This was done using 289 presence-absence datasets

161    available in the nestedness temperature program (Atmar & Patterson 1993). Two

162    datasets were not analysed from the original collection of 291, since for these it was not

163    possible to randomise species co-occurrences. Note that many of these datasets relate

164    to distributions of species across islands, although some are from continuous areas of

165    habitat. The C-score was used to assess the degree of structure in these matrices. This

166    metric is a standardised measure of the number of checkerboard units present in a

167    matrix, i.e. the number of pairs of sites and pairs of species for which each species

168    occurs only once and at a different site (Stone & Roberts 1990). If the C-score is higher

169    than expected by chance, then pairs of species tend to be segregated across sites. The

170    C-score was calculated for each matrix (*nestedchecker* function in the R package

171    *vegan*; Oksanen et al. 2018). Each matrix was then randomised 5000 times while

172    maintaining occurrences per species, but assigning species occurrences to different

173    sites with equal probability. This was done with algorithm *c0* in the R *oecosimu* function

174    (Oksanen et al. 2018) and results in matrices with identical species occurrence

175    distributions and identical matrix fill to the observed dataset, but with each individual

176    species occurrence being assigned to a site at random (unless that site is already

177    occupied). P-values and standardised effect sizes were calculated on the basis of these

178    null distributions. For each matrix I also calculated the mean absolute deviation (MAD)

179    in species richness between sites as an easily interpretable measure of variability in

180    species richness. The species richness MAD was then used to predict C-score

181    standardised effect sizes (SES values) using a linear model. In the model I also

182    included the matrix fill, the number of sites, and the number of species, since these are

183    known to affect probability of pattern detection (Fayle & Manica 2010; Pitta et al. 2012;

184    Ulrich & Gotelli 2007), and MAD is likely to be greater in larger matrices, all other things

185    being equal. To test for any nonlinearity in the relationships between predictors, which

186    might make them unsuitable for use in this kind of model, I plotted MAD against number

187    of sites and number of species, and failed to find any nonlinear patterns (Supplementary

188    Figure 1). To achieve normally distributed residuals in this analysis and in all following

9

189    analyses, the response (C-score SES) and all predictors with the exception of matrix fill

190    were log transformed.

191    I carried out a second set of analyses to assess whether detection of non-random

192    patterns of species co-occurrence could arise solely from high variability in species

193    richness with species otherwise co-occurring at random. To do this I generated a series

194    of 289 matrices for which the row and column sums (and therefore also matrix fill) were

195    fixed to those from the real 289 matrices, but in which species occurrences were

196    randomised. This meant that for each matrix the distribution of species richness

197    between sites, occurrences between species, matrix fill, and matrix size were all

198    identical to the corresponding real matrix. This was done by running the *quasiswap*

199    algorithm in the R package *vegan* (Oksanen et al. 2018) on the corresponding real

200    dataset, and then conducting null model analyses in the same way as described in the

201    preceding paragraph. I also ran all of the analyses presented here using the *curveball*

202    randomisation algorithm in *vegan* (Strona et al. 2014), to check that these two

203    commonly used fixed-fixed algorithm give similar results. The results of these analyses

204    are almost identical to those using the *quasiswap* algorithm and hence only the latter

205    are presented in the main text (see Supplementary Figures 2 and 3 for plots using

206    *curveball*), I then used the same linear modelling framework as described above for the

207    real datasets to assess the effects of species richness MAD, matrix fill, the number of

208    sites, and the number of species on C-score standardised effect sizes, in the absence

209    of any interactions between species. Null model tests of these pre-randomised matrices

210    should give significant results (i.e. result in type 1 errors, since species co-occurrence is

211    known to be random) at a rate equal to the critical p-value. In order to test these type 1

212 error rates, the total number of matrices (out of 289) giving significantly non-random

213 results was tested against a binomial distribution with P(significant)=0.1 (note that

214 *oecosimu* uses p=0.1 as a critical value for two-tailed tests).

215 *What is the effect of including variability in species richness between sites within a*

216 *dataset in null models on detection of co-occurrence patterns?*

217 Including observed variability in species richness in null models should decrease the

218 probability of detecting non-random species co-occurrence. This is because, first, for

219 communities in which differences between species have generated non-uniform

220 patterns of species richness, the power to detect this effect will be decreased (the

221 Narcissus effect, see Introduction). Second, in communities in which species richness

222 deviates from the distribution expected with equal probabilities of species occupation

223 across sites as a result of processes unrelated to species co-occurrence, for which an

224 equiprobable model therefore incorrectly detected non-random co-occurrence patterns,

225 these errors will be corrected. I tested the degree to which accounting for differences in

226 species richness between sites in null models altered the outcomes of analyses of real

227 and simulated communities. The null models of species co-occurrence and linear

228 models linking MAD and standardised effect sizes were rerun on the real and pre-

229 randomised datasets, but null matrices for were generated using a fixed-fixed

230 randomisation algorithm that does not allow species richness in the null matrices to

231 deviate from that in the observed matrix (algorithm *quasiswap* in R *oecosimu* function,

232 see Supplementary Figure 3 for parallel analysis using the *curveball* algorithm). Note

233 that the pre-randomised datasets were the same as those used in the first section,

234 being generated using the *quasiswap* algorithm that constrains both row and column

235    sums. Hence the second of these analyses involving matrices pre-randomised using the

236    *quasiswap* algorithm, and then analysed using the *quasiswap* algorithm as a null model

237    are expected to give significant results no more than would be expected at random, and

238    are presented here for completeness only. Type 1 error rates were tested using

239    binomial tests, as described above.

240    **Results**

241    *Does variability in species richness between sites within a dataset correlate with*

242    *detection of deterministic species co-occurrence patterns?*

243    Of the 289 real matrices tested, 218 (75%) showed significant deviations in the C-score

244    from null models for which observed richness variability between sites was not included

245    (Table 1). The C-score standardised effect size (SES) was more negative with

246    increasing variation in species richness between sites (Linear model: mean absolute

247    deviation: $t_{4,284}=-3.54$, P<0.001, Figure 1A,B). That is, detecting non-random species

248    co-occurrence was more likely if there was high variation in species richness between

249    sites within a dataset (Figure 1A). This was still the case after taking into account

250    negative relationships between SES and number of species, number of sites, and

251    matrix fill (species: $t_{4,284}=-28.75$, P<0.001; sites: $t_{4,284}=-3.03$, P=0.003, fill: $t_{4,284}=-9.50$,

252    P<0.001). The majority of observed matrices that deviated significantly from null models

253    had fewer checkerboard units than would be expected at random (205 of 218 non-

254    random results; 94%), meaning that there were low levels of segregation between

255    species.

256  For randomly generated matrices with identical species richness and species

257  occurrence distributions to the real matrices, 223 of the 289 sites (77%) showed non-

258  random species co-occurrence (one tailed binomial test against P(significant)=0.1:

259  P<0.001). As was the case with the real matrices, the C-score standardised effect size

260  was more negative with increasing variation in species richness between sites (Linear

261  model: mean absolute deviation: $t_{4,284}$=-3.94, P<0.001, Figure 1C,D). This was still the

262  case after taking into account negative relationships between SES and number of

263  species, number of sites, and matrix fill (species: $t_{4,284}$=-31.69, P<0.001; sites: $t_{4,284}$=-

264  3.43, P=0.001, fill: $t_{4,284}$=-9.75, P<0.001). The high proportion of significant results

265  indicates that analyses of these datasets, in which species co-occurred at random,

266  suffered from extremely high levels of type 1 errors. If non-random co-occurrence

267  between species had generated variation in species richness, then this analysis of

268  randomised datasets and the first set of analyses should have given different results.

269  This is because the signature of non-random co-occurrence that generated differences

270  in species richness should still be present in the first set of analyses, but would have

271  been removed by pre-randomisation for the second set of analyses. However, this was

272  not the case, and results of the two sets of analyses were similar (205 real matrices vs.

273  210 pre-randomised matrices with C-score lower than expected at random, 13 real

274  matrices vs. 13 pre-randomised matrices with C-score higher than random).

275  *What is the effect of including variability in species richness between sites within a*

276  *dataset in null models on detection of co-occurrence patterns?*

277  When using a randomisation algorithm that preserves the number of species per site in

278  addition to the number of occurrences per species (fixed-fixed), for real matrices, 111 of

13

279 the 289 sites (38%) showed non-random species co-occurrence. There was no longer a

280 relationship between C-score SES and variation in species richness (Linear model:

281 $t_{4,284}=0.17$, P=0.866, Figure 2A,B). There were still positive relationships between C-

282 score SES and number of species, number of sites, and matrix fill (species: $t_{4,284}=-7.70$,

283 P<0.001; sites: $t_{4,284}=2.86$, P=0.005, fill: $t_{4,284}=2.64$, P=0.009). The majority of datasets

284 for which there was significantly non-random co-occurrence had a higher C-score than

285 would be expected (111 of 113 non-random results, 98%), in contrast with the results

286 for the fixed-equiprobable algorithm (not accounting for observed variation in species

287 richness), which gave most matrices as having significantly lower C-score than would

288 be expected (205 of 218 non-random results, 94%; Figure 1A, Figure 2A, Table 1).

289 As expected, for randomly generated matrices with identical species richness and

290 species occurrence distributions to the real matrices, fixing the number of species per

291 site in null models resulted in the loss of the relationship between C-score standardised

292 effect size and variation in species richness between sites, although the relationship

293 was only marginally non-significant (Linear model: mean absolute deviation: $t_{4,284}=-1.87$,

294 P=0.062, Figure 2C,D; compare to Figure 1C,D). Furthermore, there were also no

295 relationships between SES and number of species, number of sites, and matrix fill

296 (species: $t_{4,284}=-1.02$, P=0.309; sites: $t_{4,284}=-0.09$, P=0.173, fill: $t_{4,284}=-0.06$, P=0.956).

297 From the 289 sites, 12 (4%) showed non-random species co-occurrence (one tailed

298 binomial test against P(significant)=0.1: P>0.999), indicating that type 1 error rates were

299 within acceptable limits (10%) for these analyses. These results were expected because

300 the same algorithm was used to generate the "observed" datasets and to create

301 corresponding null models.

**Discussion**

302

303 My results show that for this set of 289 real matrices, if null models are run without

304 including observed variation in species richness, the majority of tests are significant,

305 with most matrices having a smaller C-score (fewer checkerboard units) than would be

306 expected at random. Furthermore, the greater the variation in species richness across

307 sites, the larger the effect size. Such a pattern might be caused in two different ways.

308 First, matrices with high variation in species richness might be those for which particular

309 pairs of species co-occur non-randomly. In this case, since C-scores were lower than

310 expected, this could be interpreted as evidence for ecological mechanisms that might

311 lead to aggregation, such as mutualism, facilitation, or shared environmental

312 preferences. It is easy to see how such mechanisms might give rise to sites with

313 variable species richness. For example, if one "aggregator" species was present, which

314 made the co-occurrence of many other species more likely, then this could generate

315 sites with higher species richness. Conversely, if a strongly competitive species were

316 present, which made co-occurrence with other species less likely, then this could

317 generate sites with lower species richness. However, there is a second possible

318 explanation for the tendency of matrices with high variance in richness to give

319 significant results.

320 This alternative explanation is that variation in species richness is not related to non-

321 random co-occurrence between species, and hence this variation has been incorrectly

322 excluded from the null model, giving rise to false positive results (Type 1 errors). I

323 explored this possibility by rerunning the same null model analyses, but with matrices

324 pre-randomised, while maintaining species richness per site, and occurrences per

15

325    species. These tests gave similar results to those conducted on the real datasets. Of

326    218 real datasets that tested significant using the fixed-equiprobable randomisation

327    algorithm, 188 of these (86%) remained significant even when all structure other than

328    species richness differences had been removed through a preliminary fixed-fixed

329    randomisation (Figures 1a vs Figure 1c). It is therefore unclear whether the significant

330    results for real datasets are driven entirely by variation in species richness between

331    sites (unrelated to co-occurrence patterns between particular pairs of species), or if

332    species really do co-occur non-randomly, and in the process of doing so, create

333    variation in species richness. If the latter were the case, then one would expect that for

334    matrices showing significantly low C-scores in the unconstrained analysis (e.g. where

335    some hypothetical multi species mutualist had caused aggregations of species at high

336    richness sites), then these significantly low C-scores should still be detectable even

337    when richness distributions are fixed. I did not find this to be the case, since of the 205

338    matrices found to have a lower C-score than expected at random in the unconstrained

339    analysis, only 2 (1%) also gave the same result in the fixed-fixed analysis. Furthermore,

340    the results of the second set of tests are cause for concern, since they show that it is

341    possible to incorrectly detect non-random co-occurrence patterns in random datasets if

342    there is variation in richness between sites that is not accounted for in the null model.

343    Hence for any particular null modelling exercise, if species richness is not included in

344    the model, any significant results might be due either to non-random co-occurrence

345    between species (the usual interpretation) or to variation in species richness that has

346    not been included in the model. The fact that my results were similar between real and

16

347    pre-randomised datasets indicates that the latter explanation is more likely to be correct

348    in this case.

349    Why does variation in species richness (in the absence of non-random species co-

350    occurrence) give rise to significantly different C-scores in the observed datasets from

351    those predicted by the null model? I suggest that this relates to the way that variation in

352    species richness between sites constrains the number of possible checkerboard units

353    (Figure 3). For a given matrix, if variation in species richness is high, there are few

354    possibilities for the formation of checkerboard units. However, species richness

355    resulting from equal probability of site occupation (in this case represented as an

356    extreme situation in which species richness is uniform, although note that if site

357    occupancy is equiprobable, this would still lead to a case with variation in richness

358    between sites) results in larger numbers of checkerboard units (on average). This holds

359    for both structured and randomised matrices. Hence if a real matrix with high variation in

360    species richness is randomised to create a null model under the assumption that

361    species richness has lower variation, then the null model will have more checkerboard

362    units than the observed dataset, independently of any non-random co-occurrence

363    between species. In support of this speculation, an increase in C-scores for the real

364    matrices used in this study was also observed when randomised using the c0 algorithm,

365    but not when using fixed-fixed algorithms (Figure 4). Consequently, the real dataset will

366    be interpreted as having a lower C-score than would be expected at random. Such

367    results have been observed previously (Table 2 in Ulrich & Gotelli 2007), although have

368    not been interpreted specifically in terms of the effects of variation in species richness

369    between sites within a dataset. This potentially explains the large proportion of matrices

17

370   for which the C-score was significantly lower (i.e. fewer checkerboard units) than would

371   be expected under random community assembly in my analyses (Figure 1 A,C).

372   Therefore, finding a significant result for a null model that does not explicitly account for

373   variation in species richness could mean either that species co-occur non-randomly with

374   respect to each other, or simply that species richness is more variable between sites

375   than has been accounted for in the model (note that these explanations are not mutually

376   exclusive).

377   In order to avoid this issue, one approach is to constrain the null model species

378   richnesses to those from the observed dataset. This allows detection of non-random co-

379   occurrence within the pre-existing species richness distribution. Using this approach, I

380   found that standardised effect size no longer correlated with variance in species

381   richness, the proportion of significant results was lower, and the majority of those that

382   were significant indicated that the observed dataset had a higher C-score than would be

383   expected at random. Worryingly, of the 289 real datasets, 96 (33%) showed significant

384   but *opposing*, patterns of species co-occurrence when the two different randomisation

385   algorithms were used. For all of these, the C-score was lower than expected when

386   using the fixed-equiprobable algorithm, and higher than expected when using the fixed-

387   fixed algorithm. This indicates that some patterns of segregation between species were

388   previously masked by not accounting for variation in species richness, which instead

389   resulted in reporting of patterns of species aggregation. Effectively, assuming that

390   species occurrence probability does not differ between sites in a null model results in

391   real high richness sites being interpreted as species aggregations, while constraining

392   null model species richness to that observed allows for the possibility that within a

393   varying richness distribution, there might be some pairs of species that are segregated

394   (or aggregated). This mechanism potentially explains previous results in which the

395   direction of the reported species co-occurrence pattern depended on the randomisation

396   algorithm used (Hétérier et al. 2008; Rooney 2008). My result also emphasizes the

397   importance of specifying the null hypothesis under consideration (and hence choice of

398   algorithm use), since the two most commonly used choices (fixed-equiprobable and

399   fixed-fixed) give strong effects in the opposite direction for one in three datasets.

400   Taken together, these results are a cause for concern, because they suggest that the

401   outcome of null modelling of species co-occurrence can depend greatly on the

402   randomisation algorithm used. The same data could hence be interpreted as indicating

403   entirely different patterns, for example species segregation as a result of competition,

404   rather than species aggregation as a result of facilitation. Matrices that show high

405   variation in species richness between sites are always likely to give significant results

406   when tested against a null model that assumes low variance in richness, regardless of

407   the way that pairs of species co-occur. This kind of bias may also explain the high type

408   1 error rates of algorithms that do not constrain species occurrences (73-76%; Gotelli

409   2000), since observed matrices may have high variation in species occurrences (i.e.

410   they have skewed relative occurrence distributions). Although I only tested these effects

411   with the C-score metric, it is likely that these differences in richness variation between

412   observed and simulated matrices will affect the calculation of other metrics of co-

413   occurrence. Researchers need to think carefully about how the value of their co-

414   occurrence metric of choice is likely to change when matrices are randomised.

415   However, the possibility that variation in species richness might be generated by

19

416    species interactions should also not be discounted, and understanding the links

417    between pairwise species interactions and larger scale community patterns is likely to a

418    fruitful avenue of future research. Indeed, an alternative approach that avoids the

419    problems described here is to construct a series of hypothesised mechanistic models

420    (as well as a null model) and use the results of these models as support for presence or

421    absence of particular species assembly rules (Fayle et al. 2015).

422    In line with previous recommendations (Gotelli 2000; Ulrich & Gotelli 2007) researchers

423    should have a strong rationale for using any randomisation algorithm other than one

424    that fixes both the species richness distribution, and the number of occurrences per

425    species. The analyses here demonstrate that use of multiple randomisation algorithms

426    on the same datasets can be useful, if outputs are interpreted carefully. More broadly,

427    these results illustrate how the comparison of a single alternative hypothesis to a null

428    model is not a strong analytical strategy, since multiple mechanisms might cause

429    deviations from the null model, including those not considered as part of the alternative

430    hypothesis. A more powerful approach is the postulation of multiple competing

431    alternative hypotheses, and the construction of mechanistic models of species co-

432    occurrence based on these hypotheses to allow direct comparison of their explanatory

433    power (Gotelli & Ulrich 2012).

434    **Acknowledgements**

437     the anonymous reviewers who have provided feedback on this manuscript. Funding

438     was provided by a standard grant from the Czech Science Foundation (19-14620S).

439     **Data Accessibility**

440     All datasets used in this paper were downloaded as part of the nestedness temperature

441     programme (to which TMF is not affiliated), which is no longer maintained, but remains

442     available in archived form

443     (https://web.archive.org/web/20160819062742/https://www.fieldmuseum.org/nestednes

444     s-temperature-calculator-program). All model outputs and code are available online at

445     www.zenodo.org (https://doi.org/10.5281/zenodo.4244215).

446     **References**

447     Atmar W, and Patterson BD. 1993. The measure of order and disorder in the distribution of

448         species in fragmented habitat. *Oecologia* 96:373-382.

449     Bell JR, King RA, Bohan DA, and Symondson WOC. 2010. Spatial co-occurrence networks

450         predict the feeding histories of polyphagous arthropod predators at field scales.

451         *Ecography* 33:64-72.

452     Colwell R, and Winkler D. 1984. A null model for null models in biogeography. *Ecological*

453         *communities: conceptual issues and the evidence* 440:344-359.

454     Connor EF, and Simberloff D. 1979. The Assembly of Species Communities: Chance or

455         Competition? *Ecology* 60:1132-1140.

456     De los Ríos P, Rivera N, and Galindo M. 2008. The use of null models to explain crustacean

457         zooplankton associations in shallow water bodies of the Magellan region, Chile.

458         *Crustaceana* 81:1219-1228.

459    Fayle TM, Eggleton P, Manica A, Yusah KM, and Foster WA. 2015. Experimentally testing and

460         assessing the predictive power of species assembly rules for tropical canopy ants.

461         *Ecology Letters* 18:254-262.

462    Fayle TM, and Manica A. 2010. Reducing over-reporting of deterministic co-occurrence patterns

463         in biotic communities. *Ecological Modelling* 221:2237-2242.

464    Fayle TM, and Manica A. 2011. Bias in null model analyses of species co-occurrence: A

465         response to Gotelli and Ulrich (2011). *Ecological Modelling* 222:1340-1341.

466    Gotelli NJ. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81:2606-2621.

467    Gotelli NJ, and Ulrich W. 2011. Over-reporting bias in null model analysis: A response to Fayle

468         and Manica (2010). *Ecological Modelling* 222:1337-1339.

469    Gotelli NJ, and Ulrich W. 2012. Statistical challenges in null model analysis. *Oikos* 121:171-180.

470    Hétérier V, David B, De Ridder C, and Rigaud T. 2008. Ectosymbiosis is a critical factor in the

471         local benthic biodiversity of the Antarctic deep sea. *Mar Ecol Prog Ser* 364:67-76.

472    Holdaway RJ, and Sparrow AD. 2006. Assembly rules operating along a primary riverbed-

473         grassland successional sequence. *Journal of Ecology* 94:1092-1102.

474    Holt BG, Lessard J-P, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, Fabre P-H, Graham

475         CH, Graves GR, Jønsson KA, Nogués-Bravo D, Wang Z, Whittaker RJ, Fjeldså J, and

476         Rahbek C. 2013. An Update of Wallace's Zoogeographic Regions of the World. *Science*

477         339:74-78.

478    Hubbell SP. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton:

479         Princeton University Press.

480    Miklós I, and Podani J. 2004. Randomization of presence-absence matrices: comments and

481         new algorithms. *Ecology* 85:86-92.

482    Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Dan McGlinn, Minchin PR, O'Hara

483         B, Simpson GL, Solymos P, Stevens MHH, Szoecs E, and Wagner H. 2018. vegan:

484     Community Ecology Package. R package version 2.5-1.    https://CRAN.R-

485     project.org/package=vegan.

486     Pitta E, Giokas S, and Sfenthourakis S. 2012. Significant Pairwise Co-occurrence Patterns Are

487     Not the Rule in the Majority of Biotic Communities. *Diversity* 4:179-193.

488     Rooney TP. 2008. Comparison of co-occurrence structure of temperate forest herb-layer

489     communities in 1949 and 2000. *acta oecologica* 34:354-360.

490     Stone L, and Roberts A. 1990. The checkerboard score and species distributions. *Oecologia*

491     85:74-79.

492     Strona G, Nappo D, Boccacci F, Fattorini S, and San-Miguel-Ayanz J. 2014. A fast and

493     unbiased procedure to randomize ecological binary matrices with fixed row and column

494     totals. *Nature Communications* 5:4114.

495     Ulrich W, Baselga A, Kusumoto B, Shiono T, Tuomisto H, and Kubota Y. 2017. The tangled link

496     between β-and γ-diversity: a Narcissus effect weakens statistical inferences in null

497     model analyses of diversity patterns. *Global Ecology and Biogeography* 26:1-5.

498     Ulrich W, and Gotelli NJ. 2007. Disentangling community patterns of nestedness and species

499     co-occurrence. *Oikos* 116:2053-2061.

500     Ulrich W, Kubota Y, Kusumoto B, Baselga A, Tuomisto H, and Gotelli NJ. 2018. Species

501     richness correlates of raw and standardized co-occurrence metrics. *Global Ecology and*

502     *Biogeography* 27:395-399.

503     Wilson JB. 1995. Null models for assembly rules - the Jack-Horner effect is more insidious than

504     the Narcissus effect. *Oikos* 72:139-144.

505
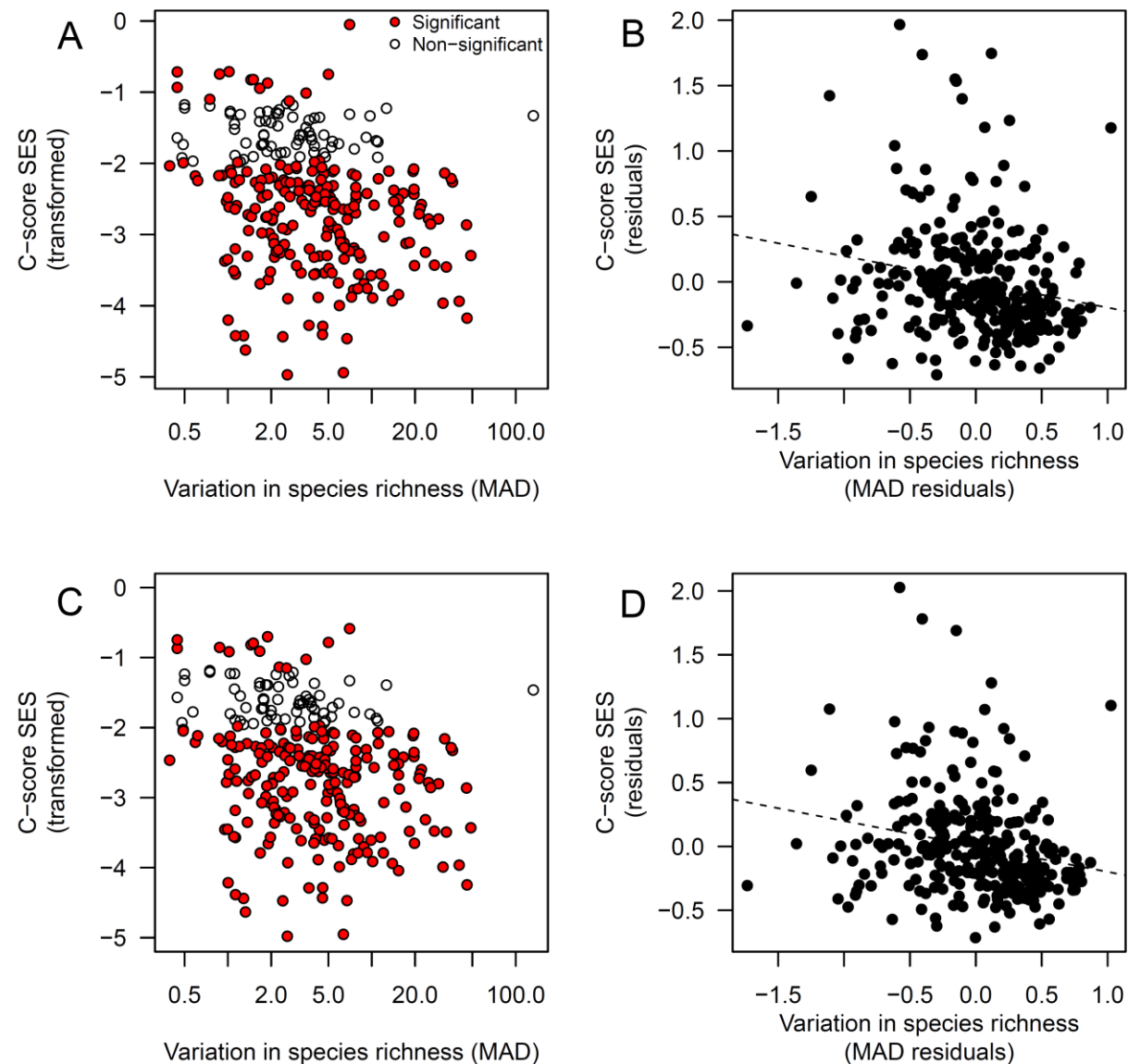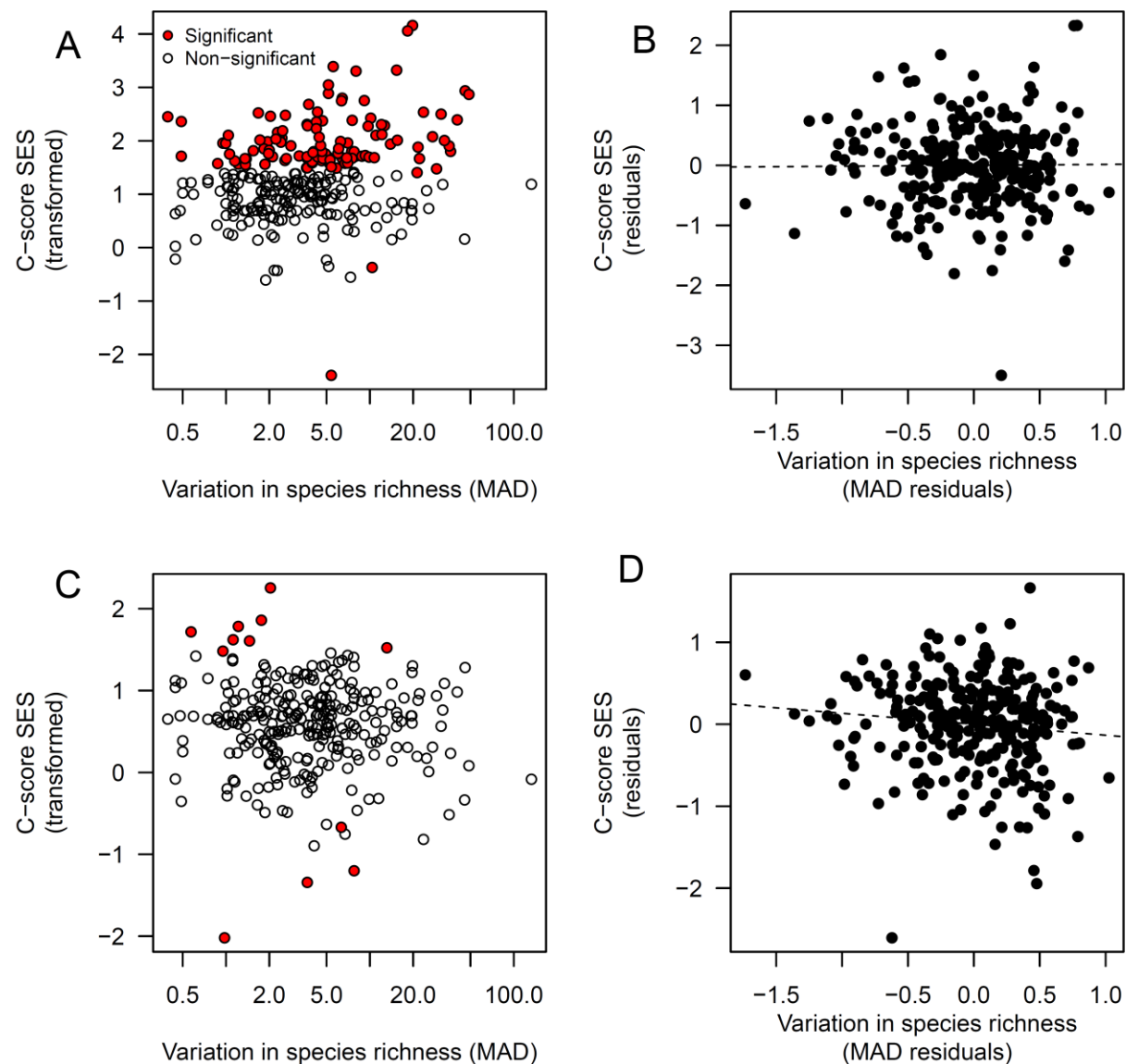
506

507  **Figures**



508

509  Figure 1. Relationship between variability in species richness and deviation from

510  random species assembly when observed variation in species richness is not accounted

511  for in the null model. (A) With increasing variability in species richness between

512  locations within a site (Mean Absolute Deviation, MAD), the detected strength of

513  deviation from random species assembly in real matrices increases when this variation

514     in species richness is not included in the model (fixed-equiprobable randomisation

515     algorithm). Red points denote statistically significant deviations from random species

516     assembly. (B) This is still the case when the effects of number of sites, number of

517     species and matrix fill are taken into account (partial regression plot). (C) When the

518     species richness and number of sites are fixed to those from the 289 real matrices, but

519     with species being placed at random to each other to create datasets, the same

520     negative relationship between MAD and SES is observed. (D) Partial regression plot

521     displaying this pattern.

522

523

Figure 2. Relationship between variability in species richness and deviation from

random species assembly when observed variation in species richness is accounted for

in the null model. (A) When variation in species richness is included explicitly in the

model of species co-occurrence (fixed-fixed randomisation algorithm) then for real

matrices there is no relationship between variability in species richness between

locations and the detected strength of deviation from random species assembly. (B)

530    There is also no relationship when the effects of number of sites, number of species and

531    matrix fill are taken into account (partial regression plot). (C) This is also true for the

532    datasets for which the species richness and number of sites were fixed to those from

533    the 289 observed matrices, but with species being placed at random to each other. (D)

534    Partial regression plot displaying this pattern. Note that p-values are calculated from the

535    raw null distribution, while standardised effect sizes are calculated from the variance in

536    the null distribution. This means that occasionally some data points with more extreme

537    SES values may be non-significant, while those with less extreme values are significant

538    (e.g. in panel A).

539

540
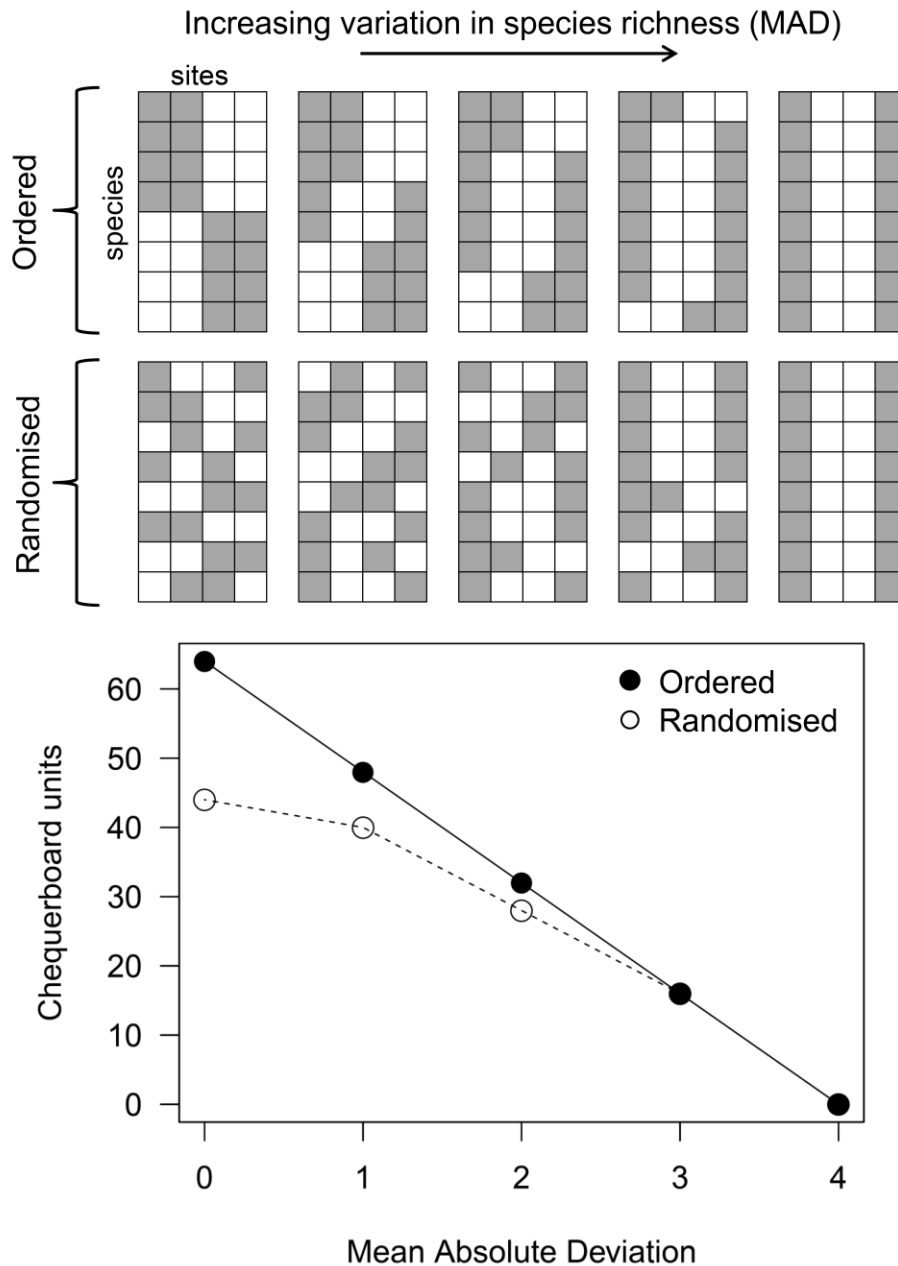
Figure 3. A demonstration for some small hypothetical datasets of the impact of variation in species richness on the possible number of checkerboard units. Assume that we have eight species distributed across four sites, and that each species always occurs at only two sites. Variation in site species richness can vary to give a range of

545    values of mean absolute deviation (MAD), from zero to four. For both completely

546    segregated matrices and randomised matrices the number of checkerboard units

547    decreased with increasing variation in species richness. Note that median numbers of

548    checkerboard units for 10,000 replications are plotted for randomised matrices, while

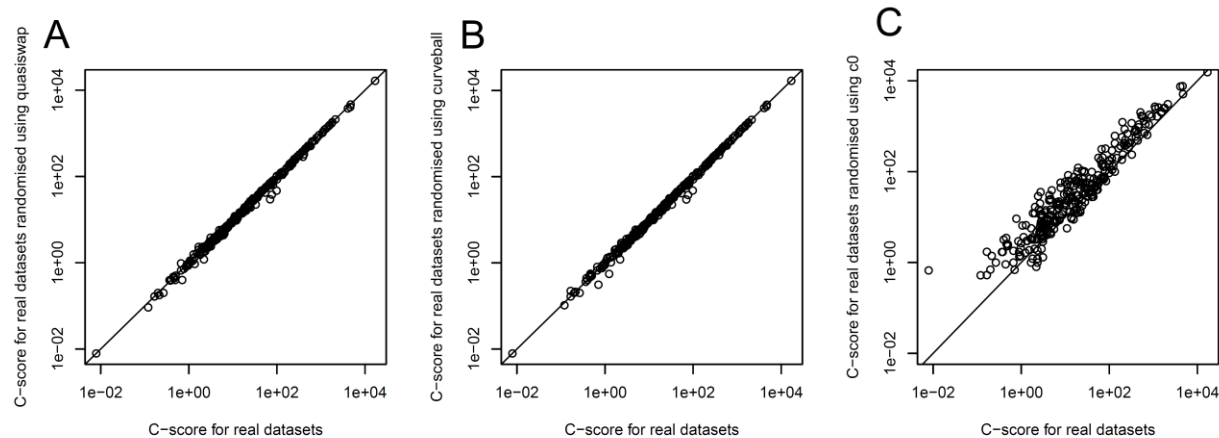549    only single example randomised matrices are plotted above.

550

551

Figure 4. Impacts of randomisation on C-scores for real datasets for three algorithms.
Relationship between C-scores for real datasets and single randomisations carried out
using the two fixed-fixed algorithms quasiswap (A) and curveball (B), and the fixed-
equiprobable algorithm c0 (C). Lines denote the scenario in which randomisation does
not affect the C-score. Note the large proportion of points above the line when datasets
are randomised using the c0 algorithm.

558

559  Table 1. Influence of restricting randomisation algorithm to observed species richness

560  across sites on significance and direction of results from null models of co-occurrence.

561  The effects of including species richness in a null model of species co-occurrence

562  (fixed-fixed randomisation algorithm) or not (fixed-equiprobable algorithm) on

563  frequencies of significant test results for real datasets. Where the observed C-score is

564  significantly higher than the null distribution of C-scores ($C_{obs}>C_{null}$), then there is more

565  segregation between species than expected at random, and vice versa for $C_{obs}<C_{null}$.

| | | Species richness included in null model | | | Totals |
|---|---|---|---|---|---|
| | | $C_{obs}<C_{null}$ (P<0.05) | P>0.05 | $C_{obs}>C_{null}$ (P<0.05) | |
| Species richness not included in null model | $C_{obs}<C_{null}$ (P<0.05) | 2 | 107 | 96 | 205 |
| | P>0.05 | 0 | 58 | 13 | 71 |
| | $C_{obs}>C_{null}$ (P<0.05) | 0 | 11 | 2 | 13 |
| Totals | | 2 | 176 | 111 | 289 |

566