1 **Cryptic matters: Overlooked species generate most butterfly beta-diversity**

2

3

4 Raluca Vodă^{a,b}, Leonardo Dapporto^c, Vlad Dincă^{a,d,e}, Roger Vila^{a,*}

5

6

7 ^aInstitut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la

8 Barceloneta 37, 08003, Barcelona, Spain

9 ^bDepartament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, 08193, Bellaterra,

10 Spain

11 ^cDepartment of Biological and Medical Sciences, Oxford Brookes University, Headington, Oxford,

12 OX3 0BP, UK

13 ^dDepartment of Zoology, Stockholm University, S-106 91, Stockholm, Sweden

14 ^eBiodiversity Institute of Ontario, University of Guelph, Guelph, N1G 2W1, Ontario, Canada

15

16 * roger.vila@csic.es.

17

18

19

20 **Abstract**

21

22 The cryptic fraction of biodiversity is composed of morphologically similar species that are or have
23 been overlooked by scientists. Although current research is increasingly documenting new cases,
24 cryptic species are frequently ignored in large-scale studies and monitoring programs, either
25 because they have not yet been discovered, or because of the practical difficulties involved in
26 differentiating them. However, it is unknown if this could represent a bias extending beyond the
27 number of missed species. By analyzing the butterfly fauna of the West Mediterranean (335
28 species), we defined cryptic species based on the current consensus of the scientific community,
29 compared their properties to other congeneric species and investigated the consequences of their
30 inclusion/exclusion in beta-diversity analyses. We show that, as defined, the cryptic fraction of
31 butterfly diversity represents about 25% of the West Mediterranean fauna and is overwhelmingly
32 composed by groups of species that are not sympatric. Our results show that co-occurrence among
33 cryptic species is significantly lower than among congeneric non-cryptic species. Accordingly,
34 albeit the frequency of cryptic species is homogenously distributed over the study area, their
35 distribution pattern accounts for most beta-diversity turnover over sea (from 50 to 100%). Beta-
36 diversity turnover, a direct measure of the frequency of species replacement from site to site, is
37 recognized as a fundamental parameter in ecology and is widely used to detect biogeographic
38 patterns. These findings represent a change of paradigm in showing that cryptic diversity comprises
39 original qualitative aspects in addition to merely quantitative ones. This highlights the importance
40 of differentiating cryptic species for various research fields and opens the door to the study of
41 further potential particularities of cryptic diversity.

42

43 **Keywords:** cryptic species, beta-diversity turnover, island biogeography, Lepidoptera

44 **Introduction**

45 Recent molecular studies unveiled a substantial fraction of biodiversity by recognizing a plethora of
46 cryptic species in numerous taxonomic groups and across biogeographical regions (Pfenninger and
47 Schwenk 2007). Cryptic species (Steyskal 1972) are those that have been erroneously treated as
48 conspecific because of high morphological resemblance, and sometimes also behavioral or
49 ecological similarity (Bickford et al. 2007). Thus, the concept of cryptic species is anthropocentric
50 by definition because it is based on human perception of species. Irrespective of the evolutionary
51 significance of delimiting such a group of species, it is justified to ask about the practical
52 consequences of their discovery or omission. To date, no specific studies directly compared the
53 properties of cryptic and non-cryptic diversity, perhaps due to the difficulty of unambiguously
54 identifying cryptic species and their precise distributions for a sufficiently large area and taxonomic
55 group. Consequently, researchers frequently simplify their datasets by not distinguishing cryptic
56 species (Dennis 1997, Isaac et al. 2004, Pfenninger and Schwenk 2007, Descimon and Mallet 2009,
57 Devictor et al. 2012).

58 Due to their similar natural history and appearance, cryptic species are predicted to compete for
59 resources and to reproductively interfere between them (Pigot and Tobias 2013). Results for a series
60 of cryptic butterfly species revealed only restricted contact zones on mainland and chequered
61 distributions over narrow (3-10 km) sea straits (Dincă et al. 2011, Dapporto et al. 2012, Sañudo-
62 Restrepo et al. 2013). If chequered distributions are overrepresented among cryptic species, they
63 could encompass a disproportionately high fraction of beta-diversity turnover, a direct measure of the
64 frequency of species replacement from site to site. Spatial turnover is recognized as a fundamental
65 parameter in ecology and is increasingly used to detect biogeographic patterns (Holt et al. 2013).
66 To assess the effects of ignoring cryptic diversity, we studied the biogeographical properties of
67 cryptic groups of species uncovered since 1970, after the cornerstone work of Higgins and Riley
68 (1970), that are now generally accepted by the scientific community. In particular, we explored the

69 prevalence of chequered distribution patterns among groups of cryptic taxa in West Mediterranean
70 butterflies, which represent one of the best-studied invertebrate groups, especially regarding their
71 taxonomy and distribution. We also assessed if cryptic taxa account for an unbalanced fraction of
72 the beta-diversity turnover and concluded that, although they are frequently excluded from large-
73 scale surveys and studies, their contribution may be far more important than the mere percentage of
74 species they represent.

75

76 **Methods**

77 **Presence data and co-occurrence analyses**

78 We analyzed butterfly presence data from 82 areas comprising 27 small islands and 55 sites of 25
79 km radius located on mainland (43) and on large islands (12) (Supplementary material, Dataset 1).
80 An objective identification of species is a crucial step for comparative studies at continental level,
81 especially on a diverse group such as butterflies. Although European butterflies have been among
82 the most studied invertebrates in the last centuries, recognition of species is not always
83 straightforward and it has been based on different concepts for various taxa. In some cases species
84 are readily recognizable based on their external morphology, while in other cases, in depth
85 morphological and molecular analyses, as well as the assessment of reproductive barriers, have
86 been needed (see Descimon and Mallet 2009 for a review). For these reasons, we used a taxonomic
87 species concept (see Mayden 1997 for a review) recognizing species as ‘whatever a competent
88 taxonomist chooses to call a species’ (Wilkins 2009). We applied this criterion and compared past
89 taxonomy with the current consensus of the scientific community in the following way:

90 1. We produced a list of species based on the widely accepted checklist of the Fauna Europaea
91 Project (Karsholt and Nieuwerkerken 2013, <http://www.faunaeur.org>), funded by the European
92 Commission (EVR1-1999-2001). We complemented this list with the North African species
93 recognized in the recent synthesis works that include this region (Tennent 1996, Tarrier and Delacre

2008, Tshikolovets 2011). We also added four species (*Zerynthia cassandra*, *Polyommatus celina*,
Coenonympha lyllus, *Coenonympha darwiniana*) and eliminated another one (*Polyommatus galloi*)
based on recent molecular studies, generally combined with morphometrics, mate choice
experiments and/or ecological data (see Appendix 1 for references). These changes have not yet
been introduced to Fauna Europaea but will most likely be in the next version.

2. Within the list of species obtained, we considered as cryptic those pairs or groups of species that
have been considered as a single species in at least one of the most widely used synthesis works
regarding Europe and/or North Africa published after the reference work of Higgins and Riley
(1970) (see Appendix 1 for discussions on each case). The references provided should be regarded
as examples and usually cover general works that likely represent widely accepted trends in
European butterfly taxonomy. In specific cases where relevant studies were available, these have
been cited as well.

Co-occurrence between pairs of species has been computed as the fraction of shared areas by using
the Sørensen index $QS=2C/(A+B)$, where C is the number of areas where the two species co-occur
and A and B are the number of areas where the first and second species are present, respectively.
We computed the QS between pairs of cryptic species and between pairs of non-cryptic species
belonging to the same genus. To test for significant differences in QS values we performed the
Mann-Whitney U test with 1000 Monte Carlo permutations.

Beta-diversity turnover

To evaluate the Contribution of Cryptic Species to the beta-diversity Turnover (CCST), we
calculated the Simpson index among the 82 study areas. Subsequently, for each presence submatrix
composed of each cryptic group, we computed a matrix (sm) for areas by using the Simpson
dissimilarity denominator. Finally, we added all sm and divided this sum matrix by the total

118 Simpson matrix (also containing the contribution of non-cryptic species). Following the previous
119 description, CCST can be computed by the formula:

$$120 \quad CCST = \frac{\sum_{sm=1}^{36} (A_{sm} + \min(B_{sm}, C_{sm}))}{A - \min(B, C)},$$

121 where A_{sm} is the number of shared species and B_{sm} and C_{sm} are the number of non-shared species
122 among areas for each of the 36 cryptic groups considered separately (sm). A is the overall number
123 of shared species and B and C the overall number of non-shared species among the same pairs of
124 the 82 areas. With the “deldir” R package we calculated a Delaunay triangulation among areas
125 using their center GPS coordinates and plotted the CCST index on a map. The CCST among pairs
126 of areas were attributed to the midpoints connecting those areas and were interpolated through
127 Inverse-Distance-Weighting using Spatial Analyst (www.esri.com).

128 To evaluate the importance of cryptic species in determining the overall turnover pattern, we
129 calculated the mean value of the ratio between the Simpson and the Sørensen dissimilarity matrices
130 (β_{simp}/β_{sor}) among areas. This ratio is considered to be the best measure describing the
131 contribution of turnover to the overall beta-diversity pattern (Dobrovolski et al. 2012). We then
132 removed all cryptic taxa and recalculated the same mean value. Finally, we created 10,000 matrices
133 by removing a random number of species equal to the number of cryptic ones and calculated their
134 mean β_{simp}/β_{sor} . We tested if the matrix lacking all cryptic taxa shows a significantly lower ratio
135 than the random ones (if at least 95% of the random matrices showed a β_{simp}/β_{sor} higher than the
136 matrix without all cryptic species).

137

138 **Results**

139 **Co-occurrence analyses for cryptic and non-cryptic diversity fractions**

140 In total, 86 species out of 335 (25.67%) are involved in 36 cryptic groups. In most groups (20),
141 cryptic species have been separated only on the basis of differences in morphology (mostly male
142 genitalia), while in six groups species were recognized based exclusively on molecular evidence
143 (DNA, allozyme or number of chromosomes) because no relevant differences in morphology have
144 been detected. In nine groups morphological and molecular evidence have been coupled, and in one
145 group (*Leptidea*) morphological and molecular analysis have been combined with mating
146 experiments (Appendix 1). The Mann-Whitney test ($U = 15319.5$, Monte Carlo $P < 0.001$) revealed
147 that cryptic species pairs have significantly lower QS than non-cryptic pairs of species (Fig. 1),
148 which means that they share a smaller fraction of their distributions.

149

150 **Contribution of non-sympatric cryptic species to the overall beta-diversity pattern**

151 Although the fraction of cryptic diversity was relatively constant among the 82 study areas ($23.02 \pm$
152 0.04 , Fig. 2a), the fraction of beta-diversity turnover due to replacements between cryptic species
153 showed pronounced variation over the West Mediterranean (Fig. 2b). As a general rule, the CCST
154 was higher than 50% only over sea areas, with especially high values of more than 80% between
155 Sardinia-Corsica-Italian mainland, Sicily-Italian mainland, Sicily-North Africa and Balearics-
156 Catalonia-France. By contrast, the CCST was lower (about 10%) between sites separated by land
157 (Fig. 2b). The complete presence matrix revealed a mean $\beta_{\text{simp}}/\beta_{\text{sor}}$ value of 0.450 (Fig. 2c). When
158 all the 86 cryptic species have been removed the mean value decreased to 0.346, while the mean
159 value for cryptic species alone was 0.666. None of the 10,000 matrices generated by removing 86
160 random species showed a $\beta_{\text{simp}}/\beta_{\text{sor}}$ mean value lower than the matrix lacking the 86 cryptic
161 species ($P < 0.001$, Fig. 2c).

162

163 **Discussion**

164 Due to their morphological similarity and complex taxonomy, cryptic groups are frequently ignored
165 or treated as a unit in large-scale studies and monitoring schemes. In this study we objectively
166 defined cryptic species and showed that co-occurrence within cryptic groups over the West
167 Mediterranean is much lower than among congeneric non-cryptic species. This result could be
168 explained if we assume that cryptic groups are composed by entities that recently evolved in
169 allopatry and maintained their distribution pattern simply because they did not have enough time to
170 disperse. However, many cryptic taxa in our study are not sister species and show substantial
171 genetic divergence, while other morphologically differentiated species are their sister taxa; in most
172 such cases a chequered distribution still occurs (Dincă et al. 2011, Sañudo-Restrepo et al. 2013).
173 The mechanisms maintaining chequered distributions across specific narrow sea straits are still
174 unknown, but a convincing hypothesis can be formulated. Recent studies have shown that
175 butterflies can experience range expansions/contractions even over relatively short time periods
176 (Dapporto and Bruschini 2012, Dapporto et al. 2012). It is thus likely that, in addition to temporal
177 and physical constraints, the dispersal of most cryptic species has been hampered by the presence of
178 other members of the same group in the recipient areas (Waters et al. 2013), a phenomenon recently
179 demonstrated in birds (Pigot and Tobias 2013). Presumably, no strong precopulatory barriers exist
180 between many cryptic butterfly species (Wickman 2009) and hybrids between related taxa often
181 display reduced fitness (Descimon and Mallet 2009). In the absence of specific mate recognition
182 mechanisms, contact areas may act as population sinks unlikely to enlarge, although they are
183 predicted to shift until areas allowing low dispersal and low population densities are reached
184 (Dasmahapatra et al. 2010). A concentration of boundaries between cryptic species over sea straits
185 perfectly matches this hypothesis. Noteworthy, the fraction of cryptic diversity is homogeneous
186 across the study region and almost all cryptic taxa that we analyzed occur on both mainland and
187 islands. For this reason, the striking contribution to island divergence is not due to an unbalanced
188 occurrence of island endemics among cryptic taxa. These mechanisms are generally applicable to

189 most organisms, and thus it is probable that the phenomenon here discovered for butterflies holds
190 for other taxa as well, albeit the exact impact on the overall patterns may vary.

191 We conclude that unveiling the cryptic fraction of biodiversity is of primary importance for a
192 comprehensive description of biogeographical patterns. The finding that cryptic groups of species
193 present characteristic geographical distributions highlights the need to study further potential
194 particularities they may display, as well as to assess their implications in multiple research fields.

195 We showed that ignoring cryptic species results in losing a significant fraction of spatially
196 structured taxonomic variation, which hampers the opportunity to reveal phylogeographic suture
197 zones (Moritz et al. 2009) and to investigate historical and ecological processes determining
198 distribution range dynamics (Marske et al. 2013). Such phenomena underlie fundamental processes
199 like dispersal, extinction, environmental filtering, competition, sexual interference, hybrid
200 depression, and "founder takes all" dynamics (e.g. Marske et al. 2013, Waters et al. 2013).

201 Moreover, it is likely that cryptic species respond differently to climatic conditions (e.g. Pearman et
202 al. 2010) and geographical shifts in their contact zones correlated with climate change have been
203 documented (Mercader and Scriber 2008, Mallet et al. 2011). This suggests that studies quantifying
204 the effects of global climatic changes on organisms may be missing part of the signal stemming
205 from cryptic biodiversity. The same could apply to life history and other ecological preferences that
206 may be differentially expressed in morphologically similar species. Morphological differentiation
207 may well be the result of diverging selective pressures, which could generally lack in these groups
208 of species. Different evolutionary and speciation factors and processes may result in varying
209 degrees of morphological differentiation (*Heliconius* genome consortium 2012), and thus
210 phylogenetic reconstruction and macroecological inference could be affected due to
211 underrepresented cryptic diversity.

212 Case studies typically rely on well-assessed taxonomy, but macroecology and comparative studies
213 are usually based on data that encompass a great number of species and are more susceptible of

214 missing cryptic taxa. The costs and benefits of taking into consideration cryptic species have been
215 rarely discussed due to a reductionist view of their importance limited to alpha-taxonomy, i.e.
216 species numbers (Isaac et al. 2004). From this perspective, our study not only provides novel
217 insights into biogeography, but it also serves as a heads-up about the importance of including
218 the cryptic diversity in ecological, biodiversity and biogeographic studies to avoid biased
219 results.

220

221 **Acknowledgments**

222 Funding was provided by the Spanish Ministerio de Ciencia e Innovación (Project CGL2010-
223 21226/BOS and a predoctoral fellowship to R. Vodă, ref. AP-2010-5409), by the project
224 "Barcoding Italian butterflies" and by a postdoctoral fellowship from the Wenner-Gren Foundation
225 Sweden and a EU Marie Curie International Outgoing Fellowship for Career Development (project
226 no. 625997) to V. Dincă.

227

228 **References**

229

- 230 Bickford, D. et al. 2007. Cryptic species as a window on diversity and conservation. – Trends Ecol.
231 Evol. 22: 148–155.
- 232 Dapporto, L. and Bruschini, C. 2012. Invading a refugium: post glacial replacement of the ancestral
233 lineage of a Nymphalid butterfly in the West Mediterranean. – Org. Divers. Evol. 12: 39–49.
- 234 Dapporto, L. et al. 2012. Identifying zones of phenetic compression in West Mediterranean
235 butterflies (Satyrinae): refugia, invasion and hybridization. – Divers. Distrib. 18: 1066–1076.
- 236 Dasmahapatra, K.K. et al. 2010. The anatomy of a 'suture zone' in Amazonian butterflies: a
237 coalescent-based test for vicariant geographic divergence and speciation. – Mol. Ecol. 19: 4283–
238 4301

239 Dennis, R.L.H. 1997. An inflated conservation load for European butterflies: increases in rarity and
 240 endemism accompany increases in species richness. – J. Insect Cons. 1: 43–62
 241 Descimon, H. and Mallet, J. 2009. Bad species. – In: Settele, J. et al. (eds), Ecology of Butterflies in
 242 Europe. Cambridge University Press, Cambridge, pp 219–249.
 243 Devictor, V. et al. 2012. Differences in the climatic debts of birds and butterflies at a continental
 244 scale. – Nat. Clim. Change 2: 121–124.
 245 Dincă, V. et al. 2011. A combined genetic-morphometric analysis unravels the complex
 246 biogeographical history of *Polyommatus icarus* and *Polyommatus celina* Common Blue
 247 butterflies. – Mol. Ecol. 20: 3921–3935.
 248 Dobrovolski, R. et al. Melo 2012. Climatic history and dispersal ability explain the relative
 249 importance of turnover and nestedness components of beta diversity. – Global Ecol. Biogeogr.
 250 21:191–197.
 251 *Heliconius* Genome Consortium 2012. Butterfly genome reveals promiscuous exchange of mimicry
 252 adaptation among species. – Nature 487: 94–98.
 253 Higgins L.G. and Riley N.D. 1970. A field guide to the butterflies of Britain and Europe. – 1st
 254 American edn. Houghton Mifflin, Boston.
 255 Holt, B. et al. 2013. An Update of Wallace's Zoogeographic Regions of the World. – Science 339:
 256 74–78.
 257 Isaac, N.J.B. et al. 2004. Taxonomic inflation: its influence on macroecology and conservation. –
 258 Trends Ecol. Evol. 19:464–469.
 259 Mallet, J., et al. 2011. Hybridisation and climate change: brown argus butterflies in Britain
 260 (*Polyommatus* subgenus *Aricia*). – Insect Conserv. Divers. 4: 192–199.
 261 Marske, K.A. et al. 2013. Phylogeography: spanning the ecology-evolution continuum. –
 262 Ecography 36: 1169–1181.

263 Mayden, R.L. 1997. A hierarchy of species concepts: the denouncement in the saga of the species
 264 problem – In: Claridge, M.F. et al. (eds), *Species: The Units of Biodiversity*, Chapman & Hall,
 265 London (1997), pp. 381–424.

266 Mercader, R. J. and Scriber, J. M. 2008. Asymmetrical thermal constraints on the parapatric species
 267 boundaries of two widespread generalist butterflies. – *Ecol. Entomol.* 33: 537–545.

268 Moritz, C. et al. 2009. Identification and dynamics of a cryptic suture zone in tropical rainforest. –
 269 *Proc. R. Soc. B* 276: 1235–1244.

270 Pearman, P. B., et al. 2010. Within-taxon niche structure: niche conservatism, divergence and
 271 predicted effects of climate change. – *Ecography* 33: 990–1003.

272 Pfenninger, M. and Schwenk, K. 2007. Cryptic animal species are homogeneously distributed
 273 among taxa and biogeographical regions. – *BMC Evol. Biol.* 7.

274 Pigot, A.L. and Tobias, J.A. 2013. Species interactions constrain geographic range expansion over
 275 evolutionary time. – *Ecol. Lett.* 16: 330–338.

276 Sañudo-Restrepo, C.P. et al. 2013. Biogeography and systematics of *Arícia* butterflies (Lepidoptera,
 277 Lycaenidae). – *Mol. Phylogenet. Evol.* 66: 369–379

278 Steyskal, G.C. 1972. The meaning of the term "sibling species". – *Syst. Zool.* 21: 239.

279 Waters, J.M. et al. 2013 Founder takes all: density dependent processes structure biodiversity. –
 280 *Trends Ecol. Evol.* 28: 78-85.

281 Wickman, P.-O. 2009. Mating behaviour in butterflies. – In: Settele, J. et al. (eds), *Ecology of*
 282 *Butterflies in Europe*. Cambridge University Press, Cambridge, pp 17–28.

283 Wilkins, J.S. 2009. *Species: A History of the Idea*. – University of California Press, Berkeley.
 284
 285

286 **Figure Legends**

287

288 Figure 1. Boxplot indicating median values of the Sørensen index (QS) for pairs or groups of
289 cryptic species (Cryptic) and congeneric pairs of non-cryptic species (Non-cryptic). Cryptic species
290 share a significantly smaller fraction of their distributions compared with the non-cryptic ones.

291

292 Figure 2. (a) Distribution map (equirectangular projection) of the percentage that cryptic butterfly
293 species represent of the total diversity. (b) Landscape analysis for the Contribution of Cryptic
294 Species to the beta-diversity Turnover (CCST). (c) Mean $\beta_{\text{simp}}/\beta_{\text{sor}}$ value of the complete matrix
295 (red arrow) compared with the value obtained by removing all 86 cryptic species (black arrow) and
296 by using only cryptic species (cyan arrow). The histogram represents frequencies of $\beta_{\text{simp}}/\beta_{\text{sor}}$
297 mean values for 10,000 matrices generated by removing 86 random species.

298

299 **Supplementary Material:**

300 Appendix 1. Defining groups of cryptic species.

301 Dataset 1. Presence data for the species used in the analysis

302