Host plant use drives genetic differentiation in syntopic populations of *Maculinea alcon*

András Tartally, Andreas Kelager, Matthias A Fürst, David R Nash

The rare socially parasitic butterfly *Maculinea alcon* occurs in two forms, which are characteristic of hygric or xeric habitats, and which exploit different host plants and host ants. The status of these two forms has been the subject of considerable controversy. Populations of the two forms are usually spatially distinct, but at Răscruci in Romania both forms occur on the same site (syntopically). We examined the genetic differentiation between the two forms using eight microsatellite markers, and compared with a nearby hygric site, Şardu. Our results showed that while the two forms are strongly differentiated at Răscruci, it is the xeric form there that is most similar to the hygric form at Şardu, and Bayesian clustering algorithms suggest that these two populations have exchanged genes relatively recently. We found strong evidence for population substructuring, caused by high within host ant nest relatedness, indicating very limited dispersal of most ovipositing females, but not association with particular host ant species. Our results are consistent with the results of larger scale phylogeographic studies that suggest that the two forms represent local ecotypes specialising on different host plants, each with a distinct flowering phenology, providing a temporal rather than spatial barrier to gene flow.

1   **Host plant use drives genetic differentiation in syntopic populations of**

2   ***Maculinea alcon***

4   Authors

5   András Tartally [1,2*], Andreas Kelager[2,3*], Matthias A. Fürst[2,4], David R. Nash[2†].

7   Affiliations

8   1. University of Debrecen, Department of Evolutionary Zoology and Human Biology,

9   Hungary, H-4032, Debrecen, Egyetem tér 1.

10   2. Centre for Social Evolution, Department of Biology, University of Copenhagen.

11   Universitetsparken 15, DK-2100 Copenhagen, Denmark.

12   3. Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark,

13   University of Copenhagen. Universitetsparken 15, DK-2100 Copenhagen, Denmark.

14   4. IST Austria (Institute of Science and Technology Austria), Am Campus 1, 3400

15   Klosterneuburg, Austria.

17   * These authors contributed equally to this work

18   † Corresponding author.   E-mail: DRNash@bio.ku.dk

19   Tel: +45 353 21323

20   Fax: +45 353 21250

23   **Abstract**

24   The rare socially parasitic butterfly *Maculinea alcon* occurs in two forms, which are

25   characteristic of hygric or xeric habitats, and which exploit different host plants and host ants.

26   The status of these two forms has been the subject of considerable controversy. Populations of

27   the two forms are usually spatially distinct, but at Răscruci in Romania both forms occur on the

28   same site (syntopically). We examined the genetic differentiation between the two forms using

29   eight microsatellite markers, and compared with a nearby hygric site, Șardu. Our results showed

30   that while the two forms are strongly differentiated at Răscruci, it is the xeric form there that is

31   most similar to the hygric form at Șardu, and Bayesian clustering algorithms suggest that these

32   two populations have exchanged genes relatively recently. We found strong evidence for

33   population substructuring, caused by high within host ant nest relatedness, indicating very

34   limited dispersal of most ovipositing females, but not association with particular host ant species.

35   Our results are consistent with the results of larger scale phylogeographic studies that suggest

36   that the two forms represent local ecotypes specialising on different host plants, each with a

37   distinct flowering phenology, providing a temporal rather than spatial barrier to gene flow.

39   **Introduction**

40   Larvae of *Maculinea* van Eecke (Lepidoptera: Lycaenidae) butterflies start their development on

41   specific host plants. A few weeks later they are adopted into the nests of suitable *Myrmica*

42   Latreille (Hymenoptera: Formicidae) colonies, where they act as social parasites of the ants

43   (Thomas et al., 1989). This unusual life cycle has shaped their evolution, as different populations

44   are strongly selected to adapt to different initial host plants and *Myrmica* species depending on

45   their availability (Thomas et al., 1989; Witek et al., 2008).

46   Larvae of the *Maculinea alcon* Denis & Schiffermüller group follow a rather specialised

47   development compared to other *Maculinea* species, as they are not simply predators of ant brood,

48   but are fed by *Myrmica* workers in preference to their own brood – a behaviour that has been

49   described as a "cuckoo" strategy (Thomas & Elmes, 1998). Because they are constantly

50   interacting with worker ants, this means that they need to adapt precisely to the local host ant

51   species, e.g. by mimicking the odour (Akino et al., 1999; Nash et al., 2008; Thomas & Settele,

52   2004) and sounds (Barbero et al., 2009) of the ants, in order to be accepted by a suitable

53   *Myrmica* colony. While the initial host plants of this group are all species of gentian (*Gentiana*

54   L. and *Gentianella* Mönch), they can occur in very different open habitats, such as lowland and

55   mountain meadows or wet and dry swards (Munguira & Martín, 1999; Oostermeijer, Vantveer &

56   Dennijs, 1994; Settele, Kühn & Thomas, 2005; Tartally, Koschuh & Varga, 2014). Based on

57   these different habitat types, several forms or (sub)species of the *M. alcon* group have been

58   described. The most widely accepted separation within this group is that the nominotypic *M.*

59   *alcon* occurs on humid meadows and there is another xerophilous form which has usually been

60   referred to as *M. rebeli* Hirschke (Thomas et al., 2005; Thomas & Settele, 2004; Wynhoff,

61   1998). Both forms are patchily distributed (Wynhoff, 1998) and have been considered as

62   endangered in many European countries (Munguira & Martin, 1999), with the xerophilous form

63   considered to be a European endemic (Munguira & Martin, 1999). However, several papers

64   (Habeler, 2008; Kudrna & Belicek, 2005; Kudrna & Fric, 2013) have made the case that the

65   xerophilous form is most likely not synonymous with the nominotypic *M. rebeli*, which is found

66   at higher altitude, and has a unique host plant and host ant usage (Tartally, Koschuh & Varga,

67   2014). Furthermore, recent molecular phylogenetic studies (Als et al., 2004; Ugelvig et al.,

68   2011b; Bereczki et al., 2015) suggest that the hygrophilous and xerophilous forms of *M. alcon*,

69   while distinct from other congeners, are not two distinct lineages, and show very little variation

70   in genes normally used for phylogenetic inference. This has been confirmed by several regional

71   population genetic studies (Bereczki et al., 2005; Bereczki, Pecsenye & Varga, 2006; Sielezniew

72   et al., 2012; Bereczki et al., 2015), where there is no consistent separation of the two forms. This

73   has led to the current situation where xerophilous and hygrophilous *M. alcon* are not

74   distinguished for conservation purposes, and the species is now considered as "of least concern"

75   in Europe (Van Swaay et al., 2010). To avoid confusion, we will refer to the "typical"

76   hygrophilous form of *M. alcon* as '*M. alcon H*' and the xerophilous form as '*M. alcon X*'

77   throughout the rest of this manuscript, following Tartally, Koschuh & Varga (2014).

78   The host plant and host ant usage of the two *M. alcon* forms are different, because different

79   gentian and *Myrmica* species are available on the hygric sites of *M. alcon H* and xeric sites of *M.*

80   *alcon X*. While *M. alcon H* starts development typically on the marsh gentian (*Gentiana*

81   *pneumonanthe* L.), *M. alcon X* typically uses the cross gentian (*G. cruciata* L.), and there is some

82   evidence that enzyme systems related to host plant use may be diverging in the two forms

83   (Bereczki et al., 2015). The development of *M. alcon X* typically continues in nests of *Myrmica*

84   *schencki* Viereck and *My. sabuleti* Meinert but *M. alcon H* most often uses *My. rubra* L., *My.*

85   *ruginodis* Nylander or *My. scabrinodis* Nylander as host ants. Furthermore, some other minor or

86   locally important host plant and host ant species have been recorded for both forms (summarized

87   in Witek, Barbero & Markó, 2014).

88   Despite these differences in the host plant and ant usage of *M. alcon H* and *M. alcon X,*

89   phylogenetic reconstruction using morphological and ecological characters suggests that western

90   Palaearctic *M. alcon H* are closer to European *M. alcon X* than Asian *M. alcon H* (Pech et al.,

91   2004). In combination, all these results suggest local ecological but not genetic differentiation of

92   the two forms between hygric and xeric sites. Until recently this could only be tested by

93   comparing sites that were separated by tens of kilometres or more, but in the last decade a site

94   has been recorded from Răscruci (Transylvanian basin, Romania) where patches supporting *M.*

95   *alcon H* and *M. alcon X* occur in a mosaic separated by tens of meters. The two forms use

96   different host plants and mostly different host ants on this site (Tartally et al., 2008), and their

97   flying periods are largely separated based on the phenology of their host plants (Czekes et al.,

98   2014; Timuş et al., 2013). In addition, most previous genetic studies have been based on

99   collecting samples from either flying adults or caterpillars as they emerge from the host plant,

100   which means that any separation by host ant species could not be examined directly. Our aim

101   was therefore to investigate the genetic differentiation between the two forms of *M. alcon* at this

102   unique syntopic site, to relate this to differences in both host plant and host ant use, and to make

103   recommendations for the conservation of the forms based on their shared and predicted future

104   histories (c.f. Bowen & Roman, 2005).

105

Materials and Methods

Field methods

Two sites in Transylvania (Figure 1) were visited in the summers of 2007 and 2009 to record host plant and host ant usage and to collect genetic samples of *M. alcon*. Host ant specificity results from 2007 have already been published in Tartally et al. (2008). The first site is at Răscruci (46°54' N; 23°47' E; 485 m a.s.l.), which is predominantly an extensively grazed tall-grass meadow steppe with *Gentiana cruciata* (the host plant of *M. alcon X*), but also with numerous small marshy depressions with tall-forb vegetation in which *G. pneumonanthe* (the host plant of *M. alcon H*) is common (Czekes et al., 2014). This site gave the unique possibility to compare the host ant specificity and population genetics of *M. alcon H* and *M. alcon X* within the same site. To collect samples, two nearby patches were chosen within this mosaic site where *G. pneumonanthe* and *G. cruciata* were well separated from each other (there was a ca. 20 m wide zone without gentians). In other parts of this site border effects (because of the potential migration of *Myrmica* colonies) or the co-occurrence of the two gentians made it difficult to find *M. alcon* larvae originating clearly from *G. pneumonanthe* or *G. cruciata*. The patch with *G. pneumonanthe* will henceforth be referred to as 'Răscruci wet' (*M. alcon H* patch), while the patch with *G. cruciata* will be referred to as 'Răscruci dry' (*M. alcon X* patch). The nearest known *M. alcon* site (a *M. alcon H* site) to Răscruci is at Șardu (46°52' N; 23°24' E; 480 m a.s.l.), 29 km west of Răscruci, which was chosen as a control site. Șardu is a tall-grass, tall-sedge marshy meadow with locally dense stands of *G. pneumonanthe*. The two sites are separated by a range of hills without suitable *M. alcon* habitat (Figure 1).

127   To obtain data on the host ant specificity and to get samples for genetic analysis, *Myrmica* nests

128   were searched for within 2 m of randomly selected *Gentiana* host plants, which is considered to

129   be the approximate foraging zone of worker ants of the genus *Myrmica* (Elmes et al., 1998).

130   Searches were made no earlier than four weeks before the flying period of *M. alcon* at both sites,

131   so that any *M. alcon* caterpillars or pupae found must have survived the winter in the ant nest,

132   and hence have become fully integrated (Thomas et al., 2005). Nests were excavated carefully

133   but completely, after which the ground and vegetation were restored to as close to the original

134   conditions as possible. All *M. alcon* caterpillars, pupae and exuviae were counted, placed in 98%

135   ethanol, and stored at -20 °C until DNA could be extracted. Five to ten worker ants were also

136   collected from each ant nest and preserved in 70% ethanol for later identification in the

137   laboratory using keys by Seifert (1988) and Radchenko and Elmes (2010). For further details, see

138   Tartally et al. (2008).

139

140   *Host ant specificity*

141   Host ant specificity (deviation from random occurrence in nests of different *Myrmica* species)

142   was calculated based on the number of fully grown butterfly larvae, pupae and exuviae in two

143   ways: *P1* is the 2-tailed probability from a Fisher exact test of heterogeneity in infection of host

144   ant nests (as implemented at http://www.quantitativeskills.com/sisa/), and *P2* is the probability

145   from a randomization test of ant nests between species, using the software MACSAMP (Tartally et

146   al., 2008). Published (Tartally et al., 2008) and newly-collected data on host ant specificity were

147   combined for these analyses. In the case of Răscruci, host ant specificity results were calculated

148   separately for Răscruci wet and Răscruci dry and also based on the combined data from both

149   patches ('Răscruci both' below).

150

151   *DNA extraction and microsatellite analysis*

152   DNA was extracted from approximately 1-2 mm^3 of tissue from caterpillars or pupae using a

153   10% Chelex-10mM TRIS solution with 5 µl Proteinase K. Samples were incubated at 56 °C for

154   minimum 3.5 hours or overnight and boiled at 99.9 °C for 15 min. The supernatant was collected

155   and stored at 5 °C or -20 °C for short or long term storage, respectively. For each sample, nine

156   polymorphic nuclear microsatellite loci developed for *Maculinea alcon* were amplified: Macu20,

157   Macu26, Macu28, Macu29, Macu30, Macu31, Macu40, Macu44, and Macu45 (Table 1; Ugelvig

158   et al., 2011a; Ugelvig et al., 2012) using a REDTaq® ReadyMix™ PCR Reaction Mix (Sigma-

159   Aldrich). These primer pairs (see concentrations in table 1) were amplified using standard PCR

160   conditions: initial denaturation for 5 min at 95 °C followed by 30 cycles of 30 s at 95 °C, 30 s at

161   locus-specific annealing temperature (see table 1) and 30 s extension at 72 °C, finishing with

162   elongation of 15 min at 72 °C run on a Thermo PCR PXE 0.2 Thermal Cycler. Total reaction

163   volume was 10 µl of which 1 µl was template DNA. PCR products were run on a 3130xl Genetic

164   Analyzer with GeneScan 500 LIZ (Life Technologies) as internal size standard and analyzed

165   with GENEMAPPER® Software version 4.0 (Applied Biosystems). Locus Macu40 could not be

166   scored consistently (excessive stutter bands) and was omitted from all further analysis. The

167   overall proportion of alleles that could not be amplified was 4.6 % (see table S1).

168

169 *Tests for Hardy-Weinberg and linkage disequilibrium*

170 The eight microsatellite loci analysed were tested for linkage disequilibrium (genotypic

171 disequilibrium) between all pairs of loci in each sample and for deviations from Hardy-Weinberg

172 proportions using exact tests in FSTAT version 2.9.3.2 (Goudet, 1995) based on 480 and 1260

173 permutations, respectively. The software package MICRO-CHECKER version 2.2.3 (Van

174 Oosterhout et al., 2004) using 1,000 iterations and a Bonferroni corrected 95% confidence

175 interval, was employed to test for possible null-alleles.

176

177 *Population structure, genetic differentiation and kinship*

178 We studied the genetic clustering of individual genotypes using the Bayesian algorithm

179 implemented in STRUCTURE version 2.3.4 (Falush, Stephens & Pritchard, 2003, 2007; Hubisz et

180 al., 2009; Pritchard, Stephens & Donnelly, 2000). The most likely number of genetically distinct

181 clusters (K) was estimated for each K in the range 2 to 12, allowing for sub-structuring of

182 samples. A burn-in length of 50,000 MCMCs was used to secure approximate statistical

183 stationarity, followed by a simulation run of 500,000 MCMCs using an admixture model with

184 correlated allele frequencies as recommended by Pritchard, Stephens & Donnelly (2000). No

185 location prior was used, and LnP(D) values were averaged over 20 iterations. The most likely

186 value of K (number of clusters) was estimated using the ΔK method of Evanno, Regnaut &

187 Goudet (2005). To check whether the assumptions inherent in STRUCTURE were biasing our

188 genetic clustering, we also used the Bayesian genetic clustering programs BAPS version 5.2

189 (Corander et al., 2008) and INSTRUCT version 1.0 (Gao, Williamson & Bustamante, 2007), which

190 gave essentially identical results (see Additional Analysis S1, Figures S1, S2).

191   For more detailed population differentiation, samples were explored individually as well as in

192   four different partitions: (a) pre-defined populations (**POP**: Răscruci dry, Răscruci wet and

193   Şardu) which also relates to host plant use (Răscruci wet and Şardu: *G. pneumonanthe*. Răscruci

194   dry: *G. cruciata*), (b) host ant use (**ANT**: *Myrmica scabrinodis*, *My. sabuleti*, *My. schencki* and

195   *My. vandeli*), (c) host ant nests (**NEST**: specific nest ID within **POP**), and (d) year of sampling

196   (**YEAR**: 2007 and 2009), the latter to test for potential temporal differences.

197   We studied the overall population differentiation between pre-defined populations (**POP**)

198   calculating Weir and Cockerham's (1984) estimate of F_{ST} (θ) using FSTAT version 2.9.3.2 based

199   on 1,000 permutations. As the magnitude of the value of θ is related to the allelic diversity at the

200   marker loci applied, we further calculated the standardized G''_{ST} (Meirman & Hedrick, 2011),

201   and the estimator D_{EST} (Jost, 2008) as alternative quantifications of genetic differentiation,

202   making comparisons with studies based on other marker loci possible (Meirman & Hedrick,

203   2011). G''_{ST} and Jost's D_{EST} for pairs of **POP** samples were calculated using GENODIVE version

204   2.0 b27 (Meirmans & Van Tienderen, 2004). Hierarchical AMOVA (Analysis of Molecular

205   Variance: Excoffier, Smouse & Quattro, 1992) was calculated for **POP**, **ANT**, **NEST**, and

206   separately for **POP** and **YEAR** using the R-package HIERFSTAT (Goudet, 2005) with 9,999

207   permutations to estimate the variance components and their statistical significance. Individual-

208   based Principal Coordinate Analysis (PCOA) with standardized covariances was employed to

209   obtain a multivariate ordination of individual samples based on pairwise genetic distances, as

210   implemented in the software GENALEX version 6.502 (Peakall & Smouse, 2012). The PCoA

211   were explored for **NEST** within **ANT** within **POP** across **YEAR** using nested MANOVA based on

212   the sum of the variances of the different coordinates, as implemented in JMP 12.02 (SAS

213   Institute).

214   To examine whether the low dispersal ability of *Maculinea alcon* females could lead to high

215   relatedness between samples of individuals collected in the same nest, pairwise measures of

216   kinship (Loiselle et al., 1995) and relatedness (Queller & Goodnight 1989) between samples

217   were estimated using GENODIVE and GENALEX respectively. Both of these values estimate the

218   probability that samples share alleles by descent, based on the distribution of alleles in the whole

219   set of samples, with possible values ranging from -1 to +1. Negative values show that the two

220   individuals compared are less similar in the alleles they share than two randomly picked

221   individuals. Values of kinship and relatedness were then compared between the different

222   partitions of the data using stratified Mantel tests as implemented in GENODIVE. To further test

223   the hypothesis that individuals found in the same nest were likely to derive from eggs laid by the

224   same female, the program COLONY (v 2.0.6.1; Jones and Wang 2010) was used to give a

225   maximum likelihood estimate of the probability that any two sampled individuals were likely to

226   be either full or half siblings. Since males of *M. alcon* are much more mobile than females, who

227   tend to oviposit in a limited area (Kőrösi et al., 2008; D R Nash, unpublished data), this was

228   based on a mating system assuming female polygyny and male monogyny, with other parameters

229   kept at their default values.

230

231

232 **Results**

233 *Host ant specificity*

234 A total of 135 *Myrmica* ant nests were found within 2 m of host gentian plants on the two sites,

235 and 90 *Maculinea* larvae, pupae and exuviae were found in 26 infested nests (Table 2) including

236 87 nests and 56 *Maculinea* already published in Tartally et al. (2008). Altogether four *Myrmica*

237 species were found. Only *My. scabrinodis* was present at all sites, and was the most abundant ant

238 species (59% of all ant colonies found). This species was used as a host on all three sites. Only a

239 single *M. alcon X* was found in a nest of *My. scabrinodis* at Răscruci dry despite the dominance

240 of this ant there and its frequent usage by *M. alcon H* on Răscruci wet (Fisher's exact test, $P =$

241 0.032). The much greater exploitation rates of *My. sabuleti* and *My. schencki* led to significant

242 overall host ant specificity at Răscruci dry (Table 2).

243

244 *Genetic diversity and inbreeding*

245 Measures of genetic diversity and *F*-statistics generated by FSTAT for each locus are listed in

246 Table S2 in the supporting information. Analysis with MICRO-CHECKER revealed that Macu29

247 had a highly significant ($P < 0.001$) excess of homozygotes and cases of non-amplification

248 consistent with the presence of a relatively high proportion (>20%) of null-alleles, and was

249 therefore excluded from further analysis. All other loci showed no significant deviations from

250 Hardy-Weinberg proportions. Tests for linkage disequilibrium revealed only a few sporadic

251 significant results showing no overall pattern (Table S3), so all loci were retained in further

252 analysis, which was thus based on 7 polymorphic loci. All three of the pre-defined populations

253   showed no evidence of inbreeding (Table 3), and in fact showed negative values for the

254   inbreeding coefficient F_{IS} (meaning an excess of heterozygotes), although not significantly so

255   (Table 3).

256

257   *Population structure*

258   STRUCTURE analysis revealed rather invariable log-likelihood values for partitioning of the data

259   into genetic clusters, but the highest change in log-probability value was for K = 2, with lower

260   maxima at K = 4 and K = 10 (Figure 1). There was a clear overall distinction between samples

261   from Răscruci wet in one genetic group and Răscruci dry and Şardu in another group. Levels of

262   admixture between genetic clusters were generally low, but four individuals from Răscruci wet

263   (from two different *My. scabrinodis* nests) showed high affinity to the Răscruci dry-Şardu group,

264   irrespective of the value of *K*. One individual found in a *My. sabuleti* nest at Răscruci dry

265   (sample code: DA14) appeared genetically more similar to those from Răscruci wet. For values

266   of K higher than 2 there was no additional partitioning between the pre-defined populations, but

267   some substructure in Răscruci wet became apparent for K = 3, with two partitions that were

268   relatively dissimilar, while for K ≥ 4 no additional grouping of individuals was apparent (Figure

269   1).

270

271   *Genetic differentiation*

272   We found significant overall genetic differentiation between pre-defined populations (θ = 0.090,

273   D_{EST} = 0.215; Table S2). Pairwise genetic differentiation measures θ, G''_{ST} and D_{EST} were

274   significant for all population comparisons after Bonferroni adjustment ($P<0.003$; Table 3). There

275   was no evidence of inbreeding, either overall (F_{IS} = -0.074, Table S2), or within pre-defined

276   populations (Table 3).

277   Hierarchical AMOVA (Table 4) revealed that most genetic variance (93.4%) was within

278   individuals, but that significant variation was also explained by **POP** and **NEST**. The proportion

279   of variation and inbreeding coefficients for individuals within **NEST** and **ANT** were both

280   negative, indicating that there was greater heterozygosity between individuals in the same nest

281   and between samples across different *Myrmica* species than between randomly chosen samples

282   from the data set. Samples from different years explained only 0.12% of the genetic variance in a

283   separate AMOVA ($P = 0.354$). The Principal Coordinate Analysis retained a total of six principal

284   coordinates with eigenvalues greater than 1, which together explained 52% of the variance in

285   genetic distance (13.5%, 10.3%, 8.4%, 8.0%, 7.2% and 4.5% for coordinates 1-6 respectively;

286   see figure S3 for more details). These showed a similar result to the AMOVA where **YEAR**

287   samples (2007 and 2009) overlapped completely in genetic ordination space ($F_{1,36} = 8.91 \times 10^{-16}$,

288     $P = 0.999$), while samples from Răscruci wet were separated from those from Răscruci dry and

289   Şardu (Figure 2; $F_{2,36} = 6.08$, $P = 0.005$). We found a pronounced structuring of samples when

290   examining nests within pre-defined populations ($F_{18,36} = 3.59$, $P < 0.001$), with samples from the

291   same nest clustering together, but there was no consistent clustering of samples from the nests of

292   the same host ant species ($F_{3,36} = 0.887$, $P = 0.457$). The single sample from Răscruci dry that

293   was collected from a *My. scabrinodis* nest (sample code DB15) had a first principal component

294   that was more characteristic of samples from Răscruci wet (which all used this host ant species;

295   figure 2), but was not assigned to this population in the Bayesian analysis (figure 1).

296

297   *Kinship and parentage analysis*

298   Overall pairwise relatedness of individuals (Figure 3) sampled from the same nest (0.311) was

299   significantly higher than that of those sampled from different nests within the same site (-0.032;

300   Stratified Mantel test: $r^2 = 0.069$, $P = <0.0001$). Looking at individual sites, the same pattern was

301   found at both *M. alcon H* sites (Răscruci wet: within-nest relatedness = 0.32, between nests =

302   0.10, $r^2 = 0.130$, $P = 0.0002$; Șardu: within-nest = 0.249, between nests = 0.065, $r^2 = 0.126$, $P =$

303   0.002), but relatedness was not significantly different within and between nests at Răscruci dry

304   (within-nest = 0.003, between nests = 0.081, $r^2 = 0.001$, $P = 0.448$). Similar results were found

305   when analysing pairwise kinship (see figure S4 and Additional Analysis S2). Maximum-

306   likelihood analysis using COLONY identified 38 pairs of individuals as potential full siblings

307   (with probabilities ranging from 0.002 to 0.935; Figure 3), and 161 as potential half siblings

308   (with probabilities ranging from 0.002 to 0.742; Figure S4). For the 21 pairs with high (>0.5)

309   probability of being full siblings, 13 (62%) were from the same nest, six were from nests of the

310   same ant species at the same site, and only two were from different sites, both including

311   individual SCA86-2 from Răscruci wet, which appears closely related to two individuals (DA14

312   and SAB67-1) from different *Myrmica* nests from Răscruci dry (Figure 3). This last result

313   probably reflects the non-amplification of characteristic loci for these individuals (see table S1),

314   so that they share common alleles without being related. Within sites, a high proportion of

315   individuals from both Răscruci wet (60.4%) and Șardu (37.5%) from multiply-infested nest had

316   individuals estimated to be full siblings in the same nest (overall 52.9%, Generalized linear

317   model with binomial errors and Firth corrected maximum likelihood, comparing sites:

318   Likelihood-ratio $\chi^2 = 0.73$ *d.f.* = 1, $P = 0.391$). However, none of the *M. alcon X* individuals

319   sharing nests at Răscruci dry were estimated to be full siblings (*M. alcon X* vs. *M. alcon H*:

320   Likelihood-ratio $\chi^2 = 9.72$, *d.f.* $= 1$, $P = 0.002$).

321   Although there was evidence for strong within-nest relatedness of individuals, the patterns of

322   genetic diversity and differentiation were not strongly affected by this, and were unaffected when

323   analyses were repeated with only a single individual from each nest (see table S4).

324

325

326   **Discussion**

327   This study gives the first comparison of the host ant specificity and genetic composition of *M.*

328   *alcon H* and *M. alcon X* within the same site.

329   The host ant specificity found in this study confirm the earlier results of Tartally et al. (2008) that

330   these populations use the typical host ants found in other Central European studies (Höttinger,

331   Schlick-Steiner & Steiner, 2003; Sielezniew & Stankiewicz, 2004; Steiner et al., 2003; Tartally

332   et al., 2008; Witek et al., 2008). *M. alcon H* was found exclusively with *My. scabrinodis* at

333   Răscruci wet and also with *My. vandeli* at Șardu, but *M. alcon X* was found mainly with *My.*

334   *sabuleti* and *My. schencki* at Răscruci dry. Interestingly only one *M. alcon X* was found with *My.*

335   *scabrinodis*, despite this *Myrmica* species being the most numerous at Răscruci dry (Table 2) and

336   being the main host of *M. alcon X* in two other sites in the Carpathian-Basin (Tartally et al.,

337   2008). *My. scabrinodis* usage could therefore be a potential link between the *M. alcon H* and *M.*

338   *alcon X* populations at Răscruci (and probably in some other regions), but *M. alcon X* shows a

339   clear separation from the *M. alcon H* in the proportional usage of this host ant. The background

340   of this separation in the host ant specificity of *M. alcon H* and *M. alcon X* at Răscruci is not

341   clear, but could reflect the dynamic arms race between the different genetic lineages of *M. alcon*

342   and local host ants (Nash et al., 2008).

343   Our genetic results (Figures 1, 2) show strong genetic differentiation between *M. alcon H* and *M.*

344   *alcon X* at Răscruci, indicating limited gene flow between these two groups, although it is

345   interesting to note that a few individuals had genotypes more characteristic of the other

346   population. This differentiation is likely due to separation in time rather than space because of

347   the different phenology of the host plants, which results in largely non-overlapping flying

348   seasons of *M. alcon H* and *M. alcon X* (Timuş et al., 2013). This may be reinforced by lowered

349   fitness of any hybrid individuals that would emerge during the approximately 2-week gap when

350   neither host plant is suitable for oviposition.

351   The lowest level of between-population differentiation, on the other hand, was between *M. alcon*

352   *X* from Răscruci and *M. alcon H* from Şardu, and Bayesian population assignment suggests that

353   these are so similar that they have almost certainly been part of a single population. This

354   supports previous findings of no overall phylogenetic differentiation between the two forms of

355   *M. alcon* (Als et al., 2004; Fric et al., 2007; Ugelvig et al., 2011b), and that the two forms tend to

356   be more genetically similar regionally than either is to more distant populations that use the same

357   host plant (Bereczki et al., 2005; Bereczki, Pecsenye & Varga, 2006; Pecsenye et al., 2007).

358   Hence the two forms cannot be regarded as host races (Drès & Mallet, 2002), since they do not

359   fulfil the criterion of spatial and temporal replicability. Genetic analysis of several Polish and

360   Lithuanian *M. alcon* populations using microsatellite markers (Sielezniew et al., 2012) gave

361   similar results to ours (Figures 1, 2) in that there was no clear pattern reflecting genetic division

362   into two ecotypes. Sielezniew et al. also found that the *M. alcon X* ecotype was less polymorphic,

363   and its populations more differentiated than those of the *M. alcon H* ecotype. Their data also

364   suggest that *M. alcon H* populations form a single clade but *M. alcon X* can be split into more

365   clades, suggesting that *M. alcon H* is an ancestral form and that *M. alcon X* represents a group of

366   independently evolved *M. alcon H* populations that have switched to use dryer habitats with the

367   locally available *Gentiana* and *Myrmica* species. They propose that the background of this

368   pattern may be independent specialisations on different host ant species, since in their study

369   clades of *M. alcon X* largely reflected host ant use. However, we find no evidence of genetic

370   differentiation associated with host-ant usage at Răscruci or Şardu (Figures 1, 2, 3), and no

371 difference in genetic diversity in populations of the two ecotypes (table 3). Due to the relatively

372 large distances and potential barriers between Răscruci and Şardu it is unlikely that there has

373 been recent gene-flow between the two sites, which suggests that Răscruci was likely colonized

374 at least twice from two different gene pools, and that the ancestors of the Răscruci wet

375 population are no longer locally extant (or have evaded detection).

376 The higher within-nest than between-nest relatedness between individuals of *M. alcon H* is

377 consistent with observations of limited dispersal of ovipositing females (Kőrösi et al., 2008)

378 which is likely to lead to substantial within-population substructure between nests, as found here.

379 The additional grouping of individuals from K=2 to K=4 in our population assignment analysis

380 groups families of relatives within populations. Our parentage analysis confirms the relatively

381 high probability that *M. alcon H* caterpillars infesting the same *Myrmica* nest are relatives, and in

382 around 50% of cases may be full siblings. The lack of any association between relatedness and

383 infestation of nests for *M. alcon X* from the Răscruci dry site is consistent with the difference in

384 oviposition strategy and mobility of butterflies from this population compared with those from

385 the Răscruci wet site (Czekes et al., 2014; Timuş et al., 2013).

386 Given the small size of these *M. alcon* populations (Timuş et al., 2013) and their low dispersal

387 (Kőrösi et al. 2008), it is interesting to note that there is no evidence of inbreeding among the

388 individuals that we examined, and that estimated inbreeding coefficients (F_{IS} and G_{IS}) were

389 negative. This means that individuals were more heterozygous than expected if mating were

390 random (albeit not significantly so), which is probably a result of difference in mating strategy of

391 males and females. Males are highly mobile, and tend to patrol a large area while seeking

392 females, whereas females are rather sedentary, and are often mated immediately after emerging

393 from the *Myrmica* nest in which they developed, and then go on to lay eggs on host plants

394   relatively close by (Kőrösi et al., 2008). This means that females are unlikely to be related to

395   their mates, and may in fact be less related to them than to a randomly chosen male, because

396   males pupate and emerge several days before females (Meyer-Hozak, 2000). This, together with

397   the observation that caterpillars that develop in the same *Myrmica* nest may be offspring of the

398   same female, can easily lead to the negative inbreeding indices observed both overall, and for

399   individuals within *Myrmica* nests and host ant species within pre-defined populations within our

400   hierarchical AMOVA.

401   Regardless of its origin, it is clear that the *M. alcon X* population at Răscruci is ecologically

402   highly differentiated from the local *M. alcon H* populations in terms of its host plant and host ant

403   use, as well as in its behaviour (Czekes et al., 2014; Timuş et al., 2013). The pattern of

404   differentiation we see is not that typically associated with speciation via host race formation

405   (Drès & Mallet, 2002), since the two forms do not fulfil the criterion of spatial and temporal

406   replicability (i.e. "*are more genetically differentiated from populations on another host in*

407   *sympatry (and at the same time) than at least some geographically distant populations on the*

408   *same host*"; Drès & Mallet, 2002, p. 473-4). *M. alcon X* cannot, therefore, represent an

409   evolutionarily significant unit in conservation terms as usually defined, but we would argue that

410   it should still be regarded as a functional conservation unit (Casacci, Barbero & Balletto, 2013).

411   The site at Răscruci represents the only known area where both forms of *M. alcon* occur

412   syntopically, and so is of particular value to research on speciation, and has great potential for

413   examining adaptation at non-neutral genetic markers. This is enhanced by the occurrence of two

414   other *Maculinea* species on the same site; *M. teleius* Bergsträsser (Tartally & Varga, 2008) and

415   *M. nausithous kijevensis* Sheljuzhko (Rákosy et al., 2010; Tartally & Varga, 2008), as well as the

416   *Myrmica* parasites *Microdon myrmicae* Schönrogge et al. (Diptera: Syrphidae; Bonelli et al.,

417   2011) and *Rickia wasmannii* Cavara (Ascomycota: Laboulbeniales; Tartally, Szűcs & Ebsen,

418   2007). The *Maculinea* spp. parasitoid *Ichneumon eumerus* Wesmael (Hymenoptera:

419   Ichneumonidae) is also present (Tartally, 2008). Most of these species are also found at Şardu

420   (except *M. alcon X* and *M. nausithous*: Tartally, 2008). It should be emphasized that all of these

421   species can be found in the nests of, and ultimately depend on, *My. scabrinodis* (as well as other

422   *Myrmica* species: see Witek, Barbero & Markó (2014) for a review), providing a unique

423   opportunity to examine a complex set of parasitic interactions revolving around a single keystone

424   ant species.

425

426

427 **Conclusion**

428 Our analysis of *Maculinea alcon* from a unique site where both the xerophilous and hygrophilous

429 forms of this butterfly are found within tens of meters of each other has demonstrated strong

430 genetic differentiation between the two forms. However, the xerophilous form was not

431 significantly differentiated from the next closest known population of the hygrophilous form.

432 This supports other recent work suggesting that the hygrophilous and xerophilous forms are not

433 separate species or even subspecies, and that the name *M. rebeli* has frequently been applied to

434 the xerophilous form incorrectly. There is some overlap in host ant species use between the two

435 forms, so the most likely proximate reason for the local genetic differentiation found is

436 differences in host plant phenology. We suggest that since the two different forms of *M. alcon* do

437 not have a separate evolutionary history, they cannot be regarded as "evolutionarily significant

438 units" for conservation as the term was originally used (Bowen & Roman, 2005; Casacci,

439 Barbero & Balletto, 2013). However, since they represent current ecological diversification of

440 the group, and may have different evolutionary potentials (e.g. through selection on different

441 enzyme systems; Bereczki et al., 2015), they should continue to be treated as separate

442 management units for long-term conservation (Bowen & Roman, 2005; Casacci, Barbero &

443 Balletto, 2013). Hence we support the continued separation of the two forms in future studies to

444 further explore their evolutionary trajectories and conservation potential.

445

446

447   **Acknowledgements**

448   We thank Tibor-Csaba Vizauer, László Rákosy and Zoltán Varga for assistance in the field and

449   Shukriya Barzinci, Maria Mikkelsen and Sylvia Mathiasen for assistance in the laboratory.

450   Zoltán Varga, Enikő Tóth, Simona Bonelli and Robert Toonen provided valuable comments on

451   earlier versions of the manuscript.

452 **References**

453 Akino T, Knapp JJ, Thomas JA, Elmes GW. 1999. Chemical mimicry and host specificity in the

454   butterfly *Maculinea rebeli*, a social parasite of *Myrmica* ant colonies. *Proceedings of the*

455   *Royal Society of London, Series B: Biological Sciences* 266: 1419-1426. DOI:

456   10.1098/rspb.1999.0796.

457 Als TD, Vila R, Kandul NP, Nash DR, Yen SH., Hsu YF, Mignault AA, Boomsma JJ, Pierce

458   NE. 2004. The evolution of alternative parasitic life histories in large blue butterflies.

459   *Nature* 432: 386-390. DOI: 10.1038/nature03020.

460 Barbero F, Thomas JA, Bonelli S, Balletto E, Schönrogge K. 2009. Queen ants make distinctive

461   sounds that are mimicked by a butterfly social parasite. *Science* 323: 782-785. DOI:

462   10.1126/science.1163583.

463 Bereczki J, Pecsenye K, Peregovits L, Varga Z. 2005. Pattern of genetic differentiation in the

464   *Maculinea alcon* species group (Lepidoptera, Lycaenidae) in Central Europe. *Journal of*

465   *Zoological Systematics and Evolutionary Research* 43: 157-165. DOI: 10.1111/j.1439-

466   0469.2005.00305.x.

467 Bereczki J, Pecsenye K, Varga Z. 2006. Geographical versus food plant differentiation in

468   populations of *Maculinea alcon* (Lepidoptera: Lycaenidae) in Northern Hungary.

469   *European Journal of Entomology* 103: 725-732. DOI: 10.14411/eje.2006.096.

470 Bereczki J, Rácz R, Varga Z, Tóth J. P. 2015. Controversial patterns of *Wolbachia* infestation in

471   the social parasitic *Maculinea* butterflies (Lepidoptera: Lycaenidae). *Organisms Diversity*

472   *and Evolution 15*: 591-607. DOI:10.1007/s13127-015-0217-7

473   Bonelli S, Witek M, Canterino S, Sielezniew M, Stankiewicz-Fiedurek A, Tartally A, Balletto E,

474   Schönrogge K. 2011. Distribution, host specificity, and the potential for cryptic speciation

475   in the hoverfly *Microdon myrmicae* (Diptera: Syrphidae), a social parasite of *Myrmica*

476   ants. *Ecological Entomology* 36: 135-143. DOI: 10.1111/j.1365-2311.2010.01253.x.

477   Bowen, BW, Roman JOE. 2005. Gaia's Handmaidens: The Orlog Model for Conservation

478   Biology. *Conservation Biology* 19: 1037-1043. DOI: 10.1111/j.1523-1739.2005.00100.x

479   Casacci LP, Barbero F, Balletto E. 2013. The "evolutionarily significant unit" concept and its

480   applicability in biological conservation. *Italian Journal of Zoology* 81: 182-193.

481   doi:10.1080/11250003.2013.870240

482   Corander J, Marttinen P, Sirén J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software

483   for learning genetic structures of populations. *BMC Bioinformatics* 9: 539. DOI:

484   10.1186/1471-2105-9-539.

485   Czekes Z, Markó B, Nash DR, Ferencz M, Lázár B, Rákosy L. 2014. Differences in oviposition

486   strategies between two ecotypes of the endangered myrmecophilous butterfly *Maculinea*

487   *alcon* (Lepidoptera: Lycaenidae) under unique syntopic conditions. *Insect Conservation*

488   *and Diversity* 7: 122-131. DOI: 10.1111/icad.12041.

489   Drès, M, Mallet J. 2002. Host races in plant-feeding insects and their importance in sympatric

490   speciation. *Philosophical Transactions of the Royal Society of London, Series B:*

491   *Biological Sciences* 357: 471-492. DOI: 10.1098/rstb.2002.1059

492   Elmes GW, Thomas JA, Wardlaw JC, Hochberg ME, Clarke RT, Simcox DJ. 1998. The ecology

493   of *Myrmica* ants in relation to the conservation of *Maculinea* butterflies. *Journal of Insect*

494   *Conservation* 2: 67-78. DOI: 10.1023/A:1009696823965.

495 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the

496   software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620. DOI:

497   10.1111/j.1365-294X.2005.02553.x.

498 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric

499   distances among DNA haplotypes - application to human mitochondrial-DNA restriction

500   data. *Genetics* 131: 479-491.

501 Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus

502   genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

503 Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus

504   genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7: 574-578.

505   DOI: 10.1111/j.1471-8286.2007.01758.x.

506 Fric Z, Wahlberg N, Pech P, Zrzavy J. 2007. Phylogeny and classification of the *Phengaris-*

507   *Maculinea* clade (Lepidoptera: Lycaenidae): total evidence and phylogenetic species

508   concepts. *Systematic Entomology* 32: 558-567. DOI: 10.1111/j.1365-3113.2007.00387.x.

509 Gao H, Williamson S, Bustamante CD. 2007. A Markov chain Monte Carlo approach for joint

510   inference of population structure and inbreeding rates from multilocus genotype data.

511   *Genetics* 176: 1635-1651. DOI: 10.1534/genetics.107.072371.

512 Goudet J. 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of*

513   *Heredity* 86: 485-486.

514 Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics.

515   *Molecular Ecology Notes* 5: 184-186. DOI: 10.1111/j.1471-8278 .2004.00828.x

516 Habeler H. 2008. Die subalpin-alpinen Lebensräume des Bläulings *Maculinea rebeli* (Hirscke,

517   1904) in den Ostalpen (Lepidoptera, Lycaenidae). *Joannea - Zoologie* 10: 143-164.

518   Höttinger H, Schlick-Steiner BC, Steiner FM. 2003. The Alcon blue *Maculinea alcon*

519   (Lepidoptera: Lycaenidae) in eastern Austria: status and conservation measures. *Ekologia*

520   *(Bratislava)* 22: 107-118.

521   Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with

522   the assistance of sample group information. *Molecular Ecology Resources* 9: 1322-1332.

523   DOI: 10.1111/j.1755-0998.2009.02591.x.

524   Jones OR, Wang J. 2010. COLONY: a program for parentage and sibship inference from

525   multilocus genotype data. *Molecular Ecology Resources* 10: 551-555.

526   Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* 17: 4015-

527   4026. DOI: 10.1111/j.1365-294X.2008.03887.x.

528   Kőrösi Á, Örvössy N, Batáry P, Kövér S, Peregovits L. 2008. Restricted within-habitat

529   movement and time-constrained egg laying of female *Maculinea rebeli* butterflies.

530   *Oecologia* 156: 455-464. DOI: 10.1007/s00442-008-0986-1.

531   Kudrna O, Belicek J. 2005. On the "Wiener Verzeichnis", its authorship and the butterflies

532   named therein. *Oedippus* 23: 1-32.

533   Kudrna O, Fric ZF. 2013. On the identity and taxonomic status of *Lycaena alcon rebeli*

534   HIRSCHKE, 1905 — a long story of confusion and ignorance resulting in the fabrication

535   of a "ghost species" (Lepidoptera: Lycaenidae). *Nachrichten des Entomologischen Vereins*

536   *Apollo* 34**:** 117-124.

537   Loiselle BA, Sork VL, Nason J, Graham C. 1995. Spatial genetic structure of a tropical

538   understory shrub, Psychotria officinalis (Rubiaceae). *American Journal of Botany* 82:

539   1420-1425.

540   Maes D, Vanreusel W, Talloen W, Van Dyck H. 2004. Functional conservation units for the

541   endangered Alcon Blue butterfly *Maculinea alcon* in Belgium (Lepidoptera: Lycaenidae).

542   *Biological Conservation* 120: 229-241. DOI: 10.1016/j.biocon.2004.02.018.

543   Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the

544   analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* 4: 792-794

545   Meirmans PG, Hedrick PW. 2011. Assessing population structure: F_{ST} and related measures.

546   *Molecular Ecology Resources* 11**:** 5-18. doi: 10.1111/j.1755-0998.2010.02927.x

547   Meyer-Hozak C. 2000. Population biology of *Maculinea rebeli* (Lepidoptera: Lycaenidae) on the

548   chalk grassland of Eastern Westphalia (Germany) and implications for conservation.

549   *Journal of Insect Conservation* 4: 63-72. DOI: 10.1023/A:1009695031802.

550   Munguira ML, Martín J, eds. 1999. Action Plan for the *Maculinea* butterflies in Europe. *Nature*

551   *and Environment* 97. Strasbourg: Council of Europe Publishing.

552   Nash DR, Als TD, Maile R, Jones GR, Boomsma JJ. 2008. A mosaic of chemical coevolution in

553   a large blue butterfly. *Science* 319: 88-90. DOI: 10.1126/science.1149180.

554   Oostermeijer JGB, Vantveer R, Dennijs JCM. 1994. Population structure of the rare, long lived

555   perennial *Gentiana pneumonanthe* in relation to vegetation and management in the

556   Netherlands. *Journal of Applied Ecology* 31: 428-438.

557   Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic

558   software for teaching and research – an update. *Bioinformatics* 28: 2537-2539. DOI:

559   10.1093/bioinformatics/bts460

560   Pech P, Fric Z, Konvicka M, Zrzavy J. 2004. Phylogeny of *Maculinea* blues (Lepidoptera :

561   Lycaenidae) based on morphological and ecological characters: evolution of parasitic

562   myrmecophily. *Cladistics* 20: 362-375. DOI: 10.1111/j.1096-0031.2004.00031.x.

563 Pecsenye K, Bereczki J, Tihanyi B, Toth A, Peregovits L, Varga Z. 2007. Genetic differentiation

564   among the *Maculinea* species (Lepidoptera: Lycaenidae) in eastern Central Europe.

565   *Biological Journal of the Linnean Society* 91: 11-21. DOI: 10.1111/j.1095-

566   8312.2007.00781.x.

567 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus

568   genotype data. *Genetics* 155: 945-959.

569 Queller DC, Goodnight KF. 1989. Estimating relatedness using genetic markers. *Evolution* 43:

570   258-275.

571 Radchenko AG, Elmes GW. 2010. Myrmica *ants (Hymenoptera: Formicidae) of the Old World*.

572   Warsaw: Natura Optima Dux Foundation.

573 Rákosy L, Tartally A, Goia M, Mihali C, Varga Z. 2010. The Dusky Large Blue - *Maculinea*

574   *nausithous kijevensis* (Sheljuzhko, 1928) in the Transylvanian basin: new data on

575   taxonomy and ecology. *Nota Lepidopterologica* 33: 31-37.

576 Seifert B. 1988. A taxonomic revision of the *Myrmica* species of Europe, Asia Minor, and

577   Caucasus (Hymenoptera, Formicidae). *Abhandlungen und Berichte des*

578   *Naturkundemuseums Görlitz* 62: 1-75.

579 Settele J, Kühn E, Thomas JA, eds. 2005. *Studies on the ecology and conservation of butterflies*

580   *in Europe - Vol. 2: species ecology along a European gradient:* Maculinea *butterflies as a*

581   *model*. Sofia: Pensoft.

582 Sielezniew M, Rutkowski R, Ponikwicka-Tyszko D, Ratkiewicz M, Dziekanska I, Svitra G.

583   2012. Differences in genetic variability between two ecotypes of the endangered

584   myrmecophilous butterfly *Phengaris* (= *Maculinea*) *alcon* - the setting of conservation

585   priorities. *Insect Conservation and Diversity* 5: 223-236. DOI: 10.1111/j.1752-

586   4598.2011.00163.x.

587 Sielezniew M, Stankiewicz AM. 2004. Simultaneous exploitation of *Myrmica vandeli* and *M.*

588   *scabrinodis*. *European Journal of Entomology* 101: 693-696. DOI: 10.14411/eje.2004.091.

589 Steiner FM, Sielezniew M, Schlick-Steiner BC, Höttinger H, Stankiewicz A, Górnicki A. 2003.

590   Host specificity revisited: New data on *Myrmica* host ants of the lycaenid butterfly

591   *Maculinea rebeli*. *Journal of Insect Conservation* **7**: 1-6. DOI: 10.1023/A:1024763305517.

592 Tartally A. (2008) Myrmecophily of *Maculinea* butterflies in the Carpathian Basin (Lepidoptera:

593   Lycaenidae). PhD thesis, University of Debrecen.

594 Tartally A, Koschuh A, Varga Z. 2014. The re-discovered *Maculinea rebeli* (Hirschke, 1904):

595   Host ant usage, parasitoid and initial food plant around the type locality with taxonomical

596   aspects (Lepidoptera, Lycaenidae). *ZooKeys* 406: 25-40. DOI: 10.3897/zookeys.406.7124.

597 Tartally A, Nash DR, Lengyel S, Varga Z. 2008. Patterns of host ant use by sympatric

598   populations of *Maculinea alcon* and *M. 'rebeli'* in the Carpathian Basin. *Insectes Sociaux*

599   55: 370-381. DOI: 10.1007/s00040-008-1015-4.

600 Tartally A, Szűcs B, Ebsen JR. (2007) The first records of *Rickia wasmannii* Cavara, 1899, a

601   myrmecophilous fungus, and its *Myrmica* Latreille, 1804 host ants in Hungary and

602   Romania (Ascomycetes: Laboulbeniales; Hymenoptera: Formicidae). *Myrmecological*

603   *News* **10**: 123.

604   Tartally A, Varga Z. 2008. Host ant use of *Maculinea teleius* in the Carpathian-Basin

605   (Lepidoptera: Lycaenidae). *Acta Zoologica Academiae Scientiarum Hungaricae* 54: 257-

606   268.

607   Thomas JA, Elmes GW. 1998. Higher productivity at the cost of increased host-specificity when

608   *Maculinea* butterfly larvae exploit ant colonies through trophallaxis rather than by

609   predation. *Ecological Entomology* 23: 457-464. DOI: 10.1046/j.1365-2311.1998.00153.x.

610   Thomas JA, Elmes GW, Schönrogge K, Simcox DJ, Settele J. 2005. Primary hosts, secondary

611   hosts and 'non-hosts': common confusions in the interpretation of host specificity in

612   *Maculinea* butterflies and other social parasites of ants. In: Settele J, Kuhn E, Thomas JA,

613   eds. *Studies on the ecology and conservation of butterflies in Europe - Vol. 2: species*

614   *ecology along a European gradient:* Maculinea *butterflies as a model*. Pensoft, Sofia. 99-

615   104.

616   Thomas JA, Elmes GW, Wardlaw JC, Woyciechowski M. 1989. Host specificity among

617   *Maculinea* butterflies in *Myrmica* ant nests. *Oecologia* 79: 452-457. DOI:

618   10.1007/BF00378660.

619   Thomas JA, Settele J. 2004. Evolutionary biology: butterfly mimics of ants. *Nature* 432: 283-

620   284. DOI: 10.1038/432283a.

621   Timuş N, Craioveanu C, Sitaru C, Rus A, Rákosy L. 2013. Differences in adult phenology,

622   demography, mobility and distribution in two syntopic ecotypes of *Maculinea alcon*

623   (cruciata vs. pneumonanthe) (Lepidoptera: Lycaenidae) from Transilvania (Romania).

624   *Entomologica Romanica* 18: 21-30.

625 Ugelvig LV, Andersen A, Boomsma JJ, Nash DR. 2012. Dispersal and gene flow in the rare,

626   parasitic Large Blue butterfly *Maculinea arion*. *Molecular Ecology* 21: 3224-3236. DOI:

627   10.1111/j.1365-294X.2012.05592.x.

628 Ugelvig LV, Nielsen PS, Boomsma JJ, Nash DR. 2011a. Reconstructing eight decades of genetic

629   variation in an isolated Danish population of the large blue butterfly *Maculinea arion*.

630   *BMC Evolutionary Biology* 11:201. DOI: 10.1186/1471-2148-11-201.

631 Ugelvig LV, Vila R, Pierce NE, Nash DR. 2011b. A phylogenetic revision of the *Glaucopsyche*

632   section (Lepidoptera: Lycaenidae), with special focus on the *Phengaris-Maculinea* clade.

633   *Molecular Phylogenetics and Evolution* 61: 237-243. DOI: 10.1016/j.ympev.2011.05.016.

634 Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software

635   for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology*

636   *Notes* 4: 535-538. DOI: 10.1111/j.1471-8286.2004.00684.x.

637 Van Swaay C, Cuttelod A, Collins S, Maes D, Munguira ML, Šašić M, Settele J, Verovnik R,

638   Verstrael T, Warren M, Wiemers M, Wynhof I. 2010. *European Red List of Butterflies*.

639   Brussels: IUCN**.** DOI:10.2779/83897

640 Weir BS, Cockerham CC. 1984. Estimating *F*-statistics for the analysis of population structure.

641   *Evolution* 38: 1358-1370. DOI: 10.2307/2408641.

642 Witek M, Barbero F, Markó B. 2014. *Myrmica ants* host highly diverse parasitic communities:

643   from social parasites to microbes. *Insectes Sociaux* 61: 307-323. DOI: 10.1007/s00040-

644   014-0362-6.

645   Witek M, Śliwińska EB, Skórka P, Nowicki P, Wantuch M, Vrabec V, Settele J, Woyciechowski

646   M. 2008. Host ant specificity of large blue butterflies *Phengaris* (*Maculinea*) (Lepidoptera:

647   Lycaenidae) inhabiting humid grasslands in East-central Europe. *European Journal of*

648   *Entomology* 105: 871-877. DOI: 10.14411/eje.2008.115.

649   Wynhoff I. 1998. The recent distribution of the European *Maculinea* species. *Journal of Insect*

650   *Conservation* 2: 15-27. DOI: 10.1023/A:1009636605309.

Details of microsatellites used in this study

SR = product size range (base pairs), T_A = Annealing Temperature (ºC), Primer dye and Multiplex group, N_a = Number of alleles. Ref. = Reference source (U11 = Ugelvig et al. 2011a, U12 = Ugelvig et al. 2012, New = This study). Primer concentration in the PCR mix is given below the table.

Primer	SSR motif	Primer sequences 5' - 3'	SR	T_A	Dye and Multiplex	N_a	Genbank accession	Ref.
Macu20	$(CT)_n(AT)_n(CT)_n$	F: TGGCCCGATTTCCTCTAAAC R: TGCGTGTTTATTTTCATTTTAACAG	92-122	57	Fam 1	9	HM535963	U12
Macu26	$(CA)_n$	F: CTCCCGGGATAGCATTGAC R: CATTGTCGCGGTCGTAATTC	92-128	57	Ned 2	7	HM535964	U12
Macu28	$(CA)_n(CGCA)_n(CA)_n$	F: TTTTAATCAAAATCGGTTCATCC R: TCAACCACAAAGCAAGTGAGTC	195-223	57	Fam2	12	KT851400	New
Macu29	$(TC)_n$	F: AAACGCGCTTATGGCTAAAC R: CGGTATGTCCCGTTACATCG	81-143	57	Vic 3	15	KT851401	New
Macu30	$(TG)_n$	F: GACGCGCTGTTATGTATTGC R: CGTCTAGCGTGACCGTAACA	93-109	57	Pet 4	5	HM586096	U11
Macu31	$(GTA)_n(GTC)_n(GTA)_n$	F: GTTCTGTCCCCCGAACTAGG R: AAACCTGGGATTGGTTAAAAAC	110-173	62	Ned 5	5	HM586097	U11
Macu40	$(CA)_n(GA)_n(CA)_n(GA)_n(CA)_n(GA)_n(CA)_n$	F: CCGTTTGGGAGATACGATGT R: CGCGTGTGCGTATATGTGAT	110-220	57	Pet 1	-	KT851402	New
Macu44	$(AC)_n$	F: ATAAGTCAGCACGTCAAAGCTG R: TGCAAATACTCCGAATAAATAACTG	170-220	57	Ned 3	10	HM535965	U12
Macu45	$(AC)_n(GC)_n(AC)_n$	F: TGTGTGACTGCGGTTCTTATC R: TGTAATCGCAGGAGAGATGTG	145-217	57	Vic 4	20	HM535966	U12

2   Primer concentration in PCR mix: 0.1 ng/µl: Macu20, Macu26, Macu 29, Macu30; 0.2 ng/µl: Macu 28, Macu 31, Macu 40, Macu 44, Macu 45

Details of sampled *Myrmica* nests

The number of nests found within 2 m of gentians at each site, their infection with *M. alcon H* or *M. alcon X*, the number of individual *M. alcon* used for genetic analysis ("Genetic samples": listed in table S1), and statistical tests of host ant specificity within each site: *P1* = probability from Fisher exact test and *P2* = probability from a randomization test of ant nests between species. Significant *P*-values ($P < 0.05$) are marked in bold.

Site	*Maculinea*	*Myrmica*	N°. nests	N°. with *M.alcon*	*P1*	N°. of *M. alcon*	Range	*P2*	Genetic samples
Răscruci dry	*alcon X*	*sabuleti*	10	5	**0.004**	17	1-8	**0.002**	13
		schencki	6	2		18	1-15		5
		scabrinodis	23	1		1			1
Răscruci wet	*alcon H*	*scabrinodis*	31	9	-	30	1-7	-	28
Răscruci both	both	as above			0.078			**0.021**	
Şardu	*alcon H*	*vandeli*	27	2	0.147	9	2-7	0.495	2
		scabrinodis	38	7		15	1-4		11

Pairwise differentiation between, and inbreeding and genetic diversity within pre-defined populations

Values above the diagonal in the matrix (with blue background) are θ (F_{ST}), values along the diagonal (with green background) are F_{IS}, values below the diagonal (with yellow background) are G''_{ST} / D_{EST}. Values in bold differ significantly from zero ($P < 0.001$). Below the matrix are mean values (±SE) of four different measures of within-population genetic diversity. The effective number of alleles per locus (A_E), the observed heterozygosity (H_O), the expected heterozygosity (H_E) and the unbiased expected heterozygosity (uH_E). P-values for comparisons between pre-defined populations based on mixed model comparison across loci are shown on the right.

	Răscruci dry	Răscruci wet	Șardu	
Răscruci dry	-0.050	**0.093**	**0.059**	
Răscruci wet	**0.302 / 0.255**	-0.106	**0.103**	
Șardu	**0.207 / 0.151**	**0.330 / 0.221**	-0.052	

				P
A_E	4.000 ± 0.606	3.365 ± 0.418	4.594 ± 0.909	0.131
H_O	0.707 ± 0.067	0.717 ± 0.056	0.733 ± 0.090	0.944
H_E	0.707 ± 0.046	0.674 ± 0.034	0.709 ± 0.063	0.781
uH_E	0.729 ± 0.048	0.687 ± 0.035	0.738 ± 0.066	0.613

Hierarchical Analysis of Molecular Variance

Calculated using HierFStat (Goudet, 2005). The *F*-coefficient gives the estimated inbreeding coefficient (excess of homozygotes) at each hierarchical level. *P*-values are based on 1000 re-samplings of the data.

Source	d.f.	Variance component	%variance	F-coefficient	P
Between **POP**	2	0.577	10.6	0.106	0.027
ANT within **POP**	3	-0.257	-4.7	-0.053	0.956
NEST within **ANT**	18	0.609	11.2	0.119	<0.001
Individuals within **NEST**	36	-0.569	-10.5	-0.126	>0.999
Within Individuals	60	5.067	93.4		
Total	119	5.427			

3

5   **Figure legends**

6   **Figure 1: Site layout and Bayesian clustering of samples. A)** Map showing the relative

7   position of the two sample sites. Shaded areas in the detailed map correspond to 50 m contour

8   lines. **B)** Initial food plants of the hygric (*G. pneumonanthe*) and xeric (*G. cruciata*) forms of

9   *Maculinea alcon*. Images modified from WikiMedia commons. **C)** Posterior probabilities of the

10   number of clusters (K) identified by the Bayesian population assignment program Structure. The

11   solid line shows the mean posterior probability for each value of K, with error bars representing

12   the standard deviation across simulations. The dashed line shows the ΔK values of the posterior

13   probabilities from Structure using the method of Evanno, Regnaut & Goudet (2005) shown

14   relative to the maximum value of ΔK. Peaks in the value of ΔK may represent different levels of

15   population substructure. **D)** Comparison of genetic clustering of samples into two and four

16   groups using the Bayesian clustering program STRUCTURE without a location prior. Each column

17   represents an individual, and is divided according to its probability of membership of cluster 1

18   (orange), 2 (blue), 3 (black) or 4 (green).

20   **Figure 2: Ordination of samples based on principal coordinate analysis**. Each symbol

21   represents an individual, coloured according to its pre-defined population (blue = Răscruci wet,

22   orange = Răscruci dry, purple = Șardu). Coloured lines are convex hulls enclosing all samples

23   from each pre-defined population, while coloured regions are convex hulls enclosing samples

24   collected from the same nest. The single individual (sample DB15) collected from a *My.*

25   *scabrinodis* nest at Răscruci dry is shown with a larger symbol.

26 **Figure 3: Relatedness and parentage analysis of samples**. The pairwise matrix show the

27 estimated Queller & Goodnight (1989) relatedness of each pair of individuals (excluding those

28 with negative relatedness) above the diagonal, and the probability that each pair are full siblings

29 based on maximum likelihood estimates from COLONY (Jones & Wang, 2010) below the diagonal.

30 Comparisons between samples from the same pre-defined population are shaded according to the

31 same colour scheme as Figure 2 (blue = Răscruci wet, orange = Răscruci dry, purple = Șardu).

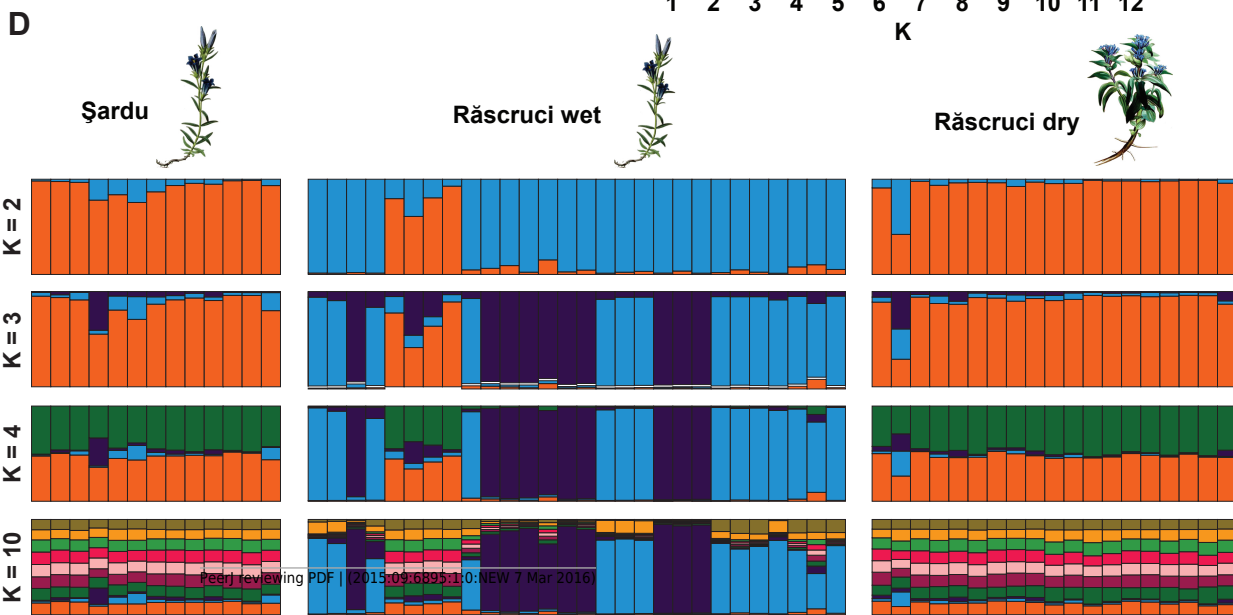
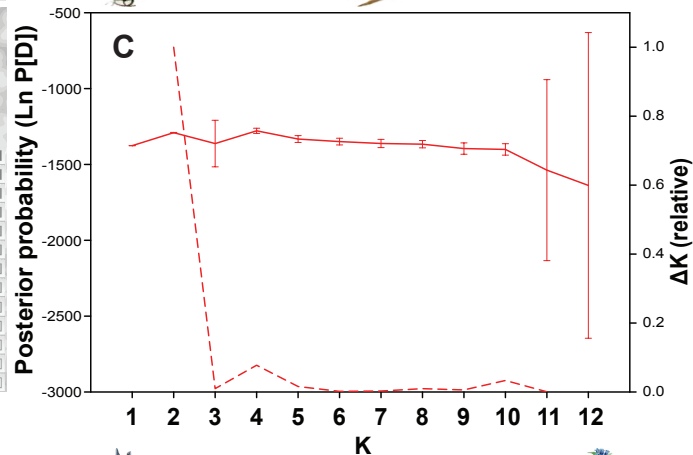
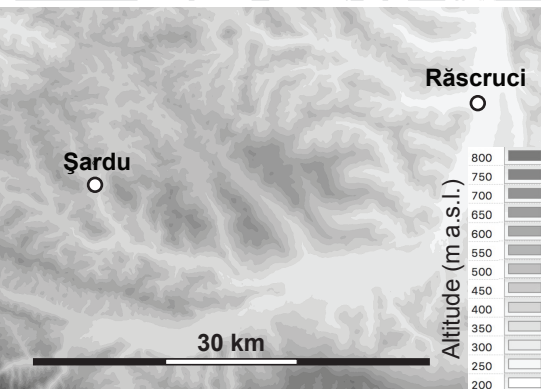
32 Individuals sharing the same ant nest are outlined with lines in these same colours, and those

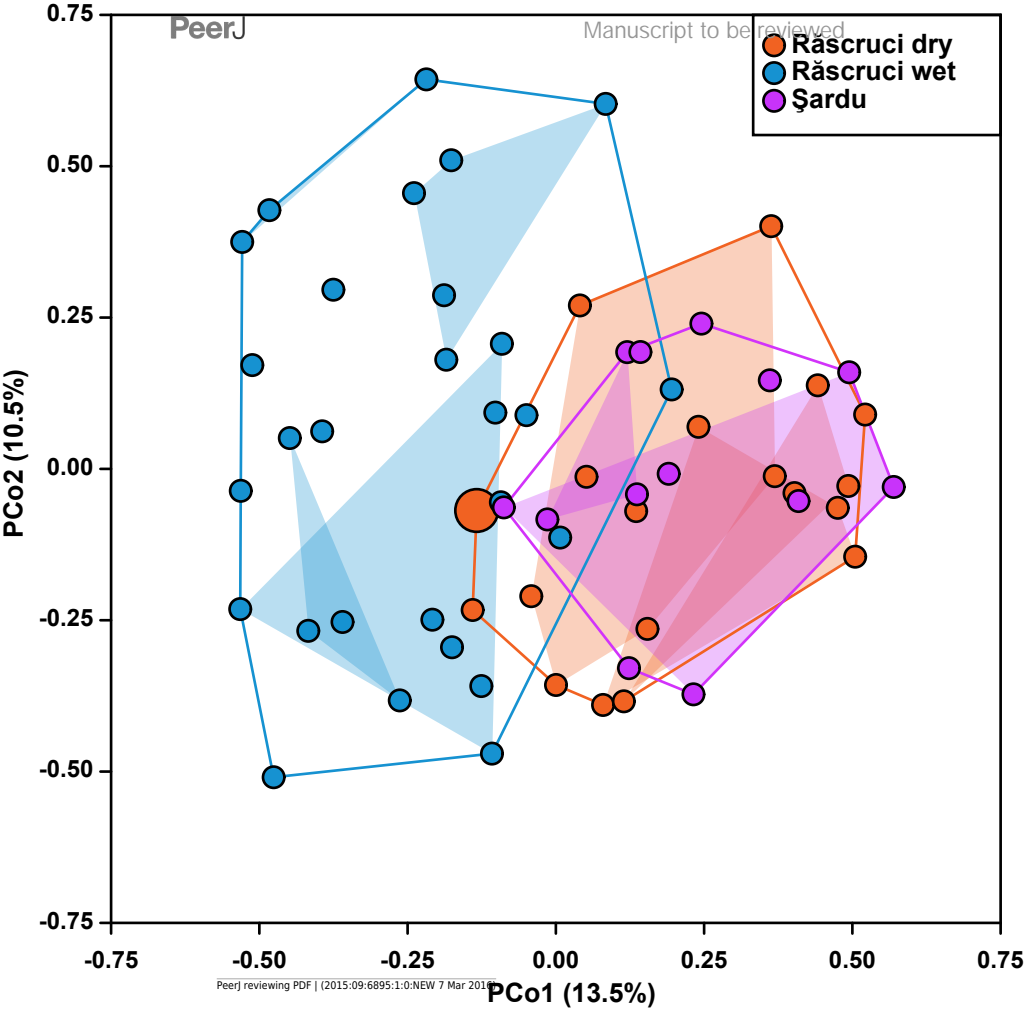
33 sharing the same *Myrmica* species as host are outlines with black lines. The area and shade of each

34 data point is proportional to the relatedness or probability of being full siblings for that pair of

35 individuals.

Site layout and Bayesian clustering of samples

A) Map showing the relative position of the two sample sites. Shaded areas in the detailed map correspond to 50 m contour lines. **B)** Initial food plants of the hygric (*G. pneumonanthe*) and xeric (*G. cruciata*) forms of *Maculinea alcon*. Images modified from WikiMedia commons. **C)** Posterior probabilities of the number of clusters (K) identified by the Bayesian population assignment program Structure. The solid line shows the mean posterior probability for each value of K, with error bars representing the standard deviation across simulations. The dashed line shows the ΔK values of the posterior probabilities from Structure using the method of Evanno, Regnaut & Goudet (2005) shown relative to the maximum value of ΔK. Peaks in the value of ΔK may represent different levels of population substructure. **D)** Comparison of genetic clustering of samples into two and four groups using the Bayesian clustering program Structure without a location prior. Each column represents an individual, and is divided according to its probability of membership of cluster 1 (orange), 2 (blue), 3 (dark purple) 4 (green), or 5-10 (other colours).

Ordination of samples based on principal coordinate analysis

Each symbol represents an individual, coloured according to its pre-defined population (blue = Răscruci wet, orange = Răscruci dry, purple = Şardu). Coloured lines are convex hulls enclosing all samples from each pre-defined population, while coloured regions are convex hulls enclosing samples collected from the same nest. The single individual (sample DB15) collected from a *My. scabrinodis* nest at Răscruci dry is shown with a larger symbol.

Relatedness and parentage analysis of samples

The pairwise matrix show the estimated Queller & Goodnight (1989) relatedness of each pair of individuals (excluding those with negative relatedness) above the diagonal, and the probability that each pair are full siblings based on maximum likelihood estimates from Colony (Jones & Wang, 2010) below the diagonal. Comparisons between samples from the same pre-defined population are shaded according to the same colour scheme as Figure 2 (blue = Răscruci wet, orange = Răscruci dry, purple = Şardu). Individuals sharing the same ant nest are outlined with lines in these same colours, and those sharing the same *Myrmica* species as host are outlines with black lines. The area and shade of each data point is proportional to the relatedness or probability of being full siblings for that pair of individuals.

Relatedness / Probability of being full siblings