1 **Cryptic diversity and multiple origins of the widespread mayfly species group *Baetis rhodani***

2 **(Ephemeroptera: Baetidae) on northwestern Mediterranean islands**

3

4

5

6 Roberta Bisconti [a], Daniele Canestrelli [a,*], Roberta Tenchini [a,b], Carlo Belfiore [a], Andrea Buffagni [b],

7 Giuseppe Nascetti [a].

8

9 [a] Dipartimento di Scienze Ecologiche e Biologiche, Università della Tuscia. Viale dell'Università s.n.c., I-

10 01100 Viterbo, Italy.

11 [b] CNR – IRSA, Water Research Institute, Via del Mulino 19, 20861 Brugherio (MB), Italy

12

13

14

15 *Correspondence address:

16 Dipartimento di Scienze Ecologiche e Biologiche, Università della Tuscia. Viale dell'Università s.n.c., I-

17 01100 Viterbo, Italy.

18 e-mail: canestrelli@unitus.it

19 phone: +39-0761-357414

20

21 **Abstract**

22 How the often highly endemic biodiversity of islands originated has been debated for decades and it remains

23 a fervid research ground. Here, using mitochondrial and nuclear gene sequence analyses, we investigate the

24 diversity, phylogenetic relationships, and evolutionary history of the mayfly *Baetis gr. rhodani* on the three

25 largest north-western Mediterranean islands (Sardinia, Corsica, Elba). We identify three distinct, largely co-

26 distributed, and deeply differentiated lineages, with divergences tentatively dated back to the Eocene-

27 Oligocene transition. Bayesian population structure analyses reveal a lack of gene exchange between them,

28 even at sites where they are syntopic, indicating that these lineages belong to three putative species. Their

29 phylogenetic relationships with continental relatives, together with the dating estimates, support a role for

30 three processes contributing to this diversity: (1) vicariance, primed by microplate disjunction and oceanic

31 transgression; (2) dispersal from the continent; and (3) speciation within the island group. Thus, our results

32 do not point toward a prevailing role for any of the previously invoked processes. Rather, they suggest that a

33 variety of processes equally contributed to shape the diverse and endemic biota of this group of islands.

34

35

36 **Keywords:** Island biogeography, Mediterranean basin, *Baetis*, Tyrrhenian islands, Vicariance, Dispersal.

37

38 **Introduction**

39

40 Which processes best explain patterns of diversity on islands has been a central topic of island biogeography

41 since its inception, and it continues to be a hot-topic for research and debate (Whittaker and Fernàndez-

42 Palacios, 2007). The geological origin of an island system has been traditionally regarded as a major factor in

43 explaining the structure of its biota. Accordingly, the highly distinctive biota of continental islands was often

44 interpreted as the long-term product of divergence in isolation following habitat fragmentation, more or less

45 complemented by neoendemic forms (Watson, 2009). This traditional view has been substantially enriched

46 thanks to the advent of sharper research tools (e.g. population genetic, molecular phylogenetic, and

47 phylogeographic approaches). Several of the best-studied island systems of continental origin have appeared

48 to have an 'oceanic' nature, that is, they are dominated by neoendemic forms (e.g. Grandcolas et al., 2008;

49 Goldberg et al., 2008). Vicariance and long-distance dispersal, long regarded as opposing explanations

50 (Heads, 2009), are now increasingly perceived as often interacting processes in the assembly of an insular

51 biota (Lomolino et al., 2010). In addition, microevolutionary processes priming divergence among

52 populations within island have been recognized as substantial contributors to the biodiversity of islands,

53 regardless of those islands' geological origin and the mode of species settlement on them (e.g. Bisconti et al.,

54 2013a, 2013b; Emerson et al., 2006; Holland and Hadfield, 2002; Thorpe and Malhotra, 1998; Villacorta et

55 al., 2008; Wallis and Trewick, 2009). Although a few iconic island systems have dominated the scene of

56 such reappraisal, it is noteworthy that as more islands and taxa are investigated, the range of processes seen

57 as contributing to the assembly and diversity of current insular biota becomes wider (e.g. Bauzà-Ribot et al.,

58 2011; Bisconti et al., 2011a; Salvi et al., 2014).

59 The Tyrrhenian (northwestern Mediterranean) islands belong to the largest continental fragment

60 within the Mediterranean basin. This fragment originated by detachment from the Iberian Peninsula, the

61 counterclockwise rotation of the microplate (initiated 21 to 15-18 Ma ago; Gattaccecca et al., 2007), and

62 subsequent oceanic spreading (Brandano & Policicchio, 2012), while the disjunction of the two main islands,

63 Sardinia and Corsica, was completed about 9 Ma ago (Alvarez, 1972; Bellon et al., 1977; Bonin et al., 1979).

64 Subsequent land connections between islands and with the continent formed repeatedly, as a consequence of

65 Mediterranean sea-level oscillations (Hsü et al., 1973; Krijgsman et al., 1999; Shackleton et al., 1984; Van

66   Andel and Shackleton, 1982). This island system shows an extremely wide range of climates, landscapes,

67   and habitats, and both its paleoclimates and its palaeoenvironments have been investigated in detail

68   (Vogiatzakis et al., 2008). Consequently, in light of the wealth of background information available, it offers

69   an ideal system for exploring both traditional and new hypotheses about the processes leading to the

70   assembly and uniqueness of continental island biotas, an opportunity that has not escaped attention in the

71   recent past.

72   Several hypotheses on the origin of the Tyrrhenian islands biota have been put forward during the

73   last century. At first, it was proposed that such a distinctive biota formed via hologenesis, a process implying

74   a prominent role for vicariance events (Monti, 1915; Monterosso, 1935; Luzzatto et al., 2000). Subsequently,

75   more emphasis was given to dispersal, and a three-step process of colonization was hypothesized (Baccetti,

76   1964): a pre-Miocene step through a supposed land bridge with the Baetic region (southern Spain); a

77   Miocene step which brought warm-temperate species from North Africa and Italy; and a Quaternary step

78   which brought temperate species from the Apennines (Italy) via a land bridge connecting Tuscany with

79   northern Corsica. Once the theory of plate tectonics spread (Alvarez, 1972), a reappraisal of vicariance as the

80   leading force occurred, particularly to explain the most distinctive (endemic) components of this insular biota

81   (Baccetti, 1980). With the advent of molecular phylogenetic approaches, data on several taxonomic groups

82   began to accumulate (Grill et al., 2007; Ketmaier and Caccone, 2013), leading to a reconsideration of

83   dispersal events, although vicariance events primed by the microplate disjunction are still regarded by

84   several authors as the main triggers of endemic species formation on the islands (Ketmaier and Caccone,

85   2013). Finally, in more recent times, the application of the phylogeographic toolbox to the study of

86   evolutionary histories within-islands (i.e. post-settlement), is shedding light on a plethora of previously

87   unappreciated evolutionary processes contributing to the diversity of the island biota above and below the

88   species level (Bisconti et al., 2011a, 2011b; Bisconti et al., 2013a, 2013b; Falchi et al., 2009; Gentile et al.,

89   2010; Ketmaier et al., 2010; Salvi et al., 2009; Salvi et al., 2010).

90   In this paper, we focus on mayflies of the *Baetis rhodani* (Pictet, 1843) species group, which is one

91   of the most common mayflies in the Western Palaearctic region, and one of the most abundant insects in

92   freshwater running environments (Brittain, 1982). Phylogenetic and phylogeographic investigations of this

93   species group have proliferated in recent years. In spite of the homogeneity of morphological characters,

94 such studies have revealed unexpectedly deep divergences. Williams and colleagues (2006) investigated the

95 genetic structure of *B. rhodani* populations in the Western Palaearctic, and identified multiple divergent

96 lineages, probably belonging to distinct cryptic species. Lucentini et al. (2011), expanded the analysis to

97 other European countries (UK and Italy), increasing the number of lineages. More recently, populations from

98 the Canary Islands have also been investigated (Rutschmann et al., 2014), again showing deep divergences

99 and several endemic lineages, likely resulting from a complex evolutionary history. Unfortunately, these

100 studies were based on a single mitochondrial DNA marker (mostly cytochrome oxidase I), preventing sound

101 inferences about the species status and the evolutionary processes involved. Nonetheless, results of these

102 studies point to these mayflies as potentially excellent models for evolutionary and historical biogeographic

103 studies in the Western Palaearctic.

104   Within the Tyrrhenian islands, a single endemic species, *Baetis ingridae* (Thomas and Soldàn, 1987)

105 has been described from the *B. rhodani* group, based on a sample from northeastern Corsica. This species

106 shows subtle morphological differences with respect to continental *B. rhodani*. Their larvae are very similar,

107 with differences affecting colour patterns, the shape of the femoral bristles, and the shape and number of the

108 teeth on the inner margin of the paraproct (Bauernfeind and Soldàn, 2012). So far, the diversity (including

109 genetic diversity) and phylogenetic relationships of this mayfly within the Mediterranean islands is still

110 virtually unexplored.

111   Here we employ a set of nuclear (nDNA) and mitochondrial (mtDNA) gene regions with the aim of

112 investigating (1) the diversity of the *B. rhodani* species group in the Tyrrhenian islands; (2) the phylogenetic

113 relationships with its continental relatives; (3) its evolutionary history, with special reference to the mode of

114 settlement on islands; and ultimately (4) to aid shedding more light on the historical origin of these islands'

115 highly endemic diversity.

116

117 **Materials and Methods**

118

119 **Sampling**

120 Larval individuals of the *Baetis rhodani* species group were collected from aquatic habitats at 28 localities:

121 23 localities from the two main Tyrrhenian islands (Sardinia and Corsica); 2 samples from the Elba island,

5

122 which is close to the northeastern side of Corsica and shows faunal and floral affinities with that island; and

123 3 samples from peninsular Italy (i.e. the nearest mainland), added to this study for comparative purposes.

124 Taxonomic assignment of the sampled individuals to the *B. rhodani* species groups was carried out using

125 well established morphological characters of diagnostic value (Müller-Liebenau, 1969). The geographic

126 location of the sampling sites and sample details are shown in Table 1 and Figure 1B. Individuals of the

127 *Baetis vernus* species group (sampling coordinates: 42° 12.72' N; 12°24.95' E) were added to the sampling

128 for use as an outgroup in phylogenetic analyses. The sampled individuals were transported to the laboratory

129 and preserved in 95% ethanol until DNA extraction.

130

131 **DNA extraction, amplification, and sequencing**

132 We extracted whole genomic DNA using proteinase K digestion followed by a standard phenol-chloroform

133 protocol with RNase treatment (Sambrook et al., 1989). Polymerase chain reactions (PCR) were carried out

134 to amplify fragments of two mitochondrial genes (cytochrome oxidase 1, hereafter referred to as CO1;

135 NADH dehydrogenase subunit 1, hereafter referred to as ND1), and two nuclear genes (phosphoenol

136 pyruvate carboxykinase, hereafter referred to as PEPCK; ryanodine receptor 44f, hereafter referred to as

137 RYA). PCR primers and protocols for the gene fragments CO1 and PEPCK were drawn from the literature

138 (Simon et al. 1994; Pereira-da-Conceicoa et al., 2012). PCR primers for the ND1 gene fragment were

139 designed using PRIMER 3, based on the complete genome alignment of *Parafronurus youi*, *Ephemera*

140 *orientalis*, and *Baetis* sp., downloaded from the GenBank database (accession numbers EU349015.1,

141 EU591678.1, and GU936204.1, respectively). For the RYA gene fragment, the PCR procedure was newly

142 developed for the present study using the EvolMarkers bioinformatic tool (Li et al., 2010). With this tool, we

143 searched for exon-primed intron-crossing (EPIC) markers, using annotated genomes of *Drosophila*

144 *melanogaster* ('query'), *Aedes aegypti* ('subject'), and *Apis mellifera* ('subject'), since annotated genomes of

145 mayflies were not available. PCR primers were designed for 10 putative EPIC markers using PRIMER 3.

146 After preliminary PCR trials, we focused on five putative markers which were consistently amplified in a

147 sample of 10 randomly selected individuals, and which provided single and clean PCR bands in a standard

148 (1.5x) agarose-gel electrophoresis. Among these markers, the RYA gene fragment returned high-quality

149   sequence electropherograms, and was thus retained and used to analyse the entire pool of sampled

150   individuals.

151   PCR cycling conditions were the same for all genes: 5 min at 92°C followed by 30 cycles of 1 min at

152   92°C, 1 min at an annealing temperature specific for each gene, and 90 s at 72°C, followed by a single final

153   step of extension of 10 min at 72°C. Amplifications were conducted in 25 µL containing the following:

154    $MgCl_2$ (2.5 mM), the reaction buffer (1X; Promega), 4 dNTPs (0.2 mM each), 2 primers (0.2 µM each), the

155   enzyme Taq polymerase (1 unit; Promega), and 2 µl of DNA template. PCR primers and annealing

156   temperature specific for each gene fragment analysed are given in Table S1.

157   PCR products were purified and sequenced by Macrogen Inc. (http://www.macrogen.com) using an ABI

158   PRISM 3700 sequencing system. The electropherograms were visually checked using CHROMAS 2.31

159   (Technelysium ltd.). All sequences were deposited in the GenBank database. Multiple sequence alignments

160   were produced using CLUSTALX (Thompson et al., 1997), under the default settings.

161

162   **Data analysis**

163   Sequence variation, substitution patterns, and divergence between haplotypes and between the main

164   haplotype groups were assessed using MEGA 5.1 (Tamura et al., 2011).

165   Since this study is the first to investigate patterns of genetic diversity and affinities of mayflies of the *B.*

166   *rhodani* group throughout the Tyrrhenian islands, we began our data exploration by verifying their purported

167   endemicity, that is, whether they belong to well supported and divergent clades, geographically restricted to

168   the Tyrrhenian islands. It is worth emphasizing, however, that we refrained from further interpreting the

169   results of this analysis, since it relied on a final alignment of limited extension (see Results) and on a single

170   gene fragment. This task was undertaken through a comparison of our CO1 data (the most widely used

171   genetic marker in previous studies) with previously published sequences of *B. rhodani*. To this end, we

172   retrieved one representative haplotype sequence of each main clade previously observed by Williams et al.,

173   (2006), Lucentini et al., (2011), and Murria et al., (2014), in order to have the widest possible coverage of

174   continental Europe, and we built a Bayesian phylogenetic tree as described below.

175   Gene trees for the CO1 dataset, as well as for the concatenated mtDNA, PEPCK, and RYA datasets,

176   were estimated by means of the Bayesian inference procedure implemented in MRBAYES v.3.2.1 (Ronquist

7

177 et al., 2012). The nuclear gene sequences were phased using the software PHASE 2.1 (Stephens et al., 2003)

178 with the default settings, while the possible occurrence of recombination was assessed using the pairwise

179 homoplasy index (PHI statistic, Bruen et al., 2006) implemented in SPLITSTREE v.4.11 (Huson and Bryant,

180 2006). The best partitioning strategy for the concatenated mtDNA was selected by means of

181 PARTITIONFINDER v1.0.1 (Lanfear et al., 2012), using the Bayesian information (BI) criterion and the

182 'greedy' search strategy. This analysis suggested TrN + I + G as the best model for this dataset. The best-fit

183 model of nucleotide substitution for nuclear gene fragments (nDNA) was selected among 88 alternative

184 models using the BI criterion implemented in JMODELTEST 2.1.3 (Darriba et al., 2012). This analysis

185 indicated TrNef + G and K80 + G as the best-fit substitution models for the PEPCK and the RYA genes,

186 respectively. For gene tree estimations, all the analyses consisted of two independent runs of four Markov

187 chains, each run consisting of 10 million generations sampled every 1000 generations. Stationarity of the

188 analyses was verified by examining log-likelihood scores using TRACER v1.6 (Rambaut et al., 2014), and

189 chain convergence between simultaneous runs was evaluated by means of the average standard deviation of

190 split frequencies (all values <0.001). For the nuclear datasets, which showed substantially lower levels of

191 variation than the mtDNA dataset (see Results), gene trees were converted into haplotype networks using

192 HAPLOTYPEVIEWER (Guidon et al., 2004).

193   Nuclear DNA data were further analysed using the non-hierarchical Bayesian clustering procedure

194 implemented in BAPS 6.0 (Corander et al., 2008), which allows us to identify the best clustering option

195 based on multilocus datasets, to assign individuals to clusters, and, importantly, to identify individuals of

196 mixed ancestry among groups. We ran five replicates of this clustering analysis to check its consistency,

197 based on the 'clustering with linked loci' option, the 'independent loci' model, and with the maximum

198 number of clusters (KMAX) set to 20. The admixture analyses were run for 200 iterations, with 100

199 reference individuals and 50 iterations per reference individual.

200   Finally, we used the Bayesian inference procedure implemented in *BEAST (Heled and Drummond,

201 2010), in order to co-estimate individual gene trees for the four gene fragments analysed (CO1, ND1, RYA,

202 PEPCK), and to embed them into a single multilocus species-tree. With this method, terminal taxa have to be

203 indicated a priori, and so we used the results of previous phylogenetic inferences based on mtDNA and the

204 non-hierarchical cluster analysis of nDNA data. Substitution, clock, and tree models were unlinked, and each

205   dataset was assigned the appropriate ploidy level and the substitution model previously estimated by means

206   of PARTITIONFINDER and JMODELTEST. The analysis was run using the Yule species-tree prior, and a

207   strict molecular clock model was assumed. In order to get an approximate temporal context for the main

208   divergences inferred, and in the absence of internal calibration points or fossil records, we used the CO1

209   specific substitution rate of 0.0115 substitutions per site per Ma (Brower, 1994) which has been extensively

210   used to date divergences in insects, including mayflies, and which falls in the middle of the ranges estimated

211   for this gene fragment in insects. Substitution rates for the remaining gene fragments were estimated using

212   the CO1 specific rate as a reference, and the 1/x prior distribution. Five independent *BEAST analyses were

213   performed to check for convergence, each run for 100 million generations, and sampling every 10,000.

214   Convergence of the analyses, the appropriate number of samples to discard as burn-in (10%), stationarity,

215   and the achievement of appropriate ESS values ($\gg 200$) were inspected using TRACER. The final species

216   tree was estimated as a maximum clade credibility tree (MCC) summarizing results of the post-burn-in trees,

217   with nodes annotated with ages and the corresponding 95% highest posterior densities (HPD), by means of

218   the BEAST module TREEANNOTETOR. In addition, the entire posterior distribution of species-trees

219   obtained with *BEAST was inspected using DENSITREE.

220

221   **Results**

222

223   For the 112 individuals of the *B. rhodani* species group used in this study, we obtained fully resolved

224   sequences (i.e., without missing data) for all the gene fragments analysed. Sequence length was as follows:

225   413 bp for CO1, 673 bp for ND1, 290 bp for PEPCK, and 178 bp for RYA. Within the combined mtDNA

226   dataset (overall 1086 bp), 496 variable positions were found (422 excluding the outgroup), of which 390

227   were parsimony informative (381 excluding the outgroup). No indels, stop codons, nonsense codons, or

228   heterozygous positions were observed within the CO1 and ND1 genes, consistent with the expectation for

229   coding regions of mitochondrial origin. Within the nuclear gene sequences, indels were found of length 1-3

230   bp. For all downstream analyses, indels were removed from the alignments. The PEPCK gene presented 54

231   variable positions of which 45 were parsimony informative (41 and 32, respectively, excluding the

232   outgroup), whereas the RYA gene showed 76 variable positions of which 61 were parsimony informative (54

9

233   and 32, respectively, excluding the outgroup). No statistically significant indications of recombination events

234   were found with the PHI tests carried out on the nuclear gene fragments.

235   A CO1 alignment of length 412 bp was generated by pooling our CO1 sequence with those retrieved

236   from the literature. The Bayesian tree estimation carried out on this dataset (Figure S1) clearly identified our

237   individuals as belonging to 3 well supported clades (clades A, B, C) geographically restricted to the

238   northwestern Mediterranean islands, and divergent from all other clades. Two of these clades (A, B) showed

239   sister relationships, whereas the third clade (C) was nested within continental haplotypes. Thus, we found no

240   evidence against the purported endemicity of the insular representatives of the *B. rhodani* species group,

241   although they did not appear to be a monophyletic group.

242   The Bayesian phylogenetic analysis performed on the concatenated mtDNA (Figure 1A) identified a

243   well resolved topology, with most nodes having a posterior probability ≥0.99. Five main clades were found:

244   3 observed within the Tyrrhenian islands (clades A, B, and C, as above), and 2 only found on the

245   neighbouring mainland (clades D, E). Within clade A, two subclades were observed, one geographically

246   restricted to Elba (haplotypes H20, H21), the other restricted to Corsica. Clade B was observed both in

247   Corsican and Sardinian samples, whereas it was not observed on Elba. Instead, haplotypes of clade C were

248   carried by individuals from all the sampled islands. Clades A, B, and C together were not monophyletic, the

249   latter being nested within Italian clades. The mean uncorrected sequence divergence among clades (Table 2)

250   ranged between 0.12 (clade A vs. B) and 0.269 (clade A vs. D), whereas mean within-clade divergence

251   ranged between 0.006 (clade C) and 0.011 (clade A).

252   The non-hierarchical cluster analysis of the nDNA dataset carried out with Baps indicated K = 4 as the

253   option best fitting the data (posterior probability = 1.00). Individuals from insular samples were assigned to

254   three distinct clusters, fully corresponding to clades A, B, and C identified from the mtDNA dataset.

255   Contrary to what was previously found from the mtDNA, all individuals from the mainland were grouped

256   into a single cluster. Moreover, with this analysis, no individuals of mixed ancestry between clades were

257   found, even within population samples where distinct mtDNA clades and nuclear clusters were observed to

258   be syntopic (samples 7, 12, 20, 25).

259   The phylogenetic networks built among haplotypes found at the nuclear gene fragments PEPCK and

260   RYA are shown in Figure 2. Within both networks, haplotypes from individuals carrying the mtDNA clades

261 A, B, and C can be identified as groups of closely related nDNA haplotypes. Heterozygote individuals

262 carrying haplotypes from two of these groups were not observed, not even within sites of syntopy. Haplotype

263 from these groups were never observed In addition, a closer affinity between the nDNA group C (yellow in

264 Figure 2) and the Italian haplotypes was observed. However, the genealogical relationships among Italian

265 haplotypes and between them and haplotypes of group C appeared much less clearly resolved than with the

266 mtDNA dataset. Within the insular populations where the three haplogroups from distinct groups groups

267 were found, heterozygotes for haplotypes from the three distinct groups.

268   The multilocus species-tree analysis with *BEAST was carried out using three alternative strategies for

269 the a priori definition of terminal taxa. First, following results based on the CO1 dataset, we added to our

270 multilocus dataset two individuals from the closest relatives of our insular clades. We defined a set of 7

271 ingroup taxa, 5 resulting from the Bayesian analysis of the concatenated mtDNA data, plus 2 groups

272 retrieved from the literature (G1-G2 [Lucentini et al., 2011] and G5-Brho2hap8 [Lucentini et al., 2011;

273 Murria et al., 2014], here referred to as clades G and H). The *BEAST analysis was then run with the non-

274 CO1 data from these 2 latter groups encoded as missing. Second, the same analysis was re-run after the

275 exclusion of CO1 sequences derived from the literature (i.e. with 5 terminal taxa). Third, individuals were

276 assigned to four terminal taxa defined according to results of the non-hierarchical cluster analysis of the

277 nDNA data, that is, with the mainland samples grouped together. With 7 groups (see Figure 3), the *BEAST

278 analysis confirmed with high support a closer affinity between clade C and the continental clade G (rather

279 than between clade C and the other insular clades). The divergence between the two main clades was

280 estimated to have occurred at 31.4 Ma (95% HPD: 21.0–42.4), whereas the divergence between the two

281 insular clades A and B was dated at 7.4 (95% HPD: 4.6–10.3). Divergence between clade A + B and its

282 closest continental relative (H) was estimated at 19.9 (95% HPD: 12.0–29.3). Instead, the divergence

283 between the insular clade C and clade G was dated at 2.8 (95% HPD: 0.5–4.8). Carrying out the analysis

284 with 4 terminal taxa did not appreciably affect branching patterns, nodal support, or node age estimates,

285 whereas node ages were 1 to 4 Ma older with 5 groups than with the other grouping options, although they

286 fell well within the respective 95% HPDs (see Figure S2).

287

288 **Discussion**

11

289

290   With the ever increasing application of barcoding procedures to the study of biological diversity, a wealth of

291   animal and plant species, including mayflies of the *Baetis rhodani* group, are revealing previously

292   unsuspected levels and depths of variation, with potentially far-reaching implications for their taxonomy,

293   systematics, and conservation, as well as for our understanding of their ecologies and evolutionary pathways.

294   Unfortunately, however, in most cases mtDNA is the only component of the genetic variation analysed,

295   which prevents a full appraisal of the biological meaning of the observed patterns of diversity (Balloux,

296   2010; Hudson and Coyne, 2002).

297   Within the *B. rhodani* group, all previous studies identified deeply divergent mtDNA (CO1)

298   lineages, often occurring in syntopy. Nonetheless, since mtDNA provides no information about the

299   occurrence and extent of gene exchange between diverging lineages, conclusions about the many

300   implications of these results remained open. In the present study, we took advantage, for the first time in the

301   *B. rhodani* group, of combined mitochondrial and nuclear perspectives, which allowed us to overcome the

302   many pitfalls associated to the use of mtDNA alone when addressing patterns of genetic variation and

303   species' evolutionary histories.

304   Within the Tyrrhenian islands and neighbouring Elba, we found three endemic lineages (Figure S1)

305   that were deeply divergent in their mtDNA. Indeed, the observed degree of divergence (Table 2) largely

306   exceeded the suggested threshold for putative species, in one case (haplogroups A + B vs. C) exceeding also

307   the one frequently observed among congeneric mayflies (18%; Ball et al., 2005). Furthermore, based on the

308   inferred phylogenetic pattern, they appeared to be independently derived from continental relatives (see

309   below). Nuclear DNA data confirm this pattern and, when analysed in a Bayesian population genetics

310   framework, clearly indicate the lack of admixture between lineages, even where the lineages are sympatric

311   (see Figure 1 and results of BAPS analysis). Thus, overall, our data clearly indicate that the three lineages we

312   observed belong to three distinct species, under both phylogenetic and biological species concepts (Hausdorf,

313   2011).

314   Our sampling localities in northeastern Corsica (samples 18 and 19) belong to the same drainage

315   system, and are located close to the type locality of *Baetis ingridae* (Le Bevinco, northeastern Corsica;

316   Thomas and Soldàn, 1987). Since only lineage B was observed at these sites, it seems plausible to

317 hypothesize that lineage B belongs to *B. ingridae*, whereas lineages A and C belong to two previously

318 undescribed species. Nonetheless, this hypothesis deserves further confirmation in conjunction with genetic

319 characterization of the type material.

320   There are several sources of uncertainty in our time estimates, including our reliance on a fixed and

321 non-organism-specific molecular rate, and the lack of internal fossil calibration points. In addition, we

322 refrained from using the disjunction of the plate from the continent as a dating point, since we had no data

323 supporting this choice, and we cannot firmly exclude subsequent overseas dispersal for this mayfly

324 (Monaghan et al., 2005). In spite of these major limitations, our dating exercise yielded a time estimate for

325 the divergence between lineage A + B and its closest continental relative clade (19.9 Ma; see Figure 3) that

326 fits comfortably with the oceanic spreading in the Liguro-Provençal basin (~21–18 Ma; Gattaccecca et al.,

327 2007; Brandano & Policicchio, 2012; Tomassetti et al., 2013), separating the Corsica-Sardinia block from

328 the Iberian peninsula. Since intervening marine barriers may in fact fragment mayfly populations, this time

329 coincidence appears strongly suggestive of a causal relation.

330   In line with previous studies (Ketmaier and Caccone, 2013), our results support an important role for

331 the microplate disjunction and the consequent paleogeographic events in the origin of the highly endemic

332 biota of this island complex. Nonetheless, they also emphasize that this paleogeographic process has not

333 been the only or the major forge of endemic taxa. At least two further processes contributed: (1) speciation

334 within the island complex, and (2) colonization from the continent.

335   Speciation within the island complex, which is invoked here to explain the occurrence of the

336 divergent clades A and B, is a necessary consequence of the inferred origin of their common ancestor via

337 microplate disjunction. Our time estimate of the divergence between clades A and B (and the associated

338 uncertainty) would suggest a role for Corsica and Sardinia in this process. Nevertheless, in this case, we

339 refrain from indicating this as the most plausible scenario, for three main reasons. First, there is no strong

340 consensus on the time of final separation between Corsica and Sardinia (Alvarez, 1972; Bellon et al., 1977;

341 Bonin et al., 1979; Gattaccecca et al., 2007). Second, both clades have been found in Corsica, and our data

342 are not appropriate to carry out a sound analysis of the respective ancestral areas; more data will be needed

343 for that purpose. Third, and more importantly, without clear evidence to invoke specific scenarios, we cannot

344 exclude non-allopatric processes of speciation from the set of hypotheses (Coyne and Orr, 2004), especially

345   considering that knowledge of the evolutionary ecology of this group of mayflies is still virtually null. Thus,

346   how clades A and B originated, and how they reached reproductive isolation, will necessarily be a subject for

347   future research.

348   Lineage C was the most differentiated among the three endemic lineages of the *B. rhodani* group on

349   the island complex. The mean estimate of the divergence time between the respective clades (31.4 Ma) lies

350   close to the Eocene-Oligocene transition (the so-called *Grande Coupure* or 'great break'), an epoch of major

351   and synchronous floral, faunal, and climatic turnover (Liu et al., 2009; Hooker et al., 2004). Nonetheless, our

352   estimate has much uncertainty, and any hypothesis about the underling processes would require a wealth of

353   weakly justified assumptions so that, again, we prefer not to set them at all. On the other hand, it is worth

354   noting that both for this event and for the divergence between lineage C and its closest continental relative

355   clade, our mean time estimates (2.9 Ma in the latter case) lie close to major and parallel climatic transitions

356   toward cooler and drier climates. Furthermore, the Plio-Pleistocene transition (~2.6 Ma) also introduced a

357   marked seasonality in the Mediterranean region, and led to an abrupt sea level drop that could have made

358   environmental conditions more favourable to a colonization of the Tyrrhenian islands from neighbouring

359   continental areas (Thompson, 2005).

360   Finally, although the long polarized vicariance-dispersal debate has now been largely replaced by a

361   more integrated view that sees them as variously interacting processes, it continues to form the basis of

362   historical biogeographic investigations in several instances. In the context of the Tyrrhenian islands, the view

363   that the microplate disjunction (i.e. vicariance) was the leading process responsible for the islands' highly

364   endemic biota has historically prevailed (Ketmaier & Caccone, 2013). Our results do not support this view,

365   and add to some previous studies (e.g. Nascetti et al., 1996) in showing that not only both vicariance and

366   dispersal played a part, but also that other processes, which acted within the island complex, were involved.

367   These results thus invite us, once more (e.g. Whittaker and Fernandez-Palacios, 2007), to a sharper shift from

368   an either-or logical framework to a more integrative one, which is increasingly appearing to provide a better

369   perspective on the complexity of the processes that have moulded patterns of biological diversity.

370

371   **Concluding remarks**

372

14

373   Patterns of genetic variation above and below the species level have remained largely unexplored in baetid

374   mayflies until recent times. This study adds to several recent ones in showing that, in spite of their limited

375   morphological variation, they usually show high levels of deeply structured genetic diversity. Besides

376   retaining the genetic footprints of past evolutionary processes over a wide time scale, they are rather easy to

377   sample (during the aquatic larval stages), are usually common and abundant at sites of occurrence, and given

378   the very limited duration of the adult life stages (in the order of hours), they show comparatively limited

379   dispersal abilities with respect to other flying insects. All these features make these organisms excellent

380   models for historical biogeographic and evolutionary studies.

381   Here, we have identified the occurrence of three distinct and deeply divergent species within the

382   *Baetis rhodani* group in the northwestern Mediterranean islands. For the time being, we refrained from

383   making formal taxonomic descriptions, since we judged our data suboptimal for that purpose. This task shall

384   await the examination (genetic and morphological) of the available type material, and following

385   comparisons, the designation of new type material.

386   Finally, although we did not carry out statistical phylogeographic and historical demographic

387   analyses on our samples, given the limited sample size per lineage, it did not escape our notice that the three

388   lineages we found showed substantial intraspecific variation. Thus, as soon as a larger sample becomes

389   available, a thorough analysis of this level of variation will allow us to shed light also on the evolutionary

390   processes that followed settlement on the island system and the achievement of reciprocal isolation, and so

391   will contribute to the ongoing debate concerning the relevance of intraisland microevolutionary processes.

392

393

References

395   Alvarez, W., 1972. Rotation of the Corsica-Sardinia microplate. Nature Physical. *Science* **235**:103-105.

396   Baccetti, B., 1964. Considerazioni sulla costituzione e l'origine della fauna di Sardegna. *Archivio*

397   *Botanico e Biogeografico Italiano* **40:**217-283.

398   Baccetti, B., 1980. Biogeografia sarda venti anni dopo. *Biogeographia* **8**:859-870.

399   Ball, S.L., Hebert, P.D.N., 2005. Biological identifications of mayflies (Ephemeroptera) using DNA

400   barcondes. *Journal of the North American Benthological Society* **24:**508–524.

401   Ballaux F., 2010. The worm in the fruit of the mitochondrial DNA tree. *Heredity* **104:**419-420.

402   Bauernfeind, E., Soldan, T., 2012. The Mayflies of Europe (Ephemeroptera). Apollo Books, Ollerup,

403   Denmark.

404   Bauzà-Ribot, M.M., Jaume, D., Fornos, J.J., Juan, C., Pons, J., 2011. Islands beneath islands:

405   phylogeography of a groundwater amphipod crustacean in the Balearic archipelago. *BMC Evolutionary*

406   *Biology* **11**:211.

407   Bellon, H., Coulon, C., Edel, J.B., 1977. Le déplacement de la Sardaigne. Synthèse des données

408   géochronologiques, magmatiques et paléomagnétiques. *Bulletin de la Société Géologique de France*

409   **7:**825-831.

410   Bisconti, R., Canestrelli, D., Nascetti, G., 2011a. Multiple lines of evidence for demographic and range

411   expansion of a temperate species *Hyla sarda*. during the last glaciation. *Molecular Ecology* **20:**5313-

412   5327.

413   Bisconti, R., Canestrelli, D., Nascetti, G., 2011b. Genetic diversity and evolutionary history of the

414   Tyrrhenian treefrog *Hyla sarda* (Anura: Hylidae): adding pieces to the puzzle of Corsica–Sardinia biota.

415   *Biological Journal of the Linnean Society* **103:**159-167.

416   Bisconti, R., Canestrelli, D., Nascetti, G., 2013a. Has living on islands been so simple? Insights from the

417   insular endemic frog *Discoglossus montalentii*. *PloS ONE* **8:**e55735.

418   Bisconti, R., Canestrelli, D., Salvi, D., Nascetti, G., 2013b. A geographic mosaic of evolutionary lineages

419   within the insular endemic newt *Euproctus montanus*. *Molecular Ecology* **22:**143-156.

420   Bonin B., Chotin, P., Giret, A., Orsini, J.B., 1979. Etude du bloc corso-sarde sur documents satellites: le

421   probleme des mouvements differentiels entre les deux iles. *Revue de Geologie Dynamique et De*

422   *Geographie Physique* **21**:147-154.

423   Brandano, M., Policicchio, G., 2012. Strontium stratigraphy of the Burdigalian transgression in the Western

424   Mediterranean. *Lethaia* **45:**315-328.

425   Brittain, J.E., 1982. Biology of Mayflies. *Annual Review of Entomology* **27**:119-197.

426   Brower, A.V.Z., 1994. Rapid morphological radiation and convergence among races of the

427   butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA Evolution. *Proceedings of the*

428   *National Academy of Sciences* **91:**6491-6495.

429   Bruen, T.C,, Hervé, P., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of

430   recombination. *Genetics* **172:**2665-2681.

431   Corander, J., Siren, J., Arjas, E., 2008. Bayesian spatial modelling of genetic population structure.

432   *Computational Statistics* **23**:111–129.

433   Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

434   Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and

435   parallel computing. *Nature Methods* **98:**772.

436   Emerson, B.C., Forgie, S.U., Goodacre, S., Oromì, P., 2006. Testing phylogeographic predictions on an

437   active volcanic island: *Brachyderes rugatus* (Coleoptera: Curculionidae) on La Palma Canary Islands.

438   *Molecular Ecology* **15**:449-458.

439   Falchi, A., Paolini, J., Desjobert, J.M., Melis, A., Costa, J., Varesi, L., 2009. Phylogeography of *Cistus*

440   *creticus* L. on Corsica and Sardinia inferred by the TRNL-F and RPL32-TRNL sequences of cpDNA.

441   *Molecular Phylogenetics and Evolution* **52:**538-543.

442   Gattaccecca, J., Deino, A., Rizzo R., Jones D.S., Henry, B., Beaudoin, B., Vadeboin, F., 2007. Miocene

443   rotation of Sardinia: new paleomagnetic and geochronological constraints and geodynamic implications.

444   Earth Planet. *Science Letters Journal* **258**:359-377.

445   Gentile, G., Campanaro, A., Carosi, M., Sbordoni, V., Argano, R., 2010. Phylogeography of *Helleria*

446   *brevicornis* Ebner 1868 (Crustacea, Oniscidea): Old and recent differentiations of an ancient lineage.

447   *Molecular Phylogenetics and. Evolution* **54:**640–646.

17

448 Goldberg, J., Trewick, S.A., Paterson, A.M., 2008. Evolution of New Zealand's terrestrial fauna: a review of

449   molecular evidence. *Philosophical Transactions of the Royal Society B* **363**:3319-3334.

450 Grandcolas, P., Murienne, J., Robillard, T., Desutter-Grandcolas, L., Jourdan, H., Guilbert, E., Deharveng,

451   L., 2008. New Caledonia: a very old Darwinian island? *Philosophical Transactions of the Royal Society B*

452   **363**:3309-3317.

453 Grill, A., Casula, P., Lecis, R., Menken, S., 2007. Endemism in Sardinia, in: Weiss, S., Ferrand, N. (Eds.),

454   Phylogeography of Southern European Refugia. CIBIO, University of Porto, Portugal, pp. 273- 296.

455 Guindon, S., Lethiec, F., Duroux, P., Gascuel, O. 2004. PHYML Online—a web server for fast maximum

456   likelihood-based phylogenetic inference 33 suppl 2.:W557-W559.

457 Hausdorf, B., 2011. Progress toward a general species concept. *Evolution* **65:**923-31.

458 Heads, M., 2009. Inferring biogeographic history from molecular phylogenies. *Biological Journal of the*

459   *Linnean Society* **98**:757–774.

460 Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Molecular*

461   *Biology and Evolution* **27**:570-580.

462 Holland, B.S., Hadfield, M.G., 2002. Islands within an island: phylogeography and conservation genetics of

463   the endangered Hawaiian tree snail *Achatinella mustelina*. *Molecular Ecolology* **11**:365-376.

464 Hooker, J. J., Collinson, M. E. & Sille, N. P., (2004). Eocene-Oligocene mammalian faunal turnover in the

465   Hampshire Basin, UK: calibration to the global time scale and the major cooling event. *Journal of the*

466   *Geological Society* **161**:161–172.

467 Hsü, K.J., Ryan, W.B.F., Cita, M.B., 1973. Late Miocene desiccation of the Mediterranean. *Nature* **242**:240-

468   244.

469 Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. *Evolution*

470   **56:**1557-1565.

471 Huson, D.H., Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular*

472   *Biology and Evolution* **23:**254-267.

473 Ketmaier ,V., Manganelli, G., Tiedemann, R., Giusti, F., 2010. Peri-Tyrrhenian Phylogeography in the Land

474   Snail *Solatopupa guidoni* (Pulmonata). *Malacologia* **52**:81-96.

475   Ketmaier,V., Caccone, A., 2013. Twenty years of molecular biogeography in the Wet Mediterranean Islands

476   of Corsica and Sardinia: lessons learnt and future prospects, in: Silva-Opps, M., (Eds.), Current Progress

477   in Biological Research. InTech, Croatia, pp. 71–93.

478   Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S. 1999. Chronology, causes and progression of

479   the Messinian salinity crisis. *Nature* **400:**652-655.

480   Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning

481   schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**:1695-

482   1701.

483   Li, C., Riethoven, J.J.M., Ma, L., 2010. Exon-primed intron-crossing EPIC. markers for non-model teleost

484   fishes. *BMC Evolutionary Biology* **10**:90.

485   Liu, Z.H., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S.R., Leckie, M.R.,

486   Pearson, A., (2009). Global cooling during the Eocene- Oligocene climate transition. *Science* **323**:1187–

487   1190.

488   Lomolino, M.K., Riddle, B.R., Whittaker, R.J., Brown, J.H., 2010. Biogeography Fourth Edition. Sinauer

489   Associates Inc. Massachusetts, USA.

490   Lucentini. L., Rebora. M., Puletti. M.E., Gigliarelli. L., Fontaneto. D., Gaino. E., Panara, F., 2011.

491   Geographical and seasonal evidence of cryptic diversity in the *Baetis rhodani* complex (Ephemeroptera,

492   Baetidae) revealed by means of DNA taxonomy. *Hydrobiologia* **673**:215–228.

493   Luzzatto, M., Palestrini, C., D'entrèves, P.P., 2000. Hologenesis: the last and lost theory of evolutionary

494   change. *Italian Journal of Zoology* **67**:129-138.

495   Monaghan, M.T., Gattolliat, J.L, Sartori, M., Elouard, JM., James, H., Derleth, P., Glaizot, O., de Moor, F.,

496   Vogler, A.P., 2005. Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera,

497   Baetidae) of Madagascar. *Proceedings of the Royal Society of London A* **272**:1829–1836.

498   Monterosso, B., 1935. Lo stato attuale del problema zoogeografico sardo. Atti XII

499   Congresso Geografico italiano, Cagliari, 3-31.

500   Monti, R., 1915. La fisionomia faunistica della Sardegna e le nuove idee circa le origini e la distribuzione

501   geografica delle specie. *Natura* **6**:109-128.

502   Müller-Liebenau I., 1969. Revision der europäischen Arten Gattung *Baetis* Leach, 1815 (Insecta,

503   Ephemeroptera). *Gewässer und Abwässer* **48-49**:1-214.

504   Murria, C., Morante, M., Rieradevall, M., Carles, R., Narcis, P., 2014. Genetic diversity and species richness

505   patterns in Baetidae (Ephemeroptera) in the Montseny Mountain range North-East Iberian Peninsula.

506   *Limnetica* **33**:313-326.

507   Nascetti, G., Cimmaruta, R., Lanza, B., Bullini, L., 1996. Molecular taxonomy of European plethodontid

508   Salamanders (Genus Hydromantes). *Journal of Herpetology* **30**:161-183.

509   Pereira-da-Conceicoa, L.L., Price, B.W., Barber-James, H.M., Barker, N.P., de Moor, F.C., Villet, M.H.,

510   2012. Cryptic variation in an ecological indicator organism: mitochondrial and nuclear DNA sequence

511   data confirm distinct lineages of *Baetis harrisoni* Barnard (Ephemeroptera: Baetidae) in southern Africa.

512   *BMC Evolutionary Biology* **12**:26.

513   Picet, F.J., 1843. Histoire naturelle générale et particulière des Insectes Névroptères. Famille des

514   Ephémérines. Chez J Kessmann & A Cherbuliz Eds., Geneva. Texte: pp. 1-1300.

515   Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from

516   http://beast.bio.ed.ac.uk/Tracer.

517   Ronquist, F., Teslenko, M., Van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L.,

518   Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and

519   model choice across a large model space. *Systematic Biology* **61**:539-542.

520   Rutschmann, S., Gattolliat, J.L., Hughes, S.J., Baez, M., Sartori, M., Monaghan, M.T., 2014. Evolution and

521   island endemism of morphologically cryptic *Baetis* and *Cloeon* species (Ephemeroptera, Baetidae) on the

522   Canary Islands and Madeira. *Freshwater Biology* doi:10.1111/fwb.12450.

523   Salvi, D., Capula, M., Bombi, P., Bologna, M.A., 2009. Genetic variation and its evolutionary implications

524   in a Mediterranean Island endemic lizard. *Biological Journal of the Linnean Society* **98**:661-676.

525   Salvi, D., Harris, D. J., Bombi, P., Carretero, M.A., Bologna, M.A., 2010. Mitochondrial phylogeography of

526   the Bedriaga's rock lizard, *Archaeolacerta bedriagae* (Reptilia: Lacertidae) endemic to Corsica and

527   Sardinia. *Molecular Phylogenetics and Evolution* **56**:690-697.

bioRxiv preprint doi: https://doi.org/10.1101/056556; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

528 Salvi, D., Schembri, P.J., Sciberras, A., Harris, D.J., 2014. Evolutionary history of the Maltese wall lizard

529   *Podarcis filfolensis*: Insights on the 'Expansion-Contraction' model of Pleistocene biogeography.

530   *Molecular Ecology* **23**:1167-1187.

531 Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, 2nd Edition. Cold

532   Spring Harbor Laboratory Press.

533 Shackleton, J.C., Van Andel, T.H., Runnels, C.N. 1984., Coastal paleogeography of the central and western

534   Meditterranean during the last 125,000 years and its archeological implications. *Journal of Field*

535   *Archaeology* **11**:307-314.

536 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and

537   phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain

538   reaction primers. *Annals of the Entomological Society of America* **87**:651–701.

539 Stephens, M., Donnelly, P., 2003. A comparison of Bayesian methods for haplotype reconstruction from

540   population genotype data. *The American Journal of Human Genetics* **73**:1162-1169.

541 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S., 2011. MEGA5: Molecular

542   Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum

543   Parsimony Methods. *Molecular Biology and Evolution* **28**:2731-2739.

544 Thomas, A.G.B., Soldan, T., 1987. *Baetis ingridae* n. sp., Ephemeroptere nouveau de Corse (Baetidae).

545   *Annales de Limnologie - International Journal of Limnology* **23:** 23–26.

546 Thompson, J.D., 2005. Plant evolution in the Mediterranean. Oxford University Press. New York.

547 Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows

548   interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic*

549   *Acids Research* **25**:4876-4882.

550 Thorpe, R.S., Malhotra, A., 1998. Molecular and morphological evolution within small islands. In. Evolution

551   on islands Eds. Grant P.R. pp 67-82. Oxford: Oxford University UK.

552 Tomassetti, L., Bosellini, F.R., Brandano, M., 2013. Growth and demise of a Burdigalian coral

553   bioconstruction on a granite rocky substrate (Bonifacio Basin, southeastern Corsica). *Facies* **59**:703-716.

554 Van Andel, T.H., Shackleton, J.C., 1982. Late Paleolithic and Mesolithic coaslines of Greece and the

555   Aegean. *Journal of Field Archaeology* **9**:445-454.

bioRxiv preprint doi: https://doi.org/10.1101/056556; this version posted October 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

556   Villacorta, C., Jaume, D., Oromì, P., Juan, C., 2008. Under the volcano: phylogeography and evolution of the

557   cave-dwelling *Palmorchestia hypogaea* (Amphipoda, Crustacea) at La Palma Canary Islands*. BMC*

558   *Biology* **6**:doi.10.1186.

559   Vogiatzakis, I.N., Pungetti, G., Mannion, A.M., 2008. Mediterranean island landscapes: natural and cultural

560   approaches. Springer Publishing. New York.

561   Wallis, G.P., Trewick, S.A., 2009. New Zealand phylogeography: evolution on a small continent. *Molecular*

562   *Ecology* **18**:3548– 3580.

563   Watson, D.M., 2009. Continental islands, in: Gillespie, R.G. Clague, D.A. (Eds.) , Encyclopedia of Islands,

564   University of California Press. Berkeley and Los Angeles, California, pp.180-187.

565   Whittaker, R.J., Fernández-Palacios, J.M., 2007. *Island Biogeography: ecology, evolution, and conservation*.

566   Oxford University Press. Oxford, UK.

567   Williams, H.C., Ormerod, S.J., Bruford M.W., 2006. Molecular systematics and phylogeography of the

568   cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and*

569   *Evolution* **40**:370–382.

570

571

578

579   **Data accessibility**

580   Genbank accession numbers: KX707651-KX707786.

581

582   **Conflict of interest**

583   The authors have no conflict of interests to declare.

584

585   **Tables**

586   **Table 1.** Geographical location and sample size of the 28 population samples of the *Baetis rhodani* species

587   group used in this study. The geographic distribution of the haplotypes found and their abundance among

588   populations is also given for each gene marker analysed.

	Site	Latitude N	Longitude E	n	Haplotypes (n)		
					mtDNA	PEP	RYA
Sardinia	1	40° 46.235'	9° 32.380'	2	H22(2)	P5(2); P7(2)	R34(2); R38(2)
	2	40° 22.751'	9° 26.065'	2	H5(2)	P11(4)	R2(2); R5(2)
	3	40° 2.752'	9° 31.012'	2	H23(2)	P5(2); P25(2)	R41(2); R43(2)
	4	39° 55.590'	9° 38.244'	2	H26(2)	P7(4)	R34(2); R35(2)
	5	39° 30.284'	9° 8.095'	4	H6(4)	P10(6); P11(2)	R2(6): R4(2)
	6	39° 23.525'	8° 40.131'	6	H24 (2); H25(2); H31(2)	P5(6); P7(6)	R34(8); R35(2); R36(2)
	7	39° 49.494'	9° 12.106'	4	H6(2); H27(2);	P5(4); P1(4)	R2(4); R34(2); R43(2)
	8	40° 8.863'	8° 32.295'	2	H33(2)	P5(4)	R34(2); R44(2)
	9	40° 24.421'	8° 37.614'	6	H6(6)	P10(6); P11(6)	R2(12)
	10	40° 31.236'	8° 52.058'	4	H7(2); H8(2)	P10(2); P1 (6)	R2(4); R6(4)
Corsica	11	41° 37.541'	9° 4.962'	2	H15(2);	P3(2); P4(2)	R28(2); R29(2)
	12	41° 39.831'	9° 0.889'	8	H15(2); H17(2); H28(2); H30(2)	P3(2); P5(8); P26(2); P22 (4)	R16(2); R17(2); R31(2); R32(2); R34(4); R40(4)
	13	42° 10.170'	8° 49.200'	2	H11(2)	P16(4)	R2(2); R8(2)
	14	42° 21.865'	8° 48.113'	2	H19(2)	P3(4)	R16(2); R26(2)
	15	42° 29.176'	8° 48.176'	8	H15(4); H18(2); H27(2)	P3(6); P5(6); P8(4)	R16(4); R23(2); R25(2); R30(4); R34(4)
	16	42° 28.076'	9° 6.385'	2	H15(2)	P23(4)	R16(4)
	17	42° 36.138'	9° 8.313'	6	H10(2); H12(2); H13(2)	P12(2); P13(8); P14(2)	R2(6); R7(6);
	18	42° 34.182'	9° 18.256'	6	H29(6)	P5(12)	R34(10); R40(2)
	19	42° 35.515'	9° 21.799'	2	H32(2)	P6(4)	R34(2); R39(2)
	20	42° 26.069'	9° 13.433'	6	H15(2); H29(4)	P3(4); P5(8)	R16(2); R33(2); R34(6); R42(2)
	21	42° 16.504'	9° 6.438'	2	H29(2)	P5(2); P7(2)	R34(2); R37(2)
	22	42° 6.153'	9° 14.679'	2	H15(2)	P24(4)	R19(2); R24(2)

24

	23	41° 49.292'	9° 15.599'	4	H15(2); H16(2)	P3(8)	R16(2); R18(2); R22(2); R27(2);
Elba	24	42° 47.064'	10° 9.962'	6	H20(2); H21(4)	P3(12)	R21(12)
	25	42° 44.298'	10° 10.604'	8	H9(6); H20(2)	P3(4); P13(8); P15(2); P16(2)	R2(6); R3(4); R7(2);R20(2); R21(2)
Italy	26	42° 38.672'	11° 44.156'	6	H1(2); H2(2); H3(2)	P9(2); P17(4); P18(2); P20(2); P21(2)	R10(2); R11 (2); R12 (2); R13(2); R14(2); R15(2)
	27	42° 15.212'	12° 4.478'	2	H14(2)	P1(2); P2(2)	R1(4)
	28	42° 23.927'	13° 2.641'	2	H4(2)	P19(4)	R9(4)

589

590

591

25

592   **Table 2.** Mean pairwise uncorrected sequence divergence among (below diagonal) and within

593   (diagonal) the five main haplotype groups identified in the concatenated mtDNA dataset. Standard

594   errors of the between-group divergence estimates are given above the diagonal. Haplotype groups

595   are encoded as in Figure 1.

	A	B	C	D	E
A	**0.011**	0.010	0.013	0.014	0.014
B	0.135	**0.008**	0.013	0.014	0.015
C	0.263	0.238	**0.006**	0.012	0.012
D	0.280	0.256	0.166	**0.009**	0.012
E	0.264	0.249	0.194	0.190	**n/c**

596

597

26
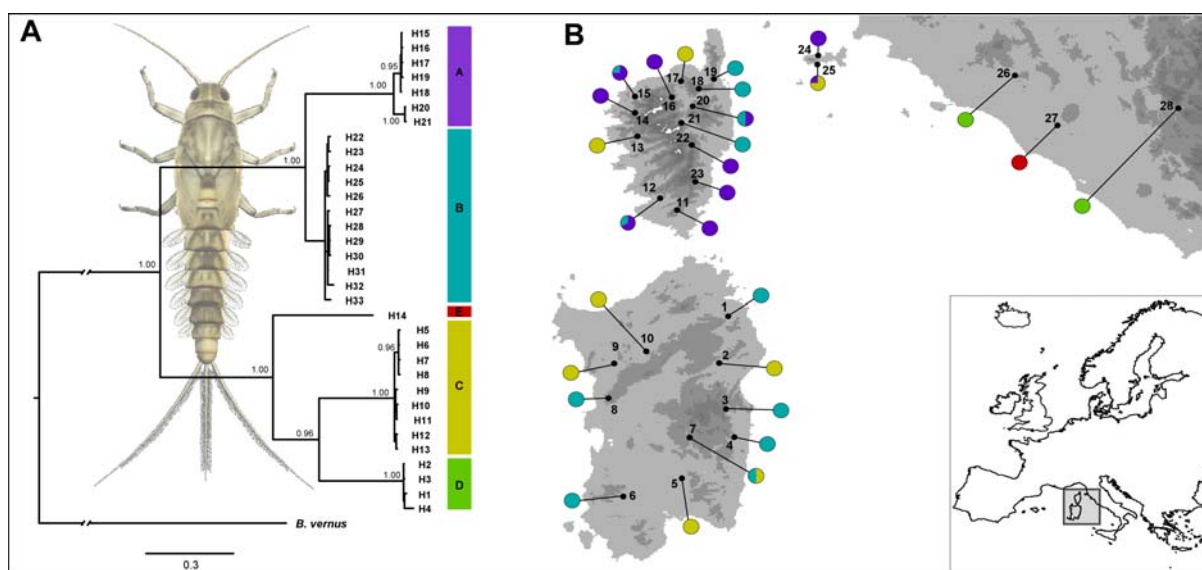
598   **Figures**

599

600

601

602

603   **Figure 1. (A)** Bayesian phylogenetic tree based on the 33 mtDNA haplotypes found in *Baetis rhodani*

604   populations from the Tyrrhenian islands and neighbouring areas, estimated using MRBAYES. Posterior

605   probabilities are shown at the nodes when ≥0.90. **(B)** Geographic locations of the 28 *Baetis rhodani*

606   populations sampled, numbered as in Table 1. Pie-diagrams show the geographic location and frequency

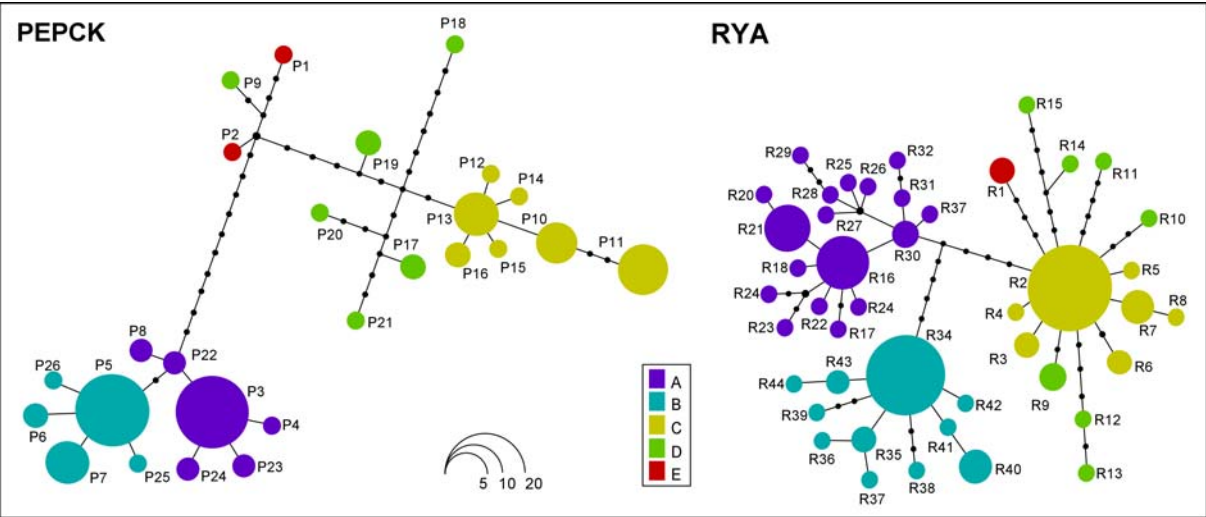607   within populations of the main mtDNA haplogroups found.

608

Figure 2. Phylogeographic networks generated with HAPLOTYPEVIEWER, based on the Bayesian phylogenetic trees estimated for each nDNA gene fragment using MRBAYES. Circle size is proportional to the frequency of the corresponding haplotype across the dataset. Each individual was given the same colour as the respective mtDNA haplogroup, for comparative purposes.
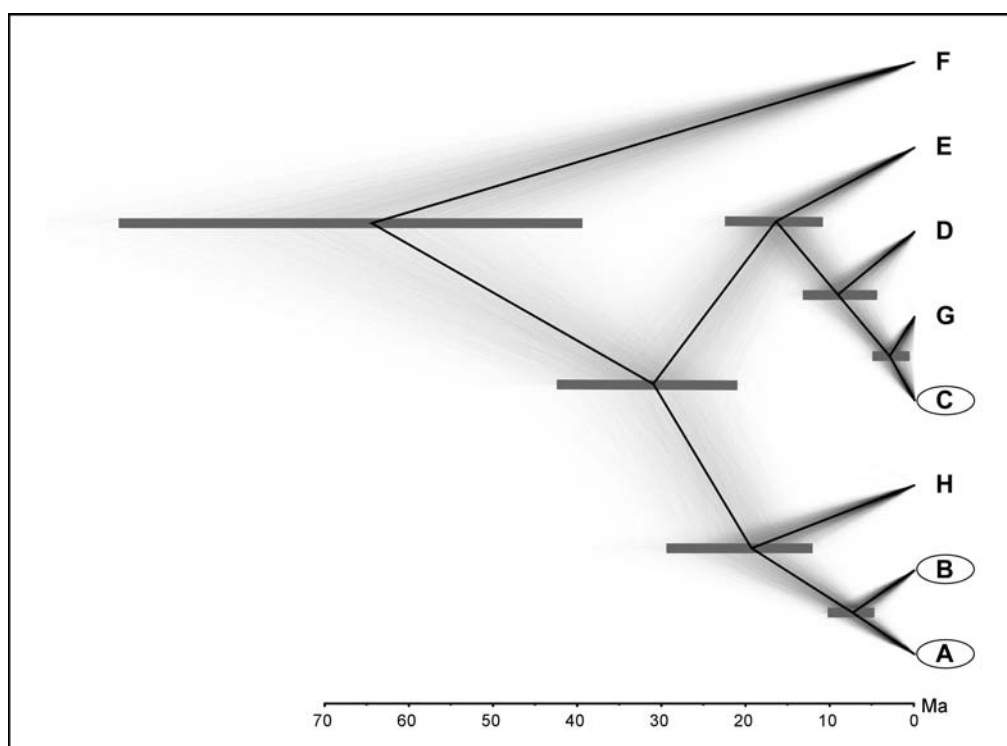
Figure 3. Maximum clade credibility multilocus species-tree (black line) superimposed on a cloudogram of the entire posterior distribution of species-trees obtained with *BEAST. Horizontal grey bars are 95% highest posterior densities of the estimated node ages. X-axis is in units of million years (Ma). All posterior probabilities were ≥0.99. The analysis was run with 7 terminal taxa, 5 resulting from the Bayesian analysis of the concatenated mtDNA data, plus 2 groups retrieved from the literature (clades G and H) which turned out to be the closest relatives of our insular ingroups (clades A, B, C) during previous analysis of the CO1 dataset.

625 **Supporting information**

626

627 **Table S1.** PCR primers and annealing temperatures used to amplify the two mitochondrial and the two

628 nuclear DNA fragments used in this study.

629

630

Gene	Primer name	Primer sequence 5'-3'	Annealing T. (°C)	Source
CO1	C1-N-2191	CCCGGTAAAATTAAAATATAAACT TC	56	Simon et al. 1994
	C1-J-1718	GGAGGATTTGGAAATTGATTGTTCC		
ND1	ND1F	TAAAGTTAGCAGGTTCATACCC	58	This study
	ND1R	CACCTATATTTGTACTTTGAAGG		
PEPCK	PepFb12	GGAACTTCAAACAGCACCAAT	53	Pereira-da-Conceicoa et al. 2012
	PepRb45	ACCTTGTGTTCTGCAGCT		
RYA	RYAf	GGCGAGAAGAACATGCACG	60	This study
	RYAr	CCAGTAGAAGTCCATGATGG		

631   **Figure S1.** Bayesian phylogenetic tree, estimated using MRBAYES, based on the CO1 haplotypes found in

632   this study, and 2 representatives of each main clade previously identified by Williams et al., (2006) Lucentini

633   et al., (2011), and Murria et al., (2014). Posterior probabilities are shown at the nodes when ≥0.90.
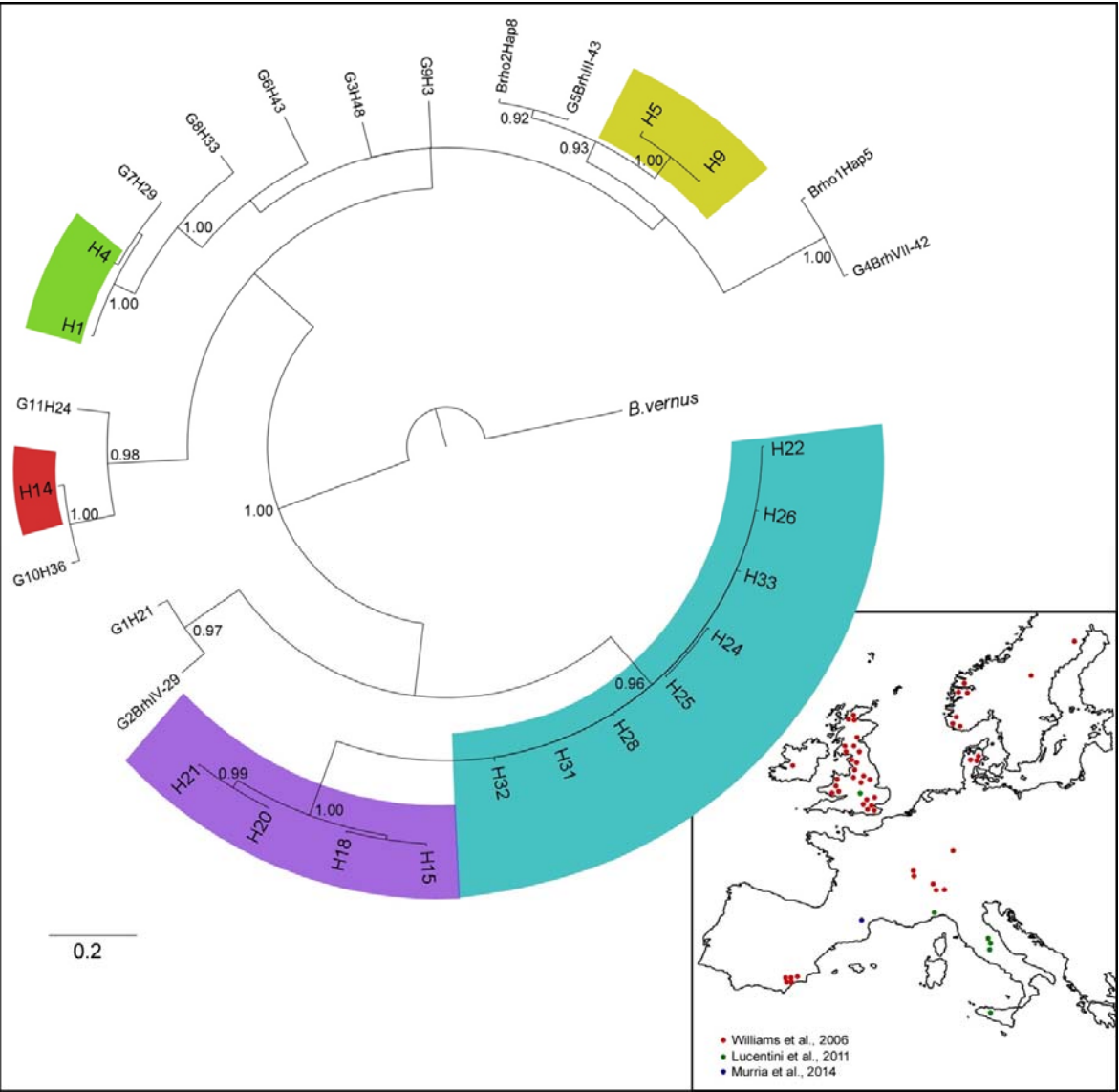
634

635

636

637

638   **Figure S2.** Maximum clade credibility multilocus species-trees obtained with *BEAST. Horizontal grey bars

639   are 95% highest posterior densities of the estimated node ages. X-axis is in units of million years. All

640   posterior probabilities were $\geq$0.99. (A) Analysis run with 5 terminal taxa, following results based on the

641   concatenated mtDNA dataset (see main text Figure 1) (B) Analysis run with 4 terminal taxa, following

642   results with BAPS based on nDNA data.

643

644

645

646