University of Plymouth

PEARL https://pearl.plymouth.ac.uk

Faculty of Science and Engineering School of Biological and Marine Sciences

2018-08

Introgression in native populations of Apis mellifera mellifera L: implications for conservation

Ellis, JS

http://hdl.handle.net/10026.1/11543

10.1007/s10841-018-0067-7
Journal of Insect Conservation
Springer (part of Springer Nature)

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

1 **Introgression in native populations of *Apis mellifera mellifera* L: implications
2 for conservation**

3 Jonathan S. Ellis*[1,2], Gabriele Soland-Reckeweg[3], Victoria G. Buswell[1,2]; J. Vanessa
4 Huml[1,2], Andrew Brown[4], and Mairi E. Knight [1,2]
5
6 Affiliations and addresses
7 1. School of Biological and Marine Sciences, Plymouth University, Drake Circus,
8 Plymouth, PL4 8AA
9 2. Sustainable Earth Institute, Plymouth University, Drake Circus, Plymouth, PL4
10 8AA
11 3. Apigenix, Gaicht 19, 2513 Twann, Switzerland
12 4. B4, Newton Farm Metherell, Callington, Cornwall, PL17 8DQ
13
14 Email addresses and ORCID identifiers:
15 Corresponding author: Jonathan Ellis, School of Biological and Marine Sciences,
16 Plymouth University, Drake Circus, Plymouth, PL4 8AA, tel: +44(0)1752585946
17 Jonathan Ellis: jonathan.s.ellis@plymouth.ac.uk; orcid.org/0000-0002-6655-4685
18 Other authors:
19 Gabriele Soland: info@apigenix.com
20 Victoria Buswell: victoria.buswell@plymouth.ac.uk
21 J. Vanessa Huml: vanessa.huml@plymouth.ac.uk
22 Andrew Brown: andrew@aglbrown.co.uk
23 Mairi Knight: mairi.knight@plymouth.ac.uk; orcid.org/0000-0001-9388-4850
24
25 **Compliance with ethical standards**

No human participants were involved in the work. Ethical standards were followed for the sampling of honey bee drones by the removal of antennae by beekeepers. The article has not been submitted elsewhere. All authors consent to submission of this work.

Funding

The work was funded by the organisation 'B4: bringing back black bees'. Grant number: N/A

Conflict of interest

The authors are aware of no potential conflict of interest (financial or non-financial).

Abstract

Hybridisation and introgression can have negative impacts on regional biodiversity through the potential erosion of locally adapted lineages. The honey bee (*Apis mellifera* L.) occurs in twenty-seven subspecies across Europe, is an extremely economically important insect, yet threatened by multifarious impacts. Transhumance of the most commercially appealing varieties threatens native honey bee diversity by introgression and subsequent loss of locally adapted traits, or even by complete removal of some subspecies from parts of the range. Here levels of admixture and introgression are examined in UK honey bees suspected to be from hives of the dark European honey bee (*Apis mellifera mellifera*). Microsatellite DNA and STRUCTURE analyses reveal that the studied populations are generally admixed, and discriminant analysis of principal components shows them to be intermediate between *A. m. mellifera* and *A. m. carnica* populations. However, examining mitochondrial haplotype data (COI-COII intergenic spacer region) and nuclear DNA reveal that some hives are relatively pure (from four to fifteen hives, depending on the Q-value threshold). Genetic diversity is relatively high in comparison with other European populations. Implications for conservation and management are discussed.

Keywords:
Introgression, honey bee, subspecies, microsatellite, mitochondrial DNA, conservation

58 **Introduction**

59 Hybridisation is widely defined as interbreeding between individuals from distinct
60 lineages, subspecies or species. Globally, rates of hybridisation are increasing due to
61 movement of organisms by humans and habitat alteration (Allendorf et al. 2001).
62 Introgression is the incorporation of genetic material from one lineage into the
63 background of another (Anderson 1949) and can occur following hybridisation and
64 repeated back-crossing. Hybridisation and introgression are often considered to be
65 problematic for conservation because these processes can lead to the loss of
66 combinations of alleles that have resulted from long periods of adaptive evolution. This
67 can disrupt local adaptation, leading to outbreeding depression (Templeton et al. 1986)
68 and can even lead to genomic extinction (the loss of a lineage by introgression or
69 anthropogenic displacement (Epifanio and Philipp 2001; Allendorf et al. 2004)). Overall,
70 hybridisation and introgression may thus be considered to have a negative impact on
71 regional biodiversity (Allendorf et al. 2004). Conversely, hybridisation is sometimes
72 regarded as a positive management option because it augments genetic diversity,
73 conserves evolutionary potential as a consequence, and sometimes the fitness of
74 admixed genotypes is increased (Hamilton and Miller 2016). From this point of view,
75 hybridisation and introgression can increase the overall capacity for adaptation, which
76 is important in a changing environment (Hamilton and Miller 2016).

77 Honey bees provide an interesting study system to investigate issues arising
78 from hybridisation and introgression. They are amongst the most important insect
79 pollinators, especially for the pollination of crop monocultures (Delaplane and Mayer
80 2000; van Engelsdorp and Meixner 2010). Insect pollination itself has been estimated
81 as worth €153 billion annually (Gallai et al. 2009) and worth €505 million annually in
82 the UK (POST 2010). The value of honey bee pollination in the USA alone has been

83 estimated at $14.6 billion (Morse and Calderone 2000). Despite this importance,
84 honey bees face various threats. For example, in Europe between 2008-2012 average
85 winter losses by country varied from 7% to 30% (OPERA 2013). These unexplained
86 winter losses of honey bees may be attributable to interacting underlying factors such
87 as the spread of diseases and parasites (*Varroa destructor, Nosema* spp., bacterial
88 pathogens, deformed wing virus and acute bee paralysis virus), autumn colony
89 strength and winter severity (Genersch 2010; Highfield et al. 2009; Lee et al. 2015;
90 Meixner et al. 2010; OPERA 2013; van Engelsdorp et al. 2012). The increasing use of
91 pesticides and the role of neonicotinoids in particular is another potentially important
92 factor contributing to declines, and is a subject of ongoing debate (reviewed in the
93 OPERA report 2013).

94 In addition to these issues, many beekeepers are now concerned about
95 the potential loss of locally adapted forms that occur in subspecies, regional varieties
96 and ecotypes (Meixner et al. 2010). Ten out of twenty-seven subspecies of honey bee
97 are present in Europe (Meixner et al. 2013). Early morphometric analyses classified
98 these into M, A, C and O lineages, which owe their origin to the glacial history of
99 Europe (Ruttner 1988). The M lineage occurs in the west Mediterranean area and
100 north-western Europe and includes *Apis mellifera mellifera* and *A. m. iberiensis.* The
101 C lineage occurs in south-eastern Europe and includes the subspecies *A. m. ligustica,*
102 *A. m. carnica, A. m. macedonica, A. m. cecropia, A. m. cypria* and *A. m. adami* (Ruttner
103 1988; Meixner et al. 2013). The O lineage occurs in the near East and western Asia
104 and includes the subspecies *A. m. caucasia, A. m. anatolica, A. m. syriaca, A. m.*
105 *meda, A. m. armeniaca, A. m. jemenitica* and *A. m. pomonella* (Meixner et al. 2013;
106 Ruttner 1988). The A lineage represents a further seven African subspecies (Ruttner

107 1988; Meixner et al. 2013;). There is also a Y lineage in Ethiopia (Franck et al. 2001)
108 and a Z lineage in Lybia (Alburaki et al. 2013).
109

110 Maintaining the diversity distributed across these subspecies is considered necessary
111 to ensure future resilience of honey bees to environmental change (Pinto et al. 2014).
112 Yet, transhumance of commercial varieties (by importation of queens and movement
113 of hives) that are favoured for characteristics that make them amenable to beekeeping,
114 may cause 'genetic pollution' of these varieties by introgression (Garnery et al. 1998a).
115 The subspecies most favoured commercially are *A. m. ligustica* and *A. m. carnica* (van
116 Engelsdorp and Meixner 2010). In some areas, importation of these subspecies has
117 seen complete replacement of local subspecies, e.g. the replacement of *A. m.*
118 *mellifera* by *A. m. carnica* in Germany (Kauhausen-Keller and Keller 1994; Maul and
119 Hähnle 1994).

120 As part of the effort to conserve native bee diversity, there is a movement to
121 protect the dark European honey bee, *A. m. mellifera* (Meixner et al. 2010, 2013). The
122 range of this subspecies has been much reduced (see Meixner et al. 2010) and for the
123 purpose of its conservation, the *Societas Internationalis pro Conservatione Apis*
124 *melliferae melliferae* was established in 1995 (Pinto et al. 2014). Dark European honey
125 bees can occur in ecotypes with distinct colony population cycles (Louveaux 1966,
126 cited in Strange et al. 2007) that still persist today (Strange et al. 2007). Genetic
127 methods can identify these local varieties, although specific ecotypes within these
128 varieties may not be clearly delineated (Strange et al. 2008; Soland-Reckeweg et al.
129 2009). In general, local-origin colonies have been shown to have longer colony
130 survivorship than non-local colonies (Büchler et al. 2014).

131 The identification of native honey bee subspecies and varieties is aided by the
132 study of mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) (Cornuet et al. 1991;
133 Cornuet and Garnery 1991; De la Rúa et al. 1998; Garnery et al. 1998a, 1998b; Muñoz
134 et al. 2017). Mitochondrial DNA is ideal as a colony-level marker (Garnery et al. 1998b)
135 as all individuals in a colony share the same haplotype since mtDNA is maternally
136 inherited. Cornuet et al. (1991) outlined the structure of the mitochondrial COI-COII
137 intergenic spacer in honey bees. This is based on copy number variation and
138 sequence variation of 'P' and 'Q' sequences in the intergenic spacer region between
139 these genes (Cornuet et al. 1991). Haplotypes are named similarly to the
140 morphometric lineages, but there is not complete consistency between the systems,
141 for example *A. m. iberiensis* is in the M morphometric lineage, but can have M and A
142 mtDNA haplotypes (Meixner et al. 2013). Dark European honey bees (*A. m. mellifera*)
143 are in the M morphopmetric lineage and have M haplotypes (Meixner et al. 2013). A
144 comprehensive review and description of COI-COII haplotypes in *A. m. mellifera* has
145 been published by Rortais et al. (2011). This diversity may be interrogated by the use
146 of restriction fragment length polymorphism analyses known as the *Dra*I test,
147 (validated by Garnery et al. 1993). Nuclear markers like microsatellites are also useful
148 as they may demonstrate different levels of introgression to those inferred from mtDNA
149 (Ballard and Whitlock 2004; Garnery et al. 1998a). For example, Garnery et al. (1998a)
150 observed asymmetrical levels of introgression for mtDNA versus nDNA markers in
151 parts of France and the Iberian peninsula. Mitochondrial DNA is most commonly
152 inherited uniparentally and generally does not undergo recombination (Ballard and
153 Whitlock 2004). In haplodiploid and diploid taxa the mtDNA effective population size
154 is usually smaller than for nDNA, and mtDNA also represents only a small proportion
155 of the whole genome (Ballard and Whitlock 2004). Consequently it is prudent to utilise

156 information from both DNA sources when assessing the history of a species using
157 molecular data.

158 Previous studies have examined rates of introgression in *A. m. mellifera*.
159 Soland-Reckeweg et al. (2009) quantified introgression and hybridization between M
160 and C lineages of honey bees in Switzerland. Considerable hybridization was
161 observed, even in colonies managed for pure breeding by apiculturalists interested in
162 conservation (Soland-Reckeweg et al. 2009). Pinto et al. (2014) examined the integrity
163 of protected populations using single nucleotide polymorphisms (SNPs) and mtDNA.
164 Despite their protection, introgression was detected in these populations, although
165 introgression was higher in unprotected than protected colonies (Pinto et al. 2014).
166 Honey bees from England and Scotland were included in this analysis. Jensen et al.
167 (2005) also included English and Scottish samples in their earlier analysis of
168 introgression in north-west European populations of *A. m. mellifera*. Microsatellite data
169 and *Dra*I tests revealed varying levels of introgression, but also demonstrated the
170 persistence of this subspecies in northwestern Europe. More recently, Parejo et al.
171 (2016) examined introgression in Swiss and French populations of *A. m. mellifera*
172 using whole genome sequence information and were able to detect admixture as well
173 as population structuring by subspecies and geographic origin.

174 Here, local populations of *A. m. mellifera* from Cornwall in the South-West of
175 the UK are examined. As mentioned above, subspecies of honey bee, including the
176 dark European honey bee, may show evidence of local adaptation (Louveaux 1966;
177 Strange et al. 2007). Populations of dark European honey bee (*A. m. mellifera*) are
178 likely to have been native to the UK for at least 4000 years (Carreck 2008) and occur
179 in the South-West of the UK, but have been neglected in previous studies, which have
180 sampled elsewhere in the UK or continental Europe (Costa et al. 2012; Ilyasov et al.

181 2016; Jensen et al. 2005; Muñoz et al. 2015; Pinto et al. 2014). However, local
182 beekeepers believe that relict hives occur in the region and that these show local
183 adaptations including winter hardiness, a maritime brood cycle, longevity of workers
184 and queens, activity in cold weather, and possible hardiness against *Varroa* (see
185 http://www.b4project.co.uk/). These beekeepers have initiated the 'B4 project: bringing
186 back black bees' for beekeepers interested in conserving local diversity of the dark
187 bee, *A. m. mellifera* in this region. We emphasise that the focus of our study is at the
188 regional level because of a real need to identify introgression for the practical
189 conservation of dark European honey bees by beekeepers in the 'B4' organisation.
190 These beekeepers suspect their colonies to be dark European bees and have set up
191 a voluntary reserve in the area where only dark European hives are to be kept. It is
192 not possible to identify relatively pure hives or hybrid individuals confidently on the
193 basis of morphometric data, thus there is a practical conservation need on the part of
194 these beekeepers to accurately identify and know the state of introgression in their
195 hives. Our research therefore uses genetic techniques and modern analytical methods
196 to bridge a gap between scientific research and the practical conservation of insects,
197 an approach which is especially important for sound conservation practice.
198
199
200
201 **Materials and Methods**
202
203 ***Sampling***
204 Bees were sampled from forty-three hives across thirty-four apiaries managed by ten
205 beekeepers in Cornwall, England, during summer 2015 in the vicinity of Truro and to
206 the west of Plymouth. Colonies were chosen by the beekeepers where they suspected
207 an unhybridized dark bee, thus sampling aimed to detect remaining population

208 fragments of *A. m. mellifera*. Members of the B4 network were supplied with 5mL
209 sterile sample tubes and ~2 mL absolute ethanol. Queens were indirectly sampled
210 using a pool of antennae of 30 drones. DNA can be efficiently extracted from antennae
211 (Issa et al. 2013). Drone brood were sampled by removing the cell lid with a clean
212 sharp tool. Beekeepers were instructed to sample thirty individuals of the drone brood
213 that were quite well-developed with antennae. The right antenna of each of the thirty
214 drones was then removed using college pliers and placed in a 1.5mL centrifuge tube
215 in absolute ethanol. Samples were posted to Apigenix (Biel, Switzerland) for genetic
216 analysis. Pools of drones from each hive sampled were genotyped to establish the
217 queen genotype. DNA was isolated from the pools. In the authors' experience it is
218 better to ask beekeepers to supply drone antennae because it is easy to then use a
219 standard amount of tissue per individual when extracting the DNA. The use of larvae
220 gives variously sized tissue samples from the individuals sampled. Furthermore, the
221 use of drone antennae makes it more probable that worker-produced individuals have
222 been removed by this stage. This means the estimation of the queen genotype is not
223 'contaminated' by alleles from the patrilines that would be present if the worker
224 offspring were accidentally included. Regarding whether pools of 30 drones per hive
225 are sufficient to establish the queen's genotype at a given heterozygous locus, the
226 probability a haploid male has either one of the queen's alleles is 0.5, on average. The
227 probability of only detecting a single one of these alleles can therefore be modelled as
228 a binomial distribution where the probability of success is 0.5 and the number of trials
229 equals the number of males sampled, in this case 30. In this case, the probability of
230 all trials detecting a single allele at a given locus is $9.3x10^{-8}$. This assumes an equal
231 contribution to the DNA pool across males and the absence of null alleles. All mtDNA
232 sequencing and genotyping was conducted by Apigenix (Switzerland).

233

234 ***Investigation of admixture***

235 DNA was isolated from the drone samples using a Qiagen DNEasy Blood and Tissue
236 kit following the manufacturer's protocol. PCR amplificiation of 12 microsatellite loci
237 was performed in two multiplex reactions in a 10 µl reaction volume containing 2-10 ng
238 of genomic DNA, 5 µl HotStarTaq Master Mix, double distilled water, and 10 µM of
239 forward and reverse primers each. (Multiplex 1: FAM A43, FAM A273, FAM AC306,
240 FAM Ap33, ATTO565 Ap226, ATTO565 B24; multiplex 2: FAM A76; ATTO550 A28,
241 ATTO550 Ap289, ATTO532 A007, ATTO532 AP1, ATTO565 A29, Solignac et al.
242 2003). The following cycling protocol on a TC-412 programmable thermal controller
243 (Techne) was used: 40 cycles with 94°C for 30 s, 56°C for 90 s, and 72°C for 60 s.
244 Before the first cycle, a prolonged denaturation step (95°C for 15 min) was included
245 and the last cycle was followed by a 30 min extension at 72°C. Fragments were run
246 on a ABI3730 Prism Genetic Analyser (Applied Biosystems) using GeneScan TM-500
247 LIZ size standard. Fragments were scored using the software GeneMarker 3.0 (ABI).

248

249 Additional samples from Italy for *A. m. ligustica*, Austria and Slovenia for *A. m. carnica*
250 and Sweden, France, Norway, Switzerland and Ireland for *A. m. mellifera* were
251 included for testing admixture and introgression in Cornish bees (see Soland-
252 Reckeweg et al. 2009 for more information including a map of sampling locations).
253 These sample locations include *A. m. mellifera* conservation areas in Norway and
254 Sweden, and areas where least introgression is expected. Hybrids have been
255 previously removed from this reference dataset of genotypes (Soland-Reckeweg et al.
256 2009). Microsatellite genotyping was also conducted using a set of pre-typed
257 individuals of known genotypes to create an allele ladder across the size range of

258 alleles at each locus. This approach allows to confidently assign microsatellite
259 genotypes and avoid errors due to size shifts which can problematic, especially if
260 different genotyping equipment is used (e.g. Ellis et al. 2011)
261
262 After genotyping, MICROCHECKER (van Oosterhout et al. 2004) was used to
263 investigate the presence of null alleles and other common sources of genotyping error
264 (e.g. stutter). Samples were grouped by population in this analysis. Estimates of
265 linkage disequilibrium and departures from Hardy-Weinberg equilibrium were made in
266 Arlequin 3.5 (Excoffier and Lischer 2010). Again, samples were grouped by population
267 for these analyses. For pairwise tests of linkage disequilibrium the number of
268 permutations was 10,000. The selected significance level was P=0.05, but strict
269 Bonferroni corrections were applied to pairwise tests by population, thus the revised
270 level of significance was P = 0.0008 (there were 66 tests per population). For exact
271 tests of Hardy-Weinberg equilibrium the number of steps in the Markov chain was
272 1,000,000 and the number of dememorization steps was 100,000. Strict Bonferroni
273 corrections were again applied to tests done by population (adjusted P varied as some
274 loci were monomorphic in some populations, minimum adjusted P = 0.004). Some loci
275 were removed after these steps, prior to downstream analyses (see Results).
276 Standard measures of genetic diversity were estimated in Arlequin 3.5 (observed and
277 expected heterozygosity; Excoffier and Lischer 2010) and FSTAT (allelic richness,
278 Goudet 2001). Allelic richness was calculated across loci per population and was
279 based on a minimum sample size of ten individuals.
280
281 To investigate admixture, two complementary approaches were used, as has been
282 recommended (Janes et al. 2017). For the first approach, STRUCTURE (Pritchard et

283 al. 2000) was used. A burn-in period of 50,000 steps was used followed by 500,000
284 MCMC steps. *K* values of 1 to 12 were tested, with three iterations of each *K* value. A
285 correlated allele model (Falush et al. 2003) was applied. The admixture model was
286 used, but LOCPRIOR was not applied (LOCPRIOR is usually used where the
287 expected signal is too weak for standard structure models and makes use of location
288 information with each individual to assist clustering, Pritchard et al. 2010). Iterations
289 were examined for consistency (by examining similarity of alpha values and 'ln Prob.
290 of data' across the iterations). The best *K* value was investigated using the original
291 method recommended in STRUCTURE and using the Evanno et al. (2005) method.
292 The standard method infers the most probable value of *K* based on the 'log probability
293 of the data' (or where this value begins to reach a plateau). The Evanno et al. (2005)
294 method is based on the rate of change in values of 'log probability of data' for
295 successive values of *K*. *Structure Harvester* was used to generate the relevant plots
296 for inference of *K* (Earl 2012). Although these methods can be used to estimate the
297 'best' *K* value, multiple *K* values were interpreted as this is recommended (Janes et al.
298 2017). Barplots were produced using the online application STRUCTURE PLOT
299 (Ramasamy et al. 2014). The degree of introgression of sampled colonies was
300 investigated through inspection of mean Q values and their standard deviations (from
301 the three iterations of the analysis) for $K = 3$ (further explanation in the Results). A
302 population cluster which included the Cornish honey bee samples and the other *A. m.*
303 *mellifera* samples was then investigated separately. Analysis parameters and steps
304 were conducted as described above. Finally, the relationship between degree of
305 admixture and observed heterozygosity and allelic richness were tested at the
306 population level. Mean coefficients of admixture (i.e. mean Q value across individuals)
307 were calculated for membership to the 'dominant' cluster for each population.

308 Correlations between mean Q-value and mean observed heterozygosity (calculated
309 in Arlequin 3.5) and mean allelic richness (calculated in FSTAT 2.9.3.2 (Goudet 2001))
310 were tested using Spearman's rank method in R 3.4.1 (R Foundation for Statistical
311 Computing).

312

313 In the second approach, investigation of admixture was carried out using the R
314 package *adegenet* 1.3-0 (Jombart 2008) using the *dapc* functions (discriminant
315 analysis of principal components, DAPC, Jombart 2011). Preliminary analysis showed
316 that Italian bees were distant from the other samples, so this analysis was performed
317 only for *A. m. carnica* and *A. m. mellifera* samples (however, analysis of all samples
318 is included in the supplementary material). First the *find.clusters* function was used to
319 identify the most likely number of population clusters. A test DAPC analysis (for the
320 most likely number of population clusters identified using *find.clusters*) was then run
321 retaining all principal components and linear discriminants. The *a.score* function was
322 used to select the ideal number of principal components (PCs) to avoid overfitting. The
323 *a.score* function was run four times to examine convergence in the recommended
324 number of PCs to retain. The DAPC analysis was then repeated with the reduced set
325 of twenty PCs and four linear discriminants, for the most likely number of population
326 clusters identified in the first step. Scatter plots were produced for visual inspection of
327 clusters. Group memberships of individuals across source populations to the identified
328 clusters were tabulated. Membership probability of individual Cornish bees to the
329 identified clusters was plotted using the *compoplot* function.

330

331 ***Assignment of mtDNA haplotypes***

332 The COI-COII region was sequenced using the primers E2 (GGC AGA ATA AGT GCA
333 TTG) and H2 (CAA TAT CAT TGA TGA CC) (Garnery et al. 1993). The following
334 cycling protocol on a TC-412 Programmable Thermal Controller (Techne) was used:
335 35 cycles with 94°C for 60 s, 54°C for 45 s, and 62°C for 120 s. Before the first cycle,
336 a prolonged denaturation step (95°C for 15 min) was included and the last cycle was
337 followed by a 10 min extension at 72°C. Sanger sequencing was then conducted
338 (Sanger et al. 1977) using fluorescent dyes (Ansorge et al. 1987; Middendorf et al.
339 1992), specialized DNA polymerases (Taq-polymerase; Carothers et al. 1989) and
340 modified nucleotides to avoid problems with DNA secondary structure (Frederick
341 1999). Capillary electrophoresis was performed on an ABI3730 using Dye Chemistry
342 Software Data Collection Version 3; Sequencing Analysis 5.2 (Applied Biosystems).
343 Sequences were aligned using ClustalW (Thompson 1994) in Bioedit (Hall 2004).
344 Mitochondrial haplotypes were identified on the basis of the presence of P and Q
345 repeats (Cornuet et al. 1991). *A. m. mellifera* and *A. m. iberiensis* are in the M lineage
346 and are indicated by the presence of a P repeat and one or more Q repeats (Cornuet
347 et al. 1991, Achou et al. 2015) although *A. m. iberiensis* can also have A haplotypes
348 and present two types of P sequence (P0 and P). The common C lineage commercial
349 subspecies, *A. m. ligustica* and *A. m. carnica* lack a P repeat and have only a single
350 Q repeat (Cornuet et al. 1991).
351
352
353 **Results**
354 ***Quality control***
355 Three loci (A76, Ap001, A29) were not genotyped in Italian and French population
356 samples. Locus A43 was implicated twice as showing evidence of null alleles in

357 MICROCHECKER. Loci Ap226, A76, Ap289 were implicated once in showing
358 evidence of null alleles. Departure from Hardy-Weinberg equilibrium was shown for
359 loci A43 (Austria 2015), Ap001 (Austria) and for A76 and A28 (Austria Würm). Pairwise
360 linkage disequilibrium was not consistent for loci across populations apart from loci
361 Ac306 and Ap226 in the Swedish and Slovenian samples. Consequently loci A43 and
362 A76 (showed null alleles and departure from Hardy-Weinberg equilibrium), Ap001
363 (showed departure from Hardy-Weinberg equilibrium and had poor coverage across
364 populations) and A29 (poor coverage across all populations) were removed from the
365 dataset prior to downstream analyses. Standard estimates of genetic diversity are
366 shown in Table 1.

367

368 ***Investigation of admixture***

369 STRUCTURE analysis of all populations showed *K* = 3 and 5 as numbers of clusters
370 likely to be useful to describe the population structure (one should be careful with
371 interpreting the 'correct' *K* (Pritchard et al. 2000; Janes et al. 2017) hence results for
372 both are presented; see supplementary data figures 1 and 2). *K* = 3 clearly delineates
373 all three subspecies, with admixture shown for the Cornish population (Figure 1a). *K*
374 = 5 again separates *A. m. ligustica* from the other species; *A. m. carnica* show
375 membership to two clusters and *A. m. mellifera* again show a separate signal for the
376 Cornish sample in comparison with other populations of this subspecies (Figure 1b).

377 For *A. m. mellifera* examined separately in STRUCTURE, *K* = 2 and *K* = 3 are
378 useful descriptions of the population structure (supplementary data figures 3 and 4).
379 Both analyses show the Cornish population showing some distinction from the other
380 *A. m. mellifera* populations (Figure 2a and b).

381 There was a significant negative relationship between mean Q-value and mean
382 observed heterozygosity at the population level (rho = -0.68, n = 12, P<0.05; Figure
383 3a). There was also a significant negative relationship between mean Q-value and
384 mean allelic richness (corrected for sample size) at the population level (rho = -0.62,
385 n=12, P<0.05; Figure 3b)
386
387 Discriminant analysis of principal components showed five clusters as providing a
388 useful description of the population structure (supplementary figure 5). Twenty PCs
389 were retained after inspecting four iterations of *a.score* optimisation (the range of
390 recommended PCs to retain was 14-26; supplementary data figure 6). In addition, four
391 linear discriminants were used to model the population structure, visualised in a
392 scatterplot (Figure 4). Examination of membership of individuals to the identified
393 clusters (Table 2) shows that clusters 2, 3 and 5 mostly consist of *A. m. carnica*
394 individuals, cluster four consists mostly of Cornish *A. m. mellifera* samples and cluster
395 one represents other populations of *A. m. mellifera.* Individuals of Cornish samples
396 that did not group with cluster four were assigned to clusters two and five (four of forty-
397 three samples [9.3%]; Table 2, Figures 4 and 5). Analysis including the Italian bees
398 can be seen in the supplementary material (supplementary figure 7 and Table S1) and
399 also shows *A. m. mellifera* from Cornwall to be intermediate between *A. m. mellifera*
400 from continental Europe and *A. m. carnica*.
401
402 ***Admixture and mtDNA haplotype assignment***
403 Examination of Q-values from the STRUCTURE analysis of all populations for K = 3
404 in combination with mtDNA haplotype assignment give an indication of the degree of
405 introgression across the Cornish hives included in the analyses (Table 3; Table 4).

406 Interpretation is considered for Q threshold values of 80%, 90%, 95% and 99%; i.e for
407 a threshold of 0.99 an individual has to meet or exceed this value to be deemed 'pure'
408 (Table 4, see discussion). When lower values of Q-threshold are specified to indicate
409 a 'pure' bee, agreement between nuclear and mtDNA signal improves (i.e. more of the
410 M lineage samples are deemed to be 'pure' *A. m. mellifera* samples). Applying the
411 strictest threshold to deem a queen as 'pure' reveals only four individuals to be so and
412 also have an M haplotype (Table 4). Applying the lowest threshold, fifteen bees are
413 deemed to be pure on the basis of agreement between nDNA and mtDNA data (Table
414 4). All mtDNA sequences are available in GenBank (accession numbers MF197320-
415 197363).
416
417 **Discussion**
418 STRUCTURE analyses and a discriminant analysis of principal components both
419 indicate that samples of *A. m. mellifera* from beekeepers involved in the B4 project in
420 the south-west of the UK are clearly distinct to other *A. m. mellifera* populations. This
421 is most likely a consequence of admixture with imported lines rather than these
422 apiaries representing a naturally differentiated lineage of *A. m. mellifera.* The bees
423 sampled showed admixture from *A. m. carnica* introgression (STRUCTURE analyses)
424 and were intermediate between clusters of *A. m. mellifera* and *A. m. carnica* (DAPC
425 plots). This result is hardly surprising given the history of beekeeping in the UK. Local
426 beekeepers report that after the First World War and the 'Isle of Wight disease' (when
427 widespread losses of bees were attributed (incorrectly) to a single infectious disease
428 (Bailey 1964)), bees were brought into Cornwall from the Netherlands (dark European
429 honey bees), but also from Italy. Since this time, there have also been many imports
430 of other subspecies of honey bee to the UK. Cornwall is not far from Buckfast in Devon

431 where Brother Adam developed the hybrid line that became known as "the Buckfast
432 bee[TM]". Imports of honey bee into the UK increased in the period 2013-2016 (Learner
433 2017) and current advice to beekeepers from the National Bee Unit is that importing
434 bees 'is neither difficult nor a chore' (Learner 2017). Previous studies examining *A. m.*
435 *mellifera* have shown that there is admixture in unprotected English populations and
436 that English samples showed both M and C lineages (Jensen et al. 2005; Pinto et al.
437 2014). Scottish samples from protected areas showed only M lineages (Jensen et al.
438 2005; Pinto et al. 2014). Elsewhere in Europe, and for other subspecies, hybrids have
439 been found in populations of dark bees in Poland (Oleksa et al. 2011), admixed
440 ancestry is reported in Serbian bees between *A. m. carnica* and *A. m. macedonica*
441 (Nedić et al. 2014), but there are also places where lineages are relatively pure, e.g.
442 *A. m. mellifera* in parts of the Urals and Volga region (Ilyasov et al. 2016) and *A. m.*
443 *carnica* in Hungary (Péntek-Zakar et al. 2015). Clearly, transhumance of colonies
444 frequently leads to introgression, but there are also places where *A. m. mellifera*
445 remains relatively intact (Byatt et al. 2015)
446
447 Regarding the identification of relatively pure hives for conservation efforts in South-
448 West England, the power in the dataset to effectively detect hybrids and the effect of
449 threshold values on the designation of 'pure' individuals needs to be considered. Vähä
450 and Primmer (2006) investigated the use of STRUCTURE and NEWHYBRIDS to do
451 so and show that the number of loci for efficient and accurate determination of hybrids
452 depends on the amount of genetic differentiation between the parental populations.
453 F_{ST} values between the subspecies studied here are quite large (e.g. in the range 0.3-
454 0.6 for *Aml* and *Amm*, 0.2-0.6 for *Amc* and *Amm*, and 0.3-0.4 for *Aml* and *Amc* for the
455 loci used here, data not shown), so the use of relatively few loci will still be suitable for

identification of hybrids. However, it should be remembered that here, we are not detecting hybrid individuals between two parent lines where K=2, but rather trying to identify the degree of admixture from populations with a long history of crosses and back-crosses, where the useful number of clusters to describe the populations is probably from 3-5. Consequently, caution should be drawn when considering the relative purity of individuals using the approach described here. Vähä and Primmer (2006) showed that a stricter Q-value reduced the misclassification of back-crossed individuals as purebred individuals in their simulations. These authors note that as Q-value thresholds are increased there is a trade-off between the efficiency of detecting hybrids (proportion of individuals in a group correctly identified as hybrids) and the accuracy (proportion of an identified group truly belonging in that category). More stringent thresholds improve the accuracy of identifying hybrids, but decrease the efficiency (Vähä and Primmer 2006). Essentially, what needs to be decided is whether accurate hybrid identification (all the individuals in the hybrid group are hybrids, but some of the individuals in the purebred group are hybrids) or accurate purebred identification (all the individuals in the purebred group are purebred, but some of the individuals in the hybrid group are purebreds) is required. To conserve dark European honey bees, the purity of the dark European honey bee stock is of course the most desired outcome. However, to be certain of purity, the founding stock will be small (Table 4). Only four hives sampled showed an M lineage and a Q-value of >99% to *A. m. mellifera* cluster when K = 3. This value increases to 12 for a Q-value of 0.9. Further work investigating the phenotypic traits of Cornish bees for hives of differing admixed ancestry will help elucidate what is a useful threshold Q-value. We also note here that samples are limited and only eight microsatellites are analysed so the results should be interpreted with caution.

481

482 Depending on the Q-value (Table 4), some hives were observed to have an M lineage
483 haplotype and show nuclear introgression. This pattern would be expected in a
484 controlled population threatened by hybridization. Between one to three samples
485 showed limited nuclear introgression, but had C haplotypes. This suggests an
486 historical intake of foreign queens. This could be from swarms of unknown origin or
487 purchase of queens from uncontrolled breeding programmes. Recurrent backcrosses
488 with native bees subsequent to this historical introgression would give rise to a
489 situation where foreign nDNA cannot be detected with the applied marker set. No
490 samples were classified as pure *A. m. carnica* or *A. m. ligustica* (for Q>0.9, one
491 individual was assigned to *A. m. carnica* at Q>0.8) but the sampling method used here
492 particularly targeted beekeepers believing they likely had dark European honey bees.
493 This was intentional as we aimed to investigate the level of admixture in bees of this
494 type and identify potential hives for conservation management of dark European
495 honey bees in the South-West. Sampling was also limited to mostly East and West
496 Cornwall. It is likely that other keepers of dark European honey bees in the area could
497 have been missed in the current study; our research was conducted through the local
498 organisation 'B4' and only included the dark Europen honey bee beekeepers known
499 to this organisation.

500

501 The conservation implications of these findings are either to accept a degree of foreign
502 introgression, or to look to set up breeding programmes with other UK hives in order
503 to 'stock' reserves for South-West dark European honey bees. Although four samples
504 could be classified as pure *A. m. mellifera,* clearly, breeding from a founding stock of
505 only four colonies would lead to inbreeding and significant loss of genetic diversity

506 which may increase extinction risk (Frankham 2005). Much broader sampling of hives
507 in the region needs to be undertaken to identify other dark European honey bee hives
508 in the area (we specifically sampled hives from beekeepers involved in the B4 project,
509 but in total in the region in the year of sampling there were 4966 hives registered on
510 Beebase. In April 2018, there were 1140 beekeepers registered in Cornwall, with 5538
511 colonies). Although it is possible that much of the genetic load can be purged by
512 selection on the haploid sex in haplodiploids (Henter 2003), female sex-limited traits
513 may not be affected (Tien et al. 2015) and there is evidence that haplodiploids can still
514 show inbreeding depression (Henter 2003). This is especially important in systems
515 where single locus complementary sex determination exists (Whitehorn et al. 2009).
516 In honey bees, within-colony genetic diversity is also known to be important for disease
517 resistance (Brown and Schmid-Hempel 2003).
518
519 The argument for conservation of locally adapted varieties makes sense from a
520 viewpoint that maintaining a network of locally adapted forms (i) maximises genetic
521 variation across the species as a whole; (ii) maintains co-adapted locus complexes
522 and allows the persistence of locally adapted forms which already exist and are
523 assumed to be most resilient to local environmental stochasticity; (iii) maintains the
524 possibility of allowing human-mediated migration of particularly resistant forms (e.g. in
525 the event of disease outbreaks or climate change). Nevertheless, and as already
526 mentioned, action should be taken to minimise erosion of genetic diversity from these
527 local populations through inadequate breeding population sizes and consequent
528 genetic drift. In contrast, counter arguments could be made (Harpur et al. 2012) in the
529 sense that admixture will increase within population genetic variance. Selection is also
530 more efficient in large populations (because low frequency advantageous *de novo*

531 mutations are less likely to be lost by drift (see Olson-Manning et al. 2012 for an up-
532 to-date review)). Although a preliminary analysis included here shows a significant
533 negative correlation between Q-values and observed heterozygosity and allelic
534 richness (more admixed populations are more genetically diverse), our sample sizes
535 were small in several populations and only twelve population samples were included.
536 In contrast, de la Rua et al. (2013) in their critique of Harpur et al. (2012), found that
537 even where ongoing introduction of foreign queens takes place, genetic diversity is not
538 necessarily increased (Muñoz et al. 2014). In Italian honey bee populations, large-
539 scale queen breeding has reduced genetic diversity (see Dall'Olio et al. 2007). The
540 argument (de la Rua et al. 2013 cf. Harpur et al., 2012) about which scenario best
541 maximises evolutionary potential depends on the relative importance of increased
542 within population variation (resulting from hybridization/introgression) versus loss of
543 between population variance (that conservation of locally adapted forms seek to
544 minimize).
545
546 Currently, legislation regarding honey bees in England and Wales relates to screening
547 of colonies for diseases and parasites (Bee Diseases and Pests Control (England)
548 Order 2006; Bee Diseases and Pests Control (Wales) Order 2006), health certification
549 (Council Directive 92/65/EEC) and regards countries from outside the EU whence
550 bees may be imported (Commision regulation (EU) 206/2010) as well as foods
551 standards laws regarding the composition of honey for sale. There is an argument that
552 the National Pollinators Strategy (DEFRA) should be extended to give greater
553 protection to the native honey bee diversity that exists. Strict protection would be
554 necessary to avoid hybridization of native colonies, as has occurred in other protected

areas in the past (Jensen et al. 2005). Urgent action is needed to characterise local adaptation before further erosion of these forms occurs.

Considering all results, immediate action is recommended to (i) more extensively sample both the South-West population and the UK populations to detect any pure uncompromised breeding stock; (ii) obtain more accurate assessment of introgression using ancestry informative SNPs which are known to outperform microsatellites (Muñoz et al. 2017); (iii) measure local adaptation of dark European honey bee colonies across the UK using genome-wide data aiming to detect recent and historical selection; (iv) start conservation actions to protect locally adapted varieties identified; (v) bring together networks of *A. m. mellifera* beekeepers from across the UK at the appropriate geographic scales identified.

Acknowledgements

Many thanks to all the beekeepers of the B4 organisation for providing samples for the project.

References

Achou M, Loucif-Ayad W, Legout H, Alburaki M, Garnery L (2015) An insightful molecular analysis reveals foreign honey bees among Algerian populations (*Apis mellifera L.*) Data Mining in Genomics and Proteomics 6:1

Alburaki M, Bertrand B, Legout H, Moulin S, Sheppard WS, Garnery L (2013) A fifth major genetic group among honeybees revealed in Syria. BMC Genetics 14: 117

580

581 Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids:
582 setting conservation guidelines. Trends Ecol Evol 16: 613-622
583
584 Allendorf FW, Leary RF, Hitt NP, Knudsen KI, Lundquist LI, Spruell P (2004)
585 Intercrosses and the U.S. Endangered Species Act: should hybridized populations be
586 included as Westslope Cutthroat Trout? Conserv Biol 18: 1203-1213
587
588 Anderson E (1949) Introgressive Hybridization. John Wiley and Sons, Inc, New York;
589 Chapman and Hall, Limited, London
590
591 Ansorge W, Sproat B, Stegemann J, Schwager C, Zenke M. (1987) Automated DNA
592 sequencing: ultrasensitive detection of fluorescent bands during electrophoresis.
593 Nucleic Acids Res15: 4593–4602.
594
595 Bailey L (1964) The 'Isle of Wight Disease': the origin and significance of the myth.
596 Bee World 45: 32-37
597
598 Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria.
599 Mol. Ecol. 13: 729–744
600
601 Bee Diseases and Pests Control (England) Order 2006 Available at:
602 http://www.legislation.gov.uk/uksi/2006/342/pdfs/uksi_20060342_en.pdf
603

604 Bee Diseases and Pests Control (Wales) Order 2006 Available at:
605 http://www.legislation.gov.uk/wsi/2006/1710/pdfs/wsi_20061710_mi.pdf
606
607 Brown MJF, Schmid-Hempel P (2003) The evolution of female multiple mating in social
608 Hymenoptera. Evolution 57: 2067-2081
609
610 Büchler R, Costa C, Hatjina F Andonov S, Meixner MD, Le Conte Y et al. (2014) The
611 influence of genetic origin and its interaction with environmental effects on the survival
612 of *Apis mellifera L.* colonies in Europe. J Apicult Res 53: 205-214
613
614 Byatt MA, Chapman NC, Latty T, Oldroyd BP (2015) The genetic consequences of the
615 anthropogenic movement of social bees. Insectes Soc 63: 15-24
616
617 Carothers A, Urlaub G, Mucha J, Grunberger D, Chasin L (1989) Point mutation
618 analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of
619 cDNA, and Taq sequencing by a novel method. BioTechniques 7: 494–6, 498–9
620
621 Carreck NL (2008) Are honey bees (*Apis mellifera* L.) native to the British Isles? J
622 Apicult Res & Bee World 47: 318-322
623
624 Commision regulation (EU) 206/2010 of 12 March 2010 laying down lists of third
625 coutnries, territories or parts thereof authorised for the introduction into the European
626 Union of certain animals and fresh meat and the veterinary certification requirements
627 (Text with EEA relevance). Available at:
628 http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32010R0206

629

630 Cornuet J-M, Garnery L (1991) Mitochondrial DNA variability in honeybees and its
631 phylogeographic implications. Apidologie 22: 627-642
632
633 Cornuet J-M, Garnery L, Solignac M (1991) Putative origin and function of the
634 intergenic region between COI and COII of *Apis mellifera* L. mitochondrial DNA.
635 Genetics 128: 393-403
636
637 Costa C, Büchler R, Berg S, Bienkowska M, Bouga M, Bubalo D, et al. (2012) A
638 Europe-wide experiment for assessing the impact of genotype-environment
639 interactions on the vitality and performance of honey bee colonies: experimental
640 design and trait evaluation. J Apicult Sci 56: 147-158
641
642 Council Directive 92/65/EEC of 13 July 192 laying down animal health requirements
643 governing trade in and imports into the Community of animals, semen, ova and
644 embryos not subject to animal health requirements laid down in specific Community
645 rules referred to in Annex A (I) to Directive 90/425/EEC. Available at:
646 http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31992L0065
647
648
649 Dall'Olio R, Marino A, Lodesani M, Moritz RFA. (2007) Genetic characterization of
650 Italian honeybees, *Apis mellifera ligustica,* based on microsatellite DNA
651 polymorphisms. Apidologie 38: 207-217
652

653 Delaplane KS, Mayer DF (2000) *Crop pollination by bees.* CABI, ISBN 085199783X,
654 9780851997834, Oxford UK and New York, USA
655
656 De la Rúa P, Serrano J, Galián J (1998) Mitochondrial DNA variability in the Canary
657 Islands honeybees (*Apis mellifera* L.) Mol Ecol 7: 1543-1547
658
659 De la Rúa P, Jaffé R, Muñoz I, Serrano J, Moritz RF, Kraus FB (2013) Conserving
660 genetic diversity in the honeybee: comments on Harpur et al., (2012). Mol Ecol 22:
661 3208-3210
662
663 Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualising
664 STRUCTURE output and implementing the Evanno method. Conserv Genet Resour
665 4: 359-361
666
667 Ellis JS , Gilbey J , Armstrong A, Balstad T, Cauwelier C, Cherbonnel C, Consuegra S,
668 Coughlan J, Cross TF, Crozier W, Dillane E, Ensing D, García de Leániz C, García-
669 Vázquez E, Griffiths AM, Hindar K, Hjorleifsdottir S, Knox D, Machado-Schiaffino G,
670 McGinnity P, Meldrup D, Nielsen EE, Olafsson K, Primmer CR, Prodohl P, Stradmeyer L,
671 Vähä J-P, Verspoor E, Wennevik V, Stevens J R (2011) Microsatellite standardization and
672 genotyping error in a large multi-partner research programme for conservation of Atlantic
673 salmon (Salmo salar L.) Genetica 139: 353-367
674
675 Epifanio J, Philipp D (2001) Simulating the extinction of parental lineages from
676 introgressive hybridization: the effects of fitness, initial proportions of parental taxa
677 and mate choice. Rev Fish Biol Fisher 10: 339-354
678

679 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of
680 individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-
681 2620
682
683 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to
684 perform population genetics analyses under Linux and Windows. Mol EcolResour
685 10: 564-567
686
687 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using
688 multilocu genotype data: linked loci and correlated allele frequencies. Genetics 164:
689 1567-1587
690
691 Franck P, Garnery L, Loiseau A, Oldroyd BP, Hepburn HR, Solignac M, Cornuet JM
692 (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial
693 data. Heredity 86: 420-430
694
695 Frankham R (2005) Genetics and extinction. Biol Conserv 126: 131-140
696
697 Frederick MA (1999) *Short protocols in molecular biology: a compendium of methods*
698 *from current protocols in molecular biology 4th Ed* (FM Ausubel, R Brent, RE Kingston,
699 et al., Eds,). John Wiley and Sons, Inc., Hoboken, NJ, USA.
700
701 Gallai N, Salles J-M, Settele J, Vassière BE (2009) Economic value of the vulnerability
702 of world agriculture confronted with pollinator decline. Ecol Econ 68: 810-821
703

704 Garnery L, Solignac M, Celebrano G, Cornuet J-M (1993) A simple test using restricted
705 PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L.
706 Experientia 49: 1016
707
708
709 Garnery L, Franck P, Baudry E, Vautrin D, Cornuet J-M, Solignac M. (1998a) Genetic
710 diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*)
711 II. Microsatellite loci. Genet Sel Evol 30: S49-74
712
713 Garnery L, Franck P, Baudry E, Vautrin D, Cornuet J-M, Solignac M (1998b) Genetic
714 diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*)
715 I. Mitochondrial DNA. Genet Sel Evol 30: S31-S47
716
717 Genersch L. (2010) Honey bee pathology: current threats to honey bees and
718 beekeeping. Appl Microbiol Biot 87: 87-97
719
720 Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and
721 fixation indices (version 2.9.3). Available from
722 http://www.unil.ch/izea/softwares/fstat.html
723
724 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and
725 analysis program for Windows 95/98/NT. Nucl Acid S 41:95-98.
726
727 Hamilton JA, Miller JM (2016) Adaptive introgression as a resource for management
728 and genetic conservation in a changing climate. Conserv Biol 30: 33-41

729

730 Harpur BA, Minaei S, Kent CF, Zayed A (2012) Management increases genetic
731 diversity of honey bees via admixture. Mol Ecol 21: 4414-4421
732
733 Henter HJ (2003) Inbreeding depression and haplodiploidy: experimental measures in
734 a parasitoid and comparisons across diploid and haplodiploid insect taxa. Evolution
735 57: 1793-1803
736
737 Highfield AC, Nagar AE, Mackinder LCM, Noël JM-LL, Hall MJ, Martin SJ, Schroeder
738 DC (2009) Deformed wing virus implicated in overwintering honey bee colony losses.
739 Appl Environ Microb 75: 7212-7220
740
741 Issa MRC, Figueiredo VLC, De Jong D, Sakamoto CH, Simões ZLP (2013) Rapid
742 method for DNA extraction from the honey bee *Apis mellifera* and the parasitic bee
743 mite *Varroa destructor* using lysis buffer and proteinase K. Genetics and Molecular
744 Research 12: 4846-4854
745
746 Ilyasov RA, Poskruakov AV, Petukhov AV, Nikolenko AG (2016) New approach to the
747 mitotype classification in black honeybee *Apis mellifera mellifera* and Iberian
748 honeybee *Apis mellifera iberiensis.* Russ J Genet+ 52: 281-291
749
750 Janes JKJ, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew
751 RL (2017) The $K = 2$ conundrum. Mol Ecol 26: 3594-3602
752

753 Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005) Varying degrees of *Apis*
754 *mellifera ligustica* introgressionin protected populations of the black honey bee, *Apis*
755 *mellifera mellifera,* in northwest Europe. Mol Ecol 14: 93-106
756
757 Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic
758 markers. Bioinformatics 24: 1403-1405
759
760 Jombart T (2011) A tutorial for discriminant analysis of principal components (DAPC)
761 using *adegenet* 1.3-0. Available at:
762 https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/inst/doc/adegenet-
763 dapc.pdf?revision=887&root=adegenet&pathrev=887
764
765
766 Kauhausen-Keller D, Keller R (1994) Morphometrical control of pure race breeding in
767 the honey bee (*Apis mellifera* L). Apidologie 25: 133-143
768
769 Learner J (2017) Imports and exports of honey bees. Bee Craft 99: 7-10
770
771 Lee KV, Steinhauer N, Rennich K, Wilson ME, Tarpy DR, Caron DM, et al. (2015) A
772 national survey of the managed honey bee 2013-2014 annual colony losses in the
773 USA. Apidologie 46: 292-305
774
775 Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for
776 connecting population genetics and genomics programs. Bioinformatics 28: 298-299.
777

778 Louveaux J, Albisetti M, Delangue M, Theurkauff M (1966) Les modalités de
779 l'adaptation des abeilles (*Apis mellifica* L.) au milieu naturel. Ann. Abeille 9: 323–350
780
781
782 Maul V, Hähnle A (1994) Morphometric studies with pure bred stock of *Apis mellifera*
783 *carnica* Pollman from Hessen. Apidologie 25: 119-132
784
785 Meixner MD, Costa C, Kryger P, Hatjina F, Bouga M, Ivanova E, Büchler R (2010)
786 Conserving diversity and vitality for honey bee breeding. J Apicult Res 49: 85-92
787
788 Meixner MD, Pinto MA, Bouga M, Kryger P, Ivanova E, Fuchs S (2013) Standard
789 methods for characterising subspecies and ecotypes of *Apis mellifera.* J Apicult Res
790 52: The COLOSS BEEBOOK Part 2
791
792 Middendorf LR, Bruce JC, Bruce RC et al. (1992) Continuous, on-line DNA sequencing
793 using a versatile infrared laser scanner/electrophoresis apparatus. Electrophoresis 13:
794 487–494.
795
796 Morse RA, Calderone NW (2000) The value of honey bees as pollinators of U.S,
797 crops in 2000. Cornell University, Ithaca, New York
798
799 Muñoz I, Henriques D, Spencer Johnston J, Chavez-Galarza J, Kryger P, Pinto M
800 (2015) Reduced SNP panels for genetic identification and introgression analysis in the
801 dark honey bee (*Apis mellifera mellifera*). PLOS ONE 10, e0124365
802

803 Muñoz IM, Henriques D, Jara L, Spencer Johnston J, Chávez-Galarza J, de la Rua
804 P, Pinto M.A. (2017) SNPs selected by information content outperform randomly
805 selected microsatellite loci for delineating genetic identification and introgression in
806 the endangered dark European honey bee (*Apis mellifera mellifera*). Mol Ecol
807 Resour 17: 783-795
808
809 Nedić N, Francis RM, Stanisavljević L, Pihler I, Kezić N, Bendixen C, Kryger P
810 (2014) Detecting population admixture in honey bees of Serbia. J Apicult Res 53:
811 303-313
812
813
814 Oleksa A, Chybicki I, Tofilski A, Burczyk J (2011) Nuclear and mitochondrial patterns
815 of introgression into native dark bees (*Apis mellifera*) in Poland. J Apicult Res 50:
816 116-129
817
818 Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution:
819 evaluating empirical support for theoretical predictions. Nat Rev Genet 13: 867-877
820
821 OPERA Research Center (2013) Bee health in Europe – Facts and figures 2013.
822 Compendium of the latest information on bee health in Europe. Università Cattolica
823 del Sacro Cuore.
824
825 Parejo M, Wragg D, Gauthier L, Vignal A, Neumann P, Neuditschko M (2016) Using
826 whole-genome sequence information to foster conservation efforts for the European
827 dark honey bee, *Apis mellifera mellifera.* Front Ecol Evol 4: 140

828

829 Péntek-Zakar E, Oleksa A, Borowik T, Kusza S (2015) Population structure of honey
830 bees in the Carpathian Basin (Hungary) confirms introgression from surrounding
831 subspecies. Ecol Evol 5: 5456-5467
832
833 Pinto MA, Henriques D, Chávez-Galarza J, Kryger P, Garnery L, et al. (2014) Genetic
834 integrity of the dark European honey bee (*Apis mellifera mellifera*) from protected
835 populations: a genome-wide assessment using SNPs and mtDNA sequence data. J
836 Apicult Res 53: 269-278
837
838 Parliamentary Office of Science and Technology (POST) (2010) Postnote: Insect
839 Pollination. Number 348. Parliamentary Office of Science and Technology, London
840
841 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
842 multilocus genotype data Genetics 155: 945-959
843
844 Pritchard JK, Wen X, Falush D. (2010) Documentation for *structure* software:
845 Version 2.3. From http://pritch.bsd.uchicago.edu/structure.html
846
847 Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a
848 program for drawing elegant STRUCTURE bar plots in user friendly interface.
849 SpringerPlus 3: 431
850

851 Rortais A, Arnold G, Alburaki M, Legout H, Garnery L (2011) Review of the Dra*I COI-*
852 *COII* test for the conservation of the black honeybee (*Apis mellifera mellifera*). Conserv
853 Genet Resour 3: 383-391
854
855 Ruttner F (1988) Biogeography and taxonomy of honey bees. Springer-Verlag, Berlin,
856 Germany. ISBN 0387177817
857
858 Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating
859 inhibitors. P Natl Acad Sci USA 74: 5463–5467.
860
861
862
863 Soland-Reckeweg G, Heckel G, Neumann P, Fluri P, Excoffier L (2009) Gene flow in
864 admixed populations and implications for the conservation of the Western honey bee,
865 *Apis mellifera.* J Insect Conserv 13: 317-328
866
867 Solignac M, Vautrin D, Loiseau A, Mougel F, Baudry E, Estoup A, Garnery L, Haberl
868 M, Cornuet JM (2003) Five hundred and fifty microsatellite markers for the study of the
869 honeybee (Apis mellifera L.) genome. Mol Ecol Notes 3: 307-311
870
871 Strange JP, Garnery L, Sheppard WS (2007) Persistence of the Landes ecotype of
872 *Apis mellifera mellifera* in southwest France: confirmation of a locally adaptiveannual
873 brood cycle trait. Apidologie 38: 259-267
874

875 Strange JP, Garnery L, Sheppard WS (2008) Morphological and molecular
876 characterization of the Landes honey bee (*Apis mellifera* L.) ecotype for genetic
877 conservation. J Insect Conserv 12: 527-537
878
879
880 Templeton AR, Hemmer H, Mace G, Seal US, Shields WM, Woodruff DS (1986) Local
881 adaptation, coadaptation, and population boundaries. Proceedings of the Workshop
882 on Genetic Management of Captive Populations 5: 115-125
883
884 Tien NSH, Sabelis MW. Egas M (2015) Inbreeding depression and purging in a
885 haplodiploid: gender-related effects. Heredity 114: 327-332
886
887 Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W—improving the sensitivity of
888 progressive multiple sequence alignment through sequence weighting, position-
889 specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
890 Vähä J-P, Primmer C.R. (2006) Efficiency of model-based Bayesian methods for
891 detecting hybrid individuals under different hybridization scenarios and with different
892 numbers of loci. Mol Ecol 15: 63-72
893
894 van Engelsdorp D, Meixner MD (2010) A historical review of managed honey bee
895 populations in Europe and the United States and the factors that may affect them. J
896 Invertebr Pathol 103: S80-S95
897

898 van Engelsdorp D, Caron D, Hayes J, Underwood R, Henson, M, Rennich, K et al.
899 (2012) A national survey of managed honey bee 2010-11 winter colony losses in the
900 USA: results from the Bee Informed Partnership. J Apicult Res 51: 115-124
901
902 van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER:
903 software for identifying and correcting genotyping errors in microsatellite data. Mol
904 Ecol Resour 4: 535-538
905
906 Whitehorn PR, Tinsley MC, Brown MJF, Darvill B, Goulson D (2009) Impacts of
907 inbreeding on bumblebee colony fitness under field conditions. BMC Evol Biol 9: 152
908
909
910
911
912

913 **Figure legends**

914

915 **Fig. 1**

916 Group membership to clusters identified using STRUCTURE with inference based
917 on all populations (*A. m. ligustica* = 1 Italy, *A. m. carnica* = 2 Austria, 3 Austria
918 Würm, 4 Slovenia, 11 Austria 2015, *A. m. mellifera* = 5 Sweden, 6 France, 7 Norway,
919 8 Switzerland Glarus, 9 Switzerland Schistal, 10 Ireland, 12 Cornwall), **(a)** K = 3, **(b)**
920 K = 5. (Note that the coloured bar at the bottom of the figure illustrates the population
921 of origin only)

922

923 **Fig. 2** Group membership to clusters identified using STRUCTURE with inference
924 based on *A. m. mellifera* clusters only (numbering of populations is retained as in Fig
925 2 for comparison), **(a)** K = 2 **(b)** K = 3. (Note that the coloured bar at the bottom of
926 the figure illustrates the population of origin only)

927

928 **Fig. 3 (a)** Correlation between Q-values (lower values indicate increased admixture)
929 and observed heterozygosity across populations, **(b)** correlation between Q-values
930 and allelic richness (Ar)

931

932 **Fig. 4** Discriminant analysis of principal components for all populations sampled,
933 apart from *A. m. ligustica* (Italy; see text). Group four represents the putative dark
934 European honey bees from Cornwall. Clusters 2, 3 and 5 mostly consist of *A. m.*
935 *carnica* individuals. Cluster one represents continental European populations of *A.*
936 *m. mellifera.* (Membership of individuals from each population to each group is

937 shown in table 2). Analysis is based on retention of 20 principal components (Fig S6)

938 and all linear discriminants (four)

939

940 **Fig. 5** Group membership to clusters identified using the *find.clusters* function, for

941 Cornish honey bees only, based on the proportion of successful assignments to the

942 identified clusters. Groups 2,3 and 5 represent *A. m. carnica* populations (see Table

943 2) and group 4 *A. m. mellifera* populations

944

945

946

947 **Tables**

948 **Table 1** Genetic diversity of population samples included in the study. Populations

949 for each subspecies are: *A. m. ligustica* – Italy; *A. m. carnica* – Austrian and

950 Slovenian samples; *A. m. mellifera* all other samples

951

Population	**Sample size**	**Allelic richness** (± standard deviation)	**Observed heterozygosity** mean	standard deviation	standard error	**Expected heterozygosity** mean	standard deviation	standard error
Italy	55	3.5±2.7	0.35	0.35	0.12	0.38	0.37	0.13
Austria	44	4.4±2.0	0.54	0.24	0.08	0.57	0.20	0.07
Austria Würm	36	3.9±0.57	0.57	0.15	0.05	0.59	0.12	0.04
Austria 2015	102	4.3±1.9	0.51	0.22	0.08	0.52	0.24	0.08
Slovenia	212	3.6±1.8	0.45	0.26	0.09	0.44	0.27	0.09
Sweden	10	3.0±1.4	0.38	0.27	0.10	0.46	0.33	0.12
France	24	3.8±3.6	0.38	0.31	0.11	0.39	0.34	0.12
Norway	18	3.6±2.8	0.41	0.38	0.13	0.40	0.35	0.12
Switzerland Glarus	10	3.9±3.0	0.41	0.34	0.12	0.43	0.33	0.12
Switzerland Schistal	12	3.6±2.4	0.43	0.33	0.12	0.40	0.32	0.11
Ireland	22	3.7±2.6	0.36	0.26	0.09	0.39	0.32	0.11
Cornwall	43	5.1±2.8	0.60	0.26	0.09	0.63	0.21	0.07

952

953 **Table 2** Membership of individuals from each sampled population to the clusters
954 inferred in adegenet 1.3-0 using the *find.clusters* function. Numbers are the number
955 of individuals assigned to each cluster. Five clusters were inferred on the basis of
956 BIC estimates (Supplementary data, Figure 4). Subspecies are indicated in
957 superscript (*c* = *A. m. carnica*; *m* = *A. m. mellifera*)

958

Populations	**Inferred cluster**				
	1	**2**	**3**	**4**	**5**
Austria^{c}	3	7	17	1	16
Austria Wurm^{c}	0	29	2	0	5
Austria 2015^{c}	0	12	43	0	44
Slovenia^{c}	0	1	15	0	5
Sweden^{m}	10	0	0	0	0
France^{m}	24	0	0	0	0
Norway^{m}	18	0	0	0	0
Switzerland Glarus^{m}	9	0	0	1	0
Switzerland Schistal^{m}	12	0	0	0	0
Ireland^{m}	22	0	0	0	0
Cornwall^{m}	0	1	0	39	3

959
960
961

Table 3 Membership of Cornish honey bees to the clusters identified in STRUCTURE for K = 3. The Q-values are the mean admixture coefficients from three iterations of each K value. Standard deviations are also indicated. Blank cells under ‘mtDNA haplotype’ were not sequenced.

	Apis mellifera mellifera		***Apis mellifera ligustica***		***Apis mellifera carnica***		
Individual	Q	s.d	Q	s.d.	Q	s.d.	**mtDNA haplotype**
15-001	0.94	0.000	0.04	0.001	0.02	0.000	
15-002	0.97	0.000	0.01	0.000	0.01	0.000	
15-003	0.91	0.002	0.04	0.001	0.05	0.001	
15-004	0.98	0.000	0.01	0.000	0.01	0.000	
15-1321	0.94	0.001	0.04	0.000	0.03	0.000	
15-1401	0.41	0.001	0.03	0.000	0.57	0.001	
15-1407	0.99	0.000	0.00	0.000	0.01	0.000	
15-1409	0.99	0.000	0.01	0.000	0.01	0.000	
15-1410	0.99	0.000	0.01	0.000	0.01	0.000	
15-1411	0.97	0.000	0.01	0.000	0.01	0.000	
15-1137	0.94	0.001	0.03	0.001	0.03	0.000	M
15-1188	0.91	0.001	0.00	0.000	0.09	0.001	M
15-1204	0.91	0.001	0.01	0.000	0.08	0.001	M
15-1310	0.82	0.001	0.01	0.000	0.17	0.001	M
15-1311	0.99	0.000	0.00	0.000	0.01	0.000	M
15-1312	0.56	0.001	0.03	0.001	0.41	0.002	C
15-1313	0.93	0.001	0.01	0.000	0.07	0.001	M
15-1315	0.37	0.002	0.01	0.000	0.62	0.002	C
15-1317	0.98	0.000	0.01	0.000	0.01	0.000	
15-1322	0.62	0.003	0.04	0.001	0.34	0.004	
15-1323	0.99	0.000	0.00	0.000	0.01	0.000	M
15-1327	0.95	0.001	0.01	0.000	0.04	0.000	M
15-1330	0.86	0.002	0.01	0.000	0.14	0.002	M
15-1332	0.55	0.000	0.01	0.000	0.45	0.000	C
15-1334	0.63	0.003	0.01	0.000	0.36	0.003	M
15-1335	0.99	0.000	0.00	0.000	0.01	0.000	C
15-1336	0.43	0.001	0.02	0.000	0.55	0.000	C
15-1345	0.99	0.000	0.00	0.000	0.01	0.000	
15-1348	0.98	0.000	0.01	0.000	0.01	0.000	C
15-1349	0.94	0.000	0.03	0.000	0.03	0.000	M
15-1351	0.16	0.001	0.01	0.000	0.84	0.001	C
15-1352	0.83	0.000	0.02	0.000	0.15	0.001	C
15-1355	0.99	0.000	0.00	0.000	0.00	0.000	M
15-1356	0.57	0.001	0.01	0.000	0.42	0.001	C

15-1359	0.86	0.000	0.02	0.000	0.13	0.000	M
15-1360	0.35	0.001	0.07	0.001	0.58	0.000	C
15-1364	0.74	0.002	0.02	0.001	0.24	0.001	M
15-1372	0.49	0.001	0.00	0.000	0.51	0.001	C
15-1377	0.99	0.000	0.01	0.000	0.01	0.000	M
15-1422	0.58	0.001	0.08	0.001	0.34	0.002	C
15-1423	0.94	0.000	0.01	0.000	0.06	0.000	M
15-1426	0.98	0.000	0.01	0.000	0.01	0.000	C
15-1427	0.98	0.000	0.01	0.000	0.02	0.000	M

Table 4 Assignment of mtDNA haplotypes (M or C lineage) and nuclear introgression considered together. Individuals are assigned as 'pure' based on differing Q-values of >99%, >95%, >0.90% and >0.80% for $K = 3$.

		mtDNA		**Not sequenced**
Q-threshold	**nDNA**	M	C	
0.99	'pure'	4	1	4
	introgressed	13	12	9
0.95	'pure'	6	3	8
	introgressed	11	10	5
0.90	'pure'	12	3	11
	introgressed	5	10	2
0.80	'pure'	15	3	11
	introgressed	2	10	2

Fig 1

(a)

(b)

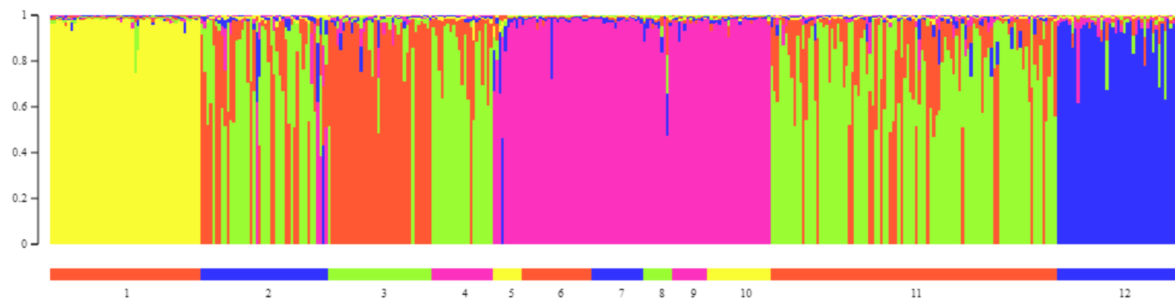
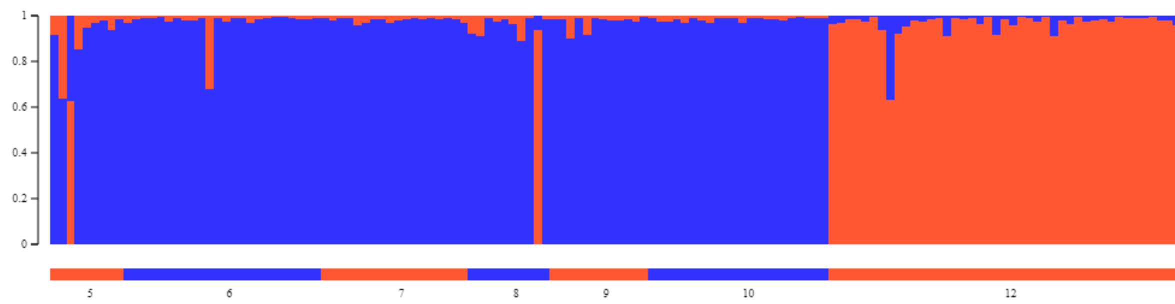


Fig 2

(a)



(b)

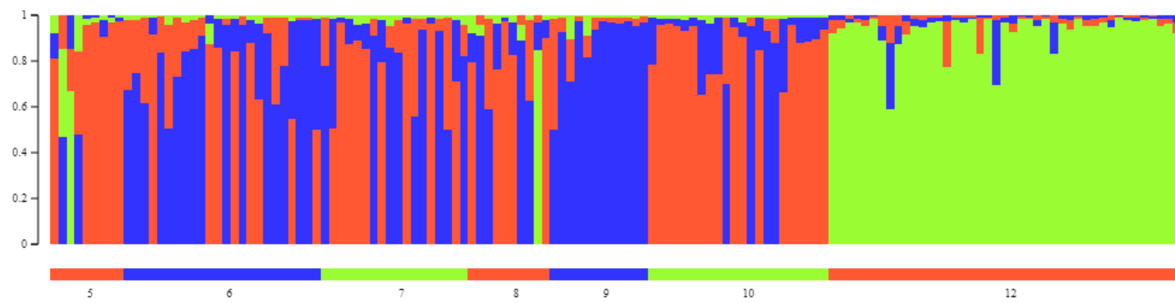
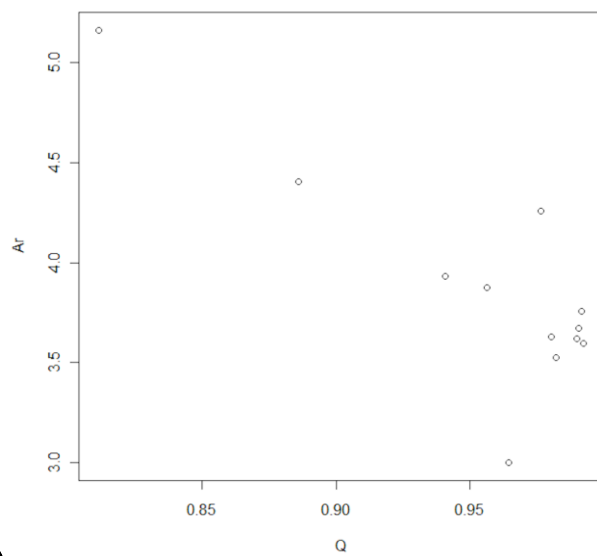


Fig 3

(a)



(b)

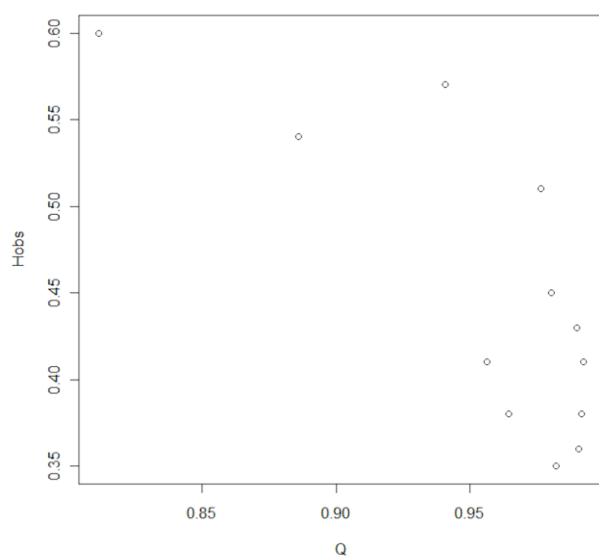


Fig 4

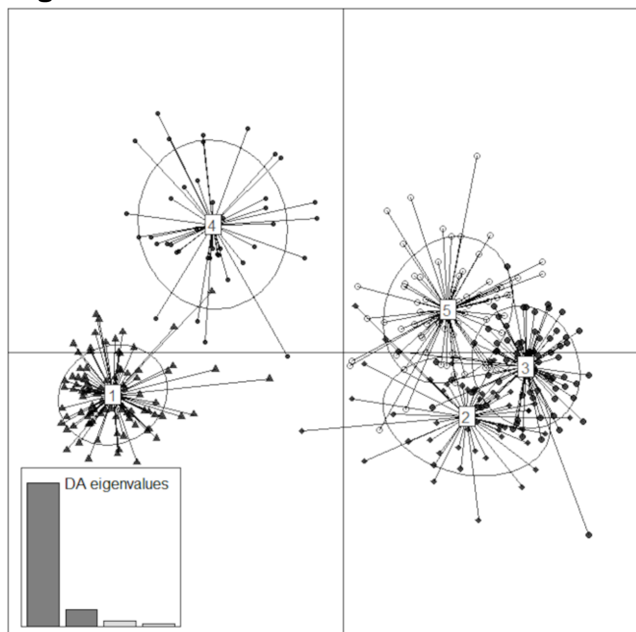


Fig 5