Microsatellite analysis supports the existence of three cryptic species within the bumble bee *Bombus lucorum sensu lato*

Lorraine McKendrick[1], Jim Provan[2], Úna Fitzpatrick[3], Mark J. F. Brown[4], Tómas E. Murray[3], Eckart Stolle[5], Robert J. Paxton[1,6 *]

[1] School of Biological Sciences, Queen's University Belfast, Belfast BT9 7BL, UK

[2] Institute of Biological, Environmental and Rural Sciences, Penglais, Aberystwyth University, Aberystwyth SY23 3DA, UK

[3] National Biodiversity Data Centre, Carriganore, Waterford, Ireland

[4] School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK

[5] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

[6] Institute for Biology, Martin Luther-University Halle-Wittenberg, Hoher Weg 8, 06120 Halle (Saale), Germany and German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig (iDiv), Deutscher Platz 5e, 04103 Leipzig, Germany

[*]Corresponding author: Robert Paxton

E-mail: robert.paxton@zoologie.uni-halle.de

Tel: +49 (0)345 55 26500

Fax: +49 (0)345 55 27428

Abstract

Mitochondrial cytochrome oxidase I (COI) partial sequences are widely used in taxonomy for species identification. Increasingly, these sequence identities are combined with modeling approaches to delineate species. Yet the validity of species delineation based on such DNA 'barcodes' is rarely tested and may be called into question by phenomena such as ancestral polymorphisms in DNA sequences, phylogeographic divergence, mitochondrial introgression and hybridization, or distortion of mitochondrial inheritance through such factors as *Wolbachia* infection. The common and widespread European bumble bee *Bombus lucorum s. lato* contains three distinct mitochondrial DNA lineages that are assumed to represent three cryptic species, namely *Bombus cryptarum*, *B. lucorum s. str.* and *Bombus magnus*. To test whether nuclear gene pools of the three putative species were differentiated, we genotyped 304 sympatric members of the *lucorum* complex (54 *B. cryptarum* females, 168 *B. lucorum s. str.* females and 82 *B. magnus* females, as defined using mtDNA COI haplotypes) from 11 localities spread across the island of Ireland at seven nuclear microsatellite loci. Multilocus genotypes clustered into three discrete groups that largely corresponded to the three mtDNA lineages: *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*. The good fit of mitochondrial haplotype to nuclear (microsatellite) genotypic data supports the view that these three bumble bee taxa are reproductively isolated species, as well as providing a vindication of species identity using so-called DNA barcodes.

Keywords DNA barcode; *cryptarum*; *magnus*; mitochondrial cytochrome oxidase I; STRUCTURE software; PCoA; DAPC, sympatry

Introduction

Bumble bees (Hymenoptera: Apidae, genus *Bombus*) are of great ecological and economic importance as major pollinators of both crops and wild flowers in the Northern Hemisphere, yet they are in decline (e.g. Fitzpatrick et al. 2007; Goulson 2009; Cameron et al. 2011). Though members of the subgenus *Bombus sensu stricto* (=*Terrestribombus* Vogt) are the most abundant and widespread of all bumble bees, exhibiting a holarctic distribution (Hines 2008; Williams et al. 2008, 2012a), they can be difficult to identify in the field using the minor morphological differences that separate species (Carolan et al. 2012; Bossert 2015); the apparent abundance of members of the subgenus may mask the rarity of its morphologically indistinguishable, or cryptic, species.

In Europe, five species of *Bombus s. str*. are recognised: *B. cryptarum* (Fabricius), *B. lucorum* (L.), *B. magnus* Vogt, *B. terrestris* (L.) and *B. sporadicus* (Nylander). The taxonomic status of *B. terrestris* and *B. sporadicus* is widely accepted (Williams 1998). Difficulties arise over the other three species: *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*, which are generally grouped to form the *lucorum* complex (*B. lucorum sensu lato*). They are cryptic species that appear very similar in colour and form, particularly as workers or males, and that are often difficult to differentiate morphologically from *B. terrestris* even as queens (Figure 1; Rasmont 1984; Rasmont et al. 1986).

Species classification based on morphological characters may not be suitable for cryptic species and genetic methods may help support species identification. Correct identification is of conservation importance because the taxonomic status of a species must be accurately established in order to assign status and direct conservation efforts (Ryder, 1986; Crandall et

al. 2000). The presence of cryptic species, however, has potentially detrimental implications since reproductively isolated groups should be managed independently of each other (Riddle et al. 2000; Palsbøll et al. 2007). To facilitate identification of cryptic species, molecular methods such as DNA barcoding can be used as a means of designating species on the basis of sequence similarity (Hebert et al. 2003). Such approaches have confirmed the view that cryptic species are particularly common in insects (e.g. Berkov 2002; Hebert et al. 2004).

In pre-DNA based studies, allozyme polymorphisms and variation in male cephalic odour bouquet supported the view that the *lucorum* complex of bumble bees comprise two or three species (reviewed in Bossert 2015). Bertsch et al. (2005) subsequently used mitochondrial cytochrome oxidase I (COI) gene sequences of specimens morphologically well-characterised as queens to show that the *lucorum* complex of bumble bees contained three distinct mitochondrial DNA (mtDNA) lineages in Europe, albeit sampling of two putative species was limited to two specimens apiece at each of two sites in Europe. Using a far larger number of samples from across Europe, including >300 from the island of Ireland, Murray et al. (2008) showed that the three mtDNA lineages exhibited considerable interspecific DNA sequence divergence (≥ 2.3%) at COI compared to intra-taxon sequence variability (≤ 1.3%), with overwhelming support for each lineage, supporting the idea that the three mtDNA lineages represent species: *B. cryptarum*, *B. lucorum s. str*. and *B. magnus*. Williams et al. (2012a) gave the distribution of these three taxa in Europe, Asia and, for *B. cryptarum*, even into North America based on COI sequences, again demonstrating overwhelming statistical support for each of the three COI lineages representing the *lucorum* complex. Murray et al. (2008) also developed a relatively quick and economic restriction enzyme based mtDNA COI marker system based on restriction fragment length polymorphisms (RFLPs) that differentiated among lineages.

Notwithstanding the success of DNA barcoding in separating species and even entire regional bee faunas (Sheffield et al. 2009; Magnacca and Brown 2010a, 2012; Schmidt et al. 2015), mitochondrial lineages may not represent independent, reproductively isolated species. Reasons include the retention of ancestral mitochondrial sequence polymorphisms, mitochondrial introgression and biased inheritance of maternal genetic markers by such factors as *Wolbachia* infection, known to be widespread in bees (Gerth et al. 2011, 2013, 2014; Gerth and Bleidorn 2013; but see Stahlhut et al. 2012), heteroplasmy (e.g., Magnacca and Brown 2010a), and associated tissue segregation of haplotypes (Magnacca and Brown 2010b). Moreover, the use of DNA barcoding or any other mitochondrial DNA marker system does not permit the detection of hybrids between taxa. This is all the more relevant for the *lucorum* complex of bumble bees, in which Carolan et al. (2012) found apparent mis-match between widely employed species-characteristic, discriminatory morphological traits of queens and mtDNA lineage among Irish specimens.

A resolution to this problem is to incorporate multilocus sequence typing into DNA barcoding studies (Gerth and Bleidorn 2013; Bossert 2015), an approach we report here for the *lucorum* complex. We used microsatellites, biparentally inherited nuclear markers, to determine the extent to which the nuclear gene pools of the *lucorum* complex taxa concur with the three mtDNA lineages of Bertsch et al. (2005) and Murray et al. (2008), with the aim of reducing taxonomic uncertainty in this group.

Materials and Methods

Sample collection and DNA extraction

Females (queens and workers) of *B. lucorum s.l.* were collected in 2005 and 2006 from 11 localities spread across the island of Ireland from both rural and urban environments while foraging on flowers (Table 1, Figure 2); they are a subset of the same Irish dataset originally presented in Murray et al. (2008). Individuals were either frozen or stored in 99% ethanol at 4°C prior to DNA extraction from a single leg using 10% Chelex (Walsh et al. 1991) or from half a thorax using a standard high salt protocol (Paxton et al. 1996).

Microsatellite genotyping and species identification by mitochondrial haplotyping

Individuals were genotyped at seven nuclear microsatellite loci (Supplementary Table S1) described in Stolle et al. (2011) and developed for *B. terrestris*. Forward primers included a 19 bp M13 5' tail (CACGACGTTGTAAAACGAC) and reverse primers included a 7 bp 5' tail (GTGTCTT). PCRs were carried out in a total volume of 10 μL containing 1-10 ng genomic DNA, 1 μM of 6-FAM-, TET- or HEX-labelled M13 primer (see Supplementary Table S1), 0.1 μM tailed forward primer, 1 μM reverse primer, 1x PCR reaction buffer (Promega), 200 μM each dNTP, 2.5 mM $MgCl_2$ and 0.25 U GoTaq Flexi DNA polymerase (Promega). PCRs were carried out on a MWG Primus thermal cycler using the following parameters: initial denaturation at 96 °C for 3 min followed by 35 cycles of denaturation at 96 °C for 45 s, annealing at 57 °C for 45 s, extension at 72 °C for 45 s, and a final extension at 72 °C for 3 min. Genotyping was carried out on an AB3730xl capillary genotyping system (Life Technologies; Carlsbad, California, USA). Allele sizes were scored using 500 LIZ size standards and were checked by comparison with previously sized samples.

Microsatellite genotypes were obtained at 5-7 loci for 304 females (54 *B. cryptarum*, 168 *B. lucorum s. str.* and 82 *B. magnus*; Table 1). All had already been classified to mt haplotype by RFLP analysis of a mitochondrial partial COI gene sequence by Murray et al. (2008), the results of which we use (and update by Sanger sequencing of the COI 'barcode' of eight samples) here.

Data analysis

GENEPOP (version 3.4; Raymond and Rousset 1995) was used to test for linkage disequilibrium between nuclear microsatellite loci and for deviation from Hardy-Weinberg equilibrium (HWE) at these loci. We also tested for the presence of null alleles using MICRO-CHECKER (Van Oosterhout et al. 2004). Genetic differentiation within each putative species at microsatellite loci was calculated in MICROSATELLITE ANALYSER (MSA, version 4.05 for OSX) (Dieringer and Schlötterer 1997) and isolation by distance tested using the online web service IBDWS version 3.23 (Jensen et al. 2005).

We used three approaches to determine the fit between mtDNA lineage and multilocus nuclear genotype. In the first approach, genetic clustering of individuals was assessed using a Bayesian procedure implemented in the STRUCTURE software package (version 2.3.3; Pritchard et al. 2000). The program was run without priors, and with or without the admixture ancestry model. Twenty independent runs were carried out for each model and value of K, the number of genetic clusters, from $K = 1$ to $K = 3$. Our rationale was to test the hypothesis of $K = 3$ clusters (representing the three species: *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*) versus a null hypothesis of $K = 1$ or $K = 2$ clusters (species). Each Markov chain Monte Carlo analysis used a burn-in of 50,000 followed by a further 500,000 iterations. STRUCTURE's Q

149 value, a probability of group membership, was calculated for each individual at $K = 3$ using
150 the admixture ancestry model.
151
152 Because our dataset suggested deviation from HWE (see results) whereas HWE is an
153 assumption of STRUCTURE, we employed two distance-based methods to test for the
154 association between genotypes and mitochondrial haplotypes, methods that do not make
155 assumptions about mating structure and that do not make *a priori* assumptions about group
156 membership. In the second approach, we visualised relationships among multilocus
157 microsatellite genotypes of the 304 females of the *lucorum* complex using principle
158 coordinate analysis (PCoA) in GENALEX version 6.5 (Peakall and Smouse 2006). In the third
159 approach, we used Discriminant Analysis of Principal Components (DAPC; Jombart et al.
160 2010) to cluster genotypes independently of *a priori* haplotype designation using the R
161 package *adegenet* version 1.4.2 (Jombart 2008) in R version 3.1.0 (R Core Team 2014). For
162 DAPC, we examined results after extracting 5, 10, 20 or 40 principal components from the
163 genotype data.
164
165 DNA sequencing to improve mt RFLP-based haplotyping
166
167 STRUCTURE *Q* values suggested that three individuals were in a different genotypic cluster to
168 that of the other individuals with the same mt RFLP haplotype (Supplementary Table S2).
169 Preliminary visualisation of the PCoA suggested that two of these three individuals and five
170 additional individuals were in a different genotypic cluster to those with the same mt RFLP
171 haplotype (Supplementary Table S2). All eight aberrant individuals were sequenced at the
172 COI 'barcode' (Hebert et al. 2003) and identified by a web-based BLASTn search against the
173 entire NCBI nucleotide database.

The original (in Murray et al. 2008) mitochondrial RFLP classification for four of these eight individuals was incorrect; two individuals with *lucorum* RFLP patterns had *cryptarum* COI DNA sequences and two individuals with *cryptarum* RFLP patterns had *magnus* COI DNA sequences (Supplementary Table S2). Though error rates in defining an individual's mt lineage using RFLPs were likely low, they nevertheless call into question the value of using RFLPs to define unambiguously the mt haplotype, as has been proposed for the *lucorum* complex of bumble bees (Murray et al. 2008; Versterlund et al. 2014). We suggest that DNA sequencing of the COI barcode is a more reliable method of defining the mt lineage in the *lucorum* complex of bumble bees in Europe and likely in other taxa, too. We recommend Sanger sequencing rather than RFLP-based inference of haplotypes in future studies of the *lucorum* complex.

The final, updated dataset is presented in Table 1 and in the Results section below.

Results

Approximately 10% of samples were duplicated per 96-well plate for PCR; duplicates gave identical microsatellite genotypes, suggesting very low rates of error in amplifying and calling genotypes. No consistent linkage disequilibrium (i.e. involving the same loci) was detected between any of the seven nuclear microsatellites analysed across the three putative species of the *lucorum* complex (Supplementary Table S3).

When individuals from all 11 populations were lumped together into their three putative species, *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*, loci 255 and 278 exhibited deviation

199 from Hardy-Weinberg equilibrium in all three putative species and loci 198, 331 and 554
200 deviated in two putative species (Supplementary Table S4). In most of these cases, there was
201 evidence from MICRO-CHECKER for null alleles as the cause of the deviation (Supplementary
202 Table S4).
203
204 Lumping individuals from different populations into a single group could lead to deviation
205 from HWE and evidence for null alleles due to the Wahlund effect. To explore this
206 possibility, we tested for deviation from HWE using GENEPOP and for evidence of null
207 alleles using MICRO-CHECKER by testing each locus in each putative species at each sampling
208 location separately (Supplementary Tables S4.1-S4.11). In the majority of cases (147 of 172
209 locus by species by locality combinations), genotypes did not deviate from HWE and there
210 was no evidence of null alleles. These results suggest that all three putative species are
211 regular outbreeders and that deviation from HWE was a consequence of having lumped
212 individuals from different populations.
213
214 Interestingly, when we analysed deviation from HWE and sought evidence for null alleles by
215 lumping individuals from different putative species into a single taxon, *B. lucorum s. lato*,
216 either across all sampling localities (Supplementary Table S4) or for each sampling locality
217 separately (Supplementary Tables S4.1-S4.11), loci were often out of HWE and showed
218 evidence of null alleles (57 of 83 locus by location comparisons). These results provide a hint
219 that the three putative species are differentiated in sympatry.
220
221 We tested for population genetic differentiation for each putative species separately across
222 sampling localities for sampling site with $n \geq 5$ individuals. For each putative species,
223 differentiation across Ireland (Figure 2) was subtle, not significantly different from zero for

224 *B. cryptarum* (6 locations, global F_{ST} = 0.015, P = 0.109) but significant for *B. lucorum s. str.*
225 (9 locations, global F_{ST} = 0.008, P = 0.036) and *B. magnus* (8 locations, global F_{ST} = 0.018, P
226 = 0.023). For each putative species, Isolation by Distance was not significant (statistics and
227 population pairwise F_{ST}, Supplementary Table S5), probably due to low statistical power
228 (lack of sampling sites).
229
230 Differentiation between the three putative species was high (global F_{ST} = 0.268, P < 0.001);
231 all three putative species pairs were significantly differentiated (for all three pairwise
232 comparisons, F_{ST} > 0.2, P < 0.001). There was no suggestion in the F_{ST} values that *B.*
233 *cryptarum* was closer to *B. magnus* than either was to *B. lucorum s. str.*. However, when we
234 lumped the three putative species into one taxon, *B. lucorum s. lato*, differentiation across our
235 11 sampling sites in Ireland was insignificant (F_{ST} = 0.046, n.s.). These results suggest that
236 the three putative species are genetically well differentiated.
237
238 Differences in allele frequencies of *B. lucorum s. str.* to the other two taxa were particularly
239 evident at locus 327 (Figure 3). Allele frequencies also differed markedly in *B. magnus*
240 compared to the other two putative species at locus 331 (Figure 3). Yet not one allele at any
241 of the loci was both private (restricted to a putative species) and at a sufficiently high
242 frequency within that species to allow it to be used to discriminate readily between species.
243
244 STRUCTURE analysis at K = 2 (mean likelihood Ln Pr (X|K) = -5668.6; admixture ancestry
245 model) revealed the mitochondrial lineage corresponding to *B. lucorum s. str.* to be well
246 differentiated from *B. cryptarum* and *B. magnus* (Figure 4), which may reflect the closer
247 phylogenetic affinity of the latter pair of species than either of them to *B. lucorum s. str.*
248 (Murray et al. 2008). The nuclear gene pools of the three putative species were clearly

249 separated at $K = 3$ (mean likelihood Ln Pr (X|K) = -5232.6**;** Figure 4), with greater model
250 support than for $K = 1$ (mean likelihood Ln Pr (X|K) = -6880.2) or $K = 2$. Results were
251 qualitatively the same using STRUCTURE's non-admixture model (Supplementary Figure S1).
252
253 Multilocus microsatellite genotypes of one out of the 304 individuals did not concur with the
254 COI mtDNA species delineation; an individual with a *magnus* mitochondrial haplotype was
255 assigned to the *cryptarum* nuclear gene pool cluster (individual Ff53, Supplementary Table
256 S6). Two additional individuals exhibited a major split in their nuclear gene pool assignments
257 between two putative species (individuals Ff19 and Ff27: Q value ≤ 0.5; Supplementary
258 Tables S6 and S7). STRUCTURE assignment of the other 301 individuals to their correct
259 mitochondrial lineage was generally with high posterior probability (Q value >0.93); only 21
260 of the other 301 individuals (~8%) analysed exhibited a major assignment (Q) value of < 0.9.
261 These results lend weight to the hypothesis that the *lucorum* complex comprises three
262 species, with good fit of multilocus nuclear genotypes to mitochondrial haplotypes and few
263 exceptions.
264
265 Because STRUCTURE makes the strong assumption that genotypic data are in HWE, we
266 repeated analyses by removing three loci that suggested marked deviation from HWE: loci
267 255, 331 and 198 (Supplementary Table S4). Results from STRUCTURE analyses with only
268 four loci gave largely similar results to those with the entire dataset (Supplementary Figure
269 S2).
270
271 Genotypes of the three mtDNA lineages each formed a separate cluster when mapped in
272 multivariate space by PCoA, with only slight overlap at the edges of clusters (Figure 5).
273 Seven of 304 individuals did not concur with COI mtDNA species delineation. These

included the three individuals (Ff19, Ff27, Ff53) whose STRUCTURE assignment suggested their genotypes did not fit with other members of the same mitochondrial lineage (Supplementary Table S7).

Clustering genotypes by DAPC also revealed three discrete clusters that largely concurred with mitochondrial lineages (20 PCs extracted from the genotype data, Figure 6), providing additional support for the hypothesis that *B. cryptarum*, *B. lucorum s. str.* and *B. magnus* are discrete species. Five of 304 individuals were at the multivariate edge of mtDNA lineages. These also included the same three aberrant individuals (Ff19, Ff27, Ff53) highlighted by STRUCTURE and PCoA (Supplementary Table S7). The same five individuals were identified as outliers when 5, 10 or 40 PCs were extracted from the genotype data for DAPC (generating 7, 7 and 5 outlier individuals respectively).

Though STRUCTURE analysis suggested *B. cyrptarum* and *B. magnus* are genetically closer to each other than either is to *B. lucorum s. str.* (Figure 4), PCoA and DAPC analyses did not support this view. Using the multivariate distance-based approaches, all three putative taxa were similarly differentiated (Figures 5 and 6).

Discussion

Our multilocus nuclear (microsatellite) data of the *lucorum* complex of bumble bees collected in Ireland were grouped into three discrete multivariate clusters that corresponded well to the three mitochondrial lineages formerly assigned to the putative species *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*. Our data lend weight to the hypothesis, based on mtDNA COI

298 partial sequences, that the *lucorum* complex comprises three morphologically cryptic but
299 reproductively isolated species: *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*.
300
301 Mallet (1995, 2007) has persuasively argued that, in the age of genetics and genomics, a
302 robust species definition for sexually reproducing organisms is that, when in sympatry, they
303 form discrete genotypic clusters. The multilocus genetic differentiation of the three *lucorum*
304 complex mitochondrial lineages that we found in our STRUCTURE analyses at $K = 3$ and using
305 PCoA and DAPC is consistent with reciprocal monophyly of three species in sympatry.
306
307 Carstens et al. (2013) have argued that robust species delimitation should be based upon
308 multiple, independent analyses. Here we employed three methods to differentiate nuclear
309 gene pools of the *lucorum* complex, all of which were consistent in their identification of
310 three discrete groups, with few outliers. STRUCTURE in particular makes the assumption of
311 HWE, which, in our case, was violated at several loci when individuals from multiple
312 locations were lumped together within a putative species. That STRUCTURE nevertheless
313 identified the same outliers as PCoA and DAPC suggests it may be robust to given degree of
314 violation of HWE in our dataset.
315
316 Though our nuclear microsatellite marker dataset suggests reproductive isolation between the
317 three taxa of the *lucorum* complex, eight of 304 individuals possessed a multilocus genotype
318 that did not concur with their mitochondrial haplotype. For three of these eight individuals,
319 all three methods (STRUCTURE, PCoA and DAPC) placed them in a cluster different to that of
320 other members of their mitochondrial lineage. One explanation for this incongruence is that
321 genotypic clusters of different species may overlap at their edges when based on a reduced
322 number of loci; use of additional microsatellite loci or genome-wide markers may resolve this

323 'lack of data' issue (Lozier and Zayed 2016). Secondly, a low degree of hybridization
324 between species may take place in the field. Individuals Ff19 and Ff27 could represent
325 hybrids because their probability of assignment (STRUCTURE *Q* value) was intermediate
326 between their mitochondrial lineage and that of another lineage. Artificial crosses between
327 members of the *lucorum* complex are needed to support this suggestion. Thirdly, aberrant
328 individuals might be a product of mitochondrial introgression; individual Ff53 could
329 represent such a case, with a *magnus* mitochondrial haplotype confirmed by Sanger
330 sequencing and a multilocus nuclear genotype assigned to *cryptarum*, *cryptarum/lucorum* and
331 *cryptarum* by STRUCTURE, PCoA and DAPC respectively. It may be of significance that all
332 three consistently aberrant individuals (Ff19, Ff27, Ff53) were collected at one site, Slieve
333 Gullion, as workers (Table 1). We note, though, that the three discrete multilocus genotypic
334 clusters we detected are unlikely to be maintained if hybridization or mtDNA introgression
335 were common and widespread. Thus our second and third explanations seem unlikely to
336 account for the mis-assigned individuals. Alternatively, if they do occur, they may not lead to
337 fertile sexual descendants (queens and males).
338
339 If, as seems likely, our first explanation is correct, it suggests that considerable effort will be
340 required for microsatellites to be used to separate among cryptic species or to detect
341 hybridization. Within European bees, there are many putative cryptic species pairs or cryptic
342 species complexes that share COI barcodes (Schmidt et al. 2015). Interspecific DNA
343 sequence divergence at COI of the *lucorum* complex in Ireland is >2.3% (Murray et al. 2008),
344 yet the 7 microsatellite loci we used to analyse 54-168 individuals per taxon were insufficient
345 to resolve unambiguously all 304 individuals. For comparison, in the cryptic Neotropical
346 orchid bee sister species *Euglossa dilemma* and *Euglossa viridissima*, species-typical alleles
347 at one locus have been found to differentiate between taxa (Eltz et al. 2011), though not

348 unambiguously. Indeed, as the number of analysed individuals per taxon increases, so too is
349 the likelihood of detecting greater allelic diversity, reducing the probability of finding private,
350 species-diagnostic alleles at highly variable loci.
351
352 Lack of resolution in separating between reproductively isolated nuclear gene pools using
353 microsatellites might be due to shared ancestral polymorphisms and homoplasy caused by
354 high mutation rates at microsatellite loci (e.g. Schlötterer 1998). Sequence divergence of
355 other cryptic species complexes of bee at COI is considerably less than 2.3%, suggesting a
356 more recent common ancestor than that of the *lucorum* complex; for example, three members
357 of the *Colletes succinctus* complex (species *hederae, halophilus* and *succinctus*) share the
358 same COI barcode (Kuhlmann et al. 2007). For these and other closely related species, a
359 larger sample size of sympatric individuals and, possibly more importantly, a larger number
360 of nuclear loci may be needed to separate unambiguously among reproductively isolated
361 nuclear gene pools (e.g. using deep sequencing via genome skimming, Cossiac et al. 2016),
362 calling into question the feasibility of using microsatellites for multilocus sequence typing so
363 as to differentiate readily among cryptic species.
364
365 Species-specific pheromones may play an important role in mate-recognition, presumably
366 decreasing the incidence of interspecific mating (Paterson 1985). *Bombus* male sex
367 pheromones contain over 50 different volatile compounds derived from the labial glands that
368 are used to scent-mark substrates and that are thought to attract unmated conspecific queens
369 (Bergström et al. 1981; Ayasse et al. 2001). It has been previously demonstrated that males of
370 *B. cryptarum*, *B. lucorum s. str.* and *B. magnus* differ in their cephalic secretions to the extent
371 that their multivariate chemical composition has been used to support specific classification
372 of the three taxa (Bertsch 1997a; Bertsch et al. 2004, 2005; Pamilo et al. 1997). It is plausible

373 that mate recognition based on male labial secretions plays a significant role as a prezygotic
374 isolating mechanism in maintaining species boundaries between members of the *lucorum*
375 complex.
376
377 The existence of three cryptic species within the *lucorum* complex of *Bombus* has several
378 implications for conservation and management. In terms of conservation *per se*, members of
379 the *lucorum* complex, like *B. terrestris*, are classified as of 'least concern' in Europe (Nieto et
380 al. 2014). Nevertheless, it is highly likely that these cryptic species have been previously
381 overlooked, and treatment of them as such should be borne in mind when addressing
382 conservation status, particularly in light of the on-going declines in bumble bees in Europe
383 (Goulson et al. 2005; Fitzpatrick et al. 2007; Rasmont et al. 2015). Indeed, in Ireland, *B.*
384 *lucorum s. str.* is classified as 'least concern' whereas *B. cryptarum* and *B. magnus* are more
385 cautiously and appropriately classified as 'data deficient' (Fitzpatrick et al. 2006). The three
386 *lucorum* complex species may exhibit ecological specialization. With regard to altitude, *B.*
387 *cryptarum* is has more of an upland distribution in Ireland (Murray et al. 2008) whereas *B.*
388 *magnus* is considered a lowland species in Germany (von Hagen 2003). More recent analysis
389 of their ecological associations has suggested that *B. cryptarum* and *B. magnus* prefer cooler
390 sites in comparison to *B. lucorum* (Walters et al. 2010; Scriven et al. 2015). Phenological
391 differences also exist; *B. cryptarum* is an early species that precedes *B. lucorum s. str.* and *B.*
392 *magnus* (Bertsch 1997b; Bertsch et al. 2004), which may also play a role in preventing
393 hybridization.
394
395 From a management perspective, it is necessary to be able to identify taxa correctly because
396 bumble bees are important managed crop pollinators and colonies are increasingly being
397 transplanted long distances to provide pollination services (Goulson 2003). This has become

of particular conservation concern since bumble bee translocations have been implicated in colony declines, including through pathogen spillover, and in the replacement of native *Bombus* species (Inoue et al. 2008; Meeus et al. 2011; Schmid-Hempel et al. 2014). Indeed, in Asia, confusion remains over the identification of bumble bee species imported for crop pollination (Williams et al. 2012b). Importation of non-native bumble bee species not only brings risks associated with the introduction of a competitively superior congener and of a non-native's pathogens, there is also the risk of hybridization between native and introduced species. We note that inter-specific mating and hybridization can only be detected by using codominant nuclear markers such as microsatellites or SNPs; the former have been employed to demonstrate interspecific mating and hybrid inviability in crosses between native Japanese bumble bees and imported European *B. terrestris* (Kanbe et al. 2008; Tsuchida et al. 2010).

The utility of DNA barcoding versus morphological-based taxonomy in biodiversity inventorying has been hotly debated (Packer et al. 2009; Stahlhut et al. 2012; Gerth and Bleidorn 2013). Yet DNA barcodes continue to be used for species identification and even the characterization of new *Bombus* species (Williams et al. 2016). Our data vindicate their use for species identification within the *lucorum* complex.

Acknowledgements We thank friends and colleagues who helped to collect bumble bees across Ireland: D. Cookson, D. Dominoni, M. Kelly and S. Roos; Andreas Bertsch for use of his photographs, comments on this manuscript and encouragement to engage with the *lucorum* complex; and Robin Moritz for laboratory and intellectual support. We also thank two anonymous reviewers and editor-in-chief as well as Shalene Jha and Christophe Praz for many insightful comments that helped improve the manuscript.

Funding This work was supported by a grant from the Higher Education Authority of Ireland as part of its North-South Research Programme for Peace and Reconciliation. L McKendrick thanks DARD for their financial support (a PhD stipend) and patience.

Conflict of Interest The authors declare that they have no conflict of interest.

References

Ayasse M, Paxton RJ, Tengö J (2001) Mating behavior and chemical communication in the order Hymenoptera. Ann Rev Entomol 46:31-78.

Bergström G, Svensson BG, Appelgren M, Groth I (1981) Complexity of bumble bee marking pheromones: biochemical, ecological and systematic interpretations. pp.175-183 in: Hows PE, Clemet J-L (eds.) Biosystematics of Social Insects Vol. 19, Academic Press, London, UK.

Berkov A (2002) The impact of redefined species limits in *Palame* (Coleoptera: Cerambycidae: Lamiinae: Acanthocinini) on assessments of host, seasonal, and stratum specificity. Biol J Linn Soc 76:195-209.

Bertsch A (1997a) Abgrenzung der Hummel-Arten *Bombus cryptarum* und *B. lucorum* mittels männlicher Labialdrüsen-Sekrete und morphologischer Merkmale (Hymenoptera, Apidae). Entomologia Generalis 22:129-145.

Bertsch A (1997b) Wieviele Arten der Untergattung *Terrestribombus* (Hymenoptera, Apidae) gibt es in Nordhessen; die Abgrenzung von *Bombus cryptarum* und *B. lucorum* mittels

männlicher Labial-Drüsen-Sekrete und morphologischer Merkmale. Marburger Entomologische Publikationen 2:1-28.

Bertsch A, Schweer H, Titze A (2004) Discrimination of the bumblebee species *Bombus lucorum, B. cryptarum* and *B. magnus* by morphological characters and male labial gland secretions. Beiträge zur Entomol 54:365-386.

Bertsch A, Schweer H, Titze A, Tanaka H (2005) Male labial gland secretions and mitochondrial DNA markers support species status of *Bombus cryptarum* and *B. magnus* (Hymenoptera, Apidae). Insectes Soc 52:45-54.

Bossert S (2015) Recognition and identification of bumblebee species in the *Bombus lucorum*-complex (Hymenoptera, Apidae) – A review and outlook. Dtsch Entomol Z 62:19-28.

Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL (2011) Patterns of widespread decline in North American bumble bees. Proc Nat Acad Sci USA 108:662–667.

Carolan JC, Murray TE, Fitzpatrick Ú, Crossley J, Schmidt H, Cederberg B, McNally L, Paxton RJ, Williams PH, Brown MJF (2012) Colour patterns do not diagnose species: quantitative evaluation of a DNA barcoded cryptic bumblebee complex. PLoS ONE 7:e29251.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. Mol Ecol 22:4369-4383.

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol 25:1423-1428.

467 Crandall KA, Binindaemonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary
468 processes in conservation biology. Trends Ecol Evol 15:290-295.

469 Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform
470 independent analysis tool for large microsatellite data sets. Mol Ecol Notes 3:167-169.

471 Eltz T, Fritzsch F, Pech JR, Zimmermann Y, Ramírez SR, Quezada-Euan JJG, Bembé B
472 (2011) Characterization of the orchid bee E*uglossa viridissima* (Apidae: Euglossini) and a
473 novel cryptic sibling species, by morphological, chemical, and genetic characters. Zool J
474 Linn Soc 163:1064-1076.

475 Fitzpatrick Ú, Murray TE, Byrne A, Paxton RJ, Brown MJF (2006) Regional red list of Irish
476 bees. Dublin, Ireland: National Parks and Wildlife Service (Republic of Ireland) and
477 Environment and Heritage Service (Northern Ireland). 1-38.

478 Fitzpatrick Ú, Murray TE, Paxton RJ, Breen J, Cotton D, Santorum V, Brown MJF (2007)
479 Rarity and decline in bumblebees - a test of causes and correlates in the Irish fauna. Biol
480 Conserv 136:185-194.

481 Gerth M, Bleidorn C (2013) A multilocus sequence typing (MLST) approach to diminish the
482 problems that are associated with DNA barcoding: A reply to Stahlhut et al. (2012).
483 Systematics and Biodiversity 11:1-3.

484 Gerth M, Gansauge M-T, Weigert A, Bleidorn C (2014) Phylogenomic analyses uncover
485 origin and spread of the *Wolbachia* pandemic. Nature Comm 5:5117.

486 Gerth M, Geißler A, Bleidorn C (2011) *Wolbachia* infections in bees (Anthophila) and
487 possible implications for DNA barcoding. Systematics and Biodiversity 9:319-327.

488 Gerth M, Röthe J, Bleidorn C (2013) Tracing horizontal *Wolbachia* movements among bees

(Anthophila): a combined approach using multilocus sequence typing data and host phylogeny. Mol Ecol 22:6149-6162.

Goulson D (2003) Effects of introduced bees on native ecosystems. Ann Rev Ecol Evol Syst 34:1-26.

Goulson D (2009) Bumblebees. Behaviour, Ecology and Conservation. 2nd edn. Oxford University Press, Oxford, UK.

Goulson D, Hanley ME, Darvill B, Ellis JS, Knight ME (2005) Causes of rarity in bumblebees. Biol Conserv 122:1-8.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Roy Soc Lond B 270:313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Nat Acad Sci USA 101:14812-14817.

Hines HM (2008) Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). Syst Biol 57:58-75.

Inoue MN, Yokoyama J, Washitani I (2008) Displacement of Japanese native bumblebees by the recently introduced *Bombus terrestris* (L.) (Hymenoptera: Apidae). J Insect Cons 12:135-146.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:13.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

511 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a
512 new method for the analysis of genetically structured populations. BMC Genet 11:94.

513 Kanbe Y, Okada I, Yoneda M, Goka K, Tsuchida K (2008) Interspecific mating of the
514 introduced bumblebee *Bombus terrestris* and the native Japanese bumblebee *Bombus*
515 *hypocrita sapporoensis* results in inviable hybrids. Naturwiss 95:1003-1008.

516 Kuhlmann M, Else GR, Dawson A, Quicke DLJ (2007) Molecular, biogeographical and
517 phenological evidence for the existence of three western European sibling species in the
518 *Colletes succinctus* group (Hymenoptera: Apidae). Organisms Divers Evol 7:155-165.

519 Lozier JD, Zayed A (2016) Bee conservation in the age of genomics. Conserv Genet this
520 special issue.

521 Magnacca KN, Brown MJF (2010a) Mitochondrial heteroplasmy and DNA barcoding in
522 Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). BMC Evol Biol
523 10:174.

524 Magnacca KN, Brown MJF (2010b) Tissue segregation of mitochondrial haplotypes in
525 heteroplasmic Hawaiian bees: implications for DNA barcoding. Mol Ecol Res 10:60-68.

526 Magnacca KN, Brown MJF (2012) DNA barcoding a regional fauna: Irish solitary bees. Mol
527 Ecol Res 12:990-998.

528 Mallet J (1995) A species definition for the Modern Synthesis. Trends Ecol Evol 10:294-299.

529 Mallet J (2007) Species, concepts of. In: Levin SA et al., eds. Encyclopedia of Biodiversity.
530 2nd ed. San Diego, California, USA: Academic Press. 427–440.

531 Meeus I, Brown MJF, De Graaf DC, Smagghe G (2011) Effects of invasive parasites on
532 bumble bee declines. Conserv Biol 25:662-671.

533 Murray TE, Fitzpatrick Ú, Brown MJF, Paxton RJ (2008) Cryptic species diversity in a
534 widespread bumble bee complex revealed using mitochondrial DNA RFLPs. Conserv
535 Genet 9:653-666.

536 Nieto A, Roberts SPM, Kemp J, Rasmont P, Kuhlmann M, García Criado M, Biesmeijer JC,
537 Bogusch, Dathe HH, De la Rúa P, De Meulemeester T, Dehon M, Dewulf A, Ortiz-
538 Sánchez FJ, Lhomme P, Pauly A, Potts SG, Praz C, Quaranta M, Radchenko VG,
539 Scheuchl E, Smit J, Straka J, Terzo M, Tomozii B, Window J, Michez D (2014) European
540 Red List of Bees. Publication Office of the European Commission: Luxembourg.

541 Packer L, Gibbs J, Sheffield CS, Hanner R (2009) DNA barcoding and the mediocrity of
542 morphology. Mol Ecol Res 9:42-50.

543 Palsbøll PJ, Berube M, Allendorf FW (2007) Identification of management units using
544 population genetic data. Trends Ecol Evol 22:11-16.

545 Pamilo P, Tengö J, Rasmont P, Pirhonen K, Pekkarinen A, Kaarnama E (1997) Pheromonal
546 and enzyme genetic characteristics of the *Bombus lucorum* species complex in northern
547 Europe. Entomol Fennici 14:187-194.

548 Paterson HE (1985) The recognition concept of species. In: Vrba ES (ed.) Species and
549 Speciation. Transvaal Museum Monographs 4:21-29.

550 Paxton RJ, Thoren PA, Tengo J, Estoup A, Pamilo P (1996) Mating structure and nestmate
551 relatedness in a communal bee, *Andrena jacobi* (Hymenoptera, Andrenidae), using
552 microsatellites. Mol Ecol 5:511-519.

553 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic
554 software for teaching and research. Mol Ecol Notes 6:288-295.

555 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
556 multilocus genotype data. Genetics 155:945-959.

557 R Core Team (2014). R: A language and environment for statistical computing. R Foundation
558 for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

559 Rasmont P (1984) Les bourdons du genre *Bombus* Latreille *sensu stricto* en Europe
560 occidentale et centrale. Spixiana 7:135-160.

561 Rasmont P, Franzén M, Lecocq T, Harpke A, Roberts S, Biesmeijer JC, Castro L, Cederberg
562 B, Dvorak L, Fitzpatrick Ú, Gonseth Y, Haubruge E, Mahé G, Manino A, Michez D,
563 Neumayer J, Ødegaard F, Paukkunen J, Pawlikowski T, Potts S, Reemer M, Settele J,
564 Straka J, Schweiger O (2015) Climatic risk and distribution atlas of European bumblebees.
565 BioRisk 10:1-236.

566 Rasmont P, Scholl A, de Jonghe R, Obrecht E, Adamski A (1986) Identité et variabilité des
567 mâles de bourdons du genre *Bombus* Latreille *sensu stricto* en Europe occidentale et
568 centrale (Hymenoptera, Apidae, Bombinae). Rev Suisse Zool 93:661-682.

569 Raymond M, Rousset F (1995) GENEPOP (v. 1.2): Population genetic software for exact tests
570 and ecumenicism. J Hered 86:248–249.

571 Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical
572 assembly of a Baja California Peninsular Desert biota. Proc Nat Acad Sci USA 97:14438-
573 14443.

574 Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. Trends
575 Ecol Evol 1:9-10.

576 Schlötterer C (1998) Microsatellites. In: Molecular Genetic Analysis of Populations. A

577 Practical Approach (ed. Hoelzel AR), pp. 237-261. Oxford University Press, Oxford, UK.

578 Schmid-Hempel R, Eckhardt M, Goulson D, Heinzmann D, Lange C, Plischuk S, Escudero
579 LR, Salathé R, Scriven JJ, Schmid-Hempel P (2014) The invasion of southern South
580 America by imported bumblebees and associated parasites. J Anim Ecol 83:823-837.

581 Scriven JJ, Woodall LC, Tinsley MC, Knight ME, Williams PH, Carolan JC, Brown MJF,
582 Goulson D (2015) Revealing the hidden niches of cryptic bumblebees in Great Britain:
583 implications for conservation. Biol Conserv 182:126-133.

584 Schmidt S, Schmid-Egger C, Morinière J, Haszprunar G, Hebert PDN (2015) DNA barcoding
585 largely supports 250 years of classical taxonomy: identifications for Central European
586 bees (Hymenoptera, Apoidea *partim*). Mol Ecol Res 15:985-1000.

587 Sheffield CS, Hebert PDN, Kevan PG, Packer L (2009) DNA barcoding a regional bee
588 (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol Ecol Res
589 9:196-207.

590 Stahlhut JK, Gibbs J, Sheffield CS, Alex Smith M, Packer L (2012) *Wolbachia*
591 (Rickettsiales) infections and bee (Apoidea) barcoding: a response to Gerth et al.
592 Systematics and Biodiversity 10: 395-401.

593 Stolle E, Wilfert L, Schmid-Hempel L, Schmid-Hempel P, Kube M, Reinhardt R, Moritz
594 RFA (2011) A second generation genetic map of the bumblebee *Bombus terrestris*
595 (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. BMC
596 Genomics 12:48.

597 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular
598 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
599 maximum parsimony methods. Mol Biol Evol 28:2731-2739.

600 Tsuchida K, Ito Kondo N, Inoue MN, Goka K (2010) Reproductive disturbance risks to
601 indigenous Japanese bumblebees from introduced *Bombus terrestris.* Appl Entomol Zool
602 45:49-58.

603 Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER:
604 software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol
605 Notes 4:535-538.

606 Vesterlund SR, Sorvari J, Vasemägi A (2014) Molecular identification of cryptic bumblebee
607 species from degraded samples using PCR–RFLP approach. Mol Ecol Res 14:122-126.

608 von Hagen E (2003) Hummeln: Bestimmen, Ansiedeln, Vermehren, Schützen. Fauna-Verlag,
609 Nottuln, Germany.

610 Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of
611 DNA for PCR-based typing from forensic material. Biotechniques 10:506-513.

612 Waters J, Darvill B, Lye GC, Goulson D (2011) Niche differentiation of a cryptic bumblebee
613 complex in the Western Isles of Scotland. Insect Conserv Div 4:46-52.

614 Williams PH (1998) An annotated checklist of bumble bees with an analysis of patterns of
615 description (Hymenoptera: Apidae, Bombini). Bull Natural History Mus Lond (Entomol)
616 67:79-152.

617 Williams PH, Cameron SA, Hines HM, Cederberg B, Rasmont P (2008) A simplified
618 subgeneric classification of the bumblebees (genus *Bombus*). Apidol 39:46-74.

619 Williams PH, Brown MJF, Carolan JC, An J, Goulson D, Aytekin AM, Best LR, Byvaltsev
620 AM, Cederberg B, Dawson R, Huang J, Ito M, Monfared A, Raina RH, Schmid-Hempel
621 P, Sheffield CS, Šima P, Xie Z (2012a) Unveiling cryptic species of the bumblebee

622 subgenus *Bombus s. str.* worldwide with COI barcodes (Hymenoptera: Apidae).
623 Systematics and Biodiversity 10:21-56.

624 Williams PH, An J, Brown MJF, Carolan JC, Goulson D, Huang J, Ito M (2012b) Cryptic
625 bumblebee species: consequences for conservation and the trade in greenhouse pollinators.
626 PLoS ONE 7:e32992.

627 Williams PH, Cannings SG, Sheffield CS (2016) Cryptic subarctic diversity: a new
628 bumblebee species from the Yukon and Alaska (Hymenoptera: Apidae). J Nat Hist:1-13.
629 DOI: 10.1080/00222933.2016.1214294

630 **Figure legends**

631

632 **Fig. 1** Queens of the four Irish members of the *Bombus s. str.* group, namely *B. terrestris* and
633 the three members of the *lucorum* complex: *B. cryptarum*, *B. lucorum s. str.* and *B. magnus*
634 (photo credit: Andreas Bertsch 2004)

635

636 **Fig. 2** Sample sites (numbers correspond to those in Table 1) and numbers of individuals
637 genotyped at seven microsatellite loci of the *lucorum* complex of bumble bees. Species
638 designation was based on mtDNA COI RFLPs and updated by Sanger sequencing

639

640 **Fig. 3** Barplot of STRUCTURE (using the admixture ancestry model) output showing
641 percentage assignment of individuals of the three *lucorum* complex species of bumble bees
642 genotyped at seven microsatellite loci to a given putative species for $K = 2$ and $K = 3$ (species
643 designation based on mtDNA COI haplotypes)

644

645 **Fig. 4** Bubble plots of allele frequencies at seven microsatellite loci of 304 individuals of the
646 three putative species of the *lucorum* complex of bumble bees from 11 sites in Ireland;
647 species designation was based on mtDNA COI haplotypes updated by Sanger sequencing and
648 bubble diameters reflect allele frequencies which, within a species (column), sum to one

649

650 **Fig. 5** PCoA of the multilocus microsatellite genotypes of 304 *lucorum* complex bumble bees
651 from Ireland; each of the three mtDNA lineages is coded by a different shading; the three
652 individuals (Ff19, Ff27, Ff53) with low multilocus group membership (low STRUCTURE Q
653 value) to others of the same mtDNA lineage are circled in grey

654

Fig. 6 DAPC of the multilocus microsatellite genotypes of 304 *lucorum* complex bumble bees from Ireland; each of the three mtDNA lineages is coded by a different shading; the three individuals (Ff19, Ff27, Ff53) with low multilocus group membership (low STRUCTURE *Q* value) to others of the same mtDNA lineage are circled in grey