1 Molecular identification and distribution of leatherjackets (Diptera:Tipulidae) in UK
2 agricultural grassland.
3
4 Running title: Identification of leatherjackets in grassland
5
6 Benefer, C.M.[1*], D'Ahmed, K.S.[2], Murray, P.J.[3], Blackshaw, R.P.[4]
7 *Corresponding author: Email: carly.benefer@plymouth.ac.uk, telephone: +44 1752
8 584483, fax: +44 1752 584605.
9
10
11
12
13
14
15 [1]School of Biological Sciences, Plymouth University, Drake Circus, Plymouth, Devon,
16 PL4 8AA, UK.
17 [2]Department of Plant Protection, Salahaddin University–Erbil, Erbil, Iraq.
18 [3]Sustainable Soils and Grassland Systems Department, Rothamsted Research -
19 North Wyke, Okehampton, Devon, EX20 2SB, UK.
20 [4]Blackshaw Research and Consultancy, Parade, Chudleigh, Devon, TQ13 0JF, UK.

Abstract

1. DNA barcoding is useful for the identification of morphologically cryptic invertebrates. An important application is for pest species, for which it is critical to determine the distribution, biology and ecology of damaging life-stages in order to target management effectively.

2. *Tipula paludosa* (Meigen) and *Tipula oleracea* (Linnaeus) leatherjackets, the larvae of crane flies (Diptera: Tipulidae), are pests of agricultural grassland in Europe and the USA, and difficult to distinguish morphologically.

3. We used cytochrome c oxidase subunit I (COI) barcoding to identify leatherjackets from 19 permanent grassland fields over 2 sampling seasons on the Rothamsted Research North Wyke Farm Platform, south west UK, to assess species-level distribution and genetic diversity.

4. Most larvae (94%) were found to be *T. paludosa*, comprising 18 haplotypes which were spread across the sampling site in a panmictic population. However, *T. oleracea* were found in low abundance (3% of larval samples) and only in the second year of sampling. Other morphologically similar Dipteran larvae (3%) were also found.

5. This dominance of one species suggests there may be underlying differences in species biology, such as choice of oviposition site and dispersal ability, in agreement with other studies, and has implications for monitoring and management.

Keywords: DNA sequencing, COI barcoding, soil insects, soil pests, belowground, grassland, haplotype analysis.

43 **1. Introduction**

44 *Tipula paludosa* (Meigen) and *Tipula oleracea* (Linnaeus) are polyphagous
45 agricultural pests, mainly inhabiting permanent pasture but also causing damage to a
46 variety of crops. They are known to cause significant damage in the UK (Blackshaw
47 and Coll 1999; Peck and Olmstead 2009) but have also recently become a problem
48 in the USA following their accidental introduction and spread, causing cosmetic and
49 economic damage; larvae feed on the roots and stems of plants as well as decaying
50 vegetation (Rao et al. 2006). *Tipula paludosa* mainly damages pastures and spring
51 cereals, causing characteristic bare patches due to feeding on roots, and sometimes
52 leaves (Blackshaw and Coll 1999; White and French 1968). In contrast, *T. oleracea*
53 have been important pests of winter cereals planted after oilseed rape (Blackshaw
54 and Coll 1999). Other crops including lettuce, celery, sugar beet, sweet corn, berries,
55 tobacco, various vegetables and ornamentals are also targeted (Blackshaw and Coll
56 1999; Campbell 1975; Emmet 1992).

57 Adults can be easily separated by morphology, but larvae are more difficult to
58 distinguish and as such identification is challenging. The pattern of microscopic hairs
59 on the abdomen as well as sclerotisation on the final abdominal segment have been
60 suggested as key morphological features for separating these species (Brindle 1959),
61 but this has been disputed as a reliable character (Gelhaus 2005). Isoelectric
62 focusing, a protein based method, has also been used (Humphreys et al. 1993), but
63 it is not able to distinguish between these and other tipulid species. Rao et al. (2006)
64 used sequences of the mitochondrial Cytb gene to identify *T. paludosa* and *T.*
65 *oleracea* as well as other native North American crane flies, but there is a lack of
66 data at this DNA region for many tipulid species and so it does not have universal
67 applicability which is important in community studies. This lack of a reliable, widely

68 relevant and cost effective technique often means that larvae are not identified,
69 despite that this is important for applying pest management strategies; there are
70 distinct differences in life cycle, feeding behaviour and the damage period throughout
71 the year (Rao et al. 2006), which can result in potentially costly misuse of biological
72 control applications. It also means that at present we have limited knowledge of
73 populations in areas with a high proportion of grassland (e.g. the south west UK).
74 Since tipulids are also an important food source for birds (Pearce-Higgins et al. 2005,
75 Rhymer et al. 2012), understanding their distribution is critical for predicting their
76 population dynamics.

77 Molecular genetic methods offer a new approach to this problem. DNA barcoding
78 (Hebert et al. 2003), sequencing a section of the cytochrome c oxidase subunit I
79 (COI) mitochondrial protein coding gene, has become a standard for identification of
80 closely related taxa across a range of taxonomic groups, and as such has been used
81 to distinguish morphologically cryptic species in many taxa. However, to date, the
82 application of molecular genetic techniques to soil insects has lagged behind their
83 aboveground counterparts (Benefer and Blackshaw 2013). Where they have been
84 used, at the COI region and other diagnostic regions of DNA, it has often proven to
85 be invaluable in elucidating previously unclear aspects of the biology and ecology of
86 a range of species (e.g. in wireworms; Benefer et al. 2010; Benefer et al. 2012;
87 Benefer et al. 2013; Staudacher et al. 2011).

88 In this study, we used DNA barcoding to identify larvae of *T. paludosa* and *T.*
89 *oleracea* collected using a systematic sampling approach in agricultural grassland
90 on the Rothamsted Research North Wyke Farm Platform (NWFP) in south west UK.
91 The samples collected here were used as a baseline survey to assess the
92 distribution of soil insects across the site (described in Benefer et al. 2016). Using

93 DNA barcoding to identify leatherjackets collected to the species level, our specific
94 objectives were to:

95 1. Determine the relative proportion of leatherjacket species present
96 2. Investigate the intra- and inter-specific genetic variation of the species identified
97 3. Assess the spatial relationships between tipulid haplotypes.

98

99 **2. Methods**

100 **2.1 Sample collection**

101 Adult *T. paludosa* and *T. oleracea* were collected randomly from agricultural
102 grassland on the North Wyke Farm Platform (NWFP), Okehampton, Devon, UK (Lat:
103 50.73237; Long: -3.99635; Fig. 1) in May 2013 using hand nets and light traps. The
104 species were separated according to morphological characters (Blackshaw and Coll
105 1999; Hoebeke and Klass 2005), and samples were stored at -20°C before use in
106 DNA extraction. Leatherjackets recovered in the surveys described in Benefer et al.
107 (2016) were used as source material. Briefly, larvae were heat extracted from soil
108 cores that were collected using a systematic sampling approach (a 25m grid across
109 19 fields) in two sampling periods over two years. Larvae collected in April-May 2012
110 (106 samples) were stored in 70% ethanol, whilst those collected in April-May 2013
111 (117 samples) were stored at -20°C before subsequent processing (it was intended
112 to use stable isotope analysis on the 2013 samples and this required that no alcohol
113 be present, hence a different preservation method was used in 2012; no difference in
114 DNA quality or quantity was noticed between these preservation techniques). A
115 subset of these larvae was subsequently analysed.

116 **2.2 DNA extraction, amplification and sequencing**

117 DNA was extracted using a DNeasy® Tissue Kit (Qiagen, Hilden, Germany)
118 following the spin-column protocol for purification of total DNA from animal tissue.
119 For both adult tipulids and larvae, one of the central abdominal segments was cut
120 into small pieces using a sterilized blade and an approximately 1mm^3 section used
121 for extraction. All 117 larvae collected in 2013 were processed, but due to time and
122 resource limitations only 59 of the 106 larvae (i.e. 53%) collected in 2012 were
123 processed. Extracted DNA was preserved at 4°C for later PCR.

124 PCR was performed in individual 0.2mL PCR tubes using a Qiagen Taq PCR Core
125 Kit (Qiagen). Each 25uL PCR contained (final concentrations given in brackets): 10X
126 PCR buffer (1X), $MgCl_2$ (3mM), dNTPs (200μM of each), forward primer (1uM),
127 reverse primer (1uM), BSA (500ng; larval amplification was unsuccessful without this
128 additive), Taq (0.5 units), DNA grade water and 1uL DNA. Original Folmer primers
129 (Folmer et al. 1994) gave variable results, so we used modified Folmer primers
130 (Ramirez-Gonzalez et al. 2013), which resulted in a greater success rate. The tubes
131 were subjected to the following PCR cycle: initial denaturation at 94°C (3 min), 35
132 cycles of denaturation at 94°C (1 min), annealing at 52°C (30s) and extension at
133 72°C (1 min), followed by a final extension at 72°C (10 min). PCR products were
134 stored at 4°C in the short-term (<1 month) before subsequent processing.

135 PCR success and specificity was determined by running 5uL of PCR product on a 2%
136 agarose gel stained with SYBR® Safe DNA gel stain (Thermo Fisher Scientific,
137 Paisley, UK), and the remaining 20uL cleaned to remove excess primers and dNTPs
138 using Exonuclease I (Exo, 10 units; New England Biolabs, Hitchin, UK) and Shrimp
139 Alkaline Phosphatase (SAP, 1 unit; New England Biolabs, Hitchin, UK). The cleaned

140 PCR products were sent to Macrogen Inc. (Europe) for sequencing in the forward
141 direction only. 157 larvae out of the original 176 processed (plus 25 adults) were
142 sequenced in total; six 2012 larvae and thirteen 2013 larvae, i.e. 11% of the samples
143 processed, failed to amplify even when PCR was repeated and so were not sent for
144 sequencing.

145

146 **2.3 Sequence analysis**

147 The sequence trace files were visually checked and sequences manually edited
148 where applicable (i.e. in cases of ambiguous base calling) using BioEdit (Version
149 7.1.9.0; Hall 1999). Each sequence was translated to protein using Expasy
150 (http://web.expasy.org/translate/) to ensure that they complied with an open reading
151 frame (Buhay 2009). In order to achieve larval identification, known adult and larval
152 sequences were aligned and compared using ClustalW in MEGA 5.1 (Tamura et al.
153 2011), and all sequences were compared to the Barcode of Life Database (BOLD)
154 (BOLD Systems V3; http://www.boldsystems.org/).

155 Intra- and inter-specific genetic characteristics were assessed for each year
156 separately and combined using *P*-distance in MEGA 5.1 (Tamura et al. 2011) and by
157 calculating the number of haplotypes, nucleotide diversity and haplotype diversity in
158 DNAsp. v.5 (Librado and Rozas 2009) for an aligned region of 456bp for *T. paludosa*
159 only (*T. oleracea* contained <5 larval individuals, which can produce variable and
160 potentially inaccurate results; Goodall-Copestake et al. 2012).

161 A locus-by-locus analysis of molecular variance (AMOVA) with 1000 permutations
162 was carried out in Arlequin ver. 3.5.2 (Excoffier and Lischer 2010) to test for

significant differentiation within and between fields, for both years combined, and only for the 14 fields with >5 individuals (N=14), and a Minimum Spanning Network (MSN) was calculated in Network (Fluxus Technology Ltd. 2010; Bandelt et al. 1999), using median joining (MJ) to assess the relationships between haplotypes and their distribution across the study site.

3. Results

3.1 Species identification

In total, good quality sequences were obtained for 151 larval samples (49 from 2012, 102 from 2013) and all 25 of the adult specimens, which matched the original morphological identification (GenBank accession numbers KX698438 - KX698604). Sequences ranged from 590bp – 662bp in length, reflecting the differing quality of the sequence reads at either end.

Eighteen adults and 142 larvae (all 49 samples in 2012 and 93 samples in 2013; 94% of the total larvae processed) were identified as *T. paludosa*. Seven adults and five larvae (3% of the total larvae processed), all of which were found in 2013, were identified as *T. oleracea* in BOLD, with a 92-100% match (eight of these samples had a <97% match). There were seven larval sequences, which, with the exception of one, were a high match to either *T. paludosa* or *T. oleracea* (99-100%), but for which the protein sequence was incorrect (it did not adhere to a reading frame); although these are included in the identification data above since they contain a high match, they were excluded from further genetic analyses to avoid analysis and interpretation of potentially incorrect haplotypic/genetic diversity data.

186 There were four larval sequences which did not match either of these species in
187 2013 (3% of the total larvae processed): two matched other related crane fly families,
188 one with >99% match to *Tricyphona immaculate* (Meigen) (Family Pediciidae) and
189 one with >99% match to an unknown species in the Family Limoniidae. The third
190 was a 100% match to *Rhagio scolopaceus* (Linnaeus) (Family Rhagioniidae), and
191 the fourth was >99% match to an unknown Gammaproteobacteria.

192

193 **3.2 Genetic diversity and haplotype analyses**

194 For *T. paludosa* there were 18 haplotypes, containing between 1 and 79 sequences
195 (Table 1). Six of these haplotypes were shared by both adult and larval samples.
196 Apart from the seven more divergent samples (one was excluded due to having an
197 incorrect protein sequence), which comprised three of these haplotypes with a *p*-
198 distance of between 3.07% and 8.77% to all other non-divergent samples, the intra-
199 specific genetic distance was low – between 0.22% and 1.75% (Table 1). Some of
200 the *T. paludosa* haplotypes found in 2012 were not found in 2013 and vice versa,
201 with six shared between samples collected in different years, but the majority
202 belonged to four haplotypes while the rest were comprised of 1-3 sequences (Fig.
203 2a,b; Fig. 3). Genetic diversity indices for *T. paludosa* larvae are reported in Table 2.

204 There were three *T. oleracea* larval haplotypes (one sample was excluded due to
205 having an incorrect protein sequence), and there were an additional three adult
206 haplotypes; intra-specific *p*-distance was 0.19% – 1.1%. Interspecific *p*-distance was
207 between 4.61% - 6.36% (Table 1).

208

209 **3.3 Haplotype distribution**

210 The AMOVA (Table 3) suggested that most of the variation (>99%) was within rather
211 than between fields, suggesting no population structuring is present at this scale in
212 this study site and that this is a panmictic population. This is further supported by the
213 MSN (Fig. 3), which shows the relationships between the haplotypes and the fields
214 they were collected from. It can be seen here that there is no clear clustering of
215 haplotypes by field. There was also no clear pattern in terms of the distribution of
216 haplotypes between years, with some fields which contained relatively large
217 numbers of larvae one year containing none or reduced numbers the following year
218 (e.g. field 19 and field 10, Fig. 2). However, haplotype 4 in particular was widely
219 distributed in both years.

220 **4. Discussion**

221 This study set out to determine the relative proportions, genetic variation and
222 distribution of two pest tipulid species in a south-west UK grassland using DNA
223 barcoding. We found that a 456bp region of the COI gene is diagnostic for these two
224 species, and the method described is a quick and easy approach to identification of
225 these morphologically cryptic pests. Given the extensive and ever expanding
226 collection of publically available COI sequences stored in databases such as
227 GenBank and BOLD, the method and data presented here will be of use to others
228 investigating the ecology of *T. paludosa* and *T. oleracea* worldwide.

229 The vast majority (94%) of larvae collected were *T. paludosa*, with *T. oleracea*
230 representing only 3% of the sequences obtained. This is in agreement with
231 Humphreys et al. (1993), who reported 4% of larvae to be this species in a survey of
232 75 agricultural grass fields in Northern Ireland. Adults were not directly surveyed in
233 this current study, though based on observational data both species were abundant
234 aboveground, which could suggest that there are species-specific oviposition
235 preferences and/or differences in dispersal ability (Blackshaw et al. 1996; Benefer et
236 al. 2016). Although the limited data from surveys in UK agricultural grassland
237 suggest *T. oleracea* may not be present in damaging proportions, in North America
238 this species is becoming more common in turfgrass (Rao et al. 2006) and so it is
239 important to be able to separate damaging species reliably. As for other pests,
240 differences in the biology and ecology of damaging species has implications for the
241 monitoring of populations and targeting of control methods. Our specific observation
242 (concurring with that of Humphreys et al. 1993) that adult *T. oleracea* were widely
243 distributed over this grassland area whereas the larvae were relatively scarce shows
244 the importance of understanding how the presence and abundance of above-ground

245 adult stages relates to belowground larval populations. A similar dissonance in adult
246 and larval *Agriotes lineatus* (Linnaeus) wireworm distributions has also been
247 reported (Benefer et al. 2012).

248 That other related cranefly species in the Tipuloidea superfamily were found here
249 (albeit in very low abundance suggesting they are unlikely to be pests) and that they
250 were initially identified as tipulid larvae indicates how difficult it is to identify some
251 soil-dwelling larval species and emphasises the role of molecular methods for
252 species identification. Validated barcoding approaches, such as that described here,
253 which improve the efficiency and reliability of species identification for cryptic species
254 are therefore vital for future studies.

255 The level of genetic variation between *T. oleracea* and *T. paludosa* found in this
256 study (4.61% - 6.36%) is in accordance with that described by Rao et al. (2006), who
257 reported an average inter-specific *p*-distance of ~7% at the mitochondrial cytb region
258 and suggested that these species are sister groups. Intra-specific variation was
259 generally less than 2% (excluding the five more divergent samples), meaning
260 individuals could be unambiguously assigned to the correct species. Although we did
261 not screen all of the samples collected at our study site using this molecular
262 technique, we did assess a substantial proportion (67%) of them, comprising a large
263 collection of individuals from spatially separated locations as well as morphologically
264 identified adults. We also validated our sequences against those available in
265 GenBank/BOLD, which provided a >97% match for most individuals. While we can
266 therefore be relatively confident that we have accounted for variation adequately and
267 that this approach would be suitable for other populations, we did also find four larval
268 individuals that had an intra-specific *p*-distance of 4% and one that had a *p*-distance
269 of 8%. This could indicate that intra-specific variation is higher than the standard 3%

270 sequence divergence threshold used in many COI barcoding studies (e.g. Hebert et
271 al. 2004a) and/or that cryptic species are present (e.g. Hebert et al. 2004b).
272 Alternatively, there may be issues with contamination through the use of universal
273 COI primers that amplify across many species (e.g. we did find one individual with
274 a >99% match to a Gammaproteobacteria), or nuclear copies of mitochondrial DNA
275 (numts) may be present, increasingly found to be an issue in barcoding studies that
276 use mitochondrial genes for species identification (Song et al. 2008; Buhay 2009;
277 Benefer et al. 2012). Here, 11% of larvae failed to amplify and six of the sequences
278 (4%) contained additional underlying sequence leading to ambiguous base calling.
279 However, by using the validation steps here (manually checking trace files carefully
280 for errors, checking the protein sequence for stop codons, comparing sequences to
281 public sequence databases) we were able to highlight and omit low quality and
282 potentially erroneous sequences, leaving us with a high quality dataset. The use of
283 specific tipulid primers, more than one region of DNA and sequencing of
284 morphologically identified adults from grassland may aid in improving the efficacy of
285 the method and the identification of cryptic species.

286

287 Across both years, 17 larval *T. paludosa* haplotypes were found, with greater
288 nucleotide and haplotype diversity and haplotype abundance in the second year
289 (which may be somewhat expected as there was a bigger sample size this year).
290 Differences between haplotypes were generally small, but there were five more
291 divergent haplotypes, four of which were found in a single field each (Fig. 3).
292 Interestingly, haplotypes 3 and 9, two of the more divergent haplotypes that were
293 more closely related to each other (Table 1) than all other haplotypes (apart from an
294 adult; haplotype 18), were also found in neighbouring fields. This might suggest that

295 there are limitations to dispersal in this area, which can isolate and maintain local
296 populations through which little gene flow takes place (Schowalter 2011). Depending
297 on the temporal scale of isolation, this may eventually lead to divergent groups with
298 different behavioural phenotypes (Edelaar et al. 2008), reducing interaction with
299 other individuals even further. Given that some haplotypes (e.g. haplotype 4) are
300 common throughout the study site, and that AMOVA results suggest a panmictic
301 population with little evidence of spatial structuring (in accordance with Benefer et al.
302 (2016) who in a broader assessment found little evidence of scale or space affecting
303 tipulid larvae presence/absence in general), lack of dispersal ability in *T. paludosa* in
304 general does not seem to be the case, at this scale of study at least. For example,
305 field 1 and field 19 are separated by a distance of approximately 1.6km but share
306 haplotype 1. Bearup et al. (2013) have suggested that directional, wind-assisted
307 dispersal may account for synchronisation in *T. paludosa* dynamics over landscape
308 distances, and so a lack of genetic spatial structure at this scale is not unexpected.
309 The present study, along with those of Bearup et al. (2013) and Blackshaw and
310 Petrovskii (2007), suggest that *T. paludosa* would be a useful model organism for
311 investigating the relationship between spatial and genetic variation with respect to
312 adult dispersal, and interactions with the landscape and environment.

313

314 In conclusion, DNA barcoding of the COI region has proven to be reliable in
315 separating *T. paludosa* and *T. oleracea*, with the former being by far the most
316 prevalent species in this study site. The data suggest a relatively high level of
317 genetic variation and dispersal at this study scale, as well as the potential presence
318 of cryptic species, and future more extensive studies might use the method

319 described here to address whether tipulid species exist as a complex and the
320 implications of this for damage, monitoring and control.
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337

338 **References**

339 Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring
340 intraspecific phylogenies. Mol Biol Evol 16:37–48

341 Bearup D, Petrovskii S, Blackshaw R, Hastings A (2013) Synchronized dynamics of
342 *Tipula paludosa* metapopulation in a southwestern Scotland agroecosystem: Linking
343 pattern to process. Amer Nat 182(3):393-409

344 Benefer, CM, D'Ahmed, KS, Blackshaw, RP, Sint, HM, Murray, PJ (2016) The
345 distribution of soil insects across three spatial scales in grassland. Front Ecol Evol. 4:
346 (41). doi:10.3389/fevo.2016.00041

347 Benefer CM, Blackshaw RP (2013) Molecular approaches for studying root
348 herbivores. Adv Insect Physiol. 4:219–255. doi:10.1016/B978-0-12-417165-7.00005-
349 2

350 Benefer CM, Andrew P, Blackshaw RP, Ellis JS, Knight, ME (2010) The spatial
351 distribution of phytophagous insect larvae in grassland soils. Appl Soil Ecol 45:269-
352 274. doi:10.1016/j.apsoil.2010.05.002

353 Benefer CM, Knight ME, Ellis JS, Hicks H, Blackshaw RP (2012). Understanding the
354 relationship between adult and larval *Agriotes* distributions: the effect of sampling
355 method, species identification and abiotic variables. Appl Soil Ecol 53:3948.
356 doi:10.1016/j.apsoil.2011.11.004

357 Blackshaw RP, Coll C, Humphreys IC, Stewart RM (1996) The Epidemiology of a
358 New Leatherjacket Pest (*Tipula oleracea*) of Winter Cereals in Northern Britain.
359 HGCA Project Report 120.

360 http://cereals.ahdb.org.uk/media/369259/project_report_120.pdf. Accessed 23
361 January 2016.

362 Blackshaw R, Coll C (1999) Economically important leatherjackets of grassland and
363 cereals: biology, impact and control. Integrated Pest Manag Rev 4(2):145-162.

364 Blackshaw RP, Petrovskii SV (2007) Limitation and regulation of ecological
365 populations: a meta-analysis of *Tipula paludosa* field data. Math Model Nat Phenom
366 2(4):46-62

367 Brindle A (1959) Notes on the larvae of the British Tipulinae (Dipt.,Tipulidae). Part 6
368 The larvae of the *Tipula oleracea* group. Entomologist's Monthly Magazine, 95:176-
369 177.

370 Buhay JE (2009) "COI-like" sequences are becoming problematic in molecular
371 systematic and DNA barcoding studies. J Crustacean Biol 29(1):96–110.

372 Campbell LR (1975) Insecticidal control of European crane fly in Washington. J Econ
373 Entomol 68(3): 386-388.

374 Edelaar P, Siepielski AM, Clobert J (2008) Matching habitat choice causes directed
375 gene flow: a neglected dimension in evolution and ecology. Evolution 62(10): 2462-
376 2472.

377 Emmet B (1992) Pests of cruciferous crops. In: McKinlay RG (ed) Vegetable pest
378 crops. Macmillan Press, London, pp 74-87.

379 Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to
380 perform population genetics analyses under Linux and Windows. Mol Ecol Resour
381 10(3):564–567

382 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for
383 amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan
384 invertebrates. Mol Mar Biol Biotechnol 3(5):294-9.

385 Gelhaus JK (2005) The crane-fly *Tipul*a (*Tipula*) *Oleracea* (Diptera: Tipulidae)
386 reported from Michigan; a new pest of turfgrass in eastern North America. *Great*
387 *Lakes Entomologist* 38(1/2):97.

388 Goodall-Copestake WP, Tarling GA, Murphy EJ (2012) On the comparison of
389 population-level estimates of haplotype and nucleotide diversity: a case study using
390 the gene *cox1* in animals. Heredity 109(1):50-56. doi:10.1038/hdy.2012.12.

391 Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and
392 analysis program for Windows 95/98/NT. Nucl Acid S 41:95–98.

393 Hebert PDN, Cywinska A, Ball SL (2003) Biological identifications through DNA
394 barcodes. P Roy Soc Lond B Bio, 270(1512): 313-321.

395 Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species
396 in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly
397 *Astraptes fulgerator*. Proc Natl Acad Sci USA 101(41):14812-7.

398 Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. (2004b) Identification of birds
399 through DNA Barcodes. PLoS Biol 2(10):e312. doi:10.1371/journal.pbio.0020312.

400 Hoebeke ER, Klass C (2005) *Tipula paludosa* Meigen and *T. oleracea* Meigen,
401 European Crane Flies New to the Eastern United States: Potentially Serious
402 Turfgrass and Pasture Pests. Proceedings, 16th U.S. Department of Agriculture
403 interagency research forum on gypsy moth and other invasive species 2005; 2005

404 January 18-21; Annapolis, MD. Gen. Tech. Rep. NE-337. Newtown Square, PA: U.S.
405 Department of Agriculture, Forest Service, Northeastern Research Station, pp 43.

406 Humphreys I, Blackshaw R, Stewart R, Coll C (1993) Differentiation between larvae
407 of *Tipula paludosa* and *Tipula oleracea* (Diptera: Tipulidae) using isoelectric focusing,
408 and their occurrence in grassland in northern Britain. Ann Appl Biol 122(1):1-8.

409 Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA
410 polymorphism data. Bioinformatics 25(11):1451-1452.

411 Pearce-Higgins J, Yalden D, Whittingham M (2005) Warmer springs advance the
412 breeding phenology of golden plovers *Pluvialis apricaria* and their prey (Tipulidae).
413 Oecologia 143(3): 470-476.

414 Peck DC, Olmstead D (2009) Invasive *Tipula* (Diptera: Tipulidae) in turfgrass of the
415 northeast United States: geographic distribution and local incidence three years after
416 detection. J Econ Entomol 102(2):652-658.

417 Ramirez-Gonzalez R, Yu DW, Bruce C, Heavens, D, Caccamo M, Emerson BC
418 (2013) PyroClean: Denoising pyrosequences from protein-coding amplicons for the
419 recovery of interspecific and intraspecific genetic variation. PLOS ONE 8(3): e57615.
420 DOI: 10.1371/journal.pone.0057615.

421 Rao, S., Liston, A., Crampton, L. & Takeyasu, J. (2006) Identification of larvae of
422 exotic *Tipula paludosa* (Diptera: Tipulidae) and *T. oleracea* in North America using
423 mitochondrial cytB sequences. Ann Entomol Soc Am, 99(1):33-40.

424 Rhymer CM, Devereux CL, Denny MJH, Whittingham, MJ (2012) Diet of Starling
425 *Sturnus vulgaris* nestlings on farmland: the importance of Tipulidae larvae. Bird
426 Study 59(4):426. DOI: 10.1080/00063657.2012.725026

427 Schowalter TD (2011) Insect ecology: an ecosystem approach, 3rd edn. Academic
428 Press.

429 Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA
430 barcoding overestimates the number of species when nuclear mitochondrial
431 pseudogenes are co-amplified. PNAS 105:13486–13491.

432 Staudacher K, Pitterl P, Furlan L, Cate PC, Traugott M (2011a) PCR based species
433 identification of *Agriotes* larvae. Bull Entomol Res 101:201–210

434 Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary
435 genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

436 White J, French N (1968) Leatherjacket damage to grassland. Grass Forgae Sci
437 23(4):326-329.

438

439

440

441

442

443

444

445

446

447

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1. *T. paludosa* hap 1		*0.3*	*0.81*	*0.22*	*0.3*	*0.29*	*0.36*	*0.36*	*0.82*	*0.3*	*0.31*	*0.31*	*1.22*	*0.3*	*0.57*	*0.32*	*0.3*	*0.84*	*0.97*	*0.97*	*1*	*0.99*	*0.99*	*0.99*	*2.3*	*2.26*	*2.29*	*2.33*
2. *T. paludosa* hap 2	0.44		*0.8*	*0.21*	*0.29*	*0.29*	*0.35*	*0.35*	*0.81*	*0.3*	*0.3*	*0.3*	*1.23*	*0.29*	*0.56*	*0.3*	*0.3*	*0.83*	*0.98*	*0.98*	*1.02*	*1*	*1*	*1*	*2.29*	*2.26*	*2.29*	*2.33*
3. *T. paludosa* hap 3	3.29	3.29		*0.78*	*0.81*	*0.75*	*0.78*	*0.79*	*0.3*	*0.81*	*0.81*	*0.82*	*1.3*	*0.81*	*0.93*	*0.81*	*0.8*	*0.36*	*1.1*	*1.1*	*1.14*	*1.07*	*1.08*	*1.11*	*2.33*	*2.27*	*2.27*	*2.34*
4. *T. paludosa* hap 4	0.22	0.22	3.07		*0.21*	*0.2*	*0.29*	*0.29*	*0.79*	*0.21*	*0.22*	*0.22*	*1.21*	*0.21*	*0.53*	*0.22*	*0.21*	*0.81*	*0.95*	*0.95*	*0.98*	*0.98*	*0.97*	*0.97*	*2.29*	*2.26*	*2.29*	*2.33*
5. *T. paludosa* hap 5	0.44	0.44	3.29	0.22		*0.3*	*0.36*	*0.35*	*0.82*	*0.3*	*0.31*	*0.3*	*1.19*	*0.29*	*0.58*	*0.3*	*0.3*	*0.85*	*0.92*	*0.92*	*0.96*	*0.95*	*0.95*	*0.94*	*2.31*	*2.27*	*2.3*	*2.31*
6. *T. paludosa* hap 6	0.44	0.44	2.85	0.22	0.44		*0.21*	*0.21*	*0.75*	*0.28*	*0.3*	*0.29*	*1.2*	*0.29*	*0.56*	*0.31*	*0.28*	*0.78*	*0.94*	*0.94*	*0.98*	*0.97*	*0.96*	*0.96*	*2.29*	*2.26*	*2.28*	*2.33*
7. *T. paludosa* hap 7	0.66	0.66	3.07	0.44	0.66	0.22		*0.29*	*0.78*	*0.35*	*0.36*	*0.36*	*1.22*	*0.35*	*0.6*	*0.37*	*0.35*	*0.8*	*0.96*	*0.96*	*1*	*0.98*	*0.98*	*0.97*	*2.29*	*2.26*	*2.28*	*2.33*
8. *T. paludosa* hap 8	0.66	0.66	3.07	0.44	0.66	0.22	0.44		*0.79*	*0.34*	*0.36*	*0.36*	*1.22*	*0.36*	*0.61*	*0.37*	*0.34*	*0.81*	*0.96*	*0.96*	*0.99*	*0.99*	*0.98*	*0.98*	*2.29*	*2.26*	*2.28*	*2.33*
9. *T. paludosa* hap 9	3.29	3.29	0.44	3.07	3.29	2.85	3.07	3.07		*0.82*	*0.82*	*0.83*	*1.31*	*0.82*	*0.93*	*0.82*	*0.81*	*0.21*	*1.1*	*1.1*	*1.14*	*1.07*	*1.08*	*1.11*	*2.31*	*2.27*	*2.26*	*2.33*
10. *T. paludosa* hap 10	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29		*0.3*	*0.3*	*1.22*	*0.3*	*0.57*	*0.3*	*0.29*	*0.84*	*0.97*	*0.97*	*1*	*0.99*	*0.98*	*0.98*	*2.3*	*2.25*	*2.3*	*2.34*
11. *T. paludosa* hap 11	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29	0.44		*0.3*	*1.23*	*0.3*	*0.57*	*0.31*	*0.31*	*0.84*	*0.97*	*0.97*	*1.01*	*1*	*1*	*0.99*	*2.3*	*2.26*	*2.29*	*2.32*
12. *T. paludosa* hap 12	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29	0.44	0.44		*1.21*	*0.3*	*0.58*	*0.31*	*0.3*	*0.85*	*0.95*	*0.95*	*0.98*	*0.98*	*0.97*	*0.97*	*2.3*	*2.26*	*2.29*	*2.33*
13. *T. paludosa* hap 13	7.68	7.68	8.33	7.46	7.24	7.24	7.46	7.46	8.33	7.68	7.68	7.46		*1.21*	*1.3*	*1.24*	*1.2*	*1.33*	*0.76*	*0.81*	*0.82*	*0.86*	*0.77*	*0.79*	*2.33*	*2.25*	*2.33*	*2.29*
14. *T. paludosa* hap 14	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29	0.44	0.44	0.44	7.68		*0.48*	*0.3*	*0.29*	*0.84*	*0.96*	*0.96*	*1*	*0.96*	*0.98*	*0.98*	*2.3*	*2.27*	*2.29*	*2.33*
15. *T. paludosa* hap 15	1.54	1.54	4.39	1.32	1.54	1.54	1.75	1.75	4.39	1.54	1.54	1.54	8.77	1.1		*0.57*	*0.56*	*0.95*	*1.07*	*1.07*	*1.1*	*1.06*	*1.09*	*1.08*	*2.28*	*2.27*	*2.28*	*2.32*
16. *T. paludosa* hap 16	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29	0.44	0.44	0.44	7.68	0.44	1.54		*0.3*	*0.84*	*0.98*	*0.98*	*1.02*	*1.01*	*1*	*1*	*2.3*	*2.25*	*2.3*	*2.35*
17. *T. paludosa* hap 17	0.44	0.44	3.29	0.22	0.44	0.44	0.66	0.66	3.29	0.44	0.44	0.44	7.68	0.44	1.54	0.44		*0.83*	*0.95*	*0.95*	*0.99*	*0.98*	*0.98*	*0.97*	*2.29*	*2.25*	*2.29*	*2.32*
18. *T. paludosa* hap 18	3.51	3.51	0.66	3.29	3.51	3.07	3.29	3.29	0.22	3.51	3.51	3.51	8.55	3.51	4.61	3.51	3.51		*1.11*	*1.11*	*1.16*	*1.08*	*1.1*	*1.13*	*2.31*	*2.27*	*2.26*	*2.33*
19. *T. oleracea* hap 1	5.04	5.04	6.14	4.82	4.61	4.61	4.82	4.82	6.14	5.04	5.04	4.82	3.07	5.04	6.14	5.04	5.04	6.36		*0.21*	*0.29*	*0.37*	*0.22*	*0.2*	*2.32*	*2.29*	*2.3*	*2.3*
20. *T. oleracea* hap 2	5.04	5.04	6.14	4.82	4.61	4.61	4.82	4.82	6.14	5.04	5.04	4.82	3.29	5.04	6.14	5.04	5.04	6.36	0.22		*0.35*	*0.42*	*0.31*	*0.29*	*2.32*	*2.3*	*2.31*	*2.3*
21. *T. oleracea* hap 3	5.48	5.48	6.58	5.26	5.04	5.04	5.26	5.26	6.58	5.48	5.48	5.26	3.51	5.48	6.58	5.48	5.48	6.8	0.44	0.66		*0.46*	*0.37*	*0.36*	*2.32*	*2.29*	*2.3*	*2.3*
22. *T. oleracea* hap 4	5.26	5.26	5.92	5.04	4.82	4.82	5.04	5.04	5.92	5.26	5.26	5.04	3.73	4.82	5.92	5.26	5.26	6.14	0.66	0.88	1.1		*0.43*	*0.41*	*2.33*	*2.31*	*2.3*	*2.3*
23. *T. oleracea* hap 5	5.26	5.26	5.92	5.04	4.82	4.82	5.04	5.04	5.92	5.26	5.26	5.04	3.29	5.26	6.36	5.26	5.26	6.14	0.22	0.44	0.66	0.88		*0.31*	*2.32*	*2.29*	*2.31*	*2.29*
24. *T. oleracea* hap 6	5.26	5.26	6.36	5.04	4.82	4.82	5.04	5.04	6.36	5.26	5.26	5.04	3.29	5.26	6.36	5.26	5.26	6.58	0.22	0.44	0.66	0.88	0.44		*2.31*	*2.29*	*2.29*	*2.3*
25. *Tricyphona immaculata*	47.4	47.4	46.3	47.1	47.4	47.1	47.1	47.4	46.7	47.4	47.4	46.9	49.6	46.9	48	47.4	46.9	46.7	46.9	47.1	47.1	46.5	46.7	46.7		*2.25*	*1.68*	*1.61*
26. Gammaproteobacteria	57.2	57	56.8	57	57.2	56.8	56.8	56.8	57	57.2	57	57.2	58.6	56.8	57.2	57.2	56.8	57	56.4	56.6	56.6	55.9	56.4	56.4	43.6		*2.31*	*2.3*
27. Limoniidae	44.1	44.1	44.3	44.1	44.3	43.9	43.9	43.9	44.7	44.3	44.3	43.9	46.7	43.9	45	44.3	44.3	44.7	43.9	43.6	44.1	44.1	43.6	43.6	15.6	43.9		*1.62*
28. *Rhagio scolopaceus*	47.1	46.9	45.8	46.9	46.7	46.9	46.9	46.9	46.3	47.1	46.7	47.1	49.1	46.9	48	47.1	46.7	46.3	46.7	46.5	46.9	46.7	46.5	46.7	14.3	43.2	16.2	

448 Table 1. Genetic distance (p-distance, %) for haplotypes of all species, for adults and larvae combined, identified by DNA barcoding.

449 One sequence haplotype was used in a 452bp alignment (excluding gaps and missing data). Standard error estimates (%) are

450 given above in italics on the diagonal.

451

452

	2012	2013	All
Number of individuals	46	90	136
Number of haplotypes	10	14	17
Haplotype diversity	0.688 (0.071)	0.703 (0.049)	0.695 (0.040)
Nucleotide diversity	0.00478 (0.00171)	0.00607 (0.00192)	0.00562 (0.00141)

453

454 Table 2. Genetic diversity indices for a 456bp section of the COI gene of *T. paludosa*
455 larvae collected from the North Wyke Farm Platform, UK, in 2012, 2013, and these
456 years combined. Figures in brackets are ± standard deviation.

457

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Among fields	13	18.325	0.00799	0.59
Within fields	114	152.441	1.3372	99.41
Total	127	170.766	1.34519	

458

459 Table 3. Locus-by-locus AMOVA for a 456bp region of the COI gene for *T. paludosa*
460 larvae (both years combined) collected on the North Wyke Farm Platform, UK.

461

462

463

464

465

466

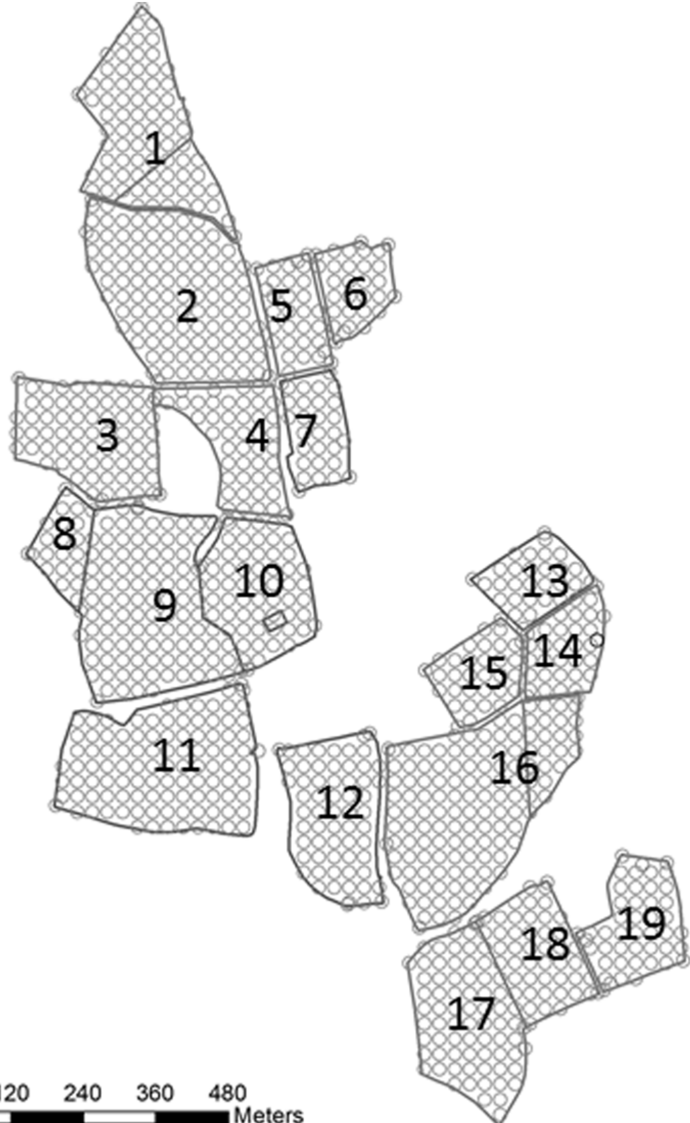
467

468 **Fig. 1** The study site at the North Wyke Farm Platform, Okehampton, Devon, UK.
469 Fields are numbered and small open circles represent the individual sampling points,
470 which are on a 25m grid.

471 **Fig. 2** The number of sequences per haplotype for *T. paludosa* larvae collected
472 across 19 fields in a) 2012 (N=46) and b) 2013 (N=90) on the North Wyke Farm
473 Platform, UK.

474 **Fig. 3** Minimum spanning network of *T. paludosa* adult and larval COI sequences
475 (456bp; N = 154). The size of the pie charts, which represent the field(s) the
476 haplotype was found in, is proportional to the number of sequences of that haplotype
477 in the population. The length of the connectors is proportional to the number of
478 nucleotide changes between haplotypes (the longest branches were contracted; the
479 number of mutations is indicated next to the connector)

480



481

482

483

484

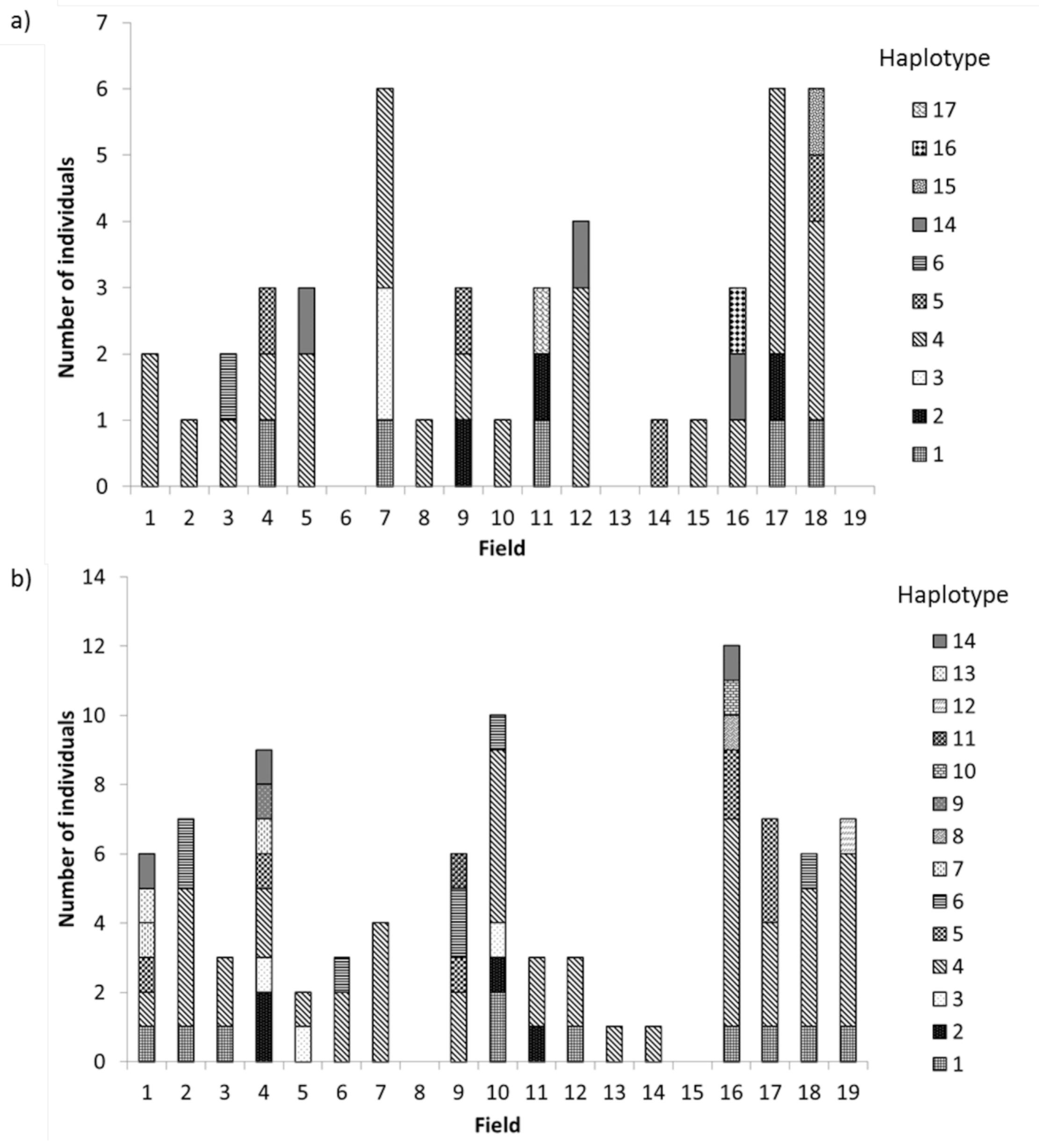
485

486

487

488

489



490

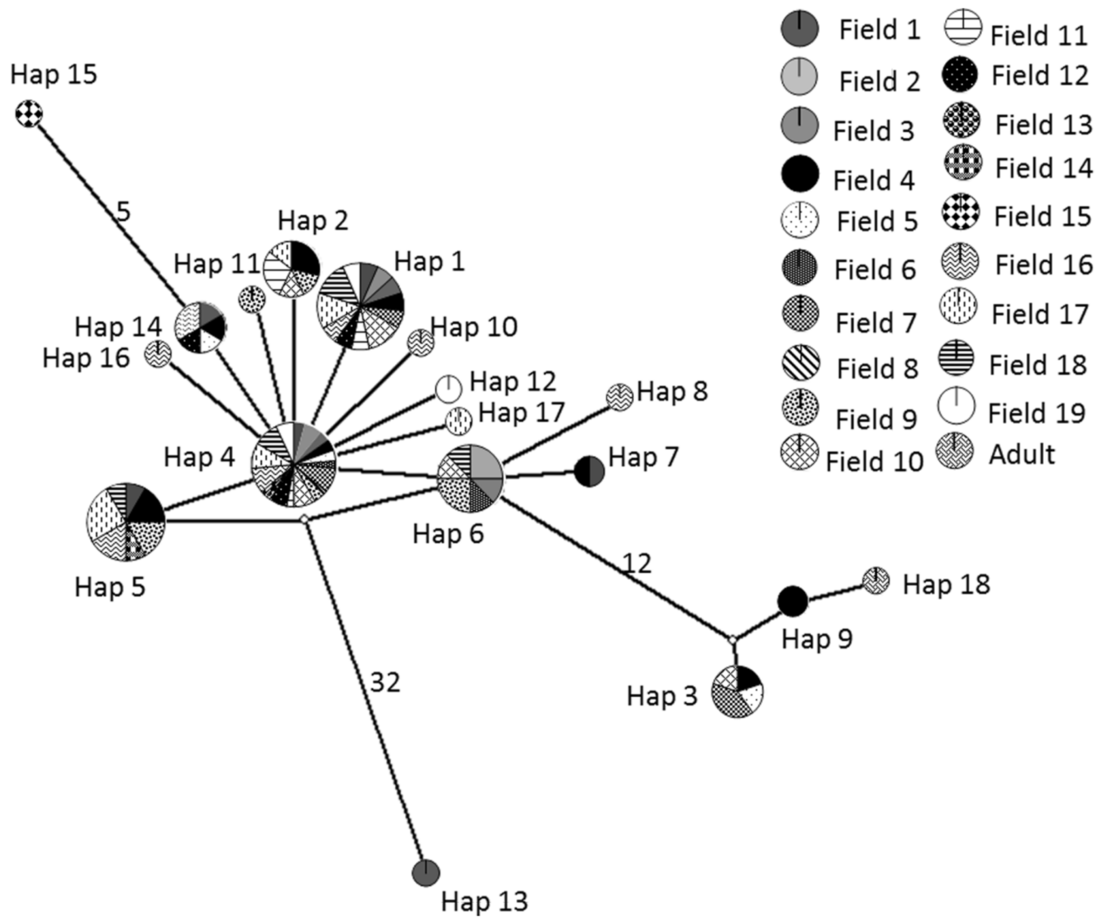
491

492

493

494

495



496

497

498

499

500

501

502

503

504

505

506 **Acknowledgements**

507 This study was part of a ResM awarded to KD and funded by the Kurdistan Region
508 Government and the Ministry of Higher Education through the Human Capacity
509 Development Program–Round1. The North Wyke Farm Platform is a BBSRC
510 supported National Capability (BB/J004308). The authors would like to thank Jane
511 Akerman and Peter Smithers (Plymouth University) for their help in collecting,
512 processing and identifying samples and Bruce Griffiths (Rothamsted Research–
513 North Wyke) for providing GPS and maps.

514

515 **Conflict of interest statement**

516 There are no disputes over the ownership of the data presented in the paper and all
517 contributions have been attributed appropriately.