1 **Genetic structure of regional water vole populations and footprints of reintroductions: a case study from**
2 **Southeast England**

3 Rowenna J Baker¹; Dawn M Scott²; Peter J King³; Andrew DJ Overall^{1*}

5 *Affiliations and addresses:*

6 ¹ Ecology, Conservation and Zoonosis Research Group, Huxley Building, University of Brighton, Brighton BN2 4GJ,
7 UK.

8 ² School of Life Sciences, Keele University, Keele, Staffordshire, ST5 5BG

9 ³ Ouse & Adur River's Trust, Barcombe, East Sussex BN8 5BW, UK.

10

11 * Corresponding author: Email: a.d.j.overall@brighton.ac.uk. Telephone: +441273642099

12 ORCID:

13 R.Baker: 0000-0002-5751-4999; D.Scott: 0000-0002-9570-2739; A.Overall: 0000-0001-9766-1056

14

15 **Abstract**

16 An important consideration when implementing species management is preserving genetic variation, which is
17 fundamental to the long-term persistence of populations and adaptive potential of the species. The European water vole
18 *Arvicola amphibius* is of high conservation importance in the United Kingdom due to its documented decline in both
19 distribution and abundance. Conservation strategies for this species include protecting source populations, increasing
20 habitat availability and connectivity, non-native predator control and reintroduction. We used mtDNA control region
21 sequences and eight microsatellite markers from samples collected from 12 localities in southeast England, to determine
22 how genetic variation is structured amongst regional water vole populations and to what extent population structure has
23 been influenced by reintroductions. We found high haplotype diversity (*h*) across native populations in the southeast
24 region and evidence that divergent lineages had been introduced to the region. We detected significant structure
25 between watersheds with mtDNA from native populations and evidence of finer scale structure between populations
26 within watersheds with both mtDNA and microsatellites. We suggest that management strategies should aim to
27 conserve genetic diversity at a population level and that watersheds are a practicable unit for prioritising management
28 within regions. We propose that introductions to restore or augment water voles within watersheds should consider the
29 genetic composition of regional populations and highlight that genetic data has an important role in guiding future
30 conservation options that secure adaptive potential in the face of future environmental change.

31 **Key words**

32 Water vole, genetic structure, reintroductions, microsatellites, mitochondrial DNA, landscape

33 INTRODUCTION

34 Current patterns in genetic variation are shaped by historical and contemporary factors that have influenced the
35 distribution, dispersal, and demography of a species. For example, Pleistocene glaciations have played a major role in
36 shaping the phylogenetic structure of most European species, generating distinct intraspecific lineages that result from
37 the isolation of populations between glacial refugia (Schmitt 2007). At a finer scale, landscape heterogeneity and
38 geographic barriers can restrict dispersal and gene flow, resulting in genetic structuring amongst populations (Manel et
39 al. 2003). These hierarchical patterns in genetic structure are typically referred to as evolutionary significant units
40 (ESUs) and management units (MUs) and are regarded as important biological entities for conservation below the
41 species level (Ryder 1986; Moritz 1994; Allendorf and Luikart 2007; Pasbøll et al. 2007).

42 The European water vole (*Arvicola amphibius*) is a species of high conservation importance in the UK where it
43 has undergone one of the fastest documented declines of any British mammal during the past century (Jefferies et al.
44 1989; Strachan and Jefferies 1993; Strachan et al. 2000). The causes of the decline have primarily been attributed to the
45 widespread loss of wetland habitats for agricultural and urban development and the predation of populations by feral
46 American mink (*Neovison vison*) (Woodroffe et al. 1990; Baretto et al. 1998; Macdonald and Strachan 1999).
47 Consequently, conservation strategies have been widely implemented in recent decades to preserve remaining
48 populations and to restore water voles to their former distribution (JNCC 1995; Strachan et al. 2011). These include
49 increasing suitable habitat availability and functional connectivity of populations, mink control and the protection of
50 key source populations to facilitate range expansion. Owing to the relatively limited dispersal capabilities of water voles
51 (1–2 km) (Stoddart 1970; Telfer et al. 2003; Aars et al. 2006) and their range contraction over the past century (Mcquire
52 and Whitfield, 2017), the reintroduction and translocation of water voles has also become popular for re-establishing
53 populations in areas where they are unlikely to achieve natural recolonization (Strachan et al. 2011), a strategy that has
54 been exacerbated by the need to relocate water voles as part of mitigation work driven by development. These
55 reintroductions and translocations are undertaken in line with best practice guidance developed for water voles in the
56 UK (Strachan et al. 2011; Dean et al. 2016) and outlined for all species by the IUCN (IUCN/SSC 2013). However,
57 information on the provenance or location of these activities is rarely accessible to organisations other than those
58 involved in release programmes, such that genetic consequences are largely unknown.

59 As genetic variation is fundamental to the long-term persistence of populations and adaptive potential of
60 species (Moritz 2002; Reed and Frankham 2003; Frankham 2005), preservation of genetic diversity is an important
61 consideration when implementing species conservation management strategies. Genetic analysis of population structure
62 and origin is central to this as it provides conservation practitioners with information on the processes and factors that
63 generate, preserve, and maintain genetic variation, which can be integrated into effective management plans (Moritz

2002; Schwartz et al. 2005; Waples and Gaggiotti 2006; Lowe and Allendorf 2010). Genetic data on population structure is thus vital to the process of defining meaningful units for targeting conservation. This is particularly critical for directing captive breeding and translocations that may inadvertently homogenise otherwise genetically divergent, or locally adapted, entities, or cause outbreeding depression if populations of divergent evolutionary lineages are admixed (Edmands 2007; Huff et al. 2011); a specific explanation being the Dobzhansky-Muller Model of hybrid incompatibility which, although originally conceived to explain species incompatibility, offers an explanation for how long-separated populations accumulate different recessive mutations within linked genes (Johnson 2008). This highlights the fine balance that often needs to be struck between negating the negative consequences of both inbreeding and outbreeding depression (Houde et al. 2011).

Previous research on the phylogeographic structure of water voles in the UK identified a major division between the Scottish and English/Welsh mitochondrial (mtDNA) lineages (Piertney et al. 2005). This is believed to have resulted from the isolation of refugial populations during the late glacial (Piertney et al. 2005) and a temporally distinct post glacial recolonization which saw the initial Scottish lineage of colonisers later displaced by the English up to the current Scottish border (Brace et al. 2016). Despite insufficient morphological variation, the two lineages are consequently managed as separate ESUs (Piertney et al. 2005; Strachan et al. 2011). At a finer scale, it has also been shown that significant divergence exists between regional populations, which is characteristic of restricted dispersal, suggesting that appropriate MUs are likely to be at a watershed level (Piertney et al. 2005). It is reasonable to hypothesise that water vole populations, with a semi-aquatic lifestyle in the UK, would be structured by river or drainage basins. These habitats are fragmented by their nature and similar patterns have been identified in other organisms associated with aquatic environments, including amphibians (Lind et al. 2011), fish (Ferguson 1989; Costello et al. 2003) and mammals (Igea et al. 2013; Querejeta et al. 2017). However, as water voles have also been shown to disperse overland between waterways and watersheds (Telfer et al. 2003; Aars et al. 2006; Fisher et al. 2008) there is some doubt on whether watersheds do in fact present genetically divergent entities of conservation importance.

Our aims were to 1) determine how natural genetic variation is structured amongst regional populations with the expectation that watersheds would show genetic population divergence and 2) examine to what extent population structure has been influenced by reintroductions. As we wanted to distinguish between historical and contemporary processes, we used both microsatellites, which have very fast mutation rates (e.g., in vertebrates, the typical mutation rate is 1×10^{-4} substitutions/locus/generation (Hancock 1999, Verkuil et al. 2014)) and mtDNA sequence variation, which mutates more slowly (e.g., $2 \times 10^{-8} - 1.5 \times 10^{-7}$ (Verkuil et al. 2014)) and is therefore more likely to preserve historical population subdivision.

Furthermore, mtDNA is more useful for distinguishing unique haplotypes resulting from introductions. Given the widespread conservation efforts to secure and re-establish water voles using reintroductions, our findings can be used to determine conservation MUs and inform translocation and reintroduction practice with the aim to maintain natural patterns in genetic variation.

MATERIALS AND METHODS

Study area and sites

This study was carried out in southeast England within the counties of East and West Sussex, Kent and Greater London which are located between the coordinates 50.4° to 51.4° N and 0.5° W to 1.25° E. Sampling was not inclusive of all known populations within the region but represented 12 populations from three main watershed groups: Manhood Peninsular (MHP), Central Sussex (CS) and North Kent (NK) (see Fig 1). There are seven documented sites within the region where water voles have been reintroduced (Table S1) (McGuire and Whitfield 2017). Two of these were sampled for this study and included sites LWr and AWr which had received introductions of 192 and 231 water voles between 1999 and 2005. The remaining 10 study sites comprised natural populations of presumed wild origin. A map showing water vole distribution, the location of study populations and reintroduction sites within each watershed group is given in Fig 1.

Sample collection and DNA extraction

Water vole tissue and hair samples were collected between 2010 and 2013. Hair samples were obtained from plucks taken from individuals (n = 453) live captured at nine sites and from remote hair capture tubes (n = 7) that were placed at an additional three sites. The tubes consisted of small sections of 3" (7.62 cm) drainpipe lined with adhesive flypaper on the inside roof (Baker 2013). Traps were baited with carrot and apple (live capture traps) or just apple (hair tubes) and placed at 30 m intervals along water vole occupied wetland habitats (watercourses, reedbeds and fen). Tissue samples (n = 2) were from the tails of dead individuals encountered during survey work. All surveys were carried out under Natural England Protected Species Licence (No.'s 2010505, 20120780 and 20123101). Hair and tissue samples were stored in dry tubes at -20°C for a maximum of four weeks until DNA extraction. DNA was extracted using a Qiagen DNeasy Blood and Tissue kit following manufacturer's protocol for tissue but with a slight modification involving the initial digestion of hair samples using an extraction buffer (Pfeiffer et al. 2004).

FIGURE 1 HERE

Mitochondrial DNA sequencing and microsatellite genotyping

124 A 736 bp fragment of the mitochondrial control region was amplified from samples obtained across all 12 sites
125 using the primers F15708 and R92 and using cycling conditions described by Piertney et al. (2005). PCR reactions were
126 carried out in a final volume of 25µl comprising approximately 5ng of genomic DNA, 1x PCR buffer (Invitrogen),
127 0.2mM of each dNTP, 2mM MgCl₂, 0.2µM of each primer, 1U Platinum Taq polymerase (Invitrogen) and 1.5µg BSA
128 (Invitrogen). PCR products were sequenced at Source BioScience (UK) using an Applied Biosystems 3730 series DNA
129 Analyser. Electropherograms were visually inspected, aligned and trimmed to remove ambiguous end regions using the
130 default parameters in SeqTrace (Stucky 2012). Any novel haplotypes that appeared only once were verified through re-
131 sequencing.

132 Hair pluck samples from nine of the 12 populations were individually genotyped using eight microsatellite
133 markers (AV7, AV8, AV9, AV11, AV12, AV13, AV14, AV15) developed by Stewart et al. (1998) and using modified
134 primer sequences (AV8, AV9 and AV11) described by Berthier et al. (2005). Forward primers were labelled with the
135 fluorescent reference dyes FAM, HEX or NED and PCR amplifications were carried out in a final volume of 15µl using
136 components described above but with 1.5 mM MgCl₂. Cycling conditions included an initial polymerase activation step
137 of 94°C for 30sec, followed by 35 cycles of denaturation at 94°C for 10sec, 30sec of primer specific annealing (54°C
138 for AV7 and AV13, 58°C for AV8, AV11–12 and AV14–15 and 60°C for AV9) and 30sec of primer extension with a
139 final extension step of 2 min at 72°C. PCR products were separated by fragment size at Source BioScience (UK) using
140 capillary electrophoresis on an ABI3730xl using a ROX500 size standard. Fragment analysis was performed using
141 PeakScanner V1.0 (Applied Biosystems). Any ambiguous genotypes, in terms of stutter and peak-height ratios were
142 genotyped twice and all loci were checked for null alleles and other technical artefacts using Micro-Checker (Van-
143 Oosterhaut et al. 2004). Only individuals successfully genotyped for four or more microsatellite loci were used in the
144 analyses.

145 **Mitochondrial DNA variation and phylogenetic structure**

146 Variation in the mtDNA control region was assessed for each study population by calculating the number of
147 haplotypes (N_H), the number of private haplotypes (P_H), haplotype diversity (h) and nucleotide diversity (π) (Nei 1987,
148 Tajima 1989) using ARLEQUIN v3.5.1.2 (Excoffier et al. 2010). Genealogical relationships among haplotypes were
149 first assessed using a haplotype network constructed using a 95% maximum parsimony connection criterion
150 implemented in the programme TCS v1.21 (Clement et al. 2000). Connections among haplotypes were resolved using
151 the approach outlined in Crandall and Templeton (1993), which optimizes the construction of *within*-species
152 phylogenies. Phylogenetic relationships were then further investigated between water vole control region haplotypes
153 obtained in this study and those obtained by Piertney et al. (2005) that were trimmed to contain the same regions.
154 Phylogenetic analysis was performed using Bayesian inference methods implemented in MRBAYES v3.2 (Ronquist et

155 al. 2012). A standard substitution model was employed using the software's default settings, which concurred with the
156 optimal model of sequence evolution found by Piernney et al. (2005). Chains were run for 1 million generations and
157 samples were drawn every 1,000 generations until multiple trees reached convergence, measured by a standard
158 deviation of split frequencies of <0.01 . Consensus trees were visualised using FigTree v1.4.0 (Rambaut 2006–2014).

159 **Microsatellite variation and Bayesian analysis of population structure**

160 Genetic variation in microsatellite loci was analysed for each population and watershed by determining the
161 number of alleles averaged across loci (N_A), allelic richness (A_R), observed heterozygosity (H_O), and expected
162 heterozygosity (H_E) using FSTAT v2.9.3.2 (Goudet 1995, 2002). Significant departures of genotype frequencies from
163 Hardy-Weinberg expectations were determined for each population from F_{IS} values calculated using a randomization
164 procedure (72000 permutations). Bonferroni procedures were applied to the unbiased probability estimates that were
165 generated across loci within each sample from each population (Holm 1979; Rice 1989). Where significant F_{IS}
166 estimates were found, the R script ConStruct (Overall, 2016) was used to identify the relative contributions of close kin
167 inbreeding and cryptic subpopulations (Wahlund effect) to the excess homozygosity. The number of private alleles (P_A)
168 within each population was calculated using GenAlEx v6.5 (Peakall and Smouse 2006, 2012).

169 Genetic population structure was investigated using STRUCTURE v2.3.4 (Pritchard et al. 2000). Analysis was
170 performed assuming admixture and correlated allele frequencies and included prior information on sampling location to
171 test whether the ancestry of individuals correlated with sampling area. The most likely number of populations (K) was
172 determined using the ΔK statistic (Evanno et al. 2005) using results from 20 runs of $K = 1$ to 10, each with 100,000
173 burn-in period and Markov chain Monte Carlo randomisations using STRUCTURE HARVESTER (Earl and vonHolt
174 2011). Population assignment coefficients (Q_P) and the average individual assignment probability (Q_i) for the identified
175 K value were performed using the software CLUMPAK (Kopelman et al. 2015). Subsequently, separate re-analysis of
176 each inferred cluster was carried out to identify finer levels of structure (Pisa et al. 2015; Vergara et al. 2015).

177 **Spatial structure in genetic variation**

178 Global estimates of genetic differentiation were examined for mtDNA sequences using Φ_{ST} and for
179 microsatellite data using F_{ST} , as defined by Weir and Cockerham (1984), implemented in Arlequin v3.5 and FSTAT
180 software. Hierarchical analysis of molecular variance (AMOVA) was used to investigate the partitioning of genetic
181 variance (using both Φ_{ST} and F_{ST}) by watershed (Fig 1), by populations within each watershed group and within
182 populations using Arlequin. Analyses were performed with and without reintroduced populations and compared in
183 terms of relative model performance to assess the contribution of reintroductions to spatial patterns of genetic structure.

184 To explore the potential relationship between genetic variation and geographic distances between study sites, a
185 spatial principal component analysis (sPCA) (Jombart 2008a; Jombart et al. 2008b) was carried out using the R

186 software packages adegenet (Jombart 2008) and ade4 (Dray and Dufour 2007). Unlike STRUCTURE, this approach
187 does not assume Hardy Weinberg or linkage equilibrium and detects spatial genetic patterns based on both the genetic
188 variance among sampling locations and its correlation to distances among locations (Jombart et al. 2008b). The
189 resulting components are differentiated by positive and negative eigenvalues that correspond respectively to global
190 (positive spatial autocorrelation, e.g. clines) and local (negative spatial autocorrelation, e.g. clusters) patterns in genetic
191 structure. Analysis was performed using an inverse distance-weighting network so that all populations were considered
192 neighbours. Population lagged scores for components showing the highest level of variation and spatial autocorrelation
193 were mapped to reveal spatial patterns of interest. A Monte Carlo simulation (spca_randtest from the Adegenet
194 package) of 1000 permutations was used to test for significant patterns of global and local structure (Montano and
195 Jombart 2017).

196 **RESULTS**

197 **Mitochondrial DNA variation and phylogenetic structure**

198 A total of 14 novel water vole haplotypes, defined by 34 polymorphic nucleotide sites, were detected across 68
199 mtDNA sequences obtained from the 12 sampling locations. Two haplotypes, S02 and S08, were most common,
200 occurring at frequencies of 0.34 and 0.24 respectively. Five populations shared haplotype S02, which was the most
201 widely distributed, spanning all three watersheds, whilst haplotype S08 was found in four populations that were
202 distributed across the CS and NK watersheds. The other 12 haplotypes were observed at frequencies between 0.01 and
203 0.1 including three haplotypes that were represented each by a single individual (S07, S10 and S14 from sites CC, PL
204 and RD respectively). Over half (64%) of the haplotypes were restricted to a single population and 86% were private to
205 a single watershed, with the highest number being observed in NK, where five of the seven NK haplotypes were unique
206 to this watershed (Fig 2). Final consensus sequences were obtained for 68 individuals (one to 16 individuals per
207 population) and have been deposited in GenBank under accession numbers MH734256–MH734270.

208

209 **FIGURE 2 HERE**

210

211 Overall, haplotype diversity (h) across the whole southeast region appears to be high, $h = 0.8196$, and by
212 watershed ranged from 0.66 in CS to 0.79 in NK. Discounting populations with a sample size of one ($N_{MT} = 1$), four
213 populations were monomorphic, and these included one of the two known reintroduction sites (LWr), whilst haplotype
214 diversity across the remaining populations of the whole region ranged from 0.33 to 0.80. Nucleotide diversity (π) for
215 these same populations ranged from 0.0005 to a high of 0.0069, which was observed at the reintroduction site (AWr).

216 Amongst watersheds, the highest nucleotide diversity was observed in haplotypes sampled from NK ($\pi = 0.0048$).
217 Summary molecular diversity indices are presented in Table 1.

218

219 TABLE 1 HERE

220

221 Statistical parsimony successfully resolved relationships among 13 of the 14 haplotypes, showing most
222 haplotypes (84%) differing by a single mutational step, whilst two haplotypes (S11 and S13) showed higher levels of
223 divergence with the number of mutational steps ranging from three to nine. The remaining haplotype (S14) was
224 parsimoniously connected at the 95% level only when including the sequences of UK water voles described by Piernney
225 et al. (2005). This resulted in two separate networks and showed the S14 haplotype to be more closely associated with
226 Scottish water vole populations than the English/Welsh (Fig 3a). To bring this divergence into context, phylogenetic
227 analysis, incorporating the UK sequences described by Piernney et al. (2005), was conducted (Fig 3b). This revealed the
228 well-supported assignment of haplotype S14 to the Scottish clade and of haplotype S11 to an English/Welsh sub-clade
229 consisting of haplotypes originating from the southwest region of England. These assignments do not represent the
230 geographical affinity of either haplotype S11, which was derived from the reintroduced population AWr or haplotype
231 S14, which originated from site RD in the NK watershed. The remaining haplotypes form a well-supported regional
232 subclade (posterior probability = 0.89), which joins sequences from the neighbouring counties of Oxfordshire and
233 Hampshire.

234

235 FIGURE 3 HERE

236

237 **Microsatellite variation and Bayesian analysis of population structure**

238 Summary statistics for microsatellite variation are shown in Table 2. Overall, a total of 96 alleles were detected
239 across all eight microsatellite loci with the average number of alleles across loci ranging from 2.5 to 9.13. A total of 18
240 private alleles were observed, half ($n = 9$) of which were restricted to a single watershed (MHP), where both
241 populations (CC and MM) had the highest number of private alleles compared with the other study sites. Allelic
242 richness was highest in the NK watershed ($A_R = 2.74$) and, between populations, ranged from 2.03 (RD) to 2.93 (AWr).
243 Observed heterozygosity was also highest at AWr ($H_O = 0.73$), contributing to the high heterozygosity observed within
244 the CS watershed and was lower than expected in eight populations, of which five showed significant heterozygosity
245 deficit (F_{IS} range: 0.065 to 0.166). The ConStruct analysis of the relative contributions of close kin breeding and cryptic

246 population substructure indicated that excess homozygosity in all five populations was consistent with both inbreeding
247 and cryptic substructure (Fig S1A–E, Supplementary Material).

248

249 TABLE 2 HERE

250

251 The number of clusters calculated by STRUCTURE was $K = 2$ based on the ΔK method (Evanno et al. 2005)
252 (Fig 4a and b). At $K = 2$, populations from the MHP watershed were assigned to one cluster, C1, with respective
253 membership coefficients (Q_P) of 0.99 for both populations (MM and CC). The second cluster (C2) included populations
254 from the CS and NK watersheds of which only one population HB had population membership coefficient (Q_P) of >0.9 .
255 The remaining six populations showed varying levels of admixture (Q_P range = 0.53 to 0.86) (Fig 4c). Further
256 substructure was apparent as the Delta K showed a relatively large decline after $K = 3$ (Fig 4a). At $K = 3$ the C2 cluster
257 was separated into two groups, one comprising population HB ($Q_P = 0.94$) and the other including all other populations
258 from the CS and NK watershed (Q_P range = 0.73 to 0.99) (Fig 4b). Within this third cluster, only 13% of individuals, all
259 from site PL, were assigned with a posterior probability of $Q_i > 0.9$ suggesting the remaining populations (AWr, LWr,
260 EM, RD and SM) have admixed ancestry.

261

262 FIGURE 4 HERE

263

264 Separate population structure analysis for clusters C1 and C2 at $K = 2$ were performed in order to identify finer
265 levels of structure. The analysis of the C1 group, which included individuals from population MM and CC, assigned all
266 individuals to their population of origin with Q_P values of 0.99 and 0.97 respectively (Fig S2a). Analysis of the C2
267 group also identified two clusters and separated population HB from the remaining populations, confirming the results
268 of the major cluster at $K = 3$ (Fig S2b). Corresponding membership coefficients to each cluster were > 0.85 for each
269 population.

270 Spatial structure in genetic variation

271 Significant genetic structure was identified across the study region for mtDNA, $\Phi_{ST} = 0.198$ ($p = <0.001$) and
272 microsatellite data, $F_{ST} = 0.126$ ($p = <0.001$), which increased when discounting reintroduced populations for both
273 mtDNA, $\Phi_{ST} = 0.494$ ($p = <0.001$) and microsatellite data $F_{ST} = 0.151$ ($p = <0.001$).

274 The analyses of molecular variance showed significant structure among watersheds for mtDNA from natural
275 populations which was not evident when haplotypes from reintroduced populations (AWr, LWr and RD) were included
276 in the model (Table 3). Significant partitioning of both mtDNA and microsatellite variation was also present among

277 populations within watershed groups regardless of the inclusion of reintroduced populations, yet more mtDNA variation
278 was attributable to this level of structure when considering reintroduced populations in the analyses (“All populations”).
279 For both models, the majority of microsatellite variation (>80%) was present within each sampled population.
280

281 TABLE 3 HERE

282

283 The results of the sPCA analysis revealed that the first two positive components, sPC-1 and sPC-2, showed the
284 greatest magnitude and accounted for 54% of the total variance (Fig 5a). The first eigenvalue ($\lambda_1 = 0.19$) which showed
285 the largest variance and spatial autocorrelation confirmed the differentiation between the MHP watershed and the
286 remaining populations observed in the STRUCTURE analysis (Fig 4c and 5b). The second eigenvector ($\lambda_2 = 0.11$) was
287 weaker and showed no clear structure in the dataset. Permutation tests on the sPCA found no evidence for global ($\lambda =$
288 $0.29, p = 0.180$) or local ($\lambda = 0.29, p = 0.998$) structuring across populations in southeast England.
289

289

290 FIGURE 5 HERE

291

292 DISCUSSION

293 Our study aimed to determine how natural genetic variation is structured amongst regional water vole
294 populations and to what extent population structure has been influenced by reintroductions. We found that genetic
295 structure exists among populations at a regional scale and, for the most part, this appears to represent natural population
296 subdivision. However, reintroductions have left a genetic footprint that is evident in both mitochondrial haplotypes and
297 nuclear genotypes.

298 At a national scale our phylogenetic analysis showed 12 of the 14 control region haplotypes formed a well-
299 supported regional subclade within the English and Welsh water voles previously published by Pierniey et al. (2005)
300 (Fig 3b). This corresponds with the significant regional genetic subdivision of water voles identified by Pierniey et al.
301 (2005) and provides further support that regional population subdivision exists between water vole populations in the
302 UK. The short branch lengths within this subclade and pattern of common, widespread haplotypes (S02 and S08)
303 connecting to other population specific haplotypes shown in the network analysis (Fig 3a) is also consistent with the
304 post-glacial range expansion of water voles in the UK proposed by both Pierniey et al. (2005) and Brace et al. (2016).
305 Two haplotypes, S11 from population AWR and S14 from population RD, however, were anomalous, showing closer
306 affiliation to haplotypes retrieved from other regions of the UK (Pierniey et al. 2005). The most striking of these was
307 the well supported assignment of haplotype S14 to the Scottish clade of water voles. Brace et al. (2016) also found a

308 Scottish haplotype present in a population of water voles on the Humber Estuary located >200 km south of the Scottish
309 border, indicating this is not an isolated occurrence. Whilst it is not impossible that these represent relic populations of
310 the Scottish ancestors that previously occupied the whole of the UK (ca 12–8 kyr BP, Brace et al. 2016), we consider it
311 more plausible that this anomalous haplotype was introduced. The closest known reintroduction site to RD is located 5
312 km upstream and occurred 2 years prior to our study (Table S1). Gene flow at this scale, over this time period, can be
313 expected for water voles as this is equivalent to 2–3 generations in which dispersal of both sexes is frequent for
314 distances up to 2 km (Telfer et al. 2003; Aars et al. 2006). The distribution of haplotype S11, however, appears to
315 remain restricted to site AWR where it was introduced ≥ 7 years prior to our study. This is despite the proximity of
316 neighbouring study sites (HB and PB, located 3.5 and 10 km upstream) being within an expected range and timescale
317 for natural dispersal between sites. As we found some evidence suggesting genetic exchange between AWR and
318 neighbouring population HB at a nuclear level (Fig 4b and 5b), it is reasonable to assume that this haplotype has spread
319 to local populations and was either undetected in our samples or has been eliminated due to genetic drift.

320 Despite the introduction of non-native haplotypes, there remains a legacy of native mtDNA diversity in the
321 region which appears to be structured by watershed. Of the 11 haplotypes found in native populations, (discounting
322 private haplotypes S11, S12 and S14 which we assume to have been introduced into sites AWR and RD), 82% were
323 private to a single watershed. The majority (88%) of these were observed in the MHP ($n = 4$) and NK ($n = 4$)
324 watersheds, which, interestingly, correspond to ‘key areas’ for water voles that are recognised for supporting a
325 significant proportion of viable source populations, regionally for MHP and nationally for NK (Strachan et al. 2011;
326 McGuire et al. 2014). Our data supports this, suggesting that the current criteria for selecting key areas in which to
327 prioritise conservation resources (Strachan et al. 2011) is appropriate for safeguarding important reservoirs of genetic
328 diversity in this species. The geographical pattern of haplotypes within these watersheds suggest that river topology is
329 an important landscape component that has contributed to the genetic structure in maternal lineages of water voles. This
330 was confirmed by the AMOVA model for “Natural populations” (Table 3) which revealed a significant proportion ($\sigma^2 =$
331 0.36) of native mtDNA variation was attributable to differences between watersheds ($\Phi_{CT} = 0.356$, $p = 0.01$), and that
332 this proportion was twice ($\sigma^2 = 0.18$) that of the variation among populations within watersheds ($\Phi_{CT} = 0.276$, $p = 0.01$).
333 As most native haplotypes within watershed groups differed by a single mutational step, this pattern appears to reflect
334 historical isolation and subsequent divergence of populations between watersheds, supporting the notion of Pierniey et
335 al. (2005) that appropriate MUs are likely to be at this scale.

336 When considering microsatellite genotypes, only MHP was found to represent a differentiated watershed group
337 following STRUCTURE analysis (Fig 4c) and this is likely to be contributed by the high number of private alleles
338 observed in this watershed (Table 2). Although the sPCA analysis did not reveal any significant global, or local patterns

in our data, this division of the MHP watershed was nonetheless present (Fig 5b). The remaining populations that span the NK and CS watershed did not show any differentiation at $K = 2$, and this was confirmed by the sPCA analysis which showed relatively small differences in the lagged scores between these same populations (Fig 5b). Given the fast mutation rate of microsatellites (Hancock 1999) and limited dispersal capabilities of water voles (<2 km) (Stoddard 1970; Telfer et al. 2003; Aars et al. 2006), homogenisation of nuclear DNA at this scale was surprising and suggests factors, other than geographic distribution, are influencing these patterns. Previous large-scale analysis of water vole populations inhabiting upland areas in Scotland found genetic similarity reduced at a scale of 20 km, suggesting the gene pool in upland areas are at least at this scale (Aars et al. 2006). There has been no comparable study for lowland populations of water voles, however, our findings based on the STRUCTURE and sPCA analysis suggests genetic similarities at a nuclear level exist between populations distributed across two discrete watersheds and in some instances, over distances in excess of 100 km. One factor that may have contributed is the inclusion of reintroduced populations AWr (from CS) and LWr (from NK). Individuals of Kent origin formed part of the mixed stock released into AWr between 1999 and 2005 and this may also be the case for LWr where 60 of 192 individuals introduced a decade prior to our study were of unknown origin (Table S1). This notion is partly supported by our mtDNA results which showed these populations share a common haplotype (S02), though not exclusively, with a natural population (SM) in the North Kent watershed. For these populations to have retained similarities at a nuclear level, they would have to have remained isolated and large enough following their release to have avoided a substantial change in allele frequencies that result from migration and/or drift. This is certainly possible for both reintroduction sites which appear to have relatively high heterozygosity and allelic richness compared to native populations (Table 2). Individuals at both these sites were released into high quality, vacant habitat that was mink free, which are factors that have previously been shown to increase the survivorship of reintroduced water voles (Moorhouse et al. 2009). It is, therefore, not inconceivable that the reintroduced populations are similar to each other and to the other natural populations in the NK watershed due to populations being similar by common decent, rather than homogenisation by contemporary gene flow. It is unclear why the remaining natural populations (EM, SM and PL) were genetically similar as they are distributed across two disjunct watersheds that are separated by downland. This is particularly notable, given that finer-scale population differentiation was observed from our STRUCTURE analysis between neighbouring populations CC and MM (in MHP) and HB and AWr (in CS). It is possible that these populations have also been influenced by human-mediated translocations, given that documentation of translocations is often absent. However, it is possible that genetic subdivision was masked by the high proportion of variation (86%) in nuclear DNA that was attributable to differences within populations in our AMOVA analysis (Jost 2008). Rosenberg (2001), for example, showed that the most useful markers for clustering populations of known origin are those that display larger variation across, rather than within,

populations. Their study also showed true assignment increased when using loci with high expected heterozygosity, yet in our study, over half of the populations showed heterozygosity deficits (Table 2 and S1, supplementary material). As previous studies of water voles have found, Hardy-Weinberg disequilibrium is common and variable in magnitude and direction between populations (Stewart et al. 1999; Aars et al. 2006), increasing the number of loci may only resolve genetic clustering for populations where close-kin inbreeding and/or substructure is absent.

The genetic markers used in this study gave partially congruent results in the sense that they both showed separation between the MHP populations and other populations from the remaining study region. However, the mtDNA control region dataset, provided higher resolution for identifying regional structure and reintroduction histories, than microsatellite markers that showed relatively little geographic structure. Discordance between mtDNA and microsatellites is common (Johnson et al. 2003). Genetic drift between populations, as measured by F_{ST} , proceeds at a rate inversely proportional to the effective population size and the rates of migration and mutation. Because mtDNA has one fourth of the N_e than nuclear DNA (for any pair of parents, there is one mitochondrial allele for every four nuclear alleles), mtDNA is expected to reach drift-migration equilibrium at a faster rate than nuclear DNA, particularly when nuclear markers have a higher mutation rate. Therefore, mtDNA is more likely display isolation by distance across populations than it is with nuclear markers such as microsatellites. As discussed in Gariboldi et al. (2016), different temporal genetic patterns can emerge from microsatellite and mtDNA markers for a variety of reasons other than differences in effective population sizes, including different rates of mutation (Wan et al. 2004), the fact that multiple microsatellite markers are evaluated compared with a single mtDNA marker (Larsson et al. 2009) and, finally, through biased sex dispersal (Lawson et al. 2007). We conclude from our study that mtDNA markers are likely to be more appropriate for defining large scale patterns in genetic structuring of water voles, however, to evaluate introgression between reintroduced and natural populations, microsatellite markers may be more appropriate.

Implications for conservation

As water voles continue to decline across the UK (McGuire and Whitfield 2017), the role of conservation measures that include safeguarding extant populations and reintroducing new viable populations, are integral to securing the future of water voles nationally. In southeast England there remains a legacy of native mtDNA diversity which appears to be structured by watersheds. This population subdivision suggests that appropriate management units for water voles are at the watershed level and that introductions are a necessary conservation tool to re-establish populations in watersheds where natural re-colonisation is unlikely to occur.

The current IUCN guidelines state that individuals selected for reintroduction programmes should be of the same genetic provenance to what would occur naturally (IUCN/SCC 2013). Although arguably conservative (Weeks et al. 2011), this guidance aims to preserve locally adapted genotypes, reduce the risk that maladapted gene complexes

will fail to persist under different ecological conditions and reduce the risk of outbreeding depression (lowering of reproductive fitness in offspring), which generally increases with genetic, geographic and environmental distance (Frankham et al. 2011). We revealed, however, that divergent lineages have been introduced into the region, including a haplotype of Scottish provenance. Given the presumably thousands of years of independent evolution that has occurred between the Scottish and English/Welsh water voles (Piertney et al. 2005; Brace et al. 2016), the translocation of individuals between these ESUs is concerning, particularly as this is not an isolated occurrence (Brace et al. 2016). Mixing genetically distinct populations exhibiting low levels of genetic diversity, such as in captive-bred stocks, can restore genetic vigour and population fitness (Frankham 2015), however, evidence also suggests this may still be the case when mixing only slightly diverged sub-populations (Biebach and Keller 2012). As genetic footprints of past introductions have been demonstrated for a number of mammalian species including red squirrels, *Sciurus vulgaris* (Simpson et al. 2013), American mink, *Neovison vison* (Garcia et al. 2017) and orang-utans, *Pongo spp.* (Banes et al. 2016), consideration needs to be given to the introduction of appropriate founder populations to avoid these risks. An example being the genetic assessment of the Eurasian beaver, *Castor fiber*, for informing source population(s) for reintroductions to Scotland (Senn et al. 2014).

Given that our study has shown that regional populations can, collectively, exhibit high levels of genetic diversity, even in the presence of possibly high degrees of close-kin inbreeding, we argue that regional stocks, at least in the south east, are currently likely to provide adequate diversity for the purposes of “local” reintroductions. It may, however, be beneficial to avoid populations containing non-native haplotypes to mitigate their spread and protect native genotypes within the region. Nevertheless, this situation may need to be reconsidered in the future should their environment present novel pressures as a consequence of, for example, climate change, which, historically, has been identified as a significant influence on water vole colonization events (Brace et al. 2016). The potentially high levels of close-kin inbreeding identified in some populations certainly warrant some future investigation. Although close-kin inbreeding, *per se*, is not in itself expected to diminish genetic diversity at the level of the population in the absence of inbreeding depression, the potential for inbreeding to diminish census number in vulnerable populations cannot be ignored in future conservation efforts. With this in mind, reintroductions and translocations are nonetheless becoming increasingly used for water vole conservation in the UK and, genetic data, as considered here, will be a useful approach in determining appropriate source populations and for assessing the genetic effects on populations following their release.

ACKNOWLEDGEMENTS

We are grateful to all the landowners and volunteers for support with field work. In particular, we would like to thank Chloe Sadler and Fran Southgate from the Wildlife Trusts, Paul Stevens and Adam Salmon from the Wildfowl

432 and Wetlands Trust, and Jane Reeves and volunteers from the Manhood Wildlife Trust and Heritage Group. We would
 433 also like to acknowledge Hampshire Biodiversity Information Centre, Hampshire Mammal Group, Kent Biodiversity
 434 Records Centre, Kent Mammal Group and Greenspace Information for Greater London for kindly providing
 435 distributional data on water voles. Lastly, we would also like to thank Inga Zeisset from the University of Brighton for
 436 her valuable comments on the manuscript. This project was part funded by the Arun & Rother Connections Heritage
 437 Lottery Fund, the Environment Agency and the Royal Geographical Society with IBS Postgraduate Research Award.

438 REFERENCES

- 439 Aars J, Dallas JF, Piertney SB, Marshall F, Gow JL, Telfer S, Lambin X (2006) Widespread gene flow and high genetic
 440 variability in populations of water voles *Arvicola terrestris* in patchy habitats, *Molecular Ecology* 15(6):1455–
 441 1466
- 442 Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Malden,
 443 Massachusetts, USA
- 444 Baker R (2013) Demographic and genetic patterns of water voles in human modified landscapes: implications for
 445 conservation. PhD Thesis, University of Brighton, UK.
- 446 Banes GL, Galdikas BMF, Vigilant L (2016) Reintroduction of confiscated and displaced mammals risks outbreeding
 447 and introgression in natural populations, as evidenced by organ-utans of divergent subspecies. *Scientific*
 448 *Reports* 6: 22–26
- 449 Barreto GR, Rushton SP, Strachan R, Macdonald DW (1998) The role of habitat and mink predation in determining the
 450 status of water voles in England. *Animal Conservation* 2:53–61
- 451 Berthier K, Galan M Foltête JC, Charbonnel N, Cosson JF (2005) Genetic structure of the cyclic fossorial water vole
 452 (*Arvicola terrestris*): Landscape and demographic influences. *Molecular Ecology* 14(9):2861–2871
- 453 Biebach I and Keller LF (2012) Genetic variation depends more on admixture than number of founders in reintroduced
 454 Alpine ibex populations. *Biological Conservation* 147(1):197–203
- 455 Brace S, Ruddy M, Miller R, Schreve DC, Stewart JR, Barnes I (2016) The colonization history of British water vole
 456 (*Arvicola amphibius* (Linneus, 1758)): Origins and development of the Celtic Fringe. *Proceedings of the Royal*
 457 *Society B* 283:20160130
- 458 Costello AB, Down TE, Pollard SM, Pacas CJ, Taylor EB (2003) The influence of history and contemporary stream
 459 hydrology on the evolution of genetic diversity within species: an examination of microsatellite DNA variation
 460 in bull trout, *Salvelinus confluentus* (Pisces: Salmonidae). *Evolution* 57(2):328–345
- 461 Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular ecology*
 462 9(10):1657–1659

463 Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to
464 intraspecific phylogeny reconstruction. *Genetics* 134(3):959–969

465 Dean M, Strachan R, Gow D, Andrews R (2016) *The Water Vole Mitigation Handbook* (The Mammal Society
466 Guidance Series). Eds F. Mathews and P. Chanin. The Mammal Society, London

467 Dray S, Dufour A (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical*
468 *Software* 22(4):1-20

469 Earl DA, vonHoldt BM (2011) Structure Harvester: A website and program for visualizing structure output and
470 implementing the Evanno method. *Conservation Genetics Resources* 4(2):359–361

471 Edmands S (2007) Between and rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for
472 conservation and management. *Molecular Ecology* 16(3):463–475

473 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: A
474 simulation study. *Molecular Ecology* 14(8):2611–2620

475 Excoffier L, Lischer HEL (2010) ARLEQUIN Suite Ver 3.5: A new series of programs to perform population genetics
476 analyses under Linux and Windows. *Molecular Ecology Resources* 10(3):564–567

477 Ferguson A (1989) Genetic differences among brown trout, *Salmo trutta*, stocks and their importance for the
478 conservation and management of the species. *Freshwater Biology* 21:35–46

479 Frankham R (2015) Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of
480 gene flow. *Molecular Ecology* 24: 2610–2618

481 Frankham R, Ballou JD, Eldridge MD, Lacy RC, Ralls K, Dudash MR, Fenster CB (2011) Predicting the Probability of
482 Outbreeding Depression. *Conservation Biology* 25:465–475

483 Fisher DO, Lambin X, Yletyinen SM (2008) Experimental translocation of juvenile water voles in a Scottish lowland
484 metapopulation. *Population Ecology* 51(2):289–295

485 Frankham R (2005) Genetics and extinction. *Biological Conservation* 126:131–140

486 Garcia K, Melero Y, Palazón, Gosálbec J, Castresana J (2017) Spatial mixing of mitochondrial lineages and greater
487 genetic diversity in some invasive populations of the American mink (*Neovison vison*) compared to native
488 populations. *Biological invasions* 19(9):2663–2673

489 Gariboldi MC, Túnez JI, Failla M, Hevia M, Panebianco MV, Viola MNP, Vitullo AD, Cappozzo HL (2016) Patterns
490 of population structure at microsatellite and mitochondrial DNA markers in the franciscana dolphin
491 (*Pontoporia blainvillei*) *Ecology and Evolution* 6: 8764–8776.

492 Goudet J (1995) FSTAT (Version 1.2) A computer program to calculate F-Statistics, *Journal of Heredity* 86(6):485–486

493 Goudet J (2002) FSTAT, A program to estimate and test gene diversities and fixation indices (Version 2.9.3.2).
 494 Available: <http://www.unil.ch/izea/software/fstat>

495 Hancock JM (1999) Microsatellites and other simple sequences: Genomic context and mutational mechanisms. In
 496 Goldstein DB, Schlotterer C (eds) Microsatellites: Evolution and Applications. Oxford University Press. New
 497 York

498 Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6(2):65-70

499 Houde ALS, Fraser DJ, O'Reilly P, Hutchings JA (2011) Relative risks of inbreeding and outbreeding depression in the
 500 wild in endangered salmon. Evolutionary Applications 4(5):634–647

501 Huff DD, Miller LM, Chizinski CJ, Vondracek B (2011) Mixed-source reintroductions lead to outbreeding depression
 502 in second-generation descendants of a native North American fish. Molecular Ecology 20:4246–4258

503 Igea J, Aymerich P, FernándeX-González A, González-Esteban J, Gómez A, Alonso R, Castresana J (2013)
 504 Phylogeography and postglacial expansion of the endangered semi-aquatic mammal *Galemys pyrenaicus*.
 505 BMC Evolutionary Biology 13:115

506 IUCN/SSC (2013) Guidelines for Reintroductions and Other Conservation Translocations. Version 1.0. Gland,
 507 Switzerland: IUCN Species Survival Commission viiii + 57pp.

508 Jefferies DJ, Morris PA, Mulleneux JE (1989) An enquiry into the changing status of the water vole *Arvicola terrestris*
 509 in Britain. Mammal Review 19:111–131

510 Johnson JA, Toepfer JE, Dunn PO (2003) Contrasting patterns of mitochondrial and microsatellite population structure
 511 in fragmented populations of greater prairie-chickens Molecular Ecology 12: 3335–3347.

512 Johnson, N. (2008) Hybrid incompatibility and speciation. Nature Education 1(1):20.

513 Jombart, T, (2008a) adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24,
 514 1403e1405.

515 Jombart T, Devillard S, Dufour AB, Pontier D (2008b) Revealing cryptic spatial patterns in genetic variability by a new
 516 multivariate method. Heredity 101: 92–103.

517 Jost L (2008) G(ST) and its relatives do not measure differentiation. Molecular Ecology 17:4015–26

518 JNCC (1995) Biodiversity: The UK Steering Group Report. Volume 2: Action Plans. HMSO. London

519 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: A program for identifying
 520 clustering modes and packaging population structure inferences across K. Molecular Ecology Resources
 521 15(5):1179–1191

522 Larsson LC, Charlier J, Laikre L, Ryman N (2009) Statistical power for detecting genetic divergence—organelle versus
 523 nuclear markers. Conservation Genetics, 10, 1255–1264.

524 Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular*
525 *Ecology*, 16, 1559–1578.

526 Lind AJ, Spinks, PQ, Fellers GM, Bradley Shaffer H (2011) Rangewide phylogeography and landscape genetics of the
527 Western U.S. endemic frog *Rana boylei* (*Ranidae*): implications for the conservation of frogs and rivers.
528 *Conservation Genetics* 12:269–284

529 Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology* 19:3038-
530 3051

531 Macdonald DW, Strachan R (1999) The mink and the water vole: analyses for conservation. Wildlife and Conservation
532 Research Unit, Oxford

533 Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population
534 genetics. *Trends in Ecology and Evolution* 18(4):189–197

535 McGuire C, Whitfield D (2017) National Water Vole Database and Mapping Project: Part 1 Project Report 2006–2015.
536 Hampshire and Isle of Wight Wildlife Trust, Curdridge

537 McGuire C, Whitfield D, Perkins H, Owen C (2014) National Water Vole Database and Mapping Project: Guide to the
538 Use of Project Outputs to End of 2012. Available:
539 [http://hiwwt.org.uk/sites/default/files/files/Reports/Guide%20to%20the%20Use%20of%20Project%20Output](http://hiwwt.org.uk/sites/default/files/files/Reports/Guide%20to%20the%20Use%20of%20Project%20Outputs%202014%20comp.pdf)
540 [ts%202014%20comp.pdf](http://hiwwt.org.uk/sites/default/files/files/Reports/Guide%20to%20the%20Use%20of%20Project%20Outputs%202014%20comp.pdf) Accessed 4 February 2014

541 Montano V and Jombart T (2017) An Eigenvalue test for spatial principal component analysis. *BMC Bioinformatics*
542 18:562

543 Moorhouse TP, Gelling M, Macdonald DW (2009) Effects of habitat quality upon reintroduction success in water voles:
544 Evidence from a replicated experiment. *Biological Conservation* 142:53–60

545 Moritz C (1994) Defining ‘Evolutionarily Significant Units’ for conservation. *Trends in Ecology and Evolution*
546 9(10):373–375

547 Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic*
548 *Biology* 51(2):238–254

549 Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York

550 Overall, ADJ (2016) ConStruct 1.0: An R Script to distinguish between substructure and consanguinity within a
551 population using multilocus microsatellite data. *Cogent Biology*, 2: 1128317.

552 Pasbøll PJ, Berube M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends*
553 *in Ecology and Evolution* 22:11–16

554 Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in excel. Population genetic software for teaching and
555 research. *Molecular Ecology Notes* 6(1):288–295

556 Peakall R, Smouse PE (2012) GENALEX 6.5: Genetic analysis in excel. Population genetic software for teaching and
557 research. *Bioinformatics* 28(19):2537–2539

558 Pfeiffer I, Völkel I, Täubert H, Brenig B (2004) Forensic DNA-typing of dog hair: DNA-extraction and PCR
559 amplification. *Forensic Science International* 141(2):149–151

560 Piertney SB, Stewart WA, Lambin X, Telfer S, Aars ON, Dallas JF (2005) Phylogeographic structure and postglacial
561 evolutionary history of water voles (*Arvicola terrestris*) in the United Kingdom. *Molecular Ecology*
562 14(5):1435–1444

563 Pisa G, Orioli V, Spilotros G, Fabbri E, Randi E, Bani L (2015) Detecting a hierarchical genetic population structure:
564 The case study of the fire salamander (*Salamandra salamandra*) in Northern Italy. *Ecology and Evolution*
565 5(3):743–758

566 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data,
567 *Genetics* 155(2):945–95

568 Querejeta M, Fernández-González A, Romero R, Castresana J (2017) Postglacial dispersal patterns and mitochondrial
569 genetic structure of the Pyrenean desman (*Galemys pyrenaicus*) in the northwestern region of the Iberian
570 Peninsula. *Ecology and Evolution* 7:4486–4495

571 Rambaut A (2006–2014) Figtree: Tree figure drawing tool v1.4.0. Institute of Evolutionary Biology Web.
572 <http://tree.bio.ed.ac.uk/software/figtree>. Accessed 9 July 2014.

573 Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology* 17:230–237

574 Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43(1):223–225

575 Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck
576 JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model
577 space. *Systematic Biology* 61(3):539–542

578 Rosenberg NA, Burke T, Elo K et al. (2001) Empirical evaluation of genetic clustering methods using multilocus
579 genotypes from 20 chicken breeds. *Genetics* 159(2): 699–713.

580 Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution*
581 1:9–10

582 Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*
583 4:11

584 Schwartz MK, Luikart G, Waples RS (2005) Genetic monitoring as a promising tool for conservation and management.
585 Trends in Ecology and Evolution 22:25–33

586 Senn H, Ogden R, Frosch C et al. (2014) Nuclear and mitochondrial genetic structure in the Eurasian beaver (*Castor*
587 *fiber*) – implications for future reintroductions. Evolutionary Applications 7: 645–662.

588 Simpson S, Blampied N, Peniche G, Dozières, Blankett, T, Coleman, S, Cornish N, Groombridge JJ (2013) Genetic
589 structures of introduced populations: 120-year-old DNA footprint of historic introduction in an insular small
590 mammal population. Ecology and Evolution 3(3):614–628

591 Stewart WA, Piertney SB, Dallas JF (1998) Isolation and characterization of highly polymorphic microsatellites in the
592 water vole, *Arvicola terrestris*. Molecular Ecology 7:1258–1259

593 Stoddart M (1970) Individual range, dispersion and dispersal in a population of water voles (*Arvicola terrestris* (L.).
594 Journal of Animal Ecology 37:403–424.

595 Strachan R, Jefferies DJ (1993) The water vole *Arvicola terrestris* in Britain 1989-1990: its distribution and changing
596 status. The Vincent Wildlife Trust. London

597 Strachan R, Moorhouse T, Gelling M (2011) Water vole conservation handbook 3rd Edition. The Wildlife Conservation
598 Unit, Abingdon, UK

599 Strachan C, Strachan R, Jefferies DJ (2000) Preliminary report on the changes in the water vole as shown by the
600 National Surveys of 1989–1990 and 1996–1998. The Vincent Wildlife Trust. London

601 Stewart WA, Dallas JF, Piertney SB, Marshall F, Lambin X, Telfer S. (1999) Metapopulation genetic structure in the
602 water vole, *Arvicola terrestris*, in NE Scotland. Biological Journal of the Linnean Society, 68, 159–171.

603 Stucky BJ (2012) Seqtrace: A graphical tool for rapidly processing DNA sequencing chromatograms. Journal of
604 Biomolecular Techniques 23(3):90–93

605 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics
606 123:585–595

607 Telfer S, Piertney SB, Dallas JF, Stewart WA, Marshal F, Gow J, Lambin X (2003) Parentage assignment reveals
608 widespread and large-scale dispersal in water voles. Molecular Ecology 12:1939–1951

609 Van-Oosterhout C, Hutchinson WF, Wills DPD, Shipley P (2004) Micro-checker: Software for identifying and
610 correcting genotyping errors in microsatellite data, Molecular Ecology Notes 4(3):535–538

611 Verkuil YI, Juillet C, Lank DB, Widemo F, Piersma T (2014) Genetic variation in nuclear and mitochondrial
612 markers supports a large sex difference in lifetime reproductive skew in a lekking species. Ecology and
613 Evolution
614 4(18): 3626-3632.

615 Vergara M, Basto MP, Madeira MJ, Gómez-Moliner BJ, Santos-Reis M, Fernandes C, Ruiz-González A (2015)
616 Inferring population genetic structure in widely and continuously distributed carnivores: The stone marten as a
617 case study. PLoS ONE 10(7):e0134257

618 Wan QH, Wu H, Fujihara T, Fang SG (2004) Which genetic marker for which conservation genetics issue?
619 Electrophoresis, 25, 2165–2176.

620 Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying
621 the number of gene pools and their degree of connectivity. Molecular Ecology 15(6):1419–1439

622 Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, Byrne M, Coates DJ, Eldridge MDB,
623 Sunnucks P, Breed MR, James, EA, Hoffmann AA (2011) Assessing the benefits and risks of translocations in
624 changing environments: a genetic perspective. Evolutionary Applications 4(6):709–725

625 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–
626 1370

627 Woodroffe GL, Lawton JH, Davidson WL (1990) The impact of American mink *Mustela vison* on water voles *Arvicola*
628 *terrestris* in the North Yorkshire Moors National Park. Biological Conservation 51:49–62

629

630 **Fig 1** Map of study area showing location of genetic sampling sites symbolised by watershed groups, where NK =
631 North Kent, CS = Central Sussex and MHP = Manhood Peninsular, water vole distribution by 5 km grid squares based
632 on biological records from the last 10 years, county boundaries, river topology and known reintroduction sites
633 (McGuire and Whitfield 2017). Codes and the purported origin (where R = reintroduced and N = natural) for each site
634 are shown in inset table and main UK regions (where E = East, SE = South West, Wa = Wales, Mid = Midlands, Y =
635 Yorkshire, NW = North West, NE = North East and Sc = Scotland) are shown in inset map.

636

637 **Fig 2** Map showing distribution of southeast water vole haplotypes which are represented by colour, their proportional
638 occurrence across sample sites and sized according to the number of individuals sampled per site (N = 1–16)

639

640 **Fig 3** Genealogical relationships amongst novel haplotypes identified in this study (S01-02, S04-15) and the UK (H#
641 shown in grey, as described by Piertney et al. (2005)) where: **a)** shows the final haplotype networks derived from
642 maximum parsimony of S14 with Scottish haplotypes (top) and the remaining 13 haplotypes identified in southeast
643 England (S#) (bottom), and where **b)** shows the final consensus tree based on Bayesian analyses of phylogenetic
644 structure amongst all haplotypes described. Far right shows sampling locations for southeast haplotypes and regional
645 origin of closest affiliated haplotypes (where SW = South West, SWa = South Wales, Mid = Midlands, NW = North
646 West and SE = South East, as shown in Fig 1)

647

648 **Fig 4** Population structure of water voles in southeast England based on average assignment probabilities from
649 microsatellite data using STRUCTURE where: **a)** shows plot of Delta K with peak at K = 2 (most likely main structure)
650 and a high rate of change between Delta K = 3 and K = 4 suggesting further substructure at K = 3; **b)** histogram where
651 each bar represents an individual's probability of assignment (Qi) to each major group for K = 2 (top) and K=3
652 (bottom) and where **b)** shows a map of the proportion of individuals assigned to each cluster. Unique colours (black or
653 white) in b and c represent the same major cluster (K = 2) derived using 20 iterations of K = 1 to 10

654

655 **Fig 5** Results obtained by spatial principal component analysis (sPCA) using inverse distance-weighting and
656 microsatellite variation among nine water vole populations in southeast England. **(a)** Barplot of eigenvalues, where
657 positive eigenvalues (on the left) correspond to global structures and negative eigenvalues (on the right) correspond to
658 local structures. **(b)** The population lagged scores associated with the first eigenvalues of the sPCA. The size of each
659 square is proportional to the absolute value of the population's score and represents each population's position relative to
660 the overall genetic structure of the population. The colour (black or white) of the square corresponds to the sign of the
661 score (positive or negative, respectively)

662

663

664

665

666

667

668

Contains National Statistics data © Crown copyright and database right 2018. Contains Ordnance Survey data © Crown copyright and database right 2018. County water vole data for Hampshire [© Hampshire Biodiversity Information Centre Partnership], Greater London [© Greenspace Information for Greater London], Kent [© Kent and Medway Biological Records Centre], Surrey [© Surrey Wildlife Trust] and Sussex [© Sussex Biological Records Centre].

Contains National Statistics and Ordnance Survey data © Crown copyright and database right 2018. Contains Natural England data © Natural England copyright.

Contains National Statistics and Ordnance Survey data © Crown copyright and database right 2018. Contains Natural England data © Natural England copyright.

690 Figure 5
691

Contains National Statistics and Ordnance Survey data © Crown copyright and database right 2018.
Contains Natural England data © Natural England copyright.

692
693

694 **Table 1** Molecular diversity estimates based on mitochondrial DNA (mtDNA) control region sequences for each study
695 site, watershed (bold) and overall (SE) (where N_{MT} = sample size, N_H = number of haplotypes, P_H = number of private
696 haplotypes, h = haplotype diversity and π = nucleotide diversity). Standard deviation for diversity indices are provided
697 in parenthesis
698

Site	Mitochondrial diversity				
	N_{MT}	N_H	P_H	h	π
LWr	5	1	0	0.000	0.000
RD	1	1	1	0.000	0.000
DM	2	1	1	0.000	0.000
EM	5	3	1	0.800 (0.164)	0.003 (0.002)
OM	5	2	0	0.600 (0.175)	0.003 (0.003)
SM	6	2	0	0.333 (0.215)	0.001 (0.001)
NK	24	7	5	0.786 (0.065)	0.005 (0.003)
AWr	7	3	2	0.667 (0.160)	0.007 (0.004)
PB	5	1	0	0.000	0.000
HB	5	1	0	0.000	0.000
PL	6	2	1	0.333 (0.215)	0.001 (0.001)
CS	23	5	3	0.557 (0.109)	0.003 (0.002)
CC	5	3	1	0.700 (0.218)	0.002 (0.001)
MM	16	4	2	0.692 (0.074)	0.002 (0.002)
MHP	21	5	4	0.719 (0.066)	0.002 (0.001)
SE	68	14	14	0.820 (0.032)	0.004 (0.002)

699

700 **Table 2** Genetic diversity estimates based on eight microsatellite markers for each study site, watershed (bold) and
701 overall (SE) (where N_{MS} = sample size, N_A = average number of alleles across eight loci, P_A = number of private alleles,
702 A_R = allelic richness, H_O = observed heterozygosity, H_E = expected heterozygosity and F_{IS} with the proportion of
703 randomisations that gave a larger F_{IS} than the observed, based on 72000 randomisations. Indicative adjusted nominal
704 level (5%): 0.0007. Standard deviation for diversity indices are provided in parenthesis

Site	Microsatellite diversity							
	N_{MS}	N_A	P_A	A_R	H_O	H_E	F_{IS}	Prop _{larger}
LWr	18	5.00 (1.60)	0	2.68 (0.27)	0.70 (0.17)	0.73 (0.08)	0.039	0.2186
RD	4	2.50 (1.07)	1	2.03 (0.72)	0.47 (0.36)	0.47 (0.29)	0.001	0.5453
EM	34	6.63 (1.19)	0	2.75 (0.33)	0.64 (0.15)	0.74 (0.09)	0.126	0.0002
SM	24	6.38 (2.92)	0	2.67 (0.57)	0.66 (0.15)	0.70 (0.17)	0.057	0.0781
NK	80	9.38 (2.56)	2	2.54 (0.57)	0.65 (0.11)	0.81 (0.07)		
AWr	129	9.13 (2.80)	1	2.93 (0.16)	0.73 (0.06)	0.78 (0.04)	0.065	0.0001
HB	97	7.88 (2.47)	2	2.64 (0.34)	0.59 (0.09)	0.71 (0.09)	0.166	0.0000
PL	30	5.50 (1.31)	1	2.58 (0.46)	0.57 (0.23)	0.68 (0.15)	0.164	0.0001
CS	236	10.13 (3.23)	7	2.71 (0.36)	0.66 (0.05)	0.79 (0.06)		
CC	36	6.75 (1.65)	3	2.71 (0.41)	0.64 (0.11)	0.72 (0.12)	0.111	0.0004
MM	81	7.50 (1.60)	2	2.80 (0.25)	0.70 (0.11)	0.75 (0.07)	0.068	0.0007
MHP	117	9.00 (1.51)	9	2.76 (0.33)	0.68 (0.10)	0.77 (0.09)		
SE	453	12.00 (2.93)		2.68 (0.27)	0.66 (0.03)	0.82 (0.04)		

705 **Table 3** Hierarchical analysis of molecular variance (AMOVA) for mtDNA and microsatellite data showing
 706 significance of genetic structure among watersheds (F_{CT}), among populations within watersheds (F_{SC}) and among all
 707 populations (F_{ST}). Results are shown for all populations and for natural populations, where reintroduced populations
 708 (lineages) have been omitted from analysis

Source of Variation	mtDNA			Microsatellites		
	df	% Variation	F-statistics	df	% Variation	F-statistics
All populations						
Among watersheds	2	3.29	$F_{CT} = 0.033$	2	2.15	$F_{CT} = 0.022^*$
Among populations within watersheds	9	50.27	$F_{SC} = 0.536^{**}$	6	11.04	$F_{SC} = 0.113^{**}$
Within populations	56	46.44	$F_{ST} = 0.520^{**}$	897	86.81	$F_{ST} = 0.132^{**}$
Natural populations						
Among watersheds	2	35.55	$F_{CT} = 0.356^{**}$	2	1.39	$F_{CT} = 0.014$
Among populations within watersheds	6	17.77	$F_{SC} = 0.276^{**}$	4	13.91	$F_{SC} = 0.141^{**}$
Within populations	46	46.68	$F_{ST} = 0.533^{**}$	605	84.70	$F_{ST} = 0.153^{**}$

*Significant at $P = 0.05$. **Significant at $P = 0.01$.

729 **Genetic structure of regional water vole populations and footprints of reintroductions: a case study from**
730 **Southeast England**
731
732 Conservation Genetics
733
734 Rowenna J Baker¹; Dawn M Scott²; Peter J King³; Andrew DJ Overall^{1*}

735 *Affiliations and addresses:*

736 ¹ Ecology, Conservation and Zoonosis Research Group, Huxley Building, University of Brighton, Brighton BN2 4GJ,
737 UK.

738 ² School of Life Sciences, Keele University, Keele, Staffordshire, ST5 5BG

739 ³ Ouse & Adur River's Trust, Barcombe, East Sussex BN8 5BW, UK.

740
741 * Corresponding author: Email: a.d.j.overall@brighton.ac.uk. Telephone: +441273642099
742

743 **Supporting Information**

744
745 **Table S1** Documented reintroductions of water voles within the southeast England study region showing year of
746 release, number (N) origin of individuals introduced (where known) and watershed where they were released.
747 Locations of release sites are mapped in Figure 1.
748

Location	Watershed	Year	N	Origin
London Wetlands (LWr) ^{1,2}	MHP	2001-2003	192	Berkshire, Hampshire & individuals of unknown origin
Dartford Park ^{1,2}		2009-2010	221	Unknown
West Malling ¹		Unknown	Unknown	Unknown
Ruxley Gravel Pits ^{1,2}		2002	200	From development site
Rainham Marsh ¹		Unknown	Unknown	Unknown
Arundel Wetlands ^{1,2}	CC	1999 - 2005	231	Kent, Hampshire, Derbyshire, Wales & 60 individuals of unknown origin
Pagham Harbour ^{1,2}	MHP	2002	10	Kent, Hampshire, Derbyshire, Wales

749 ¹ McGuire and Whitfield 2017

750 ² Unpublished material gathered by authors
751
752

753 **S1: Analysis of inbreeding**

754
755 F_{IS} is a measure of the homozygosity in excess of Hardy-Weinberg expectations within a single, randomly mating
756 population. This is often taken to be a measure of close-kin inbreeding, or consanguinity. However, if the
757 population being analysed is in fact made up of genetically divergent, but cryptic, subpopulations, then excess
758 homozygosity results as a consequence of this and F_{IS} should be interpreted as F_{ST} . This is known as the
759 Wahlund effect (Hartl & Clark, 2007). Estimates of F_{IS} were significantly greater than expectations for five
760 populations identified in Table 1. To distinguish between the contribution consanguinity and cryptic population
761 substructure make to excess homozygosity, these populations were analysed using the software ConStruct
762 (Overall, 2016). The ConStruct software initially estimates the excess homozygosity (F) using maximum
763 likelihood. For example, for the AW population sample, the maximum likelihood value of $F = 0.0625$. This
764 could be interpreted as 100% of the individuals having parents related as 1st cousins and none of the excess due
765 to cryptic substructure (i.e., $F_{ST} = 0$ and $C_g = 1.0$ (the proportion of the sample inbred to degree r , where here r
766 $= 0.0625$)). However, as the contour plot in Figure S1.A shows, the joint maximum likelihood estimates of F_{ST}
767 and C_g are 0.02264826 and 0.46, respectively. This is interpreted as evidence for both consanguinity *and* cryptic
768 substructure within this sample. The outermost envelope of the contour plots corresponds with the support limit
769 for the maximum likelihood value exceeding $F = 0$ (Overall, 2016).
770
771
772

775
776 **Figure S1.A:** Joint maximum likelihood estimates of F_{ST} and C_g for population AW, where $F = 0.0625$.

777
778 **Figure S1.B:** Joint maximum likelihood estimates of F_{ST} and C_g for population HB, where $F = 0.166$.
779

780
781 **Figure S1.C:** Joint maximum likelihood estimates of F_{ST} and C_g for population CC, where $F = 0.09$.

782
783 **Figure S1.D:** Joint maximum likelihood estimates of F_{ST} and C_g for population PL, where $F = 0.159$.
784

Figure S1.E: Joint maximum likelihood estimates of F_{ST} and C_g for population EM, where $F=0.09$.

Figure S2 Substructure of water vole populations within two major genetic clusters ($K = 2$) identified from microsatellite data using STRUCTURE where **a)** Independent runs for cluster 1 (C1) showing $K = 2$ and **b)** independent runs for cluster 2 (C2) showing $K = 2$. Each vertical bar represents an individual and each colour represents the probability of assignment to each cluster.

References

- Hartl D, Clark A (2007) Principles of Population Genetics, Ed. 4. Sinauer Associates, Inc., Sunderland, MA.
- McGuire C, Whitfield D (2017) National Water Vole Database and Mapping Project: Part 1 Project Report 2006-2015. Hampshire and Isle of Wight Wildlife Trust, Curdridge.
- Overall, ADJ (2016) ConStruct 1.0: An R Script to distinguish between substructure and consanguinity within a population using multilocus microsatellite data. *Cogent Biology*, **2**: 1128317.