



**HAL**  
open science

# Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits

M. Szulkin, P.-A. Gagnaire, N. Bierne, A. Charmantier

► **To cite this version:**

M. Szulkin, P.-A. Gagnaire, N. Bierne, A. Charmantier. Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Molecular Ecology*, 2016, 25 (2), pp.542-558. 10.1111/mec.13486 . hal-02326694

**HAL Id: hal-02326694**

**<https://hal.science/hal-02326694v1>**

Submitted on 22 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**MOLECULAR ECOLOGY****Population genomic footprints of fine-scale differentiation  
between habitats in Mediterranean blue tits**

Journal:	<i>Molecular Ecology</i>
Manuscript ID	MEC-15-0774.R3
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Szulkin, Marta; CEFE, UMR 5175 CNRS Gagnaire, Pierre-Alexandre; ISEM, UMR 5554 CNRS Bierne, Nicolas; Institut des Sciences de l'Evolution, Integrative Genomics Charmantier, Anne; CEFE, UMR 5175 CNRS
Keywords:	blue tit, Population Genetics - Empirical, Landscape Genetics, local adaptation, RAD sequencing, genomic differentiation

1 **Population genomic footprints of fine-scale differentiation between**  
2 **habitats in Mediterranean blue tits**

3

4 ***M. Szulkin*<sup>1§\*</sup>, *P.-A. Gagnaire*<sup>2,3\*</sup>, *N. Bierne*<sup>2,3</sup> & *A. Charmantier*<sup>1</sup>**

5 <sup>1</sup> Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 Campus CNRS, 1919 Route de Mende,  
6 34293 Montpellier cedex 5, France

7 <sup>2</sup> Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

8 <sup>3</sup> ISEM - CNRS, UMR 5554, SMEL, 2 rue des Chantiers, 34200 Sète, France

9 \* Joint first authors

10 § corresponding author

11

12 **Keywords:** blue tit, population genomics, landscape genetics, RAD sequencing, local adaptation,  
13 genetic differentiation

14 **Address of corresponding author:**

15 Marta Szulkin

16 CEFE CNRS

17 1919, route de Mende

18 34293 Montpellier cedex 5, France

19 Fax number: +33 4 67 61 33 36

20 Email: marta.szulkin@zoo.ox.ac.uk

21 **Running title:** population genomics of wild blue tits

## 22 **Abstract**

23 Linking population genetic variation to the spatial heterogeneity of the environment is of  
24 fundamental interest to evolutionary biology and ecology, in particular when phenotypic differences  
25 between populations are observed at biologically small spatial scales. Here, we applied restriction-  
26 site associated DNA sequencing (RAD-Seq) to test whether phenotypically differentiated populations  
27 of wild blue tits (*Cyanistes caeruleus*) breeding in a highly heterogeneous environment exhibit  
28 genetic structure related to habitat type. Using 12106 SNPs in 197 individuals from deciduous and  
29 evergreen oak woodlands, we applied complementary population genomic analyses, which revealed  
30 that genetic variation is influenced by both geographical distance and habitat type. A fine-scale  
31 genetic differentiation supported by genome- and transcriptome-wide analyses was found within  
32 Corsica, between two adjacent habitats where blue tits exhibit marked differences in breeding time  
33 while nesting less than 6 km apart. Using redundancy analysis (RDA), we show that genomic variation  
34 remains associated with habitat type when controlling for spatial and temporal effects. Finally, our  
35 results suggest that the observed patterns of genomic differentiation were not driven by a small  
36 proportion of highly differentiated loci, but rather emerged through a process such as habitat choice,  
37 which reduces gene flow between habitats across the entire genome. The pattern of genomic  
38 isolation-by-environment closely matches differentiation observed at the phenotypic level, thereby  
39 offering significant potential for future inference of phenotype-genotype associations in a  
40 heterogeneous environment.

## 41 Introduction

42 The evolutionary tug-of-war between local adaptation and counteracting gene flow is a fascinating  
43 biological process which plays a key role in shaping genetic and phenotypic diversity of natural  
44 populations. In the absence of gene flow and other evolutionary constraints (such as genetic  
45 correlations, or a lack of adaptive genetic variation), divergent selection should cause each local  
46 population to evolve traits providing an advantage in its local habitat (Kawecki & Ebert 2004).  
47 However, local adaptation may be limited by gene flow, especially if habitat patch size is small  
48 relative to the scale of dispersal (Lenormand 2002; Slatkin 1973, 1987).

49 Genetic evidence for local adaptation is usually inferred indirectly by searching for molecular  
50 signatures of selection, with the implicit expectation that selection varies across environments  
51 (Barrett and Hoekstra 2011). Another important prediction is that genetic differentiation should  
52 correlate with environmental variables independently of geographic distance - a pattern commonly  
53 referred to as Genetic-Environment Association (GEA; Hedrick *et al.* 1976) or Isolation-by-  
54 Environment (IBE; Wang & Bradburg 2014).

55 Gene flow mediated through dispersal is a key element opposing the effect of local adaptation.  
56 Consequently, gene flow across habitats has long been assumed to preclude adaptive differentiation,  
57 thereby preventing the evolution of marked intraspecific phenotypic differences at small spatial  
58 scales in highly mobile organisms such as birds (Garant *et al.* 2007; Slatkin 1987) or marine species  
59 (Palumbi 1994). More recently, the importance of non-random gene flow through matching habitat  
60 choice has received increased theoretical and empirical attention (see Edelaar *et al.* 2012; Edelaar &  
61 Bolnick 2012; Edelaar *et al.* 2008; Ravigne *et al.* 2009). However, the extent to which individuals  
62 choose to settle in the habitat that maximizes their fitness with respect to their phenotype remains

63 poorly understood, and so are the consequences of matching habitat choice on the evolution of local  
64 adaptation (Edelaar *et al.* 2008).

65 The long-term monitoring of several populations of a small passerine bird, the blue tit *Cyanistes*  
66 *caeruleus* breeding in a highly heterogeneous habitat in Southern France (Blondel *et al.* 2006)  
67 revealed multiple lines of evidence offering scope for non-random dispersal and habitat-dependent  
68 selection: marked phenotypic differences in lay date, clutch size, number of fledglings and  
69 morphometric traits can be observed not only between the Southern French mainland and Corsica  
70 (Figure 1), but also between two Corsican populations residing 27 km apart (Blondel *et al.* 1999;  
71 Blondel *et al.* 2006; Lambrechts *et al.* 1997) (Figure 1). Even more strikingly, these differences were  
72 also observed at a finer scale between two Corsican populations located only 5.6 km apart within the  
73 same valley (Blondel *et al.* 2006, Figure 1). In addition, quantitative genetic models revealed that all  
74 of these traits harbour significant genetic variation (reviewed in Charmantier *et al.* in press, Blondel  
75 *et al.* 2006).

76 The phenotypic differences observed in this study system are expected to be driven by habitat  
77 heterogeneity, and in particular by the type of oak species dominating the habitat where blue tits  
78 breed (Blondel *et al.* 2001; Blondel *et al.* 2006; Lambrechts *et al.* 2004). Indeed, Mediterranean  
79 habitats are interspersed with distinct patches of either evergreen (holm oak *Quercus ilex*) or  
80 deciduous (downy oak *Quercus pubescens*) oak populations. Oak type influences the entire food  
81 chain blue tits depend on to feed their young: First, the ca. one month time-lag in leaf development  
82 between evergreen and deciduous oaks translates in a time-lag in oak leaf-feeding caterpillar hatch  
83 dates. Second, temporally contrasted caterpillar availability (the primary food source of blue tit  
84 nestlings) triggers shifts in the distribution of blue tit breeding time between habitat patches. As a  
85 result, blue tit populations breeding 27 km from each other in a heterogeneous environment  
86 including evergreen and deciduous habitat patches, with no clear-cut boundaries limiting dispersal

87 (such as open spaces generated by large crop fields), start to breed on average at a one month  
88 difference from each other (Blondel *et al.* 2006). This temporal breeding shift recorded between  
89 “early” deciduous habitats and “late” evergreen habitats at a small geographical scale is recurrently  
90 noted at larger, but also at smaller spatial scales when blue tit populations from several evergreen  
91 and deciduous oak habitats are compared (Blondel *et al.* 2001; Blondel *et al.* 2006; Lambrechts *et al.*  
92 1997; Porlier *et al.* 2012a; Szulkin *et al.* in press). This metapopulation blue tit study system thus  
93 offers a particularly suitable model to test for Isolation-by-Environment (IBE) over short geographical  
94 distances.

95 Available evidence for a genetic basis to differences in habitat-specific laying date was originally  
96 deduced from common-garden experiments (Blondel *et al.* 1990; Lambrechts *et al.* 1997). In nature,  
97 genetic differences between habitats were also found over large (Corsica vs French mainland) and  
98 small (27 km in Corsica) spatial scales using microsatellite markers (Porlier *et al.* 2012b). However, no  
99 differentiation could be evidenced between habitat patches located 5.6 km apart within the same  
100 Corsican valley. This suggests that either gene flow at such a fine spatial scale homogenises allelic  
101 frequencies within the valley, or that genetic inference made from a limited number of neutral  
102 markers was underpowered to detect biologically significant fine-scale population structure visible at  
103 the phenotypic level (Figure 1). In this context, the potential of large single nucleotide  
104 polymorphisms (SNPs) datasets may be of particular interest to increase the power and resolution in  
105 the detection of fine-scale genetic structure and IBE.

106 Here, we used Restriction-site associated DNA sequencing (RAD seq) to generate a high density SNP  
107 dataset covering the entire blue tit genome and characterise genetic patterns of diversity in blue tit  
108 populations from Southern France (mainland and the island of Corsica). First, we present a general  
109 overview of the *de novo* strategy to obtain genome-scale polymorphism data in this wild passerine  
110 bird with no reference genome. We emphasise the usefulness of reporting checkpoints for data

111 validation throughout the analysis pipeline, by matching genome-wide estimates of relatedness with  
112 field and microsatellite-based pedigrees, and by including blind-sequencing control samples to  
113 estimate genotyping repeatability. Second, we investigated whether the previously described  
114 habitat-specific phenotypic differentiation is corroborated by genetic differentiation between  
115 populations at the genomic and transcriptomic levels. We evaluated the robustness of our results by  
116 controlling for the effect of rare variants, family relatedness, sample size and variation in individual  
117 birth year. We also took advantage of earlier molecular work in the study population to test whether  
118 microsatellite and SNP data concur in population genetic estimates of population differentiation. In  
119 particular, we aimed to confirm or refute (i) earlier reports suggesting a lack of genetic  
120 differentiation between two phenotypically contrasted blue tit populations located 5.6 km apart with  
121 no dispersal barrier between them, as well as (ii) the general role of the environment in creating  
122 habitat-dependent genetic structuring that is independent of geographical distance.

123

## 124 **Materials & Methods**

### 125 **Study system and data collection**

126 The blue tit *Cyanistes caeruleus* is a small resident passerine bird of the tit (Paridae) family, breeding  
127 throughout temperate Europe and western Asia in deciduous or mixed woodlands (Snow 1954). In  
128 this study, we sampled 197 blue tits breeding in nestboxes as part of a long-term monitoring survey  
129 (Blondel *et al.* 2006; Charmantier *et al.* in press). Study sites include a forest in southern French  
130 mainland near Montpellier (“D-Rouviere”, with “D” for deciduous habitat), where blue tits belong to  
131 the continental nominal subspecies (*C. caeruleus caeruleus*), and 3 locations on Corsica (“E-Muro”,  
132 “D-Muro” and “E-Pirio”, with “E” for evergreen habitat – see Figure 2 and Table 1). Corsican blue tit  
133 populations belong to the subspecies *C. caeruleus ogliastreae*, which is *ca.* 15% smaller compared to



134 its mainland relative (Martin 1991). All four populations breed in a mosaic of heterogeneous habitats  
135 containing in majority, among other tree species, interspersed patches of deciduous downy oak  
136 (*Quercus pubescens*) and evergreen holm oak (*Quercus ilex*) (Figure 2, Table 1). In Corsica, habitat  
137 type is known to be associated with marked differences in the timing of breeding and reproductive  
138 success at a small geographic scale (Blondel *et al.* 1999; Blondel *et al.* 2006; Porlier *et al.* 2012a)  
139 (Figure 1). These populations are described in further detail in previous studies (Blondel *et al.* 2006;  
140 Charmantier *et al.* in press; Porlier *et al.* 2012b).

141 Birds were captured in the nestboxes when offspring were between 9 and 15 days old; their identity  
142 and morphometric measurements were recorded, and 7-30  $\mu$ l of blood was taken from a small neck  
143 vein – a method deemed safer relative to the risk of hematomas and flight impairment caused by  
144 sampling in the wing. A total of 197 birds were selected; all were residents - i.e. they were born and  
145 later recruited as breeding individuals in one of the 4 locations of interest (Table 1). Birth year varied  
146 between 1991 and 2008, with an average birth year in 2002. Maternal and paternal identities were  
147 obtained from field observations, which enabled us to identify 93.5% and 95% of social fathers and  
148 mothers of birds in our dataset, respectively. GPS coordinates were measured for most nestboxes  
149 using a handheld GPS device (Garmin GPSMAP 62S). Missing nestbox coordinates were retrieved  
150 using annotated maps of the study sites.

151

#### 152 **RAD libraries construction and sequencing**

153 Blue tit blood samples used in this study were stored in Queen's buffer, and DNA extraction was  
154 performed using Qiagen DNeasy Blood & Tissue kits. DNA extractions were quantified using a  
155 NanoDrop ND8000 spectrophotometer and a Qubit 2.0 fluorometer with the DNA HS assay kit (Life  
156 Technologies), and checked for DNA quality after migration on agarose gel to select samples with  
157 appropriate DNA concentration ( $>20\text{ng}/\mu\text{l}$ ) and molecular weight ( $>10\ 000\text{bp}$ ).

158 In order to assess the repeatability of library construction and to evaluate the rate of genotyping  
159 errors, DNA from 5 of the 197 individuals were replicated as follows: DNA from 4 individuals were  
160 extracted twice and the DNA extract of one individual was split into two samples which were  
161 analysed independently. In total, 202 DNA extracts were sent to Floragenex Inc. for library  
162 preparation and single-end sequencing according to the original protocol (Baird *et al.* 2008). We used  
163 the restriction enzyme *Sbf*I, which targets an 8-bp cutting site (5'CCTGCAGG3'). Each individual RAD  
164 library was ligated to a unique molecular identifier (a 6bp DNA barcode) before sample multiplexing  
165 was performed in equimolar proportions by groups of 29 individuals per pool. Each pool was then  
166 sequenced on one lane of an Illumina HiSeq 2000 instrument, generating 101-bp single reads which  
167 were further automatically trimmed to 91bp reads. As recommended by Meirmans *et al.* (2015),  
168 samples were assigned to sequencing lanes in a randomised fashion, and samples from each of the 4  
169 populations were present in all 7 or 8 (out of 8) sequencing lanes used in the study.

170

## 171 **Bioinformatics**

172 Short sequence reads were quality filtered and demultiplexed using individual barcode information.  
173 We used the *Stacks* pipeline (Catchen *et al.* 2013; Catchen *et al.* 2011) to identify loci *de novo*,  
174 discover SNPs and infer individual genotypes. Preliminary runs were performed to determine the  
175 most appropriate parameter combination for *UStacks*. SNPs were detected at each locus using the  
176 maximum likelihood approach under the 'snp' model. We empirically determined an optimal  
177 minimum depth of coverage of 5 reads per allele ( $m = 5$ ) and a maximum of 3 nucleotide mismatches  
178 between alleles ( $M = 3$ ). Increasing the number of mismatches between alleles did not allow to  
179 retrieve many more SNPs while increasing the risk of merging paralogs, as detected by HWE tests. A  
180 catalog of loci found across all individuals was then built using *CStacks*, allowing a maximum number  
181 of 3 mismatches between two homozygous individuals at a same locus ( $n = 3$ ). *De novo* loci

182 constructed with *UStacks* were then searched against the catalog of loci using *SStacks*. Finally, we  
183 used the *Stacks* module *populations* to retain only loci that were successfully genotyped in at least  
184 50% of the individuals from at least 2 populations. Individual genotypes were outputted as a VCF file  
185 which was submitted to further downstream filtering.

186 The increasing sequencing error rate toward the end of reads produced an elevation in the total  
187 number of SNPs being called from position 85 to 91bp. Therefore, all variable sites located after  
188 position 84 of the reads were removed from the VCF file. We further filtered the SNP dataset based  
189 on several quality and population-genetic criteria to only retain highly reliable SNPs using *VCFtools*  
190 (Danecek *et al.* 2011) (Table 2). After removing the five individual replicates, we excluded SNPs  
191 showing strong deviations to Hardy-Weinberg equilibrium (HWE) within at least one of the three  
192 locations (D-Rouviere, Muro and E-Pirio) using a p-value threshold of 0.01 (D-Muro and E-Muro were  
193 pooled together due to low sample size in E-Muro and close physical proximity). This filtering step  
194 aimed to remove poor-quality SNPs and artefactual variation due to the merging of paralogous  
195 sequences, but was insensitive to small deviations from HWE resulting from subtle within-population  
196 structure. The dataset was then filtered to only retain loci that were genotyped in at least 90% of all  
197 samples (90% call rate), and with a global Minor Allelic Frequency (MAF) of at least 2%. Details on the  
198 number of SNPs retained at subsequent bioinformatics filtering steps are presented in Table 2.

199

#### 200 **Genome-wide relatedness between individuals**

201 To infer SNP-based relatedness structure within populations, we calculated pairwise identity-by-state  
202 (IBS) coefficients between all possible pairs of individuals within the mainland (D- Rouviere) and  
203 within Corsica, as well as for the 5 pairs of individual replicates using the R package *SNPRelate* (Zheng  
204 *et al.* 2012). Contrasting RAD-seq derived pairwise IBS estimates with independently acquired  
205 information about the relatedness between any two given individuals can be used as a valuable data

206 processing check point throughout the analysis pipeline. For example, such contrasts can be used (1)  
207 as a quality control check to confirm that samples were not mixed-up in the lab preparation and  
208 sequencing stages (this complements the repeatability values of extracts sequenced independently),  
209 (2) to confirm that RAD-seq data, microsatellite data (if available) and pedigree data concur, and  
210 finally (3) to verify that RAD-seq derived IBS values are linearly related to independently established  
211 relatedness values. Genome-wide relatedness values estimated from IBS coefficients were therefore  
212 compared with (i) expected relatedness values for 5 full siblings ( $r = 0.5$ ) confirmed with  
213 microsatellite genotyping data (Charmantier *et al.* 2004), (ii) 10 mother-offspring pairs ( $r=0.5$ ) and (iii)  
214 5 maternal half-siblings ( $r=0.25$ ) established using social pedigree information. Individual replicates  
215 were further used to evaluate the overall genotyping repeatability of the RAD marker dataset.  
216 Because absolute (unscaled) values of IBS are influenced by the population's allele frequency  
217 spectrum, we provide the folded allele frequency spectrum along with the genome-wide relatedness  
218 distributions of mainland and Corsica separately (Figure 3).

219

## 220 **Analyses of population genomic structure**

221 We used a combination of complementary population and landscape genetics analyses to evaluate  
222 the extent of genetic structure within and among Mediterranean blue tit populations. Table 2  
223 summarises the different nested datasets used for analysis and their associated number of markers.

224 We first evaluated the extent of population differentiation between the French mainland and  
225 Corsica, as well as between Corsican populations. Principal Component Analysis (PCA) implemented  
226 in the R package SNPRelate (Zheng *et al.* 2012) was used to illustrate population structure at  
227 different scales (*i.e.* Mainland-Corsica, between samples within Corsica, between habitats within the  
228 Muro valley).

229 We estimated the genome-wide average genetic differentiation between each pair of populations  
230 and habitats by computing Nei's pairwise  $F_{ST}$  using the *adegenet* R package (Jombart 2008; Jombart &  
231 Ahmed 2011) after applying an MAF threshold of 5% ( $n=3159$  SNPs, see Table 2). The significance of  
232 pairwise  $F_{ST}$  values was tested through 500 random permutations of the genotypes among  
233 populations.

234 To further investigate the spatial scale of genetic variation within populations, we performed spatial  
235 PCA analyses, a method for detecting spatial patterns that are not always associated with the  
236 principal components of genetic variation detected in standard PCA. Spatial PCA uses connection  
237 networks to separate the product of the genetic variance between individuals and their spatial  
238 autocorrelation into negative and positive components (Jombart *et al.* 2008). For instance, global  
239 structures, such as two spatial groups, or a cline, will display positive spatial autocorrelation (Moran's  
240  $I$ , Moran 1950) that can be inferred from allelic frequency data. We applied the "Global Test"  
241 (Jombart *et al.* 2008) to test for global structures against the null hypothesis of no genetic structure  
242 in the population. Spatial PCA was performed using the *adegenet* package in R (Jombart 2008;  
243 Jombart & Ahmed 2011), with a connection network of 15 nearest neighbours for Muro, Pirió and  
244 Rouvière (and 12 neighbours in the "no family ties dataset"). Because (i) the power of principal  
245 component based methods usually scales with the product between the number of individuals and  
246 the number of markers (Patterson *et al.* 2006), and because (ii) the detection of fine-scale population  
247 structures may benefit from the inclusion of rare variants (O'Connor *et al.* 2015), we used an MAF  
248 threshold of 2% to maximise dataset size for this analysis ( $n= 12\ 106$  SNPs, see Table 2).

249 Genetic variation among blue tit populations may be confounded by several factors including habitat  
250 type, geography, individual birth year and individual sequencing depth (expected to reflect  
251 genotyping accuracy). We therefore used constrained ordination to specifically test the marginal  
252 effect of each factor on the distribution of samples genotypes. Here, we used redundancy analysis

253 (RDA), a constrained ordination method implemented in the *Vegan* package (Oksanen *et al.* 2014) in  
254 *R* to infer the extent to which available environmental, but also experiment-dependent variables are  
255 influencing SNP genotypic variation in the dataset (see Meirmans *et al.* 2015). A key strength of this  
256 analysis is to provide a statistical means for inferring the effect of partially confounded variables  
257 separately. The following initial model was used:  $Y$  (individual genotype)  $\sim$  Latitude + Longitude +  
258 Habitat + Birth year + Number of Reads. To assess whether the different variables significantly  
259 influenced allele frequencies, we first used permutation tests to assess the global significance of the  
260 RDA by performing 1000 permutations where the genotypic data were permuted randomly and the  
261 model was refitted. Second, the significance of each individual variable was tested by running an RDA  
262 marginal effects permutation test (with 1000 permutations) where we removed each term one by  
263 one from the model containing all other terms. Non-significant effects were removed from the final  
264 model. This procedure was implemented both for all birds in the dataset and for Corsican birds only.

265 To establish the role of habitat independently from other sources of genetic variation (i.e. the  
266 remaining explanatory variables in the final model), we performed conditioned (partial) RDA where  
267 the effects of all significant explanatory variables but habitat were removed from the ordination by  
268 using the *condition* function:  $Y \sim \text{habitat} + \text{condition}(\text{remaining significant variables in the final}$   
269  $\text{model})$ .

270 Finally, the distribution of SNP contributions to the single RDA habitat axis after conditioning on  
271 remaining variables was compared to that obtained for conditioned RDA estimating the specific  
272 effect of geography or birth year (equivalently conditioned by all other significant variables in the  
273 final model). It is expected that directional selection on loci conferring adaptation to habitat type will  
274 generate outlier SNPs in the distribution of SNP contributions to the effect of habitat. Therefore, the  
275 distribution of SNP contributions to the conditioned effect of habitat should differ from the  
276 conditioned effect of geography or birth year if the habitat-dependent IBE pattern is mainly driven by

277 directional selection.

278

### 279 **Controlling for family relatedness**

280 The genetic sampling of free-living animals is frequently made without considering the underlying  
281 family genetic structure, which is often unknown during sampling. Moreover, it is often difficult to  
282 specify what constitutes a bias when sampling relatives (Szulkin *et al.* 2013): while relatives are part  
283 of a null distribution of genetic relatedness for a given animal system, the distribution of relatedness  
284 in a sampled population may become skewed because of field work protocols, for example due to  
285 site and nestbox fidelity. To control for a possible bias induced by relatedness in our inference of blue  
286 tit genetic structuring in Corsica and mainland France, we reran PCA, sPCA and Fst analyses using a  
287 “no family ties” dataset. Out of the 197 individuals in the original dataset, we removed closely  
288 related individuals (offspring or siblings), reducing the original dataset to 119 individuals. In the case  
289 of full siblings, we conserved the sibling with the largest number of reads (e.g. of best genotyping  
290 quality). The “no family ties” sample sizes per population are as follows: D-Muro = 36, E-Muro = 8, E-  
291 Pirio = 43, D-Rouviere = 32; see also Part II of Sup. Mat. for more details).

292

### 293 **Controlling for inequalities in sample size, sex composition and birth year**

294 Because sample size varied among the four sampled populations, reaching the lowest value of 9  
295 individuals in E-Muro, we generated a “symmetrical minimal dataset” where all individuals from E-  
296 Muro were complemented by 9 individuals from each of the remaining 3 populations, matching E-  
297 Muro individuals in terms of birth year ( $\pm 1$  year difference) and sex. This resulted in a dataset of 36  
298 individuals in which each population was equally represented and was homogeneous in terms of  
299 birth year and sex composition (see also Part III of Sup. Mat for more details).

### 300 **Transcriptomic variation analyses**

301 To estimate what proportion of genomic RAD tags could be identified as transcriptomic sequences,  
302 we used the full length (91 bp-long) consensus sequence of each polymorphic RAD locus to perform  
303 Blast searches against transcriptome databases. To maximise the number of annotated sequences,  
304 we used 19 760 RAD loci with a global call rate > 80% and a MAF > 1%. All loci were blasted against  
305 the three following transcriptomes with *Blastx*, using an e-value threshold of  $10^{-7}$  to retain significant  
306 matches:

307 (1) the blue tit *Cyanistes careuleus* transcriptome. RNA from blood of ten blue tits (including 4  
308 *Cyanistes c. ogliastreae* individuals from Corsica, 3 *Cyanistes c. careuleus* from D-Rouviere) was used  
309 to synthesise and sequence cDNA fragments on 454 and Illumina sequencers using a previously  
310 described protocol (Cahais *et al.* 2012; Romiguier *et al.* 2014). (2) the great tit *Parus major*  
311 transcriptome (Santure *et al.* 2011); divergence time from *Cyanistes caeruleus*: 19 million years  
312 (onezoom.org). (3) the zebra finch *Taeniopygia guttata* transcriptome, where both (a) ab-initio  
313 predicted genes and (b) cDNA transcripts, available at [www.ensembl.org](http://www.ensembl.org), were inspected. Divergence  
314 time from *Cyanistes caeruleus*: 72 million years (Hedges *et al.* 2006).

315 To determine the percentage of genomic RAD tags mapping to the transcriptome, we estimated the  
316 proportion of blue tit RAD sequences matching (i) either a blue tit or great tit transcriptomic  
317 sequence, and (ii) a blue tit, great tit or zebra finch transcriptomic sequence. In cases where more  
318 than one tag matched the same transcriptomic contig, we selected the RAD tag with the lowest E-  
319 value. Transcriptomic SNPs used to compute pairwise *Fst* values were extracted from the previously  
320 described dataset (Table 2), using only SNPs derived from RAD sequences that fully matched the blue  
321 tit transcriptome on at least 80bp (Table 2).



## 322 **Results**

323 Sequencing RAD-Tags from blue tit populations generated an average of *ca.* 4,7 million sequences  
324 per sample (median 4.5 million reads). Overall, sequencing quality control reports revealed mostly  
325 uniform, high quality sequencing across samples. One sample had a lower than expected number of  
326 reads, which translated into a lower number of RAD tags (Figure S1). On the other hand, over-  
327 sequencing resulted in an increased number of variable RAD-tags produced by sequencing errors  
328 (Figure S1). Applying HWE tests, genotyping call rate and MAF thresholds efficiently removed poorly  
329 sequenced tags and artefactual SNPs originating from sequencing errors or paralogous tags. The  
330 resulting dataset is characterised by a minimum of 90% genotype call rate (and an average of 96%), a  
331 2% MAF threshold, and contains 12 106 SNPs with an average sequencing depth of 73X per  
332 individual.

333

### 334 **Relatedness distribution and repeatability of control samples**

335 The analysis of the identity-by state (IBS) matrix calculated for both Corsican and mainland birds  
336 revealed unimodal distributions of IBS coefficients flanked by right-hand tails of high IBS values  
337 (Figure 3A&B). These distributions were further annotated with independently confirmed family links  
338 (full siblings and mother-offspring pairs), which showed that the right tails reflect the presence of  
339 close relatives in the dataset. The genotyping repeatability of RAD loci assessed with sample  
340 replicates averaged 97%, a value which was well above the range of IBS values observed in the  
341 dataset (Figure 3A&B). Genome-wide relatedness measured with IBS coefficients increased linearly  
342 with expected relatedness inferred from microsatellite and pedigree data (Figure S2). The folded  
343 allele frequency spectrum differed between the mainland and Corsica, revealing a deficit of rare  
344 variants (<10%) on the mainland (Figure 3C) compared to Corsica (Figure 3D).

345 **Strong genome- and transcriptome-wide signals of between-population differentiation**

346 Large scale as well as small scale genetic differentiation was confirmed by pairwise  $F_{st}$  analyses  
347 (Table 4). At the genomic level, we detected highly significant differentiation between Corsican  
348 populations and Southern French mainland. No significant genetic differentiation was detected  
349 between Pirio and Muro sites (27 km apart; Table 4). At the same time, a clear signal of genetic  
350 differentiation was found between Muro Evergreen and Muro Deciduous sites, two sites with  
351 contrasted vegetation cover located 5.6 km apart (Table 1). These  $F_{st}$  values were qualitatively and  
352 quantitatively similar in the “no family ties” dataset albeit 23% higher on average than in the entire  
353 dataset (Table 4). This is not surprising, since removing close relatives inflates total genetic variance  
354 more strongly than it increases within-population variance, which causes the  $F_{st}$  to increase.  
355 Importantly,  $F_{st}$  measures applied to the “no family ties” and “symmetrical minimal” datasets yielded  
356 qualitatively the same results as the full dataset containing 197 individuals (Table 4, Table S4).  $F_{st}$   
357 values derived from microsatellites (data from Porlier *et al.* 2012b) were 11% lower on average  
358 relative to those measured from genome-wide data, which is an intrinsic consequence of higher  
359 polymorphism in microsatellite markers (Edelaar *et al.* 2011; Jakobsson *et al.* 2013).

360 We further measured genetic differentiation values at the transcriptomic level. A summary of  
361 matched RAD sequences against each of the three transcriptomes (blue tit, great tit, zebra finch) is  
362 presented in Table 3. Overall, 6.1% of the RAD sequences (1202 out of 19760 sequences) in this study  
363 were matched to either great tit or blue tit transcriptome sequences, and 11.4% (2251 out of 19760  
364 sequences) aligned to either one of four available transcriptomes (blue tit, great tit, zebra finch  
365 cDNA, zebra finch *ab initio* genes). Out of the 326 RAD loci that aligned on the blue tit transcriptome  
366 (1.65%, Table 3), 179 SNPs were retained for pairwise  $F_{st}$  tests. Transcriptomic  $F_{st}$  values strongly  
367 corroborated those found genome-wide, although they were stronger on average by 24% (Table 4).

368

369 **PCA and sPCA analyses of genetic distinctiveness**

370 Genetic distinctiveness between Corsican birds (*C. caeruleus ogliastreae*) and Southern French blue  
371 tits from the mainland (*C. caeruleus caeruleus*) was evidenced by the projection of individuals on PCA  
372 axis 1, which encompasses 6.68% of the entire genetic variance (Figure 4). Genetic differentiation  
373 between the Corsican sites of Muro and Pirió was explained by PCA axis 2, which captured 1.61% of  
374 the genetic variance. PCA analyses within Corsica showed a fine scale genetic differentiation with  
375 both spatial and habitat components (Figure S3).

376 Inferring spatially explicit within-population structure using the spatial PCA method corroborated  $F_{st}$   
377 and PCA results since it revealed a significant differentiation within Muro reflecting habitat structure  
378 (Global P-value: 0.028, N=57, Table S2). In addition, we found no evidence for spatial structure within  
379 the Muro deciduous habitat (Global P-value: 0.214, N=48), the Pirió evergreen habitat (Global P-  
380 value: 0.103, N=83), or within the Rouvière habitat (Global P-value: 0.116, N=57). PCA and sPCA  
381 analyses applied to the “no family ties” dataset yielded qualitatively the same results as in the full  
382 dataset containing 197 individuals (Figure S5, Table S1, Table S2).

383

384 **Redundancy analysis reveals significant habitat and spatial components of differentiation**

385 When the 4 sites were analysed together (n=197), the proportion of constrained variance explained  
386 by the redundancy analysis (RDA) was highly significant (Table 5), thus confirming the  
387 informativeness of the constraining variables used in the full RDA model. After removing the single  
388 non-significant term (i.e. the number of reads), four constrained axes explained 9.6% of the total  
389 genotypic variance and the first two RDA axes received a large contribution of both habitat and  
390 spatial variables (Table 5, Table 6 & Table S3). Geographical location (latitude and longitude) was  
391 largely represented by RDA axis 1, whereas habitat type (deciduous or evergreen) was mainly

392 captured both by RDA axis 1 and 2, and birth year mainly by RDA axis 2 (Figure 5A).

393 Because geographical coordinates explained the largest amount of variance among individuals, we  
394 further restricted the RDA analysis to Corsican birds only (n=140) to test for habitat effects  
395 independently of the geographical distance between continent and Corsica. With the 3 Corsican sites  
396 included in the analysis (D-Muro, E-Muro, E-Pirio), the RDA was highly significant (Table 5). After  
397 removing correlated and non-significant terms (longitude and number of reads, respectively), the  
398 three constrained axes explained 3.6% of the total genotypic variance and the first two RDA axes  
399 received a large contribution of habitat and latitude (Table 6, Figure 5B).

400 The partial habitat RDA conditioned on geography and birth year revealed a significant effect of  
401 habitat after removing variation caused by the other significant factors, both for the full dataset  
402 (Table 5), the Corsican dataset (Table 5) and for the “symmetrical minimal” dataset (Table S3). Thus,  
403 habitat type (deciduous vs. evergreen oaks) was a significant predictor of genotypic variation  
404 independently of geographical distance and birth date. Interestingly, when Corsican genotypes were  
405 projected on the single habitat RDA axis conditioned for other variables (the direction of the habitat  
406 vector, Figure S4), we observed greater genetic distinctiveness between D-Muro and E-Muro than  
407 between D-Muro and E-Pirio, which reflected the  $F_{st}$  values between these populations presented in  
408 table 4. Also, a non-explained source of genotypic variance in E-Pirio was captured by the first  
409 principal component of the three partial RDAs (Figure S4). These analyses revealed the genetic  
410 distinctiveness of 17 individuals from E-Pirio (those with the most negative coordinates on PC1),  
411 which already occupied extreme positions on the axis 2 of the PCA (Fig. 4).

412 Finally, we compared the distributions of SNP contributions to the single conditioned RDA axis of 3  
413 partial RDAs that independently captured the effect of habitat, latitude and birth date in Corsica. We  
414 found that the three distributions largely overlapped, and that none of them was driven by SNPs  
415 showing large contributions (Figure S4).

416 **Discussion**

417 Dense SNP genotyping obtained by RAD sequencing and applied to a phenotypically well  
418 characterised study system of free living birds revealed significant fine-scale genetic structuring at a  
419 small spatial scale (5.6 km), a distance considered to be within the natal dispersal range of  
420 continental blue tits as estimated by Tufto *et al.* (2005) and discussed by Charmantier *et al.* (in press).  
421 Moreover, RDA analyses confirmed that genomic differentiation between populations was  
422 significantly driven by the type of oaks blue tits reproduced in, independently of geographical  
423 distance. A previous study (Porlier *et al.* 2012b) using microsatellite genetic characterisation of the  
424 two populations inhabiting the Muro valley did not reach enough power to detect a significant  
425 differentiation signal between D-Muro and E-Muro, although the order of pairwise  $F_{st}$  values was the  
426 same both here as in Porlier *et al.* (2012b). Here, we demonstrated that the two populations  
427 significantly differ from each other in terms of their allelic frequencies at the genomic and  
428 transcriptomic level, independently of the presence of individuals with close family ties in the dataset  
429 or other confounding effects such as sample size or temporal variation in birth year.

430 The possibility to identify genetic distinctiveness at such a small spatial scale in a highly mobile  
431 species unequivocally suggests that the large number of SNPs identified through RAD sequencing  
432 brings unprecedented explanatory power in elucidating weak yet distinct genetic signals harboured  
433 by populations breeding in heterogeneous habitats. The fine-scale genetic structuring coincides with  
434 evergreen and deciduous habitat patchiness in the valley of Muro, and agrees with earlier reports  
435 identifying habitat type to be instrumental in creating structure at the genetic (Porlier *et al.* 2012b)  
436 and phenotypic (Blondel *et al.* 1999; Blondel *et al.* 2006; Charmantier *et al.* in press; Lambrechts *et al.*  
437 1997) level. At the same time, the population genetic landscape of Corsican blue tit populations  
438 was found to be more complex than expected, and required the use of complementary analytical  
439 methods to unravel the potential of habitat-dependent genetic structuring. In this context, RDA

440 analysis proved particularly suitable to identify and test the effect of individual variables influencing  
441 genomic variability, while also offering the potential to detect collinearity between them. Below we  
442 discuss in detail both methodological aspects as well as key biological findings of the study.

443

#### 444 **Genetic diversity, relatedness distribution and repeatability**

445 RAD-sequencing and subsequent bioinformatic analyses resulted in identifying c. 12 000 SNPs with a  
446 2% MAF (and c. 6500 SNPs with a 5% MAF) genotyped on average in 96% of birds. Overall, single-end  
447 RAD sequencing confirmed its suitability for dense genotyping in a natural bird population with no  
448 available reference genome.

449 Allele frequency spectra differed markedly between Corsica and the mainland, due to high-levels of  
450 low-frequency polymorphisms on Corsica contrasting with a deficit of low-frequency variants on the  
451 mainland. These differences in the distributions of allele frequencies probably reflect contrasting  
452 demographic histories between mainland and Corsica. By contrast, analysing the distribution of  
453 genome-wide similarity between birds in the dataset revealed highly similar population relatedness  
454 composition on the mainland and in Corsica. These relatedness structures were characterised by a  
455 mode of unrelated individual and a right-hand tail of close relatives, that were already validated by  
456 parentage analysis based on microsatellite data from Charmantier *et al.* (2004) and field  
457 observations. The presence of family members in the dataset, sometimes associated with field  
458 sampling limitations in time and space, always constitutes an inherent yet unknown fraction of  
459 populations sampled at random when no pedigree is available. These unknown family links can thus  
460 be straight-forwardly revealed when dense SNP genotyping is available, without the need for  
461 computationally intensive genetic pedigree reconstruction. Importantly, the population genetic  
462 patterns observed in this study were confirmed using concurrent “no family ties” and “symmetrical

463 minimal” datasets, yielding qualitatively and quantitatively comparable results (see Part II and III of  
464 supplementary material).

465 Finally, genotyping repeatability scores not only provided a useful analytical control step in sample  
466 processing, but it also generated insight into the limits of genotype-by-sequencing accuracy. While  
467 next-generation sequencing repeatability scores have been reported at intermediate analytical  
468 stages (Sharma *et al.* 2012), genotyping error rates estimated at the final stages of bioinformatic  
469 analyses are rarely reported in RAD-seq studies (but see Mastretta-Yanes *et al.* 2015). Here, our  
470 strategy for inferring individual genotypes *de novo* allowed to keep the genotyping error rate below  
471 3%, a value that is close to the lowest error rates estimated by Mastretta-Yanes and colleagues  
472 (2015) in RAD-seq studies.

473

#### 474 **Population genetic structure**

475 When Corsica and the mainland population of Rouvière were compared,  $F_{st}$  values observed in this  
476 study were well within the range of genetic differentiation observed in other blue tit and great tit  
477 (*Parus major*) populations. Indeed, these studies reported  $F_{st}$  indexes that can be as low as 0.01 for  
478 the great tit between a Dutch and a UK population (Van Bers *et al.* 2012), and as high as 0.79  
479 between two insular blue tit populations in the Canary Island system (Hansson *et al.* 2014). Genetic  
480 differentiation between Corsican populations and the mainland site of Rouvière was nearly 4 times  
481 stronger than  $F_{st}$  values within Corsica, thereby confirming the genetic distinctiveness of the  
482 Corsican blue tit sub-species, and the very limited gene flow between the island and the mainland (as  
483 in Porlier *et al.* 2012b).

484 Within the island of Corsica however, genetic differentiation does not only scale with geographical  
485 distance: the strongest, and highly significant genetic differentiation between Corsican populations

486 was found between two populations inhabiting two different oak habitats within the same Muro  
487 valley, with an average nestbox distance between the two populations of 5.6 km, and 4 km between  
488 the closest nestboxes from each habitat. While the signal of genetic differentiation between the two  
489 populations is surprising, this result is particularly robust since it was not only confirmed by four  
490 complementary population genomic analyses ( $F_{st}$ , PCA, sPCA and RDA), but also when tested at  
491 genome-wide and transcriptome-wide levels, and when controlling for family structure, sample size  
492 and birth year (Table 4, Supplementary material part II and III). It is worth noting that the strength of  
493 genetic differentiation (pairwise  $F_{st}$ ) was strongest using transcriptome-derived SNPs, followed by  
494 genomic SNPs, and weakest when using microsatellites (Table 4). High-dimensional SNP datasets  
495 undoubtedly provide an increased precision and a more powerful detection of small  $F_{st}$  values than  
496 those derived from a small number of markers (Waples 1998). At the same time, higher  
497 transcriptome  $F_{st}$  values relative to those calculated genome-wide likely reflect the effect of  
498 selection at linked sites, causing local reductions in effective population size in coding regions due to  
499 purifying selection (Charlesworth *et al.* 1993). Further analyses will be required to gain better insight  
500 into which genomic, and in particular transcriptomic regions co-vary with the phenotypic differences  
501 in the study system.

502 Spatial PCA and RDA analysis explicitly tested and provided support for small-scale differentiation  
503 and the possible role of habitat in generating genomic structuration (Figure 5, Table 5, Table S2).  
504 Concurrent efforts to include a greater number of sampling sites with replicated oak habitats and  
505 nestbox-specific indicators of environmental heterogeneity (Szulkin *et al.* in press) would be valuable  
506 to fully validate the role of IBE in generating genomic structuring in this study system. Moreover, the  
507 Mediterranean blue tit study system offers decades of individual life-history and fitness measures in  
508 the 4 sites studied here, thus offering considerable potential for complementary analyses of  
509 covariation between genomic, phenotypic and environmental data.



510

511 **Isolation-by-Environment despite high capacities for gene flow**

512 One important question that needs to be further addressed by integrating ecological, behavioural  
513 and genetic data, is whether isolation-by-environment results from reduced dispersal through  
514 habitat choice, local selection against maladapted genotypes, or a combination of both. The average  
515 natal dispersal distance of blue tits on the continent is crudely estimated to range between ~330m  
516 and 4 km (depending on population and dispersal distance estimation method, (Ortego *et al.* 2011;  
517 Tufto *et al.* 2005)). However, there is variation around these average values, and there are known  
518 records of much larger blue tit natal dispersal distances (see Charmantier *et al.* in press and  
519 discussion therein). The scale of natal dispersal in Corsica is currently unknown and may be smaller  
520 than in the rest of the species range due to the insular nature of the population.

521 Given that there is no barrier to dispersal in the Corsican landscape (such as important mountain  
522 ridges or open spaces birds would be reluctant to fly over (Blondel *et al.* 2006; Porlier *et al.* 2012b)),  
523 a significant habitat-driven genetic differentiation at a 5.6 km scale in such a highly mobile species as  
524 the blue tit suggests either strong local selection capable of counteracting gene flow, or non-random  
525 dispersal of genotypes with respect to habitat type. The fact that we could not detect loci with  
526 extreme contributions to fine-scale differentiation between habitats in our partial RDA analysis  
527 (Figure S4) suggests that the signal of differentiation is genome-wide rather than driven by a subset  
528 of loci strongly influenced by selection. Admittedly, polygenic selection acting on a large suite of  
529 complex quantitative traits could generate correlated but minor allele frequency changes (Latta  
530 1998; Le Corre & Kremer 2012), but it remains unclear whether polygenic selection in the face of  
531 gene flow could translate into a detectable IBE pattern in the RDA analysis. Thus, it is possible that  
532 the fine-scale genetic structure is more likely to be the outcome of a migration-drift balance (here  
533 associated with habitat choice) than that of a migration-selection balance (concurrent with local

534 adaptation). This interpretation may also account for the surprising finding of small and non-  
535 significant  $F_{st}$  values between D-Muro and E-Pirio (located 25 km apart), contrasting with the  
536 significant differentiation found between D-Muro and E-Muro over a much smaller spatial scale.  
537 Indeed, it is likely that the blue tit population from E-Muro has an overall smaller population  
538 effective size than E-Pirio. The E-Muro population also represents an isolated patch of evergreen  
539 habitat surrounded by deciduous habitat populations while E-Pirio is well connected to other  
540 evergreen habitat populations; hence the possibly higher rate of drift in E-Muro. Therefore, for  
541 similar migration rates between D and E habitats, we expect a stronger genetic differentiation at  
542 equilibrium between D-Muro and E-Muro than between D-Muro and E-Pirio, which is what we  
543 observed. Our results thus suggest that there is limited dispersal across habitats resulting most  
544 probably from matching habitat choice (Edelaar *et al.* 2008); but currently on-going cross-fostering  
545 between habitats, coupled with larger spatial sampling schemes will be instrumental to understand  
546 in greater detail the relative importance of local adaptation and habitat choice.

547 While genetic structuring at small spatial scales is known to occur in some vertebrate species, three  
548 avian studies (Garcia-Navas *et al.* 2014; Postma *et al.* 2009; Senar *et al.* 2006) mirror the fine-scale  
549 pattern of genetic differentiation reported here. However, the reported genetic differences are not  
550 always supported by evident environmental differences, and in all three cases comparisons are based  
551 on 2 populations. Interestingly, Postma and van Noordwijk (2005) and Postma *et al.* (2009) found  
552 that phenotypic differences in great tit clutch size on Vlieland (an island in The Netherlands 19km  
553 long, 2km wide and 25km from the mainland) were coupled with microsatellite genetic  
554 differentiation at a similar geographical scale as in this study. The authors argued that such genetic  
555 differences could arise thanks to highly restricted gene flow to some parts of the island and selection  
556 against immigrants. While there are no physical barriers to dispersal for Corsican blue tits (Blondel *et al.*  
557 2006; Porlier *et al.* 2012b), limited dispersal may act as a component of the “insular syndrome”  
558 (Adler & Levins 1994; Blondel *et al.* 2006); see also Bertrand *et al.* (2014) and Komdeur *et al.* (2004)),

559 contributing to enhance the genetic differentiation and contributing to local adaptation at small  
560 spatial scales in island settings.

561

## 562 **Conclusions and Perspectives**

563 It is undisputable that ongoing improvements in high-throughput sequencing, SNP chip development  
564 and genotyping-by-sequencing approaches facilitate the creation of a rapidly growing number of  
565 large population genomic datasets in wild animal populations, and are in consequence impacting our  
566 understanding of the factors influencing genomic structuration of natural populations (Ellegren *et al.*  
567 2012; Poelstra *et al.* 2014). Here we have validated the potential for RAD sequencing to study small  
568 scale genomic differentiation in an avian system (see also Bertrand *et al.* (2014)). Results in this  
569 study, but also in those of Bertrand *et al.* (2014) and Postma *et al.* (2009) contribute to undermine  
570 generally held assumptions regarding the homogenising effect of gene flow at small spatial scales in  
571 terrestrial vertebrates and birds in particular. Moreover, our study suggests that habitat may play a  
572 key role in generating genome-wide IBE patterns, which is concomitant to habitat-dependent  
573 phenotypic variation reported earlier (Figure 1). In the next phase of genomic exploration of the  
574 Mediterranean blue tit study system, we plan to apply finer-scale axes of environmental variation  
575 (Garroway *et al.* 2013), in particular by focusing on high-resolution satellite imagery (Szulkin *et al.* in  
576 press) and quantitative genetic analyses of phenotypic trait variation, thereby providing a robust  
577 framework to test hypotheses of habitat-dependent adaptation at the genetic level.

578

579

580

581 **Acknowledgements**

582 This study was funded by an IEF Marie Curie Fellowship to M.S., an ANR BioAdapt grant (ANR-12-  
583 ADAP-0006-02-PEPS) to A.C., an APEGE funding to A.C. & M.S., and by OSU-OREME funding for the  
584 long-term monitoring of tits in Corsica and La Rouviere sites. We thank Pascal Marrot for discussion  
585 and for providing nestbox GPS coordinates and oak data, and Marie-Pierre Dubois and Max Galan for  
586 help in the lab. We thank Nicolas Galtier for access to blue tit transcriptome data and Anna Santure  
587 for the great tit transcriptome. Bioinformatic analyses were conducted at the ISEM platform PGPM7  
588 at the Station Méditerranéenne de l'Environnement Littoral (OSU OREME) and through access to the  
589 Montpellier LabEx CeMEB computation facilities. We thank Jason Boone for advice on RAD analyses,  
590 and Jacques Blondel, Dany Garant and Charles Perrier for valuable feedback on the manuscript. We  
591 also whole-heartedly thank the many generations of researchers, students and field assistants who  
592 contributed to blue tit population monitoring and sampling and who made this study possible.

593 **References**

594

- 595 Adler GH, Levins R (1994) The island syndrome in rodent populations. *Quarterly Review of Biology* **69**,  
596 473-490.
- 597 Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using  
598 Sequenced RAD Markers. *Plos One* **3**.
- 599 Bertrand JAM, Bourgeois YXC, Delahaie B, *et al.* (2014) Extremely reduced dispersal and gene flow in  
600 an island bird. *Heredity* **112**, 190-196.
- 601 Blondel J, Dias PC, Ferret P, Maistre M, Lambrechts MM (1999) Selection-based biodiversity at a  
602 small spatial scale in a low-dispersing insular bird. *Science* **285**, 1399-1402.
- 603 Blondel J, Perret P, Dias PC, Lambrechts MM (2001) Is phenotypic variation of blue tits (*Parus*  
604 *caeruleus* L.) in Mediterranean mainland and insular landscapes adaptive? *Genetics Selection*  
605 *Evolution* **33**, S121-S139.
- 606 Blondel J, Perret P, Maistre M (1990) On the genetic basis of the laying date in an island population  
607 of blue tits. *Journal of Evolutionary Biology* **3**, 469-475.
- 608 Blondel J, Thomas DW, Charmantier A, *et al.* (2006) A thirty-year study of phenotypic and genetic  
609 variation of blue tits in Mediterranean habitat mosaics. *Bioscience* **56**, 661-673.
- 610 Cahais V, Gayral P, Tsagkogeorga G, *et al.* (2012) Reference-free transcriptome assembly in non-  
611 model animals from next-generation sequencing data. *Molecular Ecology Resources* **12**, 834-  
612 845.
- 613 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for  
614 population genomics. *Molecular Ecology* **22**, 3124-3140.
- 615 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and  
616 Genotyping Loci De Novo From Short-Read Sequences. *G3-Genes Genomes Genetics* **1**, 171-  
617 182.
- 618 Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral  
619 molecular variation. *Genetics* **134**, 1289-1303.
- 620 Charmantier A, Blondel J, Perret P, Lambrechts MM (2004) Do extra-pair paternities provide genetic  
621 benefits for female blue tits *Parus caeruleus*? *Journal of Avian Biology* **35**, 524-532.
- 622 Charmantier A, Doutrelant C, Dubuc Messier G, Fargevieille A, Szulkin M (in press) Mediterranean  
623 blue tits as a case study of local adaptation. *Evolutionary Applications*.
- 624 Danecek P, Auton A, Abecasis G, *et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**,  
625 2156-2158.
- 626 Edelaar P, Alonso D, Lagerveld S, Senar JC, Bjorklund M (2012) Population differentiation and  
627 restricted gene flow in Spanish crossbills: not isolation-by-distance but isolation-by-ecology.  
628 *Journal of Evolutionary Biology* **25**, 417-430.
- 629 Edelaar P, Bolnick DI (2012) Non-random gene flow: an underappreciated force in evolution and  
630 ecology. *Trends in Ecology & Evolution* **27**, 659-665.
- 631 Edelaar P, Burraco P, Gomez-Mestre I (2011) Comparisons between Q(ST) and F-ST-how wrong have  
632 we been? *Molecular Ecology* **20**, 4830-4839.
- 633 Edelaar P, Siepielski AM, Clobert J (2008) Matching habitat choice causes directed gene flow: a  
634 neglected dimension in evolution and ecology. *Evolution* **62**, 2462-2472.
- 635 Ellegren H, Smeds L, Burri R, *et al.* (2012) The genomic landscape of species divergence in *Ficedula*  
636 flycatchers. *Nature* **491**, 756-760.
- 637 Garant D, Forde SE, Hendry AP (2007) The multifarious effects of dispersal and gene flow on  
638 contemporary adaptation. *Functional Ecology* **21**, 434-443.

- 639 Garcia-Navas V, Ferrer ES, Sanz JJ, Ortego J (2014) The role of immigration and local adaptation on  
640 fine-scale genotypic and phenotypic population divergence in a less mobile passerine.  
641 *Journal of Evolutionary Biology* **27**, 1590-1603.
- 642 Garroway CJ, Radersma R, Sepil I, *et al.* (2013) Fine-scale genetic structure in a wild bird population:  
643 the role of limited dispersal and environmentally based selection as causal factors. *Evolution*  
644 **67**, 3488-3500.
- 645 Hansson B, Ljungqvist M, Illera J-C, Kvist L (2014) Pronounced Fixation, Strong Population  
646 Differentiation and Complex Population History in the Canary Islands Blue Tit Subspecies  
647 Complex. *Plos One* **9**.
- 648 Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among  
649 organisms. *Bioinformatics* **22**, 2971-2972.
- 650 Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous environments.  
651 *Annual Review of Ecology and Systematics* **7**, 1-32.
- 652 Jakobsson M, Edge MD, Rosenberg NA (2013) The Relationship Between FST and the Frequency of  
653 the Most Frequent Allele. *Genetics* **193**, 515-528.
- 654 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.  
655 *Bioinformatics* **24**, 1403-1405.
- 656 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.  
657 *Bioinformatics* **27**, 3070-3071.
- 658 Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic  
659 variability by a new multivariate method. *Heredity* **101**, 92-103.
- 660 Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- 661 Komdeur J, Piersma T, Kraaijeveld K, Kraaijeveld-Smit F, Richardson DS (2004) Why Seychelles  
662 Warblers fail to recolonize nearby islands: unwilling or unable to fly there? *Ibis* **146**, 298-302.
- 663 Lambrechts MM, Blondel J, HurtrezBousses S, Maistre M, Perret P (1997) Adaptive inter-population  
664 differences in blue tit life-history traits on Corsica. *Evolutionary Ecology* **11**, 599-612.
- 665 Lambrechts MM, Caro SP, Charmantier A, *et al.* (2004) Habitat quality as a predictor of spatial  
666 variation in blue tit reproductive performance: a multi-plot analysis in a heterogeneous  
667 landscape. *Oecologia* **141**, 555-561.
- 668 Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally  
669 adaptive traits. *American Naturalist* **151**, 283-292.
- 670 Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local  
671 adaptation. *Molecular Ecology* **21**, 1548-1566.
- 672 Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* **17**,  
673 183-189.
- 674 Martin JL (1991) Patterns and significance of geographical variation in the blue tit (*Parus caeruleus*).  
675 *Auk* **108**, 820-832.
- 676 Mastretta-Yanes A, Arrigo N, Alvarez N, *et al.* (2015) Restriction site-associated DNA sequencing,  
677 genotyping error estimation and de novo assembly optimization for population genetic  
678 inference. *Molecular Ecology Resources* **15**, 28-41.
- 679 Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23.
- 680 O'Connor TD, Fu W, Turner E, *et al.* (2015) Rare variants facilitates inferences of fine scale population  
681 structure in humans. *Molecular Biology and Evolution* **32**, 653-660.
- 682 Oksanen J, Blanchet GG, Kindt R, *et al.* (2014) vegan: Community Ecology Package.
- 683 Ortego J, Garcia-Navas V, Ferrer ES, Sanz JJ (2011) Genetic structure reflects natal dispersal  
684 movements at different spatial scales in the blue tit, *Cyanistes caeruleus*. *Animal Behaviour*  
685 **82**, 131-137.
- 686 Palumbi SR (1994) Genetic divergence, reproductive isolation and marine speciation. *Annual Review*  
687 *of Ecology and Systematics* **25**, 547-572.

- 688 Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *Plos Genetics* **2**, 2074-  
689 2093.
- 690 Poelstra JW, Vijay N, Bossu CM, *et al.* (2014) The genomic landscape underlying phenotypic integrity  
691 in the face of gene flow in crows. *Science* **344**, 1410-1414.
- 692 Porlier M, Charmantier A, Bourgault P, *et al.* (2012a) Variation in phenotypic plasticity and selection  
693 patterns in blue tit breeding time: between- and within-population comparisons. *Journal of*  
694 *Animal Ecology* **81**, 1041-1051.
- 695 Porlier M, Garant D, Perret P, Charmantier A (2012b) Habitat-Linked Population Genetic  
696 Differentiation in the Blue Tit *Cyanistes caeruleus*. *Journal of Heredity* **103**, 781-791.
- 697 Postma E, Den Tex R-J, Van Noordwijk AJ, Mateman AC (2009) Neutral markers mirror small-scale  
698 quantitative genetic differentiation in an avian island population. *Biological Journal of the*  
699 *Linnean Society* **97**, 867-875.
- 700 Postma E, van Noordwijk AJ (2005) Gene flow maintains a large genetic difference in clutch size at a  
701 small spatial scale. *Nature* **433**, 65-68.
- 702 Ravigne V, Dieckmann U, Olivieri I (2009) Live Where You Thrive: Joint Evolution of Habitat Choice  
703 and Local Adaptation Facilitates Specialization and Promotes Diversity. *American Naturalist*  
704 **174**, E141-E169.
- 705 Romiguier J, Gayral P, Ballenghien M, *et al.* (2014) Comparative population genomics in animals  
706 uncovers the determinants of genetic diversity. *Nature* **515**, 261-U243.
- 707 Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J (2011) Characterisation of the transcriptome  
708 of a wild great tit *Parus major* population by next generation sequencing. *Bmc Genomics* **12**.
- 709 Senar JC, Borrás A, Cabrera J, Cabrera T, Bjorklund M (2006) Local differentiation in the presence of  
710 gene flow in the citril finch *Serinus citrinella*. *Biology Letters* **2**, 85-87.
- 711 Sharma R, Goossens B, Kun-Rodrigues C, *et al.* (2012) Two Different High Throughput Sequencing  
712 Approaches Identify Thousands of De Novo Genomic Markers for the Genetically Depleted  
713 Bornean Elephant. *Plos One* **7**.
- 714 Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733-756.
- 715 Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-  
716 792.
- 717 Snow DW (1954) The habitats of eurasian tits (*Parus* spp.). *Ibis* **96**, 565-585.
- 718 Szulkin M, Stopher KV, Pemberton JM, Reid JM (2013) Inbreeding avoidance, tolerance, or  
719 preference in animals? *Trends in Ecology & Evolution* **28**, 205-211.
- 720 Szulkin M, Zelazowski P, Marrot P, Charmantier A (in press) Application of high temporal and spatial  
721 resolution satellite imagery to characterise individual-based environmental heterogeneity in  
722 a wild bird. *Remote Sensing*.
- 723 Tufto J, Ringsby TH, Dhondt AA, Adriaensen F, Matthysen E (2005) A parametric model for estimation  
724 of dispersal patterns applied to five passerine spatially structured populations. *American*  
725 *Naturalist* **165**, E13-E26.
- 726 Van Bers NEM, Santure AW, Van Oers K, *et al.* (2012) The design and cross-population application of  
727 a genome-wide SNP chip for the great tit *Parus major*. *Molecular Ecology Resources* **12**, 753-  
728 770.
- 729 Wang JJ, Bradburg GS (2014) Isolation by environment. *Molecular Ecology* **23**, 5649-5662.
- 730 Waples RS (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high  
731 gene flow species. *Journal of Heredity* **89**, 438-450.
- 732 Zheng XW, Levine D, Shen J, *et al.* (2012) A high-performance computing toolset for relatedness and  
733 principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328.

734

735

736 **Data Accessibility**

737 Demultiplexed RAD sequencing read data is available at the NCBI Short Read Archive under accession  
738 number SRP065946 . The raw VCF file (209 860 SNPs), the filtered VCF file (12 106 SNPs), a list of  
739 consensus RAD sequences and sample details are available on Dryad (doi:10.5061/dryad.713v1).

740

741 **Author contributions**

742 M.S. and A.C. designed the study; M.S. performed preliminary lab work; M.S. and P.-A. G. analysed  
743 the data with input from N.B.; M.S. wrote the manuscript with input from P-A. G., N. B. and A. C.

744



745 **Tables and Figures**

746

747 **Table 1.**

748

Location & coordinates (long, lat)	Breeding site	Predominant Oak species	N birds (N females)	Average geographical distances between (in km):
Corsica (42.5893, 8.9667)	E-Muro	Evergreen holm oak <i>Quercus ilex</i> (100%)	9 (3)	E-Muro and D-Muro: 5.6
Corsica (42.5509, 8.9233)	D-Muro	Deciduous downy oak <i>Quercus pubescens</i> (100%)	48 (19)	
Corsica (42.3763, 8.7497)	E-Pirio	Evergreen holm oak <i>Quercus ilex</i> (100%)	83 (42)	E-Muro and E-Pirio: 29.6 D-Muro and E-Pirio: 24.1
Continental France (43.6639, 3.6658)	D-Rouviere	Deciduous downy oak <i>Quercus pubescens</i> (81%)	57 (32)	Corsican sites and D-Rouviere: 441.2

749

750

751 **Table 1.** Spatial and vegetation-based habitat characteristics of the different sampling sites in  
752 Southern French mainland and on Corsica Island. For oak species, values in brackets indicate the  
753 mean proportion (in %) of predominant oak species (Evergreen “E” vs. Deciduous “D”) within a 50m  
754 radius of each nestbox in the dataset. Sampling site coordinates were calculated as the average of  
755 nestbox coordinates for all breeding birds sampled in the study.

756

757 **Table 2.**

758

Step in the bioinformatic pipeline	Number of SNPs	Analysis applied to that dataset
Raw VCF file from <i>Stacks</i>	209 860	Quality filtering
Trimming bases $\geq$ to position 85 of reads	189 356	Quality filtering
Excluding loci that are not in within-population HWE at $p < 0.01$ within D-Rouvière D, E&D-Muro or E-Pirio.	166 410	Quality filtering
Working dataset MAF 2%, 90% call rate	12 106	IBS analysis, Spatial PCA, RDA
Working dataset MAF 5%, 90% call rate	6 555	PCA
Working dataset MAF 5%, 95% call rate	3 159	Genome-wide pairwise $F_{st}$ Transcriptome pairwise $F_{st}$

759

760 **Table 2.** RAD sequencing bioinformatic analysis pipeline, starting from the VCF file generated with  
761 *Stacks*, and detailing the filtering steps undertaken in *VCFTools*. The resulting changes in the number  
762 of SNPs are indicated at each step. The genotype call rate threshold represents the minimum  
763 proportion of genotypes called for each locus across all birds in the dataset. The MAF represents the  
764 minimum allele frequency threshold applied to the rare variant for each locus. Number of birds  
765 sampled: 197.

766

767 **Table 3.**

768

transcriptome	N of sequences present in the transcriptome	N of aligned RAD tags	% of mapped RAD tags	Average E value	% identical
Blue tit	7120	326	1.6	1.02 E-9	97
Great tit	95979	1007	5.1	1.56 E-8	96
Zebra finch <i>cDNA</i>	59816	722	3.7	4.89 E-10	96
Zebra finch <i>ab-initio</i>	18610	1135	5.7	2.14 E-9	95

769

770 **Table 3.** Blue tit *Blastx* outputs obtained from searches against blue tit, great tit and zebra finch

771 transcriptomes. Total number of consensus RAD tags blasted = 19 760 (global call rate &gt; 80%, MAF &gt;

772 1%); all E-values assessing the significance of blast matches were  $\leq 1E-7$ .

773 **Table 4.**

774

	D-Rouviere	D-Muro	E-Muro	E-Pirio
D-Muro	<b>0.0541 (p≤0.002**)</b> <u>0.0572 (p≤0.002**)</u> 0.0591 (p≤0.002**) <i>0.0487</i>	-	-	-
E-Muro	<b>0.0335 (p≤0.002**)</b> <u>0.0480 (p≤0.002**)</u> 0.0379 (p≤0.002**) <i>0.0415</i>	<b>0.0156 (p=0.004**)</b> <u>0.0182 (p=0.006**)</u> 0.0246 (p≤0.002**) <i>0.0065</i>	-	-
E-Pirio	<b>0.0520 (p≤0.002**)</b> <u>0.0531 (p≤0.002**)</u> 0.0600 (p≤0.002**) <i>0.0413</i>	<b>0.0099 (p=0.347)</b> <u>0.0111 (p=0.762)</u> 0.0102 (p=0.333) <i>0.0043</i>	<b>0.0102 (p=0.403)</b> <u>0.0160 (p=0.168)</u> 0.0151 (p=0.015*) <i>0.0050</i>	-

775

776 **Table 4.** Fst values for SNPs retained after filtering with 5% MAF and a 95% call rate. In bold: all SNPs777 *n*=197 individuals, 3159 SNPs. Underlined: all SNPs in a “no family ties” dataset”, *n*=119 individuals,778 2816 SNPs. Normal type: transcriptome-derived SNPs, *n*=197 individuals, 179 SNPs. Empirical p-

779 values were computed using 500 permutations (lowest p-values are therefore bounded by 0.002).

780 After Bonferroni correction, a significant signal of genetic differentiation in RADseq derived Fst values

781 was found in all Corsica-mainland comparisons and between D-Muro and E-Muro in each of the

782 inspected datasets. In italic: Fst values from Porlier *et al.* 2012 (data averaged across several years,783 *n*=607 adults (see manuscript for sampling details), 6-10 microsatellite markers).

784 **Table 5.**

785

	<b>Full dataset (continent+Corsica) N=197</b>	<b>Conditioned RDA full dataset N=197</b>	<b>Corsican birds only N=140</b>	<b>Conditioned RDA Corsican birds N=140</b>
	Variance (D.f), p-value	Variance (D.f.), p-value	Variance (D.f), p-value	Variance (D.f), p-value
<b>Global Analysis</b>	56.86 (4), 0.001	-	18.15 (3), 0.001	-
<i>Residual</i>	534.87 (192)		485.27 (136)	
<b>Marginal Test</b>				
Latitude	3.83 (1), <b>0.020</b>	3.83 (1), <b>0.002</b>	5.41 (1), 0.001	5.41 (1), 0.001
Longitude	4.21 (1), <b>0.010</b>	4.21 (1), <b>0.001</b>	-	-
Habitat	3.50 (1), <b>0.035</b>	3.50 (1), <b>0.002</b>	4.89 (1), 0.001	4.89 (1), 0.001
Birth Year	3.48 (1), <b>0.040</b>	3.48 (1), <b>0.001</b>	4.52 (1), 0.001	4.52 (1), 0.001
<i>Residual</i>	534.87 (192)	534.87 (192)	485.27 (136)	485.27 (136)

786

787 **Table 5.** Results of RDA significance tests (variance, d.f. and p-values obtained through 1000

788 permutations; significant p-values are in bold), detailed for the global RDA analysis (model with non-

789 significant terms removed) and the marginal effect of each constraining variable in the model. RDA

790 were performed on the full dataset ( $n=197$ ) and on Corsican birds only ( $n=140$ ). The marginal effect

791 of each constraining variable was tested through permutation tests by removing each term one by one

792 from the model containing all other terms. The second and fourth results column report partial RDA

793 significance tests (variance, d.f. and p-values) for each term, after conditioning on other constraining

794 variables to remove their confounding effects. Longitude was not used as a constraining variable in

795 Corsican RDA due to its strong correlation with Latitude.

796 **Table 6.**

797

RDA axis	Full dataset (continent+Corsica)				Corsican birds only		
	RDA1	RDA2	RDA3	RDA4	RDA1	RDA 2	RDA 3
% of variance explained	7.39%	1.06%	0.58%	0.57%	1.77%	0.96%	0.88%
<b>Constraining variables</b>							
Latitude	-0.9926	0.1146	0.0404	0.0041	-0.9647	-0.2371	0.1144
Longitude	0.9988	0.0476	0.0003	0.0091	-	-	-
Habitat	0.6084	-0.7227	0.1196	0.3053	0.9119	-0.3577	-0.2014
Birth year	0.0212	-0.4572	0.7455	-0.4846	0.5012	-0.2634	0.8243

798

799 **Table 6.** Summary of RDA analysis for the full dataset and Corsican birds only. The proportion of  
800 genotypic variance explained by each RDA axis is provided, along with the vector coordinates of each  
801 constraining variable in the RDA space (these vectors are represented in blue colour, Figure 5). For  
802 each RDA axis, the longest vector projection indicates the most important variable explaining  
803 variation along that axis.

804

## 805 **Figure Legends**

806 **Figure 1.** Phenotypic trait values (means and 95% confidence intervals) for (A) fitness traits and (B)  
807 morphological traits for the four study sites. All traits depicted have a significant genetic basis  
808 established with quantitative genetic models (see Charmantier et al. in press for more details). All  
809 traits show high similarity within habitat type (deciduous or evergreen oaks – see illustrative tab  
810 below each graph), or (non-exclusively) carry continental distinctiveness (D-Rouviere), followed by  
811 intermediate values of E-Muro relative to D-Muro and E-Pirio. All traits were recorded annually  
812 between 1991 and 2014 for D-Rouviere, 1993 and 2014 for D-Muro, 1998 and 2014 for E-Muro and  
813 1976 and 2014 for E-Pirio.

814 **(A)** Fitness traits: egg laying date is represented as filled triangles (1=1<sup>st</sup> March), clutch size (from first  
815 broods only) as circles and the number of fledglings as squares (total n=5566, 5555, 4367,  
816 respectively).

817 **(B)** Morphological traits: female and male body mass are represented as filled and open circles,  
818 female and male tarsus length are represented as filled and open squares (total n=4962, 4559,  
819 3068, 2792 respectively).

820

821 **Figure 2.** Map of the 4 study sites: D-Rouviere on the French mainland, E-Muro, D-Muro and E-Pirio  
822 in Corsica, complemented with an illustration of predominant oak species (Deciduous or Evergreen)  
823 for each site.

824

825

826

827

828 **Figure 3.** Population-specific distribution of pairwise relatedness coefficient and folded-allele  
829 frequency spectrum.  
830  
831 **(A&B):** Identity-by-state (IBS) pairwise relatedness distribution for (A) the mainland (D-Rouviere) and  
832 (B) Corsica (D-Muro, E-Muro and E-Pirio). The right hand tails of both continental and Corsican  
833 populations reflect family structure in the datasets, further validated with IBS values of full siblings  
834 (black crosses, established with microsatellite data from Charmantier et al. 2003) and mother-  
835 offspring pairs (black diamonds, established with pedigree data). Black vertical lines are IBS values of  
836 sample replicates (one replicate from the mainland, four replicates from Corsica). **(C&D):** Minor allele  
837 frequency (MAF) spectrum for (C) the continent and (D) Corsica.

838

839 **Figure 4.** Principal Component Analysis (Axes 1 & 2, explaining 6.7% and 1,6% of the variance,  
840 respectively – see also Table S1) of the 4 blue tit populations ( $n=197$  individuals, 6555 SNPs, MAF 5%,  
841 90% call rate). Colour legend: D-Rouviere (red circles), D-Muro (orange circles), E-Muro (green  
842 triangles), E-Pirio (violet triangles).

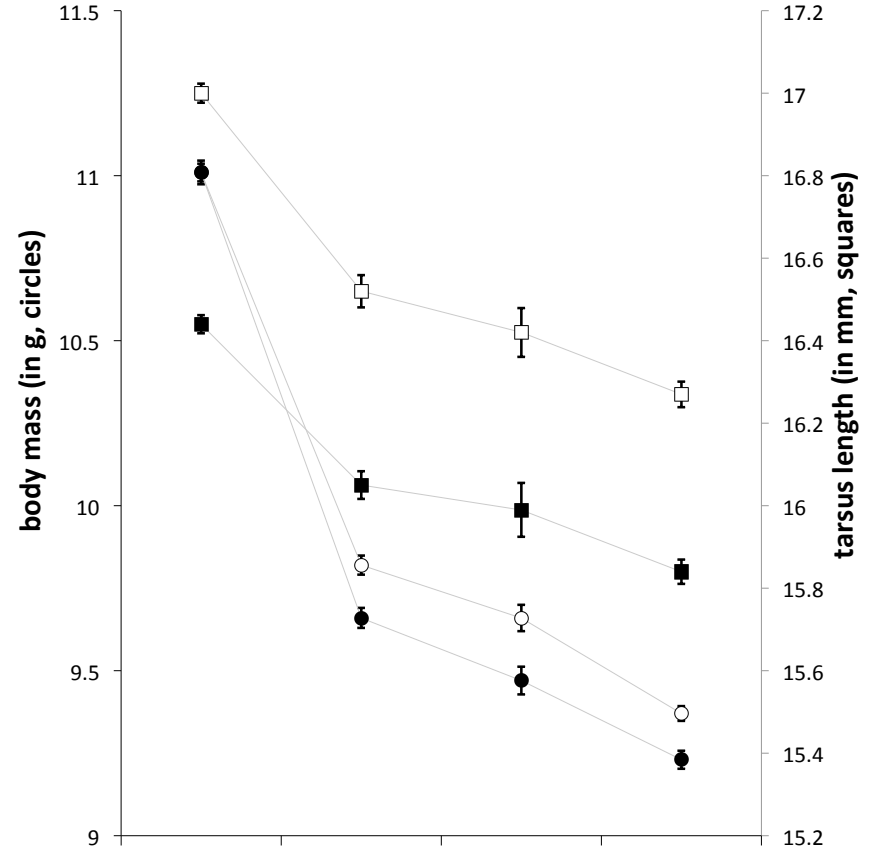
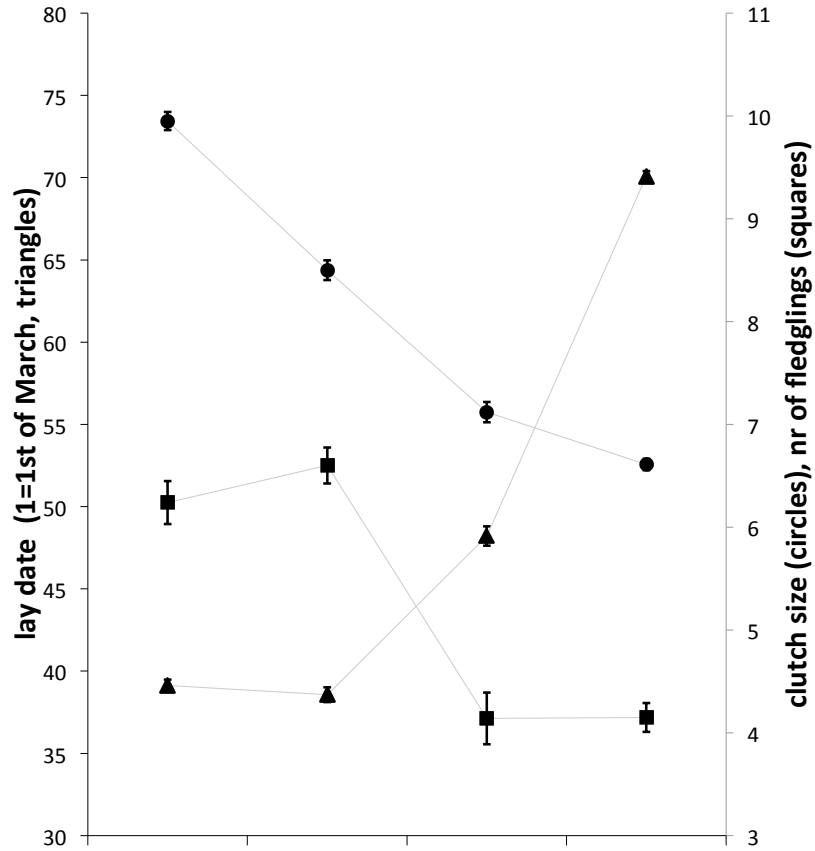
843

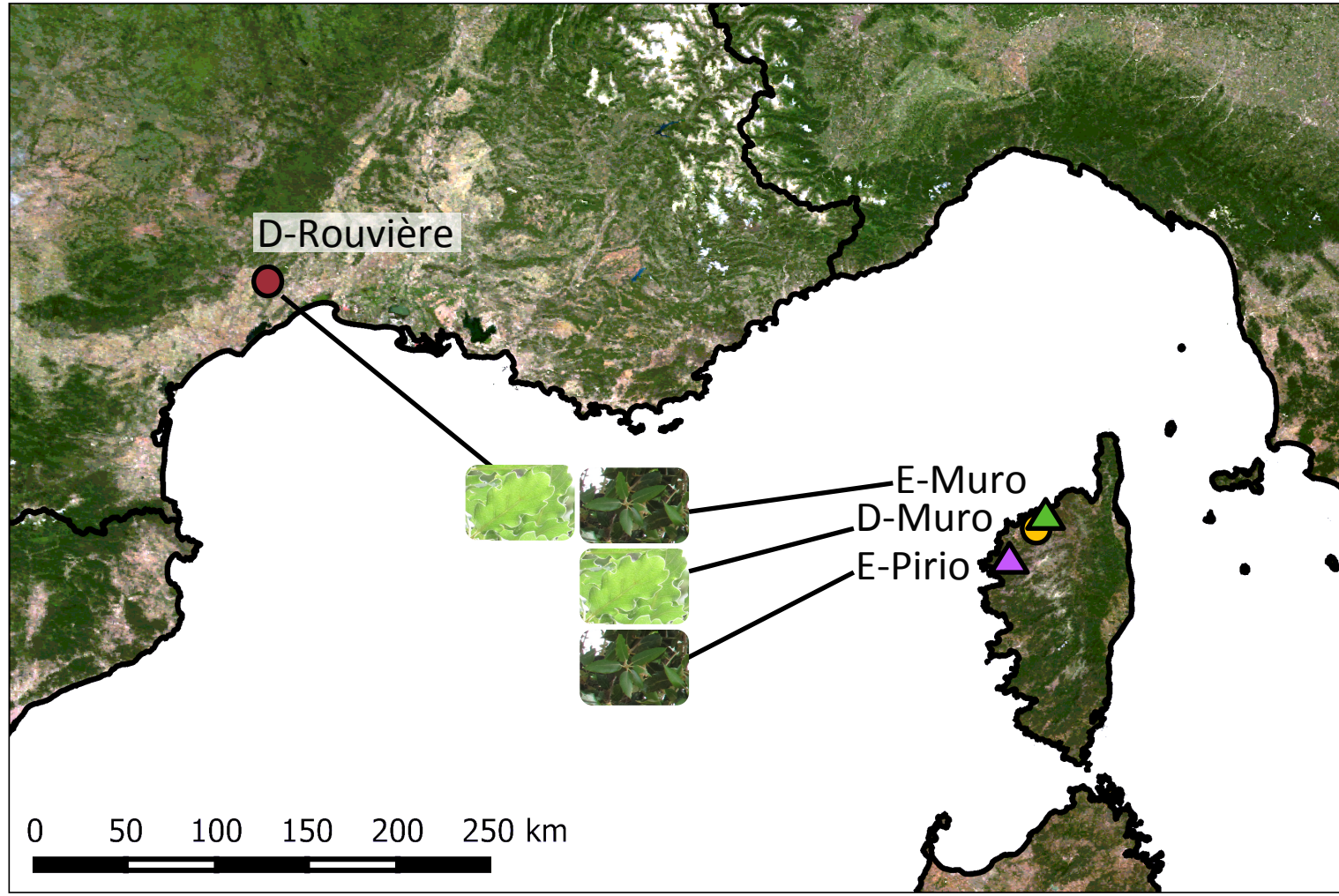
844 **Figure 5.** RDA analysis (12106 SNPs, MAF 2%, 90% call rate) for **(A)** the entire dataset or **(B)** Corsican  
845 birds only. Data points correspond to the projection of individual genotypes on RDA axes 1 and 2,  
846 explaining cumulatively 8.5% of the total genotypic variance in (A) and 2.7% in (B). Vectors of  
847 constraining variables (latitude, longitude, habitat, birth year) are projected on RDA axes 1 and 2;  
848 their arrows points to the direction of strongest gradient of variation, and their projected lengths  
849 indicate the strength of their contribution to each axis. The projection of a factor's vectors was  
850 rescaled (right and upper blue scales) to facilitate their interpretation. The value of each individual

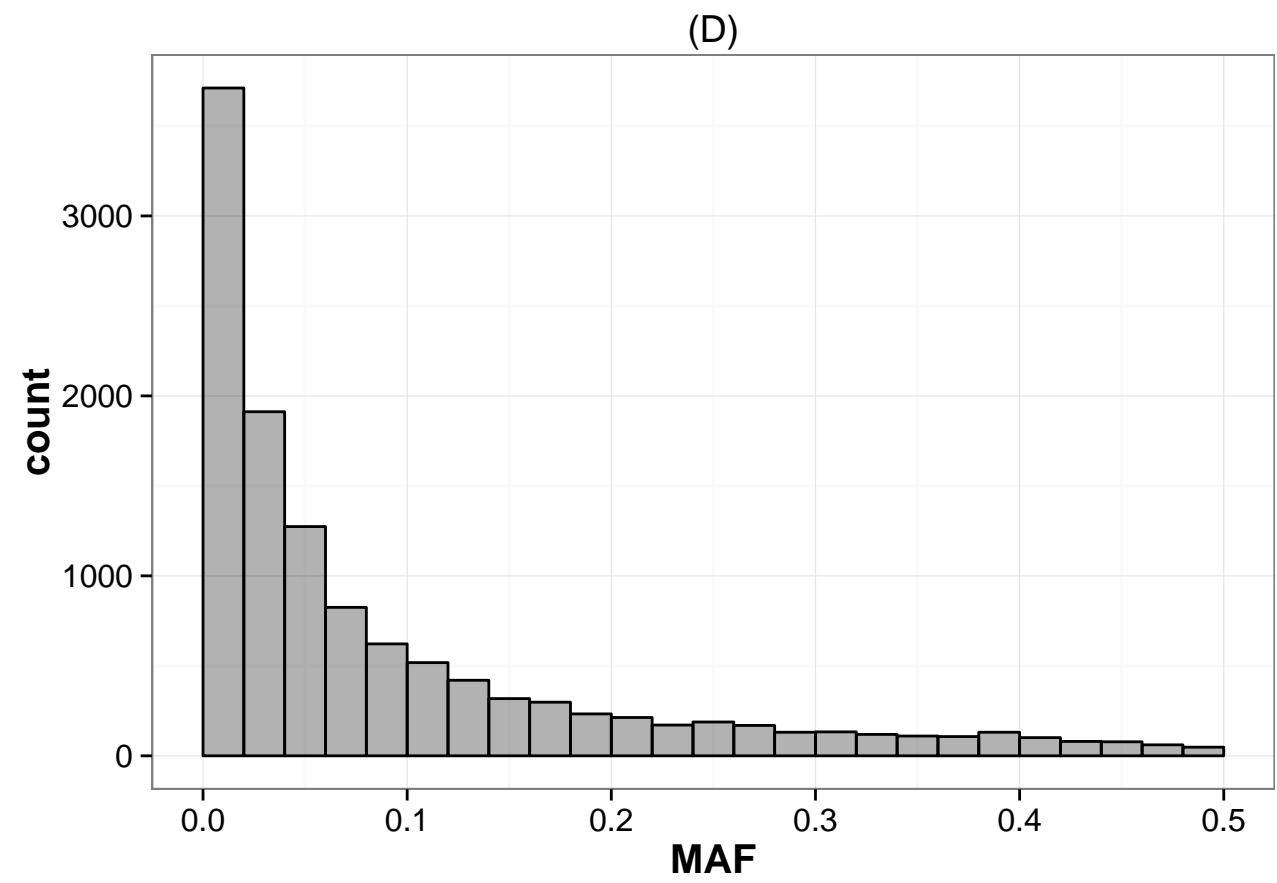
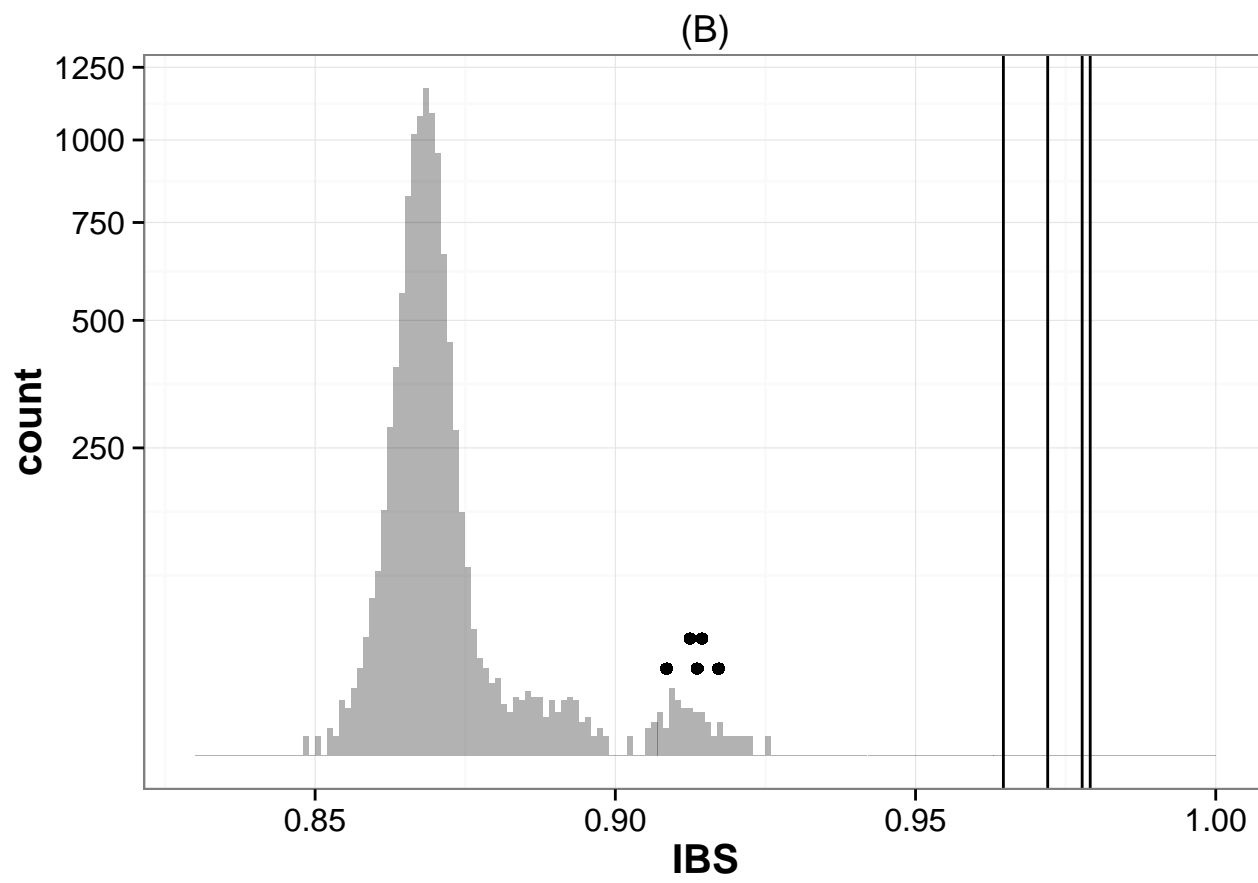
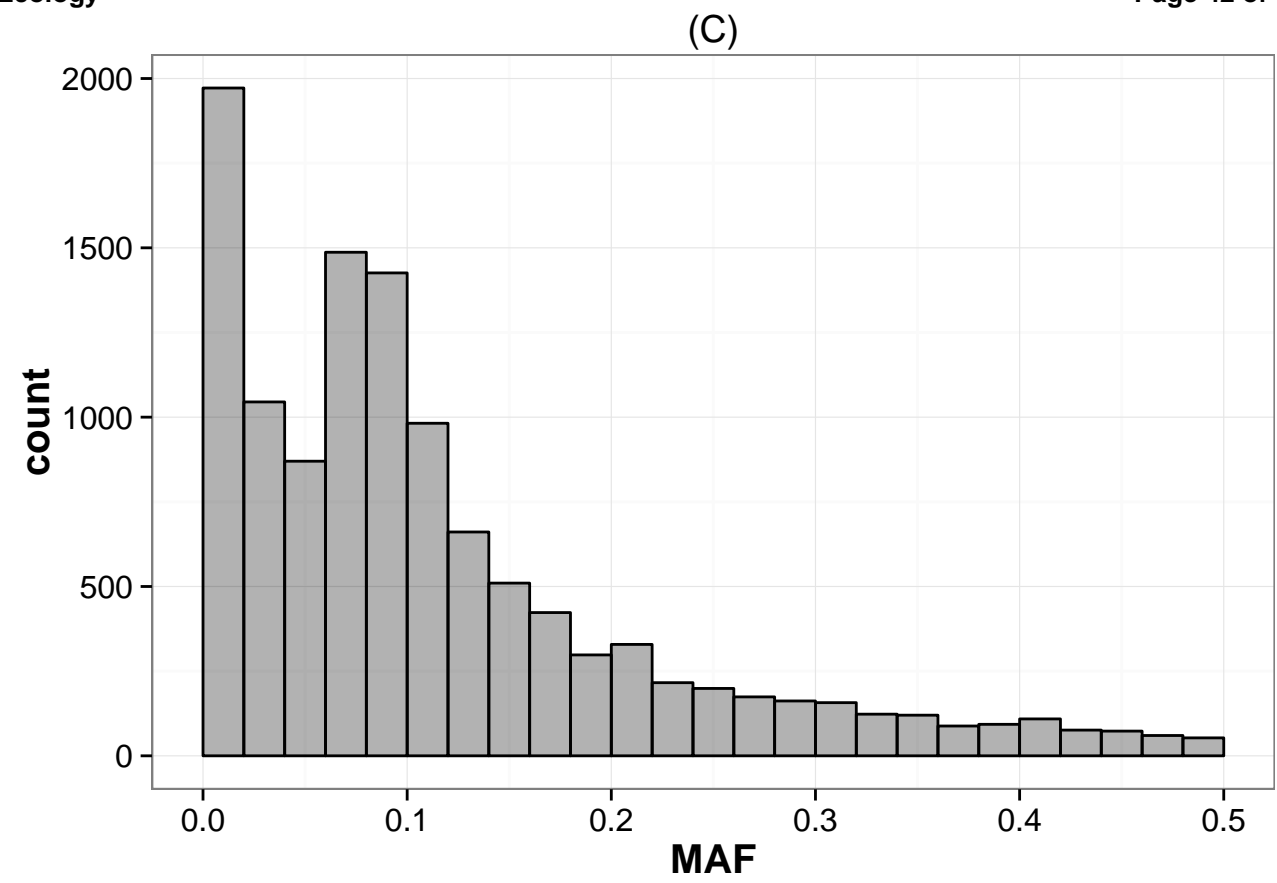
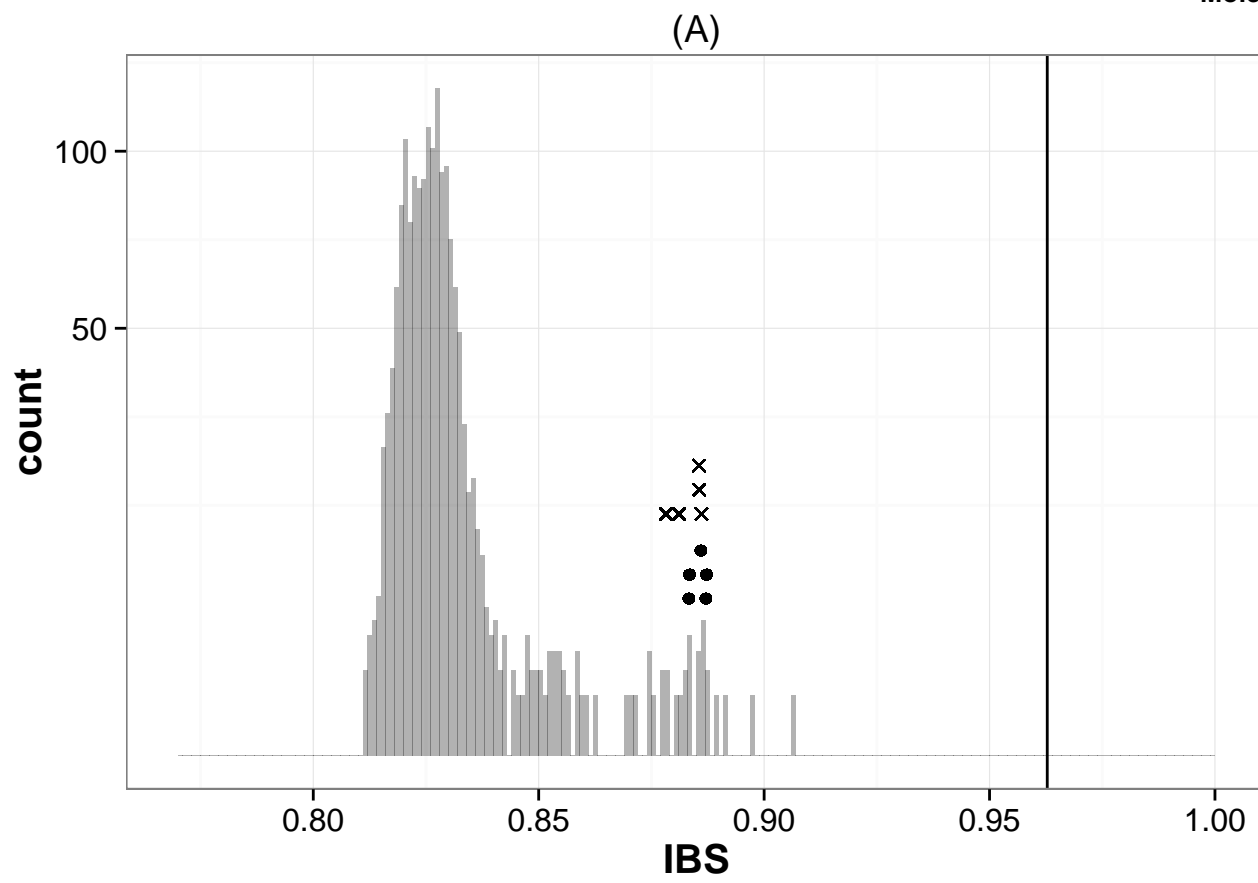


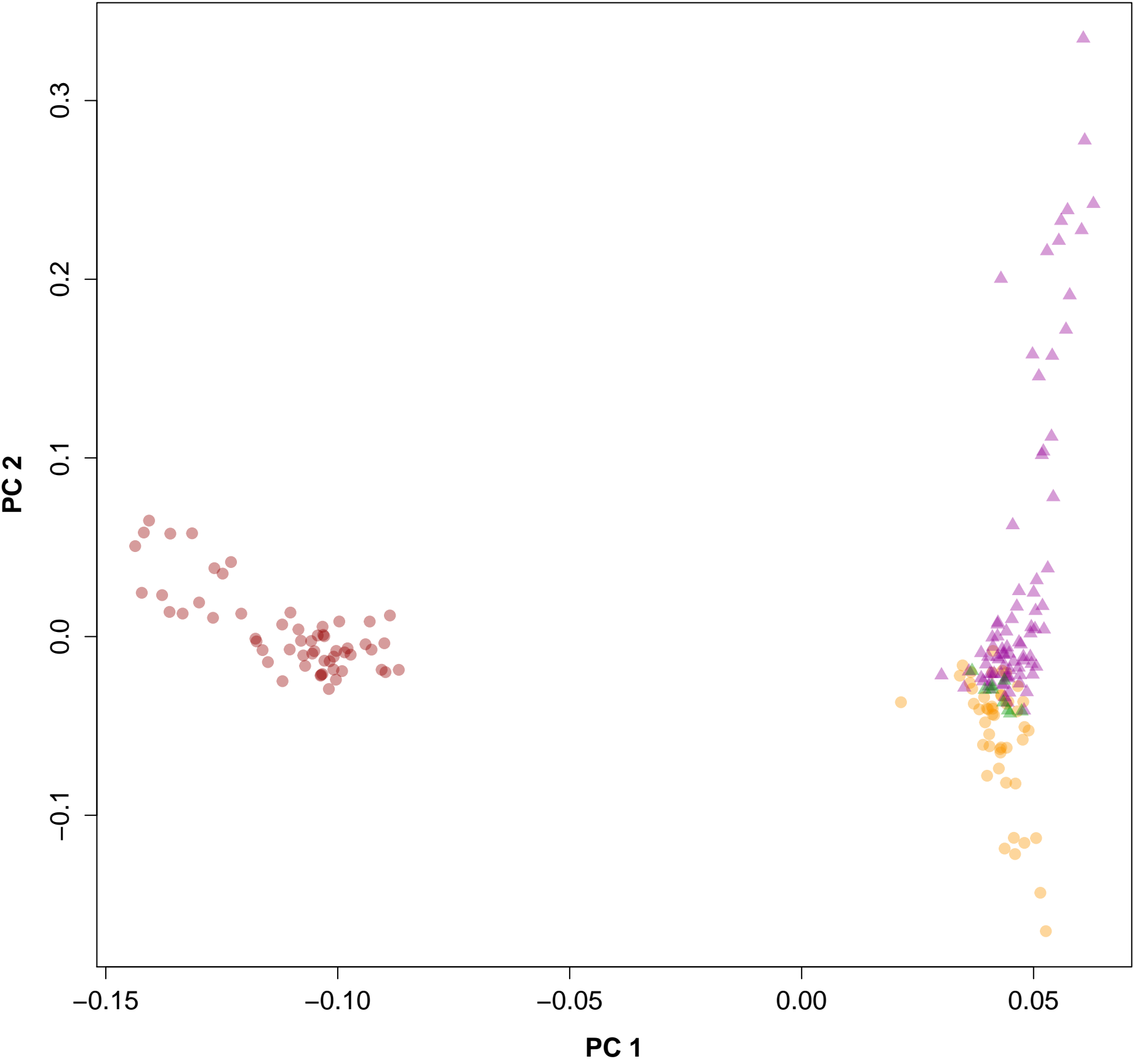
851 data point on any factor vector can be inferred by performing an orthogonal projection of that point  
852 on any chosen vector (for example in Figure 4B, E-Muro and E-Pirio data points project on the same  
853 space of the habitat vector, while D-Muro data points are shifted to the left for that vector. On the  
854 latitude vector, E-muro data points are grouped mostly to the left, D-Muro values have intermediate  
855 values, and E-Pirio data points are grouped to the right – this projection of points on the latitude  
856 vector is concordant with their geographical positioning (Figure 2).

For Review Only

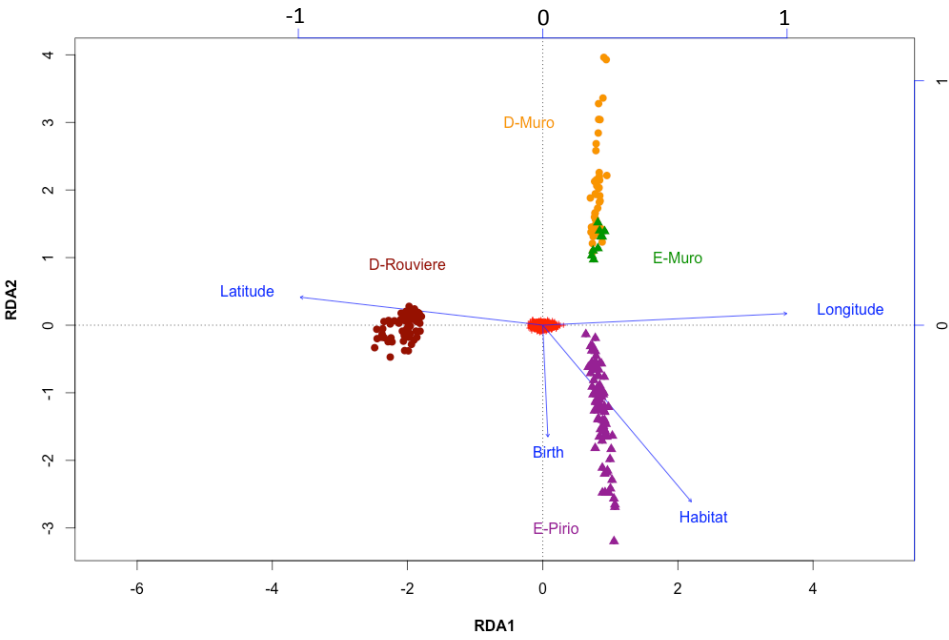








(A)



(B)

