



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/161669/>

Version: Accepted Version

Article:

Radersma, R., Garroway, C.J., Santure, A.W. et al. (2017) Social and spatial effects on genetic variation between foraging flocks in a wild bird population. *Molecular Ecology*, 26 (20). pp. 5807-5819. ISSN: 0962-1083

<https://doi.org/10.1111/mec.14291>

This is the peer reviewed version of the following article: Radersma, R, Garroway, CJ, Santure, AW, et al. Social and spatial effects on genetic variation between foraging flocks in a wild bird population. *Mol Ecol*. 2017; 26: 5807– 5819, which has been published in final form at <https://doi.org/10.1111/mec.14291>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

MOLECULAR ECOLOGY**Social and spatial effects on genetic variation between foraging flocks in a wild bird population**

Journal:	<i>Molecular Ecology</i>
Manuscript ID	MEC-17-0368.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Radersma, Reinder; Lund University, Biology Garraway, Colin; University of Oxford, Department of Zoology Santure, Anna; The University of Auckland, School of Biological Sciences De Cauwer, Isabelle; Université de Lille – Sciences et Technologies, Unité Evolution, Ecologie et Paléontologie Farine, Damien; University of Oxford, Department of Zoology Slate, Jon; University of Sheffield, Animal & Plant Sciences; Sheldon, Ben; University of Oxford, Zoology
Keywords:	Behavior/Social Evolution, Birds, Ecological Genetics, Landscape Genetics, Population Dynamics, Population Genetics - Empirical

Social and spatial effects on genetic variation between foraging flocks in a wild bird population

REINDER RADERSMA*†, COLIN J. GARROWAY*‡, ANNA W. SANTURE§¶, ISABELLE DE CAUWERS§**, DAMIEN R. FARINE*††‡‡, JON SLATE§ and BEN C. SHELDON*

- 1 *Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK, †
- 2 Department of Biology, Lund University, 22362 Lund, Sweden, ‡ Department of Biological
- 3 Sciences, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada, §Department of Animal
- 4 and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK, ¶School of Biological
- 5 Sciences, The University of Auckland, Auckland, New Zealand, ** Univ. Lille, CNRS, UMR 8198 -
- 6 Evo-Eco-Paleo, F-59000 Lille, France, †† Department of Collective Behaviour, Max Planck Institute
- 7 for Ornithology, Konstanz, Germany, ‡‡ Chair of Biodiversity and Collective Behaviour,
- 8 Department of Biology, University of Konstanz, Germany .

Correspondence: Reinder Radersma

Email: reinder.radersma@biol.lu.se

Running title: Social & spatial effects on genetic structure

Key words: population genetics, social networks, SNPs, *Parus major*, eigenvector maps

10 **Abstract (word count: 258 [max 250])**

11 Social interactions are rarely random. In some instances animals exhibit homophily or
12 heterophily, the tendency to interact with similar or dissimilar conspecifics respectively.
13 Genetic homophily and heterophily influence the evolutionary dynamics of populations, because
14 they potentially affect sexual and social selection. Here we investigate the link between social
15 interactions and allele frequencies in foraging flocks of great tits (*Parus major*) over three
16 consecutive years. We constructed co-occurrence networks which explicitly described the
17 splitting and merging of 85,602 flocks through time (fission-fusion dynamics), at 60 feeding
18 sites. Of the 1711 birds in those flocks we genotyped 962 individuals at 4701 autosomal single-
19 nucleotide polymorphisms (SNPs). By combining genome-wide genotyping with repeated field
20 observations of the same individuals we were able to investigate links between social structure
21 and allele frequencies at a much finer scale than was previously possible. We explicitly
22 accounted for potential spatial effects underlying genetic structure at the population level. We
23 modelled social structure and spatial configuration of great tit fission-fusion dynamics with
24 eigenvector maps. Variance partitioning revealed that allele frequencies were strongly affected
25 by group fidelity (explaining 27-45% of variance) as individuals tended to maintain associations
26 with the same conspecifics. These conspecifics were genetically more dissimilar than expected,
27 shown by genome-wide heterophily for pure social (*i.e.* space-independent) grouping
28 preferences. Genome-wide homophily was linked to spatial configuration, indicating spatial
29 segregation of genotypes. We did not find evidence for homophily or heterophily for putative
30 socially relevant candidate genes or any other SNP markers. Together, these results
31 demonstrate the importance of distinguishing social and spatial processes in determining
32 population structure.

33 **Introduction**

34 In many animal species, individuals interact repeatedly with particular individuals while
35 avoiding or ignoring others (Krause & Ruxton 2002; Krause *et al.* 2007, 2014; Croft *et al.* 2008).
36 Repeated interactions can be with the same individuals (e.g. social interactions in breeding pairs
37 or stable groups) and with counterparts of a particular pheno- or genotype. Preference to
38 interact with similar counterparts (*i.e.* homophily) is common, can evolve under a wide variety
39 of conditions (Fu *et al.* 2012) and can be reinforced by assortative social learning (the
40 preference to learn from specific individuals; Katsnelson *et al.* 2014). Homophily has been found
41 for various phenotypic traits such as age and sex in bottlenose dolphins (*Tursiops spp.*; Lusseau
42 & Newman 2004), sex in Grevy's zebra (*Equus grevyi*; Sundaresan *et al.* 2007), body size in
43 guppies (*Poecilia reticulata*; Croft *et al.* 2005), and personality in great tits (Aplin *et al.* 2013)
44 and chimpanzees (*Pan troglodytes*; Massen & Koski 2014). Genotypic homophily has been
45 reported in humans (Fowler *et al.* 2011; Christakis & Fowler 2014).

46 On the other hand, the tendency to interact with dissimilar individuals (*i.e.* heterophily)
47 is not very often observed outside the context of reproduction, in which the two sexes interact
48 to produce and raise offspring. However, heterophily can evolve under particular conditions (Fu
49 *et al.* 2012). Heterophily has been found for sex in foraging great tits (Farine *et al.* 2015), which
50 may be linked to future mate choice. Genotypic heterophily has also been reported, but it was
51 only apparent in particular regions of the genome. For instance, disassortative mating with
52 respect to genes of the major histocompatibility complex (MHC) has been reported for many
53 animal species, but meta-analyses showed that overall support is present but weak (Winternitz
54 *et al.* 2013; Kamiya *et al.* 2014). In humans, social interactions are more common between
55 individuals who differ in genomic regions associated with the immune system. This might be
56 adaptive if interacting with individuals which are resistant to different pathogens, rather than
57 similar pathogens, reduces infection risk (Christakis & Fowler 2014).

58 Genetic homophily and heterophily can have profound effects on the evolutionary
59 dynamics of populations. For example, if mates are selected locally, any type of preference for
60 particular genotypes (either homophily or heterophily) will result in a non-random pool of
61 potential mates. Genetic homophily might therefore result in inbreeding or local adaptation.
62 Individual fitness can be affected by homophily and heterophily, because fitness does not only
63 depend on an individual's own phenotype, but is also affected by the phenotypes it is interacting
64 with. For instance, the outcome of competition for food depends on the competitiveness of
65 others. In this context, selection is shaped by the so-called social environment. These additional
66 selection forces, called social selection (West-Eberhard 1983), can accelerate or counteract
67 natural selection (Wolf *et al.* 1999). To understand non-random social associations or social
68 selection, it is critical to disentangle social processes (*i.e.* sexual and social selection) from
69 spatial processes (*e.g.* the phenotype-dependent response to local variation in ecological
70 conditions, or those arising from limited dispersal), because spatial process could otherwise be
71 misinterpreted as social processes. For our purpose, we define spatial processes as all processes
72 that affect the location and movement of individuals and are spatially stable over the time
73 period of our study. Social processes are all processes that affect the spatial location and
74 movement of individuals and depend on the movements or spatial locations of others.

75 In this study we test for a relationship between social interactions outside breeding
76 seasons and genetic variation in the great tit, a seasonally breeding passerine. Between
77 breeding seasons, individuals join foraging flocks consisting of about 2 to 50 individuals (Ekman
78 1989). Individuals are more likely to join the same foraging flock when they are born in close
79 spatial proximity and when they are siblings (Grabowska-Zhang *et al.* 2016); however the
80 consequences for the spatial distribution of genetic diversity remain unknown. The foraging
81 flocks show fission-fusion dynamics (Farine *et al.* 2015), the process of changing flock
82 composition and sizes over time, due to single or multiple individuals joining or leaving these
83 flocks (Krause & Ruxton 2002). The movement of individuals between flocks potentially
84 homogenizes the genetic structure, but such homogenization might be prevented by either

85 homophily or heterophily at a broader spatial scale. We tested whether allele frequencies in
86 foraging flocks were affected by the tendency to be associated with preferred flock members in
87 general (which we call group fidelity from this point onwards) and genome-wide homophily or
88 heterophily specifically. We partitioned social and spatial effects to investigate whether allele
89 frequencies resulted from the spatial distribution of animals or whether there were additional
90 social forces driving preference for group fidelity, and whether preference was due to
91 homophily or heterophily. We also tested whether homophily or heterophily were present for
92 three categories of candidate genes which are likely to affect social interactions (See
93 supplementary materials, Table S2). We tested candidate genes for personality, because
94 phenotypic homophily for personality has been found in our population (Aplin *et al.* 2013),
95 circadian timing because we expect individuals to flock with others with similar circadian
96 rhythms, and novelty seeking, because we expect individuals with similar exploratory
97 tendencies to flock. Lastly we tested a panel of markers across the genome to determine
98 whether heterophily or homophily were present for single loci. As shown by Christakis and
99 Fowler (2014) regions of the genome may differ, with some specific regions showing
100 heterophily while most of the others show homophily. An overview of our research questions
101 and hypotheses can be found in Table 1.

102 We explicitly modelled the fission-fusion dynamics of foraging flocks as well as their
103 spatial configuration with techniques which have recently been developed for spatial analysis.
104 We deployed Asymmetric Eigenvector Maps (AEMs; Blanchet *et al.* 2011) and distance-based
105 Moran's Eigenvector Maps (db-MEMs; Dray *et al.* 2006) to model fission-fusion dynamics of the
106 foraging flocks and their spatial configuration respectively. Both methods decompose network
107 structures into uncorrelated components (eigenvectors), which can be used to describe network
108 patterns. AEMs decompose directed networks which we used to model changes of foraging
109 flocks over time (i.e. fission-fusion dynamics). Db-MEMs decompose undirected networks which
110 we used to model the spatial configuration of those flocks. A major advantage of using
111 eigenvector maps to describe spatial patterns and fission-fusion dynamics is that eigenvectors

112 are orthogonal and therefore they can be used as predictors in regression or redundancy
113 analyses. By simultaneously introducing AEMs and db-MEMs into the same model we can
114 partition variance into social and spatial processes and estimate their relative contributions to
115 the variance observed in the dependent variables. We partitioned the variance in allele
116 frequencies of 4701 autosomal SNPs between foraging flocks to AEMs and db-MEMs. This
117 enables us to quantify the relative importance of pure social processes such as preference or
118 avoidance of certain conspecifics, and the distribution of individuals in space in driving allele
119 frequencies across social groups (Fig. 1). Simulations to assess this methodology can be found in
120 the supplementary materials. To our knowledge this is the first study to disentangle social and
121 spatial processes in the genetic structure of animal groups and the first study in non-human
122 animals to investigate whole genome heterophily and homophily.

123 **Materials and methods**

124 *Study system*

125 This study was conducted in the great tit population of Wytham Woods, a 385 ha mixed
126 deciduous woodland near Oxford, U.K. (51°46'N, 1°20'W; Fig. 2), over three years. In this
127 population 250–450 great tit pairs breed annually. Breeding pairs occupy exclusive territories,
128 but when their offspring fledge these territories break down. After fledging, offspring roam
129 around with their parents and those families typically break apart after a few weeks (Naef-
130 Daenzer *et al.* 2001), while the parents often stay together (Culina *et al.* 2015, Firth *et al.* 2015).
131 The offspring assimilate into the population, but dispersal is spatially restricted, meaning that
132 individuals interact more with others born in close proximity, and also slightly more with
133 siblings (Grabowska-Zhang *et al.* 2016). As part of a long-term monitoring project, all nestlings
134 and most breeding birds are ringed for individual identification and breeding performance has
135 been recorded systematically since the early 1960s (*e.g.* Lack 1964). Over winter, birds are
136 caught by mist netting at the feeding stations at regular intervals to ring immigrating birds. Up
137 to 90% of the birds in the population are estimated to be ringed and tagged (Aplin *et al.* 2013).

138 *Identifying foraging flocks*

139 Since 2007, we have equipped birds with PIT-tags (passive integrated transponder tags)
140 encased in plastic rings, which are used for automated radio frequency identification (RFID).
141 Additional catching with mist nets was undertaken during the winter to mark part of the
142 immigrating birds. Between September 2007 and March 2010, i.e. the non-breeding seasons of
143 2007-8, 2008-9 and 2009-10, we concurrently placed 20 feeding stations in the woodland
144 equipped with RFID readers (Francis Instruments, Cambridge, UK) to register the identity and
145 time and date of PIT-tagged birds visiting those feeding stations (see supplementary video in
146 Farine *et al.* 2014). The feeding stations were rotated over 60 approximately equally spaced
147 locations every 3 days (Fig. 2). We used Gaussian mixture models (GMMs) to assign records
148 (detections on the feeder) of individuals into flocks (bursts or 'waves' of activity on the feeders).
149 The Gaussian mixture models identify instances of individuals visiting the same feeding station
150 close in time, which are defined as 'gathering events' (Psorakis *et al.* 2012). Those gathering
151 events serve as snapshots of the composition of foraging flocks and have been demonstrated to
152 outperform other flock-detection methods (Psorakis *et al.* 2015). We used the movements of
153 individuals between gathering events to quantify fission-fusion dynamics across the three
154 winters. See Fig. S2 in the supplementary materials for more details on the movements between
155 different feeding stations.

156 *Genotypes and minor allele frequencies*

157 We collected blood samples for genotyping from breeding birds from 2001 onwards.
158 These individuals were genotyped on a SNP chip with 9193 markers. This SNP chip was
159 developed based on transcriptome sequencing of great tits from Wytham Woods and genomic
160 sequencing of great tits from populations in the Netherlands (van Bers *et al.* 2010; Santure *et al.*
161 2011; van Bers *et al.* 2012). Of those 9193 markers, 7032 passed quality control (using the
162 criteria genotype call rates >95%, minor allele frequency >0.05, and Hardy-Weinberg
163 equilibrium $P > 0.001$, calculated using PLINK v1.06; Purcell *et al.* 2007). 4878 markers were

164 incorporated into a linkage map for our population (van Oers *et al.* 2014), and we focussed
165 subsequent analysis on this subset of markers. Owing to sex-biased dispersal, with females
166 dispersing longer distances than males (Verhulst *et al.* 1997), gene flow differs between
167 markers on the autosomes and the sex chromosomes. Therefore we only used the 4701 SNPs
168 which were located on the autosomes. The markers were used to genotype 2652 great tits,
169 primarily focussing on adults with life history and morphological data (van Bers *et al.* 2012); of
170 these, 962 were recorded at the feeding stations in 2007 ($N = 757$), 2008 ($N = 727$) and 2009
171 ($N = 743$). 339 genotyped individuals were recorded in two winter seasons and 88 in all three
172 winter seasons.

173 *Modelling fission-fusion dynamics*

174 To model fission-fusion dynamics we used Asymmetric Eigenvector Maps (AEMs;
175 Blanchet *et al.* 2008). AEMs belong to a family of statistical methods which are based on
176 calculating eigenvectors for adjacency or incidence matrices (Legendre & Legendre 2012).
177 These matrices describe graphs (*i.e.* networks) of spatial structure in which nodes represent the
178 spatial locations and edges a measure of distance between them. Adjacency matrices are node-
179 by-node matrices in which x_{ij} is the spatial distance between nodes i and j . Incidence matrices
180 are node-by-edge matrices in which x_{ij} is one when node i is connected by edge j to another
181 node and zero if not. These spatial configuration matrices can, however, be replaced by other
182 adjacency or incidence matrices. In our case, we used matrices that describe the fission-fusion
183 dynamic of birds moving between flocks. The nodes in these networks were the gathering
184 events (*i.e.* flocks) and the edges were movements of birds between them. AEMs have been
185 developed to model directional spatial processes, for example the distribution and abundance of
186 a species in riverine systems (Blanchet *et al.* 2011). The social structure in our population is also
187 directional, because individuals can only move from one gathering event to another where the
188 latter is later in time.

189 We first constructed fission-fusion networks for each year. Edges connected two
190 gathering events (*i.e.* the nodes) when at least one bird moved between those gathering events,
191 without having been present at any other gathering event in the meantime. We also only used
192 gathering events with at least two genotyped individuals present. Next we transformed the
193 fission-fusion networks into incidence matrices for each year. For every node (*i.e.* gathering
194 event) we gave edges (*i.e.* movements of birds) a value of 1 when the edge was part of the path
195 connecting the node to the origin and a zero when it did not (Legendre & Legendre 2012). See
196 Fig. 3 for an example fission-fusion network with corresponding incidence matrix. The origin is
197 a fictitious gathering event in which all individuals were present before any other gathering
198 event. The origin and edges connecting directly to the origin were removed before further
199 analyses. More information on the construction of the incidence matrices can be found in
200 Blanchet *et al.* (2008), Borcard *et al.* (2011) and Legendre & Legendre (2012).

201 Next we performed partial singular value decomposition of the matrices to estimate the
202 first 500 eigenvectors with the R package “irlba”, which makes use of the implicitly-restarted
203 Lanczos bidiagonalization algorithm (Baglama & Reichel 2012). We used this method to
204 estimate a subset of eigenvectors rather than the singular value decomposition method present
205 in the base package of “R”, which calculates all eigenvectors, because our incidence matrices
206 were too large for calculating all eigenvectors. We kept the 500 estimated eigenvectors for
207 further analysis; those axes described broad-scale social patterns.

208 *Modelling spatial configuration*

209 We modelled space with distance-based Moran’s Eigenvector Maps (db-MEMs; Dray *et al.*
210 2006), based on a distance matrix which we computed from the spatial coordinates of the
211 feeding stations. From the distance matrix, we built a neighbour network which linked all
212 feeding stations within the minimum distance which was necessary to keep the network fully
213 connected (498.7m; *i.e.* the longest edge of the minimum spanning tree; Fig. 2). For all feeding
214 stations which were not neighbouring (more than 498.7m apart) we replaced the distance in

215 the distance matrix by 4 times the threshold value ($4 \times 498.7 = 1998.4\text{m}$) as suggested by
216 Borcard & Legendre (2002). Next we performed a principal coordinate analysis on the
217 truncated distance matrix, resulting in 20 db-MEMs (Fig. S1 in supplementary materials). We
218 used those db-MEMs to describe spatial patterns. The db-MEM analyses were performed with
219 the R package “PCNM” (Legendre *et al.* 2013).

220 *Variance partitioning*

221 For all gathering events we calculated the allele frequencies across all 4701 mapped
222 autosomal SNPs (all SNPs were biallelic) among all individuals present at the gathering events.
223 We used variance partitioning to estimate the fractions of variation in allele frequencies
224 between gathering events explained by social (modelled with fission-fusion dynamics) and
225 spatial (based on the distance between logging sites) structure. In variance partitioning,
226 redundancy analysis is used to estimate the fractions of variation in a set of multivariate
227 response variables explained by two or more sets of explanatory variables as well as the
228 fractions in which they overlap (Peres-Neto *et al.* 2006). This collinear fraction, which we call
229 here the fraction explained by socio-spatial structure, is not simply a fraction for which we are
230 uncertain whether it resulted from spatial or social structure, but is the product of the inherent
231 role space plays in social interactions. The fraction explained by social structure minus the
232 collinear fraction will be called pure social structure and the fraction explained by spatial
233 structure minus the collinear fraction will be called pure spatial structure. Because foraging
234 flocks also move between feeding stations, we were able to separate the effects of pure social
235 structure from socio-spatial and pure spatial structure. Variance partitioning was performed
236 with the R package “vegan” (Oksanen *et al.* 2013).

237 *Group fidelity*

238 To test whether individuals interacted repeatedly with the same flock members, we
239 tested for group fidelity. We define social behaviour as different from gregarious behaviour in
240 the sense that individuals act socially when they show a preference to interact with particular

241 individuals (either specific individuals or specific phenotypes, compared to a preference to
242 interact with any conspecific in the case of gregariousness). We used randomization tests to test
243 for group fidelity. We produced null reference distributions by shuffling all individuals within
244 each year, at every gathering event in order of time. Individuals would therefore not only follow
245 the path of a random individual through space and time, but can switch paths at every gathering
246 event given that there were other individuals present which took a different path from this
247 particular gathering event (Fig. 3a). Genes would therefore follow a path through space and
248 time that could have been the path of an individual, if this individual was not socially or spatially
249 restricted. However, unrealistic movements (*e.g.* too large distances in too little time or through
250 unfavourable habitats) were excluded, because only movements which were actually made by
251 an individual were included (i.e. no new edges were created in the fission-fusion network, only
252 the identities were swapped). To test for group fidelity we recalculated the minor allele
253 frequencies for all SNPs in every gathering event. We repeated the variance partitioning for the
254 randomized data sets and compared the relative and absolute variance in minor allele
255 frequencies of the observed data sets to the values from the randomized data sets. We
256 performed 999 randomizations, included the observed data and calculated two-sided *P*-values.
257 All *P*-values reported in the results are *P*-values for all three years combined with the Fisher's
258 combined probability test.

259 *Genome wide homophily and heterophily*

260 To test whether individuals tended to associate with genetically similar or dissimilar
261 individuals we performed additional randomization tests. We randomized the data in such a
262 way that it would produce null reference distributions under which individuals did not express
263 homophily or heterophily. At every iteration, we shuffled the identities of the individuals within
264 each year while keeping the movement patterns of individuals the same (Fig. 3b). Genes would
265 therefore follow the path of a random individual through space and time and appear at the
266 gathering events this random individual was present. To test for homophily, we recalculated the

267 minor allele frequencies for all SNPs in every gathering event. We performed variance
268 partitioning on the minor allele frequencies, with the hypothesis that if similar individuals flock
269 together, the proportion of variance in minor allele frequencies attributed to social structure is
270 larger than expected. To test for heterophily we used the same randomization test however,
271 with the hypothesis that if genetically dissimilar individuals flock together, the observed
272 gathering events will show less variance in minor allele frequency than the randomized data.
273 We performed 999 randomizations, included the observed data and calculated two-sided *P*-
274 values.

275 *Candidate SNPs*

276 Of the selected SNPs for this study, 93 are known to be linked with 45 candidate genes
277 for ecologically relevant traits (van Bers *et al.* 2012). These SNPs were selected mainly based on
278 the zebra finch (*Taeniopygia guttata*) and chicken (*Gallus gallus*) genomes, but also from
279 association studies in great tit, human, house mouse (*Mus musculus*), blue tit (*Cyanistes*
280 *caeruleus*) and starling (*Sturnus vulgaris*). We selected three gene categories that might
281 potentially affect social behaviour, namely candidate genes associated with “personality”,
282 “circadian timing” and “novelty seeking”; which have 10, 6 and 13 SNPs associated with them
283 respectively. We calculated the variation in allele frequency for those particular SNPs explained
284 by pure social structure, socio-spatial structure and pure spatial structure. To test whether
285 those SNPs were significantly more or less affected by social, socio-spatial or spatial structure
286 than a random set of SNPs, we performed a randomization test. We randomly selected the same
287 number of SNPs 999 times, calculated the fractions of variance in allele frequencies explained by
288 social, socio-spatial and spatial structure and used those as null reference distributions.

289 *Single SNP homophily and heterophily*

290 To test for the presence of homophily or heterophily for particular SNPs (which are not
291 linked to candidate genes, but may be linked to other, uncharacterised, genes determining
292 population structure) we repeated the variance partitioning for all SNPs separately and

293 repeated the randomization tests for homophily and heterophily for all SNPs across the genome.
294 We only focused on the variance explained by pure social structure. To reduce computation
295 time we used dynamic stopping rules for the randomization tests. For every SNP we performed
296 at least 99 randomizations. After 99 randomizations we either stopped producing iterations
297 when the reference distribution had at least 10 randomized values higher and 10 values lower
298 than the observed value (*i.e.* sequential sampling; Besag & Clifford 1991), or after 9999
299 iterations, whichever was satisfied first. For all SNPs with 9999 iterations we estimated its P -
300 value by modelling the tail on the reference distribution with the generalized Pareto
301 distribution as described in Knijnenburg *et al.* (2009) and using the Anderson-Darling goodness
302 of fit test of the R package “ADGofTest” (Gil Bellosta 2011).

303 *Loci under selection*

304 To detect loci under selection, the leading eigenvectors of the SNP genotype matrix can
305 be used, because those eigenvectors can be interpreted as F_{st} metrics for each SNP (Weir 1996).
306 High values of F_{st} can indicate both population stratification or selection, however by controlling
307 for the background population structure, one can focus on selection (Chen *et al.* 2016). We used
308 the EigenGWAS method (Chen *et al.* 2016) to find loci under selection by first calculating the 10
309 leading eigenvectors on the SNP matrix for all individuals combined (using the EigenGWAS
310 function in the R-package “sommer”; Covarrubias-Pazarán 2016). Next, we calculated the
311 genomic inflation factor (λ_{GC}) to control for background population structure by taking the ratio
312 between the median observed χ^2 value (calculated with the “estlambda” function in the R-
313 package “GenABEL”; Aulchenko *et al.* 2007) and the median of the χ^2 distribution (*i.e.* 0.455). We
314 selected loci that were significant after controlling for the genomic inflation factor and
315 Bonferroni correction for the eigenvectors with 4 or more significant SNPs. To test whether
316 those loci were differentially affected by the social, socio-spatial or spatial structure than
317 random we performed the same randomization tests as we did for the candidate SNPs.

318

319 Results

320 In 2007, 2008 and 2009, we detected 757, 727 and 743 marked birds visiting the feeding
321 stations respectively. Those birds participated in 39 740, 36 493 and 9 531 gathering events,
322 which consisted of 1 to 37, 1 to 22 and 1 to 20 birds with medians of 3, 2 and 3 respectively. Of
323 those individuals 551 (73%), 485 (67%) and 341 (49%) were genotyped. The genotyped birds
324 participated in 27 968 (70%), 22 319 (61%) and 4 573 (48%) gathering events with two or
325 more participants being genotyped, which consisted of 2 to 32, 2 to 18 and 2 to 15 genotyped
326 birds respectively, all with medians of 3. On average there were 0.060 pairs of first-degree
327 relatives (parent-offspring or sibling pairs) present in gathering events with two or more
328 participants.

329 *Variance partitioning*

330 Across years, pure social structure explained 52.7-62.3% (range of variance across
331 years) of the variation in allele frequencies between gathering events. A further 10.3-18.0% was
332 explained by socio-spatial structure and 1.2-1.8% was explained by pure spatial structure (Fig.
333 4). The fractions of variance explained were fairly similar between the three years.

334 *Group fidelity*

335 When testing for group fidelity we found that the variance explained by pure social
336 structure and by socio-spatial structure were 2.0-4.6 ($P < 0.001$) and 1.5-3.7 ($P < 0.001$) times
337 higher than expected by chance. The effect of spatial structure on group fidelity varied between
338 years and was between 1.12 times lower and 1.34 times higher than expected by chance (Fig.
339 5a; Table S3). Group fidelity therefore explained a substantial part of the variation we found in
340 allele frequencies. This was not only the result of individuals sharing the same spatial vicinity,
341 but also of individuals specifically interacting with preferred group members.

342 *Genome-wide homophily and heterophily*

343 When we performed the randomizations to test for genome-wide homophily and
344 heterophily, the variance explained by social structure alone was 1.01-1.02 times lower than the
345 null prediction ($P < 0.001$). Socio-spatial and spatial structure explained 1.05-1.15 ($P < 0.001$)
346 and 1.08-1.18 ($P < 0.001$) times more variation than expected (Fig. 5b; Table S4). Social
347 structure explained 1-2% less variation than expected by chance; indicating that allele diversity
348 was slightly higher than expected as a result of social processes. Hence, individuals tended to
349 associate with genetically dissimilar conspecifics given their local pool of potential associates.
350 Spatial and socio-spatial structure explained 5-18% more variation than expected by chance;
351 indicating that genotypes were not randomly distributed in space, but rather clustered.

352 *Candidate SNPs*

353 Variation in SNPs associated with personality was not explained by pure social structure
354 ($P = 0.130$) or pure spatial structure ($P = 0.075$), but marginally by socio-spatial structure
355 ($P = 0.049$; Fig. 5c; Table S5). Variation in SNPs associated with circadian timing was not
356 explained by pure social structure ($P = 0.161$), socio-spatial structure ($P = 0.137$) or pure
357 spatial structure ($P = 0.187$; Fig. 5d; Table S6). Variation in SNPs associated with novelty
358 seeking was not explained by pure spatial structure ($P = 0.204$), but marginally and
359 inconsistently by pure social structure ($P = 0.010$) and socio-spatial structure ($P = 0.018$; Fig.
360 5e; Table S7). Particularly, the inconsistent patterns between years and marginal P -values,
361 despite large sample sizes, weaken any of the support for effects of the social and spatial
362 structure on variance in SNPs associated with candidate genes we found.

363 *Single SNP homophily and heterophily*

364 None of the single SNPs explained significantly more or less variance than expected in
365 both homophily (Fig. 6a) and heterophily (Fig. 6b) after Bonferroni correction and correction
366 for the false discovery rate (Benjamini & Hochberg 1995).

367 *Loci under selection*

368 After controlling for the genomic inflation factor and Bonferroni correction the principal
369 eigenvector did not show any loci under selection but for the 2nd to the 10th eigenvectors we
370 found 16, 7, 5, 5, 4, 1, 2, 2, and 2 loci respectively (Fig S10 in supplementary materials). Those
371 loci were not concentrated in particular parts of the genome, but spread seemingly random. We
372 tested whether the loci found for 2nd to the 6th eigenvectors were non-randomly distributed in
373 the social and spatial structure, but this was not the case (Fig S11 in supplementary materials).

374

375 **Discussion**

376 In this study we explored the relative importance of social interactions and space for
377 structuring allele frequencies in foraging flocks. We found that individuals tend to non-
378 randomly associate with the same flock mates during winter and those flock mates tended to be
379 genetically more diverse than expected by chance. We showed that this was not exclusively due
380 to individuals sharing spatial vicinities, but also due to some space-independent social
381 preferences which promoted both group fidelity and genome-wide heterophily. By contrast, we
382 found that genome-wide homophily primarily arose from limited spatial movements resulting
383 in spatial variation in allele frequencies in the population. None of the alleles associated with
384 candidate genes for personality, circadian timing and novelty seeking nor any other SNP-
385 markers were substantially affected by social structure.

386 *Group fidelity*

387 The great tits in our study population showed significant group fidelity. This means that
388 individuals not only congregate passively, for example to dilute predation risk (Krause & Ruxton
389 2002), confuse predators (Landeau & Terborgh 1986), benefit from selfish herd effects
390 (Hamilton 1971) or improve the detectability of predators (Caraco 1979) and food (Krebs *et al.*
391 1972), but also showed a preference for repeated interactions with particular individuals.
392 Although this was partly the result of spatial limitations on the movements of birds, for instance

393 due to territoriality or site preference, as shown by the significant effect of the socio-spatial
394 component of the model, there was also a substantial pure social component. Intraspecific
395 mutualism is arguably the simplest explanation for this (Clutton-Brock 2002). Individuals could
396 for instance alternate between foraging and vigilance behaviour, which promotes the evolution
397 of group fidelity. Great tits do produce more alarm calls in the presence of familiar individuals
398 than in the presence of unfamiliar individuals, which is in line with intraspecific mutualism
399 (Krams *et al.* 2006), and are also more likely to join known territorial neighbours when
400 mobbing predators (Grabowska-Zhang *et al.* 2012). Alternatively, intraspecific mutualism could
401 be the result of spatial hierarchy dominance. Individuals tend to be dominant in the area of their
402 former breeding territories (Ekman 1989). Further, a previous study on wintering social
403 structure in this population has demonstrated that individuals' fitness is related to the relative
404 time that they dispersed into the population relative to their social associates (Farine & Sheldon
405 2015), suggesting that local population structure has evolutionary implications.

406 We can exclude kin selection as a cause of group fidelity, because kin selection would
407 have resulted in pure social effects for homophily as well, which we did not find. Interestingly,
408 another study in the same population found that the composition of gathering events were
409 largely unstable within days and only marginally differed from random after more than ten
410 minutes (Farine *et al.* 2015). The combination of this and our study suggests that, although
411 there was a high turnover rate of individuals in flocks, individuals did regularly encounter
412 particular individuals over winter. This has also been confirmed by state-dependent modelling
413 of re-encounter rates in Aplin *et al.* (2013). Our study shows that this was not exclusively driven
414 by the spatial configuration of individuals.

415 *Genome-wide homophily and heterophily*

416 Although the effect was rather small, genome-wide heterophily was present as the result
417 of pure social processes. This suggests that individuals tend to interact more with genetically
418 dissimilar individuals. This genetic dissimilarity is probably present at many loci, albeit with

419 weak effects on the individual loci, because we did not find any signals in the analyses for single
420 SNPs. This finding seems at odds with conclusions of previous studies: for instance, great tits do
421 not show preference to mate or associate with genetically distant individuals (Szulkin *et al.*
422 2009) and they tend to interact more with their siblings in their first non-breeding season
423 (Grabowska-Zhang *et al.* 2016). However, both those processes are not purely social and are
424 more likely to be represented by the genome-wide homophily we found for socio-spatial
425 processes. Perhaps parents and offspring – though they live in close proximity and therefore
426 have a higher probability than random to encounter – display active avoidance. However, the
427 underlying social processes resulting in genome-wide heterophily remain unclear. While
428 genome-wide homophily was not the result of pure social processes, but rather socio-spatial
429 and pure spatial processes, it does nonetheless affect the genetic structure of the population.
430 This is in line with a previous study in our study population which showed that limited
431 dispersal and natural selection due to environmental conditions resulted in fine scale spatial
432 structure in genotypes (Garroway *et al.* 2013). The fact that we found genome-wide homophily
433 resulted from spatial and social-spatial processes suggests that it was not the result of
434 individuals actively being homophilous (at least at the genome-wide level), but rather caused by
435 exogenous processes such as limited dispersal.

436 *Candidate genes and Single SNP homophily and heterophily*

437 We did not find convincing evidence for social and spatial processes affecting the allele
438 frequencies of candidate genes or single SNPs. However, we cannot conclude that social
439 structure did not affect the distributions of particular social genotypes, because of the following
440 reasons. First, even though we analysed a limited number of SNPs, the probability of detecting
441 effects is low given our sample size. Thousands of individuals are needed to have sufficient
442 power to reliably detect or refute correlations between phenotypes and SNPs (Wray *et al.*
443 2013). Second, we expect social traits to be complex, so will be affected by interactions between
444 many genetic and environmental factors (Robinson *et al.* 2005). Third, the Great tit genome

445 turned out to have low linkage disequilibrium (Laine *et al.* 2016), therefore the 4701 SNP
446 markers might not sufficiently cover the whole genome to pick up signals of particular genes. A
447 denser SNP-chip would improve the coverage over the genome. Finally, the spatial distribution
448 of birds could arise from social processes occurring at a broader scale, such as when individuals
449 make decisions about where to settle during dispersal.

450 *Evolutionary perspective*

451 As shown in this study, allele frequencies in foraging flocks are affected by non-random
452 social and spatial processes. Those processes have been recognized to drive spatial
453 autocorrelation of pheno- and genotypes (Sokal & Oden 1978; Fortin & Dale 2005), but this is
454 the first study to separate social from spatial processes. If the non-random distribution of
455 genotypes between foraging flocks also translates into non-random mating, it will potentially
456 affect the evolutionary dynamics of the population. Whether flock composition affects pair
457 formation remains to be tested. It has been shown that mate choice is at least restricted in space
458 (Szulkin *et al.* 2009) and is likely to be affected by social structure as well. Since social interactions
459 can affect the strength of selection (Wolf *et al.* 1999), we also investigated whether social and
460 spatial structure affected the distribution of loci under selection. We did not find evidence for
461 this, but this might have been caused by the low linkage disequilibrium between our SNPs.

462 *Temporal and spatial effects*

463 Dealing with and describing the effects of temporal dynamics and spatial heterogeneity
464 on social networks are a current challenge in the study of animal social networks (Blonder *et al.*
465 2012; Pinter-Wollman *et al.* 2014; Farine & Whitehead 2015). Both aspects are important for
466 investigating evolutionary consequences of social interactions, because one needs to either look
467 at changes over time or compare networks (or parts of a network) in different environments.
468 Suggested methods to deal with temporal dynamics are either using discrete methods in which
469 different networks for different time periods are produced (time-aggregated networks) or
470 continuous methods in which the temporal aspects of the data is maintained (time-ordered

471 networks; Blonder *et al.* 2012; Pinter-Wollman *et al.* 2014), as in our analyses. The analytical
472 tools for continuous networks are less well developed than the discrete networks (Blonder *et al.*
473 2012; Pinter-Wollman *et al.* 2014). Taking space into account is limited to using spatially
474 restricted randomization techniques and calculating network properties for the observed and
475 randomized networks (Pinter-Wollman *et al.* 2014). The method presented here can be used for
476 continuous analysis of temporal network data, but its most important virtue lies in the
477 combination of temporal and spatial effects. The current up-scaling of social network analysis
478 (Krause *et al.* 2014) will result in the increase of the relative importance of dealing with spatial
479 effects, for which we presented here an attractive method. This method has previously been
480 used for analysing spatial data together with univariate covariates (e.g. Lasky *et al.* 2012), but
481 here we extend the use to social network data, which is multivariate. One must however note
482 that the method we used here focuses on network properties and not so much on particular
483 individuals or dyads, which would be an interesting area for future research.

484

485 **Acknowledgements**

486 We thank the many people who have collected data from the Wytham Woods great tit
487 population over the past decades and the many funding sources, too numerous to list here, that
488 supported the collection of the data. We specifically would like to thank Teddy Wilkin and Ada
489 Grabowska-Zhang for collecting most of the data on feeding station visits. We thank C. Blancher
490 and G. Buck at the Wellcome Trust Centre for Human Genetics in Oxford for genotyping. We
491 would like to acknowledge the use of the Advanced Research Computing (ARC) in Oxford in
492 carrying out this work. The work was in part funded by Natural Environment Research Council
493 grants (NER/A/S/2002/00877 and NE/F005725/1) to BCS, and by an European Research
494 Council (ERC) Advanced Investigator Award to BCS (AdG 250164) and ERC grant, project Avian-
495 EGG, number 202487, awarded to JS.

496

497 **References**

- 498 Aplin LM, Farine DR, Morand-Ferron J *et al.* (2013) Individual personalities predict
499 social behaviour in wild networks of great tits (*Parus major*). *Ecology Letters*, **16**,
500 1365–72.
- 501 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R package for
502 genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
- 503 Baglama J, Reichel L (2012) *irlba: Fast partial SVD by implicitly-restarted Lanczos*
504 *bidiagonalization*. R package.
- 505 Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and
506 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series*
507 *B*, **57**, 289–300.
- 508 van Bers NEM, van Oers K, Kerstens HHD *et al.* (2010) Genome-wide SNP detection in
509 the great tit *Parus major* using high throughput sequencing. *Molecular Ecology*, **19**
510 **Suppl 1**, 89–99.
- 511 van Bers NEM, Santure AW, van Oers K *et al.* (2012) The design and cross-population
512 application of a genome-wide SNP chip for the great tit *Parus major*. *Molecular*
513 *Ecology Resources*, **12**, 753–70.
- 514 Besag J, Clifford P (1991) Sequential Monte Carlo p-values. *Biometrika*, **78**, 301–304.
- 515 Blanchet FG, Legendre P, Borcard D (2008) Modelling directional spatial processes in
516 ecological data. *Ecological Modelling*, **215**, 325–336.
- 517 Blanchet FG, Legendre P, Maranger R, Monti D, Pepin P (2011) Modelling the effect of
518 directional spatial ecological processes at different scales. *Oecologia*, **166**, 357–68.
- 519 Blonder B, Wey TW, Dornhaus A, James R, Sih A (2012) Temporal dynamics and
520 network analysis. *Methods in Ecology and Evolution*, **3**, 958–972.
- 521 Borcard D, Gillet F, Legendre P (2011) *Numerical ecology with R*. Springer Verlag, New
522 York.
- 523 Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of
524 principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
- 525 Caraco T (1979) Time budgeting and group size: a test of theory. *Ecology*, **60**, 618–627.
- 526 Chen G-B, Lee SH, Zhu Z-X, Benyamin B, Robinson MR (2016) EigenGWAS: finding loci
527 under selection through genome-wide association studies of eigenvectors in
528 structured populations. *Heredity*, **117**, 1–11.

- 529 Christakis N, Fowler J (2014) Friendship and natural selection. *Proceedings of the*
530 *National Academy of Sciences of the United States of America*, **111**, 10796–10801.
- 531 Clutton-Brock T (2002) Breeding together: kin selection and mutualism in cooperative
532 vertebrates. *Science*, **296**, 69–72.
- 533 Covarrubias-Pazaran G (2016) Genome-Assisted prediction of quantitative traits using
534 the r package sommer. *PLoS One* **11**, 1–15.
- 535 Croft DP, James R, Krause J (2008) *Exploring Animal Social Networks*. Princeton
536 University Press, Princeton.
- 537 Croft DP, James R, Ward a JW *et al.* (2005) Assortative interactions and social networks
538 in fish. *Oecologia*, **143**, 211–9.
- 539 Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive
540 framework for principal coordinate analysis of neighbour matrices (PCNM).
541 *Ecological Modelling*, **196**, 483–493.
- 542 Ekman J (1989) Ecology of non-breeding social systems of *Parus*. *The Wilson Bulletin*,
543 **101**, 263–288.
- 544 Farine DR, Aplin LM, Garroway CJ, Mann RP, Sheldon BC (2014) Collective decision
545 making and social interaction rules in mixed-species flocks of songbirds. *Animal*
546 *Behaviour*, **95**, 173–182.
- 547 Farine DR, Firth JA, Aplina LM *et al.* (2015) The role of social and ecological processes in
548 structuring animal populations: a case study from automated tracking of wild birds.
549 *Royal Society Open Science*, **2**, 150057.
- 550 Farine DR, Sheldon BC (2015) Selection for territory acquisition is modulated by social
551 network structure in a wild songbird. *Journal of Evolutionary Biology*, **28**, 547–556.
- 552 Farine DR, Whitehead H (2015) Constructing, conducting, and interpreting animal social
553 network analysis. *Journal of Animal Ecology*, **84**, 1144–1163 .
- 554 Fortin M-J, Dale M (2005) *Spatial Analysis: A Guide for Ecologists*. Cambridge University
555 Press, Cambridge.
- 556 Fowler JH, Settle JE, Christakis N a (2011) Correlated genotypes in friendship networks.
557 *Proceedings of the National Academy of Sciences of the United States of America*,
558 **108**, 1993–7.
- 559 Fu F, Nowak MA, Christakis NA, Fowler JH (2012) The evolution of homophily. *Scientific*
560 *Reports*, **2**, 845.
- 561 Garroway CJ, Radersma R, Sepil I *et al.* (2013) Fine-scale genetic structure in a wild bird
562 population: the role of limited dispersal and environmentally-based selection as
563 causal factors. *Evolution*, **67**, 3488–3500.

- 564 Gil Bellosta CJ (2011) ADGofTest: Anderson-Darling GoF test. R package version 0.3.
- 565 Grabowska-Zhang AM, Hinde CA, Garroway CJ, Sheldon BC (2016). Wherever I may
566 roam: social viscosity and kin affiliation in a wild population despite natal
567 dispersal. *Behavioral Ecology*, **27**, 1263–1268.
- 568 Grabowska-Zhang AM, Sheldon BC, Hinde CA (2012) Long-term familiarity promotes
569 joining in neighbour nest defence. *Biology Letters*, **8**, 544–546.
- 570 Hamilton WD (1971) Geometry for the selfish herd. *Journal of Theoretical Biology*, **31**,
571 295–311.
- 572 Kamiya T, O'Dwyer K, Westerdahl H, Senior a., Nakagawa S (2014) A quantitative review
573 of MHC-based mating preference: the role of diversity and dissimilarity. *Molecular
574 Ecology*, **23**, 5151–5163.
- 575 Katsnelson E, Lotem A, Feldman MW (2014) Assortative social learning and its
576 implications for human (and animal?) societies. *Evolution*, **68**, 1894–1906.
- 577 Knijnenburg TA, Wessels LFA, Reinders MJT, Shmulevich I (2009) Fewer permutations,
578 more accurate *P*-values. *Bioinformatics*, **25**, i161–i168.
- 579 Krams I, Krama T, Igaune K (2006) Alarm calls of wintering great tits *Parus major*:
580 warning of mate, reciprocal altruism or a message to the predator? *Journal of Avian
581 Biology*, **37**, 131–136.
- 582 Krause J, Croft DP, James R (2007) Social network theory in the behavioural sciences:
583 potential applications. *Behavioral Ecology and Sociobiology*, **62**, 15–27.
- 584 Krause J, James R, Franks D, Croft D (2014) *Animal Social Networks*. Oxford University
585 Press, Oxford.
- 586 Krause J, Ruxton GD (2002) *Living in groups*. Oxford University Press, New York.
- 587 Krebs JR, MacRoberts MH, Cullen JM (1972) Flocking and feeding in the great tit *Parus
588 major* - an experimental study. *Ibis*, **114**, 507–530.
- 589 Lack D (1964) A long-term study of the great tit (*Parus major*). *The Journal of Animal
590 Ecology*, **33**, 159–173.
- 591 Laine VN, Gossmann TI, Schachtschneider KM *et al.* (2016) Evolutionary signals of
592 selection on cognition from the great tit genome and methylome. *Nature
593 Communications*, **7**, 10474.
- 594 Landeau L, Terborgh J (1986) Oddity and the “confusion effect” in predation. *Animal
595 Behaviour*, **34**, 1372–1380.

- 596 Lasky JR, Des Marais DL, McKay JK *et al.* (2012) Characterizing genomic variation of
597 *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, **21**,
598 5512–29.
- 599 Legendre P, Borcard D, Blanchet FG, Dray S (2013) PCNM: MEM spatial eigenfunction
600 and principal coordinate analyses. R package.
- 601 Legendre P, Legendre L (2012) *Numerical ecology*. Elsevier, Amsterdam.
- 602 Lusseau D, Newman MEJ (2004) Identifying the role that animals play in their social
603 networks. *Proceedings of the Royal Society B*, **271 Suppl**, S477–81.
- 604 Massen JJM, Koski SE (2014) Chimps of a feather sit together: chimpanzee friendships
605 are based on homophily in personality. *Evolution and Human Behavior*, **35**, 1–8.
- 606 van Oers K, Santure AW, De Cauwer I *et al.* (2014) Replicated high-density genetic maps
607 of two great tit populations reveal fine-scale genomic departures from sex-equal
608 recombination rates. *Heredity*, **112**, 307–316. Oksanen J, Blanchet FG, Kindt R *et al.*
609 (2013) vegan: Community Ecology Package. R package.
- 610 Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species
611 data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–25.
- 612 Pinter-Wollman N, Hobson EA, Smith JE *et al.* (2014) The dynamics of animal social
613 networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology*, **25**,
614 242–255.
- 615 Psorakis I, Roberts SJ, Rezek I, Sheldon BC (2012) Inferring social network structure in
616 ecological systems from spatio-temporal data streams. *Journal of the Royal Society*,
617 *Interface / the Royal Society*, **9**, 3055–66.
- 618 Psorakis I, Voelkl B, Garroway CJ *et al.* (2015) Inferring social structure from temporal
619 data. *Behavioral Ecology and Sociobiology*, **69**, 857–866.
- 620 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome
621 association and population-based linkage analyses. *American Journal of Human*
622 *Genetics*, **81**, 559–75.
- 623 Robinson GE, Grozinger CM, Whitfield CW (2005) Sociogenomics: social life in
624 molecular terms. *Nature Reviews Genetics*, **6**, 257–70.
- 625 Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J (2011) Characterisation of the
626 transcriptome of a wild great tit *Parus major* population by next generation
627 sequencing. *BMC Genomics*, **12**, 283.
- 628 Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological*
629 *Journal of the Linnean Society*, **10**, 199–228.

- 630 Sundaresan SR, Fischhoff IR, Dushoff J, Rubenstein DI (2007) Network metrics reveal
631 differences in social organization between two fission-fusion species, Grevy's zebra
632 and onager. *Oecologia*, **151**, 140–9.
- 633 Szulkin M, Zelazowski P, Nicholson G, Sheldon BC (2009) Inbreeding avoidance under
634 different null models of random mating in the great tit. *Journal of Animal Ecology*,
635 **78**, 778–788.
- 636 Verhulst S, Perrins CM, Riddington R (1997) Natal dispersal of great tits in a patchy
637 environment. *Ecology*, **78**, 864–872.
- 638 Weir BS (1996) *Genetic Data Analysis*. 2nd edition. Sinauer Associates, Sunderland, MA,
639 USA.
- 640 West-Eberhard M (1983) Sexual selection, social competition, and speciation. *Quarterly*
641 *Review of Biology*, **58**, 155–183.
- 642 Winternitz JC, Minchey SG, Garamszegi LZ *et al.* (2013) Sexual selection explains more
643 functional variation in the mammalian major histocompatibility complex than
644 parasitism. *Proceedings of the Royal Society B*, **208**, 20131605.
- 645 Wolf JB, Brodie III ED, Moore AJ (1999) Interacting phenotypes and the evolutionary
646 process. II. Selection resulting from social interactions. *The American Naturalist*,
647 **153**, 254–266.
- 648 Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM (2013). Pitfalls of
649 predicting complex traits from SNPs. *Nature Reviews Genetics*, **14**, 507–515.

650

651

652 RR, CJG, JS and BCS designed this study. The phenotypic data and blood sampling was organized
653 by BCS, with contributions to the data collection by RR, CJG and DRF. RR performed most data
654 analyses, with involvement of CJG in the eigenvector analyses. The SNP genotyping and quality
655 control was organised and performed by JS, AWS and IDC. Gathering events were quantified by
656 DRF. RR drafted the manuscript. All authors provided comments on the manuscript, and read
657 and approved the final manuscript.

658

659 **Data accessibility**

660 The individual by gathering events matrices, the spatial weighting matrix and individual
661 genotypes will be uploaded to Dryad.

662

663 **Figures:**

664 **Figure1:** Schematic presentation of exemplar fission-fusion networks and associated Venn
665 diagrams, showing the relative contributions of social and spatial structure to allele frequency
666 variation between flocks given various degrees of group and site fidelity by individuals. A, B, C
667 and D represent different spatial locations. The open circles represent observations of foraging
668 flocks at those locations and the continuous lines between the open circles are movements of
669 individuals from one observation to another. The Venn diagrams below the fission-fusion
670 networks represent the relative contributions of social structure (in red) and space (in blue)
671 and are all similarly scaled. In (a) individuals move completely at random, not restricted by
672 group or site. In (b) individuals are perfectly site and group faithful. In (c) flocks are site faithful,
673 but some individuals change flocks. In (d) individuals stay continuously with the same flock, but
674 those flocks are largely unconstrained in space. (e) is similar to (d), but there are some
675 occasions in which individuals change groups.

676

677 **Figure2:** The outline (in grey) of Wytham Woods, Oxfordshire, United Kingdom. Black dots
678 represent the 60 locations at which 20 feeding stations rotated about in this study. Lines
679 connecting the dots identify feeding stations which are considered neighbouring sites (i.e. sites
680 that are less than 498.7m apart; the longest edge of the minimum spanning tree). Dashed lines
681 are a 1 by 1 km grid.

682

683 **Figure 3:** A graphical representation of the randomizations used for testing (a) group fidelity
684 and (b) homophily and heterophily. The numbered circles represent gathering events at three
685 different feeding stations (A, B, and C). The coloured lines are the movements of individuals
686 between gathering events (every colour represents a different individual). To test for group
687 fidelity we randomly shuffled movements of individuals. For example from gathering event 1
688 the movement of the red individual is replaced by the black individual and the movement of the

689 blue individual by the yellow individual. From gathering event 3, which now consists of
690 individuals black and yellow, the black individual follows the original path of the blue individual
691 and the yellow individual replaced the movement of the red individual. Individuals follow a path
692 which was physically possible for individuals to follow, but was not necessarily done by any of
693 the individuals (*e.g.* the red path after the randomization [gathering events 1,4,6 and 5]). To test
694 for homophily and heterophily we randomly shuffled all identities of individuals. For example
695 the black individual replaced the red individual and the red individual replaced the blue
696 individual. Individuals only followed paths which were completely followed by others. (c) The
697 incidence matrix of the networks drawn in (a) and (b). Rows are nodes (*i.e.* gathering events)
698 and columns are edges (*i.e.* movements). *E.g.* the one in row 1 and column a indicates that node
699 1 is connected to another node by edge a. This other node is 2, because row 2 also has a 1 in
700 column a.

701

702 **Figure 4:** Venn diagrams showing the proportion of variation in all 4 701 allele frequencies
703 between gathering events explained by social structure (in red) and space (in blue) for (a) 2007,
704 (b) 2008 and (c) 2009 (39 740, 36 493 and 9 531 gathering events respectively). Circles are
705 scaled within years, but not between years.

706

707 **Figure 5:** Results of the randomization test for (a) group fidelity (b) genome-wide homophily
708 and heterophily and the SNPs associated with (c) personality, (d) circadian timing and (e)
709 novelty seeking expressed as the average amount of variance explained by one SNP at one
710 gathering event. Black dots are the observed values, while coloured dots are the values of the
711 null distributions. To ease comparison between all tests and years, the amount of variance in
712 allele frequencies explained by each component was divided by the number of SNPs (to account
713 for the variable number of SNPs involved in the different randomization tests) as well as the

714 number of gathering events (ge) (to account for the difference in the number of gathering
715 events between years).

716

717 **Figure 6:** Manhattan plots of genome-social environment association studies for homophily and
718 heterophily. SNPs explaining less variance than expected reveal heterophily and are blue. SNPs
719 explaining more variance than expected reveal homophily and are red. Dashed line is the
720 Bonferroni corrected significance level. Dot radius is linear to $-\log^{10}(p)$.

721

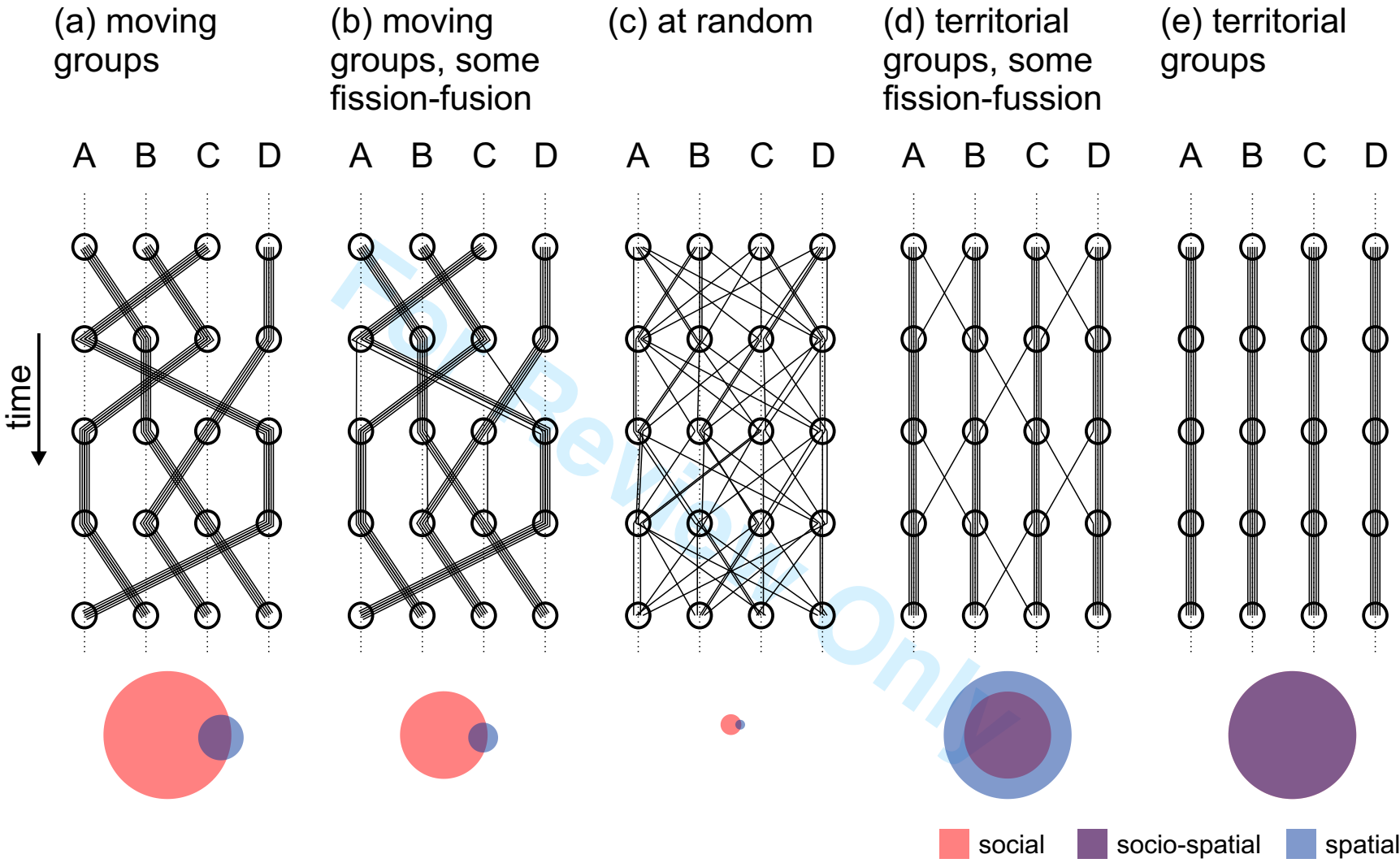
For Review Only

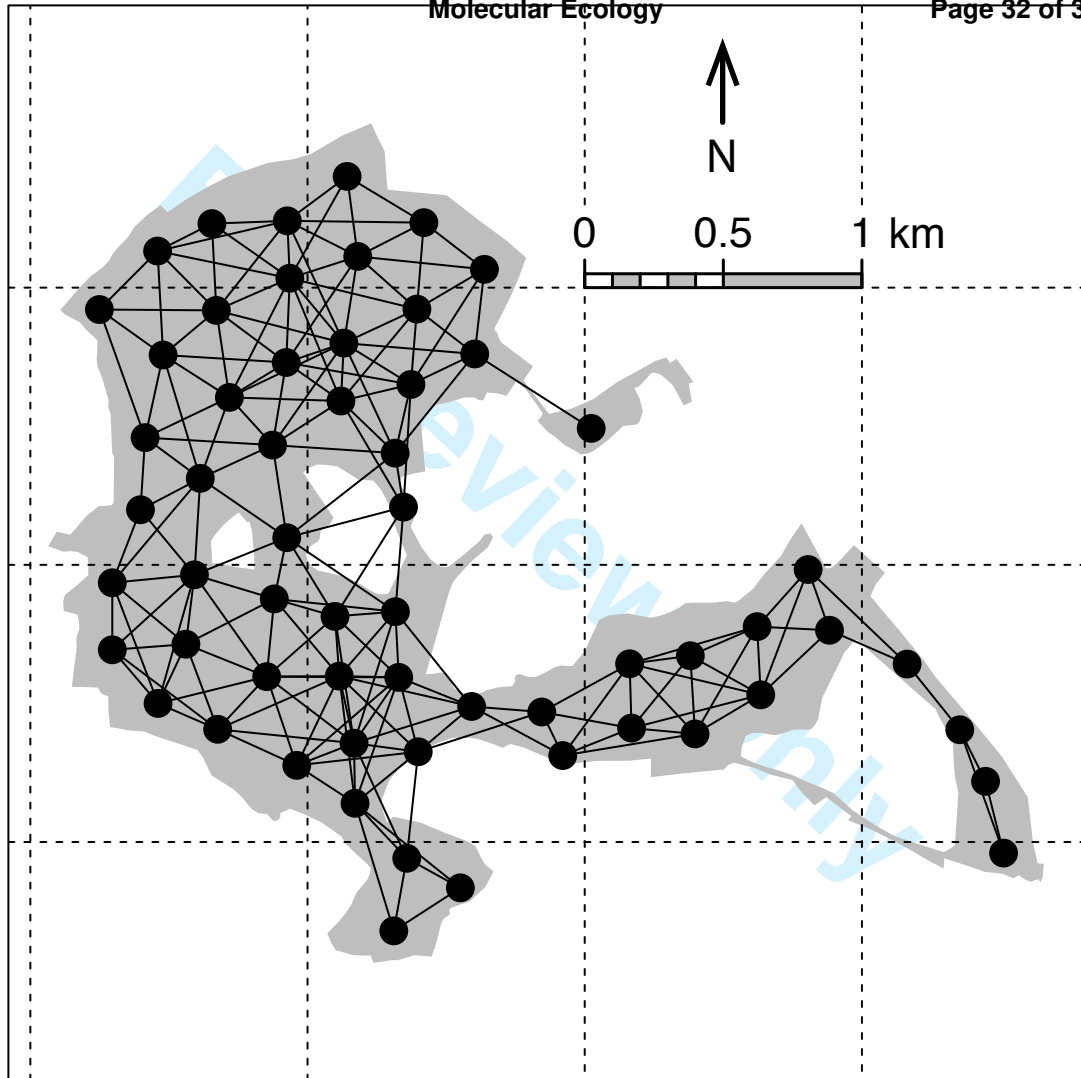
722 Table 1:

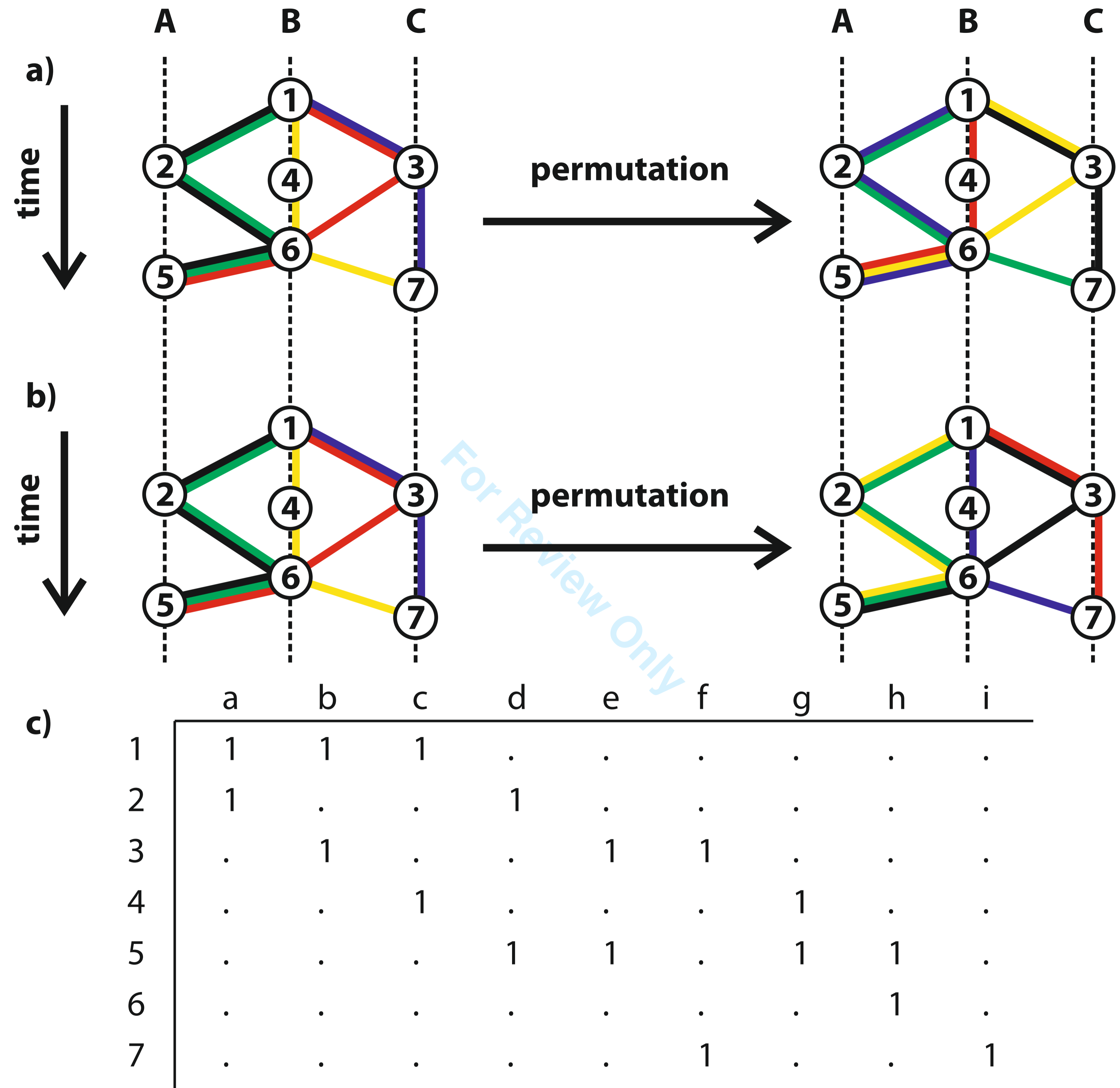
723 Overview of the questions and hypotheses addressed in this study. The unit of interest is the variance which our models try to explain with the social
 724 and spatial components. For the social and spatial components we note whether the observed proportion of the variance should be equal ($O = E$),
 725 larger ($O > E$) or smaller ($O < E$) than expected by chance. The randomized unit is the unit which was randomized in the randomization protocol.

Question	Unit of interest	Hypotheses	Social component	Spatial component	Randomized unit
Is the genetic structure at the population level affected by non-random movements of individuals?	Total variance in allele frequencies of flocks for all SNPs	H ₀ : No, random movements	$O = E$	$O = E$	Identities of individuals leaving a gathering event (Fig 3a)
		H ₁ : Yes, group fidelity	$O > E$	$O = E$	
		H ₂ : Yes, spatially restricted movements	$O = E$	$O > E$	
Is the genetic structure at the population level affected by clustering of genetically similar or dissimilar individuals?	Total variance in allele frequencies of flocks for all SNPs	H ₀ : No, no preference	$O = E$	$O = E$	Identities of all individuals within a year (Fig 3b)
		H ₁ : Yes, genome-wide homophily	$O > E$	$O = E$	
		H ₂ : Yes, genome-wide heterophily	$O < E$	$O = E$	
		H ₃ : Yes, spatial clustering of genetically similar (or related) individuals	$O = E$	$O > E$	
Is the distribution of particular candidate genes in the population differentially affected by social structure?	Total variance in allele frequencies of flocks for SNPs associated with the candidate genes	H ₀ : No, not different	$O = E$	N.A.	Identities of all SNPs associated with candidate genes
		H ₁ : Yes, homophily for candidate genes	$O > E$	N.A.	
		H ₂ : Yes, heterophily for candidate genes	$O < E$	N.A.	
Is the distribution of particular single loci in the population differentially affected by social structure?	Total variance in allele frequencies of flocks for single SNPs	H ₀ : No, not different	$O = E$	N.A.	Identities of single SNPs
		H ₁ : Yes, homophily for particular single loci	$O > E$	N.A.	
		H ₂ : Yes, heterophily for particular single loci	$O < E$	N.A.	

726



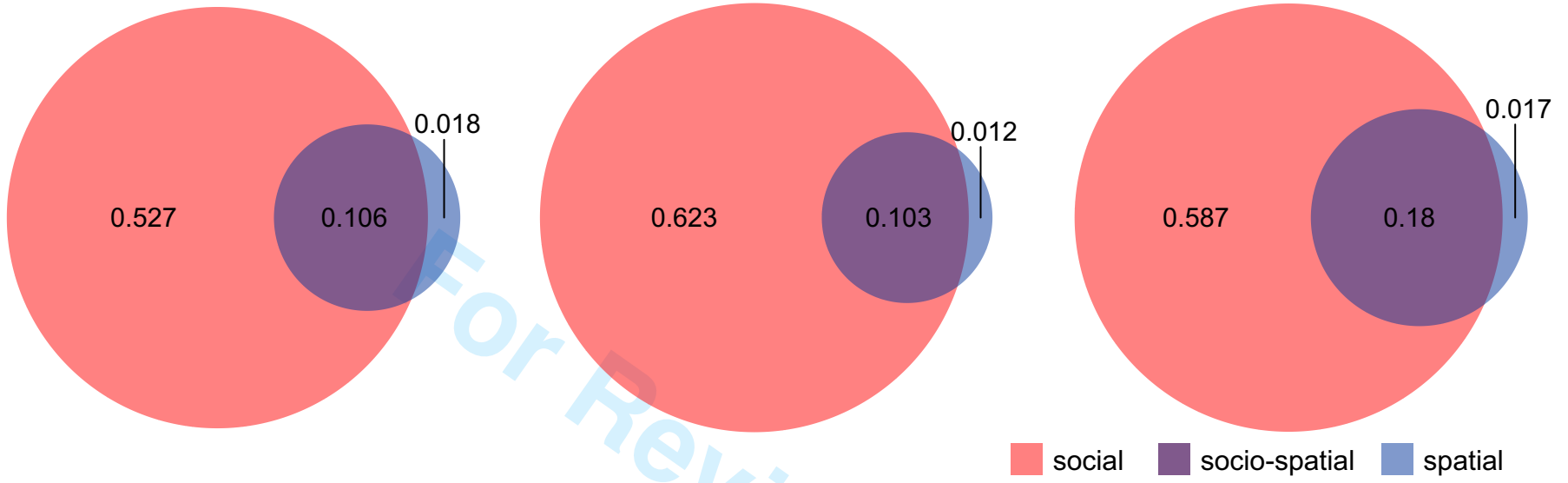




(a) 2007

(b) 2008

(c) 2009



For Review Only

