Effects of social organisation and elevation on spatial genetic structure in a montane ant

Amaranta Fontcuberta[1], Martin Kapun [2,3], Patrick Tran Van[1], Jessica Purcell[1, 4]* and Michel Chapuisat[1]*

1. Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

2. Center for Anatomy and Cell Biology, Department of Cell and Developmental Biology, Medical University of Vienna, 1090 Vienna, Austria

3. Natural History Museum of Vienna, 1010 Vienna, Austria

4. Department of Entomology, University of California, CA 9252, Riverside, USA

Correspondence: Amaranta.Fontcuberta@unil.ch; Michel.Chapuisat@unil.ch

* Co-senior authors

ABSTRACT

Studying patterns of population structure across the landscape sheds light on dispersal and demographic processes, which helps to inform conservation decisions. Here, we study how social organisation and landscape factors affect spatial patterns of genetic differentiation in an ant species living in mountainous regions. Using genome-wide SNP markers, we assess population structure in the Alpine silver ant, *Formica selysi*. This species has two social forms controlled by a supergene. The monogyne form has one queen per colony, while the polygyne form has multiple queens per colony. The two social forms co-occur in the same populations. For both social forms, we found a strong pattern of isolation-by-distance across the Alps. Within regions, genetic differentiation between populations was weaker for the monogyne form than for the polygyne form. We suggest that this pattern is due to higher dispersal and effective population sizes in the monogyne form. In addition, we found stronger isolation-by-distance and lower genetic diversity in high elevation populations, compared to lowland populations, suggesting that gene flow between *F. selysi* populations in the Alps occurs mostly through riparian corridors along lowland valleys. Overall, this survey highlights the need to consider intraspecific polymorphisms when assessing population connectivity and calls for special attention to the conservation of lowland habitats in mountain regions.

Keywords: social polymorphism, dispersal, population genetics, landscape genetics, mountain – valley model

INTRODUCTION

Sociality has profound effects on population processes (Wilson, 1992; Szathmary and Maynard Smith, 1995). Because few individuals reproduce in each group, eusocial insects (ants, bees, wasps and termites) have comparatively smaller effective population sizes than solitary insects (Pamilo and Crozier, 1997), which translates into lower genetic diversity within populations, and potentially higher inbreeding (Hedrick and Parker, 1997; Chapman and Bourke, 2001). Social insects often form sessile and perennial colonies, and philopatry of reproductive individuals is common (Le Galliard, Ferrière and Dieckmann, 2005; Seppä, 2008), so that local population genetic structure is expected (Ross, 2001). Together, small effective population size, low diversity and strong population structure reduce selection efficiency and adaptive capacity (Romiguier *et al.*, 2014; Settepani, Bechsgaard and Bilde, 2016; Weyna and Romiguier, 2021). Therefore, some aspects of sociality may hamper the capacity of social insects to respond to rapid environmental change, presenting an added challenge for their conservation (Chapman and Bourke, 2001; Seppä, 2008; Fisher *et al.*, 2019).

The ability of social insects to disperse and cope with environmental change depends on their social organisation. Ant colonies can have a single queen (= "monogyne") or multiple queens (= "polygyne"). The monogyne and polygyne social forms generally differ in several traits, including colony size and lifespan, sex allocation, dispersal, and colony founding strategy (Keller, 1993). Across species, queens of the monogyne social form disperse on the wing and establish novel colonies independently. In contrast, queens of the polygyne social form have the additional options of staying in their natal nests and establishing new polygyne colonies by dispersing on foot with workers ("colony budding", Bourke and Franks 1995). Because of higher long-range dispersal, population genetic structure is generally weaker in monogyne species, compared to polygyne species (e.g. Seppä and Pamilo 1995; Chapuisat et al. 1997; Ross 2001). This pattern has also been documented between monogyne and polygyne populations of polymorphic species (e.g. Ross and Shoemaker 1997; Sundström et al. 2005; Huszár et al. 2014). Yet, when social forms are allopatric, the effects of social organisation, geography and ecology are confounded. Socially polymorphic species in which monogyne and polygyne colonies occur in sympatry offer the opportunity to study the direct effects of social organisation on dispersal and population genetic structure.

Several landscape factors tend to restrict gene flow and lead to population structure. First, gene flow may be constrained by geographic distance, causing distant populations to diverge through drift ("isolation-by-distance", Wright 1943). The process is exacerbated by barriers to movement, such as water bodies, high mountains or urbanised areas. Second, populations may experience ecological isolation, leading to divergent selection and local adaptation ("isolation-by-environment", Wang and Bradburd 2014). These factors, alone or in combination, act at multiple spatial scales, and may lead to complex population genetic patterns in heterogenous landscapes (Cushman *et al.*, 2006; Meirmans, 2012). Mountains encompass a great range of elevation, climate and ecosystems within small regions. Thus, mountain regions are prime areas to investigate how social organisation and landscape factors interact in shaping dispersal and population structure.

Here, we study the population genetic structure of a montane ant species, *Formica selysi*. This socially polymorphic ant is a pioneer species colonizing floodplains along mountain rivers (Lude, Reich and Plachter, 1999; Chapuisat, Bocherens and Rosset, 2004; Zahnd *et al.*, 2021). Natural floodplains are among the most diverse ecosystems on earth, but are highly threatened: up to 90% of natural European floodplains have disappeared as a result of human activity (Tockner and Stanford, 2002). Although *F. selysi* can be locally common (Zahnd *et al.*, 2021), it is considered a threatened species in certain parts of the European Alps (Glaser, 2007).

Most well-sampled populations of *F. selysi* have both monogyne and polygyne colonies (Chapuisat, Bocherens and Rosset, 2004; Purcell, Pellissier and Chapuisat, 2015). Colony social organisation is controlled by a large supergene with two haplotypes, *M* and *P* (previously called *Sm* and *Sp*; Purcell et al. 2014). Queens and workers in monogyne colonies are homozygous for the *M* haplotype, whereas queens and workers in polygyne colonies are homozygous for the *P* haplotype or heterozygous (*MP* genotype; Purcell et al. 2014; Avril et al. 2019). Outside of the supergene, there is little genetic differentiation between social forms (Chapuisat, Bocherens and Rosset, 2004; Purcell and Chapuisat, 2013; Purcell *et al.*, 2014), suggesting extensive gene flow.

The monogyne and polygyne social forms of *F. selysi* differ in a suite of traits, including sex allocation, dispersal and mode of colony founding. Monogyne colonies produce 90% of the alate females (the future queens) dispersing by flight (Fontcuberta *et al.*, 2021). These females of monogyne origin are larger and more successful at independent colony founding than females produced by polygyne colonies (Rosset and Chapuisat, 2007; Reber, Meunier and Chapuisat, 2010; De Gasperin *et al.*, 2020). Some females of polygyne origin also disperse by flight and found colonies independently (Rosset and Chapuisat, 2006; Reber, Meunier and Chapuisat, 2010; De Gasperin *et al.*, 2020; Blacher, De Gasperin and Chapuisat, 2021; Fontcuberta *et al.*, 2021). Females from polygyne colonies tend to mate with slightly related males (Avril *et al.*, 2019), which suggests that some of the polygyne females mate inside or close to their natal nest and forgo dispersal.

Restricted dispersal of polygyne females is expected to result in stronger population genetic structure and isolation-by-distance in the polygyne social form, compared to the monogyne form (Ross, 2001; Sundström, Seppä and Pamilo, 2005). However, male-mediated gene flow within and between social forms might erode population genetic structure (Avril *et al.*, 2019). Previous studies did not detect strong differences between *F. selysi* social forms in the degree of isolation-by-distance among colonies within populations (Chapuisat, Bocherens and Rosset, 2004; Avril *et al.*, 2019). Whether genetic structure differs between social forms at a larger geographical scale has not been investigated so far.

A previous genetic survey of several populations in the Alps revealed that large river drainage basins have a strong influence on spatial genetic differentiation in *F. selysi* (Purcell, Pellissier and Chapuisat, 2015). Little genetic differentiation was detected between populations within mountain valleys, suggesting high gene flow along elevation gradients. Dispersal success depends on the ability to cross geographical barriers and availability of suitable habitat within flying distance (Hakala, Seppä and Helanterä, 2019). The ability of *F. selysi* to fly over long distances and cross mountain ridges is unknown. Its habitats consist of gravel and sandy floodplains along rivers, which are rare in steep mountains valleys and become more and more fragmented with increasing active management of water courses (Ballinger, Lake and Mac Nally, 2007). Thus, more research is needed to understand how this riverine ant species disperses and colonizes its discontinuous mountain habitats, and how elevation affects population connectivity.

We used genome-wide ddRAD-seq markers to infer the population genetic structure of *F. selysi* across several valleys in the European Alps. The sampling scheme covers a large portion of the species range. We sampled populations in three well-separated geographical regions belonging to two drainage basins (Rhône and Rhine, Purcell et al. 2015) and comprising strong elevation contrasts in independent valleys. Our goals were first, to identify landscape factors affecting gene flow in mountains; and second, to investigate how intraspecific variation in social organisation affects patterns of population structure. Overall, this study sheds light on factors affecting population connectivity and dispersal in social insects, which can prove valuable for conservation management.

METHODS

Sampling and genotyping

Formica selysi lives in riverine ecosystems throughout the European Alps and the Pyrenees mountains (Seifert, 2002). We sampled workers in 152 colonies from 13 localities ranging from 180 m to 1450 m in elevation (1-32 colonies per locality, **Table S1**). In each locality monogyne and/or polygyne colonies were sampled within a 1 km^2 area (**Table S1**). The sampling localities were situated along the Rhine River or tributaries (3 localities, east Switzerland and west Austria), along the Upper Rhône River or tributaries (6 localities, west Switzerland) and along tributaries of the Lower Rhône River (4 localities, France; **Figure 1**, **Table S1**). Each locality represents a separate population.

We genotyped one worker per colony. We extracted DNA from the head and thorax of each worker using the Qiagen DNeasy Blood and Tissue kit, following the protocol for insect tissue. We obtained double-digest RAD sequence data by following the ddRAD-seq protocol described in Brelsford et al. (2016). In brief, we digested genomic DNA using restriction enzymes *EcoRI* and *MseI*, ligated inline barcoded adapters, removed DNA fragments shorter than 250 bp using AMPure magnetic beads, carried out PCR amplification of each individual in triplicate, during which we added a second unique adapter for each independent plate, and carried out a final size selection on the pooled libraries, to retain sequences in the 400-500 bp range. The resulting libraries were sequenced on the Illumina 2500 Hi Seq platform of the Lausanne Genomic Technologies Facility.

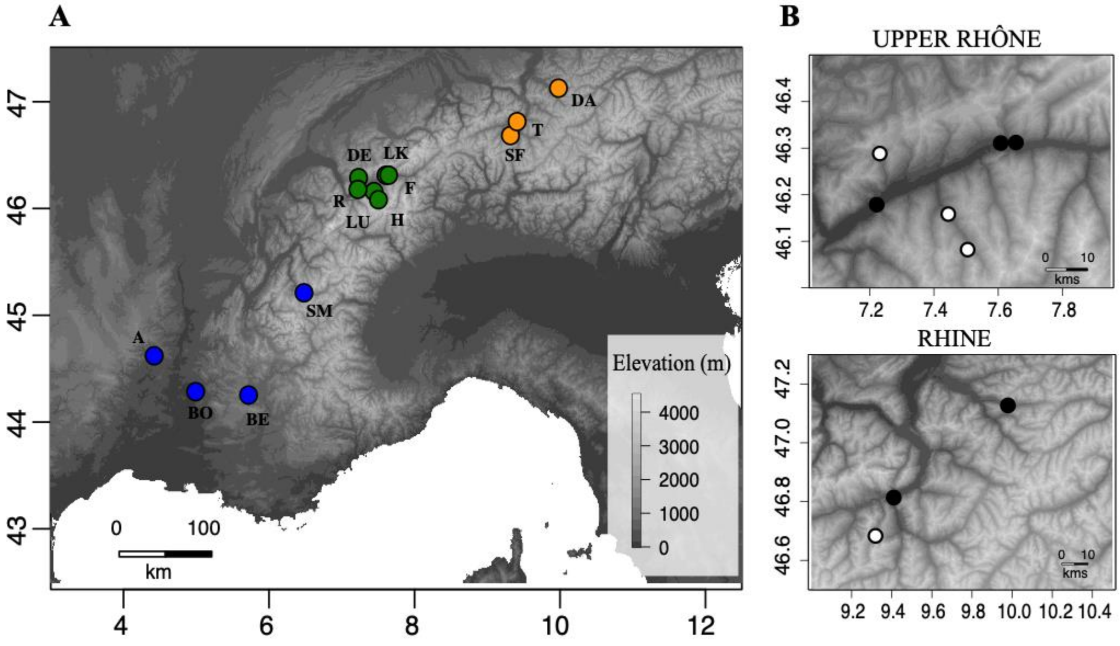

Figure 1. Map of sampling localities (= populations). A) Colours indicate three well-separated regions, blue: Lower Rhône region (A: Aubenas, BO: Buisson, BE: Bussets, SM: St-Michel), green: Upper Rhône region (F: Finges, LK: Leuk, R: Riddes, LU: Luette, H: Les Haudères, DE: Derborence), orange: Rhine region (T: Tamins, SF: Safien, DA: Dalaas). B) Zoom of the localities in the Upper Rhône and Rhine regions. White circles are highland populations (> 1000 m), and black circles are lowland populations (< 1000 m). Axes show the longitude and latitude. The grey

tones in the background maps show elevation, based on SRTM elevation raster data with 30 m resolution.

Bioinformatics

Demultiplexing and quality control of raw sequences were done with the *process_radtags* pipeline in STACKS v. 2.2 (Catchen *et al.*, 2013). Clean reads were aligned to an upgraded version of the reference genome of *Formica selysi* (Brelsford et al. 2020, NCBI, GenBank accession number: GCA_009859135.1)*,* using BWA v. 0.7.17 (Li and Durbin, 2009). Single Nucleotide Polymorphisms (SNPs) and genotypes were called with the *ref_map* pipeline in STACKS, using default parameters. The initial consensus output catalogue from the *populations* program contained 628,232 RAD loci, with average length of 84.7 bp and average sample coverage of 28.9x. In total, 323,797 SNPs were retained, distributed across 99,299 polymorphic RAD loci.

Further SNP filtering was done using the VCFtools (Danecek *et al.*, 2011) and the "VcfR" *R* package (Knaus BJ, 2017). Genotypes with quality score lower than 20 and sequencing depth lower than three-fold were considered missing data. We retained one random polymorphic site per RAD locus, to avoid bias due to linkage disequilibrium. We removed sites with heterozygosity higher than 0.70, to exclude merging paralogous loci (Paris, Stevens and Catchen, 2017). We only retained SNPs with minor allele frequency higher than 0.01 and mapping to one of the 27 chromosome-length scaffolds of the reference genome. We further removed individuals with more than 30% of missing data and selected SNPs present in 95% of the individuals retained. The resulting dataset had 13,421 SNPs, of which 923 were on chromosome 3, which contains the non-recombining social supergene (Purcell *et al.*, 2014), and 12,498 were in the remaining 26 chromosomes.

Determination of social form

We inferred the social form of each individual from their social supergene genotype (Purcell *et al.*, 2014; Brelsford *et al.*, 2020). Individuals in monogyne colonies are homozygous for the *M* haplotype, whereas individuals in polygyne colonies are either homozygous for the *P* haplotype or heterozygous (*MP* genotype; Purcell et al. 2014; Avril et al. 2019). Worker genotypes were perfectly associated with colony queen number across hundreds of individuals from both types of colonies, suggesting that worker drifting between social forms is unlikely (Purcell *et al.*, 2014; Avril *et al.*, 2019; Fontcuberta *et al.*, 2021; Zahnd *et al.*, 2021). To determine the supergene genotype of each individual, we ran a PCA on SNPs in chromosome 3, using the "adegenet" R package (Jombart and Ahmed, 2011). The first component (32.5% of variance) distinguishes the three supergene genotypes. The inbreeding coefficient (F_{IS}), calculated with VCFtools, distinguishes homozygous from heterozygous individuals (**Figure S1**). Overall, 106 individuals belonged to the monogyne social form, whereas 46 individuals belonged to the polygyne social form (**Table S1**). We will refer to them as monogyne and polygyne individuals, respectively.

Population genetic analyses

All analyses were carried out in R v. 2.4.01 (R Core Team, 2020), using the 12,498 SNPs located in chromosomes other than chromosome 3, since the supergene evolves independently from the rest of the genome and including the non-recombining supergene haplotypes would not reflect population genetic structure. Genetic variation among individuals was investigated by clustering individuals with DAPC (discriminant analysis of principal components; Jombart et al. 2010) based on allele frequencies, using the "adegenet" package. To best identify the number of genetic clusters, we ran K-means algorithm with the function *find.clusters*, with K ranging from 1 to 15, and selected the number of clusters K with the lowest Bayesian information criteria (BIC). We further inferred population genetic structure with hierarchical *F*-statistics analyses, and obtained 95% confidence intervals (CI) by bootstraping over loci, as implemented in the R package "hierfstat" (Goudet, 2005). The hierarchical levels were regions, populations within regions, and social forms within populations.

We tested for isolation-by-distance (IBD) and isolation-by-environment (IBE) between pairs of populations, excluding two populations in which fewer than three individuals were sampled (Aubenas and Dalaas, **Figure 1, Table S1**). Genetic distances between population pairs were calculated with the function *betas* in "hierfstat". This function uses the Weir and Goudet estimator of F_{ST}, which is robust to unequal sample sizes and appropriate for SNPs markers with allele dosage information (Weir and Goudet, 2017). We calculated geographic great-circle distance, elevation distance, and four multivariate environmental distances, namely temperature, precipitation, soil, and vegetation (**Table S2**). Environmental variables were estimated using raster data from public databases (**Table S2**). They were scaled and centred to account for differences in magnitude (Lichstein, 2007). The environmental distances were then calculated as euclidian dissimilarities, using the R package "ecodist" (Goslee and Urban, 2007). We used separate Mantel tests to examine the association between genetic distance (F_{ST}) and each of the other distances. Next, we ran a multiple regression of distance matrices (MRM, Lichstein 2007) with the genetic distance (F_{ST}) as response variable and geographic distance, elevation distance and the four environmental distances as predictors. These tests were run in "ecodist", and the significance of the associations tested with 1000 permutations.

We investigated if elevation and social organisation affected isolation-by-distance and population differentiation at a local scale, within regions. For that, we used maximum-likelihood population-effects models, which are linear mixed effect regression models (LMER) that include a random term to account for correlation of pairwise distances involving a common population (Clarke, Rothery and Raybould, 2002; Yang, 2004; Van Strien, Keller and Holderegger, 2012). To test if elevation impacts genetic differentiation between populations, we focused on the upper Rhône and Rhine regions, since they comprise populations close to each other and differing strongly in elevation (**Figure 1B, Table S1**). We classified populations in two categories: lowland (< 1000 m, range 473 – 631 m) or highland (> 1000 m, range 1045 – 1455 m). We included F_{ST} between each pair of populations as the response variable in a LMER. The geographic distance, elevation category combination (lowland-lowland, lowland-highland, and highland-highland), and interaction between the two factors were included as fixed explanatory factors, while a random term accounted for correlation of pairwise distances. Additionally, we tested for the effect of elevation on genetic diversity. Genetic diversity (H_s, expected heterozygosity, averaged across loci) within each population was estimated with the "hierfstat" package. We ran a linear model with genetic diversity (H_s) as response variable and elevation category (lowland or highland) as well as region as explanatory variables.

To test if social organisation affects genetic differentiation between populations within regions, we focused on pairs of populations less than 100 km apart. These pairs comprise all populations with three or more individuals, except St. Michel in the Lower Rhône region, which is distant from all other populations (**Figure 1**, **Table S1**). We calculated F_{ST} between individuals belonging to the monogyne (M) or the polygyne (P) social form in each population, resulting in distances corresponding to three social form combinations between each population pair (M-M, P-P, M-P). To control for sample size bias, we calculated F_{ST} with rarefaction and 1000 iterations of resampling, taking as sample size the smallest number of individuals belonging to one social form in one population of this pair. For example, for the F_{ST} between "Derborence-M" (N = 14) and "Finges-M" (N = 22), we resampled 10 individuals from each of the two groups, corresponding to the smallest sample size for one social form in this population pair, which is "Finges-P" (N = 10; **Table S1**). We ran a LMER model with pairwise F_{ST} as a response variable. We included as fixed explanatory factors the geographic distance between two populations, the social form combination (M-M, P-P or M-P), and the interaction between the two factors. We also included a random term to account for correlation of pairwise distances.

We checked for normality, homoscedasticity and absence of overdispersion of residuals in all statistical models by visual inspection of plots, as well as tests implemented in the "DHARMa" package (Hartig, 2018). LMER models were ran with the package "lme4" (Bates *et al.*, 2014). ANOVA type III estimates and *p*-values for the LMER models were obtained using the Kenward Roger approximation with the function *KRmodcomp* in the "pbkrtest" package (Halekoh and Højsgaard, 2014), and with the *drop1* function for the linear model. We performed post-hoc Tukey tests on estimated marginal means, as implemented in the "emmeans" package (Lenth *et al.*, 2020).

RESULTS

The clustering analysis and hierarchical *F*-statistics revealed high genetic differentiation among regions, moderate differentiation among populations within regions, and low differentiation between social forms within populations (**Figure 2**). These analyses were based on 12,498 SNPs spanning the entire genome except chromosome 3, which contains the social supergene. Individuals clustered in three geographically concordant genetic groups, distinguishing individuals from east Switzerland and Austria (Rhine region), central Switzerland (Upper Rhône region), and southeast France (Lower Rhône region; **Figure 1 and 2**). DAPC clustering confirmed that the best number of clusters was K = 3, separating the three regions (**Figure 2**).

In hierarchical *F*-analyses, differences among regions explained most of the genetic variance across landscape ($F_{\text{region-total}} = 0.084$, 95% CI = [0.082,0.087]). Differentiation among populations within regions was moderate ($F_{\text{population-region}} = 0.047$, 95% CI = [0.046,0.049]), while differentiation between social forms within populations was low ($F_{\text{social form-population}} = 0.017$, 95% CI = [0.015, 0.018]), indicating extensive gene flow between social forms. Overall, individuals clustered by regions, but not by population or social form, at SNPs located outside of the social supergene.

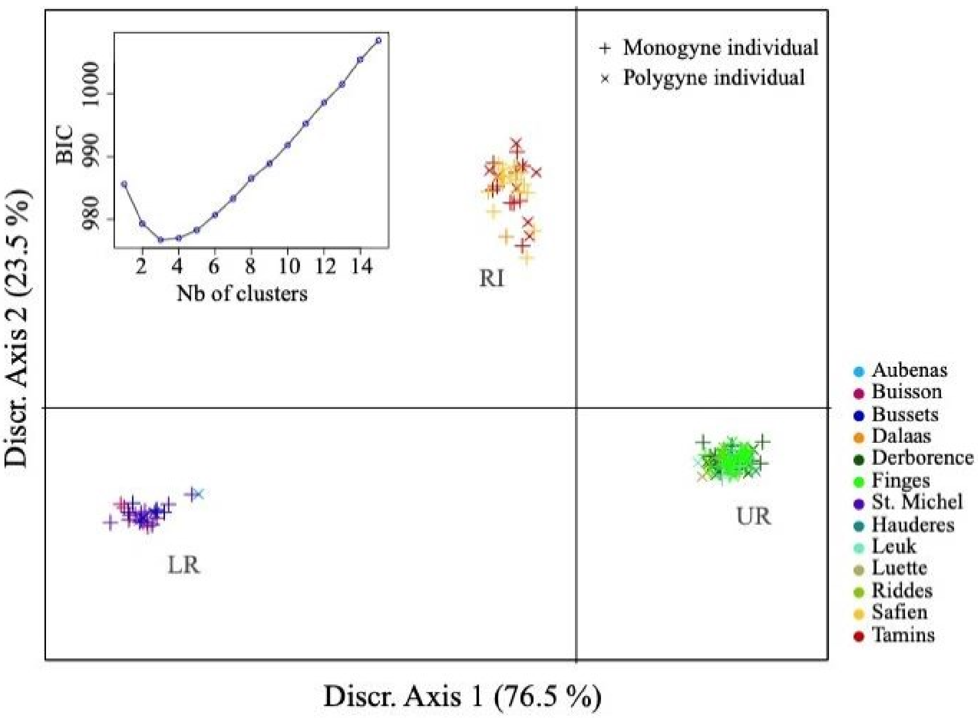

Figure 2. Genetic clustering of individuals by DAPC, with K = 3 clusters. Each cross represents a monogyne (+) or polygyne (x) individual, coloured by population. LR cluster: individuals from Lower Rhône region; UR cluster: individuals from Upper Rhône region; RI cluster: individuals from Rhine region. *Caption:* Bayesian information criteria for K = 1 to 15, used for determining K. Includes individuals from all populations sampled.

Isolation-by-distance and isolation-by-environment

There was a very strong pattern of isolation-by-distance at a range-wide scale (Mantel test: $R = 0.83$, $p < 0.001$; **Figure 3**). Genetic distance between populations was also significantly correlated with temperature distance (Mantel test: $R = 0.54$, $p < 0.001$, **Figure S2**), but not with any of the other environmental or elevation distances. In a multiple regression matrix (MRM) that included geography, elevation, and the four environmental distances, only geographic distance was significantly associated with genetic distance (MRM: $R^2 = 0.72$; geography: $p = 0.001$; elevation: $p = 0.77$;

temperature: $p = 0.1$; precipitation: $p = 0.27$; soil: $p = 0.61$; vegetation: $p = 0.84$). This suggests that the effect of temperature is due to its correlation with geography, and that geography accounts for most of the genome-wide genetic differentiation across the range.

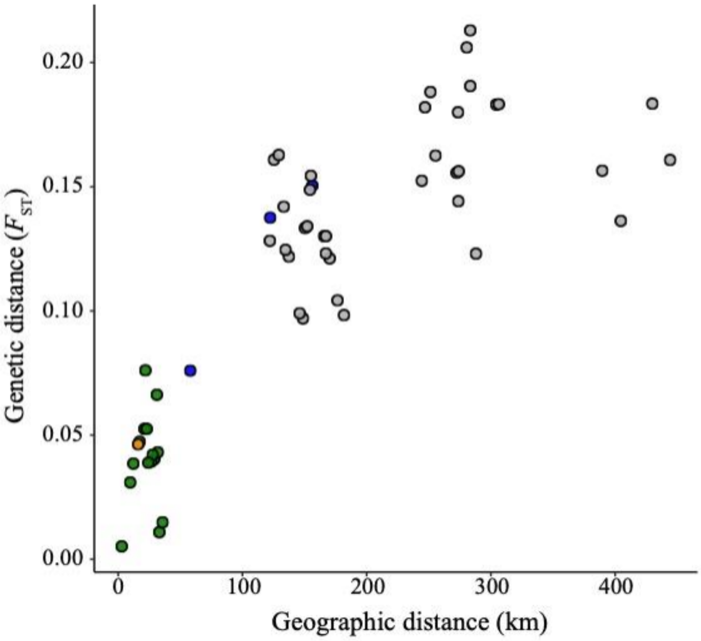

Figure 3. Isolation-by-distance. Relation between genetic distance (F_{ST}) and geographic distance across pairs of populations. Coloured dots are population comparisons within regions (blue: Lower Rhône, green: Upper Rhône, orange: Rhine), and grey dots represent comparisons between populations from different regions. Includes populations BO, BE and SM in Lower Rhône, all populations in Upper Rhône, and T and S populations in Rhine region (Table S1).

Effects of elevation and social organisation on population differentiation

Isolation-by-distance between populations within regions varied with elevation. The association between genetic and geographic distance was stronger among highland populations than among lowland populations (LMER, interaction "geographic distance" and "elevation": $F = 133.97$, df = 2, 3.1, $p < 0.001$; **Figure 4**). Genetic distance was higher between pairs of highland populations than between pairs of lowland populations (LMER, “elevation”: $F = 11.46$, df = 2, 6.3, $p = 0.008$; Tukey post-hoc test, estimate “lowland-lowland” vs “highland-highland” = -0.051, SE = 0.01, df = 6.2, $t = -4.97$, $p = 0.006$), while genetic distances between lowland and highland populations were intermediate (post-hoc tests, estimate “lowland-lowland” vs “highland-lowland” = -0.035, SE = 0.0086, df = 4.8, $t = -4.05$, $p = 0.024$; estimate “highland-highland” vs “highland-lowland” = -0.016, SE = 0.006, df = 7.3, $t = -2.56$, $p = 0.082$). Genetic diversity was higher in lowland populations than in highland populations (LM, estimate “elevation” = -0.007, SE = 0.002, $t = -3.65$, $p = 0.015$; **Figure S3**). Populations in the Rhine region were genetically more diverse than populations in the Upper Rhône region (estimate “region” = -0.022, SE = 0.003, $t = -6.9$, $p < 0.001$).

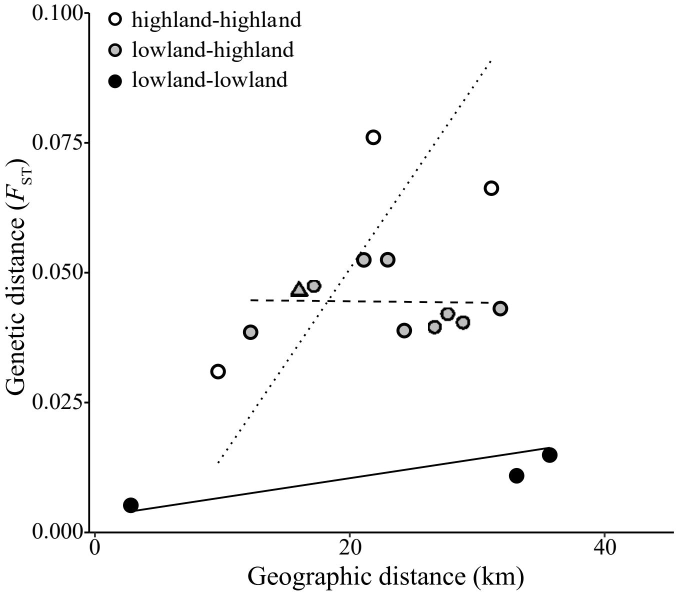

Figure 4. Isolation-by-distance according to elevation. Isolation-by-distance (genetic distance relative to geographic distance) between pairs of lowland populations (black dots, thick line), pairs of highland populations (white, dotted line) or pairs of lowland and highland populations (grey, dashed line) from the same region (circles: Upper Rhône, triangle: Rhine). Lines represent predicted values from the LMER model, which includes as explanatory variables geographic

distance, elevation category, and their interaction. Includes all populations in Upper Rhône region, and populations T and S in Rhine region (Table S1).

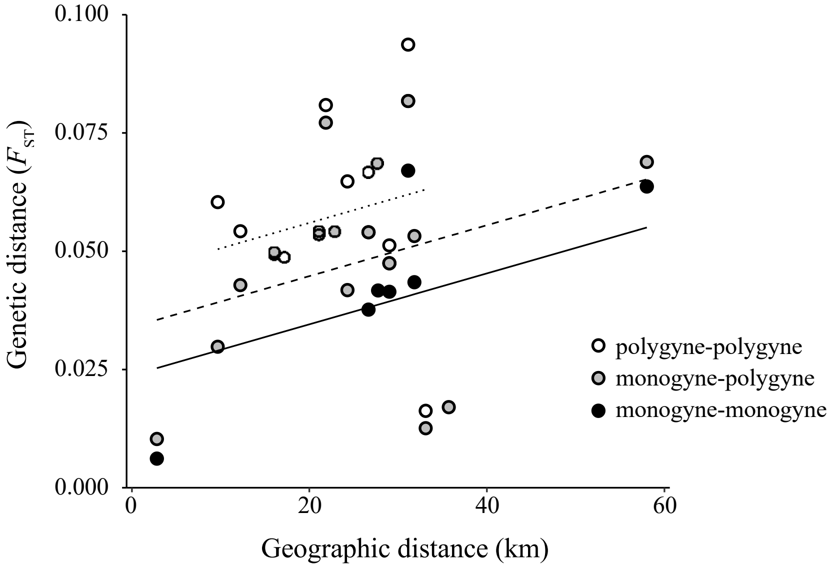

Figure 5. Isolation-by-distance according to social organisation. Isolation-by-distance within and between social forms, across pairs of populations. Each dot represents the distance between a pair of populations, considering only individuals belonging to the monogyne social form (black circles, thick line), only individuals belonging to the polygyne social form (white circles, dotted line), or monogyne individuals in one population and polygyne individuals in the other population (grey circles, dashed line). Lines represent predicted values from the LMER model, which includes as explanatory variables geographic distance and social form. Includes all populations in Upper Rhône region; populations T and S in Rhine region and populations BO and BE in Lower Rhône region (Table S1).

Social organisation affected the genetic distance between populations within regions (LMER, “social form”: F = 17.28, df = 2, 18.6, p < 0.001; **Figure 5**). For the same population pairs, genetic distances were higher when considering individuals belonging to the polygyne social form than when considering individuals belonging to the monogyne social form (Tukey post-hoc test, estimate “P-P” vs “M-M” = -0.021, SE =

0.0037, df = 19, $p < 0.001$). Distances between monogyne individuals of one population and polygyne individuals of the other were intermediate (post-hoc tests, estimate “M-P” vs “P-P” = -0.010, SE = 0.003, df = 18.7, $p = 0.002$; estimate “M-P” vs “M-M” = -0.011, SE = 0.003, df = 18.5, $p = 0.006$). The degree of isolation-by-distance did not differ according to the social form considered (interaction “social form” and “geographic distance”: $F = 1.5$, df = 2, 16.9, $p = 0.26$).

DISCUSSION

Patterns of population genetic structure depend on landscape structural features, but also on species-specific traits that determine how organisms respond to geographical constraints (Baguette *et al.*, 2013). In this population genomics survey, we investigated how topographic and environmental factors affect spatial genetic patterns in a montane ant species, and whether these patterns vary between alternative genetically-determined social forms within this species. We detected a strong pattern of isolation-by-distance at a range-wide scale, but only moderate genetic structure within regions, especially among lowland populations. Moreover, spatial genetic structure differs between social forms.

Such strong pattern of isolation-by-distance between populations (IBD) is uncommon in ants (but see Flucher et al. 2021). In a review of 14 species of the *Formica* genus, Sundström et al. (2005) found IBD at inter-population scale in only one species. IBD is more common at a local scale, that is, between colonies within populations (reviewed in Sundström et al. 2005; Johansson et al. 2018). Differentiation between populations from distinct regions (pairwise $F_{ST} = 0.1 – 0.2$) was high compared to measures over similar geographical scales in other ant species.

High population genetic differentiation and isolation-by-distance in *F. selysi* across the European Alps may be explained by the ecology of this riverine species. Suitable habitats – natural floodplains – tend to be discontinuous along river valleys, which restricts the possibilities of successful colony founding and limits gene flow. Moreover, distant regions might correspond to independent glacial refugia (Schmitt, 2009; Purcell, Pellissier and Chapuisat, 2015; Trettin, Agrawal and Heinze, 2016). Low connectivity of riverine ecosystems between regions and colonization of regions from distinct sources are not mutually exclusive, and can together account for the strong genetic differentiation detected between distant populations, across the species range.

Intraspecific variation in social organisation affected population structure within regions, irrespective of geographical distance. Population differentiation was stronger for the polygyne social form than for the monogyne social form. Previous studies within one large *F. selysi* population found that spatial genetic differentiation above the colony level was similar in the two social forms, at a local scale (Chapuisat, Bocherens and Rosset, 2004; Avril *et al.*, 2019). Our new results reveal that social organisation affects spatial genetic structure at a larger, inter-population spatial scale.

Stronger genetic differentiation between polygyne populations than between monogyne ones has been documented in other ant species (Seppä and Pamilo, 1995; Ross and Shoemaker, 1997; Seppä *et al.*, 2004; Sundström, Seppä and Pamilo, 2005). Yet, in these species, polygyne and monogyne colonies occur in geographically separated populations, so that differences in spatial genetic patterns may be explained by other environmental correlates. In *F. selysi,* monogyne and polygyne colonies co-occur within the same locations. Therefore, the association between social form and spatial genetic structure is due to differences in social organisation, and not to other correlated geographic effects.

Strong genetic differentiation in the polygyne social form could be caused by restricted female dispersal, recurrent founder effect and/or smaller effective population size. Each of these factors tends to reduce genetic diversity and increase F_{ST} (Ross, 2001). In *F. selysi,* monogyne colonies produce numerous females that disperse on the wing, while polygyne colonies produce very few females (Rosset and Chapuisat, 2006; Fontcuberta *et al.*, 2021). Moreover, monogyne females are larger (by 59% in dry weight, and 2% in head width), more fertile and more successful at independent colony founding, while polygyne females are smaller, less fertile, and more philopatric (Rosset and Chapuisat, 2007; Reber, Meunier and Chapuisat, 2010; Avril *et al.*, 2019; De Gasperin *et al.*, 2020; Fontcuberta *et al.*, 2021). Most of the monogyne females (~ 80%) mate with monogyne males and yield monogyne colonies (Fontcuberta *et al.*, 2021). Thus, females that manage to reach distant populations and establish novel colonies independently are much more likely to belong to the monogyne social form.

Monogyne females mated to monogyne males and producing monogyne colonies are probably the main dispersers and founders across populations, resulting in high effective population sizes and high gene flow across populations for the monogyne form. Yet, about 20% of monogyne females mate with polygyne males, and this cross probably yields polygyne colonies (Fontcuberta *et al.*, 2021). Hence, the monogyne and polygyne social forms appear to follow a source–sink dynamics, with asymmetrical gene flow from the monogyne to the polygyne social form (Ross and Shoemaker, 1997; Seppä *et al.*, 2004; Avril *et al.*, 2019). Rare independent colony founding after dispersal flight by polygyne females (Blacher, De Gasperin and Chapuisat, 2021) or by monogyne females mated to polygyne males (Fontcuberta *et al.*, 2021), followed by local budding of polygyne colonies, likely explain the higher inter-population genetic differentiation in the polygyne social form.

Elevation was a major determinant of genetic structure within regions. First, population differentiation was about six times higher among highland populations than among lowland populations (average F_{ST} = 0.058 and 0.01 for highland-highland and lowland-lowland comparisons, respectively). This difference persisted when considering geographic distance: isolation-by-distance was significantly stronger among highland populations than among lowland populations. Sample size was small, and pairwise genetic distances were variable, so further research including more highland and lowland population pairs from additional independent valleys will be needed to confirm this pattern. Second, highland populations were genetically less diverse than lowland populations. Restricted gene flow, founder effects and small effective population sizes at high elevations might explain this pattern (Funk *et al.*, 2005; Polato *et al.*, 2017). High ridges of unsuitable habitat separating alpine valleys probably restrict dispersal between highland populations. Founder effects and small effective population sizes are also expected, given harsh climate conditions characterizing high elevation montane habitat (Catalan, Ninot and Aniz, 2017).

Highland populations are nevertheless connected to nearby lowland populations, as indicated by the lack of effect of elevation distance on genetic differentiation. Gene flow is likely asymmetrical from lowland to highland populations, since strong bidirectional gene flow would homogenise allele frequencies and mask the contrast in connectivity among lowland versus highland populations, respectively. Low genetic diversity and asymmetrical gene flow are consistent with high elevation sites acting as sink populations (Pulliam, 1988; Pannell and Charlesworth, 2000). Overall, our results suggest that gene flow among *F. selysi* populations mostly occurs along lowland valleys, and, to a lesser extent, from low to high elevations along secondary steep valleys.

Such pattern of elevated gene flow in lowland areas of mountain regions has been called the "mountain-valley model" (Funk *et al.*, 2005). It has been found in a variety of montane species, such as chickadees (Branch *et al.*, 2017), frogs (Funk *et al.*, 2005) and mayflies (Polato *et al.*, 2017). Accessible and arable lowland valleys are rare in mountain regions. They are highly exploited for agriculture, industry, roads and urbanisation, which may jeopardise population connectivity of mountain species. Mountains harbor one third of the terrestrial biodiversity in the world (Spehn and Körner, 2005) and are a priority for conservation programs (CBD, 2010; Catalan, Ninot and Aniz, 2017). Yet, protected areas in mountains still fail to cover biodiversity-important sites (Rodríguez-Rodríguez *et al.*, 2011). The fact that many montane species rely on lowland riparian corridors for dispersal highlights the need for conserving not only high elevation ecosystems, but also lowland montane habitats.

LITERATURE CITED

Avril, A. *et al.* (2019) 'Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization', *Molecular Ecology*, 28(6), pp. 1428–1438. doi: 10.1111/mec.14793.

Baguette, M. *et al.* (2013) 'Individual dispersal, landscape connectivity and ecological networks', *Biological Reviews*, 88(2), pp. 310–326. doi: 10.1111/brv.12000.

Ballinger, A., Lake, P. S. and Mac Nally, R. (2007) 'Do terrestrial invertebrates experience floodplains as landscape mosaics? Immediate and longer-term effects of flooding on ant assemblages in a floodplain forest', *Oecologia*, 152(2), pp. 227–238. doi: 10.1007/s00442-006-0648-0.

Bates, D. *et al.* (2014) 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*, 67(1), pp. 1–48.

Blacher, P., De Gasperin, O. and Chapuisat, M. (2021) 'Cooperation by ant queens during colony-founding perpetuates alternative forms of social organization', *Behavioral Ecology and Sociobiology*, 75(165).

Bourke, A. F. G. and Franks, N. R. (1995) *Social evolution in ants*, *Monographs in behavior and ecology*. Edited by J. R. Krebs and T. Clutton-Brock. Princeton, NJ: Princeton University Press.

Branch, C. L. *et al.* (2017) 'Absence of population structure across elevational gradients despite large phenotypic variation in mountain chickadees (*Poecile gambeli*)', *Royal Society Open Science*, 4(3). doi: 10.1098/rsos.170057.

Brelsford, A. *et al.* (2020) 'An ancient and eroded social supergene is widespread across *Formica* ants', *Current Biology*, 30(2), pp. 304–311. doi: 10.1016/j.cub.2019.11.032.

Brelsford, A., Dufresnes, C. and Perrin, N. (2016) 'High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex', *Heredity*, 116, pp. 177–181.

Catalan, J., Ninot, J. M. and Aniz, M. M. (2017) 'High mountain conservation in a changing world', in Catalan, J., Ninot, J. M., and Aniz, M. M. (eds) *Advances in Global Change Research*. Cham, Switzerland: Springer Open. doi: 10.1007/978-3-319-55982-7_16.

Catchen, J. *et al.* (2013) 'Stacks: An analysis tool set for population genomics', *Molecular Ecology*, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.

CBD (2010) *COP 10 Decision X/2: Strategic Plan for Biodiversity 2011–2020*. Available at: http://www.cbd.int/decision/cop/?id=12268 (Accessed: 21 September 2020).

Chapman, R. E. and Bourke, A. F. G. (2001) 'The influence of sociality on the conservation biology of social insects', *Ecology Letters*, 4(6), pp. 650–662. doi: 10.1046/j.1461-0248.2001.00253.x.

Chapuisat, M., Bocherens, S. and Rosset, H. (2004) 'Variable queen number in ant colonies: no impact on queen turnover, inbreeding, and population genetic differentiation in the ant *Formica selysi*.', *Evolution*, 58(5), pp. 1064–1072. doi: 10.1111/j.0014-3820.2004.tb00440.x.

Chapuisat, M., Goudet, J. and Keller, L. (1997) 'Microsatellites reveal high population viscosity and limited dispersal in the ant *Formica paralugubris*', *Evolution*, 51(2), pp. 475–482.

Clarke, R. T., Rothery, P. and Raybould, A. F. (2002) 'Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance', *Journal of Agricultural, Biological, and Environmental Statistics*, 7(3), pp. 361–372. doi: 10.1198/108571102320.

Cushman, S. A. *et al.* (2006) 'Gene flow in complex landscapes: Testing multiple hypotheses with causal modeling', *American Naturalist*, 168(4), pp. 486–499. doi: 10.1086/506976.

Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp. 2156–2158.

Fisher, K. *et al.* (2019) 'Are societies resilient? Challenges faced by social insects in a changing world', *Insectes Sociaux*, 66(1), pp. 5–13. doi: 10.1007/s00040-018-0663-2.

Flucher, S. M. *et al.* (2021) ‘Effect of social structure and introduction history on genetic diversity and differentiation’, *Molecular Ecology*, 30, pp. 2511–2527. doi: 10.1111/mec.15911.

Fontcuberta, A. *et al.* (2021) ‘Disentangling the mechanisms linking dispersal and sociality in supergene- mediated ant social forms’, *Proceedings of the Royal Society B*, 288, p. 20210118. doi: https://doi.org/10.1098/rspb.2021.0118.

Funk, W. C. *et al.* (2005) ‘Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape’, *Molecular Ecology*, 14(2), pp. 483–496. doi: 10.1111/j.1365-294X.2005.02426.x.

Le Galliard, J.-F., Ferrière, R. and Dieckmann, U. (2005) ‘Adaptive evolution of social traits: origin, trajectories, and correlations of altruism and mobility’, *The American Naturalist*, 165(2), pp. 206–224. doi: https://doi.org/10.1086/427090.

De Gasperin, O. *et al.* (2020) ‘Winter is coming: harsh environments limit independent reproduction of cooperative-breeding queens in a socially polymorphic ant’, *Biology Letters*, 16(1), p. 20190730. doi: 10.1098/rsbl.2019.0730.

Glaser, F. (2007) ‘Ants (Hymenotptera, Formicidae) in alpine floodplains- ecological notes and conservation aspects.’, in *Internationales LIFE-symposium, riverine landscapes, restoration—flood protection—conservation. Natur in Tirol 13—Naturkundliche Beiträge der Abteilung Umweltschutz*, pp. 147–163.

Goslee, S. and Urban, D. (2007) ‘The ecodist package for dissimilarity-based analysis of ecological data.’, *Journal of Statistical Software*, 22, pp. 1–19.

Goudet, J. (2005) ‘HIERFSTAT, a package for R to compute and test hierarchical F-statistics’, *Molecular Ecology Notes*, 5(1), pp. 184–186.

Hakala, S. M., Seppä, P. and Helanterä, H. (2019) ‘Evolution of dispersal in ants (Hymenoptera: Formicidae): a review on the dispersal strategies of sessile superorganisms’, *Myrmecological News*, 29, pp. 35–55. doi: 10.25849/myrmecol.news.

Halekoh, U. and Højsgaard, S. (2014) ‘A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R Package pbkrtest.’, *Journal of Statistical Software*, 59(9), pp. 1–30.

Hartig, F. (2018) *DHARMa: Residual diagnostics for hierarchical (multi-level/ mixed) regression models. R package version 0.2.6.*

Hedrick, P. W. and Parker, J. D. (1997) ‘Evolutionary genetics and genetic variation of haplodiploids and X-linked genes.’, *Annual Review of Ecology and Systematics*, 28, pp. 55–83.

Huszár, D. B., Larsen, R. S. and Carlsen, S. (2014) ‘Convergent development of ecological, genetic, and morphological traits in native supercolonies of the red ant *Myrmica rubra*’, *Behavioral Ecology and Sociobiology*, 68, pp. 1859–1870. doi: 10.1007/s00265-014-1795-7.

Johansson, H. *et al.* (2018) ‘Weak population structure in the ant *Formica fusca*’, *PeerJ*, 6, p. e5024. doi: 10.7717/peerj.5024.

Jombart, T. and Ahmed, I. (2011) ‘adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.’, *Bioinformatics*, 27(21), pp. 3070–3071.

Jombart, T., Devillard, S. and Balloux, F. (2010) ‘Discriminant analysis of principal components: a new method for the analysis of genetically structured populations’, *BMC Genetics*, 11, p. 94.

Keller, L. (1993) *Queen number and sociality in insects*. Oxford: Oxford University Press.

Knaus BJ, G. N. (2017) ‘VCFR: a package to manipulate and visualize variant call format data in R’, *Molecular Ecology Resources*, 17(1), pp. 44–53.

Lenth, R. *et al.* (2020) ‘Package “emmeans.” R Package Version 1.5.’

Li, H. and Durbin, R. (2009) ‘Fast and accurate short read alignment with Burrows–Wheeler transform’, *Bioinformatics*, 25(14), pp. 1754–1760.

Lichstein, J. W. (2007) ‘Multiple regression on distance matrices: A multivariate spatial analysis tool’, *Plant Ecology*, 188(2), pp. 117–131. doi: 10.1007/s11258-006-9126-3.

Lude, A., Reich, M. and Plachter, H. (1999) ‘Life strategies of ants in unpredictable floodplain habitats of Alpine rivers (Hymenoptera: Formicidae)’, *Entomologia Generalis*, 24(2), pp. 75–91.

Meirmans, P. G. (2012) ‘The trouble with isolation by distance’, *Molecular Ecology*, 21(12), pp. 2839–2846. doi: 10.1111/j.1365-294X.2012.05578.x.

Pamilo, P. and Crozier, R. H. (1997) ‘Population biology of social insect conservation’, *Memoirs of the Museum of Victoria*, 56, pp. 411–419.

Pannell, J. R. and Charlesworth, B. (2000) ‘Effects of metapopulation processes on measures of genetic diversity.’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 355(1404), pp. 1851–1864. doi: 10.1098/rstb.2000.0740.

Paris, J. R., Stevens, J. R. and Catchen, J. M. (2017) ‘Lost in parameter space: a road map for stacks’, *Methods in Ecology and Evolution*, 8(10), pp. 1360–1373. doi: 10.1111/2041-210X.12775.

Polato, N. R. *et al.* (2017) ‘Genetic diversity and gene flow decline with elevation in montane mayflies’, *Heredity*, 119(2), pp. 107–116. doi: 10.1038/hdy.2017.23.

Pulliam, H. R. (1988) ‘Sources, sinks, and population regulation’, *The American Naturalist*, 132(5), pp. 652–661.

Purcell, J. *et al.* (2014) ‘Convergent genetic architecture underlies social organization in ants’, *Current Biology*, 24, pp. 2728–2732. doi: 10.1016/j.cub.2014.09.071.

Purcell, J. and Chapuisat, M. (2013) ‘Bidirectional shifts in colony queen number in a socially polymorphic ant population', *Evolution*, 67(4), pp. 1169–1180. doi: 10.1111/evo.12010.

Purcell, J., Pellissier, L. and Chapuisat, M. (2015) 'Social structure varies with elevation in an Alpine ant', *Molecular Ecology*, 24(2), pp. 498–507. doi: 10.1111/mec.13042.

R Core Team (2020) 'R: A language and environment for statistical computing.', *R Foundation for Statistical Computing*. Vienna, Austria: http://www.R-project.org. Available at: http://www.r-project.org/.

Reber, A., Meunier, J. and Chapuisat, M. (2010) 'Flexible colony-founding strategies in a socially polymorphic ant', *Animal Behaviour*, 79(2), pp. 467–472. doi: 10.1016/j.anbehav.2009.11.030.

Rodríguez-Rodríguez, D. *et al.* (2011) 'Progress towards international targets for protected area coverage in mountains: a multi-scale assessment', *Biological Conservation*, 144(12), pp. 2978–2983. doi: 10.1016/j.biocon.2011.08.023.

Romiguier, J. *et al.* (2014) 'Population genomics of eusocial insects: the costs of a vertebrate-like effective population size', *Journal of Evolutionary Biology*, 27(3), pp. 593–603. doi: 10.1111/jeb.12331.

Ross, K. G. (2001) 'Molecular ecology of social behaviour: analyses of breeding systems and genetic structure', *Molecular Ecology*, 10(2), pp. 265–284. doi: 10.1046/j.1365-294X.2001.01191.x.

Ross, K. G. and Shoemaker, D. D. (1997) 'Nuclear and mitochondrial genetic structure in two social forms of the fire ant *Solenopsis invicta*: insights into transitions to an alternate social organization', *Heredity*, 78, pp. 590–602. doi: 10.1038/hdy.1997.98.

Rosset, H. and Chapuisat, M. (2006) 'Sex allocation conflict in ants: when the queen rules', *Current Biology*, 16(3), pp. 328–331. doi: 10.1016/j.cub.2005.12.036.

Rosset, H. and Chapuisat, M. (2007) 'Alternative life-histories in a socially polymorphic ant', *Evolutionary Ecology*, 21(5), pp. 577–588. doi: 10.1007/s10682-006-9139-3.

Schmitt, T. (2009) 'Biogeographical and evolutionary importance of the European high mountain systems', *Frontiers in Zoology*, 6(1), pp. 1–10. doi: 10.1186/1742-9994-6-9.

Seifert, B. (2002) 'A taxonomic revision of the Formica cinerea group (Hymenoptera, Formicidae)', *Abh. Ber. Naturkundemus*, 74(2), pp. 245-272.

Seppä, P. *et al.* (2004) 'Coexistence of the social types: genetic population structure in the ant *Formica exsecta*.', *Evolution*, 58(11), pp. 2462–2471. doi: 10.1554/04-295.

Seppä̈ , P. (2008) 'Do ants (Hymenoptera: Formicidae) need conservation and does ant conservation need genetics?', *Myrmecological News*, 11(August), pp. 161–172.

Seppä̈ , P. and Pamilo, P. (1995) 'Gene flow and population viscosity in *Myrmica* ants', *Heredity*, 74(June 1994), pp. 200–209. doi: 10.1038/hdy.1995.28.

Settepani, V., Bechsgaard, J. and Bilde, T. (2016) 'Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection', *Ecology and Evolution*, 6(2), pp. 469–477. doi: 10.1002/ece3.1886.

Spehn, E. M. and Körner, C. (2005) 'A global assessment of mountain biodiversity and its function', in Huber, U.M., Bugmann, H.K.M., Reasoner, M. . (ed.) *Global change and mountain regions: an overview of current knowledge*. Dordrecht, Netherlands: Springer, pp. 393–400.

Van Strien, M. J., Keller, D. and Holderegger, R. (2012) 'A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models', *Molecular Ecology*, 21(16), pp. 4010–4023. doi: 10.1111/j.1365-294X.2012.05687.x.

Sundström, L., Seppä̈ , P. and Pamilo, P. (2005) 'Genetic population structure and dispersal patterns in *Formica* ants — a review', *Annales Zoologici Fennici*, 42(3), pp. 163–177. doi: 10.1007/BF00172934.

Szathmary, E. and Maynard Smith, J. (1995) 'The major evolutionary transitions', *Nature*, 374, pp. 227–231.

Tockner, K. and Stanford, J. A. (2002) ‘Riverine flood plains: present state and future trends’, *Environmental Conservation*, 29(3), pp. 308–330. doi: 10.1017/S037689290200022X.

Trettin, J., Agrawal, S. and Heinze, J. (2016) ‘Phylogeography of social polymorphism in a boreo-montane ant’, *BMC Evolutionary Biology*, 16(1), pp. 1–14. doi: 10.1186/s12862-016-0711-3.

Wang, I. J. and Bradburd, G. S. (2014) ‘Isolation by environment’, *Molecular Ecology*, 23(23), pp. 5649–5662. doi: 10.1111/mec.12938.

Weir, B. S. and Goudet, J. (2017) ‘A unified characterization of population structure’, *Genetics*, 206(August), pp. 2085–2103.

Weyna, A. and Romiguier, J. (2021) ‘Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees’, *bioRxiv*, (2020.04.14.038893, ver.5. peer-reviewed and recommended by PCI Evol Biol.). doi: 10.1101/2020.04.14.038893.

Wilson, E. (1992) ‘The effects of complex social life on evolution and biodiversity’, *Oikos*, 63(1), pp. 13–18.

Wright, S. (1943) ‘Isolation by distance’, *Genetics*, 28, pp. 114–138.

Yang, R. C. (2004) ‘A likelihood-based approach to estimating and testing for isolation by distance’, *Evolution*, 58(8), pp. 1839–1845. doi: 10.1111/j.0014-3820.2004.tb00466.x.

Zahnd, S. *et al.* (2021) ‘Fine-scale habitat heterogeneity favours the coexistence of supergene-controlled social forms in Formica selysi’, *BMC Ecology and Evolution*, 21(24). doi: 10.1186/s12862-020-01742-0.

Data Accessibility:

Single Nucleotide Polymorphism data in VCF format will be deposited in Dryad upon acceptance.

Authors contributions

JP, AF and MC planned the study, JP performed the field sampling and laboratory work, AF analysed the data. AF and MC wrote the manuscript, with contribution from JP. We declare we have no competing interests.

Acknowledgements

We thank Alan Brelsford and Daniel Jeffries for extensive discussions and advice on analyses; Alan Brelsford, Amaury Avril and Dorin Pirogan for help in the field; Livio Ruzzante for help with bioinformatic analyses; Susana Freitas, Herbert C. Wagner and an anonymous reviewer for comments on the manuscript. Special thanks to Jérôme Goudet for advice on population genetics and for customizing the hierfstat package. This work was supported by grant 31003A-173189 from the Swiss National Science Foundation to MC, United States NSF CAREER DEB-1942252 grant to JP and a grant from the Fondation pour l’Université de Lausanne to AF.

APPENDIX

Table S1. Sampling localities, number of colonies sampled (N) and individual genotypes at the social supergene. FR stands for France, CH for Switzerland, and AT for Austria. Lat = Latitude, Lon = Longitude. N = Number of colonies sampled, one worker per colony was genotyped. Dataset 1 was used for Fig. 1, Fig. 2 and Fig. S1; Dataset 2 was used for Fig. 3 and Fig. S2; Dataset 3 was used for Fig. 4 and Fig. S3; and Dataset 4 was used for Fig. 5.

						Supergene genotype			
Sampling locality	**Region**	**Lat**	**Lon**	**Elevation**	**N**	***MM***	***MM***	***MM***	**Dataset**
Aubenas (A)	Lower Rhône (FR)	44.6208	4.4220	300	1	0	0	0	1
Buisson (BO)	Lower Rhône (FR)	44.2846	4.9917	180	8	8	8	8	1,2,4
Bussets (BE)	Lower Rhône (FR)	44.2526	5.7188	644	8	5	5	5	1,2,4
St. Michel (SM)	Lower Rhône (FR)	45.2103	6.4812	710	10	10	10	10	1,2
Finges (F)	Upper Rhône (CH)	45.2103	6.4812	565	32	22	22	22	1,2,3,4
Leuk (LK)	Upper Rhône (CH)	46.3121	7.6443	631	14	12	12	12	1,2,3,4
Riddes (R)	Upper Rhône (CH)	46.1786	7.2221	473	4	1	1	1	1,2,3,4
Luette (LU)	Upper Rhône (CH)	46.1583	7.4446	1045	3	0	0	0	1,2,3,4
Les Haudères (H)	Upper Rhône (CH)	46.0821	7.5047	1455	10	7	7	7	1,2,3,4
Derborence (DE)	Upper Rhône (CH)	46.2883	7.2315	1360	27	14	14	14	1,2,3,4
Tamins (T)	Rhine (CH)	46.8137	9.4100	630	18	10	10	10	1,2,3,4
Safien (SF)	Rhine (CH)	46.6835	9.3191	1305	16	16	16	16	1,2,3,4
Dalaas (DA)	Rhine (AT)	47.1270	9.9791	835	1	1	1	1	1
Total					**152**	**106**	**32**	**14**	

687

Table S2. Environmental variables for isolation by environment analyses. Multivariate "temperature distance" was based on the "Bioclim" variables 1 to 11, "precipitation distance" based on the "Bioclim "variables 12 to 19, "soil distance" based on the five topsoil variables and "vegetation distance" based on two vegetation indexes.

Environmental raster	**Database**	**Resolution**
BIO1 = Annual Mean Temperature	WorldClim v.1.4 (1)	1 km
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))	WorldClim v.1.4 (1)	1 km
BIO3 = Isothermality (BIO2/BIO7) (* 100)	WorldClim v.1.4 (1)	1 km
BIO4 = Temperature Seasonality (standard deviation *100)	WorldClim v.1.4 (1)	1 km
BIO5 = Max Temperature of Warmest Month	WorldClim v.1.4 (1)	1 km
BIO6 = Min Temperature of Coldest Month	WorldClim v.1.4 (1)	1 km
BIO7 = Temperature Annual Range (BIO5-BIO6)	WorldClim v.1.4 (1)	1 km
BIO8 = Mean Temperature of Wettest Quarter	WorldClim v.1.4 (1)	1 km
BIO9 = Mean Temperature of Driest Quarter	WorldClim v.1.4 (1)	1 km
BIO10 = Mean Temperature of Warmest Quarter	WorldClim v.1.4 (1)	1 km
BIO11 = Mean Temperature of Coldest Quarter	WorldClim v.1.4 (1)	1 km
BIO12 = Annual Precipitation	WorldClim v.1.4 (1)	1 km
BIO13 = Precipitation of Wettest Month	WorldClim v.1.4 (1)	1 km
BIO14 = Precipitation of Driest Month	WorldClim v.1.4 (1)	1 km
BIO15 = Precipitation Seasonality (Coefficient of Variation)	WorldClim v.1.4 (1)	1 km
BIO16 = Precipitation of Wettest Quarter	WorldClim v.1.4 (1)	1 km
BIO17 = Precipitation of Driest Quarter	WorldClim v.1.4 (1)	1 km
BIO18 = Precipitation of Warmest Quarter	WorldClim v.1.4 (1)	1 km
BIO19 = Precipitation of Coldest Quarter	WorldClim v.1.4 (1)	1 km
% Bulk density	LUCAS Topsoil (2)	500 m
% Silt Extra	LUCAS Topsoil (2)	500 m
% Coarse fragments extra	LUCAS Topsoil (2)	500 m
% Clay extra	LUCAS Topsoil (2)	500 m
% Sand extra	LUCAS Topsoil (2)	500 m
Normalised Difference Vegetation Index (NDVI) *	MODIS NASA (3)	1 km
Enhanced Vegetation Index (EVI)*	MODIS NASA (3)	1 km
Elevation (SRTM)	SRTM, CIAT (4)	30m

We extracted environmental values from raster data for each population coordinates, using the R package "raster". *We averaged MODIS rasters for the summer months June, July, August of years 2011 to 2013, to match as close as possible the vegetation during reproductive ant season for the year of sampling (2013) and previous years.

References environmental raster data

[1] Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005; 25:1965–1978.

[2] Panagos P, Van Liedekerke M, Jones A, Montanarella L. “European Soil Data Centre: Response to European policy support and public data requirements.” Land use policy. 2012;29(2):329–38.

[3] Didan K. MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006. NASA EOSDIS Land Processes DAAC. 2015.

[4] Jarvis A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled seamless SRTM data- V4, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org.

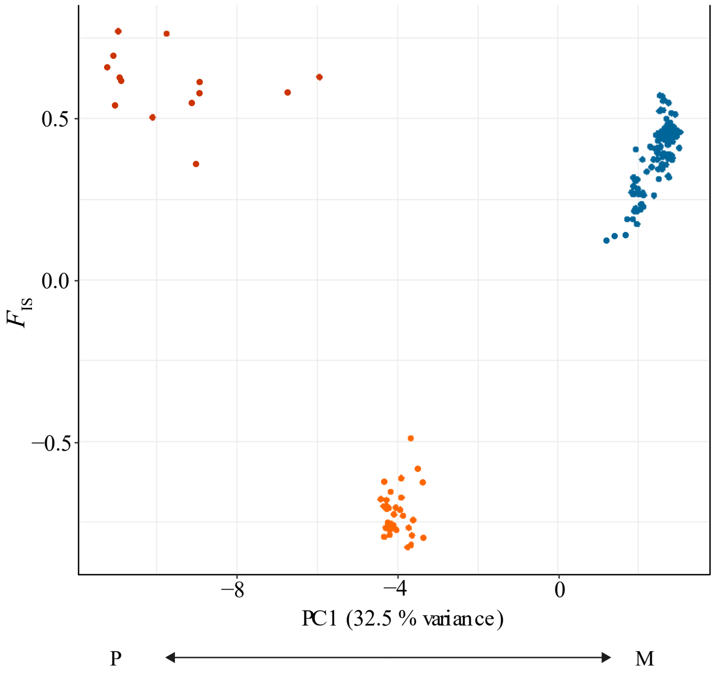

Figure S1. Determination of the social supergene genotype. X-axis displays the first component (PC1) of a PCA of SNPs on chromosome 3, which contains the social supergene. Y-axis represents the F_{IS} per individual: negative values indicate heterozygous genotypes and positive values indicate homozygous genotypes. The three groups correspond to the social supergene genotypes *PP* (red), *PM* (orange) and *MM* (blue), respectively. *PP* and *PM* individuals belong to the polygyne social form, whereas *MM* individuals belong to the monogyne social form (Purcell et al. 2014; Avril et al. 2019).

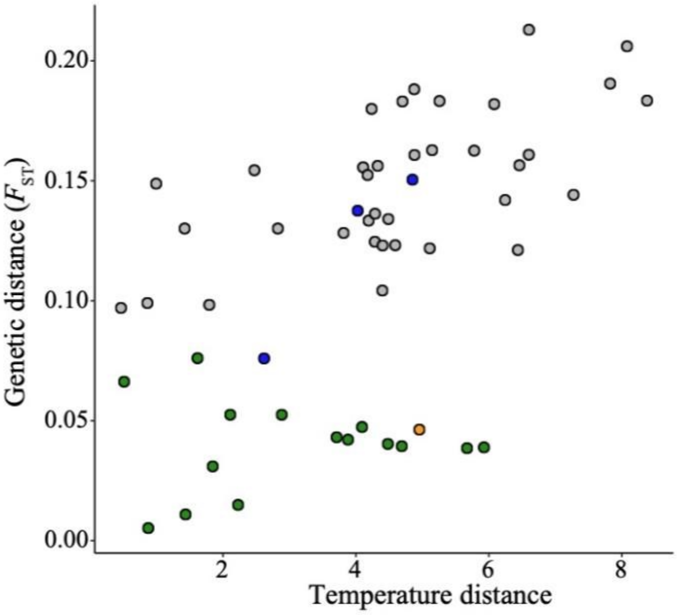

Figure S2. Isolation by environment. Relation between genetic distance (F_{ST}) and multivariate temperature distance. Coloured dots are population pairs within regions (blue: Lower Rhône, green: Upper Rhône, orange: Rhine), and grey dots represent population pairs from different regions. Includes populations BO, BE and SM in Lower Rhône, all populations in Upper Rhône, T and S populations in Rhine region (Table S1).

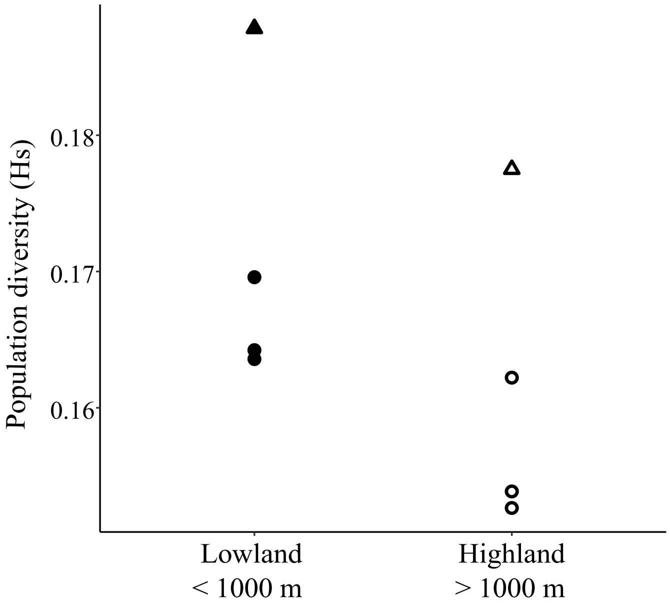

Figure S3. Genetic diversity in lowland (black) and highland (white) populations. Each dot represents a population (circles: Upper Rhône, triangles: Rhine). Includes all populations in Upper Rhône region, and populations T and S in Rhine region (Table S1).