

RESEARCH

Open Access



# Environmental differences explain subtle yet detectable genetic structure in a widespread pollinator

Marcel Glück<sup>1\*</sup> , Julia C. Geue<sup>2</sup>  and Henri A. Thomassen<sup>1</sup> 

## Abstract

**Background:** The environment is a strong driver of genetic structure in many natural populations, yet often neglected in population genetic studies. This may be a particular problem in vagile species, where subtle structure cannot be explained by limitations to dispersal. Consequently, these species might falsely be considered quasi-panmictic and hence potentially mismanaged. A species this might apply to, is the buff-tailed bumble bee (*Bombus terrestris*), an economically important and widespread pollinator, which is considered to be quasi-panmictic at mainland continental scales. Here we aimed to (i) quantify genetic structure in 21+ populations of the buff-tailed bumble bee, sampled throughout two Eastern European countries, and (ii) analyse the degree to which structure is explained by environmental differences, habitat permeability and geographic distance. Using 12 microsatellite loci, we characterised populations of this species with  $F_{st}$  analyses, complemented by discriminant analysis of principal components and Bayesian clustering approaches. We then applied generalized dissimilarity modelling to simultaneously assess the informativeness of geographic distance, habitat permeability and environmental differences among populations in explaining divergence.

**Results:** Genetic structure of the buff-tailed bumble bee quantified by means of  $F_{st}$  was subtle and not detected by Bayesian clustering. Discriminant analysis of principal components suggested insignificant but still noticeable structure that slightly exceeded estimates obtained through  $F_{st}$  analyses. As expected, geographic distance and habitat permeability were not informative in explaining the spatial pattern of genetic divergence. Yet, environmental variables related to temperature, vegetation and topography were highly informative, explaining between 33 and 39% of the genetic variation observed.

**Conclusions:** In contrast to previous studies reporting quasi-panmixia in continental populations of this species, we demonstrated the presence of subtle population structure related to environmental heterogeneity. Environmental data proved to be highly useful in unravelling the drivers of genetic structure in this vagile and opportunistic species. We highlight the potential of including these data to obtain a better understanding of population structure and the processes driving it in species considered to be quasi-panmictic.

**Keywords:** Landscape genetics, Population genetics, Quasi-panmixia, Environmental gradients, Isolation by environment, Microsatellites, Generalized dissimilarity modelling, Eastern Europe, Bumblebee, *Bombus terrestris*

## Background

Detecting genetic population structure and its underlying causes is crucial to better understand basic evolutionary ecological processes and how these are affected by human actions, as well as to improve conservation strategies.

\*Correspondence: marcel.glueck@uni-tuebingen.de

<sup>1</sup> Comparative Zoology, Institute of Evolution and Ecology, Tübingen University, Tübingen, Germany

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Population structure is often inferred using methods that are only based on genetic data, and do not take into account the geographic relationships between populations [1]. These methods perform well when population structure is strong, but may fail to correctly detect weak structure. Considering spatial relationships can help to improve the detectability of weak structure [2–4], but mainly when there are small but distinct genetic breaks in geographic space [5]. Gradually changing genetic population structure is, however, notoriously difficult to detect, which might even be further complicated when structure is subtle (e.g. [5, 6]). As a consequence, species exhibiting such patterns might incorrectly be considered quasi-panmictic.

Population structure results of a balance between gene flow, genetic drift, natural selection, and mutation. When dispersal becomes difficult, gene flow is reduced. This may be the case with increasing distance, or when habitats become difficult to traverse, resulting in patterns of genetic divergence known as isolation by distance (IBD, [7]) and isolation by resistance (IBR, [8]). Moreover, gene flow may also be reduced due to decreased fitness of dispersing individuals that are maladapted to new environmental conditions they encounter [9]. Following a decrease in genetic connectivity, populations may start to diverge by means of genetic drift, eventually resulting in selective processes becoming apparent in per se neutral markers as a pattern of isolation by environment/ecology (IBE, [10, 11]) or isolation by adaptation (IBA, [12]). Approaches that neglect environmental heterogeneity as a driver of population structure may thus be overly simplistic and result in an incomplete picture of the processes that structure natural populations (e.g. [13, 14]).

To this end, the field of landscape genetics [15–17] provides the tools and data to not only analyse genetic information in a spatially explicit context, but to also consider local environmental conditions and those of the habitat matrix between populations in explaining non-random gene flow across the landscape [9]. Thus, it has become possible to study the full range of evolutionary ecological processes driving population divergence and to tease apart their relative importance.

Integrating environmental dissimilarities into the analysis of population structure is particularly promising in species capable of dispersing widely. Typically, these species do not show genetic patterns consistent with IBD or IBR, and exhibit divergence levels close to panmixia [18, 19], which we define as random mating [20] and, consequently, the absence of genetic subdivision. Such a species, which is deemed quasi-panmictic at the subspecies level, is the buff-tailed bumble bee (*Bombus terrestris*). This highly polymorphic pollinator occurs across various environmental gradients [21, 22], and morphological

differences have prompted a division in nine subspecies [23]. Throughout the European continent [21, 24] and Tasmania [22], the species exhibits a remarkable niche breadth and is present across a wide range of habitats that differ strikingly in precipitation patterns, altitude and vegetation. Yet, studies on mainland populations have inferred little to no genetic divergence at various spatial scales, irrespective of the genetic marker applied. Using polymorphic microsatellites, no significant structure was observed at either fine [25] or broad scales [26]. A comparable picture emerged for broad-scale studies using mitochondrial (mt) DNA [27, 28], and single nucleotide polymorphisms [27]. These results are unlikely to be artefacts of the molecular markers used, as populations separated by strong oceanic barriers between islands or continents exhibited significant divergence in microsatellites [26, 28, 29], mtDNA [29] and phenotypic traits [30, 31]. Quasi-panmixia in this species may be the result of its high vagility. For instance, after its introduction to Tasmania, it only took seven years for the species to spread across the entire island [32] and buff-tailed bumble bee queens from introduced populations in Chile have been shown to disperse up to 200 km per year from their hibernation site [33]. Although gene flow in this haplodiploid species, where workers are not reproducing and males are haploid, is largely a function of queen dispersal [34], males have also been reported to disperse up to 10 km [35]. Only large water bodies and strong winds have so far been implied to limit gene flow [28]. Interestingly, however, the potential for subtle population structure and how it might be influenced by environmental heterogeneity remain unelucidated.

This might pose a problem in the face of climate change, where species might show pronounced distribution shifts (e.g. [24, 36]). Currently, the future distribution of species is predicted using known associations between today's species presence and environmental variables. Undoubtedly, this approach holds value as it might allow the delineation of areas that protect both current and future suitable habitats (e.g. [37, 38]). Yet, these models are uninformative about landscape-induced changes in genetic composition that might accompany distribution shifts. Hence, species persistence is only modelled correctly if the influence of the environment on the genetic composition of populations is known. This maxim might also apply to widespread and quasi-panmictic species such as the buff-tailed bumble bee, where climate change-induced range losses might isolate populations genetically that are currently well connected [24].

Here we aimed to quantify population structure and its drivers in the buff-tailed bumble bee across two countries that exhibit pronounced landscape heterogeneity that is readily exploited by this species [21, 24]. Indeed, as

pronounced environmental heterogeneity might increase the likelihood of individuals to experience post-dispersal environmental conditions they are maladapted to, we hypothesised that this species shows genetic structuring consistent with a scenario of IBE [14, 39]. Further, as the buff-tailed bumble bee is highly vagile [22, 35], we expected geographic distance and landscape resistance to play only minor roles in explaining population divergence.

We used both spatially implicit and explicit analyses to identify the most likely drivers of population structure. Deploying 12 highly polymorphic microsatellite loci, we did not detect population differentiation using spatially implicit Bayesian clustering. However, *F<sub>st</sub>* analyses demonstrated subtle yet marginally significant genetic divergence among populations, a finding that was partially corroborated using discriminant analysis of principal components (DAPC, [40]). We further used generalized dissimilarity modelling (GDM, [41]) to simultaneously assess how geographic distance, landscape resistance, and environmental dissimilarities among populations shape genetic divergence. Most notably, only environmental dissimilarities proved to be informative in detecting and unravelling the drivers of genetic structure in this widespread and abundant pollinator.

## Results

### Genotyping and exclusions

A total of 376 out of 385 buff-tailed bumble bees were successfully genotyped at 12 microsatellite loci (Additional file 1: Table S1), with locus-specific error rates ranging between 0 and 5.56%. Using GenAEx [42, 43] and Colony [44], we detected clones and full siblings that might interfere with obtaining accurate estimates of genetic population structure. In total, we excluded 8 clones, 19 full siblings, and 2 individuals inferred as being clones and full siblings at the same time. To prepare the data set for subsequent tests for central population-genetic assumptions, we identified putative males based on multilocus heterozygosity scores. In the end, we excluded 36 putative males of which 34 were homozygous across all 12 loci and 2 across 10 or 11 loci.

### Descriptive statistics

Using Micro-Checker [45], we detected null alleles at all except for two loci (ms66 and ms86; Additional file 1: Table S1) with numbers ranging from one to ten. Signals of stuttering were present at five loci (ms39, 80, 81, 85, and 41) in one to eight sampling sites (Additional file 1: Table S1). We did not detect signals of large allele dropout. Deviations from Hardy–Weinberg equilibrium (HWE) were observed using Genepop on the Web [46], with 8 out of 12 loci showing significant departure from

HWE in one to five populations. We inferred overall significant ( $P=0.027$ ) linkage disequilibrium (LD) for the locus pair ms39–ms80, which was, however, not supported by population pairwise analyses, where significant LD was detected only in the populations Drăgusani and Valea Hotarului. As null alleles, stuttering, and deviations from both HWE and LD were not consistent across either loci or populations, we retained all loci in the data set.

Observed heterozygosity ( $H_O$ ) ranged from 0.46 to 0.60 in a data set containing diploid females only (diploid data set; ‘dpds’) (Additional file 1: Table S2). After correcting for unequal sample sizes by rarefaction for seven individuals, the number of alleles ranged from 2.70 to 3.09 for ‘dpds’ and from 2.26 to 3.08 for a data set comprising haploid males and diploid females (mixed-ploidy data set; ‘mpds’) (Additional file 1: Table S2). Overall population divergence measured as *F<sub>st</sub>* was low in both data sets; 0.006 ( $P=0.02$ ) and 0.041 (upper 95% CI: 0.047) for ‘dpds’ and ‘mpds’, respectively. *F<sub>st</sub>* values among pairs of sampling sites ranged from 0 to 0.07 (‘dpds’, Additional file 1: Table S3) and from 0.01 to 0.12 (‘mpds’, Additional file 1: Table S4). Following false discovery rate (FDR, [47]) correction, five pairwise comparisons remained marginally significant ( $P=0.053–0.059$ ) in ‘dpds’ with *F<sub>st</sub>* values between 0.046 and 0.065. None of the ‘mpds’ *F<sub>st</sub>* values surpassed its corresponding upper 95% confidence interval. Pairwise and global *F<sub>st</sub>* estimates in ‘dpds’ calculated with polysat [48] were higher (global = 0.039, pairwise: 0.01–0.11; Additional file 1: Table S5) than corresponding estimates computed in GenAEx. However, neither global nor pairwise estimates exceeded their corresponding 95% confidence intervals, indicating still non-significant population differentiation. This observed discrepancy between *F<sub>st</sub>* estimates calculated with GenAEx and polysat seems to be an artefact of the different computational approaches. polysat calculates Wright’s *F<sub>st</sub>* [49], based on allele frequencies derived with the ‘simpleFreq’ function that might underestimate common and overestimate rare allele frequencies (polysat manual, [48]). GenAEx, however, estimates *F<sub>st</sub>* using an approach that among others corrects for finite sample sizes and the number of populations sampled [50].

### Genetic clustering

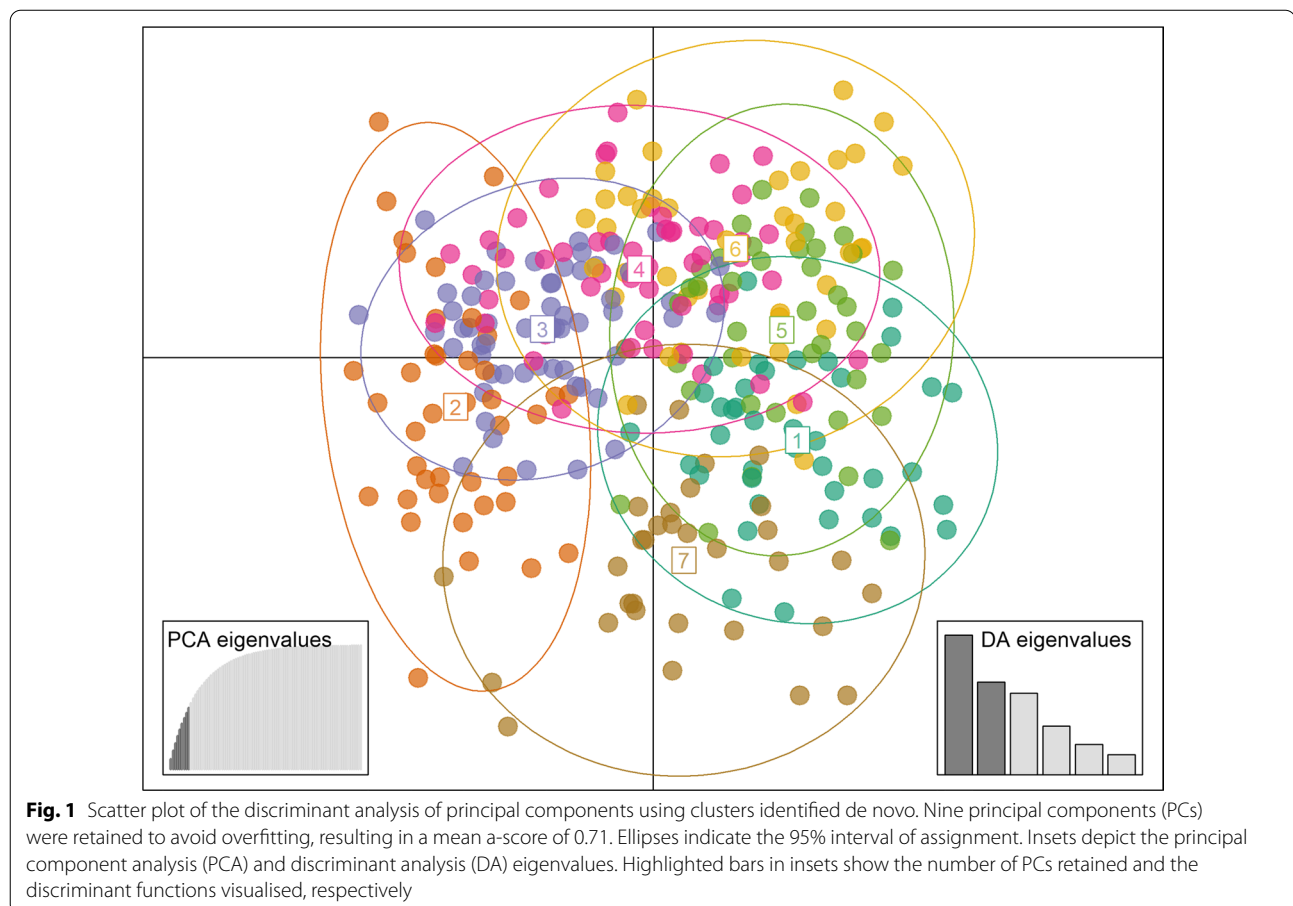
To complement *F<sub>st</sub>* analyses in inferring population divergence, we used the Bayesian clustering algorithm implemented in the program Structure [1]. Structure reconstructs patterns of genetic differentiation by assigning individuals to a specified number of clusters ( $K$ ). Visual interpretation of the results computed using the ‘Correlated Allele Frequency model’ [51] with population IDs as priors suggested the absence of clear genetic

structure ( $K=1$ ), irrespective of whether or not admixture was assumed. This pattern was consistently inferred across all independent runs for each  $K$ , ranging from 2 to 25 (Additional file 2: Figs. S1, S2). Excluding populations with fewer than 10 individuals ('pop10') for the 'Admixture' runs did not change this conclusion, regardless of whether individuals were initialised from their respective populations or not (results available on Dryad).

Because the performance of Bayesian clustering depends on how well the genetic data conform to explicit population genetic models [40] and to further assess the potential for weak genetic structure, we performed discriminant analysis of principal components (DAPC). In a run with populations as a priori groups, results obtained using DAPC were in line with the ones obtained using Structure. Following a-score optimisation to avoid overfitting, we retained 27 principal components (PCs) (Additional file 2: Fig. S3), resulting in a mean a-score of 0.04. Clusters produced overlapped significantly (Additional file 2: Fig. S4), suggesting the absence of clear genetic structuring when populations were used as prior. However, this picture changed slightly when clusters were inferred de novo; from 700 independent cluster

runs performed, 648 converged successfully and deemed  $K=7$  as the most meaningful number of clusters (Additional file 2: Fig. S5). Following a-score optimisation, nine PCs were retained (Additional file 2: Figure S3), resulting in a mean a-score of 0.71. Generally, even though clusters still showed noticeable overlap (Fig. 1), structuring was more pronounced than in the analysis that had used populations as priors. Individuals within clusters originated from many populations throughout Romania and Bulgaria (Additional file 1: Table S6), suggesting admixture on a large spatial scale. Still, as DAPC with populations as prior recreate  $F_{st}$ -derived patterns of genetic structure [52], we concluded that clusters inferred de novo detected slight but unique genetic subdivision that had not been inferred by  $F_{st}$  analyses.

Despite the potential occurrence of two subspecies (*B. t. dalmatinus* and *B. t. terrestris*) native to Romania and Bulgaria [53], it seems unlikely that *B. t. terrestris* was present among our samples, as its known range is limited to the western border regions. Additionally, we would expect its presence to result in pronounced clustering, something we did not observe here. In summary, while Bayesian clustering suggested that the buff-tailed bumble



bee resembles a quasi-panmictic population across Romania and Bulgaria, DAPC using clusters identified de novo suggested weak but still noticeable genetic subdivision undetected by  $F_{st}$  analyses.

### Landscape genetic analyses

We estimated the relative effects of potential drivers of structure using generalized dissimilarity modelling (GDM). Local habitat conditions were characterised using a set of topographical, climate and vegetation variables, and resistance distances were based on a species distribution model (SDM) generated previously [21]. GDMs explained ~33.5 and 39.2% of the divergence observed for 'dpds' and 'mpds', respectively (Table 1). We compared these values to those of 1000 random models. Although the random models performed surprisingly well and explained ~24.3 (lower CI–upper CI: 23.8–24.8) and 31.8% (31.2–32.3) of the divergence for 'dpds' and 'mpds', respectively, they were outperformed by the full models. Environmental dissimilarities contributed most to explaining differences in genetic composition. In contrast, geographic distance and the SDM-derived resistance distances performed poorly; they were not retained in the full models and only explained 0.07 and 0.00% of the total variation for 'dpds', and 0.00 and 0.01% for 'mpds' when analysed in isolation.

The environmental variables retained included temperature, precipitation, topography, measures of surface moisture, and vegetation density. Response curves (splines) visualise how the retained environmental variables contributed to the observed genetic differences and which variables were most informative in explaining the spatial pattern of divergence (Additional file 2: Figs. S6, S7). The splines produced for predictors deemed

significant were highly variable, ranging from nearly no allelic turnover across the respective environmental gradient to rapid turnover at particular gradient positions (Fig. 2; Additional file 2: Figs. S6, S7). The most important variables in both 'dpds' and 'mpds' were slope, mean Leaf Area Index (LAI) and mean temperature of the coldest quarter (Bio 11). Even among the most informative variables, pronounced differences in the splines' shapes were apparent (Fig. 2); following an initial plateau, Bio 11 and Slope produced splines that indicated strong positive allelic turnover starting at a temperature of 9 °C and a slope of 2°, respectively. These responses differed markedly from the one shown by LAI where initially pronounced allelic turnover levelled off at a mean LAI of 13. Seasonality in surface moisture, isothermality, and mean temperature of the wettest quarter contributed as well, but their importance varied between the two data sets. Congruent to the negligible percentage of genetic differentiation explained, allelic turnover caused by geographic distance was quasi non-apparent. The map of the predicted allelic turnover across Romania and Bulgaria (Fig. 3(b)) shows that it is most pronounced along elevation gradients, such as between the Danube Delta and the Carpathians in Romania and the Balkan Mountains in Bulgaria.

### Discussion

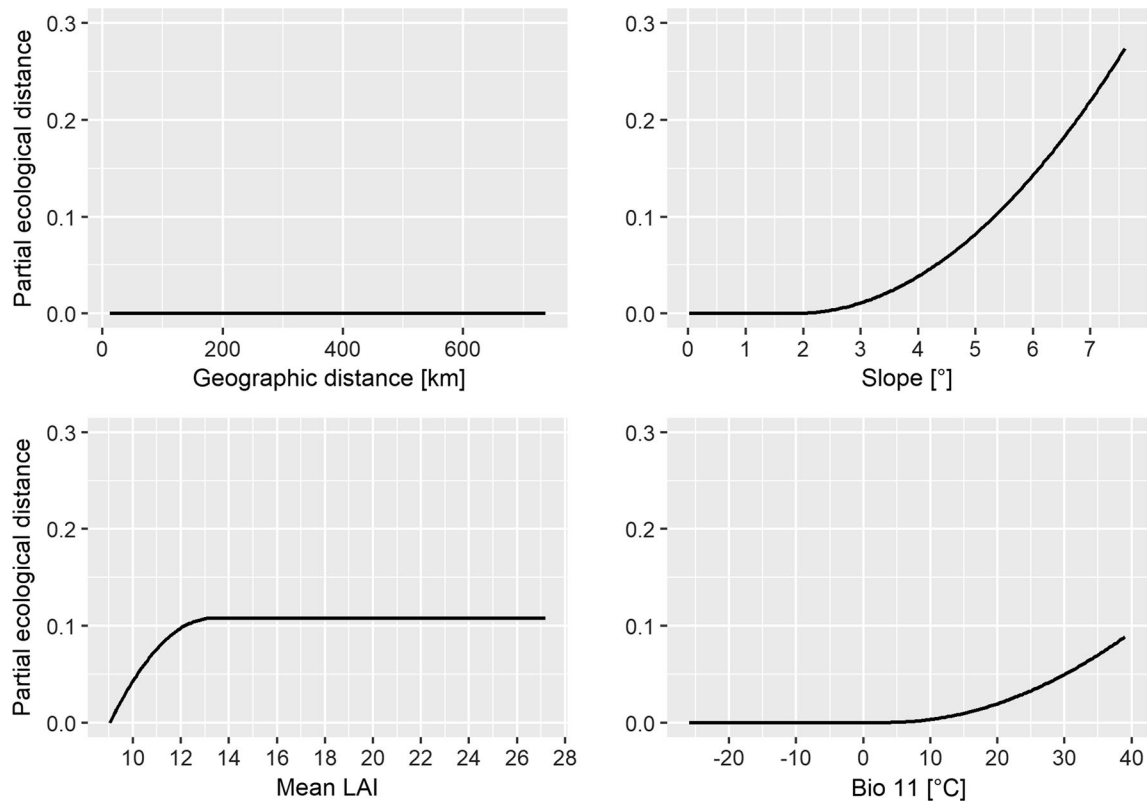
Incorporating environmental data in population genetic studies might help to explain subtle population structure in vagile species, whose dispersal is usually not constrained by geographic distance or the habitat through which dispersal takes place. To assess the informativeness of environmental data, we quantified population structure in the widespread and abundant buff-tailed bumble bee (*Bombus terrestris*). This species is usually considered to show subtle to no structure at mainland continental scales. Using genetic data only, we inferred weak structure across Romania and Bulgaria, two countries that harbour pronounced landscape heterogeneity. Harnessing a landscape genetics approach, we related genetic divergence to this heterogeneity in the environment. Geographic distance was not informative, nor were resistance distances derived from a species distribution model that quantifies resistance to dispersal. Population structure in buff-tailed bumble bees thus follows a pattern of isolation by environment, where differences in habitat conditions reduce the fitness of dispersing individuals [9]. Thus, environmental selection against dispersers contributes to a disruption of genetic connectivity that gives rise to divergence in neutral markers through genetic drift.

The subtle level of divergence observed among Romanian and Bulgarian populations is in agreement with a previous study demonstrating weak but significant

**Table 1** Percentage of variance explained by the generalized dissimilarity models for the data sets encompassing diploid ('dpds'), and both haploid and diploid ('mpds') individuals

|          | dpds  | mpds  |
|----------|-------|-------|
| Full     | 33.48 | 39.18 |
| Env only | 33.48 | 39.18 |
| Geo only | 0.07  | 0.00  |
| Res only | 0.00  | 0.01  |
| Random   | 24.30 | 31.78 |
| Lower CI | 23.77 | 31.24 |
| Upper CI | 24.82 | 32.32 |

Models were run with five different input data sets. (1) Full: geographic distance, resistance distance and environmental variables were included; (2) Env only, (3) Geo only, (4) Res only: contained environmental variables, geographic distance, and resistance distance only, respectively. (5) Random: mean of 1000 models with randomly generated environmental variables. Lower/Upper CI: lower/upper 95% confidence interval of the random models



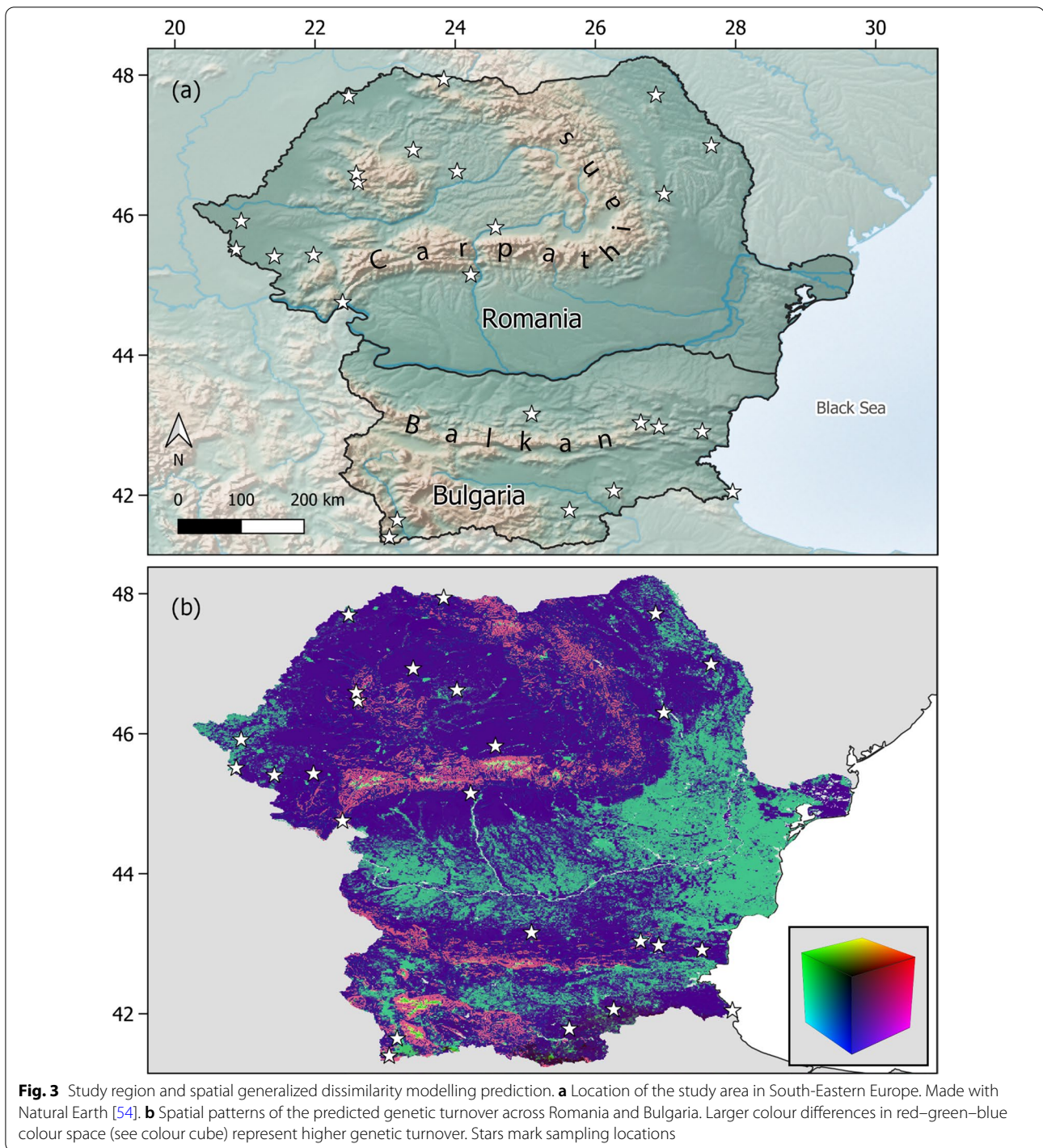
**Fig. 2** Allelic turnover across each of the environmental gradients depicted, as derived from a generalized dissimilarity model performed using environmental variables, geographic and resistance distance as predictors. Alongside geographic distance, which can be considered the baseline model of genetic structure, the splines of the three most influential variables in shaping turnover are shown; LAI: Leaf Area Index, Bio 11: mean temperature of the coldest quarter

genetic structure among buff-tailed bumble bee populations in continental Europe [28]. Interestingly, some of the genetic divergence observed in the aforementioned study might simply be attributed to the inclusion of three subspecies [53]. Nonetheless, contrasting with previous work [25–27], these studies highlight the possibility of weak but significant genetic structure in this species.

More generally, low population structure has not only been detected in *B. terrestris*, but seems fairly common among *Bombus* species, including but not limited to *B. lapidarius* [55], *B. hortorum*, *B. ruderarius*, *B. soroeensis* [56] and *B. ignitus* [57]. Quasi-panmixia in these species is likely driven by extensive gene flow [22, 26, 35], strong enough to override most of the divergence caused by genetic drift and divergent natural selection. This hypothesis is in line with the absence of detectable levels of isolation by distance in our study as well as mainland populations of many other *Bombus* species, such as *B. pascuorum* [58], *B. vosnesenskii* [59], *B. lapidarius*, *B. hortorum*, *B. ruderatus* [60] and *B. flavifrons* [61]. Dispersal, and hence gene flow, may not follow a straight line, but rather a path of least resistance through suitable

habitat. Resistance distances are thus often considered to be a better proxy of between-population dispersal than geographic distance. Nevertheless, our analyses suggest that gene flow is not constrained by variation in habitat permeability either, a finding potentially caused by the absence of strong oceanic barriers in the study area [62–64].

Subtle genetic structure might also result from the buff-tailed bumble bee's generalist foraging behaviour and the presence of many workers [65–67]. The latter might increase the probability of detecting forage, while the former might increase the likelihood of it being deemed exploitable. Hence, both traits may allow this species to efficiently exploit natural and semi-natural habitats that might be devoid of related but more specialised species [68, 69]. Freed from interspecific competition, establishment success of colonies might be increased [70]. The species' broad niche might thus translate to a more continuous distribution of nests across the landscape, with ample opportunity for gene flow among populations. However, with increasing nest densities, the benefit of a more continuous distribution might increasingly



be cancelled out by strong density-induced intraspecific competition for forage [71, 72]. This competition may reduce overall nest performance and thus reproductive output [73, 74]. Hence, compared to related species, the lower nest densities observed in *B. terrestris* [60, 75, 76] might even benefit gene flow, as a relaxation

of competition might allow more nests to contribute to genetic connectivity, resulting in overall weak population structure.

The subtle divergence we observed was best explained by environmental dissimilarities, in particular in the mean temperature of the coldest quarter, Leaf Area

Index (LAI) and slope. The importance of temperature did not come as a surprise, as it is a predictor of the distribution of *Bombus* species [21, 77], and known to govern the emergence time of queens from hibernation [78]. Hence, queens hibernating in warmer areas might emerge earlier than those in colder areas. In fact, asynchronous emergence times may translate to phenological mismatches and thus reproductive isolation between early and late colonies. Even though laboratory experiments with *B. perplexus* and *B. lucorum* did not support this hypothesis [79, 80], the complex environment these species experience throughout their life cycles is unlikely to be fully reproduced in the laboratory [81]. Indeed, emergence patterns of sexuals in natural populations of several species, including *B. flavifrons* and *B. lucorum/terrestris*, differed strongly [82, 83], highlighting the potential for asynchronous emergence to give rise to reproductive isolation. As buff-tailed bumble bee gynes usually only mate once [84, 85], the asynchronous emergence of sexuals allows early emerging males to effectively monopolise queens, thus promoting population divergence.

Even though differences in local temperatures might thus facilitate genetic divergence, the influence of this environmental variable on the species' life cycle is more nuanced. Facultative endothermy [86] and the ability for collective thermoregulation [87] allow nests to survive even under low temperatures [88], enabling great phenological flexibility [89]. Populations may even change from univoltinism to bivoltinism in regions hitherto considered incompatible with the latter [74]. Notably, this change in phenology seems to spread northwards into colder regions [74], a pattern coinciding with northward range shifts in this species [90]. As doubling the number of reproductive cycles per year effectively doubles the number of potential admixture events, bivoltinism might contribute to genetic homogeneity. Nevertheless, generally cold winters in Romania and Bulgaria [91, 92], and the absence of sufficient forage during the cold season might limit bivoltinism to a few inland and coastal regions in both countries [92].

Another important variable retained in the final models was mean LAI. The mechanism through which this measure of greenness is associated with genetic structure remains unclear. Yet, mean LAI was highly correlated with percent tree cover ( $r=0.83$ , Additional file 1: Table S7), and although forests do not seem to limit bumble bee movement [93, 94], buff-tailed bumble bee queens prefer open habitat for nesting [95]. Forests may thus reduce genetic connectivity by constraining the amount of available nesting habitat [62, 96]. Moreover, assuming a negative effect of woodland on the range expansion of this species allowed to accurately model its invasion

pattern in Japan [97], suggesting that queen-borne range expansion might indeed be limited by forests.

The last variable retained in our analyses was the slope of the terrain. Slope has previously been suggested to be an important determinant of suitable hibernation habitat [78, 98]. However, hibernation locations might only be of secondary importance to gene flow. Instead, the presence of suitable nesting habitat where populations are established is more likely to influence gene flow and hence genetic divergence among populations. To this end, slope might only shape genetic turnover if hibernation locations coincide with the locations where colonies are established (i.e. when queen dispersal after hibernation is strongly limited [99]). Given that buff-tailed bumble bee queens are instead highly vagile [22, 100], the mechanism of how slope governs genetic divergence remains unclear.

Although we included a large set of environmental variables, about 60% of the genetic variation remained unexplained. As we aimed to unravel the factors shaping and maintaining large-scale genetic structure, divergence explained by small-scale processes, temporal fluctuations or colony-intrinsic traits were beyond the scope of this work. Genetic divergence might also result from demographic processes, such as bottlenecks [57], which we could not cover here. Future studies might investigate the influence of the abundance and the spatial arrangement of plant species producing nectar or pollen in high quantities [81, 101, 102]. Finally, habitat alterations such as intensified farming practices might also structure populations, in particular through a synergy with natural stressors [103, 104].

## Conclusions

Seemingly panmictic populations might exhibit subtle genetic structure that can only be understood when considering the environment as a potential driver of divergence. We inferred subtle differentiation in a widespread and abundant quasi-panmictic pollinator that was not explained by geographic distance or variation in habitat permeability. Yet, using a suite of environmental variables, we showed that environmental dissimilarities are informative in explaining the observed spatial patterns of genetic structure in a highly vagile species.

## Methods

### Study species and study area

The buff-tailed bumble bee (*Bombus terrestris*) is a widespread and abundant pollinator species. Its native distribution covers much of the Palearctic realm, including Europe, North Africa, and the British and most Mediterranean and Atlantic islands [26, 29]. Its polylectic foraging and high pollination efficiency in various crops [23, 105] has rendered this species a prime candidate



for pollination in the greenhouse industry where it is deployed in high numbers [106]. Upon introduction outside its native range, this species has spread rapidly and established itself in many countries including Chile, Japan and New Zealand [22, 105, 107–109], which highlights its potential to rapidly adapt to novel environments.

We conducted this study in Romania and Bulgaria, neighbouring countries in South-Eastern Europe that exhibit high heterogeneity in topography, climate, and land use. On a large scale, extensive mountainous areas with peaks up to 2500 m shape the face of both countries. In Romania, the Carpathian Mountains predominate [110], whereas in Bulgaria the landscape is structured by alternating bands of high and low terrain that stretch from east to west across the country [111]. The topography gives rise to various climatic zones ranging from alpine and subarctic to humid subtropical [110, 112]. The landscape is a mosaic of natural areas such as plains, open woodland, and extensive forests, as well as inhabited and in- and extensively used agricultural areas. This pronounced environmental heterogeneity provides an ideal study ground to identify subtle population structure and its drivers in a species believed to be quasi-panmictic.

#### Field sampling

Over five consecutive years, from 2013 to 2017, we obtained 385 individuals from 25 locations across Romania and Bulgaria (Fig. 3(a)). Sampling sites (Additional file 1: Table S8) were at least 20 km apart to avoid overlapping foraging ranges [93, 113, 114]. Locations spanned a wide range of habitat conditions, encompassing both natural and semi-natural habitats, as well as extensive environmental gradients with respect to climate, vegetation and altitude. Sites were visited only once by a small team of 2–3 people for approximately 1.5 h each. Individuals regardless of sex were caught using insect nets and sacrificed in a jar with ethyl acetate [115]. Subsequently, specimens were transferred to individual tubes containing 96% ethanol and stored at  $-20^{\circ}\text{C}$  after returning to Tübingen University.

#### Marker choice, DNA extraction, and genotyping

Even though single nucleotide polymorphisms are increasingly being used to address a plethora of evolutionary and ecological questions [116], microsatellites remain a powerful yet time and cost-efficient way of detecting population structure [117]. Indeed, the markers' high mutation rates might render it particularly suited to infer the influence of recent ecological events on shaping spatial patterns of genetic diversity [118–120]. Additionally, genetic connectivity might be disrupted through reduced fitness of dispersers [9, 121]. Even though increasing genetic divergence is then caused

by genetic drift, it is environmental differences that drive this differentiation in the first place. To this end, correlations between environmental variables and genetic divergence in neutral genetic markers, known as patterns of IBE/IBA, can be used to uncover hidden selective variation. Hence, microsatellites can be used to study the influence of the environment on shaping population structure in a widespread species such as *B. terrestris*. We acknowledge that this marker with its high allelic diversity within populations might depress  $F_{st}$  estimates [122]. Size homoplasy, where alleles are identical in state but not by descent might decrease  $F_{st}$  estimates even more but especially when dispersal among populations is low [123], which is unlikely to be the case in the highly vagile buff-tailed bumble bee. However, and regardless of the strength of the latter, both limitations render  $F_{st}$  values a conservative estimate of the true genetic divergence.

From each individual, we extracted DNA from one or two legs using the DNeasy Blood and Tissue Kit and QIAamp DNA Micro Kit (Qiagen, Hilden, Germany). We followed the manufacturer's protocols except for adding 20  $\mu\text{l}$  of 1 M dithiothreitol solution to each sample, which facilitates the extraction of DNA from chitinous materials [124]. Individuals of *B. terrestris*, specifically workers [125], can be difficult to distinguish morphologically in the field from another closely related bumble bee species, the white-tailed bumble bee (*Bombus lucorum*) [126]. We therefore confirmed species identity using a 1043 bp long fragment of the cytochrome c oxidase subunit I (CO1) gene [21].

We then amplified 12 previously developed [127] microsatellite loci in 3 multiplex reactions (PM1–PM3, Additional file 1: Table S1). PCR amplification was run in a total volume of 10  $\mu\text{l}$  consisting of 5  $\mu\text{l}$  PCR master mix (Qiagen), 2.1  $\mu\text{l}$  HPLC  $\text{H}_2\text{O}$ , 0.4  $\mu\text{l}$  bovine serum albumin (10 mg/ml), 1  $\mu\text{l}$  primer solution (100  $\mu\text{M}$ , Applied Biosystems) and 1.5  $\mu\text{l}$  sample DNA. Samples were initially denatured at  $95^{\circ}\text{C}$  for 15 min, followed by 25 cycles of denaturation ( $94^{\circ}\text{C}$ , 30 s), annealing (PM1:  $56^{\circ}\text{C}$ , PM2/3:  $60^{\circ}\text{C}$ , both for 90 s) and extension ( $72^{\circ}\text{C}$ , 60 s). An additional 20 cycles were run using the following settings: denaturation ( $94^{\circ}\text{C}$ , 30 s), annealing ( $44^{\circ}\text{C}$ , 90 s) and extension ( $72^{\circ}\text{C}$ , 60 s). PCR products were visualised on an ABI3730XL capillary DNA sequencer (Applied Biosystems) using a GeneScan 500 LIZ size standard (Applied Biosystems) at MacroGen Europe (The Netherlands). Results were analysed using GeneMarker v.2.4.0 (SoftGenetics, State College, PA). Samples that had not amplified successfully or for which scoring had not yielded conclusive results were re-amplified and re-scored. Individuals that repeatedly failed to amplify or yielded inconclusive results for the second time were excluded. The presence of genotyping errors was assessed

by re-amplifying and re-scoring 36 randomly selected samples, representing approximately 9.5% of all individuals. Scoring results were compared between the first and second run and the mean error rate for each locus was calculated in Microsoft Excel.

## Data analysis

### Identification and removal of clones, sibling workers and drones

As the presence of clones and full siblings is known to distort estimates of population structure, we first identified clones using the ‘Multilocus Matches’ analysis in GenAEx v.6.503 [42, 43]. Second, full siblings and additional clones were detected using the maximum likelihood approach implemented in Colony v.2.0.6.5 [44], which has been deemed the most reliable method for assigning sibship in bumble bees [34]. In brief, we assumed a polygamous mating system to allow Colony to infer the relationship among all individuals entered as offspring, as well as inbreeding [128], the presence of clones, and dioecious reproduction with haplodiploidy. Two runs, differing in their seed values, were conducted with medium length using the full likelihood method with medium precision. Dropout rate was set to 0.001 and locus-specific mean error rates ranged from 0 to 5.56%. Individuals were considered clones or full siblings when they were inferred in both runs with probabilities >0.8 and originated from the same population. For each inferred clone or full sibling pair, one randomly selected individual was retained, resulting in the so-called ‘complete data set’ (‘cpds’, Additional file 1: Table S8). As subtle genetic distances might not be inferred accurately for small populations [129], we excluded the populations ‘Billed’ and ‘Coastra’, resulting in a ‘mixed-ploidy data set’ (‘mpds’, Additional file 1: Table S8). Additionally, because haplodiploidy in *Bombus* (i.e. haploid males, diploid females) deflates measures of heterozygosity, we identified and excluded putative males in ‘mpds’ based on observed multilocus genotypes. More specifically, individuals were considered males if they were heterozygous for a maximum of two loci, resulting in a data set encompassing diploid individuals only (‘dpds’, Additional file 1: Table S8) from 21 populations throughout Romania and Bulgaria.

### Population genetic analyses

After excluding putative clones, full siblings and haploid males, we tested for null alleles, stuttering and large allele dropout in Micro-Checker v.2.2.3 [45], applying 3000 randomisations and a Bonferroni correction while omitting missing data. Null alleles are those that do not reliably amplify in PCR, usually due to non-ideal conditions and/or mutations at primer-binding regions and

result in heterozygotes appearing as homozygotes while homozygotes typically do not show alleles at the respective loci at all [130]. Furthermore, if the alleles in a heterozygote differ strongly in size, the shorter allele may be preferentially amplified during PCR, at the expense of the larger one. Hence, the signal of the larger allele might be too weak to be confidently detected during genotyping, resulting in the individual to be erroneously called as homozygous [131]. Hardy–Weinberg equilibrium (HWE) and genotypic linkage disequilibrium (LD) were assessed through Genepop on the Web v.4.2 using the Markov chain method with 10,000 dememorisations, 1000 batches and 10,000 iterations per batch. Subsequently, we used GenAEx v.6.503 (‘dpds’) and SPAGeDi v.1.5 [132] (‘mpds’) to compute genetic diversity indices, including observed and unbiased expected heterozygosity. Rarefied allelic richness [133] was obtained in hp-rare v.1.1 [134] (‘dpds’) and SPAGeDi (‘mpds’). We then computed population pairwise and global genetic distances ( $F_{st}$ ) and corresponding  $P$  values for ‘dpds’ using GenAEx with 9999 permutations. Negative  $F_{st}$  values were converted to zero and  $P$  values were adjusted for multiple testing through false discovery rate (FDR) correction [47] using the ‘p.adjust’ function in R v.3.6.0 [135]. For ‘mpds’ both the  $F_{st}$  matrix and global  $F_{st}$  value were computed with the ‘calcPopDiff’ function based on allele frequencies calculated with the ‘simpleFreq’ function in the polysat v.1.7.4 R package [48]. Additionally, for both global and pairwise values, we computed the 95% confidence intervals of 10,000 bootstrap replicates.  $F_{st}$  values were considered significant if they surpassed the upper 95% confidence interval. To explore whether potential differences in  $F_{st}$  estimates between ‘dpds’ and ‘mpds’ are simply an artefact of the different computational approaches used, we also calculated pairwise and global  $F_{st}$  estimates for ‘dpds’ in polysat. Significance testing was performed as described above.

Following the computation of estimates of  $F_{st}$ , we assessed genetic structure in ‘cpds’ through Bayesian clustering in Structure v.2.3.4 [1] using the ‘Admixture’ and ‘Correlated Allele Frequency’ models with population IDs as priors [51]. We set the number of clusters ( $K$ ) from 2 to 25 (number of populations sampled) and computed 5 iterations with a burnin period of 100,000 and 1,000,000 Markov chain Monte Carlo (MCMC) repetitions. Relative admixture between populations was estimated by Structure (INFERALPHA = 1). Because previous studies suggested that buff-tailed bumble bees are quasi-panmictic, and they are highly vagile, we hypothesised that admixture is common, and the corresponding model in Structure was biologically most meaningful. Yet, we also explored the performance of Structure’s ‘No admixture’ model by running two iterations with the aforementioned

burnin period and MCMC repetitions. Additionally, as uneven sampling sizes among populations might result in an incorrect number of clusters [136], two further Structure analyses were run after excluding populations with fewer than 10 individuals ('pop10', Additional file 1: Table S8). After setting the maximum  $K$  to 23, the first run was computed with the settings specified above while for the second run we additionally set STARTATPOP-INFO to 1 to initialise each individual at its own population [137]. Structure output was visualised with the pophelper v.2.3.1 R package [138].

As complex patterns of population subdivision might not be detected reliably using only one clustering approach, we complemented Structure runs with discriminant analysis of principal components (DAPC, 40) implemented in the adegenet v.2.1.5 package [139] in R v.4.1.1. In fact, by maximising between-group variability, DAPC might be well-suited to detect subtle and complex patterns of population subdivision, such as hierarchical clustering or clines [40]. We performed two DAPC runs. Briefly, the first one was computed using populations as a priori groups. For the second analysis, we inferred clusters de novo using the 'find.clusters' function with 1,000,000 iterations and 700 random starting points, and selected the number of clusters that minimised the BIC value. To avoid overfitting, we determined the optimal number of PCs to retain in both analyses using the 'optim.a.score' function with 1000 simulations, respectively. The results were visualised in scatter plots.

### **Mechanisms of divergence**

As the likelihood of successful dispersal between populations decreases with geographic distance, gene flow starts to become limited and populations diverge through genetic drift. Under this scenario, a positive relationship between geographic and genetic distance is anticipated, a pattern termed isolation by distance (IBD, 7). Moreover, heterogeneous conditions of the habitat through which dispersal takes place impose varying levels of resistance to dispersing individuals. They may therefore not follow a straight line, but instead the path of least resistance. Least-cost path [140] and isolation by resistance (IBR, 8) analyses aim to quantify this heterogeneity in habitat permeability and its effect on population structure. Both IBD and IBR describe processes resulting in neutral population divergence, and are jointly coined isolation by dispersal limitation [141]. In addition, species experience heterogeneous environmental conditions that may exert strong selection pressures on populations, potentially leading to local adaptation and hence population divergence [142, 143]. Interestingly, the prolonged reduced fitness of dispersers [9] that are maladapted to newly encountered conditions might result in the disruption of

genetic connectivity. As a consequence, populations may differentiate by means of genetic drift in neutral markers, a pattern termed isolation by environment/ecology (IBE, [10, 11]) or, alternatively, isolation by adaptation (IBA, [12]).

We considered all three potential mechanisms of population divergence in our analyses. In addition to straight-line geographic distance, we also included a measure of habitat permeability using resistance distances based on a previously published species distribution model [21], as well as a set of environmental variables.

### **Environmental variables and permeability of the habitat matrix to dispersal**

We characterised environmental conditions across Romania and Bulgaria using a set of 16 environmental variables compiled for previous species distribution and landscape genetic studies [21, 39] (Additional file 1: Table S9). This set included variables related to climate, vegetation and topography at 30 arcseconds resolution, which roughly converts to a 1 km resolution at the equator. Briefly, we initially considered 19 climate variables from WorldClim v.2 [144], including temperature and precipitation variables based on a 30-year climatology from 1970 to 2000 [145]. Elevation data originated from the Shuttle Radar Topography Mission (SRTM, [146]), and were also used to compute aspect and slope. We included spatial and temporal vegetation patterns derived from satellite data: percent tree cover and Leaf Area Index (the one-sided green leaf area per unit ground area), both obtained from the Global Land Cover Facility database [147]. Information on vertical forest structure, i.e. canopy height, was derived from space borne LiDAR from 2011 [148]. Finally, to incorporate information about surface moisture, we included annual mean, minimum, maximum, and seasonality, computed from raw QuikSCAT data [39]. To do so, we used daily raw backscatter measurements downloaded from the BYU Scatterometer Climate Record Pathfinder database [149] over the period the instrument was online (2000–2008). This initial data set was reduced by excluding highly correlated variables. We did so using their variance inflation factor (VIF) and excluded those with a score  $\geq 10$  in a stepwise fashion in the usdm R package v.1.1–18 [150] in R v.3.6.1. To facilitate the discussion of our findings, we also quantified environmental similarities among population locations by means of Pearson correlation coefficients (Additional file 1: Table S7), calculated using the 'cor' function in R v.3.6.2. More information on the environmental data and processing can be found elsewhere [21, 39].

To obtain a measure of habitat permeability, we further calculated population pairwise resistance distances based

on the species distribution model (SDM) of *B. terrestris* [21]. Computations were carried out in Circuitscape v.4.0.5 [151] with the SDM surface being treated as a conductance map, a cell size of 0.0083333° (i.e. 30 arcseconds) and a cell connection scheme of eight neighbours. Although species distribution models are not directly informative about genetic connectivity (i.e. functional connectivity, [152]), they represent the spatial configuration of putatively suitable habitat (i.e. structural connectivity, [153]). One caveat of this approach is that habitat unsuitable for breeding or foraging may nevertheless be easily crossed by vagile species. Yet, it may be reasonable to suspect that more suitable habitat can be crossed more easily than less suitable habitat, and thus that SDMs provide a useful estimate of relative habitat permeability. We therefore considered it meaningful to use SDM-derived population pairwise resistance distances (Additional file 1: Table S10) as additional predictors in our analyses.

### Landscape genetic analyses

We established associations between genetic data and environmental variables using generalized dissimilarity modelling implemented in the *gdm* R package v.1.3.11 [154]. This extension of the classical matrix regression allows to fit non-linear relationships between predictors and response variables. Additionally, the contribution of predictors can be analysed simultaneously [41]. In a landscape genetics framework this means that the effects of geographic distance and the environment on genetic divergence can be analysed at the same time, and estimates of variable importance are provided through permutation [155–157]. Predictors inferred as informative in explaining genetic turnover yield I-splines that provide two pieces of information. The maximum height of the curve indicates the amount of biological change along the gradient, while the spline's shape informs about the rate of genetic turnover [156]. In this study, we used this framework to infer the relationship between pairwise genetic distances ( $F_{st}$ , scaled between zero and one) and environmental variables, geographic distance and resistance distance. In total, we computed five different models, that included (1) environmental variables, geographic and resistance distances (full model), (2) environmental variables only, (3) straight-line geographic distances only, (4) resistance distances only, and (5) 1000 models, each using 16 random environmental variables to evaluate the performance of the full model. We considered the full model significant if its variation explained surpassed the 95% confidence interval of the random models. Subsequently, following the approach by Fitzpatrick and Keller [156], we used the inferred relationships between predictor and response variables at sampling sites to predict genetic turnover across Romania and Bulgaria. First,

for each retained environmental variable, we extracted its value across the study area at 30 arcseconds resolution. Using the 'gdm.transform' function from the *gdm* package, we then transformed environmental variables into 'genetic' importance values. We selected the three most influential and uncorrelated factors through principal component analyses (PCA) using the 'princomp' function with calculations performed on the covariance matrix. We purposely centred, but not scale-transformed the PCA to preserve differences in the magnitude of the genetic importance among environmental variables. With the 'rasterize' function in QGIS v.3.16.2 [158] we converted the obtained point values to rasters, merged them with the 'r.composite' function to a composite and assigned the corresponding PC scores to a RGB palette. The result (Fig. 3(b)) visualises differences in genetic composition, where increasingly dissimilar colours represent higher predicted genetic turnover.

### Abbreviations

$A_R$ : Allelic richness; CI: Confidence interval; CO1: Cytochrome c oxidase subunit I; cpds: Complete data set; a data set containing both haploid and diploid individuals before any populations were excluded due to low sample sizes; DAPC: Discriminant analysis of principal components; dpds: Diploid data set; a data set containing diploid individuals only;  $H_O$ : Observed heterozygosity;  $K$ : Number of Structure clusters;  $k$ : Number of individuals used for calculating rarefied allelic richness; mpds: Mixed-ploidy data set; a data set containing both haploid and diploid individuals; msX: Microsatellite loci X;  $N$ : Number of individuals;  $N_A$ : Number of alleles;  $N_E$ : Number of effective alleles; PMX: Primer mix X; pop10: Complete data set after excluding populations with fewer than 10 individuals;  $T_m$ : Annealing temperature;  $uH_E$ : Unbiased expected heterozygosity.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-01963-5>.

**Additional file 1: Table S1.** Microsatellite loci used in this work, with original locus names [127], the primer mixes they were assigned to, repeat units (motifs), size ranges, annealing temperatures ( $T_m$ ), forward and reverse primer sequences (5'–3') as well as information on null alleles and stuttering. **Table S2.** Basic population genetic statistics for each sampling location encompassing diploid ('dpds') and both haploid and diploid individuals ('mpds').  $N$ : number of individuals;  $N_A$ : number of alleles;  $N_E$ : number of effective alleles;  $A_R$ : allelic richness, rarefied for seven individuals ( $k$ );  $uH_E$ : unbiased expected heterozygosity;  $H_O$ : observed heterozygosity. Observed heterozygosity values between 'dpds' and 'mpds' are largely congruent, as SPAGeDi, even though capable of working with mixed-ploidy data sets, does not include haploid individuals in calculating this metric. **Table S3.** Below the diagonal: Multilocus pairwise  $F_{st}$  values calculated with GenAlEx and based on a data set containing diploid individuals only. Above the diagonal: Corresponding FDR-corrected  $P$  values. Marginally significant  $F_{st}$  values and  $P$  values are indicated in bold. See Table S8 for population names. **Table S4.** Above the diagonal: Multilocus pairwise  $F_{st}$  values computed with polysat and based on a data set containing both haploid and diploid individuals. Below the diagonal: Corresponding bootstrap-derived upper 95% confidence intervals. See Table S8 for population names. **Table S5.** Above the diagonal: Multilocus pairwise  $F_{st}$  values calculated with polysat and based on a data set containing diploid individuals only. Below the diagonal: Corresponding bootstrap-derived upper 95% confidence intervals. See Table S8 for population names. **Table S6.**

Discriminant analysis of principal components (DAPC) result showing the assignment of individuals to clusters identified de novo. **Table S7.** Pearson correlation coefficients of environmental variables at population locations harbouring both haploid and diploid individuals. See Table S9 for the abbreviations used. **Table S8.** Populations sampled, including their abbreviations, locations, and the number of individuals. 'cpds': full data set after exclusion of full siblings and clones; 'pop10': 'cpds' after excluding populations with fewer than 10 individuals; 'mpds': mixed-ploidy data set encompassing both haploid males and diploid females; 'dpds': data set containing diploid individuals only. **Table S9.** Environmental variables used. Those in bold were retained after stepwise elimination with a variance inflation factor  $\geq 10$ . **Table S10.** Resistance matrix derived from Circuitscape, where higher values indicate higher predicted resistance to inter-population dispersal. See Table S8 for population names.

**Additional file 2:** Structure output based on the 'complete data set' ('cpds') for selected values of  $K$ ; result of the a-score optimisation for the DAPC runs performed as well as the DAPC result obtained using populations as a priori groups; a curve depicting model performance for different numbers of clusters inferred de novo; GDM splines produced from data sets encompassing haploid and diploid individuals (mixed-ploidy data set, 'mpds') and diploid individuals only ('diploid data set', 'dpds').

#### Acknowledgements

We thank Tatjana Stooß for assisting with DNA extractions. MG would like to extend his gratitude to the Studienstiftung des deutschen Volkes (<https://www.studienstiftung.de/en/>) for financial and ideational support throughout his studies.

#### Authors' contributions

JCG and HAT conceived the study with contributions from MG. HAT and JCG secured funding. JCG and HAT conducted the field sampling. MG performed the laboratory procedures with occasional help from JCG and HAT. Data analysis was led by MG in correspondence with JCG and HAT. MG led the writing process of the manuscript with contributions from JCG and HAT. All authors read the final manuscript and approved its publication.

#### Funding

This project was funded by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007–2013, Grant/Award Number: 293886 granted to HAT. Following funding was awarded to JCG: State Postgraduate Fellowship Programme of Baden-Württemberg, Germany, 2016–2018, Award Number: I 1.3-7631.2; LBBW-Stiftung; Beatrice Nolte Stiftung für Natur- und Umweltschutz. We acknowledge support from Projekt DEAL and the Open Access Publishing Fund of Tübingen University. The funding bodies played no role in the design of the study, the collection, analysis, and interpretation of data as well as writing the manuscript.

#### Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Dryad repository <https://doi.org/10.5061/dryad.gb5mkkwq8>. The R scripts used for data processing are available from the corresponding author upon reasonable request.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Comparative Zoology, Institute of Evolution and Ecology, Tübingen University, Tübingen, Germany. <sup>2</sup>Department of Ecology and Environmental Sciences, Umeå University, Umeå, Sweden.

Received: 2 July 2021 Accepted: 18 January 2022

Published online: 01 February 2022

#### References

- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945–59.
- Guillot G, Mortier F, Estoup A. Geneland: a computer package for landscape genetics. *Mol Ecol Notes*. 2005;5(3):712–5.
- Jombart T, Devillard S, Dufour A-B, Pontier D. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*. 2008;101(1):92–103.
- Chen C, Durand E, Forbes F, François O. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes*. 2007;7(5):747–56.
- Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet*. 2005;1(6):e70.
- Younger JL, Clucas GV, Kao D, Rogers AD, Gharbi K, Hart T, et al. The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Mol Ecol*. 2017;26(15):3883–97.
- Wright S. Isolation by distance. *Genetics*. 1943;28(2):114.
- McRae BH. Isolation by resistance. *Evolution*. 2006;60(8):1551–61.
- Hendry AP. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evol Ecol Res*. 2004;6(8):1219–36.
- Wang JJ, Summers K. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol*. 2010;19(3):447–58.
- Edelaar P, Alonso D, Lagerveld S, Senar J, Björklund M. Population differentiation and restricted gene flow in Spanish crossbills: not isolation-by-distance but isolation-by-ecology. *J Evol Biol*. 2012;25(3):417–30.
- Nosil P, Egan SP, Funk DJ. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evol Int J Org Evol*. 2008;62(2):316–36.
- Hantak MM, Page RB, Converse PE, Anthony CD, Hickerson CAM, Kuchta SR. Do genetic structure and landscape heterogeneity impact color morph frequency in a polymorphic salamander? *Ecography*. 2019;42(8):1383–94.
- Myers EA, Xue AT, Gehara M, Cox CL, Davis Rabosky AR, Lemos-Espinal J, et al. Environmental heterogeneity and not vicariant biogeographic barriers generate community-wide population structure in desert-adapted snakes. *Mol Ecol*. 2019;28(20):4535–48.
- Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol*. 2003;18(4):189–97.
- Holderegger R, Wagner HH. A brief guide to Landscape Genetics. *Landscape Ecol*. 2006;21(6):793–6.
- Storfer A, Murphy M, Evans J, Goldberg C, Robinson S, Spear S, et al. Putting the 'landscape' in landscape genetics. *Heredity*. 2007;98(3):128.
- Kraus RH, Van Hooff P, Megens HJ, Tsvey A, Fokin SY, Ydenberg RC, et al. Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Mol Ecol*. 2013;22(1):41–55.
- Troast D, Suhling F, Jinguji H, Sahlén G, Ware J. A global population genetic study of *Pantala flavescens*. *PLoS ONE*. 2016;11(3):e0148949.
- King RC, Mulligan PK, Stansfield WD. A dictionary of genetics. 8th ed. Oxford: Oxford University Press; 2013.
- Geue JC, Thomassen HA. Unraveling the habitat preferences of two closely related bumble bee species in Eastern Europe. *Ecol Evol*. 2020;10:4773–90.
- Hingston AB, Marsden-Smedley J, Driscoll DA, Corbett S, Fenton J, Anderson R, et al. Extent of invasion of Tasmanian native vegetation by the exotic bumblebee *Bombus terrestris* (Apoidea: Apidae). *Austral Ecol*. 2002;27(2):162–72.

23. Rasmont P, Coppée A, Michez D, De Meulemeester T, editors. An overview of the *Bombus terrestris* (L. 1758) subspecies (Hymenoptera: Apidae). *Ann Soc Entomol Fr*; 2008: Taylor & Francis.
24. Rasmont P, Franzén M, Lecocq T, Harpke A, Roberts SP, Biesmeijer JC, et al. Climatic risk and distribution atlas of European bumblebees: Pensoft Publishers; 2015.
25. Chapman R, Wang J, Bourke A. Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Mol Ecol*. 2003;12(10):2801–8.
26. Estoup A, Solignac M, Cornuet J, Goudet J, Scholl A. Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. *Mol Ecol*. 1996;5(1):19–31.
27. Silva SE, Seabra SG, Carvalheiro LG, Nunes VL, Marabuto E, Mendes R, et al. Population genomics of *Bombus terrestris* reveals high but unstructured genetic diversity in a potential glacial refugium. *Biol J Linn Soc*. 2020;129(2):259–72.
28. Moreira AS, Horgan FG, Murray TE, Kakouli-Duarte T. Population genetic structure of *Bombus terrestris* in Europe: isolation and genetic differentiation of Irish and British populations. *Mol Ecol*. 2015;24(13):3257–68.
29. Widmer A, Schmid-Hempel P, Estoup A, Scholl A. Population genetic structure and colonization history of *Bombus terrestris* sl (Hymenoptera: Apidae) from the Canary Islands and Madeira. *Heredity*. 1998;81(5):563.
30. Lecocq T, Vereecken NJ, Michez D, Dellicour S, Lhomme P, Valterova I, et al. Patterns of genetic and reproductive traits differentiation in mainland vs. Corsican populations of bumblebees. *PLoS ONE*. 2013;8(6):e65642.
31. Chittka L, Ings TC, Raine NE. Chance and adaptation in the evolution of island bumblebee behaviour. *Popul Ecol*. 2004;46(3):243–51.
32. Schmid-Hempel P, Schmid-Hempel R, Brunner P, Seeman O, Allen G. Invasion success of the bumblebee, *Bombus terrestris*, despite a drastic genetic bottleneck. *Heredity*. 2007;99(4):414–22.
33. Schmid-Hempel R, Eckhardt M, Goulson D, Heinzmann D, Lange C, Plischuk S, et al. The invasion of southern South America by imported bumblebees and associated parasites. *J Anim Ecol*. 2014;83(4):823–37.
34. Lepais O, Darvill B, O'Connor S, Osborne JL, Sanderson RA, Cussans J, et al. Estimation of bumblebee queen dispersal distances using sibship reconstruction method. *Mol Ecol*. 2010;19(4):819–31.
35. Kraus F, Wolf S, Moritz R. Male flight distance and population substructure in the bumblebee *Bombus terrestris*. *J Anim Ecol*. 2009;78(1):247–52.
36. Kelly AE, Goulden ML. Rapid shifts in plant distribution with recent climate change. *Proc Natl Acad Sci*. 2008;105(33):11823–6.
37. Wang C-J, Wan J-Z, Zhang G-M, Zhang Z-X, Zhang J. Protected areas may not effectively support conservation of endangered forest plants under climate change. *Environmental Earth Sciences*. 2016;75(6):466.
38. Yun J-H, Nakao K, Tsuyama I, Matsui T, Park C-H, Lee B-Y, et al. Vulnerability of subalpine fir species to climate change: using species distribution modeling to assess the future efficiency of current protected areas in the Korean Peninsula. *Ecol Res*. 2018;33(2):341–50.
39. Geue JC, Vágási CI, Schweizer M, Pap PL, Thomassen HA. Environmental selection is a main driver of divergence in house sparrows (*Passer domesticus*) in Romania and Bulgaria. *Ecol Evol*. 2016;6(22):7954–64.
40. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11(1):1–15.
41. Ferrier S, Manion G, Elith J, Richardson K. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers Distrib*. 2007;13(3):252–64.
42. Peakall R, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 2006;6(1):288–95.
43. Peakall R, Smouse PE. GenAlix 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 2012;28(19):2537–9.
44. Jones OR, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour*. 2010;10(3):551–5.
45. Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes*. 2004;4(3):535–8.
46. Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour*. 2008;8(1):103–6.
47. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc: Ser B (Methodol)*. 1995;57(1):289–300.
48. Clark LV, Jasieniuk M. POLYSAT: an R package for polyploid microsatellite analysis. *Mol Ecol Resour*. 2011;11(3):562–6.
49. Wright S. The genetical structure of populations. *Ann Eugen*. 1951;15(1):323–54.
50. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
51. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164(4):1567–87.
52. Miller JM, Cullingham CI, Peery RM. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity*. 2020;125(5):269–80.
53. Lecocq T, Rasmont P, Harpke A, Schweiger O. Improving international trade regulation by considering intraspecific variation for invasion risk assessment of commercially traded species: the *Bombus terrestris* case. *Conserv Lett*. 2016;9(4):281–9.
54. Kelso NV, Patterson T. Introducing natural earth data—natureearthdata.com. *Geographia Technica*. 2010;5(8289):25.
55. Theodorou P, Radzevičiūtė R, Kahnt B, Soro A, Grosse I, Paxton RJ. Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (*Bombus lapidarius* L.). *Proc R Soc B Biol Sci*. 1877;2018(285):20172806.
56. Maebe K, Karise R, Meeus I, Mänd M, Smagge G. Pattern of population structuring between Belgian and Estonian bumblebees. *Sci Rep*. 2019;9(1):1–8.
57. Shao Z, Mao H, Fu W, Ono M, Wang D, Bonizzoni M, et al. Genetic Structure of Asian Populations of *Bombus ignitus* (Hymenoptera: Apidae). *J Hered*. 2004;95(1):46–52.
58. Widmer A, Schmid-Hempel P. The population genetic structure of a large temperate pollinator species, *Bombus pascuorum* (Scopoli) (Hymenoptera: Apidae). *Mol Ecol*. 1999;8(3):387–98.
59. Lozier JD, Strange JP, Stewart IJ, Cameron SA. Patterns of range-wide genetic variation in six North American bumble bee (Apidae: *Bombus*) species. *Mol Ecol*. 2011;20(23):4870–88.
60. Dreier S, Redhead JW, Warren IA, Bourke AF, Heard MS, Jordan WC, et al. Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Mol Ecol*. 2014;23(14):3384–95.
61. Koch JB, Looney C, Sheppard WS, Strange JP. Patterns of population genetic structure and diversity across bumble bee communities in the Pacific Northwest. *Conserv Genet*. 2017;18(3):507–20.
62. Jha S. Contemporary human-altered landscapes and oceanic barriers reduce bumble bee gene flow. *Mol Ecol*. 2015;24(5):993–1006.
63. Goulson D, Kaden J, Lepais O, Lye G, Darvill B. Population structure, dispersal and colonization history of the garden bumblebee *Bombus hortorum* in the Western Isles of Scotland. *Conserv Genet*. 2011;12(4):867–79.
64. Darvill B, Ellis JS, Lye GC, Goulson D. Population structure and inbreeding in a rare and declining bumblebee, *Bombus muscorum* (Hymenoptera: Apidae). *Mol Ecol*. 2006;15(3):601–11.
65. Amin MR, Than KK, Kwon YJ. Copulation duration of bumblebee *Bombus terrestris* (Hymenoptera: Apidae): Impacts on polyandry and colony parameters. *J Asia-Pac Entomol*. 2009;12(3):141–4.
66. Zhang H, Zhou Z, Huang J, Yuan X, Ding G, An J. Queen traits and colony size of four bumblebee species of China. *Insectes Soc*. 2018;65(4):537–47.
67. Duchateau M, Velthuis H. Development and reproductive strategies in *Bombus terrestris* colonies. *Behaviour*. 1988;107:186–207.
68. Connop S, Hill T, Steer J, Shaw P. The role of dietary breadth in national bumblebee (*Bombus*) declines: Simple correlation? *Biol Conserv*. 2010;143(11):2739–46.
69. Wood TJ, Gibbs J, Graham KK, Isaacs R. Narrow pollen diets are associated with declining Midwestern bumble bee species. *Ecology*. 2019;100(6):e02697.

70. Meeus I, Parmentier L, Pisman M, de Graaf DC, Smagghe G. Reduced nest development of reared *Bombus terrestris* within apiary dense human-modified landscapes. *Sci Rep*. 2021;11(1):1–9.
71. Fontaine C, Collin CL, Dajoz I. Generalist foraging of pollinators: diet expansion at high density. *J Ecol*. 2008;96(5):1002–10.
72. Avarguès-Weber A, Lachlan R, Chittka L. Bumblebee social learning can lead to suboptimal foraging choices. *Anim Behav*. 2018;135:209–14.
73. Müller CB, Schmid-Hempel P. Correlates of reproductive success among field colonies of *Bombus lucorum*: the importance of growth and parasites. *Ecol Entomol*. 1992;17(4):343–53.
74. Goulson D. *Bumblebees: behaviour, ecology, and conservation*. 2nd ed. Oxford: Oxford University Press; 2009.
75. Knight ME, Martin AP, Bishop S, Osborne JL, Hale RJ, Sanderson RA, et al. An interspecific comparison of foraging range and nest density of four bumblebee (*Bombus*) species. *Mol Ecol*. 2005;14(6):1811–20.
76. Darvill B, Knight ME, Goulson D. Use of genetic markers to quantify bumblebee foraging range and nest density. *Oikos*. 2004;107(3):471–8.
77. Ghisbain G, Lozier JD, Rahman SR, Ezray BD, Tian L, Ulmer JM, et al. Substantial genetic divergence and lack of recent gene flow support cryptic speciation in a colour polymorphic bumble bee (*Bombus bifarius*) species complex. *Syst Entomol*. 2020;45(3):635–52.
78. Alford D. A study of the hibernation of bumblebees (Hymenoptera: Bombidae) in Southern England. *J Anim Ecol*. 1969;38:149–70.
79. Pomeroy N, Plowright R. The relation between worker numbers and the production of males and queens in the bumble bee *Bombus perplexus*. *Can J Zool*. 1982;60(5):954–7.
80. Müller C, Schmid-Hempel P. Variation in life-history pattern in relation to worker mortality in the bumble-bee, *Bombus lucorum*. *Funct Ecol*. 1992;6:48–56.
81. Rundlöf M, Persson AS, Smith HG, Bommarco R. Late-season mass-flowering red clover increases bumble bee queen and male densities. *Biol Conserv*. 2014;172:138–45.
82. Bowers MA. Resource availability and timing of reproduction in bumble bee colonies (Hymenoptera: Apidae). *Environ Entomol*. 1986;15(3):750–5.
83. Schmid-Hempel P, Durrer S. Parasites, floral resources and reproduction in natural populations of bumblebees. *Oikos*. 1991;62:342–50.
84. Baer B, Morgan ED, Schmid-Hempel P. A nonspecific fatty acid within the bumblebee mating plug prevents females from remating. *Proc Natl Acad Sci*. 2001;98(7):3926–8.
85. Duvoisin N, Baer B, Schmid-Hempel P. Sperm transfer and male competition in a bumblebee. *Anim Behav*. 1999;58(4):743–9.
86. Heinrich B. *Bumblebee economics*. Cambridge: Harvard University Press; 1979.
87. Vogt FD. Thermoregulation in bumblebee colonies. I. Thermoregulatory versus brood-maintenance behaviors during acute changes in ambient temperature. *Physiol Zool*. 1986;59(1):55–9.
88. Owen EL, Bale JS, Hayward SA. Can winter-active bumblebees survive the cold? Assessing the cold tolerance of *Bombus terrestris audax* and the effects of pollen feeding. *PLoS ONE*. 2013;8(11):e80061.
89. Stelzer RJ, Chittka L, Carlton M, Ings TC. Winter active bumblebees (*Bombus terrestris*) achieve high foraging rates in urban Britain. *PLoS ONE*. 2010;5(3):e9559.
90. Martinet B, Rasmont P, Cederberg B, Evrard D, Ødegaard F, Paukunen J, et al., editors. Forward to the north: two Euro-Mediterranean bumblebee species now cross the Arctic Circle. *Annales de la Société entomologique de France (NS)*. Routledge: Taylor & Francis. 2015.
91. Gherghescu D-Ş, Dabija A-M, editors. *The Challenges of the Bioclimatic Architecture in Romania*. IOP Conference Series: Materials Science and Engineering. Bristol: IOP Publishing. 2020.
92. Malcheva K, Pophrystov V, Marinova T, Trifonova L, editors. *Complex approach for classification of winter severity in Bulgaria*. AIP Conference Proceedings. Melville: AIP Publishing LLC. 2019.
93. Kreyer D, Oed A, Walther-Hellwig K, Frankl R. Are forests potential landscape barriers for foraging bumblebees? Landscape scale experiments with *Bombus terrestris* agg. and *Bombus pascuorum* (Hymenoptera, Apidae). *Biol Conserv*. 2004;116(1):111–8.
94. Mola JM, Miller MR, O'Rourke SM, Williams NM. Forests do not limit bumble bee foraging movements in a montane meadow complex. *Ecol Entomol*. 2020;45(5):955–65.
95. Svensson B, Lagerlöf J, Svensson BG. Habitat preferences of nest-seeking bumble bees (Hymenoptera: Apidae) in an agricultural landscape. *Agr Ecosyst Environ*. 2000;77(3):247–55.
96. Jha S, Kremen C. Urban land use limits regional bumble bee gene flow. *Mol Ecol*. 2013;22(9):2483–95.
97. Kadoya T, Washitani I. Predicting the rate of range expansion of an invasive alien bumblebee (*Bombus terrestris*) using a stochastic spatio-temporal model. *Biol Conserv*. 2010;143(5):1228–35.
98. Liczner AR, Colla SR. A systematic review of the nesting and overwintering habitat of bumble bees globally. *J Insect Conserv*. 2019;23(5):787–801.
99. Plath O. Notes on the hibernation of several North American bumblebees. *Ann Entomol Soc Am*. 1927;20(2):181–92.
100. Makinson JC, Woodgate JL, Reynolds A, Capaldi EA, Perry CJ, Chittka L. Harmonic radar tracking reveals random dispersal pattern of bumblebee (*Bombus terrestris*) queens after hibernation. *Sci Rep*. 2019;9(1):4651.
101. Kämper W, Werner PK, Hilpert A, Westphal C, Blüthgen N, Eltz T, et al. How landscape, pollen intake and pollen quality affect colony growth in *Bombus terrestris*. *Landsc Ecol*. 2016;31(10):2245–58.
102. Redhead JW, Dreier S, Bourke AF, Heard MS, Jordan WC, Sumner S, et al. Effects of habitat composition and landscape structure on worker foraging distances of five bumble bee species. *Ecol Appl*. 2016;26(3):726–39.
103. Straub L, Williams GR, Vidondo B, Khongphinitbunjong K, Retschnig G, Schneeberger A, et al. Neonicotinoids and ectoparasitic mites synergistically impact honeybees. *Sci Rep*. 2019;9(1):1–10.
104. Botías C, Jones JC, Pamminger T, Bartomeus I, Hughes WO, Goulson D. Multiple stressors interact to impair the performance of bumblebee *Bombus terrestris* colonies. *J Anim Ecol*. 2021;90(2):415–31.
105. Dafni A, Kevan P, Gross CL, Goka K. *Bombus terrestris*, pollinator, invasive and pest: an assessment of problems associated with its widespread introductions for commercial purposes. *Appl Entomol Zool*. 2010;45(1):101–13.
106. Velthuis HH, Van Doorn A. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie*. 2006;37(4):421–51.
107. Semmens T, Turner E, Buttermore R. *Bombus terrestris* (L.) (Hymenoptera: Apidae) now established in Tasmania. *J Aust Entomol Soc*. 1993;32(4):346.
108. Matsumura C, Yokoyama J, Washitani I. Invasion status and potential ecological impacts of an invasive alien bumblebee, *Bombus terrestris* L. (Hymenoptera: Apidae) naturalized in Southern Hokkaido Japan. *Glob Environ Res*. 2004;8(1):51–66.
109. Macfarlane R, Gurr L. Distribution of bumble bees in New Zealand. *N Z Entomol*. 1995;18(1):29–36.
110. Romanian Statistical Yearbook. National Institute of Statistics. Bucharest; 2018.
111. Library of Congress. Country profile: Bulgaria. <https://www.loc.gov/rr/frd/cs/profiles/Bulgaria.pdf>. Accessed 20 May 2019.
112. Milewski P, Szmyd J. Biothermal contrasts while travelling in or between Poland and Bulgaria. *EUROPA XXI*. 2015;29:73–84.
113. Goulson D, Stout JC. Homing ability of the bumblebee *Bombus terrestris* (Hymenoptera: Apidae). *Apidologie*. 2001;32(1):105–11.
114. Wolf S, Moritz RF. Foraging distance in *Bombus terrestris* L. (Hymenoptera: Apidae). *Apidologie*. 2008;39(4):419–27.
115. Smithers CN. *The handbook of insect collecting: Their collection, preparation, preservation, and storage*. Melbourne: Angus & Robertson; 1988.
116. Garvin M, Saitoh K, Gharrett A. Application of single nucleotide polymorphisms to non-model species: a technical review. *Mol Ecol Resour*. 2010;10(6):915–34.
117. Narum S, Banks M, Beacham T, Bellinger M, Campbell M, Dekoning J, et al. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Mol Ecol*. 2008;17(15):3464–77.
118. Putman AI, Carbone I. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol Evol*. 2014;4(22):4399–428.
119. Karl SA, Toonen R, Grant W, Bowen B. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol Ecol*. 2012;21(17):4171–89.

120. Goldstein D, Pollock D. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J Hered*. 1997;88(5):335–42.
121. Lowe WH, Allendorf FW. What can genetics tell us about population connectivity? *Mol Ecol*. 2010;19(15):3038–51.
122. Jakobsson M, Edge MD, Rosenberg NA. The relationship between *F<sub>ST</sub>* and the frequency of the most frequent allele. *Genetics*. 2013;193(2):515–28.
123. Estoup A, Jarne P, Cornuet JM. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol*. 2002;11(9):1591–604.
124. Campos PF, Gilbert TM. DNA extraction from keratin and chitin. *Ancient DNA*: Springer; 2012. p. 43–9.
125. Williams PH. Phylogenetic relationships among bumble bees (*Bombus* Latr.): a reappraisal of morphological evidence. *Syst Entomol*. 1994;19(4):327–44.
126. Wolf S, Rohde M, Moritz RF. The reliability of morphological traits in the differentiation of *Bombus terrestris* and *B. lucorum* (Hymenoptera: Apidae). *Apidologie*. 2010;41(1):45–53.
127. Stolle E, Rohde M, Vautrin D, Solignac M, Schmid-Hempel P, Schmid-Hempel R, et al. Novel microsatellite DNA loci for *Bombus terrestris* (Linnaeus, 1758). *Mol Ecol Resour*. 2009;9(5):1345–52.
128. Bogo G, De Manincor N, Fisogni A, Galloni M, Zavatta L, Bortolotti L. No evidence for an inbreeding avoidance system in the bumble bee *Bombus terrestris*. *Apidologie*. 2018;49(4):473–83.
129. Kalinowski S. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*. 2005;94(1):33–6.
130. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett*. 2006;9(5):615–29.
131. Chakraborty R, Andrade MD, Daiger S, Budowle B. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann Hum Genetics*. 1992;56(1):45–57.
132. Hardy O, Vekemans X. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2002;2:618–20.
133. Hurlbert SH. The nonconcept of species diversity: a critique and alternative parameters. *Ecology*. 1971;52(4):577–86.
134. Kalinowski ST. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 2005;5(1):187–9.
135. R Core Team. R: a language and environment for statistical computing (Version 3.6). R Foundation for Statistical Computing". R Foundation for Statistical Computing. 2019.
136. Puechmaile SJ. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol Ecol Resour*. 2016;16(3):608–27.
137. Pritchard JK, Wen W, Falush D. Documentation for STRUCTURE software: Version 2. 2003.
138. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017;17(1):27–32.
139. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403–5.
140. Adriaensen F, Chardon J, De Blust G, Swinnen E, Villalba S, Gulinck H, et al. The application of 'least-cost' modelling as a functional landscape model. *Landsc Urban Plan*. 2003;64(4):233–47.
141. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol*. 2013;22(24):5983–99.
142. Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol*. 2016;25(1):104–20.
143. Sexton JP, Hangartner SB, Hoffmann AA. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*. 2014;68(1):1–15.
144. WorldClim v.2 (cited Jan 2018). <http://www.worldclim.com/version2>.
145. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;37(12):4302–15.
146. Shuttle Radar Topography Mission (cited Mar 2015). <https://www2.jpl.nasa.gov/srtm/>.
147. The Global Land Cover Facility (GLCF) [cited Nov 2014 (LAI/percent tree cover); Nov 2017 (canopy height)]. Service discontinued.
148. Simard M, Pinto N, Fisher JB, Baccini A. Mapping forest canopy height globally with spaceborne lidar. *J Geophys Res Biogeosci*. 2011;116(G4).
149. BYU Center for Remote Sensing (cited Aug 2015). <https://www.scp.byu.edu/data/Quikscat/SIRv2/qush/Eur.html>.
150. Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG. Where is positional uncertainty a problem for species distribution modelling? *Ecography*. 2014;37(2):191–203.
151. McRae BH, Beier P. Circuit theory predicts gene flow in plant and animal populations. *Proc Natl Acad Sci*. 2007;104(50):19885–90.
152. Pe'er G, Henle K, Dislich C, Frank K. Breaking functional connectivity into components: a novel approach using an individual-based model, and first outcomes. *PLoS ONE*. 2011;6(8):e22355.
153. LaPoint S, Balkenhol N, Hale J, Sadler J, van der Ree R. Ecological connectivity research in urban areas. *Funct Ecol*. 2015;29(7):868–78.
154. Manion G, Lisk M, Ferrier S, Lugilde KM, Fitzpatrick MC. Package 'gdm'. A toolkit with functions to fit, plot, and summarize Generalized Dissimilarity Models: CRAN Repository, R. 2017.
155. Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK, Smith TB. Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Mol Ecol*. 2010;19(17):3532–48.
156. Fitzpatrick MC, Keller SR. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett*. 2015;18(1):1–16.
157. Fitzpatrick MC, Sanders NJ, Normand S, Svenning J-C, Ferrier S, Gove AD, et al. Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proc R Society B Biol Sci*. 2013;280(1768):20131201.
158. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Version 3.16. 2020.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

