Report on drivers of genetic diversity of key ecosystem service providing insects in agricultural landscapes across Europe

Deliverable D19 (D3.3)

31 October 2022

Ignasi Bartomeus[1], Joaquin Ortego[1]

[1] Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain

SHOWCASE

SHOWCASing synergies between agriculture, biodiversity and Ecosystem services to help farmers capitalising on native biodiversity

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862480.

Prepared under contract from the European Commission

Grant agreement No. 862480
EU Horizon 2020 Research and Innovation action

Project acronym:	**SHOWCASE**
Project full title:	**SHOWCASing synergies between agriculture, biodiversity and Ecosystem services to help farmers capitalising on native biodiversity**
Start of the project:	November 2020
Duration:	60 months
Project coordinator:	Prof. David Kleijn Wageningen University
Deliverable title:	Planned: Report on drivers of genetic diversity of key ecosystem service providing insects in agricultural landscapes across Europe
Deliverable n°:	D3.3
Nature of the deliverable:	Report
Dissemination level:	Public
WP responsible:	WP3
Lead beneficiary:	Consejo Superior de Investigaciones Científicas (CSIC)
Citation:	Bartomeus, I. and Ortego, J. (2022). *Report on drivers of genetic diversity of key ecosystem service providing insects in agricultural landscapes across Europe*. Deliverable D3.3 EU Horizon 2020 SHOWCASE Project, Grant agreement No 862480.
Due date of deliverable:	Month n°24
Actual submission date:	Month n°24

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	27 October 2022	Bartomeus & Ortego CSIC
2.0	Final version	31 October 2022	Bartomeus & Ortego CSIC

Table of contents

Preface 4
Key takeaway messages 4
Summary 5
List of abbreviations 5
1 Introduction 5
2 Methods 6
  2.1 Population sampling 6
  2.2 Genomic library preparation and processing 7
  2.3 Quantifying genetic structure 7
  2.4 Testing alternative models of gene flow 8
3 Results 9
  3.1 Genomic data 9
  3.2 Genetic diversity of populations 10
  3.3 Quantifying genetic structure 10
  3.4 Testing alternative models of gene flow 11
4 Discussion 14
5 Policy recommendations 16
6 Acknowledgements 16
7 References 16

Preface

Despite consistent claims to value multiple dimensions of biodiversity, including not only species diversity, but also functional or genetic diversity, most studies up-to-date focus on simple species diversity measures such as richness and abundance. By contrast, very little is known about the genetic diversity and resilience to population fragmentation and local extinction of ecosystem service providing insects such as pollinators or natural enemies of crop pests. The genetic diversity of the remnant populations of beneficial insects in agricultural sites can be a good indicator of their conservation status, however, this has been rarely evaluated. Furthermore, gene flow between agricultural and natural areas may help maintain this genetic diversity, but the directionality of gene flow between natural and agricultural areas has never been assessed. Thus, a pan-European sampling protocol has been carried out by WP3 "Increasing the evidence base for synergies between agriculture and biodiversity", task 3.3 "Drivers and functional importance of genetic diversity of ecosystem service providing insects in agricultural landscapes". Here we provide a report on the main findings.

Key takeaway messages

- The studied bee pollinator populations are genetically structured at the European level indicating that its populations are genetically differentiated across the continent, however, natural enemy populations show no genetic structure and are hence more homogeneous across Europe.
- There is a high genetic connectivity between populations established in crop fields and nearby natural landscapes. As a consequence, we do not found higher levels of genetic diversity in populations from agricultural landscapes with respect to those established in nearby natural areas.
- Demographic models show an increased gene flow from natural to agricultural landscapes. Hence, natural landscapes may serve as genetic reservoirs of both natural enemies and pollinators for surrounding agricultural areas, act as sources for recurrent recolonization and, potentially, contribute to enhance ecosystem services and crop production.

Summary

Very little is known about the genetic diversity and resilience to population fragmentation and local extinction of ecosystem service providing insects such as pollinators or natural enemies of crop pests. We determined the levels of genetic diversity of two model service providing species in nine countries; a key crop pollinator, the solitary bee *Andrena flavipes* and an important natural enemy species of crop pests the ladybird beetle *Coccinella septempunctata*. In each region, at least one of the taxa was sampled across an agricultural land use intensity gradient in order to analyse genetic flow among natural and agricultural areas. Genetic variability was analysed centrally by CSIC, using a standardized genotyping protocol (ddRADseq). Overall we collected 290 specimens. The studied bee pollinator populations are genetically structured at the EU level indicating that its populations are well differentiated within the continent, however, natural enemy populations show no genetic structure. Genetic diversity levels are similar across regions and habitats due to a high genetic connectivity between populations established in crop fields and nearby natural landscapes. Gene flow goes mostly from natural to agricultural areas for both pollinator and pest control studied species, indicating the importance of natural areas as sources for maintaining agricultural populations. Hence, natural landscapes may serve as genetic reservoirs of both natural enemies and pollinators for surrounding agricultural areas, act as sources for recurrent recolonization and, potentially, contribute to enhance ecosystem services and crop production.

List of abbreviations

EU	European Union
GPS	Global Positioning System
AIC	Akaike’s Information Criterion
SFS	Site Frequency Spectrum
FDR	False Discovery Rate
PCA	Principal Component Analysis
FST	Fixation index
SNP	Single Nucleotide Polymorphisms
PCR	Polymerase Chain Reaction
DNA	Deoxyribonucleic Acid
ddRADseq	double-digestion estriction-fragment-based procedure

1 Introduction

Biodiversity conservation is no longer an activity restricted to natural areas. With more than 80% of land under the direct influence of agricultural systems, the focus has shifted on also conserving the wide array of species that can coexists within humanized landscapes (Garibaldi et al. 2021). Indeed, protecting those species is becoming pivotal as a good fraction of this biodiversity is acting as free ecosystem service providers, enhancing crop yield by contributing to pollinate crop plants (Klein et al. 2007) and control pests (Garratt et al. 2011). While advances in the right direction are being done, transitioning to a lower intensity agriculture, and following an ecological intensification approach (Kleijn et al. 2019), most efforts are focused on halting the loss of species. For example, it has been shown that actions such as planting flower strips on field margins can increase pollinator richness and abundance (Scheper et al. 2013). This coarse species diversity metrics have been core to show effective conservation measures are possible, but lack the detail to help us understand better the deeper complexity of biodiversity.

Biodiversity is a multi-faceted concept. Despite consistent claims to value multiple dimensions of biodiversity, most studies up-to-date focus on simple species diversity measures such as richness and abundance. However, biodiversity affects many levels, from community species composition and functional diversity, to species population dynamics and genetic diversity, In particular, genetic diversity is an often neglected aspect with far reaching implications for long-term conservation. More genetically divers populations may better withstand global change pressures, as this populations have a larger genetic pool from which to adapt to new situations. However, the relationships between agricultural land use and the genetic diversity of insect species that provide important ecosystem functions and services are poorly known.

Insect populations providing ecosystem services are constituted by mobile organisms (Kremen et al. 2007). This implies that its conservation does not only rely on the focal habitat quality, but also on the landscape level habitat configuration (Martin et al. 2019). A general pattern observed is that focal crops surrounded by larger proportion of natural habitats tend to host more beneficial insect diversity and abundance, and in turn, ecosystem services such as pollination and pest control are increased (Dainese et al. 2019). However, we still do not know how source and sink process work in agricultural systems. On one hand, agricultural habitats can host a wide variety of species, which may be self-sustainable. On the other hand, some species may not be able to persist in simplified agricultural habitats, and require from immigrants coming from nearby natural areas acting as sources. In any way, the benefits of having this gene flow between natural and agricultural areas can be important, even when populations would be able to persist in isolation. However, how natural areas contribute to agricultural populations genetic diversity is not explored yet.

In task 3.3, we use a pan-European sampling design to characterize the genetic structure and diversity of two key ecosystem service providers, a pollinator species and a pest control agent, across agricultural and natural habitats. In particular, we ask which is the direction of gene flow goes between natural and agricultural sites as a way to better understand the importance of natural habitats for genetic diversity conservation in Europe.

Main Objective: Evaluating the direction of the Ecosystem Services providers' gene flow between natural and agricultural areas across the EU.

2 Methods

2.1 Population sampling

In spring 2021, we sampled pairs of populations of *C. septempunctata* and *A. flavipes* in agricultural areas and nearby natural/seminatural landscapes separated by <25 km (mean: 11.9 km; range: 4.6-24.0 km; Table 1). Specimens were collected in patches of native vegetation in natural/seminatural plots and within crop fields, paths, and field margins in agricultural areas. This sampling scheme was replicated in six countries for *C. septempunctata* and four countries in *A. flavipes* (Table 1). Note that we enlarged the number of partner/countries involved in this task from the initial six countries listed in the proposal, to nine countries because more partners were willing to collect specimens. One of these partners (CEG: Center for Ecological Genetics, Arhus University), located in Denmark, is not part of the SHOWCASE project, but participates in a related nationally funded project with similar aims (Sustainscapes) and was interested in being part of this task and funded their contribution through their own resources. We targeted to sample and genotype 8-10 specimens per species and plot, but small number of individuals in some areas often reduced sample sizes. This was especially the case of *A. flavipes*, which could not be collected in some countries due to unusual low numbers in the sampling year (Table 1). Specimens were collected by hand (*C. septempunctata*) or sweep-netting (*A. flavipes*) and preserved at -20 °C in 1,500 µL ethanol 96% until needed for genomic analyses. We registered spatial coordinates using a Global Positioning System (GPS). Further details on sampling locations are provided in Table 1. We selected two large sites per country (Table 1), separated ~8-10 km. One site is located in a natural area and another in an agricultural area. Each site can be an area of ~ 2 km radius. History matters, so we carefully checked that the natural and agricultural areas have been present for > 100 years. Within each site, we walked freely along the 2 km radius area to find the target species (see below). Species in the natural areas were collected in nice patches of native vegetation. Species in agricultural areas could be caught within the crop fields or on paths, field margins, and even small semi-natural areas within the agricultural area. We spread the collections across the site area and on different visits to avoid sampling closely related individuals (i.e., sibling or half-siblings). That is, we did not sample two individuals simultaneously from the same place and time.

2.2 Genomic library preparation and processing

We extracted and purified DNA from each sampled specimen used NucleoSpin Tissue kits (Macherey-Nagel, Düren, Germany). We processed genomic DNA into different genomic libraries using the double-digestion restriction-fragment-based procedure (ddRADseq) described in Peterson et al. (2012). In brief, we digested DNA with the restriction enzymes MseI and EcoRI (New England Biolabs, Ipswich, MA, USA) and ligated Illumina adaptors including unique 7-bp barcodes to the digested fragments of each individual. We pooled ligation products and size-selected them between 350-450 bp with a Pippin Prep machine (Sage Science, Beverly, MA, USA). We amplified the fragments by PCR with 12 cycles using the iProofTM High-Fidelity DNA Polymerase (BIO-RAD, Veenendaal, The Netherlands) and sequenced the library in a single-read 201-bp lane on an Illumina NovaSeq6000 platform at the Genomics Core Facility from Centro Andaluz de Biología Molecular & Medicina Regenerativa (CABIMER, Seville, Spain). We used the different programs distributed as part of the stacks v. 1.35 pipeline (Catchen et al,. 2013) to filter and assemble our sequences into de novo loci, call genotypes, calculate genetic diversity statistics, and export input files for all downstream analyses. We exported only the first SNP per RAD locus, and retained loci with a minimum stack depth ≥ 5 (m = 5) and that were represented in at least 75% of the populations (p = 9 for *C. septempunctata* and p = 6 for *A. flavipes*) and 50% of the individuals within each population (r = 0.5).

2.3 Quantifying genetic structure

We calculated genetic differentiation (FST) between each pair of populations in arlequin v. 3.5 (Excoffier & Lischer, 2010), determining their statistical significance with Fisher’s exact tests after 10 000 permutations. We applied a false discovery rate (FDR) adjustment (5%, q < 0.05) to control for multiple tests. We quantified population genetic structure and admixture using the Bayesian Markov chain Monte Carlo clustering method implemented in the program STRUCTURE v. 2.3.3 (Pritchard et al., 2000). We conducted STRUCTURE analyses hierarchically, initially analysing data from all populations of each taxon jointly and, subsequently, running independent analyses for subsets of populations assigned to the same genetic cluster in the previous hierarchical level analysis (e.g., González-Serna et al., 2020). To make analyses computationally tractable, we ran STRUCTURE using a single random subset of 10,000 unlinked SNPs. We ran STRUCTURE analyses assuming correlated allele frequencies and admixture and without using prior population information (Hubisz et al., 2009). We conducted 15 independent runs for each value of K (from K = 1 to K = 10) to estimate the most likely number of genetic clusters with 200,000 MCMC cycles, following a burn-in step of 100,000 iterations. We retained the ten runs having the highest likelihood for each value of K and determined the number of genetic clusters that best describes our data according to log probabilities of the data (LnPr(X|K) ; Pritchard et al., 2000) and the ΔK method (Evanno et al., 2005), as implemented in STRUCTURE harvester (Earl & vonHoldt, 2012). We used clumpp v. 1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same K value (Jakobsson & Rosenberg, 2007) and distruct v. 1.1 (Rosenberg, 2004) to visualize the individuals´ probabilities of population membership in bar plots. Complementarily, we performed a principal component analysis (PCA) as implemented in the R v. 4.2.1 (R Core Team, 2022) package adegenet (Jombart, 2008). Before running PCAs, we replaced missing data by the mean frequency of the corresponding allele estimated across all samples (Jombart, 2008).

2.4 Testing alternative models of gene flow

We used the simulation-based approach implemented in fastsimcoal2 to evaluate alternative migration models and test the hypothesis of asymmetric gene flow between populations established in agricultural areas and nearby natural/seminatural landscapes (Figure 1). Specifically, we tested four alternative demographic models for each taxon (*C. septempunctata* and *A. flavipes*) and spatial replicate (i.e., country), including a null-model considering no gene flow between populations from agricultural and natural/seminatural landscapes (Model A), a model of symmetric gene flow (Model B), and two models of unidirectional gene flow, one fitting unidirectional gene flow from agricultural to natural/seminatural landscapes (Model C) and another exclusively considering unidirectional gene flow in the opposite direction (Model D) (Figure 1B). We used the site frequency spectrum (SFS) and fastsimcoal2 to estimate the composite likelihood of the observed data given a specified model (Excoffier et al., 2013). For each taxon and pair of populations, we calculated a folded joint site frequency spectrum (SFS) considering a single SNP per locus to avoid the effects of linkage disequilibrium. Because we did not include invariable sites in the SFS, we fixed the contemporary effective population size for one of the demes (θAGR; Figure 1B) to enable the estimation of other parameters in fastsimcoal2 (Excoffier et al., 2013). The effective population size fixed in the model was calculated from the level of nucleotide diversity (π) and estimates of mutation rate per site per generation (μ). Nucleotide diversity (π) was estimated from polymorphic and non-polymorphic loci using stacks (Table 1). We used the mutation rate per site per generation (2.8 × 10-9) estimated for Drosophila melanogaster (Keightley et al., 2014), which is similar to the spontaneous mutation rate estimated for the butterfly Heliconius melpomene (2.9 × 10−9; Keightley et al., 2015). To remove all missing data for the calculation of the SFS, minimize errors with allele frequency estimates, and maximize the number of retained SNPs, each population group was downsampled using a Python script written by Qixin He and available on Dryad (Papadopoulou & Knowles, 2015).

Each model was run 100 independent times considering 100,000-250,000 simulations for the calculation of the composite likelihood, 10-40 expectation-conditional maximization (ECM) cycles, and a stopping criterion of 0.001 (Excoffier & Foll, 2011; Excoffier et al., 2013). We used an information-theoretic model selection approach based on the Akaike’s information criterion (AIC) to determine the probability of each model given the observed data (Burnham & Anderson, 2002; e.g., Nater et al., 2015; Thome & Carsterns, 2016). After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores as detailed in Thome & Carsterns (2016). AIC values for each model were rescaled (ΔAIC), calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models (i.e., the best model has ΔAIC = 0). Point estimates of the different demographic parameters for the best supported model were selected from the run with the highest maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates (based on the percentile method; e.g., de

Manuel et al., 2016) from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al. 2013).

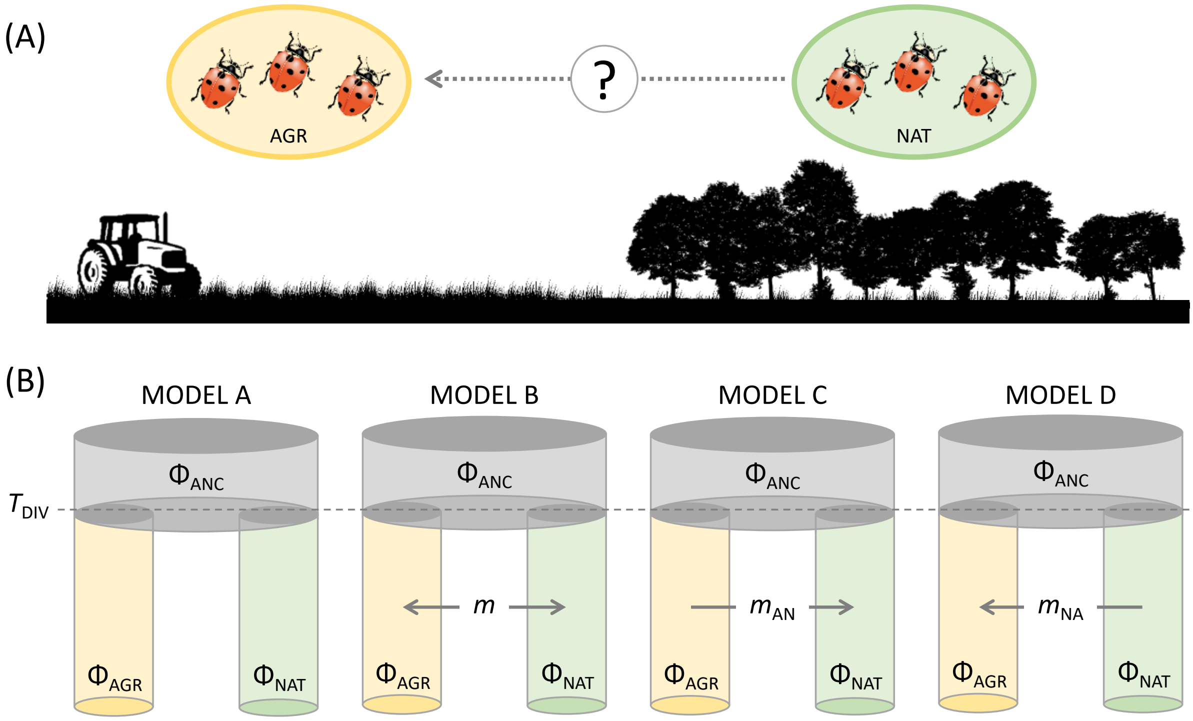

FIGURE 1 (A) Schematic illustrating the hypothesis of unidirectional gene flow from natural/seminatural habitats to nearby agricultural areas for populations of one natural enemy (*Coccinella septempunctata*; shown in panel A) and one pollinator (*Andrena flavipes*). Panel (B) shows the four alternative gene flow models tested using FASTSIMCOAL2. Parameters include mutation-scaled ancestral effective population sizes (θ_{ANC}), contemporary effective population sizes for populations from agricultural areas (θ_{AGR}) and natural/seminatural habitats (θ_{NAT}), timing of population split (T_{DIV}), and symmetric (m) and unidirectional (m_{AN} and m_{NA}) migration rates.

3 Results

Overall, seven countries sampled *C. septempunctata* and five *Andrena flavipes*, with three countries sampling both species (Table 1). In Portugal, unusual low numbers of both species precluded sampling enough specimens, but given the high success collecting specimens in other countries, this is not a problem in our experimental design.

3.1 Genomic data

The average number of reads retained per individual after the different quality filtering steps was 3,409,688 (range= 236,373-9,557,006 reads) for *C. septempunctata* and 2,103,765 (range= 420,468-6,255,835 reads) for *A. flavipes*. On average, this represented 86% (range = 64-90%) and 73% (range = 60-81%) of the total number of reads recovered for each individual of *C. septempunctata* and *A. flavipes*, respectively. After filtering loci, the final datasets including all populations of *C. septempunctata* and *A. flavipes* contained 10,896 and 18,554 SNPs, respectively.

3.2 Genetic diversity of populations

Levels of genetic diversity in populations of *C. septempunctata* and *A. flavipes* are presented in Table 1. In *C. septempunctata*, individual genetic diversity (i.e., observed heterozygosity) did not significantly differ among countries (one-way ANOVA: F5, 96 = 0.99, p = 0.423) or between populations established in agricultural areas and nearby natural/seminatural landscapes in comparisons within each country (all p-values > 0.3). In *A. flavipes*, individual genetic diversity significantly differed among countries (one-way ANOVA: F3, 56 = 5.85, p = 0.002) and post hoc Tukey's tests revealed that such differences resulted from comparisons involving populations from Denmark, which showed lower levels of genetic diversity than populations from the rest of the countries (p < 0.06 in all cases; Table 1). As in *C. septempunctata*, individual genetic diversity did not significantly differ between populations of *A. flavipes* established in agricultural areas and nearby natural/seminatural landscapes in comparisons within each country (all p-values > 0.3).

3.3 Quantifying genetic structure

In *C. septempunctata*, all pairwise FST values were negative and non-significantly different from zero. In contrast, pairwise FST values in *A. flavipes* ranged between -0.023 and 0.444 and several pairs of populations showed significant genetic differentiation. Except comparisons involving populations from the Netherlands and Switzerland, all pairs of populations from different countries were significantly differentiated. Populations established in agricultural areas and nearby natural/seminatural landscapes within each country did not show significant genetic differentiation. The only exception was the comparison involving the two populations from Denmark, albeit with a small FST value (FST = 0.03; Table S2). structure analyses for *C. septempunctata* showed LnPr(X|K) reached a plateau at K = 1 (Figure S3a), indicating lack of genetic structure (Figure 2a). STRUCTURE analyses for *A. flavipes* considering all populations identified the most likely number of clusters as K = 2 according to the ΔK criterion and LnPr(X|K) reached a plateau at the same K value. For K = 2, the two genetic clusters separated the populations from Spain from the rest of European populations. STRUCTURE analyses at a lower hierarchical level and excluding Spanish populations identified the most likely number of clusters as K = 2 according to the ΔK criterion and LnPr(X|K) reached a plateau at the same K value. For K = 2, the two genetic clusters separated the populations from Denmark from the populations from the Netherlands and Switzerland, which grouped together. Principal component analysis (PCA) confirmed the results yielded by STRUCTURE at the different hierarchical levels (Figure 3). For *C. septempunctata*, PCA did not reveal any genetic clustering and genotyped individuals collected in different localities were interspersed across the two first principal components (Figure 3a). For *A. flavipes*, PCA separated Iberian populations from the rest of European populations along the first principal component (PC1) (Figure 3b). The second principal component (PC2) separated the populations from Denmark from those sampled in the Netherlands and Switzerland, which grouped together (Figure 3b).

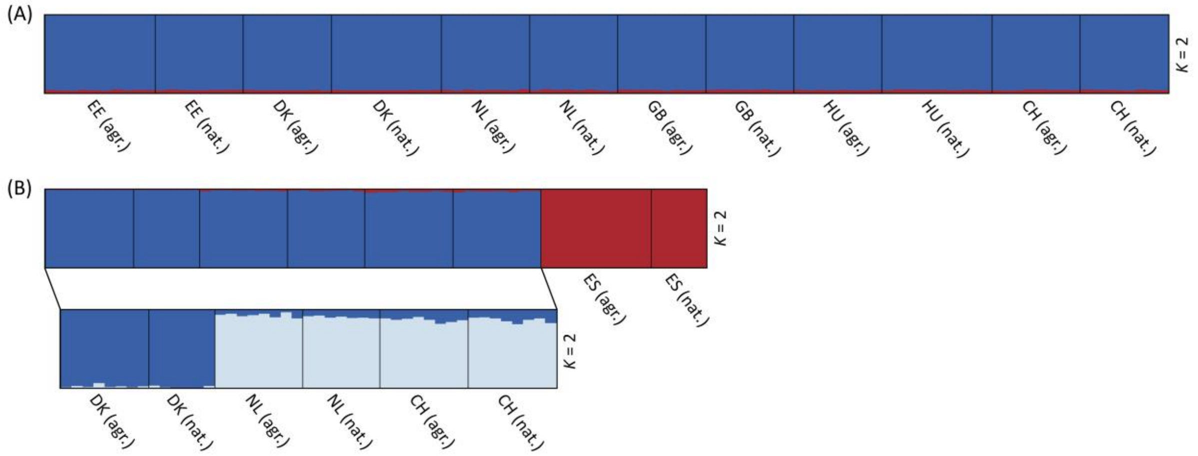

FIGURE 2 Results of genetic assignments based on the Bayesian method implemented in the program STRUCTURE for populations of (A) *Coccinella septempunctata* and (B) *Andrena flavipes* sampled in agricultural areas (agr.) and nearby natural/seminatural (nat.) landscapes from different countries. Analyses are based on a random subset of 10,000 SNPs. Each individual is represented by a vertical bar, which is partitioned into *K* colored segments showing the individual's probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different populations. Population codes as described in Table 1.

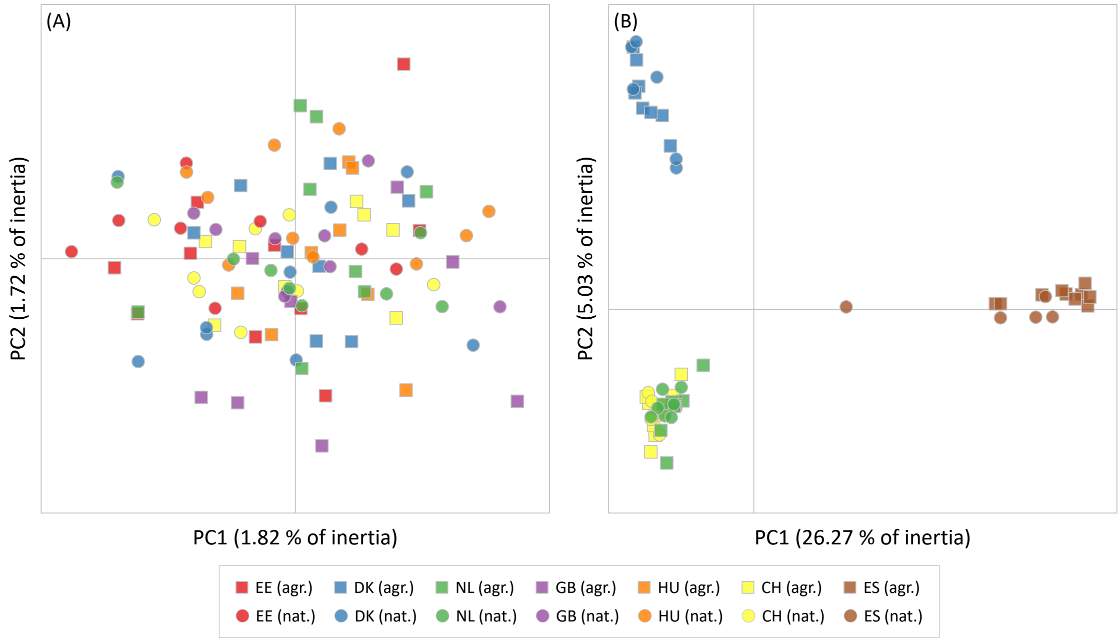

FIGURE 3 Principal component analyses (PCA) of genetic variation for populations of (A) *Coccinella septempunctata* and (B) *Andrena flavipes* sampled in agricultural areas (agr.; squares) and nearby natural/seminatural landscapes (nat.; circles) from different countries. Analyses are based on random subset of 10,000 SNPs. Population codes as described in Table 1.

3.4 Testing alternative models of gene flow

Demographic analyses in fastsimcoal2 supported unidirectional gene flow from populations established in natural/seminatural landscapes to those from agricultural areas (Model D; Figure 1B) in five *C. septempunctata* (EE, DK, NL, HU and CH) and two *A. flavipes* (NL, ES) pairwise comparisons (Table 2). In the rest of the comparisons (GB in *C. septempunctata*, and DK and CH in *A. flavipes*), the most supported model was a scenario of unidirectional gene flow from populations established in agricultural areas to those from nearby natural/seminatural landscapes (Model C; Table 2). In the case of populations from Great Britain (GB) of *C. septempunctata*, the second most supported model (Model D) was statistically indistinguishable from the most supported model (Model C) (ΔAIC = 0.79), indicating that the two alternative models of unidirectional gene flow are similarly probable (Table 2). Estimates of gene flow between populations were fairly similar across all pairwise comparisons, ranging from 2.35 × 10-4 to 5.79 × 10-4 in *C. septempunctata* and from 2.27 × 10-4 to 5.67 × 10-4 in *A. flavipes* (Table 3). In both *C. septempunctata* and *A. flavipes*, estimates of gene flow from populations established in agricultural areas to those from nearby natural/seminatural landscapes (i.e., Model C) were smaller than estimates obtained when the most supported model was the scenario of unidirectional gene flow in the opposite direction (i.e., Model D) (Table 3).

TABLE 1 Geographical location, number of genotyped individuals (*n*), and genetic diversity statistics (π, H_O, and H_E) for populations of *Coccinella septempunctata* and *Andrena flavipes* sampled in agricultural areas and nearby natural/seminatural landscapes. Genetic diversity statistics were calculated in STACKS for all positions (polymorphic and nonpolymorphic) and only variant (polymorphic) positions. Average values across loci are presented for nucleotide diversity (π), and observed (H_O) and expected (H_E) heterozygosity.

						All positions			Variant positions		
Country	Code	Site	Latitude	Longitude	*n*	π	H_O	H_E	π	H_O	H_E
(a) *Coccinella septempunctata*											
Estonia	EE	Agricultural	58.82436	23.59144	10	0.0003	0.0002	0.0003	0.0541	0.0299	0.0504
Estonia	EE	Natural	58.79168	23.51234	8	0.0003	0.0002	0.0003	0.0537	0.0298	0.0489
Denmark	DK	Agricultural	56.29121	10.40226	8	0.0003	0.0002	0.0003	0.0539	0.0318	0.0489
Denmark	DK	Natural	56.27138	10.46697	10	0.0003	0.0002	0.0003	0.0551	0.0310	0.0514
Netherlands	NL	Agricultural	50.79752	5.79459	8	0.0003	0.0002	0.0003	0.0564	0.0309	0.0517
Netherlands	NL	Natural	50.82924	5.87514	8	0.0003	0.0002	0.0003	0.0546	0.0311	0.0497
Great Britain	GB	Agricultural	50.65034	-2.28968	8	0.0003	0.0002	0.0003	0.0527	0.0288	0.0482
Great Britain	GB	Natural	50.64681	-2.00065	8	0.0003	0.0002	0.0003	0.0557	0.0308	0.0509
Hungary	HU	Agricultural	46.93410	19.10340	8	0.0003	0.0002	0.0003	0.0541	0.0297	0.0495
Hungary	HU	Natural	47.11210	19.28140	10	0.0003	0.0002	0.0003	0.0565	0.0309	0.0526
Switzerland	CH	Agricultural	47.06575	7.24068	8	0.0003	0.0002	0.0003	0.0532	0.0326	0.0486
Switzerland	CH	Natural	47.19120	7.31626	8	0.0003	0.0002	0.0003	0.0541	0.0308	0.0493
(b) *Andrena flavipes*											
Denmark	DK	Agricultural	56.29121	10.40226	8	0.0003	0.0003	0.0003	0.0608	0.0493	0.0558
Denmark	DK	Natural	56.27138	10.46697	6	0.0003	0.0003	0.0003	0.0555	0.0482	0.0505
Netherlands	NL	Agricultural	50.79752	5.79459	8	0.0004	0.0003	0.0004	0.0711	0.0557	0.0654
Netherlands	NL	Natural	50.82924	5.87514	7	0.0004	0.0003	0.0004	0.0738	0.0599	0.0672
Switzerland	CH	Agricultural	47.06575	7.24068	8	0.0004	0.0004	0.0004	0.0771	0.0652	0.0710
Switzerland	CH	Natural	47.02291	7.31958	8	0.0004	0.0003	0.0004	0.0777	0.0625	0.0716
Spain	ES	Agricultural	37.32987	-6.19802	10	0.0004	0.0003	0.0004	0.0800	0.0572	0.0747
Spain	ES	Natural	37.23660	-6.18450	5	0.0004	0.0003	0.0004	0.0766	0.0561	0.0662

TABLE 2 Alternative demographic models (detailed in Figure 1b) tested using FASTSIMCOAL2 for pairs of populations of *Coccinella septempunctata* and *Andrena flavipes* sampled in agricultural areas and nearby natural/seminatural landscapes. Best supported scenarios are highlighted in bold. Number of variable SNPs retained to calculate the site frequency spectrum is indicated in parentheses. Population codes as described in Table 1.

Model	lnL	*k*	AIC	ΔAIC	ω_i	lnL	*k*	AIC	ΔAIC	ω_i
(a) *Coccinella septempunctata*										
	Estonia (EE) (3860 SNPs)					Denmark (DK) (5127 SNPs)				
A	-4405.22	3	8816.44	108.65	0.00	-5498.61	3	11003.21	156.03	0.00
B	-4354.34	4	8716.67	8.88	0.01	-5423.18	4	10854.36	7.17	0.03
C	-4352.98	4	8713.97	6.17	0.04	-5425.60	4	10859.20	12.01	0.00
D	**-4349.90**	**4**	**8707.79**	**0.00**	**0.95**	**-5419.59**	**4**	**10847.18**	**0.00**	**0.97**
	Netherlands (NL) (2586 SNPs)					Great Britain (GB) (2933 SNPs)				
A	-2852.51	3	5711.03	79.03	0.00	-3263.30	3	6532.60	47.78	0.00
B	-2813.05	4	5634.09	2.10	0.21	-3239.54	4	6487.07	2.25	0.16
C	-2813.02	4	5634.05	2.05	0.21	**-3238.41**	**4**	**6484.82**	**0.00**	**0.50**
D	**-2812.00**	**4**	**5631.99**	**0.00**	**0.59**	-3238.81	4	6485.62	0.79	0.34
	Hungary (HU) (5822 SNPs)					Switzerland (CH) (2769 SNPs)				
A	-6373.50	3	12753.00	170.70	0.00	-3046.27	3	6098.53	51.44	0.00
B	-6291.05	4	12590.09	7.78	0.02	-3021.12	4	6050.24	3.15	0.15
C	-6289.01	4	12586.01	3.70	0.13	-3021.22	4	6050.44	3.35	0.13
D	**-6287.15**	**4**	**12582.31**	**0.00**	**0.85**	**-3019.54**	**4**	**6047.09**	**0.00**	**0.72**
(b) *Andrena flavipes*										
	Denmark (DK) (3389 SNPs)					Netherlands (NL) (6668 SNPs)				
A	-8192.57	3	16391.14	8.39	0.01	-8193.79	3	16393.59	161.97	0.00
B	-8191.80	4	16391.60	8.85	0.01	-8115.47	4	16238.93	7.31	0.02
C	**-8187.38**	**4**	**16382.75**	**0.00**	**0.97**	-8113.91	4	16235.81	4.19	0.11
D	-8192.53	4	16393.06	10.31	0.01	**-8111.81**	**4**	**16231.62**	**0.00**	**0.87**
	Switzerland (CH) (7492 SNPs)					Spain (ES) (5623 SNPs)				
A	-8787.18	3	17580.35	169.26	0.00	-5389.38	3	10784.75	103.39	0.00
B	-8703.52	4	17415.04	3.95	0.10	-5339.27	4	10686.54	5.17	0.06
C	**-8701.55**	**4**	**17411.09**	**0.00**	**0.75**	-5339.14	4	10686.27	4.91	0.07
D	-8703.21	4	17414.42	3.33	0.14	**-5336.68**	**4**	**10681.36**	**0.00**	**0.86**

lnL, maximum likelihood estimate of the model; *k*, number of parameters in the model; AIC, Akaike's information criterion value; ΔAIC, difference in AIC value from that of the strongest model; ω_i, AIC weight.

TABLE 3 Parameters inferred from coalescent simulations with FASTSIMCOAL2 under the most supported demographic model (Model C or Model D, as indicated in Table 2) for pairs of populations of *Coccinella septempunctata* and *Andrena flavipes* sampled in agricultural areas and nearby natural/seminatural landscapes. Table shows point estimates and lower and upper 95% confidence intervals for each parameter, including mutation-scaled ancestral effective population sizes (θ_{ANC}), contemporary effective population sizes for populations from natural/seminatural habitats (θ_{NAT}), timing of population split (T_{DIV}), and unidirectional migration rates (m_{AN} or m_{NA}, depending on the most supported model; Table 2). Contemporary effective population size for populations from agricultural areas (θ_{AGR}) was calculated from their respective levels of nucleotide diversity (π; Table 1) and fixed in FASTSIMCOAL2 analyses to enable the estimation of other parameters (see the Materials and Methods section for further details). Superscripts indicate the most supported model (Model C or Model D, detailed in Figure 1b). Population codes as described in Table 1.

	θ_{ANC}		θ_{NAT}		T_{DIV}		m_{NA} / m_{AN}	
Code	Point	95% CI	Point	95% CI	Point	95% CI	Point	95% CI
(a) *Coccinella septempunctata*								
EE^{D}	24,889	19,358 – 33,796	6579	5637 – 7291	6661	5626 – 8002	5.57×10^{-4}	$4.17 \times 10^{-4} – 7.08 \times 10^{-4}$
DK^{D}	82,132	67,815 – 109,399	8191	6885 – 8985	10,385	9313 – 12,554	2.81×10^{-4}	$2.34 \times 10^{-4} – 3.41 \times 10^{-4}$
NL^{D}	29,795	22,444 – 40,894	5519	4383 – 6360	7090	5913 – 8789	5.34×10^{-4}	$4.03 \times 10^{-4} – 7.31 \times 10^{-4}$
GB^{C}	113,536	88,194 – 169,780	13077	10278 – 16,041	10,098	7670 – 13,554	2.35×10^{-4}	$1.95 \times 10^{-4} – 2.91 \times 10^{-4}$
HU^{D}	61,597	51,126 – 75,089	8179	6804 – 9029	9830	8959 – 11,936	2.76×10^{-4}	$2.35 \times 10^{-4} – 3.12 \times 10^{-4}$
CH^{D}	21,203	13,913 – 30,169	5419	4551 – 6369	4445	3927 – 5945	5.79×10^{-4}	$3.90 \times 10^{-4} – 8.21 \times 10^{-4}$
(b) *Andrena flavipes*								
DK^{C}	51,303	20,647 – 66,866	118,487	59,169 – 126,548	8338	1857 – 17,100	2.27×10^{-4}	$1.67 \times 10^{-4} – 2.86 \times 10^{-4}$
NL^{D}	39,466	31,734 – 51,439	13,751	11,544 – 14,635	15,015	12,642 – 18,271	4.79×10^{-4}	$3.91 \times 10^{-4} – 5.73 \times 10^{-4}$
CH^{C}	127,634	99,263 – 170,344	18,113	13,911 – 21,774	12,537	10,399 – 16,085	3.74×10^{-4}	$3.04 \times 10^{-4} – 4.42 \times 10^{-4}$
ES^{D}	20,900	15,414 – 26,253	4381	3832 – 4712	4847	4376 – 5551	5.67×10^{-4}	$4.46 \times 10^{-4} – 7.42 \times 10^{-4}$

4 Discussion

Our data supported contrasting patterns of genetic structure and gene flow in the two focal taxa, the natural enemy *C. septempunctata* and the pollinator *A. flavipes*. A panmictic structure characterizes populations of *C. septempunctata* across both local and regional scales, whereas populations of *A. flavipes* present a marked structure at a European scale but a very limited or null genetic differentiation between nearby habitat patches within landscape replicates. We found no support for the hypothesis of higher levels of genetic diversity in populations from natural/seminatural landscapes with respect to those stablished in agricultural areas. However, demographic model testing and coalescent-based estimates of migration rates lent support to the hypothesis of increased gene flow from natural to agricultural landscapes (Figure 1a): models considering anisotropic gene flow in the predicted direction received comparatively higher statistical support than the rest of the scenarios across most landscape replicates for *C. septempunctata*, although evidence on this respect for *A. flavipes* was limited to 50% of our pairwise comparisons.

Genetic diversity and structure at regional and local scales
Analyses of genetic differentiation (FST) and structure (STRUCTURE and PCAs) revealed marked differences in the spatial distribution of genomic variation in the two focal taxa (Figures 2-3). Populations of *C. septempunctata* were panmictic and presented no evidence for genetic differentiation at both European and local scales, a phenomenon likely explained by frequent long-distance migration behaviour linked to seasonal aggregations charactering many Coccinellids (Hagen 1962). Long-distance dispersal in this taxon is exemplified by unrestricted gene flow between British and mainland European populations, which indicates the limited effects of seawater as a barrier to dispersal. These results are in line with the findings of a previous microsatellite-based study covering the entire distribution of the species (Lecompte et al., 2016). This study revealed lack of genetic structure, very limited differentiation (FST < 0.05), and a weak pattern of isolation-by-distance across populations sampled from Portugal to India. Only eastern Asia (China, Japan) and North Africa populations (Algeria) were assigned to different genetic clusters and presented a significant genetic differentiation (FST = 0.10-0.30), which suggests either long-term isolation of these peripheral populations (e.g., in different glacial refugia) or the evolution of reproductive isolation and incipient processes of cryptic speciation (Lecompte et al., 2016). In contrast, analyses for *A. flavipes* revealed a marked genetic structure and genetic differentiation (FST = 0.10-0.44) at a European scale, with more peripheral populations from Jutland (DK) and Iberian (ES) peninsulas forming distinct genetic clusters with limited gene flow between them as well as with the rest of analysed populations (Figures 2-3). Yet, central European populations (NL and CH; separated by >400 km) clustered together and populations from agricultural and natural/seminatural areas within landscape replicates revealed very low levels of genetic differentiation (FST < 0.03). Only the two sites within Denmark showed significant differentiation, albeit with a very low estimate (FST = 0.03). Such weak genetic structure at local/regional spatial scales has been previously reported in other Andrenidae (*Andrena fuscipes*: Exeler et al., 2010; *Andrena vaga*: Exeler et al., 2008; Černá et al., 2013), bumblebees (*Bombus terrestris*; Silva et al., 2020), stingless bees (*Trigona spinipes*; Jaffé et al., 2016), and carpenter bees (*Xylocopa virginica*; Ballare and Jha, 2020) and is in line with the findings of spatially-explicit landscape genetic studies suggesting a limited impact of habitat fragmentation and structure on the distribution of spatial patterns of genetic variation in different pollinators (Exeler et al., 2010; Jha & Kremen, 2013; Jackson et al., 2018; Barbosa et al. 2022). Aligned with inferences from analyses of genetic structure, testing of alternative demographic models strongly rejected the scenario of strict isolation (i.e., lack of gene flow; Model A) between populations from agricultural and nearby natural/seminatural areas in the two taxa. High levels of gene flow between nearby habitat patches could prevent genetic drift and loss of genetic diversity and explain similar levels of genetic variation across populations of the two taxa.

Anisotropic gene flow between agricultural and natural landscapes
Our model-based approach provided mixed support to the hypothesis of directional gene flow from populations established in natural/seminatural landscapes to those sustained in nearby agricultural areas (Figure 1). Coalescent-based model testing the natural enemy *C. septempunctata* supported unidirectional gene flow in the predicted direction (Model D) across most landscape replicates, suggesting that natural habitat patches genetically subsidize conspecific populations established in nearby crop fields. Although analyses in *A. flavipes* consistently supported models fitting asymmetric migration rates over those considering symmetric gene exchange, the scenario of unidirectional gene flow from agricultural to natural landscapes was the most supported only in a half of our spatial replicates. Remarkably, the

bidirectional migration model received in all cases a low statistical support, suggesting that asymmetric gene flow and source-sink dynamics govern the demographic dynamics of the focal taxa in natural-agricultural mosaic landscapes.

Conclusions

Collectively, our results support a high genetic connectivity between populations established in crop fields and nearby natural landscapes, indicating that the latter may serve – at least in some cases – as reservoirs of both natural enemies and pollinators for surrounding agricultural areas, act as sources for recurrent recolonization and, potentially, contribute to enhance ecosystem services and crop production.

5 Policy recommendations

We highlight the importance of conserving complex agricultural landscapes which embed natural areas within productive regions, as those can enhance gene flow, increasing the genetic diversity of agricultural populations.

6 Acknowledgements

We would like to thank all data contributors, and lab personnel.

7 References

Ballare, K. M., & Jha, S. (2021). Genetic structure across urban and agricultural landscapes reveals evidence of resource specialization and philopatry in the Eastern carpenter bee, Xylocopa virginica L. Evolutionary Applications, 14(1), 136-149. doi:10.1111/eva.13078

Barbosa, M. D., Jaffe, R., Carvalho, C. S., Lanes, E. C. M., Alves-Pereira, A., Zucchi, M. I., . . . Alves, D. A. (2022). Landscape influences genetic diversity but does not limit gene flow in a Neotropical pollinator. Apidologie, 53(4). doi:10.1007/s13592-022-00955-0

Cerna, K., Straka, J., & Munclinger, P. (2013). Population structure of pioneer specialist solitary bee Andrena vaga (Hymenoptera: Andrenidae) in central Europe: the effect of habitat fragmentation or evolutionary history? Conservation Genetics, 14(4), 875-883. doi:10.1007/s10592-013-0482-y

Dainese, M., Martin, E.A., Aizen, M.A., Albrecht, M., Bartomeus, I., Bommarco, R., Carvalheiro, L.G., Chaplin-Kramer, R., Gagic, V., Garibaldi, L.A. and Ghazoul, J., 2019. A global synthesis reveals biodiversity-mediated benefits for crop production. Science advances, 5(10), p.eaax0121.

Exeler, N., Kratochwil, A., & Hochkirch, A. (2008). Strong genetic exchange among populations of a specialist bee, Andrena vaga (Hymenoptera : Andrenidae). Conservation Genetics, 9(5), 1233-1241. doi:10.1007/s10592-007-9450-8

Exeler, N., Kratochwil, A., & Hochkirch, A. (2010). Does recent habitat fragmentation affect the population genetics of a heathland specialist, Andrena fuscipes (Hymenoptera: Andrenidae)? Conservation Genetics, 11(5), 1679-1687. doi:10.1007/s10592-010-0060-5

Garibaldi, L.A., Oddi, F.J., Miguez, F.E., Bartomeus, I., Orr, M.C., Jobbágy, E.G., Kremen, C., Schulte, L.A., Hughes, A.C., Bagnato, C. and Abramson, G., 2021. Working landscapes need at least 20% native habitat. Conservation Letters, 14(2), p.e12773.

Garratt, M.P.D., Wright, D.J. and Leather, S.R., 2011. The effects of farming system and fertilisers on pests and natural enemies: a synthesis of current research. Agriculture, Ecosystems & Environment, 141(3-4), pp.261-270.

Hagen, K. S. (1962) Biology and Ecology of Predaceous Coccinellidae. Annual Review of Entomology, 7, 289-326

Jackson, J. M., Pimsler, M. L., Oyen, K. J., Koch-Uhuad, J. B., Herndon, J. D., Strange, J. P., . . . Lozier, J. D. (2018). Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. Molecular Ecology, 27(14), 2926-2942. doi:10.1111/mec.14735

Jaffe, R., Castilla, A., Pope, N., Imperatriz-Fonseca, V. L., Metzger, J. P., Arias, M. C., & Jha, S. (2016). Landscape genetics of a tropical rescue pollinator. Conservation Genetics, 17(2), 267-278. doi:10.1007/s10592-015-0779-0

Jha, S., & Kremen, C. (2013). Urban land use limits regional bumble bee gene flow. Molecular Ecology, 22(9), 2483-2495. doi:10.1111/mec.12275

Kleijn, D., Bommarco, R., Fijen, T.P., Garibaldi, L.A., Potts, S.G. and Van Der Putten, W.H., 2019. Ecological intensification: bridging the gap between science and practice. Trends in ecology & evolution, 34(2), pp.154-166.

Klein, A.M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C. and Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proceedings of the royal society B: biological sciences, 274(1608), pp.303-313.

Kremen, C., Williams, N.M., Aizen, M.A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., Packer, L., Potts, S.G., Roulston, T.A., Steffan-Dewenter, I. and Vázquez, D.P., 2007. Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. Ecology letters, 10(4), pp.299-314.

Lecompte, E., Bouanani, M. A., Magro, A., & Crouau-Roy, B. (2016). Genetic diversity and structuring across the range of a widely distributed ladybird: focus on rear-edge populations phenotypically divergent. Ecology and Evolution, 6(15), 5517-5529. doi:10.1002/ece3.2288

Martin, E.A., Dainese, M., Clough, Y., Báldi, A., Bommarco, R., Gagic, V., Garratt, M.P., Holzschuh, A., Kleijn, D., Kovács-Hostyánszki, A. and Marini, L., 2019. The interplay of landscape composition and configuration: new pathways to manage functional biodiversity and agroecosystem services across Europe. Ecology letters, 22(7), pp.1083-1094.

Scheper, J., Holzschuh, A., Kuussaari, M., Potts, S.G., Rundlöf, M., Smith, H.G. and Kleijn, D., 2013. Environmental factors driving the effectiveness of European agri-environmental measures in mitigating pollinator loss–a meta-analysis. Ecology letters, 16(7), pp.912-920.

Silva, S. E., Seabra, S. G., Carvalheiro, L. G., Nunes, V. L., Marabuto, E., Mendes, R., . . . Paulo, O. S. (2020). Population genomics of Bombus terrestris reveals high but unstructured genetic diversity in a potential glacial refugium. Biological Journal of the Linnean Society, 129(2), 259-272. doi:10.1093/biolinnean/blz182