Mapping the biodiversity of today to protect it for the future - The influence of the environment on biodiversity, and its use in conservation -

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Julia Geue

aus Stuttgart

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	18.12.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	PD Dr. Henri Thomassen
2. Berichterstatter:	Prof. Dr. Oliver Bossdorf

Erklärung

Zulassung als Doktorandin im Sinne von §7 der Promotionsordnung vom 24. April 2015 erfolgte am 04. April 2016. Diese Dissertation wurde im Sinne von §6 von PD Dr. Henri Thomassen betreut.

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass diese Dissertation von mir selbstständig – abgesehen von der Beratung und Hilfe meines Betreuers – und ohne unerlaubte Hilfsmittel erarbeitet wurde.
Andere als die angegebenen Quellen und Hilfsmittel wurden nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen wurden als solche kenntlich gemacht.

Diese Dissertation wurde an keiner anderen Prüfungsbehörde vorgelegt.

Tübingen, 26.10.2020

.....
(Julia Geue)

Dissertation eingereicht am: 26.10.2020
1. Gutachter: PD Dr. Henri Thomassen
2. Gutachter: Prof. Dr. Oliver Bossdorf

"It is that range of biodiversity that we must care for – the whole thing – rather than just one or two stars."

David Attenborough

Acknowledgements

First and foremost, I want to express my deepest gratitude to Henri. Thank you for giving me the opportunity to do my PhD with you, for motivating me to learn new things and helping me to gain more confidence in myself. You always supported me with your advice and patience and believed in me, even in times when I doubted myself the most. You have not only been a great mentor and supervisor, but also a great listener and friend. Thank you for all the opportunities you gave me to join you at conferences and meetings. I am very grateful for seeing all the new places and the fun times we had during fieldwork.

I wish to thank Oliver for the financial support in the last stretch of my PhD and for agreeing to be my second reviewer. Katharina, thank you so much for supporting me financially and always being on hand to offer me good advice when I lost my way. I want to extend my gratitude to the Comparative Zoology group who have helped and supported me in one way or the other over the past years. *Linda, Nicole, Marcel, Bianca, Joscha, Melli, Martina und Paula, Ich habe unsere Kaffeepausen und vor allem den berühmt-berüchtigten Cappuccino-Donnerstag sehr genossen. Ich habe mich immer sehr wohl mit euch gefühlt und werde euch wirklich sehr vermissen.*

My thanks go to all the Students helping me in one way or the other with this work. Especially, Paula and Marcel, who performed an excellent Masters thesis which led to many engaging scientific discussions. Thanks Paula, for the fun time in Romania during our fieldwork (especially when we accidently sank 'Erwin' in the mud).

Many thanks go to my colleagues in Switzerland at the University of Fribourg, who introduced me to the wonderful intricacies of bioinformatics, which I used in my fourth chapter. Thank you Daniel, Madleina, Vivian, Marco and Liam.

I would like to acknowledge the "Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg" for the financial support by the "Landesgraduiertenförderung" and the Teufel-Stiftung, that also made it possible for me to attend several courses and conferences.

Furthermore, I am very grateful to all my friends who supported and encouraged me all the way. Ellie, you have become such a close friend, and always believed in me. This gave me immense support and I want to thank you for that! The "Bären-girls", Deike, Julia, Anna and Gillian. I would not have liked to miss our regular evenings in the Bären with swabian food and yummy Aperol Spritz and our fun times during the PhD retreats. Deike, unfortunately I had to finish my PhD at a time when you had already gone, but especially during the first years of my PhD I was always happy to have you as a colleague and friend. We supported each other and had a lot of fun. Julia, I really enjoyed our time together (especially our festival experience), thank you for listening to all my stories and giving great advice. Anna, I want to thank you for

all the nice evenings we had this last summer and especially our regular “Tübinger Feierabend” together was unforgettable. Thank you for being there and I am glad we could support each other in the last stretches of our PhDs. And lastly, Gillian thank you for all the fun times and believing in me and my capability of sounding like an owl.

Ein ganz besonderer Dank geht an meine beste Freundin Judit. Ich konnte immer meine Freude, aber auch meinen Ballast mit dir teilen. Du hast immer an mich geglaubt und mich in Allem unterstützt. Ich will dich nie mehr missen und bin unbeschreiblich dankbar, dich in meinem Leben zu haben.

Natürlich bin ich auch unfassbar dankbar für die uneingeschränkte Liebe und Unterstützung die ich durch meine Familie in all den Jahren erfahren habe. Ihr habt mich mit Rat und Tat unterstützt und immer an mich geglaubt. Ohne euch wäre das Leben nur halb so schön!
Mama, Joachim, Phillip und Desi, Ich habe euch sehr lieb.

Declaration of author contribution

The thesis entitled "Landscape genomics as a useful tool for conservation prioritization: a multi-species study in Eastern Europe" is based on the work I did during my PhD at the University of Tübingen, supervised by PD Dr. Henri Thomassen. In this thesis, Chapters I–IV include four independent scientific manuscripts, each with co-authors, and has been (or will be) published. The contribution of the authors for each chapter is stated as following:

Chapter I

Julia C. Geue, Henri A. Thomassen: UNRAVELLING THE HABITAT PREFERENCES OF TWO CLOSELY RELATED BUMBLE BEE SPECIES IN EASTERN EUROPE

Status in publication process: Published in Ecology and Evolution, Volume 6, Issue 22. DOI: 10.1002/ece3.6232

JG and HT planned and designed the research; JG and HT conducted fieldwork; JG and HT analyzed the data and wrote the manuscript.

Chapter II

Julia C. Geue, Paula J. Rotter, Caspar Gross, Zoltán Benko, István Kovács, Ciprian Fântânâ, Judit Veres-Szászka, Cristi Domsa, Emanuel Baltag, Szilárd J. Daróczi, Gábor M. Bóné, Viorel D. Popescu and Henri A. Thomassen: LIMITED RECIPROCAL SURROGACY OF BIRD AND HABITAT DIVERSITY AND INCONSISTENCIES IN THEIR REPRESENTATION IN ROMANIAN PROTECTED AREAS

Status in publication process: submitted to Diversity and Distributions

JG and HT planned and designed the research; ZB, IK, CF, JV, CD, EB, SD and GB provided data; JG, PR and HT analyzed the data with the support of CG; JG, PR and HT wrote the manuscript in collaboration with VP.

Chapter III

Julia C. Geue, Csongor I. Vágási, Mona Schweizer, Peter L. Pap, Henri A. Thomassen: ENVIRONMENTAL SELECTION IS A MAIN DRIVER OF DIVERGENCE IN HOUSE SPARROWS (*Passer domesticus*) IN ROMANIA AND BULGARIA

Status in publication process: Published in Ecology and Evolution, Volume 10, Issue 11. DOI: 10.1002/ece3.2509.

HT planned and designed the research; JG, MS and HT conducted fieldwork; CV and PP provided additional samples; JG and MS performed laboratory work; JG and HT analyzed the data and wrote the manuscript.

Chapter IV

Julia C. Geue, Madleina Caduff, Vivian Link, Daniel Wegmann and Henri A. Thomassen: SIGNS OF LOCAL ADAPTATIONS IN THE BUFF-TAILED BUMBLE BEE (*Bombus terrestris*)

Status in publication process: in preparation

JG and HT planned and designed the research; JG and HT conducted fieldwork; JG analyzed the data with support of MC, VL and DW; JG wrote the manuscript.

Abstract

Biodiversity comprises life in all forms and functions. Its complexity is difficult to grasp in one glance, and to aid in this, biodiversity is commonly divided into three broad categories: ecosystems, species, and genes. Although it defines the world we live in, and provides fundamental ecosystem services, human activities increasingly threaten biodiversity at all levels, and the intensification of efforts to protect it is urgently needed. Such efforts are often impeded by a lack of knowledge on the distribution of biodiversity. Mapping biodiversity at all levels of organization is therefore an essential step in better informing conservation decisions.

One of the key factors in determining where biodiversity is richest, how it is distributed or even how well biodiversity can bounce back from changes and threats it might experience, is the environment. Because environmental data have become widely available, knowledge on how the environment is related to the occurrence of biodiversity can greatly facilitate creating continuous maps of its distribution. Moreover, because human activities severely modify environmental conditions, disentangling their individual effects in driving spatial patterns of biodiversity plays a central in evolutionary biology, ecology and conservation. Within a changing environment, habitats may become unfavorable, and species may respond by shifting their ranges. The current level of habitat fragmentation, however, severely limits range shifts. Thus, species must respond adaptively to the modified or new selection pressures in order to persist in the long run. Such rapid evolutionary responses rely on standing genetic variation, representing adaptations to the environment. It is therefore indispensable to map the spatial distribution of adaptive genetic variation in order to maximize species' evolutionary potential. This equips scientists to understand how changes in the environment may affect biodiversity and can eventually lead to a more fitting and adjusted conservation effort and better management practices.

In this thesis, I aimed to (1) assess the relative influence of environment on different ecological and evolutionary processes, which themselves influence spatial patterns of biodiversity and to (2) evaluate how this understanding can be used for mapping and ultimately protecting biodiversity. I focused on several different components of biodiversity: habitats, species and genes. To cover multiple categories of biodiversity, I investigated (i) the distributional patterns of two closely related bumble bee species (the buff-tailed (*Bombus terrestris*) and the white-tailed bumble bee (*Bombus lucorum*)), and how those patterns are determined by the environment; (ii) the use in spatial conservation prioritization of environmental heterogeneity as a surrogate for species distributions and vice versa; (iii) the role of environment in shaping population divergence in the house sparrow (*Passer domesticus*); and (iv) signals of local adaptations in the buff-tailed bumble bee (*Bombus terrestris*).

This work was conducted in two eastern European countries, Romania and Bulgaria. These countries comprise a highly heterogeneous environment, representing a suitable area for evaluating the effect of the environment on the distribution of biodiversity (species and genes). Parts of my work resulted in specific recommendations for conservation. These are very timely, since the European Union set new biodiversity targets for 2030, requesting member states to increase their protected areas network to 30% of the total land area.

I found that: (i) both vegetation and climatic variables play a role in determining the distributions of the two bumble bee species, in particular vegetation cover and elevation-correlated climatic variables; (ii) environmental heterogeneity is not as a sufficient surrogate for species (based on bird species data) as the other way around, confirming the recommendation to use more than one type of surrogate in spatial conservation prioritization; (iii) and (iv) there is evidence of environmental selection and patterns of 'isolation by environment' determining population structure in both bumble bees and sparrows. I also identified a set of genes that may be adapted to local conditions in bumble bees.

This thesis shows that the environment largely determines how biodiversity is distributed, and yet is a poor surrogate in spatial conservation prioritization. It also provides strong evidence that it is one of the main drivers shaping the genetic structure of species. These findings should encourage scientists to continue mapping spatial patterns of biodiversity and particularly focus on the genetic level of biodiversity. Understanding the drivers and patterns of adaptive genetic variation in populations is providing insight in evolutionary processes and helps ensure that the evolutionary potential of species can be maximized.

Zusammenfassung

Biodiversität umfasst Leben in allen Formen und Funktionen. Ihre Komplexität ist auf einen Blick schwer zu erfassen und um dies zu unterstützen, wird die biologische Vielfalt üblicherweise in drei große Kategorien unterteilt: Ökosysteme, Arten und Gene. Obwohl sie die Welt in der wir leben definiert und grundlegende Ökosystemleistungen erbringt, bedrohen menschliche Aktivitäten zunehmend die biologische Vielfalt auf allen Ebenen. Eine Intensivierung der Bemühungen um ihren Schutz ist daher dringend erforderlich. Solche Bemühungen werden oft durch mangelndes Wissen über die Verteilung der Biodiversität behindert. Die Kartierung der biologischen Vielfalt auf allen Organisationsebenen, ist daher ein wesentlicher Schritt, um bessere und erfolgreichere Entscheidungen zum Naturschutz treffen zu können.
Die Umwelt ist einer der Schlüsselfaktoren die bestimmt, wo die biologische Vielfalt am reichsten ist, wie sie verteilt ist oder sogar wie gut sich die biologische Vielfalt von Veränderungen und Bedrohungen erholen kann. Da Umweltdaten inzwischen in großem Ausmaß zur Verfügung stehen, kann das Wissen darüber, wie die Umwelt mit dem Vorkommen der biologischen Vielfalt zusammenhängt, die Erstellung kontinuierlicher Karten ihrer Verteilung erheblich erleichtern. Da menschliche Aktivitäten zudem die Umweltbedingungen stark verändern, spielt die Entflechtung ihrer individuellen Auswirkungen auf die räumlichen Muster der Biodiversität eine zentrale Rolle in der Evolutionsbiologie, Ökologie und Naturschutz. Innerhalb einer sich verändernden Umwelt können Lebensräume unbewohnbar werden, und Arten können darauf mit einer Verschiebung ihrer Verbreitungsgebiete reagieren. Der gegenwärtige Grad der Habitatfragmentierung schränkt jedoch die Reichweite solcher „Verschiebungen“ stark ein. Daher müssen die Arten anpassungsfähig sein, um auf den veränderten oder neuen Selektionsdruck reagieren und langfristig überleben zu können. Solche raschen evolutionären Reaktionen basieren auf vorhandener genetischer Variation, die eine Anpassung an die Umwelt ermöglicht. Es ist daher unerlässlich, die räumliche Verteilung der adaptiven genetischen Variation zu kartieren, um das evolutionäre Potenzial der Arten zu maximieren. Auf diese Weise können Wissenschaftler verstehen, wie sich Veränderungen in der Umwelt auf die biologische Vielfalt auswirken können und schließlich zu angemesseneren Erhaltungsmaßnahmen und besseren Managementpraktiken greifen.
In dieser Arbeit zielte ich darauf ab, (1) den relativen Einfluss der Umwelt auf verschiedene ökologische und evolutionäre Prozesse abzuschätzen, die ihrerseits die räumlichen Muster der Biodiversität beeinflussen, und (2) zu bewerten, wie dieses Verständnis für die Kartierung und letztlich für den Schutz der Biodiversität genutzt werden kann. Ich habe mich mit verschiedenen Komponenten der Biodiversität, wie Lebensräume, Arten und Gene auseinandergesetzt. Um mehrere Kategorien der Biodiversität abzudecken, untersuchte ich (i) die Verbreitungsmuster zweier eng verwandter Hummelarten (die dunkle Erdhummel (*Bombus terrestris*) und die helle Erdhummel (*Bombus lucorum*)) und wie diese Muster durch die Umwelt bestimmt werden; (ii) die Verwendung der Heterogenität der Umwelt als Stellvertreter für die Verteilung der Arten und umgekehrt; (iii) die Rolle der Umwelt bei der Gestaltung der Populationsdivergenz beim Haussperling (*Passer domesticus*); und (iv) Signale für lokale Anpassungen bei der dunklen Erdhummel (*Bombus terrestris*).

Diese Arbeiten wurden in zwei osteuropäischen Ländern, Rumänien und Bulgarien, durchgeführt. Diese Länder umfassen eine sehr heterogene Umwelt, die ein geeignetes Gebiet für die Bewertung der Auswirkungen der Umwelt auf die Verteilung der Biodiversität (Arten und Gene) darstellt. Teile meiner Arbeit führten zu spezifischen Empfehlungen für die Erhaltung. Diese kommen genau zum richtigen Zeitpunkt, da die Europäische Union neue Ziele für die biologische Vielfalt bis 2030 festgelegt und die Mitgliedstaaten aufgefordert hat, ihr Netz von Schutzgebieten auf 30% der Gesamtfläche zu erweitern.
Meine Ergebnisse zeigen, dass: (i) sowohl Vegetations- als auch Klimavariablen bei der Bestimmung der Verteilung der beiden Hummelarten eine Rolle spielen, insbesondere die Vegetationsbedeckung und die höhenkorrelierten Klimavariablen; (ii) die Heterogenität der Umwelt nicht als ausreichender Stellvertreter für Artenvielfalt (auf der Grundlage von Vogelartdaten) dient, sondern umgekehrt. Das bestätigt die Empfehlung, bei der Festlegung von Prioritäten für Schutzgebiete, mehr als eine Art als Stellvertreter für Biodiversität im Allgemeinen zu verwenden; (iii) und (iv) es Hinweise auf eine Selektion durch die Umwelt (genannt: „Isolation by distance“) gibt, die die Populationsstruktur sowohl bei Hummeln als auch bei Spatzen bestimmen. Ich habe auch eine Reihe von Genen identifiziert, die an die lokalen Bedingungen bei Hummeln angepasst sein könnten.
Diese Arbeit zeigt, dass die Umwelt weitgehend bestimmt, wie die biologische Vielfalt verteilt ist und dennoch ein schlechter Ersatz für die Prioritätensetzung bei Schutzgebieten ist. Sie liefert auch überzeugende Beweise dafür, dass die Umwelt eine der Haupttriebkräfte ist, die die genetische Struktur der Arten prägt. Diese Erkenntnisse sollten Wissenschaftler dazu ermutigen, die Kartierung der räumlichen Muster der Biodiversität fortzusetzen und sich insbesondere auf die genetische Ebene der Biodiversität zu konzentrieren. Das verstehen der genetischen Muster die Populationen zu Anpassung an die Umwelt benötigen, ermöglicht Einblicke in evolutionäre Prozesse. Mit diesem Verständnis, und dem Wissen über potenzielle Triebkräfte kann das evolutionäre Potenzial von Arten maximiert werden.

Contents

1 General Introduction

Biodiversity and its spatial patterns

The term 'biodiversity' is often used casually in the media, and common perception usually involves the variability in species. However, biodiversity is much more complex, and defined as the "variability of life in all forms, levels and combinations" (Glowka et al., 1994). It comprises structural, functional and compositional components, each structured along different organizational levels, ranging from landscapes through ecosystems and habitats to species and populations; and on the finest scale to genes and genetic processes (Noss, 1990). This complexity is often compressed into three main components: ecosystems, species, and genes. The ecosystem is the highest level of organization and can be seen as a functional unit comprising biotic and abiotic factors as well as their interactions. The population and species level of biodiversity comprises the diversity of all species and variability between populations. The diversity of species within a geographical area is often referred to as 'species richness', a commonly used measure in biodiversity research and conservation. Finally, the genetic level of biodiversity is considered to be the genetic variation within organisms as well as the genetic differences among individuals or populations (Glowka et al., 1994; Noss, 1990). The hierarchical organization of these three levels of biodiversity demonstrates their connectedness. At the same time it becomes clear that one level of biodiversity cannot capture the entirety of biodiversity as a whole (Pereira et al., 2010).

Biodiversity at all levels is heterogeneously distributed across the world, with some areas being more diverse than others, and no single area capturing all existing diversity (Gaston, 2000). This naturally raises the question: what shapes the spatial patterns of biodiversity? A variety of factors have been hypothesized to be key determinants of the distribution of biodiversity (reviewed by Fine, 2015; Peters et al., 2016, including temperature and latitude (e.g. (Mittelbach et al., 2007), the strength of biotic interactions (e.g. competition, predation) (e.g. (Mittelbach et al., 2007), the size of the area species occur in (Rosenzweig, 1995), the amount of available water (Hawkins et al., 2003), or the level of plant diversity present (e.g. (Novotny et al., 2006). Some studies suggest an interplay of different drivers, such as the size of an area and fluctuations in the amount of available habitat, in combination with temperature (Belmaker and Jetz, 2015). Indeed, it is apparent that at global scales no single driver can attribute for the heterogeneous distribution of biodiversity, but rather a multitude of drivers interacting in complex ways. At smaller scales, however, a few dominant drivers may determine the distribution of biodiversity on multiple levels (Belmaker and Jetz, 2015; Peters et al., 2016). Most of the potential drivers are related to the abiotic environment, such as climate, water, the geographical location of an area, or even unpredictable, environmentally related processes such as natural disturbance regimes (like floods, fires or droughts). It is important to bear in mind that the environment drives ecological processes, setting the stage for biotic interactions, and at the same time provides the basis for micro- and macro-evolutionary processes (Jetz et al., 2012).

An ecological process is a widely used and relatively broad term entailing for instance climatic processes, interactions between organisms (and their environment), hydrological processes or the movement of organisms (Bennett et al., 2009). These processes influence the distribution of biodiversity, e.g. by determining where species occur, live in competition, or migrate (Belmaker and Jetz, 2015; Gaston, 2000; Pavé, 2007; Peters et al., 2016). Some of these processes are unpredictable in the magnitude, location and timing of their occurrence. Such chance ecological processes, like ecological drift (=random change in species abundance), random extinction or random historical events may have large effects on the distribution of biodiversity on earth (Bennett et al., 2009; Hubbell, 2001).

Micro- and macro-evolutionary processes such as genetic mutations, gene flow (the transmission of genetic variation from one population to another), genetic drift (the random change in allele frequencies in a population) and natural selection are partially influenced by the environment themselves and in addition influence biodiversity (on the species and genetic level) (Gaggiotti et al., 2009). The nature, magnitude, and influence of the variety of potential micro- and macro-evolutionary processes on generating and maintaining biodiversity comprise a major field of study and continue to be debated (reviewed by Dietrich, 2010; Li et al., 2018; Simons, 2002). Briefly, genetic variation within and between populations is a balance between mutation rates, genetic drift, gene flow, and selection. Selection by the environment may increase or decrease genetic variation within and between populations, depending on the heterogeneity of the environment and the associated differential fitness consequences. In contrast, genetic drift is a neutral process, resulting in random allele frequency changes, decreased variation within populations, and increased population divergence. Drift is balanced by new mutations and gene flow, which itself may be influenced by neutral factors, such as landscape barriers or the geographic distance between populations, resulting in genetic patterns of 'isolation by distance' (= positive relationship between genetic differentiation and geographic distance) (IBD; Wright, 1943). Interestingly, however, the reduction in fitness of maladapted dispersers in a new location may also limit gene flow, and thus enhance the effects of genetic drift. Thus, the signature of local adaptations may be detectable across the entire genome, resulting in genetic patterns of 'isolation by adaptation' (= relationship between adaptive phenotypic divergence and genetic differentiation) (IBA; Nosil et al., 2009) and 'isolation by environment' (= environmental heterogeneity shaping genetic structure) (IBE; Wang and Bradburd, 2014).

In a nutshell, environmental factors influence both ecological as well as evolutionary processes, which themselves determine the distribution of biodiversity across multiple levels of organization (from ecosystems to genes). The influence of human activities on the environment - and hence biodiversity - is bigger than ever before. The successful conservation of biodiversity requires a thorough understanding of how changes in the environment affect biodiversity. In addition, areas for protection should be identified in a way that they most effectively and efficiently conserve biodiversity. Mapping the spatial distribution of biodiversity and relating this to the prevailing environmental conditions are thus key components in informing conservation decisions.

The main goals of this thesis are therefore to assess to what extent biodiversity dependents on the environment and how we can use this knowledge to map and protect biodiversity?

Biodiversity in the 'Anthropocene'

Life on earth is dominated by human actions (Lewis and Maslin, 2015), and to reflect the pervasiveness of our influence, the time period we are living in has been coined the 'Anthropocene' (Crutzen, 2006). Human activities severely alter the environment and constitute the biggest threat to biodiversity. Anthropogenic stressors are diverse, and include -but are not limited to- habitat alteration and loss (through fragmentation or destruction), pollution, overexploitation, the introduction of potentially invasive species (Hoffmann et al., 2010; Pereira et al., 2012; Primack, 2002), the spread of pathogens (Smith et al., 2006), and anthropogenically-induced climate change (Scheffers et al., 2016).

These threats have for the last couple of hundred years severely altered the world around us, disrupting evolutionary ecological processes, and changing the distribution and abundance of species, ultimately even leading to extinctions (review by Bellard et al., 2012; Pereira et al., 2012. Among the diverse array of threats to biodiversity, habitat change or loss is the most critical (Pereira et al., 2012), resulting in decreasing population sizes, isolation of populations and hence restricted gene flow, all affecting genetic variation in natural populations (DiBattista, 2008; Hoffmann and Sgrò, 2011; Kremer et al., 2012; Willi et al., 2007). Environmental change also alters selection pressures (Hoffmann and Sgrò, 2011), and populations and species need to respond to these changes to persist in the long term. Species' responses can be ecological, such as shifting their ranges to match their optimal habitat (Anderson et al., 2012; Harrisson et al., 2014), or evolutionary, by adapting to the new conditions (Anderson et al., 2012). These responses are not mutually exclusive, and most species may need to respond in both ways in order to persist in the face of human influences (Anderson et al., 2012). Thus, in order to effectively protect biodiversity it is crucial to understand how it is distributed, what factors determine its distribution, and how it is responding to environmental change (De Mazancourt et al., 2008).

Several factors influencing species distributions have been identified, including the abiotic environment (Costa et al., 2008; Elith and Leathwick, 2009) as well as biotic interactions in this environment (Pearson and Dawson, 2003; Wisz et al., 2013) and chance (demographic) processes (Roland Pitcher et al., 2012). The relative contributions of these factors may vary, and one might ask to what extent the environment determines the distribution of a given species? To answer this question (*Chapter I*), so called Species Distribution Models (SDMs) are often used. SDMs correlate species' occurrence and abundance data with the environmental conditions they live in, and can accurately predict species' ranges (Elith and Leathwick, 2009). With improvements in remote sensing and spatial analyses, a suit of environmental, ecologically meaningful variables have over the past 20 years become available (Penado et al., 2016), enabling us to gain insight in species-specific habitat suitability (Elith and Leathwick, 2009; Guisan and Thuiller, 2005) and to predict the effects of environ-

mental change (Keitt et al., 2002). Moreover, given the dearth of high-quality survey data in many countries, it also provides a cost-effective and relatively quick means of mapping where species occur. Such knowledge about a species' geographical distribution and its abundances is crucial for adjusting practices with respect to land management, climate change policies, and conservation (Guisan et al., 2013; Jetz et al., 2012; Pearson, 2007). As such, species distribution models play an important role in systematic conservation planning and spatial conservation prioritization approaches (Margules and Pressey, 2000). Under the assumption of niche conservatism, future predictions (for instance with the incorporation of future climatic conditions) can also help to understand the potential for or limitations to range shifts over time (Wiens et al., 2010).

When populations cannot shift their ranges, however, their only means to survive over the long run is by responding plastically or adaptively to environmental change (Anderson et al., 2012; Harrisson et al., 2014). Phenotypic plasticity is the expression of various phenotypes by one genotype and may occur when species experience 'short-term' environmental changes, such as predictable seasonal events (Oostra et al., 2018). Evolutionary adaptation on the other hand is a common response to 'long-term' changes (Barrett and Schluter, 2008; Hoffmann and Sgrò, 2011; Savolainen et al., 2013; Sgro et al., 2011). Although the magnitude by which populations can respond plastically may have a genetic basis (Bijlsma and Loeschcke, 2012), and phenotypic plasticity may be crucial for the immediate future, it is likely too limited to be an adequate long-term response to the large environmental changes inflicted by human actions (Anderson et al., 2012; Bijlsma and Loeschcke, 2012; DeWitt et al., 1998; Hendry et al., 2008). To this end, adaptive responses may be most fundamental for the persistence of species (Hoffmann and Sgrò, 2011), but highly depend on the amount of available standing genetic variation (Etterson and Shaw, 2001). Thus, protecting intra-specific genetic variation that express adaptations to the environment is central to protecting biodiversity (Hoffmann and Sgrò, 2011; Miraldo et al., 2016).

Biodiversity conservation – from landscapes to genes

Protected areas are a vital component of conservation efforts, having the potential to protect biodiversity at multiple levels of organization. Historically, protected areas have often been identified based on low economic values, landscape features, the presence of a particular species of concern, or expert knowledge on the available biodiversity (Brooks et al., 2006; Myers et al., 2000; Sarkar et al., 2006). Systematic conservation planning aims to provide more objective decision support for the allocation of resources in biodiversity conservation and the implementation of conservation actions (Margules and Pressey, 2000; McIntosh et al., 2017). It is considered to be an unbiased and evidence-based approach in order to provide accountable and transparent advice when important decisions need to be taken (McIntosh et al., 2017). Different strategies, such as including stakeholders in the planning region, thorough collection of biological and socioeconomic data, and multi-level analyses to satisfy the opposing socioeconomic and biological goals of the stakeholders in the planning area are implemented in systematic conservation planning (Sarkar et al., 2006). The

main tool of systematic conservation planning is spatial conservation prioritization, aiming to identify a set of reserves that protect as much biodiversity as possible most effectively and efficiently, i.e. in the smallest area possible (Margules et al., 2002). The optimization of a network of reserves, embedded within current or future spatial and socio-economic constraints is a highly complex problem for which several algorithms have been developed (McIntosh et al., 2017; Sarkar et al., 2006). Whereas the conceptual framework of these so-called spatial conservation practices has been relatively well developed, its practical implementation is typically limited by the amount and the quality of the available data on biodiversity.

Due to the above described complexity of biodiversity it is not possible to survey and map all levels of biodiversity at once (Williams et al., 2006). Limited resources (financial, personnel) and different interests of conservation stakeholders add to the difficulty of this issue. Conservation planners must therefore revert to 'biodiversity surrogates' as a shortcut in spatial conservation prioritization (Margules and Pressey, 2000; Sarkar et al., 2005). Surrogates are ecological processes or elements (e.g. species, ecosystems or abiotic factors) representing another aspect of an ecological system (Hunter et al., 2016). Different surrogates have been used to identify areas of high conservation value, such as taxonomic (species) or environmental surrogates (Grantham et al., 2010; Oliver et al., 2004; Sarkar et al., 2005), but the concept can be used at any level of biodiversity. The most often used surrogate is species occurrence, identifying hotspots of species richness, rarity, and complementarity (e.g. Arponen et al., 2008). However, although species level data (Gomes et al., 2018) is increasingly becoming available, extensive high-quality data on species distributions are still lacking for most taxonomic groups and most areas of the world (Arponen et al., 2008; Beier et al., 2015). Environmental data representing habitat structure, complexity, and heterogeneity, on the other hand can be quite easily and cheaply acquired, aiding spatial conservation planning (Arponen et al., 2008; Beier et al., 2015; Grantham et al., 2010; Rodrigues and Brooks, 2007). It was recognized early on that it is also important to not only protect species, but also biodiversity at the higher levels of organization, i.e. habitats and ecosystems themselves (Noss, 1990). The assumption is that by protecting environmental heterogeneity one might also protect many species living in that variety of different habitats (e.g. Arponen et al., 2008; Bonn and Gaston, 2005. Yet, studies on the power of environment as a surrogate for biodiversity (mostly at the species level) remain inconclusive, demonstrating adequate (Sarkar et al., 2005; Trakhtenbrot and Kadmon, 2005), but also insufficient 'surrogacy power' (Araújo et al., 2007; Bonn and Gaston, 2005). More research is thus needed to clarify whether biodiversity surrogates are adequately representing one another and thus can be reliably used in conservation prioritization (*Chapter II*).

Environmental diversity also constitutes the arena in which species adapt to different conditions and exhibit spatial and temporal differences in their phenotype and genotype. Environmental heterogeneity is therefore expected to harbor phenotypic and genetic variation, for instance across environmental gradients, relevant for the long-term persistence of species (Conover et al., 2009; Smith et al., 2001). This is very important when species face environmental changes, since they will most likely need to rely on a combination of ecological and evolutionary responses (Anderson et al., 2012). Many species have already been document to have shifted their ranges

poleward and to higher elevations because of climate change (Parmesan, 2006). One problem however is that responding just ecologically might not do the trick, since range shifts may not be possible, and are likely to come with some degree of novel abiotic and biotic conditions and novel interactions, which might require long-term adaptive changes (Anderson et al., 2012; Etterson and Shaw, 2001). And what will happen if even those northern or higher regions become warmer and unsuitable? In order to avoid extinction in this scenario, species have to respond evolutionarily to these changes and adapt locally (Anderson et al., 2012; Hoffmann and Sgrò, 2011). Local adaptation occurs when a population shows higher fitness in its native habitat compared to any other population that would be introduced and is largely influenced by natural selection and gene flow. Looking at signs of local adaptation in populations with ongoing gene flow provides an opportunity to identify selective forces imposed by particular environmental differences (Kawecki and Ebert, 2004). The spatial extent of local adaptation is therefore depending on the strength of selection, gene flow (and migration) between populations (Kawecki and Ebert, 2004), and most importantly the underlying genetic diversity of these populations (Vincent et al., 2013). This so called 'standing genetic variation' (which is the presence of alternative alleles at a given locus) is the basis for rapid genetic adaptations (Barrett and Schluter, 2008; Biodiversity, Council, et al., 1999; Etterson and Shaw, 2001; Thomassen et al., 2011). Several studies revealed a correlation between genetic variation and the long-term viability of species (Arenas Busto et al., 2014; Laikre, 2010; Vangestel et al., 2012) as well as the expression of higher rates of adaptability (Barrett and Schluter, 2008).

The importance of genetics in conservation was already recognized decades ago by Frankel, 1974, but just protecting species will not automatically protect its associated genetic diversity (Glowka et al., 1994). Genetic diversity is facing the exact same threats as other levels of biodiversity and is decreasing worldwide (Leigh et al., 2019). The loss of genetic diversity is permanent and even though species might recover in numbers again, the amount of genetic diversity remains low for many generations, leading to lowered adaptive potential and increased extinction risks (Brondizio et al., 2019). However, to protect biodiversity on the genetic level as well, we need to understand how it is influenced by human-induced threats and global changes (Miraldo et al., 2016; Palumbi, 2001). At the broad scale, regions more disturbed and influenced by humans were shown to harbor less genetic diversity than regions with lower human influences (DiBattista, 2008). Knowledge like this is important, and mapping genetic variation and its distribution is crucial (Miraldo et al., 2016; Pereira et al., 2013). In particular that part of genetic variation that is correlated to environmental characteristics (also known as environmentally associated variation (EAV)) may be most relevant to adapting to novel environmental conditions and changed habitats. Hence, a prudent strategy to protect biodiversity in the face of environmental and climate change is to maximize species' adaptive potential through mapping and protecting the current distribution of EAV (Frankham, 2010; Sgro et al., 2011; Vandergast et al., 2008; Vandersteen Tymchuk et al., 2010).

The problem however is that despite its widely recognized importance in conservation, genetic diversity is still lacking in many conservation policies and practical management (Laikre, 2010). Indeed, the distribution of genetic variation (Pereira et al., 2012) as well as the rate of its loss is unknown for most of the world (Leigh et al.,

2019). There is thus an urgent need to map the genetic level of biodiversity in order to detect and monitor its distribution and loss.

Landscape genetics as a tool for mapping spatial patterns of (genetic) diversity

A set of tools was developed over the years that have proven extremely useful to map environmentally associated genetic variation (EAV), investigate its distribution, and identify important drivers in shaping it. These so called landscape genetic approaches combine population genetic methods with landscape ecology (Manel et al., 2003). Landscape genetics enhances our understanding of how the environment is shaping genetic variation of individuals (or entire populations) by looking at the interaction between the environment and the genetic make-up of individuals and populations (Balkenhol et al., 2017; Guillot et al., 2005a; Manel et al., 2003). Molecular markers (e.g. microsatellites, single-nucleotide polymorphisms (SNPs)) genotyped in individuals distributed across a geographic area are used to identify genetic patterns (e.g. population structure, or distribution of genetic variation) and are then correlated to landscape features (e.g. barriers) and environmental variables (*Chapter III*) (Manel et al., 2003; Storfer et al., 2018). Potential selective drivers of genetic variation, such as 'isolation by adaptation' (IBA; Nosil et al., 2009) and 'isolation by environment' (IBE; Wang and Bradburd, 2014) can be identified by these methods, and teased apart from neutral evolutionary processes, such as 'isolation by distance' (IBD; Wright, 1943).

With technological advances in recent years, such as high throughput sequencing and new analytical tools, landscape genetics developed into landscape genomics, further advancing our understanding of the contribution of the environment on evolutionary processes (Henriques et al., 2018; Storfer et al., 2018). So called gene-environment associations (GEAs) help to identify potential selection pressures driving local adaptation by scanning the genome for loci that are significantly associated to the environment or profusely differentiated between populations (*Chapter IV*) (Capblancq et al., 2018; Frichot et al., 2013; Hoban et al., 2016; Joost et al., 2007). In Moroccan sheep populations, for example, candidate loci (located in a gene responsible for wax secretion) were associated to precipitation indicating potential adaptations to 'wetter' environmental conditions (Duruz et al., 2019). In Californian oak trees, climate-associated functional genes were identified, with differences between populations also demonstrating adaptations to different environmental settings (Sork et al., 2016).

Such landscape genetic or genomic approaches can also benefit biodiversity conservation directly by informing us, for example, which (combination of) populations possess high levels of standing genetic variation, which populations might be exposed to higher risks of anthropogenic threats and/or climate change (e.g. Jia et al., 2020; Martins et al., 2018), and which populations may require adjusted management practices (Gugger et al., 2018; Harrisson et al., 2014). The incorporation of evolutionary processes such as adaptation and gene flow in spatial conservation planning is thus hypothesized to decrease biodiversity loss under rapid environmental change

(Hoffmann and Sgrò, 2011). Particularly understanding the relative roles of the environment in shaping the distribution of genetic diversity as well as other levels of biodiversity, can advance biodiversity conservation practices and eventually protect what is left of the precious diversity on our planet.

Objectives of this thesis

As outlined above, the relationship between the environment and patterns of biodiversity is important to understand in order to effectively protect biodiversity on earth. With this work I want to help fill in the knowledge gap existing around the spatial patterns of biodiversity, and demonstrate some of the methods one can use to map biodiversity, especially the so far rather neglected genetic level of biodiversity. Work for this thesis was conducted in two Eastern-European countries: Romania and Bulgaria (Figure 1). Both countries cover an area of approximately 350 000 km^2 and comprise distinct climatic zones, such as the continental and temperate climatic zones in Romania, and the continental and Mediterranean climatic zones in Bulgaria. This leads to highly divergent climatic conditions on a relatively small scale and is supplemented by a variety of habitats with differing intensity of human-influence such as agricultural and industrial areas, as well as natural areas like mountain ranges, river valleys, forests and grasslands. The Danube River, represents a natural border between these countries, located in the South of Romania and the North of Bulgaria. Both countries contain large mountainous areas; in Romania the Carpathian mountain region is predominant, whereas the Balkan, Rhodope, Rila and Pirin mountains merge to a large mountainous area in Bulgaria. This variety of habitats resulted in the recognition of different biogeographical regions: the continental, alpine, steppic, Black Sea, and pannonian regions ((CoE), 2015). This high environmental heterogeneity potentially imposes divergent selection pressures on species and populations and is therefore the ideal testbed to investigate the relationship between environment (and its change) and spatial patterns of biodiversity and assess the influence of environment on biological diversity.

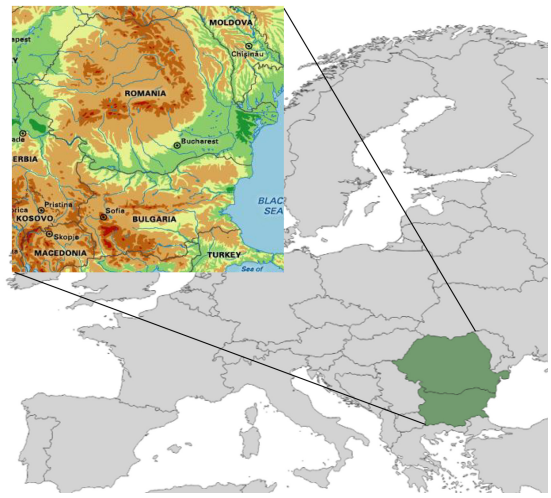

FIGURE 3: Study region

Since 2007 both countries are members of the European Union (EU) and are also part of the NATURA 2000 network, which required the identification and implementation of protected areas. The Natura 2000 network is the most important large-scale

biodiversity conservation measure in Europe (Gaston et al., 2008). Different types of protected areas are implemented in the network, on the basis of different biodiversity features: the terrestrial Sites of Community Importance and Special Areas of Conservation (SCI and SAC, for habitats and/or species) and terrestrial Special Protection Areas (SPA, for bird protection only) (Commission, 2007). Approximately 20% of the total landscape area of Romania and Bulgaria is dedicated to nature conservation, however under the new European Union Biodiversity Strategy for 2030 (Commission, 2020) the percentage of protected areas should be increased to 30%. This hands us the opportunity to scrutinize the current status of biodiversity and its conservation status quo in both countries.

In **CHAPTER I** of this thesis, I correlated species occurrence and abundance data with environmental data in order to evaluate the influence of environmental variation on the distributional patterns of two closely related bumble bee species (the buff-tailed (*Bombus terrestris*) and the white-tailed bumble bee (*Bombus lucorum*)). The partial overlap in their occurrence patterns posed the question about ecological drivers shaping those distributions and how the habitat requirements differ between both species. I used an ensemble of different species distribution modelling techniques to combine the occurrence and abundance data I collected with a set of ecologically important environmental variables.

In **CHAPTER II**, I investigated if environmental heterogeneity, which is playing a major role in determining species distributions and occurrences (e.g. Elith and Leathwick, 2009), could be used as a conservation surrogate for the species level of biodiversity. Here I used a measure of bird species richness and classified environmental diversity to evaluate their use as surrogates for one another. With the help of a spatial conservation prioritization method, I designed protected areas based on both measures of biodiversity and additionally assessed how those two measures are captured by already existing protected areas.

In **CHAPTER III** of this thesis, I investigated the influence of the environment on the genetic level of biodiversity and if population divergence of house sparrows (*Passer domesticus*) can be explained by environmental variation. Since species might likely respond in several ways to environmental change I also included phenotypic data. With the help of a landscape genetics approach, I correlated phenotypic and genetic diversity with a set of environmental variables. The goal was to detect signatures of environmental selection hinting towards local adaptations.

In **CHAPTER IV**, I took the landscape genetics towards the landscape genomics level by using adaptive genetic markers (single-nucleotide polymorphisms, SNPs) to further investigate the broad ecological patterns influencing the occurrence of the buff-tailed bumble bee (*Bombus terrestris*) I discovered in *Chapter I*. Here, I identified gene-environment associations (GEAs) between a whole genome SNP data set and the same set of ecological important environmental variables (as in *Chapter I* I aimed to characterize potential selection pressures driving local adaptation in *B. terrestris* and identify loci under selection which could underline the occurrence patterns found in *Chapter I*.

2 CHAPTER I

UNRAVELING THE HABITAT PREFERENCES OF TWO CLOSELY RELATED BUMBLE BEE SPECIES IN EASTERN EUROPE

Julia C. Geue and Henri A. Thomassen

Abstract

Co-occurrence of closely related species is often explained through resource partitioning, where key morphological or life-history traits evolve under strong divergent selection. In bumble bees (genus *Bombus*), differences in tongue lengths, nest sites, and several life-history traits are the principal factors in resource partitioning. However, the buff-tailed and white-tailed bumble bee (*Bombus terrestris* and *Bombus lucorum* respectively) are very similar in morphology and life history, but their ranges nevertheless partly overlap, raising the question how they are ecologically divergent. What little is known about the environmental factors determining their distributions stems from studies in Central and Western Europe, but even less information is available about their distributions in Eastern Europe, where different subspecies occur. Here, we aimed to disentangle the broad habitat requirements and associated distributions of these species in Romania and Bulgaria. First, we genetically identified sampled individuals from many sites across the study area. We then not only computed species distributions based on presence-only data, but also expanded on these models using relative abundance data. We found that *B. terrestris* is a more generalist species than previously thought, but that *B. lucorum* is restricted to forested areas with colder and wetter climates, which in our study area are primarily found at higher elevations. Both vegetation parameters such as annual mean Leaf Area Index and canopy height, as well as climatic conditions, were important in explaining their distributions. Although our models based on presence-only data suggest a large overlap in their respective distributions, results on their relative abundance suggest that the two species replace one another across an environmental gradient correlated to elevation. The inclusion of abundance enhances our understanding of the distribution of these species, supporting the emerging recognition of the importance of abundance data in species distribution modeling.

KEYWORDS
Bombus lucorum, *Bombus terrestris*, Eastern Europe, pollinators, random forests, relative abundance, species distribution

Introduction

The co-occurrence of closely related species has long puzzled evolutionary ecologists. Closely related species are expected to occupy similar niche space through niche conservatism, and thus occur in the same regions, but also to compete more strongly with one another than more distantly related species (Anacker and Strauss, 2016). Strong competition may result in the exclusion of the weaker competitor, as well as in rapid divergent evolution of key life-history traits or phenotypes as a result of resource partitioning (Gause, 1934). In bumble bees (genus *Bombus*), a classical theory is that species evolved a range of different tongue lengths, allowing them to specialize on different floral resources, and to occur sympatrically as a result (Goulson and Darvill, 2004). Other mechanisms of resource partitioning include differences in nest sites, foraging distances, and the spatial use of habitat (Stanley et al., 2013; Westphal et al., 2006). However, two of the most common bumble bee species in Europe, the buff-tailed bumble bee (*Bombus terrestris*) and the white-tailed bumble bee (*Bombus lucorum*), co-occur (pers. obs.; (Goulson, 2010; Goulson et al., 2008a; Kells et al., 2001; Stanley et al., 2013)), despite being very similar in their morphology, choice of nest sites, and life-history (Goulson, 2010; Stanley et al., 2013). They even possess the same tongue lengths (Goulson et al., 2005; Stanley et al., 2013) and hence occupy a very similar dietary niche space (Goulson et al., 2008a).This begs the question to what extent their ecological niches overlap and conversely how they are divergent. Despite many studies into their ecology and behavior (e.g. (Bossert et al., 2016; Scriven et al., 2015; Stanley et al., 2013; Walther-Hellwig and Frankl, 2000), and broad-scale evidence that their ranges only partially overlap (e.g. (Rasmont et al., 2015a), their distributions at smaller scales remain equivocal. One reason for this ambiguity may be the fact that these species are morphologically highly variable within species, yet very similar between species (Bossert, 2014; Murray et al., 2007; Waters et al., 2011). As a consequence, they may be difficult to distinguish in the field, depending on where they occur and whether queens, males, or workers are compared. In mainland Europe, identification can be complex because both species possess a white abdomen (Gammans et al., 2018; Rasmont et al., 2013), in contrast to those in Great Britain (Murray et al., 2007). Queens and males can be distinguished (Bertsch et al., 2004; Goulson, 2010), but workers (especially of some subspecies of *B. terrestris*) are difficult to discriminate (Williams, 1994). Indeed, in Central Europe, only 45.5% of *B. lucorum* workers could be correctly identified and distinguished from *B. terrestris* workers using the most up-to-date morphological key (Wolf et al., 2009). As a consequence, many studies focusing on the ecology or behavior of European bumble bees group these taxonomically well-recognized species together (Bommarco et al., 2011; Carvell, 2002; Goulson et al., 2005; Meek et al., 2002; Pywell et al., 2005; Walther-Hellwig and Frankl, 2000), leading to imprecise information on habitat preferences (Murray et al., 2007; Scriven et al., 2015) and other life-history traits (Stanley et al., 2013). One of the few studies that investigated the ecological preferences of

those species separately identified differences in nesting site selection at small spatial scales in Sweden (Svensson et al., 2000) where *B. terrestris* preferred more open habitat, such as fields in agricultural landscapes, and *B. lucorum* preferentially built nests close to forest boundaries. Also in Austria, *B. lucorum* appeared to frequently occur in forested areas (Bossert et al., 2016). At the scale of Europe, *B. lucorum* occurs at higher latitudes than *B. terrestris*, suggesting a differentiation based on temperature (Rasmont et al., 2015a). Most studies on these species focused on Western and Central Europe, but little attention has been paid to Eastern Europe, where the situation is complicated by the occurrence of two subspecies of *B. terrestris, B. t. terrestris and B. t. dalmatinus*, that are morphologically variable (Lecocq et al., 2013; Rasmont et al., 2013). Hence, their distribution patterns in Eastern Europe remain equivocal.

Here, we investigate the distributions and broad habitat characteristics of buff-tailed (*B. terrestris*) and white-tailed bumble bee (*B. lucorum*) in Bulgaria and Romania, where they are the two most common bumble bee species. *B. lucorum* is one of three cryptic species, the other two being *B. cryptarum* and *B. magnus*. Here, we only include *B. lucorum*, because we did not find any individuals of the latter two species, despite extensive sampling efforts. We hypothesized that the differential habitat use at small scales can be scaled up to landscape scale habitat preferences across a spatial extent of hundreds of kilometers. We first genetically identified the species at sites where multiple individuals were sampled, providing a reliable tool for species identification (Bossert et al., 2016; Murray et al., 2007; Stanley et al., 2013; Waters et al., 2011; Williams et al., 2012; Wolf et al., 2009). We subsequently created species distribution models (SDMs), which have been used previously in quantitative ecological studies of bumble bees (Casey et al., 2015; Herrera et al., 2014; Kadoya et al., 2009; Pradervand et al., 2014; Rasmont et al., 2015a). They are usually based on presence-only or presence–absence data, with the assumption that the modeled probabilities of occurrence are indicative of abundance (e.g. (Dallas and Hastings, 2018)). However, recent work across multiple species suggests that these so-called abundance–suitability relationships are often weak (Dallas and Hastings, 2018; Howard et al., 2014; Mi et al., 2017). For that reason, the collection and use of abundance data to improve the accuracy of species distribution models was highly recommended (Howard et al., 2014), but still not commonly applied to date. To improve our distribution models, and to specifically investigate their co-occurrence and the associated abiotic factors driving variation in abundance patterns, we therefore also collected and modeled the relative abundance of these species.

Methods

Study species and study area

Bumble bees (*Bombus* sp.) are important pollinators for crops and wild plants, in particular in temperate ecosystems (Corbet et al., 1991; Murray et al., 2007). Their body is covered in a dense, colored fur that enables them to be endothermic (Heinrich, 1993), and hence adapt to cold climates, such as alpine and arctic environments. Their distribution extends much further north than that of other bees, and their colonies have been found in the extreme northern regions of the northern hemisphere (Goulson, 2010). *Bombus terrestris* and *Bombus lucorum* are two of the most common bumble bee species in Europe. These species have very similar life cycles and are often found in the same areas. Both species use underground nest sites and often choose already existing holes, previously used by small rodents (Goulson, 2010). They possess similar tongue lengths, and as a result forage primarily on the same short corollas and daisy type of flowers (personal observation and (Goulson et al., 2008b). In addition, their workers almost perfectly resemble each other, and only the queens and males can be identified reliably (Murray et al., 2007; Wolf et al., 2009), but field identification remains complicated due to the subtlety of morphological differences.

We conducted our study in Bulgaria and Romania, two neighboring countries in southeastern Europe, covering an area of approximately 350.000 km^2. These countries present a heterogeneous landscape, comprising continental, Mediterranean, and temperate climatic zones, consisting of natural areas such as mountains, river valleys, forests, open woodlands, and grasslands, as well as areas inhabited and influenced by humans, including extensive agricultural lands. The Danube River forms a natural border along much of its length between Romania in the north and Bulgaria in the south. Large mountainous areas are present in both countries: the Carpathians in Romania, and the Balkan, Rila, Rhodope, and Pirin mountains in Bulgaria. As a result of this variety of habitats, different biogeographical regions are recognized: the continental, alpine, steppic, Black Sea, and pannonian regions **((CoE), 2015)**. This high habitat heterogeneity represents an interesting area for evaluating habitat preferences and niche differentiation within and among species.

Sampling

We collected 743 individuals compromising queens and workers of *Bombus terrestris* and *Bombus lucorum* over a timeframe of 4 years (2013, 2014, 2015, and 2017), in each between April and July. We sampled 44 locations in total (Figure 1a and Table 1), which were selected to span a broad range of habitat conditions in both entirely natural and semi-natural or cultivated environments, as well as along environmental gradients (altitude, vegetation, and climate). We visited additional locations where we searched for, but did not find any bumble bees. These locations were not included

as true absences in our species distribution models, but served in computing a sampling bias map (see below).

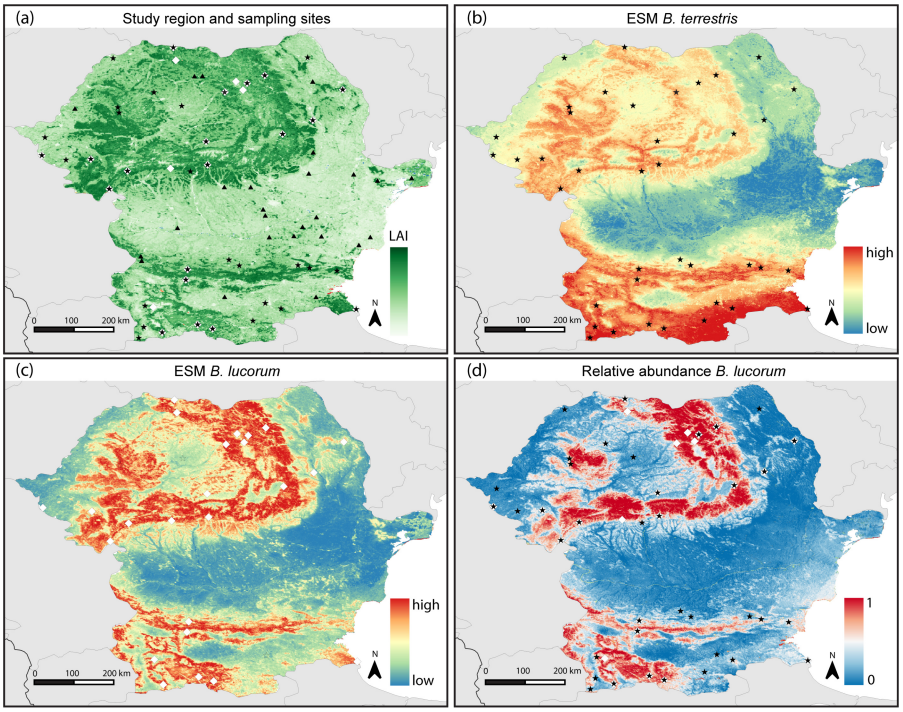

FIGURE 2: Study region with sampling sites, species distribution modeling results based on ensembles of small models, and relative abundance modeling results. (a) Study area with sampling sites indicated in black stars for *Bombus terrestris* and in white diamonds for *B. lucorum*. Sites where we searched for bumble bees, but did not find any are indicated in black triangles. The background map is annual mean Leaf Area Index (LAI mean), a measure of vegetation greenness. (b) Ensemble of small models for *B. terrestris* (c) Ensemble of small models for *B. lucorum*. The colors in panel (b) and (c) indicate the probability of occurrence, with warmer colors indicating higher probabilities. Black stars (b) and white diamonds (c) indicate the sampling populations. (d) Machine learning ensemble for the relative abundance of *B. lucorum*. Warmer colors indicate a higher abundance of *B. lucorum* relative to *B. terrestris*.

Sampling locations were located at least 20 km apart to rule out the possibility of overlapping foraging ranges (Chapman et al., 2003; Westphal et al., 2006) and were visited only once. At each sampling location, capturing efforts were undertaken by 2–3 people for 1.5–2 hrs between 1 hr after sunrise and 1 hr before sunset. Individuals were collected on suitable forage patches with a radius of 100 m, using an entomological net. Individuals were visually identified as one of the two study species, anesthetized in a killing jar with a 1.5 cm layer of plaster of paris saturated with ethyl acetate, and immediately upon cessation of movement stored in 96% ethanol (Smithers, 1988). After fieldwork, specimens were stored frozen at $-20\,^{\circ}\mathrm{C}$ in the laboratory at the University of Tübingen.

TABLE 1: Sampling locations and sample sizes of *Bombus terrestris* and *Bombus lucorum*

	Location	Latitude	Longitude	No. of Individuals *Bombus terrestris*	*Bombus lucorum*
1	Baita Plai	46.46871	22.61674	21	0
2	Billed	45.91412	20.94701	9	0
3	Blandesti	47.71380	26.86323	33	0
4	Brebu	45.42815	21.97966	20	1
5	Burja	43.02797	25.32507	5	0
6	Carei	47.69646	22.48073	24	0
7	Cerna	45.15962	22.80671	4	16
8	Coastra	45.14758	24.22260	9	0
9	Corbeni	45.29905	24.60912	5	8
10	Dobrovat	46.99043	27.65404	24	1
11	Drăgusani	46.29929	26.97973	22	2
12	Föen	45.51085	20.87627	31	1
13	Golitsa	42.90956	27.52514	13	0
14	Gothal	45.40790	21.42069	29	0
15	Grohotno	41.70118	24.38684	10	25
16	Gura Glodului	47.13575	25.50107	2	20
17	Gura Haitii	47.17505	25.25018	0	15
18	Handal	47.65028	23.89441	0	21
19	Hlyabovo	42.06055	26.26459	11	0
20	Iesle	47.31038	25.89774	1	12
21	Iod	46.93652	25.00172	0	12
22	Kamenitsa	41.64449	23.17299	12	0
23	Koevtsi	43.15832	25.09082	21	0
24	Levochevo	41.60707	24.72302	2	28
25	Mengishevo	43.03566	26.64753	13	0
26	Ojdula	45.98988	26.29976	1	15
27	Orsova	44.75420	22.39480	14	2
28	Pastra	42.12283	23.20023	3	0
29	Pietroasa	46.58998	22.58807	12	0
30	Pirin	41.52480	23.58790	4	11
31	Poienita	45.82299	24.57591	19	1
32	Polovragi	45.21492	23.77486	0	12
33	Razdelna	42.18144	25.90854	4	0
34	Rilski Manastir	42.09243	23.38633	0	3
35	Rish	42.97442	26.90731	32	0
36	Sinemorets	42.04499	27.95808	11	1
37	Stambolovo	41.78435	25.63166	15	0
38	Strumeshnitsa	41.39833	23.06046	20	0
39	Topa Mica	46.92851	23.40238	21	0
40	Toplita	46.98115	25.40812	0	2
41	Valea Hotarului	47.93870	23.83761	20	1
42	Valea Pădurii	46.62236	24.02727	12	0
43	Zdravets	42.94361	24.15964	6	9
44	Zlatitza	42.70908	24.12053	3	6
	total			**518**	**225**

Species identification

Because of the previously described difficulties in distinguishing species based on external morphology, we used a 1,043 bp long fragment of the cytochrome c oxidase subunit I (COI) gene, which is known for its relatively fast mutation rate, and is used across many taxa for genetic identification purposes, including bumble bees (Bertsch et al., 2004; Bossert et al., 2016; Hebert et al., 2004; Murray et al., 2007; Waugh, 2007; Williams et al., 2012). This long fragment is completely overlapping with an 890 bp region used by Bertsch et al., 2004 to distinguish between the closely related *Bombus cryptarum, B. magnus,* and *B. lucorum* and was therefore used here to distinguish between *B. lucorum* and *B. terrestris*. DNA was extracted from one or both middle legs using the QIAmp DNA Micro Kit (Qiagen) following the manufacturer's protocol. PCR was performed with primers originally developed for *Apis* (Tanaka et al., 2001): forward 5'-ATAATTTTTTTTATAGTTATA-3' and reverse 5'-GATATTAATCCTAAAAAATGTTGAGG-3'. They were used to amplify a 1,043 bp long fragment of the COI gene. The PCR reaction mix consisted of 2.5 µl of 10× PCR Buffer S (Genaxxon), 15.4 µl HPLC water, 1.0 µl dNTP's, 1.0 µl MgCl2 (25 mM), 1.0 µl BSA, 1.0 µl of each primer (0.1 mM), 0.1 µl Taq polymerase, and 2 µl extracted DNA. PCRs were performed with a Mastercycler epgradient (Eppendorf) with the following conditions: an initial denaturation step at 95 °C for 1 min, followed by 55 cycles of a 3-step process: denaturation for 40 s at 95 °C, annealing for 1 min at 45 °C, and extension for 2 min at 60 °C with a final extension step at 60 °C for 4 min. PCR products were visualized using agarose gel electrophoresis to check for the amplification of the fragment. Successfully amplified PCR products were cleaned up using the Promega Wizard SV Gel and PCR Clean-Up System according to the manufacturer's protocol. Cleaned up samples were then sent to LGC Genomics for sequencing with the reverse primer only. Sequences were visualized and edited with Geneious R8 (Biomatters,(Kearse et al., 2012)). None of the obtained sequences showed any signs of cross-contamination (e.g., double peaks in the chromatograms or ambivalent species identification). We used two methods to assign a species to the sequenced samples. First, we blasted sequences in GenBank (NIH genetic equences database) and assigned the species with the highest identity (range 95%–100%) to the corresponding sample (Table S1). In addition, we created a phylogenetic tree (Figure S1), which included reference sequences for various Bombus species obtained from Genbank (Table 2). We included reference sequences from various geographic origins, because we expected that genetic intraspecific variability should be smaller than interspecific differences, and thus that a well-supported clustering of our samples with the reference sequences suggests high confidence in the species identification. To construct the phylogenetic tree, we first identified the most likely substitution model in MEGA-X (Kumar et al., 2018). We then created a Bayesian phylogenetic tree in Geneious R8 with the MrBayes module (Huelsenbeck and Ronquist, 2001), using one cold and four heated chains with a temperature of 0.2 and a chain length of 1.1 million and

a burn-in of 100,000, five gamma categories, a sampling frequency of 200, unconstrained branch lengths (GammaDir [1,0.1,1,1]), and exponential shape distribution. We used a reference sequence of *B. pascuorum* as an outgroup (Table 2). Individuals that were included in a monophyletic group with reference sequences of *B. terrestris* or *B. lucorum* were considered members of the corresponding target species. Comparisons to the blast results revealed no differences, and all individuals of *B. terrestris* and *B. lucorum* were included in subsequent species distribution models.

TABLE 2: Reference COI sequences obtained from GenBank and used for the phylogenetic tree.

Species	GenBank accession No.	Author
B. terrestris	AY181171	Pedersen, 2002
	AY181170	Pedersen, 2002
	AY181169	Pedersen, 2002
	KT164618	Tang et al., 2015
B. lucorum	AY181121	Pedersen, 2002
	AY181119	Pedersen, 2002
	AY181117	Pedersen, 2002
	KT164681	Tang et al., 2015
B. sporadicus	AF279500	Tanaka et al., 2001
	AY181163	Pedersen, 2002
	MF409659	Yu et al., 2017
B. cryptarum	AY181100	Pedersen, 2002
	AY530011	Bertsch et al., 2004
	AF279485	Tanaka et al., 2001
B. magnus	AY181123	Pedersen, 2002
	AY530014	Bertsch et al., 2004
	KC192046	Vesterlund et al., 2014
B. hortorum	AY181105	Pedersen, 2002
B. pratorum	AY181145	Pedersen, 2002
B. pascuorum	AY181141	Pedersen, 2002

Environmental variables

In order to create species distribution models, we used a set of environmental variables at 30 arcsec resolution (Table 3). We initially considered 19 climate variables from the WorldClim database (http://www.world clim.org), which included temperature and precipitation variables based on a 30-year climatology from 1970 to 2000 (Fick and Hijmans, 2017). Additionally, elevation data were acquired from the Shuttle Radar Topography Mission (SRTM; https://www2.jpl.nasa.gov), and used directly, as well as to compute aspect and slope. Because distribution patterns and habitat preferences of bumble bees have previously been suggested to relate to vegetation characteristics and forest cover, we also included spatial and temporal vegetation

patterns derived from satellite data. We used percent tree cover and Leaf Area Index (LAI, the one-sided green leaf area per unit ground area), which were both obtained from the Global Land Cover Facility database (http://www.glcf.umd.edu/data/). Information on the vertical forest structure, that is, canopy height, was derived from space-borne LiDAR from 2011 (Simard et al., 2011). Canopy height was found to be a better predictor for species distributions than other remote sensing variables such as canopy cover or land-use variables (Ficetola et al., 2014; Goetz et al., 2007), and we hypothesized that it may be related to forest understory flower availability and the presence and abundance of flowering tree species relevant for these bumble bee species. Finally, to include information about surface moisture, we included annual mean, minimum, maximum, and seasonality, computed from raw Quickscat data (Geue et al., 2016). To do so, we used daily raw backscatter measurements downloaded from the BYU Scatterometer Climate Record Pathfinder database (http://www.scp.byu.edu/data/Quiks cat/SIRv2 /qush/Eur.html) over the period the instrument was online (2000–2008). We excluded highly correlated variables, which we identified by means of their variance inflation factor (VIF). To do so, we used the automated stepwise exclusion procedure implemented in the "usdm" package v. 1.1-18 in R 3.6.1 (R Development Core Team, 2008), keeping only those variables with VIF < 10. The final data set consisted of 16 variables (Table 3).

Species distribution modeling

Spatial autocorrelation and sampling bias

Spatial autocorrelation is a major statistical challenge in spatial analyses, causing inflated measures of predictive power and incorrect distribution models (e.g. Guélat and Kéry, 2018; Segurado et al., 2006). There are two main causes for spatial autocorrelation in species distribution modeling.

First, there is often a spatial clustering of sampling sites. Reasons for such clustering are manifold and may be related to the sampling design (for instance ease of access, or issues with the logistics of evenly spaced sampling), or to the biotic and abiotic drivers of species distributions themselves, such as gaps in a species' range due to unsuitable habitat. Many approaches have been proposed to correct for sampling bias, among which subsampling locations to acquire a more even distribution of known presences are optimal in most cases (Fourcade et al., 2014). As a first step, we therefore removed one of the sites of pairs that were located within 20 km from one another. However, because in our study the number of locations is rather limited, further subsampling would result in an even smaller data set. Hence, in a second step, we instead weighted each location based on the density of known presences within a given radius, which was shown to be a good alternative to subsampling as a correction method (Fourcade et al., 2014; Stolar and Nielsen, 2015). To do so, we created a bias grid in QGIS 3.4.4. (Team et al., 2016) at 30 arcsec resolution, with

each grid cell representing the density of sampling locations within a 50 km radius, and kernel densities following a Gaussian distribution (Balestrieri et al., 2016). We used the inverse of the density to weight each presence and background location (see below), thus downweighting clustered locations. We not only included locations of known presence in this bias grid, but also locations where we searched for bumble bees with similar effort, but did not find any. We restricted these putative absence locations to those that were at least 50 km apart from known presences. We specifically only included these sites in our sampling bias map, and not in our models, because we cannot be sure that these represent true absences.

The second cause of spatial autocorrelation in SDMs is the often inherent spatial autocorrelation of habitat conditions, in particular climate variables. In this case, species occurrences are spatially dependent on the underlying habitat variables and thus represent a true association between species presence and local conditions. It is often impossible and undesirable to a priori remove spatial autocorrelation due to spatial dependence. Instead, spatial autocorrelation is expected to be absent in model residuals, regardless of the presence of initial spatial dependence. Models should correctly predict the presence or absence of a species at any given location, independent of its spatial relation to other locations. We thus tested for spatial autocorrelation in the probabilities of occurrence at known presence locations using global Moran's I in the R package ''lctools" v.0.2-7. We used four neighbors and tested the significance of correlations with resampling and randomization procedures.

Presence-only data

To model species distributions based on presence-only data, we used an ensemble method, which has been shown to perform better than any given individual modeling method (e.g. Araújo et al., 2007; Elith and Leathwick, 2009; Marmion et al., 2009. Because the number of known locations of species presence was limited, we employed the ensemble of small models approach implemented in the "ecospat" R package (Breiner et al., 2015; Breiner et al., 2018; Di Cola et al., 2017). Ecospat fits bivariate models of presence/(pseudo-)absence with two predictor variables at a time, creating an ensemble of "small" models weighted by each bivariate model's performance. It can do so for multiple modeling approaches, using the "Biomod2" package for R (Thuiller et al., 2009). Hence, for each modeling approach, bivariate (small) models are combined into a model ensemble, and model ensembles are in turn combined into an overall ensemble. We used ecospat v.3.1 and Biomod2 v.3.3-19 to run Maxent models (specifically the MAXENT.Phillips models, as implemented by Phillips et al., 2006, generalized linear models (GLM), classification tree analysis (CTA, also known as classification and regression trees (CART); Breiman et al., 1984, and artificial neural networks (ANN; Ripley, 2007). In a recent study comparing 10 different modeling approaches implemented in ecospat and Biomod2, these were shown to be

the top performing ones, while keeping computation times manageable (Breiner et al., 2018). We used model tuning to optimize the parameter settings for each model. We generated input files using presence-only sites and 5,000 background points that were sampled randomly at a minimum distance of 20 km from known presences. To correct for sampling bias, we extracted weights for all locations, which were implemented using the Yweights argument in ecospat.

TABLE 3: Environmental variables used for species distribution modeling and random forest analyses. Variables marked in bold were selected for our models after stepwise removal of variables with a variance inflation factor > 10.

Variable	Meaning	Source
Bio 1	Annual mean temperature	http://www.worldclim.org
Bio 2	Mean diurnal range mean of monthly (max temp – min temp)	http://www.worldclim.org
Bio 3	**Isothermality (Bio2/Bio7) * 100**	**http://www.worldclim.org**
Bio 4	**Temperature seasonality standard deviation*100**	**http://www.worldclim.org**
Bio 5	Max temperature of warmest month	http://www.worldclim.org
Bio 6	Minimum temperature of coldest month	http://www.worldclim.org
Bio 7	Temperature annual range (Bio5-Bio6)	http://www.worldclim.org
Bio 8	**Mean temperature of wettest quarter**	http://www.worldclim.org
Bio 9	**Mean temperature of driest quarter**	**http://www.worldclim.org**
Bio 10	Mean temperature of warmest quarter	http://www.worldclim.org
Bio 11	**Mean temperature of coldest quarter**	**http://www.worldclim.org**
Bio 12	Annual precipitation	http://www.worldclim.org
Bio 13	Precipitation of wettest month	http://www.worldclim.org
Bio 14	**Precipitation of driest month**	**http://www.worldclim.org**
Bio 15	**Precipitation seasonality (coefficient of variation)**	**http://www.worldclim.org**
Bio 16	Precipitation of wettest quarter	http://www.worldclim.org
Bio 17	Precipitation of driest quarter	http://www.worldclim.org
Bio 18	Precipitation of warmest quarter	http://www.worldclim.org
Bio 19	**Precipitation of coldest quarter**	**http://www.worldclim.org**
Elevation	Elevation	https://www2.jpl.nasa.gov/srtm/
Aspect	**Aspect**	**https://www2.jpl.nasa.gov/srtm/**
Slope	**Slope**	**https://www2.jpl.nasa.gov/srtm/**
LAI sd	**Leaf Area Index standard deviation across the year**	**http://www.glcf.umd.edu/data/**
LAI min	Leaf Area Index annual minimum	http://www.glcf.umd.edu/data/
LAI mean	**Leaf Area Index annual mean**	**http://www.glcf.umd.edu/data/**
LAI max	Leaf Area Index annual maximum	http://www.glcf.umd.edu/data/
Tree	**Percent tree cover**	**http://www.glcf.umd.edu/data/**
Canopy	**Canopy height**	**Simard et al., 2011**
QSCAT mean	**Surface moisture (mean)**	**http://www.scp.byu.edu, Geue et al., 2016**
QSCAT min	Surface moisture (min)	http://www.scp.byu.edu, Geue et al., 2016
QSCAT max	Surface moisture (max)	http://www.scp.byu.edu, Geue et al., 2016
QSCAT season	**Surface moisture (coefficient of variation)**	**http://www.scp.byu.edu, Geue et al., 2016**

To evaluate model performance, we computed various evaluation scores and used K-fold cross-validation with subsets of training and testing data. The Boyce index (Hirzel et al., 2006) is specifically designed and hence a particularly appropriate evaluation score for presence-only models. It is limited between -1 and 1, with 0 indicating model performance no better than random, and values close to 1 indicating high performance. We used the Boyce index to assess model performance, but also report the area under the receiver operator curve (AUC; Swets, 1988), Cohen's kappa (Cohen, 1960; Hirzel et al., 2006), and the true skill statistic (TSS; (Allouche et al., 2006)). To create training and testing data partitions for K-fold cross-validation, we used spatial blocking. Partitioning the data into spatial blocks has the advantage over random allocation of sites that it is better suited to evaluate model performance in the potential presence of spatial autocorrelation (e.g. Roberts et al., 2017). If a model performs well, it is expected to correctly predict occurrences in both distant as well as nearby locations (Telford and Birks, 2009). We generated spatial blocks of training and testing data with the R package ''blockCV'' v.2.0.0. (Valavi et al., 2019). We created fivefold and set the size of the spatial blocks to the median of the spatial autocorrelation range across the input environmental variables, which were sampled at 5,000 random locations. To find evenly dispersed folds, we ran 100 iterations.

Finally, to visually inspect species occurrence as a function of environmental conditions, we created two types of response curves. In the first, we plotted the response as a function of one environmental variable, while letting all other variables covary. These curves are particularly useful to understand the spatial patterns of species distributions. The curves cover the full range of responses, where the model takes advantage of sets of variables changing together. Second, we also plotted marginal response curves, where we plotted the effect of one environmental variable, while keeping all other variables at their sampled median. These curves are informative with respect to the individual contributions of each environmental variable.

Relative abundance data

To test whether the relative abundance of *B. lucorum* compared to *B. terrestris* is associated with environmental conditions, we used a machine learning approach implemented in the "SuperLearner" (v.2.0-25) package for the R statistical framework. SuperLearner uses model tuning to optimize model parameter settings and cross-validation to estimate the performance of multiple models. It creates optimized ensembles, weighted by the performance of the individual models. Where possible, we ran models similar to those for the presence-only data: generalized additive models (GAM; Hastie and Tibshirani, 1990), generalized linear models (GLM), Bayesian addi-tive regression trees (BART; Chipman et al., 2010), random forests (RF; Breiman et al., 1984; Breiman, 1996; Breiman, 2001), and neural networks (ANN; Ripley, 2007). We also ran a very simple mean-of-abundance model as a baseline. We corrected for

sampling bias using the weighting method described above, but we also fitted models to uncorrected abundance. We ran models on the full data set, where bagging and randomization were done internally, as well as on a partial data set, where we omitted 20% of the data, which were used as test data for independent cross-validation. For each model, we report its associated risk (a measure of model performance) and coefficient (the weight with which it is included in the ensemble). Response curves were created as described above.

To subsequently create a map of the predicted relative abundance of *B. lucorum* across the entire study area, we extracted the values for environmental variables at all 30 arcsec gridcells within Bulgaria and Romania. We then used the "predict.SuperLearner" function to project the known relationship between relative abundance and environmental conditions onto the entire landscape. These values were imported and converted to a raster format in QGIS 3.4.4 (Team et al., 2016).

Results

Species identification

The most likely substitution model was the General Time Reversible (GTR) model with gamma distribution, which we implemented to create the phylogenetic tree. We found that 514 individuals clustered with reference sequences of *B. terrestris* and 220 with those of *B. lucorum* (Table 1).

Presence-only data

Boyce indices for individual K-fold cross-validated models for *B. terrestris* ranged between 0.434 and 0.878 (median 0.751), suggesting overall decent to good model performance, except for those based on classification trees (CTA; Table 4). CTA models were therefore not included in the final ensemble. Boyce indices for ensemble cross-validated models ranged between 0.133 and 0.869. For *B. lucorum*, Boyce indices for individual cross-validated models ranged between 0.594 and 0.936 (median 0.766), and for ensembles between 0.650 and 0.870 (Table 4). CTA models performed as poorly as those for *B. terrestris* and were not included in the ensembles. Overall, models for *B. lucorum* performed slightly better than those for *B. terrestris*.

Spatial autocorrelation in the predicted occurrences was absent for *B. lucorum* (Moran's I = 0.08, expected I = -0.04, resampling z = 1.08, resampling p = 0.280, randomization z = 1.09, randomization p = 0.276). However, for *B. terrestris* we still found significant spatial autocorrelation, despite correcting for sampling bias (Moran's I = 0.41, expected I = -0.03, resampling z = 4.44, resampling $p < 0.001$, randomization z = 4.47, and randomization $p < 0.001$). We visually inspected the predictive map and

compared it to maps of important environmental variables. We found that particularly high probability of occurrence was predicted for sites in Mediterranean Bulgaria, which is consistent with the pattern of seasonality in surface moisture (QSCAT season), the most important variable in predicting the species' distribution. We suspected that the remaining spatial auto-correlation was the result of spatial dependence rather than of sampling bias. We further tested for residual spatial autocorrelation in a second analysis, where we also extracted the predictions for sites where we searched for bumble bees, but did not find any, despite making the same sampling effort. These sites were the same as those used to generate a sampling bias grid and were located at least 50 km from known presences. Although these sites were not included in the models as true absences, we expected that a well-performing model should predict low probability of occurrence for these sites. Indeed, this time we found no evidence for spatial autocorrelation (Moran's I = 0.12, expected I = -0.02, resampling z = 1.67, resampling p = 0.096, randomization z = 1.65, and randomization p = 0.099), and we concluded that sampling bias was sufficiently well corrected for.
Interestingly, the most important variables in limiting each species' distribution overlapped between species. The top four variables for *B. terrestris* were seasonality in surface moisture (QSCAT season), mean temperature of the wettest quarter (Bio 8), mean leaf area index (LAI mean), and temperature seasonality (Bio 4; Table 5). For *B. lucorum*, these variables comprised mean leaf area index (LAI mean), canopy height, seasonality in surface moisture (QSCAT season), and percent tree cover (Tree; Table 6), subsequently followed by mean temperature of the wettest quarter (Bio 8). For both species, the ranking of variables by their importance was largely consistent between modeling approaches. The main difference in the response curves between the two species is that those for *B. lucorum* are generally much steeper than those for *B. terrestris*, suggesting a stronger influence of the environment on *B. lucorum* (Figures 2 and 3). This difference is particularly pronounced for the top most important variables that were not overlapping between species, that is, percent tree cover and canopy height.
Consistent with the response curves, *B. terrestris* was predicted to be widely distributed, with medium suitability in lowland areas (in the north of Bulgaria and south of Romania) and low suitability in the Danube Delta and at the highest elevations (Figure 1b). Very high suitability was predicted for Mediterranean Bulgaria, south of the Balkan Mountains. Conversely, the range of *B. lucorum* was predicted to be much more restricted to higher elevations (the Carpathian Mountains and surrounding lowlands and the Balkan, Rila, Rhodope, and Pirin Mountains; Figure 1c).

TABLE 4: Performance scores of ESMs using presence-only data. Five cross-validated models were run based on spatial blocks generated with the R package 'blockCV'. MAXENT.P is the MAXENT.Phillips model. ENS is the ensemble of small models.

	B. terrestris				*B. lucorum*			
Model	**Boyce**	**AUC**	**Kappa**	**TSS**	**Boyce**	**AUC**	**Kappa**	**TSS**
RUN1_ANN	0.751	0.696	0	0	0.701	0.782	0	0
RUN1_CTA	-	0.5	0	0	-	0.5	0	0
RUN1_GLM	0.871	0.708	0	0	0.841	0.754	0	0
RUN1_MAXENT.P	0.553	0.69	0.014	0.36	0.716	0.774	0.027	0.594
RUN1_ENS	0.632	0.69	0.012	0.332	0.725	0.774	0.021	0.561
RUN2_ANN	0.737	0.599	0	0	0.825	0.796	0	0
RUN2_CTA	-	0.5	0	0	-	0.5	0	0
RUN2_GLM	0.694	0.665	0	0	0.936	0.862	0.179	0.327
RUN2_MAXENT.P	0.467	0.65	0.009	0.32	0.837	0.791	0.01	0.654
RUN2_ENS	0.555	0.65	0.027	0.262	0.87	0.796	0.008	0.611
RUN3_ANN	0.814	0.755	0.067	0.149	0.644	0.838	0	0
RUN3_CTA	-	0.5	0	0	-	0.5	0	0
RUN3_GLM	0.833	0.736	0.285	0.167	0.721	0.839	0	0
RUN3_MAXENT.P	0.434	0.71	0.007	0.413	0.766	0.818	0.024	0.638
RUN3_ENS	0.133	0.714	0.006	0.367	0.779	0.82	0.023	0.636
RUN4_ANN	0.764	0.782	0.136	0.27	0.808	0.688	0	0
RUN4_CTA	-	0.5	0	0	-	0.5	0	0
RUN4_GLM	0.636	0.785	0.13	0.355	0.755	0.662	0	0
RUN4_MAXENT.P	0.476	0.724	0.013	0.369	0.844	0.684	0.007	0.515
RUN4_ENS	0.635	0.74	0.015	0.368	0.854	0.685	0.005	0.459
RUN5_ANN	0.852	0.761	0	0	0.819	0.779	0	0
RUN5_CTA	-	0.5	0	0	-	0.5	0	0
RUN5_GLM	0.821	0.779	0.068	0.214	0.721	0.829	0.11	0.334
RUN5_MAXENT.P	0.878	0.768	0.021	0.458	0.594	0.815	0.055	0.677
RUN5_ENS	0.869	0.77	0.017	0.414	0.65	0.816	0.051	0.666

TABLE 5: Variable importance scores for ESMs based on presence-only data for *B. terrestris*. Scores for CTA are not included, because of its low model performance. MAXENT.P is the MAXENT.Phillips models. ENS is the ensemble of small models. See Table 1 for the meaning of the variables.

Variable	**ANN**	**GLM**	**MAXENT.P**	**ENS**
QSCAT season	0.140	0.099	0.088	0.109
Bio 8	0.104	0.088	0.098	0.096
LAI mean	0.083	0.067	0.088	0.079
Bio 4	0.055	0.080	0.085	0.073
Bio 3	0.047	0.080	0.091	0.073
LAI sd	0.068	0.057	0.060	0.062
Slope	0.071	0.053	0.062	0.062
Bio 11	0.049	0.059	0.064	0.057
Bio 9	0.063	0.055	0.050	0.056
Bio 19	0.045	0.063	0.059	0.056
QSCAT mean	0.079	0.042	0.032	0.051
Bio 14	0.056	0.048	0.046	0.050
Canopy height	0.038	0.053	0.054	0.049
Tree	0.036	0.062	0.037	0.046
Bio 15	0.040	0.049	0.040	0.043
Aspect	0.023	0.045	0.046	0.038

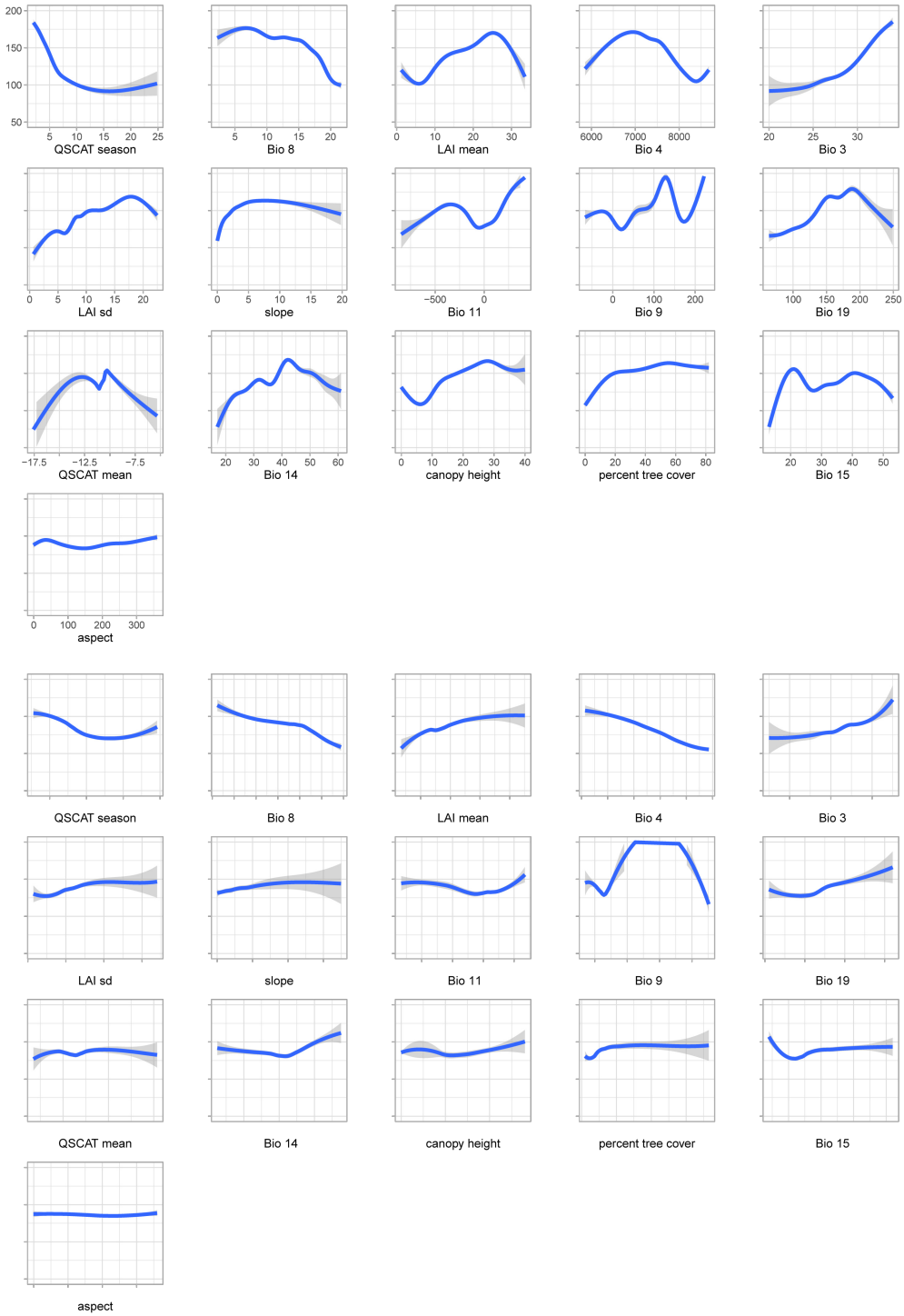

FIGURE 3: Overall (top panels) and marginal (bottom panels) response curves for presence-only model predictions for *Bombus terrestris*. Overall response curves were generated for each variable, while letting all other variables covary. In contrast, marginal response curves were created for each variable, while keeping all other variables at their median observed values

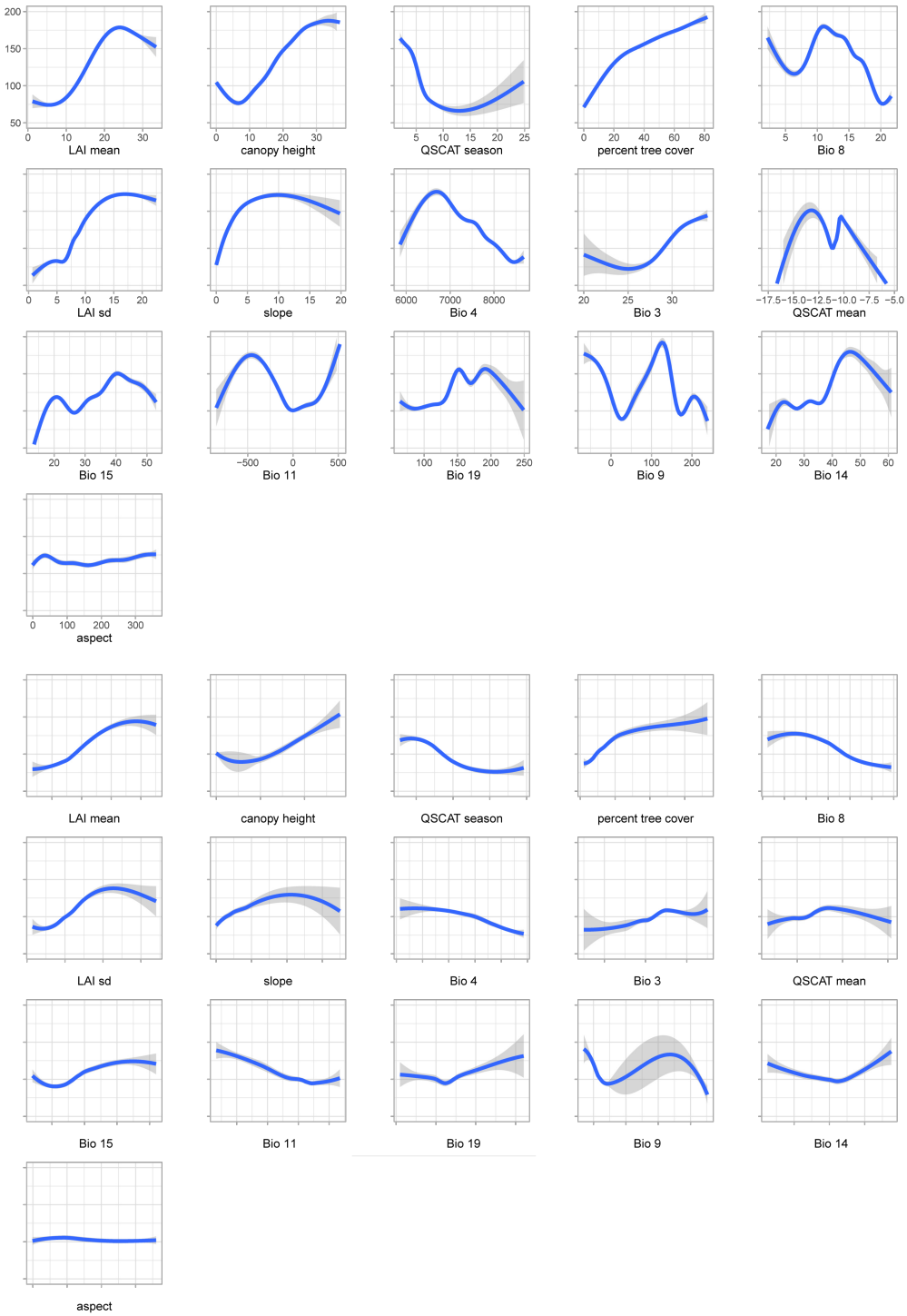

FIGURE 4: Overall (top panels) and marginal (bottom panels) response curves for presence-only model predictions for *Bombus lucorum*.Overall response curves were generated for each variable, while letting all other variables covary. In contrast, marginal response curves were created for each variable, while keeping all other variables at their median observed values.

TABLE 6: Variable importance scores for ESMs based on presence-only data for *B. lucorum*. Scores for CTA are not included, because of its low model performance. MAXENT.P is the MAXENT.Phillips models. ENS is the ensemble of small models. See Table 1 for the meaning of the variables.

Variable	ANN	GLM	MAXENT.P	ENS
LAI mean	0.091	0.088	0.092	0.090
Canopy height	0.082	0.089	0.083	0.085
QSCAT season	0.087	0.079	0.080	0.082
Tree	0.086	0.075	0.080	0.081
Bio 8	0.068	0.081	0.073	0.074
LAI sd	0.078	0.069	0.075	0.074
Slope	0.080	0.055	0.066	0.067
Bio 4	0.060	0.062	0.055	0.059
Bio 3	0.063	0.054	0.045	0.054
QSCAT mean	0.072	0.048	0.042	0.054
Bio 15	0.048	0.052	0.056	0.052
Bio 11	0.041	0.051	0.054	0.049
Bio 19	0.040	0.053	0.052	0.048
Bio 9	0.040	0.047	0.052	0.046
Bio 14	0.038	0.045	0.046	0.043
Aspect	0.027	0.051	0.048	0.042

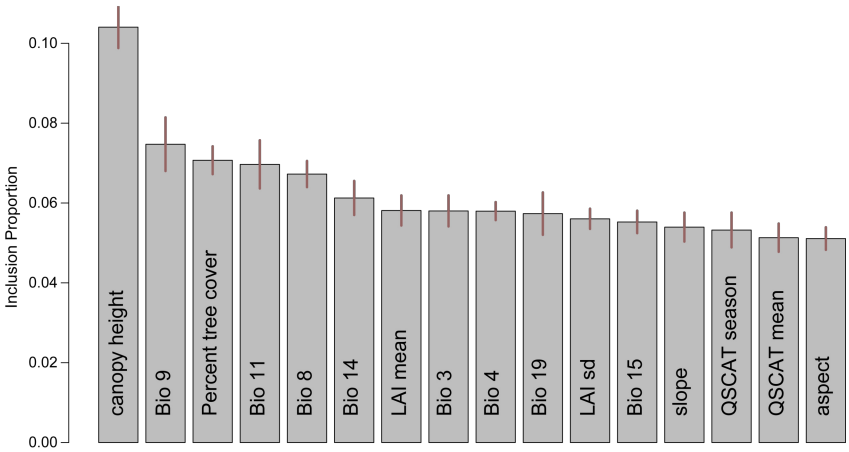

FIGURE 5: Variable importance inferred from a BART model for the relative abundance of *Bombus lucorum*. This model had a coefficient > 0.9 in the ensemble model, and it was the single best performing one in a nested crossvalidation analysis. We therefore used its robust estimate of variable importance to assess the contribution of each variable in the overall ensemble.

Relative abundance data

Models of relative abundance that were corrected for sampling bias performed considerably worse than uncorrected models (corrected models: BART coefficient = 0.885, BART risk = 0.404, GLM coefficient = 0.115, and GLM risk = 0.682). We therefore report results for uncorrected models from here onwards. The only two models included in the ensemble were GAM (coefficient = 0.084, risk = 0.199) and BART (coefficient = 0.916, risk = 0.093). K-fold nested cross-validation with fivefold suggested that the single best model was BART, which performed even slightly better than the ensemble model, yet statistically nonsignificant (Table 7). Because of the high weight of the BART model, we evaluated variable importance based on BART only, providing a robust posterior importance score (Chipman et al., 2010; Hernández et al., 2018). The top most important variable was canopy height, subsequently followed by percent tree cover and three temperature variables (Figure 4), which is broadly consistent with the results for the presence-only data. Overall and marginal response curves suggest that *B. lucorum* is more abundant in more densely vegetated, wet and cool areas (Figure 5).

To gain further insight in how our two target species differ in their preferred habitat conditions, we visually inspected scatter-plots of the relative abundance of *B. lucorum* as a function of the most important variable, canopy height. We noted that the major mountain ranges in Romania and Bulgaria are a prominent feature in our distribution maps, which is consistent with previous descriptions of occurrence patterns. Although we dropped elevation from our analyses because of its high VIF, we also created a scatter plot of relative abundance versus elevation. Visual inspection of these plots suggested that *B. lucorum* does not occur in unforested areas with a canopy height under 20 m (Figure 6a). Yet, the dichotomy between species is particularly striking for elevation, where *B. lucorum* is almost completely absent below 600 m, but makes up the majority of the two species at higher elevations (Figure 6b). Hence, elevation captures the combined influence of correlated environmental variables on limiting the distribution of *B. lucorum* particularly well.

TABLE 7: Risk scores of five-fold cross-validated models of relative abundance, with mean, standard error, minimum, and maximum values. The lower the risk, the better the model performance. The single best model was the BART model. SuperLearner is the ensemble of all models.

Algorithm	Mean	SE	Min	Max
SuperLearner	0.096303	0.020817	0.04578	0.15106
MEAN	0.162742	0.022342	0.130919	0.22429
GLM	0.137423	0.033966	0.058933	0.22817
GAM	0.137423	0.033966	0.058933	0.22817
BART	0.095774	0.019007	0.052584	0.1495
RF	0.096686	0.021774	0.043806	0.15481
ANN	0.162742	0.022342	0.130919	0.22429

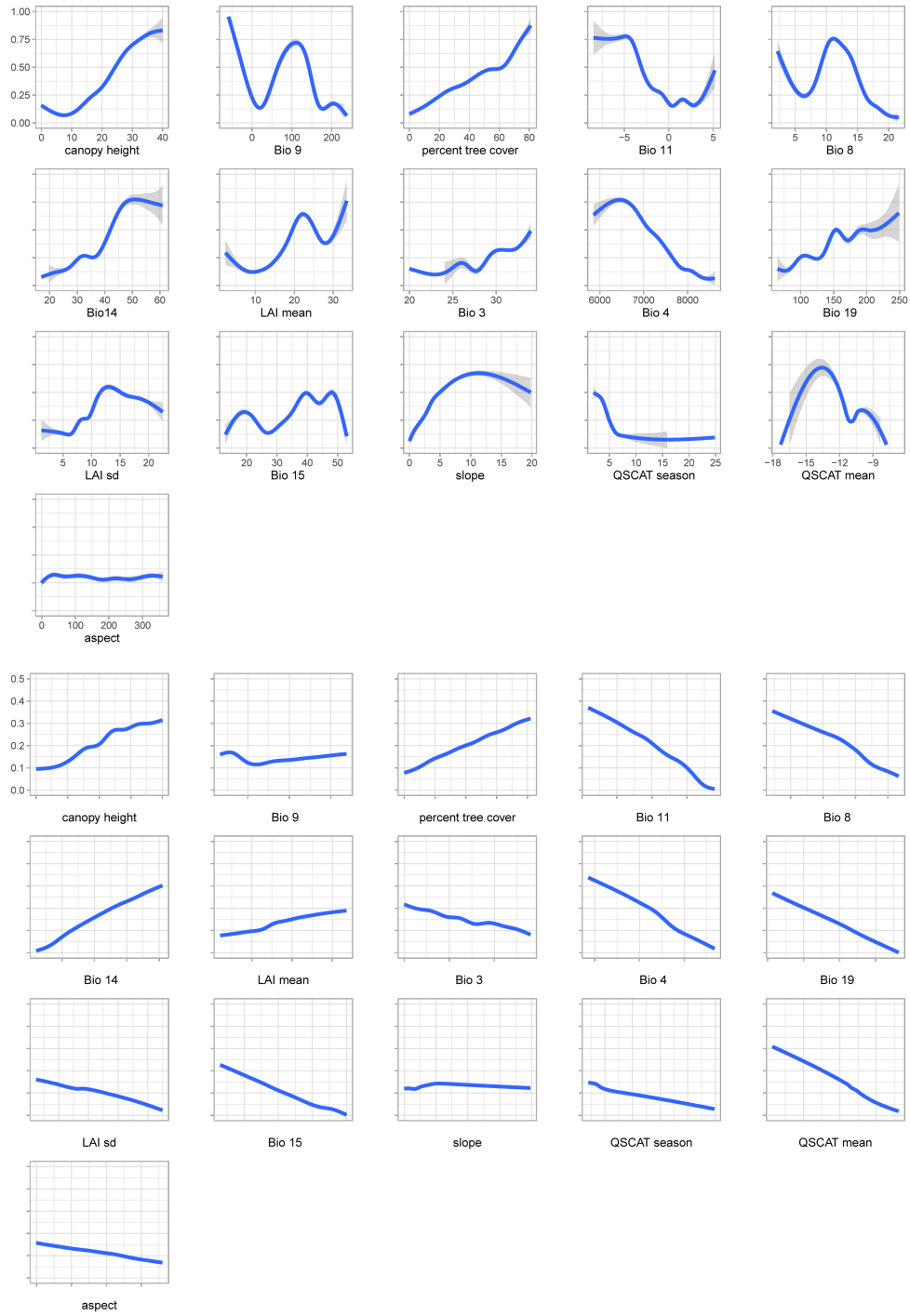

FIGURE 6: Overall (top panels) and marginal (bottom panels) response curves for relative abundance model predictions. Overall response curves were generated for each variable, while letting all other variables covary. In contrast, marginal response curves were created for each variable, while keeping all other variables at their median observed values.

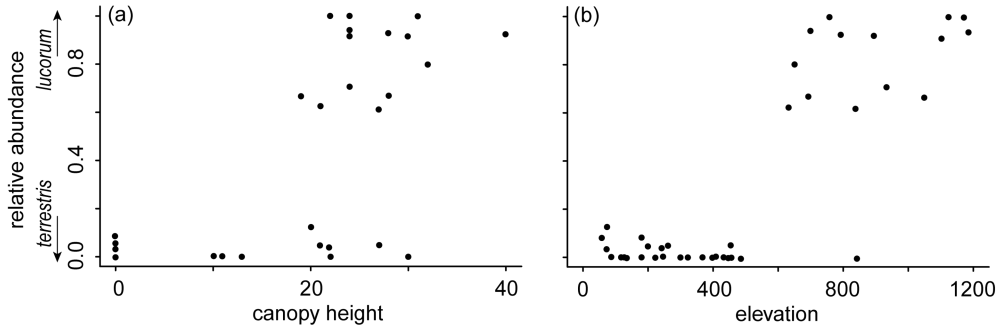

FIGURE 7: Scatterplots of the observed relative abundance of *Bombus lucorum* as a function of (a) canopy height, the most important variable in the BART model, and (b) elevation. Elevation was not entered in our models, because it was correlated to many environmental variables.

Discussion

Here, we modeled the distributions and relative abundance of two cryptic bumble bee species in Bulgaria and Romania from samples that were assigned to one of the species using a long fragment of the COI gene for genetic identification. We demonstrated that even though there is a certain degree of overlap between the ranges of the two species, *B. lucorum* has a much more restricted distribution than *B. terrestris*. Our models suggested that both vegetation and climate variables are key factors in determining their distributions. These results are concordant with previous studies for *B. lucorum* suggesting that it prefers closed habitat near forests (Bossert et al., 2016; Svensson et al., 2000). Our findings also support observations that *B. lucorum* occurs at higher elevations (Ban-Calefariu and Sárospataki, 2007; Bossert et al., 2016; Goulson et al., 2008a; Ploquin et al., 2013; Tomozei, 2006), which was suggested to be the result of an adaptation to colder climates (Benton, 2006). Indeed, in northern Europe, *B. lucorum* generally occurs in colder areas, where it at least partly substitutes *B. terrestris* (Rasmont et al., 2015a). Many environmental variables change along an elevation gradient, and elevation itself is unlikely to determine the distribution of these species, but rather its covariates (Bossert et al., 2016). In our study, mean temperature of the coldest quarter (Bio 11), canopy height, percent tree cover, and mean leaf area index (LAI mean) were particularly highly correlated with elevation (Pearson correlation coefficient > 0.6). Hence, the distribution of *B. lucorum* is clearly restricted to the mountainous areas in Bulgaria and Romania (Figure 1c,d), where temperatures are lower, precipitation is higher, and where most of the forest is remaining.

In contrast, our findings for *B. terrestris* suggest that it is not as restricted to open habitat as previously thought (Bossert et al., 2016; Svensson et al., 2000), but rather is a generalist species, occurring in open as well as more densely vegetated areas. This notion is also apparent in our maps, showing a wide distribution for *B. terrestris*.

Interestingly, our presence-only species distribution models showed a considerable overlap between the ranges of *B. terrestris* and *B. lucorum*, but analyses of their relative abundance evoke a much stronger separation between these species. Although the use of relative abundance does not allow for conclusions regarding the absolute abundance of either one of the species, the large range of relative abundance values for *B. lucorum*, spanning from 0 to 0.94, suggests that the two species replace one another across an environmental gradient. Thus, the inclusion of abundance enhances our understanding of the distribution of these species based on presence-only models.

Although the conservation status of our study species across the European continent is ''least concern" (Rasmont et al., 2015b; Rasmont et al., 2015c), they are ''vulnerable" or ''nearly threatened" in a few countries (Kosior et al., 2007). A previous cross-continent study suggested that both *B. terrestris* and *B. lucorum* may suffer from range contractions under future climate change (Rasmont et al., 2015a). The study by Rasmont et al., 2015a provides a great overview of overall distributions and risks posed by climate change. Yet, such large-scale models of species distributions, spanning major latitudinal and environmental gradients, and based on climate variables only, may be of limited use at intermediate to smaller spatial extents. Indeed, we found that vegetation characteristics were among the most important variables explaining the distribution and relative abundance of our study species, and it will be difficult to predict how these variables will change in the future, both as a result of climate change, as well as due to changes in forest management. We did not proceed with an attempt to predict the distribution of *B. lucorum* onto future climate conditions, because a model based on only current climate conditions failed to even broadly resemble that based on both climate and vegetation variables (not shown). Moreover, populations are likely adapted to local and regional conditions, and may not respond the same to changing environmental conditions. Our study provides further insight by teasing apart the habitat preferences of these species in southeastern Europe, providing higher resolution range maps that are probably more relevant for the region, where the distribution of *B. lucorum* is assumed to be rather disjunct. Despite the complexity of predicting future changes in vegetation characteristics, the difference in habitat requirements between these species is expected to have implications for the way they respond to changing climate conditions. Our finding that *B. lucorum* is rather restricted in its suitable habitat conditions compared to *B. terrestris*, may suggest that it is more vulnerable to climate change than the latter.

We genetically identified a large number of individuals of two closely related bumble bee species sampled at many sites and modeled their distributions and gained insight into their habitat requirements. We showed that *B. terrestris* is rather a generalist species, whereas *B. lucorum* is restricted to colder and wetter climates in forested areas, which in our study area primarily occur at higher elevations. We support the emerging recognition of the importance of abundance data in species distribution

modeling. Despite the overlap in occurrence suggested by presence-only data, their relative abundance gradually changes along a major environmental gradient, with one of the species being virtually absent at the extreme ends of this gradient. Our study contributes to the urgent need to fill a major gap of knowledge in the distribution and ecology of these species that can help facilitate the assessment of their conservation status and the development of management plans where necessary.

Acknowledgements

We are particularly grateful to Frank Breiner, who advised on ESM modeling and implemented new functionality in ecospat in response to our questions. We thank Isabel Fraccaroli for help in the molecular laboratory, and two anonymous reviewers, whose contribution much improved a previous version of this manuscript. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013 under REA grant agreement number 293886 to HT. JG was supported by the State Postgraduate Fellowship Programme of Baden-Würrtemberg, Germany, 2016-2018, Award Number: I 1.3-7631.2.

3 CHAPTER II

LIMITED RECIPROCAL SURROGACY OF BIRD AND HABITAT DIVERSITY AND INCONSISTENCIES IN THEIR REPRESENTATION IN ROMANIAN PROTECTED AREAS

Julia C. Geue, Paula J. Rotter, Caspar Gross, Zoltán Benkő, István Kovács, Ciprian Fântânâ, Judit Veres-Szászka, Cristi Domșa, Emanuel Baltag, Szilárd J. Daróczi, Gábor M. Bóné, Viorel D. Popescu and Henri A. Thomassen

Abstract

Because it is impossible to comprehensively characterize biodiversity at all levels of organization, conservation prioritization efforts need to rely on surrogates. As species distribution maps of relished groups as well as high-resolution remotely sensed data increasingly become available, both types of surrogates are commonly used. A good surrogate should represent as much of biodiversity as possible, but it often remains unclear to what extent this is the case. Here, we aimed to address this question by assessing how well bird species and habitat diversity, two frequently used biodiversity surrogates, represent one another. We conducted our study in Romania, a species-rich country with high landscape heterogeneity where bird species distribution data have only recently started to become available. First, we prioritized areas for conservation based on either 137 breeding bird species or 36 habitat classes, and then evaluated their reciprocal surrogacy performance. Second, we examined how well these features are represented in already existing protected areas. Finally, we identified target regions of high conservation value for the potential expansion of the current network of reserves (as planned under the new EU Biodiversity Strategy for 2030). We found that bird species were a better surrogate for habitat diversity than vice versa. Highly ranked areas based on habitat diversity were represented better than areas based on bird species, which varied considerably between species. Our results highlight that taxonomic and environmental data may be poor surrogates for one another and that different types of biodiversity surrogates should be combined in spatial conservation prioritization.

KEYWORDS
biodiversity surrogate, bird species, habitat type, spatial conservation prioritization, Zonation software, Romania

Introduction

The ultimate goal of conservation prioritization is the protection of biodiversity at all levels of organization (Pressey, 2004). However, limited financial resources and competing stakeholder interests constrain the area that can reasonably be protected. The process of identifying potential regions for designation as protected area (PA) should therefore be undertaken thoroughly and strategically (Bottrill et al., 2008; Joseph et al., 2009), see Margules and Pressey, 2000 for a review). The striking obstacle is however that biodiversity is very complex and difficult to characterize (Noss, 1990), and surveying biodiversity in its entirety is nearly impossible. Shortcuts necessarily need to be taken to quicken the prioritization process and make it more feasible (Andelman and Fagan, 2000). One of these shortcuts is using a biodiversity or environmental indicator as a conservation surrogate (see Margules and Pressey, 2000 for a review; (Sarkar et al., 2005), which is: "An ecological process or element (e.g., species, ecosystem, or abiotic factor) that [should] [...] represent (i.e., serve as a proxy for) another aspect of an ecological system" (Hunter et al., 2016). The efficacy and efficiency of surrogates for overall biodiversity (known and unknown) have progressively been evaluated (Araújo et al., 2007; Di Minin and Moilanen, 2014; Gaston et al., 2008; Oliver et al., 2004; Sarkar et al., 2005; Sauberer et al., 2004), and appear to be influenced by factors such as the size of the study area, type of surrogate, and the spatial resolution of surrogate data (e.g. Di Minin and Moilanen, 2014; Franco et al., 2009; Grantham et al., 2010). Nevertheless, it often remains ambiguous to what extent a surrogate represents other levels of biodiversity, in particular across different levels of organization.

Biodiversity surrogates are usually subdivided into taxonomic and environmental surrogates (Grantham et al., 2010; Oliver et al., 2004; Sarkar et al., 2005). Many studies have evaluated the efficacy of taxonomic surrogates for other taxonomic groups (e.g. see Andelman and Fagan, 2000; Caro and O'Doherty, 1999 for a review; (Lawler et al., 2003; Lund and Rahbek, 2002; Rozylowicz et al., 2011; Sibarani et al., 2019; Wiens et al., 2008). The general consensus is that one taxonomic group alone might not be an adequate surrogate for others (Bertrand et al., 2006; Billeter et al., 2008; Di Minin and Moilanen, 2014; Franco et al., 2009; Moritz et al., 2001); see Rodrigues and Brooks, 2007 for a review), and the identification of PAs should include more than one species or taxonomic group (Franco et al., 2009; Larsen et al., 2012). Yet again, for many areas in the world accurate species distribution data is scarce. However, one of the taxonomic groups for which rich datasets are available are birds, because they are of interest to many people and are therefore one of the best surveyed taxa in the world (Garson et al., 2002; Larsen et al., 2012; Veríssimo et al., 2009). As such, birds are often used as biodiversity indicators and conservation surrogates, and their surrogacy effectiveness varies from representing overall species diversity well (other taxa than birds) (Juutinen and Mönkkönen, 2004; Larsen et al., 2012; Rodrigues and Brooks, 2007; Sauberer et al., 2004), or threatened birds being adequate surrogates for

non-threatened bird species (Franco et al., 2009), to being poor surrogates for other taxa (Lund and Rahbek, 2002; Moore et al., 2003; Williams et al., 2006). Adding more taxa (Larsen et al., 2012) or even different biodiversity features, such as environmental diversity (Bonn and Gaston, 2005; Di Minin and Moilanen, 2014), increased the overall surrogacy of birds for other levels of biodiversity.

Environmental diversity, in particular habitat diversity, has the potential to be a powerful surrogate and represent other levels of biological organization, because habitat data can be generated quickly and relatively inexpensively from remotely sensed or extrapolated ground data (Arponen et al., 2008; Beier et al., 2015; Grantham et al., 2010; Rodrigues and Brooks, 2007; Sarkar et al., 2005). Furthermore, environmental surrogates may capture interactions between species and their environment (Bonn and Gaston, 2005), and compensate for a potential lack of congruence between taxonomic surrogates (Moritz et al., 2001). However, compared to taxonomic surrogates, the application of environmental surrogates received less attention. Results suggest that continuously distributed environmental variables (e.g. climate variables such as temperature and precipitation, or vegetation characteristics such as percent tree cover) may not be adequate (Araújo et al., 2001; Di Minin and Moilanen, 2014; Rodrigues and Brooks, 2007) or at most better than random surrogates for species occurrence (Sarkar et al., 2005). Categorical environmental data in the form of pre-classified information (e.g. land classes, ecological vegetation classes or habitat types) may be better surrogates than continuous environmental data. However, habitat or land cover categories vary in their representation of other levels of biodiversity, for instance weak for plant species (Bonn and Gaston, 2005; Carmel and Stoller-Cavari, 2006), but better for plants than for vertebrates (Di Minin and Moilanen, 2014; Grantham et al., 2010; Lombard et al., 2003; Mac Nally et al., 2002; Oliver et al., 2004). Yet, such contrasting results could also result from differences in the spatial extent and resolution of the study area, as well as the type of environmental data used as a surrogate (vegetation or climate-based) (Grantham et al., 2010; Hess et al., 2006; Margules and Pressey, 2000; Reyers et al., 2002).

Given uncertainties surrounding the potential for categorical habitat data to serve as a surrogate for biodiversity, the goal of our study was to evaluate its representation of one of the most frequently used biodiversity surrogates, bird species distributions, and vice versa. We implemented this analysis for Romania, a country within the European Union exhibiting high bird species and habitat diversity, likely caused by the variety of biogeographic regions it comprises (Ioras, 2003; Schmitt and Rákosy, 2007). While 23% of Romania is protected, either under the pan-European Natura 2000 network or as natural or national parks or biosphere reserves (Niculae et al., 2017), and despite its high levels of biodiversity, efforts to identify conservation priorities and evaluate the efficacy of the network of reserves to protect biodiversity have been sparse (mentioned by Gaston et al., 2008 but not examined; Iojă et al., 2010; Miu et al., 2018; Niculae et al., 2017; Popescu et al., 2013; Rozylowicz et al., 2011). One reason

for this disparity is that species distribution data have only recently become widely available. As such, PA management could greatly benefit from prioritization efforts using systematic conservation planning principles and the latest available data, particularly when establishing new PAs (Iojă et al., 2010; Niculae et al., 2017). The implementation of such scientific research in the establishment and governance process of PAs is, however, often limited (Opermanis et al., 2014; Popescu et al., 2014). This is not a unique situation, as for instance Natura 2000 sites consist of a diverse array of reserves designed for particular species, but not to protect biodiversity as a whole, so they often represent species and habitat diversity only to a limited extent (Araújo et al., 2001; D'Amen et al., 2013; Iojă et al., 2010; Pechanec et al., 2018). Furthermore, the European Commission decided to set new targets for 2030 and increase the percentage of protected areas in EU member states to 30% (Commission, 2020). Hence, there is a need to identify additional areas for protection, which is best done using the principles of systematic conservation planning (Miu et al., 2020).

In this study, we first evaluated whether breeding bird species and habitat diversity based on remote-sensing data are adequate surrogates for one another. We assessed surrogacy of the two datasets using high-resolution data (1km) of (a) 137 modelled breeding bird species distributions and (b) 36 classes of mapped habitat types from the "Ecosystem Types of Europe" (ETE) data set (Agency, 2016). Second, we evaluated whether existing protected areas (national and natural parks, biosphere reserves, wetland reserves and SPAs (as part of the Natura 2000 network)) in Romania are effective in representing areas of conservation concern for both birds and habitats. Finally, we identified additional areas that could be prioritized in an effort to expand the current PA network to more comprehensively protect bird and habitat diversity.

Methods

Study region

Romania is located in Eastern Europe, at the western shores of the Black Sea. It covers 238 397 km^2 and natural landmarks and borders are dominated by the Carpathian Mountains and the Danube River. Five biogeographical regions have been characterized across Romania: Pannonian, Continental, Alpine, Steppic, and Black Sea. The heterogeneous landscape consists of an alternation between intensive and extensive agricultural areas and (semi-) natural areas, such as forest, open woodland, and grassland.As a member of the European Union, Romania is bound to the directives of the Natura 2000 network, and dedicated about 23% of its total landscape to conservation. The Natura 2000 network is an important biodiversity conservation measure (Gaston et al., 2008), and consists of different types of protected areas: the terrestrial Special Protection Areas (SPA, for bird protection only), the terrestrial Sites of Community Importance, and Special Areas of Conservation (SCI and SAC, for habitats

and/or species) (Commission, 2007)). In addition to, but partly overlapping with the Natura 2000 network, Romania also implemented protected areas designated as natural and national parks as well as biosphere reserves (Niculae et al., 2017).

Biodiversity features

Bird species distributions

Bird species occurrence data (from the years 2006-2018) were obtained from the forthcoming Romanian Breeding Bird Atlas (Fântânâ and Kovács, 2020, in preparation), a scheme run by Milvus Group Association and the Romanian Ornithological Society. We modelled the distributions of 137 breeding bird species using MaxEnt v.3.4.1 (Phillips et al., 2017) at two different resolutions (1km and 2km), depending on the species' ecology or in some cases by the available data (Table A.2, Appendix A). Appendix B provides in-depth details on the species distribution modeling approach.

Habitat types

We used the published maps of habitat types classified in the "Ecosystem Types of Europe" (ETE) data set (version 3.1)(Agency, 2016). ETE is a combination of the non-spatially referenced EUNIS (European Nature Information System) habitat classification scheme and a spatially explicit habitat data set, the Corine-based "Mapping and Assessment of Ecosystem and their services (MAES)" ecosystem classes (Weiss and Banko, 2018). In Romania, 42 ETE habitat classes are mapped (level 2 classification) at 100 m resolution (Table A.3, Appendix A). Habitat classes including highly built-up areas (six classes) were excluded in the subsequent spatial conservation mapping. These built-up areas where selected according to the ETE classification category "J" (J1-J6, see Table A.3, Appendix A), which include buildings in cities and villages, industrial sites, transport networks, artificial water structures and waste deposits. To produce maps of habitat types that match the spatial resolution of those for the bird species, we split the ETE data set into single data layers per class (36 in total) and calculated the proportion of each habitat type within 1 km^2 grid cells.

Data handling

All spatial data layers were re-projected to the Dealul Piscului 1970/ Stereo 70 projection and processed at a 1 km resolution containing a total number of 381 248 grid cells. Species distribution models at 2 km resolution were resampled to 1 km grid cell size. Preparation of input maps and post-processing of results was done in R (version 3.6.1), using the packages (zonator, raster, rgdal, rgeos, sp, maptools, tiff, data.table, plyr, dplry, ggplot2, zoo). Maps were visualized in QGIS (version 3.10.6 'A Coruña').

Spatial conservation prioritization

We prioritized areas for conservation using the software Zonation 4.0 (Moilanen and Kujala, 2006). Zonation can handle large data sets (Kremen et al., 2008) and provides a priority ranking over the entire landscape rather than satisfying a specific target. The ranking is produced by iteratively discarding locations (grid cells) with the lowest conservation values, retaining the ones with the highest conservation value throughout the process (Di Minin and Moilanen, 2012; Moilanen et al., 2005).

We used the additive-benefit function (ABF), which directly sums up the conservation value across features (Moilanen et al., 2011) and results in a reserve network with high average performance across all features (Arponen et al., 2005). The ABF algorithm is appropriate for our study since we aim to identify areas representing overall richness rather than core areas that lead to the equal representation of both common and rare species or habitats. The algorithm accounts for the total and remaining distributions of features, and optional feature-weights can be implemented (Di Minin and Moilanen, 2014). We equally weighted habitat types and bird species distributions at the aggregate level to avoid prioritization biases due to the different numbers of features contained within (e.g., combined weights for 137 bird species or for 36 habitat types summed to 1). To exclude land uses that for administrative or ecological reasons did not contribute to either overall conservation value or to the expansion of protected areas (six classes of built-up area), we applied a cumulative negative weight of -1 to these layers (Moilanen et al., 2011). These build up areas where selected according to the ETE classification category "J" (J1-J6), which include buildings in cities and villages, industrial sites, transport networks, artificial water structures and waste deposits.

Performance curves were produced with the R package 'zonator' (Lehtomäki, 2016-2018). These curves show the proportion of the original occurrence of features remaining in the landscape as a function of the proportion of the landscape that is lost (Lehtomäki and Moilanen, 2013). The curves start at 1.0, where the entire distribution of features is represented in the full landscape, and end at 0.0, where the entire landscape is lost. Because we observed a wide spread in the performance curves of the bird species, we explored potential underlying patterns related to their broad habitat requirements. We grouped species into their preferred breeding habitat types (Table A.2, Appendix A) to assess differences between groups and their performance when the prioritization is accounting for all bird species. We also suspected that the range size of feature types, in particular within bird species, influences their performance in the prioritization. Specifically, we assumed that range restricted species would perform better, since these species might be retained throughout many prioritization iterations. Yet, this may only be the case when range-restricted features largely overlap with more widespread features. To explore this further, we calculated the AUC (area under the curve) of each feature performance curve, and plotted these as a function of range size (Figure 4). For bird species we calculated range sizes by summing

the Maxent probabilities, and for habitat types we summed the area in km^2.

Surrogacy analyses

We evaluated the reciprocal surrogacy of bird species and habitat types, and assessed the efficacy of the existing network of protected areas to protect these biodiversity features. To test the surrogacy of the two feature types, we ran separate analyses using one feature type as the surrogate and the other as the target. To do so, bird species and habitat types were both included in each run, but positive weights (=1) were only assigned to the surrogate, while the target was assigned a weight of 0.

We evaluated the surrogacy power of each feature type using the performance curves. A performance curve by itself provides, however, little information, and for correct interpretation it should be compared to an optimal and a random curve (Rodrigues and Brooks, 2007). For instance, when testing whether habitat types are a good surrogate for bird species, the optimal curve is equivalent to the surrogacy of bird species for themselves. The random curve in this scenario reflects the representation of bird species expected in the absence of biological data, when 'area' is used as a surrogate (Rosenzweig, 1995). Qualitatively the surrogacy value can be assessed visually by comparing the three curves. The closer the target curve is to the optimal curve, the higher the surrogacy value. To quantify the surrogacy power, we calculated an equivalent to the species accumulation index (SAI; Ferrier, 2002):

$$SAI = (S\text{-}R)/(O\text{-}R),$$

where S is the area under the target curve, R is the area under the random curve, and O is the area under the optimal curve. The optimal curve was extracted from the runs when targets were used as a surrogate themselves. To create the random curve, we executed 100 surrogacy runs with randomly, uniformly distributed data as a surrogate and bird species and habitat types as targets. We used the mean of the corresponding target curves to calculate SAI.

Evaluation and potential expansion of protected area network

To evaluate the representation of habitat and birds in existing reserves, we specifically focussed on SPAs, national and natural parks, and biosphere reserves. We thus excluded the SCI and SAC areas (Natura 2000 sites), since they are designed to protect specific species or habitats, but do not necessarily protect others - or even biodiversity as a whole. To evaluate the effectiveness of the current network in Romania, we tested 1) how well current PAs represent areas of conservation concern for bird species and habitat types, and 2) how much of the individual feature type's distributions are represented within the current network. Furthermore, we 3) assessed which areas should be prioritized when expanding the current conservation network.

The analyses for 1) and 3) were based on a Zonation prioritization outputs, where both bird species and habitat types had been considered simultaneously. We did not differentiate between protection levels of the existing PAs. If PAs had been selected indiscriminately, we expected that Zonation values within PAs would be uniformly distributed, as they are across the entire study region. We thus tested the frequency of Zonation values within PAs against a uniform distribution using a Chi-square test. For 2) we summarized the distribution of bird species and habitat types within current PAs as a proportion of their total distribution via boxplots (Figure A.3, Appendix).

To identify potential areas that should be prioritized when expanding the current network of PAs, we performed a mask analysis (Moilanen et al., 2014). In this analysis, current PAs are included as a mask layer, and are assigned a high rank (=1) in the final prioritization map. As such, the next highly ranked areas outside protected areas can be identified as potential expansion areas that represent bird and habitat diversity well.

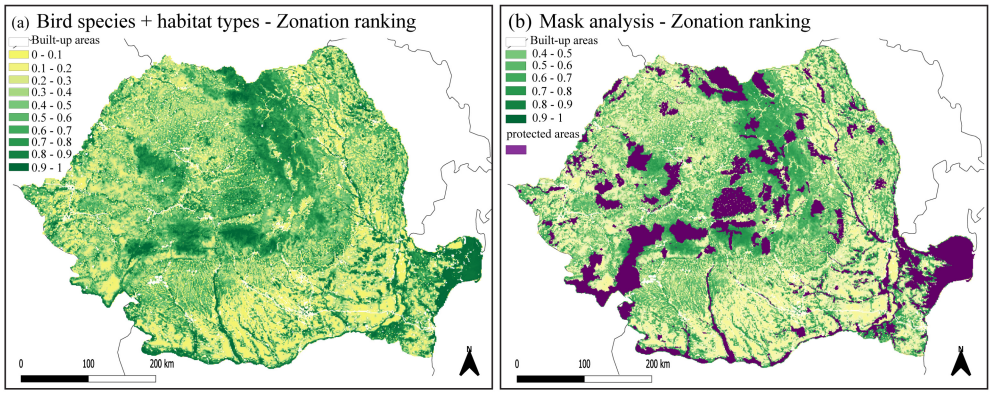

FIGURE 1: Study region with Zonation ranking based on bird species and habitat type data without (a) and with (b) considering currently protected areas (mask analysis). Colors indicate importance ranking scores for conservation, with 0 meaning lowest importance and 1 meaning highest importance. Built-up areas are indicated in white and were excluded from prioritization. Purple in panel (b) indicates current protected areas.

Results

Spatial conservation prioritization

Both the separate and combined prioritization using bird species and habitat types resulted in broadly similar patterns, with highly ranked areas in the Carpathian Mountains, river valleys and parts of the Danube Delta. However smaller-scale differences are apparent, in particular with respect to the size and clustering of those areas (Figure A.1, Appendix A, Figure 1a).

The overall performance of bird species for themselves was rather low (AUC=0.65, area under the bird performance curve) (Figure 2a, Table A.1, Appendix A), but we observed considerable differences between groups based on breeding habitat (Table A.2, Appendix A). Wetland and shore-breeders were best retained through the ranking process, followed by those breeding in "forest to (dense) woodland" areas (Figure 3). In contrast, birds breeding in "arable land, open woodland to grassland" or being "generalist and close to humans" were lost much more quickly (Figure 3). To explore this further, we plotted each species' performance as a function of its range size (Figure 4, Table A.2, Appendix A), and found a clear negative trend. "Wetland and shore" breeders include more range-restricted species compared to other groups and at the same time performed best in the prioritization, whereas forest, generalist and grassland birds overall have larger ranges, and performed worst in the prioritization. In addition, the distributions of wetland and shore breeders often overlap with those of other groups, those resulting in areas of high species richness that are preferentially prioritized by the ABF algorithm (Figure A.2, Appendix A).

Habitat types were generally retained well throughout the prioritization process (AUC=0.9, area under the habitat surrogate curve) (Figure 2b, Table A.1, Appendix A). We observed that features with smaller ranges were retained the longest (Figure 4, Table A.3, Appendix A).

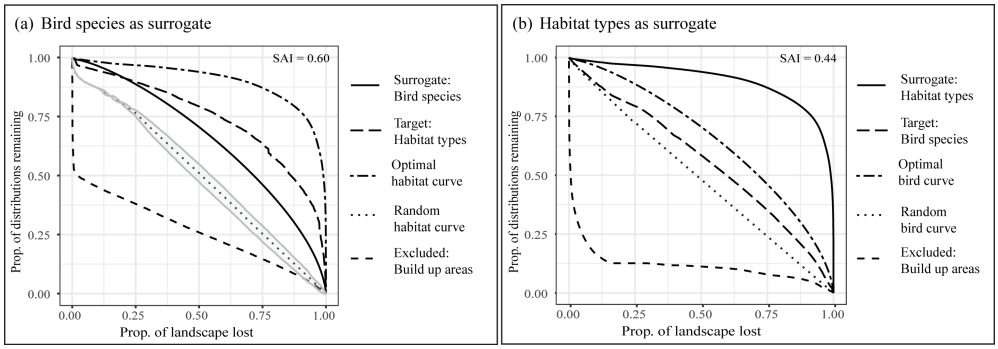

FIGURE 2: Performance and surrogacy curves quantifying the average proportion of original feature distributions represented as landscape is lost. Built-up areas were negatively weighted and hence excluded from the prioritization (lower dashed line). The area between the target curve and the random curve divided by the area between the optimal curve and random curve represents the efficacy of the surrogate (SAI; Species accumulation index). In panel (a) bird species were used as a surrogate for habitat types and in (b) habitats were used as a surrogate for birds.

Surrogacy analyses

Birds were a moderately good surrogate for habitats (SAI = 0.60). Interestingly, birds represented habitats better than themselves (Figure 2a), although as shown above this is only true for the representation of all birds combined, and there are large differences between bird groups (Figure 3). The reciprocal representation of habitats for birds was less effective (SAI = 0.44; Figure 2b).

Evaluation of protected areas and identification of expansion regions

We found that the Zonation values within current PAs, when both habitat types and bird species were considered, differ significantly from a uniform distribution, with an overrepresentation of higher values (Chi2 test, Chi2 = 29289, df = 9, p-value < 2.2e-16) (Figure A.3a, Appendix A).

These results suggest that current PAs generally comprise areas of high conservation value better than would be expected based on a random assignment of areas for conservation. However, current PAs also comprise a considerable amount of land surface area with relatively low conservation values based on bird and habitat diversity, suggesting that improvements could be made.

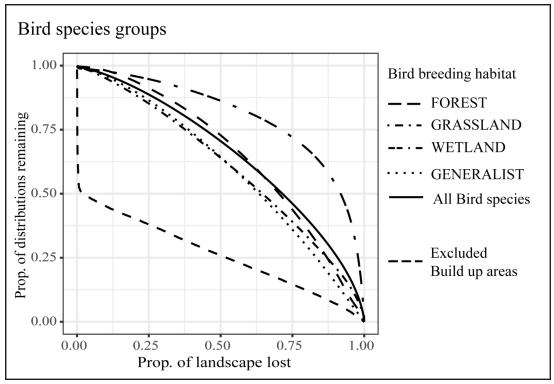

FIGURE 3: Performance curves for bird species split by breeding habitats. The solid line is the average performance curve of all bird species used in the surrogacy approach. Built-up areas were negatively weighted and hence excluded from the prioritization (lower dashed line).

Habitat types are relatively well represented in the current protected areas network (Figure A.3c, Appendix A), with the exception of grassland, heathland and woodland habitats. Among the breeding groups, generalist and grassland breeders are on average represented less well than expected under a random assignment, although in the grassland breeding group much variation between the species can be observed (Figure A.3b, Appendix A). The mask analysis highlighted transition areas from highland to lowland regions, such as along the northern Carpathian Mountains, the eastern foothills of the Carpathian Mountains, and the eastern part of the Apuseni Mountains (Figure 1b) as particularly important expansion sites for bird and habitat conservation.

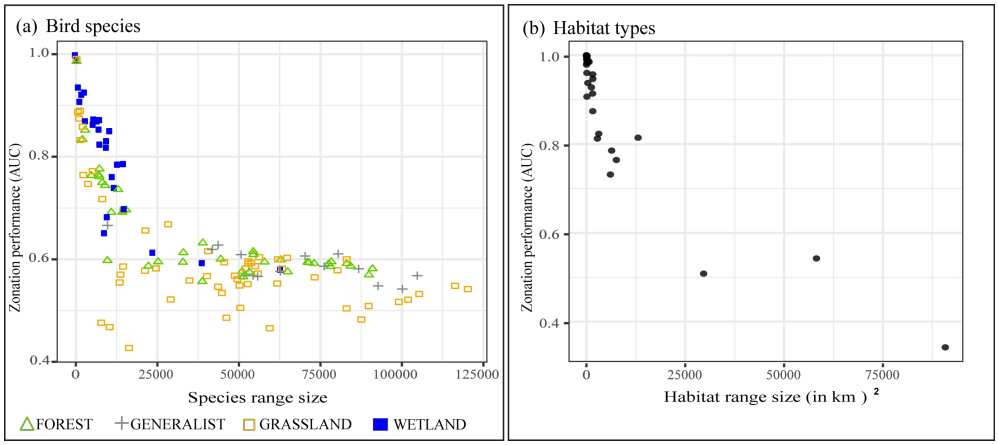

FIGURE 4: The Zonation performance of individual features (AUC) as a function of its corresponding range size. (a) Individuals bird species, belonging to one of the four breeding habitat groups. Green triangle = forest to (dense) woodland; grey cross = generalist and close to humans; yellow square = arable land, open woodland to grassland; blue square = wetland and shores. The values for the range sizes of bird species were computed by adding up Maxent species distribution values. (b) Individual habitat types.

Discussion

The necessity to rely on surrogates for conservation prioritization raises the question of how effective they are. Here we evaluated the mutual surrogacy power of bird species and habitat types in Romania, an area in Europe with high biodiversity, and demonstrated that neither birds nor habitat types are effective surrogates of one another. Birds represented 60% of habitat conservation priorities, while habitats were less effective at representing bird conservation priorities (44%). These results are concordant with studies in other regions suggesting to use more than one type of surrogate for conservation prioritization (Bonn and Gaston, 2005; Di Minin and Moilanen, 2014; Lombard et al., 2003). We also found that existing protected areas in Romania capture areas of high conservation value for both biodiversity features better than expected at random, but could potentially be designed more effectively and more efficiently. Finally, we identified additional areas that should be prioritized in case the existing network were to be expanded under the European Union Biodiversity Strategy to 2030, or where conservation strategies for conserving avian and habitat diversity on private lands could be incentivized.

Bird species as a surrogate

The effectiveness of 137 breeding bird species as a surrogate for habitats was $\sim$ 60% of that of habitats for themselves (Figure 2a). Thus, in the absence of other data,

birds could represent habitat types better than random, but only to a limited extent. These results appear robust because we included many bird species, breeding in a wide variety of habitats (Table A.1, Appendix A), thus covering the existing habitat diversity quite well. Our results corroborate other studies that found that taxonomic groups are poor surrogates for one another (for other taxonomic groups, e.g. Billeter et al., 2008; Di Minin and Moilanen, 2014; Franco et al., 2009; Larsen et al., 2012), and should be used cautiously as surrogates for habitat diversity.

Interestingly, when prioritizing bird species only (Figure 3) we found that wetland and shore birds were much better represented than forest, grassland, and generalist species. This unexpected result corroborates the focal areas of the Bird Directive, which demands particular attention to wetland species (Commission, 2009), Art 4 (2)). A potential explanation for this representation bias is the emphasis of the additive-benefit function (ABF) on high average performance across all features - in the case of bird species, areas with high species richness (Arponen et al., 2005) - combined with differences in range sizes between the bird groups (Franco et al., 2009; Moilanen et al., 2005). We found that species richness was highest in areas where the distributions of wetland-breeding species overlapped with those of species breeding in other types of habitat (Figure A.2, Appendix A). Because wetland birds generally have small ranges due to the limited availability of suitable habitat (Tozer et al., 2010), Zonation prioritized the species-rich wetlands over areas with fewer species, where more widely distributed species occur (Figure 4). These results are in line with similar patterns in small versus large-range moths (Lund and Rahbek, 2002), butterflies, reptiles, and amphibians (Franco et al., 2009). The representation bias in our study may be exacerbated by associations of generalist species to human-dominated landscapes. Because we negatively weighted and hence excluded built-up areas from the prioritization, species occurring in those areas may be underrepresented in the final results.

Habitats as a surrogate

Habitats as a surrogate for birds were only 44% as effective as the maximum possible. This result is consistent with other studies showing that environmental diversity may not a good proxy for the diversity of small vertebrates (including bird species) (Bonn and Gaston, 2005; Popescu et al., 2020). Yet, habitats represented birds better than random (Figure 2b), potentially due to the influence of habitat structure on bird species occurrence and distributions (Mac Nally et al., 2002), and may therefore have merit for prioritization when no other data are available. It remains unclear whether higher spatial and thematic resolutions – in particular more detailed habitat classifications – could improve the mutual representation.

Previous studies suggested that pre-classified environmental data such as the ETE dataset (Agency, 2016) perform better as a surrogate for species diversity than con-

tinuous environmental variables (e.g. Bonn and Gaston, 2005; Grantham et al., 2010; Lombard et al., 2003; Oliver et al., 2004). Thus, easily obtained environmental data could act as a biodiversity surrogate for other levels of biodiversity (Beier et al., 2015; Engelbrecht et al., 2016). Our results suggest however that habitat classes performed relatively poorly at representing bird biodiversity, and ideally should not be used on their own in prioritization efforts. Instead, combining taxonomic and environmental surrogates could increase the surrogacy power for the protection of overall biodiversity (Di Minin and Moilanen, 2014; Lombard et al., 2003), but a single taxonomic group may not suffice. For instance, habitats and birds did not perform well in representing amphibians and reptiles in other areas (Araújo et al., 2001; Grantham et al., 2010; Mac Nally et al., 2002). Thus, we recommend to combine environmental and taxonomic surrogates, preferentially from multiple taxonomic groups.

Representation in existing protected areas and conservation implications

We found that a considerable fraction of PAs is located in areas with high conservation values. It is important to stress, however, that our evaluations by no means suggest that the current network of PAs is sufficient. Around 23% of Romania's land surface area is currently under protection, and improvements to the protected area network may be necessary (Iojă et al., 2010; Niculae et al., 2017). Large ecoregions and several widespread bird and mammal species may be protected sufficiently well, but smaller ecoregions, as well as invertebrate and plant species are for example underrepresented in the existing Natura 2000 network (Iojă et al., 2010). The current network of PAs consists of reserves designed for various purposes. In our evaluation, we specifically focused on those that have been designed to protect birds, habitats, or biodiversity as a whole, i.e. SPAs, national and natural parks, and biosphere reserves. We found that these PAs represent areas of high bird or habitat conservation value better than a random assignment of areas for protection. However, habitats were better represented than birds (Figure A.3b and c, Appendix A). We also found that rare habitats are well represented, which is consistent with results for the Czech Republic (Pechanec et al., 2018). These habitats typically are wetlands and shores, large areas of which are protected in the Danube Delta. Surprisingly, the representation of grassland and woodland habitats was rather poor. A likely reason for this result is the large area of wood- and grassland habitats in Romania, only part of which can be represented in PAs (Figure A.4, Appendix A). In contrast, rare habitats such as littoral areas are represented at high percentages, because they can be entirely contained within a fraction of the total land surface area. Despite the fact that current PAs capture important areas for conservation relatively well, a tail of areas with low conservation value can also be observed (Fig. A.3A). It remains unclear whether these areas may be important for other reasons, such as for other taxonomic groups, or as corridors between areas of high conservation value. Yet, the presence of areas with

low conservation value also suggests that improvements in both the efficiency and efficacy of the network may be possible. To this end, we identified areas that should be prioritized based on bird and habitat diversity in a scenario of future expansions of the current network. A recent study suggests that such improvements may best be developed at the level of Biogeographical Regions rather than at the national level (Miu et al., 2020). Protected areas are a crucial component of conservation, but the identification and designation of PAs is often a lengthy and difficult process. In addition, even when the new targets for the EU Biodiversity Strategy are met, 70% of the land surface area will remain unprotected. Hence, effective conservation also depends on the protection of biodiversity outside of PAs. To do so, the development of incentives for targeted management practices to retain high diversity of species and habitats should be prioritized (Manolache et al., 2020), yet scientific research that can support management decision is largely lacking (Nita et al., 2019).

Our study adds to the body of evidence that taxonomic and environmental surrogates represent one another only to a limited extent. Hence, the use of just one type of surrogate likely does not capture the broad patterns of biodiversity sufficiently well. This situation is less than ideal, as conservation measures respond to the biodiversity crisis, with little time to collect data on the distribution of species or habitats. Although these data are becoming increasingly available, our results highlight the need for investing in survey and monitoring schemes in countries such as Romania, where data still remains relatively scarce. Our study also presents an example of the importance of scientific research in informing conservation strategies as a stakeholder than often remains underrated (Opermanis et al., 2014; Popescu et al., 2014).

Acknowledgements

The Romanian National Breeding Bird Atlas was financed by Milvus Group Association, the Romanian Ornithological Society and the Swiss Ornithological Institute through a grant from MAVA Foundation (KST 10103 MAVA).

Appendix A. Supplementary data

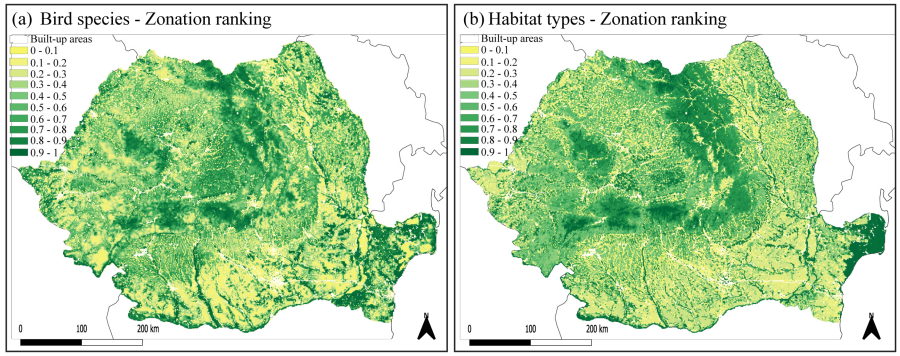

FIGURE A.1: Study region with Zonation ranking based on (a) Bird species and (b) habitat types. Colors indicate importance ranking scores for conservation, with 0 meaning lowest importance and 1 meaning highest importance. Built-up areas are indicated in white and were excluded from prioritization.

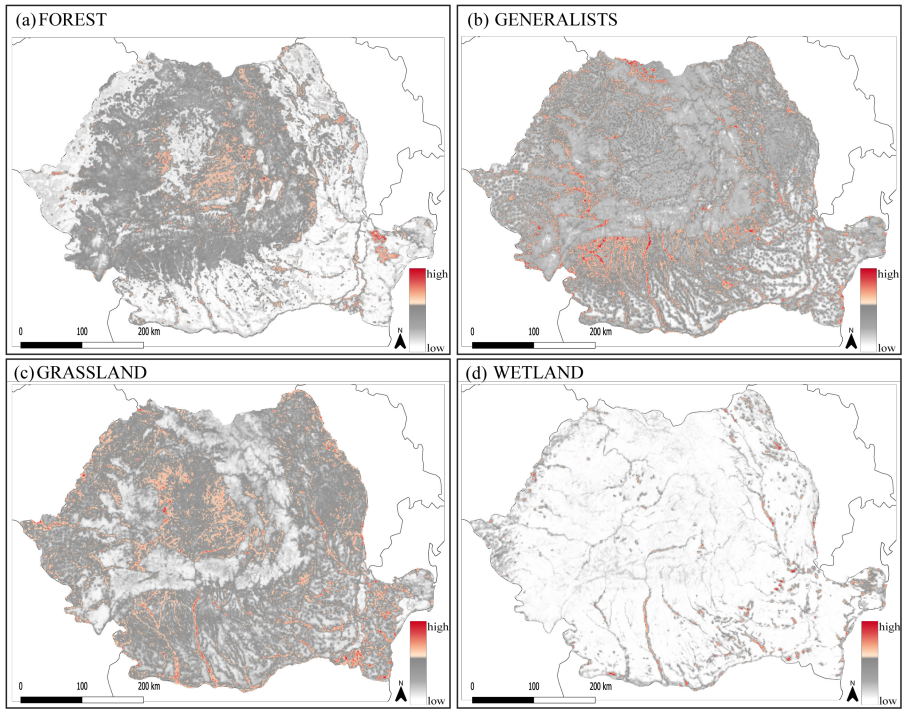

FIGURE A.2: Overlapping bird species occurrences per breeding habitat group: (a) forests to (dense) woodland, (b) generalist and close to humans, (c) arable land, open woodland to grasslands, and (d) wetlands and shores. Red indicates species-rich areas; white to grey indicate no or low overlap of species occurrences.

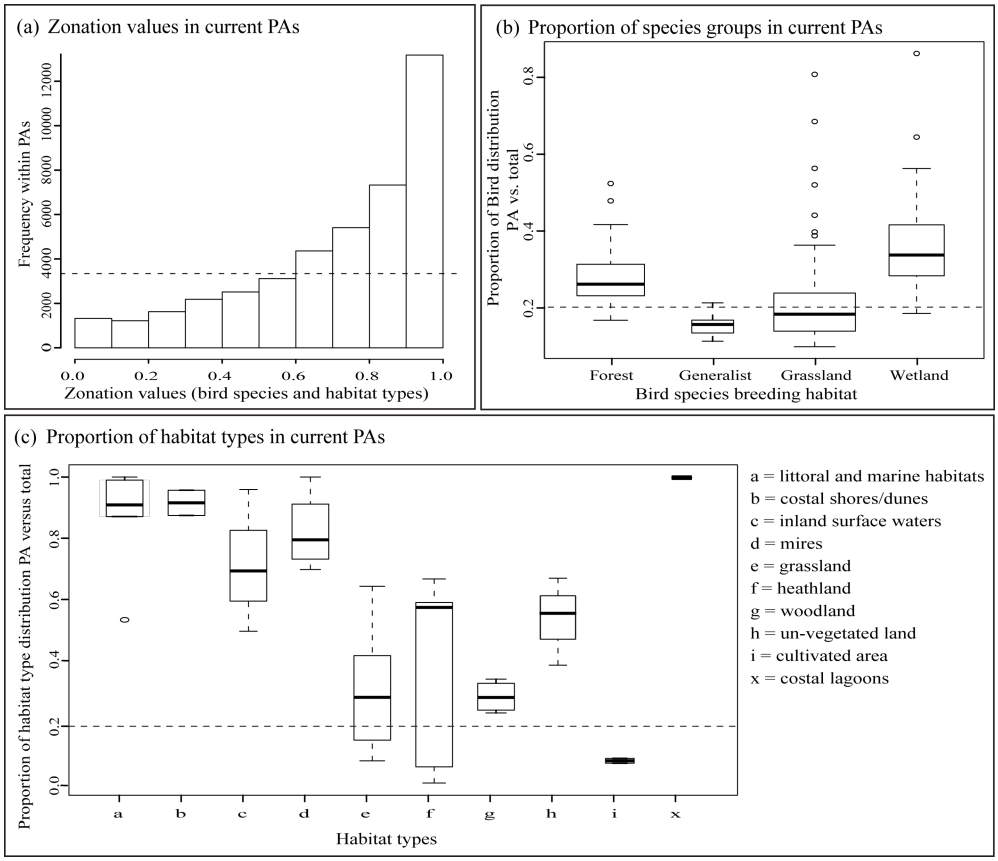

FIGURE A.3: Barplot of conservation values of areas in current reserves. The horizontal dashed line indicates the expected frequency of each conservation value (freq = 3338.6), had the current PAs be selected at random. The high frequencies of high conservation values, combined with the low frequencies of low conservation values suggest that current PAs were selected efficiently. (b-c) Box-and-whisker plots for birds (b) and habitats (c) showing the proportion of the total distribution of each group of feature types that is represented in the existing protected area network. A dotted line indicates the random expectation for the representation of each feature class based on the amount of protected area in Romania ($\sim$ 23% of land surface area).

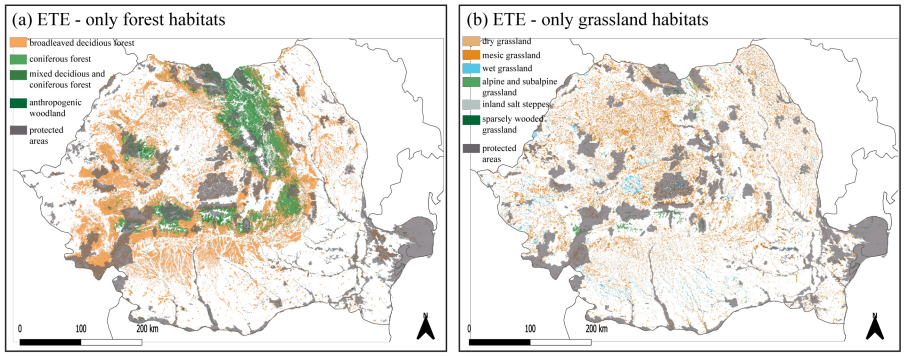

FIGURE A.4: Study region with (a) forest habitats and (b) grassland habitats highlighted. The used protected area network is highlighted in grey.

Appendix A

Table A.1: AUC values for all the performance curves within each of the surrogacy analyses.

Bird species as surrogate		Habitat types as surrogate	
Curve	AUC	Curve	AUC
Surrogate: Bird species	0.6563234	Surrogate: Habitat types	0.901305
Target: Habitat types	0.7418273	Target: Bird species	0.5609637
Optimal habitat curve	0.901305	Optimal bird curve	0.6563234
Random habitat curve	0.4992973	Random bird curve	0.4846996

Table A.2 – Bird species included in prioritization analyses, sorted by breeding habitat. For each species, the common name, breeding habitat, conservation status, range size and AUC of the Zonation performance curve, as well as the resolution of the species distribution maps and the regularization multiplier for each species in order to reduce the model complexity are provided.

species	**common name**	**breeding habitat**	**conservation status**	**range size**	**AUC**	**resolution**	**regularization multiplier**
Accipiter brevipes	Levant sparrowhawk	forest to (dense) woodland	least concern	8061.57016	0.74990883	1km	1
Accipiter gentilis	Northern goshawk	forest to (dense) woodland	least concern	77534.98407	0.59119630	1km	1
Accipiter nisus	Eurasian sparrowhawk	forest to (dense) woodland	least concern	90978.51613	0.58160768	1km	1
Acrocephalus agricola	Paddyfield warbler	wetlands and shores	least concern	121.65638	0.99784597	1km	1
Acrocephalus arundinaceus	Great reed warbler	wetlands and shores	least concern	39035.58315	0.59233947	1km	1
Acrocephalus palustris	Marsh warbler	arable land, open woodland to grassland	least concern	43637.63396	0.54630492	1km	1
Acrocephalus schoenobaenus	Sedge warbler	wetlands and shores	least concern	12035.75636	0.73894781	1km	1
Acrocephalus scirpaceus	Eurasian reed warbler	wetlands and shores	least concern	6736.38939	0.86876331	1km	1
Aegithalos caudatus	Long-tailed tit	forest to (dense) woodland	least concern	57872.27831	0.59468770	1km	1
Alauda arvensis	Eurasian skylark	arable land, open woodland to grassland	least concern	87559.61642	0.48238245	1km	1
Alcedo atthis	Common kingfisher	wetlands and shores	least concern	9597.15719	0.81748182	1km	1
Anas platyrhynchos	Mallard	wetlands and shores	least concern	63027.82016	0.57953751	2km	1
Anas strepera	Gadwall	wetlands and shores	least concern	2874.35555	0.92502952	2km	1
Anser anser	Greylag goose	wetlands and shores	least concern	2058.04150	0.92071810	2km	1

Anthus campestris	Tawny pipit	arable land, open woodland to grassland	least concern	50424.37797	0.50528700	1km	1
Anthus spinoletta	Water pipit	arable land, open woodland to grassland	least concern	968.85526	0.87441015	1km	1
Anthus trivialis	Tree pipit	arable land, open woodland to grassland	least concern	63217.79347	0.58182476	1km	2
Aquila pomarina	Lesser spotted eagle	arable land, open woodland to grassland	least concern	55005.77685	0.58658109	1km	1
Asio otus	Long-eared owl	arable land, open woodland to grassland	least concern	61755.49970	0.55281792	1km	1
Athene noctua	Little owl	generalist and close to humans	least concern	51542.02505	0.56806267	1km	1
Bonasa bonasia	Hazel grouse	forest to (dense) woodland	least concern	10857.52089	0.69164888	1km	1
Bubo bubo	Eurasian eagle-owl	arable land, open woodland to grassland	least concern	14583.92646	0.69569954	1km	1
Burhinus oedicnemus	Eurasian stone-curlew	arable land, open woodland to grassland	least concern	8085.43081	0.71756006	1km	1
Buteo buteo	Common buzzard	arable land, open woodland to grassland	least concern	116307.14361	0.54823333	1km	1
Buteo rufinus	Long-legged buzzard	arable land, open woodland to grassland	least concern	34836.83075	0.55854096	1km	1
Calandrella brachydactyla	Greater short-toed lark	arable land, open woodland to grassland	least concern	7761.67829	0.47625058	1km	1
Caprimulgus europaeus	European nightjar	arable land, open woodland to grassland	least concern	40532.75741	0.61605196	1km	1
Carduelis cannabina	Common linnet	arable land, open woodland to grassland	least concern	52845.10863	0.59140200	1km	1
Carduelis carduelis	European goldfinch	generalist and close to humans	least concern	86726.94063	0.58135138	1km	1
Carduelis chloris	European greenfinch	generalist and close to humans	least concern	70295.94246	0.60629425	1km	1
Carduelis spinus	Eurasian siskin	forest to (dense) woodland	least concern	2839.96055	0.85193304	1km	1
Certhia brachydactyla	Short-toed treecreeper	forest to (dense) woodland	least concern	2013.28922	0.83382209	1km	1
Certhia familiaris	Eurasian treecreeper	forest to (dense) woodland	least concern	44435.41463	0.60044900	1km	1
Charadrius alexandrinus	Kentish plover	wetlands and shores	least concern	402.80379	0.98700321	2km	1
Charadrius dubius	Little ringed plover	wetlands and shores	least concern	11437.19756	0.76026444	1km	1
Cinclus cinclus	White-throated dipper	wetlands and shores	least concern	9042.78134	0.65070329	1km	2
Coccothraustes coccothraustes	Hawfinch	forest to (dense) woodland	least concern	50998.87105	0.57640894	1km	1

Columba livia domestica	Domestic pigeon	generalist and close to humans	least concern	54317.06182	0.56837325	1km	1
Columba oenas	Stock dove	arable land, open woodland to grassland	least concern	21185.92176	0.57768825	1km	1
Coracias garrulus	European roller	arable land, open woodland to grassland	least concern	24555.93468	0.58215295	1km	1
Coturnix coturnix	Common quail	arable land, open woodland to grassland	least concern	83034.01921	0.50421482	1km	1
Crex crex	Corn crake	arable land, open woodland to grassland	least concern	50193.95260	0.54882257	1km	1
Cuculus canorus	Common cuckoo	arable land, open woodland to grassland	least concern	120210.34997	0.54176096	1km	1
Cygnus olor	Mute swan	wetlands and shores	least concern	7537.50276	0.87129993	2km	1
Delichon urbicum	Common house martin	generalist and close to humans	least concern	43626.40116	0.62783735	1km	1
Dendrocopos leucotos	White-backed woodpecker	forest to (dense) woodland	least concern	22250.06326	0.58679011	1km	1
Dendrocopos major	Great spotted woodpecker	forest to (dense) woodland	least concern	78391.42973	0.59522897	1km	1
Dendrocopos medius	Middle spotted woodpecker	forest to (dense) woodland	least concern	32955.40140	0.61298939	1km	1
Dendrocopos minor	Lesser spotted woodpecker	forest to (dense) woodland	least concern	38934.26162	0.63192121	1km	1
Dendrocopos syriacus	Syrian woodpecker	generalist and close to humans	least concern	50665.64839	0.60917411	1km	1
Dryocopus martius	Black woodpecker	forest to (dense) woodland	least concern	54417.02676	0.61511955	1km	1
Emberiza cia	Rock bunting	arable land, open woodland to grassland	least concern	2271.28839	0.76424958	1km	1
Emberiza cirlus	Cirl bunting	arable land, open woodland to grassland	least concern	1344.30056	0.83244702	1km	1
Emberiza citrinella	Yellowhammer	arable land, open woodland to grassland	least concern	73230.19592	0.56465259	1km	1
Emberiza hortulana	Ortolan bunting	arable land, open woodland to grassland	least concern	46165.78880	0.48603792	1km	1
Emberiza melanocephala	Black-headed bunting	arable land, open woodland to grassland	least concern	16282.96982	0.42695403	1km	1
Emberiza schoeniclus	Common reed bunting	wetlands and shores	least concern	5717.80150	0.87258025	1km	1
Erithacus rubecula	European robin	forest to (dense) woodland	least concern	73091.24488	0.59193831	1km	1
Falco subbuteo	Eurasian hobby	arable land, open woodland to grassland	least concern	80250.96856	0.57867657	1km	1
Falco tinnunculus	Common kestrel	arable land, open woodland to grassland	least concern	101869.92984	0.52123089	1km	1

Ficedula parva	Red-breasted flycatcher	forest to (dense) woodland	least concern	9690.38547	0.59766073	1km	1
Ficedula semitorquata	Semicollared flycatcher	forest to (dense) woodland	least concern	187.86025	0.98597321	1km	1
Fringilla coelebs	Common chaffinch	forest to (dense) woodland	least concern	84296.54300	0.58662739	1km	1
Fulica atra	Eurasian coot	wetlands and shores	least concern	10633.35631	0.84990775	1km	1
Galerida cristata	Crested lark	arable land, open woodland to grassland	least concern	40207.25027	0.56719285	1km	1
Gallinula chloropus	Common moorhen	wetlands and shores	least concern	14865.47427	0.78560910	1km	1
Garrulus glandarius	Eurasian jay	forest to (dense) woodland	least concern	78083.09732	0.58596724	1km	1
Hieraaetus pennatus	Booted eagle	arable land, open woodland to grassland	least concern	21325.78723	0.65608778	1km	1
Himantopus himantopus	Black-winged stilt	wetlands and shores	least concern	7601.28856	0.82358760	2km	1
Hippolais icterina	Icterine warbler	forest to (dense) woodland	least concern	4734.55775	0.76345323	1km	1
Hippolais pallida	Eastern olivaceous warbler	arable land, open woodland to grassland	least concern	636.41807	0.88715480	1km	1
Hirundo rustica	Barn swallow	generalist and close to humans	least concern	92684.57095	0.54812916	1km	1
Ixobrychus minutus	Little bittern	wetlands and shores	least concern	9672.45646	0.83022420	1km	1
Lanius collurio	Red-backed shrike	arable land, open woodland to grassland	least concern	105283.62519	0.53238743	1km	2
Lanius minor	Lesser grey shrike	arable land, open woodland to grassland	least concern	49545.80656	0.56142450	1km	1
Locustella fluviatilis	River warbler	wetlands and shores	least concern	9883.75889	0.68186192	1km	1
Locustella luscinioides	Savi's warbler	wetlands and shores	least concern	13024.20543	0.78439736	1km	1
Loxia curvirostra	Red crossbill	forest to (dense) woodland	least concern	12968.37490	0.73620076	1km	2
Lullula arborea	Woodlark	arable land, open woodland to grassland	least concern	53795.41696	0.59298697	1km	1
Luscinia luscinia	Thrush nightingale	arable land, open woodland to grassland	least concern	13674.82363	0.57008231	1km	1
Luscinia megarhynchos	Common nightingale	arable land, open woodland to grassland	least concern	51973.65397	0.58210337	1km	1
Melanocorypha calandra	Calandra lark	arable land, open woodland to grassland	least concern	10370.97087	0.46776981	1km	1
Merops apiaster	European bee-eater	arable land, open woodland to grassland	least concern	56258.58013	0.60406792	1km	1
Motacilla alba	White wagtail	generalist and close to humans	least concern	80438.03822	0.61047234	1km	1
Motacilla cinerea	Grey wagtail	wetlands and shores	least concern	23852.95797	0.61261186	1km	1

Motacilla flava	Western yellow wagtail	arable land, open woodland to grassland	least concern	59462.76328	0.46563968	1km	1
Nucifraga caryocatactes	Spotted nutcracker	forest to (dense) woodland	least concern	7249.30683	0.77635845	1km	1
Oenanthe isabellina	Isabelline wheatear	arable land, open woodland to grassland	least concern	1217.57581	0.88971704	1km	1
Oenanthe oenanthe	Northern wheatear	arable land, open woodland to grassland	least concern	45487.63231	0.59425191	1km	1
Oriolus oriolus	Eurasian golden oriole	forest to (dense) woodland	least concern	89862.92670	0.56965290	1km	1
Otus scops	Eurasian scops owl	arable land, open woodland to grassland	least concern	55827.45031	0.57201477	1km	1
Parus caeruleus	Eurasian blue tit	forest to (dense) woodland	least concern	62848.64603	0.59909183	1km	1
Parus cristatus	European crested tit	forest to (dense) woodland	least concern	8947.17371	0.74419417	1km	1
Parus lugubris	Sombre tit	arable land, open woodland to grassland	least concern	3738.45119	0.74668054	1km	1
Parus major	Great tit	generalist and close to humans	least concern	104681.26292	0.56798917	1km	1
Parus montanus	Willow tit	forest to (dense) woodland	least concern	14259.97792	0.69191956	1km	1
Parus palustris	Marsh tit	forest to (dense) woodland	least concern	51336.50316	0.56590091	1km	1
Passer domesticus	House sparrow	generalist and close to humans	least concern	62745.54466	0.57630376	1km	1
Passer hispaniolensis	Spanish sparrow	arable land, open woodland to grassland	least concern	14484.03502	0.58587226	1km	1
Phoenicurus ochruros	Black redstart	generalist and close to humans	least concern	41834.92825	0.61907826	1km	1
Phylloscopus collybita	Common chiffchaff	forest to (dense) woodland	least concern	71431.24855	0.59370802	1km	1
Phylloscopus sibilatrix	Wood warbler	forest to (dense) woodland	least concern	25237.17963	0.59518699	1km	1
Pica pica	Eurasian magpie	arable land, open woodland to grassland	least concern	99061.75709	0.51677079	1km	1
Picoides tridactylus	Eurasian three-toed woodpecker	forest to (dense) woodland	least concern	6997.02729	0.76442532	1km	1
Picus canus	Grey-headed woodpecker	forest to (dense) woodland	least concern	54437.01940	0.60940213	1km	1
Picus viridis	European green woodpecker	arable land, open woodland to grassland	least concern	52897.80818	0.59607385	1km	1
Podiceps cristatus	Great crested grebe	wetlands and shores	least concern	7371.90259	0.85279854	2km	1
Prunella collaris	Alpine accentor	arable land, open woodland to grassland	least concern	235.86942	0.99085294	1km	1
Prunella modularis	Dunnock	arable land, open woodland to grassland	least concern	5153.02586	0.77200520	1km	1

Pyrrhula pyrrhula	Eurasian bullfinch	forest to (dense) woodland	least concern	7230.06574	0.76103579	1km	1
Rallus aquaticus	Water rail	wetlands and shores	least concern	5556.81137	0.86233389	1km	1
Recurvirostra avosetta	Pied avocet	wetlands and shores	least concern	3202.83744	0.86916668	2km	1
Regulus regulus	Goldcrest	forest to (dense) woodland	least concern	15427.99648	0.69644053	1km	1
Saxicola rubetra	Whinchat	arable land, open woodland to grassland	least concern	29109.11443	0.52158472	1km	1
Saxicola torquatus	African stonechat	arable land, open woodland to grassland	least concern	52718.00252	0.55167082	1km	1
Serinus serinus	European serin	generalist and close to humans	least concern	9720.81702	0.66615576	1km	1
Sitta europaea	Eurasian nuthatch	forest to (dense) woodland	least concern	64966.35194	0.57578479	1km	1
Streptopelia decaocto	Eurasian collared dove	generalist and close to humans	least concern	55746.31333	0.56628207	1km	1
Streptopelia turtur	European turtle dove	arable land, open woodland to grassland	least concern	62021.97272	0.59983583	1km	1
Strix aluco	Tawny owl	forest to (dense) woodland	least concern	53225.06837	0.57489893	1km	1
Strix uralensis	Ural owl	forest to (dense) woodland	least concern	32850.95119	0.59425224	1km	1
Sturnus vulgaris	Common starling	generalist and close to humans	least concern	100143.48935	0.54204224	1km	1
Sylvia atricapilla	Eurasian blackcap	forest to (dense) woodland	least concern	83074.98241	0.59172058	1km	2
Sylvia communis	Common whitethroat	arable land, open woodland to grassland	least concern	89801.92144	0.50865592	1km	2
Sylvia curruca	Lesser whitethroat	arable land, open woodland to grassland	least concern	83056.52531	0.59982587	1km	1
Sylvia nisoria	Barred warbler	arable land, open woodland to grassland	least concern	28279.00220	0.66816209	1km	2
Tachybaptus ruficollis	Little grebe	wetlands and shores	least concern	15069.00180	0.69765523	2km	1
Tadorna ferruginea	Ruddy shelduck	wetlands and shores	least concern	1481.95172	0.90693381	2km	1
Tadorna tadorna	Common shelduck	wetlands and shores	least concern	1028.39293	0.93484062	2km	1
Troglodytes troglodytes	Eurasian wren	forest to (dense) woodland	least concern	38787.66051	0.55674206	1km	1
Turdus merula	Common blackbird	generalist and close to humans	least concern	76184.82882	0.58633076	1km	1
Turdus philomelos	Song thrush	forest to (dense) woodland	least concern	71071.01128	0.59438408	1km	1
Turdus torquatus	Ring ouzel	arable land, open woodland to grassland	least concern	2106.73183	0.85863618	1km	1
Turdus viscivorus	Mistle thrush	arable land, open woodland to grassland	least concern	48748.68883	0.56797913	1km	2

Tyto alba	Western barn owl	arable land, open woodland to grassland	least concern	13425.57259	0.55469077	1km	1
Upupa epops	Eurasian hoopoe	arable land, open woodland to grassland	least concern	64782.17531	0.60267177	1km	1
Vanellus vanellus	Northern lapwing	arable land, open woodland to grassland	least concern	44779.79424	0.53426956	2km	1

Table A.3 – Habitat types included in prioritization analyses (sorted by ETE abbreviation). For each habitat type, the abbreviation, the habitat name, range size and AUC of the Zonation performance curve are provided. Habitat types in bold where excluded from surrogacy analyses, due to their built-up character (with weight=0 in surrogacy analyses).

habitat type	**habitat name**	**own habitat code**	**range size**	**AUC**
A100	Littoral undetermined substrate with no sea ice presence	41	68.36000	0.99772770
A105	Littoral sand with no sea ice presence	42	0.30000	0.99983000
A200	Infralittoral undetermined substrate with no sea ice presence	32	5.90000	0.99947803
A205	Infralittoral sand with no sea ice presence	33	138.39000	0.99867043
A206	Infralittoral mud with no sea ice presence	35	1.78000	0.99981548
A306	Circalittoral mud with no sea ice presence	36	7.61000	0.99969668
B1	Coastal dunes and sandy shores	34	5.10000	0.99955978
B2	Coastal shingle	37	13.71000	0.99824092
C1	Surface standing waters	6	1560.24000	0.91355051
C2	Surface running waters	2	1597.55000	0.87395398
C3	Littoral zone of inland surface waterbodies	19	1621.58000	0.94670255
D1	Raised and blanket bogs	40	0.03000	0.99983000
D4	Base-rich fens and calcareous spring mires	28	1.72000	0.99968006
D5	Sedge and reedbeds	8	1249.69000	0.92751736
D6	Inland saline and brackish marshes and reedbeds	38	0.68000	0.99979106
E1	Dry grasslands	11	6374.13999	0.78522941
E2	Mesic grasslands	4	29647.62997	0.50813865
E3	Seasonally wet and wet grasslands	27	3118.05000	0.82308049
E4	Alpine and subalpine grasslands	22	1586.63000	0.95684484
E6	Inland salt steppes	30	115.01000	0.98268017
E7	Sparsely wooded grasslands	16	25.43000	0.97878650
F2	Arctic	23	452.03000	0.98696866

F3	Temperate and mediterranean-montane scrub	24	97.41000	0.99374123
F4	Temperate shrub heathland	25	68.68000	0.99090658
F5	Maquis	31	0.12000	0.99983000
FB	Shrub plantations	7	6060.00999	0.73127128
G1	Broadleaved deciduous woodland	3	58187.91002	0.54242909
G3	Coniferous woodland	14	13062.37999	0.81423318
G4	Mixed deciduous and coniferous woodland	10	7558.81999	0.76408771
G5	Lines of trees	5	2767.76999	0.81223679
H2	Screes	13	347.36000	0.93751919
H3	Inland cliffs	12	87.41000	0.99084328
H5	Miscellaneous inland habitats with very sparse or no vegetation	17	110.99000	0.96018654
I1	Arable land and market gardens	1	90784.89003	0.34219191
I2	Cultivated areas of gardens and parks	26	130.16000	0.90673659
J1	**Buildings of cities**	**9**	**-**	**-**
J2	**Low density buildings**	**15**	**-**	**-**
J3	**Extractive industrial sites**	**21**	**-**	**-**
J4	**Transport networks and other constructed hard-surfaced areas**	**18**	**-**	**-**
J5	**Highly artificial man-made waters and associated structures**	**20**	**-**	**-**
J6	**Waste deposits**	**29**	**-**	**-**
X2_3	Coastal lagoons	39	666.38000	0.98499521

Appendix B. Supplementary data

Bird species distribution modeling

For the species distribution models, we used presence data from the years 2006-2018 (occasional observations included) originating from three data bases: Ornitodata (Ornitodata, 2018), OpenBirdMaps (OpenBirdMaps, 2018) and Rombird (Rombird, 2018), and an initial set of 172 environmental variables. To reduce redundancy and autocorrelation between the environmental variables, we identified pairs of variables with a Pearson correlation coefficient > 0.8, and excluded the one with the highest correlations with other variables. Furthermore, habitat data was obtained from multiple sources, but in the final covariates we selected only the set that performed the best in the preliminary models. Hence our final data set consisted of 73 variables (Table B.1).

Climate variables were obtained from the CliMond repositories (Kriticos et al., 2012), percentual habitat cover data from Corine Land Cover 2012, the Ecosystem types of Europe from EUNIS (European Nature Information System). Information on pedological data were extracted from the European Soil Data Centre (Panagos et al., 2012) and remote sensing data on vegetation from the Copernicus Global Land Service. Furthermore, variables describing open areas, shrubs, woodlands and wetlands in surrounding areas, and distance to water at the landscape level were derived from Corine Land Cover 2012.

To reduce sampling bias caused by unequal coverage and the use of occasional observations, we resampled occurrence data. From every 1x1 km grid cell only one observation was used, and remaining data were further rarefied by selecting a maximum of five presences in every grid cell of the ETRS89 LAEA 10x10 km grid. To further reduce the effects of sampling bias, we created a bias file using simple spatial interpolation of all observations, and used this in the model building process in MaxEnt. We ran and averaged 100 models for each species with default parameters. For some species model complexity was reduced by increasing the regularization multiplier from 1 to 2 (Table A.2, Appendix A) (Radosavljevic and Anderson, 2014).

Finally, we truncated the resulting maps using one of the threshold values offered by MaxEnt. Because the thresholds are case specific and hard to select based only on calculation only (Liu et al., 2013), threshold selection was done for each species using expert opinion by the Romanian Atlas Committee, taking into account the omission error and the distribution area.

Table B.1 – Environmental variables used for species distribution modeling. Variables marked in bold were included in our models, after highly cross-correlated variables (with Pearson correlation coefficient > 0.8) were omitted.

Code	Description	Source
alt	altitudinea	CliMond repositories
Bio01	Annual mean temperature (°C)	CliMond repositories
Bio02	**Mean diurnal temperature range (mean(period max-min)) (°C)**	**CliMond repositories**
Bio03	Isothermality (Bio02 ÷ Bio07)	CliMond repositories
Bio04	Temperature seasonality (C of V)	CliMond repositories
Bio05	Max temperature of warmest week (°C)	CliMond repositories
Bio06	Min temperature of coldest week (°C)	CliMond repositories
Bio07	**Temperature annual range (Bio05-Bio06) (°C)**	**CliMond repositories**
Bio08	Mean temperature of wettest quarter (°C)	CliMond repositories
Bio09	**Mean temperature of driest quarter (°C)**	**CliMond repositories**
Bio10	Mean temperature of warmest quarter (°C)	CliMond repositories
Bio11	Mean temperature of coldest quarter (°C)	CliMond repositories
Bio12	Annual precipitation (mm)	CliMond repositories
Bio13	Precipitation of wettest week (mm)	CliMond repositories
Bio14	Precipitation of driest week (mm)	CliMond repositories
Bio15	**Precipitation seasonality (C of V)**	**CliMond repositories**
Bio16	Precipitation of wettest quarter (mm)	CliMond repositories
Bio17	Precipitation of driest quarter (mm)	CliMond repositories
Bio18	Precipitation of warmest quarter (mm)	CliMond repositories
Bio19	**Precipitation of coldest quarter (mm)**	**CliMond repositories**
Bio20	**Annual mean radiation (W m-2)**	**CliMond repositories**
Bio21	Highest weekly radiation (W m-2)	CliMond repositories
Bio22	Lowest weekly radiation (W m-2	CliMond repositories
Bio23	**Radiation seasonality (C of V)**	**CliMond repositories**
Bio24	**Radiation of wettest quarter (W m-2)**	**CliMond repositories**
Bio25	Radiation of driest quarter (W m-2)	CliMond repositories
Bio26	Radiation of warmest quarter (W m-2)	CliMond repositories
Bio27	**Radiation of coldest quarter (W m-2)**	**CliMond repositories**
Bio28	Annual mean moisture index	CliMond repositories
Bio29	Highest weekly moisture index	CliMond repositories
Bio30	Lowest weekly moisture index	CliMond repositories
Bio31	Moisture index seasonality (C of V)	CliMond repositories
Bio32	Mean moisture index of wettest quarter	CliMond repositories

Bio33	Mean moisture index of driest quarter	CliMond repositories
Bio34	Mean moisture index of warmest quarter	CliMond repositories
Bio35	Mean moisture index of coldest quarter	CliMond repositories
Bio36	**First principal component of the first 35 Bioclim variables**	**CliMond repositories**
Bio37	**Second principal component of the first 35 Bioclim variables**	**CliMond repositories**
Bio38	Third principal component of the first 35 Bioclim variables	CliMond repositories
Bio39	**Fourth principal component of the first 35 Bioclim variables**	**CliMond repositories**
Bio40	Fifth principal component of the first 35 Bioclim variables	CliMond repositories
aglim1	Code of the most important limitation to agricultural use of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
aglim2	Code of a secondary limitation to agricultural use of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
fao90fu	Full soil code of the STU from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
parmado	Code for dominant parent material of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
parmase	Code for secondary parent material of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
slopedo	Dominant slope class of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
slopese	Secondary slope class of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
usedo	Code for dominant land use of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
usese	Code for secondary land use of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
vs	Volume of stones.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
wm1	Code for normal presence and purpose of an existing water management system in agricultural land on more than 50% of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
wr	Dominant annual average soil water regime class of the soil profile of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
il	Code for the presence of an impermeable layer within the soil profile of the STU.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
pmh	Parent material hydrogeological type.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
dr	Depth to rock.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre

ILSWE	**Index of Land Susceptibility to Wind Erosion 1981-2010**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
desiccation	Desiccation	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
andosol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
arenosol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
cambisol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
chernozem	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
fluvisol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
gleysol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
greyzem	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
kastanozem	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
leptosol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
luvisol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
phaeozem	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
podzol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
regosol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
vertisol	Soil tipe extracted from the 1990 FAO-UNESCO Soil Legend.	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre
dist2water1km	**Distance to water**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div3_1_10	Forest in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_1_100	**Forest in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**

div3_1_20	Forest in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_1_50	Forest in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_2_10	**shrub and/or herbaceous vegetation in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div3_2_100	shrub and/or herbaceous vegetation in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_2_20	shrub and/or herbaceous vegetation in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_2_50	**shrub and/or herbaceous vegetation in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div3_3_10	**open spaces with little or no vegetation**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div3_3_100	open spaces with little or no vegetation	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_3_20	open spaces with little or no vegetation	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div3_3_50	**open spaces with little or no vegetation**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div4_1_10	inland wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_1_100	inland wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_1_20	inland wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_1_50	**inland wetlands in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div4_2_10	coastal wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_2_100	coastal wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_2_20	coastal wetlands in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
div4_2_50	**coastal wetlands in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div5_1_10	inland waters in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)

div5_1_100	**inland waters in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div5_1_20	**inland waters in surrounding area of x km**	**Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)**
div5_1_50	inland waters in surrounding area of x km	Covariat derivat from Corine Land Cover - compiled and calculated by SOVON (Netherlands)
eco_f0	**Littoral rock and other hard substrata**	**EUNIS: Ecosystem types of Europe**
eco_f10	**Coastal lagoons**	**EUNIS: Ecosystem types of Europe**
eco_f11	**Coastal dunes and sandy shores**	**EUNIS: Ecosystem types of Europe**
eco_f14	**Surface standing waters**	**EUNIS: Ecosystem types of Europe**
eco_f15	**Surface running waters**	**EUNIS: Ecosystem types of Europe**
eco_f16	**Littoral zone of inland surface waterbodies**	**EUNIS: Ecosystem types of Europe**
eco_f20	**Base-rich fens and calcareous spring mires**	**EUNIS: Ecosystem types of Europe**
eco_f21	**Sedge and reedbeds, normally without free-standing water**	**EUNIS: Ecosystem types of Europe**
eco_f23	**Dry grasslands**	**EUNIS: Ecosystem types of Europe**
eco_f24	**Mesic grasslands**	**EUNIS: Ecosystem types of Europe**
eco_f25	**Seasonally wet and wet grasslands**	**EUNIS: Ecosystem types of Europe**
eco_f26	**Alpine and subalpine grasslands**	**EUNIS: Ecosystem types of Europe**
eco_f28	**Inland salt steppes**	**EUNIS: Ecosystem types of Europe**
eco_f31	**Arctic, alpine and subalpine scrub**	**EUNIS: Ecosystem types of Europe**
eco_f32	**Temperate and mediterranean-montane scrub**	**EUNIS: Ecosystem types of Europe**
eco_f38	Riverine and fen scrubs	EUNIS: Ecosystem types of Europe
eco_f40	**Shrub plantations**	**EUNIS: Ecosystem types of Europe**
eco_f41	**Broadleaved deciduous woodland**	**EUNIS: Ecosystem types of Europe**
eco_f43	**Coniferous woodland**	**EUNIS: Ecosystem types of Europe**
eco_f44	**Mixed deciduous and coniferous woodland**	**EUNIS: Ecosystem types of Europe**
eco_f45	**Lines of trees, small anthropogenic woodlands, recently felled woodland, early-stage woodland and coppice**	**EUNIS: Ecosystem types of Europe**
eco_f47	**Screes**	**EUNIS: Ecosystem types of Europe**
eco_f48	**Inland cliffs, rock pavements and outcrops**	**EUNIS: Ecosystem types of Europe**
eco_f50	**Miscellaneous inland habitats with very sparse or no vegetation**	**EUNIS: Ecosystem types of Europe**
eco_f52	**Arable land and market gardens**	**EUNIS: Ecosystem types of Europe**
eco_f53	**Cultivated areas of gardens and parks**	**EUNIS: Ecosystem types of Europe**
eco_f54	**Buildings of cities, towns and villages**	**EUNIS: Ecosystem types of Europe**
eco_f55	**Low density buildings**	**EUNIS: Ecosystem types of Europe**
eco_f56	**Extractive industrial sites**	**EUNIS: Ecosystem types of Europe**
eco_f57	**Transport networks and other constructed hard-surfaced areas**	**EUNIS: Ecosystem types of Europe**

eco_f59	**Waste deposits**	**EUNIS: Ecosystem types of Europe**
Ecoreg	**Unitati de relief.**	**http://www.geo-spatial.org/download/harta-unitati-relief-romania**
fma_f2_1km	Forest management: close-to-nature	
fma_f3_1km	Forest management: combined objective forestry	
fma_f4_1km	Forest management: even-aged forestry	
fma_f5_1km	Forest management: short rotation forestry	
soil_clay	**Soil: clay content**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
soil_oc	**Soil: organic carbon content**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
soil_ph	**Soil: PH**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
soil_salt	**Soil: availability of salt**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
soil_silt	**Soil: silt content**	**European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre**
ndeposition_1km	**Total nitrogen deposition**	
sdeposition_1km	Total deposition of oxidized sulpher	
pet_he_yr	**Potential Evapotransporation**	
tsum	Total sum of daily average temperatures	
inhabited_land_clc	Localitati, suprafete antropizate	CorineLandCover2006
urban_habitat_clc	Habitate urbane	CorineLandCover2006
rural_habitat_clc	Habutate rurale din localitati	CorineLandCover2006
rural_landskape_clc	Peisaj rural	CorineLandCover2006
arable_clc	Teren arat, agricultura intensa neirigata	CorineLandCover2006
open_habitat_clc	Habitate deschise, pasuni	CorineLandCover2006
shrubs_clc	Tufarisuri	CorineLandCover2006
rock_clc	Suprafete de piatra si pietris	CorineLandCover2006
wetland_clc	Habitate umede	CorineLandCover2006
water_body_clc	Luciuri de apa	CorineLandCover2006
conifer_forest_wwf	Paduri rasinoase	Proiect LIFE05 NAT/RO/000176 "Habitate prioritare alpine, subalpine si forestiere din Romania"
fagus_coniferus_mix_wwf	Amestec de fag si rasinoase	Proiect LIFE05 NAT/RO/000176 "Habitate prioritare alpine, subalpine si forestiere din Romania"
fagus_forest_wwf	Paduri de fag	Proiect LIFE05 NAT/RO/000176 "Habitate prioritare alpine, subalpine si forestiere din Romania"

riparian_forest_wwf	Paduri ripariene	Proiect LIFE05 NAT/RO/000176 "Habitate prioritare alpine, subalpine si forestiere din Romania"
quercus_forest_wwf	Cvercinee	Proiect LIFE05 NAT/RO/000176 "Habitate prioritare alpine, subalpine si forestiere din Romania"
RO_population_2011	**populatia romania**	**ROStat**
ALBH_2014apr	**Bihemispheric Albedo**	**Copernicus Global Land Service**
ALDH_2014apr	**Directional Albedo**	**Copernicus Global Land Service**
DMP_2016apr	**Dry Matter Productivity**	**Copernicus Global Land Service**
FAPAR_2016apr	Fraction of Absorbed Photosynthetically Active Radiation	Copernicus Global Land Service
FCOVER_2016apr	**Fraction of green Vegetation Cover**	**Copernicus Global Land Service**
LAI_2016apr	**Leaf Area index**	**Copernicus Global Land Service**
NDVI_2014apr	**Normalized Difference Vegetation Index**	**Copernicus Global Land Service**
tcd_33	Tree Cover Density 2012 0-33% acoperire	Copernicus Global Land Service
tcd_33_66	Tree Cover Density 2012 33-66% acoperire	Copernicus Global Land Service
tcd_66_100	**Tree Cover Density 2012 66-100% acoperire**	**Copernicus Global Land Service**
tcd_total	**Tree Cover Density 2012**	**Copernicus Global Land Service**
TOCR_2014apr	**Top Of Canopy Reflectances**	**Copernicus Global Land Service**
VCI_2016apr	**Vegetation Condition Index**	**Copernicus Global Land Service**

4 CHAPTER III

ENVIRONMENTAL SELECTION IS A MAIN DRIVER OF DIVERGENCE IN HOUSE SPARROWS (PASSER DOMESTICUS) IN ROMANIA AND BULGARIA

Julia C. Geue, Csongor I Vágási, Mona Schweizer, Péter L. Pap and Henri A. Thomassen

Abstract

Both neutral and adaptive evolutionary processes can cause population divergence, but their relative contributions remain unclear. We investigated the roles of these processes in population divergence in house sparrows (*Passer domesticus*) from Romania and Bulgaria, regions characterized by high landscape heterogeneity compared to Western Europe. We asked whether morphological divergence, complemented with genetic data in this human commensal species, was best explained by environmental variation, geographic distance, or landscape resistance - the effort it takes for an individual to disperse from one location to the other - caused by either natural or anthropogenic barriers. Using generalized dissimilarity modeling, a matrix regression technique that fits biotic beta diversity to both environmental predictors and geographic distance, we found that a small set of climate and vegetation variables explained up to $\sim$ 30% of the observed divergence, whereas geographic and resistance distances played much lesser roles. Our results are consistent with signals of selection on morphological traits and of isolation by adaptation in genetic markers, suggesting that selection by natural environmental conditions shapes population divergence in house sparrows. Our study thus contributes to a growing body of evidence that adaptive evolution may be a major driver of diversification.

KEYWORDS
Eastern Europe, evolutionary process, isolation by adaptation, isolation by distance, landscape genetics, *Passer domesticus*

Introduction

It has become clear through theoretical and empirical research that neutral as well as selective evolutionary processes can result in population divergence and ultimately lead to speciation (e.g., Coyne and Orr, 2004). While neutral processes such as isolation by dispersal limitation (IBDL; Orsini et al., 2013) can lead to a pattern of isolation by distance (IBD; Wright, 1943) or isolation by landscape resistance (McRae, 2006), it is unclear how influential these forces are, and recent evidence suggests instead that divergent selection may be a major driver of evolutionary change (e.g.,Ellner et al., 2011; Hendry and Kinnison, 2001). Currently, the relative importance of each of these processes often remains unresolved (Mitchell-Olds et al., 2007).

Both neutral and selective processes have been well studied and documented (e.g., Mitchell-Olds et al., 2007), but have in many cases been investigated independently from one another. However, it is crucial to simultaneously assess the potential role of neutral divergence and that of selection in a comparative framework. Classic approaches to demonstrate the presence of selection and local adaptation in a species are common garden or reciprocal transplant experiments in which the fitness of individuals from locations with strong environmental differences are compared (Kawecki and Ebert, 2004). Advantages to this approach include the acquisition of direct evidence for local adaptation and the potential to quantify the resulting fitness consequences to then identify the specific agent of selection (Kawecki and Ebert, 2004). However, such experiments are difficult to apply to organisms with long generation times, complex ecological requirements, or life cycles that are difficult to mimic experimentally (Savolainen et al., 2013).

As an alternative, landscape genetic approaches directly associate phenotypes or genotypes with environmental variables and measures of geographic distance or topography (Manel et al., 2003; Storfer et al., 2007). While these approaches do not provide fitness estimates, their power lies within the joint processing of biological traits and a large variety of environmental variables measured on the ground and from remote sensors. To this end, morphological measurements are useful markers as they may directly represent responses to natural selection. However, whether or not such a response has an adaptive genetic basis, or is merely plastic, remains unclear. To complement morphological measurements, as a genetic marker of choice, easily obtained microsatellite repeat markers do not provide insight into specific adaptations, but are nevertheless useful in a first order assessment of the overall relative importance of neutral and selective processes in driving and maintaining population divergence (Orsini et al., 2013). Such neutral markers diverge through the process of genetic drift, effects of which are maintained by either increasing geographic distance, physical barriers or inhospitable habitat conditions between populations (landscape resistance; McRae, 2006), or by the reduced fitness of dispersing individuals that are maladapted to the conditions at new locations (Nosil et al., 2009). Thus, a correla-

tion between neutral markers and environmental variables that cannot be explained by geographic distance alone may be indicative of divergent selection driving population divergence, a phenomenon termed isolation by adaptation (IBA; Nosil et al., 2009).

House sparrows (*Passer domesticus*) are a suitable species to examine landscape-level patterns of intraspecific variation, because they are widespread and occur along a range of different environmental conditions that may pose divergent selection pressures (Kekkonen et al., 2011; MacGregor-Fors et al., 2010; Vangestel et al., 2012). Here, we studied the relative roles of neutral and selective processes on the divergence of natural house sparrow populations in Romania and Bulgaria, a still understudied region in Europe. To do so, we: (1) analyze the population genetic structure based on microsatellite markers; (2) relate morphological and genetic variation to environmental variables and measures of geographic distance and landscape resistance; and (3) compare the importance of natural habitat variables with those related to human habitation. Finally, because protecting standing intraspecific variation will help maximizing a species' evolutionary potential facing changing environmental conditions (Brooks et al., 2015; Dawson et al., 2011; Frankham, 2010; Grivet et al., 2008; Hartl et al., 2003; Matala et al., 2014; Smith et al., 2001; Thomassen et al., 2011; Vandergast et al., 2008), and intraspecific variation in common species may represent that in species of conservation concern (e.g., Thomassen et al., 2011), we also aimed to map intraspecific variation in house sparrows in Romania and Bulgaria for conservation purposes. We used morphological and genetic data collected from 691 individuals from 33 populations distributed across and along environmental gradients in temperature, precipitation, elevation, and land cover. As morphological markers, we used the size and shape components resulting from a "PCA ratio spectrum" analysis (Baur and Leuenberger, 2011) of a set of measurements describing primarily wing, tail, and tarsus sizes. We complemented our morphological dataset with twelve microsatellite markers, eight of which were found to be polymorphic. To then relate intraspecific variation to environmental variables, we used a dissimilarity-based matrix regression (generalized dissimilarity modeling; GDM) technique that - in contrast to other methods often applied - can simultaneously take into account the effects of distance and environment (Ferrier et al., 2007).

Methods

Study region

Romania and Bulgaria are located in southeastern Europe (Figure 1a) and comprise distinct climatic zones: the continental and Mediterranean climatic zones in Bulgaria, and the continental and temperate climatic zones in Romania. The Danube River forms a natural border along much of its length between Romania in the north and

Bulgaria in the south. Large mountainous areas, with peaks up to about 2,500 m, cover much of the land surface in these countries; in Romania, the Carpathian mountain region is predominant, whereas the Balkan, Rhodope, Rila, and Pirin mountains merge to a large mountainous area in Bulgaria (Figure 1b). At a smaller scale, the landscape in this region can be characterized as extensive and intensive agriculture interspersed with seminatural areas consisting of forest, open woodland, and grassland. As a result of this variation of habitats, different biogeographical regions are recognized, includ-ing the continental, alpine, steppic, black sea, and pannonian regions ((CoE), 2015). This habitat mosaic constitutes an ideal test bed to study evolutionary processes in natural populations, because of its high habitat heterogeneity across short distances, allowing for the potential of strong divergent selection pressures on natural populations.

Study species

House sparrows are a widespread, synanthropic species (Anderson, 2006). It has been suggested that factors related to human habitation and land use play a key role in the abundance and genetic diversity of house sparrow populations (Kekkonen et al., 2011; Vangestel et al., 2012). Postnatal dispersal distances are low, ranging between 1 and 1.7 km (Anderson, 2006; Paradis et al., 1998), allowing for the potential for population divergence to be driven by IBD (Kekkonen et al., 2011; Vangestel et al., 2012). Previous studies of house sparrow population structure in other regions demonstrated varying levels of divergence. For instance, Finnish populations were found to be essentially panmictic, with little evidence for a pattern of IBD (Kekkonen et al., 2011). In contrast, populations in mainland Norway and associated islands showed low- to- moderate divergence, most likely caused by IBD (Jensen et al., 2013). Similarly, weak but significant structure was observed in native populations in Belgium (Vangestel et al., 2012) and France (Liu et al., 2013), and in introduced populations in Brazil (Lima et al., 2012).

Field sampling

Samples were collected in 2007 and 2008, and 2013–2015 at 33 locations throughout Romania and Bulgaria (Figure 1b; Table S1). Sites were selected based on two key criteria: (1) the full set of sites covers as much as possible of the environmental niche breadth observed in the study area and (2) sites are located across as well as along environmental gradients, such that the potential effects of geographic distance and environmental gradients on population divergence are decoupled and can be distinguished in subsequent correlative analyses. Identification of gradients and selection of sites were performed using available climate and satellite remotely sensed habitat data at 0.25- to 1- km resolutions (see below). All sampling locations were near

anthropogenic sites. Birds were captured using mist nets, which were set up around villages and at the edges of gardens or farms. Individuals were sexed, morphological measurements were recorded, and DNA samples obtained via two tail feather and blood samples. Feathers were stored dry in envelopes and blood samples in >96% ethanol. Birds were banded and released immediately after processing at the site of their capture. In total, 691 individuals were sampled (on average $\sim$ 21 per site): 314 males, 302 females, and 75 unsexed individuals.

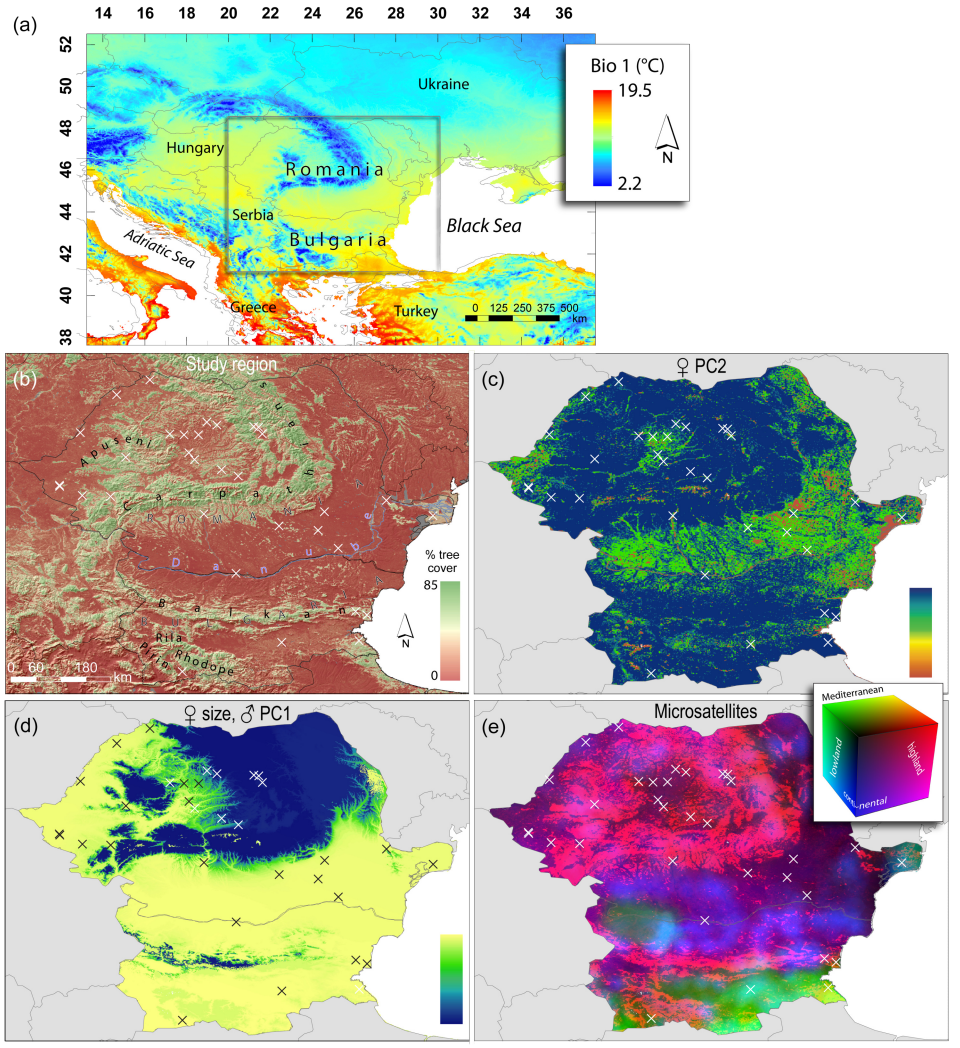

FIGURE 1: Study region, sampling sites, and generalized dissimilarity modeling results. (a) Location of the study region within Eastern Europe, with average temperature of the year (Bio 1). (b) Overview of the study area, with sampling sites (crosses) on a hillshade map and an overlay of percent tree cover. (c–e) GDM results for the second morphological shape component for females (c), the morphological size component in females and the first shape component for males (d), and microsatellites (e). The color difference between two locations along the color bar (c, d) or on the RGB color cube (e) in the GDM maps represents the magnitude of the difference in the biotic response variable, that is, morphological variable or FST.

Morphological measurements and analyses

Because samples were collected over the course of several years and additional measurements were added later in this study, a set of three morphological measurements were available for all individuals (wing length, tail length, and tarsus length), and an additional four measurements for only a subset of our samples (populations from Ognyanovo, Beli bryag, Jasna poljana, Popovits, Golica, Poiana, Berzovia, Salonta, Parta, Caransebeş, Mihăești, Hălmagiu, Măgheruș, Runc and Lăzarea): culmen length, bill depth, head width, and head length. Morphological data were analyzed for adult indi-viduals only, and because of sexual dimorphism in this species (with males being generally larger than females), for males and females independently.
Raw morphological measures are unlikely to be independent from one another due to allometric relationships, and as a result, we used the raw morphological data to create a size and several independent shape components using the "PCA ratio spectrum" method developed by Baur and Leuenberger, 2011 (Appendix S1). For the size component and each shape component that explained >10% of the total variation in the PCA, we computed population pairwise differences as follows: $|\overline{x} - \overline{y}| / \sigma_x + \sigma_y$, where x and y are the averages for populations $\overline{x}$ and $\overline{y}$ and σ_x and σ_y are their standard deviations. Because we only had partial datasets-one with three morphological variables (wing, tail, and tarsus lengths) for all locations, and one with all morphological variables (also including culmen and head lengths, head width, and bill depth) for only nine locations, we evaluated which one was the most appropriate to use. Our assessment suggested that the three- variables- all- locations dataset gave the most robust results (Appendix S1).

Laboratory methods and genotyping

DNA was extracted using the QIAGEN DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Because of potential misidentification of Spanish sparrows (*P. hispaniolensis*) as house sparrows, we genetically identified individuals to species from southern sampling sites using the cytochrome c oxidase subunit I (COI) mitochondrial gene (Appendix S1).
We genotyped house sparrow individuals for twelve published microsatellite loci (Dawson et al., 2012; Garnier et al., 2009; Griffith et al., 2007) (Table S2). Of these twelve loci, two were monomorphic in the majority of the sampling locations after initial genotyping of a subset of individuals and omitted from further analyses. Fragment length analysis was carried out on an ABI 3730 sequencer at the University of Turku, Finland. Results were analyzed with GeneMarker V2.4.1 (Softgenetics, State College, PA, USA).

Population genetic analyses

Because only a few birds could be sampled at certain locations, we calculated the geographic distances between locations, and pooled locations with small sampling sizes with those nearby (Măgheruș and Runc were 13.8 km apart; Parta 1 and Parta 2 3.6 km). Loci were checked for the presence of null alleles using MICRO-CHECKER (Van Oosterhout et al., 2004), deviations from Hardy–Weinberg Equilibrium (HWE) using GenAlEx 6.501 (Peakall and Smouse, 2006), and Linkage Disequilibrium (LD) using GENEPOP web version 4.2 (Rousset, 2008). We used COLONY version 2.0.5.9 to identify full siblings within sampling sites. Presence of full siblings would confound Bayesian clustering analyses and FST estimates. All individuals were coded as offspring. No full-sib ships were detected (results not shown); hence, all individuals were kept for subsequent analyses. To assess the level of genetic structure, we conducted two Bayesian clustering analyses: STRUCTURE (Pritchard et al., 2000) and GENELAND (Guillot et al., 2005a; Guillot et al., 2005b; Guillot et al., 2008) (Appendix S1).

For subsequent landscape genetic analyses, we calculated site pairwise FST values. Because a signal of null alleles was detected, we computed corrected FST using the "excluding null alleles" (ENA) method implemented in FreeNA with 10,000 bootstrap replicates (Chapuis and Estoup, 2007). To minimize the risk that potential correlations between FST and environmental variables are the result of demographic processes, we evaluated whether or not population divergence was simply a result of differences in genetic variation within populations (Appendix S1). We also tested whether morphological divergence and genetic population divergence were concordant using a Mantel test with 999 permutations.

Environmental variables

To describe environmental conditions across Romania and Bulgaria, we compiled a set of 34 environmental variables related to climate, topography, vegetation, and human habitation (Table S3) at 30 arc-sec resolution. Although the home-range sizes of individual birds are likely much smaller, spatial heterogeneity in climate variables within each grid cell is small compared to that between distant grid cells, and dispersal has been reported to be up to 1.7 km (Anderson, 2006; Paradis et al., 1998). The used spatial resolution of variables thus balances home-range size with dispersal distances as well as availability and computational tractability of subsequent analyses. Bioclimatic variables expressing variations in temperature and precipitation were obtained from WorldClim (http://www.worldclim.org/) (Hijmans et al., 2005). These variables are derived from a network of weather stations and are based on a 50-year climatology from 1950 to 2000. Elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) and used directly in further analyses as well as to compute slope (steepness of the terrain) and aspect (the compass direction that a

slope faces). Vegetation data included the percent tree cover from 2001 (Hansen et al., 2002) and Leaf Area Index (LAI; Myneni et al., 2002) obtained from the Global Land Cover facility database (http://www.glcf.umd.edu/data/). We also used a measure of surface moisture based on the QuikSCAT microwave instrument (QSCAT; Long et al., 2001). For areas with dense forest, QSCAT is sensitive to canopy roughness. We computed multiyear (2000-2008) averages of raw backscatter measurements at the horizontal polarization, including means, minima, maxima, and seasonality (expressed as the coefficient of variation). Further details on the computation of QSCAT variables are provided in Appendix S1.

Because sparrows are commensal with anthropogenic activity, we also included two measures of human habitation as predictors: road density and human population density. The road density layer was created out of a shape file of roads (Digital Chart of the World, downloaded from http://www.diva-gis.org/gdata on 21 November 2013), processed in ArcGIS 10.0 (ESRI, Redlands, USA) using the "line density tool." The output cell size was set to 0.0083333 degrees (i.e., 30 arcsec) to match the other environmental variables, and because of the short natal dispersal and small home-range sizes of sparrows (Anderson, 2006; Paradis et al., 1998), the search radius was set to five map units ($\sim$ 5 km). Human population density data were obtained from the Gridded Population of the World dataset, version 3 for the year 2000 at 2.5 arcmin ($\sim$ 5 km) resolution (Center for International Earth Science Information Network (CIESIN), Columbia University 2005; http://sedac.ciesin.columbia.edu retrieved 12 May 2015).

In addition to straight-line geographic distance, we included two other types of distance that may be more realistic measures of the distance dispersing individuals have to travel to reach another location. First, because of the short postnatal dispersal distances of house sparrows (1–1.7 km; Anderson, 2006; Paradis et al., 1998) and the width of the Danube River at places reaching 1.5 km, the Danube was included as a barrier to dispersal. In a GIS layer, areas north of the Danube River were coded 0, and those south 1, resulting in differences of 1 between sampling sites across the river and of 0 between those on one side of the river. Second, we computed resistance distances based on human population density in Circuitscape 3.5.8 (McRae, 2006). To do so, human population density was treated as a conductance map (i.e., higher densities are favorable to dispersal and gene flow), and a cell connection scheme of eight neighbors was used.

To reduce this set of environmental variables to a smaller suite that each provided unique information, we extracted their values at the sampling sites using ArcMap 10.2.2 (ESRI, Redlands, USA) and computed Pearson correlation coefficients (logistic regression in the case of the Danube River barrier) in R 3.1.2 (Table S4). When two variables had a Pearson correlation coefficient "$a \geq$ b"0.7 (or $p < .05$ for the logistic regression), one of them was excluded from further analyses. Of those pairs, we retained the one that is more easily interpretable (e.g., Bio 1: mean temperature of the

year versus Bio 3: isothermality).

Landscape genetic analyses

To assess correlations of morphological or genetic data with environmental variables, geographic distance, and landscape resistance, we used generalized dissimilarity modeling (GDM; Ferrier et al., 2007), implemented in the R package gdm (Manion et al., 2016). GDM has increasingly been used in landscape genetic studies, including in tests for IBA (Freedman et al., 2010; Mitchell et al., 2015; Thomassen et al., 2010). It is an iterative matrix regression method that fits dissimilarities of predictor variables to dissimilarities of a response variable. It can analyze and predict spatial patterns of beta diversity across large areas, using I- spline basis functions to adjust nonlinear relationships between environmental variables and biological variation (Ferrier et al., 2007). In this work, GDM was used to predict the relationship between a set of predictor variables and pairwise genetic distances (FST) or morphological differences as response variables. The predictor variables consisted of environmental variables, geographic distance, resistance distance, and the Danube barrier, and were selected to determine the biotic variation that is explained by IBA, isolation by distance (IBD; Wright, 1943), or resistance by the habitat matrix in between populations. The importance of predictor variables is tested by permutations, where only variables that contribute significantly to explaining variation in the response variable are retained. The relative importance of predictor variables can be evaluated by examining the maximum height that is reached in variable response curves. In total, five types of models were performed: (1) a best fit model (including all environmental variables, as well as geographic distance); (2) a model with only the environmental variables; (3) a model with only straight-line geographic distance; (4) a model with only resistance distances; and (5) a set of 1,000 models with random environmental variables to evaluate the significance of the variation explained by the best fit model. The best fit model was considered not significant if the variation explained fell below the upper 95% confidence interval of the random models. Model fit was visualized in a scatter plot of predicted versus observed response values.

In a subsequent step, the spatial distribution of the response variable can be projected across the study area using the known environmental conditions (obtained from the predictor variables) outside the sampling locations and the calculated relationship between the environment and biological variation. We visualized this variation in the response variables in three-dimensional RGB color space. To do so, we followed the computationally tractable approach from Fitzpatrick and Keller, 2015. Briefly, we first extracted the values of the retained environmental variables at a grid with 30 arcsec resolution, corresponding to the midpoints of grid cells in the 30 arcsec WorldClim dataset. We then ”transformed” the environmental variables for these sites into a set of ”genetic importance” variables (Fitzpatrick and Keller, 2015). We

conducted a principal component analysis (PCA) on these transformed variables to obtain a smaller set of independent variables. We then matched RGB values to the first three PC axes, which were subsequently combined into one multiband RGB GIS layer in ArcMap 10.2.2 (ESRI, Redlands, USA). We verified that the resulting maps were concordant with those obtained using the "predict.gdm" function with subsequent multidimensional scaling, but which was only possibly at low resolution due to computational limitations (Appendix S1).

Finally, to inform conservation practices in Romania and Bulgaria, we visually assessed whether current protected areas capture genetic and morphological variation in house sparrows sufficiently well (Appendix S1).

Results

Size and shape components of morphological measurements

We used the "PCA ratio spectrum" method to distinguish between the size and shape components of the morphological measurements. PCA results of the shape component are shown in Table S5. For both males and females, the first two extracted principal shape components explained all of the observed variation. PCA ratio spectra for wing, tail, and tarsus length are nearly identical for males and females (Figure S1) and suggest that most variation along the first axis is explained by the ratio of tail and tarsus, and along the second axis by the ratio of tail and wing. These results for the dataset with just wing, tail, and tarsus length but for all locations are supported by those for the all-variables-nine-locations dataset for the first axis. This was, however, not the case for the second axis, where tail and wing are close to one another on the axis (explaining very little of the variation), bill depth is positioned on one end of the spectrum, contrasted on the other end by tarsus and culmen in males and by tarsus and head measures in females. Along the third axis, most variation is explained by the ratio between culmen and tail in both males and females, but bill depth is also important in males, whereas it is not in females. We conducted subsequent landscape genetic analyses using the PC scores of the size component and the first two shape components in males and females separately.

Population genetic analyses

The number of effective alleles (NE) ranged from 3.184 to 6.887; HO from 0.583 to 0.846; and HE from 0.590 to 0.794 (Table S6). Two microsatellite loci were found to be out of HWE in many sampling locations: Pdo31 significantly deviated from HWE in 15 locations and Pdo7 in 25 locations. These loci were, therefore, omitted from further analyses. After Bonferroni correction, no loci were in significant LD. We found a signal for the presence of null alleles and therefore calculated ENA-corrected

(Chapuis and Estoup, 2007) FST values for the remaining eight loci to be used in subsequent landscape genetic analyses. The global population variation across all loci and all sites was FST = 0.011.

STRUCTURE analyses using the admixture model with location prior and either correlated or noncorrelated allele frequencies suggested there is no clear genetic structuring among sparrow populations in Romania and Bulgaria (K = 1). Inclusion of the spatial component using GENELAND supported this finding. When we did not use a model for null alleles, all ten independent runs inferred six clusters (K = 6). However, assignments of cluster membership were highly inconsistent between runs (not shown), and we therefore concluded that there was little evidence for significant population structure based on these analyses. A lack of clear population genetic structure, however, does not necessarily mean a lack of IBD or IBA; merely that selection pressures may be relatively low, or there is a much relatively recent or ongoing gene flow. In fact, correlation analyses between genetic divergence and environmental heterogeneity may be better suited to identify potential patterns of IBD or IBA than those purely based on genetic data. We, therefore, proceeded with landscape genetic analyses using the ENA- corrected FST values (Chapuis and Estoup, 2007).

Mantel tests between FST and morphological divergence were only significant for shape PC2 in females but with a low correlation (Z = 23.80624, r = .289, p = .001 for 999 permutations; for female size Z = 17.79662, r = .111, p = .063 for 999 permutations; for male PC1 Z = 24.01073, r = .137, p = .075 for 999 permutations).

Landscape genetic analyses

Among the morphological variables, models for the first shape com-ponent (shape PC1) in males and for the size and second shape (shape PC2) components in females performed better than random models (Table 1). For shape PC1 in males, geographic distance was included in the best fit model, but explained very little of the variation when used alone, and similar results were found for size and shape PC2 in females. Thus, IBD appears to play only a minor role in driving population divergence in morphological variables. This finding is supported by the lack of a correlation between morphological divergence and geographic distance (Figures S2c–S4c). The mean temperature of the driest quarter (Bio 9) was the most important variable explaining variation in PC1 for males and size in females (Figures S2a and S3a), whereas minimum leaf area index (LAImin) was the most important variable describing variation in shape PC2 in females (Figure S4a). Variables related to human habitat contributed little (for shape PC2 in females) to no explanatory power to help distinguish morphological variation.

The best fit generalized dissimilarity model for microsatellites, where all variables were entered in the model, explained 24.95% of the observed variation (Table 1) and

only retained environmental variables in the final model. A model with only geographic distance or resistance distance as the predictor variable explained 3.84% and 7.93% of the variation, respectively, and random models explained 4.35% of the variation, with an upper confidence level of 4.44%. These results also suggest that local environmental conditions rather than isolation by distance or isolation by resistance are important in generating house sparrow population genetic divergence, which is supported by a lack of correlation between FST and geographic distance (Figure S5c). The variables most important in explaining the observed genetic variation were annual precipitation (Bio 12), mean leaf area index (LAImean, a measure of greenness), mean temperature of the driest quarter (Bio 9), and precipitation of the driest month (Bio 14) (Figure S5a). Road density was also retained as an explanatory variable, but did not contribute as much as the above-mentioned climate and vegetation variables.

TABLE 1: Results of generalized dissimilarity models of the size and shape components of wing, tail, and tarsus length measurements and of microsatellites.Numbers represent the total observed variance (%) explained by the best fit model (Best fit) and models with only environmental variables (Env only), only geographic distance (Dist only), and the mean value of 1000 models with random environmental variables (Random) and the associated confidence intervals (Lower CI, Upper CI).

	Best fit	Env only	Dist only	Random	Lower CI	Upper CI
Males size	6.8	6.8	0.2	6.3	6.2	6.4
Males shape PC1	30.3	30.1	0.0	7.0	6.9	7.2
Males shape PC2	6.6	6.6	0.0	6.5	6.4	6.6
Females size	14.7	14.4	0.0	6.5	6.3	6.6
Females shape PC1	2.1	2.1	0.0	5.6	5.5	5.7
Females shape PC2	27.3	27.3	1.9	6.5	6.4	6.7
Microsatellites	25.0	25.0	3.8	4.4	4.3	4.4

Discussion

Landscape genetic analyses

We examined whether neutral (isolation by dispersal limitation) or selective evolutionary processes are the most important drivers of house sparrow population divergence in Romania and Bulgaria and whether measures of human habitation play a role in the divergence in this human commensal species. We found that IBDL could not explain either morphological or genetic divergence, whereas environmental variables explained a large proportion (up to 30%) of the observed variation. Our results for morphological measurements were thus consistent with a signal of selection. Although the number of polymorphic microsatellite markers was relatively low, and a large set of SNP markers will be more suited to get insight into population divergence and selection at the genetic level, results for microsatellites were nevertheless consistent with a pattern of IBA, and thus support the morphological data in the notion

that adaptive processes are more important than neutral ones in driving population divergence. Our results suggesting that divergent natural selection is a main driver of intraspecific variation in this species are in agreement with findings for populations in Norway (Holand et al., 2011), Brazil (Lima et al., 2012), and France (Liu et al., 2013). However, these studies were conducted at much smaller scales, with much fewer populations. Moreover, those in Norway and Brazil did not relate population divergence to environmental variables, but rather compared FST to estimates of morphological divergence (QST or PST). Perhaps more importantly, the study in France found fine-scale spatial autocorrelation, suggesting IBDL at short distances, but the potential effect of distance was not included in subsequent correlative analyses with environmental factors, making it difficult to assess the relative importance of IBDL versus IBA.

The spatial patterns of morphological variation in the size component in females and shape PC1 in males show a very sharp division between higher and lower elevation areas (Figure 1d) due to a large response to small differences in mean temperature of the driest quarter (Bio 9) between mountain and lowland areas, which then levels off to a flat response at larger differences (Figures S2a and S3a). In contrast, the spatial pattern of variation in shape PC2 in females is more complex (Figure 1c): the main turnover of the morphological measures occurs at smaller differences in minimum leaf area index (LAI min; Figure S4a). The potential underlying causal relationship between shape PC2 in females-dominated by wing length- and minimum leaf area index remains unclear. Wing length in birds is often related to vegetation density, where individuals from forests tend to have shorter wings than those from the open field because of the advantage of shorter wings for maneuverability in dense vegetation; however, we did not find such a relationship in our house sparrow samples (results not shown), nor did we find that leaf area index was an important factor in the shape components of males, as would be expected given that both males and females should exhibit similar selection pressures for wing length related to vegetation. As for microsatellite variation, spatial patterns roughly follow a lowland versus highland and Mediterranean versus continental subdivision (Figure 1e). Specifically, higher elevation populations are genetically similar, but lowland populations from southern Bulgaria, with a more Mediterranean climate, are distinct from those in Romania, where a more continental climate prevails. In addition, lowland populations from the Danube Delta are nearly as distinct from other lowland populations as the latter are from higher elevation populations.

Despite only subtle population differentiation at the genetic level, we found that divergence is tied to the environment, independent of geographic distance. Further support for these findings comes from visual inspection of observed versus predicted values and plots of population divergence versus geographic distance (Figures S2b–S5b and S2c–S5c, respectively). Of all variables entered into the models, only a small set was selected that explained most of the observed vari-ation (Fig-

ures S2a–S5a), notably mean temperature of the driest quarter (Bio 9), annual precipitation (Bio 12), and mean and minimum leaf area index (LAI mean, LAI min). However, visual examination of the shape of the response curves suggests that the effects of those predictor variables vary between response variables. For instance, for the shape PC1 in males (Figure S2a) and the size component in females (Figure S3a), there is a very steep response to small changes in the mean temperature of the driest quarter (Bio 9), which then quickly levels off. In contrast, for genetic variation, small differences in the mean temperature of the driest quarter (Bio 9) do not result in larger FST values (Figure S5a); larger differences, however, result in exponentially increasing divergence. These results thus suggest that divergence in morphological traits is not shaped by the same environmental variables as in microsatellites. Further insight into this issue comes from the correlations between FST and morphological divergence, as well as from a comparison of the spatial patterns of variation shown in the GDM maps (Figures 1c–e). A crude subdivision into highland and lowland populations in both morphological traits and microsatellites and a small but significant correlation between FST and female shape PC2 suggest that similar factors may underlie population divergence in phenotype and genotype. However, finer substructuring of populations and a lack of correlations between FST and female size and male shape PC1 indicate that such a pattern is not broadly supported. Thus, if genetic divergence indeed is related to IBA, the factors that limit gene flow, leading to neutral divergence in microsatellites, must be primarily physiological characteristics or morphological variables other than those measured here.

Although in our study selective processes appeared to be the most important factors underlying population divergence, most of the variation ($\sim$ 70% or more; see also the spread of points in Figures S2b–S5b) could not be explained, despite the fact that many predictor variables were considered. We can only speculate about additional factors that may cause population divergence. One explanation may be that habitat conditions other than the ones included may cause strong divergent selection or limit dispersal between populations. Such conditions should be measured at much smaller scales than those used in our study and may include microhabitat characteristics such as the grain size of crops grown, types of cattle feed used, and available to this granivorous species, or food availability, which was found to be related to population divergence in a valley in France (Liu et al., 2013). A similar result was found for rural and urban populations in Hungary, but common garden experiments suggested that food availability did not result in a short-term response in body mass (Liker et al., 2008). The high level of heterogeneity of the landscape mosaic in Romania and Bulgaria suggests that the process of local adaptation may occur at relatively small scales in those countries. If so, our estimate of the relative importance of IBA in population divergence is conservative. Another category of factors that may explain the remaining variation is related to chance events that are not linked to long-term environmental conditions or the distance between populations, such as population

demographic fluctuations or isolation by colonization (IBC; De Meester et al., 2002; Orsini et al., 2013). Under IBC, a signal of founder effects can persist over time due to monopolization, where local adaptation is based only on standing genetic variation present in the first colonizers. However, relatively high population divergence is expected under such a scenario, which does not seem to be the case in our study region. Finally, morphological characteristics may rather be shaped by sexual than by environmental selection. Most studies on sexual selection in house sparrows have focused on the size of the black patch on males' chests and on the white wing stripe, but females have also been shown to prefer larger males in some populations (Moreno-Rueda and Hoi, 2012). Although the morphological traits measured here have not been implicated in sexual selection so far, it is conceivable that at least part of the divergence -in particular in the size component in males- can be attributed to differences in mate preferences between populations.

Influence of human habitation

Measures of human habitation appear to have little effect (positive or negative) on population divergence of house sparrows in Romania and Bulgaria. First, although road density was selected as a predictor in microsatellite variation and in the shape component PC2 in females, it did not contribute much to explaining the observed variation. Similarly, human population density was among the predictors in the model for microsatellite variation, but this variable retained a comparatively low importance score. Even though we have not sampled house sparrows in cities and thus lack information on this extreme end of the range of niches available along gradients in human-dominated landscapes, our results are broadly concordant with those of Vangestel et al., 2012, who found no evidence for divergence between urban and rural house sparrow populations in Belgium (but see e.g., Liker et al., 2008 for morphological characteristics). Second, despite a lack of clear genetic structure, we expected that trends in genetic and morphological variation would be correlated to dispersal pathways facilitated by human habitation. For instance, Schrey et al., 2014 found evidence that population expansion of house sparrows in Kenya could be explained by human-mediated dispersal. However, in our study, resistance distances based on human habitation were not included in any of the models, suggesting that dispersal in these populations is not limited nor mediated by human activities.

Conservation recommendations

Although house sparrows are listed by the IUCN as of least concern (IUCN 2015), their populations are declining, most notably in their native range (Anderson, 2006; De Laet and Summers-Smith, 2007; Murgui and Macias, 2010). The underlying causes of their decline remain poorly understood, but may be related to predation, competi-

tion, disease occurrence (Kruszewicz et al., 1995), an increase in pollution (Summers-Smith, 1999), and changes in anthropogenic activity that led to a shortage in food sources (Hole et al., 2002) and nest sites (Siriwardena et al., 2002). While house sparrows currently appear to be abundant in Romania and Bulgaria, the ongoing modernization of agriculture (Ioras, 2003) and predicted climate change may thus impact their numbers and require adaptive genetic or phenotypic changes. Moreover, intraspecific variation in these house sparrows may be a surrogate for that in other, less common species. In our preliminary and qualitative assessment, we found that environmentally associated intraspecific variation is likely insufficiently protected (Figure S6; Appendix S1). Particular conservation attention is warranted for lowland areas bordering the Danube River in the west, and the elevation gradient along the southern Carpathian Mountains. The results from the current study will be incorporated in much more detail in forthcoming work aiming at prioritizing areas for conservation in this biologically rich region, unique for Europe (e.g., Iojă et al., 2010; Wilson et al., 2012).

In summary, we found that selection by environmental variables, but not IBDL, is the main driver of population divergence in Romanian and Bulgarian house sparrow populations. Variables related to climate and vegetation best explained intraspecific variation, whereas those related to human habitation contributed comparatively little. Our study thus contributes to a growing body of literature suggesting that divergent selection may be a key driver of population divergence in many species and populations, and it improves our understanding of the spatial patterns and drivers of biodiversity in an understudied region.

Acknowledgements

We thank N. Tzankov for help in obtaining permits to work in Bulgaria; D. Wegmann and R.J. Harrigan for useful discussions; and four anonymous reviewers for feedback that improved this work. We also thank Costică Adam for his kind help with data collection. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007- 2013 under REA grant agreement number 293886. CIV was financed by the Postdoctoral Fellowship Programme of the Hungarian Academy of Sciences and PLP by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Sampling was conducted under Romanian authoriza-tion number 2257/09.12.2009 and Bulgarian authorization number 520/23.04.2013.

Supplementary information

Supplementary Materials and Methods

Genetic identification of *Passer domesticus* and *P. hispaniolensis*

In the southern part of its range in Bulgaria, house sparrows (*Passer domesticus*) co-occur with the Spanish sparrow, *P. hispaniolensis*, and are even found breeding in mixed colonies. Males of the two species are easily distinguishable, but females are very similar in appearance. We, therefore, identified the species of females sampled in southern populations using molecular methods. To do so, we sequenced a 749-base pair region of the cytochrome c oxidase subunit I (COI) mitochondrial gene using the primers BirdF1 and BirdR1 (Hebert et al., 2004). The COI gene is known for its relatively fast mutation rate and its use across many taxa for barcoding purposes. The PCR reaction mix consisted of 2.5 µl 10× PCR buffer, 13.4 µl HPLC water, 1.0 µl dNTP's (10 mM), 2.0 µl MgCl2 (25 mM), 1.0 µl BSA (20 mg/ml), 1.0 µl of each primer (0.1 mM), 0.1 µl Taq DNA polymerase and 3.0 µl extracted DNA per reaction. PCR conditions were: 94°C initial denaturation for 5 minutes; 39 cycles of 94°C denaturation for 40 seconds, 50°C annealing for 40 seconds, and 72°C extension for 1 minute; and 72°C final extension for 10 minutes. PCR products were visualized using agarose gel electrophoresis to check for the amplification of the fragment. Successfully amplified PCR products were cleaned up using the Promega Wizard© SV Gel and PCR Clean-Up System according to the manufacturer's protocol. Cleaned up samples were then sent to LGC Genomics (Berlin, Germany) for sequencing. Resulting sequences with corresponding chromatograms were visualized and edited where necessary with Unipro UGENE v1.12.2 (Okonechnikov et al., 2012). We used ClustalX 2.1 (Larkin et al., 2007; Thompson et al., 1997) to align sequences with each other and with reference sequences for *P. domesticus, P. hispaniolensis,* and *P. montanus*, obtained from GenBank. We then used Mega v. 6 (Tamura et al., 2013) to construct a Maximum Likelihood tree using the Hasegawa-Kishino-Yano (HKY) substitution model with uniform rates and 1000 bootstrap replicates. Individuals that were included in a monophyletic group with reference sequences of *P. domesticus* were considered members of the target species and included in further microsatellite analyses.

Morphological measurements and analyses

Morphological measures are unlikely to be completely independent from one another. If an individual grows isometrically, i.e. without changing shape, all measures will be affected equally. For multivariate analyses, it is therefore essential to distinguish between the isometric size component and the allometric (shape) component of these measures. Several methods are available to do so, such as Procrustes analyses on morphological landmarks, or standardization using an independent measure of body size. However, neither of these were available based on our collection methods, so we used the "PCA ratio spectrum" method developed by Baur and Leuenberger, 2011. This method identifies common patterns - that could represent isomet-

ric growth - among all possible ratios of morphological measurements, and returns independent PCA scores for one size and several shape components. For the size component and each shape component that explained > 10% of the total variation in the PCA, we computed population pairwise differences as follows: $|\overline{x} - \overline{y}|/\sigma_x + \sigma_y$, where x and y are the averages for populations $\overline{x}$ and $\overline{y}$ and σ_x and σ_y are their standard deviations. Because we only had partial datasets - one with three morphological variables (wing, tail, and tarsus lengths) for all locations, and one with all morphological variables (also including culmen and head lengths, head width, and bill depth) for only nine locations - we needed to decide which of those datasets was most appropriate to use. To further explore these datasets, we in fact ran PCA ratio spectra and generalized dissimilarity models for both and compared the results (not shown). We found that the percent of total variation explained by our models was about twice as high for the all-variables-nine-location dataset as compared to the three-variables-all-locations dataset. To investigate whether this could potentially be explained by the difference in the number of locations, we reduced the three-variables-all-locations dataset to only comprise the nine locations of the other dataset, and ran GDMs with this reduced dataset. We found that this data reduction method almost doubled the percent of total variation explained in GDMs. Although we cannot rule out that this is a real pattern among those nine locations, we suspected that these results were an artifact of entering too few locations in the models, and we only present the results for the three-variables-all-locations dataset, except for the results from PCA ratio spectra.

Microsatellite analysis

The extracted DNA was used to determine intraspecific genetic variation by genotyping an initial set of twelve microsatellite loci (Pdo31, Pdo75, Pdoμ3, PdoA06, PdoA08, PdoH05, Pdo7, Pdo10, Pdo16, Pdo36, Pdo46, and PdoF05) by using the M13-hybrid primer process (Boutin-Ganache et al., 2001; Schuelke, 2000). This procedure uses three types of primers: first a hybrid primer, which consists of the forward primer with a tagged-on M13F sequence (16p: 5'-GTAAAACGACGGCCAG-3') on the 5'end; second the corresponding reverse primer; and third a dye labeled M13F primer, which is complementary to the M13F sequence. The primer mix consisted of 4 µl of the reverse primer (100 µl), 8 µl forward-M13 hybrid primer (2.5 µM), 8 µl M13 dye labeled primer (2.5 µl) and 180 µl water. To run the multiplex PCR, a PCR reaction mix was made, for each 10 µl reaction consisting of 1.0 µl primer mix, 2.1 µl water, 0.4 µl Bovine Serum Albumin, and 5.0 µl Qiagen Multiplex Mastermix added to 1.5 µl sample DNA. The PCR was then run as a two-stage cycle starting with 15 minutes at 95 °C, followed by the first cycle which consists of three steps: 30 seconds at 94 °C, 90 seconds at 55 °C/56 °C/60 °C(depending on the primer mix) and 60 seconds at 72 °C, repeated 25 times. The second cycle also consisted of three steps and was conducted as follows: 94 °C for 30 seconds, then 90 seconds at 53 °C and finally 60 seconds at 72 °C, repeated 20 times. After these cycles, a final step of 60 °C or 30 minutes was run. Fragment analysis was carried out on an ABI 3730 sequencer at the

University of Turku, Finland. Results were analyzed with GeneMarker V2.4.1 (Softgenetics, State College, PA). We used the following peak detection settings: detection range between 100 and 400 base pairs (bp); peak detection threshold with an intensity between 100 and 8000; stutter peak filter of 5% on the left side and 40% on the right side. The detected peaks were then examined visually, and edited where necessary.

Population genetic structure

To assess the level of genetic structure, we conducted two Bayesian clustering analyses. First, we ran STRUCTURE (Pritchard et al., 2000) which uses the genotypic data only and is capable of incorporating putative population origin as prior information, but does not incorporate the geographic sampling location of individuals. We ran five independent runs of 500,000 iterations after a burn-in of 50,000 iterations, exploring values of K - the assumed number of different genetic clusters - ranging from K = 1 to K = 31, which corresponds to the maximum number of sampling locations. In two different analyses, we assumed an admixture model with either correlated or uncorrelated allele frequencies and the sampling site as prior information (LOCPRIOR). The most likely number of clusters was determined using the method proposed by Evanno et al., 2005 in STRUCTURE HARVESTER (Earl et al., 2012).

The second Bayesian clustering method we implemented was GENELAND 4.0 (Guillot et al., 2005a; Guillot et al., 2005b; Guillot et al., 2008), run in R 3.1.2 (R Development Core Team, 2008). In contrast to STRUCTURE, GENELAND explicitly takes into account the spatial location and orientation of samples. Here, we also assumed an admixture model with either correlated or uncorrelated allele frequencies. Because MICRO-CHECKER results suggested the presence of null alleles, we also implemented a null allele model, which attempts to correct for the false identification of genetic structure as a result of the presence of an excess of homozygotes due to null alleles. We ran ten independent runs of 500,000 iterations, thinning of 100, and a burn-in of 50,000 for K = 1 - 31.

Influence of the number of loci on estimates of F_{ST}

The genetic data used here consists of a relatively small set of eight microsatellite markers. To get insight into the robustness of this data set, we computed F_{ST} values of further reduced data sets consisting of all possible combinations of six and seven loci, and correlated these F_{ST} values with those from the full set of eight loci. The average R2 of correlations between all sets of seven loci and the full data set was 0.87 (SD = 0.11; median R2 = 0.93); the average R2 of correlations between all 28 sets of six loci and the full data set was 0.75 (SD = 0.14; median R2 = 0.76). These results suggest that a general trend of genetic divergence is detectable in our data set, and that this trend is conveyed by different combinations of microsatellite loci. We may thus expect that this same trend continues to be found when more loci are added, and therefore that our results of subsequent landscape genetic analyses are relatively robust, even with the data set presented here.

F_{ST} versus allelic richness

The level of divergence between populations is susceptible to differences in alpha diversity (e.g. allelic richness) within populations. Lower alpha diversity may increase divergence, in particular when divergence is a result from genetic drift after for instance a bottleneck event. Such a scenario would result in spurious correlations between F_{ST} and environmental variables if bottleneck or founder events systematically occur in a certain type of environment. To assess whether GDM correlations between F_{ST} values and environmental variables could be attributed to differences in alpha diversity within populations, we used the following two approaches. We calculated overall allelic richness (ar) across all loci per sampling site. We then calculated location-pairwise differences in ar and plotted the negative log-transformed values against the corresponding F_{ST} values (Fig. S7). If alpha diversity is affecting beta diversity, we would expect a positive correlation between ar and F_{ST}. However, we did not find a significant correlation between these variables (slope = -0.001, R2 = 0.0013). Even given the results above, it is conceivable that populations in different environments exhibit different levels of allelic richness. We, therefore, also tested for linear correlations between ar and environmental variables using the "lm" function in R, and found no significant linear model ($P > 0.05$ for each variable; overall adjusted R2 = -0.117, F = 0.8255, P = 0.6537). From these results we concluded that correlations between F_{ST} and environmental heterogeneity were not the result of differences in population-wise genetic variation (i.e. alpha diversity; allelic richness).

Range expansion

Demographic processes may influence genetic diversity within and between populations, and as a result could confound tests for selection and IBD. Since the Last Glacial Maximum, house sparrows have likely expanded their range from southern refugia, or even from a refugium within the Carpathian Mountains. Range expansions are expected to result in lowered genetic variation in peripheral populations, in the direction of the expansion. Even though it appeared unlikely that a signal of past range expansion would still be present in these southern European populations after thousands of generations, we visually assessed the geographic pattern of allelic richness (which was similar between populations, ranging from 3.63 – 4.33) by plotting ar per population on a map (Fig. S8). If a signal of range expansion would still be present, we would have expected to see diminishing ar from south to north or away from the Carpathian Mountains. However, we found no such pattern, and concluded that past range expansion and associated effects on intraspecific variation is unlikely to affect our landscape genetic analyses.

Environmental variables - computation of QSCAT

A measure of surface moisture and canopy roughness (over dense forest) was obtained from the QuikScat microwave instrument (QSCAT; (Long et al., 2001)). We computed multi-year (2000 - 2008) averages of raw backscatter measurements at the horizontal polarization. To do so, daily data records for Europe for the years 2000

- 2008 were downloaded from the BYU Scatterometer Climate Record Pathfinder database (http://www.scp.byu.edu/data/Quikscat/SIRv2/qush/Eur.html).
The downloaded .SIR datafiles were converted to GeoTIFF files using "sir_utils", available from NASA SCP (http://www.scp.byu.edu/docs/geotiff.html). Further processing was then done in ArcGIS 10.2.2 (ESRI, Redlands, USA). We visually inspected daily images for potential anomalies, such as large areas with missing data and geometric patterns that could indicate an error of the sensor. Out of the 365 daily images, a minimum of 0 and maximum of 7 images showed errors and were omitted from further processing. We then computed yearly averages, minima, maxima, and seasonality for each year; we subsequently averaged these variables across the years 2000 - 2008. To focus more on climatology and less on short-term extreme weather events, we did not use the daily images to compute minima and maxima, but first calculate two-week averages, of which we then took the minima and maxima. Finally, to compute QSCAT seasonality, we used the coefficient of variation, analogous to the WorldClim variable Bio 15: precipitation seasonality (www.worldclim.org/bioclim). We first computed monthly averages and then defined seasonality as: 100 × (SD-monthly mean / |average of monthly mean|).

Visualization of biotic variation across the landscape
To visualize biotic variation across our study area, we used the approach from Fitzpatrick and Keller, 2015. In brief, the turnover functions derived from GDMs were used to transform the retained environmental variables into indices of genetic importance. These genetic importance values were subsequently reduced into orthogonal axes by a Principal Components Analysis (PCA), and the first three axes were mapped to the red, green, and blue channels in a RGB composite layer. One caveat of this approach is that not all variation in the genetic importance variables is explained by the first three PC axes.

A more proper way to visualize biotic variation is to use the 'predict.gdm' function in the 'gdm' package with subsequent multidimensional scaling to reduce the n-dimensional GDM matrix into three dimensions that can be mapped to RGB values. However, in the current set-up (R package 'gdm' version 1.1.2, run on a stand-alone PC with 8 Gb of RAM), it was only possible to do so for a maximum of 3000 locations. To evaluate whether the high resolution spatial projections from the approach from (Fitzpatrick and Keller, 2015) were broadly consistent with those at low resolution using the 'predict.gdm' function, we created maps using both approaches and visually compared the results. We extracted the values of environmental variables at 3000 locations randomly distributed across the study area. We then predicted the biotic response between those 3000 locations based on the GDM turnover functions, and subsequently used the 'cmdscale' function in R to reduce the 3000-dimensional GDM matrix into three dimensions. We created spatial interpolations for each of these dimensions using Empirical Bayesian Kriging (EBK) in ArcGIS 10.2.2 (ESRI, Redlands, USA). EBK accounts for the error in estimating the semivariogram model by estimat-

ing and using many different semivariograms and attempting to find the optimal parameters for the spatial interpolation. We used EBK with no transformation; a power semivariogram; maximum number of points in each local model: 100; local model area overlap factor: 1; number of simulated semivariograms: 100; standard circular search neighborhood; maximum 15 neighbors. Resulting maps were visually very similar to those obtained through ordinary kriging with a spherical semivariogram, variable search radius, and twelve points in the search radius. Maps for each of the three dimensions were combined into a composite RGB map. The spatial projections of biotic variation were highly consistent between methods (results not shown). As a result, we only show the high resolution maps obtained through a transformation of environmental variables into genetic importance values and a subsequent PCA.

Conservation of intraspecific variation

We visually compared the overlap of protected areas in Romania and Bulgaria with the full range of environmentally associated intraspecific variation in house sparrows. This evaluation is by no means meant to be a detailed analysis of the protection status of intraspecific variation, but can serve as a preliminary and qualitative assessment. The current study will contribute to a larger forthcoming comparative study of multiple species, integrating intraspecific variation with other measures of biodiversity. We plotted protected areas recorded in the Natura 2000 database (http://natura2000.eea.europa.eu/), updated till 2014, on the maps of GDM results (Fig. S6). We did not distinguish between the protection status of sites, or whether protection had been implemented. For both genetic and morphological variation incomplete coverage of the full range of variation (all colors in the maps) can be observed. Insufficient protection of genetic variation and variation in morphology in females (PC2) is suggested in the lowland areas bordering the Danube River, where protected areas only to a small extent overlap with the blue-black colors (microsatellites; Fig. S6a) and the green and brown colors (female PC2; Fig. S6b). In addition, female size and the first shape component in males change rapidly along the elevation gradient in the southern Carpathian Mountains, but hardly any of these areas are under protection (Fig. S6c).

Supplementary Figures

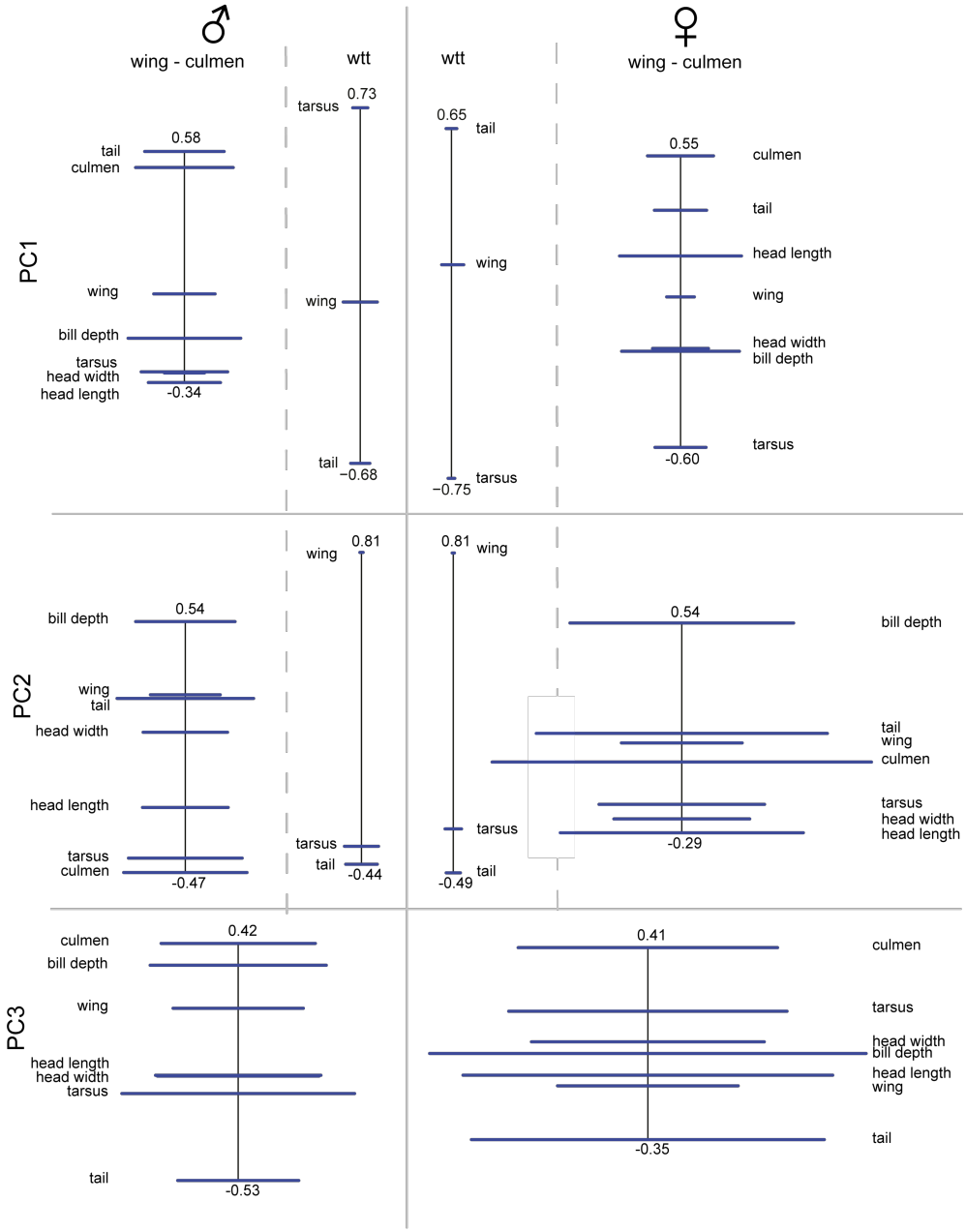

FIGURE S1: PCA ratio spectra for the shape component of males and females for a dataset including all locations but only three morphological variables (wing, tail, and tarsus length), or for only three locations but all morphological variables.

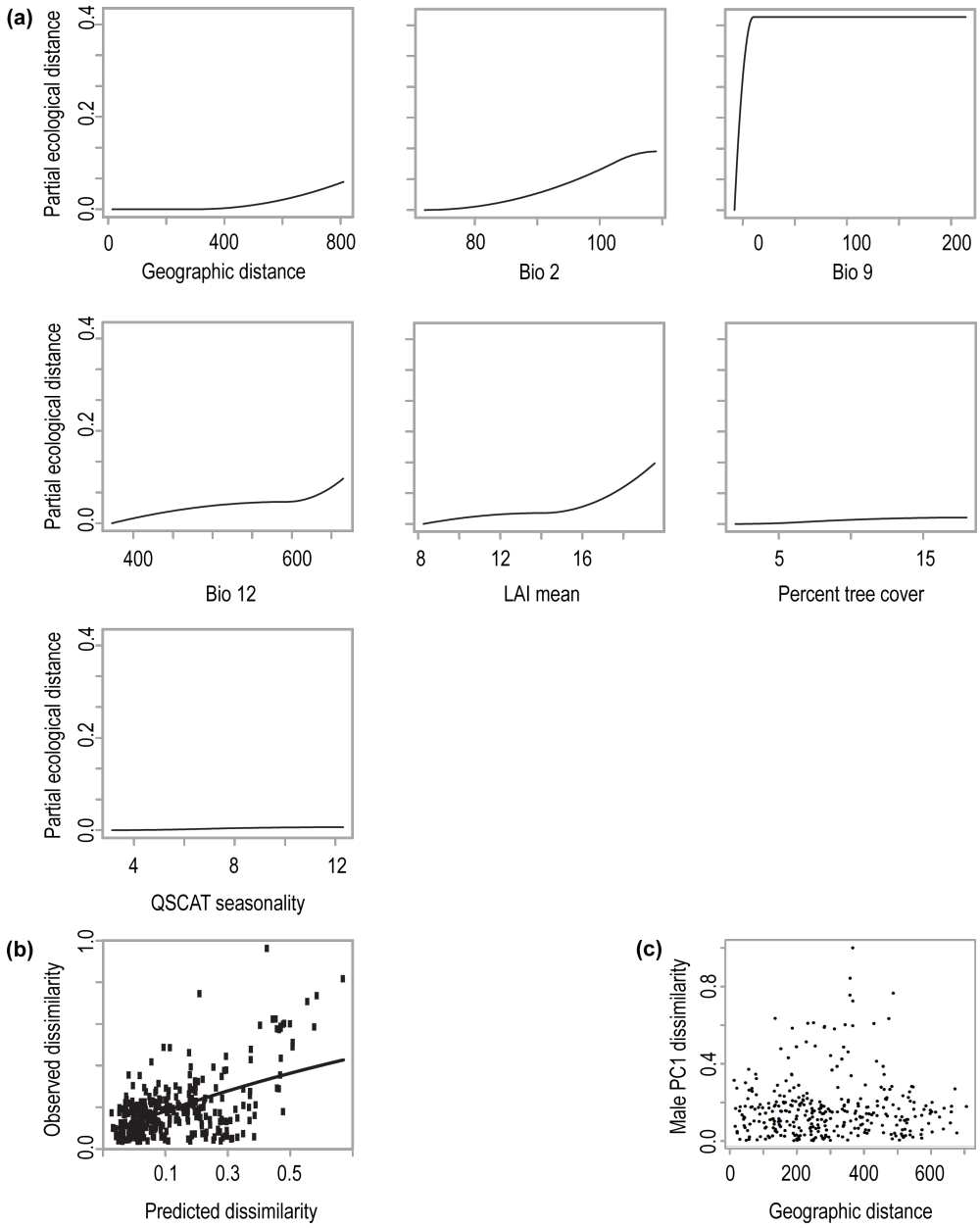

FIGURE S2: (A) Generalized dissimilarity model response curves for the best fit models for the first principal shape component of morphological variables for males (males PC1). The maximum height reached by a variable represents the relative importance of that variable compared to the others. The slope of the curves is indicative of the turnover rate of the response variable as a function of the turnover in the predictor variable. (B) Observed versus GDM-predicted dissimilarities based on a model that included all of the variables from panel (A). (C) Observed dissimilarities as a function of geographic distance.

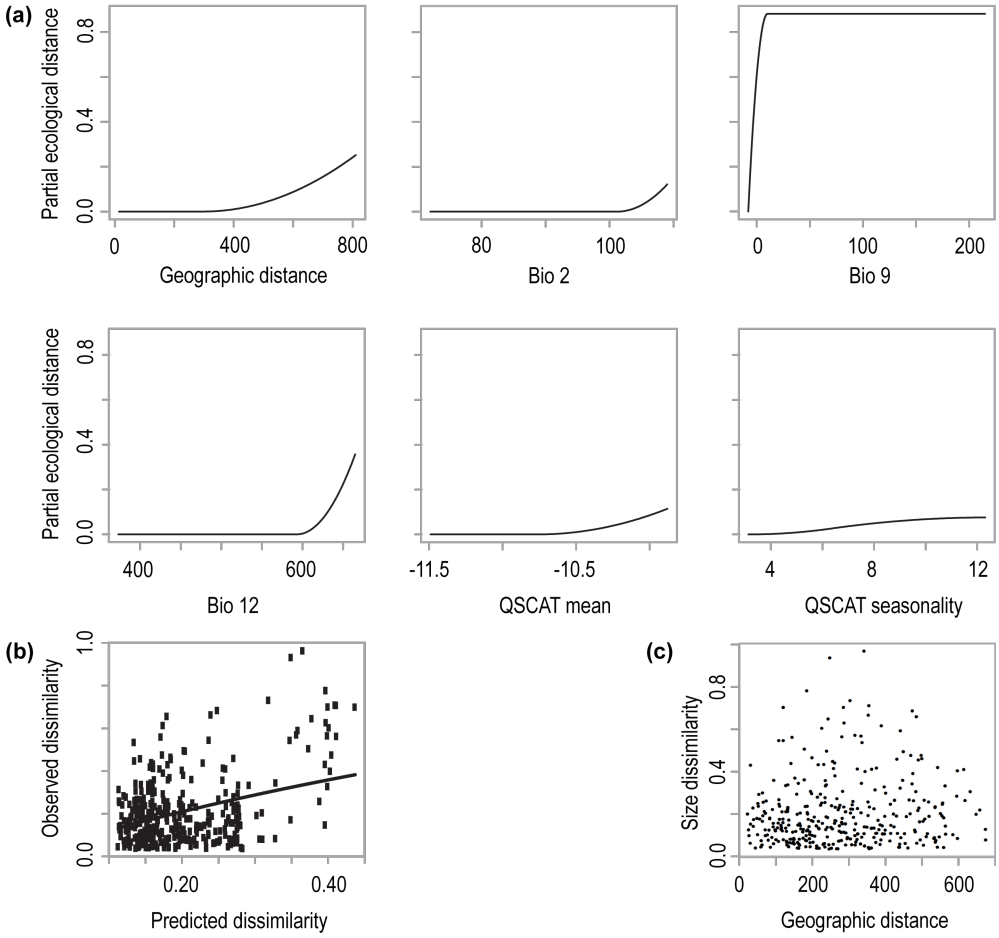

FIGURE S3: S3 Same as Fig. S2, but for the morphological size component for females.

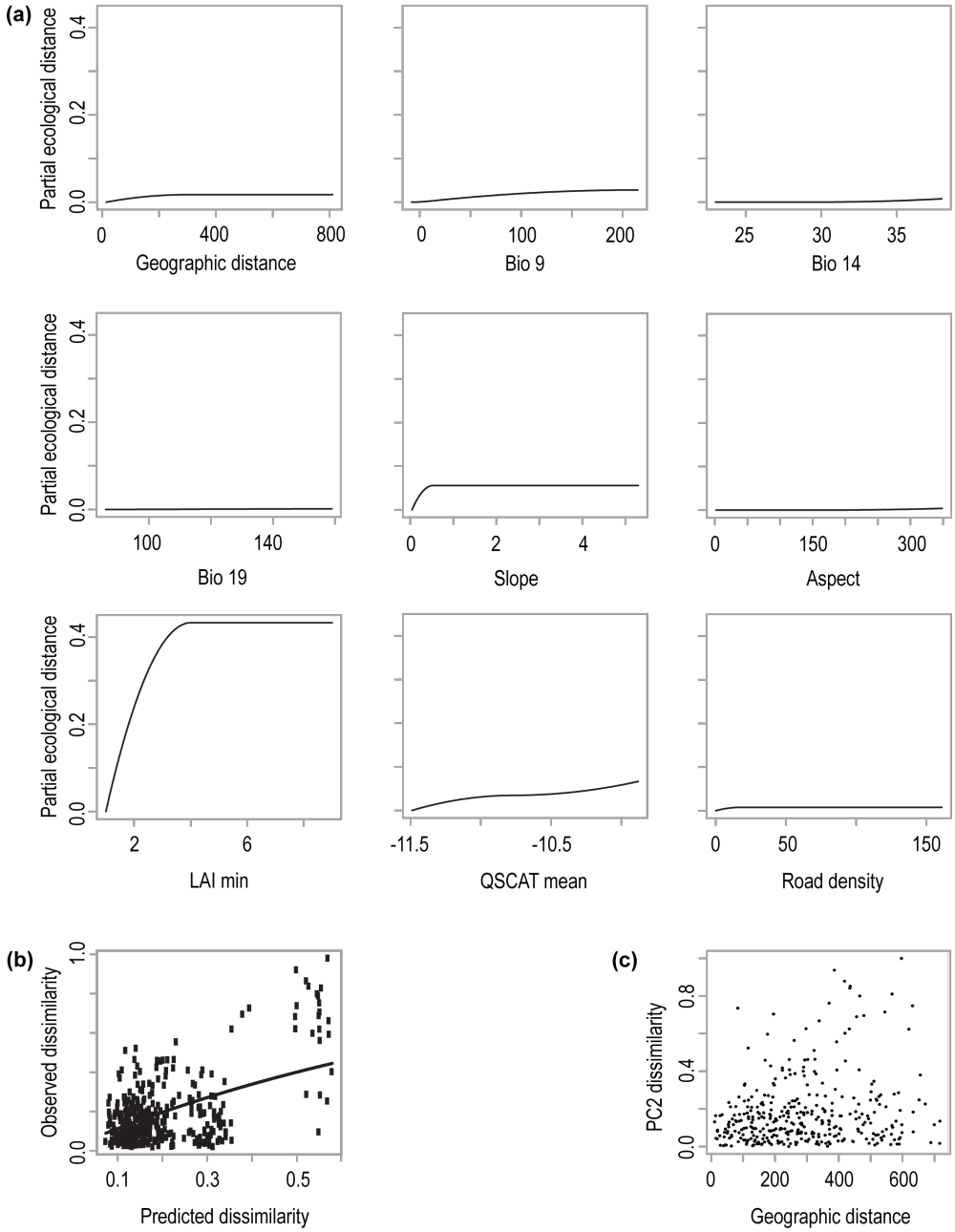

FIGURE S4: Same as Fig. S2, but for the second principal shape component for females (females PC2).

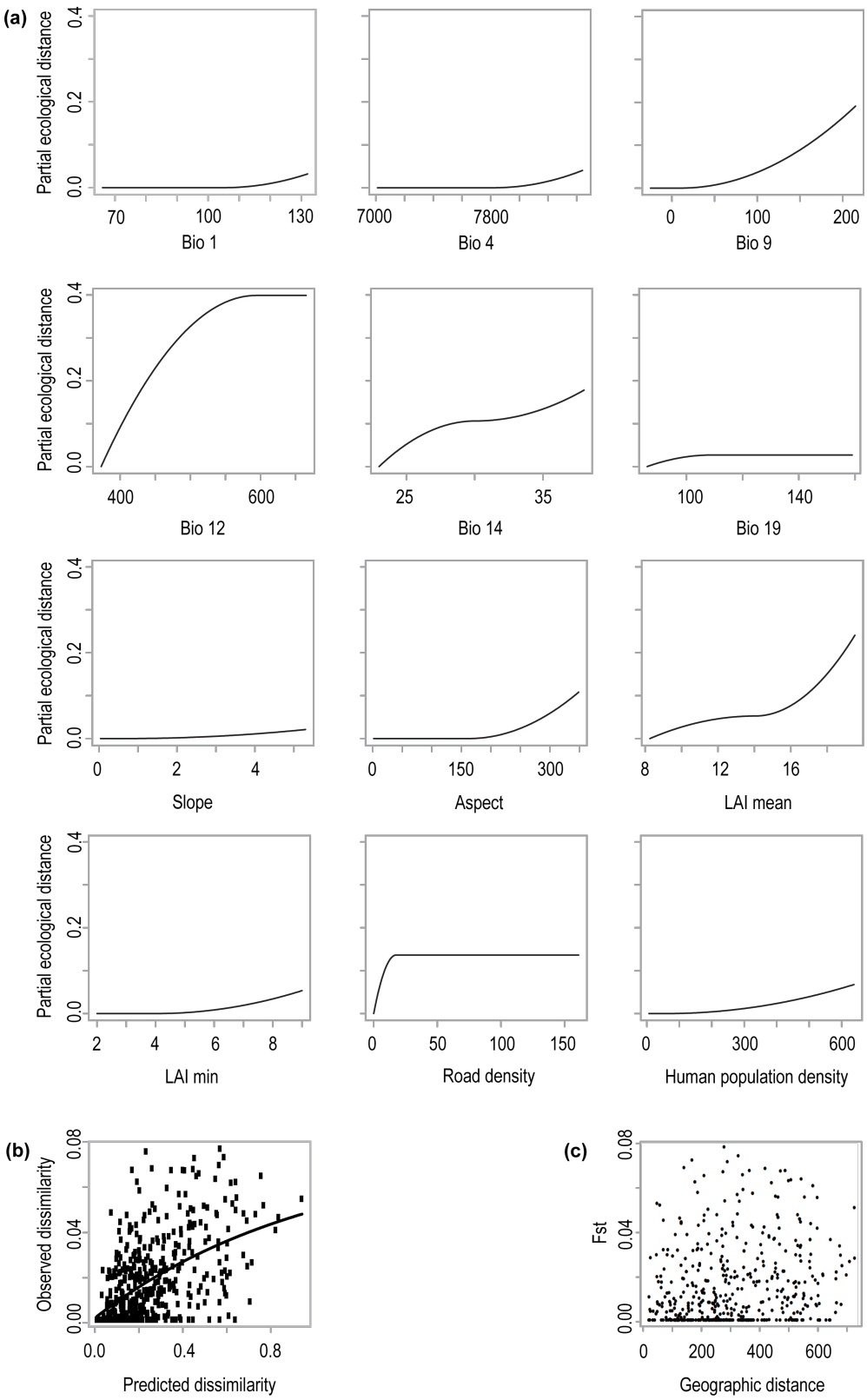

FIGURE S5: Same as Fig. S2, but for microsatellite variation.

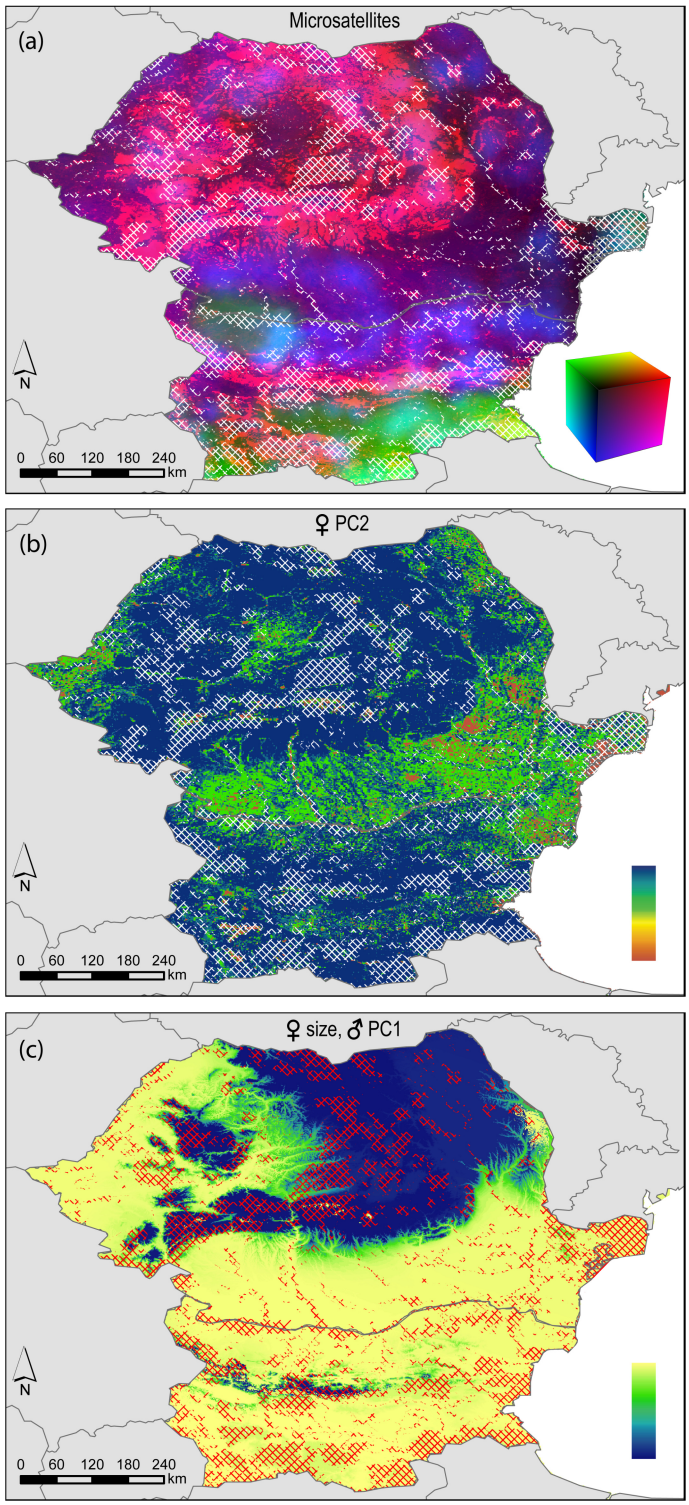

FIGURE S6: GDM results with protected areas (Natura 2000 sites as identified or implemented by 2014) indicated in cross-hatching. (a) microsatellites; (b) the second shape component (PC2) for females; (c) the size component for females and the first shape component (PC1) for males. Color differences between two locations indicated the magnitude of divergence in genetic or morphological variation on the RGB color cube (a), or along the corresponding color bar (b and c).

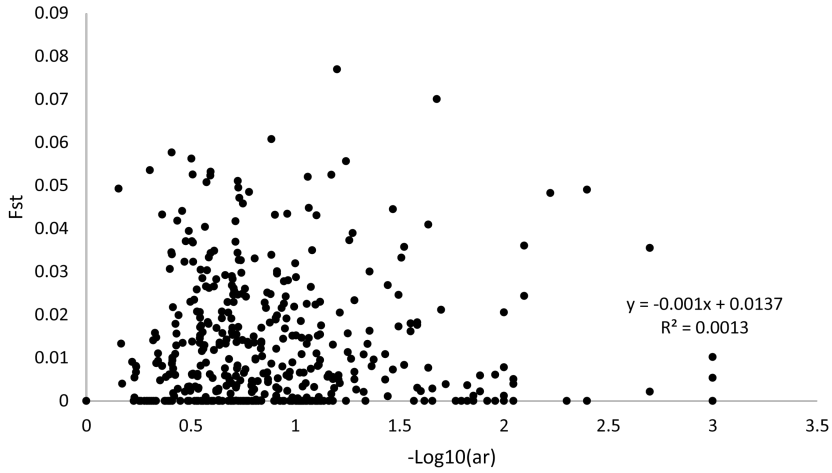

FIGURE S7: Scatterplot of FST values as a function of log-transformed differences in allelic richness (ar).

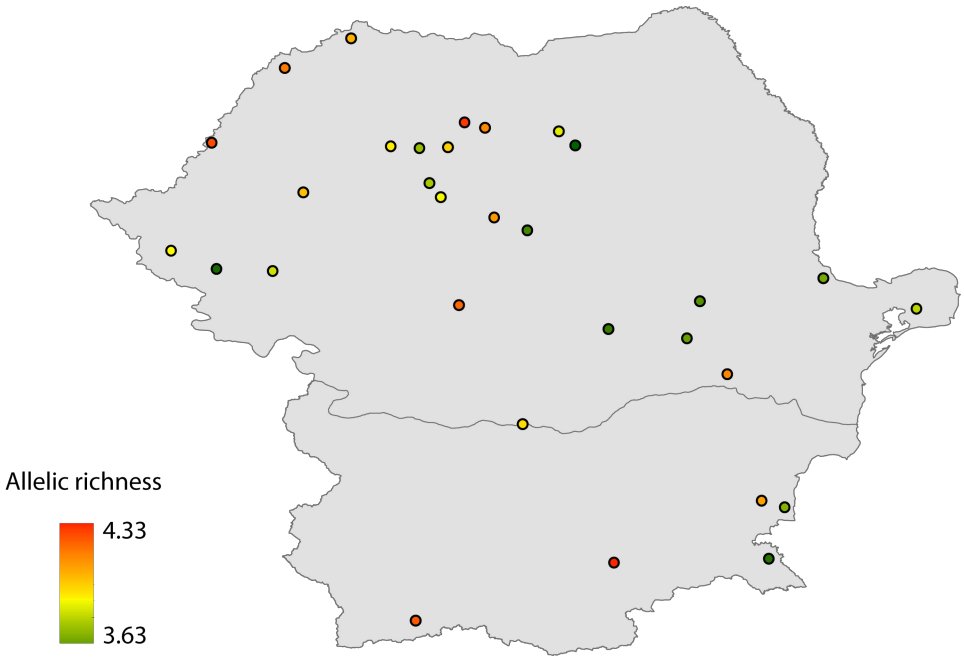

FIGURE S8: Map of allelic richness per population.

Supplementary Tables

TABLE S1: Overview of sampling locations including coordinates and number of sampled individuals. Locations 1 to 5 are situated in Bulgaria and locations 6 to 33 in Romania. Locations 24 and 25, as well as 29 and 30 (bold) were pooled together because of small sampling size and/or they were close together.

				No. of Individuals			
	Location	Latitude	Longitude	Total	Males	Females	Unknown
1	Ognyanovo	41.61487	23.79100	7	7	0	0
2	Beli bryag	42.24340	25.94131	3	3	0	0
3	Jasna poljana	42.28058	27.61904	6	0	6	0
4	Popovits	42.83530	27.78912	8	1	2	5
5	Golica	42.91410	27.54206	16	16	0	0
6	Poiana	43.73681	24.95477	3	3	0	0
7	Berzovia	45.42264	21.62602	3	3	0	0
8	Cluj Napoca	46.74912	23.52103	31	15	15	1
9	Cojocna	46.73354	23.82930	33	12	20	1
10	Vama Seăca	46.35025	23.93556	40	15	15	10
11	Bălcaciu	46.19968	24.05896	34	15	15	4
12	Agnita	45.98115	24.64165	33	15	15	3
13	Făgăraș	45.84098	25.00389	31	15	15	1
14	Independența	44.28193	27.16513	30	15	15	0
15	Văcăreni	45.31745	28.20522	40	15	15	10
16	Broșteni	44.66979	26.73272	30	15	15	0
17	Cornești	44.76584	25.88395	30	16	14	0
18	Balda	46.73661	24.14184	29	15	14	0
19	Pinticu	46.95053	24.54299	33	17	16	0
20	Lechința	47.01407	24.32037	30	15	15	0
21	Turulung	47.92441	23.08823	32	16	16	0
22	Petrești	47.59987	22.37137	31	15	16	0
23	Salonta	46.78522	21.58472	25	11	10	4
24	**Parța1**	**45.61900**	**21.14103**	**11**	**1**	**9**	**1**
25	**Parța2**	**45.64539**	**21.13206**	**19**	**4**	**7**	**8**
26	Caransebeș	45.40022	22.24125	28	14	10	4
27	Mihăești	45.02586	24.25767	31	18	13	0
28	Hălmagiu	46.25341	22.56737	14	0	6	8
29	**Măgheruș**	**46.91388**	**25.33898**	**3**	**1**	**2**	**0**
30	**Runc**	**46.89917**	**25.44515**	**2**	**0**	**2**	**0**
31	Lăzarea	46.75690	25.52250	3	2	1	0
32	Tintești	45.06982	26.86985	15	0	0	15
33	Dunavățu de Jos	44.98804	29.21840	7	4	3	0
	total			**691**	**314**	**302**	**75**

TABLE S2: Microsatellite loci used, with repeat unit (motif), primers, range of allele sizes, corresponding annealing temperatures, references, and the multiplex primer mix it was used it. Loci highlighted in bold were used in the final analyses.

Primer	Motif	Forward/reverse primers	Allele size (bp)	PCR-Temp	Reference	Primer combination
Pdo31	CTCA	GATCCACAGACGCAGACACAG CATGCTGAATACTTTGTGAACTTGC	232-279	55 °C	Dawson et al., 2012	PM 1
Pdo75	**GAAA**	**GCATGACCTACAAACAGTTGC TCCACCTATCTGATTCTGTCAAG**	**94-149**	55 °C	**Dawson et al., 2012**	**PM 1**
Pdoμ3	**CCAT**	**CTGTTCATTAACTCACAGGT AGTGAAACTTTAATCAGTTG**	**140**	55 °C	**Griffith1999**	**PM 4**
PdoA06	**GT**	**GGCTTGAAGGACAGTGTATG TTTCAAAAGGCACAGGTCT**	**103**	56 °C	**Garnier et al., 2009**	**PM 3**
PdoA08	**TG**	**AGCTTTTCAGGTCTCCTTCT CTACACCAGCAAGATCCATT**	**189**	56 °C	**Garnier et al., 2009**	**PM 3**
PdoH05	AC	CAAAGAATTTAAGGGGTGAA ATGAACAACTCTCCAGCATC	151	56 °C	Garnier et al., 2009	PM 3
Pdo7	TTTC	AAATGCAAATAAATGTGCGG GGCAAAGCCTTCCTTATCTC	162-285	60 °C	Griffith et al., 2007	PM 5
Pdo10	**CA**	**AATGTGAATCCCTCCAGAAAC ATGGAGTTTGGGGAATGG**	**113-147**	60 °C	**Griffith et al., 2007**	**PM 6**
Pdo46	**CA**	**GTGGGTGTGCCTGAAGATGTG AGCGGGTCAGGAGCCTCTC**	**191-211**	60 °C	**Dawson et al., 2012**	**PM 6**
Pdo16	**CA**	**GTGTATATGCAAATGACAAGACCAAAGC TCACGCTGACCTAGATGCTATCAGAG**	**282-297**	60 °C	**Dawson et al., 2012**	**PM 2**
Pdo36	**GT**	**GCATTCAAAAATGGCAAGAGGA GAGGCTACCCCTTTCCTGAACA**	**183-221**	60 °C	**Dawson et al., 2012**	**PM 2**
PdoF05	TG	GCATATTTCTGGCATTCTTC TCAAATAAAGTGCTCCACAA	103	60 °C	Garnier et al., 2009	PM 2

TABLE S3: Environmental variables obtained for this study. Variables highlighted in bold were included in generalized dissimilarity models, after highly cross-correlated ones (with Pearson correlation coefficients > 0.7) were omitted.

Name	Attribute	Source
Bio 1	**Annual mean temperature**	**WorldClim**
Bio 2	**Mean diurnal range [mean of monthly (max temp – min temp)]**	**WorldClim**
Bio 3	Isothermality [(Bio2/Bio7) * 100]	WorldClim
Bio 4	**Temperature seasonality (standard deviation*100)**	**WorldClim**
Bio 5	Maximum temperature of the warmest month	WorldClim
Bio 6	Minimum temperature of the coldest month	WorldClim
Bio 7	Temperature annual range (Bio5-Bio6)	WorldClim
Bio 8	Mean temperature of the wettest quarter	WorldClim
Bio 9	**Mean temperature of the driest quarter**	**WorldClim**
Bio 10	Mean temperature of the warmest quarter	WorldClim
Bio 11	Mean temperature of the coldest quarter	WorldClim
Bio 12	**Annual precipitation**	**WorldClim**
Bio 13	Precipitation of the wettest month	WorldClim
Bio 14	**Precipitation of the driest month**	**WorldClim**
Bio 15	Precipitation seasonality (coefficient of variation)	WorldClim
Bio 16	Precipitation of the wettest quarter	WorldClim
Bio 17	Precipitation of the driest quarter	WorldClim
Bio 18	Precipitation of the warmest quarter	WorldClim
Bio 19	**Precipitation of the coldest quarter**	**WorldClim**
ELV	Elevation	SRTM
SLOPE	**Slope**	**SRTM**
ASPECT	**Aspect**	**SRTM**
TREE2001	**Percent tree cover**	**Global Land Cover facility**
LAI sd	Lead Area Index standard deviation	Global Land Cover facility
LAI min	**Leaf Area Index minimum**	**Global Land Cover facility**
LAI mean	**Leaf Area Index mean**	**Global Land Cover facility**
LAI max	Leaf Area Index maximum	Global Land Cover facility
QSCAT min	Quickscat minimum	NASA SCP
QSCAT mean	**Quickscat mean**	**NASA SCP**
QSCAT max	**Quickscat maximum**	**NASA SCP**
QSCAT seasonality	**Quickscat seasonality (coefficient of variation)**	**NASA SCP**
Road density	**Road density**	**Digital Chart of the World**
Danube barrier	**Danube as barrier**	
Human pop dens	**Human population density**	**Gridded Population of the World**

Table S4 Pearson correlation coefficients for environmental variables obtained for this study.
Pearson r below the diagonal; p-values above the diagonal.

Pearson r	Bio 1	Bio 2	Bio 3	Bio 4	Bio 6	Bio 7	Bio 8	Bio 9	Bio 12	Bio 13	Bio 14	Bio 15	Bio 16	Bio 17	Bio 18
Bio 1	1	0.0089	0.0267	0.3910	0.0000	0.0333	0.2559	0.0000	0.0012	0.0000	0.1007	0.0000	0.0000	0.0034	0.0000
Bio 2	-0.45	1	0.0000	0.0169	0.0001	0.0000	0.1387	0.0372	0.0013	0.0001	0.7961	0.0004	0.0001	0.2687	0.0000
Bio 3	-0.39	0.87	1	0.7313	0.0089	0.0075	0.6719	0.6805	0.0000	0.0001	0.7795	0.0018	0.0000	0.7825	0.0008
Bio 4	-0.15	0.41	-0.06	1	0.0231	0.0000	0.0000	0.0005	0.1969	0.8913	0.9465	0.3491	0.8566	0.0492	0.1654
Bio 6	0.94	-0.63	-0.45	-0.39	1	0.0001	0.0381	0.0000	0.0032	0.0000	0.1162	0.0000	0.0000	0.0013	0.0000
Bio 7	-0.37	0.83	0.46	0.83	-0.64	1	0.0009	0.0010	0.2855	0.0335	0.9897	0.0099	0.0211	0.0643	0.0015
Bio 8	-0.20	0.26	-0.08	0.72	-0.36	0.55	1	0.0000	0.7185	0.1730	0.4276	0.1759	0.2478	0.1248	0.0178
Bio 9	0.69	-0.36	-0.07	-0.57	0.75	-0.55	-0.69	1	0.1720	0.0017	0.7810	0.0005	0.0013	0.0380	0.0000
Bio 12	-0.54	0.54	0.67	-0.23	-0.50	0.19	-0.07	-0.24	1	0.0000	0.0220	0.0024	0.0000	0.0550	0.0000
Bio 13	-0.77	0.61	0.63	0.02	-0.76	0.37	0.24	-0.53	0.88	1	0.4609	0.0000	0.0000	0.7201	0.0000
Bio 14	0.29	0.05	0.05	0.01	0.28	0.00	0.14	-0.05	0.40	0.13	1	0.0156	0.9403	0.0000	0.8948
Bio 15	-0.89	0.58	0.52	0.17	-0.89	0.44	0.24	-0.57	0.51	0.81	-0.42	1	0.0000	0.0002	0.0000
Bio 16	-0.85	0.64	0.65	0.03	-0.83	0.40	0.21	-0.54	0.84	0.98	-0.01	0.89	1	0.2651	0.0000
Bio 17	0.50	-0.20	-0.05	-0.35	0.54	-0.33	-0.27	0.36	0.34	-0.06	0.87	-0.61	-0.20	1	0.0752
Bio 18	-0.91	0.65	0.56	0.25	-0.92	0.53	0.41	-0.74	0.71	0.92	-0.02	0.88	0.95	-0.31	1
Bio 19	0.63	-0.30	-0.03	-0.58	0.70	-0.51	-0.56	0.69	0.18	-0.26	0.57	-0.69	-0.36	0.87	-0.54
Elevation	-0.85	0.33	0.47	-0.23	-0.70	0.06	-0.27	-0.30	0.52	0.63	-0.35	0.75	0.71	-0.38	0.66
Slope	-0.37	-0.10	0.09	-0.45	-0.21	-0.31	-0.50	-0.04	0.16	0.18	-0.33	0.34	0.24	-0.20	0.16
Aspect	0.05	0.11	0.24	-0.18	0.07	-0.04	-0.03	0.06	0.35	0.16	0.42	-0.16	0.09	0.40	0.08
LAI mean	-0.25	0.05	0.31	-0.59	-0.13	-0.32	-0.16	0.03	0.50	0.48	-0.04	0.34	0.46	0.07	0.32
LAI sd	-0.55	0.19	0.32	-0.36	-0.45	-0.08	0.01	-0.28	0.60	0.65	-0.04	0.55	0.66	-0.06	0.61
LAI min	0.38	-0.23	0.07	-0.65	0.47	-0.52	-0.40	0.59	0.14	-0.05	0.08	-0.21	-0.09	0.36	-0.31
LAI max	-0.31	0.04	0.24	-0.51	-0.20	-0.26	-0.13	-0.03	0.52	0.50	-0.02	0.37	0.49	0.08	0.38
Percent tree cover	-0.53	0.23	0.36	-0.18	-0.41	0.01	0.05	-0.26	0.45	0.54	-0.06	0.51	0.56	-0.15	0.54
QSCAT mean	0.46	-0.25	-0.27	-0.06	0.45	-0.21	0.06	0.13	-0.25	-0.34	0.24	-0.44	-0.39	0.24	-0.35
QSCAT min	0.34	-0.25	-0.13	-0.33	0.40	-0.37	-0.27	0.30	-0.05	-0.15	0.16	-0.27	-0.20	0.26	-0.30
QSCAT max	0.31	-0.42	-0.58	0.25	0.28	-0.11	0.34	-0.10	-0.62	-0.55	-0.08	-0.40	-0.54	-0.16	-0.35
QSCAT seasonality	0.03	-0.13	-0.39	0.52	-0.08	0.25	0.50	-0.30	-0.52	-0.36	-0.25	-0.11	-0.31	-0.40	-0.08
Road density	-0.13	0.14	0.13	0.12	-0.10	0.08	0.17	-0.15	-0.02	0.09	-0.07	0.22	0.13	-0.19	0.17
Danube barrier*	-1.88	-1.89	5.30	-0.12	-0.54	2.59	0.49	0.20	0.21	-1.15	14.18	6.93	-0.36	-1.79	-1.24
Human pop density	-0.19	0.21	0.18	0.07	-0.18	0.12	0.12	-0.18	0.15	0.25	-0.07	0.30	0.26	-0.15	0.27

* The Danube barrier was coded as 0 vs 1. Figures provided are coefficient estimates and p-values of a logistic regression.

Bio 19	Elevation	Slope	Aspect	LAI mean	LAI sd	LAI min	LAI max	Percent tree cover	QSCAT mean	QSCAT min	QSCAT max	QSCAT seasonality	Road density	Danube barrier*	Human pop density
0.0001	0.0000	0.0336	0.7627	0.1546	0.0009	0.0301	0.0775	0.0016	0.0066	0.0544	0.0768	0.8722	0.4676	1.0000	0.2930
0.0870	0.0614	0.5709	0.5557	0.7962	0.2849	0.2004	0.8135	0.1963	0.1644	0.1664	0.0163	0.4784	0.4208	1.0000	0.2465
0.8876	0.0054	0.6328	0.1753	0.0827	0.0672	0.7120	0.1771	0.0425	0.1254	0.4553	0.0004	0.0259	0.4828	1.0000	0.3238
0.0004	0.1927	0.0092	0.3226	0.0003	0.0387	0.0000	0.0026	0.3282	0.7545	0.0573	0.1552	0.0020	0.4965	1.0000	0.7092
0.0000	0.0000	0.2501	0.6929	0.4868	0.0079	0.0062	0.2695	0.0181	0.0087	0.0194	0.1121	0.6610	0.5886	1.0000	0.3228
0.0023	0.7280	0.0836	0.8141	0.0733	0.6392	0.0018	0.1429	0.9714	0.2375	0.0343	0.5524	0.1630	0.6539	1.0000	0.5126
0.0008	0.1234	0.0031	0.8736	0.3803	0.9477	0.0211	0.4617	0.8033	0.7322	0.1333	0.0565	0.0034	0.3334	1.0000	0.5141
0.0000	0.0880	0.8120	0.7553	0.8746	0.1107	0.0003	0.8653	0.1499	0.4781	0.0927	0.5713	0.0904	0.3948	1.0000	0.3177
0.3159	0.0019	0.3655	0.0476	0.0034	0.0002	0.4450	0.0018	0.0080	0.1581	0.7687	0.0001	0.0019	0.8930	1.0000	0.4100
0.1397	0.0001	0.3275	0.3719	0.0048	0.0000	0.7802	0.0033	0.0013	0.0535	0.3970	0.0009	0.0385	0.6011	1.0000	0.1572
0.0005	0.0477	0.0615	0.0152	0.8182	0.8241	0.6659	0.9009	0.7416	0.1779	0.3829	0.6466	0.1653	0.6900	1.0000	0.7103
0.0000	0.0000	0.0511	0.3787	0.0518	0.0008	0.2303	0.0357	0.0027	0.0112	0.1294	0.0204	0.5466	0.2135	1.0000	0.0896
0.0369	0.0000	0.1753	0.6204	0.0070	0.0000	0.6319	0.0037	0.0008	0.0263	0.2708	0.0013	0.0812	0.4843	1.0000	0.1436
0.0000	0.0278	0.2659	0.0204	0.7080	0.7457	0.0390	0.6552	0.3957	0.1779	0.1377	0.3611	0.0228	0.2819	1.0000	0.3965
0.0011	0.0000	0.3859	0.6393	0.0721	0.0002	0.0784	0.0309	0.0013	0.0461	0.0908	0.0469	0.6598	0.3551	1.0000	0.1286
1	0.0442	0.5377	0.0367	0.4755	0.4610	0.0011	0.6503	0.2851	0.1900	0.0788	0.4075	0.0195	0.1882	1.0000	0.2104
-0.35	1	0.0004	0.6865	0.0449	0.0013	0.4260	0.0457	0.0024	0.0143	0.5993	0.0073	0.0474	0.6238	1.0000	0.3641
-0.11	0.58	1	0.8201	0.0809	0.0219	0.6475	0.0869	0.0766	0.3239	0.4991	0.0963	0.0530	0.7058	1.0000	0.8540
0.36	-0.07	-0.04	1	0.8139	0.5355	0.9683	0.5616	0.9738	0.8230	0.4926	0.6701	0.8897	0.0703	1.0000	0.0328
0.13	0.35	0.31	0.04	1	0.0000	0.0000	0.0000	0.1622	0.8420	0.2796	0.0536	0.0129	0.9458	1.0000	0.3426
-0.13	0.54	0.40	0.11	0.81	1	0.1847	0.0000	0.0423	0.9585	0.3404	0.2332	0.0659	0.3731	1.0000	0.1559
0.54	-0.14	0.08	0.01	0.68	0.24	1	0.0002	0.4744	0.7868	0.1480	0.0801	0.0155	0.4592	1.0000	0.7077
0.08	0.35	0.30	0.10	0.94	0.87	0.60	1	0.2793	0.8817	0.4164	0.1103	0.0539	0.9412	1.0000	0.4812
-0.19	0.51	0.31	-0.01	0.25	0.36	-0.13	0.19	1	0.3566	0.8915	0.3898	0.3053	0.1307	1.0000	0.2654
0.23	-0.42	-0.18	-0.04	-0.04	0.01	0.05	-0.03	-0.17	1	0.0000	0.0000	0.8277	0.0434	1.0000	0.0457
0.31	-0.09	0.12	-0.12	0.19	0.17	0.26	0.15	-0.02	0.80	1	0.2071	0.0008	0.0688	1.0000	0.0256
-0.15	-0.46	-0.29	-0.08	-0.34	-0.21	-0.31	-0.28	-0.15	0.66	0.23	1	0.0000	0.3993	1.0000	0.9445
-0.40	-0.35	-0.34	-0.03	-0.43	-0.32	-0.42	-0.34	-0.18	-0.04	-0.55	0.66	1	0.4330	1.0000	0.1154
-0.23	0.09	0.07	-0.32	0.01	0.16	-0.13	0.01	0.27	0.35	0.32	0.15	-0.14	1	1.0000	0.0004
0.17	-0.06	0.76	0.00	0.74	-15.53	-11.13	5.74	-0.57	-29.83	4.23	33.16	-3.79	0.00	1	1.0000
-0.22	0.16	-0.03	-0.37	0.17	0.25	-0.07	0.13	0.20	0.35	0.39	0.01	-0.28	0.58	0.03	1

TABLE S5: PCA results for the shape component of wing, tail, and tarsus lengths in males and females based on the three-variables-all-locations dataset. Numbers indicate factor loadings and proportion of variance explained. The first two axes explained all shape variation. *Prop Var = Proportion of variance explained by each morphological shape component*.

		Shape PC1	**Shape PC2**
Females	Wing length	0.1049	-0.8097
	Tail length	0.6488	0.4957
	Tarsus length	-0.7537	0.3140
	Prop Var	0.8142	0.1858
Males	Wing length	-0.0396	0.8155
	Tail length	-0.6865	-0.4421
	Tarsus length	0.7261	-0.3734
	Prop Var	0.7261	0.2739

TABLE S6: Basic population genetic statistics for each sampling location. N_E = number of effective alleles; H_O = observed heterozygosity; H_E = expected heterozygosity; F = fixation index.

Population		N_E	H_O	H_E	F
Agnita	Mean	6.066	0.785	0.759	-0.044
	SE	0.857	0.087	0.085	0.034
Bălcaciu	Mean	5.094	0.846	0.771	-0.094
	SE	0.563	0.063	0.044	0.044
Balda	Mean	5.574	0.808	0.753	-0.075
	SE	0.713	0.093	0.082	0.042
Beli bryag	Mean	4.049	0.844	0.722	-0.164
	SE	0.482	0.066	0.040	0.056
Berzovia	Mean	3.220	0.708	0.639	-0.099
	SE	0.401	0.098	0.061	0.093
Broșteni	Mean	4.893	0.750	0.740	0.008
	SE	0.708	0.087	0.062	0.084
Caransebeș	Mean	4.890	0.761	0.727	-0.053
	SE	0.728	0.087	0.078	0.042
Cluj Napoca	Mean	5.479	0.808	0.766	-0.067
	SE	0.701	0.061	0.063	0.034
Cojocna	Mean	5.291	0.757	0.725	-0.043
	SE	0.649	0.117	0.104	0.055
Cornești	Mean	5.618	0.725	0.726	0.000
	SE	0.834	0.108	0.105	0.039
Dunavățu de Jos	Mean	4.558	0.625	0.737	0.207
	SE	0.581	0.117	0.052	0.142
Făgăraș	Mean	5.124	0.698	0.732	0.028
	SE	0.772	0.089	0.081	0.072
Golica	Mean	6.887	0.801	0.794	0.000
	SE	0.928	0.079	0.070	0.033
Hălmagiu	Mean	6.477	0.731	0.767	0.034
	SE	0.884	0.090	0.091	0.044

5 CHAPTER IV

SIGNS OF LOCAL ADAPTATIONS IN THE BUFF-TAILED BUMBLE BEE (BOMBUS TERRESTRIS)

Julia C. Geue, Madleina Caduff, Vivian Link, Daniel Wegmann and Henri A. Thomassen

Abstract

Insect pollinators provide one of the most important ecosystem services to humans. With ongoing global change, however, the long-term survival of these species is at stake. Species likely need to adapt to new environmental conditions, but depend on 'standing genetic variation' to do so. Thus, it is critical to map standing genetic variation correlated to the environment in order to understand how species are adapted to their current environment, how they will be affected by future changes, and what populations to protect in order to maximize their adaptive potential. So far, however, only little is known about local adaptations in pollinator species.

Here, we aimed to identify loci showing signs of local adaption in the buff-tailed bumble bee (*Bombus terrestris*) in Romania and Bulgaria, one of the most important pollinator species in Europe. First, we conducted whole genome sequencing of hundreds of individuals to create a large Single Nucleotide Polymorphism (SNP) data set. We then combined the resulting SNP markers with a set of 16 environmental variables in order to detect 'gene-environment' associations (GEAs). Using Latent Factor Mixed Modeling we identified 29 significant associations with nine of the investigated environmental variables. We found that seasonality in temperature and precipitation as well as vegetation variables were significantly associated to SNPs located in coding regions of the genome, suggesting potential local adaptations in *B. terrestris*. Interestingly, we previously found the same environmental variables to be important in determining the spatial distribution of this species.

Introduction

Environmental stressors resulting from ongoing global change through human-induced threats such as habitat alteration, industrial development, and climate change are challenging for many species and have major impacts on biodiversity (Vanbergen and Initiative, 2013). Mounting evidence suggests that one of the most affected taxonomic groups are insects, where particularly pollinator species such as wild bees have declined in recent years (e.g.Hallmann et al., 2017; Potts et al., 2010). Loss of insect diversity and a decrease in pollination activity may thus have adverse effects on global food security, floral biodiversity, and indeed entire food-webs and ecosystems (reviewed by Potts et al., 2010).

One of the biggest threats to biodiversity is habitat loss (Pereira et al., 2012), being responsible for reductions in species richness and abundance of wild bees (Goulson and Sparrow, 2008; Kennedy et al., 2013). In addition, anthropogenically-induced climate change is assumed to affect the availability of important resources for wild bees (e.g. nest sites, flower abundance), changing their distributional patterns and population sizes (Kerr et al., 2015; Pyke et al., 2016). The resulting new environmental conditions will modify natural selection pressures (Hoffmann and Sgrò, 2011), forcing populations to either shift their distributional range or respond plastically or evolutionary and adapt to those novel conditions (Anderson et al., 2012). Phenotypic plasticity is considered a 'short-term' response crucial for the immediate future (Oostra et al., 2018), but may not be sufficient for the expected magnitude of changes necessary in the more distant future (Anderson et al., 2012; Bijlsma and Loeschcke, 2012; Hendry et al., 2008). Evolutionary responses, however, help species to persist and survive over the long run (Anderson et al., 2012; Hoffmann and Sgrò, 2011). An emerging issue in a rapidly changing environment is that mutation rates are likely too low for new advantageous genotypes to arise (Barrett and Schluter, 2008; Savolainen et al., 2013), and adaptive responses are contingent upon standing genetic variation (Etterson and Shaw, 2001). It is thus crucial to protect as much standing genetic variation as possible in order to maximize a species' evolutionary potential. A practical strategy may be to map this standing genetic variation (in particular that part of genetic variation that represents adaptations to different environmental conditions). Doing so could enhance our understanding of the (underlying) processes of local adaptation and the role of genetic diversity herein, and eventually will help us to disentangle the relative influence of human-induced threats (De Mazancourt et al., 2008), and to predict species' future responses to these threats (Bay et al., 2017; Hoffmann and Sgrò, 2011).

Here we aimed to explore patterns of local adaptation to diverse environmental conditions in the buff-tailed bumble bee (*Bombus terrestris*). *B. terrestris* is a widespread and important pollinator species of crops and wild plants, experiencing diverse environmental conditions, which in return pose different selection pressures (Goulson,

2010; Rasmont et al., 2013). Such widespread species are expected to show a distinct genetic composition between populations as a result of local adaptations (Vincent et al., 2013). The buff-tailed bumble bee is well studied with respect to its behavior (e.g. Kraus et al., 2009; Walther-Hellwig and Frankl, 2000; Woodard et al., 2015), distribution (e.g Chapter I (Geue and Thomassen, 2020), Polce et al., 2018; Rasmont et al., 2015a), and habitat preference (e.g. Carvell, 2002; Scriven et al., 2015; Svensson et al., 2000). As is the case for other important pollinator species,*B. terrestris* may also be affected by human induced habitat alterations and future climatic change (Brown and Paxton, 2009; Kerr et al., 2015; Kosior et al., 2007; Rasmont et al., 2015a). However, only little attention has been paid to the genetic consequences of threats imposed by humans. Because *B. terrestris* is not threatened yet, its genetics and adaptive potential have so far been neglected. Only in recent years, studies investigated signatures of local adaptation to environmental conditions in other pollinator species, including the tropical bee *Melipona subnitida* (Jaffé et al., 2019), the honey bee *Apis mellifera* (Henriques et al., 2018), two alpine bumble bees *Bombus balteatus* and *B. sylvicola* (Miller-Struttmann et al., 2015), and the red-tailed bumble bee *Bombus lapidarius* (Theodorou et al., 2018). Both the tropical bee and the honey bee showed signs of local adaptations to climatic conditions, in particular precipitation (Henriques et al., 2018; Jaffé et al., 2019). The two bumble bee studies demonstrated the ability of these species to adapt to environmental change in both urban and alpine environments (Miller-Struttmann et al., 2015; Theodorou et al., 2018).

We implemented a landscape genomics approach to assess genomic adaptations to the environment and elucidate the ecological importance of loci putatively under selection ('selective loci' from here onwards). Landscape genomics is a quickly developing field, combining genomic data with environmental information, thus facilitating the mapping of genetic diversity, and providing a way to identify adaptations to local environmental conditions (Balkenhol et al., 2017; Guillot et al., 2005a; Holderegger et al., 2006; Manel et al., 2010; Manel et al., 2003; Storfer et al., 2018). Technological advances, such as high throughput sequencing and new analytical tools aid in these attempts (Henriques et al., 2018; Storfer et al., 2018). So called gene-environment associations (GEA) scan the genome for loci (such as single-nucleotide polymorphisms, SNPs) with allele frequencies correlated to environmental features, indicating potential selection pressures driving local adaptation providing an advantage in some environments (Frichot et al., 2013; Joost et al., 2007). Linking molecular markers with the corresponding functional genes can enhance our knowledge about key processes influenced by environmental selection (Bonin, 2008; Parisod and Holderegger, 2012). A widely-used method for this kind of analysis are latent factor mixed models (LFMM; Frichot et al., 2013). LFMM accounts for population structure (in the form of latent factors), limiting potential false positives in the detected genotype environment associations. LFMM was shown to be a good compromise between error rates and the detection power of important loci (Rellstab et al., 2015).

In this study, we specifically aimed to identify signals of local adaptations to different environmental conditions in the buff-tailed bumble bee (*Bombus terrestris*) by combining a large genomic dataset with a set of ecologically meaningful environmental variables. We whole-genome sequenced (WGS) hundreds of individuals from 25 locations at low coverage to identify single-nucleotide polymorphism (SNPs) across the genome, and correlated these to a set of 16 environmental variables. Based on the documented habitat preferences (Chapter I (Geue and Thomassen, 2020), Svensson et al., 2000) and its seasonal-dependent life-cycle (Goulson, 2010), we expected to find signatures of selection related to seasonality in temperature and precipitation as well as to vegetation parameters. We also traced back selective loci to the genome, and identified genes that may be relevant relevant for the adaptability of *B. terrestris* to its habitat.

Material and Methods

Study species and study area

The buff-tailed bumble bee, *Bombus terrestris* is one of the most common bumble bee species in Europe and an important pollinator species (Corbet et al., 1991; Murray et al., 2007). It is considered to be a generalist species showing highly polylectic feeding behavior (collecting pollen from flowers of a variety of unrelated plants) (Rasmont et al., 2013; Westphal et al., 2009) and occurring in a wide geographical range comprising different habitats (Goulson, 2010; Rasmont et al., 2013). *B. terrestris* is currently not threatened within its native range, however with ongoing human-related reduction of suitable habitat and climate change it was predicted to be highly affected and hence could experience population declines (Rasmont et al., 2015a).

This study was conducted in Bulgaria and Romania, two southeastern European countries, covering an area of 350.000 km2. Both countries are shaped mainly by mountain ranges (the Rila, Rhodope and Balkan Mountains in Bulgaria, the Carpathian mountains in Romania) and the Danube River representing the border between both countries. Bulgaria and Romania are environmentally highly heterogeneous, comprising different climatic zones (continental, mediterranean and temperate) resulting in a wide variety of habitats. These habitats include areas inhabited and influenced by humans, as well as natural areas such as mountains, river valleys, forests, open woodlands, and grasslands. The resulting habitat heterogeneity is ideal to assess the influence of the environment on biological diversity, such as genetic diversity and offer the opportunity to look for signs of local adaptations.

Sampling, WGS library preparation and sequencing

We used 411 *Bombus terrestris* samples from 25 locations in Romania and Bulgaria (Table 1), which are part of the dataset in *Chapter I*. DNA was extracted as described in *Chapter I* (Geue and Thomassen, 2020). We prepared a WGS library for those 411 samples using the 'Nextera XT DNA library Preparation Kit' (Illumina, FC 131-1096), following the previously established protocol by Baym et al., 2015. The detailed protocol can be found in the Supplementary Information.

TABLE 1: Sampling locations and sample sizes for each location. Some samples had to be excluded because of bad sequencing quality. The location Billed was removed from further analyses because most of the samples showed a bad sequencing quality and a sample size of three was not considered to be representative enough.

	Location	Latitude	Longitude	No. of Individuals sequenced	No. of individuals in downstream analyses
1	Baita Plai	46.46871	22.61674	21	20
2	Billed	45.91412	20.94701	9	-
3	Blandesti	47.71380	26.86323	20	20
4	Brebu	45.42815	21.97966	20	18
5	Carei	47.69646	22.48073	19	18
6	Coastra	45.14758	24.22260	9	8
7	Dobrovat	46.99043	27.65404	20	20
8	Drăgusani	46.29929	26.97973	20	20
9	Föen	45.51085	20.87627	20	14
10	Golitsa	42.90956	27.52514	13	13
11	Gothal	45.40790	21.42069	20	17
12	Hlyabovo	42.06055	26.26459	11	11
13	Kamenitsa	41.64449	23.17299	12	11
14	Koevtsi	43.15832	25.09082	19	19
15	Mengishevo	43.03566	26.64753	13	12
16	Orsova	44.75420	22.39480	14	12
17	Pietroasa	46.58998	22.58807	12	12
18	Poienita	45.82299	24.57591	19	17
19	Rish	42.97442	26.90731	20	20
20	Sinemorets	42.04499	27.95808	14	11
21	Stambolovo	41.78435	25.63166	15	15
22	Strumeshnitsa	41.39833	23.06046	20	16
23	Topa Mica	46.92851	23.40238	19	19
24	Valea Hotarului	47.93870	23.83761	20	19
25	Valea Pădurii	46.62236	24.02727	12	12
	total			**411**	**374**

Briefly, DNA was normalized to 2.5 ng/µl and was then fragmented with a single "tagmentation" enzymatic reaction. With help of a PCR, dual-matched index adapters with Unique Molecular Indices (UMI) (produced by Integrated DNA Technologies, Inc. (IDT)) were ligated to the DNA fragments. We specifically used unique adapters on both the 5′ and 3′ ends to make sure that potential index hopping could be detected in downstream bioinformatics analyses. Index-hopping can occur on sequencing machines when the adapters attached to different DNA fragments are partially overlapping and switch between fragments, resulting in erroneous assignments of reads to individuals after de-multiplexing the pooled samples (Ros-Freixedes et al.,

2018). PCRs were performed with a Mastercycler epgradient (Eppendorf) with the following conditions: initial denaturation steps at 72 °C for 3 minutes and followed by an additional step at 95 °C for 3 minutes; then 13 cycles of a 3-step process: denaturation for 20 seconds at 98 °C, annealing for 15 seconds at 62 °C and extension for 1 minute at 72 °C with a final extension step at 72 °C for 1 minute. A double size selection with magnetic beads (Beckman Coulter™ Agencourt AMPure XP) was then performed to get fragments in the size range of 450-700 bp. Fragment sizes of the final libraries of the individual samples were checked on a Tapestation 4150 (Agilent Technologies) and quantified using a Qubit fluorometer (Waltham, MA) with the High Sensitivity DNA assay. Samples were subsequently, in a randomized order, pooled in five different libraries, containing 25 ng/µl of each sample. Paired-end sequencing (2x 150bp) was performed on an Illumina HiSeq3000 (Max-Planck Institute, Tübingen, Germany).

Bioinformatic analyses and SNP discovery

The raw sequencing files were trimmed using Trimmomatic v0.36 (Bolger et al., 2014) and all adapter sequences and low-quality bases (quality score <20) were removed. Reads shorter than 50 nucleotides after trimming were discarded. The remaining reads were then mapped to a 'reference' genome of Bombus terrestris (NCBI, GenBank assembly accession: GCA_000214255.1) using the Burrows-Wheeler Aligner mem (BWA-MEM) with the default parameters (Li and Durbin, 2010). The resulting SAM files were transformed into BAM files and merged with 'samtools' (Li et al., 2009) so that genome-wide information was available for every individual. Using 'picardtools' (*Picard toolkit* 2018) 'readgroups' were added, which was important because the individual libraries were pooled and run on different lanes of the sequencing machine. Additionally, PCR duplicates were marked, but not removed. In a lot of cases, PCR duplicates are removed because there is concern that they can lead to false positive variant calls. We however decided to just mark them, since they seem to only have a minimal effect on the accuracy of subsequent variant calls (Ebbert et al., 2016).

For the remaining bioinformatics analyses, ATLAS, a tool mainly developed for ancient and low coverage DNA sequencing approaches was used (Link et al., 2017). As a first step, we used ATLAS to merge paired-end reads, so that in the overlapping region only one of the two reads would be considered (task=mergeReads). We randomly selected the read to be considered with the parameter "keepRandomRead". Using ATLAS (task=recal), we estimated base sequencing quality score recalibration parameters for every BAM file. The BAM files were recalibrated with the help of ultra-conserved elements (UCE) in order to remove monomorphic sites (UCE's were based on: Faircloth et al., 2015). Those recalibration parameters were then considered when estimating genotype likelihoods for every BAM file (task=GLF). We inferred

the two most likely alleles for each site and created a VCF (Variant Call File) containing the genotype likelihoods for all individuals using ATLAS (task=majorMinor). For this step, we restricted the output to contain only variant sites with the parameter "minVariantQuality=40" and used the parameter: “skotte”, where genotype frequencies for the best allelic combination with the highest likelihood are selected (Skotte et al., 2012).

For the subsequent landscape genomics analysis we split our genome into the 18 chromosomes and imputed the genotype likelihoods with STITCH (Davies et al., 2016) and created a dataset with posterior genotypes. We ran initial tests to estimate the number of ancestral haplotypes (called ‘K’). To do so, we chose 10 random samples with a higher sequencing depth (5X) and created a gen-file based on these samples to have a ‘reference’ for the certainty of potential genotypes. We tested different values for K ranging from 10-100. Larger values for K allow for more accurate imputation. Whereas it is important to assess the accuracy of the imputation, the feasibility of the procedure is constrained by computational limitations (Davies et al., 2016). Based on the 10 randomly selected samples we determined that K=60 provided the most accurate and still computationally feasible setting for the ancestral haplotypes. The final settings for STITCH where: K=60, 1333 generations and a shuffle bin radius of 1000. We combined the separate Chromosome files (pre-imputation) again by using ‘bcftools’ (Li, 2011) and removed SNPs with a minor allele frequency (MAF) < 5%.

Landscape genomics analyses

To assess the contribution of the environment in structuring genetic variation and to identify selective loci associated to the environment, we performed gene-environment associations (GEAs) with the same set of environmental variables we used to model the distribution of *Bombus terrestris* in *Chapter I* (*Chapter I*, Table 3). In order to identify SNPs, which are highly correlated to our 16 environmental variables, we used Latent Factor Mixed Models (LFMM). LFMM is one of the most commonly used methods for doing GEAs (Ahrens et al., 2018) and correlates genetic markers and environmental variables, while at the same time estimating the hidden factors of population structure and subsequently correcting for these (Frichot et al., 2013). Population structure is modelled via so called latent factors (‘K’). Prior knowledge about population structure, for example through clustering approaches such as STRUCTURE (Pritchard et al., 2010) should be directly implemented as a measure of K (=integer for the number of latent factors used in the regression model) (Frichot et al., 2013). If no prior knowledge about population structure is present, so called Principal Component Analyses (PCAs) can assist in detecting some structure and determining the number of latent factors. Studies on the buff-tailed bumble bee suggest that genetic diversity is predominantly unstructured, with a rather panmictic pattern, indicating high rates of

gene flow (Lecocq et al., 2013; Silva et al., 2020; Woodard et al., 2015). The selection of our sampling locations was guided by the dispersal ability and large foraging ranges of this species (Chapman et al., 2003) and we ensured that locations were located at least 20 km apart (see *Chapter I*, section 2.2). We considered therefore the number of sampling locations as a good estimate for population structure and ran a PCA with 25 principle component axes and determined the number of latent factors K according to the 'knee' point of the curve (Figure 1).

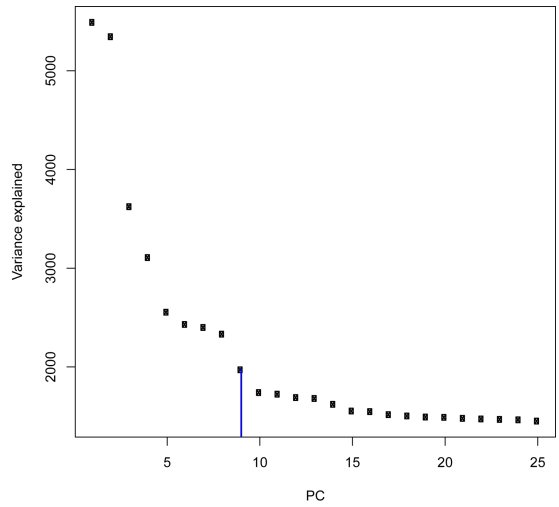

FIGURE 3: Scree plot for the percentage of variance explained by 25 PC-Axes (number of populations in analysis) in order to identify latent factor (K) for the LFMM analysis by using 'prcomp'. Blue line indicates the knee point at K = 9 with nine major genetic clusters in the data.

However, the PCA plot did not only show one clear 'knee' point at K=9, but also one less distinct 'knee' point at K=5. We thus ran the analyses with latent factors ranging between 4 and 11 and conservatively focused on those correlations that were consistent across multiple values for K. Latent factor mixed model analyses were implemented in the R package lfmm (Caye et al., 2019). We decided to run the advanced version LFMM2 (hereafter just LFMM), since this version can handle the posterior genotypes we have in our dataset. We ran LFMM with the following settings: K=4-11, lambda=1e-05, algorithm=analytical, it.max=10 000 and the default setting for the number of burn-ins (5000). Resulting P-values were adjusted for multiple tests using a false discovery rate (FDR; corrected P-value < 0.05 and < 0.01) correction with the Benjamini-Hochberg algorithm (Benjamini and Hochberg, 1995). This threshold implies that 5% or 1% of our associations are expected to be false positive. For each significantly associated environmental variable, SNP associations were plotted in 'Manhattan Plots' and patterns within and between sampling locations were explored in 'scatter plots' in R 3.6.1 (R Development Core Team, 2008). Noteworty patterns were then visualized by creating 'polar coordinated pie chart' with the "ggplot2" package in R and combined with maps produced with QGIS 3.10.6 (QGIS, 2017). These "polar coordinated pie charts" visualize the posterior genotypes of each individual in each location. Every line indicates one individual and the genotypes are shown clockwise: starting with the homozygous (0 at 12 o'clock), the heterozygous (1 at 6 o'clock) and ending with the other homozygous (2 at 12 o'clock again).

Selective loci were mapped back to the genome using the website 'Ensembl' (Hunt et al., 2018). According to their location on the genome, genes around those loci or close by were identified (±100.000bp) and where then blasted on 'NCBI' (Coordinators, 2016) to investigate possible functions.

TABLE 2: LFMM results – showing the significant associations with environmental variables. Markers are consistent between all three K values (K=9 ±1). P- and q-values are the mean between the different LFMM runs (K=9 ±1). Color coding indicates similar SNP marker between the environmental variables. Associations in bold are highly significant with a FDR (False Discovery Rate) of 1%.

Env. variable	SNP Marker	p-value	q-value (5%)	Position in genome	Chromosome
QSCATseason	V841835	2.34E-08	0.0383	NC_015770.1:973808	Chr 9
Bio 8	**V1285063**	**4.40E-09**	**0.0072**	**NC_015773.1:10980707**	**Chr 12**
	V802687	2.83E-08	0.0232	NC_015769.1:5731985	Chr 8
Bio 4	**V466493**	**1.29E-09**	**0.0009**	**NC_015766.1:6120680**	**Chr 5**
	V647421	**1.34E-09**	**0.0009**	**NC_015768.1:5413242**	**Chr 7**
Bio 3	V466493	4.99E-08	0.0412	NC_015766.1:6120680	Chr 5
	V1355085	7.63E-08	0.0412	NC_015774.1:7344328	Chr 13
	V1503835	5.60E-08	0.0412	NC_015776.1:6611877	Chr 15
	V280355	1.11E-07	0.0454	NC_015764.1:7252345	Chr 3
LAIsd	V511914	2.23E-08	0.0365	NC_015767.1:275765	Chr 6
Bio 11	**V466493**	**3.02E-10**	**0.000494**	**NC_015766.1:6120680**	**Chr 5**
	V647421	**7.47E-09**	**0.00611**	**NC_015768.1:5413242**	**Chr 7**
Bio19	**V1463790**	**1.99E-09**	**0.00325**	**NC_015776.1:2064504**	**Chr 15**
	V915653	6.88E-08	0.0427	NC_015770.1:9935769	Chr 9
	V307659	1.88E-07	0.0432	NC_015764.1:9864342	Chr 3
	V676937	1.17E-07	0.0432	NC_015768.1:9628798	Chr 7
	V1516520	1.62E-07	0.0432	NC_015776.1:8010461	Chr 15
	V1463770	1.95E-07	0.0432	NC_015776.1:2062133	Chr 15
	V1334923	2.07E-07	0.0432	NC_015774.1:4329591	Chr 13
	V560136	2.61E-07	0.0436	NC_015767.1:5787495	Chr 6
Tree cover	V1457378	1.31E-08	0.0210	NC_015776.1:1253942	Chr 15
	V69823	5.91E-08	0.0265	NC_015762.1:10211783	Chr 1
	V511914	6.49E-08	0.0265	NC_015767.1:275765	Chr 6
	V1063161	4.56E-08	0.0265	NC_015771.1:13309360	Chr 10
	V1060815	1.24E-07	0.0366	NC_015771.1:13140800	Chr 10
	V788356	1.36E-07	0.0366	NC_015769.1:4062621	Chr 8
	V441690	1.66E-07	0.0388	NC_015766.1:2462459	Chr 5
	V179495	1.99E-07	0.0407	NC_015763.1:7568447	Chr 2
Bio 15	**V466493**	**1.68E-08**	**0.0274**	**NC_015766.1:6120680**	**Chr 5**

Results

Sequence data

Our whole-genome sequencing approach generated approximately 256 million paired-end reads from 411 individuals. 248 million reads (96,8%) passed initial quality filters. The average individual coverage was 4X, when calculated across the entire genome. A total of 34 individuals with bad sequencing quality were discarded. This left one location ('Billed') with only three individuals. We decided to drop the entire location, since the significance of genetic patterns could not be assured, resulting in a final data set of 374 individuals from 24 locations (Table 1).
The VCF with the genotype likelihood scores contained 11,961,795 SNPs; and after imputation 9,752,655 SNPs. After a final filtering for MAF (5%) the number of SNPs was reduced to 1,635,950 distributed over the 18 chromosomes.

Landscape genomics

To identify associations between individual SNPs and the 16 environmental variables, we used Latent Factor Mixed Models with K (the number of latent factors) ranging between 4 and 11. LFMM detected between 37 and 60 SNPs (depending on the value for K) significantly associated with one or more environmental variable. The set of environmental variables associated with one or more SNP markers were consistent across values for K, with one exception: with latent factors ranging between K=4 and K=8, we detected additional associations with 'canopy height'. Comparing all the genotype-environment associations, we found that the set of markers in Table 2 was present in all LFMM runs. LFMM runs with K=4 – K=7 and K=11 showed some additional markers in comparison to K=8 – K=10, so we decided to act very conservatively and only consider the common SNP marker between all LFMM runs (K=4-11). We chose a tight range of latent factors (K=9 ±1), because an incorrect assumption about underlying population structure was shown to increase type I and type II errors (Cushman et al., 2010; Storfer et al., 2018). Consequently, we only considered loci detected for all three runs (K=9 ±1).

Our most conservative estimates for significance (based on K=9 ±1) resulted in 29 significant associations between SNPs and environmental variables (with seven showing an FDR-corrected p-value < 0.01). The nine environmental variables significantly associated to genetic markers were: QSCATseason (seasonality in surface moisture (coefficient of variation)), Bio 8 (mean temperature of the wettest quarter), Bio 4 (temperature seasonality), Bio 3 (isothermality), LAIsd (seasonality in Leaf Area Index (standard deviation across the year)), Bio 11 (mean temperature of the coldest quarter), Bio 19 (precipitation of the coldest quarter), percent tree cover and Bio 15 (precipitation seasonality). For each of these, a Manhattan plot was generated (Figure 2).

The significant SNPs were broadly distributed across the genome, except of Chromosomes 4, 11, 14, 16 – 18. Gene-environment associations of three loci (V466493, V647421 and V511914) were detected repeatedly across the environmental variables used in this study. They were located in Chromosome 5, 6 and 7. These associations were plotted, but no clear visual pattern could be identified (Figure 3). Four loci (V466493, V647421, V1285063 and V1463790) showed highly significant associations (FDR of 1%) (Figure 3 B-F, Figure 4). A noteworthy pattern existed (Figure 4 A) in the association of V1285063 and the mean temperature of the wettest quarter (Bio 8), which we additionally visualized on a map (Figure 5). Whereas only heterozygotes and homozygotes for the reference allele are present in populations with mean temperatures of the wettest quarter below 15 °C, in warmer regions also individuals that are homozygote for the alternative allele occur.

All SNP markers (except V441690, in Chromosome 5) are located in coding regions and are associated with 44 genes in total (either inside a gene or located within ± 100.000 bp) (Table S1). *Bombus terrestris* shares 6,768 orthologs with the honey bee (*Apis melifera*) (Woodard et al., 2011), so it was not surprising that all SNP markesr in this study were located in or close to genes which are also orthologs with genes found in Apis melifera. Nine of the genes were uncharacterized without any known function. Seven genes were orthologs with genes in the fruit fly (*Drosophila melanogaster*) and 28 of the identified genes showed a variety of functions in humans, spanning from developmental processes through DNA-repair mechanisms to other biological processes. We want to highlight two genes, which have previously been identified in wild bee genomes: LOC100647823 is the 'poly(rC)-binding protein 3', which was identified to be involved in the evolution of eusociality in bees (Woodard et al., 2011). LOC100648854, which is the 'ABC transporter G family member 22', an ABC transporter, is commonly associated to insecticide resistance in insects (Broehan et al., 2013).

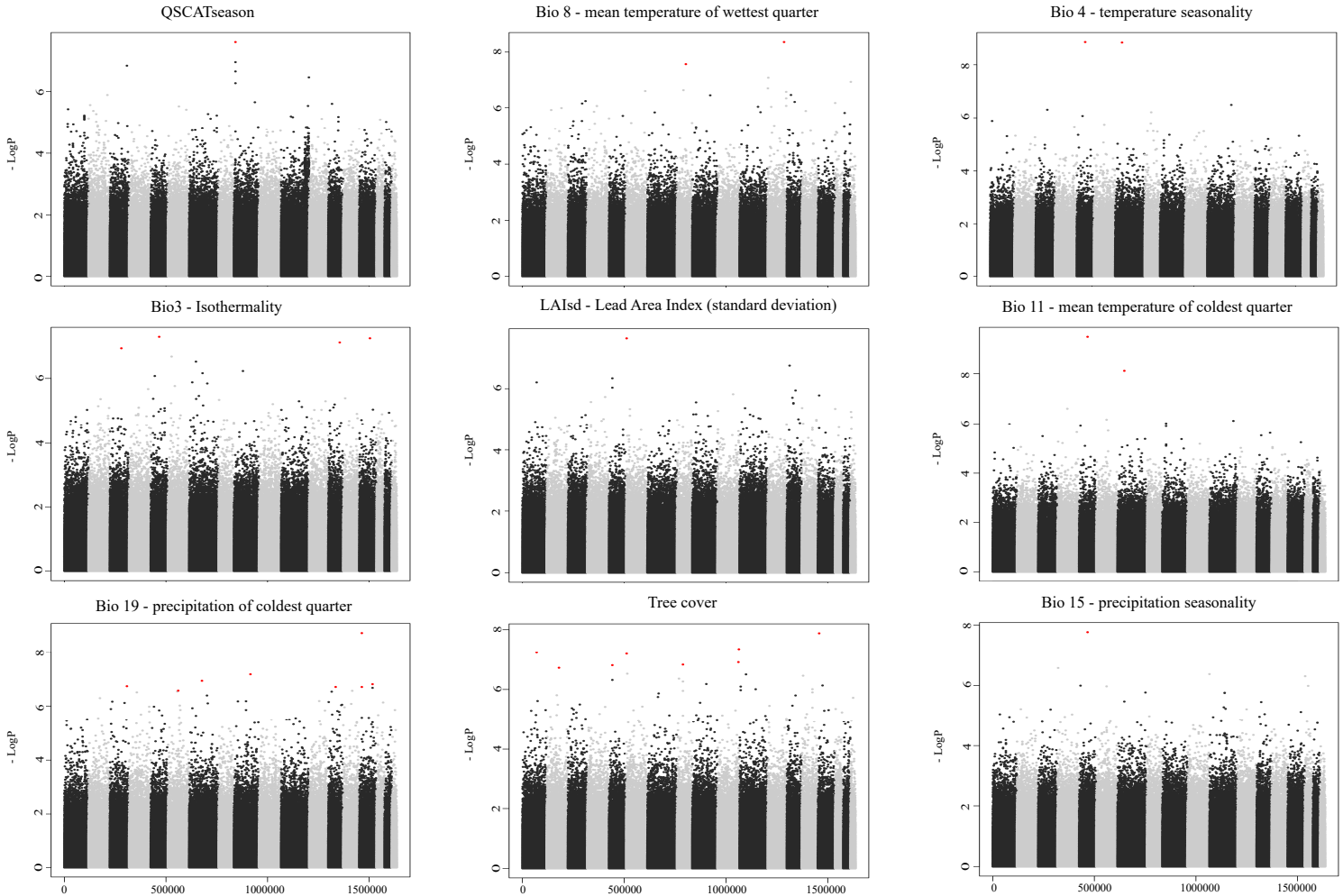

FIGURE 2: Manhattan plots of significant gene- environment associations. Indicated in red, the significant SNP marker associated to the respective environmental variable.

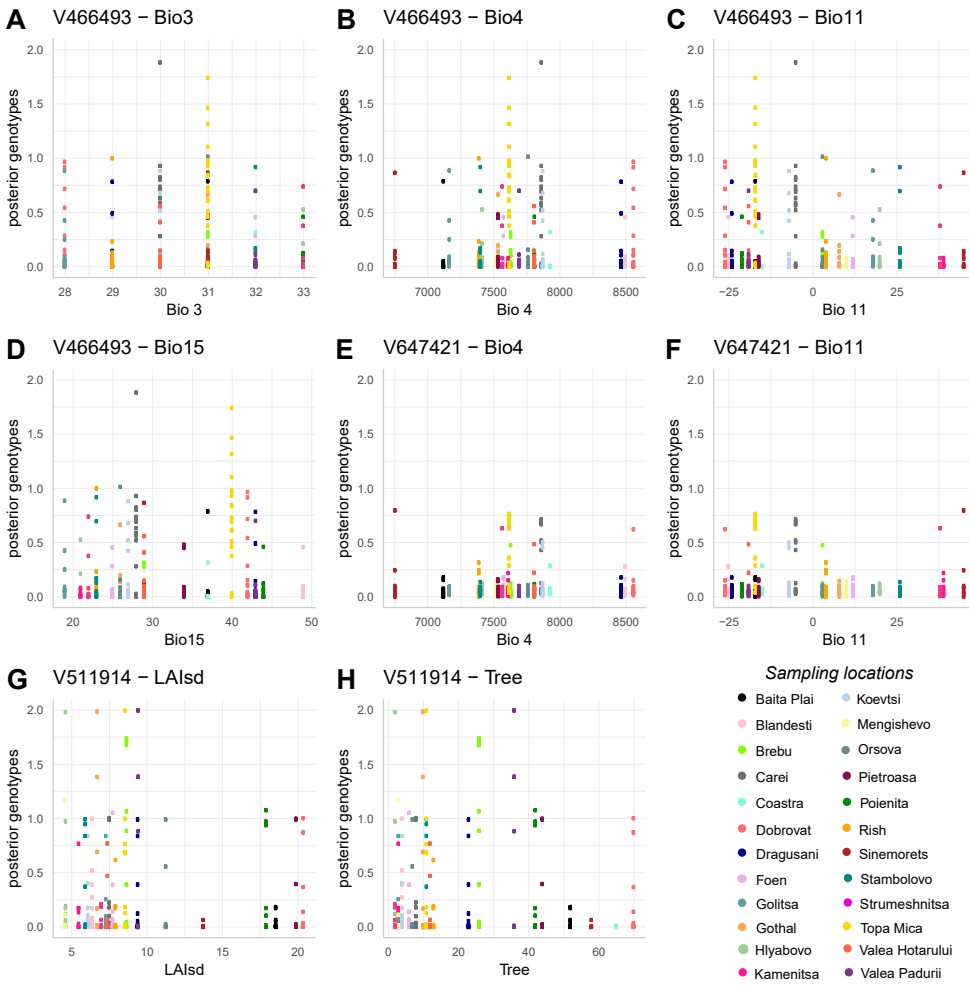

FIGURE 3: Scatter plots of the three marker (V466493 (Plot A-D), V647421 (Plot E and F) and V511914 (Plot G and H)) significantly associated to more than one environmental variable. The values for each environmental variable with the posterior genotypes of each samples individual per location is plotted here. Plot B – F show highly significant associations with a FDR of 1%. Sampling locations are indicated in different colors.

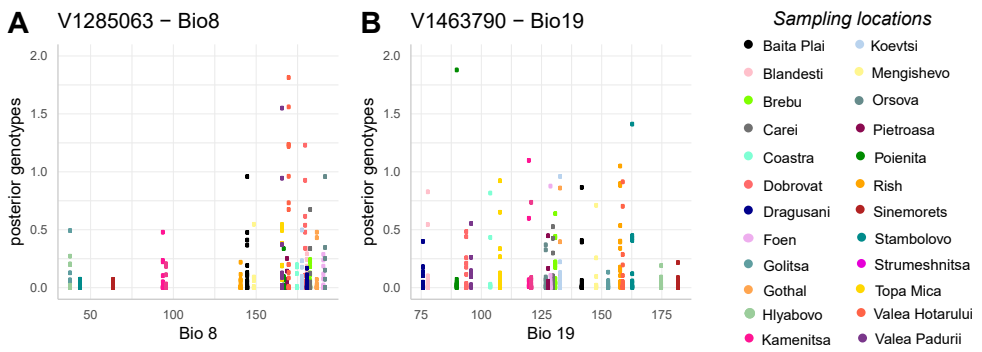

FIGURE 4: Scatter plots of two marker (V1285063 and V1463790) highly significantly associated (FDR of 1%). The values for each environmental variable with the posterior genotypes of each samples individual per location is plotted here. Sampling locations are indicated in different colors.

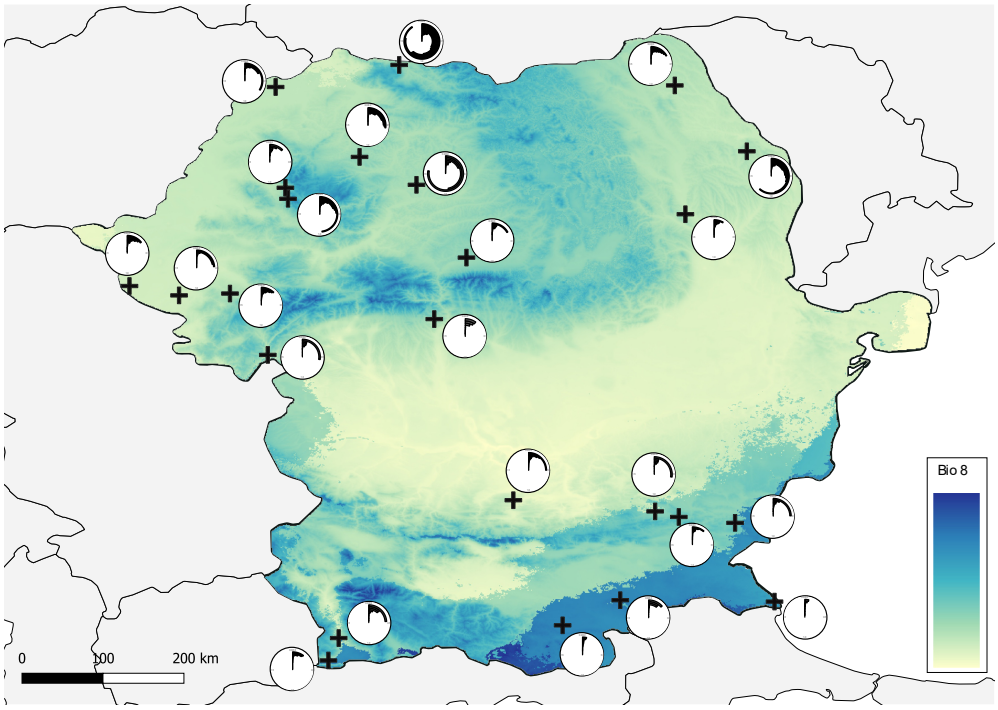

FIGURE 5: Mapped association between V1285063 with Bio 8 (mean temperature of the coldest quarter). For each location a polar coordinated pie chart is visualizing the posterior genotypes of each individual in each location. Every line indicates one individual and the genotypes are shown clockwise: starting with the homozygous (0 at 12 o'clock), the heterozygous (1 at 6 o'clock) and ending with the other homozygous (2 at 12 o'clock again). Colder colors (blue) indicate colder temperatures.

Discussion

Here, we present the first genome-wide analysis of potential local adaptations in the buff-tailed bumble bee (*Bombus terrestris*) using whole-genome sequencing. With the help of gene-environment associations (GEAs) we identified 29 loci significantly associated to 9 (out of 16) environmental variables. As hypothesized based on their habitat preferences and life-cycle, we found molecular markers to be associated with seasonality in temperature and precipitation, as well as vegetation variables. Our results suggest that these variables are key factors in structuring genetic variation and may impose selection pressures, driving local adaptation. With this work we added to the small list of studies investigating local adaptation in bumble bees, such as the study by Miller-Struttmann et al., 2015 who identified signs of local adaptation in two alpine bumble bee species as a response to a decrease in floral resources, as well as another study showing the adaptability of the red-tailed bumble bee (*Bombus lapidarius*) to the urban environment (Theodorou et al., 2018).

In *Chapter I*, we investigated the habitat preferences of *B. terrestris* with the help of species distribution modeling (SDM), using the same set of environmental variables. Among the top six (most influential) variables determining the habitat suitability for *B. terrestris*, we found seasonality in surface moisture (QSCATseason), mean temperature of wettest quarter (Bio 8), temperature seasonality (Bio 4), isothermality (Bio 3) and the standard deviation of the Leaf Area Index (LAIsd). This set is concordant with the results of the GEAs in this study, further suggesting that these environmental conditions may be important for the ecology of this species and might drive evolutionary adaptations to local conditions. *B. terrestris*, like other bumble bee species, exhibits a highly seasonal life cycle: it occurs usually as early as March and disappears between September and October (Goulson, 2010). Therefore, it may be no surprise that we found significant associations to the seasonality variables within our data set (seasonality in surface moisture, temperature and precipitation). Here we should point out that in some European countries such as Turkey, the UK and countries in the Mediterranean region, more and more evidence is found that *B. terrestris* averts diapause and is therefore considered to be 'winter-active' (Owen et al., 2013; Stelzer et al., 2010). We however do not have any indications that this might be the case for the bumble bees collected for this study, since Romania (and partially Bulgaria) still experience harsh winters.

We did not only find associations to seasonality variables but also to variables describing climate in the coldest or wettest months (Bio 8, Bio 11 and Bio 19). These results are concordant with morphological adaptations of bumble bee species, which are covered in dense fur to enable them to adapt to cold climates (Goulson, 2010), and are supported by the fact that climate is an important determinant of geographical ranges and habitat choice (Williams and Osborne, 2009). Similar genomic signatures of adaptation were found in a recent landscape genomics study of the tropical bee

Melipona subnitida, where temperature, precipitation and forest cover were among the strongest associated environmental variables (Jaffé et al., 2019). For honey bees (*Apis melifera*) from the Iberian Peninsula, precipitation was also one of the strongest associated environmental variables (Henriques et al., 2018).

Interestingly, the most important variables for explaining the occurrence patterns (QSCATseason, Bio 8, Bio 4 and Bio 3; see *Chapter I*) were also associated to one or more loci in this study (Table 1, Figure 3). Particular attention is warranted for the three loci associated to several environmental variables (V466493 with Bio 3, Bio 4, Bio 11 and Bio 15; V647421 with Bio 4 and Bio 11; V511914 with LAIsd and tree cover). The repeated detection of these loci makes it less likely that the associations are strongly influenced by population structure or the covariance between variables (Rellstab et al., 2015), and suggest an important ecological meaning of these environmental variables. Visual evaluations of those associations did however not suggest any clear patterns within and between locations (Figure 3).

In addition, a total of four markers (V466493, V647421, V1285063 and V1463790) showed particularly significant associations (q-value < 0.01). One of these associations is showing an interesting pattern (V1285063 with Bio 8; Figure 4 A), suggesting that in regions where the mean temperature of the wettest quarter is below 15 °C only individuals homozygous for the reference allele are present. In 'warmer' regions with a mean temperature over 15 °C both the heterozygous and homozygous genotype for the non-reference allele are also present, although at relatively low frequencies. These results may suggest that negative selection against the heterozygous and homozygous (for non-reference allele) genotypes in 'colder' regions. We might speculate that under climate change this negative selection pressure will be reduced, resulting in higher frequencies of the heterozygous and alternative homozygous genotypes.

For most of the potential genes we here identified, little is known about their function in insect species, and even less so in bumble bees. Drawing conclusions on the functional relevance of the identified genes with respect to the associated environmental variables is therefore extremely limited, in particular because most of them are only known for their function in humans (Table S1). There are two exceptions: LOC100647823 a 'poly(rC)-binding protein 3', which is known to be involved in the evolution of eusociality in bees (Woodard et al., 2011) and LOC100648854, an 'ABC transporter G family member 22', which is related to insecticide resistance in insects (Broehan et al., 2013). In the buff-tailed bumble bee, these genes are located in Chromosome 3 and 15, and associated to 'isothermality' (Bio 3) and 'precipitation of the coldest quarter' (Bio 19) respectively. However, how these factors may be related to eusociality and insecticide resistance remains unclear.

In general, climate and other environmental factors are known to pose selection pressures on natural populations (Joshi et al., 2001). The here used gene-environment associations help to disentangle these pressures and indicate what environmental

variables might shape genetic structure for the species being investigated (Frichot et al., 2013; Joost et al., 2007). They expand our understanding of evolutionary processes influenced by environmental selection (Bonin, 2008; Parisod and Holderegger, 2012), but may be particularly relevant for conservation.

Although the conservation status of *B. terrestris* is "least concern", it was predicted that it might suffer from population declines due to climate change and human habitat alterations (Rasmont et al., 2015a). We here found significant associations of genomic markers with several climate variables, suggesting that global warming will likely affect buff-tailed bumble bees. For example the shift in seasons is directly linked to warmer global temperatures (Carré and Cheddadi, 2017). This could result in a change in the life-cycle of this species, and might even provoke *B. terrestris* to become 'winter-active' in Romania and Bulgaria. In addition, increases in precipitation seasonality are predicted as a result of climate change (Williams and Middleton, 2008), and will likely require adaptive responses as well. Although the exact nature and magnitude of these effects and responses are difficult to forecast, our results provide insights into which regions of the genome may be particularly affected by changing habitat conditions. Moreover, because under the current rate of environmental change, species' adaptive responses are highly dependent on the already available genetic variation, it is crucial to protect a set of populations that together possess both the reference and alternative alleles of the here identified significantly associated loci. Such a network of reserves must not protect locations and populations as isolated entities, but also safeguard the potential for gene flow.
So far, little attention has been paid to the genetic structure and the adaptive potential of *B. terrestris*. With this study, we took a first step towards filling this gap. Future work is needed to better understand the functional context of genomic adaptations in *B. terrestris*, which are facilitated by the fact that the ecology of *B. terrestris* has been relatively well studied. To this end, our study provides a set of candidate loci and genes that warrant further investigation.

Supplementary information

NEXTERA low coverage library adapted from Baym (2015, Plos One)
with the help of Claudia Michel (ETH Zurich)

Illumina Nextera DNA sample prep kit (96) FC-121-1031 enough for 950 samples and IDT indexes (384 indexes)

DNA : Normalize at 2.5 ng/ul in a volume of **20ul**

Tagmentation
3ul DNA + 2.5ul Tag-Buffer + 0.5 ul Tag-Enzyme } **Tag-Mix** for samples
10 min at 55°C
Mix by pipetting 10X up & down

*Thaw **everything** on ice! Invert gently 3-5 times and centrifuge*

PCR
IDT Adapter (2 primers mixed): 10ul of adapter mix (1:2) + 30ul water + **15ul** KAPA Hifi Mix

Already mixed → 10ul

Thaw on ice for ~20min! Invert gently 3-5 times and centrifuge briefly

6 ul tagmented DNA
15 ul KAPA Hifi Mix
10 ul IDT Adapter-Mix } **31**ul total volume

72°C 3min, 95°C 3 min
98°C 20sec, 62°C 15sec, 72°C 1 min (**13** cycles)
72°C 1 min
Centrifuge down!

Double size selection: *At RT for at least 30min. Vortex for 2 min and make aliquots! Vortex between each sample*
Long fragments
28ul PCR + 22ul water + **30**ul Ampure beads (= **0,60x**); Mix by pipetting 10X up & down
Wait 10 min (vortex and centrifuge rapidly every 2-3min), then bring the tube in contact with magnet, wait **2 min**, **KEEP and TRANSFER** the supernatant
Short fragments
75ul Supernatant + **7,5** ul Ampure beads (= **0,1X**); Mix by pipetting 10X up & down
Wait 10 min (vortex and centrifuge rapidly every 2-3min), then bring the tube in contact with magnet, wait **2 min**, **DISCARD** the supernatant
Wash beads
Add 200 ul 80% ethanol, wait 30sec and remove (x2) *Prepare **freshly** every time!*
Wait 5min (if not dry another 5min) to make sure that the ethanol is evaporated

Normalization and pooling:
Remove tube of the magnet and elute in **50**ul H2O, vortex and centrifuge;
Wait 10 min (vortex and centrifuge rapidly every 2-3min), then bring the tube in contact with magnet, wait **2 min** and put the eluate in a new tube.
Measure DNA concentrations

Pool equal amounts of each library → **25ng per sample**

Table S1 – List of significant SNP Marker associated to environmental variables. Information of the position in the genome, Chromosome, if they are located in coding regions and if they are located in genes or close to genes. The name of the genes, the location and the potential function in either Humans (Homo sapiens) or if available in Fruit flies (Drosophila melanogaster) is described.

Env. variable	SNP Marker	Position in genome	Chrom-osome	Coding region	Genes	Potential function
QSCATseason	V841835	NC_015770.1:973808	Chr 9	yes	**LOC100646111** Location: 983,273-986,393 *ankyrin repeat and MYND domain-containing protein 2*	Humans: enzyme binding (involved in the trafficking of signalling proteins to the cilia)
Bio 8	V1285063	NC_015773.1:10980707	Chr 12	IN gene	**LOC100648067** Location: 10,909,367-11,030,506 *transcription factor SPT20 homolog*	Humans: Developmental protein (involved in Autophagy and Gastrulation)
	V802687	NC_015769.1:5731985	Chr 8	**IN gene**	**LOC100648283** Location: 5,688,473- 5,844,250 *aryl hydrocarbon receptor protein 1*	*D. melanogaster*: involved in the control of breathless expression and in the cellular or tissue response to oxygen deprivation
Bio 4	**V466493**	**NC_015766.1:6120680**	**Chr 5**	**yes**	**LOC100651835** Location: 6,128,724-6,163,862 ***tubulin polyglutamylase TTLL5***	Humans: Developmental protein (involved in transcription)
					LOC100642245 **Location: 6,121,046-6,128,513** ***zinc finger protein 16***	Humans: Developmental protein (involved in Cell cycle, Cell division, Transcription, Transcription regulation)
					LOC100646176 **Location: : 6,119,114-6,120,392** ***uncharacterized***	
					LOC100642448 **Location: 6,115,454-6,118,582** ***transferrin receptor protein 1***	Humans: involved in Endocytosis, Host-virus interaction
					LOC100642566 **Location: 6,114,379-6,115,885** ***uncharacterized***	
	V647421	**NC_015768.1:5413242**	**Chr 7**	**yes**	**LOC110119146** **Location: 5,414,421-5,415,904**	*D. melanogaster*: involved in sugar transport

					UDP-sugar transporter UST74c	
					LOC100649656 **Location: 5,413,929-5,419,399** ***ubiquinone biosynthesis protein COQ4 homolog, mitochondrial***	Humans: involved in ubiquinone biosynthesis
					LOC100649295 **Location: 5,409,433-5,413,456** ***protoheme IX farnesyltransferase, mitochondrial***	Humans: Transferase in heme biosynthesis
					LOC100647971 **Location: 5,417,443-5,425,368** ***discoidin domain-containing receptor 2***	Humans: involved in osteogenesis
Bio 3	V466493	NC_015766.1:6120680	Chr 5	yes	**LOC100651835** Location: 6,128,724-6,163,862 ***tubulin polyglutamylase TTLL5***	Humans: Developmental protein (involved in transcription)
					LOC100642245 **Location: 6,121,046-6,128,513** ***zinc finger protein 16***	Humans: Developmental protein (involved in Cell cycle, Cell division, Transcription, Transcription regulation)
					LOC100646176 **Location: : 6,119,114-6,120,392** ***uncharacterized***	
					LOC100642448 **Location: 6,115,454-6,118,582** ***transferrin receptor protein 1***	Humans: involved in Endocytosis, Host-virus interaction
					LOC100642566 **Location: 6,114,379-6,115,885** ***uncharacterized***	
	V1355085	NC_015774.1:7344328	Chr 13	yes	**LOC100650416** Location: 7,247,835-7,339,008 *protein tiptop*	*D. melanogaster*: developmental protein

	V1503835	NC_015776.1:6611877	Chr 15	**IN the gene**	**LOC100647823** Location: 6,594,424-6,771,007 *poly(rC)-binding protein 3*	Humans: involved in RNA-binding **Genes involved in convergent evolution of eusociality in bees (Woodard et al., 2011)**
	V280355	NC_015764.1:7252345	Chr 3	**IN the gene**	**LOC100650996** Location: 7,097,441-7,628,587 lachesin	*D. melanogaster*: developmental protein (required for normal tracheal development and maintenance of the trans-epithelial diffusion barrier)
LAIsd	V511914	NC_015767.1:275765	Chr 6	**IN the gene**	**LOC110119371** Location: 249,570-297,467 *uncharacterized*	
Bio 11	**V466493**	**NC_015766.1:6120680**	**Chr 5**		**LOC100651835** Location: 6,128,724-6,163,862 ***tubulin polyglutamylase TTLL5***	Humans: Developmental protein (involved in transcription)
					LOC100642245 **Location: 6,121,046-6,128,513** ***zinc finger protein 16***	Humans: Developmental protein (involved in Cell cycle, Cell division, Transcription, Transcription regulation)
					LOC100646176 **Location: : 6,119,114-6,120,392** ***uncharacterized***	
					LOC100642448 **Location: 6,115,454-6,118,582** ***transferrin receptor protein 1***	Humans: involved in Endocytosis, Host-virus interaction
					LOC100642566 **Location: 6,114,379-6,115,885** ***uncharacterized***	
	V647421	**NC_015768.1:5413242**	**Chr 7**		**LOC110119146** **Location: 5,414,421-5,415,904** ***UDP-sugar transporter UST74c***	*D. melanogaster*: involved in sugar transport
					LOC100649656 **Location: 5,413,929-5,419,399** ***ubiquinone biosynthesis protein COQ4 homolog, mitochondrial***	Humans: involved in ubiquinone biosynthesis

					LOC100649295 **Location: 5,409,433-5,413,456** ***protoheme IX farnesyltransferase, mitochondrial***	Humans: Transferase in heme biosynthesis
					LOC100647971 **Location: 5,417,443-5,425,368** ***discoidin domain-containing receptor 2***	Humans: involved in osteogenesis
Bio19	**V1463790**	**NC_015776.1:2064504**	**Chr 15**	**IN the gene And 1 really close**	**LOC100651020** **Location: 1,669,726-2,210,653** ***suppressor of lurcher protein 1***	
					LOC110119947 **Location: 2,055,928-2,059,531** ***uncharacterized***	
	V915653	NC_015770.1:9935769	Chr 9	**IN the gene And 2 really close**	**LOC100649027** Location: 9,891,067-10,118,404 *neuroligin-4, X-linked*	Humans: involved in cell adhesion
					LOC100648829 Location: 9,879,245-9,887,975 *ras-related protein Rab6*	*D. melanogaster*: involved in protein transport (regulator of membrane traffic from the Golgi apparatus towards the endoplasmic reticulum (ER))
					LOC100648714 Location: 9,874,683-9,878,945 *protein FAM8A1*	
	V307659	NC_015764.1:9864342	Chr 3	**IN the gene**	**LOC100648854** Location: 9,657,786-9,923,274 *ABC transporter G family member 22*	important physiological functions in all living organisms**.** **In insects, ABC transporters are of special interest because of their role in insecticide resistance. (Broehan et al., 2013)**
	V676937	NC_015768.1:9628798	Chr 7	**IN the gene**	**LOC100648397** Location: 9,524,224-9,635,260	

				And 2 really close	*ankyrin repeat domain-containing protein SOWAHB*	
					LOC100648635 Location: 9,638,884-9,666,426 *serine/threonine-protein kinase N*	Humans: involved in Apoptosis
					LOC100648514 Location: 9,636,206-9,637,441 *uncharacterized*	
	V1516520	NC_015776.1:8010461	Chr 15	**IN the gene And 1 really close**	**LOC100650939** Location: 8,002,941-8,064,830 *serine-rich adhesin for platelets*	Humans: key players in cancer biology, as well as inflammation
					LOC100651055 Location: 7,998,214-8,000,747 *selenocysteine insertion sequence-binding protein 2-like*	
	V1463770	NC_015776.1:2062133	Chr 15	**IN the gene**	**LOC100651020** **Location: 1,669,726-2,210,653** ***suppressor of lurcher protein 1***	
	V1334923	NC_015774.1:4329591	Chr 13	**IN the gene**	**LOC100650658** Location: 4,195,490-4,369,953 *metabotropic glutamate receptor 2*	*D. melanogaster:* G-protein coupled receptor, Receptor, Transducer
	V560136	NC_015767.1:5787495	Chr 6	**IN the gene And 1 really close**	**LOC100650482** Location: 5,784,209-5,787,516 *potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1-like*	Humans: involved in Ion transport, Potassium transport, Sodium transport
					LOC100644721 Location: 5,046,661-5,760,540 *teneurin-a*	*D. melanogaster:* involved in neural development, regulating the establishment of proper connectivity within the nervous system.
Tree cover	V1457378	NC_015776.1:1253942	Chr 15		**LOC110119922** Location: 1,252,102-1,253,251	

					uncharacterized	
					LOC100643343 Location: 1,246,486-1,250,242 *DNA replication licensing factor Mcm6*	Humans: involved in cell cycle, DNA replication
					LOC100645539 Location: 1,242,472-1,245,796 *uncharacterized*	
	V69823	NC_015762.1:10211783	Chr 1		**LOC110119719** Location: 10,197,277-10,197,349 *transfer rna phenylalanine*	
					LOC100646973 Location: 10,212,445-10,218,399 *uncharacterized*	
					LOC100651910 Location: B01: 10,219,602-10,226,126 *uncharacterized*	
					LOC100647090 Location: 10,226,244-10,229,784 *another transcription unit protein*	
	V511914	NC_015767.1:275765	Chr 6		LOC110119371 Location: 249,570-297,467 *uncharacterized*	
	V1063161	NC_015771.1:13309360	Chr 10	**IN the gene**	**LOC105666188** Location: 12,974,121-13,453,525 *CCR4-NOT transcription complex subunit 6-like*	Humans: involved in Transcription, Transcription regulation, Translation regulation, mRNA processing, RNA-mediated gene silencing,
	V1060815	NC_015771.1:13140800	Chr 10	**IN the gene**	**LOC105666188** Location: 12,974,121-13,453,525	Humans: involved in Transcription, Transcription regulation, Translation regulation, mRNA processing, RNA-mediated gene silencing,

					CCR4-NOT transcription complex subunit 6-like	
	V788356	NC_015769.1:4062621	Chr 8	**IN the gene**	**LOC100642492** Location: 4,012,209-4,176,608 *Fanconi anemia group J protein homolog*	Humans: involved in DNA damage, DNA repair
	V441690	NC_015766.1:2462459	Chr 5	NO		
	V179495	NC_015763.1:7568447	Chr 2	**IN the gene And 3 really close**	**LOC100644876** Location: 7,551,296-7,622,375 *LIM domain only protein 3*	Humans: involved in DNA damage, DNA repair
					LOC100643967 Location: 7,529,932-7,532,068 *cell cycle checkpoint protein RAD1*	Humans: involved in DNA damage, DNA repair
					LOC100643844 Location: 7,460,929-7,538,283 *calmodulin*	Humans: involved in ATP-binding, Nucleotide-binding
Bio 15	**V466493**	**NC_015766.1:6120680**	**Chr 5**		**LOC100651835** Location: 6,128,724-6,163,862 ***tubulin polyglutamylase TTLL5***	Humans: Developmental protein (involved in transcription)
					LOC100642245 **Location: 6,121,046-6,128,513** ***zinc finger protein 16***	Humans: Developmental protein (involved in Cell cycle, Cell division, Transcription, Transcription regulation)
					LOC100646176 **Location: : 6,119,114-6,120,392** ***uncharacterized***	
					LOC100642448 **Location: 6,115,454-6,118,582** ***transferrin receptor protein 1***	Humans: involved in Endocytosis, Host-virus interaction
					LOC100642566 **Location: 6,114,379-6,115,885** ***uncharacterized***	

6 General Discussion

With ongoing global environmental change and human-induced threats such as habitat alteration, pollution, overexploitation, and climate change, species are challenged, and biodiversity as a whole is under major pressure (Hoffmann et al., 2010; Pereira et al., 2012; Vanbergen and Initiative, 2013). It is now more important than ever to act and protect biodiversity. Its complexity, however, presents an obstacle in the efficient and effective implementation of protection and management strategies (Bartkowski et al., 2015; Noss, 1990), and an in-depth understanding of how environmental conditions and future changes influence biodiversity is urgently needed.

There is ample evidence that environment is one of the main drivers of spatial patterns of biodiversity (Fine, 2015; Jetz et al., 2012; Peters et al., 2016). Environment is acting on ecological and evolutionary processes, which themselves shape all levels of biodiversity in one way or the other; ranging from habitat types being defined by different climatic and geological conditions (e.g. presence of water, altitude) (e.g. Belmaker and Jetz, 2015; Mittelbach et al., 2007; Peters et al., 2016), through species being influenced in their distribution and abundance (e.g. Costa et al., 2008; Elith and Leathwick, 2009; Peters et al., 2016), to genetic diversity being shaped by environmental characteristics and differences (e.g. Conover et al., 2009; Wang and Bradburd, 2014). With selective pressures by the environment on natural populations being just one factor shaping the spatial distribution of biodiversity (Joshi et al., 2001; Mosca et al., 2012), it is important to disentangle its effects from those of other components. One way to do this is by first mapping the spatial distribution of biodiversity (such as the distribution of habitats, species and genes) and relating this to the prevailing environmental conditions. Additionally, insights gained by these approaches can be then used to adjust management practices and conservation efforts (Guisan et al., 2013; Jetz et al., 2012; Pearson, 2007).

In this thesis, I investigated the relationship between environment and the spatial patterns of biodiversity in in the context of biodiversity conservation. The aim of my thesis was to assess the relative influence of environment on biodiversity and to evaluate how this understanding can be used for mapping and ultimately protecting biodiversity.

To do so, I focused on several different components of biodiversity: habitats, species and genes. This work was conducted in two eastern European countries, Romania and Bulgaria. These countries are an interesting area for evaluating the effect of environment on the distribution of biodiversity because of the high levels of environmental heterogeneity present within their borders. In **Chapter I**, I investigated the distributional patterns of two closely related bumble bee species (the buff-tailed (*Bombus terrestris*) and the white-tailed bumble bee (*Bombus lucorum*)) and the influence of environmental characteristics in determining those patterns. In **Chapter II**, I examined

how environmental heterogeneity (in form of different habitat types) can be used as a conservation surrogate for the occurrence of species and vice versa. Additionally, I was interested in how well these two levels of biodiversity (habitats and species) are represented by already existing conservation areas. In **Chapter III**, I moved on to the genetic level of biodiversity, and studied the role of environment in shaping the distribution of genetic and morphological diversity in the house sparrow (*Passer domesticus*). Finally, in **Chapter IV**, I moved from using neutral markers exhibiting patterns of isolation by adaptation towards genome-wide SNP markers to home in on signs of local adaptations in coding regions in the genome of the buff-tailed bumble bee (*Bombus terrestris*).

Influence of the environment on species distributions

In **Chapter I**, focussed on the influence of environmental variation on the occurrence patterns of two closely related bumble bee species (the buff-tailed (*Bombus terrestris*) and the white-tailed bumble bee (*Bombus lucorum*)). The distributions of these species partially overlap, leading to the question how the habitat requirements and the ecological drivers shaping their distributions differ between species.

Because the majority of the samples were workers, which are difficult to distinguish just morphologically, in a first step, I genetically identified individuals to the species level (e.g. Bertsch, 2010; Gammans et al., 2018; Williams, 1994). I then correlated species occurrence and abundance data with environmental data in order to model the species' distributions. I used ensemble Species Distribution Modeling (SDM) techniques to improve the validity of my results. I found that despite their partial overlap, *B. terrestris* has a much wider distribution than *B. lucorum*, which was mainly restricted to mountainous areas. In the SDMs, both vegetation and climatic variables played a major role in determining the distributions of both species. These results are concordant with the literature, indicating that these species prefer different vegetation types and densities (Bossert et al., 2016; Svensson et al., 2000), and occur at different elevations (e.g. Bossert et al., 2016; Ploquin et al., 2013). The resulting differences in habitat requirements between these species suggest they will exhibit differential responses to future environmental and climate change. Such responses can be ecological, evolutionary or a combination of both (Anderson et al., 2012). Knowledge on the distribution and ecology of species is thus key to facilitate the assessment of their conservation status, and the development of management practices in order to protect them (Guisan et al., 2013; Jetz et al., 2012; Pearson, 2007).

Reciprocal surrogacy of habitat and species diversity

Because environmental heterogeneity at least partially determines the distribution of species (my work and for example Costa et al., 2008; Elith and Leathwick, 2009),

one might expect that it can be used as a surrogate for other measures of biodiversity (Arponen et al., 2008; Beier et al., 2015; Engelbrecht et al., 2016; Grantham et al., 2010). However, previous studies showed contradictory results in the surrogacy performance of environmental diversity (Bonn and Gaston, 2005; Sarkar et al., 2005; Trakhtenbrot and Kadmon, 2005). In **Chapter II**, I thus aimed to test the reciprocal use of classified environmental diversity (in form of habitat types) and a measure of bird species richness, in many countries the two best known and most easily obtained measures of biodiversity of high quality. Using a spatial conservation prioritization method, I identified areas of conversation concern based on both habitats and birds, and evaluated their representation for one another. I found that bird species were a better surrogate for habitat diversity than vice versa. This result was rather surprising, because one might expect that by covering a certain amount of environmental heterogeneity, a large proportion of species diversity should also be protected (Engelbrecht et al., 2016). So even though environment plays an important role for species distributions, the results of this chapter show the limitations of the use of environment as a conservation surrogate (concordant with Araújo et al., 2007; Bonn and Gaston, 2005). These results highlight that species and environmental data may be poor surrogates for one another and that different types of biodiversity measures should be combined in spatial conservation prioritization (Arponen et al., 2008; Bonn and Gaston, 2005; Di Minin and Moilanen, 2014; Lombard et al., 2003).

In a second step, I evaluated the representation of habitat types and bird species in existing protected areas and additionally identified potential expansion regions. This is of particular interest under the new targets of the European Union Biodiversity Strategy for 2030, where each country has to increase their network of protected areas up to 30% of the total land surface area (Commission, 2020). I found that habitat diversity was generally better represented than bird species. In particular elevation gradients and rural grassland regions were underrepresented and identified as potential expansion areas.

Relationship between the environment and genetic diversity

In **Chapter III** and **Chapter IV**, I investigated the influence of the environment on the genetic level of biodiversity. When species face environmental changes, they will most likely need to rely on a combination of ecological and evolutionary responses (Anderson et al., 2012). While adaptive evolution can be quick, it strongly depends on the currently available genetic variation, i.e. standing genetic variation (Moritz et al., 2001; Thomassen et al., 2011)), and in particular on that part of genetic variation that is correlated to current environmental characteristics (environmentally-associated variation, EAV), and may thus be under natural selection. It is therefore essential to understand the spatial patterns of EAV in natural populations in order to assess the impact of environmental change and predict species' responses (Miraldo et al., 2016;

Palumbi, 2001).
In **Chapter III**, I assessed the effect of environment on genetic population structure and morphological divergence of house sparrows (*Passer domesticus*). Using a landscape genomics approach, I correlated phenotypic (wing, tail and tarsus length) and genetic (pairwise genetic differences based on microsatellite marker) data with a set of environmental variables. The results suggest that 'isolation by distance' (IBD; Wright, 1943) did not have any influence on the observed morphological and genetic variation. In contrast, up to 30% of the observed population divergence could be explained by the environment (mainly climate and vegetation variables), which is hinting towards a scenario of 'isolation by environment', where signatures of selection are detectable in neutral markers (IBE; Wang and Bradburd, 2014). Our finding that local environmental conditions to a considerable extent explain genetic variation, and thus that selective processes may be a major driver of genetic diversity in house sparrows are concordant with previous results in other countries (Holand et al., 2011; Lima et al., 2012; Liu et al., 2013), and my work added to a growing body of evidence that adaptive evolution may be a major driver of diversification.
Based on this, I moved on from neutral molecular markers towards single-nucleotide polymorphisms (SNPs) that can be traced to intronic and exonic regions of the genome, and thus provide more detailed information on potential adaptations and responses to future environmental changes. In **Chapter IV** I looked for signs of local adaptations related to environmental heterogeneity in the buff-tailed bumble bee (*Bombus terrestris*). I used whole-genome data of the same individuals that I collected for **Chapter I**. Based on the confirmed influence of the environment on shaping the distribution of this species, I aimed to further investigate its genetic basis. With the help of gene-environment associations (GEAs), I correlated genotypic data with the same set of environmental variables used in **Chapter I**. I found loci putatively under selection, which were correlated to the same climatic and vegetation variables shaping *B. terrestris*' distributional patterns, suggesting that these are key factors in structuring genetic variation in this species. My results add to a better understanding of how populations are adapted to local conditions, and whether these adaptations may be relevant for future responses to changing environments (Capblancq et al., 2018; Frichot et al., 2013; Hoban et al., 2016; Joost et al., 2007).

Conclusions and Outlook

In this work, I demonstrated the tight link between environmental conditions and biodiversity, and teased apart the relative roles of environmental variables in shaping the spatial distribution of biodiversity. I investigated the influence of environment on different ecological and evolutionary processes using a multi-species approach. This gave me the advantage to look at different processes in the same species (e.g. the buff-tailed bumble bee, **Chapter I** and **Chapter IV**), but also compare similar processes between different species (the buff-tailed and white-tailed bumble bee in **Chapter I**, and the house sparrow and buff-tailed bumble bee in **Chapter III** and **IV**).

Here, I showed how two closely related species are differentially distributed across the landscape depending on environmental conditions. I was also able to add to the body of evidence that environment is a key driver in shaping the genetic structure of species. I investigated patterns of coding genetic markers (single-nucleotide polymorphisms (SNPs) located in genes) and those in neutral genetic markers (e.g. microsatellites), presenting a scenario of 'isolation by environment'. I demonstrated methods to map the spatial patterns of biodiversity, such as species distribution modeling (SDM) approaches, and landscape genomic approaches to correlate genetic data with environmental diversity. Mapping these spatial patterns is particularly important for biodiversity conservation, since protection can only succeed contingent upon high-quality data on the presence and abundance of biodiversity.

I additionally evaluated the use of environmental heterogeneity as a conservation surrogate in spatial conservation planning, and assessed its representation for the species-level of biodiversity. I found a rather limited surrogacy performance of environment for species, leading to the question the same holds true for other taxonomic groups, and in particular for the genetic level of biodiversity. Because genotyping many individuals from many locations remains prohibitively expensive and time consuming, an adequate and easy to obtain surrogate for genetic diversity would be extremely valuable.

In nature conservation, ideally all levels of biodiversity are equally well protected (Pressey, 2004). However, because of the complexity of biodiversity and financial and personnel constraints, shortcuts need to be taken (Andelman and Fagan, 2000; Williams et al., 2006). Using conservation surrogates is a commonly used 'shortcut' and shows great potential (Margules and Pressey, 2000; Sarkar et al., 2005). So far, genetic data was rather underrepresented in conservation planning, mainly because of the high costs to generate meaningful and useful data (Frankham et al., 2002; Laikre, 2010). With technological advances, it is getting cheaper to generate genetic data, in particular whole genome data (Imelfort et al., 2009; Storfer et al., 2018). Conservationists should keep in mind that the genetic level and its variation is essential for species to adapt to environmental changes (Barrett and Schluter, 2008; Etterson and Shaw, 2001), and integrating evolutionary processes such as adaptation and gene

flow in spatial conservation planning is predicted to decrease biodiversity loss (Hoffmann and Sgrò, 2011). When levels of standing genetic variation are reduced, species are likely unable to adapt to new conditions, and consequently experience elevated extinction risks (Kawecki and Ebert, 2004; Vincent et al., 2013). Using genomic data can also benefit biodiversity conservation immensely by informing us which populations showed the strongest decline in genetic diversity and therefore might be at imminent risk under ongoing environmental change (Jia et al., 2020; Martins et al., 2018). This knowledge can help to adjust conservation practices and management (Gugger et al., 2018; Harrisson et al., 2014; Segelbacher et al., 2010). In recent years, more and more efforts were made to call attention for the importance of genetic diversity in nature conservation (Hoban et al., 2020; Laikre, 2010). Initiatives such as the Genomic Biodiversity Knowledge for resilient Ecosystems (G-BiKE) Action (https://sites.google.com/fmach.it/g-bike-genetics-eu/home) were established to enable standard tools for monitoring and managing genetic diversity and the related adaptive potential of populations. This can be seen as a first step towards a better and more efficient biodiversity conservation strategy and should be continued in order to find accessible, informative and flexible ways to incorporate the genetic level of biodiversity into biodiversity conservation.

Bibliography

Agency, European Environment (2016). "Mapping and Assessing the Condition of Europe's Ecosystems: Progress and Challenges." In: (EEA).

Ahrens, Collin W, Paul D Rymer, Adam Stow, Jason Bragg, Shannon Dillon, Kate DL Umbers, and Rachael Y Dudaniec (2018). "The search for loci under selection: trends, biases and progress." In: *Molecular Ecology* 27.6, pp. 1342–1356.

Allouche, Omri, Asaf Tsoar, and Ronen Kadmon (2006). "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)." In: *Journal of applied ecology* 43.6, pp. 1223–1232.

Anacker, Brian L and Sharon Y Strauss (2016). "Ecological similarity is related to phylogenetic distance between species in a cross-niche field transplant experiment." In: *Ecology* 97.7, pp. 1807–1818. ISSN: 0012-9658.

Andelman, Sandy J and William F Fagan (2000). "Umbrellas and flagships: efficient conservation surrogates or expensive mistakes?" In: *Proceedings of the National Academy of Sciences* 97.11, pp. 5954–5959. ISSN: 0027-8424.

Anderson, Jill T, Anne Marie Panetta, and Thomas Mitchell-Olds (2012). "Evolutionary and ecological responses to anthropogenic climate change: update on anthropogenic climate change." In: *Plant physiology* 160.4, pp. 1728–1740.

Anderson, Ted R (2006). *Biology of the ubiquitous house sparrow: from genes to populations*. Oxford University Press.

Araújo, MB, CJ Humphries, PJ Densham, R Lampinen, WJM Hagemeijer, AJ Mitchell-Jones, and JP Gasc (2001). "Would environmental diversity be a good surrogate for species diversity?" In: *Ecography* 24.1, pp. 103–110. ISSN: 0906-7590.

Araújo, Miguel B, Jorge M Lobo, and Juan C Moreno (2007). "The effectiveness of Iberian protected areas in conserving terrestrial biodiversity." In: *Conservation Biology* 21.6, pp. 1423–1432. ISSN: 0888-8892.

Arenas Busto, Miguel, Stefano Mona, Audrey Trochet, Anna Sramkova Hanulova, Mathias Currat, Nicolas Ray, Lounes Chikhi, Rita Rasteiro, Dirk S Schmeller, and Laurent Excoffier (2014). "The scaling of genetic diversity in a changing and fragmented world." In:

Arponen, Anni, Risto K Heikkinen, Chris D Thomas, and Atte Moilanen (2005). "The value of biodiversity in reserve selection: representation, species weighting, and benefit functions." In: *Conservation Biology* 19.6, pp. 2009–2014. ISSN: 0888-8892.

Arponen, Anni, Atte Moilanen, and Simon Ferrier (2008). "A successful community-level strategy for conservation prioritization." In: *Journal of Applied* 45.5, pp. 1436–1445. ISSN: 0021-8901.

Balestrieri, Alessandro, Giuseppe Bogliani, Giovanni Boano, Aritz Ruiz-González, Nicola Saino, Stefano Costa, and Pietro Milanesi (2016). "Modelling the distribution of forest-dependent species in human-dominated landscapes: patterns for the pine marten in intensively cultivated lowlands." In: *PloS one* 11.7, e0158203.

Balkenhol, Niko, Rachael Y Dudaniec, Konstantin V Krutovsky, Jeremy S Johnson, David M Cairns, Gernot Segelbacher, Kimberly A Selkoe, Sophie von der Heyden, Ian J Wang, Oliver Selmoni, et al. (2017). "Landscape genomics: understanding relationships between environmental heterogeneity and genomic characteristics of populations." In: *Population genomics*. Springer, pp. 261–322.

Ban-Calefariu, Cristina and Miklós Sárospataki (2007). "Contributions to the knowledge of Bombus and Psithyrus Genera (Apoidea: Apidae) in Romania." In: *Trav. Mus. nat. d'Hist. natur."Grigore Antipa* 1, pp. 239–258.

Barrett, Rowan DH and Dolph Schluter (2008). "Adaptation from standing genetic variation." In: *Trends in ecology & evolution* 23.1, pp. 38–44.

Bartkowski, Bartosz, Nele Lienhoop, and Bernd Hansjürgens (2015). "Capturing the complexity of biodiversity: A critical review of economic valuation studies of biological diversity." In: *Ecological Economics* 113, pp. 1–14.

Baur, Hannes and Christoph Leuenberger (2011). "Analysis of ratios in multivariate morphometry." In: *Systematic Biology* 60.6, pp. 813–825.

Bay, R. A., N. Rose, R. Barrett, L. Bernatchez, C. K. Ghalambor, J. R. Lasky, R. B. Brem, S. R. Palumbi, and P. Ralph (2017). "Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures." In: *Am Nat* 189.5, pp. 463–473. ISSN: 1537-5323 (Electronic) 0003-0147 (Linking). DOI: 10.1086/691233.

Baym, M., S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, and R. Kishony (2015). "Inexpensive multiplexed library preparation for megabase-sized genomes." In: *PLoS One* 10.5, e0128036. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0128036.

Beier, Paul, Patricia Sutcliffe, Jan Hjort, Daniel P Faith, Robert L Pressey, and Fabio Albuquerque (2015). "A review of selection-based tests of abiotic surrogates for species representation." In: *Conservation Biology* 29.3, pp. 668–679. ISSN: 0888-8892.

Bellard, Céline, Cleo Bertelsmeier, Paul Leadley, Wilfried Thuiller, and Franck Courchamp (2012). "Impacts of climate change on the future of biodiversity." In: *Ecology letters* 15.4, pp. 365–377.

Belmaker, Jonathan and Walter Jetz (2015). "Relative roles of ecological and energetic constraints, diversification rates and region history on global species richness gradients." In: *Ecology Letters* 18.6, pp. 563–571.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the Royal statistical society: series B (Methodological)* 57.1, pp. 289–300.

Bennett, Andrew F, Angie Haslem, David C Cheal, Michael F Clarke, Roger N Jones, John D Koehn, P Sam Lake, Linda F Lumsden, Ian D Lunt, Brendan G Mackey, et al. (2009). "Ecological processes: a key element in strategies for nature conservation." In: *Ecological Management & Restoration* 10.3, pp. 192–199.

Benton, Ted (2006). *Bumblebees: the natural history & identification of the species found in Britain*. Vol. 98. Harper Uk. ISBN: 0007174500.

Bertrand, Y., F. Pleijel, and G. W. Rouse (2006). "Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks." In: *Systematics and Biodiversity* 4.2, pp. 149–159. ISSN: 1477-2000 1478-0933. DOI: 10.1017/s1477200005001908.

Bertsch, Andreas (2010). "A phylogenetic framework for the bumblebee species of the subgenus Bombus sensu stricto based on mitochondrial DNA markers, with a short description of the neglected taxon B. minshanicola Bischoff, 1936 n. status.(Hymenoptera: Apidae: Bombus)." In: *Beiträge zur Entomologie= Contributions to Entomology* 60.2, pp. 471–487. ISSN: 2511-6428.

Bertsch, Andreas, Horst Schweer, and Andreas Titze (2004). "Discrimination of the bumblebee species Bombus lucorum, Bombus cryptarum and Bombus magnus by morphological characters and male labial gland secretions (Hymenoptera: Apidae)." In: *Beiträge zur Entomologie= Contributions to Entomology* 54.2, pp. 365–386. ISSN: 2511-6428.

Bijlsma, R and Volker Loeschcke (2012). "Genetic erosion impedes adaptive responses to stressful environments." In: *Evolutionary Applications* 5.2, pp. 117–129.

Billeter, Regula, Jan Liira, Debra Bailey, Rob Bugter, Paul Arens, Isabel Augenstein, Stéphanie Aviron, Jacques Baudry, Roman Bukacek, and Francoise Burel (2008). "Indicators for biodiversity in agricultural landscapes: a pan-European study." In: *Journal of Applied ecology* 45.1, pp. 141–150. ISSN: 0021-8901.

Biodiversity, Economic Value of, National Research Council, et al. (1999). "Perspectives on Biodiversity: Valuing Its Role in an Everchanging World." In:

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." In: *Bioinformatics* 30.15, pp. 2114–2120.

Bommarco, Riccardo, Ola Lundin, Henrik G Smith, and Maj Rundlöf (2011). "Drastic historic shifts in bumble-bee community composition in Sweden." In: *Proceedings of the Royal Society B: Biological Sciences* 279.1727, pp. 309–315. ISSN: 0962-8452.

Bonin, Aurelie (2008). "Population genomics: a new generation of genome scans to bridge the gap with functional genomics." In: *Molecular ecology* 17.16, pp. 3583–3584.

Bonn, Aletta and Kevin J Gaston (2005). "Capturing biodiversity: selecting priority areas for conservation using different criteria." In: *Biodiversity & Conservation* 14.5, pp. 1083–1100. ISSN: 0960-3115.

Bossert, Silas (2014). "Recognition and identification of species in the Bombus lucorum-complex-A review and outlook." In: *bioRxiv*, p. 011379.

Bossert, Silas, Barbara-Amina Gereben-Krenn, Johann Neumayer, Bernhard Schneller, and Harald W Krenn (2016). "The cryptic Bombus lucorum complex (Hymenoptera: Apidae) in Austria: Phylogeny, distribution, habitat usage and a climatic characterization based on COI sequence data." In: *Zool Stud* 55, pp. 2016.55–13.

Bottrill, Madeleine C, Liana N Joseph, Josie Carwardine, Michael Bode, Carly Cook, Edward T Game, Hedley Grantham, Salit Kark, Simon Linke, and Eve McDonald-Madden (2008). "Is conservation triage just smart decision making?" In: *Ecology & evolution* 23.12, pp. 649–654. ISSN: 0169-5347.

Boutin-Ganache, Isabelle, Marco Raposo, Martine Raymond, and Christian F Deschepper (2001). "M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods." In: *Biotechniques* 31.1, pp. 25–28.

Breiman, L, J Friedman, R Olshen, and C Stone (1984). *Classification and regression trees. Pacific Grove, California: Wadsworth*.

Breiman, Leo (1996). "Bagging predictors." In: *Machine learning* 24.2, pp. 123–140.

– (2001). "Random forests." In: *Machine learning* 45.1, pp. 5–32. ISSN: 0885-6125.

Breiner, Frank T, Antoine Guisan, Ariel Bergamini, and Michael P Nobis (2015). "Overcoming limitations of modelling rare species by using ensembles of small models." In: *Methods in Ecology and Evolution* 6.10, pp. 1210–1218.

Breiner, Frank T, Michael P Nobis, Ariel Bergamini, and Antoine Guisan (2018). "Optimizing ensembles of small models for predicting the distribution of species with few occurrences." In: *Methods in Ecology and Evolution* 9.4, pp. 802–808.

Broehan, Gunnar, Tobias Kroeger, Marcé Lorenzen, and Hans Merzendorfer (2013). "Functional analysis of the ATP-binding cassette (ABC) transporter gene family of Tribolium castaneum." In: *BMC genomics* 14.1, p. 6.

Brondizio, Eduardo S, J Settele, S Díaz, and HT Ngo (2019). "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services." In: *IPBES Secretariat: Bonn, Germany*.

Brooks, Thomas M, Russell A Mittermeier, Gustavo AB da Fonseca, Justin Gerlach, Michael Hoffmann, John F Lamoreux, Cristina Goettsch Mittermeier, John D Pilgrim, and Ana SL Rodrigues (2006). "Global biodiversity conservation priorities." In: *science* 313.5783, pp. 58–61.

Brooks, TM, A Cuttelod, DP Faith, J Garcia-Moreno, P Langhammer, and S Pérez-Espona (2015). "Why and how might genetic and phylogenetic diversity be reflected in the identification of key biodiversity areas?" In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1662, p. 20140019.

Brown, Mark JF and Robert J Paxton (2009). "The conservation of bees: a global perspective." In: *Apidologie* 40.3, pp. 410–416. ISSN: 0044-8435.

Capblancq, Thibaut, Keurcien Luu, Michael GB Blum, and Eric Bazin (2018). "Evaluation of redundancy analysis to identify signatures of local adaptation." In: *Molecular Ecology Resources* 18.6, pp. 1223–1233.

Carmel, Yohay and Liron Stoller-Cavari (2006). "Comparing environmental and biological surrogates for biodiversity at a local scale." In: *Israel Journal of Ecology and Evolution* 52.1, pp. 11–27. ISSN: 2224-4662.

Caro, TM and Gillian O'Doherty (1999). "On the use of surrogate species in conservation biology." In: *Conservation biology* 13.4, pp. 805–814. ISSN: 0888-8892.

Carré, Matthieu and Rachid Cheddadi (2017). "Seasonality in long-term climate change." In: *Quaternaire. Revue de l'Association française pour l'étude du Quaternaire* 28.2, pp. 173–177.

Carvell, Claire (2002). "Habitat use and conservation of bumblebees (Bombus spp.) under different grassland management regimes." In: *Biological conservation* 103.1, pp. 33–49. ISSN: 0006-3207.

Casey, LM, H Rebelo, E Rotheray, and D Goulson (2015). "Evidence for habitat and climatic specializations driving the long-term distribution of UK and I rish bumblebees." In: *Diversity and Distributions* 21.8, pp. 864–875. ISSN: 1366-9516.

Caye, Kevin, Basile Jumentier, Johanna Lepeule, and Olivier François (2019). "LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies." In: *Molecular biology and evolution* 36.4, pp. 852–860.

Chapman, RE, J Wang, and AFG Bourke (2003). "Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators." In: *Molecular Ecology* 12.10, pp. 2801–2808. ISSN: 1365-294X.

Chapuis, Marie-Pierre and Arnaud Estoup (2007). "Microsatellite null alleles and estimation of population differentiation." In: *Molecular biology and evolution* 24.3, pp. 621–631.

Chipman, Hugh A, Edward I George, Robert E McCulloch, et al. (2010). "BART: Bayesian additive regression trees." In: *The Annals of Applied Statistics* 4.1, pp. 266–298.

(CoE), Directorate-General for Environment (DG ENV) Europe Council of (2015). Online Database.

Cohen, Jacob (1960). "A coefficient of agreement for nominal scales." In: *Educational and psychological measurement* 20.1, pp. 37–46.

Commission, European (2007). "Natura 2000 barometer." In: *Nature Newsletter* 23, pp. 11–12.

– (2009). *Directive 2009 /147 /EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wildbirds.* Government Document.

– (2020). *EU Biodiversity Strategy for 2030. Bringing nature back into our lives. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.* Government Document.

Conover, David O, Tara A Duffy, and Lyndie A Hice (2009). "The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of countergradient and cogradient variation." In: *Annals of the New York Academy of Sciences* 1168.1, pp. 100–129.

Coordinators, NCBI Resource (2016). *Database resources of the National Center for Biotechnology Information Nucleic Acids Res*.

Corbet, Sarah A, Ingrid H Williams, and Juliet L Osborne (1991). "Bees and the pollination of crops and wild flowers in the European Community." In: *Bee world* 72.2, pp. 47–59. ISSN: 0005-772X.

Costa, Gabriel C, Christina Wolfe, Donald B Shepard, Janalee P Caldwell, and Laurie J Vitt (2008). "Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient." In: *Journal of Biogeography* 35.4, pp. 637–646.

Coyne, JA and HA Orr (2004). "Polyploidy and hybrid speciation." In: *Speciation. Sunderland, MA, USA: Sinauer Associates*, pp. 321–351.

Crutzen, PJ (2006). "The "anthropocene"(pp. 13–18)." In: *Earth system science in the anthropocene. Springer Berlin Germany*.

Cushman, S. A., K. S. McKelvey, B. R. Noon, and K. McGarigal (2010). "Use of abundance of one species as a surrogate for abundance of others." In: *Conserv Biol* 24.3, pp. 830–40. ISSN: 1523-1739 (Electronic) 0888-8892 (Linking). DOI: 10.1111/j.1523-1739.2009.01396.x.

Dallas, Tad A and Alan Hastings (2018). "Habitat suitability estimated by niche models is largely unrelated to species abundance." In: *Global ecology and biogeography* 27.12, pp. 1448–1456.

D'Amen, M, P Bombi, A Campanaro, L Zapponi, MA Bologna, and F Mason (2013). "Protected areas and insect conservation: questioning the effectiveness of N atura 2000 network for saproxylic beetles in I taly." In: *Animal Conservation* 16.4, pp. 370–378. ISSN: 1367-9430.

Davies, Robert W, Jonathan Flint, Simon Myers, and Richard Mott (2016). "Rapid genotype imputation from sequence without reference panels." In: *Nature genetics* 48.8, p. 965.

Dawson, Deborah A, Gavin J Horsburgh, Andrew P Krupa, Ian RK Stewart, Sigrun Skjelseth, Henrik Jensen, Alexander D Ball, Lewis G Spurgin, MARIA-ELENA MANNARELLI, Shinichi Nakagawa, et al. (2012). "Microsatellite resources for Passeridae species: a predicted microsatellite map of the house sparrow Passer domesticus." In: *Molecular ecology resources* 12.3, pp. 501–523.

Dawson, Terence P, Stephen T Jackson, Joanna I House, Iain Colin Prentice, and Georgina M Mace (2011). "Beyond predictions: biodiversity conservation in a changing climate." In: *science* 332.6025, pp. 53–58.

De Laet, Jenny and JD Summers-Smith (2007). "The status of the urban house sparrow Passer domesticus in north-western Europe: a review." In: *Journal of Ornithology* 148.2, pp. 275–278.

De Mazancourt, C, E Johnson, and TG Barraclough (2008). "Biodiversity inhibits species' evolutionary responses to changing environments." In: *Ecology Letters* 11.4, pp. 380–388.

De Meester, Luc, Africa Gómez, Beth Okamura, and Klaus Schwenk (2002). "The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms." In: *Acta oecologica* 23.3, pp. 121–135.

DeWitt, Thomas J, Andrew Sih, and David Sloan Wilson (1998). "Costs and limits of phenotypic plasticity." In: *Trends in ecology & evolution* 13.2, pp. 77–81.

Di Cola, Valeria, Olivier Broennimann, Blaise Petitpierre, Frank T Breiner, Manuela d'Amen, Christophe Randin, Robin Engler, Julien Pottier, Dorothea Pio, Anne Dubuis, et al. (2017). "ecospat: an R package to support spatial analyses and modeling of species niches and distributions." In: *Ecography* 40.6, pp. 774–787.

Di Minin, Enrico and Atte Moilanen (2012). "Empirical evidence for reduced protection levels across biodiversity features from target-based conservation planning." In: *Biological Conservation* 153, pp. 187–191. ISSN: 0006-3207.

– (2014). "Improving the surrogacy effectiveness of charismatic megafauna with well-surveyed taxonomic groups and habitat types." In: *Journal of Applied Ecology* 51.2, pp. 281–288. ISSN: 1365-2664.

DiBattista, Joseph D (2008). "Patterns of genetic variation in anthropogenically impacted populations." In: *Conservation Genetics* 9.1, pp. 141–156.

Dietrich, Michael R (2010). "Microevolution and macroevolution are governed by the same processes." In: *Francisco J. Ayala and Robert Arp*, p. 169.

Duruz, Solange, Natalia Sevane, Oliver Selmoni, Elia Vajana, Kevin Leempoel, Sylvie Stucki, Pablo Orozco-terWengel, Estelle Rochat, Susana Dunner, NEXTGEN Consortium, et al. (2019). "Rapid identification and interpretation of gene–environment associations using the new R. SamBada landscape genomics pipeline." In: *Molecular ecology resources* 19.5, pp. 1355–1365.

Earl, Dent A et al. (2012). "STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method." In: *Conservation genetics resources* 4.2, pp. 359–361.

Ebbert, Mark TW, Mark E Wadsworth, Lyndsay A Staley, Kaitlyn L Hoyt, Brandon Pickett, Justin Miller, John Duce, John SK Kauwe, and Perry G Ridge (2016). "Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches." In: *BMC bioinformatics* 17.7, p. 239. ISSN: 1471-2105.

Elith, Jane and John R Leathwick (2009). "Species distribution models: ecological explanation and prediction across space and time." In: *Annual review of ecology, evolution, and systematics* 40, pp. 677–697. ISSN: 1543-592X.

Ellner, Stephen P, Monica A Geber, and Nelson G Hairston Jr (2011). "Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics." In: *Ecology letters* 14.6, pp. 603–614.

Engelbrecht, Ian, Mark Robertson, Marnus Stoltz, and Johan W Joubert (2016). "Reconsidering environmental diversity (ED) as a biodiversity surrogacy strategy." In: *Biological Conservation* 197, pp. 171–179. ISSN: 0006-3207.

Etterson, Julie R and Ruth G Shaw (2001). "Constraint to adaptive evolution in response to global warming." In: *Science* 294.5540, pp. 151–154.

Evanno, Guillaume, Sebastien Regnaut, and Jérôme Goudet (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." In: *Molecular ecology* 14.8, pp. 2611–2620.

Faircloth, Brant C, Michael G Branstetter, Noor D White, and Seán G Brady (2015). "Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among H ymenoptera." In: *Molecular Ecology Resources* 15.3, pp. 489–501. ISSN: 1755-098X.

Ferrier, Simon (2002). "Mapping spatial pattern in biodiversity for regional conservation planning: where to from here?" In: *Systematic biology* 51.2, pp. 331–363. ISSN: 1076-836X.

Ferrier, Simon, Glenn Manion, Jane Elith, and Karen Richardson (2007). "Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment." In: *Diversity and distributions* 13.3, pp. 252–264.

Ficetola, Gentile Francesco, Anna Bonardi, Caspar A Mücher, Niels LM Gilissen, and Emilio Padoa-Schioppa (2014). "How many predictors in species distribution models at the landscape scale? Land use versus LiDAR-derived canopy height." In: *International Journal of Geographical Information Science* 28.8, pp. 1723–1739. ISSN: 1365-8816.

Fick, Stephen E and Robert J Hijmans (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." In: *International journal of climatology* 37.12, pp. 4302–4315.

Fine, Paul VA (2015). "Ecological and evolutionary drivers of geographic variation in species diversity." In: *Annual Review of Ecology, Evolution, and Systematics* 46, pp. 369–392.

Fitzpatrick, M. C. and S. R. Keller (2015). "Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation." In: *Ecol Lett* 18.1, pp. 1–16. ISSN: 1461-0248 (Electronic) 1461-023X (Linking). DOI: 10.1111/ele.12376.

Fântânâ, Ciprian and István Kovács (2020). "The Romanian breeding bird atlas 2006-2017, a common scheme of Milvus Group Association and the Romanian Ornithological Society." In: *in preparation*.

Fourcade, Yoan, Jan O Engler, Dennis Rödder, and Jean Secondi (2014). "Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias." In: *PloS one* 9.5, e97122.

Franco, Aldina MA, Barbara J Anderson, David B Roy, Simon Gillings, Richard Fox, Atte Moilanen, and Chris D Thomas (2009). "Surrogacy and persistence in reserve selection: landscape prioritization for multiple taxa in Britain." In: *Journal of Applied Ecology* 46.1, pp. 82–91. ISSN: 0021-8901.

Frankel, Otto H (1974). "Genetic conservation: our evolutionary responsibility." In: *Genetics* 78.1, pp. 53–65.

Frankham, Richard (2010). "Challenges and opportunities of genetic approaches to biological conservation." In: *Biological conservation* 143.9, pp. 1919–1927.

Frankham, Richard, Scientist Emeritus Jonathan D Ballou, David A Briscoe, and Jonathan D Ballou (2002). *Introduction to conservation genetics*. Cambridge university press.

Freedman, Adam H, Henri A Thomassen, Wolfgang Buermann, and Thomas B Smith (2010). "Genomic signals of diversification along ecological gradients in a tropical lizard." In: *Molecular ecology* 19.17, pp. 3773–3788.

Frichot, Eric, Sean D Schoville, Guillaume Bouchard, and Olivier François (2013). "Testing for associations between loci and environmental gradients using latent factor mixed models." In: *Molecular biology and evolution* 30.7, pp. 1687–1699.

Gaggiotti, Oscar E, Dorte Bekkevold, Hanne BH Jørgensen, Matthieu Foll, Gary R Carvalho, Carl Andre, and Daniel E Ruzzante (2009). "Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study." In: *Evolution: International Journal of Organic Evolution* 63.11, pp. 2939–2951.

Gammans, N., R. Comont, S.C. Morgan, and G. Perkins (2018). *Bumblebees: An Introduction*. Bumblebee Conservation Trust. ISBN: 9780995773905.

Garnier, Stéphane, Patrick Durand, Céline Arnathau, Ange Marie Risterucci, Rodrigo ESPARZA-SALAS, Elise CELLIER-HOLZEM, and Gabriele Sorci (2009). "New polymorphic microsatellite loci in the house sparrow, Passer domesticus." In: *Molecular Ecology Resources* 9.3, pp. 1063–1065.

Garson, Justin, Anshu Aggarwal, and Sahotra Sarkar (2002). "Birds as surrogates for biodiversity: an analysis of a data set from southern Quebec." In: *Journal of Biosciences* 27.4, pp. 347–360. ISSN: 0250-5991.

Gaston, Kevin J (2000). "Global patterns in biodiversity." In: *Nature* 405.6783, pp. 220–227.

Gaston, Kevin J, Sarah F Jackson, Arnold Nagy, Lisette Cantú-Salazar, and Mark Johnson (2008). "Protected areas in Europe: principle and practice." In: *Annals of the New York Academy of Sciences* 1134.1, pp. 97–119. ISSN: 0077-8923.

Gause, GF (1934). "The struggle for existence. Baltimore: Williams and Wilkins. 163 p." In:

Geue, Julia C and Henri A Thomassen (2020). "Unraveling the habitat preferences of two closely related bumble bee species in Eastern Europe." In: *Ecology and Evolution*. ISSN: 2045-7758.

Geue, Julia C, Csongor I Vágási, Mona Schweizer, Péter L Pap, and Henri A Thomassen (2016). "Environmental selection is a main driver of divergence in house sparrows (Passer domesticus) in Romania and Bulgaria." In: *Ecology and evolution* 6.22, pp. 7954–7964. ISSN: 2045-7758.

Glowka, Lyle, Françoise Burhenne-Guilmin, and Hugh Synge (1994). *Guide to the convention on biological diversity*. IUCN–the World Conservation Union.

Goetz, Scott, Daniel Steinberg, Ralph Dubayah, and Bryan Blair (2007). "Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA." In: *Remote Sensing of Environment* 108.3, pp. 254–263. ISSN: 0034-4257.

Gomes, V. H. F., I. Jff SD, N. Raes, I. L. Amaral, R. P. Salomao, L. de Souza Coelho, F. D. de Almeida Matos, C. V. Castilho, D. de Andrade Lima Filho, D. C. Lopez, J. E. Guevara, W. E. Magnusson, O. L. Phillips, F. Wittmann, M. de Jesus Veiga Carim, M. P. Martins, M. V. Irume, D. Sabatier, J. F. Molino, O. S. Banki, J. R. da Silva Guimaraes, N. C. A. Pitman, M. T. F. Piedade, A. M. Mendoza, B. G. Luize,

E. M. Venticinque, E. M. M. de Leao Novo, P. N. Vargas, T. S. F. Silva, A. G. Manzatto, J. Terborgh, N. F. C. Reis, J. C. Montero, K. R. Casula, B. S. Marimon, B. H. Marimon, E. N. H. Coronado, T. R. Feldpausch, A. Duque, C. E. Zartman, N. C. Arboleda, T. J. Killeen, B. Mostacedo, R. Vasquez, J. Schongart, R. L. Assis, M. B. Medeiros, M. F. Simon, A. Andrade, W. F. Laurance, J. L. Camargo, L. O. Demarchi, S. G. W. Laurance, E. de Sousa Farias, H. E. M. Nascimento, J. D. C. Revilla, A. Quaresma, F. R. C. Costa, I. C. G. Vieira, B. B. L. Cintra, H. Castellanos, R. Brienen, P. R. Stevenson, Y. Feitosa, J. F. Duivenvoorden, C. Ga Aymard, H. F. Mogollon, N. Targhetta, J. A. Comiskey, A. Vicentini, A. Lopes, G. Damasco, N. Davila, R. Garcia-Villacorta, C. Levis, J. Schietti, P. Souza, T. Emilio, A. Alonso, D. Neill, F. Dallmeier, L. V. Ferreira, A. Araujo-Murakami, D. Praia, D. D. do Amaral, F. A. Carvalho, F. C. de Souza, K. Feeley, L. Arroyo, M. P. Pansonato, R. Gribel, B. Villa, J. C. Licona, P. V. A. Fine, C. Ceron, C. Baraloto, E. M. Jimenez, J. Stropp, J. Engel, M. Silveira, et al. (2018). "Species Distribution Modelling: Contrasting presence-only models with plot abundance data." In: *Sci Rep* 8.1, p. 1003. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-017-18927-1.

Goulson, Dave (2010). *Bumblebees: behaviour, ecology, and conservation*. Oxford University Press on Demand. ISBN: 0199553068.

Goulson, Dave and Ben Darvill (2004). "Niche overlap and diet breadth in bumblebees; are rare species more specialized in their choice of flowers?" In: *Apidologie* 35.1, pp. 55–63. ISSN: 0044-8435.

Goulson, Dave, Michael E Hanley, Ben Darvill, JS Ellis, and Mairi E Knight (2005). "Causes of rarity in bumblebees." In: *Biological conservation* 122.1, pp. 1–8. ISSN: 0006-3207.

Goulson, Dave, Gillian C Lye, and Ben Darvill (2008a). "Decline and conservation of bumble bees." In: *Annu. Rev. Entomol.* 53, pp. 191–208. ISSN: 0066-4170.

– (2008b). "Diet breadth, coexistence and rarity in bumblebees." In: *Biodiversity and Conservation* 17.13, pp. 3269–3288. ISSN: 0960-3115.

Goulson, David and Kate R. Sparrow (2008). "Evidence for competition between honeybees and bumblebees; effects on bumblebee worker size." In: *Journal of Insect Conservation* 13.2, pp. 177–181. ISSN: 1366-638X 1572-9753. DOI: 10.1007/s10841-008-9140-y.

Grantham, Hedley S, Robert L Pressey, Jessie A Wells, and Andrew J Beattie (2010). "Effectiveness of biodiversity surrogates for conservation planning: different measures of effectiveness generate a kaleidoscope of variation." In: *PLoS One* 5.7.

Griffith, Simon C, Deborah A Dawson, Henrik Jensen, Nancy Ockendon, Carolyn Greig, Karsten Neumann, and Terry Burke (2007). "Fourteen polymorphic microsatellite loci characterized in the house sparrow Passer domesticus (Passeridae, Aves)." In: *Molecular Ecology Notes* 7.2, pp. 333–336.

Grivet, Delphine, Victoria L Sork, Robert D Westfall, and Frank W Davis (2008). "Conserving the evolutionary potential of California valley oak (Quercus lobata Née):

a multivariate genetic approach to conservation planning." In: *Molecular Ecology* 17.1, pp. 139–156.

Guélat, Jérôme and Marc Kéry (2018). "Effects of spatial autocorrelation and imperfect detection on species distribution models." In: *Methods in Ecology and Evolution* 9.6, pp. 1614–1625.

Gugger, P. F., C. T. Liang, V. L. Sork, P. Hodgskiss, and J. W. Wright (2018). "Applying landscape genomic tools to forest management and restoration of Hawaiian koa (Acacia koa) in a changing environment." In: *Evol Appl* 11.2, pp. 231–242. ISSN: 1752-4571 (Print) 1752-4571 (Linking). DOI: 10.1111/eva.12534.

Guillot, Gilles, Arnaud Estoup, Frédéric Mortier, and Jean François Cosson (2005a). "A spatial statistical model for landscape genetics." In: *Genetics* 170.3, pp. 1261–1280.

Guillot, Gilles, Frédéric Mortier, and Arnaud Estoup (2005b). "GENELAND: a computer package for landscape genetics." In: *Molecular ecology notes* 5.3, pp. 712–715.

Guillot, Gilles, Filipe Santos, and Arnaud Estoup (2008). "Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface." In: *Bioinformatics* 24.11, pp. 1406–1407.

Guisan, Antoine and Wilfried Thuiller (2005). "Predicting species distribution: offering more than simple habitat models." In: *Ecology letters* 8.9, pp. 993–1009.

Guisan, Antoine, Reid Tingley, John B Baumgartner, Ilona Naujokaitis-Lewis, Patricia R Sutcliffe, Ayesha IT Tulloch, Tracey J Regan, Lluis Brotons, Eve McDonald-Madden, Chrystal Mantyka-Pringle, et al. (2013). "Predicting species distributions for conservation decisions." In: *Ecology letters* 16.12, pp. 1424–1435.

Hallmann, Caspar A, Martin Sorg, Eelke Jongejans, Henk Siepel, Nick Hofland, Heinz Schwan, Werner Stenmans, Andreas Müller, Hubert Sumser, and Thomas Hörren (2017). "More than 75 percent decline over 27 years in total flying insect biomass in protected areas." In: *PloS one* 12.10, e0185809. ISSN: 1932-6203.

Hansen, MC, RS DeFries, JRG Townshend, R Sohlberg, C Dimiceli, and M Carroll (2002). "Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data." In: *Remote Sensing of Environment* 83.1-2, pp. 303–319.

Harrisson, Katherine A, Alexandra Pavlova, Marina Telonis-Scott, and Paul Sunnucks (2014). "Using genomics to characterize evolutionary potential for conservation of wild populations." In: *Evolutionary Applications* 7.9, pp. 1008–1025.

Hartl, Günther B, Frank Zachos, and Karl Nadlinger (2003). "Genetic diversity in European red deer (Cervus elaphus L.): anthropogenic influences on natural populations." In: *Comptes Rendus Biologies* 326, pp. 37–42.

Hastie, Trevor J and Robert J Tibshirani (1990). *Generalized additive models*. Vol. 43. CRC press.

Hawkins, Bradford A, Richard Field, Howard V Cornell, David J Currie, Jean-François Guégan, Dawn M Kaufman, Jeremy T Kerr, Gary G Mittelbach, Thierry Oberdorff,

Eileen M O'Brien, et al. (2003). "Energy, water, and broad-scale geographic patterns of species richness." In: *Ecology* 84.12, pp. 3105–3117.

Hebert, Paul DN, Mark Y Stoeckle, Tyler S Zemlak, and Charles M Francis (2004). "Identification of birds through DNA barcodes." In: *Plos biol* 2.10, e312.

Heinrich, Bernd (1993). "Heat Managers.(Book Reviews: The Hot-Blooded Insects. Strategies and Mechanisms of Thermoregulation.)" In: *Science* 260, pp. 1155–1156. ISSN: 0036-8075.

Hendry, Andrew P, Thomas J Farrugia, and Michael T Kinnison (2008). "Human influences on rates of phenotypic change in wild animal populations." In: *Molecular ecology* 17.1, pp. 20–29.

Hendry, Andrew Paul and Michael T Kinnison (2001). "An introduction to microevolution: rate, pattern, process." In: *Genetica* 112.1, pp. 1–8.

Henriques, Dora, Andreas Wallberg, Julio Chávez-Galarza, J Spencer Johnston, Matthew T Webster, and M Alice Pinto (2018). "Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula." In: *Scientific reports* 8.1, pp. 1–14.

Hernández, Belinda, Adrian E Raftery, Stephen R Pennington, and Andrew C Parnell (2018). "Bayesian additive regression trees using Bayesian model averaging." In: *Statistics and computing* 28.4, pp. 869–890.

Herrera, José M, Emilie F Ploquin, Javier Rodríguez-Pérez, and José R Obeso (2014). "Determining habitat suitability for bumblebees in a mountain system: a baseline approach for testing the impact of climate change on the occurrence and abundance of species." In: *Journal of biogeography* 41.4, pp. 700–712. ISSN: 0305-0270.

Hess, George R, Rebecca A Bartel, Allison K Leidner, Kristen M Rosenfeld, Matthew J Rubino, Sunny B Snider, and Taylor H Ricketts (2006). "Effectiveness of biodiversity indicators varies with extent, grain, and region." In: *Biological Conservation* 132.4, pp. 448–457. ISSN: 0006-3207.

Hijmans, Robert J, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis (2005). "Very high resolution interpolated climate surfaces for global land areas." In: *International Journal of Climatology: A Journal of the Royal Meteorological Society* 25.15, pp. 1965–1978.

Hirzel, Alexandre H, Gwenaëlle Le Lay, Véronique Helfer, Christophe Randin, and Antoine Guisan (2006). "Evaluating the ability of habitat suitability models to predict species presences." In: *Ecological modelling* 199.2, pp. 142–152.

Hoban, Sean, Michael Bruford, Josephine D'Urban Jackson, Margarida Lopes-Fernandes, Myriam Heuertz, Paul A Hohenlohe, Ivan Paz-Vinas, Per Sjögren-Gulve, Gernot Segelbacher, Cristiano Vernesi, et al. (2020). "Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved." In: *Biological Conservation* 248, p. 108654.

Hoban, Sean, Joanna L Kelley, Katie E Lotterhos, Michael F Antolin, Gideon Bradburd, David B Lowry, Mary L Poss, Laura K Reed, Andrew Storfer, and Michael C

Whitlock (2016). "Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions." In: *The American Naturalist* 188.4, pp. 379–397.

Hoffmann, Ary A and Carla M Sgrò (2011). "Climate change and evolutionary adaptation." In: *Nature* 470.7335, p. 479. ISSN: 1476-4687.

Hoffmann, Michael, Craig Hilton-Taylor, Ariadne Angulo, Monika Böhm, Thomas M Brooks, Stuart HM Butchart, Kent E Carpenter, Janice Chanson, Ben Collen, Neil A Cox, et al. (2010). "The impact of conservation on the status of the world's vertebrates." In: *science* 330.6010, pp. 1503–1509.

Holand, Anna M, Henrik Jensen, Jarle Tufto, and Rune Moe (2011). "Does selection or genetic drift explain geographic differentiation of morphological characters in house sparrows Passer domesticus?" In: *Genetics Research* 93.5, pp. 367–379.

Holderegger, Rolf, Urs Kamm, and Felix Gugerli (2006). "Adaptive vs. neutral genetic diversity: implications for landscape genetics." In: *Landscape Ecology* 21.6, pp. 797–807.

Hole, David G, Mark J Whittingham, Richard B Bradbury, Guy QA Anderson, Patricia LM Lee, Jeremy D Wilson, and John R Krebs (2002). "Widespread local house-sparrow extinctions." In: *Nature* 418.6901, pp. 931–932.

Howard, Christine, Philip A Stephens, James W Pearce-Higgins, Richard D Gregory, and Stephen G Willis (2014). "Improving species distribution models: the value of data on abundance." In: *Methods in Ecology and Evolution* 5.6, pp. 506–513.

Hubbell, Stephen P (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. Princeton University Press.

Huelsenbeck, John P and Fredrik Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." In: *Bioinformatics* 17.8, pp. 754–755. ISSN: 1460-2059.

Hunt, Sarah E, William McLaren, Laurent Gil, Anja Thormann, Helen Schuilenburg, Dan Sheppard, Andrew Parton, Irina M Armean, Stephen J Trevanion, Paul Flicek, et al. (2018). "Ensembl variation resources." In: *Database* 2018.

Hunter, Malcolm, Martin Westgate, Philip Barton, Aram Calhoun, Jennifer Pierson, Ayesha Tulloch, Maria Beger, Cristina Branquinho, Tim Caro, John Gross, Jani Heino, Peter Lane, Catherine Longo, Kathy Martin, William H. McDowell, Camille Mellin, Hanna Salo, and David Lindenmayer (2016). "Two roles for ecological surrogacy: Indicator surrogates and management surrogates." In: *Ecological Indicators* 63, pp. 121–125. ISSN: 1470160X. DOI: 10.1016/j.ecolind.2015.11.049.

Imelfort, Michael, Jacqueline Batley, Sean Grimmond, and David Edwards (2009). "Genome sequencing approaches and successes." In: *Plant Genomics*. Springer, pp. 345–358.

Iojă, Cristian Ioan, Maria Pătroescu, Laurenţiu Rozylowicz, Viorel D Popescu, Mircea Vergheleţ, Mihai Iancu Zotta, and Mihaela Felciuc (2010). "The efficacy of Romania's protected areas network in conserving biodiversity." In: *Biological Conservation* 143.11, pp. 2468–2476. ISSN: 0006-3207.

Ioras, Florin (2003). "Trends in Romanian biodiversity conservation policy." In: *Biodiversity & Conservation* 12.1, pp. 9–23. ISSN: 0960-3115.

Jaffé, Rodolfo, Jamille C Veiga, Nathaniel S Pope, Éder CM Lanes, Carolina S Carvalho, Ronnie Alves, Sónia CS Andrade, Maria C Arias, Vanessa Bonatti, Airton T Carvalho, et al. (2019). "Landscape genomics to the rescue of a tropical bee threatened by habitat loss and climate change." In: *Evolutionary applications* 12.6, pp. 1164–1177.

Jensen, Henrik, Rune Moe, Ingerid Julie Hagen, Anna Marie Holand, Jaana Kekkonen, Jarle Tufto, and Bernt-Erik Sæther (2013). "Genetic variation and structure of house sparrow populations: is there an island effect?" In: *Molecular Ecology* 22.7, pp. 1792–1805.

Jetz, Walter, Jana M McPherson, and Robert P Guralnick (2012). "Integrating biodiversity distribution knowledge: toward a global map of life." In: *Trends in ecology & evolution* 27.3, pp. 151–159.

Jia, Kai-Hua, Wei Zhao, Paul Andrew Maier, Xian-Ge Hu, Yuqing Jin, Shan-Shan Zhou, Si-Qian Jiao, Yousry A El-Kassaby, Tongli Wang, Xiao-Ru Wang, et al. (2020). "Landscape genomics predicts climate change-related genetic offset for the widespread Platycladus orientalis (Cupressaceae)." In: *Evolutionary applications* 13.4, pp. 665–676.

Joost, S., A. Bonin, M. W. Bruford, L. Despres, C. Conord, G. Erhardt, and P. Taberlet (2007). "A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation." In: *Mol Ecol* 16.18, pp. 3955–69. ISSN: 0962-1083 (Print) 0962-1083 (Linking). DOI: 10.1111/j.1365-294X.2007.03442.x.

Joseph, L. N., R. F. Maloney, and H. P. Possingham (2009). "Optimal allocation of resources among threatened species: a project prioritization protocol." In: *Conserv Biol* 23.2, pp. 328–38. ISSN: 1523-1739 (Electronic) 0888-8892 (Linking). DOI: 10.1111/j.1523-1739.2008.01124.x.

Joshi, J, B Schmid, MC Caldeira, PG Dimitrakopoulos, J Good, R Harris, A Hector, K Huss-Danell, A Jumpponen, A Minns, et al. (2001). "Local adaptation enhances performance of common plant species." In: *Ecology Letters* 4.6, pp. 536–544.

Juutinen, Artti and Mikko Mönkkönen (2004). "Testing alternative indicators for biodiversity conservation in old-growth boreal forests: ecology and economics." In: *Ecological Economics* 50.1-2, pp. 35–48. ISSN: 0921-8009.

Kadoya, Taku, Hiroshi S Ishii, Reina Kikuchi, Shin-ichi Suda, and Izumi Washitani (2009). "Using monitoring data gathered by volunteers to predict the potential distribution of the invasive alien bumblebee Bombus terrestris." In: *Biological Conservation* 142.5, pp. 1011–1017. ISSN: 0006-3207.

Kawecki, Tadeusz J and Dieter Ebert (2004). "Conceptual issues in local adaptation." In: *Ecology letters* 7.12, pp. 1225–1241.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, Alex Cooper, Sidney Markowitz, and Chris Duran (2012). "Geneious Basic: an integrated and extendable desktop software

platform for the organization and analysis of sequence data." In: *Bioinformatics* 28.12, pp. 1647–1649. ISSN: 1460-2059.

Keitt, Timothy H, Ottar N Bjørnstad, Philip M Dixon, and Steve Citron-Pousty (2002). "Accounting for spatial pattern when modeling organism-environment interactions." In: *Ecography* 25.5, pp. 616–625.

Kekkonen, Jaana, P Seppä, IK Hanski, H Jensen, RA Väisänen, and JE Brommer (2011). "Low genetic differentiation in a sedentary bird: house sparrow population genetics in a contiguous landscape." In: *Heredity* 106.1, pp. 183–190.

Kells, Andrea R, John M Holland, and Dave Goulson (2001). "The value of uncropped field margins for foraging bumblebees." In: *Journal of Insect Conservation* 5.4, pp. 283–291. ISSN: 1366-638X.

Kennedy, Christina M, Eric Lonsdorf, Maile C Neel, Neal M Williams, Taylor H Ricketts, Rachael Winfree, Riccardo Bommarco, Claire Brittain, Alana L Burley, Daniel Cariveau, et al. (2013). "A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems." In: *Ecology letters* 16.5, pp. 584–599.

Kerr, Jeremy T, Alana Pindar, Paul Galpern, Laurence Packer, Simon G Potts, Stuart M Roberts, Pierre Rasmont, Oliver Schweiger, Sheila R Colla, and Leif L Richardson (2015). "Climate change impacts on bumblebees converge across continents." In: *Science* 349.6244, pp. 177–180. ISSN: 0036-8075.

Kosior, Andrzej, Waldemar Celary, Paweł Olejniczak, Jan Fijał, Wieslaw Król, Wojciech Solarz, and Piotr Płonka (2007). "The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe." In: *Oryx* 41.1, pp. 79–88. ISSN: 1365-3008.

Kraus, F. B., S. Wolf, and R. F. Moritz (2009). "Male flight distance and population substructure in the bumblebee Bombus terrestris." In: *J Anim Ecol* 78.1, pp. 247–52. ISSN: 1365-2656 (Electronic) 0021-8790 (Linking). DOI: 10.1111/j.1365-2656.2008.01479.x.

Kremen, C, A Cameron, A Moilanen, SJ Phillips, CD Thomas, H Beentje, J Dransfield, BL Fisher, F Glaw, and TC Good (2008). "Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools." In: *Science* 320.5873, pp. 222–226. ISSN: 0036-8075.

Kremer, Antoine, Ophélie Ronce, Juan J Robledo-Arnuncio, Frédéric Guillaume, Gil Bohrer, Ran Nathan, Jon R Bridle, Richard Gomulkiewicz, Etienne K Klein, Kermit Ritland, et al. (2012). "Long-distance gene flow and adaptation of forest trees to rapid climate change." In: *Ecology letters* 15.4, pp. 378–392.

Kriticos, Darren J, Bruce L Webber, Agathe Leriche, Noboru Ota, Ian Macadam, Janice Bathols, and John K Scott (2012). "CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling." In: *Methods in Ecology and Evolution* 3.1, pp. 53–64.

Kruszewicz, AG, AH Kruszewicz, R Pawiak, and M Mazurkiewicz (1995). "The international symposium of the working group on granivorous birds–granivorous

birds as agricultural pests and epidemiological vectors. Bacteria in house sparrow (Passer domesticus) and tree sparrow (Passer montanus) nestlings: Occurrence and influence of growth and mortality." In: *Nestling mortality of granivorous birds due to microorganisms and toxic substances: Synthesis*, pp. 267–280.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura (2018). "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." In: *Molecular biology and evolution* 35.6, pp. 1547–1549. ISSN: 0737-4038.

Laikre, Linda (2010). "Genetic diversity is overlooked in international conservation policy implementation." In: *Conservation Genetics* 11.2, pp. 349–354.

Larkin, Mark A, Gordon Blackshields, Nigel P Brown, R Chenna, Paul A McGettigan, Hamish McWilliam, Franck Valentin, Iain M Wallace, Andreas Wilm, Rodrigo Lopez, et al. (2007). "Clustal W and Clustal X version 2.0." In: *bioinformatics* 23.21, pp. 2947–2948.

Larsen, Frank Wugt, Jesper Bladt, Andrew Balmford, and Carsten Rahbek (2012). "Birds as biodiversity surrogates: will supplementing birds with other taxa improve effectiveness?" In: *Journal of Applied Ecology* 49.2, pp. 349–356. ISSN: 0021-8901.

Lawler, Joshua J, Denis White, Jean C Sifneos, and Lawrence L Master (2003). "Rare species and the use of indicator groups for conservation planning." In: *Conservation biology* 17.3, pp. 875–882. ISSN: 0888-8892.

Lecocq, Thomas, Nicolas J Vereecken, Denis Michez, Simon Dellicour, Patrick Lhomme, Irena Valterova, Jean-Yves Rasplus, and Pierre Rasmont (2013). "Patterns of genetic and reproductive traits differentiation in mainland vs. Corsican populations of bumblebees." In: *PLoS One* 8.6, e65642. ISSN: 1932-6203.

Lehtomäki, Joona (2016-2018). ""zonator R package."" In: R package version 0.6.0.

Lehtomäki, Joona and Atte Moilanen (2013). "Methods and workflow for spatial conservation prioritization using Zonation." In: *Environmental Modelling & Software* 47, pp. 128–137. ISSN: 1364-8152.

Leigh, Deborah M, Andrew P Hendry, Ella Vázquez-Domínguez, and Vicki L Friesen (2019). "Estimated six per cent loss of genetic variation in wild populations since the industrial revolution." In: *Evolutionary Applications* 12.8, pp. 1505–1512.

Lewis, Simon L and Mark A Maslin (2015). "Defining the anthropocene." In: *Nature* 519.7542, pp. 171–180.

Li, H. (2011). "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data." In: *Bioinformatics* 27.21, pp. 2987–93. ISSN: 1367-4811 (Electronic) 1367-4803 (Linking). DOI: 10.1093/bioinformatics/btr509.

Li, H. and R. Durbin (2010). "Fast and accurate long-read alignment with Burrows-Wheeler transform." In: *Bioinformatics* 26.5, pp. 589–95. ISSN: 1367-4811 (Electronic) 1367-4803 (Linking). DOI: 10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin (2009). "The sequence alignment/map format and SAMtools." In: *Bioinformatics* 25.16, pp. 2078–2079.

Li, Jingchun, Jen-Pen Huang, Jeet Sukumaran, and L Lacey Knowles (2018). "Microevolutionary processes impact macroevolutionary patterns." In: *BMC evolutionary biology* 18.1, pp. 1–8.

Liker, A, Z Papp, V Bókony, and AZ Lendvai (2008). "Lean birds in the city: body size and condition of house sparrows along the urbanization gradient." In: *Journal of animal ecology* 77.4, pp. 789–795.

Lima, Marcos R, Regina HF Macedo, Thais LF Martins, Aaron W Schrey, Lynn B Martin, and Staffan Bensch (2012). "Genetic and morphometric divergence of an invasive bird: the introduced house sparrow (Passer domesticus) in Brazil." In: *PLoS One* 7.12, e53332.

Link, Vivian, Athanasios Kousathanas, Krishna Veeramah, Christian Sell, Amelie Scheu, and Daniel Wegmann (2017). "ATLAS: Analysis Tools for Low-depth and Ancient Samples." In: *Bioinformatics*. DOI: 10.1101/105346.

Liu, Yang, Simone Webber, Katharine Bowgen, Lucie Schmaltz, Katharine Bradley, Peter Halvarsson, Mohanad Abdelgadir, and Michael Griesser (2013). "Environmental factors influence both abundance and genetic diversity in a widespread bird species." In: *Ecology and evolution* 3.14, pp. 4683–4695.

Lombard, Amanda T, Richard M Cowling, Robert L Pressey, and Anthony G Rebelo (2003). "Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region." In: *Biological Conservation* 112.1-2, pp. 45–62. ISSN: 0006-3207.

Long, David G, Mark R Drinkwater, Benjamin Holt, Sasan Saatchi, and Cheryl Bertoia (2001). "Global ice and land climate studies using scatterometer image data." In: *Eos, Transactions American Geophysical Union* 82.43, pp. 503–503.

Lund, Mette Palitzsch and Carsten Rahbek (2002). "Cross-taxon congruence in complementarity and conservation of temperate biodiversity." In: *Animal Conservation forum*. Vol. 5. 2. Cambridge University Press, pp. 163–171.

Mac Nally, Ralph, Andrew F Bennett, Geoff W Brown, Linda F Lumsden, Alan Yen, Simon Hinkley, Peter Lillywhite, and Darren Ward (2002). "How well do ecosystem-based planning units represent different components of biodiversity?" In: *Ecological applications* 12.3, pp. 900–912. ISSN: 1939-5582.

MacGregor-Fors, Ian, Lorena Morales-Pérez, Javier Quesada, and Jorge E Schondube (2010). "Relationship between the presence of House Sparrows (Passer domesticus) and Neotropical bird community structure and diversity." In: *Biological invasions* 12.1, p. 87.

Manel, Stephanie, Stephane Joost, Bryan K Epperson, Rolf Holderegger, Andrew Storfer, Michael S Rosenberg, Kim T Scribner, Aurelie Bonin, and MARIE-JOSÉE FORTIN (2010). "Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field." In: *Molecular Ecology* 19.17, pp. 3760–3772.

Manel, Stéphanie, Michael K Schwartz, Gordon Luikart, and Pierre Taberlet (2003). "Landscape genetics: combining landscape ecology and population genetics." In: *Trends in ecology & evolution* 18.4, pp. 189–197.

Manion, G, M Lisk, S Ferrier, D Nieto-Lugilde, and MC Fitzpatrick (2016). "gdm: Functions for generalized dissimilarity modeling." In: *Rpackage version* 1.7.

Manolache, Steluta, Andreea Nita, Tibor Hartel, Iulia V Miu, Cristiana M Ciocanea, and Laurentiu Rozylowicz (2020). "Governance networks around grasslands with contrasting management history." In: *bioRxiv*.

Margules, Chris R and Robert L Pressey (2000). "Systematic conservation planning." In: *Nature* 405.6783, p. 243. ISSN: 1476-4687.

Margules, Christopher R, RL Pressey, and PH Williams (2002). "Representing biodiversity: data and procedures for identifying priority areas for conservation." In: *Journal of biosciences* 27.4, pp. 309–326. ISSN: 0250-5991.

Marmion, Mathieu, Miska Luoto, Risto K Heikkinen, and Wilfried Thuiller (2009). "The performance of state-of-the-art modelling techniques depends on geographical distribution of species." In: *Ecological Modelling* 220.24, pp. 3512–3520.

Martins, Karina, Paul F Gugger, Jesus Llanderal-Mendoza, Antonio González-Rodríguez, Sorel T Fitz-Gibbon, Jian-Li Zhao, Hernando Rodríguez-Correa, Ken Oyama, and Victoria L Sork (2018). "Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa." In: *Evolutionary Applications* 11.10, pp. 1842–1858.

Matala, Andrew P, Michael W Ackerman, Matthew R Campbell, and Shawn R Narum (2014). "Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes." In: *Evolutionary Applications* 7.6, pp. 682–701.

McIntosh, Emma J, Robert L Pressey, Samuel Lloyd, Robert J Smith, and Richard Grenyer (2017). "The impact of systematic conservation planning." In: *Annual Review of Environment and Resources* 42.

McRae, Brad H (2006). "Isolation by resistance." In: *Evolution* 60.8, pp. 1551–1561.

Meek, Bill, Dick Loxton, Tim Sparks, Richard Pywell, Heather Pickett, and Marek Nowakowski (2002). "The effect of arable field margin composition on invertebrate biodiversity." In: *Biological Conservation* 106.2, pp. 259–271. ISSN: 0006-3207.

Mi, Chunrong, Falk Huettmann, Rui Sun, and Yumin Guo (2017). "Combining occurrence and abundance distribution models for the conservation of the Great Bustard." In: *PeerJ* 5, e4160.

Miller-Struttmann, Nicole E, Jennifer C Geib, James D Franklin, Peter G Kevan, Ricardo M Holdo, Diane Ebert-May, Austin M Lynn, Jessica A Kettenbach, Elizabeth Hedrick, and Candace Galen (2015). "Functional mismatch in a bumble bee pollination mutualism under climate change." In: *Science* 349.6255, pp. 1541–1544.

Miraldo, Andreia, Sen Li, Michael K Borregaard, Alexander Flórez-Rodríguez, Shyam Gopalakrishnan, Mirnesa Rizvanovic, Zhiheng Wang, Carsten Rahbek, Katharine

A Marske, and David Nogués-Bravo (2016). "An Anthropocene map of genetic diversity." In: *Science* 353.6307, pp. 1532–1535.

Mitchell, Matthew W, Sabrina Locatelli, Paul R Sesink Clee, Henri A Thomassen, and Mary Katherine Gonder (2015). "Environmental variation and rivers govern the structure of chimpanzee genetic diversity in a biodiversity hotspot." In: *BMC Evolutionary Biology* 15.1, p. 1.

Mitchell-Olds, Thomas, John H Willis, and David B Goldstein (2007). "Which evolutionary processes influence natural genetic variation for phenotypic traits?" In: *Nature Reviews Genetics* 8.11, pp. 845–856.

Mittelbach, Gary G, Douglas W Schemske, Howard V Cornell, Andrew P Allen, Jonathan M Brown, Mark B Bush, Susan P Harrison, Allen H Hurlbert, Nancy Knowlton, Harilaos A Lessios, et al. (2007). "Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography." In: *Ecology letters* 10.4, pp. 315–331.

Miu, Iulia V, Gabriel B Chisamera, Viorel D Popescu, Ruben Iosif, Andreea Nita, Steluta Manolache, Viorel D Gavril, Ioana Cobzaru, and Laurentiu Rozylowicz (2018). "Conservation priorities for terrestrial mammals in Dobrogea Region, Romania." In: *ZooKeys* 792, p. 133.

Miu, Iulia V, Laurentiu Rozylowicz, Viorel D Popescu, and Paulina Anastasiu (2020). "Identification of areas of very high biodiversity value to achieve the EU Biodiversity Strategy for 2030 key commitments." In: *PeerJ* 8, e10067.

Moilanen, A, FM Pouzols, L Meller, V Veach, A Arponen, J Leppänen, and H Kujala (2014). "Zonation–Spatial Conservation Planning Methods and Software. Version 4. User Manual." In: *University of Helsinki: Finland*.

Moilanen, Atte, Barbara J Anderson, Felix Eigenbrod, Andreas Heinemeyer, David B Roy, Simon Gillings, Paul R Armsworth, Kevin J Gaston, and Chris D Thomas (2011). "Balancing alternative land uses in conservation prioritization." In: *Ecological Applications* 21.5, pp. 1419–1426. ISSN: 1939-5582.

Moilanen, Atte, Aldina MA Franco, Regan I Early, Richard Fox, Brendan Wintle, and Chris D Thomas (2005). "Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems." In: *Proceedings of the Royal Society B: Biological Sciences* 272.1575, pp. 1885–1891. ISSN: 0962-8452.

Moilanen, Atte and Heini Kujala (2006). "Zonation: Spatial conservation planning framework and software v. 1.0 [user manual]." In:

Moore, Joslin L, Andrew Balmford, Thomas Brooks, Neil D Burgess, Louis A Hansen, Carsten Rahbek, and Paul H Williams (2003). "Performance of sub-Saharan vertebrates as indicator groups for identifying priority areas for conservation." In: *Conservation Biology* 17.1, pp. 207–218. ISSN: 0888-8892.

Moreno-Rueda, Gregorio and Herbert Hoi (2012). "Female house sparrows prefer big males with a large white wing bar and fewer feather holes caused by chewing lice." In: *Behavioral Ecology* 23.2, pp. 271–277.

Moritz, Craig, Karen S Richardson, Simon Ferrier, Geoffrey B Monteith, John Stanisic, Stephen E Williams, and Trevor Whiffin (2001). "Biogeographical concordance and

efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota." In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1479, pp. 1875–1881. ISSN: 0962-8452.

Mosca, ELENA, AJ Eckert, EA Di Pierro, D Rocchini, Nicola La Porta, Piero Belletti, and DB Neale (2012). "The geographical and environmental determinants of genetic diversity for four alpine conifers of the E uropean A lps." In: *Molecular ecology* 21.22, pp. 5530–5545.

Murgui, Enrique and Ana Macias (2010). "Changes in the House Sparrow Passer domesticus population in Valencia (Spain) from 1998 to 2008." In: *Bird Study* 57.3, pp. 281–288.

Murray, Tomás E., Úna Fitzpatrick, Mark J. F. Brown, and Robert J. Paxton (2007). "Cryptic species diversity in a widespread bumble bee complex revealed using mitochondrial DNA RFLPs." In: *Conservation Genetics* 9.3, pp. 653–666. ISSN: 1566-0621 1572-9737. DOI: 10.1007/s10592-007-9394-z.

Myers, Norman, Russell A Mittermeier, Cristina G Mittermeier, Gustavo AB Da Fonseca, and Jennifer Kent (2000). "Biodiversity hotspots for conservation priorities." In: *Nature* 403.6772, pp. 853–858.

Myneni, Ranga B, S Hoffman, Yuri Knyazikhin, JL Privette, J Glassy, Yuhong Tian, Y Wang, X Song, Y Zhang, GR Smith, et al. (2002). "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data." In: *Remote sensing of environment* 83.1-2, pp. 214–231.

Niculae, Mihaiță Iulian, Sorin Avram, Gabriel Ovidiu Vânău, and Maria Pătroescu (2017). "Effectiveness of Natura 2000 network in Romanian Alpine Biogeographical Region: an assessment based on forest landscape connectivity." In: *Annals of Forest Research* 60.1, pp. 19–32. ISSN: 2065-2445.

Nita, Andreea, Tibor Hartel, Steluta Manolache, Cristiana M Ciocanea, Iulia V Miu, and Laurentiu Rozylowicz (2019). "Who is researching biodiversity hotspots in Eastern Europe? A case study on the grasslands in Romania." In: *Plos one* 14.5, e0217638.

Nosil, Patrik, Daniel J Funk, and DANIEL Ortiz-Barrientos (2009). "Divergent selection and heterogeneous genomic divergence." In: *Molecular ecology* 18.3, pp. 375–402.

Noss, Reed F (1990). "Indicators for monitoring biodiversity: a hierarchical approach." In: *Conservation biology* 4.4, pp. 355–364. ISSN: 0888-8892.

Novotny, Vojtech, Pavel Drozd, Scott E Miller, Miroslav Kulfan, Milan Janda, Yves Basset, and George D Weiblen (2006). "Why are there so many species of herbivorous insects in tropical rainforests?" In: *science* 313.5790, pp. 1115–1118.

Okonechnikov, Konstantin, Olga Golosova, Mikhail Fursov, and Ugene Team (2012). "Unipro UGENE: a unified bioinformatics toolkit." In: *Bioinformatics* 28.8, pp. 1166–1167.

Oliver, Ian, Andrew Holmes, J Mark Dangerfield, Michael Gillings, Anthony J Pik, David R Britton, Marita Holley, Margaret E Montgomery, Madeline Raison, and

Vicki Logan (2004). "Land systems as surrogates for biodiversity in conservation planning." In: *Ecological Applications* 14.2, pp. 485–503. ISSN: 1939-5582.

Oostra, Vicencio, Marjo Saastamoinen, Bas J Zwaan, and Christopher W Wheat (2018). "Strong phenotypic plasticity limits potential for evolutionary responses to climate change." In: *Nature communications* 9.1, pp. 1–11.

OpenBirdMaps (2018). *An online database for bird distribution and abundance in Romania, powered by Milvus Group Association, Târgu Mureș, Romania*.

Opermanis, Otars, Brian MacSharry, Jerome Bailly-Maitre, Douglas Evans, and Zelmira Sipkova (2014). "The role of published information in reviewing conservation objectives for Natura 2000 protected areas in the European Union." In: *Environmental management* 53.3, pp. 702–712. ISSN: 0364-152X.

Ornitodata (2018).

Orsini, Luisa, Joost Vanoverbeke, Ine Swillen, Joachim Mergeay, and Luc De Meester (2013). "Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization." In: *Molecular ecology* 22.24, pp. 5983–5999.

Owen, Emily L, Jeffrey S Bale, and Scott AL Hayward (2013). "Can winter-active bumblebees survive the cold? Assessing the cold tolerance of Bombus terrestris audax and the effects of pollen feeding." In: *PLoS One* 8.11, e80061.

Palumbi, Stephen R (2001). "Humans as the world's greatest evolutionary force." In: *Science* 293.5536, pp. 1786–1790.

Panagos, Panos, Marc Van Liedekerke, Arwyn Jones, and Luca Montanarella (2012). "European Soil Data Centre: Response to European policy support and public data requirements." In: *Land use policy* 29.2, pp. 329–338. ISSN: 0264-8377.

Paradis, Emmanuel, Stephen R Baillie, William J Sutherland, and Richard D Gregory (1998). "Patterns of natal and breeding dispersal in birds." In: *Journal of Animal Ecology* 67.4, pp. 518–536.

Parisod, Christian and Rolf Holderegger (2012). "Adaptive landscape genetics: pitfalls and benefits." In: *Molecular ecology* 21.15, pp. 3644–3646.

Parmesan, Camille (2006). "Ecological and evolutionary responses to recent climate change." In: *Annu. Rev. Ecol. Evol. Syst.* 37, pp. 637–669.

Pavé, Alain (2007). "Necessity of chance: biological roulettes and biodiversity." In: *Comptes rendus biologies* 330.3, pp. 189–198.

Peakall, ROD and Peter E Smouse (2006). "GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research." In: *Molecular ecology notes* 6.1, pp. 288–295.

Pearson, Richard G (2007). "Species' distribution modeling for conservation educators and practitioners." In: *Synthesis. American Museum of Natural History* 50.

Pearson, Richard G and Terence P Dawson (2003). "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" In: *Global ecology and biogeography* 12.5, pp. 361–371.

Pechanec, Vilém, Ivo Machar, Tomáš Pohanka, Zdeněk Opršal, Frantisek Petrovič, Juraj Švajda, Lubomír Šálek, Karel Chobot, Jarmila Filippovová, and Pavel Cudlín (2018). "Effectiveness of Natura 2000 system for habitat types protection: A case study from the Czech Republic." In: *Nature Conservation* 24, p. 21. ISSN: 1314-3301.

Pedersen, Vest (2002). "European bumblebees (Hymenoptera: Bombini)-phylogenetic relationships inferred from DNA sequences." In: *Insect Systematics & Evolution* 33.4, pp. 361–386.

Penado, Andreia, Hugo Rebelo, and Dave Goulson (2016). "Spatial distribution modelling reveals climatically suitable areas for bumblebees in undersampled parts of the Iberian Peninsula." In: *Insect Conservation and Diversity* 9.5, pp. 391–401. ISSN: 1752-458X.

Pereira, Henrique M, Paul W Leadley, Vânia Proença, Rob Alkemade, Jörn PW Scharlemann, Juan F Fernandez-Manjarrés, Miguel B Araújo, Patricia Balvanera, Reinette Biggs, William WL Cheung, et al. (2010). "Scenarios for global biodiversity in the 21st century." In: *Science* 330.6010, pp. 1496–1501.

Pereira, Henrique Miguel, Simon Ferrier, M Walters, Gary N Geller, RHG Jongman, Robert J Scholes, Michael William Bruford, N Brummitt, SHM Butchart, AC Cardoso, et al. (2013). "Essential biodiversity variables." In: *Science* 339.6117, pp. 277–278.

Pereira, Henrique Miguel, Laetitia Marie Navarro, and Inês Santos Martins (2012). "Global biodiversity change: the bad, the good, and the unknown." In: *Annual Review of Environment and Resources* 37.

Peters, Marcell K, Andreas Hemp, Tim Appelhans, Christina Behler, Alice Classen, Florian Detsch, Andreas Ensslin, Stefan W Ferger, Sara B Frederiksen, Friederike Gebert, et al. (2016). "Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level." In: *Nature Communications* 7.1, pp. 1–11.

Phillips, SJ, M Dudík, and RE Schapire (2017). "Maxent software for modeling species niches and distributions v. 3.4. 1." In: *URL: http://biodiversityinformatics. amnh. org/open_source/maxent*

Phillips, Steven J, Robert P Anderson, and Robert E Schapire (2006). "Maximum entropy modeling of species geographic distributions." In: *Ecological modelling* 190.3-4, pp. 231–259. ISSN: 0304-3800.

Picard toolkit (2018). `http://broadinstitute.github.io/picard/`.

Ploquin, Emilie F, José M Herrera, and José R Obeso (2013). "Bumblebee community homogenization after uphill shifts in montane areas of northern Spain." In: *Oecologia* 173.4, pp. 1649–1660. ISSN: 0029-8549.

Polce, Chiara, Joachim Maes, Xavier Rotllan-Puig, Denis Michez, Leopoldo Castro, Björn Cederberg, Libor Dvorak, Úna Fitzpatrick, Frederic Francis, and Johann Neumayer (2018). "Distribution of bumblebees across Europe." In: *One Ecosystem*. ISSN: 2367-8194.

Popescu, Viorel D, Robin G Munshaw, Nancy Shackelford, Federico Montesino Pouzols, Evgenia Dubman, Pascale Gibeau, Matt Horne, Atte Moilanen, and Wendy J Palen

(2020). "Quantifying biodiversity trade-offs in the face of widespread renewable and unconventional energy development." In: *Scientific reports* 10.1, pp. 1–12. ISSN: 2045-2322.

Popescu, Viorel D, Laurenţiu Rozylowicz, Dan Cogălniceanu, Iulian Mihăiţă Niculae, and Adina Livia Cucu (2013). "Moving into protected areas? Setting conservation priorities for Romanian reptiles and amphibians at risk from climate change." In: *PloS one* 8.11, e79330. ISSN: 1932-6203.

Popescu, Viorel D, Laurentiu Rozylowicz, Iulian M Niculae, Adina L Cucu, and Tibor Hartel (2014). "Species, habitats, society: an evaluation of research supporting EU's Natura 2000 network." In: *PloS one* 9.11, e113648. ISSN: 1932-6203.

Potts, Simon G, Jacobus C Biesmeijer, Claire Kremen, Peter Neumann, Oliver Schweiger, and William E Kunin (2010). "Global pollinator declines: trends, impacts and drivers." In: *Trends in ecology & evolution* 25.6, pp. 345–353. ISSN: 0169-5347.

Pradervand, Jean-Nicolas, Loïc Pellissier, Christophe F Randin, and Antoine Guisan (2014). "Functional homogenization of bumblebee communities in alpine landscapes under projected climate change." In: *Climate Change Responses* 1.1, p. 1. ISSN: 2053-7565.

Pressey, RL (2004). "Conservation planning and biodiversity: assembling the best data for the job." In: *Conservation biology* 18.6, pp. 1677–1681. ISSN: 0888-8892.

Primack, Richard B (2002). "Essentials of." In: *Conservation Biology*.

Pritchard, Jonathan K, Joseph K Pickrell, and Graham Coop (2010). "The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation." In: *Current biology* 20.4, R208–R215.

Pritchard, Jonathan K, Matthew Stephens, and Peter Donnelly (2000). "Inference of population structure using multilocus genotype data." In: *Genetics* 155.2, pp. 945–959.

Pyke, Graham H, James D Thomson, David W Inouye, and Timothy J Miller (2016). "Effects of climate change on phenologies and distributions of bumble bees and the plants they visit." In: *Ecosphere* 7.3, e01267.

Pywell, RF, EA Warman, C Carvell, TH Sparks, LV Dicks, D Bennett, A Wright, CNR Critchley, and A Sherwood (2005). "Providing foraging resources for bumblebees in intensively farmed landscapes." In: *Biological conservation* 121.4, pp. 479–494. ISSN: 0006-3207.

QGIS, Development Team (2017). *QGIS Geographic Information System. V 3.10.6*. Computer Program.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria.

Radosavljevic, Aleksandar and Robert P Anderson (2014). "Making better Maxent models of species distributions: complexity, overfitting and evaluation." In: *Journal of biogeography* 41.4, pp. 629–643.

Rasmont, Pierre, Audrey Coppee, Denis Michez, and Thibaut De Meulemeester (2013). "An overview of theBombus terrestris(L. 1758) subspecies (Hymenoptera: Apidae)." In: *Annales de la Société entomologique de France (N.S.)* 44.2, pp. 243–250. ISSN: 0037-9271 2168-6351. DOI: 10.1080/00379271.2008.10697559.

Rasmont, Pierre, Markus Franzén, Thomas Lecocq, Alexander Harpke, Stuart PM Roberts, Jacobus Christiaan Biesmeijer, Leopoldo Castro, Björn Cederberg, Libor Dvorak, and Úna Fitzpatrick (2015a). *Climatic risk and distribution atlas of European bumblebees*. Vol. 10. Pensoft Publishers. ISBN: 9546427683.

Rasmont, Pierre, Stuart Roberts, Björn Cederberg, Vladimir Radchenko, and Denis Michez (2015b). "*Bombus lucorum*. The IUCN red list of threatened species." In: *IUCN*.

– (2015c). "*Bombus terrestris*. The IUCN red list of threatened species." In: *IUCN*.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger (2015). "A practical guide to environmental association analysis in landscape genomics." In: *Mol Ecol* 24.17, pp. 4348–70. ISSN: 1365-294X (Electronic) 0962-1083 (Linking). DOI: 10.1111/mec.13322.

Reyers, B, KJ Wessels, and AS Van Jaarsveld (2002). "An assessment of biodiversity surrogacy options in the Limpopo Province of South Africa." In: *African Zoology* 37.2, pp. 185–195. ISSN: 1562-7020.

Ripley, Brian D (2007). *Pattern recognition and neural networks*. Cambridge university press.

Roberts, David R, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. (2017). "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure." In: *Ecography* 40.8, pp. 913–929.

Rodrigues, Ana SL and Thomas M Brooks (2007). "Shortcuts for biodiversity conservation planning: the effectiveness of surrogates." In: *Annu. Rev. Ecol. Evol. Syst.* 38, pp. 713–737. ISSN: 1543-592X.

Roland Pitcher, C, Peter Lawton, Nick Ellis, Stephen J Smith, Lewis S Incze, Chih-Lin Wei, Michelle E Greenlaw, Nicholas H Wolff, Jessica A Sameoto, and Paul VR Snelgrove (2012). "Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems." In: *Journal of Applied Ecology* 49.3, pp. 670–679.

Rombird (2018).

Ros-Freixedes, Roger, Mara Battagin, Martin Johnsson, Gregor Gorjanc, Alan J Mileham, Steve D Rounsley, and John M Hickey (2018). "Impact of index hopping and bias towards the reference allele on accuracy of genotype calls from low-coverage sequencing." In: *Genetics Selection Evolution* 50.1, p. 64.

Rosenzweig, Michael L (1995). *Species diversity in space and time*. Cambridge University Press. ISBN: 0521499526.

Rousset, Francois (2008). "genepop'007: a complete re-implementation of the genepop software for Windows and Linux." In: *Molecular ecology resources* 8.1, pp. 103–106.

Rozylowicz, Laurenţiu, Viorel D Popescu, Maria Pătroescu, and Gabriel Chişamera (2011). "The potential of large carnivores as conservation surrogates in the Romanian Carpathians." In: *Biodiversity and Conservation* 20.3, pp. 561–579. ISSN: 0960-3115.

Sarkar, Sahotra, James Justus, Trevon Fuller, Chris Kelley, Justin Garson, and Michael Mayfield (2005). "Effectiveness of environmental surrogates for the selection of conservation area networks." In: *Conservation Biology* 19.3, pp. 815–825. ISSN: 0888-8892.

Sarkar, Sahotra, Robert L Pressey, Daniel P Faith, Christopher R Margules, Trevon Fuller, David M Stoms, Alexander Moffett, Kerrie A Wilson, Kristen J Williams, Paul H Williams, et al. (2006). "Biodiversity conservation planning tools: present status and challenges for the future." In: *Annual Review of Environment and Resources* 31.

Sauberer, Norbert, Klaus Peter Zulka, Max Abensperg-Traun, Hans-Martin Berg, Georg Bieringer, Norbert Milasowszky, Dietmar Moser, Christoph Plutzar, Martin Pollheimer, Christiane Storch, Renate Tröstl, Harald Zechmeister, and Georg Grabherr (2004). "Surrogate taxa for biodiversity in agricultural landscapes of eastern Austria." In: *Biological Conservation* 117.2, pp. 181–190. ISSN: 00063207. DOI: 10.1016/s0006-3207(03)00291-x.

Savolainen, Outi, Martin Lascoux, and Juha Merilä (2013). "Ecological genomics of local adaptation." In: *Nature Reviews Genetics* 14.11, pp. 807–820.

Scheffers, Brett R, Luc De Meester, Tom CL Bridge, Ary A Hoffmann, John M Pandolfi, Richard T Corlett, Stuart HM Butchart, Paul Pearce-Kelly, Kit M Kovacs, David Dudgeon, et al. (2016). "The broad footprint of climate change from genes to biomes to people." In: *Science* 354.6313.

Schmitt, Thomas and László Rákosy (2007). "Changes of traditional agrarian landscapes and their conservation implications: a case study of butterflies in Romania." In: *Diversity and distributions* 13.6, pp. 855–862. ISSN: 1366-9516.

Schrey, Aaron W, Andrea L Liebl, Christina L Richards, and Lynn B Martin (2014). "Range expansion of house sparrows (Passer domesticus) in Kenya: evidence of genetic admixture and human-mediated dispersal." In: *Journal of Heredity* 105.1, pp. 60–69.

Schuelke, Markus (2000). "An economic method for the fluorescent labeling of PCR fragments." In: *Nature biotechnology* 18.2, pp. 233–234.

Scriven, Jessica J, Lucy C Woodall, Matthew C Tinsley, Mairi E Knight, Paul H Williams, James C Carolan, Mark JF Brown, and Dave Goulson (2015). "Revealing the hidden niches of cryptic bumblebees in Great Britain: implications for conservation." In: *Biological Conservation* 182, pp. 126–133. ISSN: 0006-3207.

Segelbacher, Gernot, Samuel A Cushman, Bryan K Epperson, Marie-Josée Fortin, Olivier Francois, Olivier J Hardy, Rolf Holderegger, Pierre Taberlet, Lisette P Waits, and Stéphanie Manel (2010). "Applications of landscape genetics in conservation biology: concepts and challenges." In: *Conservation genetics* 11.2, pp. 375–385.

Segurado, PAGE, Miguel B Araújo, and WE Kunin (2006). "Consequences of spatial autocorrelation for niche-based models." In: *Journal of Applied Ecology* 43.3, pp. 433–444.

Sgro, Carla M, Andrew J Lowe, and Ary A Hoffmann (2011). "Building evolutionary resilience for conserving biodiversity under climate change." In: *Evolutionary Applications* 4.2, pp. 326–337. ISSN: 1752-4571.

Sibarani, Marsya C., Moreno Di Marco, Carlo Rondinini, Salit Kark, and Johan du Toit (2019). "Measuring the surrogacy potential of charismatic megafauna species across taxonomic, phylogenetic and functional diversity on a megadiverse island." In: *Journal of Applied Ecology* 56.5, pp. 1220–1231. ISSN: 0021-8901 1365-2664. DOI: 10.1111/1365-2664.13360.

Silva, Sara E, Sofia G Seabra, Luísa G Carvalheiro, Vera L Nunes, Eduardo Marabuto, Raquel Mendes, Ana SB Rodrigues, Francisco Pina-Martins, Selçuk Yurtsever, Telma G Laurentino, et al. (2020). "Population genomics of Bombus terrestris reveals high but unstructured genetic diversity in a potential glacial refugium." In: *Biological Journal of the Linnean Society* 129.2, pp. 259–272.

Simard, Marc, Naiara Pinto, Joshua B Fisher, and Alessandro Baccini (2011). "Mapping forest canopy height globally with spaceborne lidar." In: *Journal of Geophysical Research: Biogeosciences* 116.G4. ISSN: 2156-2202.

Simons, Andrew M (2002). "The continuity of microevolution and macroevolution." In: *Journal of Evolutionary Biology* 15.5, pp. 688–701.

Siriwardena, Gavin M, Robert A Robinson, and Humphrey QP Crick (2002). "Status and population trends of the house sparrow Passer domesticus in Great Britain." In: *BTO Research Report*, pp. 33–50.

Skotte, Line, Thorfinn Sand Korneliussen, and Anders Albrechtsen (2012). "Association testing for next-generation sequencing data using score statistics." In: *Genetic epidemiology* 36.5, pp. 430–437.

Smith, Katherine F, Dov F Sax, and Kevin D Lafferty (2006). "Evidence for the role of infectious disease in species extinction and endangerment." In: *Conservation biology* 20.5, pp. 1349–1357.

Smith, Thomas B, Salit Kark, Christopher J Schneider, Robert K Wayne, and Craig Moritz (2001). "Biodiversity hotspots and beyond: the need for preserving environmental transitions." In: *Trends in Ecology & Evolution* 16.8, p. 431.

Smithers, Courtenay Neville (1988). *The Handbook of Insect Collecting: Their Collection, Preparation, Preservation, and Storage*. Angus & Robertson. ISBN: 0207159114.

Sork, V. L., K. Squire, P. F. Gugger, S. E. Steele, E. D. Levy, and A. J. Eckert (2016). "Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, Quercus lobata." In: *Am J Bot* 103.1, pp. 33–46. ISSN: 1537-2197 (Electronic) 0002-9122 (Linking). DOI: 10.3732/ajb.1500162.

Stanley, D. A., M. E. Knight, and J. C. Stout (2013). "Ecological Variation in Response to Mass-Flowering Oilseed Rape and Surrounding Landscape Composition by Mem-

bers of a Cryptic Bumblebee Complex." In: *PLoS One* 8.6, e65516. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0065516.

Stelzer, Ralph J, Lars Chittka, Marc Carlton, and Thomas C Ings (2010). "Winter active bumblebees (Bombus terrestris) achieve high foraging rates in urban Britain." In: *PLoS One* 5.3, e9559.

Stolar, Jessica and Scott E Nielsen (2015). "Accounting for spatially biased sampling effort in presence-only species distribution modelling." In: *Diversity and Distributions* 21.5, pp. 595–608.

Storfer, Andrew, MA Murphy, JS Evans, CS Goldberg, S Robinson, SF Spear, R Dezzani, E Delmelle, L Vierling, and LP Waits (2007). "Putting the 'landscape'in landscape genetics." In: *Heredity* 98.3, pp. 128–142.

Storfer, Andrew, Austin Patton, and Alexandra K Fraik (2018). "Navigating the interface between landscape genetics and landscape genomics." In: *Frontiers in genetics* 9, p. 68.

Summers-Smith, JD (1999). "Current status of the house sparrow in Britain." In: *British Wildlife* 10, pp. 381–386.

Svensson, Birgitta, Jan Lagerlöf, and Bo G Svensson (2000). "Habitat preferences of nest-seeking bumble bees (Hymenoptera: Apidae) in an agricultural landscape." In: *Agriculture, Ecosystems and Environment* 77.3, pp. 247–255. ISSN: 0167-8809.

Swets, John A (1988). "Measuring the accuracy of diagnostic systems." In: *Science* 240.4857, pp. 1285–1293.

Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar (2013). "MEGA6: molecular evolutionary genetics analysis version 6.0." In: *Molecular biology and evolution* 30.12, pp. 2725–2729.

Tanaka, H, DW Roubik, M Kato, F Liew, and G Gunsalam (2001). "Phylogenetic position of Apis nuluensis of northern Borneo and phylogeography of A. cerana as inferred from mitochondrial DNA sequences." In: *Insectes sociaux* 48.1, pp. 44–51.

Tang, Min, Chloe J Hardman, Yinqiu Ji, Guanliang Meng, Shanlin Liu, Meihua Tan, Shenzhou Yang, Ellen D Moss, Jiaxin Wang, Chenxue Yang, et al. (2015). "High-throughput monitoring of wild bee diversity and abundance via mitogenomics." In: *Methods in Ecology and Evolution* 6.9, pp. 1034–1043.

Team, QGIS Development et al. (2016). "QGIS geographic information system." In: *Open source geospatial foundation project*.

Telford, RJ and HJB Birks (2009). "Evaluation of transfer functions in spatially structured environments." In: *Quaternary Science Reviews* 28.13-14, pp. 1309–1316.

Theodorou, Panagiotis, Rita Radzevičiūtė, Belinda Kahnt, Antonella Soro, Ivo Grosse, and Robert J Paxton (2018). "Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (Bombus lapidarius L.)" In: *Proceedings of the Royal Society B: Biological Sciences* 285.1877, p. 20172806.

Thomassen, Henri A, Wolfgang Buermann, Borja Mila, Catherine H Graham, Susan E Cameron, Christopher J Schneider, John P Pollinger, Sassan Saatchi, Robert K

Wayne, and Thomas B Smith (2010). "Modeling environmentally associated morphological and genetic variation in a rainforest bird, and its application to conservation prioritization." In: *Evolutionary Applications* 3.1, pp. 1–16. ISSN: 1752-4571.

Thomassen, Henri A, Trevon Fuller, Wolfgang Buermann, Borja Mila, Charles M Kieswetter, Susan E Cameron, Eliza Mason, Rena Schweizer, Jasmin Schlunegger, and Janice Chan (2011). "Mapping evolutionary process: a multi-taxa approach to conservation prioritization." In: *Evolutionary Applications* 4.2, pp. 397–413. ISSN: 1752-4571.

Thompson, Julie D, Toby J Gibson, Frédéric Plewniak, François Jeanmougin, and Desmond G Higgins (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." In: *Nucleic acids research* 25.24, pp. 4876–4882.

Thuiller, Wilfried, Bruno Lafourcade, Robin Engler, and Miguel B Araújo (2009). "BIOMOD–a platform for ensemble forecasting of species distributions." In: *Ecography* 32.3, pp. 369–373.

Tomozei, Bogdan (2006). "Preliminary data regarding Apoidea (Hymenoptera) Fauna from the Piatra Craiului National Park." In: *Research in Piatra Craiului National Park* 2, pp. 183–186.

Tozer, Douglas C, Erica Nol, and Kenneth F Abraham (2010). "Effects of local and landscape-scale habitat variables on abundance and reproductive success of wetland birds." In: *Wetlands Ecology and Management* 18.6, pp. 679–693. ISSN: 0923-4861.

Trakhtenbrot, Ana and Ronen Kadmon (2005). "Environmental cluster analysis as a tool for selecting complementary networks of conservation sites." In: *Ecological Applications* 15.1, pp. 335–345. ISSN: 1939-5582.

Valavi, Roozbeh, Jane Elith, José J Lahoz-Monfort, and Gurutzeta Guillera-Arroita (2019). "block CV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models." In: *Methods in Ecology and Evolution* 10.2, pp. 225–232.

Van Oosterhout, Cock, William F Hutchinson, Derek PM Wills, and Peter Shipley (2004). "MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data." In: *Molecular Ecology Notes* 4.3, pp. 535–538.

Vanbergen, Adam J and Insect Pollinators Initiative (2013). "Threats to an ecosystem service: pressures on pollinators." In: *Frontiers in Ecology and the Environment* 11.5, pp. 251–259. ISSN: 1540-9309.

Vandergast, Amy G, Andrew J Bohonak, Stacie A Hathaway, Joshua Boys, and Robert N Fisher (2008). "Are hotspots of evolutionary potential adequately protected in southern California?" In: *Biological Conservation* 141.6, pp. 1648–1664. ISSN: 0006-3207.

Vandersteen Tymchuk, Wendy, PATRICK O'REILLY, Jesse Bittman, Danielle Macdonald, and Patricia Schulte (2010). "Conservation genomics of Atlantic salmon: variation in gene expression between and within regions of the Bay of Fundy." In: *Molecular Ecology* 19.9, pp. 1842–1859. ISSN: 1365-294X.

Vangestel, Carl, Joachim Mergeay, Deborah A Dawson, Tom Callens, Viki Vandomme, and Luc Lens (2012). "Genetic diversity and population structure in contemporary house sparrow populations along an urbanization gradient." In: *Heredity* 109.3, pp. 163–172.

Veríssimo, Diogo, Iain Fraser, Jim Groombridge, Rachael Bristol, and Douglas C MacMillan (2009). "Birds as tourism flagship species: a case study of tropical islands." In: *Animal Conservation* 12.6, pp. 549–558. ISSN: 1367-9430.

Vesterlund, S-R, TM Lilley, T Van Ooik, and J Sorvari (2014). "The effect of overwintering temperature on the body energy reserves and phenoloxidase activity of bumblebee Bombus lucorum queens." In: *Insectes sociaux* 61.3, pp. 265–272. ISSN: 0020-1812.

Vincent, Bourret, Melanie Dionne, Matthew P Kent, Sigbjørn Lien, and Louis Bernatchez (2013). "Landscape genomics in Atlantic salmon (Salmo salar): searching for gene–environment interactions driving local adaptation." In: *Evolution* 67.12, pp. 3469–3487.

Walther-Hellwig, K and Robert Frankl (2000). "Foraging habitats and foraging distances of bumblebees, Bombus spp.(Hym., Apidae), in an agricultural landscape." In: *Journal of Applied Entomology* 124.7-8, pp. 299–306. ISSN: 1439-0418.

Wang, I. J. and G. S. Bradburd (2014). "Isolation by environment." In: *Mol Ecol* 23.23, pp. 5649–62. ISSN: 1365-294X (Electronic) 0962-1083 (Linking). DOI: 10.1111/mec.12938.

Waters, JOE, BEN Darvill, Gillian C Lye, and Dave Goulson (2011). "Niche differentiation of a cryptic bumblebee complex in the Western Isles of Scotland." In: *Insect Conservation and Diversity* 4.1, pp. 46–52. ISSN: 1752-458X.

Waugh, John (2007). "DNA barcoding in animal species: progress, potential and pitfalls." In: *BioEssays* 29.2, pp. 188–197. ISSN: 0265-9247.

Weiss, M and G Banko (2018). "Ecosystem Type Map v3. 1–Terrestrial and Marine Ecosystems." In: *European Environment Agency (EEA)—European Topic Centre on Biological Diversity*, p. 79.

Westphal, C, I Steffan-Dewenter, and T Tscharntke (2009). "Mass flowering oilseed rape improves early colony growth but not sexual reproduction of bumblebees." In: *Journal of Applied Ecology* 46.1, pp. 187–193.

Westphal, Catrin, Ingolf Steffan-Dewenter, and Teja Tscharntke (2006). "Bumblebees experience landscapes at different spatial scales: possible implications for coexistence." In: *Oecologia* 149.2, pp. 289–300. ISSN: 0029-8549.

Wiens, John A, Gregory D Hayward, Richard S Holthausen, and Michael J Wisdom (2008). "Using surrogate species and groups for conservation planning and management." In: *BioScience* 58.3, pp. 241–252. ISSN: 1525-3244.

Wiens, John J, David D Ackerly, Andrew P Allen, Brian L Anacker, Lauren B Buckley, Howard V Cornell, Ellen I Damschen, T Jonathan Davies, John-Arvid Grytnes, Susan P Harrison, et al. (2010). "Niche conservatism as an emerging principle in ecology and conservation biology." In: *Ecology letters* 13.10, pp. 1310–1324.

Willi, Yvonne, Josh Van Buskirk, Bernhard Schmid, and Markus Fischer (2007). "Genetic isolation of fragmented populations is exacerbated by drift and selection." In: *Journal of evolutionary biology* 20.2, pp. 534–542.

Williams, Paul, Dan Faith, Lisa Manne, Wes Sechrest, and Chris Preston (2006). "Complementarity analysis: Mapping the performance of surrogates for biodiversity." In: *Biological Conservation* 128.2, pp. 253–264. ISSN: 0006-3207.

Williams, Paul H (1994). "Phylogenetic relationships among bumble bees (Bombus Latr.): a reappraisal of morphological evidence." In: *Systematic Entomology* 19.4, pp. 327–344. ISSN: 0307-6970.

Williams, Paul H., Mark J. F. Brown, James C. Carolan, Jiandong An, Dave Goulson, A. Murat Aytekin, Lincoln R. Best, Alexandr M. Byvaltsev, Björn Cederberg, Robert Dawson, Jiaxing Huang, Masao Ito, Alireza Monfared, Rifat H. Raina, Paul Schmid-Hempel, Cory S. Sheffield, Peter Šima, and Zenghua Xie (2012). "Unveiling cryptic species of the bumblebee subgenusBombus s. str.worldwide with COI barcodes (Hymenoptera: Apidae)." In: *Systematics and Biodiversity* 10.1, pp. 21–56. ISSN: 1477-2000 1478-0933. DOI: 10.1080/14772000.2012.664574.

Williams, Paul H and Juliet L Osborne (2009). "Bumblebee vulnerability and conservation world-wide." In: *Apidologie* 40.3, pp. 367–387. ISSN: 0044-8435.

Williams, Stephen E and Jeffery Middleton (2008). "Climatic seasonality, resource bottlenecks, and abundance of rainforest birds: implications for global climate change." In: *Diversity and Distributions* 14.1, pp. 69–77.

Wilson, J Bastow, Robert K Peet, Jürgen Dengler, and Meelis Pärtel (2012). "Plant species richness: the world records." In: *Journal of vegetation Science* 23.4, pp. 796–802.

Wisz, Mary Susanne, Julien Pottier, W Daniel Kissling, Loïc Pellissier, Jonathan Lenoir, Christian F Damgaard, Carsten F Dormann, Mads C Forchhammer, John-Arvid Grytnes, Antoine Guisan, et al. (2013). "The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling." In: *Biological reviews* 88.1, pp. 15–30.

Wolf, Stephan, Mandy Rohde, and Robin F. A. Moritz (2009). "The reliability of morphological traits in the differentiation of Bombus terrestris and B. lucorum (Hymenoptera: Apidae)." In: *Apidologie* 41.1, pp. 45–53. ISSN: 0044-8435 1297-9678. DOI: 10.1051/apido/2009048.

Woodard, S Hollis, Brielle J Fischman, Aarti Venkat, Matt E Hudson, Kranthi Varala, Sydney A Cameron, Andrew G Clark, and Gene E Robinson (2011). "Genes involved in convergent evolution of eusociality in bees." In: *Proceedings of the National Academy of Sciences* 108.18, pp. 7472–7477.

Woodard, S Hollis, Jeffrey D Lozier, David Goulson, Paul H Williams, James P Strange, and Shalene Jha (2015). "Molecular tools and bumble bees: revealing hidden details of ecology and evolution in a model system." In: *Molecular Ecology* 24.12, pp. 2916–2936. ISSN: 0962-1083.

Wright, Sewall (1943). "Isolation by distance." In: *Genetics* 28.2, p. 114.

Yu, Filippov Boris et al. (2017). “New data on bumblebee fauna (Hymenoptera: Apidae, Bombus Latr.) of Vaygach Island and the Yugorsky Peninsula.” In: *Arctic Evironmental Research* 17.4.