EPFL

Thèse n° 7106

Species conservation in the face of global environmental changes: surface-based modelling of geo-environmental data to identify vulnerable populations

Présentée le 30 octobre 2020

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire de systèmes d'information géographique
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Estelle ROCHAT

Acceptée sur proposition du jury

Prof. A. Berne, président du jury
Dr S. Joost, Prof. F. Golay, directeurs de thèse
Prof. S. Vuilleumier, rapporteuse
Prof. G. Greub, rapporteur
Prof. T. Kohn, rapporteuse
Prof. S. Manel, rapporteuse

École polytechnique fédérale de Lausanne

2020

REMERCIEMENTS

Tout d'abord, j'aimerais remercier mes deux directeurs de thèse, Stéphane et François pour la confiance qu'ils m'ont accordée, pour leur soutien et pour la grande liberté qu'ils m'ont laissée pour mener à bien ce travail. En particulier, merci Stéphane de m'avoir proposé d'effectuer cette thèse et de m'avoir offert la possibilité de la combiner avec un emploi à MicroGIS. Parce qu'il est toujours plus facile de travailler au soleil, merci également à tous deux de m'avoir permis de travailler quelques jours par semaine à la maison.

Durant cette thèse, j'ai été impliquée dans divers projets et j'aimerais remercier toutes les personnes avec qui j'ai collaboré ou qui m'ont transmis des données. Merci également aux membres du jury qui ont accepté d'évaluer ce travail.

Durant quelques semaines, ma boite aux lettres a reçu des courriers un peu spéciaux, qu'il fallait rapidement mettre au frigo… Merci aux personnes qui m'ont transmis des tiques : Sandra et Antoine Ducommun, Natalia Rochat, Tabea Schutter, Barbara Schutter, Georges Mermillod, Christine André-Bühlmann et Matthew Parkan. Pour compléter ces données, il a fallu partir à la chasse, un tout grand merci à ma maman, Corinne, qui m'a accompagnée pour la collecte des tiques et merci à ma grand-maman, Roberte, pour l'aide à l'étiquetage !

Ces 5 années de thèse ont pu être effectuées dans une atmosphère conviviale. Merci à tous mes collègues du LASIG que j'ai côtoyés durant ces années: Sylvie, Kevin, Tim, Ivo, Solange, Céline, Anaïs, Marc, Oliver, Noemi, Elia, Annie, Thibaud, Matthew, Jessie, Heydi et Marie-Christine. En particulier, merci Oliver pour les discussions scientifiques, pour les idées et pour avoir relu une partie de cette thèse. Merci Solange, Céline, Anaïs pour l'ambiance de travail du bureau du fond. Un tout grand merci Annie pour la relecture de ce document et la correction de l'anglais.

J'aimerais également remercier mes collègues de MicroGIS avec qui j'ai travaillé en parallèle de cette thèse. Merci à tous pour l'excellente ambiance de travail. Un merci particulier à Abram pour la confiance et la grande flexibilité accordée, sans laquelle il aurait été difficile de combiner les deux emplois.

Cette thèse n'aurait pas pu être réalisée sans le soutien de mon entourage. Merci à mes amis, qui ont su me changer les idées et m'ont encouragée de près ou de loin. Un grand merci aussi à ma famille, en particulier mes parents, Corinne et Pierre-Etienne, qui m'ont offert la possibilité de faire des études, ma grand-maman Roberte, mon frère Néhémie et son amie Renata, qui m'ont toujours soutenue et encouragée.

Finalement, merci à toi Raph pout tout ce que tu m'apportes et en particulier pour ta patience, ton soutien et tes encouragements pour terminer cette thèse.

ABSTRACT

Human activities are resulting in many land-use changes, particularly due to urbanisation and intensification of agricultural practices. Because of these changes, in addition to climate change, many species are facing habitat degradation. In order to avoid extinction under these conditions, they can either move to more favourable areas or adapt to their new environment. To develop appropriate conservation measures, it is essential to identify vulnerable populations that may not be able to disperse or adapt. In this context, modelling tools can be used to predict the potential impact of environmental changes on species and populations. However, to date, few approaches take into account the degree of exposure, the possibility of dispersal and the potential for adaptation. In this thesis, we present modelling approaches based on geo-environmental data to integrate these three elements.

First, we use ecological niche models to estimate the distribution of suitable habitats for a given species as a function of environmental conditions. We propose an improvement of commonly used models by developing an approach to integrate spatio-temporal variability of environmental predictors. In addition, we develop a nested model to predict the distribution of vector-borne pathogens. This model can be used to identify populations that may be threatened by an increasing presence of pathogens in their habitat. We use it to model the evolution of the distribution of *Ixodes ricinus* ticks and their *Chlamydiales* bacterial pathogen.

We then use landscape graphs to analyse the connectivity between habitats and estimate the possibilities for threatened species to move to more favourable areas. Connectivity is also essential for maintaining gene flow and genetic diversity, which is necessary for greater adaptive capacity. We propose here an approach combining landscape graphs, simulations and empirical genetic data to identify the impact of reduced connectivity on population persistence and genetic diversity. We use it to identify butterfly populations that are threatened by increasing fragmentation in an urban landscape.

Finally, we develop the concept of "Spatial Areas of Genotypes Probability" (SPAG) to better analyse the adaptive potential of populations. SPAGs make it possible to model the probability of finding locally adapted genetic variants in a given territory, as well as to identify vulnerable populations lacking in genetic variants that would favour adaptation to future climate conditions. We use it to highlight populations of Moroccan and European goats that are poorly adapted to the climatic conditions predicted under a climate change scenario for 2070 (strong variations in precipitation or increased drought).

To conclude, we show how the three modelling approaches presented can be combined and integrated into a more general conservation framework to identify vulnerable populations facing high exposure to environmental changes, low dispersal possibilities and reduced adaptive capacity.

Keywords: Biodiversity, Conservation, Climate change, Ecological Niche Models, Species Distribution Models, Connectivity, Landscape Graphs, Genetic diversity, Urban conservation, Local adaptation, Landscape genetics, Adaptive potential, *Ixodes ricinus, Chlamydiales, Pieris rapae*

RÉSUMÉ

Les activités humaines engendrent de nombreuses modifications de l'utilisation du sol, notamment par l'urbanisation et l'intensification des pratiques agricoles. Associé au changement climatique, ces modifications entrainent une dégradation de l'habitat de nombreuses espèces. Dans ces conditions, les espèces peuvent soit se déplacer vers des zones plus favorables, soit s'adapter *in-situ*. Afin d'élaborer des mesures de conservation appropriées, il est essentiel d'identifier les populations vulnérables qui pourraient ne pas être en mesure de se disperser, ni de s'adapter. Dans ce contexte, des outils de modélisation peuvent être utilisés pour prévoir l'impact des changements environnementaux. Cependant, à ce jour, peu d'approches prennent en compte à la fois le degré d'exposition, les possibilités de dispersion et le potentiel d'adaptation. Dans cette thèse, nous présentons des outils de modélisation pour intégrer ces trois éléments.

Tout d'abord, nous utilisons des modèles de niche écologique pour estimer la répartition des habitats favorables pour une espèce donnée en fonction des conditions environnementales. Nous proposons une amélioration des modèles couramment utilisés en développant une approche pour intégrer la variabilité spatio-temporelle des prédicteurs environnementaux. Nous développons également un modèle imbriqué qui permet de prédire la distribution d'un pathogène véhiculé par un hôte. Ce modèle peut être utilisé pour identifier des populations qui pourraient être menacées par une présence croissante de pathogènes. Nous l'utilisons pour modéliser l'évolution de la distribution spatiale des tiques *Ixodes ricinus* et des bactéries pathogènes *Chlamydiales*.

Nous utilisons ensuite des graphes relatifs au paysage pour analyser la connectivité entre les habitats et estimer les possibilités pour les espèces menacées de se déplacer vers des zones plus favorables. Cette connectivité est également essentielle pour maintenir le flux de gènes et la diversité génétique permettant une plus grande capacité d'adaptation. Nous proposons une approche combinant des graphes relatifs au paysage, des simulations et des données génétiques empiriques afin d'identifier l'impact d'une connectivité réduite sur la persistance et la diversité génétique des populations. Nous l'utilisons pour identifier des populations de papillons qui sont menacées par l'augmentation de la fragmentation dans un paysage urbain.

Finalement, nous développons le concept de "*zones spatiales de probabilité des génotypes*" (SPAG) pour mieux analyser le potentiel d'adaptation des populations. Les SPAGs permettent d'estimer la probabilité de trouver dans un lieu des caractéristiques génétiques qui résultent de l'adaptation locale et d'identifier des populations vulnérables ne possédant pas les caractéristiques génétiques favorisant une adaptation aux conditions climatiques futures. Nous l'utilisons pour mettre en évidence des populations de chèvres marocaines et européennes qui sont mal adaptées aux conditions climatiques prévues par un scénario de changement climatique pour 2070 (fortes variations de précipitations ou sécheresse accrue).

Pour conclure, nous montrons comment les trois approches de modélisation présentées peuvent être combinées afin d'identifier les populations vulnérables fortement exposées aux changements environnementaux avec de faibles possibilités de dispersion et une capacité d'adaptation réduite.

Mots clés: Biodiversité, Conservation, Changement climatique, Modèles de niche écologique, Modèles de distributions des espèces, Connectivité, Graphes paysagers, Diversité génétique, Conservation urbaine, Adaptation locale, Génétique du paysage, Potentiel d'adaptation, *Ixodes ricinus, Chlamydiales, Pieris rapa*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 INTRODUCTION

Biodiversity is our most valuable
but least appreciated resource.
E. O. Wilson

This thesis focuses on the conservation of species in the face of environmental changes… But why should we conserve species? Why are they endangered? And how can we preserve them? This general introduction aims to provide some initial responses to these questions in order to allow for a better understanding of the general context in which this thesis is set. It is also the place to present the structure of the thesis and to briefly introduce each of the following chapters.

1.1 What is Biodiversity and why should we preserve it?

Species preservation is essential to maintaining biodiversity. The term biodiversity was coined by W. G. Rosen as the title of a conference held in Washington in 1986 "National Forum on BioDiversity", where it was intended as a shorthand for "Biological diversity". Following this, E.O. Wilson published a book in 1988 entitled "BioDiversity" containing a collection of papers from the forum (Wilson, 1988). Since then, the term biodiversity has gradually emerged as a common scientific concept. E.O. Wilson defined it as "*all hereditarily based variation at all levels of organization, from the genes within a single local population or species, to the species composing all or part of a local community, and finally to the communities themselves that compose the living parts of the multifarious ecosystems of the world*" (E.O. Wilson in the book "Biodiversity II" (Reaka-Kudla *et al.*, 1996)). Biodiversity thus contains three fundamental components: the diversity of ecosystems, of species and of genes.

In his 1988 book, E. O. Wilson provided several arguments to demonstrate the dependence humans have on biodiversity. Indeed, plants, animals and micro-organisms are essential actors for the maintenance of many ecosystem services (Figure 1-1). First, they are the main drivers of food production, a chain involving several interacting actors. For example, plants use sunlight to convert inorganic matter into biological tissues and are thus an essential component of primary production. However, this process could not work without the intervention of soil micro-organisms, pollinating insects, etc., which themselves depend on other species for life. Second, all organisms contribute to the creation and regulation of the environment by maintaining nutrient and water

cycles, forming and decontaminating soils, purifying air and water and protecting against natural disasters (IPBES, 2019). Finally, organisms contribute to education, social well-being, recreation and cultural heritage. These services involve many species, which interact with each other and play a key role in the chain that enables the proper functioning of ecosystems. Preserving the biological diversity is thus essential to maintaining the various contributions of nature to humans (Eriksson and Hillebrand, 2019).

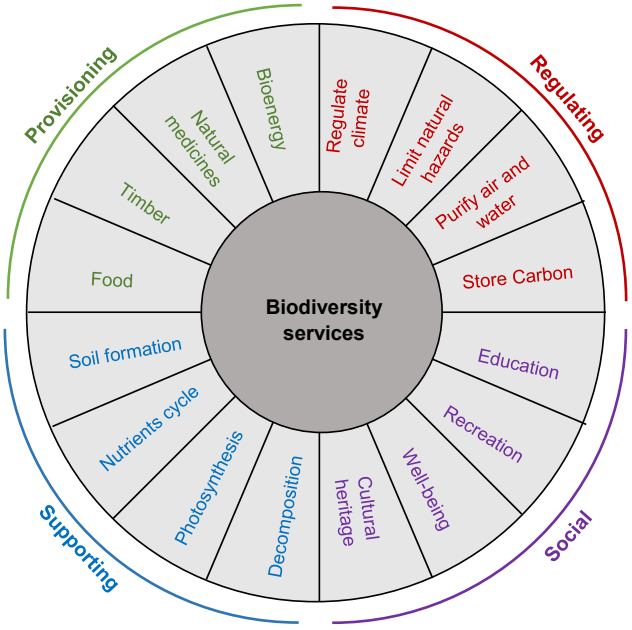


Figure 1-1 – Biodiversity services
Biodiversity provides a large range of services to humans, in various categories such as Provisioning, Supporting, Regulating and Social services (IPBES 2019, iucn.org). Preserving the biological diversity is essential to maintaining the proper functioning of all these services.

1.2 Environmental changes and biodiversity crisis

Several environmental changes can affect the conditions that species experience. Among them, four changes have been highlighted as currently having the most impact on species: land use modifications, increased pollution levels, invasion by non-native species and climate change (IPBES, 2019). Land use modifications are largely the result of agricultural and urban expansion associated with human population growth, which has led to the loss of several natural areas, particularly forests, wetlands and extensive grasslands. Human activities have also led to air, water and soil pollution, which affect the quality of many habitats, from oceans to lands and rivers. These combined effects, which have largely accelerated over the past 50 years (IPBES, 2019), have resulted in a loss of suitable habitats for many species. Last, the rate of invasion by non-native species is increasing (IPBES, 2019), modifying inter-species dynamics.

Climate change has also altered the conditions of various organisms, causing an average warming of 0.2°C per decade over the past 30 years, as well as a sea level rise of more than 3 mm per year since the start of the century and an increase in the frequency of extreme weather events such as storms, droughts and floods (IPBES, 2019). These changes are likely to continue in the future (IPCC, 2014). Indirect effects of climate change are likely to also induce variations in the spatial distribution of diseases, possibly modifying pathogen-host interactions (Hoffmann, 2010).

These changes, combined with overexploitation of natural resources by humans, have led to an increasing loss of biodiversity worldwide. With a degradation of ecosystem quality, more than 31,000 species are listed as under threat of extinction, which corresponds to 27% of all recorded species (https://www.iucnredlist.org/). Additionally, a reduction in gene flow due to a loss of connectivity between natural areas is resulting in erosion of genetic diversity. Livestock genetic resources are also weakened by modern breeding programs, which tend to replace several local breeds with a few high-producing commercial ones (Taberlet *et al.*, 2008). This overall loss of genetic diversity threatens the adaptive potential of species, and thus their ability to cope with a changing environment. Numerous observations indicate a decline in biodiversity, which has led to a focus on the development of many biological conservation projects in recent decades (Dirzo and Raven, 2003; Newbold *et al*., 2015).

1.3 Biodiversity conservation

Due to financial constraints and resource limitations, conservation and management strategies tend to rely on the identification of target populations and the delineation of priority conservation areas (McDonald *et al.*, 2019). This prioritization can occur at the landscape level (selection of priority areas), the ecosystem level (selection of target species) or the species level (selection of target populations within a species) (Vajana, 2017). Traditionally, priorities were set on the basis of habitat irreplaceability and vulnerability (Brooks *et al.*, 2006), species richness, level of endemism, vulnerability or degree of threat (Myers *et al.*, 2000). Corresponding conservation strategies aimed to prevent the loss of species and ecosystem diversity, for example by preserving refuge areas, reducing habitat loss or preventing the invasion of alien species (Williams *et al.*, 2008; Mawdsley *et al.*, 2009). To preserve endangered species, some management strategies involved translocating individuals from sites affected by environmental changes to more suitable habitats or into established captive maintenance to prevent extinction (Mawdsley *et al.*, 2009).

These traditional conservation strategies only partially address the issue of genetic diversity, the third vital component of biodiversity. Following the development of sequencing techniques, conservation genetics was developed, making it possible to integrate the genetic component into the prioritization process by taking into account the level of genetic diversity of species or populations. Indeed, the conservation of a high level of genetic diversity is crucial to preventing inbreeding and to preserve the adaptive potential of species, thus reducing vulnerability to environmental changes (Allendorf and Leary, 1986). Conservation strategies have therefore integrated genetic diversity preservation, for example by favouring gene flow between populations, through hybridization, reintroduction or cross-breeding (Frankham, 2010). Most of these applications have targeted total genetic diversity, including diversity resulting from mutations, gene flow, genetic drift or selection. However, through natural selection, some populations have developed local adaptations that have conferred a higher resistance to stressful climatic conditions, particular environmental conditions or the presence of pathogens or diseases. It is essential to preserve these local adaptations to ensure the survival of a population in a specific habitat. Particularly, some locally adapted populations may present genetic traits that are better fitted to expected future environmental conditions (e.g. by conferring them an adaptation to drought or high temperature), such that preserving these populations may increase the success of conservation outcomes. Thus, it is essential to consider local adaptation when planning conservation strategies. Despite this, few methods currently exist for considering adaptive genetic diversity in the prioritization process, although this issue is beginning to gain attention in conservation discussions (Funk *et al.*, 2019;

Hoelzel *et al.*, 2019; Mable, 2019). In addition, the transposition of academic conservation genetics findings into an applied conservation perspective is too rarely operated. This is mainly the result of insufficient genetic training of conservation managers and decision makers (Frankham, 2010) and even antipathy from conservation managers towards genetic data (Joost *et al.*, 2011). Important efforts to better integrate genetics into practical conservation are thus required, a good example being the CongressGenetics project (Hoban *et al.*, 2013).

1.4 Modelling of geo-environmental data

The data needed to develop conservation strategies cannot always be acquired through field measurements due to limitations of time, difficulties in access to terrain and restricted financial resources. Because of this, models and simulations are useful tools to support the decisions of conservation managers (Ferson and Burgman, 2006; Epperson *et al.*, 2010). Geo-environmental data are widely available, generally worldwide and are showing an increasing spatial resolution (Leempoel *et al.*, 2017). Climate grids computed from interpolations based on weather station measurements are available via global databases such as WorldClim (https://www.worldclim.org/), as well as through meteorological offices in several countries (e.g. MeteoSwiss). Satellite data provide images of the earth's surface from which several products can be derived, such as land cover classifications, vegetation indices or land surface temperature. Satellite products can also be used to compute digital elevation models, i.e. elevation grids for the entire earth. Several terrain attributes and variables can be derived from these grids, such as slope, aspect, orientation, drainage or sun exposure. These geo-environmental data can then be used to derive models that allow for a better understanding of environmental effects on species, while limiting the need for field work and the associated costs.

Once a model has been trained using current geo-environmental data, it can be projected into the past or the future. This can provide a better understanding of the influence of past environmental changes on the current status of species, or help anticipate the impacts of future environmental changes. Different climate change scenarios for the coming decades are being developed by groups of experts (IPCC, 2014) and research institutes have computed corresponding climate grids. These grids can be used to model the effects of climate change on species persistence. Modelling can also be used to estimate the impact of urban planning or other land-use changes expected in the future. Anticipating the impact of environmental changes on species persistence is essential for preparing conservation measures and ensuring their implementation before threatened species become extinct (McDonald *et al.*, 2019). This anticipation could also limit the cost and time required to plan recovery processes for endangered species in the future (McDonald *et al.*, 2019).

1.5 Thesis contribution

1.5.1 Research questions

This thesis focuses on the conservation of species and their genetic characteristics, with the aim of providing modelling tools to facilitate the identification of vulnerable populations threatened by several environmental changes. More specifically, we focus on populations facing 1) a loss of ecological niche, 2) a reduction in dispersal possibilities, 3) a decrease in genetic diversity and 4) a lack of locally adapted genetic variants favourable for future conditions.

Ecological niche

To live and reproduce, species require an ecological niche, i.e. a part of habitat that meets all the environmental conditions necessary for their long-term survival. Modifications of ecological niches associated with environmental changes is the focus of **Chapter 2**. Specifically, the suitability of a territory for a species can be estimated using modelling tools called “ecological niche models”. These models correlate records of the presence of a species with environmental variables (land-cover, climate, etc.) in order to identify suitable areas with characteristics similar to the conditions under which the species was already observed. For this purpose, environmental conditions are usually considered at the exact location where the species was recorded (sampling point). However, a species, especially with a high dispersal capacity, may be influenced by the environmental conditions of a much larger area surrounding the sampling point as it interacts with a larger part of the landscape. In addition, climatic data are usually extracted for a time period independent of the sampling date (either an annual mean value, or a summary of values over the decades prior to sampling). In Chapter 2, we thus address the question:

> *“How can we build ecological niche models that integrate the spatio-temporal variability of the environmental predictors, and does this lead to better predictive performance?”*

Furthermore, some species may become vulnerable due to the spread of a pathogen or disease in their ecological niche (Hoffmann, 2010). In this context, we investigate how ecological niche modelling can be used to estimate the distribution of host-pathogens by addressing the question:

> *“Can we use common ecological niche models to estimate the nested niche of a pathogen within the niche of its host?”*

Connectivity and genetic diversity

When species are confronted with a loss of ecological niche or a shift away from the favourable conditions necessary for their survival, in order to avoid extinction, they may either disperse to more favourable areas or adapt *in-situ* to their new conditions. Dispersal is limited by the connectivity between habitats, which is also essential for maintaining gene flow and preserving a high level of genetic diversity. The study of connectivity and its impact on genetic diversity is the focus of **Chapter 3**. Connectivity is particularly threatened in highly fragmented landscapes such as urban areas. Despite this, most conservation studies focus tends to be on natural areas, with little attention on urban landscapes. In addition, it can be difficult to collect genetic data in an urban environment due to habitat fragmentation and limited population size. In this context, Chapter 3 questions:

> *“How can we use modelisation tools using geo-environmental data to complement empirical data and help identify populations threatened by reduced dispersal opportunities and a loss of genetic diversity?”*

Locally adapted genetic variants

Preservation of genetic diversity is essential to maintaining the adaptive potential of populations. However, as previously stated, the preservation of total genetic diversity may not be sufficient, thus it is important to consider the local adaptation of species or populations to a certain environment. This is the focus of **Chapter 4.** Several methods have been developed to identify signatures

of local adaptation by studying the genetic data of individuals. However, very few tools exist for integrating this knowledge into conservation practices. Chapter 4 addresses the following question:

> *"How can we use signatures of local adaptation to identify populations threatened by climate change?"*

Conservation frameworks

This thesis focuses on several specific questions and presents both new modelling tools and improvements to existing ones, in order to identify vulnerable populations under threat from rapid environmental changes. To conclude the thesis, **Chapter 5** addresses the final question:

> *"How can the modelling tools presented be combined and implemented in a framework dedicated to the identification of vulnerable populations?"*

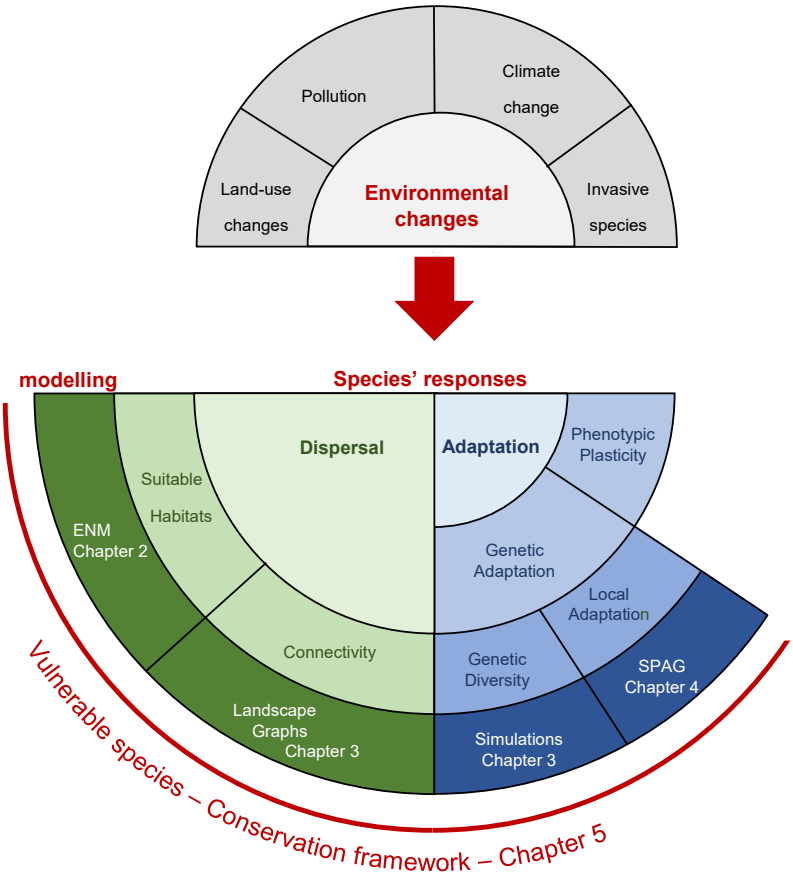


Figure 1-2 – Thesis contributions
Biodiversity is currently threatened by several environmental changes. Faced with such changes in their habitats, species can either disperse to other, more favourable areas, or adapt *in-situ*. Dispersal opportunities depend on the presence of suitable habitats, and sufficient connectivity between them. To be hereditary, adaptation should rely on genetic characteristics (genetic diversity and local adaptation). Populations of species facing high modification of their suitable habitats, limited dispersal opportunities and low adaptive potential are thus the most vulnerable to environmental changes. **ENM** stands for "Ecological Niche Modelling", whereas **SPAG** is for "SPatial Areas of Genotype Probabilities", the name of the approach we develop in Chapter 4.

1.5.2 Objectives

According to the research questions presented in the previous section, we list below the main objectives of this thesis. Detailed objectives will be presented in each chapter.

Chapter 2

- Analyse the performance of ecological niche models as a function of the spatio-temporal variability considered for the extraction of environmental variables (spatial area around the sampling point and time period before the sampling date).
- Develop nested-models to analyse the distribution of host-pathogens and to highlight risk areas where the presence of a pathogen is increasing or is likely to increase in the near future.

Chapter 3

- Use modelling tools to analyse landscape connectivity and its impact on population persistence and genetic diversity.
- Show how modelling can assist in the identification of populations threatened by limited dispersal possibilities and reduced genetic diversity.

Chapter 4

Based on signatures of local adaptation identified by existing methods, develop an approach and the related tool to:

- Predict the probability of presence of one or more locally adapted genetic variants in non-sampled areas.
- Identify areas where there is a greater probability of finding individuals better adapted to future climatic conditions.
- Identify vulnerable populations that may be threatened by climate change due to a lack of adapted variants favourable for future conditions.

Chapter 5

- Show how the various tools presented can be combined into a coherent methodological modelling framework to identify vulnerable populations.

Chapters 2, 3, and 4 are composed of three research papers that have either already been published or are submitted for publication in peer-reviewed journals. These chapters thus contain an introduction to research questions and modelling tools used in the article and are followed by the integral reproduction of the corresponding paper.

Chapter 2 SPECIES ECOLOGICAL NICHE

2.1 Research context

The conservation of a species in an environment first requires the preservation of a suitable area where the species can live and reproduce.

2.1.1 Definitions: habitat and ecological niche

The concept of **habitat** is ambiguously defined (Hall *et al.*, 1997), but the simplest definition is "the place in which species live" (Kearney, 2006). The habitat is characterised by a variety of specific conditions, such as climate and land cover and can be described without reference to a particular species (Kearney, 2006). Examples of habitats include savannah, grasslands, deserts, oceans, rivers, etc. The habitat is closely related to the **niche** of a species. The **fundamental niche** or **ecological niche** of a given species is the part of a habitat that respects all the environmental conditions needed for its long-term survival, i.e. enabling it to find food, shelter, and to reproduce (Hutchinson, 1957; Sillero, 2011). Given an environmental space, the **potential niche** is defined as the portion of the environment that respects the constraints of the fundamental niche (Sillero, 2011). However, the establishment of a population may be limited by additional biotic factors such as competition, predation, human influence, etc. (Phillips *et al.*, 2006). The **realized niche** is the part of a potential niche that is actually being occupied by the species (Sillero, 2011).

2.1.2 Threats to the ecological niche

To preserve a species in a territory, it is first necessarily to maintain the potential niche. This niche may, however, be under threat by environmental changes that alter the abiotic conditions required for the species' survival. For example, Hughes *et al*. (2000) showed that many species, including alpine plants, marine invertebrates, birds and flying insects, had to modify their living range during the 20th century in response to global warming. Similarly, Parmesan *et al*. (1999) studied 35 non-migratory butterflies species and showed that 63% of them had to expand their range northwards in response to the warming conditions of the last century. In addition, habitat loss and degradation are some of the most important factors influencing species extinction risk (Pimm and Raven, 2000; Brooks *et al.*, 2002). For example, coral reefs are threatened with global extinction due to

degradation of their potential niche resulting from global warming (causing coral bleaching) and human disturbance (increasing sedimentation and eutrophication) (Munday, 2004). Similarly, the orangutan is threatened with extinction due to deforestation associated with oil palm cultivation (Swarna Nantha and Tisdell, 2009).

In addition to threatening the preservation of appropriate potential niches, environmental changes may also lead to modifications of biotic interactions, e.g. by expanding of the potential niche of some invasive, pathogenic, competitive or predatory species. This may limit the possibility of a potential niche becoming the realized niche for a given species. For example, Musolin (2007) showed that climate change associated with milder winters led to a clear northward spread of an important crop pest between 1960 and 2000. Similarly, the review by Walther *et al.* (2009) highlights that global warming has induced an expansion of various invasive species (plants, fishes, birds) into areas where they would not have been able to survive and reproduce previously, which may reduce the diversity of native species in these areas.

To preserve the potential niche of species, it is important to understand the environmental factors that affect the presence of species, in order to attempt to predict the influence of environmental changes on the suitability of a territory or its potential to become invaded by predators, competitors or pathogens. In this context, ecological niche modelling is of great use for estimating the distribution of suitable areas for various species and how an area's suitability will evolve with time.

2.1.3 Ecological niche modelling (ENM)

Several methods exist to model species distributions based on the observed relationships between species occurrence and environmental variables. These techniques are generally referred to as "ecological niche models" (ENM) or "species distribution models" (SDM) (Peterson and Soberón, 2012), and they can be divided into two main categories: presence-absence and presence-only methods.

Presence-absence modelling techniques are based on a comparison of the environmental conditions at locations where a species is present with those locations where it is absent. Consequently, these methods require data for both presence and absence sites for the species. The latter may be very difficult to obtain as the inability to find individuals of a species at a particular site is not a confirmation of its absence. Due to this difficulty, several methods have been developed to estimate the distribution of a species based on occurrence data only. We will focus here on these presence-only modelling techniques.

The first methods to be developed for presence-only modelling were based on environmental envelopes. These envelopes correspond to a volume defined in the n-dimensional space of the environmental predictors and include all species occurrence points. The limit values observed on each axis can then be used to identify the range of environmental conditions suitable for a species. Among these methods, Busby *et al*. (1986) first suggested defining the envelope as an n-dimensional rectangle (BIOCLIM). Walker and Cooks (1991) then replaced the rectangle with a n-dimensional convex polygon. They also suggested the use of a set of sub-envelopes, each of which can have a different degree of membership in the total envelope depending on the number of occurrence points that they contain (HABITAT). Using a different approach, Carpenter *et al.* (1993) proposed deriving a habitat suitability value using a point-to-point similarity metric to compare the environmental conditions at a given site with those at the occurrences sites (DOMAIN).

Around the turn of the century, new methods emerged, proposing novel techniques based on the discrimination of presences from “background” or “pseudo-absence” points, which are resampled throughout the entire territory. Among them, Stockwell *et al*. (1999) suggested an algorithm based on a succession of rules to discriminate presences from background sites and achieve a binary prediction (GARP). Hirzel *et al.* (2002) developed a method similar to principal component analyses to compute factors that best explain the distribution of species based on marginality (how the mean observed on occurrence sites differs from the global mean in the study area) and tolerance or specialization (how the variance between occurrence sites differs from the total variance) (ENFA). Later, Phillips *et al.* (2006) developed a new algorithm to discriminate presence data from background, based on the machine-learning principle of maximum entropy (Maxent, see Chapter 2.1.4).

Several studies have compared the performance of these presence-only techniques and ranked Maxent among the most powerful ones (Elith *et al.*, 2006; Hernandez *et al.*, 2006; Guisan, Zimmermann, *et al.*, 2007; Pearson *et al.*, 2007; Tsoar *et al.*, 2007; Graham *et al.*, 2008; Huerta and Peterson, 2008; Wisz *et al.*, 2008; Hoffman *et al.*, 2010). Accordingly, we focus on the Maxent method in the following sections.

2.1.4 Maxent

When modelling species distributions, we search to estimate the probability of finding a species (y=1) based on the environmental conditions (x) of a site. Using Bayes’ theorem, we have:

$$p(y = 1|x) = \frac{p(x|y = 1) * p(y = 1)}{p(x)}$$

Formula 2-1

In this formula, *p(x|y=1)* corresponds to the probability of observing specific environmental conditions knowing that the species is present, which can be estimated from the values of the environmental conditions observed at the occurrence sites. However, *p(y=1)*, which corresponds to the prevalence of the species, and *p(x)*, which is the probability of observing specific environmental conditions, are both unknown. Nevertheless, the latter can be estimated using background sites randomly sampled across the entire territory. Indeed, these background points provide an estimation of the environmental conditions of the entire study area and can thus be used to estimate the probability of observing given conditions. As a result, only *p(y=1)* remains unknown and we can calculate a suitability index proportional to the probability *p(y=1|x)* that we were initially looking for.

$$Suitability = \frac{p(x|y = 1)}{p(x)} \sim p(y = 1|x)$$

Formula 2-2

The estimation of *p(x|y=1)* based on occurrence data requires fitting a probability distribution. To choose this distribution, Phillips *et al.* (2006) suggested first applying the constraint that the mean of the distribution has to match the mean of the environmental predictors observed in the sampled occurrences. Then, among the distributions meeting this constraint, they suggested selecting the distribution that is closest to uniform, i.e. closest to *p(x)*. Indeed, *p(x)* can be considered as a null

model for *p(x|y=1)*, as without occurrence data we could do no better than to consider that the probability of finding certain environmental conditions knowing that the species is present is proportional to the frequency of these environmental conditions in the territory (Elith *et al.*, 2010). Since the distance from *p(x|y=1)* to *p(x)* is the relative entropy of *p(x|y=1)* with respect to *p(x)*, minimising the distance between *p(x)* and *p(x|y=1)* is equivalent to minimising the relative entropy. This is also equivalent to maximizing the entropy of the probability of finding a species at a given location (see (Elith *et al.*, 2010) for a demonstration). Accordingly, Phillips *et al*. (2004) have named their modelling technique "Maxent" for "Maximum Entropy". In addition, Phillips *et al.* (2004) demonstrated that maximizing this entropy results in the fitting of a Gibbs distribution, which takes the following form:

$$p(x|y=1) = p(x) * e^{\alpha+\beta\, x}$$

Formula 2-3

where β are the coefficients of the model and α is a constant that ensures that *p(x|y=1)* sums to 1. Using Formula 2-2 and Formula 2-3, the suitability is given by:

$$Suitability = \frac{p(x) * e^{\alpha+\beta\, x}}{p(x)} = e^{\alpha+\beta\, x}$$

Formula 2-4

A log-likelihood procedure can then be used to identify the β parameters that best fit this model (Phillips *et al.*, 2004). Later, Renner and Warton (2013) demonstrated that the Maxent algorithm is also equivalent to a Poisson point process model, which can be implemented using a generalised linear model. Subsequently, Philips *et al.* (2017) proposed an open source release of Maxent, implemented through the R package "*maxnet*" and based on the "*glmnet"* package for generalized linear models. To run the model, several parameters must be chosen by the user, where these choices have been shown to strongly influence the resulting predictions (Phillips and Dudík, 2008; Merow *et al.*, 2013; Morales *et al.*, 2017; Hallgren *et al.*, 2019). Some of the options available for several parameters are presented in Annex A1.

2.1.5 Model evaluation

Several tools have been proposed to evaluate the performance of species distribution models, to estimate their accuracy and predictive power, and to compare various predictions. Most of these tools are based on estimates of sensitivity and specificity. Given the continuous suitability values predicted by models such as Maxent, the application of a threshold is necessary to obtain binary predictions corresponding to the presence or absence of the species. Once such a threshold has been chosen, one can estimate the **sensitivity** of the model, i.e. the fraction of occurrence sites that are correctly predicted as presences. If absence data are also available, **specificity**, i.e. the fraction of absence sites correctly predicted as absences, can also be calculated. However, with presence-only data, specificity cannot be computed directly. The evaluation of the model will therefore focus on the distinction of presence from background locations, and the specificity is estimated by the fraction of background locations predicted as absences.

The **Receiver Operating Curve** (**ROC**) is the plot of the sensitivity versus 1-specificity for all possible threshold values (Figure 2-1) and the **Area Under this Curve** (**AUC**) is one of the most frequently used indicators to evaluate the performance of ecological niche models (Jiménez-Valverde, 2012). This measure was first introduced to estimate the accuracy of species distribution models by Fielding and Bell (1997) and it represents the probability that a randomly selected presence site will obtain a higher suitability value than a randomly selected absence or background site (Elith *et al.*, 2006; Merow *et al.*, 2013). If the model is not better than random, the AUC is equal to 0.5. The highest theoretically achievable AUC with presence and absence data is 1.0, but it is lower for presence-background data because some of the background points can correspond to presence sites (Phillips *et al.*, 2006).

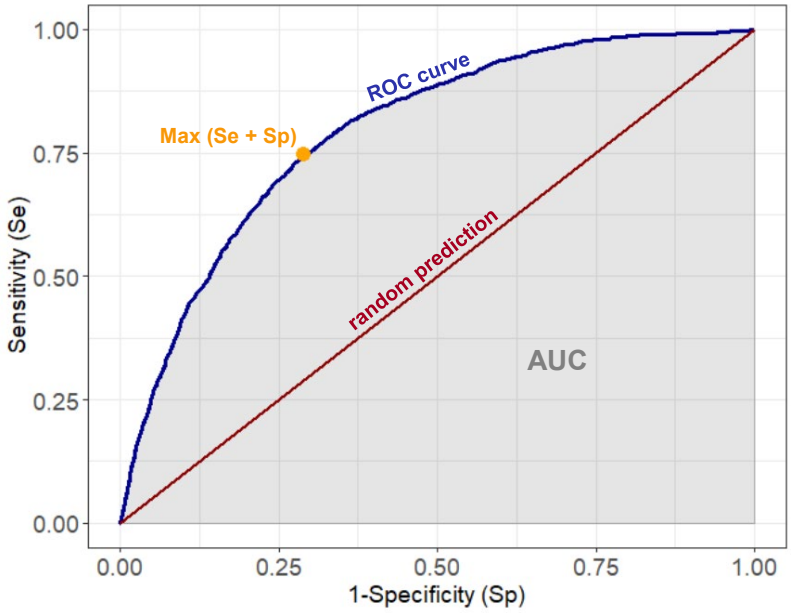


Figure 2-1 – Receiver Operating Curve (ROC)

The AUC has the advantage of being independent of any threshold choice, but it also shows some limitations (Lobo *et al.*, 2008; Jiménez-Valverde, 2012). Particularly, an important limitation is that AUC weights sensitivity and specificity equally (Lobo *et al.*, 2008). However, with presence-only data, we may want to give more weight to sensitivity since we know true presences, whereas the fraction of background sites being true absences is unknown. There is thus little reason to discriminate models with low specificity (Jiménez-Valverde, 2012). Because of its limitations, it is advisable to combine the AUC with other evaluation measures. For example, one can calculate the omission rate (1-specificity), i.e. the percentage of true presences predicted as absences. This would however need the definition of a threshold. Several methods have been suggested to choose an optimal threshold, with one frequently used being the threshold that maximises the sum of sensitivity and specificity and therefore minimises the misclassification rate (Jiménez-Valverde, 2012; Liu *et al.*, 2013).

2.1.6 Applications for conservation

Ecological niche modelling is widely used in conservation studies. First, ecological niche models can be used to study the environmental factors that influence the distribution of a species and to estimate the potential for a species to live in non-sampled areas. For example, Meentemeyer *et al.* (2008) used ENM to estimate the probability of invasion by a tree pathogen causing forest disease and they highlighted areas where early detection sampling should be carried out as a priority. Similarly, Strubbe and Matthysen (2009) used ENM to map areas suitable for an invasive species and highlighted areas not yet colonized, where a spread of the invasive species is likely in the near future, thus threatening native species with competition.

Ecological niche modelling can also be used to model the impact of projected climate or land-cover changes on habitat suitability. This can be achieved by fitting a model on current environmental conditions and then applying it to future projections. For example, Schleupner and Link (2008) used such modelling to study the impact of agricultural intensification on the availability of suitable areas for breeding bird populations in Eiderstedt, Germany. They highlighted a severe negative impact on bird populations, which can be used to argue against the projected changes. Falk *et al.* (2011) suggested using ENM as a basis for decision making in forest management planning, due to the possibility of projecting climate changes scenarios. As an example of application, they modelled the distribution of silver fir under current and future climatic conditions and highlighted risk areas where this tree species should not be introduced due to unsuitable conditions predicted in the future. Bradley *et al.* (2010) modelled the current distribution of three highly invasive plants and then used the model to project the distributions in 2100 under various climate change scenarios. From this, they highlighted areas where prompt eradication or management should be carried out as a priority.

2.2 Scientific contribution

2.2.1 Problem statement

Spatial variability

Usually, species distribution modelling studies use environmental variables extracted for the sampling point only (Elith *et al.*, 2006; Bradley *et al.*, 2010; Williams *et al.*, 2015; Raghavan *et al.*, 2016, 2019, 2020; Sage *et al.*, 2017; Minigan *et al.*, 2018; Soucy *et al.*, 2018; Eisen *et al.*, 2018; Hadgu *et al.*, 2019). However, the response of a species to its environment may involve a wider area, notably as a function of its dispersal capability, or due to various biotic interactions (e.g. the presence of competitors or predators). The area considered in the point extraction of environmental variables depends on the spatial resolution of the data used. However, this spatial resolution is often chosen based on data availability, rather than considering the species ecology (Mayer and Cameron, 2003; Meyer, 2007). For example, many studies have used Worldclim climatic data with a spatial resolution of 1 km, or other environmental layers at a similar resolution, for any species and without any justifications (Porfirio *et al.*, 2014; Manzoor *et al.*, 2018). Occasionally, the choice of the resolution is also related to the computational demand, as high resolution over a large extent can exceed the available computing power (Guisan, Graham, *et al.*, 2007; Gottschalk *et al.*, 2011). Obviously, resolution chosen based on this latter criterion may also not be consistent with the species ecology.

In this context, several studies investigated the influence of spatial scale or “grain” on the performance of the ENM predictions (Guisan, Graham, *et al.*, 2007; Gottschalk *et al.*, 2011, 2011; Connor *et al.*, 2018; Farashi and Alizadeh-Noughani, 2018; Manzoor *et al.*, 2018). They showed that the optimal grain depends notably on the species under study (Guisan, Graham, *et al.*, 2007; Connor *et al.*, 2018) and the ENM method used (Farashi and Alizadeh-Noughani, 2018). All of these authors thus concluded that the spatial scale of environmental variables should be carefully chosen, in accordance with the species ecology. Some authors also suggested the use of multi-grain approaches to consider variables affecting the presence of a species at different scales (Meyer and Thuiller, 2006; Meyer, 2007; Mertes *et al.*, 2020). In addition, occurrence data may be attached of some errors or inaccuracies, in which case it would not be consistent to extract environmental variables for the sampling point only at a high resolution (Hanberry, 2013). Guisan *et al* (2007) therefore indicated that spatial resolution should also be in accordance with the error or inaccuracy associated with occurrence records.

However, downscaling an environmental layer to a coarser resolution to account for data inaccuracies or to better fit the ecology of the species under study leads to an inevitable loss of information. For example, the aggregation of categorical variables such as land cover class by retaining the class observed in majority in the coarser cells induces an underestimation of landscape diversity and fragmentation (Saura, 2002). At present, when high-resolution data become increasingly available, notably thanks to advances in remote sensing, and computing power appears less and less limiting, it would be preferable to use methods allowing to keep the precision of the available data.

In this context, instead of downscaling the data and extracting the values for the sampling coordinates, we suggest keeping high-resolution variables, but extracting the values in different buffers surrounding the sampling point. First, this enables a better summary of the environmental conditions of the surrounding area, by using buffers centred on the point of interest instead of a fixed grid (raster) of coarser resolution. In addition, it allows the use of various statistics to summarise environmental variables in the buffer area, such as mean value, standard deviation, median, mode, or percentage for the categorical classes. The use of buffers is not new (Meyer, 2007), but its use in ecological niche modelling remains limited. In addition, an analysis of the influence of the buffer size on the results of Maxent modelling has, to our knowledge, never been presented.

Temporal variability

Similarly, environmental variables are usually extracted for a time period independent of the sampling date. For example, Worldclim data, which are among the most commonly used for ecological niche modelling (Porfirio *et al.*, 2014; Manzoor *et al.*, 2018), provide a summary of climatic conditions from 1950 to 2000. However, the suitability for a species may evolve over a much shorter temporal scale and the use of coarse temporal data may not provide a coherent picture of the current distribution, nor allow an estimate of the evolution over years.

In addition, the occurrences used for the modelling may come from different sources, with sampling corresponding to different months or years between which the climatic conditions may have changed. A common solution is to use an average of climatic conditions over the whole period covering sampling dates (Bradley *et al.*, 2010; Williams *et al.*, 2015). Once again, this leads to a loss of information since it does not allow to depict differences between the years/months under

study. In this context, we suggest extracting the environmental variables for each occurrences at an equivalent time period preceding the sampling date.

Nested-niche

Finally, as previously introduced, ecological niche models can also be useful for estimating the spread of invasive species, competitors, predators or pathogens in the context of environmental changes. In the case of a vector-borne pathogen, modelling the ecological niche of the vector may provide a first evaluation of risk areas where the pathogen's presence may increase in the future (Brownstein John S *et al.*, 2003; Illoldi-Rangel *et al.*, 2012; Vajana *et al.*, 2018). However, due to other factors influencing its presence, it is possible that the potential niche of the pathogen is smaller than that of its vector. In this context, we studied the possibility of using Maxent modelling to estimate the nested-niche of a bacterial pathogen within the niche of its vector host. We propose a two-step application of Maxent based on probability theory. More specifically, we would like to estimate the probability of simultaneously observing a host (Ho) and its pathogen (Pa). Following Baye's rule this probability can be estimated with:

$$p(Ho \cap Pa) = p(Pa\,|Ho) * p(Ho)$$

Formula 2-5

where *p(Ho)* is provided by the ENM derived for the host and *p(Pa|Ho)* can be estimated using ENM based on pathogen occurrence data. Indeed, occurrence data for pathogens are usually obtained by sampling the hosts and analysing the samples to detect those that are infected. Pathogen occurrences thus corresponds to the presence of the pathogen in sites where the host is present and their use in ENM will enable the estimation of a suitability proportional to *p(Pa|Ho)*.

2.2.2 Objectives

In this context, we aim to:

- Analyse the performance of the Maxent models as a function of the spatio-temporal variability considered for the extraction of the environmental variables (spatial area around the sampling point and time period before the sampling date).
- Elaborate an approach combining different time periods and spatial areas (buffer size) for the various environmental factors (similar to a "multi-grain" procedure).
- Analyse the possibility of using Maxent to model the nested niche of a pathogen within the predicted niche of its vector.
- Highlight risk areas where the presence of a pathogen is increasing or likely to increase in the near future.

2.2.3 Case study

In the study presented in section 2.3, we build ecological niche models to estimate the spatial distribution of the tick species *Ixodes ricinus* and its *Chlamydiales* bacterial pathogen across the whole Switzerland from 2009 to 2019. *Ixodes ricinus* is the most common tick species in Switzerland and is known to be the vector of many pathogens (Mermod *et al.*, 1973; Aeschlimann *et al.*, 1986; Pilloux *et al.*, 2015). This tick goes through four life stages before reproducing: egg, larva, nymph and adult (Figure 2-2). Before undergoing metamorphosis from one stage to the following,

the tick needs to feed once from the blood of a host. Ticks thus quest on vegetation, until they find a host to which they can attach. *Ixodes ricinus* has a very large range of potential hosts, including almost all mammals species from Switzerland (particularly rodents, hedgehogs, roe deer, livestock, dogs), but also lizards, birds, and humans (Aeschlimann, 1981). Once they have finished feeding, they fall to the ground where they metamorphose to the next life stage and start questing again. Due to their sensitivity to desiccation, *I. ricinus* ticks need high levels of humidity (McCoy and Boulanger, 2015). For this reason, they regularly interrupt their questing activity to move to the moist soil to rehydrate. The total duration of a life cycle is estimated to be two to three years (McCoy and Boulanger, 2015). During the winter months, when the temperature is too low, the ticks hide close to the ground and enter a state similar to hibernation or diapause, waiting for more favourable conditions to start questing again (Aeschlimann, 1972). If they are not in direct contact with ice, they can survive for over a month at air temperatures of -10°C and for a couple of months at -5°C (Lindgren *et al.*, 2006). However, if the unfavourable conditions persist for too long, the ticks cannot achieve their development before dying and thus the establishment of a stable population is not possible (Daniel *et al.*, 2003).

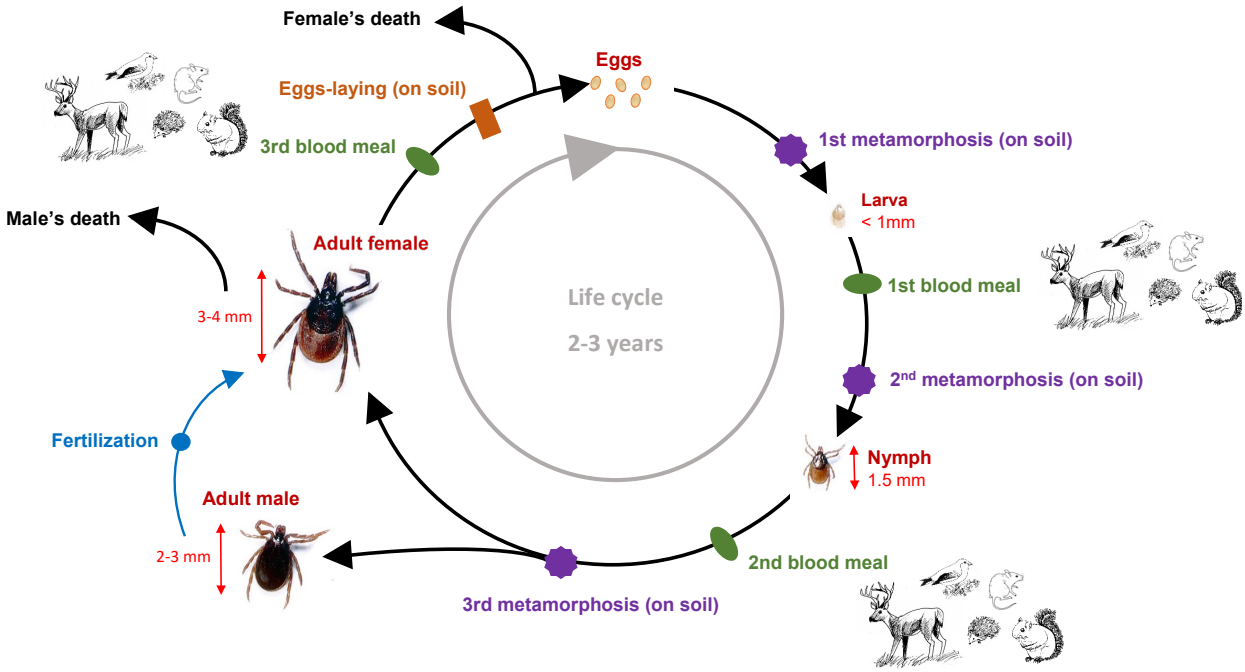


Figure 2-2 – *Ixodes ricinus* life cycle
Ixodes ricinus ticks goes through four life stages (egg, larva, nymph and adult). Before metamorphosing from one life stage to the following, ticks have to feed once from the blood of a host. The total life cycle can last two to three years.

2.2.4 Main conclusions

Our study showed that for *Ixodes ricinus*, the performance of the Maxent model was noticeably higher when considering a buffer area around the sampling point compared with extracting environmental data for sampling point coordinates only. In addition, for this species, the best performing models were obtained when extracting environmental variables in a buffer with a radius of 100 or 200 m, which corresponds to the area of dispersal of known tick hosts. Similarly, the results indicated that the time period considered before the sampling date has a significant impact on the performance of the resulting models. For *I. ricinus* the best performing models were obtained

when considering the climatic conditions of the two or three years preceding the sampling date, which corresponds to the estimated duration of the ticks' life cycle. These results thus highlighted the importance of considering the spatio-temporal variability when extracting environmental predictors for ecological niche modelling.

In addition, our results identified the environmental factors influencing the presence of the *Ixodes ricinus* tick and its *Chlamydiales* bacterial pathogen in Switzerland and allowed us to map the evolution of suitability across the country from 2009 to 2019. We thus showed an application of ecological niche models to study the nested niche of a pathogen within the ecological niche of its host, and we conducted the first investigation of the environmental factors that may influence the presence of pathogenic *Chlamydiales* in ticks. The resulting distribution maps may be used for conservation purposes, as they highlight areas at risk where the presence of a pathogen is increasing. Depending on the availability of sufficient future environmental data (not available for our study), such models may also be used to estimate the evolution of suitable areas for a species or a pathogen in relation to future climate predicted by climate change scenarios. This can be used to highlight populations that are particularly exposed to environmental changes.

Main contributions

- Demonstration of the importance to consider the spatio-temporal variability of environmental variables used in ecological niche models, using buffered areas around the sampling point and time windows preceding the sampling date.
- Illustration of the use of Maxent to model the nested-niche of a parasite within the ecological niche of its host.
- Picture of the evolution of the suitability for the tick species *Ixodes ricinus* over a decade throughout Switzerland and identification of risk areas where prevalence is largely increasing.
- First investigation of climatic factors that may influence the presence of *Chlamydiales* bacteria in ticks.

2.3 PAPER A: *Ixodes ricinus* and *Chlamydiales* Swiss distributions

Nested species distribution models of *Chlamydiales* in tick host *Ixodes ricinus* in Switzerland

Version submitted to **Applied and Environmental Microbiology –**
https://doi.org/10.1101/2020.05.26.118216

Estelle Rochat [a], Séverine Vuilleumier [b], Sebastien Aeby [c], Gilbert Greub [c,*], Stéphane Joost [a,b,d,e,*]

[a] *Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

[b] *La Source School of Nursing, University of Applied Sciences and Arts Western Switzerland (HES-SO), Lausanne, Switzerland.*

[c] *Centre for Research on Intracellular Bacteria, Institute of Microbiology, University Hospital Centre and University of Lausanne, Switzerland.*

[d] *Unit of Population Epidemiology, Division of Primary Care, Geneva University Hospitals, Switzerland*

[e] *Group of Geographic Information Research and Analysis in Population Health (GIRAPH), Switzerland*

** equally contributed*

Contributions

I collected ticks from the prospective campaign, computed the environmental variables, performed all statistical and modelling analyses and wrote the first draft of the paper.

2.3.1 Abstract

The tick *Ixodes ricinus* is the vector of various pathogens, including *Chlamydiales* bacteria, potentially causing respiratory infections. In this study, we modelled the spatial distribution of *I. ricinus* and associated *Chlamydiales* over Switzerland from 2009 to 2019. We used a total of 2293 ticks and 186 *Chlamydiales* occurrences provided by a Swiss Army field campaign, a collaborative smartphone application and a prospective campaign. For each tick location, we retrieved from Swiss federal datasets the environmental factors reflecting the topography, climate and land cover. We then used the Maxent modelling technique to estimate the suitability for *I. ricinus* and to subsequently build the nested niche of *Chlamydiales* bacteria. Results indicate that *I. ricinus* high habitat suitability is determined by higher temperature and vegetation index (NDVI) values, lower temperature during driest months and a higher percentage of artificial and forests areas. The performance of the model was increased when extracting the environmental variables for a 100 m-radius buffer around the sampling points and when considering the data over the two years previous sampling date. For *Chlamydiales* bacteria, the suitability was favoured by lower percentage of artificial surfaces, driest conditions, high precipitation during coldest months and short distances to wetlands. From 2009 to 2018, we observed an extension of tick and *Chlamydiales* suitable areas, associated with a shift towards higher altitude. The importance to consider spatio-temporal variations of the environmental conditions for obtaining better prediction was also demonstrated.

2.3.2 Importance

Ixodes ricinus is the vector of pathogens, including the agent of Lyme disease, the tick borne encephalitis virus and the less known *Chlamydiales* bacteria at the origin of some respiratory infections. In this study, we identified the environmental factors influencing the presence of *I. ricinus* and *Chlamydiales* in Switzerland and generated maps of their distribution from 2009 to 2018. We found an important expansion of suitable areas for both the tick and the bacteria during the last decade. Results provided also the environmental factors that determine the presence of *Chlamydiales* within ticks. Distribution maps as generated here are expected to bring valuable informations for decision-makers to control tick-borne diseases in Switzerland and establish prevention campaigns. The methodological framework presented could be used to predict the distribution and spread of other host-pathogen couples, to identify environmental factors driving their distribution and to develop control or prevention strategies accordingly.

2.3.3 Introduction

Ixodes ricinus is the most common tick species in Switzerland and is known to be the vector of many pathogens, including the tick-borne encephalitis virus and the bacteria *Borrelia burgdoferi*, agent of the Lyme disease (Mermod *et al.*, 1973; Aeschlimann *et al.*, 1986). In 2015, Pilloux *et al*. showed that *I. ricinus* may also have a role of vector and even reservoir for *Chlamydiales* bacteria, especially *Rhabdochlamydiaceae* and *Parachlamydiaceae*. *Chlamydiales* is an order of strict intracellular bacteria containing various bacterial pathogens or emerging pathogens associated with serious diseases for humans and animals, including respiratory tract infections and miscarriage (Corsaro and Greub, 2006; Greub, 2009; Borel *et al.*, 2018). *Parachlamydiaceae* have been largely associated to free-living amoebae (Corsaro *et al.*, 2009, 2010) and are considered as emerging agents of pneumonia in humans (Lamoth and Greub, 2010a, 2010b). They have also been associated with miscarriage in ruminants (Borel *et al.*, 2007; Deuchande *et al.*, 2010) and have been documented in roe deer and red deer, as well as in some rodents (Regenscheit *et al.*, 2012; Stephan *et al.*, 2014). *Rhabdochlamydiaceae* have been mainly described associated to arthropods, including *Porcellio scaber, Blatta orientalis* and *Ixodes ricinus* (Kostanjsek *et al.*, 2004; Corsaro *et al.*, 2007; Pillonel *et al.*, 2019). The pathogenic role of *Rhabdochlamydiaceae* is still largely unknown, but suspected to cause newborn infections (Lamoth *et al.*, 2009) and respiratory complications such as pneumonia (Lamoth *et al.*, 2011).

Considering the potential threat to human health caused by pathogens associated with the tick *Ixodes ricinus*, studies already investigated the influence of environmental factors on its presence or density. They showed that the distribution and activity of *I. ricinus* is mainly influenced by temperature and humidity (Aeschlimann, 1972; Perret *et al.*, 2000, 2003; McCoy and Boulanger, 2015). Indeed, this tick species is prone to desiccation and a relative humidity between 70 to 80% close to the soil is necessary for its survival (Aeschlimann, 1972; Kahl and Alidousti, 1997; Perret *et al.*, 2000). Its most favourable habitats may therefore be vegetation types able to maintain a high humidity level close to the soil such as woodlands with thick vegetation litter (Aeschlimann, 1972; Lindgren *et al.*, 2006; McCoy and Boulanger, 2015).

In Switzerland, several studies analysed the impact of environmental conditions on the activity or density of *Ixodes ricinus*. An early study done by Aeschlimann *et al*. (1972) indicated that *I. ricinus* distribution is mainly limited by the presence of a favourable vegetation cover, with a relative humidity close or superior to 80% and an altitude inferior to 1500 m. Perret *et al.* (2000) showed that the questing activity of ticks takes place from a temperature of 7°C and Hauser *et al.* (2018)

indicated that questing activity is largely reduced when the temperature exceeds 27°C. Jouda *et al.* (2004) showed that in the region of Neuchâtel, the density of ticks decrease with altitude, which was confirmed by Gern *et al.* (2008). However, this relationship was found opposite in the Alps (Valais), which they explained by drier conditions at lower altitude.

Bacteria communities within ticks are also known to be influenced by environmental conditions, notably through a modification of the tick density, the tick behaviour or the vector-host interactions (Carpi *et al.*, 2011; Ehrmann *et al.*, 2018; Aivelo *et al.*, 2019). For example, *B. burgdorferi* is most likely found at lower altitude (Gern *et al.*, 2008), infect more ticks collected in forests than in pastures (Halos *et al.*, 2010; Ehrmann *et al.*, 2018), and may be favoured by the forest fragmentation (Halos *et al.*, 2010; Roome *et al.*, 2018) while *Rickettsia* bacteria may be more prevalent in ticks in pasture sites showing a shrubby vegetation and a medium forest fragmentation (Halos *et al.*, 2010). Environmental factors might provide us with critical information for bacteria distribution and thus potential threats to human. However, nothing has been investigated regarding *Chlamydiales* bacteria yet.

Most studies described above analysed the impact of environmental factors on the density or questing activity of ticks. None modelled across years the spatial distribution of *Ixodes ricinus* habitat suitability at the Swiss scale nor the distribution of the *Chlamydiales* bacteria. In our study, we therefore aimed to build a model estimating the spatial distribution of the *I. ricinus* species from 2009 to 2019 in all Switzerland using the Maxent modelling technique. Beside, we also investigated, for the first time, the ecological factors that determine the distribution of *Chlamydiales* bacteria and the environmental factors that influence the presence of this bacteria within its tick host.

Modelling of *I. ricinus* distribution with Maxent has already been done at the scale of Europe (Porretta *et al.*, 2013), for an area including Europe, North Africa and Middle East (Alkishe *et al.*, 2017) and in Romania (Domsa *et al.*, 2018). Environmental data used in these studies were extracted from Worldclim climatic data at a spatial resolution of 30 arc-second (approximately 1 km). These data summarized climatic conditions from 1950 to 2000. Therefore in these studies as in many others (Williams *et al.*, 2015; Raghavan *et al.*, 2016, 2019, 2020; Sage *et al.*, 2017; Minigan *et al.*, 2018; Soucy *et al.*, 2018; Eisen *et al.*, 2018; Hadgu *et al.*, 2019) environmental data were extracted at a resolution that did not match the species ecology and more importantly the environmental conditions at sampling dates. Our goals were thus first to build a model of higher spatial resolution (100 m) for Switzerland and second to use recent climatic data to characterize in detail the distribution of *Ixodes ricinus* and its associated *Chlamydiales* bacterial pathogen over Switzerland from 2009 to 2019. To better understand the importance of the environmental conditions surrounding the sampling points, and the conditions preceding sampling date, we analysed the performance of the model 1) across buffer zones around the sampling point and 2) through different period of time before the sampling date. Finally, we investigated the potential to use the Maxent modelling to estimate the nested niche of a parasite within the ecological niche of its host.

2.3.4 Material and Methods

Species distribution can be modelled with various methods that use either records of presence and absence of the species or only presences (Elith *et al.*, 2006; Tsoar *et al.*, 2007; Huerta and Peterson, 2008; Norberg *et al.*, 2019). Among them, the Maxent algorithm (Phillips *et al.*, 2006) using presence records only has been shown to perform particularly well as compared to other presence-only modelling methods (Elith *et al.*, 2006; Huerta and Peterson, 2008). We thus chose to use this model to determine the potential ecological niche of *Ixodes ricinus* and its associated *Chlamydiales* bacterial pathogen over Switzerland. The various steps of the method detailed in the paragraphs below are summarised on a Figure in Annex A2.2.

Ticks and bacteria occurrences data

Data regarding tick occurrences were obtained from three different sources. First, ticks were collected by a field campaign conducted by the **Swiss Army** from 21st of April to 13th of July 2009. During this campaign, 172 forests were sampled with convenience sampling in forests in altitude lower than 1,500 m. 62,889 ticks were collected by flagging low vegetation using a white-cloth. The ticks were then aggregated into 8'534 pools of 5 to 10 ticks (5 nymphs or 10 adults) and each pool was analysed for the presence of *Chlamydiales* DNA by using a pan-*Chlamydiales* real-time qPCR as described by Pilloux *et al.* (2015), after extracting the DNA as described by Gäumann *et al.* (2010). Among the 8,534 pools, 543 were positive (6.4%) and they were located in 118 out of the 172 sampling sites (68.6%).

Second, data were obtained from the collaborative smartphone application "**Tick Prevention**" (zecke-tique-tick.ch) developed by A&K Strategy GmbH, a Spin-off from the Zurich University of Applied Sciences (ZHAW) in which users can indicate tick locations on a map. The application was launched in February 2015 and by the end of December 2019, 29 153 locations of tick's observations were available in Switzerland. To each observation a spatial accuracy is assigned depending on the scale (zoomed area) to which the observation was reported by the user. For our analysis, only observations with a spatial accuracy equal or higher to 100 m and only data collected from March to October were used. The final dataset corresponded to 5 781 tick's locations. Moreover, since January 2017, users bitten by a tick can send the tick removed from their body to the national centre for tick-transmitted diseases (NRZK, www.labor-spiez.ch). The ticks received are analysed by three different laboratories for detecting the presence of various bacteria, including *Chlamydiales*. In April 2019, 554 ticks from 506 sites were received and sequenced, among which 21 ticks (3.79%) were positive for *Chlamydiales* bacteria and were located in 19 sites (3.75%).

Finally, to increase the number of data, especially regarding *Chlamydiales* occurrences, a **prospective campaign** was conducted by the authors from 11th of May to 24th of June 2018. During this campaign, 95 sites were visited, mainly in west Switzerland. Those sites were chosen in areas predicted to be favourable for the presence of ticks based on a pre-analysis of the two other datasets, and such to maximise the environmental variability between visited sites (see Suppl. File 1 for more details). Whenever possible, three ticks were collected in each site, by dragging a white-cloth over the soil. For some sites however, only one or two ticks could be found. Eventually, the campaign allowed the collection of 256 ticks, each of which were placed in a sterile tube and kept at 4°C before being sent to the laboratory to be analysed for the presence of *Chlamydiales* bacteria. In the laboratory, the ticks were washed once with 70% ethanol and twice with PBS. DNA was extracted using the NucleoSpin DNA Insect Kit (Macherey-Nagel) with NucleoSpin

Bead Tubes Type E and MN Bead Tube Holder in combination with the Vortex-Genie 2. Manufacturer's protocol was slightly adapted by performing disruption during 20 min followed by a 2h incubation at 56°C in order to allow proteinase K digestion. DNA was then analysed using the pan-*Chlamydiales* qPCR developed by Lienard *et al.* (2011). A tick was considered as positive for the presence of *Chlamydiales* if either the two replicates were positive or if one of the two was highly positive (CT value < 35). As a result, 72 out of the 256 ticks were positive (28.13%), in 51 out of 95 sites (53.6%).

The characteristics of each dataset are summarized in Table 2-1.

Table 2-1 – Datasets *Ixodes ricinus* and *Chlamydiales*
Characteristics of the three data sources regarding *Ixodes ricinus* occurrences and infection by *Chlamydiales* bacteria. The data obtained via the Tick Prevention app are divided into two datasets (column 2 and 3). The first dataset (column 2) corresponds to tick locations recorded on the app. including a majority of ticks for which no information regarding *Chlamydiales* bacteria were available. This dataset was used in the modelling of the distribution of *Ixodes ricinus* only. The second dataset (column 3, which represents a subset of dataset listed in column 2) contains some ticks that were sent to laboratory for the analysis of *Chlamydiales*. This dataset was therefore used in the modelling of *Chlamydiales* distribution. Data from the two other sources (column 1 and 4) were used both for the modelling of *I. ricinus* and *Chlamydiales*.

	Swiss Army field campaign	"Tick Prevention" app. ticks recorded	"Tick Prevention" app. ticks sent for analysis	Authors' prospective campaign
Observation/Sampling dates	21.04.2009 - 13.07.2009	09.03.2015 - 30.10.2019	04.04.2017 - 07.04.2019	11.05.2018 - 24.06.2018
Number of sites	172	5,781	506	95
Number of individual ticks	62,889	5,781	554	256
Number of pools	8,534	-	-	-
Number of ticks/pools infected	543	-	21	72
Infection rate in ticks/pools	6.34%	-	3.79%	28.13%
Number of sites infected	118	-	19	51
Infection rate in sites	68.6%	-	3.75%	53.68%

Environmental data

To characterise the environmental conditions potentially influencing the spatial distribution of *Ixodes ricinus* and *Chlamydiales*, several information were retrieved for the whole Switzerland territory regarding 1) the morphometry 2) the land cover and 3) the climate.

To characterise the **morphometry** of each data point site, seven indicators were derived from the digital elevation model provided by the USGS/NASA SRTM data version 4.1, at a 90m-resolution (Jarvis *et al.*, 2008). The chosen indicators were computed using the SAGA GIS 2.3.2 software (Conrad *et al.*, 2015) and represent: slope, aspect, general curvature, morphometric protection index, terrain ruggedness, sky-view factor and topographic wetness. The definition of each of these indicators and the exact procedure followed to derive them are detailed in Annex A2.3.

To characterise the **land cover**, we first used the land cover statistics from the Swiss Federal Statistical Office (OFS, 2017). From this dataset we retrieved the classification of each Swiss hectare into six land cover types representative of the period 2004-2009: artificial areas, grass

and herb vegetation, brush vegetation, tree vegetation, bare land and watery areas. To better classify forest type, we computed in R (R Development Core Team, 2008) the percentage of coniferous in each forest based on a dataset provided by the OFS at a 25-m resolution which classifies the forests of Switzerland in four classes : pure coniferous, mixed coniferous, mixed broadleaved and pure broadleaved (OFS, 2013). Secondly, we retrieved the vector landscape model swissTLM3D 2016 from the Swiss Federal Office of Topography (O'Sullivan *et al.*, 2008) and we use the function "Proximity" in the QGIS 2.14.7 software (QGIS Development Team, 2016) to derive four indices characterising the minimal Euclidean distance to watery areas: distance to wetland, to watercourses, to stagnant water and to any watery elements. Thirdly, we retrieved the 16-days composite Normalised Difference Vegetation Index (NDVI) available in the MODIS Satellite products at a 250m-resolution (Huete *et al.*, 1999), from which we derived in R the average, minimum, maximum and range of monthly mean NDVI. More details regarding all those land cover data and the derived indicators are also available in Annex A2.3.

Finally, several indicators were computed to summarise the **climatic** conditions of each data point site. They were derived from monthly temperature (average, minimal and maximal) and sum of precipitation grids computed at a 100m-resolution by the Swiss Federal Institute for Forest, Snow and Landscape Research (www.wsl.ch), based on data from MeteoSwiss (www.meteoswiss.ch) and using the Daymet software (Thornton *et al.*, 1997). From these data, 31 indicators were derived to represent the climatic conditions during the period of interest and before sampling date (from 1 to 36 months preceding sampling date, see extraction chapter for more details). These indicators are presented in Annex A2.3 and they summarise 1) the values of the monthly mean, minimal and maximal temperature and sum of precipitation (8 indicators), 2) the variation of monthly temperature and precipitation (5 indicators), 3) the temperature of the warmest (resp. coldest) month (2 indicators) and 4) the temperature and precipitation of the three consecutive warmest (resp. coldest, wettest, driest) months (16 indicators). In addition, grids of the daily maximum and minimum temperature values at a 1km-resolution were obtained from MeteoSwiss. From these datasets, we estimated the daily saturated and ambient vapour pressure using the Tetens formula (Murray, 1966) and by approximating the temperature at dew point by the minimum temperature (Running *et al.*, 1987). We used them to compute the daily relative humidity and to derive 22 indicators summarising the monthly (9 indicators) and daily (13 indicators) values of relative humidity. All these climatic predictors were computed in R, with the detailed procedure presented in Annex A2.3. In total, this resulted in 77 environmental indicators, each of which were resampled to a final spatial resolution of 100 m.

Data extraction

The values of the 77 environmental predictors were extracted for each data point site (tick occurrence) according to their coordinates using the function "extract" from the R "raster" package. The climatic and NDVI variables were retrieved as a function of the sampling dates. To assess the influence of the conditions before sampling, we retrieved these variables for 1 month, 3 months, 6 months, 1 year, 2 years and 3 years before sampling date. For the other stable predictors such as morphometric predictors, land cover type, percentage of coniferous in forest and distances to watery areas one single extraction was used for all sampling dates over the period of analysis (from 2009 to 2019).

To assess the influence of the environmental conditions surrounding the sampling points, for each environmental predictor we also computed the mean value observed in square buffers centred on the sampling point, with radius of 100 m, 200 m, 500 m, 700 m, 1 km and 1.5 km. Raster layers were also computed for each of these indicators, with every buffer radius and time period, for June months from 2009 to 2019. For each pixel, the computation of mean values considering a square buffer around the pixel was done with a moving-window procedure implemented in R, based on the "focal" function from the "raster" package.

Finally, we also extracted all predictors for a randomly generated data set (to test it against sampling data, see hereafter). This generated data set is composed by sites with 10,000 coordinates randomly localised in Switzerland, for which dates were selected randomly within the distribution of observed sampling dates (Annex A2.4).

Ixodes ricinus modelling

Selection of environmental variables

The species distribution models were successively derived using the variables extracted for each combination of buffer radius (100 m, 200 m, 500 m, 700 m, 1 km and 1.5 km) and time period (1 month, 3 months, 6 months, 1 year, 2 years and 3 years). In addition, to select the most significant combination of buffer radius and time period individually for each variable, we performed a Student T-test to identify the variables that best discriminate the tick's presences from random points. The computation was done using the function "t.test" in R and variables were considered as significant if the p-value of the T-test was lower than 0.01 after a Bonferroni correction for multiple comparisons. For each variable, we then kept only the combination of buffer radius and time period showing the highest T-value. A "combination" model was then derived using this "combination" set of variable.

As some environmental variables considered might be correlated, we used two methods to preselect uncorrelated environmental predictors. In the first one, we run a Principal Component Analysis (PCA) on the variables to retrieved independent components. The coordinates of the PCA-components were then used as environmental predictors to run the species distribution model. In the second method, for each pair of variables showing a Pearson correlation higher than 0.8, we kept only the variable with the highest T-value in the T-test previously computed. In addition, we successively removed the variables inducing the highest inflation factor (VIF) computed with the R function "vif", until the highest VIF value was lower than 10. Only the remaining variables were used to train the model.

Maxent Modelling

Species distribution modelling was performed using the Maxent algorithm (Phillips *et al.*, 2006) implemented in the R package "maxnet" (Phillips *et al.*, 2017). Maxent estimates a suitability index which is proportional to the probability of presence of the species knowing the environmental conditions of a site of interest (Elith *et al.*, 2010). The computation requires the values of environmental predictors observed on sites where presence was recorded and on background locations (i.e. locations representative of the entire study area). The model was trained with all *Ixodes ricinus* occurrences available for years 2009 to 2017 and the occurrences from the 2018 prospective

campaign. This represents a total of 2293 presence points. The occurrences reported by the users of the Tick Prevention app. in 2018 and 2019 with 3751 presence points were kept as an independent dataset used to test the models.

Since the performance of the Maxent models is known to be influenced notably by the background point selection, environmental variable selection, features types and regularisation parameters (Lobo and Tognelli, 2011; Barbet-Massin *et al.*, 2012; Merow *et al.*, 2013; Hallgren *et al.*, 2019), we tested different alternatives regarding them. For the selection of background points, we tested two options: either we used the 10 000 points randomly selected in the Swiss territory or we used only the random points situated below 1500 m in altitude, where tick occurrence is more likely. For the environmental variables, we used the two procedures to derive uncorrelated set of variables, i.e. the coordinates of the PCA components and the variables filtered by the previously described method based on Pearson correlation and variance inflation factor. Moreover, when using the PCA components, we considered either all components of the PCA or only the components needed to retain 50% of the variance, resp. 70%, 80%, 90% or 95%. For the feature types, we tested the use of linear features only, or the combination of linear and product, linear and quadratic or linear, product and quadratic together. Finally, we varied the regularisation constant parameters with values equal to 1, 2, 5 or 10.

In order to perform a cross-validation procedure, we used 75% of the occurrences and background points to train the model and kept 25% to test it. The training and testing occurrences were selected randomly and 20 different runs were computed. All models were projected using the “cloglog” scaled output (Phillips, 2017), interpreted in terms of suitability index to avoid making assumptions regarding the prevalence of the species.

Model evaluation

The models were compared based on four criteria. First the Area under the Receiver Operating curve (AUC) (Fielding and Bell, 1997) was computed on the testing dataset. The mean value of AUC_{test} over the 20 runs was used as a measure of discrimination power. The AUC is a measure commonly used for the evaluation of species distribution models (Manel *et al.*, 2001; Elith *et al.*, 2006). It has the advantage to be threshold-independent, but needs to be used in combination with other evaluation parameters (Lobo *et al.*, 2008; Peterson *et al.*, 2008; Jiménez-Valverde, 2012). Therefore, we used as a second evaluation measure the omission error rate, which reflects the accuracy of the model. The computation of this rate requires the definition of a threshold value to classify the predictions into binary presences or absences. Based on the receiver operating curve, we chose the threshold which maximises the sum of specificity and sensitivity and therefore minimizes the misclassification rate (Kaivanto, 2008). Omission errors were computed both on the testing and independent (3751 points from 2018 and 2019) datasets. Finally, to avoid the selection of complex models, that would be difficult to interpret and probably prone to overfitting, we used a third evaluation measure that selected against models having high number of coefficients (following the principle of information criterion (Aho *et al.*, 2014)).

To combine the four evaluation parameters and select the most powerful model, we assigned four performance ranks to each model as a function of each evaluating parameter and we selected the model which minimises the sum of ranks. We then applied the best model to the raster layers to map the predicted suitability across entire Switzerland for June months from 2009 to 2019.

Identification of effective variables

In order to identify the environmental variables most contributing to the model, we implemented in R a jackknife procedure as proposed by Phillips (2017). For each environmental predictor, we computed the Maxent model with only this variable and calculated the corresponding AUC (AUC_{only}). Variables leading to high values of AUC_{only} therefore contribute a lot to the model by themselves. Similarly, we successively computed models with all variables except the one under interest and we computed the corresponding $AUC_{without}$. Predictors associated with high values of $AUC_{without}$ were identified as containing important information that is not present in the other variables.

Chlamydiales Modelling

Background dataset

To model the distribution of *Chlamydiales* bacteria within ticks, we used a similar procedure to that of *Ixodes ricinus*. The modelling was also done using Maxent, based on the 186 occurrence points available for 2009 and 2018. As for *I. ricinus*, the modelling required the definition of background data. Since we are interested by the probability to find *Chlamydiales* within ticks, background points have to represent the environmental conditions of the ecological niche for the tick. Consequently, we built a background dataset in two steps. First, we selected the points where ticks have been observed and analysed for the presence of *Chlamydiales*, but being negative (374 points). Secondly, in order to avoid a model discriminating presences from background due to differences in sampling dates, we completed the background dataset such to have a similar distribution of sampling months and sampling years as in the presence dataset (Annex A2.4). This was achieved by selecting random points within areas predicted to be suitable for ticks, based on the suitability predicted by the models previously derived for *Ixodes ricinus*. The final background dataset contains 1028 data points.

Variable selection and modelling

The same procedure was then applied as for the modelling of the tick's suitability: 1) computation of a T-test to select a "combination" dataset of environmental variables, 2) selection of uncorrelated variables with either a PCA or a correlation/VIF procedure, 3) run of Maxent models by testing various parameters (method to select uncorrelated variables, feature types and regularisation parameters). In order to build models for the suitability of *Chlamydiales* within areas suitable for ticks, the predicted suitability for *Chlamydiales* obtained by the Maxent model was then multiplied by the suitability obtained for *I. ricinus*.

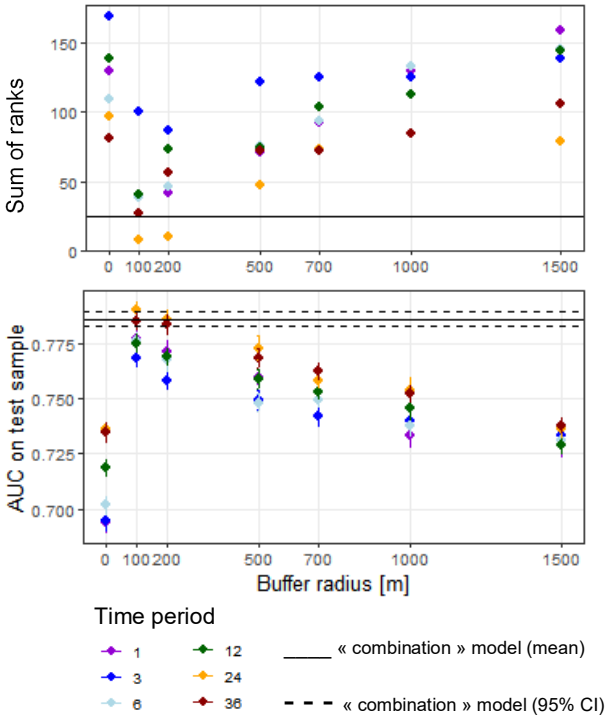
As for *I. ricinus*, twenty runs were computed for each model, using 75% of the data to train the model and 25% to test it. The ranking procedure used to evaluate the models was slightly different to the one used for the tick. The AUC_{test} and the number of coefficients were used similarly, but the omission rates on testing and independent datasets were replaced by two other indicators 1) the difference between the mean of suitability values predicted on occurrences sites in 2009 and the mean suitability predicted on sites without *Chlamydiales* in 2009 and 2) the same difference for 2018. Indeed, even if sites where no *Chlamydiales* were found could not be considered as proper absences, we suspected the probability to find *Chlamydiales* to be lower on these sites. A model showing a lower suitability in areas where *Chlamydiales* were not identified as compared to occurrence sites would therefore be considered as more performant.

2.3.5 Results

***Ixodes ricinus* modelling**

Best model

Among the 56 models tested with various parameters, the best one, according to the ranking procedure, was obtained with the following parameters: 1) background points selected below 1500 m in altitude (corresponding to 6049/10 000 points), 2) a PCA procedure to avoid correlated variables, with the components selected to retained 95% of the variance, 3) a combination of linear and quadratic features and 4) a value of 5 for the regularisation constant parameter. Details of the models tested, and their corresponding evaluation parameters, are available in Annex A2.5. These parameters were then used to test the influence of the choice of buffer radius and time period on the performance of the models. Figure 2-3 shows the AUC_{test} and sum of ranks obtained for each combination. According to these results, the best model was obtained by extracting the environmental variables in a buffer with a 100-m radius around the sampling point and for the 2 years (24 months) preceding the sampling date. Note that the performance of the "combination" model was very close, as well as the performance of models obtained with an extraction for the 3 years preceding sampling date and a buffer radius of 100 m, or for the two years preceding sampling date with a 200 m buffer. Moreover, we observed for each buffer radius, that the models were more performant when considering the variables extracted for the 2 or 3 years previous sampling date, instead of considering the conditions of the current year or even shorter time period. Similarly, the models obtained by extracting the variables within buffers of 100 m or 200 m radius always outperformed the other models. Performance of models with variables extracted at the sampling coordinates only (radius = 0m) was much lower than any buffer model, even those with a radius larger than 500 m. We retained the best model with variables extracted in a 100 m-radius buffer and for the two years preceding the sampling date (Figure 2-3). The global AUC obtained (with both the training and testing data) is 0.794 and the mean AUC_{test} obtained through the 20 runs is of 0.789. The threshold maximising the sum of sensitivity and specificity equals 0.59. Using this threshold, the average omission error on the testing dataset reach 23% and the omission rate on the independent dataset is 11%. The model estimated 31 non-negative coefficients. The median predicted suitability on all occurrences used in the model is 0.74 and the median suitability on independent occurrences from 2018 and 2019 is 0.88.



Characteristics of the best model

Buffer radius:	100 m
Time period:	24 months
Mean AUC_test:	0.789
Mean OE_test:	0.23
Mean OE_indep:	0.11

AUC global:	0.794
Number of coefficients:	31
Optimal threshold (max Se+Sp):	0.59
Median suitability on occurrences:	0.74
Median suitability on independent data:	0.88

Median suitability on occurrences data per year

Year	Number of occurrences*	Median suitability
2009	170	0.749
2015	391	0.722
2016	729	0.739
2017	908	0.737
2018	2117	0.849
2019	1729	0.906

* including independent data for 2018 and 2019

Figure 2-3 – Models performance - *Ixodes ricinus*.
Performance of models predicting the suitability for *Ixodes ricinus*. (Left) Values of the AUC_{test} and the sum of ranks as a function of the buffer radius and the time period considered for the extraction of the environmental variables. For the AUC_{test}, the points indicate the mean value computed through the 20 runs and the lines correspond to the 95% confidence intervals. (Right) Characteristics of the best model chosen according to best values on the graphics on the left. OE_test is the omission error on the test samples and OE_indep the omission errors on the independent additional data available for 2018 and 2019.

Effective variables

The four variables containing the largest amount of important information not available in the other variables (lowest $AUC_{without}$) were: the dimension 1 ($AUC_{without}$=0.748), dimension 12 (0.776), dimension 8 (0.780) and dimension 5 (0.784) (using jackknife procedure, Figure 2-4). The four variables containing the largest amount of important information by themselves (highest AUC_{only}) were: the first dimension of the PCA (AUC_{only}=0.641), the dimension 12 (0.617), dimension 21 (0.591) and dimension 8 (0.582).

The dimension 1 of the PCA is strongly positively correlated with average of the monthly mean temperatures (r=0.91) and indicates that presence of *Ixodes ricinus* is favoured by higher mean temperature. Dimension 8 is moderately correlated with the percentage of herbs and grass vegetation (r=0.57) and the mean temperature during the three consecutive driest months (r=0.40). Its negative coefficient indicates that a higher percentage of herb and grass vegetation or higher temperature values during the driest months are less favourable for the presence of ticks. Dimension 12 is moderately negatively correlated with the percentage of artificial surfaces (r=-0.51) and positively correlated with the range of monthly NDVI (r=0.35). This dimension is also negatively associated with the suitability for ticks, indicating that a higher percentage of artificial surfaces and a lower range of NDVI values are more favourable for *I. ricinus* presence. Finally, the dimension 5 is positively correlated with the mean monthly NDVI (r=0.72), the minimum and maximum NDVI (r=0.55 and 0.52) and is negatively correlated with the percentage of watery areas (r=-0.56).

Its positive coefficient indicates that the areas with higher NDVI values and less water are more favourable for ticks.

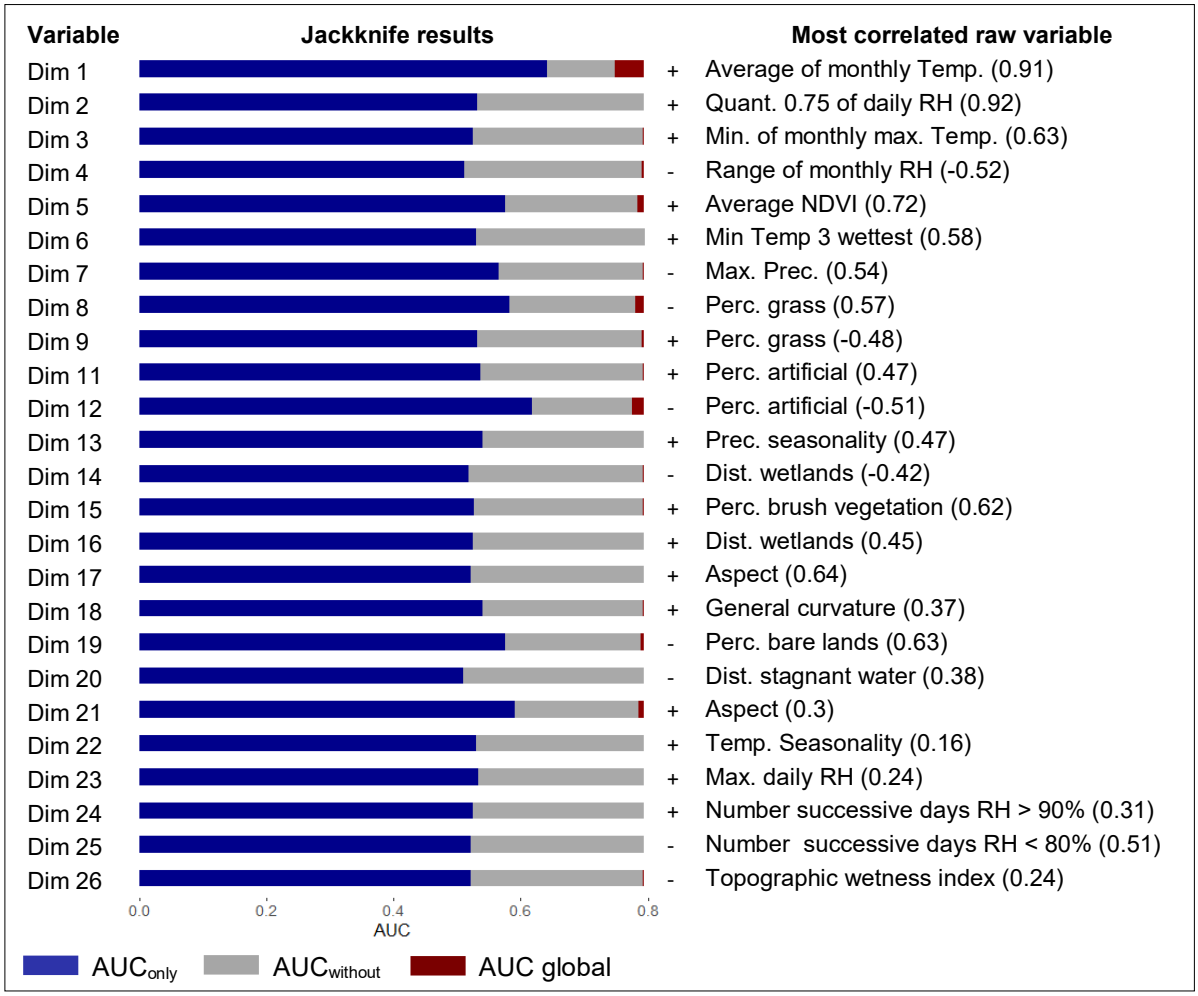


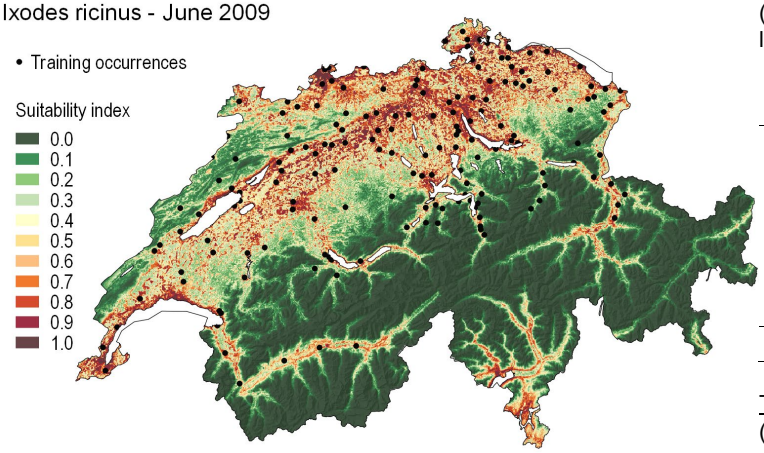
Figure 2-4 – Effective variables - *Ixodes ricinus.*
Dim1 – Dim26 correspond to the components of the PCA needed to retain 95% of the variance. The column with +/- indicates the type of association between the component and the presence of *Ixodes ricinus* (with a positive association, the higher the value of the PCA dimension, the higher the suitability for ticks). The last column shows the raw environmental variable most correlated to the PCA dimension, with the value of the correlation indicated in parenthesis (Temp. = Temperature, RH = Relative Humidity, Quant. = Quantile, Prec. = Precipitation, Perc. = Percentage).

Distribution maps

The maps of the distribution of *Ixodes ricinus* with values of suitability index predicted by the model across Switzerland for June 2009 and June 2018 are shown on Figure 2-5. The corresponding projections for June 2015, 2016, 2017 and 2019 are available in Annex A2.6. Results for June 2009 shows that 16% of the Swiss territory is predicted suitable for the presence of *Ixodes ricinus,* when using the threshold maximising the sum of specificity and sensitivity (threshold = 0.59). The suitable areas are mainly localized in land covered by tree vegetation (48.6 % of all suitable areas), however 26.6% are observed on hectares statistically classified as artificial surfaces. In addition, most of suitable area lied between 500 and 1000 m in altitude (53.04%) or below 500 m (46.5%). Only 0.46 % of the favourable area is found above 1000 m in altitude.

In June 2018, 25% of the Swiss territory is predicted suitable for *Ixodes ricinus* (considering the threshold of 0.59). Between June 2009 and 2018, the predicted suitable area increased by more than 4000 km^2 as shown in Figure 2-5 and only 31 km^2 became unsuitable. The increased suitability is particularly pronounced in the Rhône Valley (Valais), in Surselva, in Simmental, in the Jura border and in other lateral valleys of medium to high altitude (circles on the map). The evolution of the PCA components from 2009 to 2018 in these areas shows that the increase in suitability is generally associated with an increase of the values of Dimension 1 (warmer temperature), an increase of Dimension 5 (higher NDVI values), a decrease of Dimension 12 (lower range of NDVI values), and a decrease of Dimension 8 (temperature during driest months) in Valais and Jura (whereas this last dimension shows an increase of the values in Grisons). The new suitable areas concerned mainly grass and tree vegetation (40.8% each) with a large proportion (64.8%) located at an altitude between 500 and 1000 m (corresponding for example to the altitude of the suited hectares in Jura border or Rhône valley). An increase of suitable areas mainly in forests was also observed between 1000 and 1500 m (8%).The model also predicted suitable areas above 1500 m. These results therefore highlighted a spread of the favourable areas towards higher altitude.

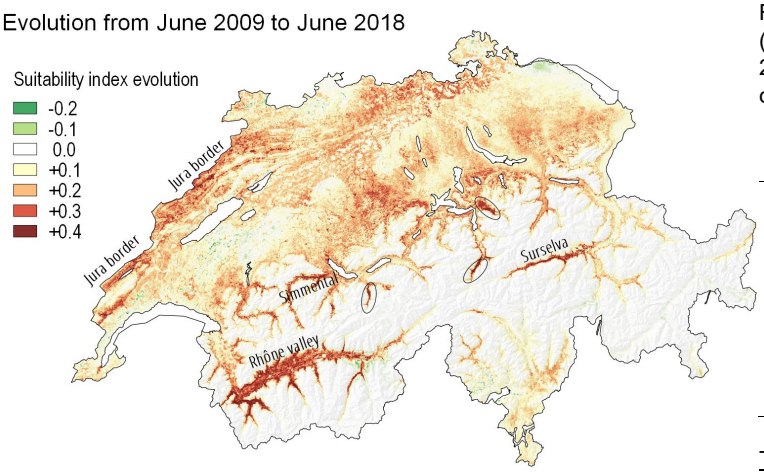
The distribution maps of *Ixodes ricinus* for the years 2015 to 2017 (Annex A2.6) indicate a constant and drastic increase in suitability which is highest between 2017 and 2018. Indeed, 15.7% of the Swiss territory was predicted as suitable in 2009, 16.8% in 2015, 16.2% in 2016, 17.6% in 2017 and 25.4% in 2018 (by considering the threshold of 0.59 for suitable areas). Moreover, the map computed for 2019 predicted important increase from 2018 to 2019, with 35% of the Swiss territory being predicted as suitable in 2019. The spread towards higher altitude was also observed between 2018 and 2019, with a maximal altitude for the favourable areas that reached 1595 m in 2019. The results indicate that since 2018, there is a relatively high probability that ticks reach such altitudes.



Repartition in % of the suitable areas (Suitability > 0.59) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	18.16	8.38	0.04	0	26.58
grass	9.73	6.44	0	0	16.18
bush	2.95	3.45	0.02	0	6.41
tree	14.15	34.02	0.39	0	48.56
bare land	0.65	0.51	0.01	0	1.17
water	0.86	0.25	0	0	1.11
Total	46.5	53.04	0.46	0	100

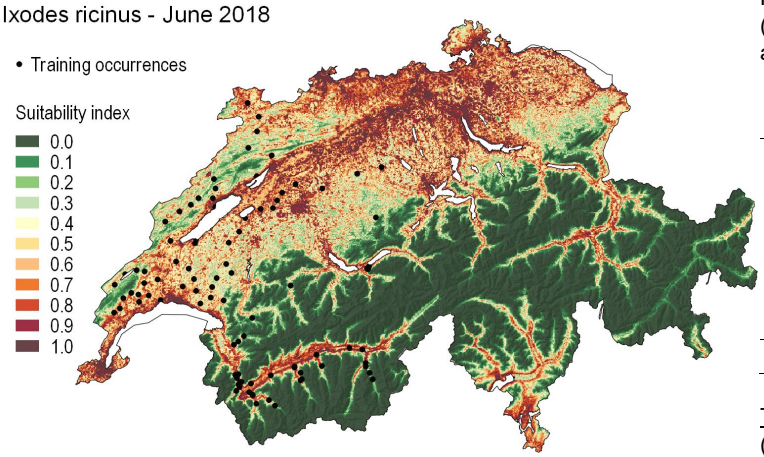
Total suitable area: 6483 km^2
(16 % of the Swiss territory)



Repartition in % of the newly suitable areas (Suitability > 0.59 in 2018 and < 0.59 in 2009) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	3.47	6.78	0.55	0	10.79
grass	18.06	22.33	0.47	0	40.86
bush	0.99	2.5	0.27	0	3.75
tree	2.56	31.43	6.83	0.0007	40.82
bare land	0.27	1.06	0.21	0	1.54
water	1.54	0.68	0.03	0	2.24
Total	26.88	64.77	8.35	0.0007	100

Total newly suitable area: 4032 km^2
Total newly unsuitable area: 31 km^2



Repartition in % of the suitable areas (Suitability > 0.59) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	12.56	7.78	0.23	0	20.57
grass	12.94	12.52	0.18	0	25.65
bush	2.20	3.08	0.11	0	5.39
tree	9.71	32.97	2.87	0.0003	45.54
bare land	0.50	0.72	0.09	0	1.31
water	1.12	0.41	0.01	0	1.54
Total	39.03	57.48	3.49	0.0003	100

Total suitable area: 10 484 km^2
(25 % of the Swiss territory)

Figure 2-5 – Suitability maps - *Ixodes ricinus*.
Suitability map for *Ixodes ricinus* in June 2009 (upper panel) and June 2018 (lower panel) as predicted by the best model (i.e. with environmental variables extracted with a 100m-radius buffer and for the two years preceding sampling date). The area concerned by the transition in suitability are represented in the intermediate panel.

Chlamydiales modelling

Best model

The best model for *Chlamydiales* bacteria, among the 60 models tested with various parameters, was obtained with the following parameters: 1) the “correlation-VIF” procedure to select uncorrelated variables, 2) a combination of linear and quadratic features and 3) a value of 1 for the regularisation constant parameter. The details of all models tested and their corresponding evaluation parameters are available in Annex A2.7. As for the modelling of *Ixodes ricinus*, we then tested the influence of the choice of buffer radius and time period on the performance of the models. Figure 2-6 shows the AUC_{test} and sum of ranks obtained for each combination. According to these results, the “combination” model outperformed the other models. Unlike the results obtained for *Ixodes ricinus* the models for *Chlamydiales* performed better when the variables are extracted for the three- or six-months preceding sampling date than when considering two or three years before sampling (Figure 2-6). In addition, the influence of buffer radius seems to be much less pronounced than for the tick models. Accordingly, we retained the “combination” model. This model used 17 uncorrelated variables selected based on the “correlation/VIF” procedure. The list of these variables, as well as the results of the T-test are available in Annex A2.8. As the “combination” model aims to retain for each variable the best combination of buffer radius and time period, not all variables are selected using the same buffer radius or time period. Interestingly, we observed that the variables used in the model involved either buffer radius smaller or equal to 200 m, or superior to 1 km (Annex A2.8). The characteristics of the model are summarised on the right of Figure 2-6. The global AUC (with both training and testing occurrences) is 0.78 and the mean AUC_{test} obtained through the 20 runs is of 0.74. The threshold maximising the sum of sensitivity and specificity equals 0.3. The mean suitability for *Chlamydiales* occurrence in 2009 is 0.47 and the mean suitability for sites where *Chlamydiales* where not identified in 2009 is 0.37. For 2018, the mean suitability on presence points is 0.46 and the suitability on sites where no *Chlamydiales* were identified is 0.15. The model estimated 35 non-negative coefficients.

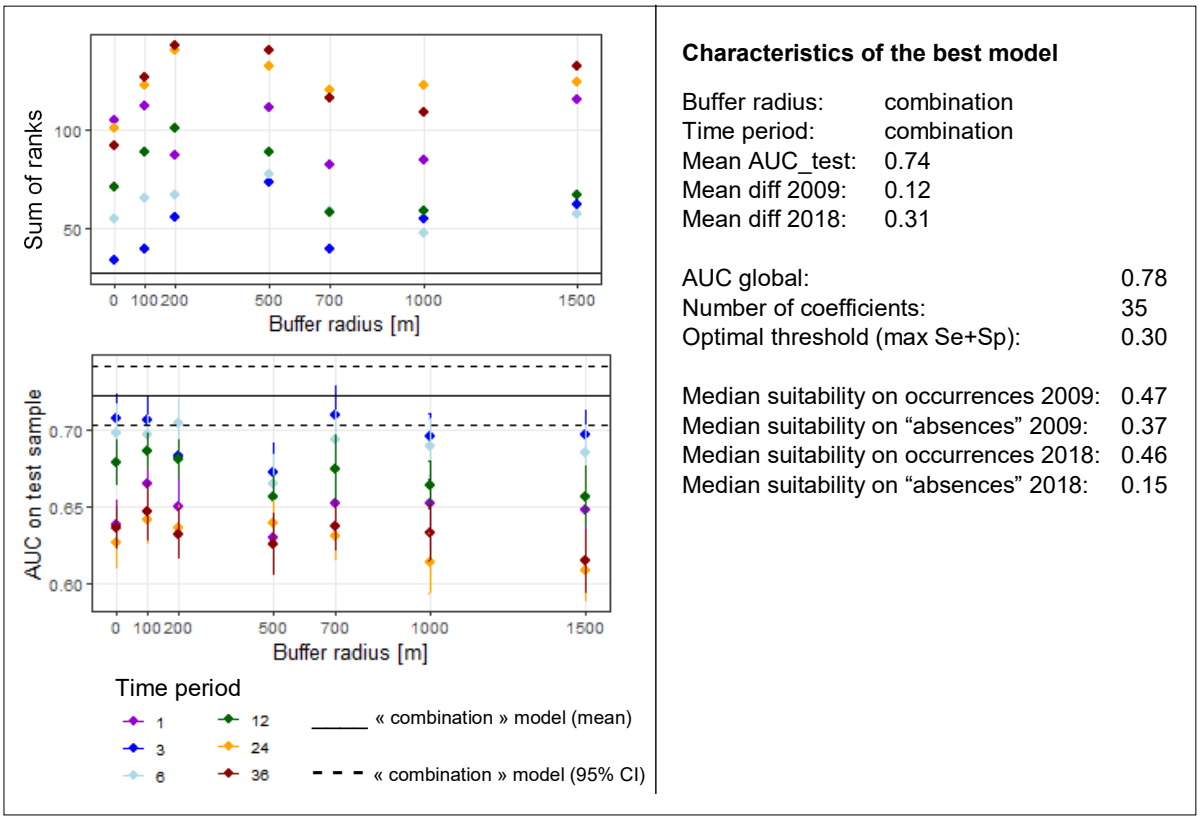


Figure 2-6 – Models performances - *Chlamydiales*.
Performance of models predicting the suitability for *Chlamydiales*. (Left) Values of the AUC_{test} and the sum of ranks as a function of the buffer radius and the time period considered for the extraction of the environmental variables. For the AUC_{test}, the points indicate the mean value computed over the 20 runs and the lines correspond to the 95% confidence intervals. (Right) Characteristics of the best model chosen according to the graphics on the left. Mean diff 2009 (resp. 2018) is the average difference between the mean suitability values predicted on Chlamydiales occurrences points and on locations where no Chlamydiales were identified in 2009 (resp. 2018).

Effective variables

The four variables containing the highest amount of important information that are not available in the other variables (lowest $AUC_{without}$) are (Figure 2-7): the percentage of tree vegetation in a 100 m buffer ($AUC_{without}$ = 0.75), the coordinates (no buffer) number of successive days with a relative humidity inferior to 80% during the 3 months preceding sampling (0.77) or inferior to 70% during the 6 months preceding sampling (0.77) and the distance to wetlands within a buffer of 1km (0.77). The four variables containing the highest amount of important information by themselves (highest AUC_{only}) are: the percentage of artificial surfaces in a 100 m buffer (AUC_{only} = 0.59), the number of days with a relative humidity superior to 90% in a 200 m buffer during the two years preceding sampling date (0.57), the precipitation of the three coldest months in a 1.5 km buffer during the two years preceding sampling (0.55) and the percentage of tree vegetation in a 100 m buffer around the sampling point (0.55).

The conditions favourable for *Chlamydiales* are thus characterised by: a lower percentage of artificial surfaces around the sampling point (7.8% in average for the occurrences locations in a 100m-buffer versus 16.8% for the background locations), a higher percentage of tree vegetation (62.8% versus 53.1%), a lower number of days with a relative humidity superior to 90% during

the two years preceding sampling date (21.1 versus 25.2), a highest amount of precipitation during the coldest months (24.15 mm versus 20.7 mm), a higher number of successive days with a relative humidity inferior to 80% during the three previous months (29.7 versus 27.1) and lower than 70% during the 6 previous months (16 versus 14.4) and finally a shorter distance to wetlands (2.5 km versus 3.1 km).

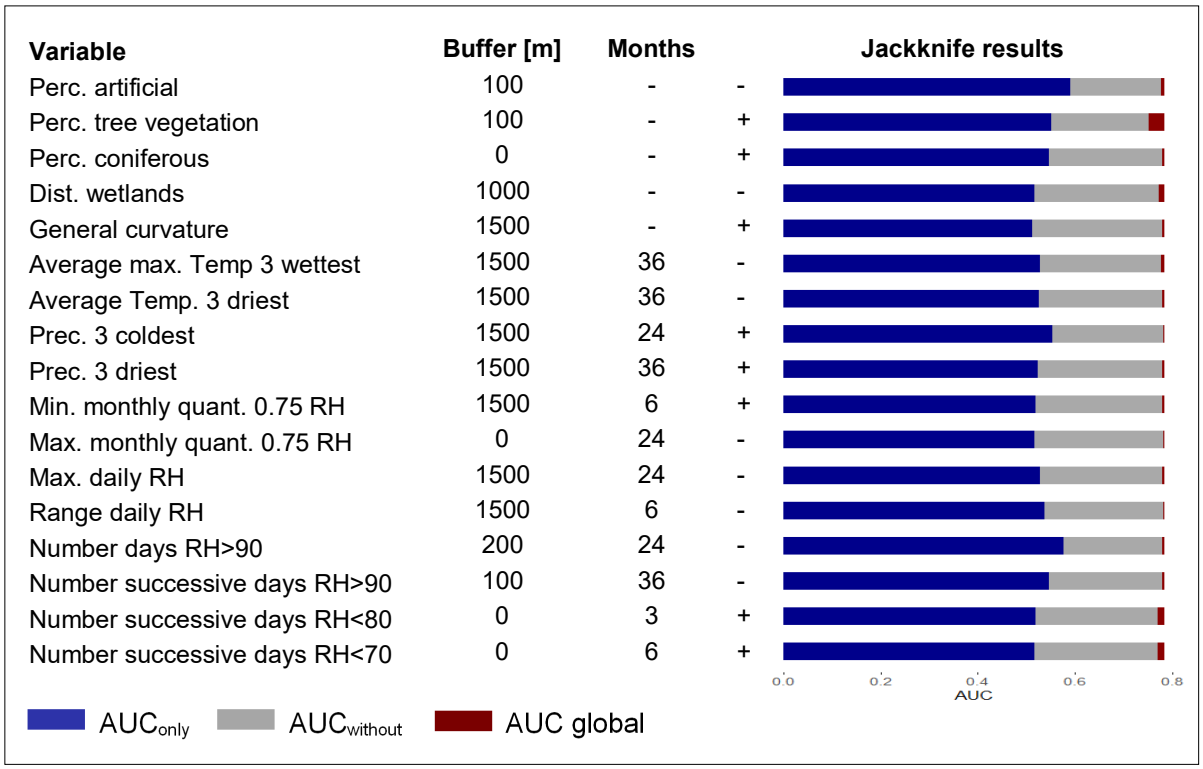


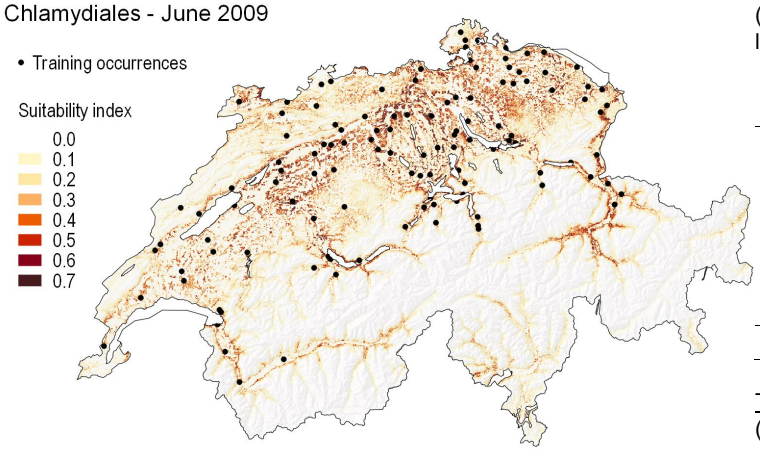
Figure 2-7 – Effective variables - *Chlamydiales*.
Jackknife results for the best model predicting the suitability of *Chlamydiales*. The column "Buffer" indicates the buffer radius around the sampling point and "Months" the number of months before sampling date. The column with +/- indicates the type of association between the variable and the presence of Chlamydiales (with a positive association, the higher the value of the variable, the higher the suitability for Chlamydiales). Perc. = Percentage, Temp. = Temperature, Prec. = Precipitation, quant. 0.75 = quantile 0.75, RH = Relative Humidity.

Distribution maps

The distribution maps of *Chlamydiales* with values of suitability predicted by the model across Switzerland for June 2009 and June 2018 are shown on Figure 2-8. In June 2009, 8% of the Swiss territory is predicted as favourable for *Chlamydiales* bacteria (using the threshold maximising the sum of sensitivity and specificity). As the niche of the bacteria is nested within the niche of the tick, modelling *Chlamydiales* bacteria suitability involved a multiplication by the suitability results for *Ixodes ricinus*. Therefore, the areas predicted to be unfavourable for the presence of the tick species are also predicted as weakly suitable for *Chlamydiales.* On the contrary, some areas predicted to be highly favourable for the presence of *Ixodes ricinus* on Figure 2-8 did not match and showed very low values on Figure 2-8. This is the case for the areas situated within urban settlements, in which a large portion was predicted to be suitable for ticks but not for *Chlamydiales*. Indeed, the distribution of the favourable areas within the various categories of land cover classes indicates that they are essentially observed in natural areas, covered either by tree

(74%) or grass (12%) vegetation, and only 4% of them are observed in regions characterised by a large portion of artificial elements. When considering the altitudinal distribution, areas favourable for *Chlamydiales* seem to be essentially predicted in forest suitable for ticks, between 500 and 1000 m in altitude. However, due to other factors influencing the model, notably the climatic conditions, 52% of those forests are also predicted to be unfavourable for the bacteria.

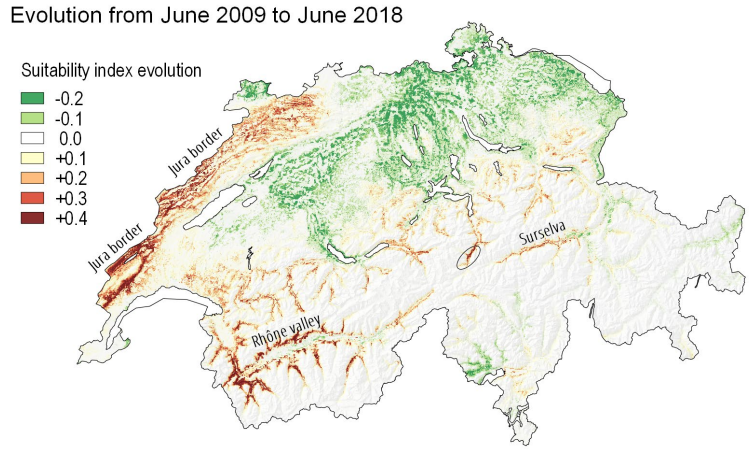
In June 2018, 9% of the Swiss territory is predicted as suitable for the presence of *Chlamydiales*. Between June 2009 and 2018, more than 1850 km^2 are newly suitable for *Chlamydiales* as shown in Figure 2-8. Some regions showing a sharp increase in suitability values (more than 0.4). However, more than 1,300 km^2 is also becoming unsuitable. In 2018, the proportion of suitable area within land cover classes is close to what observed in 2009, with however a clear spread towards higher altitude, with 23% of the favourable areas localised between 1000 and 1500 m, versus 2% only in 2009. Newly suitable area match those of *Ixodes ricinus* on Figure 2-5 (Rhône valley, Surselva, Jura border). The spread of favourable areas towards higher altitude is also predicted, with 45% of the newly suitable hectares being localised between 1000 and 1500 m. Loss of suitable area mainly occurred in the North-West part of Switzerland and appear to be associated with a decrease in precipitation during the three coldest months and a decrease of the successive number of days with a relative humidity inferior to 70% during the 6 previous months (15^{th} of December 2017 to 15^{th} of June 2018 as compared to 15^{th} of December 2008 to 15^{th} of June 2009).



Repartition in % of the suitable areas (Suitability > 0.3) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	1.64	2.34	0.08	0	4.06
grass	3.89	8.49	0.11	0	12.49
bush	1.79	5.16	0.16	0	7.11
tree	17.28	54.88	1.80	0	73.96
bare land	0.23	0.74	0.07	0	1.04
water	0.69	0.60	0.01	0	1.30
Total	25.52	72.21	2.23	0	100

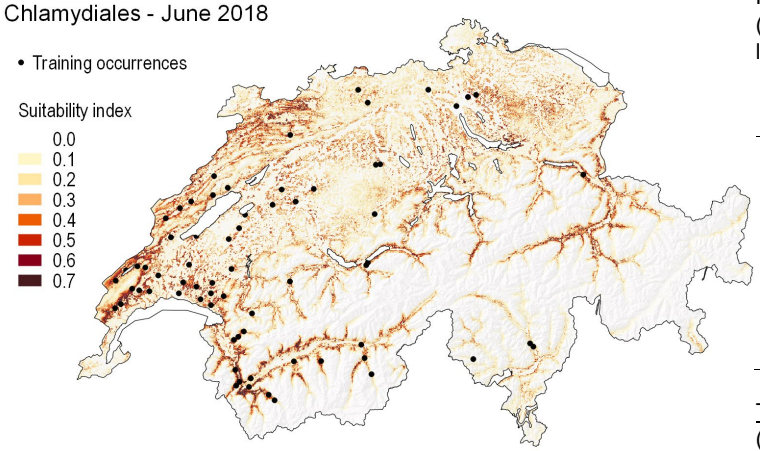
Total suitable area: 3 279 km^2
(8 % of the Swiss territory)



Repartition in % of the newly suitable areas (Suitability > 0.3 in 2018 and < 0.3 in 2009) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	0.45	1.81	1.35	0	3.61
grass	0.99	10.25	5.70	0	16.94
bush	0.29	1.46	1.44	0.02	3.21
tree	1.92	36.21	34.63	0.43	73.19
bare land	0.06	0.82	1.03	0.02	1.93
water	0.18	0.56	0.35	0	1.09
Total	3.89	51.11	44.5	0.47	100

Total newly suitable area: 1 858 km^2
Total newly unsuitable area: 1 287 km^2



Repartition in % of the suitable areas (Suitability > 0.3) within altitude (in m) and land cover classes:

	<500	500 1000	1000 1500	>1500	Total
artificial	0.72	2.08	0.71	0	4.24
grass	1.47	8.90	2.84	0	15.98
bush	0.91	3.85	0.81	0.21	5.29
tree	7.76	48.9	18.07	0.01	71.85
bare land	0.13	0.88	0.56	0	1.37
water	0.33	0.65	0.18	0	1.27
Total	11.32	65.26	23.17	0.22	100

Total suitable area: 3 850 km^2
(9.3 % of the Swiss territory)

Figure 2-8 – Suitability maps - *Chlamydiales*.
Suitability map for *Chlamydiales* in June 2009 (upper panel) and June 2018 (lower panel) as predicted by the best model (i.e. with "composition" set of environmental variables). The area concerned by the transition in suitability are represented in the intermediate panel.

2.3.6 Discussion

Expansion of *Ixodes ricinus* and *Chlamydiales* in Switzerland

Distribution maps for ticks and bacteria from 2009 to 2019 highlighted an extension of the suitable areas for both species and a spread towards higher altitude. *Ixodes ricinus* expended from 16% to 25% of the Swiss territory, and a subsequent extension for *Chlamydiales* bacteria is observed from 8% to 9.3%. *Ixodes ricinus* expansion occurred all over the Swiss Plateau and toward higher altitude in the alpine valleys and was more extended in the South-West. Newly available habitat concerned mostly grass and forest areas. Extension of *Chlamydiales* followed similar trends, restricted to forest areas. As *Ixodes ricinus* presence is favoured by higher temperature, we might expect that, in the future, this expansion might continue following global warming with some limitation by dryer conditions at lower altitude.

Our results agrees with the observed increased cases of tick-borne encephalitis (TBE) in Switzerland, that spread from eastern to western part of Switzerland (de Vallière and Cometta, 2006), leading to the extension of the vaccination recommendation (OFSP, 2013, 2019). Similar tick's expansions towards higher altitudes were observed in other European countries during the last decades (Daniel *et al.*, 2003; Skarphédinsson *et al.*, 2005; Jore *et al.*, 2011), notably in association with milder winters and extended spring and autumn seasons (Lindgren E *et al.*, 2000; Medlock *et al.*, 2013).

Variables explaining *I. ricinus* distribution

The effective variables identified by our model are related to temperature and humidity, which reflects well the tick's ecology. We found that a high temperature favours *Ixodes ricinus*, in agreement with previous studies (Estrada-peña, 1999; Porretta *et al.*, 2013). However, our analysis indicated that this relationship does not hold during driest months. This can be explained by an increased evaporation of the soil humidity under warmer temperature, thus accentuating the desiccation risk for ticks (McCoy and Boulanger, 2015). The NDVI variables, an important contribution to our model, are indicators of physiological plant activity and have often been shown to be powerful for modelling the presence of ticks as they reflect humidity conditions (Estrada-peña, 1999; McCoy and Boulanger, 2015). Nevertheless, our results indicated that the ambient relative humidity variables showed limited effect on the model. They may thus constitute a less precise predictor of soil humidity than the combination of NDVI variables with temperature and land cover indicators. Surprisingly, our results also showed that *I. ricinus* presence is favoured by a higher percentage of artificial surfaces. This might relate to an overrepresentation of ticks collected in vegetated areas situated within urban settlements or close to roads. Indeed, we expect a sampling bias as many tick occurrences comes from the Tick Prevention App., in which users provide tick locations that are likely biased towards areas closer to roads or paths and thus artificial surfaces. Moreover, the other tick occurrences, either provided by the army field campaign in 2009 or by the prospective campaign in 2018, were collected essentially in forests or close to their borders. On the contrary, grass areas, often corresponding to agricultural fields, were not sampled by the two field campaigns and were also probably less explored by the users of the application, since people are less likely to visit these areas. This might explain why our model associated a low percentage of grass vegetation as favourable for *I. ricinus* and we might have an underestimation of the suitability index in some grass areas. Nevertheless, the presence of ticks in urban and suburban areas of Switzerland has already been reported (Rizzoli *et al.*, 2014; Oechslin *et al.*, 2017) and the presence of vegetated areas in urban settlement, or close to artificial surfaces

(roads, paths, recreational areas) may constitute favourable habitats. In addition, even if we may expect some grass zones, especially at the forest border, to be highly favourable for ticks, in general, land pasture, open land and cultivated areas have been reported to be much less favourable than woodlands (Aeschlimann *et al.*, 1979; Huss and Braun-Fahrländer, 2007; McCoy and Boulanger, 2015). Finally, in agreement with previous studies (Estrada-Peña *et al.*, 2015; Hauser *et al.*, 2018), we observed that the morphometric parameters and the precipitation variables show little effect on the suitability for ticks.

Variables for *Chlamydiales* spatial distribution

Identified effective variables for the presence of *Chlamydiales* may provide novel insights to the bacteria's ecology. First, our results indicated that *Chlamydiales* are more likely present in ticks collected in forests or grass fields than in ticks collected close to artificial areas. The highest prevalence of *Chlamydiales* within natural areas could be explained by the presence of different hosts (likely rodents) on which ticks feed, with potentially a highest number of reservoir-competent hosts for *Chlamydiales* in natural areas. This may also relate to a higher tick abundance in natural areas, which is known to be associated with a higher prevalence of other pathogens in ticks (Aivelo *et al.*, 2019) but not for all tick pathogens (Oechslin *et al.*, 2017). Our results also showed that the presence of *Chlamydiales* bacteria is favoured by driest conditions (negatively associated with the number of days with a relative humidity superior to 90% and positively associated with the number of days with relative humidity inferior to 70%). High amount of precipitation during the coldest months also appeared to be favourable for the presence of *Chlamydiales*. Several suitable areas for *Chlamydiales* are predicted at an altitude higher than 1000 m, thus highest precipitation during the coldest months could be associated with largest snow amounts, preserving the soil from frost and leading to a highest tick's survival (Lindgren *et al.*, 2006). Finally, a shorter distance to wetlands was also highlighted as a factor favouring the bacteria's presence. Several *Chlamydiales* have been considered symbionts of amoebae (Kebbi-Beghdadi and Greub, 2014), which are free-living organisms usually found at the interface between water and soil, air or plants (Kebbi-Beghdadi and Greub, 2014). It is therefore likely that amoebae can be found in wetlands, which might favour the transmission of *Chlamydiales* to various animal hosts on which ticks feed.

Chlamydiales prevalence values were heterogeneous among our datasets. In 2009, ticks were collected in forests only and *Chlamydiales* were present in 68.6% of the sites visited with a low prevalence within pools (6.4%). Low prevalence was also observed in the ticks received by the users of the Tick Prevention App in 2018 and 2019 (3.79%). In 2018, the ticks sampled during the prospective campaign were also mainly collected in forest areas and *Chlamydiales* were present in 53.7% of the site but with much higher prevalence reaching 28.13%. This rate reflects values obtained in 2010 in one specific site in the Swiss Alps (Rarogne), where *Chlamydiales* prevalence rate of 28.1% was found in 192 pools collected in forests and meadows (Croxatto *et al.*, 2014). Differences between year 2009 and 2018 could be explained by a difference in the time and sampling areas (we excluded potential PCR contaminations, see Annex A2.9). As infected ticks were already present in most forest sites in 2009, spread of infection might have occurred between 2009 and 2018. Then, ticks from Tick Prevention App were collected in sites more closely related to artificial areas, which we have shown reduces the prevalence of the bacteria.

On the importance of considering the spatial and temporal scale of the environmental variables

For *I. ricinus,* the most performant models are obtained when extracting the environmental variables in a buffer with a radius of 100 or 200 m (corresponding to an area of 9 ha to 25 ha around the sampling point). This can be explained by the ecology of the species. First, the establishment of a population of ticks will probably need a suitable area that is large enough. Moreover, the presence of ticks strongly depends on the presence of hosts, which disperse across larger areas and may thus be influenced by the climatic conditions observed at some distance. Our results also indicate that buffer radius larger than 500 m (corresponding to areas larger than 121 ha) are not improving our model. This might relate to the dispersal range of tick hosts, likely rodents, which is usually smaller (among the long dispersal hosts, the roe deer dispersal is estimated to cover around 50 and 100 hectares (Cederlund and Liberg, 1995)). In addition, the most performant models are obtained when considering the climatic conditions of the two- or three- years preceding sampling date. This time period appears to be relevant as it corresponds to the estimated duration of the life cycle of ticks (McCoy and Boulanger, 2015).

For the modelling of *Chlamydiales* bacteria, small buffer (≤200m) and a short time period (one year or less) is favourable for some variables, whereas for some others, to consider a larger buffer (1 km or 1.5 km) and a longer time period (2-3 years) is better. Some variables might be influencing locally the establishment of the tick species and the ability for the bacteria to colonize and/or reproduce within it, whereas other variables may be related to the interaction of the tick with the hosts on which it feeds, that may disperse in a larger area and thus be influenced by climatic conditions at a larger scale.

Our results thus highlighted the importance of considering the environment around the sampling point for a good variables estimation in species distribution model, while single point is commonly considered (Elith *et al.*, 2006; Williams *et al.*, 2015; Raghavan *et al.*, 2016, 2019, 2020; Sage *et al.*, 2017; Minigan *et al.*, 2018; Soucy *et al.*, 2018; Eisen *et al.*, 2018; Hadgu *et al.*, 2019). Our results also showed that the time period considered before the sampling date, with sliding windows, has a significant impact on the performance of the resulting models. This should be favour over using an average of the climatic conditions over the sampling period (Bradley *et al.*, 2010; Williams *et al.*, 2015) or any larger period of time (as Worldclim climatic data from 1950 to 2000 which are commonly used for species distribution modelling (Porfirio *et al.*, 2014; Manzoor *et al.*, 2018)). Previous studies already suggested the use of multi-grain approaches involving various spatial resolutions to consider variables affecting the presence of a species at different scales (Meyer and Thuiller, 2006; Meyer, 2007; Mertes *et al.*, 2020). This adds to the recommendation of using data based on species ecology rather than on availability (Mayer and Cameron, 2003; Meyer, 2007). In addition, our results showed that the temporal scale of the environmental predictors should be accounted for.

Model performance

Ixodes ricinus distribution models are robust as they allowed a good discrimination between presences and randomly generated points and correctly predicted the presences of *I. ricinus* observed in an independent dataset. *Chlamydiales* distribution models are more difficult to validate due to the limited amount of data and poor knowledge regarding their distribution. Nevertheless, our model performed relatively well for the data collected in 2018 as most of the occurrence locations had higher suitability index than the locations where no *Chlamydiales* were identified. Year

2009 did not show such trend as many locations where no *Chlamydiales* were found were predicted as potentially suitable. This might be due to an absence of *Chlamydiales* colonisation of these sites at the sampling time despite favourable conditions.

Our investigations considered mainly environmental factors. However, other factors such as species interaction and species life history traits might influence the presence of both the ticks and their bacterial pathogens (Guisan and Zimmermann, 2000; Clay *et al.*, 2008; Estrada-Peña, 2008; Büchi and Vuilleumier, 2014; McCoy and Boulanger, 2015; Ehrmann *et al.*, 2018) . Also, additional abiotic factors might play an important role, such as landscape fragmentation and barriers that can limit dispersal of ticks hosts (Estrada-Peña, 2008; McCoy and Boulanger, 2015) or disturbances that can drive local populations to extinction (Vuilleumier *et al.*, 2007).

The precision of our predictions is limited by the precision of the data used. The interpolated climatic grids used were produced based on weather stations measurements and thus contain interpolation uncertainties that may influence the models results (Guisan and Zimmermann, 2000). Also, with interpolated grids, the inherent collinearity and autocorrelation may lower the reliability of the results (Estrada-Peña *et al.*, 2015). Finally, the occurrences data are probably prone to sampling bias and do not represent a random sample of the population being studied, which can also influence the predictions (Araújo and Guisan, 2006; Merow *et al.*, 2013), probably leading to an overestimation of suitability index in urban and artificial areas as compared to natural ones.

2.3.7 Conclusion

Both *Ixodes ricinus* and *Chlamydiales* are causing a potential threat to human health and their prevalence are currently increasing in Switzerland, with a strong expansion of ticks in forests but also in urban and suburban areas. Ticks' expansion has already recently alarmed the Public Health Services (OFSP, 2019), and this expansion is predicted to continue in the future due to global warming. In this context, our results offer a unique tool to identify precisely locations where diseases are likely to spread, to colonize new sites and to increase in prevalence. Maps as developed here, and associated methods, could thus bring critical information for decision-makers to control tick-borne diseases and target prevention campaigns.

Our methodological framework allowed a coherent identification of environmental factors influencing the presence and distribution of both *Ixodes ricinus* tick and their *Chlamydiales* bacteria in Switzerland, and enabled the mapping of suitability evolution across Switzerland from 2009 to 2019. Our results highlighted an important increase of suitable areas for both species and predicted their extension towards higher altitude. Our investigations consist in the first exploratory analysis of the environmental factors influencing the presence of *Chlamydiales* bacteria within ticks in Switzerland, showing an application of species distribution models to study the nested niche of a parasite within the ecological niche of its host. Finally, our study demonstrated the importance of considering the spatial and temporal scale of the environmental variables used for species distribution models.

Spread of pathogens through a vector is at the origin of major epidemics and infectious diseases, and affects humans, wildlife, and agriculture. We proposed a methodological framework based on geographical system able to provide deep insights on factors affecting patterns of disease emergence by providing a better characterisation of the spatial distribution of their vectors. This method can be applied to a wide range of host-pathogen association to identify their spread and

distribution, which is expected to bring critical information for a better understanding and control of pathogens.

2.3.8 Acknowledgments

We thank Dr. Dirk Shmartz from the Swiss Federal Institute for Forest, Snow and Landscape Research, for computing and providing on demand the high resolution climate grids; Werner Tischhauser, Prof. Jürg Grunder and A&K Strategy for providing an access to the data of their smartphone application (Tick Prevention, https://zecke-tique-tick.ch); Rahel Ackermann-Gäumann for the tick data from the Swiss Army field campaign and Ludovic Pilloux for advices regarding the *Chlamydiales* dataset from this same field campaign.

2.3.9 Code availability

The main R codes developed for this study are available on GitHub:
https://github.com/estellerochat/SDM-Chlamydiales.

Chapter 3 CONNECTIVITY AND GENETIC DIVERSITY

3.1 Research context

In the previous chapter, we showed how species may be confronted with a degradation of their habitat suitability due to environmental changes. We presented ecological niche modelling as a tool to identify such affected populations, threatened either by the loss of suitable conditions, or by the increased presence of competitors, predators or pathogens in their favourable habitat. Under such conditions, to survive, concerned populations may either move to more suitable areas or adapt to their new environment. However, both dispersal and adaptation are limited by landscape fragmentation.

3.1.1 Landscape fragmentation: a limit to dispersal and adaptation

During the last decades, population growth has led to an increased demand for housing, transport and infrastructure (Steffen *et al.*, 2007). Responses to this demand have involved a strong intensification of urbanization, associated with a significant increase in artificial surfaces (buildings, roads, recreational areas, etc.). In addition, agricultural practices have been modified, shifting towards more intensive production systems. These land use modifications have a negative effect on biodiversity by causing the disappearance of many habitats and creating various barriers to dispersal (Ray *et al.*, 2002). As a result, large habitat patches are divided into smaller, more isolated fragments, which is termed "**landscape fragmentation**". Due to the new barriers between fragmented habitats, some food resources become inaccessible or increasingly distant, as well as important territories, such as reproductive areas (Di Giulio *et al.*, 2009; Fu *et al.*, 2010).

As a consequence of habitat loss, the potential niche of a species may totally disappear from a territory, as discussed in the previous chapter. In other cases, the potential niche may subsist, but divided into small patches as a result of fragmentation. In this case, population survival depends on the connectivity between the remaining habitat patches, which may first allow dispersal into neighbouring habitats and colonisation of newly suitable areas that may emerge following environmental changes. In addition, connectivity is essential to maintain gene flow. Indeed, a decrease in the size of habitats is often associated with a decrease in the size of animal and plant populations living there (Fahrig, 2003a). At the genetic level, a reduced population size limits gene

flow and can lead to a loss of genetic diversity (Hitchings and Beebee, 1997; Fahrig, 2003a; Coulon *et al.*, 2006). To maintain a sufficient level of genetic mixing, and subsequently of genetic diversity, it is thus essential to maintain connectivity between habitat patches, which allows the reproduction of individuals from neighbouring populations (Figure 3-1).

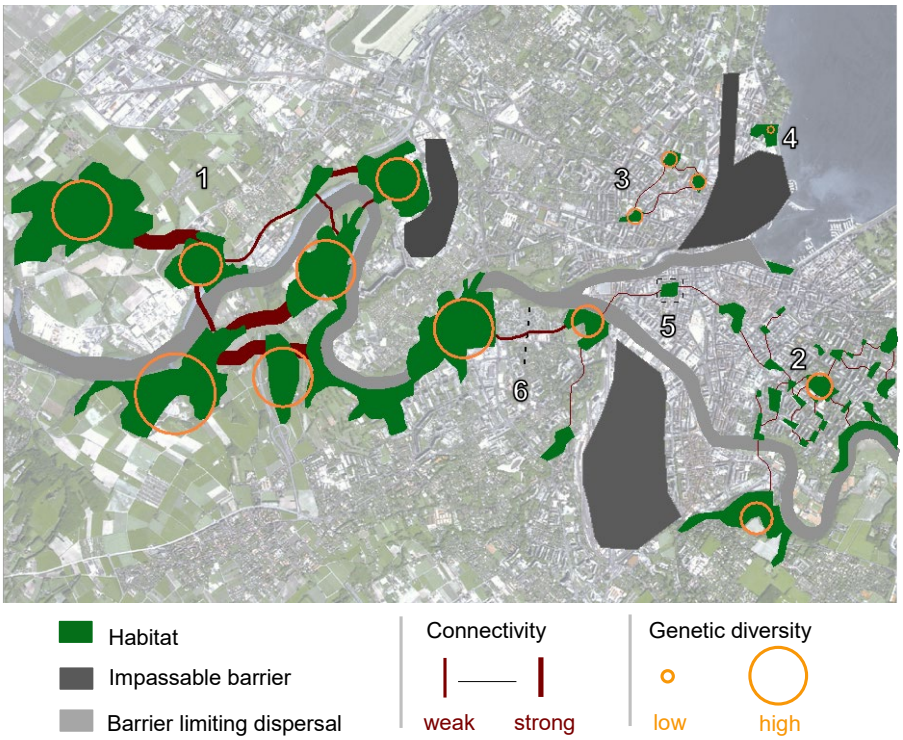


Figure 3-1 - Vulnerable populations due to fragmentation
This figure first shows a peripheral area containing large and well-connected habitats (no. 1). In this region, numerous flows of individuals are possible between habitat patches, allowing genetic exchanges favourable to maintaining of a high level of genetic diversity. In urban environments, on the other hand, fragmented habitats are small (no. 2, 3 and 4), which limits the size of populations living there. If connectivity between these small habitat patches and the larger environments of the periphery is maintained (no. 2), relatively stable genetic diversity can be preserved due to exchanges with periphery populations. On the other hand, if connectivity with the periphery is no longer ensured (no. 3), genetic diversity could be threatened. At the extreme, some small habitat patches could become totally isolated (no. 4), reducing the chances of survival of populations in these environments. When carrying out development projects (e.g. building, roads, etc.), it is therefore important to ensure that connectivity is maintained between habitat patches of dense urban centres and larger areas in the periphery. In particular, projects should avoid removing important habitats acting as relays in the urban environment (no. 5) or cutting the links ensuring connectivity with the periphery (no. 6). Note: this is a fictitious example, and it is not based on any real data.

3.1.2 Adaptive capacity

Adaptation of populations relies on phenotypic changes (modification of behaviour, morphology or physiology), which can be induced either by phenotypic plasticity or genetic evolution (Merilä and Hendry, 2014; Fox *et al.*, 2019). Phenotypic plasticity results from the capacity to change the phenotype when exposed to different environmental conditions, without any genetic modification. This can allow for rapid adaptation to changes in the environment over the lifetime of individuals (Fox *et al.*, 2019). For example, Charmantier *et al.* (2008) reported a plastic adaptation of the great tits, who advanced their breeding and egg-laying period in response to changing temperatures that induced an earlier peak of winter moth larva, an important food resource for great tits offspring. Similarly, Lu *et al.* (2018) showed that the Arabidopsis plant can modify its flowering time depending on the rhizosphere microbiota composition. However, since phenotypic plasticity does not rely on hereditary genetic variations, it will not necessarily ensure the persistence of adaptation to subsequent generations (Harrisson *et al.*, 2014). On the contrary, genetic evolution by natural selection allows the favourable traits to be preserved in populations due to a modification of hereditary genetic variants. Conserving the potential for genetic adaptation is therefore essential for population persistence.

3.1.3 Adaptive evolution and genetic diversity

The foundations of adaptive evolution theory were initiated by the work of Darwin at the end of the 19th century (Darwin, 1859). Based on his studies in the Galapagos, he observed that birds living on different islands had different beak characteristics (shape and length), which he associated to their food resources (seeds, cactus, fruits, insects, etc.). Following these and other observations, Darwin concluded that evolution was the result of **natural selection**. Due to this selection, organisms with a trait that is better adapted to their environment will have a survival advantage or better reproductive capacity, allowing favourable traits to remain in the population for future generations. Later, Mendel's studies on pea genetics highlighted the relationship between expressed traits and genetic characteristics, which demonstrated the heredity of genetic variants (Mendel, 1865). The more well-fitted individuals are thus able to transmit favourable traits to their offspring through an inherited genetic variant, which ensures the persistence of adaptation in subsequent generations.

New genetic variants in populations may arise from **mutations**, which are accidental changes in DNA. However, mutation rates are generally low and adaptation to rapid environmental changes thus largely depends on the amount of genetic variants already present in populations, i.e. standing genetic variations (Willi and Hoffmann, 2009). As greater genetic diversity indicates the presence of more genetic variants in a population, it also highlights a higher adaptive potential associated with a greater likelihood of finding a variant better suited to new conditions (Allendorf and Leary, 1986). Conversely, low genetic diversity may reduce the adaptive potential of the population, and increase the risk of extinction (Frankham, 2005). For example, it has been shown that a decline in genetic diversity of the Glanville Fritillary butterfly preceded its extinction (Saccheri *et al.*, 1998; Fountain *et al.*, 2016). Similarly, Bozzuto *et al*. (2019) showed that inbreeding substantially reduced the growth rate of Alpine ibex populations. Therefore, measuring genetic diversity can help to identify vulnerable populations.

3.1.4 Measuring the genetic diversity

Definitions: DNA structure and genetic information

The genetic information of individuals is stored in the deoxyribonucleic acid (**DNA**). DNA is composed of **nucleotides,** which are organic molecules formed by a basis of carbon and phosphate atoms, to which a nitrogenous base is attached (Figure 3-2). Four types of nucleotides exist: adenine (A), cytosine (C), guanine (G) and thymine (T). DNA molecules assemble on two parallel DNA-strands, forming a DNA-helix, according to a pre-defined assembling rule: A always assembles with T, and C with G. Some of the sequences of the DNA-helix, called **genes**, code for the formation of the amino-acids that will form **proteins**.

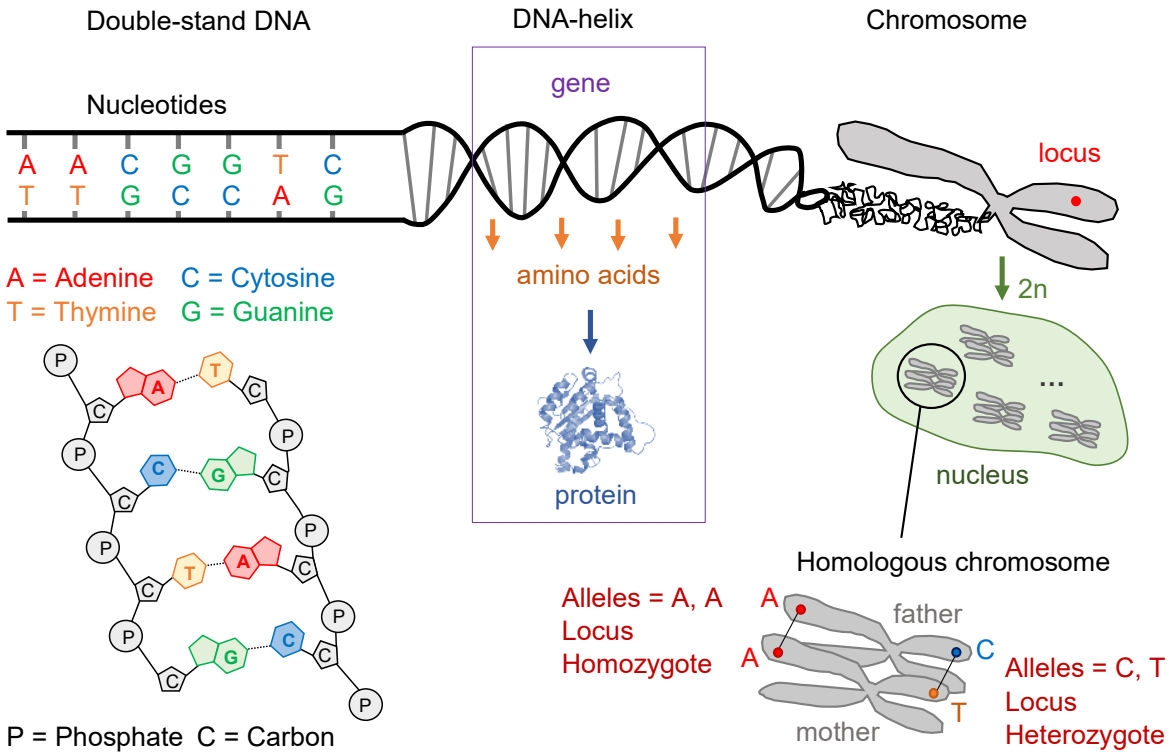


Figure 3-2 - DNA structure

The DNA-helix is subsequently enrolled into **chromosomes** stored in the cell nucleus (Figure 3-2). Each species contains a given number of chromosomes, which together constitute its **genome** that represents all of its genetic information. Many species, including humans, are **diploid**, i.e. they possess an even number (2n) of chromosomes, which are composed of n pairs of two homologous chromosomes, one obtained from the mother and the second from the father. A given position along the genome (consisting of one or a succession of nucleotides) is called a **locus** (plural: loci) and at a given locus, the nucleotide sequence observed on a chromosome is the **allele**. Following, a diploid individual possessing the same allele on the two homologous chromosomes is called **homozygote** for this locus, whereas an individual with two different alleles is **heterozygote**. In addition the allele combination (e.g. AG) for an individual is its **genotype.** When comparing several individuals of the same species, one can identify loci where different genotypes

are observed for different individuals (for example, one individual is AA and the other is AG). These variations are called **polymorphisms**.

Genetic markers

In the late 1970s, the development of sequencing techniques made it possible to analyse the genome using genetic markers. Genetic markers are DNA sequences whose position in the genome is known and can be used to identify or compare individuals, for example through the identification of polymorphisms. Several types of genetic markers exist and differ notably as regards the technical requirements for their analysis. Two of them, which are used later in the thesis, are presented below: Amplified Fragment Length Polymorphisms (AFLP) and Single Nucleotide Polymorphisms (SNP).

Amplified Fragment Length Polymorphism (AFLP)

AFLP markers (Vos *et al.*, 1995) are multi-locus markers produced by a method in which DNA is cut by two enzymes and specific fragments are amplified by a Polymerase Chain Reaction (PCR), a technique for producing multiple copies of a DNA sequence. The amplified fragments are then analysed using fluorescence techniques that allows for the detection of the presence or absence of a given allele at a polymorphic-site. AFLPs are thus dominant markers, i.e. they only enable the detection of the presence or absence of a given allele and do not provide information regarding the alternative allele. Consequently, they cannot be used to differentiate a homozygote from a heterozygote with the same allele.

Single Nucleotide Polymorphism (SNP)

SNP markers allow for the identification of single nucleotide variations (e.g. A replaced by G for some individuals). Comparing the genome of several individuals of the same species allows the identification of the position of SNPs. For some species, these known positions have been used to define chips that directly target sites of interest. These chips contain DNA fragments that correspond to the nucleotide sequence directly preceding or following a SNP. When mixed with single-stranded DNA, they associate with the sequence surrounding the SNP. The last nucleotide of the adaptor fragment corresponds to an alternative allele of the SNP and another fragment contains the other alternative (Gunderson *et al.*, 2005). By analysing the assembly results (using fluorescent methods), it is possible to identify the alleles that were present in the DNA studied.

Indexes of genetic diversity

The information provided by genetic markers can be used to quantify genetic diversity using several indexes. Three indexes used in section 2.3 are presented below.

Observed heterozygosity

The observed heterozygosity is the percentage of loci that are heterozygous. A high level of observed heterozygosity therefore indicates high genetic variability. For a population containing N individuals, the observed heterozygosity in the population can be calculated using Formula 3-1:

$$H_{obs} = \frac{1}{k}\sum_{i=1}^{k} h_i$$

Formula 3-1

where *hi* is the fraction of the *N* individuals that are heterozygous for the marker *i*, and *k* is the total number of markers considered. This value ranges from 0 (no individual heterozygous for any marker, no genetic diversity) to 1 (all individuals are heterozygous for all markers, high genetic diversity).

Expected heterozygosity

The expected heterozygosity corresponds to the heterozygosity expected under the Hardy-Weinberg equilibrium (HWE), a law that states that in an ideal population of infinite size, with random mating and without evolutionary influences (migration, natural selection, mutation, etc.), the allele frequencies will reach an equilibrium and then remain constant over generations. In the case of two alternative alleles at one locus (e.g. A and G), the equilibrium frequencies are as follows:

- p^2 is the frequency of homozygote of the first allele (for example: AA)
- q^2 is the frequency of homozygote of the second allele (for example: GG)
- 2pq is the frequency of heterozygote (for example: AG)

Consequently, the expected heterozygosity level in a population of *N* individuals can be calculated using Formula 3-2.

$$H_{exp} = \frac{1}{k}\sum_{i=1}^{k} 2p_i q_i$$

Formula 3-2

where, *pi* (respectively *qi*) is the frequency of presence of the first (respectively second) allele for the marker *i* among the *N* individuals, and *k* is the total number of markers considered.

Inbreeding coefficient

The inbreeding coefficient is defined from the comparison of observed and expected heterozygosity (Wright, 1949). When the observed fraction of heterozygotes is much lower than expected under a population with random mating, the population is facing inbreeding. The Formula 3-3 can thus be used to estimate the amount of inbreeding in a population.

$$F = \frac{H_{exp} - H_{obs}}{H_{exp}}$$

Formula 3-3

3.1.5 Conservation of genetic diversity

Following the development of genetic markers and the availability of genetic information, **conservation genetics** has been developed to define strategies to preserve genetic diversity, notably by favouring gene flow between populations, hybridization or reintroductions (Frankham, 2010).

For example, Diekmann *et al.* (2010) studied a threatened population of seagrass and used genetic data to identify an external population strand with high genetic diversity and genetically close to the threatened population. This strand could thus constitute a successful donor population for seed transplantation, thus increasing the genetic diversity of the threatened population. Raeymaekers *et al.* (2008) studied fish population genetics to determine the sections of a river that constitute the main barriers to gene flow and thus should be prioritised in restoration efforts. Piry *et al.* (2018) studied the genetic diversity of a critically endangered grasshopper species, endemic to a steppe habitat that had been highly reduced and fragmented over recent decades. They highlighted that less intensive sheep grazing may allow for gene flow to increase among the remaining populations and thus contribute to their preservation. Montero *et al.* (2019) showed that the creation of dispersal corridors aiming to restore connectivity between fragmented forest habitats allowed for an increase of gene flows for mouse lemurs in Madagascar.

Nevertheless, the application of conservation genetics to practical field projects is still very limited, and the need to develop tools for easier integration of genetics into conservation practices has often been emphasised (Shafer *et al.*, 2015; R. Taylor *et al.*, 2017; Britt *et al.*, 2018; Holderegger *et al.*, 2019). Some limitations to the use of conservation genetics by conservation professionals include a lack of genetic background, as well as financial restrictions (Holderegger *et al.*, 2019). Indeed, collecting genetic information in the field can be costly and time-demanding (Epperson *et al.*, 2010). In such situations, analysis of landscape connectivity can provide a first indication of potential gene flow and help identify areas particularly threatened by a loss of genetic diversity. In addition, landscape connectivity analyses may be used to simulate genetic data. The following sections thus present several alternatives for estimating landscape connectivity (section 3.1.6) and a software for simulating genetic data (section 3.1.7).

3.1.6 Estimating the landscape connectivity

Definitions: types of connectivity

Different types of connectivity have been defined. **Structural connectivit**y is based solely on the analysis of landscape structure (Fu *et al.*, 2010). It thus depends on physical properties such as the size and shape of habitat patches and the distance between them (Calabrese and Fagan, 2004), but is not related to any particular species and does not take into account ecological properties (Saura *et al.*, 2011), such as the dispersal capacity of individuals in the landscape. As a result, the analysis of structural connectivity requires little initial data, but its value for ecological analyses remains limited (Avon and Bergès, 2013) as it provides little indication of the real capacity of species to disperse between habitats (Calabrese and Fagan, 2004). **Potential functional connectivity**, on the other hand, incorporates information on the dispersal capacity of a given species, taking into account the Euclidean distance between habitats as well as landscape features that may constitute a barrier or, on the contrary, be favourable to the movement of individuals (Calabrese and Fagan, 2004). This connectivity does not necessarily correspond to the **real functional connectivity**, which is based on the observation of the actual movements of individuals (Avon and Bergès, 2013), for example with telemetry tools. However, potential functional connectivity can be estimated to a large extent, while requiring a limited amount of initial data (Calabrese and Fagan, 2004). One method for analysing potential functional connectivity is based on the use of raster resistance maps.

Resistance maps

Resistance maps are used to represent the influence of the landscape on species dispersal. These maps are usually defined in the form of a raster file in which each pixel is assigned a resistance value. Pixels corresponding to features that are more difficult to cross, either because of the associated high mortality risk or because of high energy demand, receive a higher resistance value (Ray *et al.*, 2002). From the resistance map, a cost can be associated with each path by summing the resistances of each pixel crossed (Figure 3-3). This cost makes it possible to define the shortest path (or least cost path) between two habitats based, among other things, on the energy consumption necessary for movement and reproduction. The cost of this path provides information on the probability of dispersal of the species between the two points. Different elements can be taken into account on the resistance maps, including land cover, slope, orientation, or any other factor that may influence the dispersion of the species studied.

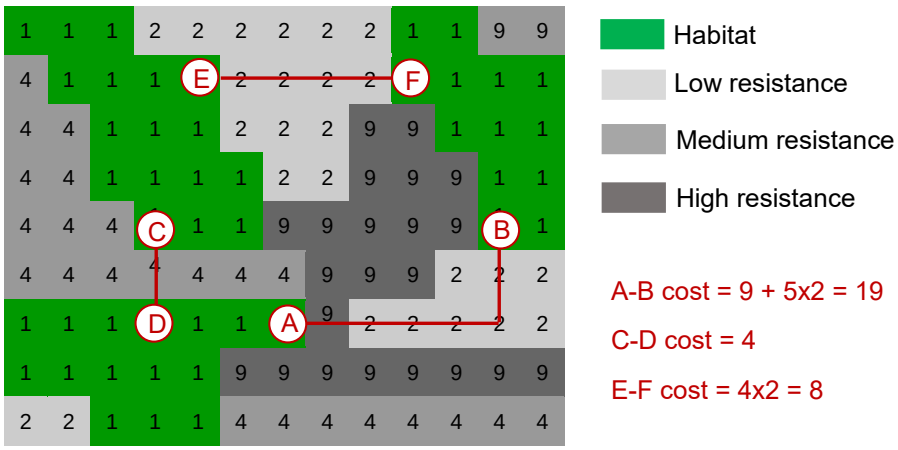


Figure 3-3 – Resistance maps
Resistance values are assigned to each pixel as a function of the difficulty to cross it by the species. The cost of the path between two points correspond to the sum of the resistance of the pixels crossed.

Defining resistance values is not easy and remains a sensitive issue that can significantly influence the results of connectivity analyses (Rayfield *et al.*, 2009; Zeller *et al.*, 2012). Very diverse ranges of resistance values can be observed in the literature, for example from 1 to 8 (Fu *et al.*, 2010), 5 to 80 (Ray *et al.*, 2002), 1 to 100 (Sutcliffe *et al.*, 2003), 1 to 1000 (Clauzel *et al.*, 2013) or 1 to 5000 (Girardet *et al.*, 2013). In general, the resistance of a favourable habitat is given as 1 (Girardet *et al.*, 2013) as the value 0 should not be used if the Euclidean distance between habitats is to be taken into account, which is also a limit to dispersion. Beside these generalities, resistance values can be assigned any real value and are generally estimated from the literature and supported by expert knowledge about dispersal and life history of the species under study (Zeller *et al.*, 2012). They can also be estimated from real measurements of individual movements or gene flow when such data are available (Cushman and Lewis, 2010; Braaker *et al.*, 2017). Finally, they can be computed from species distribution maps derived by ENMs (see Chapter 2.) that delimit favourable habitats and provide suitability values for all landscape features (Brown, 2014; Brown *et al.*, 2016; Rana *et al.*, 2019). However, this should be done in combination with raw environmental data that can provide additional information for defining resistances, such as the presence of impassable barriers. In any case, resistance values depend on the species under consideration, as resistance caused by the environment can be highly variable and a barrier for one species may constitute a dispersal corridor for another (Avon and Bergès, 2013).

Landscape graphs

Once resistance maps are defined, potential connectivity analyses can be performed using landscape graphs. Graph theory is used in many fields of geography and computer science (transport networks, mobility analyses, shorter path problems, network optimisation, etc.)(Urban and Keitt, 2001). More recently, this discipline has been extended to landscape graphs, which have emerged as a method for modelling ecological networks (Foltête, Clauzel, Girardet, *et al.*, 2012). In landscape graphs, nodes correspond to the habitat patches of interest or specific points in the territory (Galpern *et al.*, 2011). The links in the graph represent the potential flows of individuals between the different nodes. A link is therefore created between two patches if individuals of the species under consideration are able to cross the space between these two habitats, according to the defined resistances (Urban and Keitt, 2001). Landscape graphs provide a good representation of potential flows and have a low requirement in terms of initial data, they can therefore be easily applied on a large scale and for different species (Calabrese and Fagan, 2004; Avon and Bergès, 2013).

Software

Connectivity analyses using resistance maps and landscape graphs have been implemented in various software. Among these, the UNIversal CORidor (UNICOR) network simulator use resistance maps to measure potential functional connectivity between specific points in the landscape (Landguth, Hand, *et al.*, 2012). It uses the Dijkstra algorithm to calculate the least cost paths by accounting for resistance. PathMatrix (Ray, 2005) works very similarly, but also offers the alternative of using a set of polygons to represent habitat patches instead of only points. Like UNICOR and PathMatrix, the Graphab software (Girardet *et al.*, 2016) measures the potential functional connectivity and computes least-cost paths by taking into account the resistances constituted by the environment. Graphab offers the advantage of performing these calculations from a landscape map (e.g. a land cover map or a map combining land cover and relief data), thus avoiding the need to define specific points of the territory or habitat polygons. Using a different approach, CircuitScape was developed using electrical circuit algorithms (McRae *et al.*, 2013). Like UNICOR and PathMatrix, it allows for the measurement of flows between specific points in the territory. However, unlike the two previous software, CircuitScape takes into account all possible paths and not only the shortest path between two points. Nevertheless, due to the memory and the computational demand, the calculations are limited to a restricted extent.

3.1.7 Genetic simulations

The Cost Distance POPulations (CDPOP) software is a tool for simulating gene flow in the environment and the genetic evolution of populations over time (Landguth and Cushman, 2010). This individual-based model simulates the birth, death, dispersal and mating of individuals as a function of the landscape resistance and biological parameters of the species studied: mortality rate, number of offspring, mutation rate, etc. The habitats of interest must be provided in the form of points that can either be occupied by a maximum of one individual or unoccupied at the beginning of the simulations. CDPOP then simulates the dispersal of each individual, as well as the mate selection, using probabilistic functions of the cost of the shortest path between two habitats. At each generation, a maximum of one individual will be conserved at each habitat point. Additional individuals may migrate to other free points, if the cost of travel allows it, or die. Breeding is sim-

ulated with a Mendelian inheritance of genetic variants and for each generation, the results provide the genotype of the individuals present at each habitat point. These results make it possible to estimate the genetic diversity of populations and its evolution over time.

3.2 Scientific contribution

3.2.1 Problem statement

Urban biodiversity

Urban areas contain a large variety of environments, providing habitats for many native species, including rare and threatened ones (Araújo, 2003; Kantsa *et al.*, 2013; Ives *et al.*, 2016). As a result, cities can show a high level of biodiversity, that may be comparable or even higher than that of rural areas (Kowarik, 2011; Kantsa *et al.*, 2013). Since urban areas are expanding as a result of population growth, they become increasingly important for biodiversity conservation. In addition, the conservation of urban biodiversity is key to favouring a better quality of life for residents, including well-being and better health conditions (Maller *et al.*, 2006; Lee and Maheswaran, 2010; Shanahan *et al.*, 2015). Indeed, Bolund and Hunhammar (1999) highlighted six ecosystem services provided by urban biodiversity : air filtration, micro climate regulation, noise reduction, rainwater drainage, sewage treatment, and recreational and cultural values. Nature in urban settlement may thus influence human health directly, for example by reducing pollution and limiting respiratory complications (Lovasi *et al.*, 2008) or indirectly, for example by encouraging physical exercise (Timperio *et al.*, 2008). For these reasons, it is of great importance for urban authorities and planners to adapt the way they design cities and plan for urban change, particularly to identify potential impacts on biodiversity, especially for native species.

Conservation of urban biodiversity

Despite the importance of urban biodiversity, studies on ecological conservation in urban environments are very limited (Miller and Hobbs, 2002; Fazey *et al.*, 2005; Manel and Holderegger, 2013). Indeed, although Holderegger *et al.* (2019) highlighted that the study of fragmentation and connectivity is one of the two main conservation genetic topics of current interest to conservation practitioners, highly fragmented urban environments still receive lower attention than natural areas. In addition, conservation prioritisations usually allocate limited conservation value to urban environments and suggestions for urban conservation are often received with scepticism by the general population and conservation managers (Soanes *et al.*, 2019). This may derive from the common idea that urban environments cannot be suitable for long-term conservation outcomes when associated with urban activities (Soanes *et al.*, 2019). This misconception partially arises from the fact that most conservation strategies highlight the importance of large patches, while small and fragmented habitat patches are rarely protected (Kendal *et al.*, 2017). However, small patches can support a high diversity of species. Oertli *et al.* (2002), for example, showed that a set of small ponds may contain more species and have a higher conservation value than a single large pond of the same total area. Similarly, Kendal *et al.* (2017) reported that many small urban grasslands contained more species than a few large reserves. Of note, small habitats can help to maintain connectivity, e.g. by acting as stepping stone habitats (Bierwagen, 2006; Serret *et al.*, 2014) and thus maintain gene flow. Conservation measures focusing on maintaining natural habitats and connectivity within urban settlements should therefore be given greater consideration.

Impact of fragmentation

In this context, the identification of urban plant and animal species endangered by the ongoing fragmentation of habitats is essential in order to promote more sustainable urbanisation processes or conservation strategies in the future. Nevertheless, estimating the impact of urbanisation and fragmentation on species remains complex. On one hand, fragmented urban habitats can limit the survival of some species, while on the other, harbour self-sustaining populations of native (and exotic) species (Kowarik, 2011) that are able to adapt to the human-influenced environment or even take advantage of proximity to humans that may provide food sources, reduce the presence of wild predators or provide new refuges (McKinney, 2002; Shochat *et al.*, 2010). In addition, the impact of fragmentation can be highly variable depending on the dispersal mode. For example, plants pollinated by many insects may be only moderately impacted by urban fragmentation (Culley *et al.*, 2007), and may not require large patches of habitat to survive. Similarly, species that can benefit from human-mediated dispersal (attachment to clothes, vehicles, shoes, soil movements, etc.) may be particularly well adapted to urban landscapes (Banks *et al.*, 2015; Egizi *et al.*, 2016). Conversely, species with only one dispersal mode such as butterflies or plants pollinated only by specific insects may be more strongly influenced by urbanisation and endangered by the induced fragmentation (Cheptou *et al.*, 2017). The impact of urbanisation on the potential connectivity and subsequently on gene flow and genetic diversity can thus be highly dependent on the species under study, and particular attention should be paid to maintaining connectivity for particularly threatened species.

However, in urban environments, the direct observation of threatened species and the measure of their genetic diversity can be difficult due to the restricted number of individuals still present in these highly fragmented landscapes. In addition, this data collection can be costly and time-demanding, especially if several species have to be studied. In such situations, simulations are a valuable tool for analysing landscape fragmentation, structural and functional connectivity between habitats and the impact of the landscape on the genetic diversity of various species. However, simulations require the definition of many parameters that may be criticised. Consequently, without validation by empirical data, the simulation results may receive little consideration from conservation managers.

3.2.2 Objectives

In this context, we aim to:

- Simulate the impact of fragmentation on the spatio-temporal evolution of genetic diversity and population persistence for a threatened butterfly species with a high dispersal capacity.
- Analyse the impact of a simulated reduction of the dispersal ability on this evolution.
- Compare the simulated results with values of genetic diversity measured on an empirical dataset to illustrate a powerful combination of empirical and simulated data.
- Highlight at risk areas where butterfly populations may be particularly vulnerable due to a loss of connectivity and genetic diversity associated with fragmentation.

3.2.3 Case study

In the study presented in section 3.3, we combine empirical and simulated data to analyse the impact of landscape elements on the genetic diversity and persistence of *Pieris rapae* butterfly populations in the region of Marseille, France. As butterflies are an airborne species that require open-spaces to fly, they may be particularly constrained in their movements through fragmented and human-influenced landscapes, and are thus interesting species to model. In addition, butterflies play important ecological roles in urban areas, particularly due to their interactions with other wildlife groups (birds, lizards, frogs, etc.) and by pollinating a large variety of plants (Ramírez-Restrepo and MacGregor-Fors, 2017). Yet, many urban butterflies populations are threatened due to increased pollution, habitat loss and fragmentation (Dennis *et al.*, 2017).

Pieris rapae (Figure 3-4) is also called the cabbage white butterfly, as its larvae are a pest to crucifer crops such as cabbages. Adults feed on the nectar of various plant species, which enable them to colonize many different habitats. As a result, *P. rapae* is mainly found in open and sunny landscapes such as grasslands (Ohsaki, 1979), but is also one of the most common butterfly species in urban areas, where it can be found in gardens, parks, or vegetated road edges (Lafranchis, 2004).

Figure 3-4 – *Pieris rapae*
(Left) Adult. Photo Credit: Magali Deschamps Cottin. **(Right)** *P. rapae* larvae, pest of cabbage. Photo Credit: Scot Nelson “Larva and feeding injury to crucifer” (https://www.flickr.com/photos/scotnelson/)

3.2.4 Main conclusions

Our study combining empirical data and simulations demonstrated that highly urbanised areas showed lower genetic diversity for butterflies when compared with more rural environments. This reduction was explained by limited population size associated with smaller habitat patches and limited connectivity resulting from dispersal barriers caused by impervious surfaces. In addition, simulation results highlighted a decrease in the number of *P. rapae* individuals over time, especially in highly urbanized areas. This decrease suggests that population persistence is threatened in urban environments. In particular, simulations have highlighted potential habitats where populations are particularly vulnerable due to lack of connectivity with their neighbouring habitats. Our results highlighted that in order to conserve and promote genetically stable and diverse populations, it is important to 1) preserve or restore suitable habitats and 2) maintain or increase connectivity between them.

Our study illustrated how modelling tools, here landscape graphs and genetic simulations, can be used to estimate the dispersal capacities of populations and their genetic diversity. We also demonstrated some advantages of combining simulated data with empirical data in landscape

genetics. The use of simulated data allowed for the analysis to be extended to a wider study area, including zones showing a diversity of urbanization levels, over a defined period of time. In particular, this made it possible to compare several transects and highlight local differences in the Marseille metropolitan area, while empirical data were restricted to a single transect. In addition, the simulations enabled the study of a species with a shorter dispersal distance, for which it may be difficult to collect samples due to its limited presence in urban environments. Here, the simulations emphasized the threat posed by dense urban areas to low dispersal species compared with species of higher dispersal capacities. Finally, our study presented an implementation of surface-transects, where transects are commonly used to collect empirical data along a gradient of environmental conditions. We have shown here how they can be extended to a surface, enabling modelling and simulations within areas along environmental gradients.

Main contributions

- Demonstrating a reduction of genetic diversity and population persistence associated with higher levels of urbanisation for a butterfly species.
- Illustration of the advantages of combining modelling, simulations and empirical data for identifying populations threatened by a limited dispersal capacity and a reduced genetic diversity as a result of a poor landscape connectivity.

3.3 PAPER B: Fragmentation reduces persistence and genetic diversity

Persistence of butterfly populations in fragmented habitats along urban density gradients: motility helps

Postprint version of the article published in **Heredity** - https://doi.org/10.1038/hdy.2017.40

Estelle Rochat[1], Stéphanie Manel[2], Magali Deschamps-Cottin[3], Ivo Widmer[1,4*], Stéphane Joost[1,5*]

[1] Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2] EPHE, PSL Research University, CNRS, UM, SupAgro, IND, INRA, UMR 5175 CEFE, F-34293 Montpellier, FRANCE

[3] Aix Marseille University, IRD, LPED, Marseille France

[4] Swiss Academy of Sciences SCNAT, Swiss Biodiversity Forum, Bern, Switzerland

[5] Urban and regional planning community (CEAT), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

* Joint last author

Contributionse

I computed the landscape graphs and carried out the simulations, performed the analyses of simulated and empirical datasets and wrote the first draft of the paper.

3.3.1 Abstract

In a simulation study of genotypes conducted over 100 generations for more than 1600 butterfly's individuals, we evaluate how the increase of anthropogenic fragmentation and reduction of habitat size along urbanisation gradients (from 7% to 59% of impervious land cover) influences genetic diversity and population persistence in butterfly species. We show that in areas characterised by a high urbanisation rate (> 56% impervious land cover), a large decrease of both genetic diversity (loss of 60-80% of initial observed heterozygosity) and population size (loss of 70-90% of individuals) is observed over time. This is confirmed by empirical data available for the mobile butterfly species *Pieris rapae* in a sub-part of the study area. Comparing simulated data for *P. rapae* with its normal dispersal ability and with a reduced dispersal ability, we also show that a higher dispersal ability can be an advantage to survive in an urban or highly fragmented environment. The results obtained here suggest that it is of high importance to account for population persistence, and confirm that it is crucial to maintain habitat size and connectivity in the context of land-use planning.

3.3.2 Introduction

During the Anthropocene and particularly the last five decades, human population growth and migration have led to an increased demand for housing, transport and infrastructure, leading to a large expansion of cities and to a growing impact of human activities on the environment (Steffen *et al.*, 2007; EEA, 2016). The land-use transformation into dense built-up areas, associated with the intensification of agriculture practices, is mainly responsible for a loss and degradation of natural habitats (Antrop, 2000; EEA, 2016), which is an important cause of endangerment of many animal and plant species (Czech *et al.*, 2000; Dirzo and Raven, 2003, Wood and Pullin, 2002). For example, the change in agriculture practices has been associated with a decline of farmland birds across Europe between 1990 and 2000 (Donald *et al.*, 2006) and habitat loss associated with land use changes have often been reported to be an important cause of the global decline of amphibian populations (Collins and Storfer, 2003; Cushman, 2006).

In addition, due to landscape fragmentation, habitats of many species become divided by impervious surfaces (roads, buildings, etc.) or other human-influenced areas (cropland, recreational areas, etc), which reduced natural habitat size and therefore population size too (Fahrig, 2003b). In addition, functional connectivity (i.e. the movement of individuals among patches) is also affected, subsequently influencing gene flow and reducing genetic diversity (Fahrig, 2003, Hitchings and Beebee, 1997; Coulon *et al.*, 2006). In the region of Marseille, France, Schoville *et al.* (2013) measured genetic diversity in the butterfly *Pieris rapae* within four regions along a transect leading from the periphery to the city-centre and showed decreased genetic diversity in urban versus non-urban sites. Similarly, Takami *et al.* (2004) studied the genetic diversity of two butterfly species from the genus Pieris (*Pieris rapae* and *Pieris melete*) in study areas from Japan and Korea. They showed that the genetic diversity is not directly significantly different in urban areas as compared to rural ones. However, important genetic variations can be observed among seasonal subpopulations in urban areas, whereas it is not the case in rural ones. As a consequence, when considering seasonal subpopulations, the genetic diversity in cities is reduced as compared to rural environments. As the reduction of population size and genetic diversity induces a higher risk of inbreeding (Bonte *et al.*, 2012) and a lower adaptive potential (Munshi-South *et al.*, 2016), this can make species more vulnerable to extinction, particularly if exposed to further environmental change (Allendorf and Leary, 1986).

However, urban habitats can also harbour self-sustaining populations of native (and exotic) species (Kowarik, 2011) which are able to adapt to the human-influenced environment or even take advantage of the proximity with humans that may provide food sources, reduce the presence of wild predators or provide new refuges (McKinney, 2002; Shochat *et al.*, 2010). Indeed, organisms can adapt to anthropogenic fragmentation, either morphologically or behaviourally (Cheptou *et al.*, 2017). For instance, Evans *et al.* (2009) showed that the forest-specialist blackbird Turdus merula, was able to colonize and adapt to urban environments. Despite a little reduction in expected heterozygosity within urban populations of this blackbird, no reduction of observed heterozygosity was noticed and no evidence of genetic differentiation was highlighted between urban and rural populations. Similarly, Lourenço *et al.* (2017) shows no significant differences in genetic diversity between urban and rural populations of salamanders in the municipality of Oviedo (Spain). Previous studies showed that some factors could facilitate the adaptation to urban environment, notably a higher dispersal ability or a lower habitat-selectivity (Turin and den Boer, 1988; Maes and Van Dyck, 2001; Wood and Pullin, 2002; Takami *et al.*, 2004). However, estimating the

impact of urbanisation and fragmentation on species remains complex. Moreover, the direct observation of species and the measure of their genetic diversity may be difficult due to the restricted number of individuals still present in urban or highly fragmented environments. Nevertheless, the identification of plant and animal species endangered by the ongoing fragmentation of habitats is essential in order to promote more sustainable urbanisation processes or conservation strategies in the future. In such situations, simulations are a valuable tool to analyse the landscape fragmentation, the structural and functional connectivity between habitats and the impact of such landscapes on the genetic diversity of various species.

In this context, our study combines empirical and simulated data to analyse the impact of landscape elements on the genetic diversity and on the persistence of butterfly populations in the region of Marseille, France. Butterflies, as airborne species that require open-spaces to fly, are constrained in their movements in fragmented and human-influenced landscapes and are therefore interesting model species. In this study, we aim to: 1) analyse the spatio-temporal evolution of the genetic diversity and population persistence as a function of the urbanisation level from simulated genetic data (500 single nucleotide polymorphism (SNP) loci over 100 generations) for a butterfly species with high dispersal capacity (*Pieris rapae*); 2) study the impact of a simulated reduction of the dispersal ability on this evolution; and 3) compare spatially explicit simulation results with measured genetic diversity of *P. rapae* populations, estimated from an empirical dataset (366 AFLP, Schoville *et al.* 2013).

3.3.3 Material and Methods

Study area

The study area is centred on the region of Marseille, south-east France. With 855'393 inhabitants in 2013, Marseille is the second most populated French municipality, after Paris (source: www.insee.fr, population census, 2013). In order to capture the landscape heterogeneity of this region, we combined vector and raster data describing the land cover of the Marseille area (IGN BD Carto 2004, SPOT 5 2004, Lizée *et al.*, 2011) and produced a land cover classification map of eight classes (spatial resolution: 10 m): buildings (divided into four subclasses as a function of the building height), roads and other impervious surfaces, mixed surfaces (artificial and natural), grasslands, parks, forests, open areas (mainly not vegetated) and water.

To simulate the impact of urbanisation on the evolution of populations and genotypes, we then focused on twelve equally sized spatialized areas categorized into three levels of urban densities (low, medium, high). These study areas were defined along four transects, each of 18 km in length and 4 km in width, leading from the Vieux-Port of Marseille (city-centre) to the suburbs and therefore showing a high urbanisation gradient from densely populated urban to more natural areas. As a function of the urban densities, we divided lengthwise each transect into three parts: high urban density being the first 6 km along the transect from the city centre (red-colored zones in Figure 3-5), medium urban density from 6 km to 12 km (blue-colored zones) and low urban density from 12 km to 18 km (green-colored zones). We thus defined twelve rectangular areas (4 km-width, 6 km-length), partially overlapping downtown (see Figure 3-5).

In order to compare results from simulations and empirical data, we used an empirical genetic dataset for the butterfly species *Pieris rapae* (Schoville *et al.*, 2013). Consequently, one of the simulated transects (Transect B in Figure 3-5) was chosen to correspond to the sampling direction and sites of the empirical study published by Schoville *et al.* (2013). The three other directions

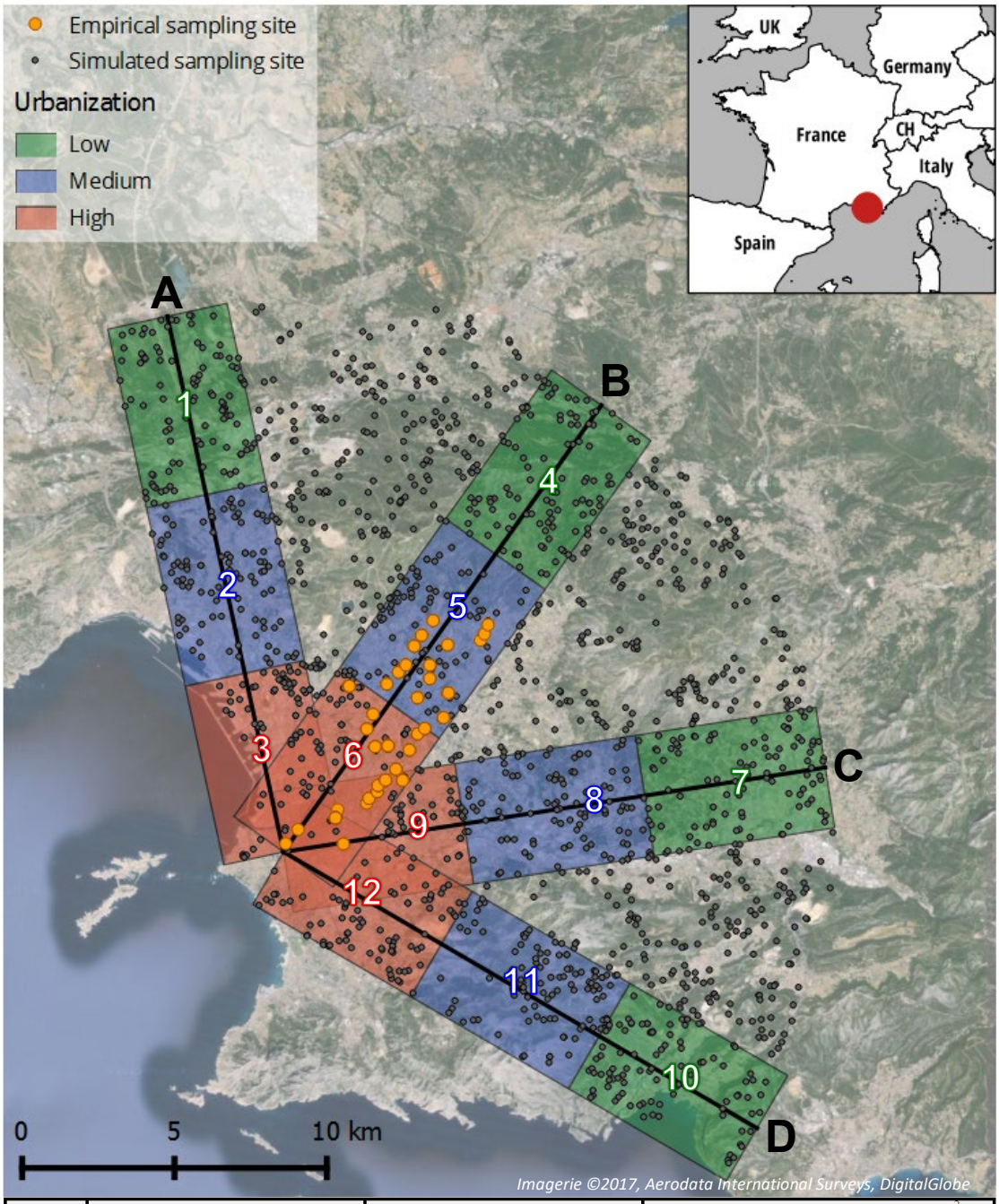
were spatially distributed such as to be representative of the variations in urbanisation level around Marseille.

Based on the land cover map, we then computed the percentage of each land cover classes in the twelve defined areas. The areas with high urbanisation level are characterised by 56-59% of impervious land cover (buildings, roads and other impervious surfaces), 3-7% of green spaces (grasslands and parks), 10-22% of forests and 15-30% of other land cover types (water, mixed surfaces, open areas). The medium urbanised areas contain 8-41% of impervious land cover, 16-26% of green spaces, 19-46% of forests and 17-32% of other land covers. Finally, the areas with a low urbanisation level show 3-13% of impervious land cover, 36-66% of green spaces, 21-37% of forests and 3-28% of other surfaces (Figure 3-5).

Resistance map

Based on expert opinion and empirical results, we assessed and assigned a relative resistance value to each class of the land cover classification, as a function of the capacity of *Pieris rapae* to disperse in each type of land cover (Table 3-1).

As *P. rapae* is a butterfly species preferring open and sunny vegetation-covered landscapes (Ohsaki, 1979), we attributed the lowest resistance value to green spaces (grasslands and parks, resistance = 1). The mean dispersal distance for females of *P. rapae* during their lifetime is about 2 km (Jones *et al.*, 1980). On this basis, we approximated and fixed their maximum dispersal distance in the most favourable land cover class to 4 km. The open areas mainly not covered by vegetation are not a barrier to dispersal but do not offer many foraging possibilities and are therefore less attractive for butterflies than the greenspaces (resistance = 3). Water surfaces, buildings under two meters height, roads and other impervious surfaces can still be potentially crossed but are not vegetated and not necessarily open and will therefore probably not be chosen preferentially as dispersal directions (resistance = 10). An intermediate resistance has been considered for mixed surfaces since they can contain each of the previous mentioned land cover classes, in various proportions (resistance = 5). Forests are not favourable at all for the dispersal of *P. rapae* (Ohsaki, 1979) and we thus assigned them a high resistance (resistance = 20). Finally, buildings constitute barriers to dispersal and received the highest resistance values, increasing as a function of building-height (4 categories, resistances respectively of 10, 20, 50, 450), with the buildings over 10 m in height considered as impossible to cross. Based on these resistance values, we produced a resistance map by assigning to each pixel of the land cover map the corresponding resistance value. The resulting map shows the dispersal-cost through each pixel and was used to compute least-cost path between two sites.



Transect	Impervious Surfaces (% of zone area)			Green spaces (% of zone area)			Green spaces (mean area of entities, m²)		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
A	12.7	41.2	56.4	60.0	22.5	3.0	3 598 980	1 938	530
B	3.3	18.2	58.5	52.2	25.5	6.5	1 565 743	1 120	430
C	9.8	24.6	57.3	65.9	16.0	7.1	47 672	622	367
D	12.0	7.7	56.3	35.9	20.2	6.2	5 757	1 592	465

Figure 3-5 - Study area and simulation transects

Four transects along urbanisation gradients leading from the city centre of Marseille (Vieux-Port) to more rural areas. Along each transect, three zones were delimited and represent different urban densities (high, medium and low). For each zone, the table indicates the percentage of land covered by impervious surfaces or green spaces, as well as the mean surface of green spaces' entities. The sampling sites for the simulation of butterfly populations and genotypes were then randomly assigned to potential habitats within the 12 zones (100 sites per zone). Transect B contains the empirical sampling sites (yellow points) of *P. rapae* used by Schoville *et al*. (2013).

Table 3-1 - Resistance values
Resistance values for the various land cover classes, used to model the dispersal of *P. rapae* butterfly in the region of Marseille, France.

Land cover class	Resistance of a 10m-pixel	Maximal dispersal distance in the land cover class [m]
Green spaces (grasslands and parks)	1	4000
Open areas (mainly not vegetated)	3	1300
Mixed surfaces (artificial and natural)	5	800
Water	10	400
Roads and impervious surfaces	10	400
Buildings : maximum height < 2m	10	400
Forest	20	200
Buildings : 2 m ≤ maximum height < 5 m	20	200
Buildings : 5 m ≤ maximum height < 10 m	50	80
Buildings : maximum height ≥ 10 m	450	0

Habitat

In order to simulate the evolution of populations and genotypes along the four transects, we first defined sites within potential habitat areas for *P. rapae*. To this end, we used the software QGIS 2.14 (function random selection within subsets) to randomly assign 100 sites to potential habitats for *P. rapae* (i.e. green spaces) within each of the 12 zones. The number of sites per zone (100) was chosen in order to obtain realistic distances between the sites, but to avoid overestimating the number of potential habitats, notably in the city centre. In order to ensure the largest habitats - which are assumed to harbour important butterfly subpopulations - to be represented in the sampling design for the simulations, we applied a stepwise procedure. We first chose to randomly position three sites in each of the green spaces showing an area of at least five hectares (ha), and we placed then one site in each of the green spaces showing an area between 1 ha and 5 ha. All the other sites - required to achieve a total of 100 sites per zone - were randomly positioned within green spaces of at least 200 m^2 (2 pixels). Due to partial overlapping between the zones (mainly in the city centre), some points were counted for two different areas. We ended up with a total of 1083 sites for simulations, with a median nearest-neighbour distance of about 320 m. In addition, in order to allow for connectivity and potential gene flow among populations from different transects, 550 sites were randomly distributed within green spaces situated between the transects, as illustrated in Figure 3-5.

Simulated data

Once the sampling sites were defined, we used the individual-based population genetics model software CDPOP 1.2.21 (Landguth and Cushman, 2010) to simulate the evolution of genotypes over 100 generations. In order to benefit from the possibility offered by the simulations to obtain a high number of genetic markers, but to avoid simulating a dataset too different from the empirical dataset available (366 AFLP markers), we decided to simulate the evolution of genotypes at 500 diploid bi-allelic single nucleotide polymorphism loci (SNPs).

CDPOP enables the user to define various demographic parameters related to the displacement of individuals between the sites, the mate-choice, the breeding with Mendelian inheritance and the mortality of individuals. Here, we started the simulations by considering a uniform distribution of butterflies over the study region, i.e. all sites previously defined were assumed to be inhabited

by one individual of *P. rapae* at the beginning of the simulations (in total 1633 *P. rapae* individuals over the study region). The initial genotypes were randomly assigned. The simulation of dispersal and mating movement between the sites was then based on a cost distance matrix indicating the cost of dispersing from one site to another. For this matrix, we defined the dispersal cost as the cumulative resistance of the least cost path, computed using the software Graphab (Foltête, Clauzel, and Vuidel, 2012) based on the resistance map previously defined. The function used to link the dispersal cost and the dispersal probability has then to be chosen between the four possibilities offered by CDPOP: linear, inverse square, nearest neighbour and random mixing. We chose here the linear one, assuming that the probability to disperse decreases linearly with the increase of the cost. Finally, we specified the maximum dispersal distance. As for the resistance maps, we approximated and fixed the maximum dispersal distance of females of *P. rapae* to 4000 m. The males are less mobile and scarcely dispersed (Ohsaki, 1980), and their maximal dispersal distance was therefore defined to approximately one third of the one of females (i.e. 1350 m).

For breeding parameters, we considered a sexual reproduction that can start from the age 0, with no selfing and no philopatry, both males and females allow to mate multiple times (Bissoondath and Wiklund, 1996) and multiple paternities possible (females can have offspring from multiple males). The number of offspring of *P. rapae* can vary between 300 and 400 eggs with about 99% mortality (Richards, 1940) and was thus simulated using a Poisson's law of parameter λ equal to 300, with a birth mortality fixed at 99%. The sex of each individual was set randomly for each generation.

Finally, generations of *P. rapae* can sometimes partially overlap (Ohsaki, 1982) and we therefore fixed the adult mortality to 95% which keeps the possibility of 5% of the individuals to live for more than one generation. In order to increase reliability of our results, we computed five runs of simulations based on these parameters.

In a second step, we computed five additional runs of simulations using exactly the same parameters, except that we reduced the maximal dispersal distance by one half, i.e. 2 km for the females and 675 m for the males. This second set of simulations corresponds to the simulation of *P.rapae* with a reduced dispersal ability and enables the analysis of the influence of the dispersal capacity on the evolution of genetic diversity and population persistence.

Our simulations finally produced a dataset of 500 SNPs markers for 1633 individuals (1083 situated along 4 transects leading from the city-centre to the periphery and divided into three level of urbanisation, and 550 individuals in-between these transects) over 100 generations. The dataset is replicated 5 times (5 runs) for *P. rapae* with normal dispersal ability, and 5 times for *P. rapae* with reduced dispersal capacity.

Genetic diversity and population persistence

Once the genetic data have been simulated over 100 generations, we analysed the level of genetic diversity within each of the twelve zones, based on measures of heterozygosity. A high heterozygosity indicates a lot of genetic variability, whereas a low heterozygosity indicates poor genetic diversity. We used here two indices: the average observed heterozygosity (H_{obs}) as well as the average expected heterozygosity (H_{exp}) assuming Hardy-Weinberg equilibrium (HWE):

$$H_{obs} = \frac{1}{k}\sum_{i=1}^{k} h_i$$

Formula 3-4

And

$$H_{exp} = \frac{1}{k}\sum_{i=1}^{k} 2p_i q_i$$

Formula 3-5

where *hi* is the frequency of individuals that are heterozygous for the marker i, *pi* (respectively *qi*) is the frequency of presence of the first (respectively second) allele for the marker i and *k* is the total number of markers. The values of these measures range from 0 (no individual heterozygous for any marker, no genetic diversity) to 1 (all individuals are heterozygous for all markers, high genetic diversity). The comparison of the values of these two indices can allow the identification of potential inbreeding. Indeed, when a population is facing high inbreeding, the fraction of heterozygotes observed will be less than what is expected under random mating. The difference between observed and expected heterozygosities can therefore be used to estimate the amount of current inbreeding (Wright, 1949).

During the simulations, habitat sites may become uninhabited, in particular if the cost of reaching sites is too high. As a result, the number of individuals per zone can change over time (starting from an initial value of 100 individuals per zone). We retrieved the number of individuals remaining in each zone at each generation (N) and used this number as an estimate of the population persistence in the respective zone.

For both dispersal abilities and for each of the twelve zones, we used the five values resulting from the five simulation runs to compute mean (μ), standard deviation (σ) and 95% confidence intervals ($\mu \pm 1.96*\sigma/\sqrt{n}$) for the three parameters (H_{obs}, H_{exp} and N) at each generation and we produced plots of their evolution over time. In order to compare the genetic diversity and number of individuals remaining in each zone at the end of the simulations, we also used the 5 values from the 5 simulation runs to compute a one-way ANOVA between the last values (generation 100) of H_{obs} (resp. H_{exp} and N) in each of the twelve zones (i.e. 12 groups, 5 measures per groups). Post-hoc testing was then performed using Scheffé test in order to highlight the significant differences between the zones. All computations were performed using the Matlab R2014b software (functions *anova1* and *multcompare*).

Empirical data

With the objective to compare the results of the genetic diversity obtained by simulations to an empirical case study, we used a published empirical dataset of *P. rapae* sampled in the same study region (Schoville et al., 2013). This dataset was composed of 366 AFLP markers for 219 *P. rapae* individuals that were sampled at 41 sites along a 100 km transect going from the Vieux-Port of Marseille to the suburbs. We here used a subset of this dataset, containing only the sampling sites present in our study area, which corresponds to 36 sites and 145 individuals (yellow points on transect B, Figure 3-5). In order to estimate genetic diversity based on this empirical data, for each site we identified the n nearest neighbours (Euclidean distance between sampling

points) for n comprised between 3 and 25. We then computed the expected heterozygosity among the individuals from the group of neighbouring sites. As AFLP markers do not allow the distinction between heterozygotes and homozygotes of the dominant allele, we can only measure the frequency of homozygotes of the recessive allele (f). Assuming Hardy-Weinberg equilibrium with p representing the allele frequencies of the dominant allele and q the allele frequencies of the recessive allele, we have: f = q^2 and $p = 1 - q = 1 - \sqrt{1-f}$. The expected heterozygosity can therefore be expressed as:

$$H_{exp} = \frac{1}{k}\sum_{i=1}^{k} 2(f - 1 + \sqrt{1-f})$$

Formula 3-6

where *k* is the total number of markers.

In order to compare simulated and empirical data, the expected heterozygosity for the simulated data was also computed along the same transect (transect B) by considering for each site the n nearest neighbours (n between 3 and 25), using Formula 3-5.

3.3.4 Results

Observed heterozygosity

The change over time of the genetic diversity, as measured by the observed heterozygosity is presented on Figure 3-6A (*P. rapae* with normal dispersal ability) and Figure 3-6B (*P. rapae* with reduced dispersal ability). The initial value at generation zero is equal to 0.5, which corresponds to the theoretical maximum value for heterozygosity expected under HWE, resulting from the random distribution of genotypes at the beginning of the simulations. For both dispersal capacities, this value rapidly decreases in all transects and for all levels of urbanisation.

In the more rural areas (green lines), a loss of 6-7% of the initial heterozygosity can be observed after ten generations for *P. rapae* with normal dispersal ability, but the decline then stabilises and more than 75% of the initial heterozygosity level is still present after hundred generations (Table 3-2). Similar evolution can be observed with the reduced dispersal, with however a more pronounced decline (15-19% lost after 10 generations, 40-58% at generation 100). The Scheffé tests computed on the values reached at the end of the simulations indicate that with reduced dispersal the loss of observed heterozygosity in area 1 (transect A, -58.3%) is significantly higher (p-values < 10^{-6}) than in the other areas with a similar urbanisation level (maximum 41.4% lost). This area is also showing the highest loss among the areas of low urbanisation levels with the normal dispersal ability.

In the areas with medium urbanisation (blue lines), a loss of 9-15% of the initial heterozygosity can be observed with the normal dispersal ability after ten generations and 20-32% at the end of the simulation. Once again, the decline is more pronounced for the reduced dispersal, where the loss already reached 25-36% at generation 10 and 56-76% at the end of the simulations. The values of observed heterozygosities in these medium urbanised areas (blue lines) are generally lower than in the areas with low urbanisation (green lines). However, for some transects, the values are close and the confidence intervals sometimes intersect (transects A and B for normal dispersal, transect A for reduced dispersal). The Scheffé tests computed on the values at generation 100 indeed indicate significant differences between the values of areas with low and medium

urbanisation only in transect C and D for the normal dispersal (p-values $<10^{-6}$) and in transects B, C, D for reduced dispersal (p-values $< 10^{-4}$).

Finally, for highly urbanized areas (red lines), more than 30% of the initial observed heterozygosity is already lost after ten generations with the normal dispersal and the decline continues until the end of the simulations, after hundred generations, where only 28 to 40% of the initial observed heterozygosity remains (Table 3-2). With the reduced dispersal, the level of observed heterozygosity remaining after hundred generations is dramatically low (between 0.1 and 0.15, representing only 4 to 28% of the initial value). For both dispersal abilities, Scheffé's tests indicate that the values reached at generation 100 in all highly urbanised areas are significantly lower than in all areas with medium or low urbanisation (p-values $< 10^{-7}$). We can also notice that with the normal dispersal ability, the value of observed heterozygosity reached at generation 100 in area 12 (transect D, -96.2%) is significantly lower than in the other highly urbanised areas (p-values $< 10^{-6}$). This area is also the one presenting the highest loss of observed heterozygosity with the reduced dispersal ability (-96.2 %).

Table 3-2 - Decline in observed and expected heterozygosity
For each zone, the table presents the mean percentage decline of observed and expected heterozygosities and the mean number of individuals lost at generation 10 and 100, computed on the basis of 5 simulation runs

Zone			Observed Heterozygosity				Expected Heterozygosity				Number of individuals			
			Generation 10		Generation 100		Generation 10		Generation 100		Generation 10		Generation 100	
			normal	reduced	normal	reduced	normal	reduced	normal	reduced	normal	reduced	normal	reduced
1	A	L	-7.3	-18.7	-23.7	-58.3	-4.6	-5.0	-19.5	-31.7	0.0	0.0	0.0	0.0
2	A	M	-15.1	-32.6	-28.6	-64.7	-3.6	-6.6	-14.0	-28.8	-0.8	-29.4	-0.8	-30.2
3	A	H	-35.9	-45.6	-65.6	-93.6	-7.3	-14.8	-30.6	-57.2	-29.3	-61.2	-40.2	-71.2
4	B	L	-7.2	-17.4	-18.5	-41.4	-3.0	-4.2	-12.1	-17.5	0.0	-1.2	0.0	-3.0
5	B	M	-8.7	-25.2	-19.5	-56.2	-3.5	-5.8	-11.9	-21.9	0.0	-3.0	0.0	-2.8
6	B	H	-36.2	-42.9	-66.1	-82.3	-6.0	-15.4	-21.1	-62.6	-18.2	-70.8	-25.8	-81.4
7	C	L	-6.0	-15.3	-16.9	-39.6	-3.4	-4.7	-13.6	-22.9	0.0	-0.8	0.0	-1.0
8	C	M	-14.6	-36.2	-32.2	-75.8	-3.9	-7.5	-14.6	-32.9	-5.0	-32.8	-3.7	-29.0
9	C	H	-30.1	-51.6	-60.9	-94.9	-6.6	-16.6	-28.1	-73.8	-17.2	-60.2	-25.5	-79.2
10	D	L	-5.9	-17.0	-19.2	-40.8	-3.6	-5.0	-17.2	-24.7	0.0	0.0	0.0	0.0
11	D	M	-12.8	-27.1	-34.3	-58.3	-4.4	-5.8	-18.1	-26.7	-3.0	-10.0	-3.0	-18.0
12	D	H	-38.0	-54.7	-82.1	-96.2	-6.4	-16.1	-26.8	-76.7	-17.8	-71.2	-28.7	-91.0

The names of the areas in the first column are indicated as follows: area number, transect and urbanisation level (L=low, M=medium and H=high). The indication 'normal' and 'reduced' are referring to the dispersal ability.

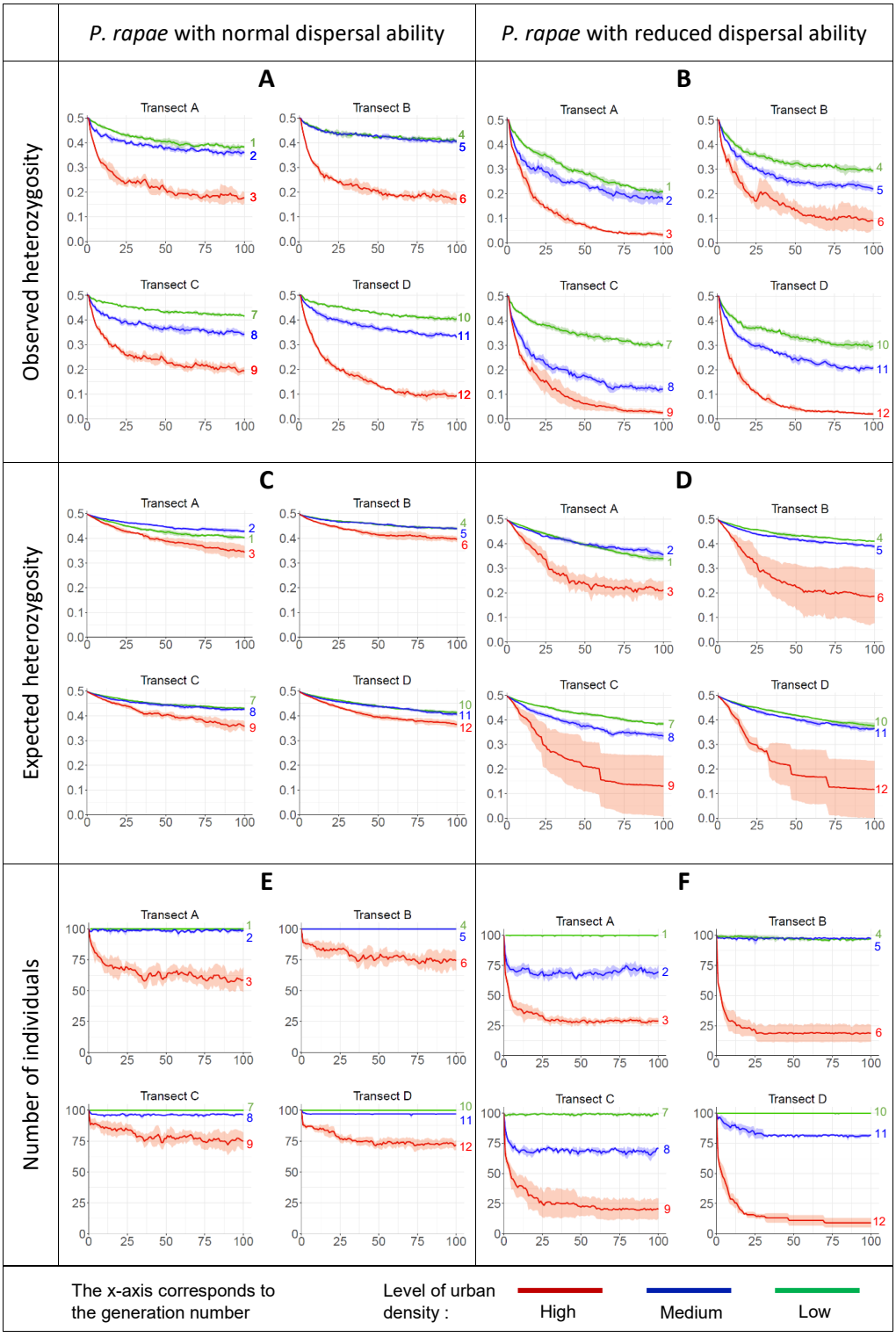


Figure 3-6 – Number of individuals and heterozygosity over generations
Simulated change over time of the number of individuals (e, f), and observed (a, b) and expected (c, d) heterozygosities within areas of different urbanisation densities. Graphs in the left column show the changes over time for P. rapae with normal dispersal ability and in the right column for reduced dispersal ability. For each transect, the green line corresponds to the more rural area (green areas in Figure 3-5), the blue line to intermediate area (blue areas in Figure 3-5) and the red line to the city-centre area (red areas in Figure 3-5). The curves present the average value and the 95% confidence intervals computed on the basis of the five simulation runs.

Expected heterozygosity

Like the observed heterozygosity, at the beginning of the simulations the expected heterozygosity is equal to 0.5 in all transects for *P. rapae* with normal dispersal ability (Figure 3-6C) and reduced dispersal (Figure 3-6D). Over time, a decrease can be observed in all transects and all areas, but this decline is less pronounced than for the observed heterozygosity, especially with the normal dispersal ability. Indeed, for this species, the highest loss of expected heterozygosity after hundred generations is of 30.6 % (Table 3-2, transect A, highly urbanised) whereas it was of 82.1% for the observed heterozygosity (Table 3-2, transect D, highly urbanised). With the reduced dispersal, the decreases are more pronounced, especially in the highly urbanised areas and the values are also less stable between the simulation runs, which is highlighted by the much larger confidence intervals.

When comparing the various levels of urbanisation, no significant difference can be highlighted for any transect between the areas of low or medium urbanisation, when considering the values of expected heterozygosity at generation 100 (p-values >> 0.5) with either dispersal abilities. The values of expected heterozygosity in the more rural areas (green lines) even occasionally drop below the value of the corresponding intermediate areas (blue lines), notably in transect A. With the normal dispersal ability, the differences between highly urbanised areas and the other levels of urbanisation are small but nevertheless significant (p-values < 0.03), except for area 6 (transect B). With the reduced dispersal, only the areas 9 and 12 (transects C and D) show significantly lower values at generation 100 as compared to the other levels of urbanisation (p-values < 0.05).

Persistence of populations

For *P. rapae* with normal (Figure 3-6E) and reduced (Figure 3-6F) dispersal abilities, the results show that the number of individuals in the less urbanised areas (green lines) remains stable throughout the study period, for all transects. Indeed, with the normal dispersal, hundred individuals are present at all time in these areas (Figure 3-6E, green lines), whereas with the reduced dispersal (Figure 3-6F, green lines), a small loss can be noticed in half of the transects (B and C), but this loss is only of 3 individuals at maximum (Table 3-2).

For the areas with a medium urbanisation (blue lines), with the normal dispersal (Figure 3-6E) the number of individuals slightly decreases but then remains also stable through time, with more than 96 individuals present at all time in all transects. However, with the reduced dispersal (Figure 3-6F), a noticeable reduction in the number of individuals (18-30 individuals) can be observed in transects A, C and D, whereas in transect B the loss of individuals is very small (3 individuals at maximum). Scheffé tests indicate that the values reached at generation 100 are not significantly different between low and medium urbanised areas for the normal dispersal ability (p-values > 0.9). However, with the reduced dispersal, the values are significantly different (p-values < 0.001) except in transect B.

Finally, for the highly urbanised areas (red lines), with the normal dispersal, the number of individuals (Figure 3-6E) rapidly drops from 100 to less than 85 during the first ten generations and this decline is even more severe with the reduced dispersal (Figure 3-6F) for which less than 40 individuals are present after ten generations. After approximately twenty generations, the number of individuals in these highly urbanised areas stabilizes for both dispersal abilities. With the normal dispersal, the stabilisation occurs around 60 (transect A) to 75 individuals (transect B, C, D) whereas with the reduced dispersal it is much lower (9 individuals in transect D to 30 individuals

in transect A). Once again, the Scheffé tests indicate that the values reached at the end of the simulations in the highly urbanised areas are always significantly lower than in the ones with medium or low urbanisation, with both dispersal abilities (p-values < $2*10^{-4}$).

Simulated versus empirical results

For both the empirical and simulated data, we can observe a significant increase of expected heterozygosity with increasing distance from the city centre (Figure 3-7). Moreover, the linear regressions fitted on the centered-reduced values show slopes that are very close for all datasets.

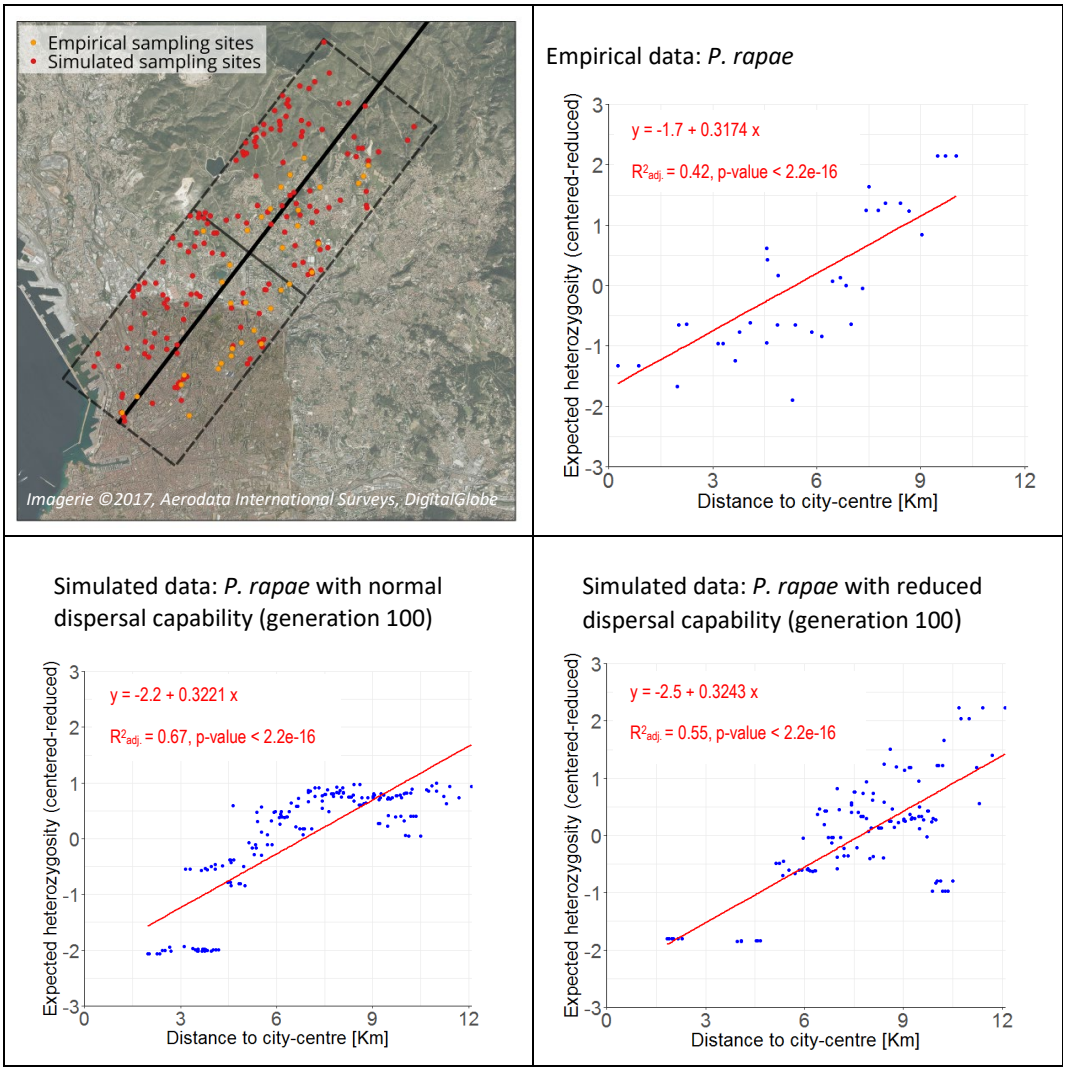


Figure 3-7 – Simulated versus empirical results
Expected heterozygosity computed for each site considering the five nearest neighbours, as a function of the distance to the city centre. For both simulated and empirical data, the values of heterozygosity have been standardised. Note that the absolute values are not directly comparable as the values of expected heterozygosity computed on the empirical data sets (366 AFLP) range from 0.07 to 0.18, whereas from simulated data (500 SNP) they are comprised between 0.012 and 0.42 (normal dispersal ability) or between 0.005 and 0.37 (reduced dispersal ability). The linear fit and the statistics were obtained using the function lm in R.

When comparing the absolute values of expected heterozygosity obtained with the simulated and empirical datasets, one can notice that the range of values obtained from the empirical dataset is much smaller than the one from simulations. Indeed the values of expected heterozygosity computed on the empirical dataset (366 AFLP) range from 0.07 to 0.18, whereas from simulated data (500 SNP) they are comprised between 0.012 and 0.42 (normal dispersal ability) or 0.005 and 0.37 (reduced dispersal ability). However, despite these differences in the absolute values, the same patterns with respect to urban density are observed in both the simulated and empirical data.

Figure 3-7 presents the evolution of the expected heterozygosity as a function of the distance to the city-centre for the computations performed considering for each site the 5 nearest neighbours. The results for the other numbers of neighbours (3-25) are not presented, but lead to the same conclusions.

3.3.5 Discussion

Potential negative impact of anthropogenic fragmentation

Results of this study illustrate that highly urbanised areas show a lower genetic diversity for butterflies, measured by both the observed and expected heterozygosities. These areas are characterized by a high percentage of impervious land cover (> 55%), a low percentage of green spaces (< 8%) and a reduced surface of green spaces entities (< 600 m^2) (Figure 3-5). In these conditions, the loss of genetic diversity observed can be explained both by the reduction in population size due to the loss of habitats and to their smaller size, and also to the limited connectivity due to the dispersal barriers caused by impervious surfaces. Indeed, during the simulations, when the resistance of the landscape is important, the cost of moving to other habitats becomes too high and eventually individuals can only reach very few congeners to reproduce. In such situations, gene flow is significantly reduced, which ultimately leads to a decline of genetic diversity. This decline has also been highlighted by the analysis of the empirical data available in the study region for *P. rapae*, which shows a decrease of the expected heterozygosity towards the city-centre. Moreover several previous studies have highlighted a similar negative influence of urban environment on dispersal (Schtickzelle and Baguette, 2003; Schtickzelle *et al.*, 2006; Dubois and Cheptou, 2017), gene flow (Keyghobadi *et al.*, 2006) and genetic diversity (Williams *et al.*, 2003; Takami *et al.*, 2004). Nevertheless, the rapidity of the decline presented here with the simulations should be interpreted with caution. Indeed, in real environments, populations are generally much larger than 100 individuals, and the reduction in genetic diversity may therefore take more time than what is presented here. However, the aim of the analysis was not to determine the time required to reach a given level of genetic diversity, but to show that simulated as well as empirical data indicate that the genetic diversity of urban populations is significantly reduced as compared to the diversity of populations living in more rural areas. Moreover, the results show that the level of observed heterozygosity is generally lower than the level of expected heterozygosity, especially in the highly urbanised areas. This difference highlighted a potential inbreeding for the populations concerned, which is a cause of extinction risk for butterfly populations (Saccheri *et al.*, 1998; Nieminen *et al.*, 2001).

Results of the simulations also highlighted a decrease in the number of *P. rapae* individuals over time, especially in highly urbanized areas. This decrease suggests that the persistence of populations is threatened in urban environments. For the simulated data, a potential site may become

unoccupied over time if the individuals living there are no longer able to find a congener to reproduce. Indeed, in this extreme case, individuals are isolated and the population is doomed to extinction. This negative impact of urban environment on the persistence of populations has already been highlighted in other urban areas and for other species (Maes and Van Dyck, 2001; Wood and Pullin, 2002; Fattorini, 2011). However, once again, when considering the number of generations after which a site potentially becomes uninhabited, the results of the simulations should be interpreted with caution. In reality, depending on the initial size of the population present in each habitat, the real extinction may take more time than what is shown here. Nevertheless, independently of the exact time required for extinction, the simulations we processed highlighted potential habitats in which populations are particularly vulnerable due to the lack of connectivity with their neighbouring habitats.

Our study was conducted in an environment, in which the fragmentation and reduced habitat size was mostly due to a high level of urbanisation. However, the negative impact on genetic diversity and population persistence highlighted here can also be observed in non-urban environments facing important fragmentation and reduction of habitat size. For example, Fountain *et al.* (2016) studied museum samples of the Glanville fritillary butterfly and showed that a decline in genetic diversity was preceding the extinction of the populations in the mainland of Finland mainly due to fragmentation and loss of suitable meadows. Similarly, loss of genetic diversity due to fragmentation and associated lack of connectivity has been highlighted for the prairie-chickens in Wisconsin (Johnson *et al.*, 2004), for the alpine chipmunk in Yosemite National Park (Rubidge *et al.*, 2012) or for a tropical rain forest tree in Costa Rica (Hall *et al.*, 1996).

Finally, we note that the impacts highlighted in this study are not relevant for all species living in urban environments. Even though similar evolutions could be most probably observed for other butterfly species which are similarly constrained in their dispersal in urban environments, other species may not be negatively impacted by urbanisation, or less impacted, as previously mentioned in the introduction.

Differences among transects: land cover and barriers to dispersal

As regards areas with low urbanisation, the highest loss of genetic diversity or number of individuals is generally observed in transect A. This area of transect A is characterised by a high percentage of green spaces (60%) and a large average surface for these entities (> 3 km^2) likely to be favourable for the species studied. However, this area also shows the highest percentage of impervious surfaces (12.7%), which could explain its disadvantage as compared to other regions of the periphery.

For the areas with medium urbanisation, transect C often seems to be the most negative, especially for the species with the lower dispersal ability. This could be explained by the lower percentage of habitat areas (grassland and parks) in this region and also to the smaller average surface of habitat entities (622 m^2) which indicates a higher fragmentation.

Finally, the highly urbanised areas of transect D often appears to be less favourable even if its land cover does not seem to be very different from the other transects. However, this transect is characterised by the lowest percentage of green spaces in the periphery, which is due to a quite high percentage of impervious surfaces (12%) but also a high percentage of water (11.8%) and forest (24.4%). These barriers may reduce the gene flow from the periphery to the city centre and

therefore threaten the viability of the populations of the city-centre. This shows that the fragmentation of the less urbanised suburb areas can also have a noticeable importance on the decrease of genetic diversity and population persistence of the urban populations.

Finally, we note that in our case for all parameters (expected and observed heterozygosity and number of individuals), the differences highlighted between the transects (differences at generation 100: 0-27.4%, 12.9% in average) remain minor relative to the differences observed between the three levels of urbanisation (differences at generation 100 between low and high urbanisation level: 9-91%, 43.7% in average).

Dispersal capacity in urban landscapes

When comparing the respective behaviours of *P. rapae* and of a butterfly with a reduced dispersal ability, results show that the reduction of genetic diversity is much more pronounced for the less mobile species and that the persistence of the latter populations is also more threatened. This underlines that a higher dispersal capacity may be an advantage for species living in urban environments, which had already been highlighted by previous studies (Maes and Van Dyck, 2001; Wood and Pullin, 2002; Duplouy *et al.*, 2013). Indeed, a higher dispersal capacity results in the ability to disperse over longer distances but also to use various dispersal modes facilitating the crossing of barriers present in urban landscapes. For example plants pollinated by many insects may be only moderately impacted by urban fragmentation (Culley et al., 2007). Similarly, species that can benefit from human-mediated dispersal (attachment to clothes, vehicles, shoes, soil movements, etc.) may be particularly adapted to urban landscapes (Banks *et al.*, 2015; Egizi *et al.*, 2016). Conversely, species with only one dispersal mode such as butterflies or plants pollinated only by specific insects may be more strongly influenced by urbanisation and endangered by the induced fragmentation (Cheptou *et al.*, 2017).

Relevance of simulations

This study's results illustrate advantages of combining simulated with empirical data in landscape genetics. Indeed, empirical genetic data reflects the current state of the genetic composition of populations, influenced by potentially unknown evolutionary processes in the past. However the collection of such data might be particularly expensive and time-demanding. In this context, simulations may offer many advantages (Epperson *et al.*, 2010).

First, as shown in this study, the use of simulated data allows for the extension of the analysis over a larger study area including zones showing a diversity of urbanization levels, and over a defined period of time. This notably makes it possible to compare several transects and to highlight local differences across the metropolitan area of Marseille, while empirical data were restricted to a single transect. Secondly, the simulations enable the study of a species with a lower dispersal distance, for which it may be difficult to collect samples due to its limited presence in urban environments. Here, the simulations permitted in particular to emphasize the threat that dense urban areas constitute for low dispersal species compared to species with a higher dispersal capacity. Finally, simulations allow the consideration of a larger genetic dataset, here based on 500 SNPs as compared to the 366 AFLPs constituting the empirical dataset. This can be particularly interesting in a context of high sequencing cost and since results may change according to the genetic data used (Landguth, Fedy, *et al.*, 2012).

However, simulations often require subjective, expert-based assumptions to be formalized (resistance values, populations' parameters, etc.) resulting in the injection of uncertainty in the results obtained. This is an important reason why the combination with empirical data is particularly powerful since the latter provide landmarks to relieve the uncertainty mentioned above and enabling a complete analysis and a more confident interpretation of the results.

Sustainable land-use planning

Our results show that butterfly species can be strongly threatened in dense urban areas, highly fragmented environment or other human-influenced areas. For these species, in order to conserve and promote genetically stable and diverse populations, it is important to 1) preserve or restore suitable habitats and 2) maintain or increase the connectivity among them in order to allow dispersal also for species with limited dispersal capacities. As increasing the connectivity by the creation of dispersal corridors may be difficult to achieve due to the numerous constraints of urban or highly fragmented environments, the creation or preservation of stepping stone habitats is promising and of special importance (Bierwagen, 2006; Serret *et al.*, 2014). In this context, the use of landscape genetic methods to assess the impact of landscape features on gene flow is a key step in designing functional ecological networks aiming to preserve genetic diversity and therefore biodiversity (Baguette *et al.*, 2013). Genetic analyses are powerful methods to estimate species' dispersal processes (Stevens *et al.*, 2010), to assess adaptive ability *(Munshi-South et al.*, 2016) and also to directly provide information about persistence of populations, which is essential to promote and plan for more sustainable land-use strategies. In urban areas, the preservation of biodiversity is also key to favour a better quality of life for the residents, including well-being related to better health conditions (Maller *et al.,* 2006). For this reason, it is of paramount importance that urban authorities and planners adapt the way they design dense city centers in particular, to identify the potential consequences on native species, and to favor the insertion of connected habitats.

3.3.6 Additionnal figure

The Figure 3-8 (unpublished) shows the landsape graphs corresponding to the dispersal paths between the simulated points for *P. rapae* (left) and for a species with a lower dispersal capacity (right). The graph components have been highlighted in this figure. These components correspond to sets of nodes that are connected by a succession of links. Two nodes thus belong to the same component if it is possible to move from one to the other, and are in different components if there is no link to connect them. At the ecological level, the presence of few large components, containing many nodes, indicates a well-connected territory in which habitats are linked by dispersal paths. On the contrary, a high number of small components containing few nodes indicates the presence of small groups of habitats that are disconnected from each other, without any possibility for exchanges.

For *P. rapae* (Figure 3-8, Left), a large component includes most of the habitat points from the periphery, indicating good opportunitites for dispersal and gene flow there. In contrast, the habitat points close to the city-centre are divided into many small components, indicating little exchange between them and a lack of connectivity with the periphery. For a species with a lower dispersal capacity (Figure 3-8, Right), the components are much more numerous, which highlights lower connectivity associated with limited dispersal and gene flow.

Such maps can thus be used to identify dispersal barriers as well as isolated habitats that are not connected to any other habitats, or only to a few of them.

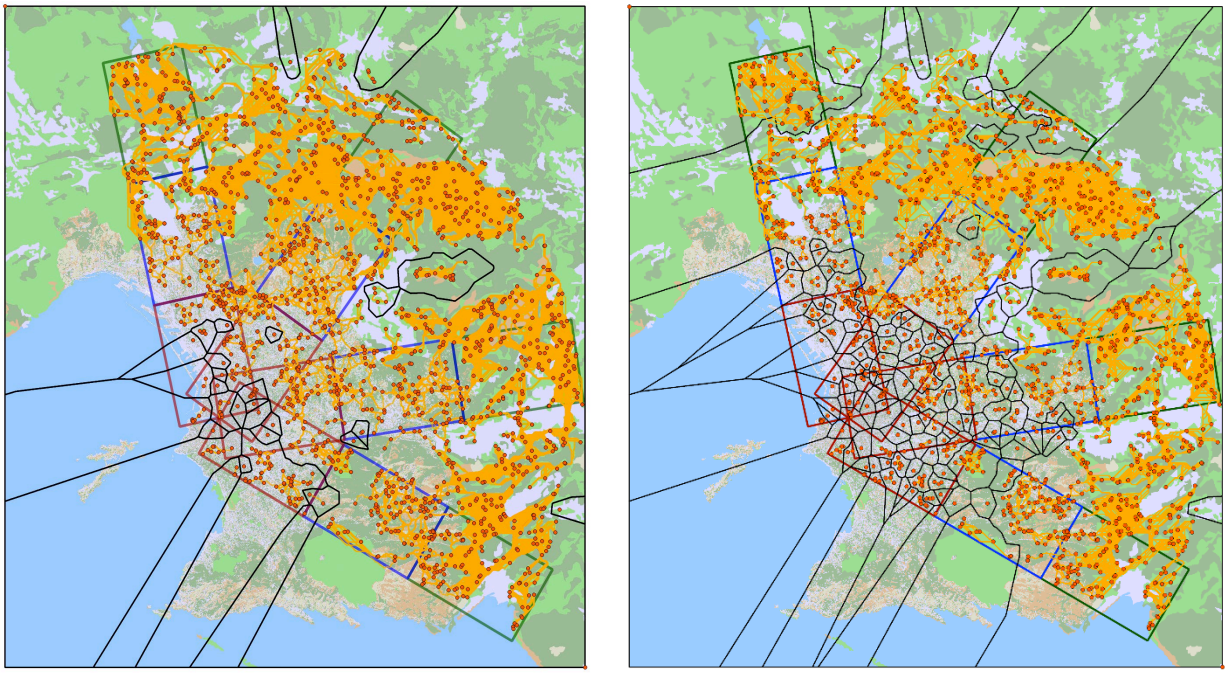


Figure 3-8 – Landscape graphs
Landscape graphs for P. rapae (left) and for a species with a lower dispersal capacity (right). The least cost paths between the simulated habitat points are shown in orange and the components of the graph are highlighted in black.

Chapter 4 LOCALLY ADAPTED GENETIC VARIANTS

4.1 Research context

In the previous two chapters, we have highlighted the importance of preserving 1) suitable habitats that may constitute the potential niche of a species and 2) sufficient connectivity between them in order to preserve the dispersal possibilities and genetic diversity of populations. A higher level of genetic diversity should indeed favour adaptation. However, in some cases, conserving total genetic diversity in populations may not be sufficient to preserve the potential for adaptation and it could be more valuable to specifically preserve locally adapted genetic variants.

4.1.1 Local adaptation

Populations of the same species living in different geographical areas may develop diverse evolutionary responses according to contrasting natural selective pressures. This process is called **local adaptation** because it does not affect all individuals of the same species similarly. Due to this local adaptation, populations will be better suited to their native environment than to other locations where individuals of the same species may live. For example, rock pocket mice were shown to adapt locally to their environment by adapting their coat colour to the colour of the rock on which they live (light desert rocks or black basaltic rocks) in order to be less visible to predators. This adaptation has been associated with a polymorphism in the melanocortin-1-receptor (MC1R) gene (Nachman *et al.*, 2003).

At the genetic level, local adaptation results in a modification of the alleles observed between individuals from different populations, i.e. the presence of a polymorphism. However, not all polymorphisms are the signature of a local adaptation since most of them are neutral, i.e. they do not confer any advantage or disadvantage. Neutral polymorphisms can notably result from mutation, gene flow or genetic drift. This neutral genetic diversity contributes to the total genetic diversity but does not indicate any local adaptation. As discussed in the previous chapter, preservation of total genetic diversity is key to preserving the adaptive potential, as it offers more opportunities to find variants that are better suited to the modified conditions. However, the survival of populations in a specific site may depend even more heavily on the preservation of genetic variants that are already locally adapted. When developing conservation strategies to increase gene flow, for

example through reintroduction, hybridization or cross-breeding, it is thus important to preserve the local adaptation of populations. In addition, some populations may already present adapted traits that may be more favourable for future climatic conditions (e.g. adaptation to drought or high temperature). Individuals from these populations might therefore be preferentially selected for conservation measures as their long-term survival capacity may be higher. They can also be used as donor populations for reintroductions. In this context, there is a need to preserve not only neutral genetic diversity, but also adaptive genetic diversity in particular (McKay *et al.*, 2001; Hoffmann and Willi, 2008; Sgrò *et al.*, 2011; Willoughby *et al.*, 2018). The development of tools to help preserve this adaptive potential is a challenging research question that is at the intersection of many fields. One of the first steps is the detection of the signatures of local adaptation in the genome.

4.1.2 Identification of signatures of local adaptation

Traditionally, local adaptation was identified using common garden experiments. However, with advances in sequencing techniques, several methods have been developed to identify loci potentially subject to natural selection, based on genetic data (Nielsen, 2005; Schoville *et al.*, 2012; Joost *et al.*, 2013; Vitti *et al.*, 2013; Rellstab *et al.*, 2015). These methods come from two main research areas: population genetics and landscape genetics.

Population genetics

Population genetics is a discipline of biology that describes the genetic variations within populations (Fisher, 1930; Wright, 1932, 1949; Haldane, 1959). In this field, many methods have been developed to identify local adaptation. A first group of methods relies on the comparison of intra- and interspecific measures. Under positive directional selection, an allele that provides greater fitness and improves the chances of survival is expected to be preferentially passed to the offspring and therefore to show a higher frequency in this population (whereas a negative selection disfavouring deleterious allele will induce a lower frequency). As a result, very different allele frequencies will be observed at the corresponding locus between populations facing different selection pressures, or between populations facing selection versus a neutral population. This genetic differentiation can be measured using the Wright's fixation index, F_{ST}, which compares the allele frequencies within a population with those existing between populations. Natural selection is thus expected to produce extreme values of F_{ST} at selected loci, relative to other loci (Beaumont and Nichols, 1996). Based on this idea, various software have been developed to compare observed frequencies with null models where no selection takes place, and to identify loci that are statistical outliers (Foll and Gaggiotti, 2008; Excoffier and Lischer, 2010). Another way to identify selection based on inter- and intraspecific divergence relies on the comparison of the ratio of non-synonymous to synonymous mutations (McDonald *et al.*, 1991). Non-synonymous mutations are mutations that modify the amino-acid sequence and the corresponding derived protein, thus resulting in a biological change in the organism. Synonymous mutations on the contrary produce the same amino acid. Natural selection affects non-synonymous mutations, where positive selection favouring an allele will increase the proportion of non-synonymous mutations when compared to synonymous ones (whereas negative selection disfavouring a deleterious allele will act inversely) (Nielsen, 2005).

Some other population genetics methods focus on the detection of selective sweeps based on the analysis of the frequency spectrum within a population. When a mutation occurs and is retained by directional selection, its frequency will increase in the population. Usually, genes that

are closely situated on the same chromosome will be transmitted together to the offspring, i.e. the copies transmitted to the offspring will come from the same homologous chromosome of the parent. As a result, the alleles of linked sites surrounding the mutation are also likely to increase their frequency, which is called genetic hitchhiking. This produces genomic regions showing reduced genetic variations as compared to the rest of the genome, which is called a selective sweep. The identification of these selective sweeps can therefore be used to identify possible loci under selection (Kim and Stephan, 2002). Some software have been developed based on this idea (Nielsen *et al.*, 2005; Chen *et al.*, 2010; Pavlidis *et al.*, 2013). All these population genetics methods focus on populations and therefore require the analysis of population structure.

Population structure analysis

There are several ways of defining populations of individuals of the same species, all of which involve a cohesive process to group individuals together (Waples and Gaggiotti, 2006). This cohesion may result from an opportunity to interact demographically (competition, interactions, etc.) or genetically (reproduction). Populations thus correspond to groups of individuals of the same species that can either interact or reproduce together. Different populations may be created due to barriers restricting dispersal or gene flow. Geographically separated individuals will then evolve differently and their allele frequencies will vary depending on their history, including the amount of genetic drift (i.e. the change in allele frequencies resulting from random sampling of organisms), potential bottlenecks (i.e. strong reductions of population size at a given time leading to an important loss of genetic diversity), or migratory events (Günther and Coop, 2013).

Several statistical methods have been developed to identify different populations based on genetic data. Pritchard *et al.* (2000) presented a Bayesian clustering method to group individuals according to their genetic characteristics (STRUCTURE). This method requires the user to define the estimated number of populations (K). Usually, several values for K are tested such to retain the most likely number. STRUCTURE then assigns each individual to one or more population(s) based on the probability of observing their allele frequencies in the population of interest. Individuals may be assigned to several populations in the case of admixture, i.e. when individuals possess recent origins from two or more distant populations. Based on the same statistical model, ADMIXTURE was then developed (Alexander and Lange, 2011), using a faster optimization algorithm and adding a cross-validation procedure to define the most likely value for K.

Principal Component Analysis (PCA) can also be used to identify population structure (Patterson *et al.*, 2006). In this case, PCA is used to reduce the dimensions of genetic data. PCA output does not directly groups individuals into populations, but instead provides their coordinates along axes of genetic variations. A clustering procedure could then be used to assign individuals into discrete populations (see for example DAPC (Jombart *et al.*, 2010)).

Landscape genomics

Identifying local adaptation with population genetics methods can be difficult when datasets show a weak genetic structure (i.e. all individuals belong to the same population) or have only a few sampled individuals. Moreover, these methods do not directly provide information on the environmental variable that may produce the selection. Finally, the computations can be very time-consuming, making them difficult to apply in genome-wide studies (Schoville *et al.*, 2012; Joost *et al.*, 2013; Stucki *et al.*, 2014). Other methods for detecting signatures of local adaptation have thus been developed, in the field of landscape genomics.

Landscape genomics is a subfield of landscape genetics, a discipline that focuses on the spatial dimension of genetic information and studies the interaction between environmental variables and evolutionary processes (Manel *et al.*, 2003). Landscape genomics studies the influence of the environment on the distribution of neutral and adaptive variations in the genome (Luikart *et al.*, 2003). Many methods have been developed in this field to identify loci under selection. These methods integrate environmental data (climate, land cover, altitude, etc.) and seek to identify genotype or allele occurrences that are significantly correlated with environmental variables, indicating that they have possibly been selected by the environment. However, high numbers of false positives can be detected using these methods if population structure is not correctly integrated. Indeed, differences in allele frequencies resulting from genetic drift, bottleneck, migratory events, etc. may be falsely attributed to local adaptation if the population structure is not taken into account (Excoffier *et al.*, 2009). Analysing the population structure is thus an essential prerequisite for landscape genomics studies.

In 1969, Johnson and Schaffer first correlated allelic frequencies with environmental variables, to look for signatures of selection in insects (Johnson *et al.*, 1969). Later, Joost et al. (2007) developed the parallel processing of a large number of logistic regressions to estimate the probability of presence of a genetic variant as a function of an environmental variable (Spatial Analysis Method, see next section). This approach enables a fast analysis of many loci in relation to many environmental variables, and the identification of possible selection signals without any prior information about loci. On this basis, to meet the requirements imposed by new high-density genomic data, Stucki *et al.* (2014) developed the Samβada software tool that allows for the rapid computation of logistic regressions with large genomic datasets. Samβada also allows for multivariate analyses, including more than one environmental variable as well as information about population structure. Several other approaches based on genotype-environment associations have been developed, in particular to more easily integrate population structure parameters (Manel *et al.*, 2010; Coop *et al.*, 2010; Günther and Coop, 2013; Gautier, 2015). In the following section, we present the SAM method, which will be used in the analysis presented in section 4.3.

4.1.3 Spatial Analysis Method (SAM)

Logistic regressions

The Spatial Analysis Method (SAM) implements logistic regressions to link the presence of an allele or genotype to environmental variables. Logistic regression is a standard method used to describe the relationship between a binary response variable, here the presence or absence of a genotype, and one or more continuous predictors, here the environmental predictors (Figure 4-1).

These logistic regressions estimate the probability *p* of observing a given genotype as a function of environmental conditions. More specifically, logistic regressions express the logarithm of the odds, which is the ratio of the probability of success (presence of the genotype = p) to the probability of failure (absence of the genotype = 1-p), as a linear regression of environmental predictors (Formula 4-1):

$$logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Formula 4-1

where *x1… xn* are the values of the environmental predictors and *β0… βn* are the coefficients of the logistic regression. These coefficients can be estimated using a maximum likelihood method.

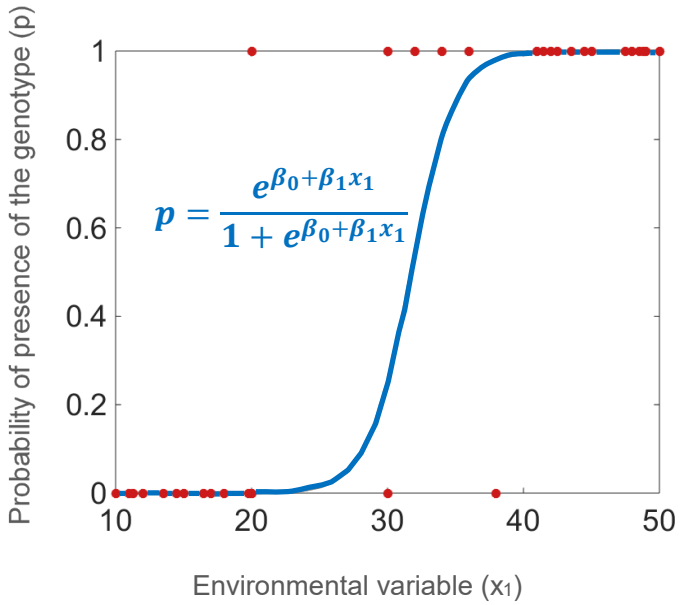


Figure 4-1 – Univariate logistic regression

As a result, the probability (*p*) of presence of the genotype (G) given the environmental conditions (x) can be obtained using Formula 4-2.

$$p = p(G|x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}$$

Formula 4-2

Covariates can also be integrated in the model, for example to consider population structure. In that case, the probability to observe a genotype, given the environmental conditions and the covariates, can be expressed using Formula 4-3.

$$p = p(G|(x \cap c)) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n+\gamma_1 c_1+\gamma_2 c_2+\cdots+\gamma_n c_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n+\gamma_1 c_1+\gamma_2 c_2+\cdots+\gamma_n c_n}}$$

Formula 4-3

where *c1….cn* are the values of the covariates and *γ1….γn* are the associated coefficients.

To identify genotypes that are significantly associated with environmental conditions, which may therefore be the signature of natural selection, independent univariate logistic regressions are computed between each genotype and each environmental predictor and the most significant associations are identified on the basis of statistical tests.

Statistical tests

Likelihood-ratio test (G score)

This statistic compares the likelihood *L* of the complete model (*M*) with the likelihood *L0* of a model excluding the environmental variable of interest (*M0*). In the univariate case, *M* thus contains a constant plus an environmental predictor, whereas *M0* contains only one constant. The *G* score is then computed using Formula 4-4.

$$G = 2\ ln\frac{L}{L0}$$

Formula 4-4

The null hypothesis H0 of this test is that the model considered is no better than a null model, i.e. the environmental variable considered does not help in predicting the probability of presence of the genotype. Assuming H0, the *G* values will follow a chi-square distribution with one degree of freedom. This distribution can thus be used to estimate the *p-value* associated with the observed *G* score.

Wald test

The Wald test is a common statistic used to estimate whether a parameter is equal to a given value (V). In our case, it is used to test the null hypothesis that the β parameter associated with an environmental variable is equal to 0, which would indicate that the environmental variable studied has no effect on the probability of finding the genotype. The Wald test is calculated from the difference between the estimated value β and the testing value (V). This difference is then expressed as a number of standard errors of the β parameter ($\sigma(\beta)$) (Formula 4-5):

$$W = \frac{\beta}{\sigma(\beta)}$$

Formula 4-5

Under the null hypothesis, the Wald ratio values will also follow a chi-squared distribution with one degree of freedom.

Functions of the marker detected

When the genetic data analysed does not correspond to the whole genome, it is possible that some of the loci detected are not under selection themselves, but are the signature of a selection that has occurred in a closely localised gene (through selective sweep). The study of the genomic areas surrounding the detected loci thus allows the identification of potential genes that may be under selection. When a reference genome is available (i.e. a complete assembly of the entire genome of a species, making it possible to locate the position of genes), the analysis of a genomic area can be carried out by using a genomic browser such as the ones available on NCBI (https://www.ncbi.nlm.nih.gov/genome/) or Ensembl (https://www.ensembl.org/index.html). Such tools enable the identification of all known genetic information located around a locus of interest.

4.2 Scientific contribution

4.2.1 Problem statement

Despite the numerous tools developed to identify the signature of local adaptation, the integration of this knowledge into conservation practices is still very limited. Several authors have thus highlighted the need to develop methods to better take into account the adaptive potential in conservation practices (Funk *et al.*, 2019; Hoelzel *et al.*, 2019; Mable, 2019). To this end, there is a need to provide tools to map the spatial distribution of locally adapted genetic variants and to project the probability of finding them in un-sampled areas or under future climatic conditions (Bay *et al.*, 2017). However, very few studies have addressed this issue. Fournier *et al*. (2011) identified locally adapted SNP alleles and then used the Maxent species distribution model (Phillips *et al.*, 2006) to estimate their spatial distribution. Similarly, Exposito-Alonso *et al*. (2018) trained one ENM (random forests) for each SNP identified as locally adapted in order to highlight which alleles are most likely given the current environmental conditions. They then projected the models into future climatic conditions to identify populations that would require significant modifications of their allele composition to be better adapted to the future climate. Fitzpatrick and Keller (2015) proposed another approach based on two community-level modelling methods (Generalised Dissimilarity Modelling and Gradient Forest). They used it to map the current and future spatial distribution of several adaptive variants and to assess the “genetic offset” of populations under climate change as a function of the mismatch between the current distributions and future predictions. Following this, Bay *et al.* (2018) defined this mismatch as the “genomic vulnerability” of populations and showed that recent climate change has already negatively affected populations with high genomic vulnerability.

However, practical applications of these methods remain limited and there is still a need to develop new tools to ease the integration of the adaptive potential into conservation practices. The SAM method presented in the previous section can be directly used to map the probability of finding a specific genetic variant in a territory by projecting the logistic regression onto the environmental layers (see case studies in Annex A3). However, since local adaptation usually involves several loci, potentially associated with different environmental conditions, the method needs to be extended to enable the simultaneous consideration of several variants and environmental variables. We will therefore consider how the SAM approach presented in the previous section can be further developed to meet this need. As compared with existing methods previously cited (Random Forests, Gradient Forest and Generalised Dissimilarity Modelling), the SAM approach relies on logistic regressions that can be understood and implemented without requiring advanced mathematical background. This could be an advantage to facilitate its practical implementation in a conservation framework involving actors with different scientific backgrounds.

4.2.2 Objectives

In this context, we aim to develop a new tool based on gene-environment associations (SAM method) to:

- Predict the probability of presence of one or many locally adapted genetic variants in non-sampled areas.
- Identify areas where there is a higher probability of finding individuals better adapted to the future climatic conditions.
- Identify vulnerable populations that may be threatened by climate change due to a lack of locally adapted genetic variant favourable for the future conditions.

4.2.3 Case study

In the study presented in section 4.3, we analyse the local adaptation of Moroccan and European goats' populations to estimate their vulnerability to climate change. Goats are able to live in very contrasting environments and are thus interesting candidates for studying local adaptation. However, like many other livestock species, their potential for local adaptation is currently threatened by selective breeding aimed at improving production value (Taberlet *et al.*, 2008).

The genetic data used in our study were provided by the NEXTGEN and AdaptMap projects. NEXTGEN, for "next generation methods to preserve farm animal biodiversity", is the first project which provided whole genome sequence data for goats (https://www.epfl.ch/labs/nextgen/). One goal of this project was to highlight the genomic regions associated with local adaptation in order to encourage a more sustainable breeding management. In the project, 161 goats were sampled in Morocco (Figure 4-2). In this county, goats live semi-wild, spending more than 8 months outdoors and the anthropic selection is relatively modest (Boujenane, 2005). The different populations, confronted with contrasting environmental conditions from the Sahara desert to the Atlas Mountains, are thus expected to have developed local adaptation.

AdaptMap (Stella *et al.*, 2018) is an international project developed to improve coordination between various independent projects collecting genetic data from goats worldwide (http://www.goatadaptmap.org/). The AdaptMap project has regrouped the genetic data collected and harmonised it into a common database. This common dataset contains geo-reference genetic data for 4563 goats sampled worldwide (Figure 4-2). We carried out a first analysis of the entire dataset with several partners from the AdapMap project and we identified several signatures of natural selection (Bertolini *et al.*, 2018).

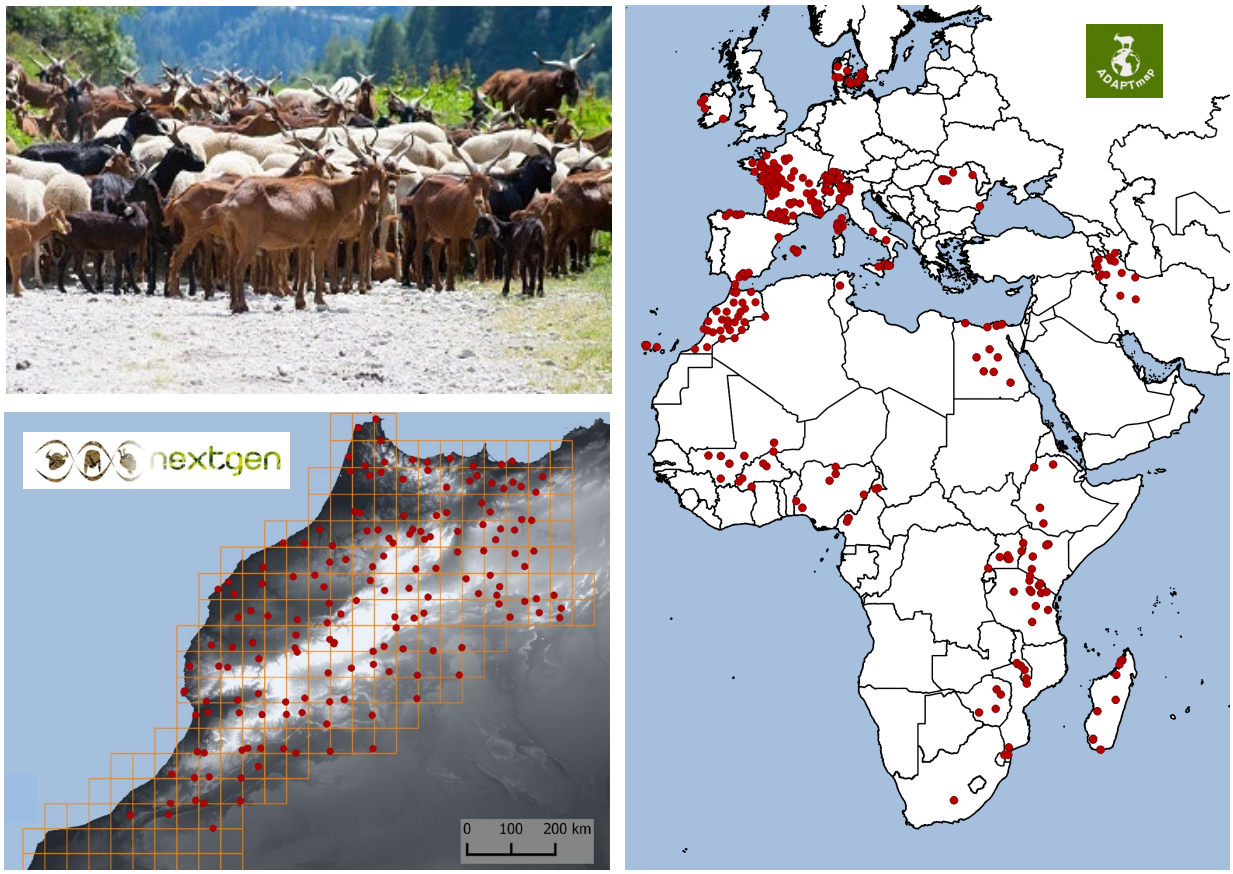


Figure 4-2 – Goat datasets from NEXTGEN and AdaptMap projects
This figure shows the localisation of goats sampled by the NEXTGEN project in Morocco (Left, bottom) and by independent projects worldwide, regrouped in the AdaptMap dataset (Right). For this latter dataset, some additional geo-referenced data are available for goats sampled in Pakistan, South America and Australia. Source for top left photo: https://ec.europa.eu/programmes/horizon2020/en/news/saving-animal-dna-future-generations.

4.2.4 Main conclusions

Using the results of the SAM method to detect locally adaptive genetic variants, our study introduced a new tool: the SPatial Areas of Genotype probability (SPAG) that maps the probability of finding beneficial variants in a study area. We presented a univariate model, that can be used to predict the spatial distribution of a single genotype, and three multivariate models allowing the integration of several genotypes, potentially associated with various environmental variables. In a second step, our study showed that the combination of the SPAG models with climate change predictions can be used to identify populations with genetic variants better suited for the future environmental conditions and vulnerable populations that may not be able to adapt. We validated the SPAG concept with one simulated dataset and two case studies on goats (Moroccan and European ones).

Several potential signatures of natural selection were identified for the goats under study. For the Moroccan population, the results obtained with the SAM method highlighted several genes strongly associated with the variation of precipitation and potentially related to skin or hair properties. Following this, the SPAGs enabled the identification of vulnerable populations currently lacking genetic variants locally adapted to the strong variations of precipitation predicted under a

2070 climate change scenario. For Europe, using logistic regressions allowed the identification of genes potentially conferring adaptation to drought, and SPAGs then made it possible to identify populations that may be particularly threatened by upcoming drought conditions.

Main contributions

- Presentation of the new SPatial Areas of Genotype Probabilities approach, to map the probability of finding locally adapted genetic variants in a study area.
- Development of a univariate model and three multivariate models making it possible to integrate several locally adapted genotypes potentially associated with different environmental variables.
- Illustration of the method with two case studies, allowing the identification of goat populations threatened by climate change due to a lack of favourable adaptive variants.

4.3 PAPER C: Spatial Areas of Genotype Probabilities

Spatial Areas of Genotype Probability (SPAG): predicting the spatial distribution of adaptive genetic variants under future climatic conditions

Version of the article revised and resubmitted in **Global Change Biology**:
https://doi.org/10.1101/2019.12.20.884114

Estelle Rochat[1, 2], Stéphane Joost[1, 2, 3]

[1] Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2] The ADAPTMAP Consortium

[3] The NEXTGEN Consortium

Contributions

I developed the theoretical basis of the three multivariate models, implemented all models as R functions, prepared the simulated dataset, computed the analyses on the case studies and wrote the first draft of the paper.

4.3.1 Abstract

In a context of rapid global change, one of the key components for the survival of species is their genetic adaptive potential. Many methods have been developed to identify adaptive genetic variants, but few tools were made available to integrate this knowledge into conservation management. We present here the SPatial Areas of Genotype probability (SPAG), using genotype-environment logistic associations to map the probability of finding beneficial variants in a study area. We define a univariate model predicting the spatial distribution of a single genotype, and three multivariate models allowing the integration of several genotypes, potentially associated with various environmental variables. We then integrate climate change projections to map the corresponding future distribution of genotypes. The analysis of the mismatch between current and future SPAGs makes it possible to identify a) populations that are better adapted to the future climate through the presence of genetic variants able to cope with future conditions, and b) vulnerable populations where genotype(s) of interest are not frequent enough for the individuals to adapt to the future climate. We validate the SPAG approach using simulations and we use it to study the potential adaptation of 161 Moroccan and 382 European goats to the bioclimatic conditions. In Morocco, using whole genome sequence data, we identify seven genomic regions strongly associated with the precipitation seasonality (WorldClim database). The predicted shift in SPAGs under a strong climate change scenario for 2070 highlights goat populations likely to be threatened by the expected increase in precipitation variation in the future. In Europe, we find genomic regions associated with low precipitation, the shift in SPAGs highlighting vulnerable populations not adapted to the very dry conditions expected in 2070. The SPAG methodology is successfully validated using cross-validations and provides an efficient tool to take the adaptive potential into account in general conservation frameworks.

4.3.2 Introduction

Climate change has altered the conditions of various organisms, causing an average warming of 0.2°C per decade over the past 30 years, as well as a sea level rise of more than 3 mm per year since the start of the century and an increase in the frequency of extreme weather events such as storms, droughts and floods (IPBES, 2019). These changes are likely to continue in the future (IPCC, 2014). When such important changes occur, many animal and plant species are confronted with a shift away from the favourable conditions necessary for their survival (IPBES, 2019). In order to avoid extinction under these conditions, they can either move to more favourable areas or adapt to their new environment (Hughes, 2000). Due to limitations in dispersal capacity, loss of favourable habitats and increased landscape fragmentation, the possibilities for dispersal to new areas are often limited (Opdam and Wascher, 2004; McGuire *et al.*, 2016). Population adaptation relies on phenotypic changes, which can be induced either by phenotypic plasticity or genetic evolution (Merilä and Hendry, 2014; Fox *et al.*, 2019). Phenotypic plasticity can allow species to rapidly evolve by changing their behaviour, physiology or morphology (Reed *et al.*, 2011; Fox *et al.*, 2019). However, it can also potentially lead to a fitness reduction (Duputié *et al.*, 2015) and, since it is not based on heritable genetic variations, it will not necessarily ensure the persistence of adaptation for the next generations. In order to preserve biodiversity, it is therefore crucial to promote the conservation of the genetic adaptive potential of populations (Hoffmann and Sgrò, 2011; Sgrò *et al.*, 2011; Nicotra *et al.*, 2015; Shafer *et al.*, 2015).

Conservation of this adaptive potential is also of major importance for livestock management in order to ensure herd persistence (Hoffmann, 2010; FAO, 2015). Indeed, many livestock populations, especially in developing countries, are breed in pastoralist systems and live most of the time outdoors, confronted to difficult production conditions (e.g. heat stress, poor food resources and the presence of parasites and diseases), and without significant food or water supplies (FAO, 2015). These populations show signatures of local adaptation to their climatic conditions (McManus *et al.*, 2009, 2011; FAO, 2015; Bertolini *et al.*, 2018), to limited food resources (Silanikove, 2000) or to the presence of parasites (Noyes *et al.*, 2011; FAO, 2015; Vajana *et al.*, 2018). However, due to the increasing demand for food production, these local breeds currently tend to be replaced by high-producing commercial breeds imported from developed countries (Rischkowsky and Pilling, 2007; Hoffmann, 2010). This has led to a loss of genetic diversity, which threatens the adaptive potential of livestock species to environmental changes (FAO, 2015). In addition, imported breeds lack in the locally adapted genetic variants, which may reduce their fitness (FAO, 2015). It is therefore essential to highlight the adaptive potential of livestock species in order to encourage farmers to conserve local traditional breeds, and to carefully design cross-breeding, translocation or artificial selection (Scherf *et al.*, 2008; Allendorf *et al.*, 2010).

One of the essential components of the adaptive capacity of populations is genetic diversity (Allendorf and Leary, 1986). Since mutation rates are generally low, adaptation to rapid environmental changes largely depends on the amount of genetic variants already present in populations, i.e. standing genetic diversity (Orr and Unckless, 2008). With the recent increase in the availability of genetic data and the development of conservation genomics, various tools have been developed to integrate genetic diversity into conservation frameworks (Bonin *et al.*, 2007; Vandergast *et al.*, 2011; Thomassen *et al.*, 2011). However, conserving neutral variation in populations may not be sufficient to allow rapid adaptation to increasingly stressful conditions (Reed and Frankham, 2001), and it could be more valuable to specifically preserve adaptive variation, i.e. variation associated with a trait involved in fitness (Hoffmann and Willi, 2008; Sgrò *et al.*, 2011; Willoughby

et al., 2018) or to combine both approaches (Funk *et al.*, 2012; Pauls *et al.*, 2013). Increasing attention is currently being paid to this issue in conservation discussions (Funk *et al.*, 2019; Hoelzel *et al.*, 2019; Mable, 2019).

Several methods have been developed to identify signatures of local adaptation, based on various assumptions and with different limitations and advantages (Schoville et al., 2012; Joost et al., 2013; Vitti et al., 2013; Hoban et al., 2016). The results have notably been used to establish prediction of future habitat range of species facing climate change (Hällfors et al., 2016; Ikeda et al., 2017; Garzón et al., 2019; Razgour et al., 2019). However, there is currently a need to integrate this knowledge in order to predict the distribution of locally adapted genetic variants along environmental gradients, and to project the probability of finding them in un-sampled areas or under future climatic conditions (Bay et al., 2017). Nevertheless, very few studies have addressed this issue. Fournier *et al.* (2011) identified locally adapted SNP alleles and then used the Maxent species distribution model (Phillips et al., 2006) to estimate their spatial distribution. Fitzpatrick and Keller (2015) proposed another approach based on two community-level modelling methods (Generalised Dissimilarity Modelling and Gradient Forest). They used it to map the current and future spatial distribution of several adaptive variants and to assess the “genetic offset” of populations under climate change as a function of the mismatch between the current distributions and future predictions. However, practical applications of these methods remain limited and there is still a need to develop new tools to ease the integration of the adaptive potential into conservation practices, especially by considering several loci, potentially non-independent and adapted to different environmental conditions.

We propose here a novel approach to predict genotype frequencies and map SPatial Areas of Genotypes Probabilities (SPAG) based on logistic genotype-environment associations *(*Joost *et* al., 2007) and conditional probability theory. SPAGs can be used to a) predict the probability of presence of one or many locally adapted genetic variants in non-sampled areas b) identify areas where there is a greater probability of finding individuals better adapted to future climatic conditions, c) identify vulnerable populations that may be threatened by climate change and d) integrate the results into conservation frameworks by means of an easy combination with other georeferenced layers. The concept of applying logistic regressions on an environmental layer to predict the probability of presence of a genotype had been sketched out several years ago (Joost 2006; page 138). Here, we formalise this concept and extend it to multivariate models. We introduce the theoretical bases of SPAGs and validate the approach with a simulated dataset. We then present an application of our approach to two case studies in order to analyse the local adaptive potential of Moroccan and European goat populations.

4.3.3 Material and Methods

SPAG’s approach

Logistic regressions (SAM)

The Spatial Analysis Method (SAM, Joost *et al.*, 2007) can be used to detect genotypes that are strongly associated with an environmental variable and are therefore potential adaptive variants (strictly additive). This method assumes a linear response of the genotype to the environmental variable and uses logistic regressions (Formula 4-6) to assess the probability of presence of a genotype *G1* as a function of an environmental variable *(x1)*,

$$p(G1) = p(G1 = 1 \mid x_1) = \frac{e^{\beta_0+\beta_1 x_1}}{1 + e^{\beta_0+\beta_1 x_1}}$$

Formula 4-6

where β0 and β1 are the parameters of the regression to be fitted. Independent univariate logistic regressions can be computed between each genotype and each environmental predictor and significant associations can be identified using statistical tests. Joost *et al.* (2007) suggested the combined use of a likelihood ratio (G score) and a Wald test. The likelihood ratio (G) compares the likelihood of a model with the likelihood of a null model without the environmental variable of interest. The null hypothesis is that the model considered is no better than the null model, i.e. the environmental variable considered does not help in predicting the probability of presence of the genotype. The Wald test is a common statistic used to estimate whether a parameter is equal to a given value. In our case, it is used to reject the null hypothesis that the β parameter associated with an environmental variable is equal to 0, which would also indicate that it has no effect on the probability of finding the genotype. The SAM approach is implemented in the Samβada software and has previously been validated against other methods for identifying signatures of natural selection (Stucki *et al.*, 2017).

Univariate SPAG

Once the genotype(s) involved into significant associations with the environmental variables have been identified, Formula 4-6 enables the estimation of the probability of presence of a genotype for any value of an environmental variable (*x1*). We consequently used it to estimate and delimit on a map the probability of presence of a genotype over the whole region of interest (Joost, 2006; Rochat *et al.*, 2016). We named such a delimited surface univariate Spatial Area of Genotype Probability (SPAG).

As more than one adaptive locus are usually identified, we also developed multivariate models to compute a single map showing the probability of presence of multiple genotypes. Three different multivariate models have been developed to date: the Intersection, Union and K-Percentage.

Intersection SPAG (I-SPAG)

The **Intersection** model (I-SPAG) is used to compute the probability that the variants of interest are all simultaneously present. Following the theory of conditional probability (Kolmogorov, 1956), the probability of simultaneous presence of *n* adaptive genotypes *Gi, i=1:n* can be computed using Formula 4-7:

$$p\left(\bigcap_{i=1}^{n} G_i\right) = p\left(\bigcap_{i=1}^{n-1} G_i\right) p\left(G_n \middle| \bigcap_{i=1}^{n-1} G_i\right)$$

Formula 4-7

where $p\left(G_n \middle| \bigcap_{i=1}^{n-1} G_i\right)$ is a conditional probability that can be estimated using a logistic regression where $\bigcap_{i=1}^{n-1} G_i$ is integrated as a covariate (Formula 4-8).

$$p(G_n \mid \bigcap_{i=1}^{n-1} G_i) = \frac{e^{\beta_0+\beta_1 x_n+\beta_2 \bigcap_{i=1}^{n-1} G_i}}{1+e^{\beta_0+\beta_1 x_n+\beta_2 \bigcap_{i=1}^{n-1} G_i}}$$

Formula 4-8

However, as we would like to use this model to predict the probability of presence of the genotypes for any point of the region of interest, i.e. also where *Gi* values are unknown, we suggested to estimate $\bigcap_{i=1}^{n-1} G_i$ by $p(\bigcap_{i=1}^{n-1} G_i)$.

Using the associative property of the intersection operator, the intersection of *n* genotypes can be computed by starting with the univariate model *p(G1)*, which is used as a covariate to compute *p(G1 ∩ G2)*, itself used to compute *p(G3 ∩ (G1 ∩ G2)),* etc.. Formula 4-8 can thus be implemented with a recursive model based on the univariate formula in which covariates are added (see Annex A4.1 for more details). We implemented this model as a function in *R*, available following the link given in the section “code availability” at the end of the paper.

Union SPAG (U-SPAG)

The **union** model (U-SPAG) is used to compute the probability of finding at least one of the adaptive genotypes of interest. We implemented it with the inclusion-exclusion principle (e.g. for two genotypes: p(G1 ∪ G2) = p(G1) + p(G2) – p(G1 ∩ G2)). We implemented the generalised formula for *n* adaptive genotypes (Formula 4-9), as a function in *R* based on the intersection model previously described (see code availability and Annex A4.1).

$$p\left(\bigcup_{i=1}^{n} G_i\right) = \sum_{i=1}^{n} p(G_i) - \sum_{i<j} p\left(G_i \bigcap G_j\right) + \sum_{i<j<k} p\left(G_i \bigcap G_j \bigcap G_k\right) + \ldots + (-1)^{n-1} p\left(\bigcap_{i=1}^{n} G_i\right)$$

Formula 4-9

K-Percentage SPAG (K-SPAG)

Finally, we developed a **K-percentage** model (K-SPAG) to estimate the probability that an individual carries K% of *n* adaptive genotypes. This probability can be computed by combining formulas from the union and intersection models (Formula 4-10, explained in more detail in Annex A4.1). Again, this formula was implemented as a function in *R* (see code availability).

$$p(K\%\ G_{i=1\ldots n}) = p\left(\bigcup_{i=1}^{n} \bigcap_{1\le i_1<i_2<\cdots<i_{(K\%*n+1)}} (G_{i_1} \cap G_{i_2} \cap \ldots \cap G_{i_k})\right)$$

Formula 4-10

Note that all multivariate models allow the integration of adaptive genotypes associated with various environmental variables since the environmental variable *xi* used to compute *p(Gi)* can be different for each *i*.

Simulation study

In order to test the SPAG's approach, we first computed a simulated dataset using the individual-based population genetics model software CDPOP 1.3 (Landguth and Cushman, 2010; Landguth *et al.*, 2020). We simulated individual genetic exchanges and natural selection across 300 non-overlapping generations among 200 individuals randomly located in a 500x500 gridded landscape. For the breeding parameters, we considered a sexual reproduction, with random mating, both male and female with replacement, no selfing, no philopatry, no multiple paternity, equal sex ratio and each mated pair producing three offspring. The movement of the individuals was linearly restricted as a function of the Euclidean distance, with a maximum dispersal corresponding to 25% of the entire landscape. We simulated 50 diallelic loci, with three loci under selection (L0, L1 and L2). The selection was implemented using three 500x500 raster gradients, the first from north to south (X0), the second from east to west (X1) and the third from northwest to southeast (X2) (see Figure 4-3A). We set the average effects bL0A0A0=10 and bL0A1A1=-10 for the locus L0 with the environmental variable X0, which indicates that the genotype A0A0 from locus L0 will be favoured in the South (where X0=1), whereas A1A1 will be favoured in the North (where X0=-1). We set similar effects for the locus L1 with the environmental variable X1 and L2 with X2. All other beta effects were set to 0, indicating no influence of the environmental variable to the distribution of genotypes. All genotypes were randomly initialised at the beginning of the simulations. The exact list of simulation parameters used is provided in Annex A4.2 .

Univariate logistic regressions were then applied to the genetic data of individuals at the 300th generation in order to identify the most significant associations, which should highlight loci under selection. Univariate and multivariate SPAGs were then computed to estimate the probability of finding these genotypes across the simulated landscape. The results were validated using a cross-validation procedure presented in Box 1 (page 105).

Moroccan and European goats

Genetic data

Two genetic datasets characterising goats (Capra hircus) were used as case studies. The first one was produced in the context of the NEXTGEN project (Alberto *et al.*, 2018) and the second was collected by the ADAPTMAP consortium (Stella *et al.*, 2018; http://www.goatadaptmap.org/).

The NEXTGEN project produced whole genome sequences data for 161 Moroccan goats from 6 different local breeds. Since goat production system in Morocco is mainly free-range, these goats are living from 8 to 12 months outdoors (Boujenane, 2005), and are confronted to contrasting environmental conditions, from the Sahara desert to the Atlas Mountains (see Figure A in Annex A4.3). The goats were sampled in 161 farms chosen such to be representative of the range of environmental conditions observed in Morocco (Stucki, 2014). The sequencing method is described by Benjelloun *et al.* (2015) and allows to genotype 31.8 M of SNPs mapped to the goat's reference genome CHIR v1.0 (Dong *et al.*, 2013).

The ADAPTMAP consortium gathered genetic data for 4'563 goats from 144 breeds, sequenced worldwide with the CaprineSNP50 BeadChip and mapped on the most recent goat reference genome ARS1 (Bickhart *et al.*, 2017). The goats were georeferenced to the place where they have been sampled. We used here a subset of these data, constituted of individuals from Switzerland, North of Italy and France. This represented 458 individuals distributed in 196 locations,

with 1 to 39 individuals per site. In order to avoid overweighting of some locations, we selected a maximum of five individuals per sampling site. These five individuals were chosen as the subset showing the highest Nei's genetic distances, computed with the function *dist.genpop* from the package *adegenet* (Jombart, 2008) in the *R* environment (R Development Core Team, 2008). The resulting dataset contains 382 individuals from 196 locations and 11 different breeds (see Figure B in Annex A4.3).

Both genetic datasets were filtered such as to keep only autosomal, bi-allelic SNPs, with a maximum missingness per individuals and per site of 0.05 and a maximum major genotype frequency of 0.9. The final datasets contain 8,497,971 SNPs for the Moroccan goats and 46,294 SNPs for the European ones.

Environmental data

The climatic conditions of the sampling locations were characterised using the 19 bioclimatic variables (Annex A4.4) from the WorldClim database (https://www.worldclim.org/), representative of the period 1960-1990 (Hijmans *et al.*, 2005). Each variable was retrieved as a raster layer with a spatial resolution of 30 arc-seconds (approx. 1km^2) and values were extracted for all sampling locations using the *extract* function from the R-package *raster* (Hijmans and van Etten, 2012). In order to get similar ranges of values for all bioclimatic variables, which makes it easier to compare the subsequently derived models, all variables were standardised for each dataset, by subtracting the mean and dividing by the standard deviation. Some of the bioclimatic variables are highly correlated. However, we choose to keep all of them to be able to identify *a posteriori* which variable had the strongest effect. Since no models computed involved more than one environmental variable simultaneously, this collinearity will not impact the results.

Population Structure

The genetic population structure was estimated with a Principal Component Analysis (Price *et al.*, 2006; Reich *et al.*, 2008) computed with the function *snpgdsPCA* from the *SNPRelate* R-package (Zheng *et al.*, 2012). In order to avoid a strong influence of SNP clusters on this analysis, we used here a pruned set of SNPs that are in approximate linkage equilibrium with each other. The pruning was performed with the function *snpgdsLDpruning* from the *SNPRelate* package, with a threshold D'=0.2. The resulting datasets contain 59,224 SNPs for the Moroccan goats and 14,571 SNPs for the European ones.

Logistic regressions and SPAGs

For the two datasets, logistic models were computed for each genotype with the 19 bioclimatic variables. The statistical significance of the model was assessed using Wald test and log-likelihood ratio (G), both corrected for the false-discovery rate due to multiple comparisons using the procedure proposed by Benjamini and Hochberg (1995), under an expected false discovery rate (FDR) of 0.05 (i.e. 5% of the results expected to be false positives).

In order to lower the number of false positive resulting from demographic processes instead of natural selection (Li *et al.*, 2012), logistic models were computed with the addition of covariates corresponding to the coordinates of individuals on the significant components of the PCA. The significance of the models with population covariates was assessed using a Wald test and a log-likelihood ratio which compares the model with environment and covariates to the model with

covariates only. An association was considered as significant if both the models without covariates and with population covariates were significant.

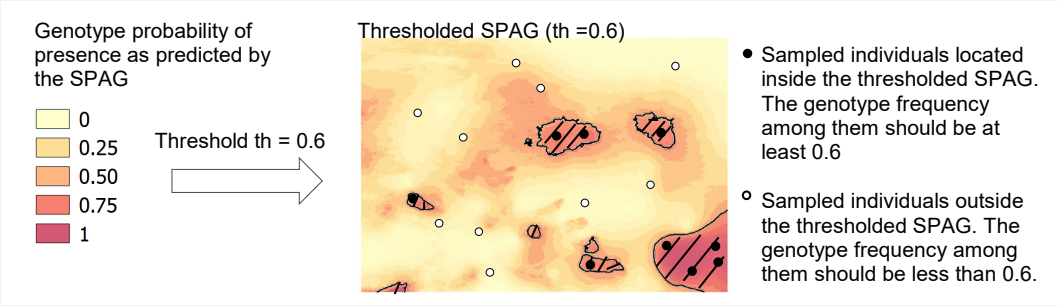
Finally, to identify potential functions of the SNPs involved into the significant associations, we used the NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/genome/?id=GCF_000317765.1) to search for the presence of annotated genes in the genomic region of 10kbp surrounding the SNPs of interest. All analyses were computed using a combination of the Samβada software (Stucki *et al.*, 2017) and a custom R-script based on the *glm* function.

Validation procedure

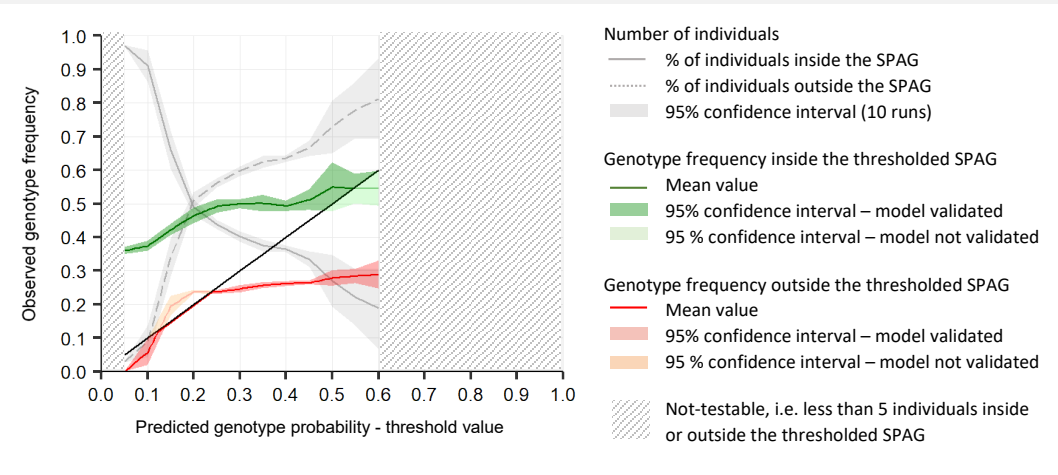
For the simulated data as well as for the two goat datasets, SPAGs were validated with a cross-validation procedure, using 25% of the individuals to compute the SPAG (i.e. 50 individuals for the simulated datasets, 41 individuals for the Moroccan goats and 96 for the European) and the remaining 75% to test it. Training individuals were selected such to represent the entire range of values of the environmental variable under study (see Annex A4.1 for the exact procedure). The model was validated using the Area Under the Receiver Operating Curve computed with the testing dataset (AUCtest, (Fielding and Bell, 1997)) and a custom validation graph presented in Box 1. The cross-validation procedure was repeated 10 times.

Box 1: Validation Procedure

SPAGs indicate the probability of finding one or more genotypes of interest in a territory (panel below). For a given threshold value (e.g. th=0.6), we can thus use the SPAG to delimit the area where the probability of finding the genotype(s) of interest is predicted to be greater or equal to this threshold (e.g. probability>=0.6). If the SPAG is valid, the frequency of the genotype(s) observed among the testing individuals located within the thresholded SPAG should effectively be greater or equal to the threshold value, whereas it should be less outside.



To validate the SPAGs, we thus calculated the observed genotype frequencies inside and outside the thresholded SPAGs for each threshold value between 0 and 1 (with a step of 0.1) and we presented the results on a graph (panel below). The green line indicates the genotype frequency observed inside the thresholded SPAG, whereas the red line shows the genotype frequency observed outside it. A black line indicates the limit case where the observed genotype frequency is equal to the threshold value. The SPAG is thus validated if the green line remains above the black line and the red line remains below it. The green and red areas around the lines indicate the 95% confidence intervals for each line, computed on the basis of the 10 cross-validation runs. We also presented on the graph the percentage of individuals located within the thresholded SPAG (grey line) and outside it (dotted grey line). The hatched grey areas indicate ranges of testing values where there was less than 5 individuals remaining inside or outside the thresholded SPAG, which was therefore considered not to be usable for the validation.



For a threshold value th=0.4, the SPAG is validated since 50% of the testing individuals located within the area SPAG≥0.4 carry the genotype of interest, whereas only 26% carry it ouside (SPAG<0.4). Inversely, the model is not validated for th=0.6, since only 54% of individuals carry the genotype of interest whithin the area where the SPAG predicted a probability of at least 0.6 (SPAG ≥0.6). The model is also not valid for a value th=0.2 since 23% of individuals carry the genotype of interest in the area SPAG<0.2.

Projections under climate change

We use the two goat datasets to present an application of the SPAG for predicting the distribution of adaptive genotype(s) under climate change. In order to predict the genotype frequency optimal for future conditions, we retrieved Worldclim data for the year 2070, corresponding to a strong climate change scenario from the Max Planck Institute Earth System Model (MPI-ESM-LR) (Giorgetta *et al.*, 2013) with a Representative Concentration Pathway equals to 8.5 (RCP 8.5). We then assume that the optimal genotype frequency for future conditions should be close to the genotype frequency currently observed in areas with climatic conditions resembling the future ones. We thus applied the current parameters of the logistic regressions on the future environmental variables in order to derive the future SPAGs for the genotypes of interest. We then study the shift between the current and future SPAGs to identify vulnerable populations for which specific genotype frequencies should be much higher so that individuals can adapt to the future conditions.

4.3.4 Results

Simulated results

The 10 most significant genotype-environment associations obtained with the simulated datasets ranked on the basis of the likelihood ratio (G) are presented in Table 4-1. We observe that the three loci simulated as under selection (L0, L1, L2) are coherently identified as the most significantly associated with the environmental variables under study.

Table 4-1 – Most significant models – Simulated datasets
10 most ignificant models obtained for the analysis of the simulated datasets, ranked based on the likelihood ratio.

ID	Marker	Locus	Geno.	Env.	G score	pG	pW	β0	β1
S1	L1A0A0	L1	A0A0	X1	92.30	7.44E-22	9.41E-14	-0.46	2.65
S2	L1A1A1	L1	A1A1	X1	90.49	1.86E-21	1.51E-12	-1.24	-2.94
S3	L2A1A1	L2	A1A1	X2	68.63	1.19E-16	9.02E-11	-1.59	-4.37
S4	L0A1A1	L0	A1A1	X0	65.69	5.29E-16	4.04E-11	-0.66	-2.35
S5	L0A0A0	L0	A0A0	X0	64.95	7.69E-16	6.92E-11	-1.58	2.61
S6	L2A0A0	L2	A0A0	X1	63.86	1.34E-15	1.37E-11	-0.55	2.04
S7	L2A1A1	L2	A1A1	X1	61.04	5.59E-15	4.12E-10	-1.56	-2.40
S8	L2A0A0	L2	A0A0	X2	60.94	5.88E-15	2.52E-10	-0.54	3.45
S9	L0A0A0	L0	A0A0	X2	43.43	4.39E-11	2.81E-08	-1.41	3.11
S10	L42A0A0	L42	A0A0	X0	42.44	7.30E-11	5.93E-09	0.06	1.67

Geno = Genotype, Env=environmental predictor, pG=p-value associated with the G score, pW=p-value associated with the Wald score, β0 and β1 are the parameters of the logistic regression.

Figure 4-3B presents the univariate SPAG for the model S1 (Locus L1, genotype A0A0 associated with the environmental variable X1). Since this locus was simulated as under selection with the east-west gradient X1, the resulting SPAG coherently shows a similar gradient. The results are validated by the validation graph for almost the entire range of probability values, except close to 0.8, where the SPAG slightly overestimate the probability of presence of the genotype (the green line falls below the black one, i.e. the observed genotype frequency is lower than what predicted by the SPAG).

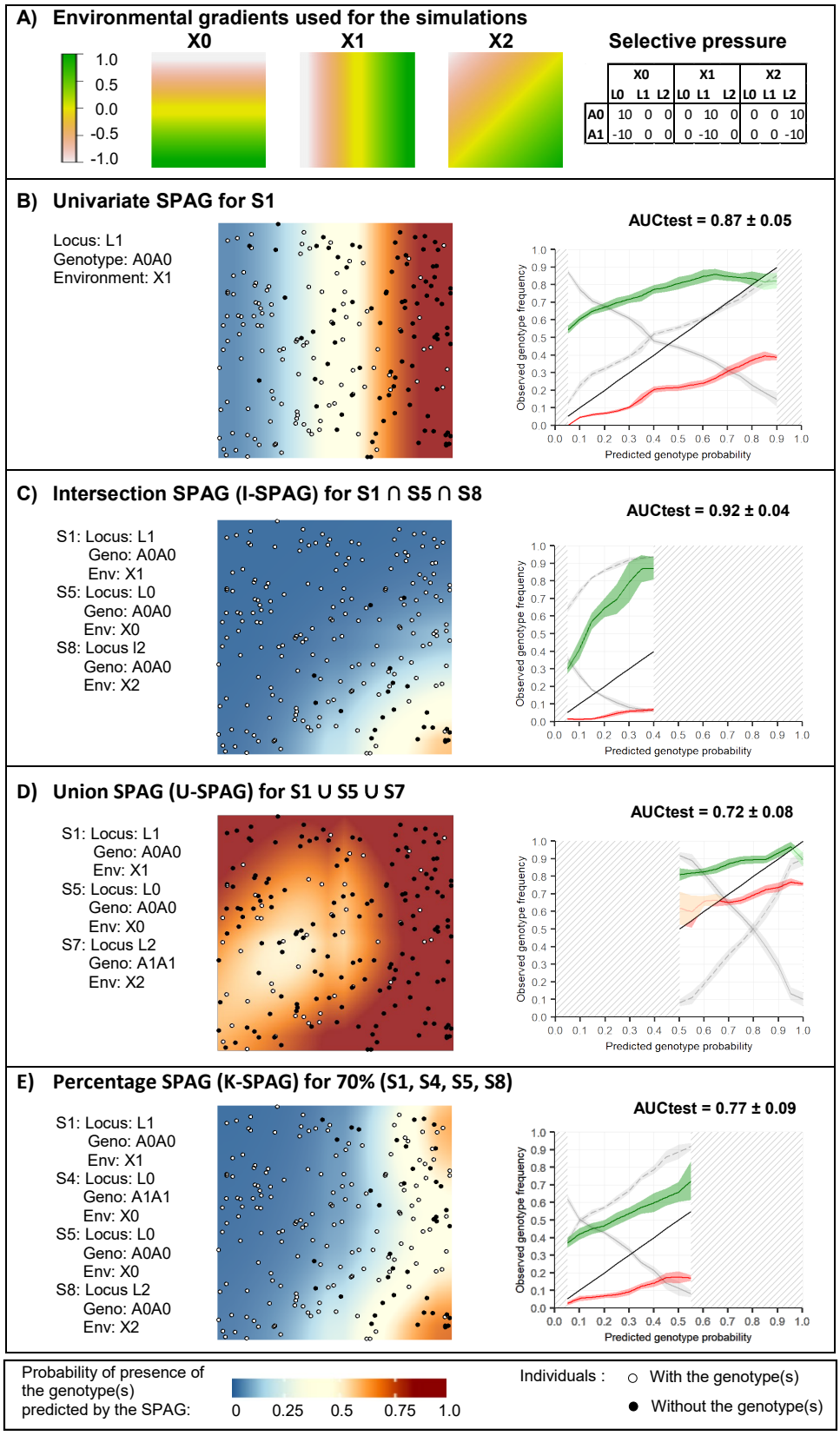


Figure 4-3 – SPAG – Simulated dataset
Univariate and Multivariate Spatial Areas of Genotypes Probability for the simulated dataset. The identifiers of the presented models (S1, S4, S5, S7, S8) refers to Table 4-1. Please refer to Box 1 to interpret the validation graphs shown on the right of each map.

Figure 4-3C shows the intersection SPAG for the genotype A0A0 of the three loci under selection. It indicates a very low probability of finding them all simultaneously, except in the South-East of the simulated area (where all environmental gradients considered had values close to 1). Figure 4-3D presents a union-SPAG of three genotypes. It shows that the probability of finding at least one of them is higher than 0.5 in most of the area. The validation graph indicates that the model tends to underestimate the probability of finding the genotypes for threshold values below 0.6 (the red line is above the black one, i.e. the genotype frequency outside the thresholded SPAG is higher than what predicted). Finally, Figure 4-3E depicts the probability of finding 70% of 4 genotypes of interest, i.e. at least 3 of them. This probability is low, except in the South-East and North-West part. Again, the validation graph indicates a good power of the SPAG to predict the probability of finding a set of genotypes of interest.

Moroccan goats

Population structure

For the Moroccan dataset, the cumulated variance explained by the 10 first PCA components on the SNP markers represents only 8.1% of the total variance and the increase in variance explained is almost proportional to the number of components, which highlights that there is no clear sub-structure. We therefore do not include any population structure on the subsequent analysis and computed only logistic regressions without any covariates.

Logistic regressions

More than 483 million logistic association models were computed. After correction for false discovery rate with a significant threshold of 5%, no model is significant according to the Wald score, but seven models are significant according to the G score (Annex A4.5). Among them, three models were strongly associated with the precipitation seasonality (bio15), which is a measure of the variation of monthly precipitation over the year. Following this initial result, we investigated in more details the adaptation to this bioclimatic variable. When considering only the associations involving bio15 (25,447,348 models), 78 models are significant after FDR-correction of G score, with a significant threshold of 5% (Annex A4.5). The SNPs involved in these models are located on seven different genomic regions (Table 4-2), corresponding to four annotated genes (DSG4, CDH2, KCTD1 and WRN) on the reference genome CHIR 1.0.

Table 4-2– Significant models – Moroccan datasets – Bio15
Significant models obtained for the analysis of Moroccan datasets with precipitation seasonality (bio15) after FDR correction.

ID	Chr	Start (BP)	End (BP)	Peak (BP)	Geno	GF	G	qG	β0	β1	Genes
M1	6	12'174'332	12'298'321	12'276'168	AA	21.74	38.74	0.004	-1.80	1.51	*(LincRNA)*
M2	13	43'436'394	43'438'732	43'436'394	GG	10.56	29.02	0.042	-3.14	1.80	-
M3	24	19'436'980	19'436'980	19'436'980	CC	76.40	34.75	0.008	1.55	1.29	-
M4	24	25'852'900	25'860'754	25'860'754	AG	38.51	34.75	0.008	-0.29	-1.07	DSG4
M5	24	28'799'029	28'833'762	28'833'253	TT	12.42	27.87	0.046	-2.72	1.58	CDH2
M6	24	30'566'869	30'584'692	30'566'869	TT	2.48	27.99	0.046	-25.69	-15.44	KCTD1
M7	27	25'930'079	25'933'133	25'930'079	GG	78.88	32.88	0.012	1.76	-1.35	WRN

Chr=Chromosome, Start=Start in base pairs of the region identified as under selection, End=End in base pairs of the region, Peak SNP = SNP of the most significant model on that region, Geno = corresponding Genotype, GF=corresponding Genotype Frequency, β0 and β1 = parameters of the logistic regression, G=G score (Log Likelihood ratio) of the model, qG=corresponding p-value corrected for FDR, Genes = Annotated genes on the genomic region.

Spatial Areas of Genotype Probability

Figure 4-4A shows the univariate SPAG for the genotype of model M1 presented in Table 4-2 (see Annex A4.7 for the other univariate SPAGs). The predicted probability of presence of the genotype is the highest in the extreme southwest of the country, near the Sahara desert. In this region, the variations of precipitation are the greatest (the standard deviation of monthly precipitation is more than 100% of the mean of monthly precipitation) and all goats carry the genotype of interest. In coastal areas, the predicted probability of finding the genotype is close to 0.5. In this area, the variations of precipitations are also high (more than 70%) and some of the sampled goats carry the genotype, while others do not. Finally, in the Atlas Mountains and in the northeast of the country, the probability of finding the genotype of model M1 is much lower (<0.2 in most areas). In these regions, variations of precipitation are less important (35-50%) and most of the goats sampled do not carry the genotype.

Two other markers positively correlated with bio15 were highlighted by models M3 and M5 (Table 4-2). Nevertheless, the I-SPAG presented in Figure 4-4B indicates that their simultaneous presence is very unlikely (probability <0.1 for most of the territory). On the contrary, the probability of finding at least one of the genotypes from the three models M1, M3 and M5, all positively associated with the coefficient of precipitation, is very high in many parts of the territory (U-SPAG, Figure 4-4C). Finally the K-SPAG presented in Figure 4-4D shows the probability that goats carry at least 50% of the four variants positively associated with the coefficient of precipitation (M1, M2, M3, M5), i.e. the probability to find at least two of them. This map is the most contrasted, showing a very high probability of presence near the coast and the Sahara desert (> 0.9) and a very low probability (<0.2) in the centre and northeast of the country. For all these multivariate cases, the mean AUC value for the testing dataset over the 10 runs is greater than 0.8. In addition, the validation graphs indicate that the SPAGs computed with 25% of the individuals generally enable a correct estimate of the probability of finding the genotype(s) of interest in the 75% remaining individuals. Only the U-SPAG (Figure 4-4C) tends to slightly overestimate the probability of presence since we observe a higher presence of the genotypes in the individuals located outside the thresholded SPAG as compare to what predicted by the SPAG (red line above the black line).

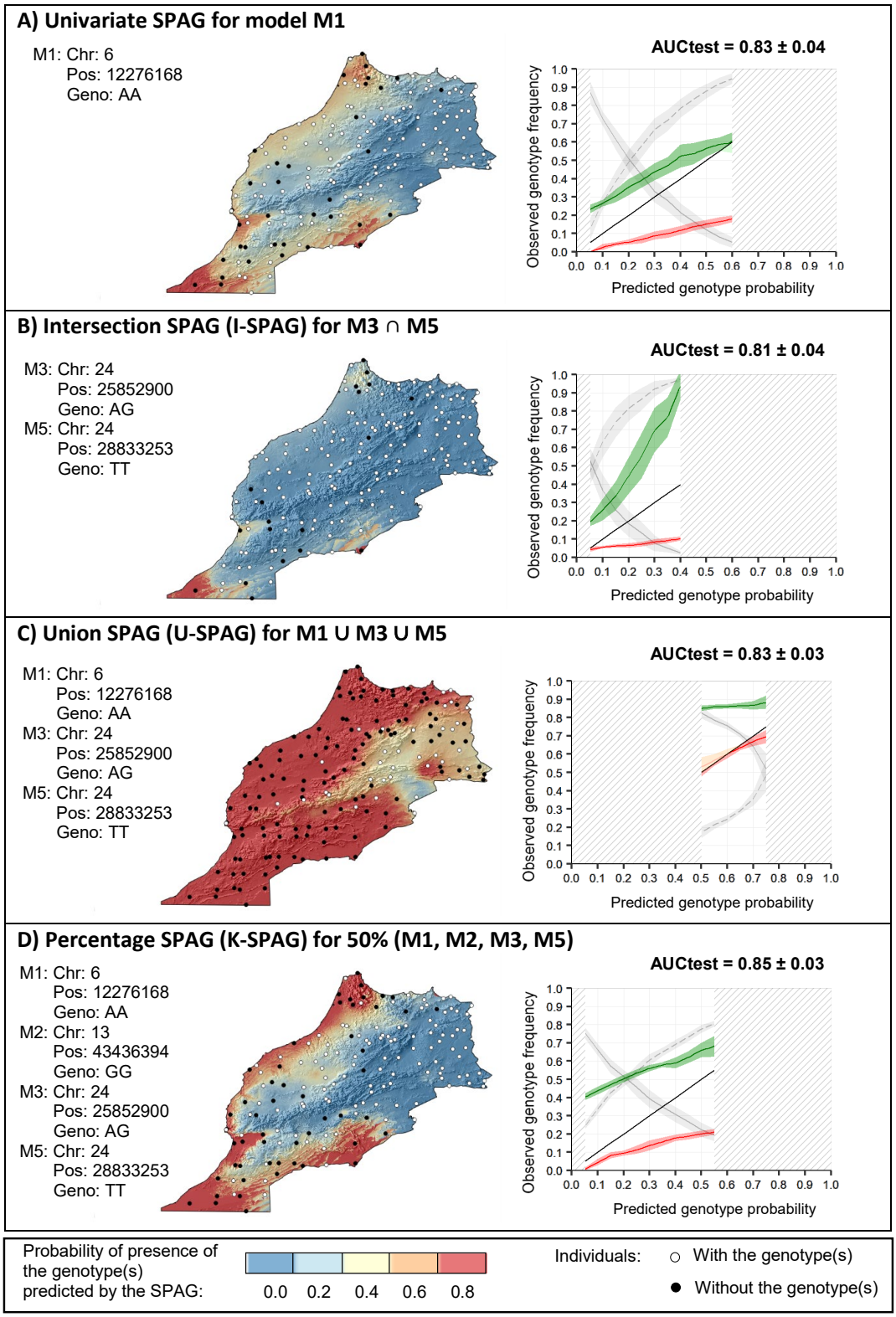


Figure 4-4 – SPAG – Moroccan dataset
Univariate and Multivariate Spatial Areas of Genotypes Probability for the Moroccan dataset. The identifiers of the presented models (M1, M2, M3, M5) refers to Table 4-2. The maps show the average genotype(s) frequency(ies) based on the 10 runs computed with different random selection of training sets containing 25% of the total number of individuals. AUCtest indicates the mean value of the AUC computed with the testing dataset over the 10 runs. Please refer to Box 1 to interpret the validation graphs shown on the right of each map.

Projections under climate change

Figure 4-5 shows the differences between the current SPAGs presented in Figure 4-4 and their corresponding projections for 2070. In Morocco, the precipitation seasonality (bio15) is predicted to increase in the northwest of the country, with a maximum increase of 5 to 10% in the extreme northwestern region (Tangier-Tetouan, see region's map in Annex A4.3) and to decrease in other areas, especially in the Atlas Mountains and near the Sahara (from -10 to -20%). The evolution of the univariate SPAG for model M1 (Figure 4-5A) consequently indicates the highest risk in the Tangier-Tetouan area, where the mismatch between the current and future SPAGs indicates that the probability of finding the genotype of interest should be 20% higher to find individuals well adapted to future conditions. However, many individuals in this area already carry the favourable genotype, and the risk for the population may thus be reduced thanks to natural gene flow. Nevertheless, this is not the case in the southwest of this area (Rabat, Casablanca) where the probability of finding the genotype should also be 10-20% higher according to the SPAGs difference, and none of the goats sampled there currently carry the adaptive variant. Similar observations can be made as regards the I-SPAG of M3 and M5 (Figure 4-5B), two other markers that may potentially confer an adaptation to high variations of precipitation. However, the U-SPAG (Figure 4-5C) highlights no vulnerable areas, which indicates that if the presence of at least one of the adaptive variants is sufficient to enable the adaptation to high variations of precipitation, no population may be at risk. Finally the K-SPAG (Figure 4-5D) also shows a risk area in the northwest of the country, where the probability of carrying the adaptive variants should be approximately 20% higher. Again, individuals in the northernmost part of this risk area may be less threatened due to the close presence of goats already carrying the favourable genotypes, whereas the population from the Rabat-Casablanca area may be more threatened due to the current much lower presence of the adapted variants.

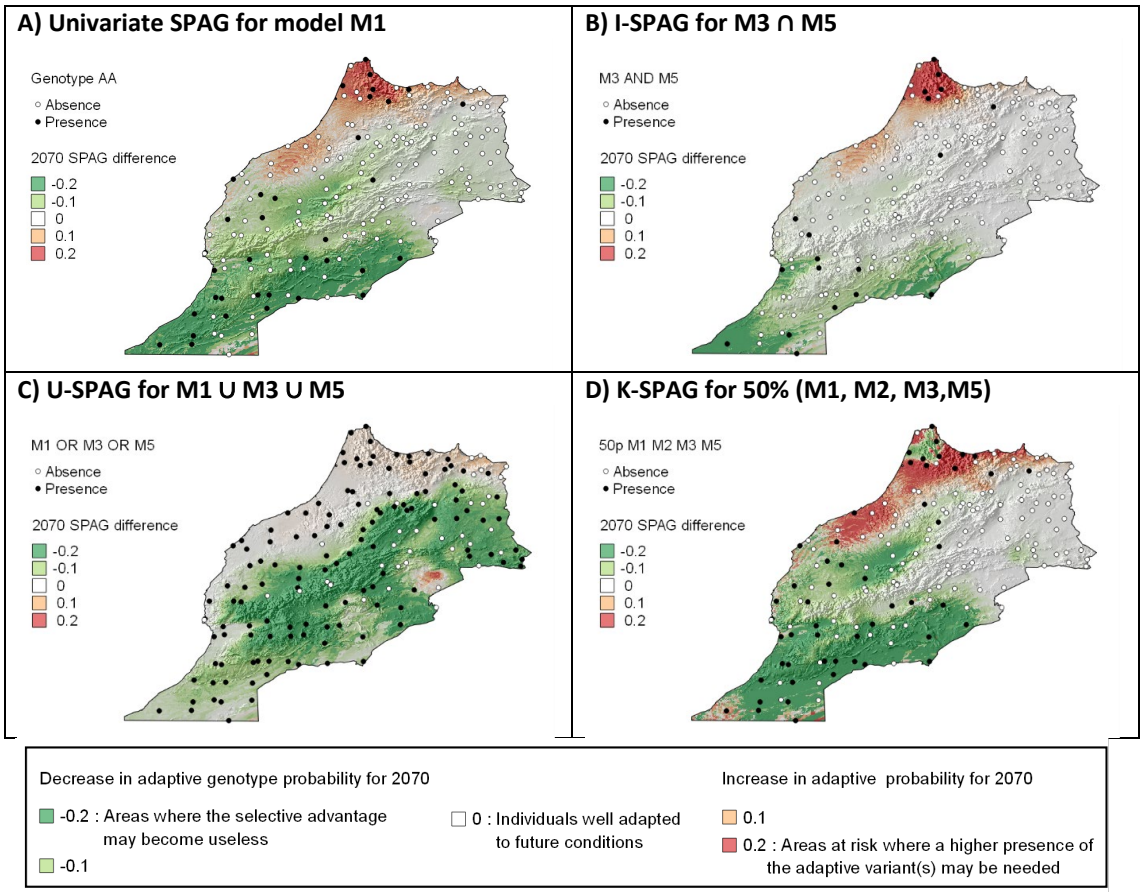


Figure 4-5 – Morocco - Predicted change in genotype probability for 2070
Predicted SPAG difference for 2070 considering the MPI-ESM-LR climate change scenario with RCP 8.5, for the Moroccan goats. The identifiers of the presented models (M1, M2, M3, M5) refer to Table 4-2. The maps show the average difference in probabilities of finding the genotype(s) based on the 10 runs computed with different selection of training sets.

European goats

Population structure

For the European goats, the first component of the PCA explains 6.2% of the total variance while the second, third and fourth components explain 2.0%, 1.7% and 1.6% respectively. The low variance explained by each PCA component indicates the absence of a clear population structure. However, since the variance explained by the first component is much higher than that explained by the next ones, it is possible that the first component is partially related to the population structure. We therefore computed logistic regressions without covariates and then logistic regressions with a covariate corresponding to the coordinates of goat individuals on the first component of the PCA.

Logistic regressions

More than 2.6 million logistic association models were computed, of which 4.9% were significant both without covariate and with the first PCA-component as covariate, according to both G score and Wald score corrected for a false positive rate of 5% (Annex A4.6). The ten genomic regions associated with the strongest G scores when computed without covariate are presented in Table 4-3. The corresponding models involved two bioclimatic variables related to precipitation (bio13 = precipitation of the wettest month, bio18 = precipitation of the warmest quarter) and two bioclimatic variables related to temperature (bio3 = isothermality, bio8 = mean temperature of the wettest quarter). Seven annotated genes correspond exactly to one of the SNPs identified: KRT12, CSN1S2, CACNB2, PRDM5, LOC102174324, PALM and NAV3.

Table 4-3– Significant models – European datasets
Models corresponding to the 10 most significant genomic regions (based on G score of the model without covariate) obtained for the analysis of the European dataset, considering all bioclimatic variables.

ID	ENV	CHR	BP	GENO	GF	qGpop	qWpop	qG0	qW0	β0	β1	Genes
E1a	bio18	19	40696776	GG	33.8	3.3E-09	3.9E-07	3.8E-17	1.2E-11	-1.41	1.30	KRT12
E1b	bio13	19	40696776	AA	40.3	4.3E-09	5.4E-07	1.4E-13	6.9E-10	-0.52	-1.07	KRT12
E2	bio18	1	38183832	AA	31.7	3.0E-09	1.3E-07	1.6E-15	1.2E-11	-0.97	1.13	-
E3a	bio3	6	86081075	CC	31.2	3.5E-11	1.2E-08	3.1E-14	4.2E-11	0.65	-1.06	CSN1S2
E3b	bio18	6	86081075	CC	31.2	9.6E-11	6.9E-08	5.7E-14	5.0E-10	0.71	1.12	CSN1S2
E4	bio8	13	32300758	GG	31.2	6.1E-10	9.0E-08	5.1E-14	5.6E-11	0.11	-1.02	CACNB2
E5	bio18	5	23213822	GG	39.0	6.0E-10	4.1E-08	7.3E-14	5.6E-11	-0.52	1.02	-
E6	bio8	6	4945809	AA	19.4	9.3E-07	7.9E-05	1.0E-13	9.2E-08	-2.06	1.55	PRDM5
E7	bio18	16	76397454	GG	22.8	1.1E-10	5.3E-07	1.0E-13	8.0E-09	1.30	1.29	LOC102174324
E8	bio18	7	67159272	CC	35.1	6.9E-10	1.3E-07	1.0E-13	2.3E-10	0.27	1.03	PALM
E9	bio3	5	7093719	GG	39.8	7.0E-11	1.8E-08	1.0E-13	9.0E-11	-0.63	1.02	NAV3
E10	bio18	14	85434737	AA	47.4	2.6E-09	1.6E-07	1.3E-13	2.2E-10	-0.16	-1.01	-

Chr=Chromosome, BP=Position in base pairs, GENO=Genotype of interest, GF=Genotype Frequency, qGpop (resp. qWpop) = FDR-corrected p-values of Gscore (resp. Wald score) of the model with the first PCA-component as covariate, qG0 (resp. qW0) = FDR-corrected p-values of Gscore (resp. Wald score) of the models without any covariate, β0 and β1 = parameters of the logistic regression without covariate, Genes = Annotated genes on the genomic region.

SPatial Areas of Genotype Probability

Figure 4-6A shows the univariate SPAG for the genotype of model E10 presented in Table 4-3 (see Annex A4.7 for the other univariate SPAGs). This genotype is negatively associated with precipitation of the warmest quarter (bio18). The predicted probability of finding this genotype is the lowest in the Swiss Alps (<0.1), slightly higher in the Jura, French Alps and Swiss Plateau (between 0.2 and 0.4) and above 0.5 everywhere else, with a maximum around 0.8 on the Mediterranean coast.

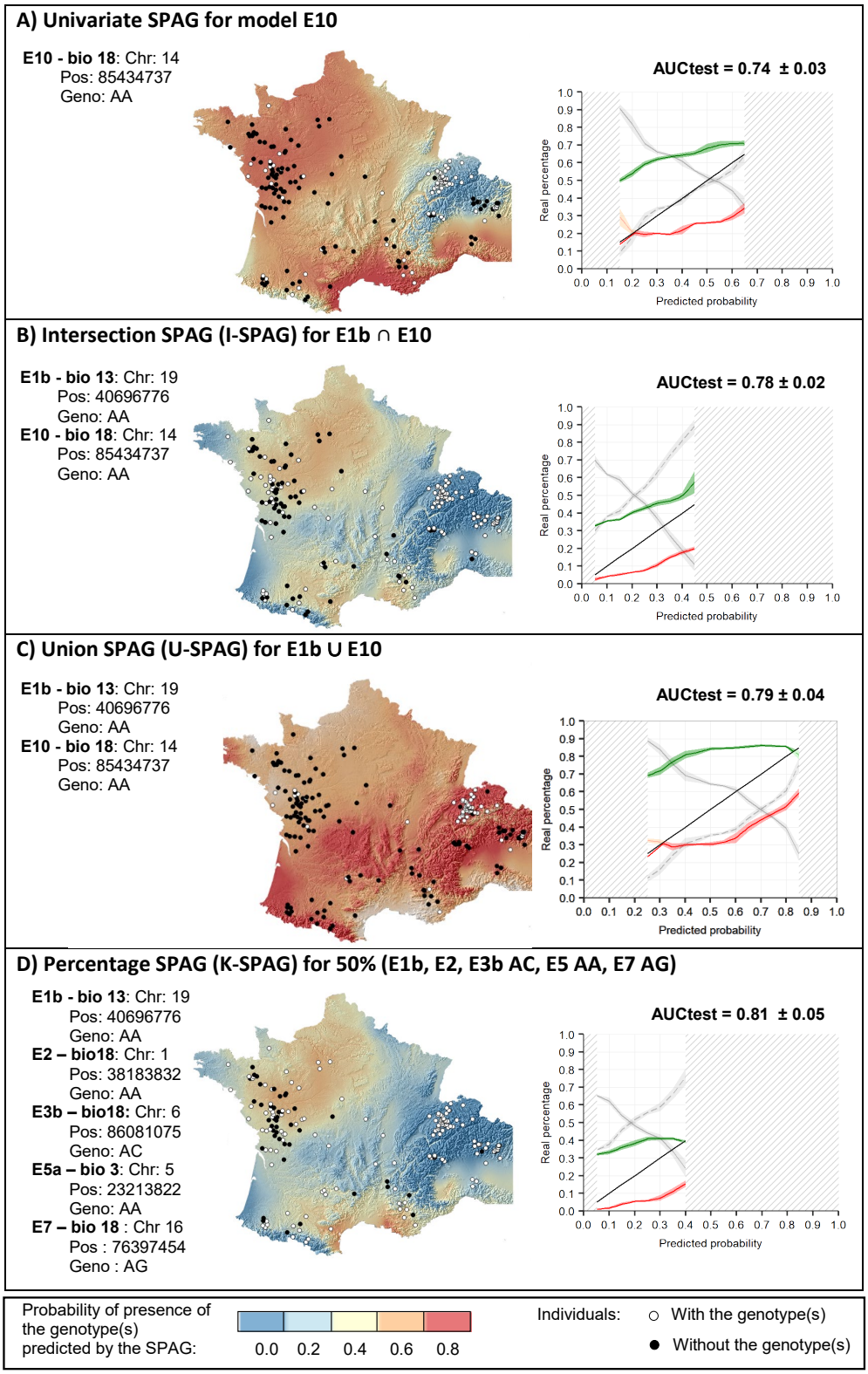


Figure 4-6 – SPAG – European dataset
Univariate and Multivariate Spatial Areas of Genotypes Probability for the European dataset. The identifiers of the presented models (E10, E1d, E2, E3b, E5, E7) refers to Table 4-3. The maps show the average genotype(s) frequency(ies) based on the 10 runs computed with different random selection of training sets containing 25% of the total number of individuals. Note that since up to five individuals can be localised on the same site, a black dot indicates a presence if at least 50% of the individuals of the site carry the marker(s). Please refer to Box 1 to interpret the validation graphs shown on the right of each map.

The I-SPAG (Figure 4-6B) shows the probability of simultaneously finding this genotype (model E10) and the genotype from model E1b, associated with low precipitation during the wettest month (bio13). The two models E1b and E10 may therefore indicate an adaptation to low values of precipitation. However, the I-SPAG indicates that their simultaneous presence is not very likely (predicted frequency < 0.6 everywhere). The predicted probability is the highest in the centre-north of France (regions Centre, Iles de France, East of Pays de la Loire, Normandie and South of Hauts de France, see Annex A4.3 for regions' map) and in the southern part (Occitanie and West of Provence), whereas in the Alps, Jura and most of Switzerland, the probability of simultaneously finding these two genotypes is close to 0 and none of the sampled goats carry them both. The probability of finding at least one of these two variants, presented on the U-SPAG in Figure 4-6C, indicates a trend similar to the probability of presence of E10 alone (Figure 4-6A), but with even stronger contrast between the Alps-Jura-Switzerland area (frequencies < 0.3) and the rest of the territory (frequencies > 0.7). Finally the K-SPAG (Figure 4-6D) shows the probability of finding at least 50% of five genotypes negatively associated with precipitation in the warmest quarter (bio18), i.e. the probability of finding at least three of them. Note that for the models positively associated with bio18 in Table 4-3, we used the alternative genotype that was the most significantly negatively correlated with bio18 (indicated after the model ID). The resulting SPAG is very close to the I-SPAG of E1b and E10 (Figure 4-6B). For all cases presented, the validation graphs indicate that the SPAGs computed with 25% of the individuals generally correctly predict the genotype frequency of the remaining 75% of individuals.

Projection under climate change

Figure 4-7 shows the differences between the current SPAGs presented in Figure 4-6 and their corresponding projections for 2070. The model E10 presented in Figure 4-6A was related to the precipitation of the warmest quarter (bio 18), which is projected to decrease by 20 to 180 mm over the whole study area until 2070. The projected precipitation loss is maximum in the Alps (-120 to -180 mm), on the Mediterranean coast (-100 to -130) and in the centre-south of France (West of Auvergne and East of Nouvelle Acquitaine, -90 to -120 mm, see region's map in Annex A4.3). As a consequence, the mismatch between current and future SPAGs for model E10 (Figure 4-7A) indicates that all populations may be threatened by climate change, since the probability of finding the adaptive genotype should be 10 to 20% higher everywhere. None of the goats living in the Alps and Switzerland currently carry this adaptive genotype and these populations may thus be particularly vulnerable. When considering the I-SPAG of E10 and E1b (Figure 4-7B), a higher risk is highlighted in the centre-south of France (West of Auvergne and its surrounding), northeast of France (Alsace in East of Grand-Est), Swiss Plateau, and northwest of Italy. In centre-south of France and Swiss-Plateau, almost none of the goats sampled carry simultaneously the two genotypes and the populations may therefore be particularly threatened. The U-SPAG of the same genotypes (Figure 4-7C) shows results very similar to the univariate SPAG for E10 alone (Figure 4-7A). If the presence of at least one of these two genotypes may be sufficient to allow adaptation to low precipitation, then the most threatened regions will again be the Alps and Switzerland, where both markers are currently absent. Finally, the evolution of the K-SPAG (Figure 4-7D) shows a high risk in a large part of France and northern Italy, where the SPAG's mismatch indicates that the probability of finding the adaptive genotypes should be more than 20% higher.

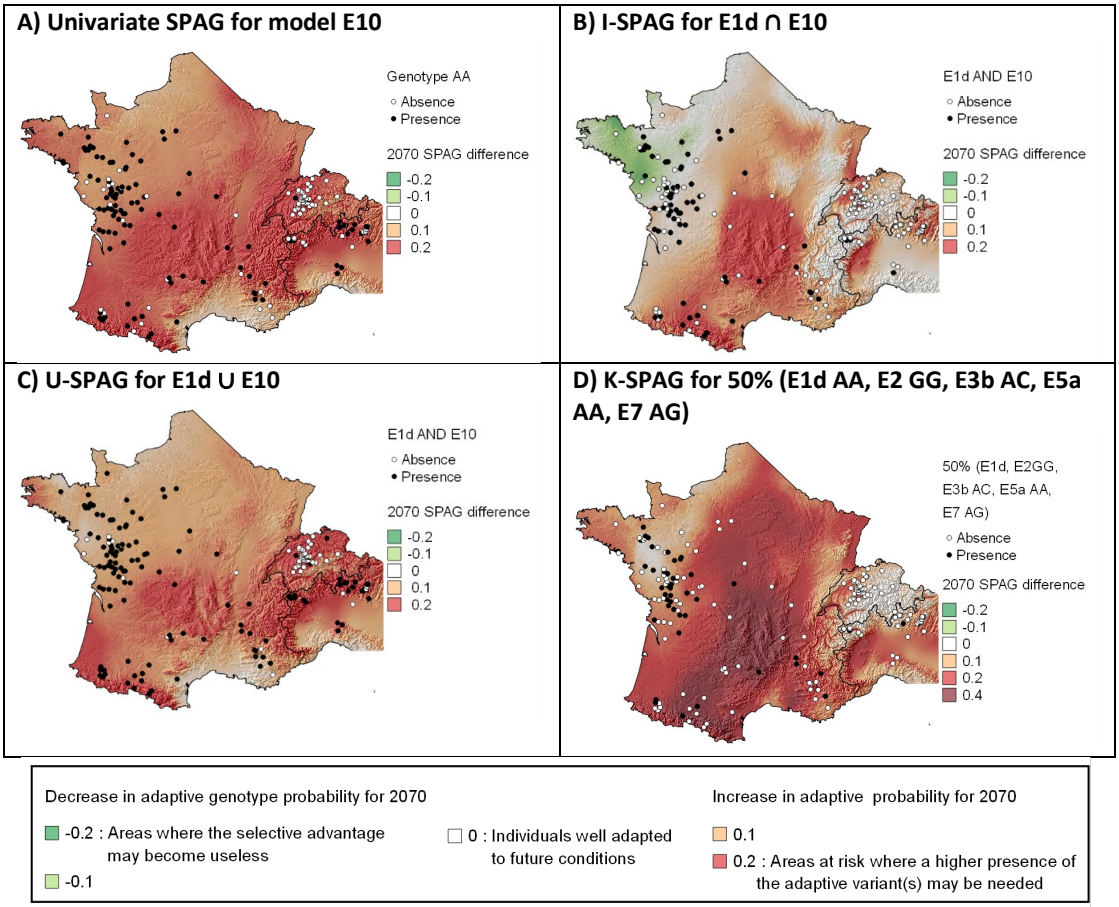


Figure 4-7 – Europe - Predicted change in genotype probability for 2070
Predicted change in genotype probability for 2070 considering the MPI-ESM-LR climate change scenario with RCP 8.5, for the European goats. The identifiers of the presented models (E10, E1b, E2, E3b, E5a, E7) refer to Table 4-3. The maps show the average difference in probability of finding the genotype(s) based on the 10 runs computed with different random selection of training sets.

4.3.5 Discussion

Mapping genotype probabilities

The first utility of SPAGs is to quantify the current probability of finding beneficial genotypes or the expression of favourable traits in plant and animal populations, even in regions where no individuals have been sampled. Our results show that with few training individuals (i.e. 50 simulated individuals, 41 Moroccan or 96 European goats), a good estimate of the probability of finding genotype(s) of interest is possible. The univariate models presented here have already been applied to map the genotype frequencies of adaptive variants of the Scandinavian brown bears (Joost, 2006), Moroccan sheep (Rochat *et al.,* 2016) and coral reefs (Selmoni *et al.*, 2019). Multivariate models are presented here for the first time and, according to the validation procedure applied on a simulated dataset and two case studies, they appear to be powerful in estimating the combined probability of finding several genotypes potentially correlated with different environmental variables. With the I-SPAG, the resulting probabilities may rapidly become very low, but this model could be used when we suspect that the simultaneous presence of some adaptive genotypes is needed to ensure the adaptation, or when we would like to highlight the probability of simultaneously finding variants that may confer an adaptive response to different environmental

variables (e.g. low precipitation and high temperature). At the other extreme, the U-SPAG may rapidly indicate high probabilities of presence in most parts of the territory, but it can be used when it is suspected that the presence of at least one of the genotypes may be sufficient to confer adaptive potential. Since it is generally difficult to know whether the simultaneous presence of adaptive genotypes is needed or whether an union is sufficient, K-SPAG offers an interesting compromise, allowing the identification of populations that retain a given percentage of variants, which can help delineate areas where there is the highest probability of finding individuals with high adaptive potential.

From SPAG to conservation

The study of the shift in SPAGs under climate change conditions can help identify 1) well-adapted populations, where individuals currently show adaptive genotypes that seem to be optimal under future conditions, 2) populations at risk where the current genotype frequency is not optimal, but where the favourable genotypes are already present in the population, thus potentially allowing a natural increase in genotype frequency through gene flow and 3) threatened populations where optimal variants are currently lacking but would be needed to ensure adaptation to future climate. These identifications may be of great value for conservation planning. Indeed, when prioritising conservation areas, success could be enhanced by choosing to preserve preadapted individuals that already carry functional variants conferring them good adaptation to future climate (Orr and Unckless, 2008). Moreover SPAGs can also be used to prevent the translocation of individuals that do not currently carry the variants favourable for future conditions at the target site, which would result in a reduction or loss of adaptive potential of the target populations (Weeks *et al.*, 2011). In addition, conservation plans can be developed to increase the survival capacity of threatened populations. This may involve assisted gene flow to import adaptive variants into a population where they are lacking (Aitken and Whitlock, 2013; Kelly and Phillips, 2016) or artificial selection of individuals already pre-adapted to future conditions (Hoffmann, 2010). However, this has to be undertaken carefully since the selection of locally adapted individuals can result in a loss of genetic diversity (Savage *et al.*, 2018), which may decrease the potential of populations to adapt to new environmental changes. Kardos and Shafer (2018) therefore proposed that gene-targeted conservation measures should only be taken with traits affecting vital processes of the species and when phenotypic variation is large enough to ensure a high probability of success. Finally, the SPAG maps presented could be integrated into decision-making frameworks considering the adaptive potential when defining the vulnerability of species (Bonin *et al.*, 2007; Williams *et al.*, 2008; Sgrò *et al*., 2011; Dawson *et al.*, 2011; Razgour *et al.*, 2018) or in more global decision frameworks that take into account other vulnerability factors such as the predation level or habitat loss.

The goats example

Several signatures of local adaptation where identified for the goats under study. In Morocco, three of the genes identified (DSG4, KCTD1 and CDH2) may be related to the development of hair (Kljuic *et al.*, 2003; Ling *et al.*, 2014; Wang *et al.*, 2017; Zhang *et al.*, 2019) or skin properties (Hayashi *et al.*, 2007). These results suggest that goats confronted with high variations of precipitation may have adapted through a modification of hair or skin traits, which could for example ensure a better water repulsion. In Europe, two of the genes highlighted as potentially conferring an adaptation to drought conditions (KRT12 and PRDM5) may be related to properties of the cornea (Kao *et al.*, 1996; Burkitt Wright *et al.*, 2011). They could thus potentially highlight an

adaptation to higher UV-radiation associated with driest conditions. The other genes identified are related the casein content of the milk (CSN1S2, Ramunno *et al.*, 2001), the calcium channel and energy pathway (CACNB2, Cardona *et al.*, 2014) or the skin properties (PALM and NAV3, Kutzleb *et al.*, 1998; Karenko *et al.*, 2005). Many of the genes highlighted on the two case studies may therefore be associated with a function that can be influenced by climate, which reinforces the potential that they are true signatures of local adaptation. However, it is known that logistic regressions such as implemented here, as most of the other methods, may lead to the identification of false positive (Stucki *et al.*, 2017), and the results should thus be confronted with other methods available to detect signatures of natural selection. In addition, although previous studies show the power of genotype-environment associations to predict phenotype (Lasky *et al.*, 2015; Vangestel *et al.*, 2018) or fitness (Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011), more investigations are needed to verify that the genotypes identified are really conferring an adaptive advantage (Funk *et al.*, 2019).

The Moroccan case study highlighted that goat populations from the surroundings of Rabat and Casablanca may lack adaptive genotypes potentially conferring an advantage to face high variation of precipitation. If the adaptive role of these genotypes is validated, goat populations in this region may be threatened. Because of the great economic and social importance of goats in Morocco, it is crucial to preserve viable populations. Indeed, in this country, agriculture contributes to 12 to 24% of the national GDP and employs 40% of the total active population (Boujenane, 2005). Livestock farming, especially small ruminants, is the most important sector of agriculture and goat farms account for 20% of the total number of farms (Boujenane, 2005). It is therefore important to consider preserving or introducing the adaptive genotypes on each vulnerable population. This could be done for example by favouring crossbreeding with individuals from the southern or north-eastern part of the country, where adaptive genotypes are currently well present, and by avoiding breeding or translocation with exotic goats or goats from the Atlas or Oriental regions. In the northern part of Morocco (Tanger-Tetouan regions), goat populations represent 12% of the national goat populations (Chentouf, 2014), and they play an important role in preserving food security (Godber *et al.*, 2016). In this region, crossbreeding with exotic breeds has been introduced to improve milk production (Boujenane, 2005; Godber *et al.*, 2016). However, our results show that the frequency of adaptive genotypes should increase in the goat populations from this region, and that it is therefore essential to maintain local individuals with the necessary adaptive genotypes.

Limitations and perspectives

The SPAG approach presented appears to be powerful for mapping the probability of finding locally adapted genetic variants in a landscape. However, the adaptation process is complex and often involves polygenic traits (Pritchard and Rienzo, 2010), for which the detection power of the genotype-environment associations may be reduced (Villemereuil et al., 2014; Harrisson et al., 2014). In this case, it may be advisable to use multivariate genotype-environment association models (Forester et al., 2017) or to integrate other methods to identify SNPs related to polygenic adaptation (Zhou et al., 2013; Lasky et al., 2015). In addition, since the results of the shift under climate change may be highly dependent on the climate change scenario considered, computations should be performed with various scenarios and less weight should be given to the conclusions not consistent within scenarios (Reside *et al.*, 2018). Finally, in order to assess the real

vulnerability of populations, an analysis of connectivity should be carried out to highlight the potential of natural gene flow to increase the probability of finding favourable genotypes in threatened populations.

SPAGs could also be used to predict the presence of genotype(s) associated with other pressures showing a spatial distribution, such as the presence of a parasite (Vajana *et al.*, 2018) or a predator (Cousyn *et al.*, 2001) or the urbanization level (Harris and Munshi-South, 2016). Very similar models can also be derived to predict allele frequencies instead of genotypes frequencies or to integrate other covariates (e.g. to consider autocorrelation or to use other indicators of population structure such as the Admixture coefficients). In addition, SPAGs are provided as maps, which enables a visual identification of threatened populations and could thus facilitate discussions between different conservation actors. SPAGs therefore constitute a valuable tool to support conservation decisions, especially under current changing climatic conditions.

4.3.6 Code availability

The main R codes developed for this study are available on GitHub:
https://github.com/estellerochat/SPAG.

Chapter 5 TOWARDS A CONSERVATION FRAMEWORK

5.1 Combining modelling tools

In the previous three chapters, we presented modelling approaches and tools using geo-environmental data to identify vulnerable populations due to environmental changes. First, Ecological Niche Modelling (ENM) integrating the spatio-temporal variability of environmental predictors can be used to identify landscape suitability for a species under current and future environmental conditions. The difference between current and projected suitability values can be used to quantify the extent to which species or populations are threatened by a reduction or degradation of their potential ecological niche. This can be referred to as the "exposure" of the species (Dawson *et al.*, 2011). This exposure may result from the direct effect of environmental changes or from indirect effects related to increased suitability for invasive species, pathogens, predators or competitors. Several studies directly used this measure as an index of threat to populations facing environmental changes (Elith and Leathwick, 2009). However, as previously mentioned, when populations are confronted with a reduction in the suitability of their territory, their vulnerability will depend on their dispersal and adaptive capacity (Dawson *et al.*, 2011; Catullo *et al.*, 2015). Despite this, few studies have addressed all these elements simultaneously (Beever *et al.*, 2016; Waldvogel *et al.*, 2020). In this chapter, we show how combining the various modelling tools presented before can help to integrate these different components for the identification of vulnerable populations.

5.1.1 ENM and connectivity

Landscape graphs can be used to estimate the potential for strongly exposed populations to move to more favourable habitats identified with ENM. Using the modelling tools presented in this thesis, an index of dispersal opportunity could be assigned to each population as a function of the number of paths leading from that population to other suitable habitats. The combination of ENM predictions with this dispersal opportunity index may enable the identification of threatened populations that are facing high exposure with limited dispersal possibilities. Some studies have already implemented similar combinations of ENMs and connectivity analyses based on resistance maps (Brown, 2014; Razgour *et al.*, 2018). Since the identified threatened populations are limited in their dispersal, they cannot move to more favourable habitats, and they are forced to adapt *in-*

situ to avoid extinction. Genetic simulations based on landscape graphs and validated with empirical genetic data can provide a first estimate of the adaptive capacity associated with the level of genetic diversity of the populations. Previous studies thus combined ENMs with genetic simulations to identify vulnerable populations (Brown *et al.*, 2016). However, such applications remained limited.

5.1.2 ENM and locally adapted genetic variants

Few studies have suggested tools to consider local adaptation when identifying vulnerable populations with ENMs. The proposed methods are usually based on a subdivision of species occurrences into populations currently associated with different climatic conditions and the computation of one ENM per population. For example, Hällfors *et al.* (2016) divided occurrence data into two populations using a clustering method based on environmental data. They computed independent ENMs for each of these populations. Comparing the results with an ENM built with all occurrences simultaneously, they highlighted large differences in suitability values, indicating that the environmental conditions important for modelling are population-specific, probably due to local adaptation. Razgour *et al.* (2019) proceeded similarly, but used signatures of local adaptation to group individuals into populations. That way, they identified two distinct populations with different signatures of local adaptation, one adapted to cold-wet and one to dry-hot climate. Again, they showed that population-ENMs lead to different results than a global ENM using all occurrence data. Their results demonstrated that when adaptive ability is not taken into account, ENMs predictions can overestimate the vulnerability of species due to the loss of suitable areas. Ruegg *et al.* (2018) used gradient forest modelling to calculate genomic vulnerability as a function of the mismatch between current and future genotype-environment associations (Fitzpatrick and Keller, 2015; Bay *et al.*, 2018). They did not combine it directly with ENM, but showed a correlation between genomic vulnerability and current species abundance.

In this thesis, we presented the SPAG tool, which can also be used to map the probability of presence of locally adapted genetic variants and to identify vulnerable populations due to a mismatch between current and future projections. SPAG output, in the form of a raster layer, could easily be combined with other rasters, such as those obtained with ENM. This combination may help to identify populations facing high exposure and limited adaptive capacity due to the lack of locally adapted genetic variants favourable for future climatic conditions.

5.1.3 Adaptation and connectivity

Analysing the difference between current and projected SPAG enables the identification of populations currently carrying genetic variants favourable to future conditions and populations lacking them. If dispersal is possible between these two populations, migration may enable the spread of favourable variants into maladapted populations, thus reducing their genomic vulnerability (Exposito-Alonso *et al.*, 2018). Analysing the connectivity between well-adapted and vulnerable populations is thus an essential step to better assess the adaptive potential of populations. In this context, the combination of landscape graphs and SPAGs can enable a better estimate of populations' vulnerability. Using landscape graphs, a favourable-dispersal index could be computed for each maladapted populations based on the number of potential paths leading to well-adapted ones. Populations facing simultaneously high exposure, low dispersal possibility, low genetic diversity, maladapted variants and low favourable-dispersal may thus be the most vulnerable to environmental changes.

5.1.4 Modelling framework

Based on the combinations of modelling tools previously discussed, Figure 5-1 presents a framework for identifying vulnerable populations in 11 steps:

1. Build ENM to predict suitable areas for competitors/pathogen-host species under current and future environmental conditions.
2. Build ENM to predict current and future suitable areas for the species under study, taking into account the predicted suitability for competitor or pathogenic species.
3. Calculate the difference between current and future suitability to identify populations most affected by environmental changes. An exposure index could be assigned to each population as a function of the amount of decrease in suitability values.
4. Use the suitability values and environmental data to develop resistance maps and landscape graphs assessing the functional connectivity for the species.
5. Compare ENMs and connectivity maps to identify the potential of highly-exposed populations to disperse to more favourable areas. An index of dispersal opportunity could be assigned to each population as a function of the number of dispersal paths leading to more favourable areas.
6. Use genetic simulations to model gene-flow and estimate the evolution of genetic diversity as a function of landscape connectivity.
7. Compute genetic diversity of populations based on current empirical data.
8. Compare current simulated and empirical results of genetic diversity to validate the simulations.
9. Identify locally adapted genetic variants and compute the corresponding SPAGs.
10. Combine SPAGs, genetic diversity and connectivity maps to identify the potential of populations to adapt *in-situ*. An index of adaptive capacity could be derived for each population at this stage. The highest adaptive potential should be assigned to populations with high genetic diversity and favourable locally adapted genetic variants. The lowest adaptive potential will define populations with low genetic diversity, a lack of locally adapted variants favourable to future conditions and an absence of connectivity with better adapted populations.
11. Combine previous results to identify vulnerable populations facing simultaneously:
    a. High exposure (e.g. based on the exposure index defined in step 3).
    b. Low dispersal opportunity (e.g. based on the index of dispersal opportunity defined in step 5).
    c. Low adaptive capacity (e.g. based on the index of adaptive capacity defined in step 10).

The integral application of such a framework to a case study has not been done in this thesis and remains to be achieved.

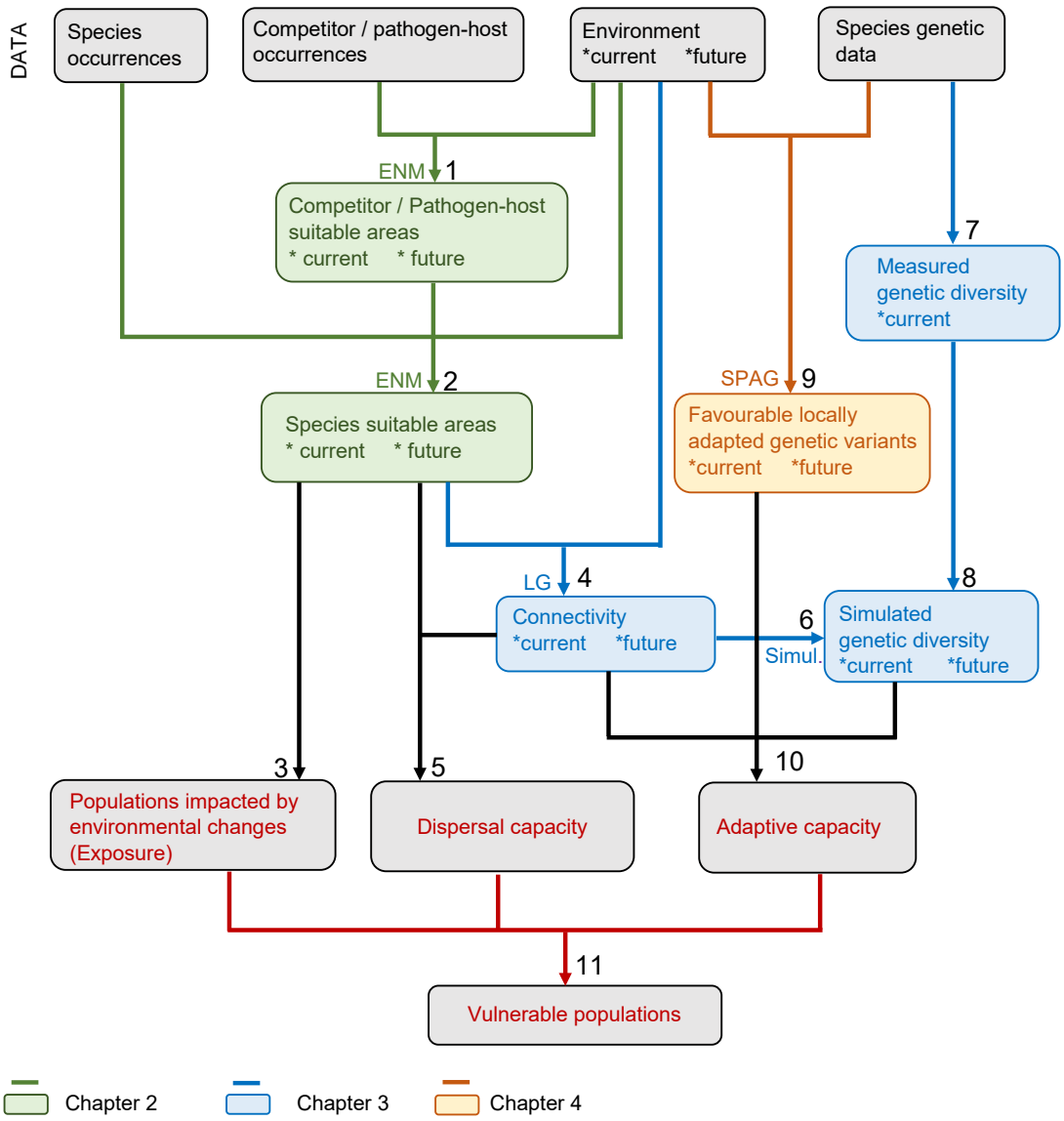


Figure 5-1 – Modelling framework
Numbers 1 to 11 correspond to the steps presented in section 5.1.4 in the main text. Colours indicate the chapter of the thesis in which the modelling tools used are presented in detail.

5.2 Applications in conservation

Once vulnerable populations have been identified, conservation measures should be defined to preserve them. The results of the modelling steps presented in Figure 5-1 provide indications on the main reasons that limit the capacity of species to cope with environmental changes. This provides useful insights for identifying conservation measures to be planned.

First, species facing high exposure may be vulnerable due to low dispersal opportunities. If ENMs results indicate the presence of favourable areas at a distance that can be reach by the species, conservation measures can focus on improving landscape connectivity to allow the migration of individuals to these more favourable areas. In this context, landscape graphs can be used to

identify barriers to dispersal, where it might be valuable to plan for the creation of dispersal corridors or the restoration of stepping stones habitats.

When possibilities to facilitate dispersal are limited, e.g. due to the lack of more favourable areas at a distance that can be reached by the species, conservation practices could focus on improving the adaptive potential of threatened populations. If genetic diversity is limited, conservation could focus first on favouring gene flow. Again, this can be done by increasing landscape connectivity. Where increasing landscape connectivity is not sufficient or not possible to favour genetic exchanges, conservation measures may include translocations or other assisted genetic rescues. In this context, genetic simulations combined with empirical data can be used to identify donor populations that are expected to present higher genetic diversity.

Populations may also be threatened due to a lack of adaptive variant favourable to future conditions. If better adapted populations are identified, conservation measures could focus on improving gene-flow between the vulnerable populations and the well-adapted, to favour the spread of favourable genetic variants. Again, this could be done by increasing landscape connectivity or through assisted genetic rescue. In this context, SPAGs can be used to identify suitable donor populations.

Finally, depending on the conservation objectives, financial and time constraints, conservation measures could also focus on preserving populations that currently show a limited vulnerability or a good potential of adaptation to future environmental conditions and thus a greater chance of long-term viability. In this case, measures should focus on maintaining the existing favourable conditions, e.g. by avoiding a reduction in landscape connectivity or destruction of habitats.

Chapter 6 CONCLUSION AND PERSPECTIVES

6.1 Answer to research questions

In this thesis, we presented modelling tools based on geo-environmental data to identify vulnerable populations threatened by environmental changes. In the following paragraphs, we summarize the main contributions of this thesis regarding the research questions we addressed in the introduction.

> *"How can we build ecological niche models that integrate the spatio-temporal variability of the environmental predictors, and does this lead to better predictive performance?"*

We presented a procedure using moving windows in the *R* programming environment to extract environmental variables in buffered areas around a sampling point and for various time periods preceding sampling date. Our results indicate that for the species studied (*Ixodes ricinus* ticks and *Chlamydiales* pathogens), model performance depends on the spatial area and time period considered. In particular, we demonstrated that models considering buffered areas around the sampling point are more powerful than models extracting environmental variables for only the sampling point. The choice of buffer size should be made in accordance with the species ecology. For *I. ricinus*, we identified that the most powerful buffer radius corresponded to the dispersal area of the tick hosts. Similarly, we showed that the performance of ENMs also depends on the period considered before the sampling date for extracting climatic variables. Again, the choice of this period should be made in accordance with the species ecology. For *I. ricinus*, the best performing period was thus obtained by considering the climatic conditions of the two or three years prior to sampling, which correspond to the tick's life cycle duration. Finally, some species can be influenced by environmental variables acting at different scales. We thus presented a procedure for combining environmental variables at different spatial and temporal scales. This approach has been identified as the most powerful for *Chlamydiales* bacteria. These results demonstrated that considering the spatio-temporal variability of environmental predictors is essential for building more powerful ENMs. When no information is available for estimating the spatial and temporal scale to consider, several values should be tested to retain the most powerful.

"Can we use common ecological niche models to estimate the nested niche of a pathogen within the niche of its host?"

We showed how Maxent models can be used to derive host-pathogens distributions using a two-step procedure. Based on the theory of conditional probabilities, we first used ENM to compute suitability values for the host (in our case, the tick *Ixodes ricinus*). Then, we computed ENM for the pathogen, by including a multiplication by the suitability values obtained for its host. This procedure enables us to automatically limit the suitability values for the pathogen in areas unsuitable for its host. Using this approach, we presented a first study of the environmental predictors affecting the presence of *Chlamydiales* bacteria in their tick host in Switzerland, and we pictured the evolution of *Chlamydiales* distribution from 2009 to 2019. Such nested-niche models could be projected onto future environmental conditions and used to identify populations that may face an increased presence of pathogen in their ecological niche in the near future.

"How can we use modelisation tools using geo-environmental data to complement empirical data and help identify populations threatened by reduced dispersal opportunities and a loss of genetic diversity?"

We presented a combination of landscape graphs and genetic simulations to model connectivity and estimate the evolution of genetic diversity and population persistence in a fragmented landscape. We validated our results with empirical data collected in the same study area. This methodology enabled us to identify populations showing low dispersal possibilities, reduced genetic diversity and limited persistence. The simulation tools used only require geo-environmental data and the definition of some parameters regarding the species' ecology. Simulated genetic data can thus be computed at low costs and for several species. This can enable a first estimation of populations' or species' vulnerability in the face of environmental change, and can be used to highlight populations for which a more comprehensive vulnerability assessment should be conducted.

"How can we use signatures of local adaptation to identify populations threatened by climate change?"

Based on logistic regressions and conditional probabilities, we have developed the new SPatial Areas of Genotypes Probability (SPAG) tool to map the frequency of locally adapted genetic variants. The projection of our models under the expected future climate enables us to identify populations lacking in adapted variants favourable to future climatic conditions. We have presented a univariate and three multivariate models that allow the consideration of several adapted variants associated with various environmental conditions. We hope that this new tool will facilitate a better consideration of the adaptive potential in conservation framework.

"How can the modelling tools presented be combined and implemented in a framework dedicated to the identification of vulnerable populations?"

We presented a conservation framework that integrates the various modelling tools to identify populations threatened by environmental changes. Currently, only a few studies consider the exposure, dispersal ability and adaptive capacity together. The modelling tools presented and the associated conservation framework can thus enable a better understanding of the potential of species to response to environmental changes. Such an understanding is essential for planning conservation measures that are better targeted to vulnerable populations and that take into account the reasons why populations are unable to cope with environmental changes.

6.2 Relevance of modelling for conservation

The modelling tools presented in this thesis show several advantages for conservation measures. First, they enable the analyses of past and future changes. The ENMs have made it possible to visualise the evolution of the suitability of the Swiss territory for ticks and *Chlamydiales* over the last decade. The results showed a clear increase of suitability for ticks over the entire country of Switzerland, including an expansion towards higher altitudes. Such nationwide evolutions are very difficult to capture using measurements alone. Modelling is thus of particular interest for understanding changes on a large spatial scale. Second, the models could be projected onto future climatic conditions to estimate the future suitability of a territory. Similarly, landscape graphs can be used to project the influence of environmental modifications on dispersal possibilities, gene flow and genetic diversity. Last, SPAGs can be used to identify the mismatch between the current presence of locally adapted genetic variants and the future needs of populations. In these contexts, modelling is essential for estimating future impacts and anticipating conservation strategies, which is crucial to ensure that measures are implemented early enough to be effective.

The methodologies presented in this thesis involve several disciplines, including GIS, informatics, statistics and genetics (approach referred to as biogeoinformatics (Duruz, 2020)), and require the processing of very large datasets (thousands of high-resolution raster layers and several millions of SNP genetic data). This could be considered as a limitation to the application of the tools presented by conservation actors. Nevertheless, all the main tools developed in this thesis were computed in the *R* programming environment that is utilised by many biologists and scientists. *R* has proven to be powerful for the efficient and fast processing of our very large datasets, including raster data, without the need for storage in an external database. The new tools developed, notably for extracting the spatio-temporal variability of the environmental predictors in ENMs and computing SPAGs (univariate and multivariate) are provided as *R* functions made available with the published papers. This should thus facilitate their use by several scientists familiar with R. It should also favour the consideration of geo-information and the processing of high resolution geo-environmental data by scientists from different fields, without much experience in GIS or computer science.

The modelling tools presented require 1) environmental data, 2) data on the occurrence of the species and eventually of some predators/competitors/pathogens, and 3) genetic data. High-resolution environmental data are becoming ever more easily accessible, and often provided free of charge, as a result of advances in remote sensing. However, the collection of occurrences and genetic data can remain time-consuming and costly. In this context, integrating citizen participation may be an effective way to collect occurrence data rapidly and at a reasonable cost (McKinley

et al., 2016). In this thesis, we used tick occurrence data collected via a participatory smartphone application. Although special attention must be paid to potential sampling bias, the collected data proved to be convenient for ENMs. In addition, occurrence data collected through citizen participation could allow for directly targeting areas for genetic data sampling, which can reduce the cost associated with sampling time.

Finally, conservation actions involve not only scientists, but also conservation practitioners who may not have knowledge in genetics, modelling or GIS, nor be familiar with *R* programming. There is thus a need to provide results formatted such as to enable effective discussions among all conservation stakeholders, including non-scientist groups or experts from other disciplines (Bickford *et al.*, 2012). The results of the modelling tools presented in this thesis can all be summarised on raster or vector maps that allow easy visual identification of the degree of threat to populations or species and direct geo-localisation of the main problems (barriers to dispersal and gene-flow, fragmentation level, isolated populations, etc.). These results provided on maps should thus facilitate discussions between different actors without a genetic or modelling background and favour their practical implementation in conservation strategies.

6.3 Perspectives

The modelling tools presented in this thesis considered exposure, dispersal opportunity and genetic adaptive capacity, but do not consider phenotypic plasticity. Even if this adaptation does not rely on genetic variations and thus does not necessarily ensure the persistence of the adaptation in future generations, it can help populations to face rapid environmental changes. This plastic adaptive capacity should thus be implemented into new modelling tools and integrated into the conservation framework presented. The degree of threat to species will also depend on the velocity of climate or environmental changes, i.e. the speed at which the changes affect the landscape. This is another parameter that has not been considered in this thesis, but which can strongly influence the rate at which species will need to move or adapt (Catullo *et al.*, 2015). This velocity should thus been taken into account, particularly when working with climate change scenarios (Brito-Morales *et al.*, 2018; Kosanic *et al.*, 2019).

Each modelling step adds a level of uncertainty to the resulting predictions. These uncertainties arise from the environmental data used, the scenarios considered for future predictions, and the modelling tool itself. Methodologies to account for these uncertainties and to efficiently integrate them in modelling still need to be developed (Brown *et al.*, 2016). This is an essential step to identify the uncertainty associated with the final predictions and the sensitivity of the results. An initial way to do this would be to perform sensitivity analyses with a large set of scenarios and modelling tools, such to estimate the robustness of the conclusions (Langford *et al.*, 2009; Kujala *et al.*, 2013; Eaton *et al.*, 2019).

Finally, the framework presented addresses the issue of identifying vulnerable populations within the same species. However, several tools presented could be used to first identify the most vulnerable species. In this context, ENMs derived for several species could be used to identify those facing the highest exposure. Landscape graphs and genetic simulations could be applied for identifying species particularly threatened by a loss of landscape connectivity and, consequently, genetic diversity. In this context, landscape graphs could also be generalized to consider groups of species, for example by using general resistance values defined according to the dispersal mode (terrestrial displacement with different classes of dispersal distances, flying, crawling, etc). This may enable a first identification of groups of species that could be most impacted by the expected

land-use changes. Analysing the level of local adaptation of different species and their corresponding SPAGs may enable to identify the environmental factors that have led to local adaptation for different species. Species that have locally adapted to certain environmental conditions that are expected to change significantly may be the species whose adaptive capacity is most likely to be affected.

To conclude, the conservation framework we presented using surface-based modelling of geo-environmental data could be extended to take into account phenotypic plasticity, velocity of changes and uncertainties. However, this framework does provide tools that should facilitate the identification of populations or species that are particularly vulnerable to environmental changes, considering exposure, dispersal opportunity and adaptive capacity. The development of these modelling tools in the common *R*-programming environment and the presentation of results with maps should facilitate the implementation of our framework for practical conservation discussions.

REFERENCES

Acevedo P, Jiménez-Valverde A, Lobo JM, Real R (2012). Delimiting the geographical background in species distribution modelling. *Journal of Biogeography* **39**: 1383–1390.

Aeschlimann A (1972). Ixodes ricinus, Linné, 1758 (Ixodoidea; Ixodidae). *Acta tropica* **29**: 321–338.

Aeschlimann A (1981). The role of hosts and environment in the natural dissemination of ticks. *Review of advances in parasitology*: 859–869.

Aeschlimann A, Burgdorfer W, Matile H, Péter O, Wyler R (1979). Aspects nouveaux du rôle de vecteur joué par Ixodes ricinus L. en Suisse. Note préliminaire. *Acta Tropica* **36**: 181–191.

Aeschlimann A, Chamot E, Gigon F, Jeanneret J-P, Kesseler D, Walther C (1986). B. burgdorferi in Switzerland. *International journal of microbiology and hygiene (Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene) A: Medical microbiology, infectiousdiseases, parasitology* **263**: 450–458.

Aho K, Derryberry D, Peterson T (2014). Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95**: 631–636.

Ai H, Yang B, Li J, Xie X, Chen H, Ren J (2014). Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics* **15**: 834.

Aitken SN, Whitlock MC (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. *Annual Review of Ecology, Evolution, and Systematics* **44**: 367–388.

Aivelo T, Norberg A, Tschirren B (2019). Bacterial microbiota composition of Ixodes ricinus ticks: the role of environmental variation, tick characteristics and microbial interactions. *PeerJ* **7**: e8217.

Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, Villemereuil P de, *et al.* (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat Commun* **9**: 1–9.

Alexander DH, Lange K (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246.

Alkishe AA, Peterson AT, Samy AM (2017). Climate change influences on the potential geographic distribution of the disease vector tick Ixodes ricinus. *PLoS One* **12**.

Allendorf FW, Hohenlohe PA, Luikart G (2010). Genomics and the future of conservation genetics. *Nat Rev Genet* **11**: 697–709.

Allendorf FW, Leary RF (1986). Heterozygosity and fitness in natural populations of animals. *Conservation biology: the science of scarcity and diversity*: 57–76.

Antrop M (2000). Background concepts for integrated landscape analysis. *Agriculture, Ecosystems & Environment* **77**: 17–28.

Araújo MB (2003). The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography* **12**: 5–12.

Araújo MB, Guisan A (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33**: 1677–1688.

Avon C, Bergès L (2013). Outils pour l'analyse de la connectivité des habitats: Test d'outils de diagnostic de la connectivité fonctionnelle potentielle de la trame forestière. *Projet J Diacofor - Convention cadre Irstea - MEDDE DEB*.

Baguette M, Blanchet S, Legrand D, Stevens VM, Turlure C (2013). Individual dispersal, landscape connectivity and ecological networks. *Biol Rev* **88**: 310–326.

Banks NC, Paini DR, Bayliss KL, Hodda M (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecol Lett* **18**: 188–199.

Banta JA, Ehrenreich IM, Gerard S, Chou L, Wilczek A, Schmitt J, *et al.* (2012). Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in Arabidopsis thaliana. *Ecology Letters* **15**: 769–777.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**: 327–338.

Bates D, Mächler M, Bolker B, Walker S (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:14065823*.

Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**: 83–86.

Bay RA, Rose N, Barrett R, Bernatchez L, Ghalambor CK, Lasky JR, *et al.* (2017). Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. *The American Naturalist* **189**: 463–473.

Beaumont MA, Nichols RA (1996). Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society of London B: Biological Sciences* **263**: 1619–1626.

Beever EA, O'Leary J, Mengelt C, West JM, Julius S, Green N, *et al.* (2016). Improving Conservation Outcomes with a New Paradigm for Understanding Species' Fundamental and Realized Adaptive Capacity. *Conservation Letters* **9**: 131–137.

Benjamini Y, Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**: 289–300.

Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, *et al.* (2015). Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Front Genet* **6**.

Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, *et al.* (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution* **50**: 57.

Bickford D, Posa MRC, Qie L, Campos-Arceiz A, Kudavidanage EP (2012). Science communication for biodiversity conservation. *Biological Conservation* **151**: 74–76.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, *et al.* (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet* **49**: 643–650.

Bierwagen BG (2006). Connectivity in urbanizing landscapes: The importance of habitat configuration, urban area size, and dispersal. *Urban Ecosyst* **10**: 29–42.

Bissoondath CJ, Wiklund C (1996). Effect of Male Mating History and Body Size on Ejaculate Size and Quality in Two Polyandrous Butterflies, Pieris napi and Pieris rapae (Lepidoptera: Pieridae). *Functional Ecology* **10**: 457–464.

Böhner J, Antonić O (2009). Chapter 8 Land-Surface Parameters Specific to Topo-Climatology. In: Hengl T, Reuter HI (eds) *Developments in Soil Science*, Geomorphometry. Elsevier Vol 33, pp 195–226.

Bolund P, Hunhammar S (1999). Ecosystem services in urban areas. *Ecological Economics* **29**: 293–301.

Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007). Population Adaptive Index: a New Method to Help Measure Intraspecific Genetic Diversity and Prioritize Populations for Conservation. *Conservation Biology* **21**: 697–708.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, *et al.* (2012). Costs of dispersal. *Biological Reviews* **87**: 290–312.

Borel N, Polkinghorne A, Pospischil A (2018). A Review on Chlamydial Diseases in Animals: Still a Challenge for Pathologists? *Vet Pathol* **55**: 374–390.

Borel N, Ruhl S, Casson N, Kaiser C, Pospischil A, Greub G (2007). Parachlamydia spp. and Related Chlamydia-like Organisms and Bovine Abortion. *Emerg Infect Dis* **13**: 1904–1907.

Boujenane I (2005). *Small Ruminant Breeds of Morocco*.

Bozzuto C, Biebach I, Muff S, Ives AR, Keller LF (2019). Inbreeding reduces long-term growth of Alpine ibex populations. *Nat Ecol Evol* **3**: 1359–1364.

Braaker S, Kormann U, Bontadina F, Obrist MK (2017). Prediction of genetic connectivity in urban ecosystems by combining detailed movement data, genetic data and multi-path modelling. *Landscape and Urban Planning* **160**: 107–114.

Bradley BA, Wilcove DS, Oppenheimer M (2010). Climate change increases risk of plant invasion in the Eastern United States. *Biol Invasions* **12**: 1855–1872.

Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, *et al.* (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* **18**: 229.

Brito-Morales I, Molinos JG, Schoeman DS, Burrows MT, Poloczanska ES, Brown CJ, *et al.* (2018). Climate Velocity Can Inform Conservation in a Warming World. *Trends in Ecology & Evolution* **33**: 441–457.

Britt M, Haworth SE, Johnson JB, Martchenko D, Shafer ABA (2018). The importance of non-academic coauthors in bridging the conservation genetics gap. *Biological Conservation* **218**: 118–123.

Brooks TM, Mittermeier RA, Fonseca GAB da, Gerlach J, Hoffmann M, Lamoreux JF, *et al.* (2006). Global Biodiversity Conservation Priorities. *Science* **313**: 58–61.

Brooks TM, Mittermeier RA, Mittermeier CG, Fonseca GABD, Rylands AB, Konstant WR, *et al.* (2002). Habitat Loss and Extinction in the Hotspots of Biodiversity. *Conservation Biology* **16**: 909–923.

Brown JL (2014). SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* **5**: 694–700.

Brown JL, Weber JJ, Alvarado-Serrano DF, Hickerson MJ, Franks SJ, Carnaval AC (2016). Predicting the genetic consequences of future climate change: The power of coupling spatial demography, the coalescent, and historical landscape changes. *American Journal of Botany* **103**: 153–163.

Brownstein John S, Holford Theodore R, Fish Durland (2003). A climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. *Environmental Health Perspectives* **111**: 1152–1157.

Büchi L, Vuilleumier S (2014). Coexistence of Specialist and Generalist Species Is Shaped by Dispersal and Environmental Factors. *The American Naturalist* **183**: 612–624.

Burkitt Wright EMM, Spencer HL, Daly SB, Manson FDC, Zeef LAH, Urquhart J, *et al.* (2011). Mutations in PRDM5 in Brittle Cornea Syndrome Identify a Pathway Regulating Extracellular Matrix Development and Maintenance. *The American Journal of Human Genetics* **88**: 767–777.

Calabrese JM, Fagan WF (2004). A comparison-shopper's guide to connectivity metrics. *Frontiers in Ecology and the Environment* **2**: 529–536.

Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, *et al.* (2014). Genome-Wide Analysis of Cold Adaptation in Indigenous Siberian Populations. *PLOS ONE* **9**: e98076.

Carpenter G, Gillison AN, Winter J (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers Conserv* **2**: 667–680.

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, *et al.* (2011). Metagenomic Profile of the Bacterial Communities Associated with Ixodes ricinus Ticks. *PLoS One* **6**.

Catullo RA, Ferrier S, Hoffmann AA (2015). Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution. *Global Ecology and Biogeography* **24**: 1192–1202.

Cederlund G, Liberg O (1995). The roe deer. Wildlife, ecology and hunting. *Swedish (R\aadjuret Viltet, ekologin och jakten) Almqvist and Wiksell, Uppsala, Sweden*.

Charmantier A, McCleery RH, Cole LR, Perrins C, Kruuk LEB, Sheldon BC (2008). Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science* **320**: 800–803.

Chefaoui RM, Lobo JM (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* **210**: 478–486.

Chen W, Kelly MA, Opitz-Araya X, Thomas RE, Low MJ, Cone RD (1997). Exocrine Gland Dysfunction in MC5-R-Deficient Mice: Evidence for Coordinated Regulation of Exocrine Gland Function by Melanocortin Peptides. *Cell* **91**: 789–798.

Chen H, Patterson N, Reich D (2010). Population differentiation as a test for selective sweeps. *Genome Res* **20**: 393–402.

Chentouf M (2014). Systèmes de production caprine au Nord du Maroc: Contraintes et propositions d'amélioration. *Opt Médit A: Médit Semin* **108**: 25–32.

Cheptou P-O, Hargreaves AL, Bonte D, Jacquemyn H (2017). Adaptation to fragmentation: evolutionary dynamics driven by human influences. *Phil Trans R Soc B* **372**: 20160037.

Clauzel C, Girardet X, Foltête J-C (2013). Impact assessment of a high-speed railway line on species distribution: Application to the European tree frog (Hyla arborea) in Franche-Comté. *Journal of Environmental Management* **127**: 125–134.

Clay K, Klyachko O, Grindle N, Civitello D, Oleske D, Fuqua C (2008). Microbial communities and interactions in the lone star tick, Amblyomma americanum. *Molecular Ecology* **17**: 4371–4381.

Collins JP, Storfer A (2003). Global amphibian declines: sorting the hypotheses. *Diversity and Distributions* **9**: 89–98.

Connor T, Hull V, Viña A, Shortridge A, Tang Y, Zhang J, *et al.* (2018). Effects of grain size and niche breadth on species distribution modeling. *Ecography* **41**: 1270–1282.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, *et al.* (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci Model Dev* **8**: 1991–2007.

Coop G, Witonsky D, Rienzo AD, Pritchard JK (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**: 1411–1423.

Corsaro D, Feroldi V, Saucedo G, Ribas F, Loret J-F, Greub G (2009). Novel Chlamydiales strains isolated from a water treatment plant. *Environmental Microbiology* **11**: 188–200.

Corsaro D, Greub G (2006). Pathogenic Potential of Novel Chlamydiae and Diagnostic Approaches to Infections Due to These Obligate Intracellular Bacteria. *Clinical Microbiology Reviews* **19**: 283–297.

Corsaro D, Pages GS, Catalan V, Loret J-F, Greub G (2010). Biodiversity of amoebae and amoeba-associated bacteria in water treatment plants. *International Journal of Hygiene and Environmental Health* **213**: 158–166.

Corsaro D, Thomas V, Goy G, Venditti D, Radek R, Greub G (2007). 'Candidatus Rhabdochlamydia crassificans', an intracellular bacterial pathogen of the cockroach Blatta orientalis (Insecta: Blattodea). *Systematic and Applied Microbiology* **30**: 221–228.

Coulon A, Guillot G, Cosson J-F, Angibault JMA, Aulagnier S, Cargnelutti B, *et al.* (2006). Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology* **15**: 1669–1679.

Cousyn C, Meester LD, Colbourne JK, Brendonck L, Verschuren D, Volckaert F (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *PNAS* **98**: 6256–6260.

Croxatto A, Rieille N, Kernif T, Bitam I, Aeby S, Péter O, *et al.* (2014). Presence of Chlamydiales DNA in ticks and fleas suggests that ticks are carriers of Chlamydiae. *Ticks and Tick-borne Diseases* **5**: 359–365.

Culley TM, Sbita SJ, Wick A (2007). Population Genetic Effects of Urban Habitat Fragmentation in the Perennial Herb Viola pubescens (Violaceae) using ISSR Markers. *Ann Bot* **100**: 91–100.

Cumming G s. (1998). Host preference in African ticks (Acari: Ixodida): a quantitative data set. *Bulletin of Entomological Research* **88**: 379–406.

Cushman SA (2006). Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation* **128**: 231–240.

Cushman SA, Lewis JS (2010). Movement behavior explains genetic differentiation in American black bears. *Landscape Ecol* **25**: 1613–1625.

Czech B, Krausman PR, Devers PK (2000). Economic Associations among Causes of Species Endangerment in the United States. *BioScience* **50**: 593–601.

Daniel M, Danielová V, Kříž B, Jirsa A, Nožička J (2003). Shift of the Tick Ixodes ricinus and Tick-Borne Encephalitis to Higher Altitudes in Central Europe. *Eur J Clin Microbiol Infect Dis* **22**: 327–328.

Darwin C (1859). *On the Origin of Species by Means of Natural Selection Or the Preservation of Favoured Races in the Struggle for Life.* H. Milford; Oxford University Press.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011). Beyond Predictions: Biodiversity Conservation in a Changing Climate. *Science* **332**: 53–58.

Dennis EB, Morgan BJT, Roy DB, Brereton TM (2017). Urban indicators for UK butterflies. *Ecological Indicators* **76**: 184–193.

Deuchande R, Gidlow J, Caldow G, Baily J, Longbottom D, Wheelhouse N, *et al.* (2010). Parachlamydia involvement in bovine abortions in a beef herd in Scotland. *Veterinary Record* **166**: 598–599.

Di Giulio M, Holderegger R, Tobias S (2009). Effects of habitat and landscape fragmentation on humans and biodiversity in densely populated landscapes. *Journal of Environmental Management* **90**: 2959–2968.

Diekmann OE, Gouveia L, Perez JA, Gil-Rodriguez C, Serrão EA (2010). The possible origin of Zostera noltii in the Canary Islands and guidelines for restoration. *Mar Biol* **157**: 2109–2115.

Dirzo R, Raven PH (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources* **28**: 137–167.

Domsa C, Mihalca AD, Sandor AD (2018). Modeling the distribution of Ixodes ricinus in Romania. *North-Western Journal of Zoology* **14**.

Donald PF, Sanderson FJ, Burfield IJ, van Bommel FPJ (2006). Further evidence of continent-wide impacts of agricultural intensification on European farmland birds, 1990–2000. *Agriculture, Ecosystems & Environment* **116**: 189–196.

Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, *et al.* (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology* **31**: 135–141.

Dubois J, Cheptou P-O (2017). Effects of fragmentation on plant adaptation to urban environments. *Phil Trans R Soc B* **372**: 20160038.

Duplouy A, Ikonen S, Hanski I (2013). Life history of the Glanville fritillary butterfly in fragmented versus continuous landscapes. *Ecol Evol* **3**: 5141–5156.

Duputié A, Rutschmann A, Ronce O, Chuine I (2015). Phenological plasticity will not help all species adapt to climate change. *Global Change Biology* **21**: 3062–3073.

Duruz S (2020). Biogeoinformatics for the management of Farm Animal Genetic Resources (FAnGR).

Eaton MJ, Yurek S, Haider Z, Martin J, Johnson FA, Udell BJ, *et al.* (2019). Spatial conservation planning under uncertainty: adapting to climate change risks using modern portfolio theory. *Ecological Applications* **29**: e01962.

EEA EEA (2016). Urban sprawl in Europe - joint EEA-FOEN report.

Egizi A, Kiser J, Abadam C, Fonseca DM (2016). The hitchhiker's guide to becoming invasive: exotic mosquitoes spread across a US state by human transport not autonomous flight. *Mol Ecol* **25**: 3033–3047.

Ehrmann S, Ruyts SC, Scherer-Lorenzen M, Bauhus J, Brunet J, Cousins SAO, *et al.* (2018). Habitat properties are key drivers of Borrelia burgdorferi (s.l.) prevalence in Ixodes ricinus populations of deciduous forest fragments. *Parasites & Vectors* **11**: 23.

Eisen RJ, Feirer S, Padgett KA, Hahn MB, Monaghan AJ, Kramer VL, *et al.* (2018). Modeling Climate Suitability of the Western Blacklegged Tick in California. *J Med Entomol* **55**: 1133–1142.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129–151.

Elith J, Leathwick JR (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* **40**: 677–697.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2010). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*: 43–57.

Epperson BK, Mcrae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin M-J, *et al.* (2010). Utility of computer simulations in landscape genetics. *Molecular Ecology* **19**: 3549–3564.

Eriksson BK, Hillebrand H (2019). Rapid reorganization of global biodiversity. *Science* **366**: 308–309.

Estrada-peña A (1999). Geostatistics as Predictive Tools to Estimate Ixodes ricinus (Acari: Ixodidae) Habitat Suitability in the Western Palearctic From AVHRR Satellite Imagery. *Exp Appl Acarol* **23**: 337–349.

Estrada-Peña A (2008). Climate, niche, ticks, and models: what they are and how we should interpret them. *Parasitol Res* **103**: 87–95.

Estrada-Peña A, Estrada-Sánchez A, Estrada-Sánchez D (2015). Methodological caveats in the environmental modelling and projections of climate niche for ticks, with examples for Ixodes ricinus (Ixodidae). *Veterinary Parasitology* **208**: 14–25.

Evans KL, Gaston KJ, Frantz AC, Simeoni M, Sharp SP, McGowan A, *et al.* (2009). Independent colonization of multiple urban centres by a formerly forest specialist bird species. *Proceedings of the Royal Society of London B: Biological Sciences*: rspb.2008.1712.

Excoffier L, Hofer T, Foll M (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285–298.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–567.

Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nat Ecol Evol* **2**: 352–358.

Fahrig L (2003a). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **34**: 487–515.

Fahrig L (2003b). Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **34**: 487–515.

Falk W, Mellert KH (2011). Species distribution models as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. *Journal of Vegetation Science* **22**: 621–634.

Farashi A, Alizadeh-Noughani M (2018). Effects of models and spatial resolutions on the species distribution model performance. *Model Earth Syst Environ* **4**: 263–268.

Fattorini S (2011). Insect extinction by urbanization: A long term study in Rome. *Biological Conservation* **144**: 370–375.

Fazey I, Fischer J, Lindenmayer DB (2005). What do conservation biologists publish? *Biological Conservation* **124**: 63–73.

Ferson S, Burgman M (2006). *Quantitative Methods for Conservation Biology*. Springer Science & Business Media.

Fielding AH, Bell JF (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**: 38–49.

Fisher RA (1930). *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. OUP Oxford.

Fitzpatrick MC, Keller SR (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* **18**: 1–16.

Foll M, Gaggiotti O (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**: 977–993.

Foltête J-C, Clauzel C, Girardet X, Tournant P, Vuidel G (2012). La modélisation des réseaux écologiques par les graphes paysagers. *Revue Internationale de Géomatique* **22**: 641–658.

Foltête J-C, Clauzel C, Vuidel G (2012). A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software* **38**: 316–327.

Forester BR, Lasky JR, Wagner HH, Urban DL (2017). Using genotype-environment associations to identify multilocus local adaptation. *bioRxiv*: 129460.

Fountain T, Nieminen M, Sirén J, Wong SC, Lehtonen R, Hanski I (2016). Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *PNAS* **113**: 2678–2683.

Fourcade Y, Besnard AG, Secondi J (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography* **27**: 245–256.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011). A Map of Local Adaptation in Arabidopsis thaliana. *Science* **334**: 86–89.

Fox RJ, Donelson JM, Schunter C, Ravasi T, Gaitán-Espitia JD (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**: 20180174.

Frankham R (2005). Genetics and extinction. *Biological Conservation* **126**: 131–140.

Frankham R (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation* **143**: 1919–1927.

Fu W, Liu S, Degloria SD, Dong S, Beazley R (2010). Characterizing the "fragmentation–barrier" effect of road networks on landscape connectivity: A case study in Xishuangbanna, Southwest China. *Landscape and Urban Planning* **95**: 122–129.

Funk WC, Forester BR, Converse SJ, Darst C, Morey S (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conserv Genet* **20**: 115–134.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**: 489–496.

Galbraith H (2010). Fundamental hair follicle biology and fine fibre production in animals. *animal* **4**: 1490–1509.

Galpern P, Manseau M, Fall A (2011). Patch-based graphs of landscape connectivity: A guide to construction, analysis and application for conservation. *Biological Conservation* **144**: 44–55.

Garzón MB, Robson TM, Hampe A (2019). ΔTraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. *New Phytologist* **222**: 1757–1765.

Gäumann R, Mühlemann K, Strasser M, Beuret CM (2010). High-Throughput Procedure for Tick Surveys of Tick-Borne Encephalitis Virus and Its Application in a National Surveillance Study in Switzerland. *Appl Environ Microbiol* **76**: 4241–4249.

Gautier M (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* **201**: 1555–1579.

Ge W, Wang S-H, Sun B, Zhang Y-L, Shen W, Khatib H, *et al.* (2018). Melatonin promotes Cashmere goat (Capra hircus) secondary hair follicle growth: a view from integrated analysis of long non-coding and coding RNAs. *Cell Cycle* **17**: 1255–1267.

Gern L, Morán Cadenas F, Burri C (2008). Influence of some climatic factors on Ixodes ricinus ticks studied along altitudinal gradients in two geographic regions in Switzerland. *International Journal of Medical Microbiology* **298**: 55–59.

Girardet X, Clauzel C, Li L, Foltête J-C (2016). Graphab: A software dedicated to the modelling of landscape networks

Girardet X, Foltête J-C, Clauzel C (2013). Designing a graph-based approach to landscape ecological assessment of linear infrastructures. *Environmental Impact Assessment Review* **42**: 10–17.

Godber OF, Laroussi BF, Chentouf M, Wall R (2016). Intensification of Mediterranean Goat Production Systems: A Case Study in Northern Morocco. *Agriculture* **6**: 16.

Goto M (1997). Hierarchical deterioration of body systems in Werner's syndrome: Implications for normal ageing. *Mechanisms of Ageing and Development* **98**: 239–254.

Gottschalk TK, Aue B, Hotes S, Ekschmitt K (2011). Influence of grain size on species–habitat models. *Ecological Modelling* **222**: 3403–3412.

Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* **45**: 239–247.

Greub G (2009). Parachlamydia acanthamoebae, an emerging agent of pneumonia. *Clinical Microbiology and Infection* **15**: 18–28.

Guisan A, Graham CH, Elith J, Huettmann F, the NCEAS Species Distribution Modelling Group (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*: 332–340.

Guisan A, Zimmermann NE (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147–186.

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007). What Matters for Predicting the Occurrences of Trees: Techniques, Data, or Species' Characteristics? *Ecological Monographs* **77**: 615–630.

Günther T, Coop G (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics* **195**: 205–220.

Hadgu M, Menghistu HT, Girma A, Abrha H, Hagos H (2019). Modeling the potential climate change- induced impacts on future genus Rhipicephalus (Acari: Ixodidae) tick distribution in semi-arid areas of Raya Azebo district, Northern Ethiopia. *j ecology environ* **43**: 43.

Haldane JBS (1959). The Theory of Natural Selection To-Day. *Nature* **183**: 710–713.

Hall LS, Krausman PR, Morrison ML (1997). The Habitat Concept and a Plea for Standard Terminology. *Wildlife Society Bulletin (1973-2006)* **25**: 173–182.

Hall P, Walker S, Bawa K (1996). Effect of Forest Fragmentation on Genetic Diversity and Mating System in a Tropical Tree, Pithecellobium elegans. *Conservation Biology* **10**: 757–768.

Hällfors MH, Liao J, Dzurisin J, Grundel R, Hyvärinen M, Towle K, *et al.* (2016). Addressing potential local adaptation in species distribution models: implications for conservation under climate change. *Ecological Applications* **26**: 1154–1169.

Hallgren W, Santana F, Low-Choy S, Zhao Y, Mackey B (2019). Species distribution models can be highly sensitive to algorithm configuration. *Ecological Modelling* **408**: 108719.

Halos L, Bord S, Cotté V, Gasqui P, Abrial D, Barnouin J, *et al.* (2010). Ecological Factors Characterizing the Prevalence of Bacterial Tick-Borne Pathogens in Ixodes ricinus Ticks in Pastures and Woodlands. *Appl Environ Microbiol* **76**: 4413–4420.

Hanberry BB (2013). Finer grain size increases effects of error and changes influence of environmental predictors on species distribution models. *Ecological Informatics* **15**: 8–13.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, *et al.* (2011). Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science* **334**: 83–86.

Harris SE, Munshi-South J (2016). Scans for positive selection reveal candidate genes and local adaptation of Peromyscus leucopus populations to urbanization. *bioRxiv*: 038141.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol Appl* **7**: 1008–1025.

Hauser G, Rais O, Morán Cadenas F, Gonseth Y, Bouzelboudjen M, Gern L (2018). Influence of climatic factors on Ixodes ricinus nymph abundance and phenology over a long-term monthly observation in Switzerland (2000–2014). *Parasites & Vectors* **11**: 289.

Hayashi R, Yamato M, Sugiyama H, Sumide T, Yang J, Okano T, *et al.* (2007). N-Cadherin Is Expressed by Putative Stem/Progenitor Cells and Melanocytes in the Human Limbal Epithelial Stem Cell Niche. *STEM CELLS* **25**: 289–296.

Hernandez PA, Graham CH, Master LL, Albert DL (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**: 773–785.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

Hijmans RJ, van Etten J (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12. http://CRAN.R-project.org/package=raster.

Hirzel AH, Hausser J, Chessel D, Perrin N (2002). Ecological-Niche Factor Analysis: How to Compute Habitat-Suitability Maps Without Absence Data? *Ecology* **83**: 2027–2036.

Hitchings SP, Beebee TJ (1997). Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity* **79**: 117–127.

Hoban S, Arntzen JW, Bertorelle G, Bryja J, Fernandes M, Frith K, *et al.* (2013). Conservation Genetic Resources for Effective Species Survival (ConGRESS): Bridging the divide between conservation research and practice. *Journal for Nature Conservation* **21**: 433–437.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, *et al.* (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist* **188**: 379–397.

Hoelzel AR, Bruford MW, Fleischer RC (2019). Conservation of adaptive potential and functional diversity. *Conserv Genet* **20**: 1–5.

Hoffman JD, Aguilar-Amuchastegui N, Tyre AJ (2010). Use of simulated data from a process-based habitat model to evaluate methods for predicting species occurrence. *Ecography* **33**: 656–666.

Hoffmann I (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics* **41**: 32–46.

Hoffmann AA, Sgrò CM (2011). Climate change and evolutionary adaptation. *Nature* **470**: 479–485.

Hoffmann AA, Willi Y (2008). Detecting genetic responses to environmental change. *Nat Rev Genet* **9**: 421–432.

Holderegger R, Balkenhol N, Bolliger J, Engler JO, Gugerli F, Hochkirch A, *et al.* (2019). Conservation genetics: Linking science with practice. *Molecular Ecology* **28**: 3848–3856.

Huerta MAO, Peterson AT (2008). Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Revista Mexicana de Biodiversidad* **79**: 205–216.

Huete A, Justice C, Van Leeuwen W (1999). MODIS vegetation index (MOD13). Algorithm theoretical basis document. **3**.

Hughes L (2000). Biological consequences of global warming: is the signal already apparent? *Trends in Ecology & Evolution* **15**: 56–61.

Huss A, Braun-Fahrländer C (2007). *Tick-borne diseases in Switzerland and climate change*. Institut für Sozial-und Präventivmedizin.

Hutchinson G (1957). Concluding remarks. *Cold Spring Harbor Symp*.

Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology* **23**: 164–176.

Illoldi-Rangel P, Rivaldi C-L, Sissel B, Trout Fryxell R, Gordillo-Pérez G, Rodríguez-Moreno A, *et al.* (2012). Species Distribution Models and Ecological Suitability Analysis for Potential Tick Vectors of Lyme Disease in Mexico. *Journal of Tropical Medicine* **2012**: e959101.

IPBES (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *E S Brondizio, J Settele, S Díaz, and H T Ngo (editors)*.

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Ives CD, Lentini PE, Threlfall CG, Ikin K, Shanahan DF, Garrard GE, *et al.* (2016). Cities are hotspots for threatened species. *Global Ecology and Biogeography* **25**: 117–126.

Jarvis A (2008). Hole-field seamless SRTM data, International Centre for Tropical Agriculture (CIAT). *http://srtm csi cgiar org*.

Jarvis A, Reuter HI, Nelson A, Guevara E (2008). Hole-field seamless SRTM data, International Centre for Tropical Agriculture (CIAT). *http://srtm.csi.cgiar.org*.

Jiménez-Valverde A (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* **21**: 498–507.

Johnson JA, Bellinger MR, Toepfer JE, Dunn P (2004). Temporal changes in allele frequencies and low effective population size in greater prairie-chickens. *Molecular Ecology* **13**: 2617–2630.

Johnson FM, Schaffer HE, Gillaspy JE, Rockwood ES (1969). Isozyme genotype-environment relationships in natural populations of the harvester ant, Pogonomyrmex barbatus, from Texas. *Biochem Genet* **3**: 429–450.

Jombart T (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**: 94.

Jones RE, Gilbert N, Guppy M, Nealis V (1980). Long-Distance Movement of Pieris rapae. *Journal of Animal Ecology* **49**: 629–642.

Joost S (2006). The geographical dimension of genetic diversity : a GIScience contribution for the conservation of animal genetic resources.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, *et al.* (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**: 3955–3969.

Joost S, Colli L, Bonin A, Biebach I, Allendorf FW, Hoffmann I, *et al.* (2011). Promoting collaboration between livestock and wildlife conservation genetics communities. *Conservation Genet Resour* **3**: 785–788.

Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S, *et al.* (2013). Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Mol Ecol* **22**: 3659–3665.

Jore S, Viljugrein H, Hofshagen M, Brun-Hansen H, Kristoffersen AB, Nygård K, *et al.* (2011). Multi-source analysis reveals latitudinal and altitudinal shifts in range of Ixodes ricinus at its northern distribution limit. *Parasites & Vectors* **4**: 84.

Jouda F, Perret J-L, Gern L (2004). Ixodes ricinus Density, and Distribution and Prevalence of Borrelia burgdorferi Sensu Lato Infection Along an Altitudinal Gradient. *J Med Entomol* **41**: 162–169.

Kahl O, Alidousti I (1997). Bodies of liquid water as a source of water gain for Ixodes ricinus ticks (Acari: Ixodidae). *Exp Appl Acarol* **21**: 731–746.

Kaivanto K (2008). Maximization of the sum of sensitivity and specificity as a diagnostic cutpoint criterion. *Journal of clinical epidemiology* **61**: 516–518.

Kantsa A, Tscheulin T, Junker RR, Petanidou T, Kokkini S (2013). Urban biodiversity hotspots wait to get discovered: The example of the city of Ioannina, NW Greece. *Landscape and Urban Planning* **120**: 129–137.

Kao WW, Liu CY, Converse RL, Shiraishi A, Kao CW, Ishizaki M, *et al.* (1996). Keratin 12-deficient mice have fragile corneal epithelia. *Invest Ophthalmol Vis Sci* **37**: 2572–2584.

Kardos M, Shafer ABA (2018). The Peril of Gene-Targeted Conservation. *Trends in Ecology & Evolution* **33**: 827–839.

Karenko L, Hahtola S, Päivinen S, Karhu R, Syrjä S, Kähkönen M, *et al.* (2005). Primary Cutaneous T-Cell Lymphomas Show a Deletion or Translocation Affecting NAV3, the Human UNC-53 Homologue. *Cancer Res* **65**: 8101–8110.

Kearney M (2006). Habitat, environment and niche: what are we modelling? *Oikos* **115**: 186–191.

Kebbi-Beghdadi C, Greub G (2014). Importance of amoebae as a tool to isolate amoeba-resisting microorganisms and for their ecology and evolution: the Chlamydia paradigm. *Environmental Microbiology Reports* **6**: 309–324.

Kelly E, Phillips BL (2016). Targeted gene flow for conservation. *Conservation Biology* **30**: 259–267.

Kendal D, Zeeman BJ, Ikin K, Lunt ID, McDonnell MJ, Farrar A, *et al.* (2017). The importance of small urban reserves for plant conservation. *Biological Conservation* **213**: 146–153.

Keyghobadi N, Unger KP, Weintraub JD, Fonseca DM (2006). Remnant populations of the regal fritillary (Speyeria idalia) in Pennsylvania: Local genetic structure in a high gene flow species. *Conserv Genet* **7**: 309.

Kim Y, Stephan W (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.

Kljuic A, Bazzi H, Sundberg JP, Martinez-Mir A, O'Shaughnessy R, Mahoney MG, *et al.* (2003). Desmoglein 4 in Hair Follicle Differentiation and Epidermal Adhesion: Evidence from Inherited Hypotrichosis and Acquired Pemphigus Vulgaris. *Cell* **113**: 249–260.

Kobayashi K, Hernandez LD, Galán JE, Janeway CA, Medzhitov R, Flavell RA (2002). IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell* **110**: 191–202.

Kolmogorov A (1956). Foundations of the Theory of Probability. *Chelsea Pub Co*.

Kopecký M, Čížková Š (2010). Using topographic wetness index in vegetation ecology: Does the algorithm matter? *Applied Vegetation Science* **13**: 450–459.

Kosanic A, Kavcic I, Kleunen M van, Harrison S (2019). Climate change and climate change velocity analysis across Germany. *Sci Rep* **9**: 1–8.

Kostanjsek R, Strus J, Drobne D, Avgustin G (2004). 'Candidatus Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod Porcellio scaber (Crustacea: Isopoda). *International journal of systematic and evolutionary microbiology* **54**: 543–549.

Kowarik I (2011). Novel urban ecosystems, biodiversity, and conservation. *Environmental Pollution* **159**: 1974–1983.

Kujala H, Moilanen A, Araujo MB, Cabeza M (2013). Conservation planning with uncertain climate change projections. *PloS one* **8**.

Kutzleb C, Sanders G, Yamamoto R, Wang X, Lichte B, Petrasch-Parwez E, *et al.* (1998). Paralemmin, a Prenyl-Palmitoyl–anchored Phosphoprotein Abundant in Neurons and Implicated in Plasma Membrane Dynamics and Cell Process Formation. *The Journal of Cell Biology* **143**: 795–813.

Lafranchis T (2004). *Butterflies of Europe: new field guide and key*. Diatheo.

Lamoth F, Aeby S, Schneider A, Jaton-Ogay K, Vaudaux B, Greub G (2009). Parachlamydia and Rhabdochlamydia in Premature Neonates. *Emerg Infect Dis* **15**: 2072–2075.

Lamoth F, Greub G (2010a). Amoebal pathogens as emerging causal agents of pneumonia. *FEMS Microbiol Rev* **34**: 260–280.

Lamoth F, Greub G (2010b). Fastidious intracellular bacteria as causal agents of community-acquired pneumonia. *Expert Review of Anti-infective Therapy* **8**: 775–790.

Lamoth F, Jaton K, Vaudaux B, Greub G (2011). Parachlamydia and Rhabdochlamydia: Emerging Agents of Community-Acquired Respiratory Infections in Children. *Clin Infect Dis* **53**: 500–501.

Landguth EL, Cushman SamuelA (2010). cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources* **10**: 156–161.

Landguth EL, Fedy BC, OYLER-McCANCE SJ, Garey AL, Emel SL, Mumma M, *et al.* (2012). Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. *Molecular Ecology Resources* **12**: 276–284.

Landguth EL, Hand BK, Glassy J, Cushman SA, Sawaya MA (2012). UNICOR: a species connectivity and corridor network simulator. *Ecography* **35**: 9–14.

Langford WT, Gordon A, Bastin L (2009). When do conservation planning methods deliver? Quantifying the consequences of uncertainty. *Ecological Informatics* **4**: 123–135.

Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, *et al.* (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances* **1**: e1400218.

Lee ACK, Maheswaran R (2010). The health benefits of urban green spaces: a review of the evidence. *J Public Health*: fdq068.

Leempoel K, Duruz S, Rochat E, Widmer I, Orozco-terWengel P, Joost S (2017). Simple Rules for an Efficient Use of Geographic Information Systems in Molecular Ecology. *Front Ecol Evol* **5**.

Li X, Chen C, Wang F, Huang W, Liang Z, Xiao Y, *et al.* (2014). KCTD1 Suppresses Canonical Wnt Signaling Pathway by Enhancing β-catenin Degradation. *PLOS ONE* **9**: e94343.

Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M (2012). Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology* **21**: 28–44.

Lienard J, Croxatto A, Aeby S, Jaton K, Posfay-Barbe K, Gervaix A, *et al.* (2011). Development of a New Chlamydiales-Specific Real-Time PCR and Its Application to Respiratory Clinical Samples. *Journal of Clinical Microbiology* **49**: 2637–2642.

Lindgren E, Tälleklint L, Polfeldt T (2000). Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick Ixodes ricinus. *Environmental Health Perspectives* **108**: 119–123.

Lindgren E, Jaenson TG, Menne B, Organization WH (2006). *Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures*. Copenhagen: WHO Regional Office for Europe.

Ling YH, Xiang, H, Zhang G, Ding JP, Zhang ZJ, Zhang YH, *et al.* (2014). Identification of complete linkage disequilibrium in the DSG4 gene and its association with wool length and crimp in Chinese indigenous sheep. *GMR | Genetics and Molecular Research | The Original by FUNPEC-RP* **13**: 5617–5625.

Liu C, White M, Newell G (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography* **40**: 778–789.

Lizée M-H, Manel S, Mauffrey J-F, Tatoni T, Deschamps-Cottin M (2011). Matrix configuration and patch isolation influences override the species–area relationship for urban butterfly communities. *Landscape Ecol* **27**: 159–169.

Lobo JM, Jiménez-Valverde A, Real R (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**: 145–151.

Lobo JM, Tognelli MF (2011). Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation* **19**: 1–7.

Lourenço A, Álvarez D, Wang IJ, Velo-Antón G (2017). Trapped within the city: integrating demography, time since isolation and population-specific traits to assess the genetic effects of urbanization. *Mol Ecol* **26**: 1498–1514.

Lovasi GS, Quinn JW, Neckerman KM, Perzanowski MS, Rundle A (2008). Children living in areas with more street trees have lower prevalence of asthma. *Journal of Epidemiology & Community Health* **62**: 647–649.

Lu T, Ke M, Lavoie M, Jin Y, Fan X, Zhang Z, *et al.* (2018). Rhizosphere microorganisms can influence the timing of plant flowering. *Microbiome* **6**: 231.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**: 981–994.

Mable BK (2019). Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conserv Genet* **20**: 89–100.

Maes D, Van Dyck H (2001). Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biological Conservation* **99**: 263–276.

Maller C, Townsend M, Pryor A, Brown P, Leger LS (2006). Healthy nature healthy people: 'contact with nature' as an upstream health promotion intervention for populations. *Health Promot Int* **21**: 45–54.

Manel S, Holderegger R (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution* **28**: 614–621.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010). Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. *Molecular Ecology* **19**: 3824–3835.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**: 189–197.

Manel S, Williams HC, Ormerod SJ (2001). Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38**: 921–931.

Manzoor SA, Griffiths G, Lukac M (2018). Species distribution model transferability and model grain size – finer may not always be better. *Sci Rep* **8**: 1–9.

Mawdsley JR, O'Malley R, Ojima DS (2009). A Review of Climate-Change Adaptation Strategies for Wildlife Management and Biodiversity Conservation. *Conservation Biology* **23**: 1080–1089.

Mayer AL, Cameron GN (2003). Consideration of grain and extent in landscape studies of terrestrial vertebrate ecology. *Landscape and Urban Planning* **65**: 201–217.

McCoy KD, Boulanger N (Eds.) (2015). *Tiques et maladies à tiques : biologie, écologie évolutive, épidémiologie*. IRD: Marseille.

McDonald JH, Kreitman M, others (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**: 652–654.

McDonald J, McCormack PC, Dunlop M, Farrier D, Feehely J, Gilfedder L, *et al.* (2019). Adaptation pathways for conservation law and policy. *WIREs Climate Change* **10**: e555.

McKay JK, Bishop JG, Lin J-Z, Richards JH, Sala A, Mitchell-Olds T (2001). Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London Series B: Biological Sciences* **268**: 1715–1721.

McKinley D, Miller-Rushing A, Ballard H, Bonney R, Brown H, Evans D, *et al.* (2016). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*.

McKinney ML (2002). Urbanization, Biodiversity, and Conservation The impacts of urbanization on native species are poorly studied, but educating a highly urbanized human population about these impacts can greatly improve species conservation in all ecosystems. *BioScience* **52**: 883–890.

McRae BH, Shah VB, Mohapatry TK (2013). Circuitscape 4 User Guide. *The Nature Conservancy*.

Mdladla K, Dzomba EF, Muchadeyi FC (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* **120**: 369–378.

Medlock JM, Hansford KM, Bormane A, Derdakova M, Estrada-Peña A, George J-C, *et al.* (2013). Driving forces for changes in geographical distribution of Ixodes ricinus ticks in Europe. *Parasites & Vectors* **6**: 1.

Meentemeyer RK, Anacker BL, Mark W, Rizzo DM (2008). Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications* **18**: 377–390.

Mendel G (1865). Experiments in plant hybridization (1865). *Verhandlungen des naturforschenden Vereins Brünn) Available online*.

Merilä J, Hendry AP (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications* **7**: 1–14.

Mermod C, Aeschlimann A, Graf J-F (1973). Ecologie et éthologie d'Ixodes ricinus Linné 1758, en Suisse (Acarina, Ixodidae). *Acarologia* **15**: 197–205.

Merow C, Smith MJ, Silander JA (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**: 1058–1069.

Mertes K, Jarzyna MA, Jetz W (2020). Hierarchical multi-grain models improve descriptions of species' environmental associations, distribution, and abundance. *Ecological Applications* **n/a**.

Meyer CB (2007). Does Scale Matter in Predicting Species Distributions? Case Study with the Marbled Murrelet. *Ecological Applications* **17**: 1474–1483.

Meyer CB, Thuiller W (2006). Accuracy of resource selection functions across spatial scales. *Diversity and Distributions* **12**: 288–297.

Miller JR, Hobbs RJ (2002). Conservation Where People Live and Work. *Conservation Biology* **16**: 330–337.

Minigan JN, Hager HA, Peregrine AS, Newman JA (2018). Current and potential future distribution of the American dog tick (Dermacentor variabilis, Say) in North America. *Ticks and Tick-borne Diseases* **9**: 354–362.

Montero BK, Refaly E, Ramanamanjato J-B, Randriatafika F, Rakotondranary SJ, Wilhelm K, *et al.* (2019). Challenges of next-generation sequencing in conservation management: Insights from long-term monitoring of corridor effects on the genetic diversity of mouse lemurs in a fragmented landscape. *Evolutionary Applications* **12**: 425–442.

Morales NS, Fernández IC, Baca-González V (2017). MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* **5**.

Munday PL (2004). Habitat loss, resource specialization, and extinction on coral reefs. *Global Change Biology* **10**: 1642–1647.

Munshi-South J, Zolnik CP, Harris SE (2016). Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evol Appl* **9**: 546–564.

Murray FW (1966). *ON THE COMPUTATION OF SATURATION VAPOR PRESSURE.* RAND CORP SANTA MONICA CALIF.

Musolin DL (2007). Insects in a warmer world: ecological, physiological and life-history responses of true bugs (Heteroptera) to climate change. *Global Change Biology* **13**: 1565–1585.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858.

Nachman MW, Hoekstra HE, D'Agostino SL (2003). The genetic basis of adaptive melanism in pocket mice. *PNAS* **100**: 5268–5273.

Newbold T, Hudson LN, Hill SLL, Contu S, Lysenko I, Senior RA, *et al.* (2015). Global effects of land use on local terrestrial biodiversity. *Nature* **520**: 45–50.

Nicotra AB, Beever EA, Robertson AL, Hofmann GE, O'Leary J (2015). Assessing the components of adaptive capacity to improve conservation and management efforts under global change. *Conservation Biology* **29**: 1268–1278.

Nielsen R (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics* **39**: 197–218.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005). Genomic scans for selective sweeps using SNP data. *Genome research* **15**: 1566–1575.

Nieminen M, Singer MC, Fortelius W, Schöps K, Hanski I (2001). Experimental Confirmation that Inbreeding Depression Increases Extinction Risk in Butterfly Populations. *The American Naturalist* **157**: 237–244.

Nix HA, Busby J (1986). BIOCLIM, a bioclimatic analysis and prediction system. *Division of Water and Land Resources: Canberra*.

Norberg A, Abrego N, Blanchet FG, Adler FR, Anderson BJ, Anttila J, *et al.* (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs* **89**: e01370.

Oechslin CP, Heutschi D, Lenz N, Tischhauser W, Péter O, Rais O, *et al.* (2017). Prevalence of tick-borne pathogens in questing Ixodes ricinus ticks in urban and suburban areas of Switzerland. *Parasites & Vectors* **10**: 558.

Oertli B, Joye DA, Castella E, Juge R, Cambin D, Lachavanne J-B (2002). Does size matter? The relationship between pond area and biodiversity. *Biological Conservation* **104**: 59–70.

OFS (2013). Waldmischungsgrad, Auflösung 100m: Geodaten | Publication. *Federal Statistical Office*.

OFS (2017). Statistique de la superficie selon nomenclature 2004 - Occupation du sol, description: métainformations sur les géodonnées | Publication. *Office fédéral de la statistique*.

OFSP (2013). Recommandation de vaccination contre l'encéphalite à tiques : actualisation et nouvelle présentation de la carte à partir d'avril 2013.

OFSP (2019). Méningo-encéphalite à tiques (FSME).

Ohsaki N (1979). Comparative population studies of threePieris butterflies,P. rapae, P. melete andP. napi, living in the same area. *Res Popul Ecol* **20**: 278–296.

Ohsaki N (1980). Comparative population studies of threePieris butterflies,P. rapae, P. melete andP. napi, living in the same area. *Res Popul Ecol* **22**: 163–183.

Ohsaki N (1982). Comparative population studies of threePieris butterflies,P. rapae, P. melete andP. napi, living in the same area III. Difference in the annual generation numbers in relation to habitat selection by adults. *Res Popul Ecol* **24**: 193–210.

Olwoch JM, Reyers B, Engelbrecht FA, Erasmus BFN (2008). Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. *Journal of Arid Environments* **72**: 108–120.

Opdam P, Wascher D (2004). Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation* **117**: 285–297.

Orr HA, Unckless RL (2008). Population Extinction and the Genetics of Adaptation. *The American Naturalist* **172**: 160–169.

O'Sullivan L, Bovet S, Streilein A (2008). TLM—the swiss 3D topographic landscape model. *ISPRS Proceeding* **37**: 1715–1719.

Pariset L, Joost S, Marsan PA, Valentini A, Econogene Consortium (EC) (2009). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet* **10**: 7.

Parmesan C, Ryrholm N, Stefanescu C, Hill JK, Thomas CD, Descimon H, *et al.* (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**: 579–583.

Patterson N, Price AL, Reich D (2006). Population Structure and Eigenanalysis. *PLoS Genet* **2**.

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013). The impact of global climate change on genetic diversity within populations and species. *Mol Ecol* **22**: 925–946.

Pavlidis P, Živković D, Stamatakis A, Alachiotis N (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol Biol Evol*: mst112.

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007). ORIGINAL ARTICLE: Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**: 102–117.

Perret J-L, Guerin PM, Diehl PA, Vlimant M, Gern L (2003). Darkness induces mobility, and saturation deficit limits questing duration, in the tick Ixodes ricinus. *Journal of Experimental Biology* **206**: 1809–1815.

Perret J-L, Guigoz E, Rais O, Gern L (2000). Influence of saturation deficit and temperature on Ixodes ricinus tick questing activity in a Lyme borreliosis-endemic area (Switzerland). *Parasitol Res* **86**: 554–557.

Peterson AT, Papeş M, Soberón J (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**: 63–72.

Peterson AT, Soberón J (2012). Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza & Conservação* **10**: 102–107.

Phillips SJ (2017). A brief tutorial on Maxent. *AT&T Research* **190**: 231–259.

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017). Opening the black box: an open-source release of Maxent. *Ecography* **40**: 887–893.

Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.

Phillips SJ, Dudík M (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161–175.

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, *et al.* (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**: 181–197.

Phillips SJ, Dudík M, Schapire RE (2004). A maximum entropy approach to species distribution modeling. In: *Proceedings of the twenty-first international conference on Machine learning*, ICML '04. Association for Computing Machinery: Banff, Alberta, Canada, p 83.

Pillonel T, Bertelli C, Aeby S, de Barsy M, Jacquier N, Kebbi-Beghdadi C, *et al.* (2019). Sequencing the Obligate Intracellular Rhabdochlamydia helvetica within Its Tick Host Ixodes ricinus to Investigate Their Symbiotic Relationship. *Genome Biol Evol* **11**: 1334–1344.

Pilloux L, Aeby S, Gaümann R, Burri C, Beuret C, Greub G (2015). The High Prevalence and Diversity of Chlamydiales DNA within Ixodes ricinus Ticks Suggest a Role for Ticks as Reservoirs and Vectors of Chlamydia-Related Bacteria. *Appl Environ Microbiol* **81**: 8177–8182.

Pimm SL, Raven P (2000). Extinction by numbers. *Nature* **403**: 843–845.

Piry S, Berthier K, Streiff R, Cros-Arteil S, Foucart A, Tatin L, *et al.* (2018). Fine-scale interactions between habitat quality and genetic variation suggest an impact of grazing on the critically endangered Crau Plain grasshopper (Pamphagidae: Prionotropis rhodanica). *Journal of Orthoptera Research* **27**: 61–73.

Porfirio LL, Harris RMB, Lefroy EC, Hugh S, Gould SF, Lee G, *et al.* (2014). Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. *PLoS One* **9**.

Porretta D, Mastrantonio V, Amendolia S, Gaiarsa S, Epis S, Genchi C, *et al.* (2013). Effects of global changes on the climatic niche of the tick Ixodes ricinus inferred by species distribution modelling. *Parasites Vectors* **6**: 271.

Pritchard JK, Rienzo AD (2010). Adaptation – not by sweeps alone. *Nat Rev Genet* **11**: 665–667.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**: 945–959.

QGIS Development Team (2016). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Raeymaekers JAM, Maes GE, Geldof S, Hontis I, Nackaerts K, Volckaert FAM (2008). Modeling genetic connectivity in sticklebacks as a guideline for river restoration. *Evolutionary Applications* **1**: 475–488.

Raghavan RK, Goodin DG, Hanzlicek GA, Zolnerowich G, Dryden MW, Anderson GA, *et al.* (2016). Maximum Entropy-Based Ecological Niche Model and Bio-Climatic Determinants of Lone Star Tick (Amblyomma americanum) Niche. *Vector-Borne and Zoonotic Diseases* **16**: 205–211.

Raghavan RK, Heath ACG, Lawrence KE, Ganta RR, Peterson AT, Pomroy WE (2020). Predicting the potential distribution of Amblyomma americanum (Acari: Ixodidae) infestation in New Zealand, using maximum entropy-based ecological niche modelling. *Exp Appl Acarol* **80**: 227–245.

Raghavan RK, Peterson AT, Cobos ME, Ganta R, Foley D (2019). Current and Future Distribution of the Lone Star Tick, Amblyomma americanum (L.) (Acari: Ixodidae) in North America. *PLoS One* **14**.

Ramírez-Restrepo L, MacGregor-Fors I (2017). Butterflies in the city: a review of urban diurnal Lepidoptera. *Urban Ecosyst* **20**: 171–182.

Ramunno L, Cosenza G, Pappalardo M, Longobardi E, Gallo D, Pastore N, *et al.* (2001). Characterization of two new alleles at the goat CSN1S2 locus. *Animal Genetics* **32**: 264–268.

Rana SK, Luo D, Rana HK, O'Neill AR, Sun H (2019). Geoclimatic factors influence the population genetic connectivity of Incarvillea arguta (Bignoniaceae) in the Himalaya–Hengduan Mountains biodiversity hotspot. *Journal of Systematics and Evolution* **n/a**.

Ray N (2005). pathmatrix: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes* **5**: 177–180.

Ray N, Lehmann A, Joly P (2002). Modeling spatial distribution of amphibian populations: a GIS approach based on habitat matrix permeability. *Biodiversity and Conservation* **11**: 2143–2165.

Rayfield B, Fortin M-J, Fall A (2009). The sensitivity of least-cost habitat graphs to relative cost surface values. *Landscape Ecol* **25**: 519–532.

Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, *et al.* (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *PNAS* **116**: 10418–10423.

Razgour O, Taggart JB, Manel S, Juste J, Ibáñez C, Rebelo H, *et al.* (2018). An integrated framework to identify wildlife populations under threat from climate change. *Molecular Ecology Resources* **18**: 18–31.

Reaka-Kudla ML, Wilson DE, Wilson EO (1996). *Biodiversity II: Understanding and Protecting Our Biological Resources*. Joseph Henry Press.

Reed DH, Frankham R (2001). How Closely Correlated Are Molecular and Quantitative Measures of Genetic Variation? A Meta-Analysis. *Evolution* **55**: 1095–1103.

Reed TE, Schindler DE, Waples RS (2011). Interacting Effects of Phenotypic Plasticity and Evolution on Population Persistence in a Changing Climate. *Conservation Biology* **25**: 56–63.

Regenscheit N, Holzwarth N, Greub G, Aeby S, Pospischil A, Borel N (2012). Deer as a potential wildlife reservoir for Parachlamydia species. *The Veterinary Journal* **193**: 589–592.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015). A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* **24**: 4348–4370.

Ren H, Wang G, Chen L, Jiang J, Liu L, Li N, *et al.* (2016). Genome-wide analysis of long non-coding RNAs at early stage of skin pigmentation in goats (Capra hircus). *BMC Genomics* **17**: 67.

Renner IW, Warton DI (2013). Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics* **69**: 274–281.

Reside AE, Butt N, Adams VM (2018). Adapting systematic conservation planning for climate change. *Biodivers Conserv* **27**: 1–29.

Richards OW (1940). The Biology of the Small White Butterfly (Pieris rapae), with Special Reference to the Factors Controlling its Abundance. *Journal of Animal Ecology* **9**: 243–288.

Riley SJ, DeGloria SD, Elliot R (1999). Index that quantifies topographic heterogeneity. *intermountain Journal of sciences* **5**: 23–27.

Rischkowsky B, Pilling D (2007). *The state of the world's animal genetic resources for food and agriculture*. Food & Agriculture Org.

Rizzoli A, Silaghi C, Obiegala A, Rudolf I, Hubálek Z, Földvári G, *et al.* (2014). Ixodes ricinus and Its Transmitted Pathogens in Urban and Peri-Urban Areas in Europe: New Hazards and Relevance for Public Health. *Front Public Health* **2**.

Rochat E, Leempoel K, Vajana E, Colli L, Ajmone-Marsan P, Joost S, *et al.* (2016). Map of genotype frequency change in autochtonous Moroccan sheep breeds due to global warming.

Roome A, Spathis R, Hill L, Darcy JM, Garruto RM (2018). Lyme Disease Transmission Risk: Seasonal Variation in the Built Environment. *Healthcare (Basel)* **6**.

R. Taylor H, Dussex N, van Heezik Y (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation* **10**: 231–242.

Rubidge EM, Patton JL, Lim M, Burton AC, Brashares JS, Moritz C (2012). Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Clim Change* **2**: 285–288.

Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, *et al.* (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters* **21**: 1085–1096.

Running SW, Nemani RR, Hungerford RD (1987). Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. *Canadian Journal of Forest Research* **17**: 472–483.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491–494.

Sage KM, Johnson TL, Teglas MB, Nieto NC, Schwan TG (2017). Ecological niche modeling and distribution of Ornithodoros hermsi associated with tick-borne relapsing fever in western North America. *PLOS Neglected Tropical Diseases* **11**: e0006047.

Saura S (2002). Effects of minimum mapping unit on land cover data spatial configuration and composition. *International Journal of Remote Sensing* **23**: 4853–4880.

Saura S, Estreguil C, Mouton C, Rodríguez-Freire M (2011). Network analysis to assess landscape connectivity trends: Application to European forests (1990–2000). *Ecological Indicators* **11**: 407–416.

Savage AE, Mulder KP, Torres T, Wells S (2018). Lost but not forgotten: MHC genotypes predict overwinter survival despite depauperate MHC diversity in a declining frog. *Conserv Genet* **19**: 309–322.

Scherf B, Rischkowsky B, Hoffmann I, Wieczorek M, Montironi A, Cardellino R (2008). Livestock Genetic Diversity in Dry Rangelands. In: Lee C, Schaaf T (eds) *The Future of Drylands*, Springer Netherlands, pp 89–100.

Schleupner C, Link PM (2008). Potential impacts on important bird habitats in Eiderstedt (Schleswig-Holstein) caused by agricultural land use changes. *Applied Geography* **28**: 237–247.

Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S (2012). Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and Systematics* **43**: 23–43.

Schoville SD, Widmer I, Deschamps-Cottin M, Manel S (2013). Morphological clines and weak drift along an urbanization gradient in the butterfly, Pieris rapae. *PloS one* **8**: e83095.

Schtickzelle N, Baguette M (2003). Behavioural responses to habitat patch boundaries restrict dispersal and generate emigration–patch area relationships in fragmented landscapes. *Journal of Animal Ecology* **72**: 533–545.

Schtickzelle N, Mennechez G, Baguette M (2006). Dispersal Depression with Habitat Fragmentation in the Bog Fritillary Butterfly. *Ecology* **87**: 1057–1065.

Selmoni O, Rochat E, Lecellier G, Berteaux-Lecellier V, Joost S (2019). Seascape genomics as a new tool to empower coral reef conservation strategies: an example on north-western Pacific Acropora digitifera. *bioRxiv*: 588228.

Serret H, Raymond R, Foltête J-C, Clergeau P, Simon L, Machon N (2014). Potential contributions of green spaces at business sites to the ecological network in an urban agglomeration: The case of the Ile-de-France region, France. *Landscape and Urban Planning* **131**: 27–35.

Sgrò CM, Lowe AJ, Hoffmann AA (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* **4**: 326–337.

Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, *et al.* (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**: 78–87.

Shanahan DF, Fuller RA, Bush R, Lin BB, Gaston KJ (2015). The Health Benefits of Urban Nature: How Much Do We Need? *BioScience* **65**: 476–485.

Shochat E, Lerman SB, Anderies JM, Warren PS, Faeth SH, Nilon CH (2010). Invasion, Competition, and Biodiversity Loss in Urban Ecosystems. *BioScience* **60**: 199–208.

Sillero N (2011). What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling* **222**: 1343–1346.

Skarphédinsson S, Jensen PM, Kristiansen K (2005). Survey of Tickborne Infections in Denmark. *Emerg Infect Dis* **11**: 1055–1061.

Soanes K, Sievers M, Chee YE, Williams NSG, Bhardwaj M, Marshall AJ, *et al.* (2019). Correcting common misconceptions to inspire conservation action in urban environments. *Conservation Biology* **33**: 300–306.

Soucy J-PR, Slatculescu AM, Nyiraneza C, Ogden NH, Leighton PA, Kerr JT, *et al.* (2018). High-Resolution Ecological Niche Modeling of Ixodes scapularis Ticks Based on Passive Surveillance Data at the Northern Frontier of Lyme Disease Emergence in North America. *Vector-Borne and Zoonotic Diseases* **18**: 235–242.

Steffen W, Crutzen PJ, McNeill JR (2007). The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. *AMBIO: A Journal of the Human Environment* **36**: 614–621.

Stella A, Nicolazzi EL, Van Tassell CP, Rothschild MF, Colli L, Rosen BD, *et al.* (2018). AdaptMap: exploring goat diversity and adaptation. *Genet Sel Evol* **50**: 61.

Stephan S, Guerra D, Pospischil A, Hilbe M, Weissenböck H, Novotný L, *et al.* (2014). Chlamydiaceae and chlamydia-like organisms in free-living small mammals in europe and afghanistan. *Journal of Wildlife Diseases* **50**: 195–204.

Stevens VM, Turlure C, Baguette M (2010). A meta-analysis of dispersal in butterflies. *Biological Reviews* **85**: 625–642.

Stockwell D (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science* **13**: 143–158.

Strubbe D, Matthysen E (2009). Predicting the potential distribution of invasive ring-necked parakeets Psittacula krameri in northern Belgium using an ecological niche modelling approach. *Biol Invasions* **11**: 497–513.

Stucki S, Orozco-terWengel P, Bruford MW, Colli L, Masembe C, Negrini R, *et al.* (2014). High performance computation of landscape genomic models integrating local indices of spatial association. *arXiv:14057658 [q-bio]*.

Stucki S, Orozco-terWengel P, Forester BR, Duruz S, Colli L, Masembe C, *et al.* (2017). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources* **17**: 1072–1089.

Sutcliffe OL, Bakkestuen V, Fry G, Stabbetorp OE (2003). Modelling the benefits of farmland restoration: methodology and application to butterfly movement. *Landscape and Urban Planning* **63**: 15–31.

Swarna Nantha H, Tisdell C (2009). The orangutan–oil palm conflict: economic constraints and opportunities for conservation. *Biodivers Conserv* **18**: 487–502.

Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, Negrini R, *et al.* (2008). Are cattle, sheep, and goats endangered species? *Molecular Ecology* **17**: 275–284.

Takami Y, Koshio C, Ishii M, Fujii H, Hidaka T, Shimizu I (2004). Genetic diversity and structure of urban populations of Pieris butterflies assessed using amplified fragment length polymorphism. *Molecular Ecology* **13**: 245–258.

Thomassen HA, Fuller T, Buermann W, Milá B, Kieswetter CM, Jarrín-V. P, *et al.* (2011). Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evolutionary Applications* **4**: 397–413.

Thorne CR, Zevenbergen LW, Burt TP, Butcher DP (1987). Terrain analysis for quantitative description of zero-order basins. *IAHS-AISH publication*: 121–130.

Thornton PE, Running SW, White MA (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of hydrology* **190**: 214–251.

Timperio A, Giles-Corti B, Crawford D, Andrianopoulos N, Ball K, Salmon J, *et al.* (2008). Features of public open spaces and physical activity among children: Findings from the CLAN study. *Preventive Medicine* **47**: 514–518.

Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* **13**: 397–405.

Turin H, den Boer PJ (1988). Changes in the distribution of carabid beetles in The Netherlands since 1880. II. Isolation of habitats and long-term time trends in the occurence of carabid species with different powers of dispersal (Coleoptera, Carabidae). *Biological Conservation* **44**: 179–200.

Urban D, Keitt T (2001). Landscape Connectivity: A Graph-Theoretic Perspective. *Ecology* **82**: 1205–1218.

Vajana E (2017). Exploring livestock evolutionary history, diversity, adaptation and conservation through landscape genomics and ecological modelling.

Vajana E, Barbato M, Colli L, Milanesi M, Rochat E, Fabrizi E, *et al.* (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front Genet* **9**.

de Vallière S, Cometta A (2006). Evidence de nouveaux foyers d'endémie de méningo-encéphalite vernoestivale en Suisse romande. *Revue Médicale Suisse*.

Vandergast AG, Perry WM, Lugo RV, Hathaway SA (2011). Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* **11**: 158–161.

Vangestel C, Eckert AJ, Wegrzyn JL, St. Clair JB, Neale DB (2018). Linking phenotype, genotype and environment to unravel genetic components underlying cold hardiness in coastal Douglas-fir (Pseudotsuga menziesii var. menziesii). *Tree Genetics & Genomes* **14**: 10.

Villemereuil P de, Frichot É, Bazin É, François O, Gaggiotti OE (2014). Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology* **23**: 2006–2019.

Vitti JJ, Grossman SR, Sabeti PC (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics* **47**: 97–120.

Vos P, Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic acids research* **23**: 4407–4414.

Vuilleumier S, Wilcox C, Cairns BJ, Possingham HP (2007). How patch configuration affects the impact of disturbances on metapopulation persistence. *Theoretical Population Biology* **72**: 77–85.

Waldvogel A-M, Feldmeyer B, Rolshausen G, Exposito-Alonso M, Rellstab C, Kofler R, *et al.* (2020). Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters* **4**: 4–18.

Walker PA, Cocks KD (1991). HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters*: 108–118.

Walther G-R, Roques A, Hulme PE, Sykes MT, Pyšek P, Kühn I, *et al.* (2009). Alien species in a warmer world: risks and opportunities. *Trends in Ecology & Evolution* **24**: 686–693.

Wan J-Z, Wang C-J, Yu F-H (2019). Effects of occurrence record number, environmental variable number, and spatial scales on MaxEnt distribution modelling for invasive plants. *Biologia* **74**: 757–766.

Wang S, Ge W, Luo Z, Guo Y, Jiao B, Qu L, *et al.* (2017). Integrated analysis of coding genes and non-coding RNAs during hair follicle cycle of cashmere goat (Capra hircus). *BMC Genomics* **18**: 767.

Waples RS, Gaggiotti O (2006). INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**: 1419–1439.

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, *et al.* (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**: 709–725.

Willi Y, Hoffmann AA (2009). Demographic factors and genetic variation influence population persistence under environmental change. *Journal of Evolutionary Biology* **22**: 124–133.

Williams BL, Brawn JD, Paige KN (2003). Landscape scale genetic effects of habitat fragmentation on a high gene flow species: Speyeria idalia (Nymphalidae). *Molecular Ecology* **12**: 11–20.

Williams HW, Cross DE, Crump HL, Drost CJ, Thomas CJ (2015). Climate suitability for European ticks: assessing species distribution models against null models and projection under AR5 climate. *Parasites & Vectors* **8**: 440.

Williams SE, Shoo LP, Isaac JL, Hoffmann AA, Langham G (2008). Towards an Integrated Framework for Assessing the Vulnerability of Species to Climate Change. *PLOS Biol* **6**: e325.

Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology* **27**: 4041–4051.

Wilson EO (1988). *Biodiversity*. National Academies Press.

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions* **14**: 763–773.

Wood BC, Pullin AS (2002). Persistence of species in a fragmented urban landscape: the importance of dispersal ability and habitat availability for grassland butterflies. *Biodivers Conserv* **11**: 1451–1468.

Wright S (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*. na.

Wright S (1949). The Genetical Structure of Populations. *Annals of Eugenics* **15**: 323–354.

Yokoyama R, Shirasawa M, Pike RJ (2002). Visualizing topography by openness: a new application of image processing to digital elevation models. *Photogrammetric engineering and remote sensing* **68**: 257–266.

Zeller KA, McGarigal K, Whiteley AR (2012). Estimating landscape resistance to movement: a review. *Landscape Ecol* **27**: 777–797.

Zhang Y, Wu K, Wang L, Wang Z, Han W, Chen D, *et al.* (2019). Comparative study on seasonal hair follicle cycling by analysis of the transcriptomes from cashmere and milk goats. *Genomics*.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.

Zhou X, Carbonetto P, Stephens M (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLOS Genetics* **9**: e1003264.

Zimmermann NE, Roberts DW (2001). Final report of the MLP climate and biophysical mapping project. *Birmensdorf: Swiss Federal Research Institute*.

ANNEXES

A1. Maxent modelling parameters

Selection of background points

The user must select background locations that are representative of the environmental conditions in the study area. However, one of the assumptions of the Maxent algorithm is that any location has the same probability of being reached by the species (Elith et al., 2010). Background locations should therefore only contain sites where the species could have dispersed and been sampled (Phillips *et al.*, 2006). Merow *et al.* (2013) indicated that background sites can be selected either throughout the entire territory or only within the species' range. Other strategies exist, such as choosing the background location at some distance (in the geographical or environmental space) from the occurrence points, or outside the environmental domain favourable to the species (Chefaoui and Lobo, 2008; Lobo and Tognelli, 2011; Barbet-Massin *et al.*, 2012). However, background locations chosen too far from the presence points or under too different environmental conditions are less informative (Lobo and Tognelli, 2011; Acevedo *et al.*, 2012; Barbet-Massin *et al.*, 2012). If possible, background locations should also present the same sampling bias as in the occurrence data (Phillips *et al.*, 2009; Merow *et al.*, 2013).

Environmental features

Instead of using only raw environmental variables, several features can be calculated and integrated into the model. First, the use of a quadratic feature (square of the environmental variable) introduces the additional constraint that the variance of the environmental variable should be close to that observed in the training dataset. This can be used to model the tolerance of species to variation from optimal conditions (Phillips *et al.*, 2006). Second, product features integrating each product of two (or more) environmental variables can be used to model complex integrations between the variables (Elith *et al.*, 2010). Finally, threshold features can also be integrated to model environmental variables for which a known tolerance limit exists for the species under study (Merow *et al.*, 2013; Wan *et al.*, 2019).

Regularisation parameter

Phillips *et al.* (2006) have implemented in Maxent a regularization procedure that makes it possible to discriminate too complex models that have a high log-likelihood but are unlikely to generalize well (Elith *et al.*, 2006). They proposed a penalized maximum likelihood procedure and the user is invited to choose the value of a regularisation constant (the higher it is, the greater the penalization for complex models). Several values need to be tested and according to the performance of the resulting models (Merow *et al.*, 2013). This penalization could allow to automatically reduce the number of environmental variables, but it is still advisable to first select a set of meaningful and uncorrelated predictors (Elith *et al.*, 2010; Fourcade *et al.*, 2018).

Output type

Several alternatives are available in Maxent to rescale the raw suitability index. A linear relationship is theoretically expected between the raw output and the local abundance, while a monotonic, but nonlinear, relationship is predicted between transformed outputs and the abundance (Phillips, 2017). Among the available transformations, logistic output is based on a logistic model instead of an exponential one and it may be interpreted as the probability to find a species in a given site (Phillips and Dudík, 2008). However, this requires an assumption on the prevalence value, the default assumption being that there is a 50% of chance to observe a species in a typically suitable area (Elith *et al.*, 2006; Phillips and Dudík, 2008). The “cloglog” transformation is closely related, but is derived from the interpretation of the Maxent model as a Poisson process, and therefore contains a stronger theoretical justification (Phillips *et al.*, 2017). This form of output is bounded between 0 and 1 and can be interpreted as a probability of presence. However, this interpretation requires the assumption that a typical presence location will have a probability of 0.63 (Phillips, 2017).

A2. Paper A: Supplementary material

A2.1 Supp. File 1 – Prospective campaign

In order to select the sites visited during the prospective campaign, we proceeded in four steps:

1. Based on the already available ticks occurrences (data from the Swiss army field campaign and data from the smartphone application for 2015 to 2017), we run a MAXENT model to obtain a first map of suitability for ticks.
2. We performed a PCA on the environmental predictors extracted in the pixels predicted as potentially suitable for ticks (suitability from step 1 greater than 0.2).
3. We computed a k-means classification on the components of the PCA. This allowed us to define 5 environmental clusters on the Swiss territory (Figure 1).
4. We manually selected the sampling sites:
    - in areas defined as potentially suitable for ticks (suitability predicted at step 1 greater than 0.2),
    - such as to sample sites in each environmental cluster defined at step 2,
    - such as to maximise the number of sites that can be visited during one day (i.e. the sites can be link together by roads or paths),
    - so as to complete the dataset already available regarding the presence of *Chlamydiales* bacteria (data from the Swiss army field campaign).

As a result, 96 sites were visited and ticks were found in 81 of them, corresponding to 228 ticks. In addition, some relatives of the authors provided ticks they had collected. By this way, 28 additional ticks were received, from 14 new sites. In total, the prospective campaign therefore provided 256 ticks from 95 sites.

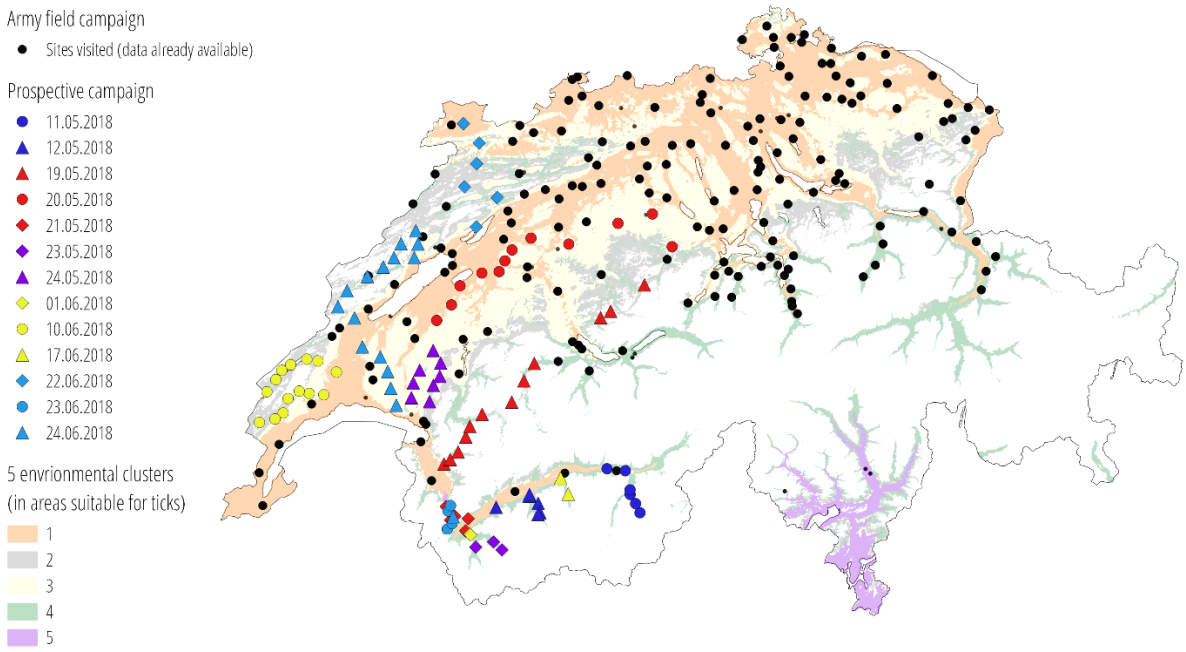
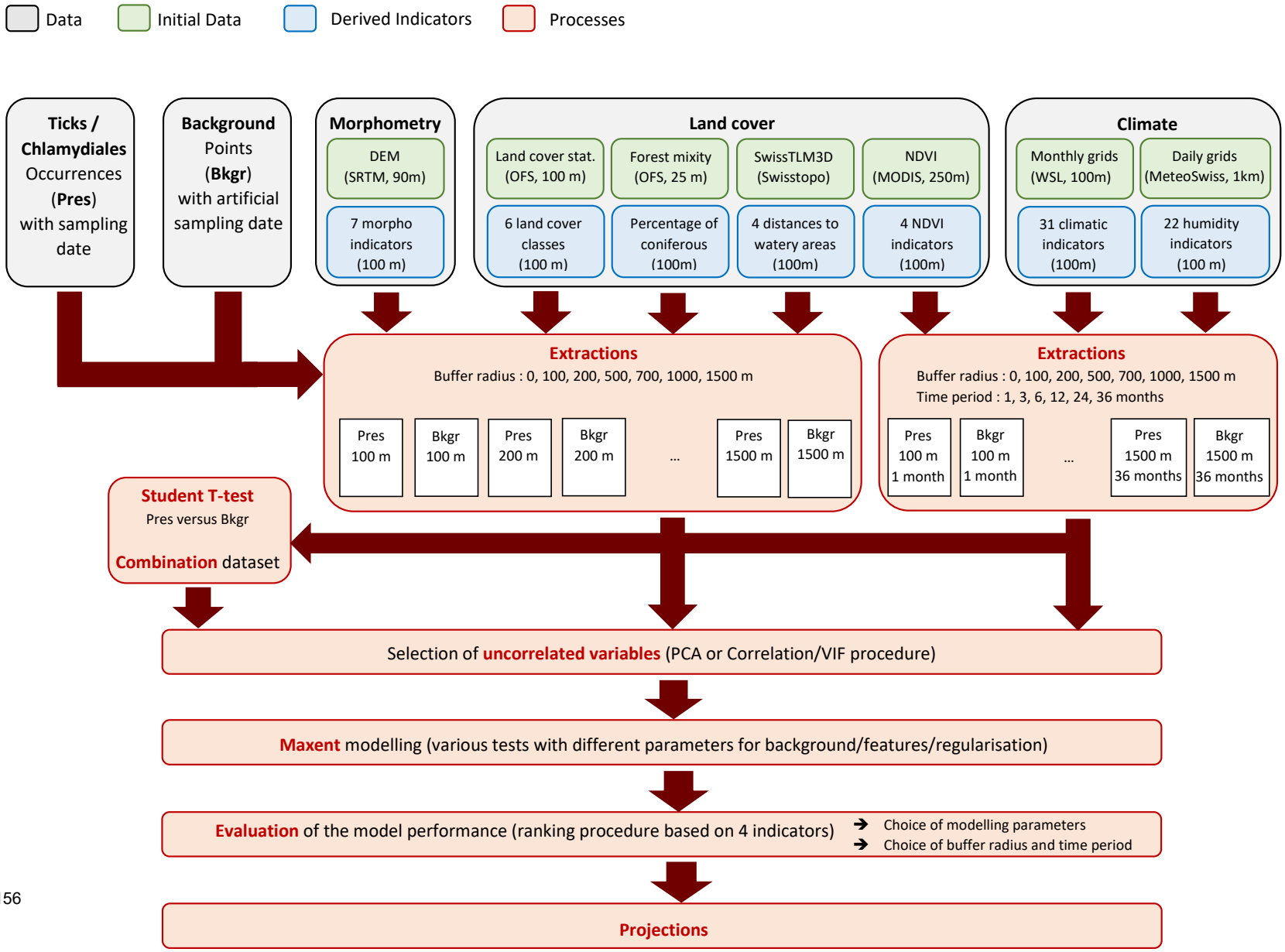


Figure 1: Map of the environmental clusters defined with the k-means performed on the PCA-components of the environmental predictors, and location of the sites visited during the prospective campaign.

A2.2 Supp. File 2 – Method



A2.3 Supp. File 3 – Environmental data

Morphometry

Initial data

- SRTM Digital Elevation Model
- **Spatial resolution**: 90 m
- **Source**: NASA Shuttle Radar Topography Mission (https://www2.jpl.nasa.gov/srtm/), with hole-filled version processed by the CIAT Agroecosystems Resilience project (Jarvis, 2008)
- **URL:** http://srtm.csi.cgiar.org/

Pre-processing

- The two tiles covering Switzerland were downloaded and merged using the QGIS 2.14.7 software (function merge from GDAL).
- The merged dataset was then resampled at a 100-m resolution and cropped to the Swiss extent using the SAGA "resampling" tool accessed from QGIS 2.14.7 and using the interpolation method: mean value (cell-area weighted).
- Null values were assigned for all pixels outside the Swiss borders.

Slope, Aspect, General Curvature (GC)

- **Method**: These indicators were computed in SAGA GIS 2.3. using the tool "Terrain Analysis > Morphometry > Slope, Aspect, Curvature", with the method "9 parameter 2nd order polynom" (Thorne *et al.*, 1987) and the units defined in degrees.

Morphometric Protection Index (MPI)

- **Definition**: This indicator provides a dimensionless index expressing how well an area is protected from the surrounding relief, based on the analysis of the environment surrounding each pixel up to a given distance. It is equivalent to the positive openness described by Yokoyama et al. (2002).
- **Method**: This indicator was computed in SAGA GIS 2.3.2 using the tool "Terrain Analysis > Morphometry > Morphometric Protection Index and the default parameters (the relief in the surrounding 2km of each pixel is taken into account).

Terrain Ruggedness Index (RI)

- **Definition**: This indicator compares the elevation of one pixel with the elevation of the neighbouring pixels to provide a measure of terrain heterogeneity (Riley *et al.*, 1999).
- **Method**: This indicator was computed in SAGA GIS 2.3.2 using the tool "Terrain Analysis > Morphometry > Terrain Ruggedness Index" with the default parameters (Radius (Cells)=1 indicating that one neighbour cell is considered in each direction).

Sky-view factor (SVF)

- **Definition**: This indicator provides an indication of the portion of sky that is obstructed by the surrounding relief: 0 = completely obstructed, 1=completely visible (Böhner and Antonić, 2009, p. 8)
- **Method**: This indicator was computed in SAGA GIS 2.3.2 using the tool "Terrain Analysis > Lighting, Visibility > Sky view factor" with the default parameters (Maximum search radius = 10 km).

Topographic Wetness Index (TWI)

- **Definition**: This indicator is defined from the ratio of the catchment area (area draining water to a given cell) to the local slope (indicator of the capacity to evacuate the water received) and is used as a proxy of soil moisture (Kopecký and Čížková, 2010).
- **Method**: First we computed the specific catchment area in SAGA GIS 2.3.2 using the tool "Terrain Analysis > Hydrology > Flow Width and Specific Catchment Area" with the default parameters (Aspect method). The TWI was then computed in SAGA GIS 2.3.2 using the tool "Terrain Analysis > Hydrology > Topographic Wetness Index" with the default parameters (Standard method).

Land Cover

Land cover classification

- Land cover classification in 6 classes : artificial areas, grass and herb vegetation, brush vegetation, tree vegetation, bare land and watery areas
- **Spatial resolution** : 100 m
- **Source**: Swiss Federal Statistical Office (OFS, 2017)
- **URL:** https://www.bfs.admin.ch/bfs/fr/home/statistiques/espace-environnement/nomenclatures/arealstatistik/nolc2004.html
- **Processing:** the only processing was to rasterise the data using the function rasterise in QGIS 2.14.7 (the initial data was available as a .csv file)

Percentage of coniferous in forest

Initial Data

- Raster file classifying the forests of Switzerland into four classes: pure coniferous, mixed coniferous, mixed broadleaved and pure broadleaved.
- **Spatial resolution**: 25 m, but with a grid translated by 12.5m as compared to the other data.
- **Source**: Swiss Federal Statistical Office (OFS, 2013)
- **URL:** https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/sol-utilisation-couverture/donnees-derivees-autres-donnees/mixite-forets.html

Processing

- First, the raster with a spatial resolution of 25m was resampled in QGIS 2.14.7 to a raster with a spatial resolution of 12.5 m using the function "resample" with the nearest neighbour method.
- A percentage of conifers was then assigned to each 12.5m pixel according to the classification proposed by OFS:
  - 0 = no-forest => 0 % coniferous
  - 1 = coniferous forest => considered 100% coniferous
  - 2 = mixed forest predominantly coniferous => considered 70% coniferous
  - 3 = mixed forest predominantly broadleaved => considered 30% coniferous
  - 4 = broadleaved forest => considered 0% coniferous
  - 9 = unclassified => considered no forest => 0% coniferous
- The 12.5 m raster was aggregated to a 100 m target grid, by computing for each target cell the average percentage of coniferous using the tool "zonal statistics" in QGIS 2.14.7.
- The resulting grid was rasterised using the "rasterise" function in QGIS 2.14.7.

Distances to water areas

Initial Data

- Vector landscape model SwissTLM3D from 2016
- **Source**: Swiss Federal Office of Topography (O'Sullivan *et al.*, 2008)
- **URL:** https://shop.swisstopo.admin.ch/en/products/landscape/tlm3D

Processing

- All the elements characterising watery areas were extracted from the landscape model
  - For running water: the lines "Fliessgewaesser" and the polygons "Fliessgewaesser" extracted from the LandCover (Bodenbedeckung) polygons
  - For stagnant water: the lines "Stehendes Gewasser" and the polygons " "Stehendes Gewasser" extracted from the land cover (Bodenbedeckung) polygons
  - For the wetlands: the polygons "Feuchtgebiet" extracted from the land cover (Bodenbedeckung) polygons
- The vector layers were rasterised using the "rasterise" function in QGIS 2.14.7.
- For each pixel, the minimal Euclidean distance to each water category was then computed using the function "Raster > Analysis > Proximity" in QGIS 2.14.7. This resulted in three raster layers, representing the minimum distance to running water elements, stagnant water and wetlands, respectively.
- Finally, the minimum of the three raster files was used as the minimal distance to any watery element.

Vegetation Indexes

Initial Data

- MODIS Terra 16-days composite NDVI
- **Definition**: The 16-day composite NDVI is produced on 16-day intervals and provide an indicator of the greenness of the vegetation during these 16 days. NDVI is derived from the reflectance in the red and near-infrared (NIR) bands obtained from the images of the MODIS satellite.

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

A large amount of red wavelengths are absorbed by the vegetation during photosynthesis, while the near infrared is reflected, in a proportion that depends in particular on the leaf area index. Land covered by vegetation will therefore show a large difference between NIR and red reflectance, resulting in high NDVI values.

- **Units**: The valid range of value is -2000 to 10'000 with a scale factor of 0.0001 (i.e. a value of -2000 correspond to a NDVI of -0.2, whereas a value of 10'000 indicates a NDVI equals to 1.0)
- **Spatial resolution**: 250 m
- **Source**: NASA Moderate Resolution Imaging Spectoradiometer (MODIS) (Huete *et al.*, 1999)
- **URL:** https://modis.gsfc.nasa.gov/data/dataprod/mod13.php
- **Download:** https://search.earthdata.nasa.gov/

Processing

- We downloaded all images for the years 2006 to 2019.
- The hdf4 files were converted to .tif format using the "gdal_translate" function in R (R Development Core Team, 2008)

- The MODIS data being in sinusoidal projection, rasters were reprojected in the CH1903/LV03 projection system using the “projectRaster” function of the “raster” package in R
- The files were cropped and resampled to a 100m resolution using the “crop” and “resample” function from the “raster” package in R
- For each pixel, the monthly mean values were then computed in R.
- Finally, remaining null values were replaced by the average value of the neighbouring pixels using the “focal” function from the “raster” package in R.

Derived variables

Four indicators were derived for the period of interest (1, 3, 12, 24 or 36 months before the sampling date).

1. Average of monthly mean NDVI (***meanNDVIm***)
2. Minimum of monthly mean NDVI (***minNDVIm***)
3. Maximum of monthly mean NDVI (***maxNDVIm***)
4. Range of monthly mean NDVI (***RgeNDVIm***)

Climate

Temperature and precipitation

Initial Data

- Monthly mean, maximum and minimum temperature and monthly sum of precipitation
- **Spatial resolution**: 100 m
- **Source**: grids computed by the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), based on data from MeteoSwiss weather stations and a 100 m resolution digital elevation model aggregated from the DHM25 of SwissTopo. The computation was performed using Daymet software (Thornton *et al.,* 1997) and the reported mean absolute error (crossvalidation) is ~1°C for temperature and ~10-15% for precipitation (personal communication from WSL).
- **URL:** https://www.wsl.ch/de/projekte/climate-data-portal.html
  DHM25 Swisstopo**:** https://shop.swisstopo.admin.ch/fr/products/height_models/dhm25
  MeteoSwiss: https://www.meteoswiss.admin.ch/home/measurement-values.html

Derived variables

First, 15 indicators were derived for the period of interest (1, 3, 12, 24 or 36 months before the sampling date). Some of these indicators are very close to the worldclim bioclimatic variables (https://worldclim.org/data/bioclim.html). They were computed in R using two custom R functions (one defined for the treatment of data frame and the other for raster layers). The two functions are available in: https://github.com/estellerochat/SDM-Chlamydiales.

1. Average of the monthly mean temperatures over the period of interest (***meantmean***)
2. Maximum of the monthly maximal temperatures over the period of interest (***maxtmax***)
3. Minimum of the monthly maximal temperatures over the period of interest (***mintmax***)
4. Maximum of the monthly minimal temperature over the period of interest (***maxtmin***)
5. Minimum of the monthly minimal temperatures over the period of interest (***mintmin***)
6. Average of the monthly range of temperatures (***meanMoRge***)
7. Global range of temperature (maxtmax-mintmin) (***tRge***)
8. Isothermality (100*meanMoRge / tRge) (***isotherm***)
9. Temperature seasonality (standard deviation*100) (***tseason***)
10. Mean temperature of the coldest month (***mintmean***)
11. Mean temperature of the warmest month (***maxtmean***)

12. Total sum of precipitation (***sumprec***)
13. Maximum of the monthly sums of precipitation over the period of interest (***maxprec***)
14. Minimum of the monthly sums of precipitation over the period of interest (***minprec***)
15. Precipitation seasonality (Coefficient of Variation CV = sd/mean*100) (***pseason***)

Secondly, 16 additional indicators were derived when the period of interest was exceeding 3 months (i.e. 6, 12, 24 or 36 months) (CM=”consecutive months”)

1. Average of the monthly mean temperature of the 3 coldest CM (***meantmean3cold***)
2. Average of the monthly minimal temperature of the 3 coldest CM (***meantmin3cold***)
3. Average of the monthly maximal temperature of the 3 coldest CM (***meantmax3cold***)
4. Sum of precipitation of the 3 coldest CM (***prec3cold***)
5. Average of the monthly mean temperature of the 3 warmest CM (***meantmean3warm***)
6. Average of the monthly minimal temperature of the 3 warmest CM (***meantmin3warm***)
7. Average of the monthly maximal temperature of the 3 warmest CM (***meantmax3warm***)
8. Sum of precipitation of the 3 warmest CM (***prec3warm***)
9. Average of the monthly mean temperature of the 3 wettest CM (***meantmean3wet***)
10. Average of the monthly minimal temperature of the 3 wettest CM (***meantmin3wet***)
11. Average of the monthly maximal temperature of the 3 wettest CM (***meantmax3warm***)
12. Sum of precipitation of the 3 wettest CM (***prec3wet***)
13. Average of the monthly mean temperature of the 3 driest CM (***meantmean3dry***)
14. Average of the monthly minimal temperature of the 3 driest CM (***meantmin3dry***)
15. Average of the monthly maximal temperature of the 3 driest CM (***meantmax3dry***)
16. Sum of precipitation of the 3 driest CM (***meantmean3dry***)

Humidity variables

Initial Data

- Daily mean, maximum and minimum temperature
- **Spatial resolution**: 1 km
- **Source**: MeteoSwiss
- **URL:** https://www.meteoswiss.admin.ch/home/climate/swiss-climate-in-detail/raeumliche-klimaanalysen.html

Processing

- The daily grids were imported in R
- The daily relative humidity was computed using the same procedure as in Zimmermann *et al.* (2001)
    - Compute the average daytime temperature following Running *et al.* (1987)

$$t_{day} = 0.394\ t_{min} + 0.606\ t_{max}$$

    - Compute ambient vapour pressure using the Tetens equation for temperatures above 0°C (Murray, 1966) and minimum temperature as an approximation of dew point temperature (Running *et al.*, 1987)

$$VP_{amb} = 610.78 \exp\left(\frac{17.269\ t_{min}}{237.3 + t_{min}}\right)$$

    - Compute the potential vapour pressure of saturated air for daytime temperature using the Tetens equation for temperatures above 0°C (Murray, 1966) and the previously computed average daytime temperature:

$$VP_{sat} = 610.78\ exp\left(\frac{17.269\ t_{day}}{237.3 + \ t_{day}}\right)$$

- Compute the relative Humidity (in %)

$$RH = \frac{VP_{amb}}{VP_{sat}} * 100$$

- The daily relative humidity grids were then aggregated to compute four monthly grids:
  1. Monthly mean of RH
  2. Monthly median of RH
  3. Monthly quantile 0.25 of RH
  4. Monthly quantile 0.75 of RH

Derived variables

22 indicators were derived for the period of interest (1, 3, 12, 24 or 36 months before sampling date). They were computed in R using two custom R functions (one defined for the treatment of data frame and the other for raster layers). The two functions are available in: https://github.com/estellerochat/SDM-Chlamydiales.

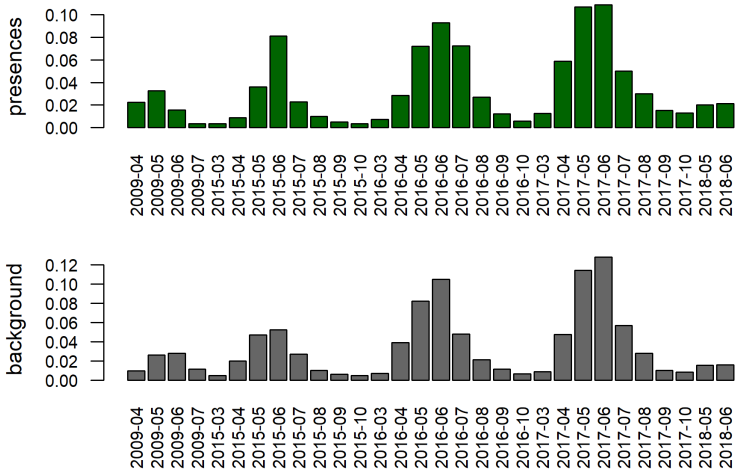
1. Average of monthly mean RH (***meanRHmean***)
2. Average of monthly median RH (***meanRHq050***)
3. Minimum of monthly mean RH (***minRHmean***)
4. Maximum of monthly mean RH (***maxRHmean***)
5. Minimum of monthly 0.25 quantile of RH (***minRHq025***)
6. Minimum of monthly 0.75 quantile of RH (***minRHq075***)
7. Maximum of monthly 0.75 quantile of RH (***maxRHq075*)**
8. Range of monthly RH (***RHrge***)
9. Average of the monthly ranges of RH (***RHMoRge***)
10. Mean daily RH (***meanRHD*)**
11. Median daily RH (***medRHD*)**
12. Minimum daily RH (***minRHD***)
13. Maximum daily RH (***maxRHD*)**
14. Range of daily RH (***rangeRHD***)
15. Quantile 0.25 of daily RH (***q025RHD***)
16. Quantile 0.75 of daily RH(***q075RHD***)
17. Number of days with RH<70% (***ndRHDinf70***)
18. Number of days with RH<80% (***ndRHDinf80***)
19. Number of days with RH>90% (***ndRHDsup90***)
20. Maximum number of consecutive days with RH< 70% (***ncdRHDinf70***)
21. Maximum number of consecutive days with RH< 80% (***ncdRHDinf80***)
22. Maximum number of consecutive days with RH>90% (***ncdRHDsup90***)

A2.4 Supp. File 4 – Background datasets

Ixodes ricinus

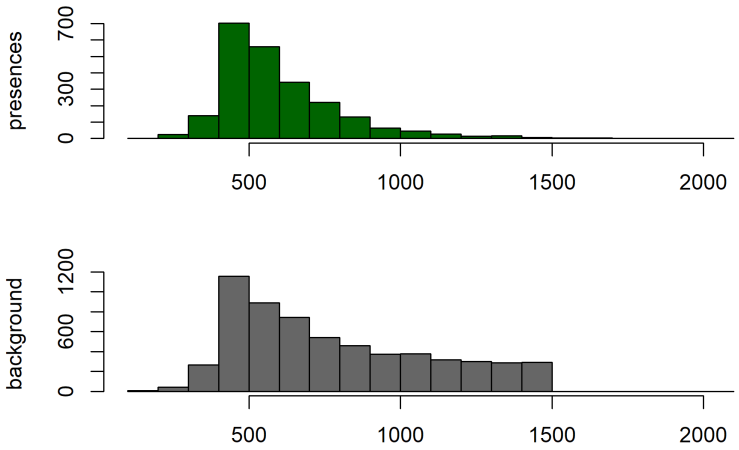
Sampling date

Distribution of sampling dates (month and year) of the occurrence dataset (presences, 2293 points) and selected background points below 1500 m (6050 points).



Altitude

Distribution of altitude values of the occurrence dataset (presences, 2293 points) selected background points below 1500 m (6050 points).



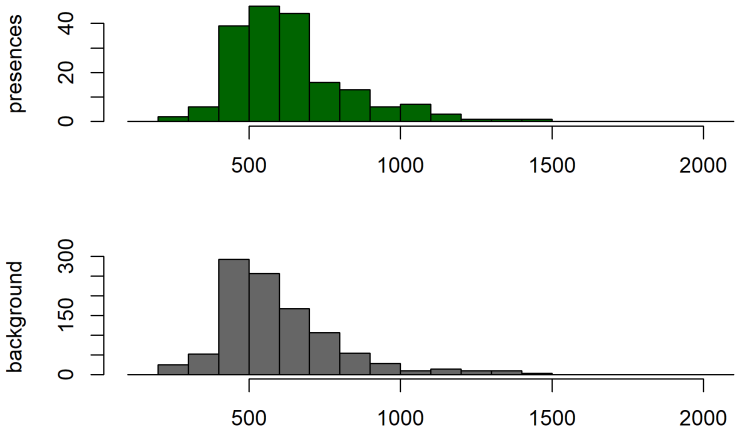
Chlamydiales

Sampling date

Distribution of sampling dates (month and year) of the occurrence dataset (presences, 186 points) and background points (1029 points).

Altitude

Distribution of altitude values of the occurrence dataset (presences, 186 points) and background points (1029 points).



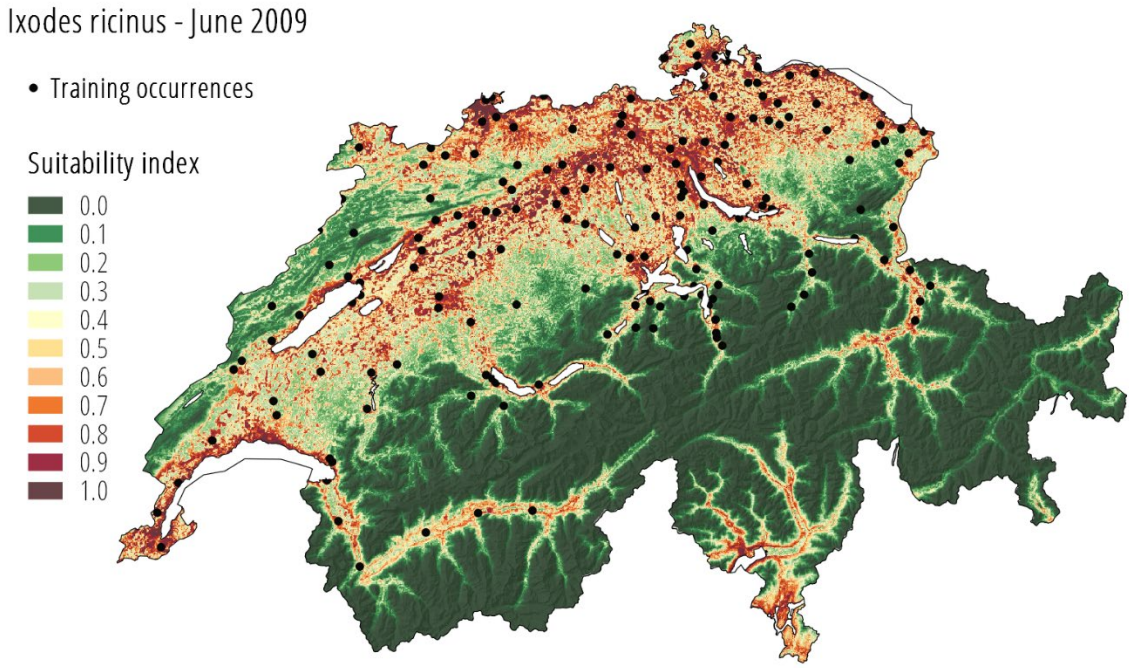
A2.5 Supp. File 5 – *Ixodes ricinus* models

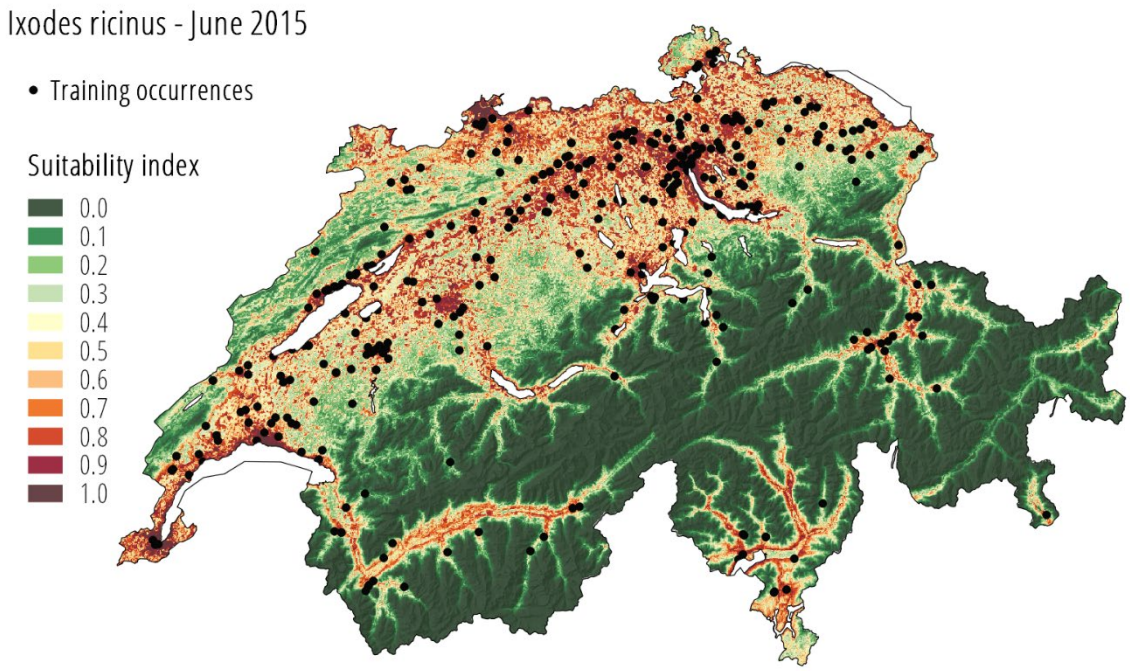
This tab provides the list of all models tested for the distribution of *Ixodes ricinus*. The mean and standard deviation (sd) values over the 20 runs are given for each of the evaluation parameters. ***reg*** is the value of the regularization multiplier. ***features*** indicate the features used (l=linear, lp=linear and product, lq = linear and quadratic, lpq=linear product and quadratic). ***OE_test*** is the omission error on the testing dataset and ***OE_indep*** on the independent dataset. **# coeff** is the number of non-zeros coefficients estimated by the model. The ranks (1-4) correspond to the ranking procedure defined in the method section. The final rank gives the final ranking of the models (1=best model, parameters selected for the final modelling).

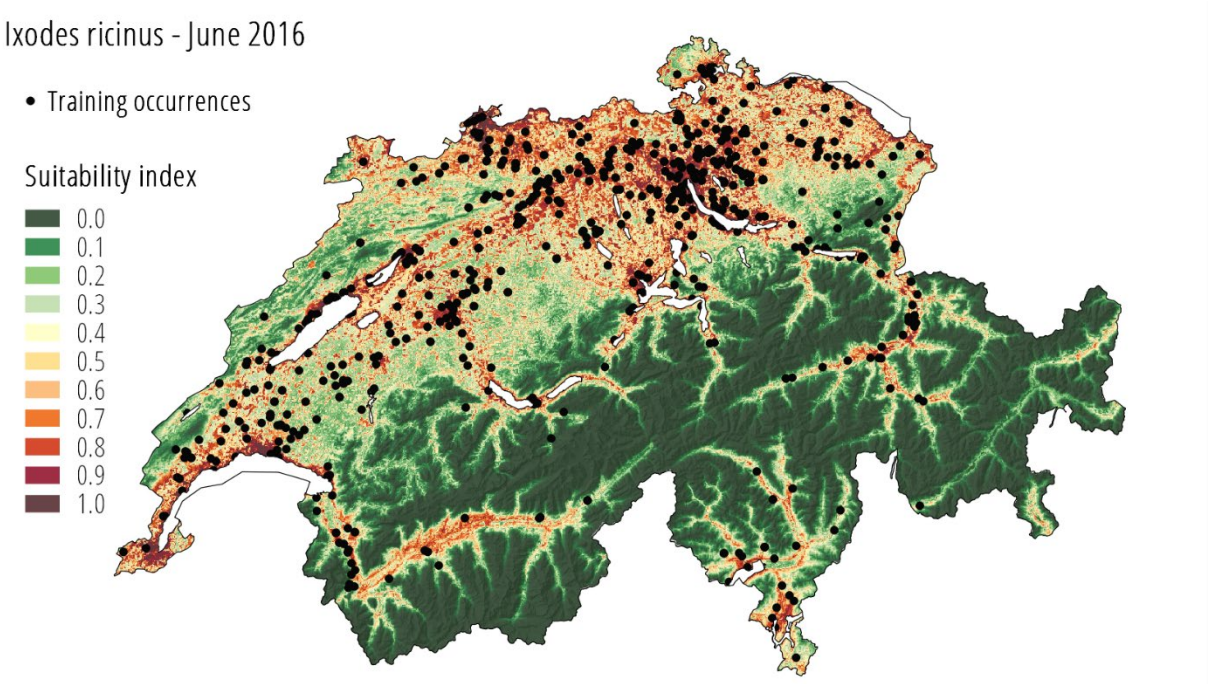
Variables Selection	PCA %variance	reg	features	# env. predictors	rank1 AUC_test	rank2 OE_test	rank3 OE_indep	rank4 #coeffs	ranks sum	final rank	mean AUC_test	mean # coeff.	mean OE_test	mean OE_indep	mean AUC_train	mean BIC	sd AUC_test	sd # coeff	sd OE_test	sd OE_indep	sd AUC_train	sd BIC
corr/VIF	-	1	l	27	48	55	54	33	190	54	0.711	23.600	0.335	0.390	0.717	29766	0.009	1.095	0.046	0.049	0.003	14.130
corr/VIF	-	1	lp	27	41	53	56	55	205	56	0.733	92.550	0.312	0.428	0.752	30235	0.012	6.194	0.049	0.073	0.004	74.669
corr/VIF	-	1	lq	27	38	49	49	40	176	51	0.738	32.600	0.282	0.344	0.742	29724	0.009	1.729	0.058	0.062	0.003	19.955
corr/VIF	-	1	lpq	27	34	44	53	54	185	53	0.747	92.200	0.270	0.377	0.768	30120	0.012	5.187	0.040	0.052	0.004	53.552
corr/VIF	-	5	l	27	49	54	50	13	166	49	0.710	14.950	0.323	0.349	0.715	29662	0.010	1.099	0.045	0.039	0.003	16.042
corr/VIF	-	5	lp	27	46	56	55	37	194	55	0.721	28.350	0.336	0.397	0.728	29696	0.011	2.368	0.044	0.052	0.003	23.684
corr/VIF	-	5	lq	27	42	51	48	27	168	50	0.730	20.950	0.303	0.331	0.732	29622	0.010	1.504	0.049	0.048	0.003	16.430
corr/VIF	-	5	lpq	27	36	52	52	41	181	52	0.741	32.700	0.305	0.360	0.746	29641	0.008	2.638	0.045	0.050	0.003	25.442
PCA	50%	1	l	3	55	2	2	1	60	3	0.663	3.000	0.147	0.161	0.659	29756	0.010	0.000	0.049	0.042	0.004	11.505
PCA	50%	1	lp	3	56	1	1	2	60	4	0.660	5.600	0.140	0.154	0.660	29766	0.009	0.503	0.057	0.057	0.003	14.468
PCA	50%	1	lq	3	52	47	46	3	148	45	0.678	5.700	0.280	0.298	0.678	29656	0.010	0.470	0.053	0.052	0.003	14.073
PCA	50%	1	lpq	3	53	5	29	5	92	18	0.674	7.650	0.213	0.228	0.679	29665	0.011	0.988	0.079	0.083	0.004	14.545
PCA	70%	1	l	6	54	3	15	4	76	8	0.666	6.000	0.185	0.211	0.667	29757	0.008	0.000	0.069	0.067	0.003	9.190
PCA	70%	1	lp	6	51	15	47	21	134	43	0.691	19.100	0.234	0.300	0.703	29725	0.007	0.852	0.048	0.071	0.003	11.890
PCA	70%	1	lq	6	50	13	45	9	117	35	0.709	11.800	0.233	0.272	0.714	29591	0.011	0.410	0.068	0.086	0.004	18.314
PCA	70%	1	lpq	6	47	32	51	29	159	48	0.720	21.750	0.253	0.358	0.729	29588	0.008	1.209	0.065	0.107	0.003	21.109
PCA	80%	1	l	10	44	41	35	8	128	40	0.724	10.000	0.266	0.244	0.723	29658	0.010	0.000	0.059	0.060	0.004	13.431
PCA	80%	1	lp	10	32	36	41	44	153	47	0.749	42.250	0.255	0.258	0.757	29766	0.010	2.403	0.042	0.052	0.003	28.257
PCA	80%	1	lq	10	19	40	38	23	120	38	0.763	19.600	0.265	0.251	0.766	29475	0.011	0.503	0.037	0.039	0.004	25.669
PCA	80%	1	lpq	10	10	10	27	50	97	22	0.780	53.000	0.230	0.224	0.793	29582	0.008	1.522	0.046	0.045	0.003	23.603
PCA	80%	2	l	10	45	42	36	7	130	41	0.723	9.900	0.268	0.247	0.724	29644	0.010	0.308	0.060	0.049	0.003	14.195
PCA	80%	2	lp	10	33	38	39	42	152	46	0.747	33.450	0.262	0.254	0.754	29681	0.007	1.504	0.045	0.048	0.003	15.620
PCA	80%	2	lq	10	23	24	30	22	99	23	0.761	19.250	0.242	0.229	0.765	29447	0.008	0.851	0.058	0.056	0.003	17.258
PCA	80%	2	lpq	10	11	17	34	46	108	29	0.777	44.150	0.238	0.232	0.785	29505	0.009	2.412	0.043	0.044	0.003	26.689
PCA	80%	5	l	10	43	34	31	6	114	33	0.724	9.400	0.254	0.229	0.722	29621	0.010	0.598	0.040	0.029	0.003	15.508
PCA	80%	5	lp	10	39	19	14	19	91	17	0.736	17.150	0.240	0.211	0.739	29605	0.011	1.785	0.048	0.047	0.004	22.917
PCA	80%	5	lq	10	30	18	22	12	82	11	0.755	13.900	0.239	0.219	0.757	29410	0.011	1.210	0.050	0.047	0.004	26.847
PCA	80%	5	lpq	10	22	27	33	26	108	28	0.761	20.050	0.247	0.231	0.764	29410	0.012	1.276	0.057	0.058	0.004	24.385
PCA	90%	2	l	17	31	26	11	18	86	14	0.750	17.000	0.244	0.205	0.750	29575	0.009	0.000	0.040	0.033	0.003	14.684
PCA	90%	2	lp	17	6	9	20	51	86	15	0.788	80.150	0.230	0.219	0.799	29805	0.010	4.171	0.041	0.034	0.004	28.721
PCA	90%	2	lq	17	12	12	37	39	100	25	0.776	31.400	0.232	0.251	0.785	29450	0.009	0.883	0.036	0.040	0.003	18.655
PCA	90%	2	lpq	17	1	7	24	52	84	12	0.804	84.500	0.221	0.223	0.820	29678	0.007	2.911	0.036	0.034	0.002	33.233
PCA	90%	5	l	17	35	22	12	16	85	13	0.746	16.450	0.242	0.207	0.748	29544	0.010	0.759	0.032	0.028	0.004	16.702
PCA	90%	5	lp	17	26	46	32	38	142	44	0.758	30.950	0.274	0.230	0.771	29503	0.012	2.259	0.057	0.053	0.004	31.500
PCA	90%	5	lq	17	17	21	42	31	111	32	0.770	22.050	0.241	0.258	0.777	29381	0.009	0.999	0.055	0.058	0.003	19.958
PCA	90%	5	lpq	17	8	4	26	43	81	10	0.783	34.800	0.209	0.224	0.789	29386	0.007	2.285	0.033	0.037	0.003	25.914
PCA	90%	10	l	17	37	29	21	10	97	20	0.739	12.200	0.252	0.219	0.740	29540	0.007	0.768	0.065	0.052	0.003	14.725
PCA	90%	10	lp	17	40	30	25	15	110	30	0.736	15.400	0.252	0.224	0.745	29531	0.011	1.353	0.050	0.041	0.003	15.650
PCA	90%	10	lq	17	21	33	43	11	108	27	0.763	13.400	0.253	0.263	0.761	29396	0.008	0.883	0.042	0.038	0.003	20.097
PCA	90%	10	lpq	17	25	37	44	14	120	37	0.759	15.150	0.259	0.267	0.764	29391	0.012	1.496	0.051	0.044	0.004	29.106
PCA	95%	5	l	24	20	43	8	30	101	26	0.763	21.950	0.270	0.202	0.773	29436	0.011	0.887	0.046	0.044	0.004	21.012
PCA	95%	5	lp	24	7	31	3	47	88	16	0.783	46.550	0.252	0.187	0.791	29492	0.011	3.605	0.041	0.036	0.004	25.573
PCA	95%	5	lq	24	9	6	5	36	56	1	0.781	26.750	0.220	0.195	0.789	29338	0.009	1.070	0.035	0.032	0.003	23.083
PCA	95%	5	lpq	24	4	16	9	48	77	9	0.795	48.700	0.235	0.203	0.806	29385	0.011	3.294	0.043	0.042	0.004	33.093
PCA	95%	10	l	24	29	48	17	17	111	31	0.756	16.900	0.280	0.214	0.762	29445	0.009	1.165	0.031	0.023	0.003	20.928
PCA	95%	10	lp	24	27	50	28	25	130	42	0.758	19.800	0.286	0.225	0.764	29451	0.010	1.322	0.051	0.046	0.003	15.838
PCA	95%	10	lq	24	16	14	7	20	57	2	0.770	18.100	0.233	0.202	0.777	29335	0.009	0.718	0.047	0.038	0.003	16.012
PCA	95%	10	lpq	24	18	23	10	24	75	7	0.770	19.650	0.242	0.204	0.778	29346	0.009	1.348	0.048	0.043	0.003	17.929
PCA	100%	5	l	69	15	35	4	45	99	24	0.773	42.550	0.255	0.187	0.784	29519	0.010	2.212	0.031	0.030	0.004	27.079
PCA	100%	5	lp	69	3	20	19	53	95	19	0.795	91.750	0.240	0.217	0.811	29721	0.011	5.129	0.047	0.037	0.004	52.228
PCA	100%	5	lq	69	5	8	6	49	68	5	0.790	48.900	0.226	0.201	0.799	29432	0.008	2.654	0.033	0.045	0.003	27.838
PCA	100%	5	lpq	69	2	25	40	56	123	39	0.803	93.800	0.243	0.255	0.824	29622	0.008	4.895	0.029	0.036	0.003	46.856
PCA	100%	10	l	69	24	45	18	28	115	34	0.760	21.700	0.272	0.215	0.764	29462	0.011	1.780	0.047	0.036	0.004	23.307
PCA	100%	10	lp	69	28	39	16	35	118	36	0.757	25.550	0.263	0.211	0.768	29467	0.010	2.911	0.036	0.030	0.003	21.902
PCA	100%	10	lq	69	14	28	23	32	97	21	0.774	23.150	0.248	0.222	0.777	29375	0.009	1.725	0.040	0.030	0.003	21.992
PCA	100%	10	lpq	69	13	11	13	34	71	6	0.774	25.450	0.230	0.210	0.779	29382	0.008	1.959	0.030	0.030	0.004	26.113

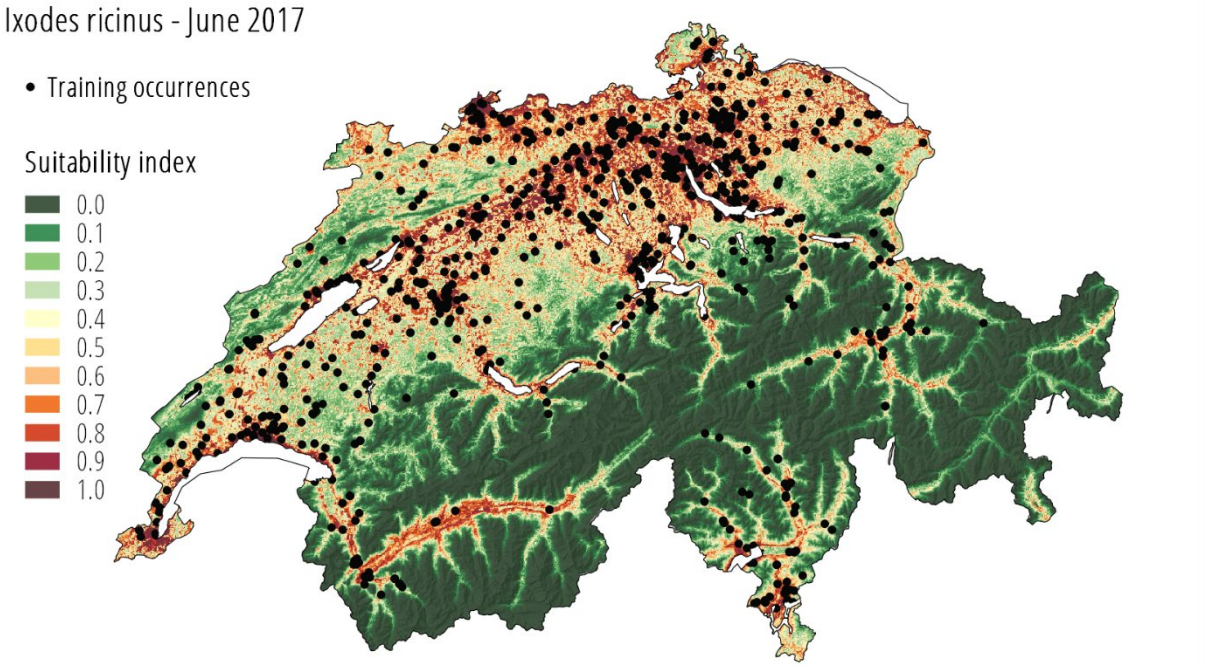
A2.6 Supp. File 6 – *Ixodes ricinus* suitability maps

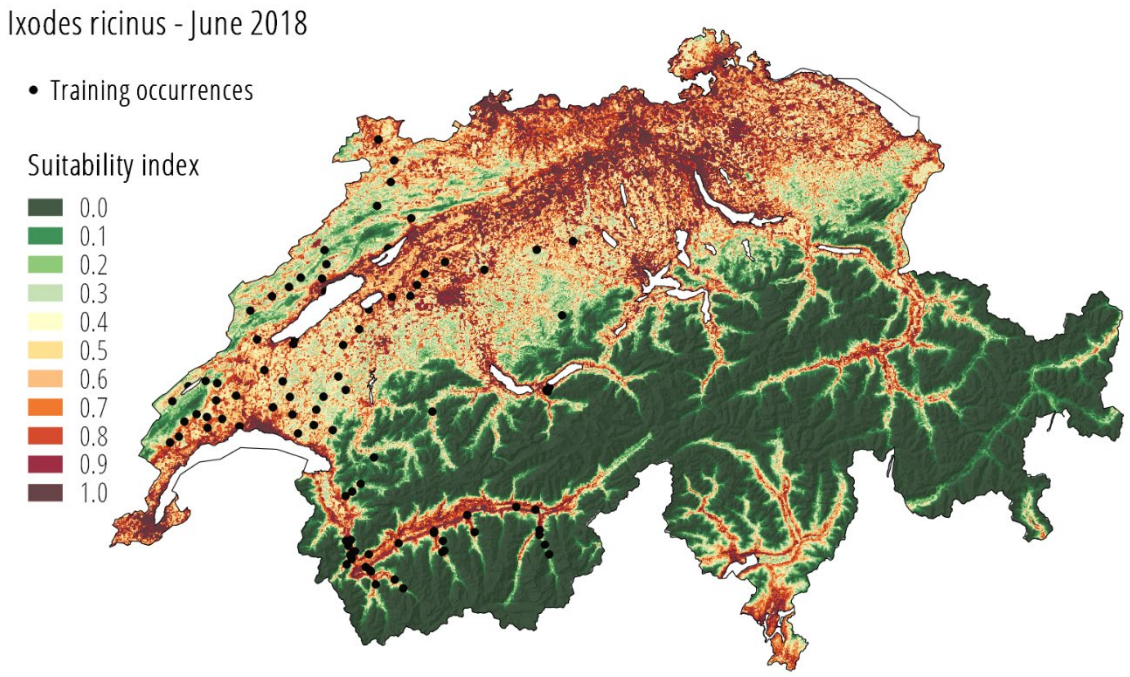
Maps of suitability predicted based on the “best” model presented in the paper.

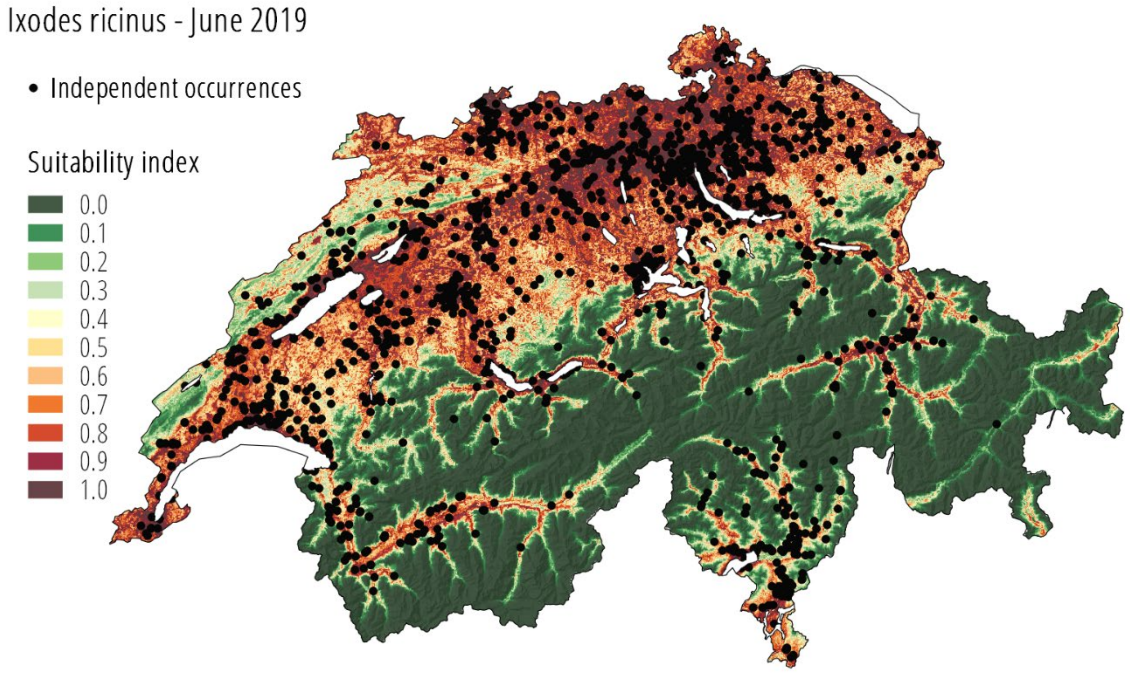





Ixodes ricinus - June 2016
• Training occurrences
Suitability index
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Ixodes ricinus - June 2017
• Training occurrences
Suitability index
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Ixodes ricinus - June 2018
• Training occurrences
Suitability index
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Ixodes ricinus - June 2019
• Independent occurrences
Suitability index
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

A2.7 Supp. File 7 – *Chlamydiales* models

This tab provides the list of all models tested for the distribution of *Chlamydiales*. The mean and standard deviation (sd) values over the 20 runs are given for each of the evaluation parameters. ***reg***. Is the value of the regularization parameter. ***feat.*** indicates the features used (l=linear, lp=linear and product, lq = linear and quadratic, lpq=linear product and quadratic). ***med suit. P 2009*** (resp. 2018) is the median of the suitability predicted on presences points from 2009 (resp. 2018). **med suit. "A" 2009** (resp. 2018) is the median of the suitability predicted at sites where no *Chlamydiales* were found ("absences") in 2009 (resp. 2018). **# coeff** is the number of non-zeros coefficients estimated by the model. The ranks (1-4) correspond to the ranking procedure defined in the method section. The final rank givse the final ranking of the models (1=best model, parameters selected for the final modelling).

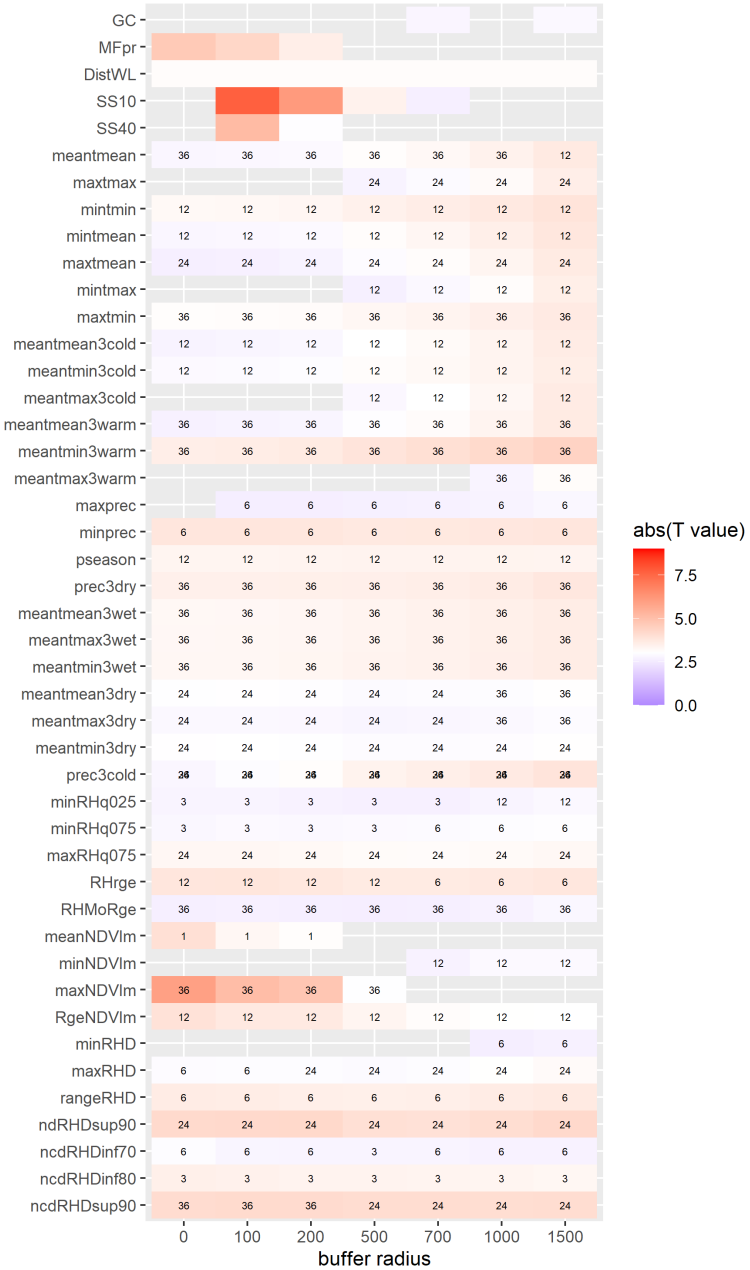
Variables Selection	PCA %var.	reg	feat.	# env	rank1 AUCtest	rank2 diff 2009	rank3 diff 2018	rank4 #coeffs	sum of ranks	final rank	mean AUCtest	mean #coeff.	mean med suit. P 2009	mean med suit. "A" 2009	diff 2009	mean med suit. P 2018	mean med suit."A"2018	diff 2018	mean AUCtrain	mean BIC
corr/VIF	-	1	l	26	15	32	32	26	105	24	0.699	23.65	0.407	0.322	0.085	0.429	0.234	0.195	0.733	1958.07
corr/VIF	-	1	lp	26	21	4	6	53	84	6	0.693	118.20	0.477	0.241	0.235	0.470	0.073	0.397	0.846	3152.50
corr/VIF	-	1	lq	26	1	16	15	35	67	1	0.744	36.20	0.467	0.346	0.121	0.458	0.150	0.308	0.789	2200.49
corr/VIF	-	1	lpq	26	23	3	4	54	84	7	0.690	121.55	0.485	0.242	0.243	0.479	0.058	0.421	0.863	3128.88
corr/VIF	-	5	l	26	32	31	34	20	117	34	0.679	17.90	0.422	0.336	0.085	0.450	0.263	0.186	0.732	1923.77
corr/VIF	-	5	lp	26	5	23	28	40	96	15	0.714	41.70	0.445	0.341	0.104	0.416	0.195	0.221	0.786	2044.63
corr/VIF	-	5	lq	26	2	33	27	25	87	10	0.730	22.60	0.460	0.376	0.084	0.470	0.246	0.223	0.771	1920.40
corr/VIF	-	5	lpq	26	3	24	24	42	93	13	0.726	46.35	0.462	0.359	0.103	0.440	0.192	0.248	0.794	2099.72
PCA	50%	1	l	4	56	52	55	1	164	55	0.584	3.00	0.407	0.358	0.049	0.490	0.424	0.066	0.604	1902.96
PCA	50%	1	lp	4	55	51	56	3	165	56	0.595	5.75	0.413	0.363	0.050	0.467	0.416	0.051	0.602	1916.03
PCA	50%	1	lq	4	54	47	49	4	154	52	0.611	5.95	0.417	0.361	0.056	0.453	0.314	0.139	0.653	1902.24
PCA	50%	1	lpq	4	52	35	53	7	147	50	0.622	8.55	0.426	0.343	0.083	0.422	0.293	0.129	0.650	1913.39
PCA	70%	1	l	6	53	42	54	2	151	51	0.622	4.95	0.414	0.348	0.066	0.490	0.369	0.121	0.640	1901.84
PCA	70%	1	lp	6	51	45	52	15	163	54	0.637	14.15	0.408	0.348	0.060	0.474	0.342	0.132	0.654	1949.90
PCA	70%	1	lq	6	39	41	42	9	131	39	0.673	9.70	0.438	0.369	0.069	0.477	0.307	0.170	0.714	1897.24
PCA	70%	1	lpq	6	35	40	41	21	137	45	0.676	18.20	0.443	0.371	0.072	0.450	0.280	0.170	0.719	1941.72
PCA	80%	1	l	9	49	34	45	6	134	41	0.652	7.90	0.417	0.334	0.083	0.468	0.317	0.151	0.672	1906.33
PCA	80%	1	lp	9	16	27	25	33	101	19	0.697	33.15	0.393	0.299	0.093	0.448	0.205	0.244	0.756	2144.63
PCA	80%	1	lq	9	25	26	31	16	98	16	0.687	15.30	0.416	0.318	0.098	0.446	0.250	0.196	0.729	2013.86
PCA	80%	1	lpq	9	6	30	16	37	89	11	0.708	39.10	0.401	0.314	0.087	0.450	0.148	0.301	0.781	2303.41
PCA	90%	1	l	14	46	38	37	13	134	42	0.665	12.60	0.412	0.333	0.079	0.452	0.273	0.179	0.694	1921.21
PCA	90%	1	lp	14	47	6	17	49	119	35	0.663	79.50	0.437	0.235	0.202	0.393	0.108	0.285	0.790	3111.37
PCA	90%	1	lq	14	27	46	33	29	135	43	0.685	24.85	0.404	0.346	0.058	0.410	0.223	0.187	0.748	2052.67
PCA	90%	1	lpq	14	12	5	8	50	75	2	0.702	87.20	0.441	0.232	0.209	0.417	0.080	0.337	0.815	3053.53
PCA	90%	2	l	14	38	39	39	12	128	37	0.674	12.25	0.418	0.344	0.074	0.455	0.284	0.172	0.692	1916.91
PCA	90%	2	lp	14	24	10	19	45	98	17	0.688	67.65	0.441	0.268	0.173	0.431	0.150	0.281	0.780	2956.55
PCA	90%	2	lq	14	11	48	38	28	125	36	0.703	24.15	0.427	0.371	0.055	0.420	0.245	0.175	0.741	1977.26
PCA	90%	2	lpq	14	34	9	14	47	104	22	0.676	76.10	0.454	0.274	0.180	0.431	0.121	0.310	0.816	2985.31
PCA	90%	5	l	14	43	37	44	11	135	44	0.668	11.55	0.435	0.356	0.080	0.463	0.306	0.157	0.686	1912.04
PCA	90%	5	lp	14	17	14	23	41	95	14	0.697	42.90	0.469	0.327	0.141	0.467	0.213	0.253	0.762	2245.18
PCA	90%	5	lq	14	26	49	43	23	141	49	0.686	21.15	0.439	0.387	0.052	0.440	0.275	0.165	0.744	1949.26
PCA	90%	5	lpq	14	4	13	21	43	81	4	0.715	49.35	0.493	0.352	0.142	0.465	0.207	0.257	0.791	2266.37
PCA	90%	10	l	14	48	36	48	8	140	48	0.663	9.45	0.448	0.368	0.081	0.466	0.324	0.141	0.676	1901.83
PCA	90%	10	lp	14	33	17	35	27	112	32	0.678	23.95	0.477	0.359	0.118	0.470	0.285	0.185	0.728	1990.70
PCA	90%	10	lq	14	30	44	47	18	139	46	0.681	15.65	0.462	0.401	0.061	0.456	0.308	0.148	0.724	1910.89
PCA	90%	10	lpq	14	19	20	36	30	105	25	0.694	26.75	0.481	0.374	0.107	0.474	0.292	0.181	0.753	2000.00
PCA	95%	5	l	19	36	29	46	17	128	38	0.675	15.35	0.435	0.348	0.088	0.425	0.277	0.149	0.700	1924.01
PCA	95%	5	lp	19	41	11	18	46	116	33	0.671	73.50	0.489	0.319	0.170	0.467	0.186	0.281	0.802	2591.90
PCA	95%	5	lq	19	18	50	40	31	139	47	0.695	29.35	0.452	0.401	0.051	0.423	0.253	0.171	0.747	2034.82
PCA	95%	5	lpq	19	28	12	11	48	99	18	0.683	78.90	0.496	0.354	0.142	0.498	0.174	0.324	0.827	2717.13
PCA	95%	10	l	19	40	28	50	14	132	40	0.672	12.90	0.453	0.366	0.088	0.452	0.316	0.137	0.689	1911.28
PCA	95%	10	lp	19	29	15	29	34	107	27	0.682	35.95	0.486	0.363	0.123	0.479	0.277	0.202	0.744	2058.26
PCA	95%	10	lq	19	37	43	51	24	155	53	0.675	21.20	0.471	0.406	0.065	0.430	0.296	0.134	0.735	1938.55
PCA	95%	10	lpq	19	13	25	30	39	107	28	0.700	40.05	0.496	0.394	0.103	0.476	0.279	0.197	0.768	2145.56
PCA	100%	5	l	47	8	19	20	36	83	5	0.704	37.90	0.435	0.323	0.112	0.458	0.177	0.281	0.771	2063.43
PCA	100%	5	lp	47	50	1	2	55	108	29	0.639	211.95	0.674	0.292	0.382	0.583	0.098	0.485	0.918	3663.34
PCA	100%	5	lq	47	7	21	13	44	85	8	0.708	59.40	0.444	0.337	0.107	0.481	0.166	0.315	0.798	2214.48
PCA	100%	5	lpq	47	44	2	1	56	103	21	0.668	219.40	0.680	0.310	0.370	0.582	0.096	0.486	0.922	3676.00
PCA	100%	10	l	47	9	18	26	32	85	9	0.704	30.55	0.473	0.356	0.116	0.482	0.247	0.235	0.754	1975.76
PCA	100%	10	lp	47	42	7	9	52	110	30	0.668	113.05	0.537	0.342	0.196	0.532	0.196	0.336	0.850	2576.08
PCA	100%	10	lq	47	10	22	22	38	92	12	0.703	39.80	0.483	0.378	0.105	0.496	0.240	0.256	0.778	2020.73
PCA	100%	10	lpq	47	31	8	12	51	102	20	0.680	110.85	0.548	0.364	0.184	0.531	0.209	0.322	0.855	2628.10

A2.8 Supp. File 8 - *Chlamydiales* : T-test and selection of variables

For the signification of the acronym names, please refer to Supp. File A2.3.

T-test

For each variable and buffer radius, the heatmap below shows the results of the T-test. Only results that were significant according to the p-value of the T-test are shown (grey area = no significant results). The numbers on the cells indicate the time period considered before sampling date (in number of months) which resulted in the highest T-value for the given combination of variable and buffer radius. Numerical values are available in the following table.



variable	buffer	time period (months)	mean1	sd1	mean0	sd0	P-value	T-value
SS10	B100m		7.76	11.79	16.76	25.282	6.49E-14	7.70
maxNDVIm	P	36	8335.76	530.03	8052.00	883.681	5.65E-09	-5.96
SS40	B100m		62.72	21.58	53.03	32.526	4.12E-07	-5.16
MFpr	P		39.37	30.52	27.86	31.863	4.21E-06	-4.70
meantmin3warm	B1500m	36	12.87	0.95	13.21	1.040	1.52E-05	4.40
ndRHDsup90	B200m	24	21.11	12.51	25.24	10.579	3.29E-05	4.23
ncdRHDsup90	B100m	36	3.12	1.35	3.57	1.328	4.44E-05	4.16
meanNDVIm	P	1	7296.71	940.89	6985.76	1156.039	8.18E-05	-3.99
RgeNDVIm	P	12	6378.89	1537.42	5883.24	1788.431	1.02E-04	-3.94
mintmin	B1500m	12	-4.82	1.25	-4.43	1.277	1.31E-04	3.88
prec3cold	B1500m	24	24.16	11.73	20.70	7.944	1.44E-04	-3.87
prec3cold	B1500m	36	24.16	11.73	20.70	7.944	1.44E-04	-3.87
RHrge	B100m	12	31.09	4.69	32.52	4.758	1.62E-04	3.83
minprec	B1500m	6	2.80	0.97	2.50	1.022	1.62E-04	-3.83
mintmean	B1500m	12	-2.23	1.07	-1.90	1.114	1.87E-04	3.79
prec3dry	B1500m	36	13.99	3.25	13.00	3.443	2.00E-04	-3.77
meantmean	B1500m	12	9.14	1.13	9.47	1.200	2.60E-04	3.70
rangeRHD	B1500m	6	44.68	4.29	45.95	4.292	2.62E-04	3.70
maxtmin	B1500m	36	15.26	1.19	15.61	1.083	2.81E-04	3.69
maxtmean	B1500m	24	18.01	1.21	18.37	1.300	2.90E-04	3.67
meantmean3warm	B1500m	36	18.08	1.19	18.43	1.293	2.90E-04	3.67
meantmax3cold	B1500m	12	2.57	1.04	2.89	1.264	2.97E-04	3.66
meantmean3cold	B1500m	12	-0.36	1.09	-0.03	1.259	3.48E-04	3.62
meantmax3wet	B1500m	36	18.00	6.71	19.84	4.413	3.78E-04	3.61
meantmin3wet	B1500m	36	9.03	5.41	10.51	3.595	3.84E-04	3.61
meantmean3wet	B1500m	36	13.29	5.98	14.93	3.945	3.85E-04	3.61
meantmin3cold	B1500m	12	-3.12	1.20	-2.77	1.354	4.18E-04	3.57
maxtmax	B1500m	24	23.61	1.45	24.02	1.518	4.49E-04	3.55
mintmax	B1500m	12	0.59	1.00	0.88	1.133	4.54E-04	3.55
ncdRHDinf80	P	3	29.71	9.04	27.14	9.929	5.31E-04	-3.51
pseason	B500m	12	47.02	10.59	44.24	8.316	8.06E-04	-3.40
maxRHq075	P	24	87.44	3.05	88.22	2.569	1.15E-03	3.29
maxRHD	B1500m	24	94.83	1.67	95.24	1.330	1.68E-03	3.18
meantmax3warm	B1500m	36	23.81	1.42	24.16	1.498	1.89E-03	3.14
DistWL	B1000m		2554.14	2311.95	3163.80	3062.125	1.89E-03	3.13
meantmean3dry	B1500m	36	4.37	3.72	5.27	3.591	2.44E-03	3.06
meantmin3dry	B1500m	24	7.54	4.17	8.55	4.132	2.53E-03	3.05
minRHq075	B1500m	6	67.80	2.81	67.12	3.235	3.19E-03	-2.97
ncdRHDinf70	P	6	16.00	6.65	14.44	6.435	3.37E-03	-2.96
meantmax3dry	B1500m	36	0.90	3.47	1.71	3.268	3.59E-03	2.94
minNDVIm	B1000m	12	1562.21	1297.28	1866.53	1461.392	4.22E-03	2.89
minRHq025	B1500m	12	56.33	3.66	55.49	3.983	4.66E-03	-2.85
GC	B1500m		0.00	0.00	0.00	0.000	4.76E-03	2.85
RHMoRge	B1500m	36	11.81	1.62	12.17	1.544	5.11E-03	2.83
maxprec	B1500m	6	16.33	6.61	14.89	4.944	5.19E-03	-2.82
minRHD	B1500m	6	49.55	3.37	48.82	3.792	7.75E-03	-2.68

mean1 is the mean of the values for occurrences points, ***mean0*** the mean of the values for background points, ***sd1*** the standard deviation of the values for occurrences points and ***sd0*** the standard deviation of the values for background points.

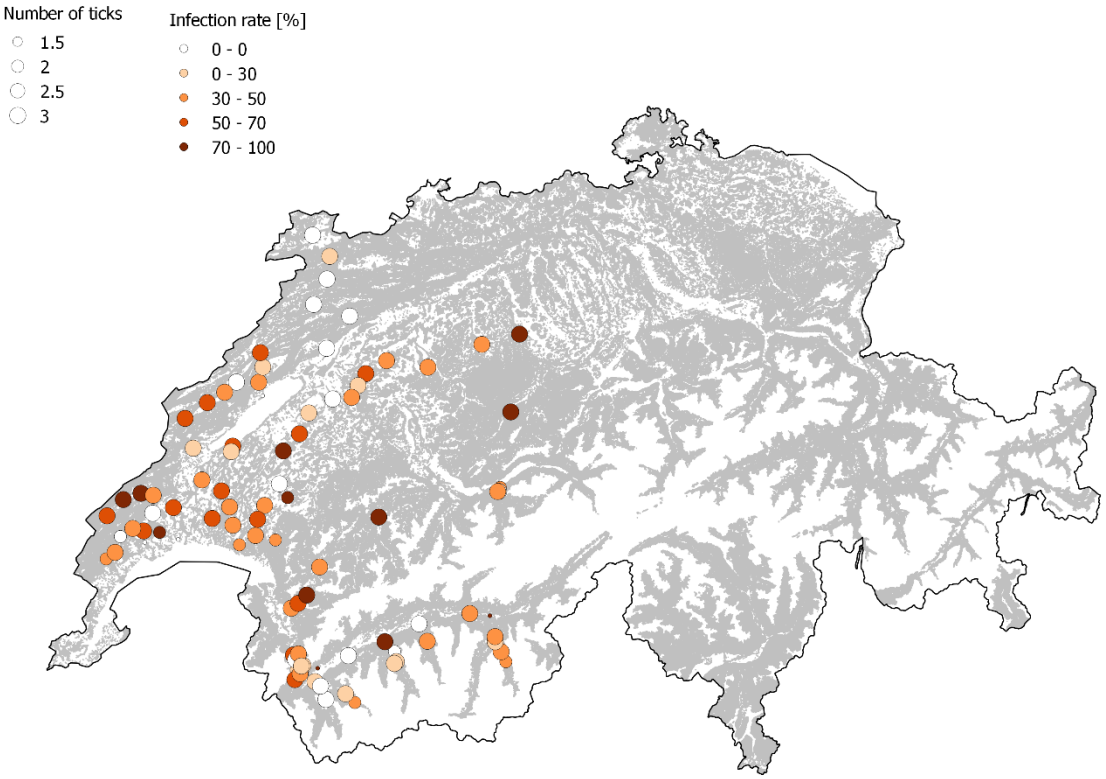
Uncorrelated variables used in the model

1. MFpr_P
2. ncdRHDinf80_P_3
3. ncdRHDinf70_P_6
4. maxRHq075_P_24
5. SS10_B100m
6. SS40_B100m
7. ncdRHDsup90_B100m_36
8. ndRHDsup90_B200m_24
9. DistWL_B1000m
10. GC_B1500m
11. minRHq075_B1500m_6
12. rangeRHD_B1500m_6
13. prec3cold_B1500m_24
14. maxRHD_B1500m_24
15. prec3dry_B1500m_36
16. meantmax3wet_B1500m_36
17. meantmean3dry_B1500m_36

A2.9 Supp. File 9 - Infection rates

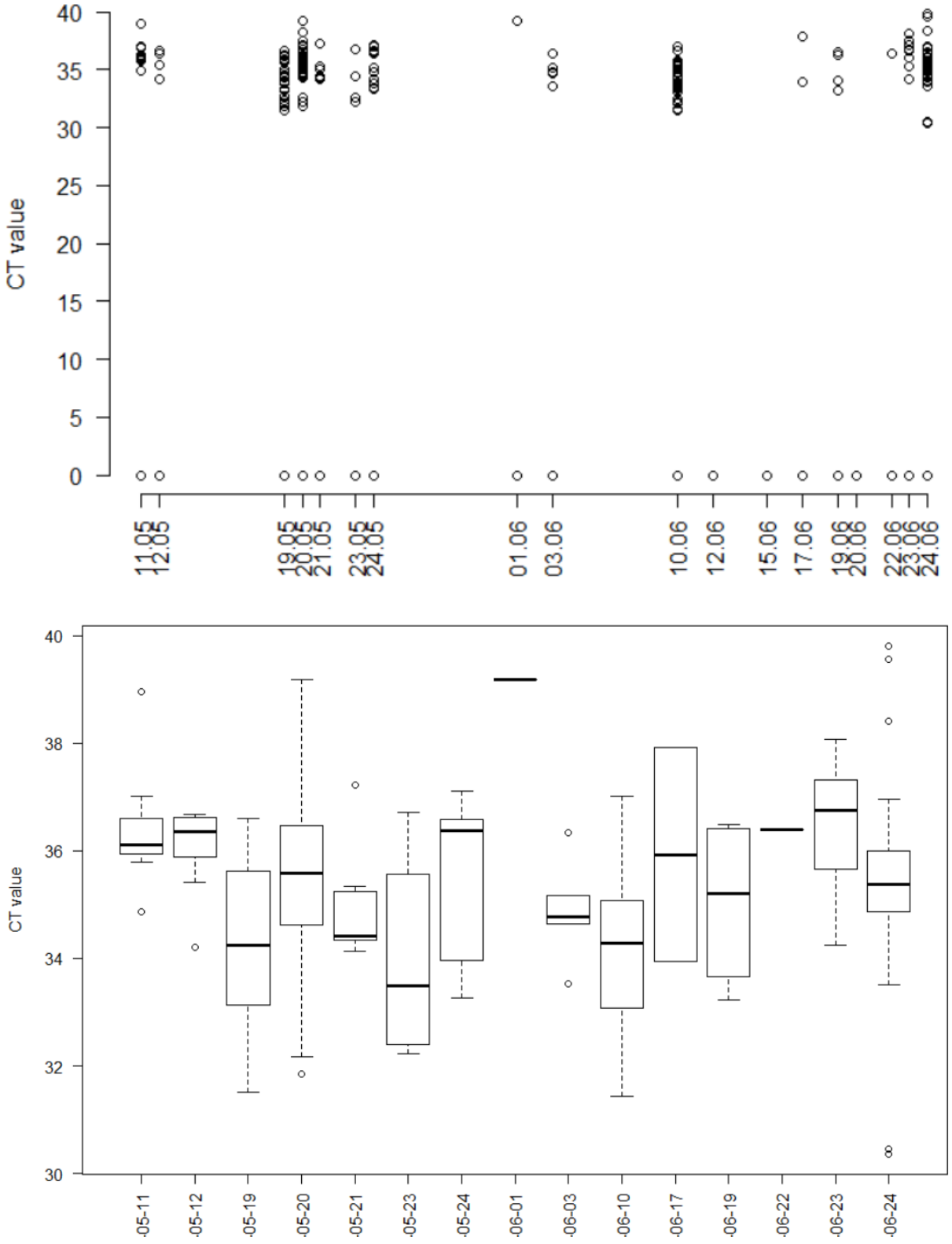
Infection rate prospective campaign: spatial distribution

The infection rate indicates no spatial clustering.



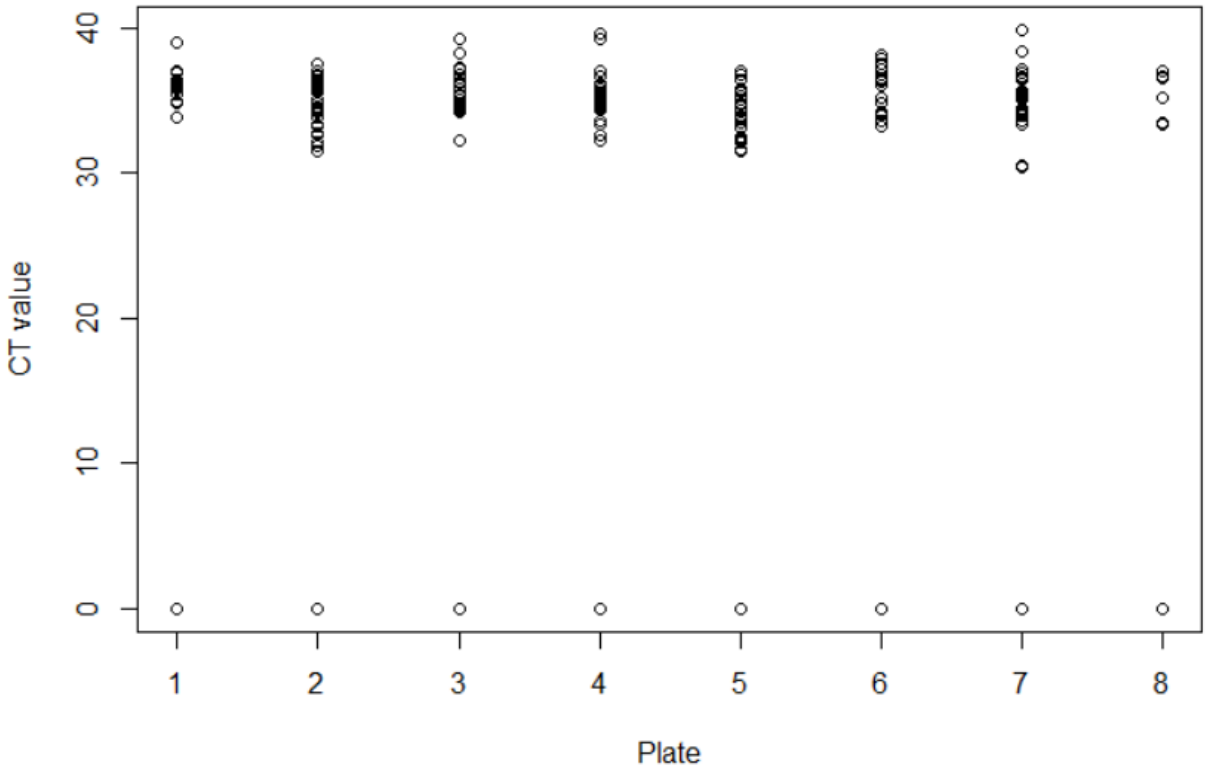
CT value as a function of sampling date

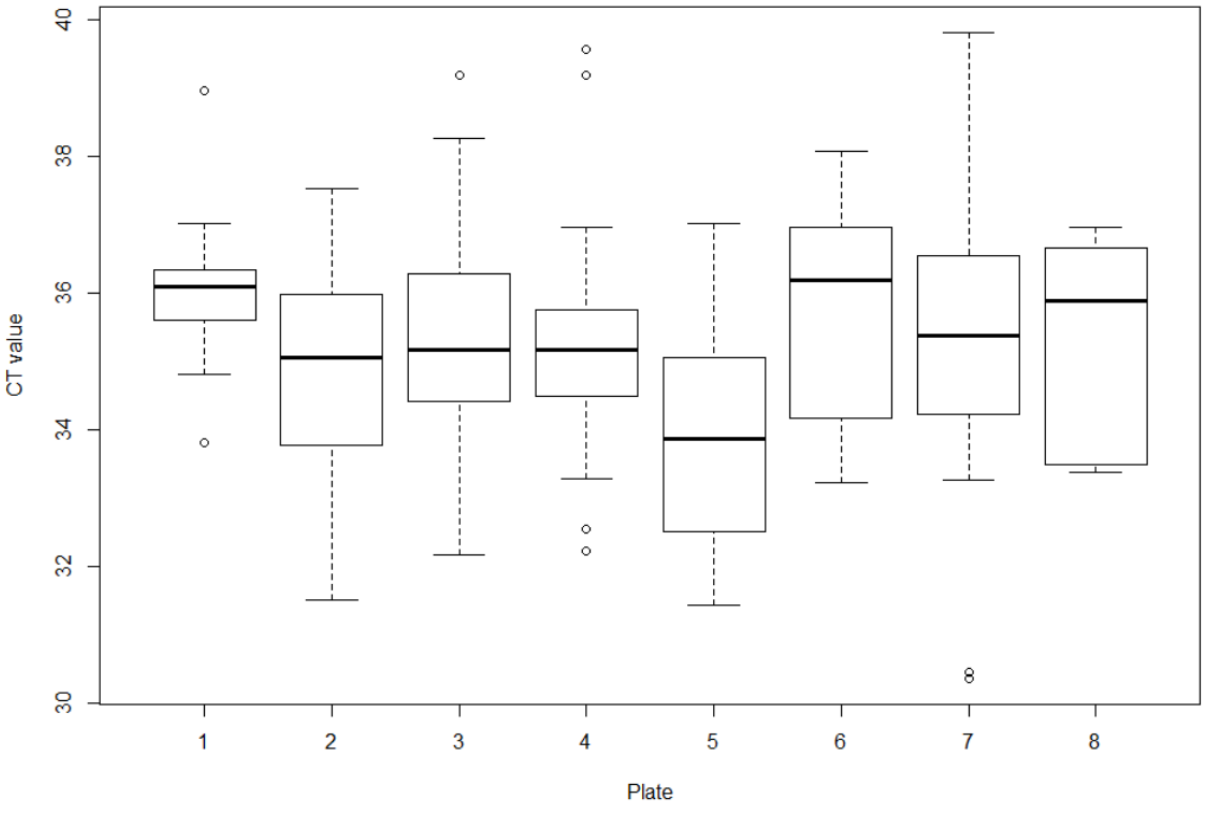
Results indicate no concentration of positive values for a given sampling date or a succession of dates. Negative results are also obtained for each sampling date.



CT value vs Plate

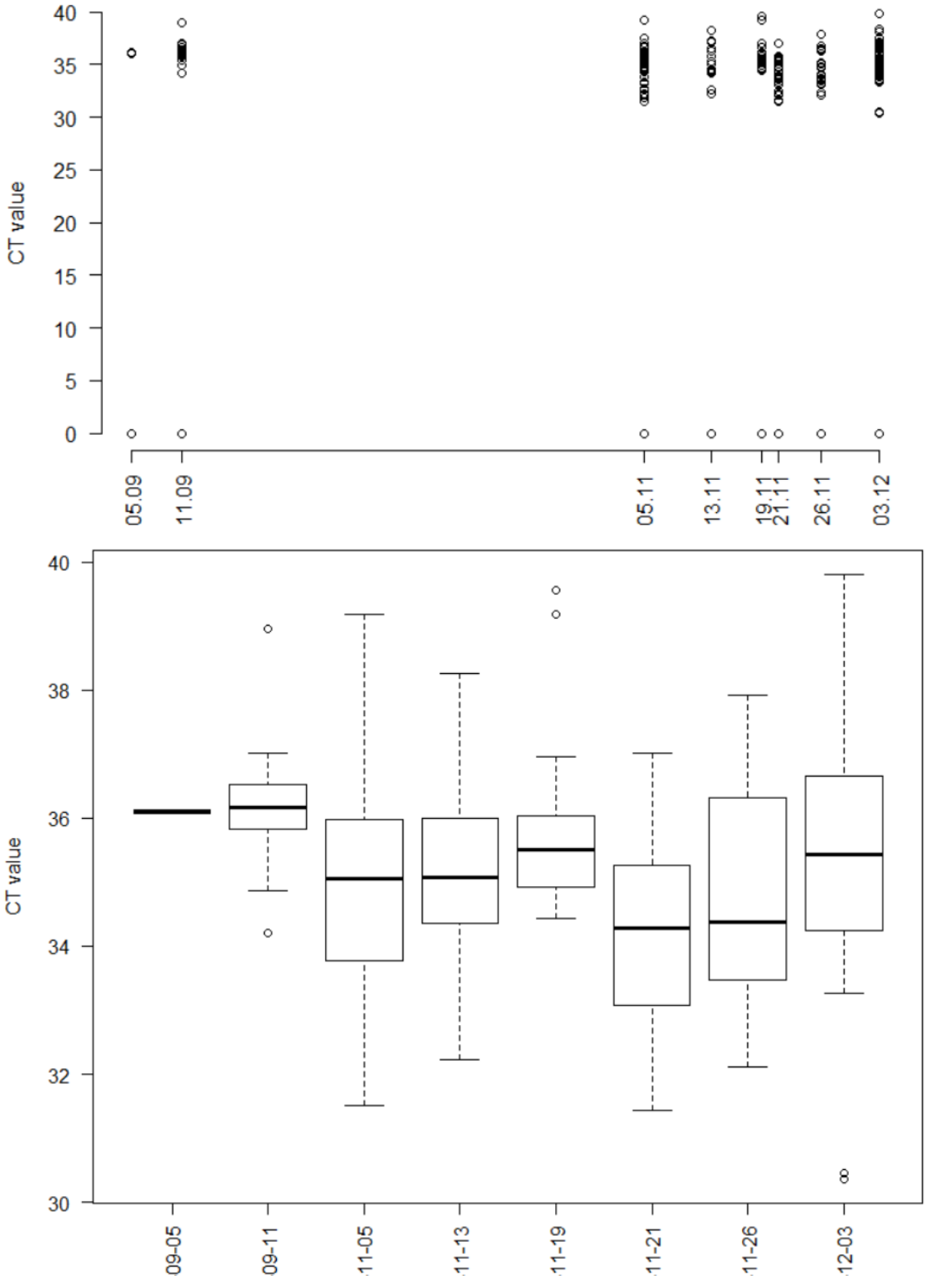
Results indicate no concentration of positive values for some plates. Negative results are also obtained on each plate.





CT value vs DNA Extraction Date

Results indicate no concentration of positive values for a given DNA-extraction date. Negative results are obtained for each extraction date.



A3. First applications of univariate SPAG

A3.1 Ugandan Cattle

Effect of climate change on the spatial distribution of genomic variants involved in the resistance to East Coast Fever in Ugandan cattle

Abstract of a talk and poster presented in ***Evolutionary Biology Meeting***, Marseille, 2015

Estelle Rochat[1*], Elia Vajana[2*], Licia Colli[2], Charles Masembe[3], Riccardo Negrini[2], Paolo Ajmone-Marsan[2], Stéphane Joost[1] and the NEXTGEN Consortium

[1] Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2] Institute of Zootechnics and BioDNA Research Centre, Faculty of Agricultural, Food and Environmental Sciences, Università Cattolica del S. Cuore, Piacenza, Italy

[3] Institute of Environment & Natural Resources, Makerere University, Kampala, Uganda

** These authors contributed equally to this work*

East Coast Fever (ECF) is a major livestock disease caused by *Theileria parva* Theiler, 1904, an emo-parasite protozoan transmitted by the tick *Rhipicephalus appendiculatus* Neumann, 1901. This disease provokes high mortality in cattle populations of East and Central Africa, especially in exotic breeds and crossbreds (Olwoch *et al.*, 2008). Here, we use landscape genomics (Joost *et al.*, 2007) to highlight genomic regions likely involved into tolerance/resistance mechanisms against ECF, and we introduce *SPatial Area of Genotype probability* (*SPAG*) to delimit territories where favourable genotypes are predicted to be present.

Between 2010 and 2012, the NEXTGEN project (nextgen.epfl.ch) carried out the geo-referencing and genotyping (54K SNPs) of 803 Ugandan cattle, among which 496 were tested for *T. parva* presence. Moreover, 532 additional *R. appendiculatus* occurrences were obtained from a published database (Cumming, 1998). Current and future values of 19 bioclimatic variables were also retrieved from the WorldClim database (www.worldclim.org/). In order to evaluate the selective pressure of the parasite, we used MAXENT (Phillips *et al.*, 2006) and a mixed logistic regression (Bates *et al.*, 2014) to model and map the ecological niches of both *T. parva* and *R. appendiculatus*. Then, we used a correlative approach (Stucki *et al.*, 2014) to detect genotypes positively associated with the resulting probabilities of presence and built the corresponding *SPAG*. Finally, we considered bioclimatic predictors representing two different climate change scenarios for 2070 - one moderate and one severe - to forecast the simultaneous shift of both *SPAG* and vector/pathogen niches.

While suitable ecological conditions for *T. parva* are predicted to remain constant, the best environment for the vector is predicted around Lake Victoria. However, when considering future conditions, parasite occurrence is expected to decrease because of the contraction of suitable environments for the tick in both scenarios. Landscape genomics' analyses revealed several markers significantly associated with a high probability of presence of the tick and of the parasite. Among them, we found the marker ARS-BFGL-NGS-113888, whose heterozygous genotype AG showed a positive association. Interestingly, this marker is located close to the gene IRAK-M, an essential component of the Toll-like receptors involved in the immune response against pathogens (Kobayashi *et al.*, 2002). If the implication of this gene into resistance mechanisms against ECF is

confirmed, the corresponding *SPAG* (Figure 1) represents either areas where the variant of interest shows a high probability to exist now, or areas where ecological characteristics are the most favorable to induce its presence under future climatic conditions.

Beyond the results presented here, the combined use of *SPAG* and niche maps could help identifying critical geographical regions that do not present the favourable genetic variant in the present, but where a parasite is likely to expand its range in the future. This may represent a valuable tool to support the identification of current resistant populations and to direct future targeted cross-breeding schemes.

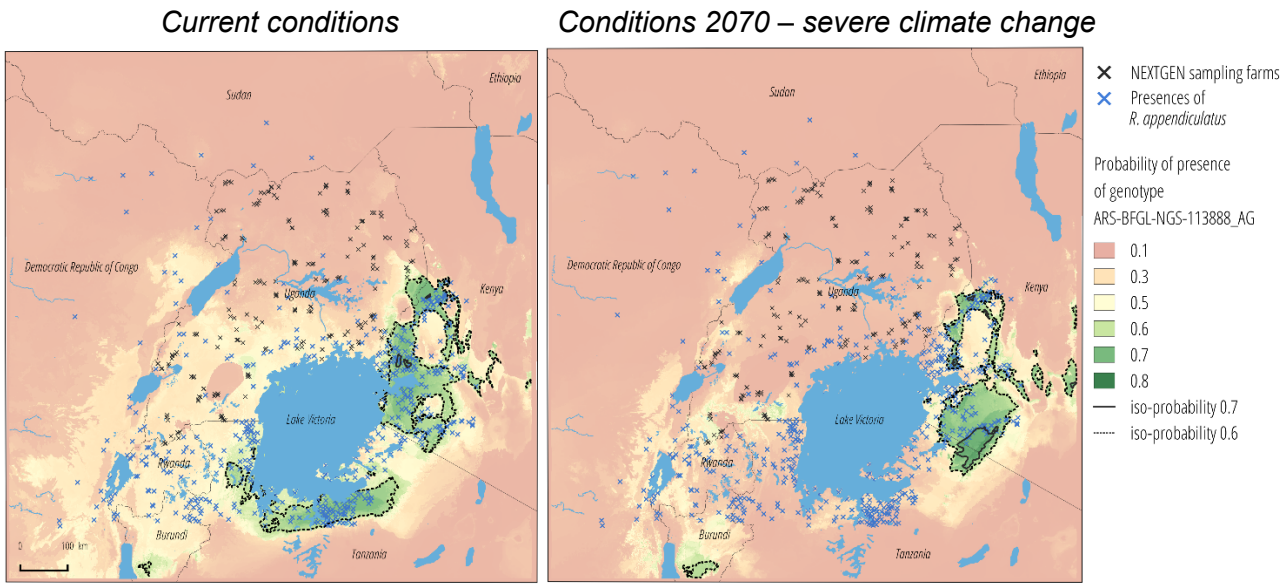


Figure 1 .- SPatial Area of Genotype probability (SPAG)
SPAG for the genotype AG of the SNP "ARS-BFGL-NGS-113888" (ARS-11), highlighting areas where this genotype shows a high probability to be present (Current Conditions), and where it may be distributed in the future (Conditions 2070). As the presence of ARS-11_AG is positively correlated with the presence of the tick R. appendiculatus (alpha = 0.01; Efron pseudo R^2 = 0.074), we can estimate the probability of presence of this genotype also in regions without sampling points and thus without genetic data. At present, the areas of high probability of presence of ARS-11_AG are mainly observed in the North-East and the South of Lake Victoria. However, when considering environmental conditions in 2070 (assuming severe climate change), these areas are expected to be mainly restricted to the North-East of Lake Victoria, where favorable conditions for the presence of R. appendiculatus are supposed to be maintained.

A3.2 Moroccan sheep

Map of genotype frequency change in autochthonous Moroccan sheep breeds due to global warming

Abstract of a talk presented in ***ConGenOmics*** conference, Vairão – Portugal, 2016

Estelle Rochat[1], Kevin Leempoel[1], Elia Vajana[2], Licia Colli[2], Paolo Ajmone-Marsan[2], Stéphane Joost[1] and the NEXTGEN Consortium

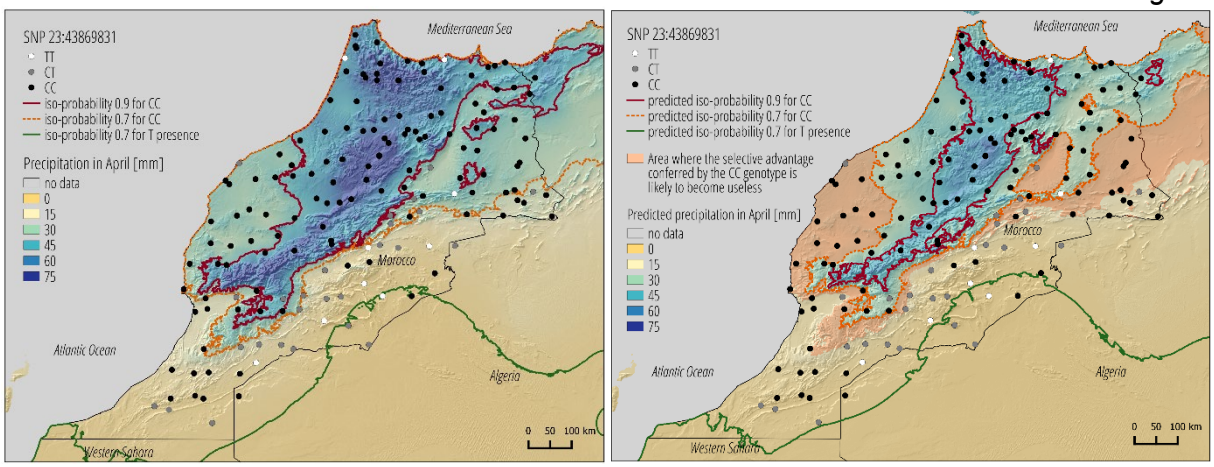
[1] *Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

[2] *Institute of Zootechnics and BioDNA Research Centre, Faculty of Agricultural, Food and Environmental Sciences, Università Cattolica del S. Cuore, Piacenza, Italy*

In developing countries, small ruminants play a key role in the livelihood of farmers and conserving traditional, locally adapted breeds is of essential cultural and economic relief. However, global warming is inducing major changes in habitat conditions and it is therefore important to assess the possible consequences of climate change for these species. We studied the local adaptation of 160 Moroccan sheep based on whole genome sequence (WGS) data from the NEXTGEN project (nextgen.epfl.ch). We used various correlative approaches and identified a locus significantly associated with precipitation. We then built the corresponding Spatial Areas of Genotype Probabilities (SPAG) to delimit the geographic regions where this genotype is currently present and finally used the Worldclim 2070 predictions for severe climate change (scenario GISS-E2-R, rcp 85, http://www.worldclim.com/) to map the predicted SPAG shift in the future.

The locus identified is located in the gene MC5R, which codes for a protein likely involved in lanolin production (Chen *et al.*, 1997). The CC genotype is positively correlated with precipitation whereas the presence of allele T shows a negative association with precipitation. CC could therefore favour the production of lanolin and prevent the lumpy wool disease while T could reduce this production and favour the adaptation to drought. In 2070, precipitation is likely to decrease in the Northern part of Morocco. The corresponding shift in CC's SPAG therefore highlights critical regions where sheep currently show this genotype, but would probably need allele T in order to adapt to forthcoming drought.

Current conditions *Conditions 2070 – severe climate change*



A4. Paper C : Supplementary material

A4.1 Supp. File 1 – Method

Multivariate models

Intersection

From the theory of conditional probabilities, we know that the probability of the simultaneous presence of two genotypes G1 and G2 can be written:

$$p(G_1 \cap G_2) = p(G_1)\, p(G_2|G_1)$$

Formula S1

where $p(G1)$ is the probability of presence of the genotype G1, which can be computed using the univariate model (Formula 4-6), and $p(G_2|G_1)$ is the conditional probability of G2 given G1. The computation of this second probability could be performed with a logistic regression where G1 is integrated as a covariate in the univariate model for G2 (Formula S2).

$$p(G_2|G_1) = \frac{e^{\beta_0+\beta_1 x_2+\beta_2 G_1}}{1+e^{\beta_0+\beta_1 x_2+\beta_2 G_1}}$$

Formula S2

where x2 is the environmental variable associated with G2 and β_0, β_1 and β_2 are the parameters of the logistic regression. However, as we would like to use this model to predict the probability of presence of the genotypes for any point of the region of interest, i.e also where values of G1 are not know, we suggest to estimate G1 by $p(G1)$ which can be computed for the entire territory from the univariate SPAG of G1. We therefore approximated $p(G_2|G_1)$ by $p(G_2|p(G_1))$ with the logistic regression in formula S3.

$$p(G_2|G_1) \approx \frac{e^{\beta_0+\beta_1 x_2+\beta_2 p(G_1)}}{1+e^{\beta_0+\beta_1 x_2+\beta_2 p(G_1)}}$$

Formula S3

The final SPAG $p(G_1 \cap G_2)$ can therefore be computed in three steps:

1. compute the SPAG corresponding to $p(G1)$ across the entire territory, using an univariate model with the environmental variable x1 associated with G1.
2. compute $p(G_2|G_1)$ using a logistic model with the G2 as the dependent variable, the environmental variable *x2* as the independent variable and the probability of presence *p(G1)* as a covariate (Formula S3).
3. multiply the results of $p(G1)$ obtained in step 1 with *p(G2|G1)* computed in step 2 to derive the final $p(G_1 \cap G_2)$.

Using the associative property of the intersection (i.e. $p(G_1 \cap G_2 \cap G_3) = p(G_3 \cap (G_1 \cap G_2))$), the procedure above can be extended to compute the probability of simultaneous presence of n genotypes of interest:

4. compute $p(G_3|(G_1 \cap G_2))$ using a logistic model with G3 as the dependent variable, the environmental variable x3 as the independent variable and the probability $p(G_1 \cap G_2)$ computed in step 3 as a covariate.

5. multiply the results of $p(G_1 \cap G_2)$ obtained in step 3 with $p(G_3|(G_1 \cap G_2))$ computed in step 4 to derive $p(G_1 \cap G_2 \cap G_3)$.

6. compute $p(G_4|(G_1 \cap G_2 \cap G_3))$ using a logistic model with G4 as the dependent variable, the environmental variable x4 as the independent variable and the probability $p(G_1 \cap G_2 \cap G_3)$ computed in step 5 as a covariate.

7. multiply the results of $p(G_1 \cap G_2 \cap G_3)$ obtained in step 5 with $p(G_1 \cap G_2 \cap G_3)$ computed in step 6 to derive $p(G_1 \cap G_2 \cap G_3 \cap G_4)$.

8. Carry on until obtaining $p(G_1 \cap G_2 \cap G_3 \cap G_4 \cap \ldots\ldots \cap G_n)$ for *n* genotypes of interest.

Note that this approach allows for the integration of adaptive genotypes associated with various environmental variables since the environmental variable *x1* used in step 1 to compute *p(G1)* can be different from the variable x2, x3, x4, etc. used in the following steps.

We implemented this recursive approach to build generalised intersection models predicting the simultaneous presence of *n* genotypes of interest as a R function *I-spag* available at : https://github.com/estellerochat/SPAG.

Union

To compute the probability of presence of at least one of the adaptive variant, we use the exclusion-inclusion principle, from which the probability of presence of genotypes *G1 OR G2* can be written:

$$p(G_1 \cup G_2) = p(G_1) + p(G_2) - p(G_1 \cap G_2)$$

Formula S4

where *p(G1)* and *p(G2)* can be computed with univariate SPAGs and *p(G1 ∩ G2)* using the intersection SPAG. For three genotypes, this becomes

$$p(G_1 \cup G_2 \cup G_3) = p(G_1) + p(G_2) + p(G_3) - p(G_1 \cap G_2) - p(G_1 \cap G_3) - p(G_2 \cap G_3) + p(G_1 \cap G_2 \cap G_3)$$

Formula S5

The generalisation for *n* genotypes is given by formula S6, and can be computed using the univariate and intersection models. This general case was implemented as an R function *U-spag* available at https://github.com/estellerochat/SPAG.

$$p\left(\bigcup_{i=1}^{n} G_i\right) = \sum_{i=1}^{n} p(G_i) - \sum_{i<j} p\left(G_i \bigcap G_j\right) + \sum_{i<j<k} p\left(G_i \bigcap G_j \bigcap G_k\right) + \ldots + (-1)^{n-1} p\left(\bigcap_{i=1}^{n} G_i\right)$$

$$\Rightarrow p\left(\bigcup_{i=1}^{n} G_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k < n} P(G_{i_1} \cap G_{i_2} \cap \ldots \cap G_{i_k})\right)$$

Formula S6

K-Percentage

To explain the K-Percentage method, we start with an example: if we have three genotypes, and would like to know the probability to carry at least 50% of them, this would be equivalent to the probability to carry at least two of the three genotypes. This can be expressed with Formula S7, which is the union of the probabilities of simultaneously carrying a combination of two genotypes chosen among the three.

$$p(50\%\, G_1, G_2, G_3) = p\big((G_1 \cap G_2) \cup (G_1 \cap G_3) \cup (G_2 \cap G_3)\big)$$

Formula S7

By developing the union operators and summarising the results, we obtain:

$$p(50\%\, G_1, G_2, G_3) = p(G_1 \cap G_2) + p(G_1 \cap G_3) + p(G_2 \cap G_3) - 3p(G_1 \cap G_2 \cap G_3)$$

By generalizing this approach, the probability to carry at least K% of *n* adaptive variant can be written:

$$p(K\%\, G_{i=1\ldots n}) = p\left(\bigcup_{i=1}^{n} \bigcap_{1 \le i_1 < i_2 < \cdots < i_{(K\% * n+1)}} (G_{i_1} \cap G_{i_2} \cap \ldots \cap G_{i_k})\right)$$

Formula S8

Again, this can be computed using univariate and intersection models and the general case was implemented as an R function *K-spag* available at https://github.com/estellerochat/SPAG.

Selection of training samples

In the cross-validation procedure, training individuals were not randomly selected, but were chosen such to represent the entire distribution of the environmental variable of interest in the study area. We thus retrieved the maximum and minimum value of the environmental variable at the sampled sites and divided this range into N uniform intervals, where N corresponds to the number of training individuals (25% of the total number of individuals). Since individuals are not necessarily sampled uniformly along the range of an environmental variable, we can obtain some intervals without any individuals and others with more than one individual. We therefore selected randomly one individual in each interval where individuals were present, and completed the training set with individuals randomly selected from all remaining individuals. We used this training set to calculate the SPAG.

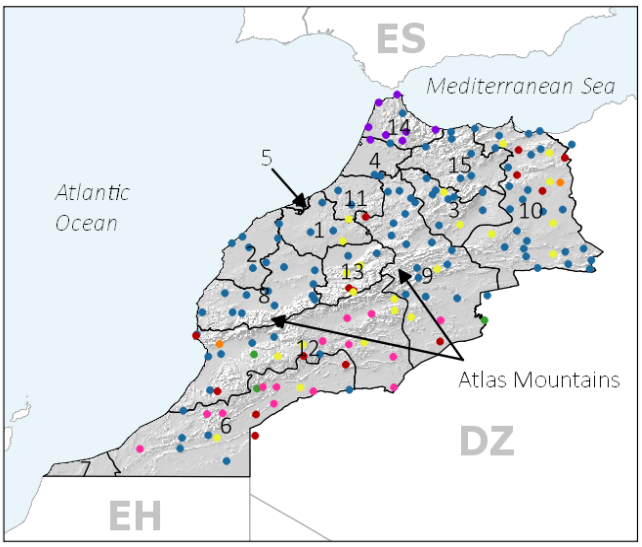
A4.2 Supp. File 2 – CDPOP simulation parameters

looptime	300
cdclimgentime	0
matemoveno	1
matemoveparA	0
matemoveparB	0
matemoveparC	0
matemovethresh	25max
sexans	Y
Freplace	Y
Mreplace	Y
philopatry	N
multiple_paternity	N
selfans	N
Fdispmoveno	1
FdispmoveparA	0
FdispmoveparB	0
FdispmoveparC	0
Fdispmovethresh	25max
Mdispmoveno	1
MdispmoveparA	0
MdispmoveparB	0
MdispmoveparC	0
Mdispmovethresh	25max
offno	2

Femalepercent	50
Equalsexratio-oBirth	N
TwinningPercent	0
popModel	exp
r	1
K_env	0
subpopmortperc	0
muterate	0
mutationtype	random
loci	50
intgenesans	random
allefreqfilename	N
alleles	2
mtdna	N
startGenes	0
cdevolveans	M_X3_L3_A2_ModelX
startSelection	0
betaFile_selection	see Figure 1 (main text)
epistasis	N
epigeneans	N
startEpigene	0
betaFile_epigene	N
cdinfect	N
transmissionprob	0

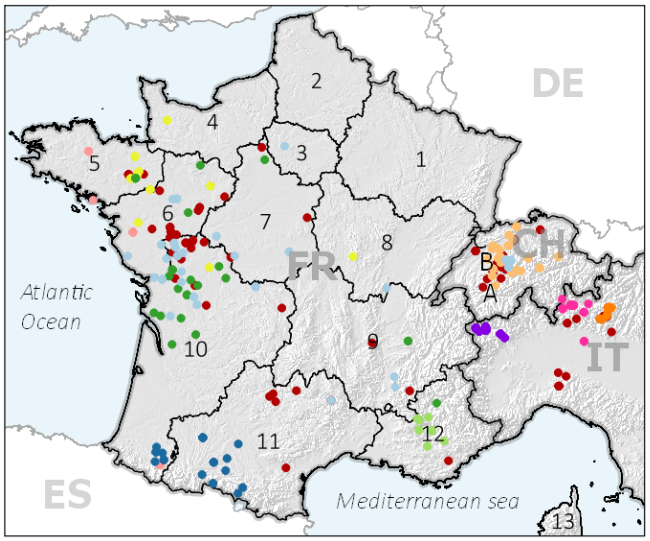
A4.3 Supp. File 3 – Genetic data

A - Moroccan Dataset



Breeds
- Atlas
- Barcha
- Black
- Draa
- Ghazalia
- local populations
- Northern

B - European Dataset



Breeds
- Alpine
- Angora
- Bionda dell'Adamello
- Boer
- Fosses
- Orobica
- Poitevine
- Provencale
- Pyrenean
- Saanen
- Valdostana

A4.4 Supp. File 4 – Bioclimatic data

BIO1 Annual Mean Temperature
BIO2 Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 Isothermality (BIO2/BIO7) (* 100)
BIO4 Temperature Seasonality (standard deviation *100)
BIO5 Max Temperature of Warmest Month
BIO6 Min Temperature of Coldest Month
BIO7 Temperature Annual Range (BIO5-BIO6)
BIO8 Mean Temperature of Wettest Quarter
BIO9 Mean Temperature of Driest Quarter
BIO10 Mean Temperature of Warmest Quarter
BIO11 Mean Temperature of Coldest Quarter
BIO12 Annual Precipitation
BIO13 Precipitation of Wettest Month
BIO14 Precipitation of Driest Month
BIO15 Precipitation Seasonality (Coefficient of Variation)
BIO16 Precipitation of Wettest Quarter
BIO17 Precipitation of Driest Quarter
BIO18 Precipitation of Warmest Quarter
BIO19 Precipitation of Coldest Quarter

A4.5 Supp. File 5 – Results logistic regressions Morocco

CHR=Chromosome, POS=Position in base pairs, GF=Genotype frequency, missGeno=Missing genotype frequency
G=Gscore, qG=p-value of Gscore corrected for FDR, W=Wald score, pW=p-value of Wald score, qW=p-value of Wald score corrected for FDR.

β0, β1=coefficients of the univariate logistic regression

Marker	CHR	POS	GF	missGeno	G	qG	W	pW	qW	β0	β1	Gene CHIR 1.0
ss1281060258_TT	6	12'174'332	17.39	0	27.93	0.046	19.97	7.89E-06	0.920	1.85	-1.26	
ss1281060321_GG	6	12'178'287	81.37	0	27.70	0.047	19.72	8.94E-06	0.920	1.90	-1.27	
ss1281060376_CC	6	12'180'068	81.37	0	27.70	0.047	19.72	8.94E-06	0.920	1.90	-1.27	
ss1281060384_CC	6	12'180'879	81.37	0	27.70	0.047	19.72	8.94E-06	0.920	1.90	-1.27	
ss1281060395_GG	6	12'181'286	81.37	0	27.70	0.047	19.72	8.94E-06	0.920	1.90	-1.27	
ss1281060520_CC	6	12'187'316	80.75	0	27.93	0.046	19.97	7.89E-06	0.920	1.85	-1.26	
ss1281061829_TT	6	12'242'353	28.57	0.01	32.02	0.016	24.60	7.06E-07	0.920	-0.60	1.06	
ss1281061965_CC	6	12'253'612	51.55	0	33.91	0.009	26.33	2.87E-07	0.824	0.08	-1.05	
ss1281061973_TT	6	12'254'244	28.57	0	37.01	0.005	25.82	3.74E-07	0.824	-1.22	1.28	
ss1281061973_AA	6	12'254'244	43.48	0	30.64	0.027	24.57	7.18E-07	0.920	-0.32	-0.99	
ss1281061986_AA	6	12'254'603	51.55	0	33.91	0.009	26.33	2.87E-07	0.824	0.08	-1.05	
ss1281062148_GG	6	12'254'883	51.55	0	33.91	0.009	26.33	2.87E-07	0.824	0.08	-1.05	
ss1281062154_CC	6	12'255'024	49.07	0	32.73	0.012	25.74	3.91E-07	0.824	-0.04	-1.03	
ss1281062169_AA	6	12'255'763	28.57	0.01	36.68	0.005	25.61	4.19E-07	0.824	-1.21	1.28	
ss1281062169_TT	6	12'255'763	43.48	0.01	29.79	0.035	24.03	9.50E-07	0.920	-0.30	-0.98	
ss1281062197_AA	6	12'256'813	28.57	0	37.01	0.005	25.82	3.74E-07	0.824	-1.22	1.28	
ss1281062197_GG	6	12'256'813	43.48	0	28.52	0.045	23.22	1.45E-06	0.920	-0.31	-0.95	
ss1281062229_GG	6	12'259'477	27.95	0.01	35.18	0.008	24.81	6.34E-07	0.920	-1.22	1.26	
ss1281062229_AA	6	12'259'477	43.48	0.01	27.66	0.047	22.62	1.97E-06	0.920	-0.31	-0.94	
ss1281062231_AA	6	12'259'667	43.48	0	28.52	0.045	23.22	1.45E-06	0.920	-0.31	-0.95	
ss1281062231_TT	6	12'259'667	28.57	0	37.01	0.005	25.82	3.74E-07	0.824	-1.22	1.28	
ss1281062238_TT	6	12'261'392	54.04	0.01	28.31	0.046	22.77	1.83E-06	0.920	0.22	-0.95	
ss1281062313_AA	6	12'275'892	19.88	0	27.83	0.046	20.05	7.55E-06	0.920	-1.79	1.23	
ss1281062317_AA	6	12'276'168	21.74	0	38.74	0.004	24.77	6.45E-07	0.920	-1.80	1.51	
ss1281062325_AA	6	12'276'649	27.95	0	30.39	0.027	22.64	1.96E-06	0.920	-1.19	1.13	
ss1281062327_TT	6	12'276'965	27.95	0	30.39	0.027	22.64	1.96E-06	0.920	-1.19	1.13	
ss1281062340_AA	6	12'277'790	27.95	0	30.39	0.027	22.64	1.96E-06	0.920	-1.19	1.13	
ss1281062347_TT	6	12'278'387	27.33	0.01	28.72	0.042	21.62	3.33E-06	0.920	-1.20	1.11	
ss1281062458_TT	6	12'285'400	21.12	0.01	36.13	0.006	23.52	1.23E-06	0.920	-1.79	1.47	
ss1281062466_CC	6	12'285'545	21.74	0	38.74	0.004	24.77	6.45E-07	0.920	-1.80	1.51	
ss1281062471_TT	6	12'285'617	21.74	0	38.74	0.004	24.77	6.45E-07	0.920	-1.80	1.51	
ss1281062473_AA	6	12'285'628	27.95	0	30.39	0.027	22.64	1.96E-06	0.920	-1.19	1.13	
ss1281062684_TT	6	12'298'321	31.68	0	27.92	0.046	22.58	2.01E-06	0.920	0.23	-0.94	
ss1370321332_GG	13	43'436'394	10.56	0	29.02	0.042	17.33	3.14E-05	0.920	-3.14	1.80	
ss1370321334_GG	13	43'436'661	10.56	0.01	28.92	0.042	17.25	3.28E-05	0.920	-3.13	1.80	

ss1370321340_CC	13	43'436'953	11.80	0	28.08	0.046	17.71	2.57E-05	0.920	-2.84	1.64	
ss1370321349_AA	13	43'437'164	11.80	0	28.08	0.046	17.71	2.57E-05	0.920	-2.84	1.64	
ss1370321350_GG	13	43'437'201	11.80	0	28.08	0.046	17.71	2.57E-05	0.920	-2.84	1.64	
ss1370321365_CC	13	43'438'109	11.80	0	28.08	0.046	17.71	2.57E-05	0.920	-2.84	1.64	
ss1370321383_AA	13	43'438'732	11.80	0	28.08	0.046	17.71	2.57E-05	0.920	-2.84	1.64	
ss1382049031_CC	24	19'436'980	76.40	0	34.75	0.008	25.60	4.21E-07	0.824	1.55	1.29	
ss1382126510_AG	24	25'852'900	43.48	0.01	33.49	0.011	26.42	2.74E-07	0.824	-0.32	-1.07	DSG4
ss1382126537_CT	24	25'854'278	39.13	0	34.54	0.008	26.92	2.12E-07	0.824	-0.32	-1.07	DSG4
ss1382126564_CT	24	25'855'578	39.13	0	34.54	0.008	26.92	2.12E-07	0.824	-0.32	-1.07	DSG4
ss1382126637_AG	24	25'860'754	38.51	0	34.75	0.008	27.04	1.99E-07	0.824	-0.29	-1.07	DSG4
ss1382166785_CC	24	28'799'029	11.80	0.01	31.08	0.022	18.46	1.73E-05	0.920	-2.97	1.79	
ss1382166820_AA	24	28'802'199	11.80	0	31.32	0.021	18.57	1.64E-05	0.920	-2.97	1.79	
ss1382166868_GG	24	28'806'275	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382166899_AA	24	28'807'953	12.42	0.01	29.11	0.042	18.38	1.81E-05	0.920	-2.74	1.63	
ss1382166919_TT	24	28'809'142	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382166921_AA	24	28'809'151	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382166933_TT	24	28'809'975	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382166966_GG	24	28'811'723	12.42	0.01	28.70	0.042	18.13	2.07E-05	0.920	-2.74	1.62	
ss1382166968_TT	24	28'811'785	11.80	0.01	29.05	0.042	17.88	2.36E-05	0.920	-2.86	1.68	
ss1382166994_CC	24	28'813'854	13.66	0	32.00	0.016	19.68	9.17E-06	0.920	-2.65	1.67	
ss1382167001_TT	24	28'814'588	13.04	0.03	30.38	0.027	19.05	1.28E-05	0.920	-2.63	1.64	
ss1382167006_CC	24	28'815'234	13.66	0.01	31.90	0.017	19.60	9.54E-06	0.920	-2.64	1.66	
ss1382167049_CC	24	28'818'344	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382167052_TT	24	28'818'565	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382167091_CC	24	28'821'168	13.04	0	31.23	0.022	19.18	1.19E-05	0.920	-2.73	1.69	
ss1382167099_TT	24	28'821'948	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382167101_CC	24	28'822'132	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	
ss1382167134_TT	24	28'825'056	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	CDH2
ss1382167154_GG	24	28'827'375	12.42	0	28.80	0.042	18.20	1.99E-05	0.920	-2.75	1.63	CDH2
ss1382167240_TT	24	28'833'253	12.42	0	27.87	0.046	17.91	2.31E-05	0.920	-2.72	1.58	CDH2
ss1382167247_GG	24	28'833'762	12.42	0	27.87	0.046	17.91	2.31E-05	0.920	-2.72	1.58	CDH2
ss1382188860_TT	24	30'566'869	2.48	0.01	27.99	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	KCTD1
ss1382188926_CC	24	30'574'708	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382188931_GG	24	30'575'137	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382188951_AA	24	30'576'834	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382188996_TT	24	30'581'083	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382189028_CC	24	30'584'345	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382189031_CC	24	30'584'536	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1382189034_CC	24	30'584'692	2.48	0	28.04	0.046	3.80	5.14E-02	0.920	-25.69	-15.44	
ss1384075360_GG	27	25'930'079	78.88	0	32.88	0.012	22.49	2.11E-06	0.920	1.76	-1.35	WRN
ss1384075361_TT	27	25'930'112	18.63	0	32.88	0.012	22.49	2.11E-06	0.920	1.76	-1.35	WRN
ss1384075370_GG	27	25'930'994	78.88	0	32.88	0.012	22.49	2.11E-06	0.920	1.76	-1.35	WRN
ss1384075403_TT	27	25'933'133	78.88	0	32.88	0.012	22.49	2.11E-06	0.920	1.76	-1.35	WRN

A4.6 Supp. File 6 - Results logistic regressions Europe

Only the 50 first results are presented here. The complete table is available in the online supplemental material of the paper.

ENV=Bioclimatic variable considered, CHR=Chromosome, POS=Position in base pairs, GF=Genotype frequency
G=Gscore, pG=p-value of Gscore, qG=p-value of Gscore corrected for FDR, qW=p-value of Wald score corrected for FDR
b0,b1=coefficients of the logistic regression

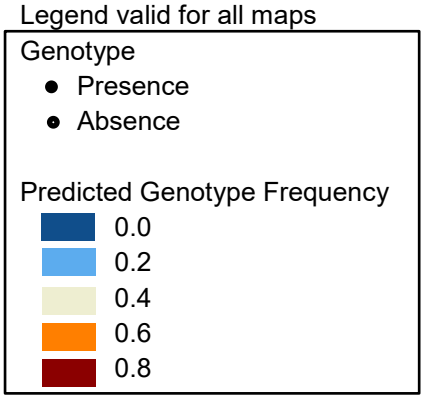
Marker	ENV	CHR	POS	GF	G	pG	qG	qG	qW	b0	b1
snp30075-scaffold33-107046_GG	bio18	19	40224821	33.77	100.09	1.46E-23	3.85E-17	4.48E-18	1.18E-11	-1.41	1.30
snp30075-scaffold33-107046_GG	bio13	19	40224821	33.77	98.57	3.14E-23	4.15E-17	1.07E-17	1.20E-11	-1.40	1.30
snp30075-scaffold33-107046_GG	bio16	19	40224821	33.77	92.49	6.78E-22	5.96E-16	4.23E-17	2.79E-11	-1.36	1.23
snp50723-scaffold731-463240_AA	bio18	1	38282037	31.68	89.92	2.48E-21	1.64E-15	1.36E-17	1.20E-11	-0.97	1.13
snp59445-scaffold980-395848_CC	bio3	6	82779273	31.15	83.66	5.89E-20	3.11E-14	8.02E-17	4.23E-11	0.65	-1.06
snp48982-scaffold7-715056_GG	bio8	13	31676938	31.15	82.33	1.15E-19	5.07E-14	1.46E-16	5.59E-11	0.11	-1.02
snp59445-scaffold980-395848_CC	bio18	6	82779273	31.15	81.78	1.52E-19	5.74E-14	4.36E-15	5.00E-10	0.71	1.12
snp38429-scaffold486-2542930_GG	bio18	5	22749602	39.01	81.03	2.22E-19	7.33E-14	1.48E-16	5.59E-11	-0.52	1.02
snp18702-scaffold189-886569_AA	bio8	6	4000730	19.37	80.01	3.72E-19	9.99E-14	1.70E-11	9.20E-08	-2.06	1.55
snp27880-scaffold299-2993589_GG	bio18	16	75413714	22.77	79.78	4.18E-19	9.99E-14	3.65E-13	8.03E-09	1.30	1.29
snp15175-scaffold1620-882429_CC	bio18	7	39507171	35.08	79.68	4.41E-19	9.99E-14	1.03E-15	2.27E-10	0.27	1.03
snp54026-scaffold822-338299_GG	bio3	5	6878482	39.79	79.62	4.55E-19	9.99E-14	3.07E-16	9.00E-11	-0.63	1.02
snp34534-scaffold406-140611_AA	bio18	14	8680762	47.38	78.93	6.45E-19	1.31E-13	9.17E-16	2.20E-10	-0.16	-1.01
snp30075-scaffold33-107046_AA	bio13	19	40224821	40.31	78.51	7.98E-19	1.45E-13	7.15E-15	6.91E-10	-0.52	-1.07
snp38429-scaffold486-2542930_GG	bio16	5	22749602	39.01	78.44	8.23E-19	1.45E-13	1.26E-15	2.32E-10	-0.51	1.02
snp32290-scaffold366-3455018_GG	bio18	8	82171564	25.92	77.89	1.09E-18	1.80E-13	9.92E-14	3.44E-09	1.05	1.18
snp38628-scaffold49-1755585_GG	bio18	4	8330032	42.67	77.52	1.31E-18	2.03E-13	6.99E-16	1.85E-10	-0.92	1.03
snp59445-scaffold980-395848_CC	bio9	6	82779273	31.15	77.40	1.40E-18	2.03E-13	1.88E-16	6.20E-11	0.64	-0.99
snp30075-scaffold33-107046_AA	bio16	19	40224821	40.31	77.31	1.46E-18	2.03E-13	2.16E-14	1.30E-09	-0.53	-1.08
snp18497-scaffold187-1352016_GG	bio13	15	67044324	39.27	76.52	2.18E-18	2.78E-13	1.90E-15	2.95E-10	-0.23	0.99
snp19007-scaffold191-2265122_AA	bio13	21	46467445	36.39	76.39	2.33E-18	2.80E-13	1.64E-15	2.70E-10	-0.66	1.01
snp23666-scaffold239-665888_GG	bio18	25	22985571	16.75	76.29	2.45E-18	2.81E-13	1.67E-13	4.84E-09	-2.16	1.34
snp50723-scaffold731-463240_AA	bio13	1	38282037	31.68	75.81	3.13E-18	3.44E-13	2.40E-15	3.52E-10	-0.93	1.03
snp47091-scaffold659-1663374_GG	bio8	3	22450724	30.63	75.52	3.61E-18	3.70E-13	1.30E-15	2.32E-10	0.04	-0.97
snp19007-scaffold191-2265122_AA	bio16	21	46467445	36.39	75.48	3.69E-18	3.70E-13	3.14E-15	3.94E-10	-0.65	1.00
snp32290-scaffold366-3455018_GG	bio4	8	82171564	25.92	75.43	3.79E-18	3.70E-13	1.26E-14	9.02E-10	0.97	1.08
snp32290-scaffold366-3455018_GG	bio3	8	82171564	25.92	75.21	4.23E-18	3.98E-13	2.93E-15	3.86E-10	0.96	-1.04
snp45539-scaffold622-316980_AA	bio18	13	80308039	38.48	74.77	5.28E-18	4.80E-13	1.32E-15	2.32E-10	-0.55	0.97
snp38429-scaffold486-2542930_GG	bio13	5	22749602	39.01	74.13	7.31E-18	6.43E-13	3.44E-15	4.13E-10	-0.51	0.98
snp18497-scaffold187-1352016_GG	bio16	15	67044324	39.27	73.44	1.04E-17	8.63E-13	9.02E-15	7.44E-10	-0.23	0.97
snp8477-scaffold1307-443595_GG	bio18	1	10438434	27.23	73.36	1.08E-17	8.63E-13	5.42E-15	5.72E-10	-1.21	1.05
snp15175-scaffold1620-882429_CC	bio13	7	39507171	35.08	73.36	1.08E-17	8.63E-13	1.45E-14	1.01E-09	0.26	0.99

snp50723-scaffold731-463240_AA	bio16	1	38282037	31.68	72.69	1.52E-17	1.17E-12	7.88E-15	6.93E-10	-0.92	1.00
snp15175-scaffold1620-882429_CC	bio16	7	39507171	35.08	72.64	1.56E-17	1.17E-12	3.53E-14	1.74E-09	0.26	1.00
snp31095-scaffold344-317295_AA	bio18	6	8456089	19.63	72.48	1.69E-17	1.24E-12	7.29E-14	2.92E-09	-1.81	1.19
snp59488-scaffold980-1135359_AA	bio16	6	83518784	41.88	72.37	1.79E-17	1.27E-12	1.06E-14	7.96E-10	-0.37	0.96
snp3939-scaffold1122-2977302_AA	bio18	9	24822549	29.84	72.20	1.94E-17	1.35E-12	5.37E-15	5.72E-10	-1.04	1.01
snp59488-scaffold980-1135359_AA	bio13	6	83518784	41.88	72.02	2.13E-17	1.44E-12	7.60E-15	6.91E-10	-0.38	0.96
snp48982-scaffold7-715056_GG	bio9	13	31676938	31.15	71.76	2.43E-17	1.60E-12	2.66E-15	3.70E-10	0.11	0.93
snp32099-scaffold362-780050_GG	bio18	4	92585577	31.68	71.35	2.99E-17	1.92E-12	3.63E-14	1.74E-09	0.50	0.99
snp12587-scaffold148-3686817_AA	bio18	6	32525693	76.18	71.21	3.21E-17	2.01E-12	2.28E-14	1.31E-09	1.45	-1.08
snp16321-scaffold1719-579549_GG	bio18	29	38690609	20.94	70.76	4.04E-17	2.48E-12	1.00E-14	7.96E-10	1.10	-1.01
snp35670-scaffold43-2702091_AA	bio18	18	21448746	37.43	70.71	4.15E-17	2.49E-12	5.74E-15	5.83E-10	-0.60	0.95
snp59445-scaffold980-395848_CC	bio11	6	82779273	31.15	70.56	4.47E-17	2.62E-12	2.62E-14	1.47E-09	0.66	-0.99
snp23703-scaffold239-2312394_AA	bio18	25	21339065	42.15	70.12	5.58E-17	3.14E-12	7.56E-15	6.91E-10	-0.36	0.93
snp394-scaffold1009-1164506_AA	bio13	1	112960571	16.75	70.05	5.79E-17	3.14E-12	4.61E-13	9.50E-09	-2.10	1.26
snp3939-scaffold1122-2977302_AA	bio13	9	24822549	29.84	70.04	5.80E-17	3.14E-12	1.89E-14	1.22E-09	-1.03	1.00
snp23698-scaffold239-2103925_AA	bio18	25	21547534	16.49	70.03	5.83E-17	3.14E-12	6.99E-13	1.21E-08	-2.14	1.27
snp30075-scaffold33-107046_AA	bio18	19	40224821	40.31	69.95	6.09E-17	3.21E-12	5.97E-14	2.50E-09	-0.51	-0.98
snp21854-scaffold2142-76410_GG	bio13	8	17321120	25.92	69.91	6.20E-17	3.21E-12	1.57E-14	1.06E-09	0.39	-0.94
snp27836-scaffold299-994978_AA	bio18	16	73415103	37.43	69.60	7.25E-17	3.65E-12	1.25E-13	3.93E-09	-0.66	-1.01

A4.7 Supp. File 7 – Univariate SPAG

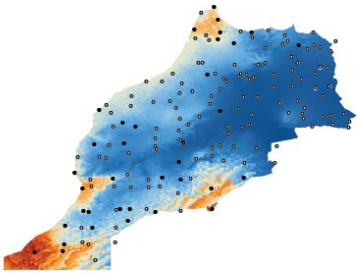
Figures on the following pages presented the Univariate Spatial Areas of Genotype Probability for all models presented in the tables in the main text. The maps show the average genotype(s) frequency(ies) based on the 10 runs computed with different random selection of training sets containing 25% of the total number of individuals. Please refer to Box 1 in main text to interpret the validation graphs shown on the right of each map and refer to Annex A4.2 for the list of bioclimatic variables.

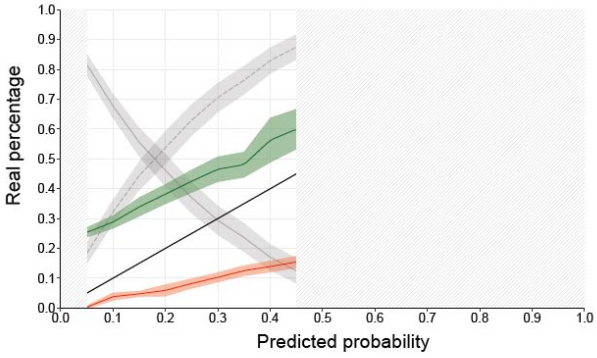
Legend valid for all maps



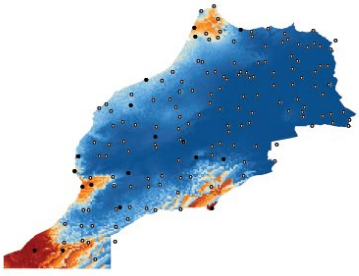
Moroccan dataset

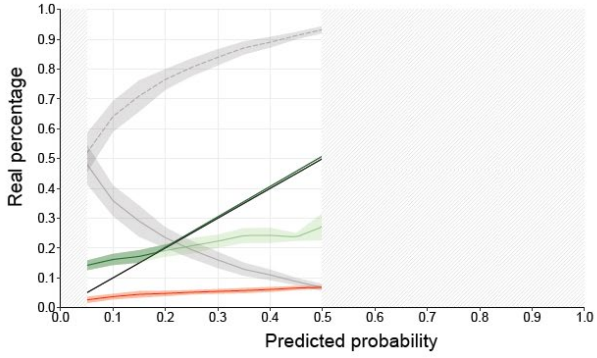
Chromosome 6: 12276168 AA



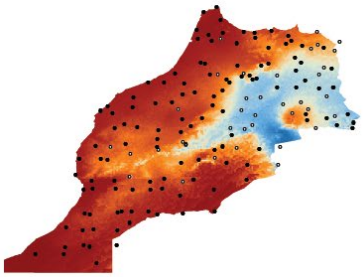


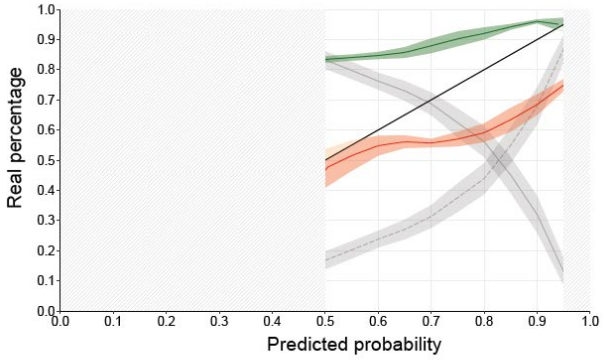
Chromosome 13: 43436394 GG



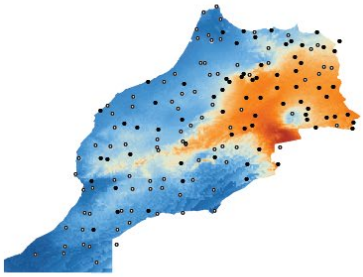


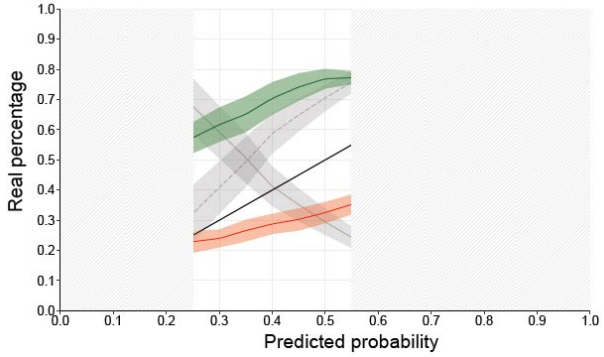
Chromosome 24: 19436980 CC



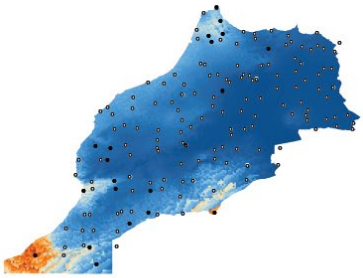


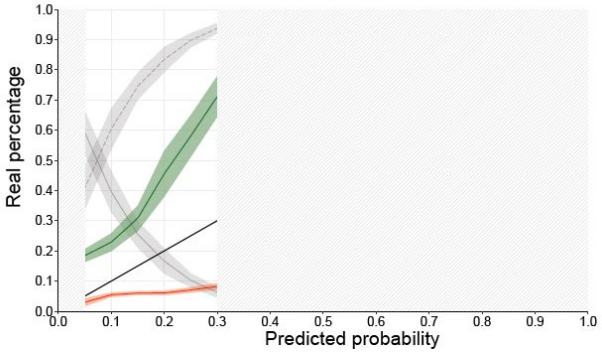
Chromosome 24: 25860754 GA



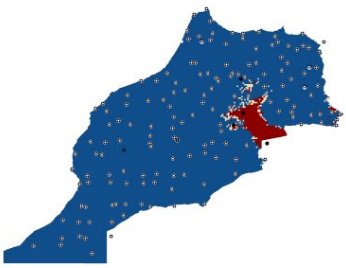


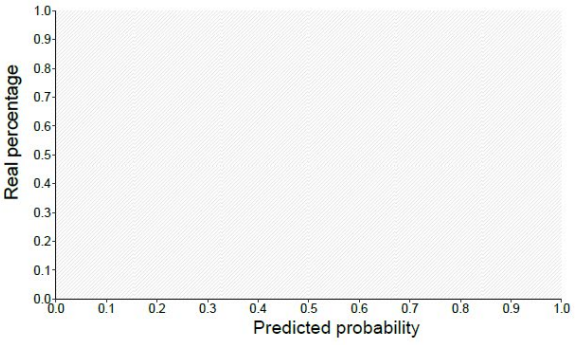
Chromosome 24: 28833253 TT



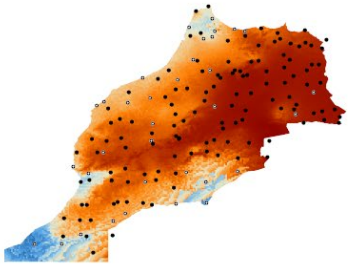


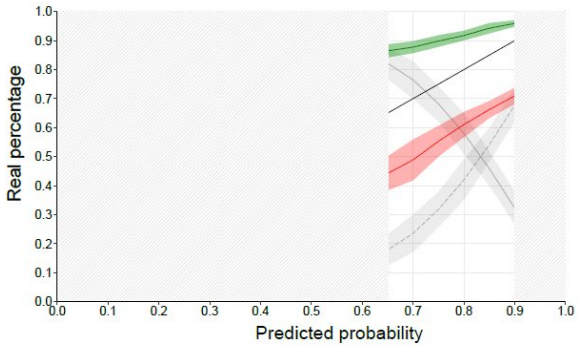
Chromosome 24: 30566869 TT





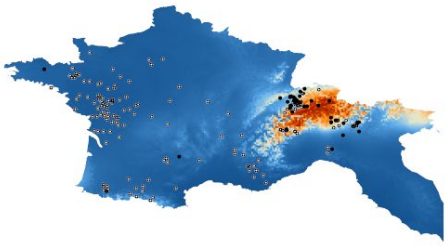
Chromosome 27: 25930079 GG

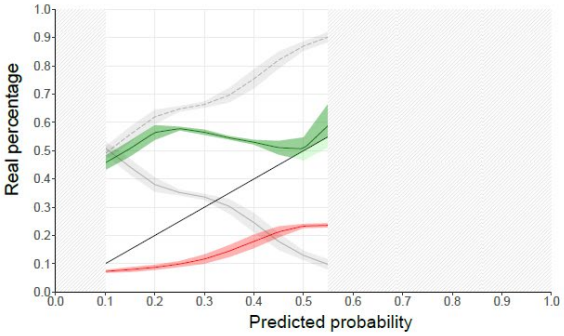




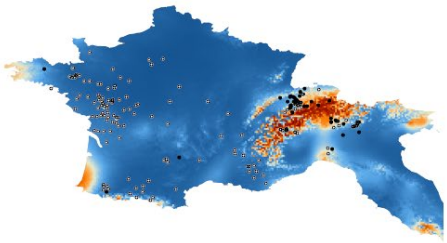
European dataset

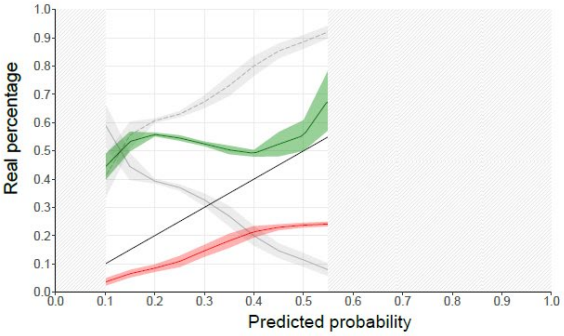
Chromosome 19: 40224821 GG Bio 18



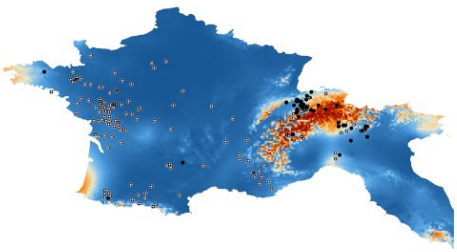


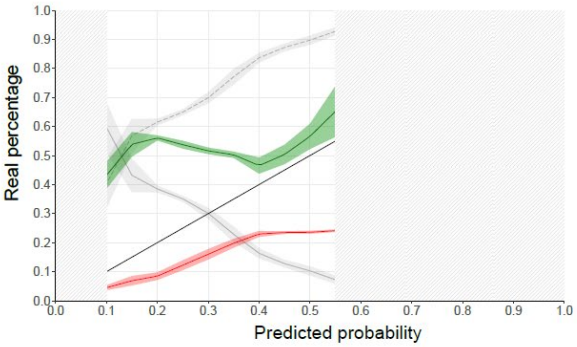
Chromosome 19: 40224821 GG Bio 13



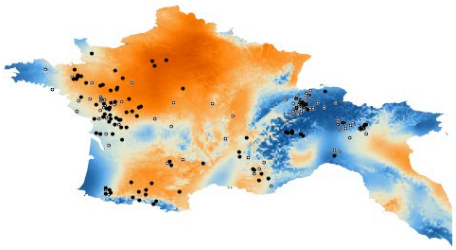


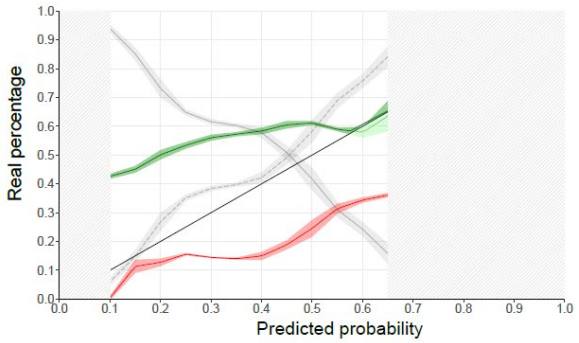
Chromosome 19: 40224821 GG Bio 16



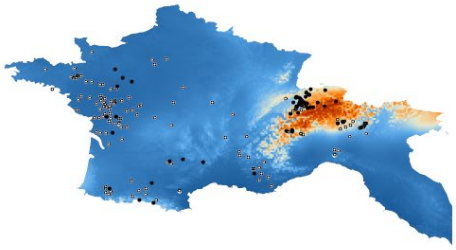


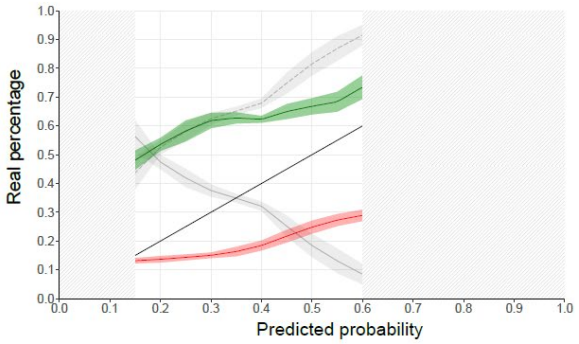
Chromosome 19: 40224821 AA Bio 13



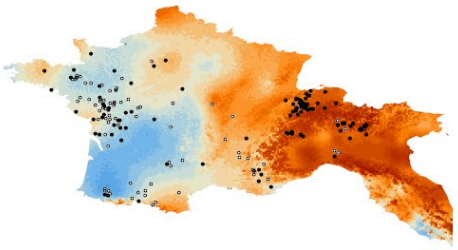


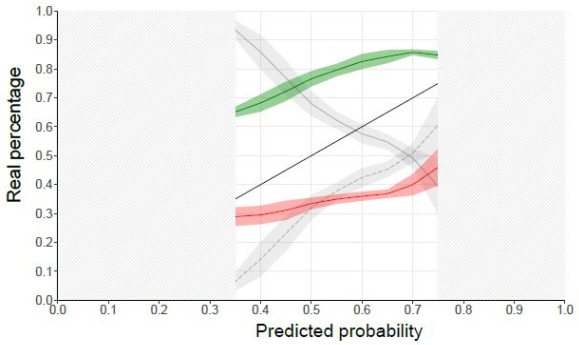
Chromosome 1: 38282037 AA Bio 18



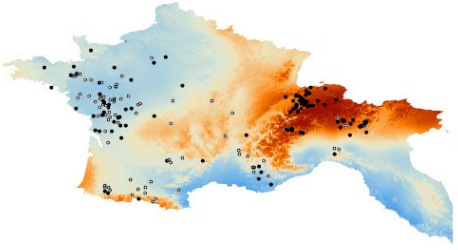


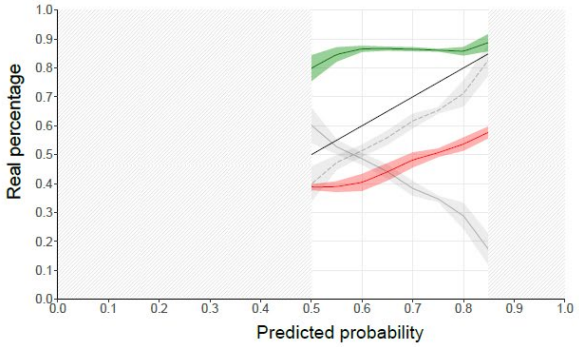
Chromosome 6: 82779273 CC Bio 3



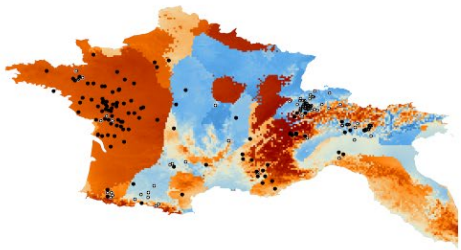


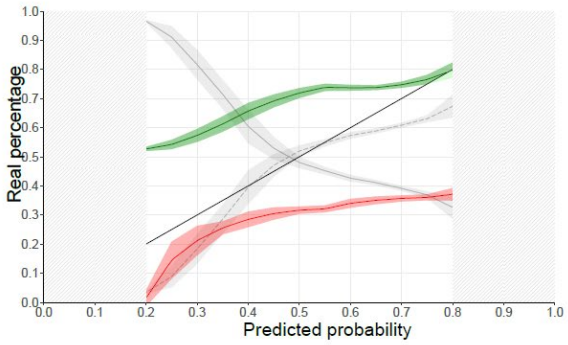
Chromosome 6: 82779273 CC Bio 18



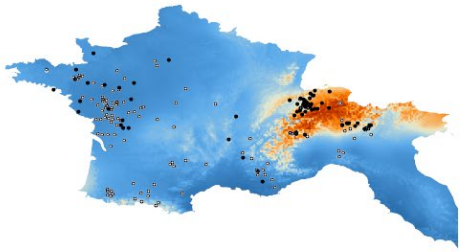


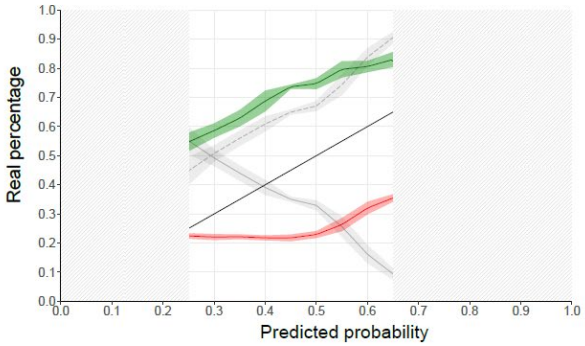
Chromosome 13: 31676938 GG Bio 8



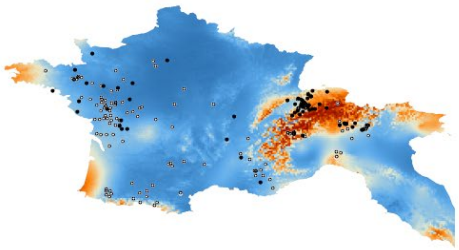


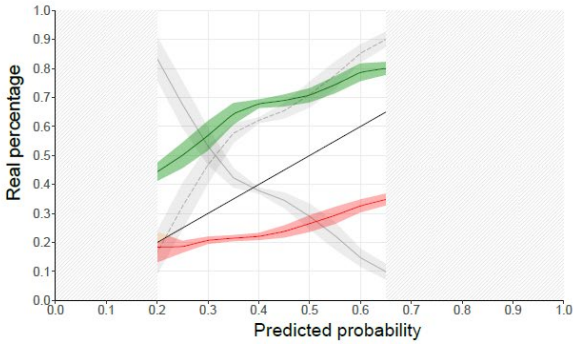
Chromosome 5: 22749602 GG Bio 18



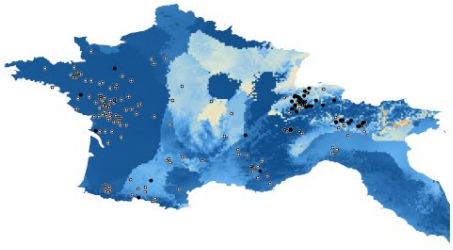


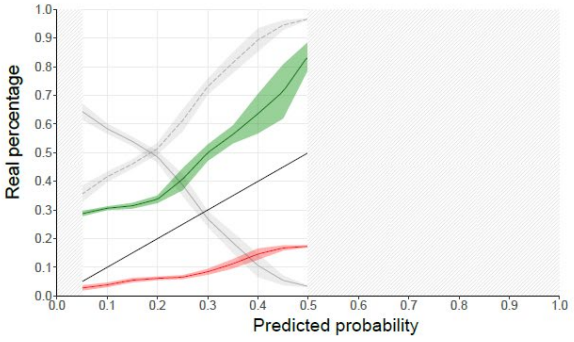
Chromosome 5: 22749602 GG Bio 16



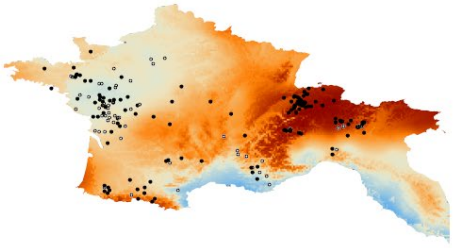


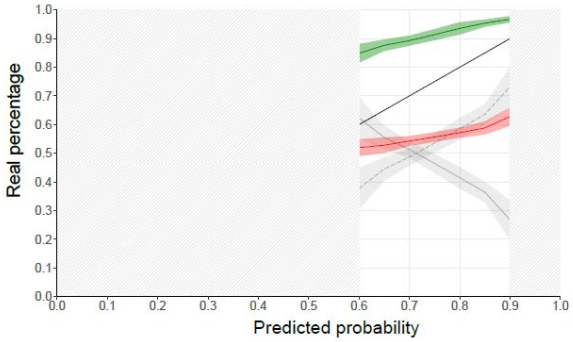
Chromosome 6: 4000730 AA Bio 8



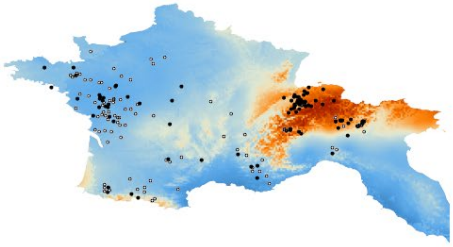


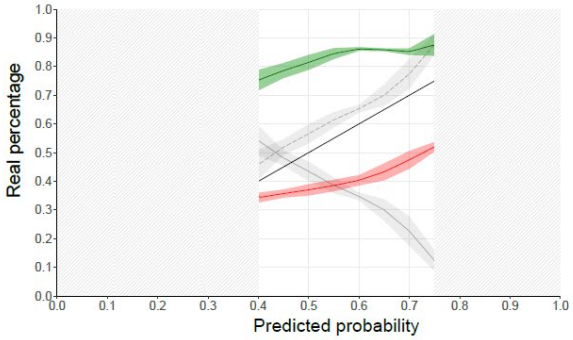
Chromosome 16: 75413714 GG Bio 18



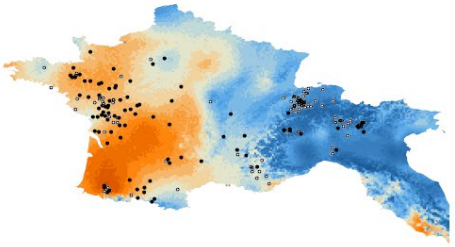


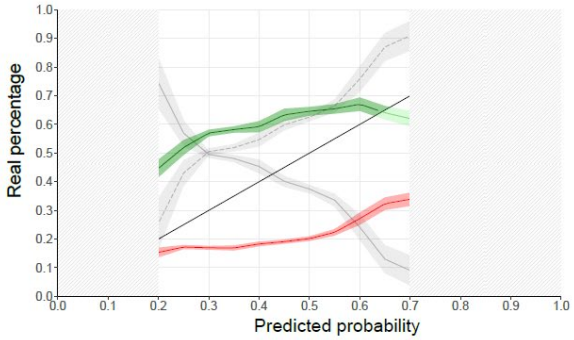
Chromosome 7: 39507171 CC Bio 18



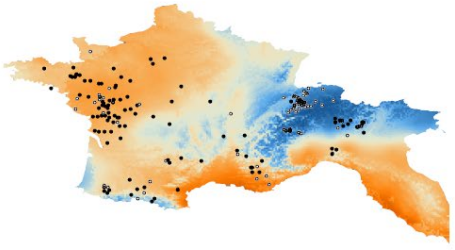


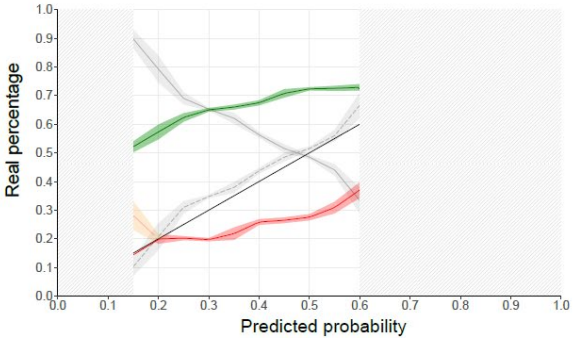
Chromosome 5: 6878482 GG Bio 3





Chromosome 14: 8680762 AA Bio 18





CURRICULUM VITAE

ROCHAT Estelle

Rue d'Allesses 2
CH-1905 Dorénaz
Switzerland

estelle.rochat1@gmail.com
+41 79 291 59 63

01.02.1990
Single, Swiss

EDUCATION

2015 - 2020	**PhD** – Swiss Federal Institute of Technology of Lausanne (EPFL) *Species conservation in the face of global environmental changes: surface-based modelling of geo-environmental data to identify vulnerable populations*
2012 - 2014	**Master** in Environmental Sciences and Engineering – EPFL Specialization in environmental monitoring and modelling
2009 - 2012	**Bachelor** in Environmental Sciences and Engineering – EPFL
2004 - 2009	**Federal Matura** – Collège de l'Abbaye, Saint-Maurice Specialization in Physics and Application of mathematics *Final project: General public booklet on fauna, flora and hydrology (E Rochat, Au fil de l'Eau d'Allesses, 2009)*

AWARDS

2015	**Geosuisse prize** for the best Master average in the Environmental section
2015	**EPFL Excellence award** for the Master degree

PROFESSIONAL EXPERIENCE

2015 - …	**MicroGIS SA**, St-Sulpice – Environmental Engineer (40%) Transport planification, Spatial analysis, Cartography
2015 - 2020	**EPFL** – Head of the EPFL Geo-database
2015 - 2018	**EPFL** – Main teaching assistant, GIS course
Summer 2012	**Silvaplus Sàrl**, Martigny – Environmental Engineering Internship Natural hazards, Forest biodiversity, Ecological compensation measures

COMPUTER SKILLS

GIS and database	QGIS, Manifold, SAGA GIS, PostgreSQL/PostGIS, OpenGeoDa, Graphab, UNICOR, ArcGIS*, Erdas Imagine*
Genetics	CDPOP, STRUCTURE, ADMIXTURE, Samβada, LFMM
Programming	R, Matlab/Octave, SQL, C*, Java*, php/html*, Python*
	* Limited knowledge

LANGUAGES

French (Mother tongue), English (Good knowledge - C1), German (Basic knowledge - B1)

PEER-REVIEWED PAPERS

- Submitted in *Applied and Environmental Microbiology*: Rochat E, Vuilleumier S, Aeby S, Greub G, Joost S (2020). Nested species distribution models of *Chlamydiales* in tick host *Ixodes ricinus* in Switzerland.

- Revised and re-submitted in *Global Change Biology*: Rochat E, Joost S (2020). Spatial Areas of Genotype Probability (SPAG): predicting the spatial distribution of adaptive genetic variants under future climatic conditions.

- Selmoni O, Rochat E, Lecellier G, Berteaux-Lecellier V, Joost S (2020). Seascape genomics as a new tool to empower coral reef conservation strategies: an example on north-western Pacific *Acropora digitifera*. *Evolutionary Applications*.

- Selmoni O, Vajana E, Guillaume A, Rochat E, Joost S (2020). Sampling strategy optimization to increase statistical power in landscape genomics: A simulation-based approach. *Molecular Ecology Resources* 20: 154–169.

- Duruz S, Sevane N, Selmoni O, Vajana E, Leempoel K, Stucki S, Orozco-terWengel P, Rochat E, et al. (2019). Rapid identification and interpretation of gene–environment associations using the new R.SamBada landscape genomics pipeline. *Molecular Ecology Resources* 19: 1355–1365.

- Vajana E, Widmer I, Rochat E, Duruz S, Selmoni O, Vuilleumier S, et al. (2019). Indication of spatially random occurrence of Chlamydia-like organisms in Bufo bufo tadpoles from ponds located in the Geneva metropolitan area. *New Microbes and New Infections* 27: 54–63.

- Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution* 50: 57.

- Colli L, Milanesi M, Talenti A, Bertolini F, Chen M, Crisà A, et al. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution* 50: 58.

- Vajana E, Barbato M, Colli L, Milanesi M, Rochat E, Fabrizi E, et al. (2018). Combining Landscape Genomics and Ecological Modelling to Investigate Local Adaptation of Indigenous Ugandan Cattle to East Coast Fever. *Front Genet* 9.

- Aebischer T, Siguindo G, Rochat E, Arandjelovic M, Heilman A, Hickisch R, et al. (2017). First quantitative survey delineates the distribution of chimpanzees in the Eastern Central African Republic. *Biological Conservation* 213: 84–94.

- Rochat E, Manel S, Deschamps-Cottin M, Widmer I, Joost S (2017). Persistence of butterfly populations in fragmented habitats along urban density gradients: motility helps. *Heredity* 119: 328–338.

- Leempoel K, Duruz S, Rochat E, Widmer I, Orozco-terWengel P, Joost S (2017). Simple Rules for an Efficient Use of Geographic Information Systems in Molecular Ecology. *Front Ecol Evol* 5.

CONFERENCE PAPERS

- Joost S, Duruz S, Rochat E, Widmer I (2016). Open computational landscape genetics. PeerJ Preprints.

- Rochat E, Duruz S, Widmer I, Clémence A, Desrichard O, Rappo D, et al. (2015). Relationship between land cover type and Body Mass Index in Geneva. In: *2015 Joint Urban Remote Sensing Event (JURSE)*.

CONFERENCE TALKS

- Rochat E, Vuilleumier S, Tischhauser W, Ackermann-Gäumann R, Pilloux l, Greub G, Joost S (2019). Current and future spatial distribution of the tick *Ixodes ricinus* - host of the *Rhabdochlamydiae* bacterial pathogens - in Switzerland. *Annual Swiss Society for Microbiology Meeting*. Lausanne, Switzerland.

- Rochat E, Leempoel K, Vajana E, Colli L, Ajmone-Marsan P, Joost S. and the NEXTGEN Consortium (2016). Map of genotype frequency change in autochthonous Moroccan sheep breeds due to global warming. *Conference on Conservation Genomics*, Vairão, Portugal.

- Rochat E, Widmer I, Manel S, Deschamps-Cottin M, Joost S (2015). Impact of the urbanization process on connectivity and genetic diversity - a spatially explicit simulation approach. *First Annual Meeting in Conservation Genetics – Science and Practice*. Birmensdorf, Switzerland.

CONFERENCE POSTERS

- Rochat E, Vuilleumier S, Tischhauser W, Ackermann-Gäumann R, Pilloux l, Greub G, Joost S (2019). Current and future spatial distribution of the tick *Ixodes ricinus* - host of the *Rhabdochlamydiae* bacterial pathogens - in Switzerland. *Swiss Microbial ecology Meeting*, Lausanne, Switzerland.

- Vajana E*, Rochat E*, Colli L, Masembe C, Negrini R, Ajmone-Marsan P, Joost S and the NEXTGEN Consortium (2016). Spatial Areas of Genotype Probability of Cattle Genomic Variants Involved in the Resistance to East Coast Fever: A Tool to Predict Future Disease-Vulnerable Geographical Regions. *Plant & Animal Genome Conference XXIV*. San Diego, California, USA. *co-first authors

- Rochat E*, Vajana E*, Colli L, Masembe C, Negrini R, Ajmone-Marsan P, Joost S and the NEXTGEN Consortium (2015). Effect of climate change on the spatial distribution of genomic variants involved in the resistance to East Coast Fever in Ugandan cattle. *19th Evolutionary Biology Meeting*, Marseille, France. *co-first authors

- Widmer I, Rochat E, Leempoel K, Clémence A, Ertz O. et al. (2015). Biodiversity dynamics and the effect of urban environment on the distribution of genetic variation in the Geneva cross-border area. *First Annual Meeting in Conservation Genetics – Science and Practice*, Birmensdorf, Switzerland.