GfÖ
GfÖ Ecological Society of Germany, Austria and Switzerland
Basic and Applied Ecology 17 (2016) 61–71
Basic and Applied Ecology
www.elsevier.com/locate/baae

Different habitat suitability models yield different least-cost path distances for landscape genetic analysis

P. Milanesi[a,b,*], R. Holderegger[b,c], R. Caniglia[a], E. Fabbri[a], E. Randi[a,d]

[a] Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Via Ca' Fornacetta 9, Ozzano dell'Emilia, 40064 Bologna, Italy
[b] WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[c] ETH Zürich, Institute of Integrative Biology, Universitätstrasse 16, CH-8092 Zürich, Switzerland
[d] Department 18/Section of Environmental Engineering, Aalborg University, Sohngårdsholmsvej 57, 9000 Aalborg, Denmark

Received 5 February 2015; accepted 16 August 2015

Abstract

Habitat suitability models (HSMs) are used to describe and predict species distributions based on multiple ecological variables and species occurrence data. HSMs may also provide a probabilistic identification of least-cost path (LCP) distances in landscape genetics. However, while several studies used HSMs for these purposes, the performance of different HSMs in landscape genetic analysis and, therefore, the consequences of model choice have not been carefully explored. In this study, we used a large dataset of wolf genotypes (*Canis lupus*; $n = 923$) that were non-invasively sampled in the central and northern Italian Apennines and western Alps, aiming (i) to estimate LCP distances derived from ten different HSMs and (ii) to quantify the correlation between inter-individual genetic and LCP distances using three statistical procedures: partial Mantel tests, multiple regression on distance matrices (MRDM) and linear mixed effect models. All LCP distances based on HSMs explained genetic distances better than Euclidean distances, irrespective of the applied landscape genetic statistical test. However, LCP distances derived by different HSMs were significantly different (paired t-test, $P \leq 0.0001$), especially between "flexible discriminant analysis" (FDA) and "boosted regression trees" (BRT) models. LCP distances derived from "factorial decomposition of Mahalanobis distances" (MADIFA) in MRDM showed the highest regression coefficient (β) with genetic distances, indicating a strong correlation between LCPs and genetic distances. Results from our case study suggest that different HSMs should be compared and model-choice procedures applied to identify the best fitting HSM in landscape genetic analysis.

Zusammenfassung

Habitateignungsmodelle (HSMs) verwenden verschiedene Umweltvariablen und Beobachtungsdaten, um das Verbreitungsgebiet von Arten zu definieren und vorherzusagen. HSMs können auch für die Identifizierung von "Least-Cost-Path"-Distanzen (LCP-Distanzen) eingesetzt werden, welche in landschaftsgenetische Analysen einfließen. Obwohl einige Studien HSMs zu diesem Zweck verwendet haben, hat bislang keine Studie den Einfluss unterschiedlicher HSMs in landschaftsgenetischen

---

[*] Corresponding author at: Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Laboratorio di Genetica, Via Ca' Fornacetta, 9, Ozzano dell'Emilia, 40064 Bologna, Italy. Tel.: +39 3288840275.
E-mail address: pietro.milanesi@isprambiente.it (P. Milanesi).

http://dx.doi.org/10.1016/j.baae.2015.08.008
1439-1791/

Analysen untersucht: die Folgen der Modellwahl sind somit unklar. In dieser Studie verwendeten wir einen großen Datensatz von Genotypen von Wölfen (*Canis lupus*; n = 923), die nicht-invasiv im zentralen und nördlichen italienischen Apennin gesammelt wurden, mit dem Ziel, (i) LCP-Distanzen basierend auf zehn verschiedenen HSMs zu berechnen und (ii) die Korrelation zwischen inter-individuellen genetischen Distanzen und LCP-Distanzen mit Hilfe von drei statistischen Verfahren zu quantifizieren: partielle Mantel-Tests, multiple Regressionen von Distanzmatrizen (MRDM) und lineare Modelle mit gemischten Effekten. Alle LCP-Distanzen basierend auf HSMs erklärten die genetischen Distanzen besser als dies euklidische Distanzen taten, unabhängig davon, welcher landschaftsgenetische statistische Test verwendet wurde. Jedoch waren die LCP-Distanzen der verschiedenen HSMs signifikant verschieden voneinander (gepaarter t-Test, $P \leq 0{,}0001$), insbesondere zwischen “flexible discriminant analysis” (FDA) und “boosted regression trees” (BRT). LCP-Distanzen von “factorial decomposition of Mahalanobis distances” (MADIFA) in MRDM zeigten die höchsten Regressionskoeffizienten (β) mit den genetischen Distanzen, was auf eine starke Korrelation zwischen LCP und genetischen Distanzen hinweist. Die Ergebnisse unserer Studie zeigen, dass verschiedene HSMs verglichen und Modellauswahlverfahren angewendet werden sollten, um das am besten passende HSM für einelandschaftsgenetische Analyse zu identifizieren.

Keywords: *Canis lupus*; Multiple regression on distance matrix; Linear mixed effect models; Partial Mantel test; Resistance surfaces

Introduction

Landscape genetics aims to identify landscape features that may influence natural selection and rates of gene flow, providing evidence for local adaptations (Segelbacher et al. 2010) and functional connectivity among populations (Storfer et al. 2007; Holderegger & Wagner 2008). Landscape genetics uses least-cost path analyses (LCPs; Cushman, McKelvey, Hayden, & Schwartz 2006; Shirk, Wallin, Cushman, Rice, & Warheit 2010; Sawyer, Epps, & Brashares 2011) to identify dispersal routes across landscapes. LCP distances typically correlate more strongly to inter-individual or population genetic distances than Euclidean (geographical) distances (Holderegger & Wagner 2008). LCPs are derived from rastered resistance surfaces, reflecting the resistance to animal movement across each raster grid. Higher and lower resistance weights indicate landscape features which are more or less likely used for movement (Adriaensen et al. 2003). Resistance values should ideally be derived from empirical data, *e.g.* direct observation of individual movement, GPS or radio-tracking data, but since empirical information is often not available, resistance weights are mostly based on expert opinions (Zeller, McGarigal, & Whiteley 2012). Nevertheless, resistance surfaces derived from expert opinions may be subjective, because they depend on the individual experiences, knowledge of the study area and the ecology of the species (Spear, Balkenhol, Fortin, McRae, & Scribner 2010; Shirk et al. 2010; Zeller et al. 2012).

A robust way to obtain resistance surfaces is the use of habitat suitability models (HSMs; also referred to as species distribution models; Guisan et al. 2013). HSMs associate locations of species occurrence with multiple ecological variables (predictors) in multivariate statistical analyses. Resistance values can be derived from HSM predictions (as 1-habitat suitability) per grid cell (Wang, Yang, Bridgman, & Lin 2008; Spear et al. 2010). Higher values indicate higher resistance to movement, while lower values represent lower resistance. HSMs have rarely been used in landscape genetics so far (but see Wang et al. 2008; Brown & Knowles 2012; Duckett, Wilson & Stow 2013; Wang, Glor, & Losos 2013; see Table A1). Resistance surfaces based on HSMs can be translated into LCPs, and a pairwise LCP distance matrix can be statistically compared with a matrix of pairwise genetic distances. For example, Wang et al. (2008) showed that LCP distances derived from one HSM better explained genetic distances among spiny rat (*Niviventer coninga*) individuals than Euclidean distances did. Brown and Knowles (2012) demonstrated the utility of a HSM to correlate genetic patterns with habitat characteristics in the American Pika (*Ochotona princeps*). Duckett et al. (2013) succesfully applied a HSM and genetic data to quantitatively predict future dispersal rates of the Australian arid zone gecko (*Gehyra variegata*) toward new areas under climate change.

HSMs encompass several approaches (see Table A1). Differences in model' performance are likely due to different assumptions of the HSMs methods and to the methods' abilities to recover true relationships between species occurrence and environmental factors (Segurado & Araújo 2004). HSMs also vary in relation to the data types they use (*i.e.* presence-only, presence/absence, or presence/pseudo-absence), the inclusion of diverse factors into the models (*i.e.* number of predictor variables or spatial autocorrelation among species locations; Thibaud, Petitpierre, Broennimann, Davison, & Guisan 2014) and species characteristics (*i.e.* species with broad or narrow niches and shape of response curve to environmental factors; Peterson et al. 2011). In principle, no HSM is supposed to outcompete all the others and model comparison is strongly suggested (Qiao, Soberón, & Peterson 2015).

Almost all studies that used HSMs in landscape genetics only applied one particular HSM, avoiding model