

A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models

MAARTEN J. VAN STRIEN,^{†*} DANIELA KELLER^{*†} and ROLF HOLDEREGGER^{*†}

^{*}WSL Swiss Federal Research Institute, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland, [†]Department of Environmental System Sciences, ETH Zurich, Universitätsstrasse 16, CH-8092 Zurich, Switzerland

Abstract

Landscape genetics aims to assess the effect of the landscape on intraspecific genetic structure. To quantify inter deme landscape structure, landscape genetics primarily uses landscape resistance surfaces (RSs) and least-cost paths or straight-line transects. However, both approaches have drawbacks. Parameterization of RSs is a subjective process, and least-cost paths represent a single migration route. A transect-based approach might oversimplify migration patterns by assuming rectilinear migration. To overcome these limitations, we combined these two methods in a new landscape genetic approach: least-cost transect analysis (LCTA). Habitat-matrix RSs were used to create least-cost paths, which were subsequently buffered to form transects in which the abundance of several landscape elements was quantified. To maintain objectivity, this analysis was repeated so that each landscape element was in turn regarded as migration habitat. The relationship between explanatory variables and genetic distances was then assessed following a mixed modelling approach to account for the nonindependence of values in distance matrices. Subsequently, the best fitting model was selected using the R^2_{β} statistic. We applied LCTA and the mixed modelling approach to an empirical genetic dataset on the endangered damselfly, *Coenagrion mercuriale*. We compared the results to those obtained from traditional least-cost, effective and resistance distance analysis. We showed that LCTA is an objective approach that identifies both the most probable migration habitat and landscape elements that either inhibit or facilitate gene flow. Although we believe the statistical approach to be an improvement for the analysis of distance matrices in landscape genetics, more stringent testing is needed.

Keywords: gene flow, landscape genetics, least-cost paths, linear mixed model, migration, resistance surface

Received 6 July 2011; revision received 2 May 2012; accepted 16 May 2012

Introduction

It is vital, for species survival, to restore gene flow (i.e. the spread of genes between demes through migration) in areas where anthropogenic land use has decreased or halted it. To help conservationists achieve this goal, the discipline of landscape genetics aims to assess the effect of landscape variables on gene flow (Manel *et al.* 2003; Storfer *et al.* 2007, 2010; Holderegger

& Wagner 2008). To achieve this, both the inter deme landscape structure and gene flow need to be quantified. Interdeme gene flow can be measured by genotyping individuals and calculating genetic (dis)similarity between demes. For a review of measures for the quantification of intraspecific genetic structure, we refer to Meirmans & Hedrick (2011).

To quantify inter deme landscape structure, several methods are currently used, which can be roughly classified into two categories: methods that use landscape resistance surfaces (RSs) or transect-based analyses (Spear *et al.* 2010). A landscape RS is a grid of a study

Correspondence: Maarten J. van Strien, Fax: +41 44 7392 215; E-mail: maarten@vstrien.nl

area in which each pixel is assigned a cost value based on the hypothesized permeability to migration of the landscape in that pixel. In many studies, interdeme least-cost paths are calculated on these RSs (Adriaensen *et al.* 2003). Either the total cost of a path (least-cost distance; Cushman *et al.* 2006; Perez-Espona *et al.* 2008; Schwartz *et al.* 2009; Davis *et al.* 2010; Hagerty *et al.* 2011) or the length of a path (effective distance; Coulon *et al.* 2004; Spear *et al.* 2005; McRae & Beier 2007) is used as a measure for interdeme landscape structure. Transect-based analyses quantify the landscape composition (i.e. the abundance of each landscape element) in a straight-line transect between two demes (e.g. Pavlacky *et al.* 2009; Angelone *et al.* 2011; Emaresi *et al.* 2011).

Both these approaches have shortcomings. The parameterization of RSs usually requires subjective decisions about the species-specific cost value of a certain landscape element (Adriaensen *et al.* 2003; Holdegger & Wagner 2008; Spear *et al.* 2010). Furthermore, least-cost paths represent only a single potential migration route and do not incorporate any potentially influential landscape elements surrounding the route (Adriaensen *et al.* 2003; Spear *et al.* 2010). Although not based on subjective decisions, nor representing a single migration path (Emaresi *et al.* 2011), straight-line transects have the disadvantage that rectilinear migration routes are assumed (Spear *et al.* 2010).

To overcome most of these shortcomings, we present an approach that combines these two techniques: least-cost transect analysis (LCTA). First, we select a number of landscape elements that may have a facilitative or inhibitive effect on gene flow. Subsequently, binary RSs representing migration habitat and matrix are created. In contrast to, for instance, breeding or foraging habitat, the migration habitat specifically refers to the habitat type through which the focal species prefers to migrate, which is not necessarily the same habitat type that the species inhabits or breeds in (Stevens *et al.* 2006; Spear *et al.* 2010). To maintain objectivity, each selected landscape element is in turn regarded as migration habitat, regardless of its hypothesized inhibitive or facilitative nature. Interdeme least-cost paths are calculated for each RS to get potential interdeme migration routes that may be more realistic than the assumed straight-line migration in transect-based approaches. We acknowledge that an interdeme least-cost path gives, at best, a rough estimate of the migration route, as not all individuals will migrate along a single least-cost path and as deviations from the path are likely. Therefore, we buffer the least-cost path to create a transect of a certain width, which is more likely to cover the real migration paths and to include

potentially important landscape elements around the least-cost path. The scale at which landscape elements influence habitat selection is a species-specific trait (Holland *et al.* 2004). We expect that the same is true for the scale of the landscape influence on migration habitat and, therefore, experiment with a range of buffer widths. We calculate the transect length and the proportion of each landscape element in the transect to form a set of complementary explanatory variables per RS. The fit of the regression between these sets of explanatory variables and genetic distance is used to determine which landscape element is the most likely migration habitat and which landscape elements are probably to act upon gene flow in a facilitative or inhibitory way.

The dependent and explanatory variables are in the form of (dis)similarity or distance matrices, in which values are not fully independent of other values in the same row or column. Therefore, Mantel tests are frequently applied in landscape genetics to test model significance (Storfer *et al.* 2010). Although successful in model testing, Mantel tests do not take into account the nonindependence of the pairwise distances when calculating the correlation coefficient between distance matrices (Yang 2004). To provide a solution to this shortcoming, Clarke *et al.* (2002) presented the maximum-likelihood population-effects (MLPE) model, of which the covariate structure is specifically designed to account for the correlated structure in regression on distance matrices. Although it has been applied in landscape ecological analysis (Bellamy *et al.* 2003), to our knowledge, Clarke's MLPE method has not been applied in other landscape genetic studies. Here, we implement this method to fit the regression models between genetic distances and sets of explanatory variables. To select the best fitting and most parsimonious set of explanatory variables, we use the R^2_β statistic described in Edwards *et al.* (2008), because information criteria, like AIC and BIC, are not suitable when comparing mixed models where parameter estimation is performed with the residual maximum-likelihood (REML) criterion (as in Clarke *et al.* 2002; Verbeke & Molenberghs 2000).

In summary, the goals of this study are twofold. First, we describe an objective procedure to quantify interdeme landscape (LCTA). Second, we present a statistical analysis approach based on MLPE models (Clarke *et al.* 2002) and the R^2_β statistic (Edwards *et al.* 2008). We demonstrate our approach on an empirical genetic dataset of the southern damselfly *Coenagrion mercuriale* and compare our approach to several other interdeme landscape measures that are calculated from RSs and are regularly used in landscape genetics (Table 1).

Table 1 For each of the applied landscape genetic methods, a short description is given on how the landscape measure was calculated. Also the input variables for the mixed effect models are shown. The dependent variable consisted of 78 pairwise F_{ST} values among 13 demes. The random effect is the same for every mixed model. The fixed effects (explanatory variables) differed per method but were always calculated from the same set of 26 binary habitat-matrix resistance surfaces

Method	Description	Dependent variable	Fixed effects (explanatory variables)	Random effect
Resistance distance analysis	Resistance distances calculated with CIRCUITSCAPE (McRae 2006)	F_{ST}	Resistance distance	Individual deme effect
Least-cost distance analysis	Total cost of least-cost path	F_{ST}	Cost distance	Individual deme effect
Effective distance analysis	Total length of least-cost path	F_{ST}	Potential effective distance (m)	Individual deme effect
Least-cost transect analysis	Total length of the least-cost path and proportions of landscape elements in a transect along the least-cost path	F_{ST}	Transect length (m) Agriculture ($m^2 \times m^{-2}$) Forest ($m^2 \times m^{-2}$) Residential ($m^2 \times m^{-2}$) Streams ($m^2 \times m^{-2}$) Roads ($m^2 \times m^{-2}$)	Individual deme effect

Methods

Parameterization of resistance surfaces

We used binary RSs that are classified as migration habitat or matrix (Fig. 1A,B). By default, migration habitat (mh) was given a resistance value of 1. The value of the matrix was 2^{sf} , where sf is the specialization factor. When sf was 0, the value of matrix and mh were both 1. In this case, migration is isotropic and the species can be regarded as a generalist with regard to the influence of landscape on its migration. The resulting least-cost path from such an RS is a straight line. Increasing sf leads to exponentially increasing matrix resistance and, consequentially, increasing reluctance of a species to move from the migration habitat into the matrix. In this study, we experimented with $sf = 0, 3, 6, 9, 12$ and 15 (Fig. 1C). Furthermore, the following landscape elements were used in the analysis of our example dataset: agricultural land, forests, residential areas, roads and streams (see the Example Dataset section).

Each of the above-mentioned landscape elements was in turn regarded as migration habitat (Fig. 1B). Thus, the two variables that defined the parameterization of each RS scenario were mh and sf . For example, RS(mh = agriculture, $sf = 6$) implied that the migration habitat in this scenario is agricultural land with a resistance value of 1, and the matrix is assigned a resistance value of $2^{sf} = 2^6 = 64$. Similarly, RS(mh = residential, $sf > 6$) refers to all RSs in which the migration habitat was residential areas and the specialization factors were larger than 6. In some scenarios, the migration habitat

was a landscape element that was suspected to be an inhibitor to gene flow. If this is correct, any landscape measure calculated from the corresponding RSs should have a low explanatory power and should have model fits that are equal or lower than average. In total, we parameterized 26 RSs (i.e. 5 migration habitats $\times$ 5 specialization factors larger than zero + 1 specialization factor of zero). Geodata were derived from a vectorized 1:25 000 land cover map of Switzerland (Vector25 © 2011 Swisstopo; DV033594), which was transformed to grids with a 10 m spatial resolution to create the RS scenarios.

Least-cost transect analysis was performed on the binary habitat-matrix RSs. We also calculated least-cost, potentially effective and resistance distances on these habitat-matrix RSs (Table 1). These methods will be explained in the following sections. Unless otherwise specified, all analyses on RSs were performed in ARC-GIS 9.3 (ESRI).

Least-cost transect analysis

For each of the 26 RSs, least-cost paths were calculated between each pair of demes (Fig. 1C). The least-cost paths were subsequently buffered with a certain distance to form a transect (Fig. 1D). We evaluated a variety of transect widths (i.e. 16, 100, 200, 400, 600, 800, 1000, 1200, 1400 and 2000 m). The narrowest transect of 16 m measured the landscape directly on the least-cost path. For each transect, we calculated the transect length (in metres; equivalent to the potential effective distance) and the proportion of each of the landscape elements in the transect ($m^2 \times m^{-2}$; Fig. 1E, Table 1),

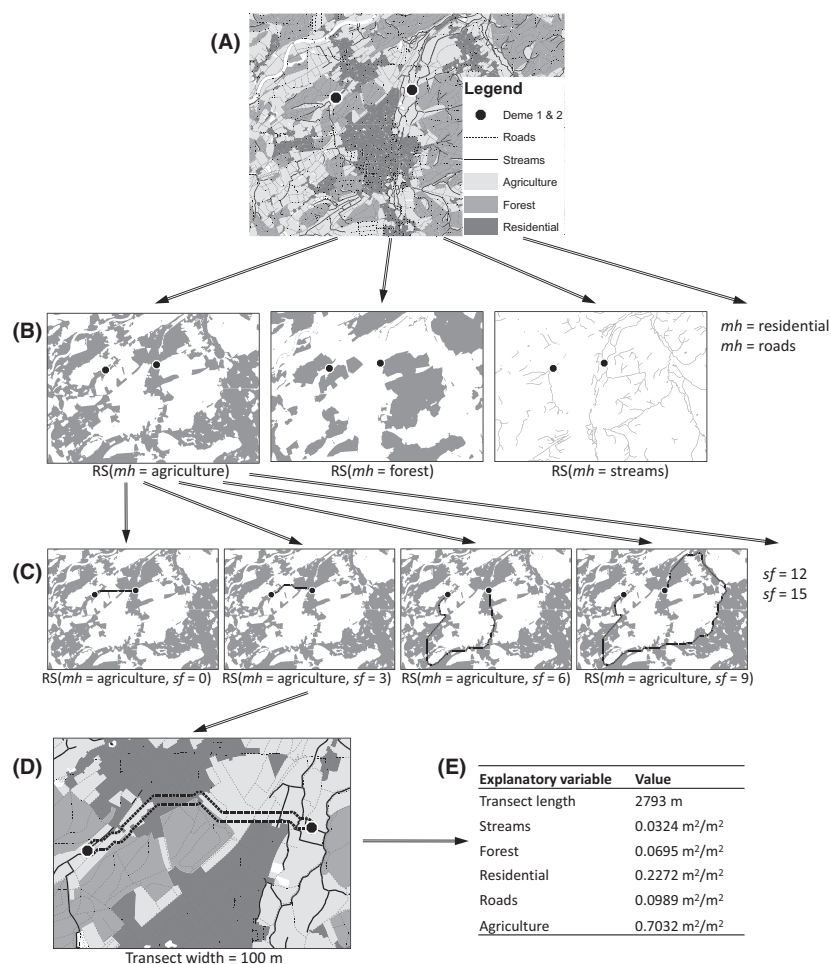


Fig. 1 Flow-chart of the parameterization of resistance surfaces (RS) and the calculation of least-cost transect analysis (LCTA). A, B and C show how the binary habitat-matrix RSs were created. D and E show the calculation of explanatory variables for LCTA. (A) Landscape analysis is demonstrated for two demes, 1 and 2. (B) Separate grids are created for each potential migration habitat (mh; Fig. 2). The resistance value of mh is set to 1. (C) The resistance value of the matrix is 2^{sf} , where sf is the specialization factor. For each migration habitat, RSs are created for sf = 0, 3, 6, 9, 12 and 15. Resistance distances and least-cost paths are calculated on each of the RSs. (D) For LCTA, the least-cost path is buffered to form a transect. (E) The transect length and the proportions of each of the landscape elements in the transect are calculated to form a complementary set of explanatory variables.

resulting in six explanatory variables. These proportions were calculated by dividing the number of grid cells that contain a certain landscape element by the total number of grid cells in a transect. It is probable that the area of a landscape element within a transect is positively correlated with the transect length. Thus, by using proportional measures, we anticipated that the transect length is less correlated with the other explanatory variables.

Least-cost paths calculated from RSs with sf = 0 should theoretically produce straight lines between demes. In practice, however, these paths may have some bends due to the rule settings in the least-cost path algorithm. Thus, for LCTA we calculated truly straight-line paths for sf = 0 to enable the comparison between traditional straight-line transects (i.e. sf = 0)

and buffered least-cost paths (i.e. sf > 0). Python-scripts for LCTA have been made available online (Van Strien *et al.* 2012).

Least-cost, effective and resistance distance analysis

We calculated traditional least-cost, potential effective and resistance distances on each of the 26 RSs. Each of these measures was used as an explanatory variable in separate regression analyses (Table 1). From interdeme least-cost paths, we calculated the total cost (i.e. least-cost distance) and the total length (i.e. potential effective distance). Potential effective distances calculated on RSs with sf = 0 provide a baseline of the level of isolation by Euclidean distance. Some studies suggested to calculate multiple interdeme least-cost paths to give a

better 'average' representation of the landscape encountered by migrating individuals (McRae 2006; Pinto & Keitt 2009; Rayfield *et al.* 2010). For instance, resistance distances are calculated by analysing all possible inter-deme paths through an RS. In some cases, this measure has proven to better explain genetic distances than Euclidean or effective distances (McRae & Beier 2007). Therefore, we also calculated resistance distances between demes with CIRCUITSCAPE 3.5.2 (McRae 2006). To overcome computational problems with CIRCUITSCAPE, the spatial resolution of RSs was transformed to 25 m.

Statistical analysis

The goals of the statistical analysis were to determine (i) the most probable representation of the landscape as experienced by migrating individuals, and (ii) the inter-deme landscape analysis (i.e. LCTA or least-cost, effective or resistance distance analysis) that explained inter-deme genetic distances (F_{ST}) best. Linear mixed effect models can be used to account for dependency between pairwise observations in a distance matrix (Yang 2004). The fixed effect terms in these models are the explanatory variables (Table 1). The random effect term indicates the dependency between the pairwise distances. Several covariate structures from the SAS[®] PROC MIXED package have been proposed to represent this dependency (Yang 2004; Pavlacky *et al.* 2009; Meeuwig *et al.* 2010; Selkoe *et al.* 2010). However, these covariate structures are not tailored for the specific dependency between values in a distance matrix (Clarke *et al.* 2002). Clarke *et al.* (2002) described a covariate structure where a proportion, ρ_τ , of the total variance, σ^2 , is the result of the correlation between two pairwise distances involving a common deme. The covariance between two pairwise distances without a common deme is 0. For example, if Y_{ij} is the genetic distance between deme i and j , then Y_{12} and Y_{13} have a covariance of $\rho\sigma^2$, whereas Y_{12} and Y_{34} have a covariance of 0. The maximum value of ρ_τ is 0.5. The intercept, slopes, ρ_τ and σ^2 of this MLPE model are estimated with REML, which is desirable for unbiased estimates of the variance components of mixed models (Clarke *et al.* 2002; Gurka 2006). We centred all explanatory variables around their mean, so that the REML estimates of the intercept were the same as estimates obtained from simple linear regression (Clarke *et al.* 2002). We fitted MLPE models with REML estimation using the 'lmer' function in the package 'lme4' (Bates *et al.* 2011) for R (R Development Core Team 2011), with a modification to account for multiple memberships (i.e. each pairwise distance was associated with two demes).

In each mixed model, the random effect (i.e. individual deme effect) was standard, and we were interested in comparing models with different fixed effects (i.e. explanatory variables). Therefore, we compared models making use of a marginal R^2 statistic, which is suitable to select the most parsimonious model (Orellien & Edwards 2008). Edwards *et al.* (2008) described the R^2_β statistic, which compares a model with fixed effects to a null model with only the random effect and an intercept. Unlike the R^2 of simple linear regression, the R^2_β can decrease when adding explanatory variables to the model. Such a decrease of R^2_β can be caused by misspecification of the new model or an increase in residual variance due to the inclusion of an explanatory variable. The R^2_β is calculated from the Kenward–Roger F and degrees of freedom estimates (Kenward & Roger 1997), which we calculated using the 'KRmodcomp' function from the R-package 'pbkrtest' (Halekoh & Højsgaard 2011).

It is often not evident from what kind of distribution (dis)similarities and distances are drawn (Smouse *et al.* 1986). Therefore, explanatory variables may show an asymmetrical distribution with highly skewed values, which hinders a standardized normalization approach. Moreover, information is often lacking to determine the most likely relationship between explanatory and dependent variables (e.g. linear, exponential, sigmoidal or logarithmic). As we were mainly interested in investigating monotonic relationships between explanatory and dependent variables, and parameter and intercept estimation were less important, we chose to utilize monotone/nonparametric regression (Legendre & Legendre 1998; Legendre & Fortin 2010). The regression was applied to the rank transformed dependent and explanatory variables, which is a robust statistical method, less influenced by outliers and asymmetrically distributed data (Conover 1999). Several studies successfully applied rank transformation for the analysis of (dis)similarity matrices (e.g. Clarke & Ainsworth 1993; Spear *et al.* 2005). As the goal of many landscape genetic studies is to gain insight into the conservation ecology of a species (Storfer *et al.* 2010), rank transformation is justified. Using landscape genetic models to make predictions of gene flow patterns in space and time will require a different transformation method.

In the first step of the statistical analysis, we regressed pairwise F_{ST} values against the explanatory variables calculated through LCTA for the range of transect widths, which resulted in 260 sets of six explanatory variables (i.e. 26 RSs $\times$ 10 transect widths). Because the full model (including all six explanatory variables) may not be the best fitting, most parsimonious model, we also calculated the R^2_β for models with all possible combinations of the six explanatory variables (=64 combinations). Whichever combination of

explanatory variables produced the highest R^2_{β} , was considered as the most representative reduced model for a certain explanatory variable set. A reduced model may, thus, include only a subset of the six explanatory variables. If a reduced model contained all six explanatory variables, it was analogous to the full model. The transect width belonging to the best fitting reduced model was regarded as the most representative transect width for our focal species.

Subsequently, pairwise F_{ST} values were regressed against the explanatory variables calculated through the least-cost, effective and resistance distance analysis (Table 1). As each of these regression models contained only one explanatory variable, no prior selection for the most parsimonious model was necessary. We compared the R^2_{β} values of these regression models to those obtained from the full and reduced LCTA models. From the best fitting overall reduced model, we used the significance and sign of the slope coefficients to draw conclusions about the influence of landscape variables on genetic distances. To compare results, for the best overall reduced model, we also calculated the slope coefficients for the full model as estimated through traditional multiple regression on distance matrices with 10 000 permutations (MRDM; Smouse *et al.* 1986).

We checked all LCTA explanatory variables obtained from each RS for multicollinearity. We arbitrarily fixed a correlation threshold of Spearman $\rho = 0.7$, above which inference from correlated variables was approached with caution (i.e. sign and value of the slope coefficient can then be unreliable, and the power to detect whether a variable is a significant predictor can be reduced; Legendre & Legendre 1998).

Example dataset

As an example dataset (Van Strien *et al.* 2012), we used data of the southern damselfly *Coenagrion mercuriale* (Charpentier, 1840, Odonata: Coenagrionidae), which is critically endangered in Switzerland (Gonseth & Monnerat 2002). The species inhabits permanently slow-flowing streams with abounding aquatic vegetation and generally avoids shaded areas (Rouquette & Thompson 2005). Although the preferred breeding habitat of the species is fairly well described, little is known about its preferred migration habitat. A mark-recapture study showed that the rate of long-distance migration (>300 m) was not affected by the connectedness of sample sites by suitable breeding habitat and that tall shrub structures along the streams acted as migration barriers (Purse *et al.* 2003). Most individuals are sedentary and only disperse a few metres during their lifetime, but some migrate over greater distances (Purse *et al.* 2003; Rouquette & Thompson 2007).

We applied the following basic landscape classification to our study area: agricultural land, forests, residential areas, roads and streams (Fig. 2). The category 'roads' included predominantly small country roads, but also a couple of busier provincial roads. The category 'streams' included all the above-ground waterways in the study area, except for a major river traversing the study area.

Towards the end of the mating season (July 2009), we sampled male damselflies from all known demes of *C. mercuriale* in an area situated on the Swiss Plateau surrounding the town of Langenthal (Fig. 2). Individuals were collected from stretches of uninterrupted streams, with an average length of 260 m. The centre point of the sampled stream-segments was taken as the geographic location of a deme. For analysis, we selected demes from which 10–45 individuals were sampled and that were at least 250 m apart, resulting in 13 demes separated by an average of 3392 m (Fig. 2). To prevent edge effects, the edges of the RSs were at least 7 km from the nearest deme.

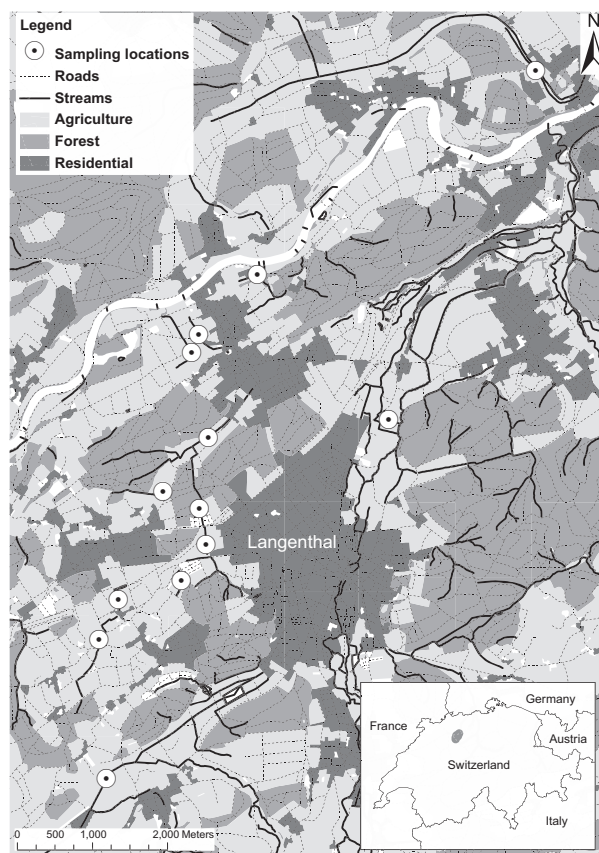


Fig. 2 Map of the study area indicating the landscape classification and the locations where individuals of *Coenagrion mercuriale* were sampled. The inset map shows the location of the study area in Switzerland in grey.

DNA extraction was carried out using the DNeasy Blood and Tissue Kit (QIAGEN) according to the supplementary protocol for insects. We amplified 12 microsatellite loci (LIST4-002, LIST4-023, LIST4-024, LIST4-031, LIST4-034, LIST4-035, LIST4-037, LIST4-042, LIST4-060, LIST4-062, LIST4-063, LIST4-066) developed by Watts *et al.* (2004). Fragments were analysed with an ABI 3730xl sequencer (Applied Biosystems) and genotyped with GENEMAPPER 3.7 (Applied Biosystems). All loci met Hardy-Weinberg equilibrium conditions (F_{IS} ; $P > 0.05$). In some demes, significant linkage disequilibrium was found between loci ($P < 0.05$), but these results were not consistent between demes. We, therefore, presumed loci to be unlinked. Positive controls in each plate always showed the same genotype. There was 0.17%

missing data in the genetic dataset. Due to very high correlations (Pearson $r > 0.984$) between pairwise F_{ST} values and other genetic distance measures (G'_{ST} , G_{ST} and D ; Meirmans & Hedrick 2011), we only use F_{ST} here (Weir & Cockerham 1984). Using GENEPOP 4.0.10 (Rousset 2008), we calculated pairwise F_{ST} values between all 13 demes, resulting in 78 F_{ST} values, which served as the dependent variable throughout all analyses (Table 1). The slightly negative F_{ST} values were set to 0.

Results

For each transect width for LCTA, we listed the three most supported reduced and full models based on R^2_β (Table 2). A transect width of 1000 m produced the

Table 2 The three reduced (left) and full (right) models with the highest R^2_β per transect width class are shown

Transect width (m)	Reduced models			Full models		
	Migration habitat (mh)	Specialization factor (sf)	R^2_β	Migration habitat (mh)	Specialization factor (sf)	R^2_β
16	Streams	6	0.753	Streams	6	0.640
16	Streams	9	0.680	Streams	3	0.626
16	Agriculture	15	0.654	Agriculture	12	0.623
100	Agriculture	3	0.692	Agriculture	3	0.680
100	Agriculture	6	0.676	Agriculture	6	0.669
100	Agriculture	12	0.671	Agriculture	12	0.664
200	Agriculture	12	0.717	Agriculture	12	0.697
200	Agriculture	15	0.716	Agriculture	15	0.697
200	Agriculture	9	0.688	Agriculture	9	0.668
400	Agriculture	12	0.697	Agriculture	12	0.678
400	Agriculture	15	0.697	Agriculture	15	0.677
400	Residential	3	0.669	Residential	3	0.641
600	Agriculture	12	0.745	Agriculture	12	0.725
600	Agriculture	15	0.745	Agriculture	15	0.724
600	Residential	3	0.732	Residential	3	0.702
800	Agriculture	12	0.786	Agriculture	12	0.761
800	Agriculture	15	0.786	Agriculture	15	0.761
800	Agriculture	9	0.758	Agriculture	9	0.731
1000	Agriculture	15	0.810	Agriculture	15	0.776
1000	Agriculture	12	0.809	Agriculture	12	0.776
1000	Agriculture	9	0.788	Agriculture	9	0.755
1200	Agriculture	15	0.805	Agriculture	15	0.766
1200	Agriculture	12	0.804	Agriculture	12	0.765
1200	Residential	3	0.793	Residential	3	0.756
1400	Agriculture	12	0.800	Agriculture	6	0.713
1400	Agriculture	15	0.800	Agriculture	15	0.712
1400	Agriculture	9	0.773	Agriculture	12	0.712
2000	Agriculture	12	0.787	Agriculture	15	0.740
2000	Agriculture	15	0.787	Agriculture	12	0.740
2000	Agriculture	9	0.742	Agriculture	9	0.675

LCTA, least-cost transect analysis.

For each width class, LCTA models were calculated for all combinations of potential migration habitats (agricultural land, forest, residential areas, roads or streams) and specialization factors (0, 3, 6, 9, 12 or 15). The LCTA model with the overall highest R^2_β is shown in bold. All R^2_β were significant at $P < 0.001$.

highest overall R^2_β (0.810) for LCTA on RS (mh = agriculture, sf = 15). A transect width of 1000 m was henceforth considered optimal for our focal species. However, regardless of transect width, the top full and reduced models were predominantly from LCTA on RS(mh = agriculture, sf > 0). An exception were the very narrow transects (width = 16 m), which showed highest R^2_β (0.753) for LCTA on RS(mh = streams, sf = 6).

In the second part of the analysis, the results of LCTA (with 1000 m transect width) were compared with those obtained from least-cost distance, effective distance and

resistance distance analysis (Table 3, Fig. 3). LCTA performed on RS(mh = residential, sf = 3) produced a high R^2_β (0.786). However, for LCTA performed on RSs with sf > 3, the migration habitat that always resulted in the highest R^2_β was agriculture ($0.745 \leq R^2_\beta \leq 0.810$; Table 3). Potential effective distances calculated on RS(mh = agriculture, sf > 0) consistently produced high R^2_β values (≥ 0.429) compared with potential effective distances calculated from other RS scenarios (Fig. 3C). Least-cost distances calculated on RS(mh = roads, sf = 6) produced the highest R^2_β (0.529) of all least-cost distances. However, model fit fluctuated substantially

Table 3 Resistance surfaces were created by varying migration habitat and the specialization factor (sf)

Specialization factor (sf)	Migration habitat (mh)	LCTA				Least-cost distance analysis	Effective distance analysis	Resistance distance analysis
		Reduced		Full		R^2_β	R^2_β	R^2_β
		R^2_β	$\rho > 0.7$	R^2_β	$\rho > 0.7$			
0		0.613		0.608	+	0.396	0.384	0.384
3	Streams	0.614		0.611		0.391	0.424	0.375
3	Forest	0.504		0.489	+	0.398	0.381	0.375
3	Residential	0.786		0.748		0.258	0.401	0.328
3	Roads	0.623		0.623		0.478	0.443	0.489
3	Agriculture	0.668		0.659	+	0.477	0.522	0.262
6	Streams	0.526		0.522		0.361	0.328	0.444
6	Forest	0.487		0.467	+	0.371	0.350	0.269
6	Residential	0.719		0.598	+	0.172	0.310	0.283
6	Roads	0.579		0.579		0.529	0.429	0.446
6	Agriculture	0.745		0.731	+	0.458	0.429	0.285
9	Streams	0.546		0.544		0.374	0.336	0.423
9	Forest	0.375		0.367	+	0.352	0.240	0.271
9	Residential	0.718		0.598	+	0.162	0.310	0.253
9	Roads	0.583		0.583		0.367	0.430	0.260
9	Agriculture	0.788		0.755	+	0.452	0.485	0.331
12	Streams	0.544		0.544		0.377	0.336	0.393
12	Forest	0.378		0.369	+	0.344	0.239	0.271
12	Residential	0.718		0.597	+	0.163	0.310	0.244
12	Roads	0.583		0.583		0.249	0.430	0.222*
12	Agriculture	0.809		0.776	+	0.327	0.547	0.286
15	Streams	0.544		0.544		0.377	0.336	0.372
15	Forest	0.383		0.374	+	0.346	0.237	0.272
15	Residential	0.719		0.597	+	0.163	0.308	0.244
15	Roads	0.583		0.583		0.225*	0.430	0.214*
15	Agriculture	0.810		0.776	+	0.354	0.546	0.257

LCTA, least-cost transect analysis; RS, resistance surface.

Sf = 0 results in a homogeneous RS. For each RS, the fit (R^2_β) of pairwise F_{ST} values regressed against the explanatory variables from different landscape genetic methods was assessed. From the LCTA, we display the R^2_β of the most parsimonious, best fitting reduced model (i.e. the model with the highest R^2_β from models with all possible combinations of explanatory variables) and the R^2_β from the full model (i.e. model including all explanatory variables). Models with the highest overall R^2_β per sf are indicated in bold. Spearman correlation coefficients ρ were calculated between all the explanatory variables of each LCTA model. Only several full models had overly correlated ($\rho > 0.7$) explanatory variables (indicated with a '+'). All models were significant at $P < 0.05$, except for R^2_β values denoted with * ($P > 0.05$).

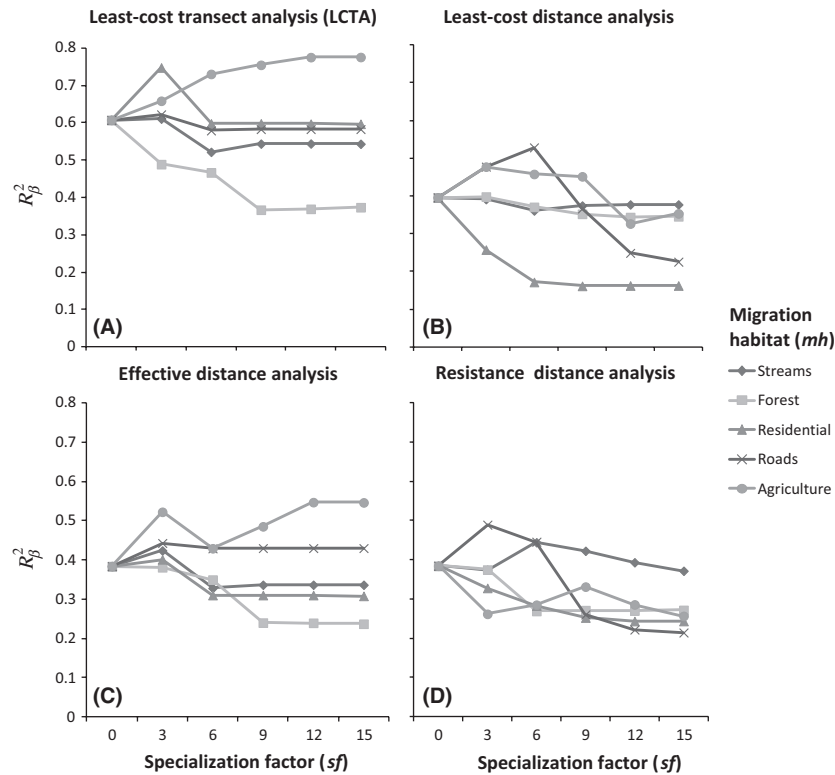


Fig. 3 The graphs show the model fit (R^2_β) of F_{ST} values monotonically regressed against explanatory variables from the different landscape genetic methods (Table 1). Binary habitat-matrix resistance surfaces (RS) were created, in which each landscape element (agricultural land, forest, residential areas, roads and streams) was in turn considered to be migration habitat, which was assigned a resistance value of 1. The resistance value of the matrix was determined by the specialization factor (sf). For each of these RSs, explanatory variables were calculated with least-cost transect analysis (LCTA) (A), least-cost distance analysis (B), effective distance analysis (C) and resistance distance analysis (D). All models were fit with linear mixed effect models, and the R^2_β was calculated from the Kenward–Roger F and degrees of freedom estimates. For LCTA the R^2_β from full models is displayed.

with increasing sf (Fig. 3B). A similar pattern was observed for the resistance distance analysis (Fig. 3D), where resistance distances calculated on RS(mh = roads, sf = 3) produced the highest R^2_β (0.489). All explanatory variables from the reduced and full LCTA models were checked for multicollinearity. Many of the full LCTA models had at least one pair of overly correlated ($\rho > 0.7$) explanatory variables, whereas none of the reduced models showed multicollinearity (Table 3).

Inference, regarding the most likely migration habitat and the gene flow inhibiting or facilitating properties of landscape elements, was drawn from LCTA on RS(mh = agriculture, sf = 15), because this model produced the highest overall R^2_β (0.810; Tables 3, 4). The explanatory variables that were selected in this reduced model were the transect length and proportions of streams, forest, residential areas and roads in the 1000-m wide transect. However, not all of these explanatory variables were significant (t -value; $P > 0.05$). The proportion of residential areas and the transect length had a significant positive monotonic correlation with F_{ST} .

The proportion of streams had a significant negative monotonic correlation with F_{ST} . The correlation between two pairwise distances involving a common deme ρ_τ was 0.053 (Table 4). A comparison of these results to an MLPE model with all six explanatory variables (i.e. full model) showed a lower R^2_β (0.776) and a higher ρ_τ (0.058). Also, the proportion of residential areas became a nonsignificant explanatory variable (Table 4). Significance and slope estimates of this full MLPE model differed slightly from those estimated by multiple regression on distance matrices (Table 4).

Discussion

With LCTA, binary RSs were created in which each selected landscape element in turn was regarded as migration habitat, which proved to be an objective method to determine the most probable migration habitat of the focal species. We found strong support for agricultural land being the most probable migration habitat of *C. mercuriale*. For most transect widths and

Table 4 Significance and slope estimates of the regression coefficients of three models

Specialization factor (sf)	Migration habitat (mh)	Statistical method	Model fit	Intercept	Streams	Forest	Residential	Roads	Agriculture	Transect length	Within-deme correlation ρ_t
15	Agriculture	Reduced MLPE model	$R^2_\beta = 0.810^{***}$	39.5***	-0.421***	0.081 ns	0.231*	-0.160 ns	-	0.7348***	0.053
15	Agriculture	Full MLPE model	$R^2_\beta = 0.776^{***}$	39.5***	-0.412***	-0.029 ns	0.194 ns	-0.156 ns	-0.125 ns	0.7479***	0.058
15	Agriculture	Full MRDM model	$R^2 = 0.701^{***}$	39.5***	-0.394***	-0.006 ns	0.188 ns	-0.172 ns	-0.089 ns	0.752***	-

LCTA, least-cost transect analysis; MLPE, maximum-likelihood population-effects; RS, resistance surface.

For all models, pairwise F_{ST} values were the dependent variable. Transect width of 1000 m was maintained for all three models. Reduced models were selected by taking the model with the highest R^2_β from models with all possible combinations of explanatory variables. Explanatory variables that were not used in a reduced model are indicated with a '-'. Full models were fitted making use of all six explanatory variables (Table 1). The first two models are fitted through MLPE models, and the significance of the explanatory variables was determined from t -values. For these models, we display the ρ_t , which is the correlation between two pairwise distances involving a common deme. The first model is the reduced MLPE model from LCTA on RS(mh = agriculture, sf = 15), which was the overall best fitting model. The second model is the full MLPE model from LCTA on RS(mh = agriculture, sf = 15). Making use of the same explanatory variables, the third model was fitted with multiple regression on distance matrices (MRDM), and significance was tested with 10 000 permutations of the dependent variable. ns = nonsignificant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

specialization factors, RSs, in which the migration habitat was agricultural land, consistently produced the most supported models. Taking into account that *C. mercuriale* has a strong preference for unshaded habitats (Purse *et al.* 2003; Rouquette & Thompson 2005), it is not surprising that it prefers to disperse through open agricultural land. Additionally, results from LCTA gave a clear indication of which landscape elements may be inhibiting or facilitating gene flow in these transects through agricultural land. We found that especially the length of the transect through agricultural land decreased gene flow (i.e. positive correlation between transect length and F_{ST} ; Table 4), which is a logical result of the sedentary character of the study species (Purse *et al.* 2003; Rouquette & Thompson 2007). The positive effect of stream density on gene flow (Table 4) can be expected for a species that is highly dependent on water for its reproduction (Rouquette & Thompson 2005). Apart from transect length, the only other inhibitor of gene flow was the proportion of residential areas along the migration route through agricultural land.

If we had used effective distance analysis, the conclusions about the most probable migration habitat would have been similar. Potential effective distances through agricultural land were on average better explanatory variables for genetic distances than Euclidean distance or potential effective distances through other landscape elements (Fig. 3C). However, least-cost and resistance distance analysis did not result in any migration habitat that consistently produced models with high R^2_β , which complicated the interpretation.

Least-cost transect analysis performed on RSs with a specialization factor of 3 suggested that residential areas were the most likely migration habitat (Fig. 3A), which demonstrates that it is also important to create RS scenarios with a range of specialization factors. For narrow transects of 16 m, we found a strong support for streams to be the most probable migration habitat, albeit with a lower R^2_β than the overall best performing model (Table 2). Many of the streams that *C. mercuriale* inhabits in the study area are either unconnected or run, at least partially, through a forest. Therefore, we presume agricultural land to be the most probable migration habitat of *C. mercuriale*. Nevertheless, this discrepancy between the results of various transect widths indicates how sensitive transect-based approaches are to transect width. Perhaps LCTA with multiple least-cost interdeme paths would provide a more realistic indication of where the main migration routes are located (McRae 2006; Pinto & Keitt 2009; Rayfield *et al.* 2010). Average landscape composition from all transects, drawn along each of the paths, would function as explanatory variables for the landscape

genetics analysis. Future research should determine whether such a setup is less sensitive to transect width and will provide a more explicit distinction between 'nonsense' migration habitats and the 'true' migration habitat.

Least-cost transect analysis can be seen as an extension to the landscape genetic method presented by Spear *et al.* (2005). Between salamander populations, the authors created several potential migration paths based on various migration hypotheses. The fit of models with Euclidean distance as explanatory variable was compared with that of models with path lengths and proportion of various landscape elements directly along these migration paths as explanatory variables. Their model R^2 increased from 0.337 to 0.828 (for the best fitting model), respectively, which is similar to the increase observed in this study (increase in R^2_β from 0.384 to 0.810). This substantiates our recommendation to use a variety of potential migration paths, whether they are straight-line, least-cost or multiple shortest paths, as a means to shape transects, in which the landscape can be quantified.

Binary habitat-matrix RSs were used in this study. An inherent assumption of least-cost paths calculated on these surfaces is that organisms move in a straight line through a homogeneous matrix and do not alter their direction to avoid or target certain landscape elements in the matrix. This is a criticized approach, because empirical studies have shown that species react differently to different matrix elements (Ricketts 2001; Prevedello & Vieira 2010). With LCTA, we accommodated for matrix elements by calculating the proportion of all landscape elements in the transects. Spear *et al.* (2010) criticized this approach because multiple landscape variables are taken into account along a least-cost path that itself is calculated on a binary RS. We acknowledge that the course of least-cost paths used in LCTA may change if RSs would represent a heterogeneous matrix. However, gene flow is a two-fold process of active migration decisions and passive events. A landscape element, like a road or a city, can be actively avoided, in which case the focal species may seek a route bypassing the element, whereby migration distance increases and other landscape elements are encountered. The same landscape element can also passively impede gene flow, by reducing the survival chances of an individual when crossing the landscape element. In both cases, gene flow is negatively influenced by the landscape element, whereas the movement paths and assigned resistance values would be quite different. Moreover, it is important to consider that RSs should incorporate the variables that a focal species would use to make a decision about its next movement step (i.e. movement from one grid cell into one of the

neighbouring grid cells). Multivariate RSs would require the focal species to make full rational decisions, where all possible landscape variables are weighed against each other, to seek an optimal solution for its next movement step. However, there is growing support among psychologists for the bounded rationality theory, which states that many decisions are based on a very small set of variables, and are made to satisfy the needs of an organism rather than to find an optimal solution (Gigerenzer & Goldstein 1996; Srinivasan & Mukherjee 2010). Taking these considerations, and other difficulties involved in translating species' behaviour into RSs (Adriaensen *et al.* 2003; Driezen *et al.* 2007; Spear *et al.* 2010; Gillies *et al.* 2011), into account, we favour a parsimonious landscape model in the form of simple binary habitat-matrix RSs.

The spatial and thematic resolution of RSs influences the outcome of landscape genetic studies that use least-cost paths (Cushman & Landguth 2010). Additionally, with LCTA, the spatial and thematic resolution of the landscape in the transects may influence the values of the explanatory landscape variables. Compared with the proportion of a landscape element in the transect, other landscape metrics may be less influenced by these resolutions (Wu 2004) and may even produce interesting additional explanatory variables. A further point of consideration is that transects between unique deme pairs may be spatially overlapping, whereby creating a spatial dependency between these deme pairs (Holland *et al.* 2004). Also, the species' response to migration habitat and other landscape elements may show temporal and spatial variations, caused by, for instance, the level of human disturbance (e.g. Hebblewhite & Merrill 2008) or by functional responses, where the strength of the behavioural response to a landscape element is proportional to the available amount of that landscape element (Mysterud & Ims 1998). The above-mentioned issues may influence the results of LCTA and other landscape genetic methods and deserve more attention in landscape genetic research (Anderson *et al.* 2010). Perhaps solutions may be presented by studies on resource selection by animals (Gillies *et al.* 2006), in which many of these issues have been addressed.

An advantage of the transect-based approach is that multiple explanatory variables are created, which are each ecologically interpretable (Emaresi *et al.* 2011). To statistically analyse several explanatory variables, we apply a multiple regression approach based on MLPE models (Clarke *et al.* 2002). The ρ_τ value, estimated through REML, indicates how strong the correlation is between pairwise distances involving a common deme. A high ρ_τ value (maximum ρ_τ value is 0.5) indicates the need to take the correlated structure of the data into account when fitting regression models (Clarke *et al.*

2002). For all 26 LCTA reduced MLPE models with a transect width of 1000 m, we found an average $\rho_\tau = 0.183$ (data not shown), which indicates a high correlated structure within demes and compels the use of MLPE models. As expected with such high ρ_τ values, we found (small) differences in slope coefficient estimates between full MLPE and MRDM models (Table 4; as in Clarke *et al.* 2002). However, for the reduced MLPE model with the highest R^2_β , we found $\rho_\tau = 0.053$ (Table 4). Apparently, most of the variance in the latter model was explained by the between-deme landscape effects, and the within-deme correlation was limited.

Few studies have focused on developing statistical tests for model selection in mixed effect models estimated through REML (Gurka 2006; Edwards *et al.* 2008; Orelie & Edwards 2008). Here, we used the R^2_β statistic for model selection (Edwards *et al.* 2008). This statistic has proven to be able to select the most parsimonious model (Orelie & Edwards 2008). We found that model selection based on R^2_β was an effective tool to remove overly correlated (i.e. Spearman $\rho > 0.7$) explanatory variables from the model (Table 3). An additional advantage of this statistic is that partial- R^2_β values can easily be calculated (Edwards *et al.* 2008). As in Edwards *et al.* (2008), we observed that the reduced model with the highest R^2_β contained nonsignificant explanatory variables (Table 4). The inclusion of (nonsignificant) explanatory variables can result in a decrease in the total variance estimate by REML, which can increase the R^2_β estimate. However, we are not aware of any guidelines to determine how much increase in statistical support for a certain model is associated with a certain increase in R^2_β (like, for instance, AIC-weights; Burnham & Anderson 2002). We stress that further research is needed to determine whether this approach to variable selection is optimally suited for landscape genetic purposes, especially in combination with the covariate structure associated with the MLPE method.

Information criteria like AIC are commonly used for variable selection in landscape genetics (e.g. Yang 2004; Spear *et al.* 2005; Pavlacky *et al.* 2009; Goldberg & Waits 2010; Meeuwig *et al.* 2010; Selkoe *et al.* 2010; Emaresi *et al.* 2011). However, in contrast to AIC values from simple linear regression or mixed models estimated with maximum likelihood, AIC calculated from REML in linear mixed models is not recommended for comparing models with different fixed effects (Verbeke & Molenberghs 2000; Orelie & Edwards 2008), although this is not always clearly stated in software documentation (Gurka 2006). When slope and intercept are estimated with ordinary least squares (e.g. MRDM), model selection with AIC is also prone to be biased by the nonindependence of pairwise distances (Goldberg &

Waits 2010). The AIC difference compared with the model with the lowest AIC determines the relative statistical support for a certain model. This difference can be inflated if observations are falsely assumed to be independent (Burnham & Anderson 2002), thereby leading to overly strong support for the model with the lower AIC. Also, because the true number of degrees of freedom are unknown in distance matrices (Legendre & Fortin 1989), measures that correct for small sample size may be biased. For instance, no valid method has yet been proposed to calculate adjusted- R^2 of a regression on distance matrices (Legendre & Fortin 2010). We assume that this is also the case for AIC values adjusted for small sample size (AICc). To be able to use AIC for model selection in landscape genetics, Goldberg & Waits (2010) proposed to remove nonindependent demes from the dataset. They removed a deme if its average genetic distance rate (genetic distance/Euclidean distance) between neighbouring demes (determined from Delaunay triangulation) was significantly different to a null distribution of rates created by 10 000 bootstrap samples from all deme pairs. A disadvantage of this approach is that not all pairwise distances can be used in the analysis (i.e. only the demes connected through Delaunay triangulation). Furthermore, we argue that a significant difference in genetic distance rates, between a random sample and a deme and its neighbours, may actually be indicative of this deme being extraordinarily isolated from its neighbours. Removing this deme from the analysis could reduce the knowledge gained from the study. As yet, we believe there is no completely unbiased AIC model selection method available for multiple regression in landscape genetics (either through MLPE or MDRM models).

Several authors have underlined the need for robust model selection techniques in landscape genetics (Selkoe *et al.* 2010; Emaresi *et al.* 2011). Here, we apply a new approach to variable selection in landscape genetics. We believe that the use of mixed effect models, with an appropriate covariance structure, can account for the nonindependence of values in distance matrices and that R^2_β statistics may be appropriate for variable selection in such models. The within-deme correlation ρ_τ can be used as a measure to determine the necessity to account for the nonindependence in distance matrices when fitting the regression model. We acknowledge, however, that these methods require more stringent testing with empirical and simulated data before becoming standard landscape genetic methods. Furthermore, we show that LCTA is a powerful and easily interpretable landscape genetic approach. In the empirical example dataset, a certain migration habitat was consistently selected as the most probable migration habitat for the focal species, which was not the case with least-

cost or resistance distance analysis. Further research should be carried out to determine whether these findings are consistent in other species. We strongly advocate the use of transect-based approaches in landscape genetics, as a separate explanatory variable is calculated for each landscape element, which greatly facilitates ecological interpretation.

Acknowledgements

M.v.S. thanks Jochen Jaeger for providing him with facilities and new insights while programming LCTA. Permits for sampling *C. mercuriale* were provided by the Cantons of Solothurn and Bern. Parts of the genetic analysis were performed at the Genetic Diversity Centre GDC of ETH Zurich. We thank Lisette Waits and five anonymous reviewers for their comments on an earlier version of the manuscript. This study is part of the ENHANCE project, which was funded by the Competence Center Environment and Sustainability of the ETH Domain.

References

- Adriaensen F, Chardon JP, De Blust G *et al.* (2003) The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, **64**, 233–247.
- Anderson CD, Epperson BK, Fortin M-J *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565–3575.
- Angelone S, Kienast F, Holderegger R (2011) Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography*, **34**, 714–722.
- Bates D, Maechler M, Bolker B (2011) lme4: linear mixed-effects models using Eigen and R. R package version 0.999375-999342.
- Bellamy PE, Rothery P, Hinsley SA (2003) Synchrony of woodland bird populations: the effect of landscape structure. *Ecography*, **26**, 338–348.
- Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer, New York.
- Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. *Marine Ecology Progress Series*, **92**, 205–219.
- Clarke RT, Rothery P, Raybould AF (2002) Confidence limits for regression relationships between distance matrices: estimating gene flow with distance. *Journal of Agricultural Biological and Environmental Statistics*, **7**, 361–372.
- Conover WJ (1999) *Practical Nonparametric Statistics*. Wiley, New York.
- Coulon A, Cosson JF, Angibault JM *et al.* (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.
- Cushman SA, Landguth EL (2010) Scale dependent inference in landscape genetics. *Landscape Ecology*, **25**, 967–979.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486–499.
- Davis ES, Murray TE, Fitzpatrick U, Brown MJF, Paxton RJ (2010) Landscape effects on extremely fragmented populations of a rare solitary bee, *Colletes floralis*. *Molecular Ecology*, **19**, 4922–4935.
- Driezen K, Adriaensen F, Rondinini C, Doncaster CP, Matthysen E (2007) Evaluating least-cost model predictions with empirical dispersal data: a case-study using radiotracking data of hedgehogs (*Erinaceus europaeus*). *Ecological Modelling*, **209**, 314–322.
- Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O (2008) An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, **27**, 6137–6157.
- Emaresi G, Pellet J, Dubey S, Hirzel AH, Fumagalli L (2011) Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. *Conservation Genetics*, **12**, 41–50.
- Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. *International Journal of Psychology*, **31**, 1315.
- Gillies CS, Hebblewhite M, Nielsen SE *et al.* (2006) Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology*, **75**, 887–898.
- Gillies CS, Beyer HL, St Clair CC (2011) Fine-scale movement decisions of tropical forest birds in a fragmented landscape. *Ecological Applications*, **21**, 944–954.
- Goldberg CS, Waits LP (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology*, **19**, 3650–3663.
- Gonseth Y, Monnerat C (2002) *Rote Liste der gefährdeten Libellen der Schweiz*. BAFU, Bern, Switzerland.
- Gurka MJ (2006) Selecting the best linear mixed model under REML. *American Statistician*, **60**, 19–26.
- Hagerty BE, Nussear KE, Esque TC, Tracy CR (2011) Making molehills out of mountains: landscape genetics of the Mojave desert tortoise. *Landscape Ecology*, **26**, 267–280.
- Halekoh U, Højsgaard S (2011) pbkrtest: parametric bootstrap and Kenward Roger based methods for mixed model comparison. R package version 0.2.0.
- Hebblewhite M, Merrill E (2008) Modelling wildlife-human relationships for social species with mixed-effects resource selection models. *Journal of Applied Ecology*, **45**, 834–844.
- Holderegger R, Wagner HH (2008) Landscape genetics. *BioScience*, **58**, 199–207.
- Holland JD, Bert DG, Fahrig L (2004) Determining the spatial scale of species' response to habitat. *BioScience*, **54**, 227–233.
- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997.
- Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.
- Legendre P, Fortin MJ (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.
- Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam, The Netherlands.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

- McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.
- McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19885–19890.
- Meeuwig MH, Guy CS, Kalinowski ST, Fredenberg WA (2010) Landscape influences on genetic differentiation among bull trout populations in a stream-lake network. *Molecular Ecology*, **19**, 3620–3633.
- Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.
- Mysterud A, Ims RA (1998) Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecology*, **79**, 1435–1441.
- Orelien JG, Edwards LJ (2008) Fixed-effect variable selection in linear mixed models using statistics. *Computational Statistics & Data Analysis*, **52**, 1896–1907.
- Pavlacky DC, Goldizen AW, Prentis PJ, Nicholls JA, Lowe AJ (2009) A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology*, **18**, 2945–2960.
- Perez-Espona S, Perez-Barberia FJ, McLeod JE *et al.* (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, **17**, 981–996.
- Pinto N, Keitt TH (2009) Beyond the least-cost path: evaluating corridor redundancy using a graph-theoretic approach. *Landscape Ecology*, **24**, 253–266.
- Prevedello JA, Vieira MV (2010) Does the type of matrix matter? A quantitative review of the evidence. *Biodiversity and Conservation*, **19**, 1205–1223.
- Purse BV, Hopkins GW, Day KJ, Thompson DJ (2003) Dispersal characteristics and management of a rare damselfly. *Journal of Applied Ecology*, **40**, 716–728.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rayfield B, Fortin MJ, Fall A (2010) The sensitivity of least-cost habitat graphs to relative cost surface values. *Landscape Ecology*, **25**, 519–532.
- Ricketts TH (2001) The matrix matters: effective isolation in fragmented landscapes. *American Naturalist*, **158**, 87–99.
- Rouquette JR, Thompson DJ (2005) Habitat associations of the endangered damselfly, *Coenagrion mercuriale*, in a water meadow ditch system in southern England. *Biological Conservation*, **123**, 225–235.
- Rouquette JR, Thompson DJ (2007) Patterns of movement and dispersal in an endangered damselfly and the consequences for its management. *Journal of Applied Ecology*, **44**, 692–701.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Schwartz MK, Copeland JP, Anderson NJ *et al.* (2009) Wolverine gene flow across a narrow climatic niche. *Ecology*, **90**, 3222–3232.
- Selkoe KA, Watson JR, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708–3726.
- Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.
- Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology*, **14**, 2553–2564.
- Spear SF, Balkenhol N, Fortin MJ, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.
- Srinivasan N, Mukherjee S (2010) Attribute preference and selection in multi-attribute decision making: implications for unconscious and conscious thought. *Consciousness and Cognition*, **19**, 644–652.
- Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh RA, Baguette M (2006) Gene flow and functional connectivity in the natterjack toad. *Molecular Ecology*, **15**, 2333–2344.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.
- Van Strien MJ, Keller D, Holderegger R (2012) Data from: a new analytical approach to landscape genetic modeling: least-cost transect analysis and linear mixed models. *Dryad Digital Repository*, DOI: 10.5061/dryad.08m6t.
- Verbeke G, Molenberghs G (2000) *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- Watts PC, Wu JH, Westgarth C, Thompson DJ, Kemp SJ (2004) A panel of microsatellite loci for the Southern damselfly, *Coenagrion mercuriale* (Odonata: Coenagrionidae). *Conservation Genetics*, **5**, 117–119.
- Weir BS, Cockerham CC (1984) Estimating F -statistics for the analysis of populations structure. *Evolution*, **38**, 1358–1370.
- Wu J (2004) Effects of changing scale on landscape pattern analysis: scaling relations. *Landscape Ecology*, **19**, 125–138.
- Yang R-C (2004) A likelihood-based approach to estimating and testing for isolation by distance. *Evolution*, **58**, 1839–1845.

This study is part of M.vS.'s PhD thesis on the development of landscape genetic methods and their application to insect species in fragmented agricultural landscapes. D.K. focuses her research on landscape and conservation genetics of Orthoptera and Odonata species. R.H. is interested in landscape genetics and its application to conservation management.

Data accessibility

Microsatellite data: Genbank accession nos BV012675, BV012676, BV012719, BV012678, BV012679, BV012680, BV012720, BV012721, BV012682, BV012683, BV012726 and BV012684.

Deme coordinates, pairwise F_{ST} values, individual genotypes, explanatory variables of the best performing overall LCTA model and Python script for LCTA: Dryad Digital Repository. doi:10.5061/dryad.08m6t