



Accounting for relatedness and spatial structure to improve plant phenotypic selection in the wild

Francisco E. Fontúrbel¹ · Pedro F. Ferrer² · Caren Vega-Retter³ · Rodrigo Medel³

Received: 22 June 2020 / Accepted: 31 October 2020
© Springer Nature Switzerland AG 2020

Abstract

Identifying natural selection in wild plant populations is a challenging task, as the reliability of selection coefficients depends, among other factors, on the critical assumption of data independence. While rarely examined, selection coefficients may be influenced by the spatial and genetic dependence among plants, which violates the independence criterion, leading to biased selection estimates. In this study, we examine the extent to which frugivore-mediated selection coefficients are influenced by spatial and genetic information. We used Generalized Additive Models to deal with spatial and relatedness issues. We compared the fit of the Lande and Arnold multivariate model with models including spatial, genetic relatedness, and spatial + genetic relatedness corrections. Our results indicate that fit in standard models was substantially increased after including the spatial structure. Likewise, the model including the genetic relatedness accounted for a variance fraction not explained by spatial structure, which permitted the identification of significant selection acting upon fruit size, a trait not detected under selection otherwise, and dealt better with autocorrelation than any other model. The model including spatial and genetic effects altogether accounted for 65% of the variance, compared to 13% of the standard model. The spatial structure and genetic relatedness played an important role in this system. As genetic effects revealed significant selection upon fruit traits otherwise hidden under standard selection estimates, field studies that control for plant dependency may provide more realistic selection estimates in natural plant populations.

Keywords *Dromiciops gliroides* · Generalized additive models · Microsatellite markers · Relatedness · Spatial autocorrelation · *Tristerix corymbosus*

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10682-020-10089-3>) contains supplementary material, which is available to authorized users.

✉ Francisco E. Fontúrbel
fonturbel@gmail.com

¹ Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

² Instituto Oncológico, Fundación Arturo López Pérez, Santiago, Chile

³ Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

Introduction

Assessing the impact of natural selection on traits involved in the persistence of wild populations is a central question in evolutionary ecology. Those effects are usually estimated from the covariance between individual fitness and the phenotypic traits of interest (Endler 1986). However, despite considerable advances in recent years, studying natural selection in the wild may be a challenging task as confounding factors often compromise data independence. In a seminal paper, Lande and Arnold (1983) proposed an approach based on multiple linear regression methodology to estimate standardized selection coefficients that provide information on the direction and magnitude of selection experienced by phenotypic traits. Since then, Lande and Arnold's equations have been extensively used to assess natural selection in various natural systems (see a review in Kingsolver et al. 2012). Despite the widespread implementation of Lande and Arnold's methodology, there are some key statistical assumptions which are rarely examined: (1) independence of fitness values, (2) lack of error in phenotype measurements, and (3) normality of fitness error. Because of the inherent complexity of ecological systems, however, most studies are unable to fully meet these requirements (Rausher 1992; Pemberton 2010), resulting in unknown error distribution of selection estimates. For example, when some individuals contribute disproportionately to the mean population fitness, the resulting error fitness distribution may differ from the normality assumption as it is often the case with the seed rain in some plant-frugivore systems (Jordano 1995; Weber and Kolb 2013). Likewise, biased selection estimates may occur from heterogeneous environments due to spurious fitness-trait covariation (Rausher 1992). As indicated by Rausher (1992), this problem may be more complex as it compromises the independence of the dataset, inflating the probability of type-I error. In some cases, this shortcoming may be avoided by having an appropriate sampling design defined a priori. However, in most situations, biological systems include an inherent degree of non-independence as a result of natural ecological heterogeneity (Pemberton 2010), which cannot be controlled by sampling design (e.g., biotic seed dispersal, host specificity among others). Perhaps the most frequent situation of data non-independence in ecological studies occurs in sessile organisms, such as plants, where the spatial distribution of abiotic resources like water, minerals, and nutrients, influence the spatial distribution of reproductive individuals hence violating the assumption of statistical independence among them (Fortin and Dale 2005; Legendre et al. 2002; Waser and Mitchell 1990). Recently, Marrot et al. (2015) proposed different alternatives to take spatial autocorrelation into account in selection analyses, such as spatially explicit models that provide more accurate selection coefficients (i.e., differentials and gradients of selection).

While accounting for spatial autocorrelation is the first step to solving data dependence problems in selection analyses, genetic relatedness is another important issue, albeit rarely considered. Individuals within a plant population share a number of alleles as a consequence of limited dispersal and gene flow during reproduction (Galloway 2005; Donohue 1999; Walsh and Lynch 2018). As individual responses to immediate microenvironmental conditions will be clearly influenced by genotype and genotype x environment interactions, it is not unreasonable to expect that genetic relatedness and its spatial distribution are important hidden factors affecting selection estimates in wild populations. These issues may be highly relevant for some plant lifeforms such as mistletoes, as they depend on host plants for nutrients and water, and on biotic vectors to safely arrive and establish on host plants (Rawsthorne et al. 2011; Watson and Rawsthorne 2013). As in many host-parasite systems, mistletoes often present clumped distributions determined by the

spatial arrangement of hosts and the behavior of biotic vectors responsible for seed dispersal (Medel et al. 2004). It has been shown that mistletoe clumps have high levels of genetic structure and relatedness (Fontúrbel et al. 2019), making mistletoe aggregations a useful study system to assess the independent contribution of spatial and genetic relatedness on selection coefficient estimates in wild plant populations.

This study aims to examine the extent to which the ability to detect selection of fruit traits involved in the seed dispersal process is influenced by the spatial structure and genetic relatedness among individuals. We compared the performance of different statistical models, including the spatial structure and relatedness covariates, using a frugivore-mistletoe system. More specifically, in this study we examined the independent and combined contribution of genetic relatedness and spatial structure effects on models used for Lande and Arnold's phenotypic selection analyses using a mistletoe-host biological model.

Materials and methods

Dataset and study model

To test our spatial-and-genetic relatedness explicit approach, we reanalyzed the dataset of Fontúrbel and Medel (2017), which is freely available from the *figshare* repository (<https://doi.org/10.6084/m9.figshare.4614769>). This dataset contains information about the seed dispersal interaction between the hemiparasitic mistletoe *Tristerix corymbosus* (Loranthaceae) and its sole legitimate disperser (southwards 37°S), the relict marsupial *Dromiciops gliroides* (Amico et al. 2011). *Tristerix corymbosus* is a hemiparasitic mistletoe that is common in central and southern Chile. While in central Chile sclerophyllous forests this mistletoe is dispersed by birds (mainly the Chilean Mockingbird *Mimus thenca*), in the temperate rainforests of southern Chile is almost exclusively dispersed by *D. gliroides* seemingly due to a fruit color variation between these habitats (fruits in the temperate forests remain green when ripe; Amico et al. 2011). This highly specialized plant-frugivore system represents a good study model to assess frugivore-mediated selection on the plant's phenotype as it excludes the potential confounding effects of other frugivore species (Fontúrbel and Medel 2017). We collected this information between 2011 and 2012 at the Valdivian Coastal Reserve (39°57'S 73°34'W), which is the largest (~50,000 ha) remnant of Valdivian temperate rainforest in Chile. The dataset contains records from 70 *T. corymbosus* individuals, their relative fitness values (estimated as the product of fruit removal and germination rates), measurements of three phenotypic traits (fruit size, dry seed weight, and sugar content), and their geographic coordinates (in UTM coordinate system, datum WGS84). The sampling area covered ~100 ha and distances between mistletoe individuals ranged from 2 to 9785 m (average distance 3264 m). More details about the study area can be found at Fontúrbel and Medel (2017).

DNA processing and genotyping

To assess inter-individual relatedness among individuals, we amplified microsatellite markers from DNA of leaf samples in the same 70 *T. corymbosus* individuals aforementioned. Eight microsatellite markers specifically developed for this species were analyzed by following the protocols of Fontúrbel et al. (2016). The complete microsatellite library is freely available from Fontúrbel et al. (2017). After getting multilocus genotype data, we

used MICROCHECKER 2.2.3 software (Van Oosterhout et al. 2004) to identify possible genotyping mistakes, along with the presence of null alleles in the microsatellite data. Due to multiple missing data, we had to remove one individual from every subsequent analysis (i.e., we also excluded this individual from the phenotypic trait dataset and conducted the analyses with $N=69$ plants).

Multidimensional scaling of genetic data

In order to include genetic relatedness into models, we first loaded the microsatellite loci database using the packages ‘adeigenet’ (Jombart 2008) and ‘PopGenReport’ (Adamack and Gruber 2014), which was converted to a Euclidean distance matrix and then performed a multidimensional scaling (MDS) to get two-dimensional coordinates (Mardia et al. 1979). MDS computes the best-fitting k -dimensional representation such that the distances between the points obtained are approximately equal to the genetic distances on the matrix. Let us suppose we have coordinates of n points in a p -dimensional Euclidean space, denoted by \mathbf{x}_r where $\mathbf{x}_r = (x_{r1}, \dots, x_{rp})^T$ with $r = 1, \dots, n$; The Euclidean distance between the r^{th} and s^{th} point is given by:

$$d_{rs}^2 = (\mathbf{x}_r - \mathbf{x}_s)^T (\mathbf{x}_r - \mathbf{x}_s)$$

We can define the inner product matrix B as follows

$$[B]_{rs} = \mathbf{x}_r^T \mathbf{x}_s$$

So, if we know the distances d_{rs} we can find matrix B , and from B obtain the unknown coordinates (Cox and Cox 2001). Mardia et al. (1979) proved that dissimilarities work as well as distances. We performed this analysis using the *cmdscale* from the ‘stats’ package in R 3.6.0 (R Development Core Team 2019).

Statistical modeling

We used a Generalized Additive Model (GAM hereafter) approach to examine the influence of spatial and genetic relatedness on phenotypic selection estimates, including spline terms containing the UTM coordinates as well as the coordinates obtained from the genetic distance MDS (as a proxy of inter-individual relatedness) for each sampled plant (Dormann et al. 2007). To compare the estimates for selection coefficients (β) we fitted: (i) a model without spline terms (i.e., a classic Lande and Arnold’s model), (ii) a model with a spline term for spatial coordinates only, (iii) a model with a spline term for genetic relatedness only and, (iv) a model with spline terms for both spatial and genetic relatedness information (error terms are denoted by ϵ).

The models selected were:

- (i) $W_i = \beta_0 + \beta_1 \text{Fruit diameter}_i + \beta_2 \text{Seed weight}_i + \beta_3 \text{Sugar content}_i + \epsilon$
- (ii) $W_i = \beta_0 + \beta_1 \text{Fruit diameter}_i + \beta_2 \text{Seed weight}_i + \beta_3 \text{Sugar content}_i + \text{spline}_1 \text{Space}_i + \epsilon$
- (iii) $W_i = \beta_0 + \beta_1 \text{Fruit diameter}_i + \beta_2 \text{Seed weight}_i + \beta_3 \text{Sugar content}_i + \text{spline}_1 \text{Genetics}_i + \epsilon$
- (iv) $W_i = \beta_0 + \beta_1 \text{Fruit diameter}_i + \beta_2 \text{Seed weight}_i + \beta_3 \text{Sugar content}_i + \text{spline}_1 \text{Space}_i + \text{spline}_2 \text{Genetics}_i + \epsilon$

We fitted GAM models in R using the package ‘mgcv’ (Wood and Scheipl 2014). Model performance was compared using R^2 , deviance, and the Bayesian Information Criterion (BIC). Then, we used univariate spline correlograms to assess the efficiency of each model to deal with residual autocorrelation. We also used partial Mantel correlograms (using 12 distance classes of equal distance intervals) to assess autocorrelation significance at each distance class (Borcard et al. 2004; Dray et al. 2006). For these analyses we used the R packages ‘spdep’ (Bivand 2014) and ‘mpmcorrelogram’ (Matesanz et al. 2011).

After fitting the models, we checked for “model health” by inspecting model overfitting and residual distribution using the function *gam.check* function of the ‘mgcv’ package. Then, we checked if models were affected by multicollinearity using two approaches: (1) examining the variance inflation factor (VIF hereafter) estimated using the *vif.gam* function of the ‘mgcv.helper’ package (available from the GitHub repository: <https://github.com/samclifford/mgcv.helper>) and (2) examining model concurvity for those models including smooth terms, using the *concurvity* function of the ‘mgcv’ package. While VIF is the most common approach to assess multicollinearity, its application on GAM models has two shortcomings: it is based on a thumb rule ($VIF > 4$ may be interpreted as strong multicollinearity) and does not take the smooth terms into account. Concurvity, on the other hand, takes the smooth term(s) into account and provides an index ranging from 0 (no problem) to 1 (total lack of identifiability).

Results

We fitted four GAM models to estimate linear selection gradients using the four model alternatives proposed. The standard model (i.e., classic Lande and Arnold’s model, without spatial or relatedness correction) showed the lowest R^2 and deviance values, explaining only 13.3% of the total variance. Next, when we add the genetic relatedness term, the variance explained raises to 33.0%, and this value rises to 62.1% when we include the spatial term instead. When we include both genetic relatedness and spatial terms in the model, R^2 and deviance values reach their maximum, explaining 65.0% of the total variance (Table 1). Whereas the spatial term had the largest contribution to improving model fit, the genetic relatedness term also contributed to explaining an additional 3% of the variance, not captured by data’s spatial structure. Examining BIC values, the spatial model was the best, followed by the spatial + genetic relatedness model, and the genetic relatedness model had the largest BIC value (larger than the standard model).

Besides the increase in model fit, accounting for space and genetic relatedness also provide valuable biological information. In the example developed here, we detected a significant negative selection on fruit size (meaning that mistletoes with smaller fruits have larger fitness values than those with larger fruits) when the genetic relatedness term was included alone, which was not possible otherwise (Fig. 1). After model comparisons it was evident that selection acting upon seed weight and sugar content was variable. However, the direction and significance of the selection gradients were consistent, albeit showing smaller errors when we included the spatial term. Examining the correlograms, the spatial and spatial + genetic relatedness models dealt better with autocorrelation than the standard model, and the genetic relatedness models (Fig. 2). The partial Mantel correlograms showed significant autocorrelation in the standard model. In contrast, no significant autocorrelation was observed in the spatial and genetic relatedness models, but the spatial + genetic relatedness model was less capable of dealing with autocorrelation (Fig. 3).

Table 1 Results of the generalized additive models fitted to obtain linear selection gradients for fruit size, seed weight, and sugar content traits, using four combinations: no terms, spatial term, genetic relatedness term, and spatial and genetic relatedness terms included as splines

Trait	Standard model	Spatial model	Genetic model	S + G model
Intercept	0.300 (0.034)***	0.298 (0.214)***	0.301 (0.307)**	0.298 (0.023)***
Fruit size	-0.061 (0.044) ^{ns}	0.016 (0.034) ^{ns}	-0.089 (0.043)*	0.035 (0.037) ^{ns}
Seed weight	0.096 (0.043)*	0.082 (0.034)*	0.119 (0.042)**	0.076 (0.035)*
Sugar content	-0.095 (0.035)**	-0.083 (0.027)**	-0.097 (0.033)**	-0.083 (0.026)**
Spatial spline	–	8.456***	–	7.004***
Genetic spline	–	–	2.814*	0.621 ^{ns}
R ² adj.	0.093	0.557	0.245	0.571
Deviance exp.	13.3%	62.1%	33.0%	65.0%
BIC	36.838	8.680	38.885	14.607

Trait results are presented as estimate plus standard error in parentheses. Spline values correspond to an F statistic. To measure model fit, we present adjusted R², deviance explained, and Bayesian Information Criterion (BIC) values

Significance abbreviations: ns, $P \geq 0.1$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

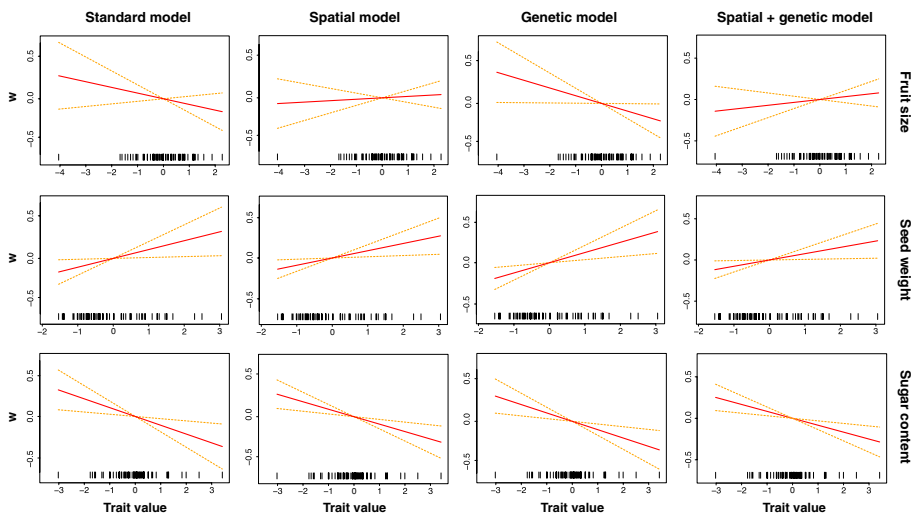


Fig. 1 Comparison of selection gradients for fruit size, seed weight, and sugar content traits estimated by **a** the standard model, **b** the spatial model, **c** the genetic relatedness model, and **d** the spatial + genetic relatedness model. Dashed lines represent one standard error. Values on the x-axis correspond to the standardized trait measurement (mean = 0, SD = 1), and values on the y-axis represent the relative fitness (estimated from individual fruit removal/mean population fruit removal, and individual seed germination rates/mean population germination rate). Plots were generated using the R function *plot.gam*

Regarding “model health”, our models showed no overfitting (Table S1 available online as Supplementary Material) or problems in residual distribution (Figure S1). Also, our models did not show multicollinearity problems in any case (VIF estimations are presented in Table 2). Regarding concavity (Table 3), the models containing smooth terms for spatial and genetic relatedness had no multicollinearity problems (the genetic relatedness

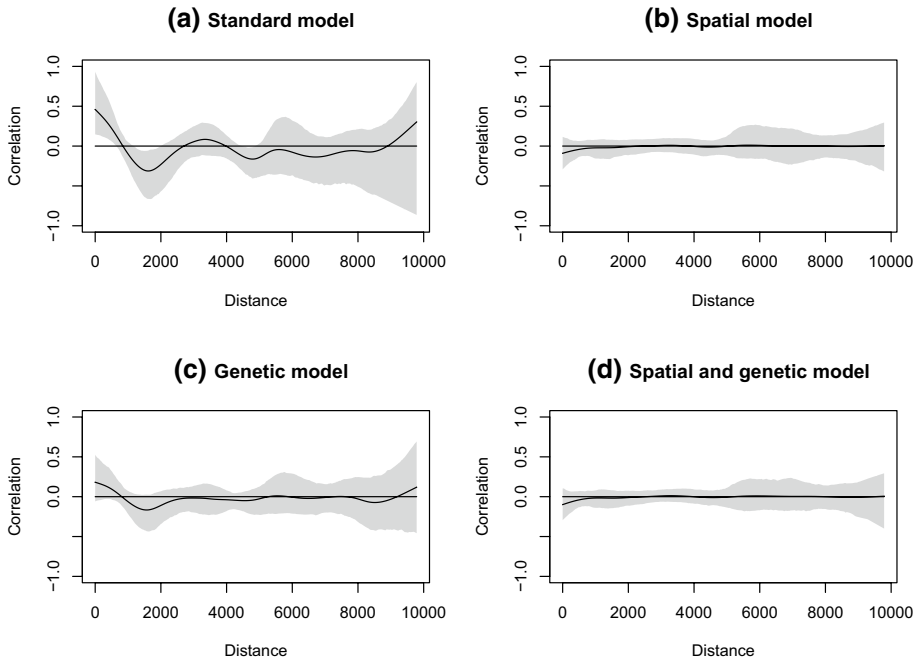


Fig. 2 Univariate spline correlograms for **a** the standard model, **b** the spatial model, **c** the genetic relatedness model, and **d** the spatial and genetic relatedness model. The y-axis represents residual correlation and shaded area represents the 95% confidence interval

model had the lowest concavity values), albeit the spatial and genetic relatedness model may be influenced by multicollinearity to some extent as it is the model with the largest number of parameters.

Discussion

Our results showed that genetic relatedness among individuals and spatial structure increased the fraction of the variance explained by the models used to estimate selection coefficients. Also, models containing spatial and genetic relatedness terms differed in the ability to detect selection on some traits (i.e., fruit size), which are likely to be influenced by data non-independence in wild populations resulting from non-random mating (influencing inter-individual genetic relatedness) and non-random spatial arrangements (influencing mistletoe spatial structure as a result of host distribution and vector behavior). The GAM approach that we used has the advantage of allowing the inclusion of linear (i.e., the measured traits) and non-linear terms (i.e., the spatial and genetic information) into the model, obtaining linear estimates, such as directional selection gradients, in the sense of Lande and Arnold (1983) approach, but taking into account both spatial and genetic relatedness terms to effectively deal with autocorrelation and improve model explanatory power. Spatial and genetic relatedness structures are likely to be non-linear as individuals are usually non-randomly distributed in the space (Dormann et al. 2007); therefore, using non-linear approaches may be more informative. In this sense, mistletoes are a good example of this

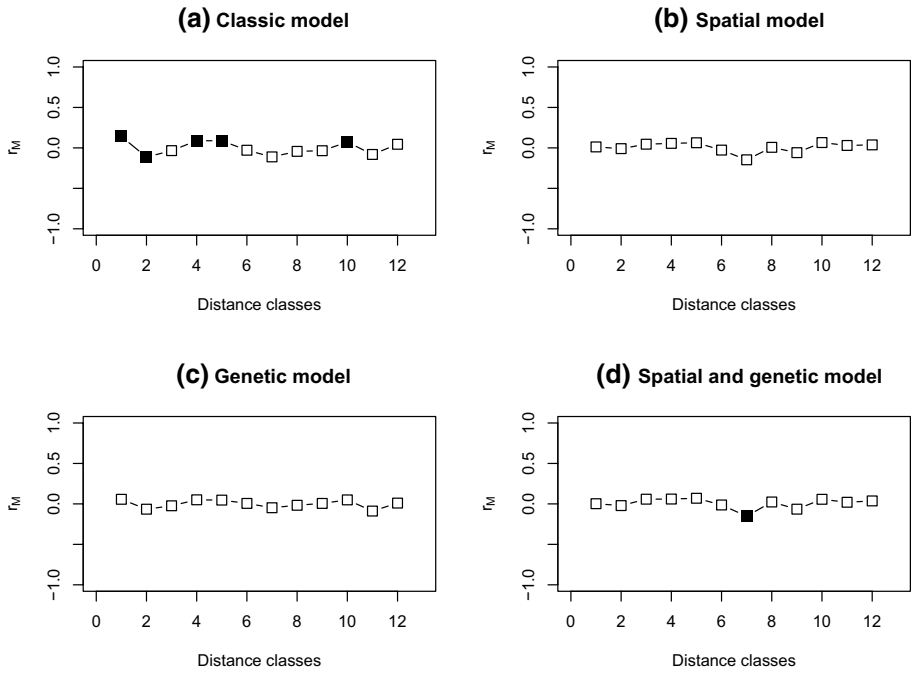


Fig. 3 Partial Mantel correlograms for **a** the standard model, **b** the spatial model, **c** the genetic relatedness model, and **d** the spatial and genetic relatedness model. Open squares denote non-significant autocorrelation, and closed squares denote significant autocorrelation at a given distance class

Table 2 Multicollinearity assessment of the models fitted using the Variance Inflation Factor (VIF) for each linear term. VIF estimates were obtained using the R function *vif.gam*

Model	Parameter	VIF
Classic	Fruit size	1.64
	Seed weight	1.64
	Sugar content	1.06
Spatial	Fruit size	2.00
	Seed weight	2.10
	Sugar content	1.22
Genetic	Fruit size	1.87
	Seed weight	1.84
	Sugar content	1.09
Spatial + genetic	Fruit size	2.43
	Seed weight	2.25
	Sugar content	1.24

situation due to their parasitic lifeform, involving complex interactions with animals (i.e., pollinators and seed dispersers) and their host plants (Watson 2004, 2009), which are usually quite specialized (Watson and Rawsthorne 2013).

The spatial structure accounted for most of the explained variance, and its inclusion significantly improved model fit and reduced estimation errors. However, although genetic

Table 3 Multicollinearity assessment of the models fitted with smooth terms using concavity estimates obtained using the R function *concavity*

Model	Measure	Parameter	Spatial	Genetic
Spatial	Worst	0.0021	0.3207	–
	Observed	0.0021	0.1146	–
	Estimate	0.0021	0.1242	–
Genetic	Worst	0.0018	–	0.1524
	Observed	0.0018	–	0.0409
	Estimate	0.0018	–	0.0368
Spatial + genetic	Worst	0.0027	0.6996	0.6691
	Observed	0.0027	0.5401	0.4803
	Estimate	0.0027	0.3517	0.3602

Concavity values range from 0 (no problem) to 1 (total lack of identifiability) and are calculated for the linear parameters and the smooth terms (spatial and genetic relatedness). Measurement meanings: ‘observed’ is the value obtained from the model, ‘estimate’ is the expected value from the F-norm distribution, and ‘worst’ is largest possible value representing a pessimistic scenario

information accounted for a small fraction of the explained variance in comparison, its inclusion into the equation contributed 20% of explained variance relative to the standard model and an additional 3% of the explained variance when included altogether with the spatial structure, but dealt very well with spatial autocorrelation. Therefore, when we include both terms, models explained most of the data variance with smaller errors. On the other hand, the model containing the genetic information was less affected by multicollinearity while the model including both spatial and genetic information may be potentially influenced by multicollinearity due to the large number of terms and redundant information. Accounting for plant spatial structure deals with most of the spatial autocorrelation, which may include part of the variance associated with genetic distances. Closer individuals are expected to be more related to each other (Fontúrbel et al. 2019). This is particularly evident in mistletoes, as they depend on dispersal biotic vectors, which often deposit seeds on already infected host trees from which they feed (Medel et al. 2004). Therefore, individuals within large mistletoe clumps may be more genetically related to each other than expected by chance.

In this study, the main outcome in biological terms was the detection of directional selection on fruit size after including the genetic relatedness term into the phenotypic selection analysis, which was not revealed otherwise. Detecting selection on fruit size in the wild could be difficult as this trait (i) covaries with seed size (Fontúrbel and Medel 2017; Palacio et al. 2016), and (ii) is likely to be strongly influenced by the parental genotype, as observed in most dispersal-related traits in plants (e.g., fruit size, seed size, sugar content, among others; Wolf and Wade 2009). Thus, detecting selection on dispersal-related traits under field conditions may be a formidable task that can be largely improved in future studies by taking genetic relatedness into account. Consideration of relatedness in this study allowed detection of selection on fruit size (which was non-significant in the standard, spatial, and spatial + genetic relatedness models). According to these results, the genetic model would be useful for detecting selection in dispersal related traits, dealing better with autocorrelation and multicollinearity, but the performance of the models including the spatial information is better. Therefore, spatial and genetic relatedness models may be telling us different parts of the story. In principle,

dispersal-related traits in wild plants result from coevolutionary processes, being a certain amount of phenotypic integration expected (i.e., traits are not expected to evolve independently). However, this does not imply that all traits evolve in the same manner. Rather, the direction and magnitude of the selection forces acting upon them may depend strongly on the ecological context (e.g., Fontúrbel and Medel 2017). In consequence, conflicting selection forces is not an unexpected result, as it is usually revealed in studies of selection for fruit size and seed size (e.g., Alcántara and Rey 2003).

The idea of combining spatial and genetic relatedness information in trait analysis in a macroecology framework has been initially proposed by Kühn et al. (2009), although there was previous work on this subject (Ritland 1990). Further, (Marrot et al. 2015) proposed several alternatives to deal with spatial autocorrelation on selection estimates. The use of a GAM approach in selection analyses is more recent (e.g., Fontúrbel and Medel 2017), but so far, it has not been combined with molecular information. While the genetic relatedness component is usually included in statistical analyses as phylogenetic distances, to the best of our knowledge, this is the first study dealing with genetic distances to account for natural selection in wild plant populations, which have particularities that limit the application of what we know from animals. As genetic effects are relevant for plant adaptation (Galloway 2005), affecting fruit and seed traits in response to a changing environment (Wang et al. 2012; Borgman et al. 2014; Fontúrbel 2020), they seem to be a critical characteristic that together epigenetic effects influence the outcome of frugivore-mediated selection.

Our analyses showed that consideration of the spatial structure and genetic relatedness improve largely the fit and explanatory power of phenotypic selection models, resulting in more realistic selection estimates in wild plant populations. Current computational power and non-linear modeling approaches (e.g., GAM) offer powerful tools to deal with the lack of data independence due to spatial structure and relatedness among sampled individuals. Instead of avoiding the ‘noise’ produced by spatial structure and genetic relatedness, we can build powerful models to deal with them. Natural variation is inherent to many (if not, most) biological systems as a result of ecological and evolutionary processes occurring in the wild. The simple methodological approach presented here, exemplified by the mistletoe *T. corymbosus*, is extensible to any wild plant population from which we can obtain geographic coordinates and DNA samples.

Acknowledgments This research was funded by the FONDECYT project 11160152 (FEF). The comments of two anonymous reviewers improved the manuscript.

Authors contributions FEF and RM conceived the idea and designed methodology. FEF collected field data. CVR analyzed molecular data. PFF developed the statistical framework. FEF and PFF analyzed the data. FEF and RM led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data availability Data from Fontúrbel and Medel (2017) is available from the *figshare* repository (<https://doi.org/10.6084/m9.figshare.4614769>). Molecular data and R scripts related to this manuscript are available from the GitHub repository: <https://github.com/fonturbel/PhenotypicSelectionGS>.

Compliance with ethical standards

Conflicts of interest The authors declare no conflicts of interest.

References

- Adamack AT, Gruber B (2014) PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol* 5:384–387. <https://doi.org/10.1111/2041-210X.12158>
- Alcántara JM, Rey PJ (2003) Conflicting selection pressures on seed size: evolutionary ecology of fruit size in a bird-dispersed tree, *Olea europaea*. *J Evol Biol* 16:1168–1176. <https://doi.org/10.1046/j.1420-9101.2003.00618.x>
- Amico GC, Rodríguez-Cabal MA, Aizen MA (2011) Geographic variation in fruit colour is associated with contrasting seed disperser assemblages in a south-Andean mistletoe. *Ecography* 34:318–326. <https://doi.org/10.1111/j.1600-0587.2010.06459.x>
- Bivand RS (2014) spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-74. <http://CRAN.R-project.org/package=spdep>
- Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85:1826–1832
- Borgman EM, Schoettle AW, Angert AL (2014) Using among-year variation to assess maternal effects in *Pinus aristata* and *Pinus flexilis*. *Botany* 92:805–814. <https://doi.org/10.1139/cjb-2014-0085>
- Cox TF, Cox MAA (2001) Multidimensional scaling, 2nd edn. Chapman and Hall, Boca Raton
- Donohue K (1999) Seed dispersal as a maternally influenced character: mechanistic basis of maternal effects and selection on maternal characters in an annual plant. *Am Nat* 154:674–689. <https://doi.org/10.1086/303273>
- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Model* 196:483–493
- Endler JA (1986) *Natural selection in the wild*. Princeton University Press, Princeton
- Fontúrbel FE (2020) Mistletoes in a changing world: a premonition of a non-analog future? *Botany* 98:479–488. <https://doi.org/10.1139/cjb-2019-0195>
- Fontúrbel FE, Medel R (2017) Frugivore-mediated selection in a habitat transformation scenario. *Sci Rep-UK* 7:45371. <https://doi.org/10.1038/srep45371>
- Fontúrbel FE, Murúa MM, Vega-Retter C (2016) Development of ten microsatellite markers from the keystone mistletoe *Tristerix corymbosus* (Loranthaceae) using 454 next generation sequencing and their applicability to population genetic structure studies. *Mol Biol Rep* 43:339–343. <https://doi.org/10.1007/s11033-016-3970-6>
- Fontúrbel FE, Murúa MM, Vega-Retter C (2017) *Tristerix corymbosus* microsatellite library. Figshare digital repository: <https://doi.org/10.6084/m9.figshare.4729954>
- Fontúrbel FE, Bruford MW, Salazar DA, Cortés-Miranda J, Vega-Retter C (2019) The hidden costs of living in a transformed habitat: ecological and evolutionary consequences in a tripartite mutualistic system with a keystone mistletoe. *Sci Total Environ* 651:2740–2748. <https://doi.org/10.1016/j.scitotenv.20148.10.125>
- Fortin MJ, Dale M (2005) *Spatial analysis: a guide for ecologists*. Cambridge University Press, Cambridge
- Galloway LF (2005) Maternal effects provide phenotypic adaptation to local environmental conditions. *New Phytol* 166:93–99. <https://doi.org/10.1111/j.1469-8137.2004.01314.x>
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- Jordano P (1995) Frugivore-mediated selection on fruit and seed size: birds and St. Lucie's cherry. *Prunus Mahaleb* *Ecol* 76:2627–2639. <https://doi.org/10.2307/2265833>
- Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol Ecol* 26:1101–1118. <https://doi.org/10.1007/s10682-012-9563-5>
- Kühn I, Nobis MP, Durka W (2009) Combining spatial and phylogenetic eigenvector filtering in trait analysis. *Global Ecol Biogeogr* 18:745–758. <https://doi.org/10.1111/j.1466-8238.2009.00481.x>
- Lande R, Arnold S (1983) The measurement of selection on correlated characters. *Evolution* 37:1210–1226. <https://doi.org/10.2307/2408842>
- Legendre P, Dale MRT, Fortin MJ, Gurevitch J, Hohn M, Myers D (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25:601–615
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic Press, Boca Raton
- Marrot P, Garant D, Charmantier A (2015) Spatial autocorrelation in fitness affects the estimation of natural selection in the wild. *Methods Ecol Evol* 6:1474–1483. <https://doi.org/10.1111/2041-210x.12448>

- Matesanz S, Gimeno TE, de la Cruz M, Escudero A, Valladares F (2011) Competition may explain the fine-scale spatial patterns and genetic structure of two co-occurring plant congeners. *J Ecol* 99:838–848
- Medel R, Vergara E, Silva A, Arroyo MK (2004) Effects of vector behavior and host resistance on mistletoe aggregation. *Ecology* 85:120–126
- Palacio FX, Girini JM, Ordano M (2016) Linking the hierarchical decision-making process of fruit choice and the phenotypic selection strength on fruit traits by birds. *J Plant Ecol*. <https://doi.org/10.1093/jpe/rtw063>
- Pemberton JM (2010) Evolution of quantitative traits in the wild: mind the ecology. *Philos Trans R Soc B* 365:2431–2438. <https://doi.org/10.1098/rstb.2010.0108>
- R Development Core Team (2019) R: a language and environment for statistical computing, reference index version 3.6.1. Foundation for Statistical Computing, Vienna, Austria
- Rausher MD (1992) The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46:616–626. <https://doi.org/10.2307/2409632>
- Rawsthorne J, Watson DM, Roshier DA (2011) Implications of movement patterns of a dietary generalist for mistletoe seed dispersal. *Austral Ecol* 36:650–655
- Ritland K (1990) Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* 44:1230–1241. <https://doi.org/10.2307/2409284>
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
- Walsh B, Lynch M (2018) Evolution and selection of quantitative traits. Oxford University Press, New York
- Wang L, Baskin JM, Baskin CC, Cornelissen JHC, Dong M, Huang ZY (2012) Seed dimorphism, nutrients and salinity differentially affect seed traits of the desert halophyte *Suaeda aralocaspica* via multiple maternal effects. *BMC Plant Biol*. <https://doi.org/10.1186/1471-2229-12-170>
- Waser NM, Mitchell RJ (1990) Nectar standing crops in *Delphinium nelsonii* flowers: spatial autocorrelation among plants? *Ecology* 71:116–123. <https://doi.org/10.2307/1940252>
- Watson DM (2004) Mistletoe: a unique constituent of canopies worldwide. In: Lowman MD, Rinker HB (eds) Forest canopies, 2nd edn. Academic Press, San Diego, pp 212–223
- Watson DM (2009) Parasitic plants as facilitators: more Dryad than Dracula? *J Ecol* 97:1151–1159. <https://doi.org/10.1111/j.1365-2745.2009.01576.x>
- Watson DM, Rawsthorne J (2013) Mistletoe specialist frugivores: latterday ‘Johnny Appleseeds’ or self-serving market gardeners? *Oecologia* 172:925–932. <https://doi.org/10.1007/s00442-013-2693-9>
- Weber A, Kolb A (2013) Population size, pollination and phenotypic trait selection in *Phyteuma spicatum*. *Acta Oecol* 47:46–51. <https://doi.org/10.1016/j.actao.2012.12.004>
- Wolf JB, Wade MJ (2009) What are maternal effects (and what are they not)? *Philos Trans R Soc B* 364:1107–1115. <https://doi.org/10.1098/rstb.2008.0238>
- Wood S, Scheipl F (2014) gamm4: Generalized additive mixed models using mgcv and lme4. R package version 0.2–3. <http://CRAN.R-project.org/package=gamm4>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.