

<https://helda.helsinki.fi>

Landscape permeability and individual variation in a
dispersal-linked gene jointly determine genetic structure in the
Glanville fritillary butterfly

DiLeo, Michelle F.

2018-12

DiLeo , M F , Husby , A & Saastamoinen , M 2018 , ' Landscape permeability and individual
variation in a dispersal-linked gene jointly determine genetic structure in the Glanville fritillary
butterfly ' , Evolution Letters , vol. 2 , no. 6 , pp. 544-556 . <https://doi.org/10.1002/evl3.90>

<http://hdl.handle.net/10138/294745>

<https://doi.org/10.1002/evl3.90>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Landscape permeability and individual variation in a dispersal-linked gene jointly determine genetic structure in the Glanville fritillary butterfly

Michelle F. DiLeo,^{1,2}  Arild Husby,^{1,3,4}  and Marjo Saastamoinen^{1,5} 

¹Organismal and Evolutionary Biology Research Programme, University of Helsinki, PO Box 65, 00014, Finland

²E-mail: michelle.dileo@helsinki.fi

³Department of Evolutionary Biology, EBC, Uppsala University, Norbyvägen 18D 75236 Uppsala, Sweden

⁴E-mail: arild.husby@ebc.uu.se

⁵E-mail: marjo.saastamoinen@helsinki.fi

Received May 4, 2018

Accepted October 29, 2018

There is now clear evidence that species across a broad range of taxa harbor extensive heritable variation in dispersal. While studies suggest that this variation can facilitate demographic outcomes such as range expansion and invasions, few have considered the consequences of intraspecific variation in dispersal for the maintenance and distribution of genetic variation across fragmented landscapes. Here, we examine how landscape characteristics and individual variation in dispersal combine to predict genetic structure using genomic and spatial data from the Glanville fritillary butterfly. We used linear and latent factor mixed models to identify the landscape features that best predict spatial sorting of alleles in the dispersal-related gene *phosphoglucose isomerase* (*Pgi*). We next used structural equation modeling to test if variation in *Pgi* mediated gene flow as measured by F_{st} at putatively neutral loci. In a year when the population was recovering following a large decline, individuals with a genotype associated with greater dispersal ability were found at significantly higher frequencies in populations isolated by water and forest, and these populations showed lower levels of genetic differentiation at neutral loci. These relationships disappeared in the next year when metapopulation density was high, suggesting that the effects of individual variation are context dependent. Together our results highlight that (1) more complex aspects of landscape structure beyond just the configuration of habitat can be important for maintaining spatial variation in dispersal traits and (2) that individual variation in dispersal plays a key role in maintaining genetic variation across fragmented landscapes.

KEY WORDS: Dispersal, eco-evolutionary dynamics, *fst*, gene flow, genetic structure, intraspecific variation, landscape genetics, landscape matrix, metapopulation.

Impact Summary

Understanding how fragmentation affects dispersal and gene flow across human-modified landscapes has long been a goal in evolutionary biology. Growing evidence suggests that individuals of the same species can vary considerably in their dispersal and movement

traits. While the effects of this individual dispersal variation on range expansions and invasions have been well characterized, knowledge of how it might mediate genetic responses to landscape fragmentation are almost entirely lacking. Here, we demonstrate that individual variation in dispersal is key to the maintenance of

genetic variation during recovery following a large decline in a butterfly metapopulation. We further show that spatial variation in dispersal is not maintained by the configuration of habitat patches alone, but by a more complex genotype–environment interaction involving the landscape matrix (i.e., landscape features found between habitat patches). This challenges the simplified landscape representations typically used in studies of dispersal evolution that ignore heterogeneity in the landscape matrix. More broadly, our results highlight the interplay of adaptive and neutral processes across fragmented landscapes, suggesting that an understanding of species vulnerability to landscape fragmentation requires consideration of both.

Introduction

Dispersal is key to the maintenance of genetic variation and adaptive potential in fragmented landscapes. Differences in species ability to maintain genetic diversity in fragmented landscapes can, in part, be explained by interspecific differences in dispersal capacity (Steele et al. 2009; Peterman et al. 2015). However, there is now clear evidence that many species across a broad range of taxa harbor extensive heritable variation in dispersal (Saastamoinen et al. 2018) and that this variation can facilitate demographic outcomes such as range expansion (e.g., Duckworth and Badyaev 2007; Ochocki and Miller 2017) and invasions (e.g., Phillips et al. 2006; Elliott and Cornell 2012; Cote et al. 2017). In comparison, few studies have considered the consequences of intraspecific variation in dispersal for genetic outcomes such as the maintenance and distribution of genetic variation across fragmented landscapes (Cheptou et al. 2017). This is an important gap given that the genetic makeup of populations can drive the trajectories of both ecological and evolutionary processes (Rius and Darling 2014; Szucs et al. 2017; Wagner et al. 2017).

This gap comes partly from a lack of integration of intraspecific variation into fields like landscape and spatial genetics (Edelaar and Bolnick 2012; Pflueger and Balkenhol 2014). Studies in these fields emphasize that gene flow across fragmented landscapes is strongly constrained by population configuration, habitat quality, and matrix heterogeneity (i.e., the landscape features intervening populations), but typically assume that dispersers respond to the landscape in the same way (Manel et al. 2003; Holderegger and Wagner 2008). However, fragmentation itself can exert strong selective pressures on dispersers, in some cases leading to the maintenance of multiple dispersal strategies across the same landscape (Cheptou et al. 2017; Cote et al. 2017;

Legrand et al. 2017). This variation might change expectations of spatial genetic structure and mask landscape genetic relationships, yet empirical tests of this are lacking (but see McDevitt et al. 2013). For example, using simulations, Palmer et al. (2014) showed that distance-based connectivity metrics underestimated the number of migrants arriving into isolated populations when individuals were allowed to vary in their dispersal ability, with the effect most severe when dispersal was rare. This spatial sorting of good dispersers into more isolated populations is expected to also impact the distribution of genetic variation, for example, by increasing genetic neighborhoods beyond expectations drawn from mean dispersal distances (DiLeo et al. 2014). Because dispersal traits often co-evolve with other aspects of morphology, physiology, and behavior (Clobert et al. 2009; Cote et al. 2017), individuals might also interact with, and respond to, the landscape matrix in different ways (Merckx and Van Dyck 2007; Delgado et al. 2010). This could mean that the effects of landscape on gene flow might be missed if intraspecific variation is ignored.

Intraspecific variation thus has the potential to play an important, but so far unexplored role in structuring species genetic response to landscape fragmentation. The maintenance of several dispersal strategies across a single landscape might allow widespread gene flow to be maintained under a broad set of ecological conditions. We test this in a model species, the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands, Finland. Importantly, individuals vary in their dispersal ability: butterflies heterozygous or homozygous for the “c” allele in a SNP associated with the gene *phosphoglucose isomerase* (*Pgi*) have higher flight metabolic rate, which translates to substantially increased dispersal propensity and dispersal distance in the field, especially at cooler temperatures (reviewed in Niitepõld and Saastamoinen 2017). Evidence for this comes from laboratory experiments linking flight metabolic rate to *Pgi* genotype (Niitepõld 2010), and from a study linking dispersal distances of butterflies tracked in the field to *Pgi* genotype (Niitepõld et al. 2009). The butterfly persists in a highly dynamic metapopulation with frequent colonizations and extinctions, and we focus on a 2-year period representing extremes of fluctuations experience by the butterfly: the year 2011 when populations had extremely low connectivity because of a large decline in the previous year, and 2012 when populations had high connectivity following the recovery of populations in the year prior (Ojanen et al. 2013). Recent work suggests that *Pgi* plays a central role in metapopulation persistence by maintaining high recolonization rates despite drastic population fluctuations (Hanski et al. 2017). However, we do not yet know to what extent *Pgi* also contributes to the maintenance of neutral genetic variation, which is important given that genetic diversity can affect persistence independently of demographic processes—e.g., populations founded by genetically diverse individuals have been observed to have higher chances of successful establishment,

population growth rates, and persistence, independent of founder population size (Ahloth et al. 2003; Szucs et al. 2017). Specifically, we ask: (1) what landscape factors drive spatial sorting of genotypes that vary in their dispersal ability? And (2) does this dispersal variation mediate the genetic response to landscape structure and population fluctuations? While previous work has found that more isolated patches have higher frequencies of good dispersers (Haag et al. 2005, Hanski and Saccheri 2006; Zheng et al. 2009), we further predict that dispersers will respond differentially to heterogeneity in the landscape matrix. We also predict that the good dispersers will facilitate genetic admixture; population with higher frequencies of the *Pgi-c* allele should be less genetically differentiated than populations with low frequencies. Finally, we predict that the effects of *Pgi* will be context dependent. Modeling predicts that the dispersive genotype will have the largest advantage when there are many open patches to colonize (Zheng et al. 2009), and thus we expect the association between landscape structure, *Pgi*, and genetic structure to be highest in 2011.

Methods

STUDY SPECIES AND SAMPLING

In Finland, the Glanville fritillary butterfly is only found in the Åland Islands where it persists in a metapopulation encompassing over 4000 meadows (hereinafter ‘patches’) that contain one or both of its host plants, *Plantago lanceolata* and *Veronica spicata*. In late summer, females lay clutches of 150–200 eggs, which develop into larvae that live gregariously until the last larval instar in the following spring (Boggs and Nieminen 2004). Before winter diapause, larvae spin nests at the base of host plants, and every fall since 1993 the number of nests have been counted in all patches in Åland allowing the quantification of long-term population dynamics (see Ojanen et al. 2013 for survey methods). In the majority of cases nests are composed of single full-sib families, although multiple families on the same host plant can also occur (Fountain et al. 2018). In any given year, only about 20% of patches on average are occupied, with frequent local recolonizations and extinctions, and large variation in the number of total larval nests (Fig. 1). Our study focused on the period 2010–2012, which is characterized by a large population decline in 2010 due to poor weather conditions, followed by a population recovery. The major recovery occurred in 2011 where a record number of re-colonizations were documented (Fig. 1B). In 2012, there were fewer colonizations but a large increase in population density, with a record number of larval nests (Fig 1A).

In 2011 and 2012, three larvae per nest were collected from patches across Åland and were genotyped at 40 neutral markers and five putatively function markers from candidate genes related to flight or dispersal traits including *Pgi* (Mc1:1687:14486), Mc1:1687:14486, Mc1:1873:36910,

Mc1:1124:71239, Mc1:752:33517 (Orsini et al. 2009; Kvist et al. 2013; Ahola et al. 2014; Kvist et al. 2015; Wong et al. 2016; Duplouy et al. 2017; Supporting Information Appendix A and B). Full details of genotyping and quality control are given in Supporting Information Appendix A. After excluding any SNP or individual with a call rate <95%, 34 neutral markers and all of the functional markers remained, and total sample sizes were 3365 larvae representing 1504 nests in 250 patches in 2011, and 8229 larvae representing 2999 nests in 322 patches in 2012. We only had a few samples from the extreme south of our study region in Lemland. As these samples were clear outliers with low connectivity and frequencies of *Pgi-c* (Figs. S1 and S2), we excluded them from downstream analysis. We further excluded 98 patches in 2011 and 129 in 2012 with only a single larval nest where the effects of genetic drift are expected to be especially strong, giving a sample size of 152 patches (74 old, 78 new) in 2011 and 193 patches (138 old, 55 new) in 2012.

DEVELOPMENT OF LANDSCAPE CONNECTIVITY HYPOTHESES

To test the effects of landscape on neutral and functional genetic variation, we developed a set of landscape resistance surfaces reflecting the permeability of the intervening landscape matrix to dispersing butterflies. We assigned each landscape feature a value representing its resistance to a dispersing *M. cinxia* individual—either a value of 1 (not resistant to movement) or 10 (restricts movement), generating a total of 20 surfaces with different combinations of resistant and nonresistant features (Table S1). Patches were always given a value of 1 as patches represent suitable habitat for the species, and continuous urban areas were always given a value of 10. For each of these surfaces, we calculated pairwise resistance distances between *M. cinxia* patches using the program CIRCUITSCAPE (McRae 2006). This program uses circuit theory to calculate effective distances among patches by taking into account the relative permeability of the intervening landscape. Further details of the landscape classifications are given in Supporting Information Appendix A.

For each of the landscape hypotheses, we calculated an index of connectivity for each occupied patch using the incidence function model (Hanski 1994):

$$S_i = \sum_{i \neq j} \exp(-\alpha d_{ij}) N_j$$

where d_{ij} is either the geographic distance (in km) or resistance distance (from CIRCUITSCAPE) between focal patch i occupied in year t and source patch j occupied in the previous year ($t - 1$), and N_j is the number of *M. cinxia* nests found in source patch j in year $t - 1$. The constant α scales the dispersal kernel and should be equal to 1/mean dispersal distance of the species, which has been estimated to be 1 km (Fountain et al. 2018). Because resistance

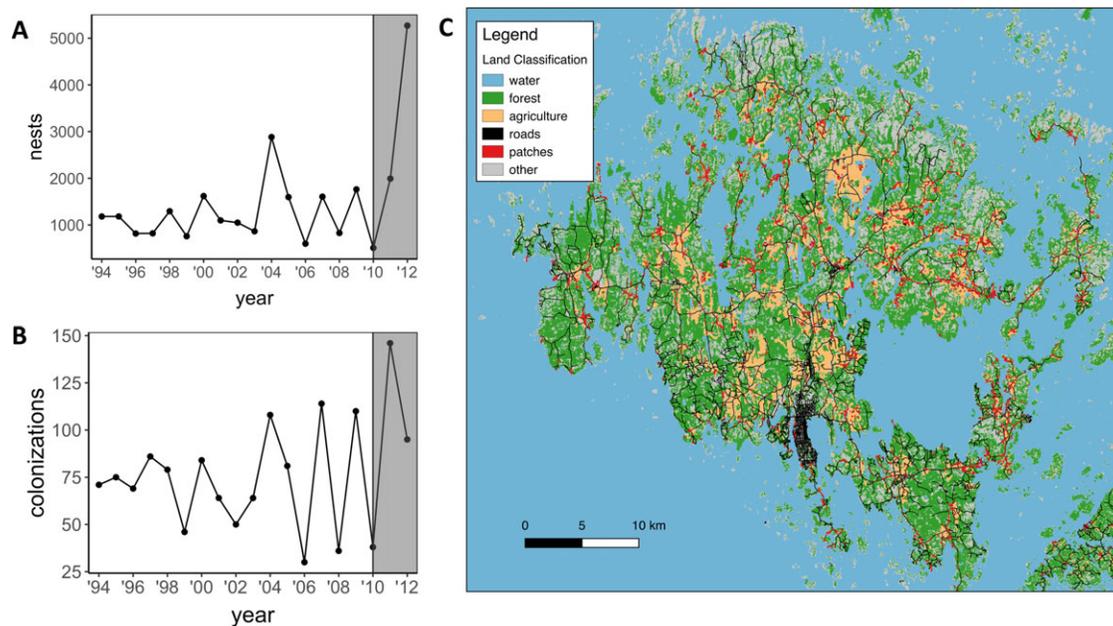


Figure 1. Yearly variation in the number of nests (A) and colonizations (B) in the Åland Islands (C). The period of study from 2010 to 2012 is highlighted in (A) and (B), and (C) shows the major land cover types across the study region. Data for population trends come from annual surveys described in detail in Ojanen et al. 2013.

distances are unit-less, mean dispersal distance for each landscape resistance hypothesis was chosen as the resistance value that roughly translated to 1 km in Euclidean distance units, which was then used to calculate α (Table S1). We did this by taking the predicted value of the resistance distance at a Euclidean distance value of 1 km using a simple linear model, after first linearizing relationships with log transformations. Many of the connectivity variables calculated from alternative landscape resistance surfaces were highly correlated, because some landscape features were only present in small proportions in our study region and thus did not have large effects on the calculations of resistance distances. For our main analyses we settled on five uncorrelated variables of landscape connectivity: $Si_{metapop}$, which is scaled by Euclidean distance among patches but assumes that the intervening landscape does not restrict dispersal, Si_{water} , Si_{forest} , and $Si_{agriculture}$, which assume that water, forest, and agriculture restrict dispersal, respectively, and Si_{roads} , which assumes that roads facilitate dispersal. The five connectivity variables had Pearson correlations below 0.6 and variance inflation factors in linear models under 2 (Table S2).

ASSOCIATIONS BETWEEN LANDSCAPE CONNECTIVITY AND *Pgi*

Our first objective was to test which landscape connectivity hypotheses best predicted the distribution of the *Pgi-c* allele. Based on previous work, we expected an interactive effect of patch age; good dispersers are expected to be at the highest

frequency in newly colonized (hereinafter “new”), isolated populations and at the lowest in old, isolated populations because dispersive individuals also have high emigration rates (Zheng et al. 2009). We conducted model selection on mixed effect models including all connectivity variables and their interaction with patch age (i.e., new or old population) as fixed effects, and genetic cluster membership as a random effect. Frequency of the *Pgi-c* allele in each patch was used as the response variable. Genetic cluster membership represents groups of individuals or populations that share a common demographic history (i.e., share common historical population dynamics and/or ancestry), and was determined for each year separately using BAPS6 (see Supporting Information Appendix A; Fig. S1; Corander et al. 2008; Cheng et al. 2013). Mixed models were run for each year separately and we applied log transformations to $Si_{metapop}$ and to Si_{roads} , and exponential transformations to Si_{water} and $Si_{agriculture}$ to linearize relationships. Each variable was scaled and mean centered. Models were implemented in the `lme4` library (Bates et al. 2015) and validated graphically by plotting residuals against fitted values and normality assumptions were checked with QQ plots. For model selection, we retained a candidate set of models with high support for further analysis and interpretation. A model was included in the candidate set based on two conditions: (1) the model was within $\Delta AICc < 2$ of the top model and (2) the model was not simply an embellishment of a higher ranked model (i.e., did not contain uninformative parameters; Arnold 2010). All predictors appearing in the resulting candidate model set were

considered as potentially important in their effects on *Pgi* and were subject to downstream analysis (see below). We used this approach since alternatives such as model averaging and summing akaike weights can lead to flawed interpretation of effects when variables are even weakly collinear (Galipaud et al. 2014; Cade 2015).

The residuals of the top model for both 2011 and 2012 showed evidence of spatial autocorrelation. To test that this did not bias our results, we compared models with and without spatial random effects implemented in *r-inla* (Lindgren & Rue 2015; Supporting Information Appendix A). In one case, we found the sign of a weak main effect changed in the spatial compared to the nonspatial model (from 0.006 to -0.003 ; Table S3). However, the major effect of this variable manifested as an interaction with a strong negative association in new populations, and the direction and strength of this interaction did not change. We found very little difference in all other estimates between the spatial and nonspatial models, and thus did not pursue spatial models further (Table S3).

As a second step, we used gene–environment association analysis (Rellstab et al. 2015) using latent factor mixed models (lfmms; Fricot et al. 2013) implemented in the *LEA* library (Fricot and François 2015) to confirm associations identified as significant in the linear mixed effect models. While the linear mixed model approach controlled for potential effects of neutral genetic structure by including genetic cluster membership (calculated from the 34 neutral loci) as a random effect, the latent factor mixed models explicitly considers all loci (*Pgi* and neutral loci) simultaneously. This applies a stronger control of background genetic structure, particularly when structure is complex and hierarchical (de Villemereuil et al. 2014), and also allowed us to confirm that we only see associations between connectivity and allele frequencies for *Pgi* and not for other loci. A drawback is that it cannot incorporate additive or interactive effects, and thus we used it only to confirm results in the partitions of the data that were found to be significant in linear mixed effect models (e.g., new populations). In addition to testing the individual S_i variables identified as important in the mixed models, we explored potential additive effects by testing for association between loci and S_i variables that incorporated the resistance of two or more of the landscape features. For example, the additive effects of $S_{i_{\text{water}}}$ and $S_{i_{\text{forest}}}$ were included in the model as the predictor $S_{i_{\text{water}+\text{forest}}}$ (Table S1). Supporting Information Appendix A describes how the latent factors were specified starting from knowledge of the number of k genetic clusters identified from *BAPS6*. Loci included in each analysis were *Pgi*, 34 neutral markers, and four additional markers previously identified as being outliers in fragmented vs. continuous landscapes or from flight induced gene expression differences (Kvist et al. 2015; Fountain et al. 2016).

DIRECT AND INDIRECT EFFECTS OF LANDSCAPE STRUCTURE ON GENETIC STRUCTURE

Our second objective was to test if intraspecific variation at the *Pgi* locus mediates the effect of landscape on the distribution of neutral genetic variation. We hypothesized that landscape structure will influence neutral genetic structure either directly (i.e., by limiting dispersal of all individuals in the same way), or indirectly through its effect on the spatial distribution of individuals with contrasting *Pgi* genotypes.

Direct effects were tested using linear mixed effect models with neutral genetic differentiation as a response variable. A patch-specific measure of genetic differentiation was calculated by taking the average pairwise Weir and Cockerham unbiased F_{st} (Weir and Cockerham 1984) for each patch following Pflueger and Balkenhol (2014). Prior to calculation of F_{st} , the data were subset to include only a single individual per nest, and a maximum of 50 nests per patch as there was large variation in sample sizes. Although Weir and Cockerham's unbiased F_{st} accounts for unequal population sizes, calculations are biased at very low sample sizes (Willing et al. 2012), and so we limited analyses to patches with more than two nests, excluding 43 patches in 2011 and 37 patches in 2012. Models were run for each year separately with the five connectivity metrics and population age as fixed predictors, and genetic cluster membership as a random factor. We did not find evidence for an interactive effect of population age and so only included it as a main effect. Predictor variables were transformed, centered, and scaled, and model selection was implemented as described above (see 'Association between landscape connectivity and *Pgi*').

We next tested for indirect effects of landscape structure on genetic differentiation in the partitions of data where we found strong association between connectivity and *Pgi* using structural equation modeling (SEM) implemented in the *lavaan* library (Rosseel 2012). Originally developed by Sewall Wright (1934), SEM allows the evaluation of complex a priori-defined relationships among variables that potentially involve direct and indirect effects. It is thus an ideal framework for testing the relative effects of connectivity vs. *Pgi* on genetic structure, while controlling for directed or residual relationships between them. We modeled two pathways: (1) connectivity directly affects genetic differentiation, and (2) connectivity indirectly affects genetic differentiation through its effects on the frequency of the *Pgi-c* allele. Connectivity was included in the model as a latent variable – i.e., a construct that is not easily measured directly but can be indicated by a number of observed, and potentially correlated, variables that have some level of measurement error (Grace et al. 2012). We included connectivity as a latent variable, with the five connectivity hypotheses $S_{i_{\text{metapop}}}$, $S_{i_{\text{water}}}$, $S_{i_{\text{forest}}}$, $S_{i_{\text{agriculture}}}$, and $S_{i_{\text{roads}}}$ as indicators. We estimated model parameters using maximum likelihood and assessed model fit with chi-square tests, where a

P-value > 0.05 indicates the model-implied covariance fits the observed covariance well (Grace et al. 2012). The relative effects of paths were evaluated from standardized path coefficients and individual significance of paths.

Results

ASSOCIATIONS BETWEEN LANDSCAPE CONNECTIVITY AND *Pgi*

For data from 2011, selection on mixed effect models identified 10 models with $\Delta AICc < 2$ (Table S4). All of these models included Si_{forest} and Si_{water} , and seven of the models were embellishments of higher ranked models—i.e., they were the same as a higher ranked model but included uninformative parameters with little effect on model fit (Arnold 2010). We thus retained three models for further analysis and interpretation: the top model including an interaction of Si_{forest} and age and main effects of Si_{water} and $Si_{agriculture}$, a model including an interaction of Si_{forest} and a main effect of Si_{water} , and a third model including an interaction of Si_{water} and main effects of Si_{forest} and $Si_{agriculture}$ (Table 1). All variables showed negative associations with *Pgi-c*, with Si_{forest} and Si_{water} having the strongest effects in new populations (Table 1). The effect of water and forest in new populations remained significant in latent factor mixed models as evidenced by a significant association of $Si_{water+forest}$ with *Pgi-c* but no other loci (Fig. S3). In contrast Si_{water} showed a significant association with a putatively neutral locus but not *Pgi-c*, Si_{forest} showed a significant association with both *Pgi-c* and a different putatively neutral locus, and $Si_{water+forest+agriculture}$ and $Si_{agriculture}$ showed no significant associations with *Pgi-c* or any other locus (Supporting Information Appendix A; Fig. S3).

The negative association between *Pgi* and connectivity in new populations switched in linear mixed models in 2012, and the effect of Si_{forest} disappeared (Fig. 2; Table 1). Model selection identified six models with $\Delta AICc < 2$ (Table S5). All of these models included an interaction of age and Si_{roads} , and either a main or interactive effect of Si_{water} . Four of the models were embellishments of higher ranked models. We thus retained two models for

further analysis and interpretation: the top model containing interactions between of both Si_{roads} and Si_{water} with patch age, and a lower ranked model with an interaction between Si_{roads} and age and a main effect of Si_{water} (Table 1). Si_{roads} showed a strong positive association with *Pgi-c* in new populations, whereas Si_{water} showed a negative association with *Pgi-c* in old populations but only a weak or nonexistent relationship in new populations in the top model (Table 1; Fig. S4). Neither Si_{roads} nor Si_{water} were found to have significant associations with *Pgi-c* or any other loci in latent factor mixed models (Supporting Information Appendix A; Fig. S3). Details on the calibration of the latent factor mixed models, including reports of genomic inflation factors can be found in Supporting Information Appendix A and Table S6.

DIRECT AND INDIRECT EFFECTS OF LANDSCAPE STRUCTURE ON GENETIC STRUCTURE

For 2011, selection on mixed effect models testing for direct effect of connectivity and age on F_{st} identified a model containing only the random intercept as the most likely model (Table S7). We also tested the model after removing a single patch that had a very high F_{st} value. This led to the selection of a model including Si_{water} as the best, however it explained only 3% of the variation in F_{st} , and the random intercept-alone model remained in the candidate set with $\Delta AICc < 2$ (Table S8). We thus conclude that there is no, or a very weak, direct effect of connectivity on genetic differentiation in 2011. For 2012, eight models had a $\Delta AICc < 2$, however five of them were embellishments of higher ranked models (Table S9). We thus retained three models: the top model including equal effects of $Si_{metapop}$ and Si_{water} , a model with just $Si_{metapop}$, and a model with equal effects of Si_{water} and Si_{roads} (Table 2).

Pgi showed a significant association with connectivity only in new populations in 2011, and thus we only tested for indirect effects of connectivity on F_{st} (mediated by *Pgi*) in this partition. The full structural equation model including all five landscape connectivity indicator variables showed poor fit as indicated by a significant deviation of the observed covariance from the

Table 1. Standardized regression coefficients showing the effects of connectivity and population age on the population frequency of the *Pgi-c* allele for the year 2011 and 2012. Candidate models with $\Delta AICc < 2$ and excluding uninformative parameters are shown (see Methods, and Table S4 and S5 for results of model selection). Variances explained by fixed effects (R^2_m) and jointly by fixed and random effects (R^2_c) are shown.

Standardized regression coefficients													
Year	(intercept)	Age	Si_{water}	Si_{forest}	$Si_{agriculture}$	Si_{roads}	Age: Si_{water}	Age: Si_{forest}	Age: Si_{roads}	R^2_m	R^2_c	AIC _c	ΔAIC_c
2011	0.23	-0.01	-0.06	-0.05	-0.04				-0.13	0.15	0.18	-148.5	0
	0.23	-0.01	-0.06	-0.04					-0.12	0.13	0.16	-147.7	0.82
	0.24	-0.01	-0.06	-0.06	-0.05		-0.12			0.14	0.17	-146.5	1.97
2012	0.24	0.006	-0.05			0.02	0.08		0.11	0.10	0.10	-216.4	0
	0.24	0.003	-0.05			0.02			0.13	0.08	0.09	-215.2	1.17

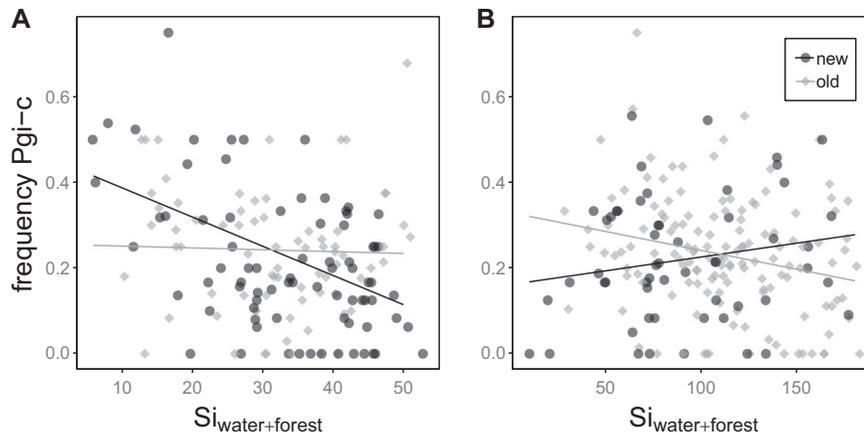


Figure 2. Scatterplots showing the relationship between the frequency of the *Pgi-c* allele and a patch connectivity metric incorporating forest and water as barriers to dispersal for the year 2011 when the metapopulation was recovering from a population decline (A) and 2012 when the metapopulation was at high density (B). Points represent local populations and old and new populations are shown in contrasting colours. Lines show predicted values of fixed effects from linear mixed models including $S_{i_{\text{water+forest}}}$ as a fixed predictor and genetic cluster as a random effect. See Fig. S2 for scatterplots showing variation across genetic clusters.

Table 2. Standardized regression coefficients showing the direct effects of connectivity on genetic differentiation (F_{st}) for the year 2011 and 2012. Candidate models with $\Delta AIC_c < 2$ and excluding uninformative parameters are shown (see Methods, and Table S7 and S9 for results of model selection). Variances explained by fixed effects (R^2_m) and jointly by fixed and random effects (R^2_c) are shown.

Standardized regression coefficients								
Year	(intercept)	$S_{i_{\text{metapop}}}$	$S_{i_{\text{water}}}$	$S_{i_{\text{roads}}}$	R^2_m	R^2_c	AIC_c	ΔAIC_c
2011	0.08					0.02	-464.5	0
2012	0.07	-0.01	-0.01		0.11	0.16	-713.3	0
	0.07	-0.01			0.08	0.16	-712.2	1.1
	0.07		-0.01	-0.01	0.10	0.14	-711.4	1.8

model-implied covariance ($\chi^2 = 37.4$, $df = 13$, $P < 0.001$). Because $S_{i_{\text{metapop}}}$ did not have strong effects in our earlier models, we removed it as an indicator variable. This model fit the data well as indicated by no significant deviation of the observed and modeled-implied covariance ($\chi^2 = 11.94$, $df = 8$, $P = 0.15$). The removal of $S_{i_{\text{metapop}}}$ did not affect the strength or significance of associations. $S_{i_{\text{water}}}$ and $S_{i_{\text{forest}}}$ were the strongest and only significant indicators of connectivity (Fig. 3). Our model supported a significant negative effect of connectivity on *Pgi-c*, and a significant negative effect of *Pgi-c* on genetic differentiation, but no significant direct effect of connectivity on genetic differentiation (Fig. 3; Fig. S5).

Discussion

Here, we show that patch connectivity metrics incorporating landscape matrix best predicted the distribution of a dispersal polymorphism during a population recovery in the Glanville fritillary butterfly. In particular, newly colonized populations that were isolated by water and forest matrix had significantly higher

frequencies of an allele associated with increased dispersal ability (*Pgi-c* allele). We further found that patch connectivity alone did not predict genetic differentiation at neutral markers, but rather the effect of landscape on genetic structure was mediated through individual variation in the *Pgi* locus; populations with higher frequencies of the *Pgi-c* allele had lower F_{st} . In the following year when the density of populations increased, these relationships disappeared, suggesting that good dispersers only have an advantage when there are many empty patches to colonize. Together our results suggest that both individual variation in dispersal traits and landscape matrix heterogeneity are important for predicting spatial patterns of genetic variation.

LANDSCAPE MATRIX PREDICTS INDIVIDUAL VARIATION IN DISPERSAL

We found that spatial sorting of individuals based on their *Pgi* genotype was best explained by a connectivity metric that incorporated the effects of water and forest matrix in 2011. Importantly, the basic metapopulation model did not show a significant association with the frequency of *Pgi-c*. While previous work showed

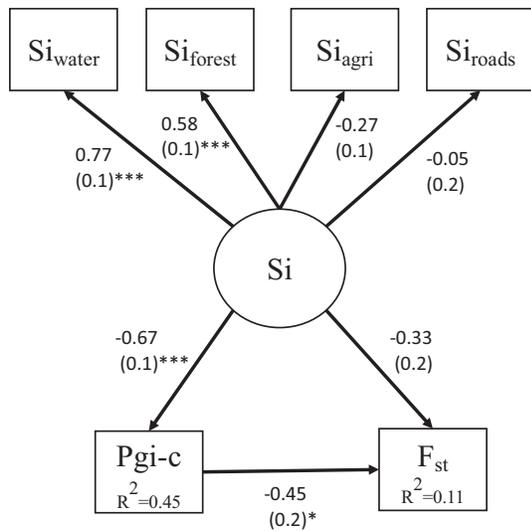


Figure 3. Results from a structural equation model test for direct and indirect effects of patch connectivity (S_i) on genetic structure (F_{st}) for new populations in 2011. Arrows show tested unidirectional relationship among variables. Connectivity (S_i) was included as a latent variable described by four observed indicator variables, S_{i_water} , S_{i_forest} , $S_{i_agriculture}$, and S_{i_roads} . Standardized coefficients and associated standard errors are shown beside their respective paths, and significant paths are marked with asterisks. Variance explained is shown in the boxes of the endogenous variables, P_{gi-c} and F_{st}

that metapopulation connectivity predicted the spatial sorting of individuals based on their *Pgi* genotype (Haag et al. 2005; Hanski and Saccheri 2006; Zheng et al. 2009), our results suggest that more complex processes are at play. There are a number of potential reasons why previous work was able to capture spatial sorting of *Pgi* with $S_{i_metapop}$ while we could not. We found evidence for a negative relationship between the frequency of *Pgi-c* and metapopulation connectivity for new populations, however, this relationship was "not" supported within individual genetic clusters (Fig. S2). In contrast, the association between *Pgi* and $S_{i_water+forest}$ held true even within genetic clusters (Fig. S2). This suggests that $S_{i_metapop}$ might, at least partly, be confounded with geography, whereas $S_{i_water+forest}$ captures connectivity at scales more relevant to dispersal. Our sample size is also much larger and we were thus able to capture a larger amount of variation in landscape structure. In comparison, previous work selected population extremes (e.g., extremely low and high connectivity), and geographic distance might have been sufficient to capture patterns. It should also be noted that previous studies tested only a single model of patch connectivity, whereas we compared several models assuming different landscape structures.

Our results hence suggest that the *Pgi* dispersal polymorphism in the *M. cinxia* system in the Åland islands is not maintained by variation in patch configuration alone (i.e., the

metapopulation model), but that the landscape matrix further influence dispersal patterns. This is an important finding given that studies investigating the drivers of dispersal evolution almost exclusively use simplified landscape models that assume a homogenous matrix (Bowler and Benton 2005; Henriques-Silva et al. 2015). Knowledge of the importance of the landscape matrix is thus lacking, and this is one avenue in which landscape genetic approaches can contribute to understanding how the matrix might modify predictions of dispersal evolution. Our results suggest that landscape features that intervene discrete habitat patches matter, and this is unsurprising given that dispersal traits are often correlated with other aspects of species biology (Saastamoinen et al. 2018). For example, *Pgi* shows a genotype-by-temperature interaction in *M. cinxia*, where heterozygotes have higher flight metabolic rate (and thus dispersal ability) at moderate and cool temperatures, but individuals without a *c* allele fly better at very warm temperatures (Niitepõld 2010). This might explain why forest is an important predictor of the spatial distribution of the dispersive allele, as it could provide a cooler environment for *Pgi-c* individuals. However, it is unclear if the association between forest and the frequency of *Pgi-c* is driven by individual differences in the use of the landscape matrix, or rather a distance effect—i.e., if *Pgi-c* individuals are able to fly further or faster through the forest.

INDIVIDUAL VARIATION IN DISPERSAL PREDICTS GENETIC DIFFERENTIATION

In our test of direct effects of connectivity on genetic differentiation, we found no evidence that any of the connectivity metrics predicted F_{st} in 2011—newly colonized patches with lower connectivity did not display higher genetic differentiation at neutral loci compared to highly connected (Table 2). Rather, patches with higher frequencies of *Pgi-c* had significantly lower F_{st} (Fig. 3; Fig. S5). Although *Pgi* explained only a small proportion of variation in F_{st} , the effects of the landscape matrix on dispersal and gene flow would have been completely missed if individual variation at the *Pgi* locus was not included in our model. This highlights the importance for integration of intraspecific variation in dispersal traits in landscape genetic studies. While sex differences have recently been considered in landscape genetic models (e.g., Paquette et al. 2014; Bertrand et al. 2017), to our knowledge, only two papers have integrated non-sex related variation in dispersal traits: DiLeo et al. (2014) found that individual variation in the number of flowers on dogwood trees influenced spatial patterns of gene flow beyond the effects of Euclidean distance; and McDevitt et al. (2013) found that genetic admixture of weasels in Poland was likely mediated by the movement of medium-, rather than large- or small-bodied weasels. While we acknowledge that variation in dispersal traits are often hard to measure, the increasing accessibility of genomic data will facilitate the identification of

candidate loci relevant to dispersal in non-model organisms (e.g., Swaegers et al. 2015; Dudaniec et al. 2018). Because dispersal is often controlled by multiple genes (Saastamoinen et al. 2018), applying landscape genomic methods that can capture the effects of polygenic adaptation will be important (e.g., redundancy and canonical correlation analysis; Rellstab et al. 2015 and reference therein). We further expect our approach to be relevant not only to systems where multiple dispersal strategies exist within a single landscape, but also the perhaps more common case of directional selection, where dispersal might be under positive or negative selection depending on broad patterns of landscape structure (Cheptou et al. 2017). An increasing number of studies have reported variability in landscape genetic relationships across replicate landscapes (e.g., Bull et al. 2011; Dudaniec et al. 2012; DiLeo et al. 2013; Balbi et al. 2018), and it will thus be interesting to see if divergent selection on dispersal genes might be a hidden source of variation contributing to these patterns (e.g., Peterson and Denno 1997).

Our study highlights the role of a single gene on the maintenance of gene flow across the landscape, and also joins a growing list of evidence that *Pgi* in particular has large effects on ecological processes in *M. cinxia* (reviewed in Niitepõld and Saastamoinen 2017). However, it is unlikely that *Pgi* is acting alone. For example, work by Wheat et al. (2011) showed that *Pgi* may epistatically interact with other genes, such as succinate dehydrogenase (*Sdh*); an allelic combination at these two loci was associated with maximal metabolic endurance in *M. cinxia*. Linkage of *Pgi* to other functional loci have not been resolved, but the low frequency and fitness of individuals homozygous for the C allele suggest possible linkage to a deleterious mutation (Orsini et al. 2009), and that the variation in *Pgi* is maintained through balancing selection via a heterozygote advantage (Wheat et al. 2009). Further, females colonizing new populations have been found to be divergent in a suite of life-history traits, many but not all are associated with variation in *Pgi* (Hanski et al. 2006; Saastamoinen 2008; Kvist et al. 2013; Wheat et al. 2011). This suggest a more complex dispersal syndrome, the genomic architecture of which remains to be characterized. We further emphasize that *Pgi* explained only 10% of variation in F_{st} , and clearly other processes are at play that may not have been captured by our model (e.g., temperature and condition-dependent dispersal). Interestingly, the four other candidate loci included in our study did not show significant associations with landscape structure, despite being found as outliers in previous work (Kvist et al. 2015; Fountain et al. 2016). In particular, Fountain et al. (2016) found that three of these loci changed more than expected under neutral expectations in contemporary vs. museum samples. Further, they found that allele frequencies changed in the same direction in comparisons of new, isolated vs. old continuous contemporary Åland populations, and in fragmented

vs. continuous replicate landscapes. In our study, these three loci (Mc1:1873:36910, Mc1:1124:71239, Mc1:1687:14486) shifted in new, isolated vs. old, connected populations in the same direction as reported by Fountain et al. (2016), although the differences were subtle (Fig. S7). This suggest that these loci may indeed be under selection, but we lack the power to detect significance given that we have samples from very different timescales compared to Fountain et al. (2016).

CONTEXT MATTERS

Associations between patch connectivity and variation in *Pgi* in new populations in 2011 disappeared in 2012. Modeling studies on *Pgi* indicates that *Pgi-c* individuals should have the greatest selective advantage when there are many empty patches to colonize (Zheng et al. 2009; Hanski et al. 2011). Thus, it was expected that our results would be much stronger in 2011—a year that marked the largest number of re-colonizations recorded in Åland following a large population decline that left many empty patches. In comparison, the metapopulation experienced a large increase in population size in 2012 but relatively fewer colonization events; all patches in 2012 had high connectivity. This appears to be driven by the much higher number of potential source patches and nests in sources in 2012, and less by difference in distances between sources and targets (Fig. S6). Observations from mark-recapture suggest that *M. cinxia* exhibit negative density dependent dispersal (Kuussaari et al. 1996), suggesting that there should be fewer dispersal events in the high density year 2012 compared to 2011. Intriguingly, effects of connectivity on *Pgi* even appeared to switch in 2012 (Table 1; Fig. S4), although these associations were not significant in latent factor mixed models. This might be suggestive of a more complex interaction between individual variation and density-dependent dispersal. Modeling work predicts that *Pgi-c* should rise to higher frequencies at very high population densities where it gains an advantage by spreading genes over more patches (Zheng et al. 2009), however this has not been empirically demonstrated. Future work should seek to resolve the drivers of yearly differences in dispersal, focusing particularly on the effects of density and weather, which may influence dispersive and nondispersive genotypes differently.

Importantly, the association between *Pgi* and F_{st} in 2011 suggests that the polymorphism plays a key role in maintaining genetic variation across the landscape following perturbation. This finding provides a more mechanistic understanding of population persistence in this highly dynamic system. Recent work showed that regions in Åland with higher long-term frequencies of *Pgi-c* maintained higher metapopulation sizes, presumably by increasing colonization rates (Hanski et al. 2017). Our results suggest that regional persistence of the metapopulation might be further facilitated through *Pgi*-mediated genetic rescue.

MODEL UNCERTAINTY AND SAMPLING EFFECTS

While some strong associations emerged from our analysis, model selection suffered from uncertainty with several likely, and sometimes non-nested models appearing to have similar support. This is a common problem with variables derived from landscape measures, which are inherently correlated (Smith et al. 2009; Prunier et al. 2015). Although our connectivity variables were well below typical collinearity thresholds (Dormann et al. 2013), it is likely that weak linear relationships still contributed to this uncertainty. This might also explain why connectivity variables with strong effects in 1 year did not emerge as important predictors in the next year, although part of this is also likely due to differences in spatial locations of populations sampled in the different years. Although it is hard to say definitively which landscape features restrict dispersal, our results make a strong case for water as it was an important predictor across years, and forest as it had strong effects across multiple methods and different partitions of the data. What is less clear are the effects of other variables that were found to be important for prediction but of weak effect, with inconsistent results across methods (e.g., $S_{\text{agriculture}}$ in 2011). Future work would benefit from fine-tuning landscape resistance surfaces to better account for these potential small additive effects (e.g., using optimization; Peterman 2014), and from testing relationships under a broader set of conditions in carefully selected landscapes where the independent effects of landscape variables can be better teased apart.

Future work is also required to determine the effects of *Pgi* in small populations (1–2 nests), which were excluded from our analysis because estimating F_{st} requires larger samples (Morin et al. 2009). *Pgi-c* individuals might be especially important to counteract drift in these small populations, and individual-based genetic approaches (Shirk et al. 2018) could be employed in the future to better quantify genetic structure including these patches. However, populations founded by a single female likely contribute little to overall population dynamics (Hanski et al. 2017), as these populations would be prone to inbreeding depression in the following generation (Haikola et al. 2001; Nieminen et al. 2001), and small, inbred populations have high observed extinction rates in the field (Saccheri et al. 1998)

Conclusions

Our work adds to growing evidence that intraspecific variation plays a key role in driving diverse biological processes (Bolnick et al. 2011; Moran et al. 2016; Des Roches et al. 2018). We showed that heterogeneity in the landscape matrix is an important predictor of spatial variation in dispersal traits, and that this individual variation mediated the effects of landscape on genetic structure. Our results therefore highlight a need for better integration of studies on dispersal evolution and landscape genetics. While studies

of dispersal evolution may need to consider more complex representations of landscape structure that captures heterogeneity in the landscape matrix, landscape geneticists should consider that key associations between landscape and genetic structure might be missed if intraspecific variation in dispersal is ignored.

ACKNOWLEDGMENTS

The authors thank Annukka Ruokolainen, Toshka Nyman, and Laura Häkkinen for sample preparation and DNA extraction and Swee Wong for post-processing of SNP data. The authors also thank Rolf Holderegger, Marie-Josée Fortin, Torsti Schulz, and Aapo Kahilainen for helpful feedback during the preparation of this manuscript. This study was funded by the Academy of Finland (decision number 265641 to M.S., and Centre of Excellence grant 284601) and the European Research Council (Independent Starting grant no. 637412 ‘META-STRESS’ to M.S.).

AUTHOR CONTRIBUTIONS

M.F.D. analyzed the data and wrote the first draft of the manuscript. A.H. and M.S. contributed substantially to the study design and revisions of the manuscript.

LITERATURE CITED

- Ahola, V., R. Lehtonen, P. Somervuo, L. Salmela, P. Koskinen, P. Rastas, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat. Commun.* 5:4737.
- Ahroth, P., R. V. Alatalo, A. Holopainen, T. Kumpulainen and J. Suhonen. 2003. Founder population size and number of source populations enhance colonization success in waterstriders. *Oecologia* 137:617–620
- Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike’s information criterion. *J. Wildlife Manage.* 74:1175–1178.
- Balbi, M., A. Ernoult, P. Poli, L. Madec, A. Guiller, M.-C. Martin, et al. 2018. Functional connectivity in replicated urban landscapes in the land snail (*Cornu aspersum*). *Mol. Ecol.* <https://doi.org/10.1111/mec.14521>.
- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67:1–48.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57:289–300.
- Bertrand, P., J. Bowman, R. J. Dyer, M. Manseau and P. J. Wilson. 2017. Sex-specific graphs: Relating group-specific topology to demographic and landscape data. *Mol. Ecol.* 26:3898–3912.
- Boggs, C. L., and M. Nieminen. 2004. Checkerspot reproductive biology. In: *On the wings of checkerspot: a model system for population biology* (Eds. Ehrlich P. R., Hanski I.). Oxford Univ. Press, New York, NY, 92–111.
- Bolnick, D. I., P. Amarasekare, M. S. Araujo, R. Burger, J. M. Levine, M. Novak, et al. 2011. Why intraspecific trait variation matters in community ecology. *Trends. Ecol. Evol.* 26:183–192.
- Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biol. Rev.* 80:205–225
- Bull, R. A. S., S. A. Cushman, R. Mace, T. Chilton, K. C. Kendall, E. L. Landguth, et al. 2011. Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Mol. Ecol.* 20:1092–1107.
- Cade, B. S. 2015. Model averaging and muddled multimodel inferences. *Ecology* 96:2370–82.

- Cheng, L., T. R. Connor, J. Siren, D. M. Aanensen, and J. Corander. 2013. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Mol. Biol. Evol.* 30:1224–1228.
- Cheptou, P. O., A. L. Hargreaves, D. Bonte, and J. Jacquemyn. 2017. Adaptation to fragmentation: evolutionary dynamics driven by human influences. *Philos. Trans. R. Soc. B* 372:20160037.
- Clobert, J., J. F. Le Galliard, J. Cote, S. Meylan, and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecol. Lett.* 12:197–209.
- Corander, J., J. Siren, and E. Arjas. 2008. Bayesian spatial modeling of genetic population structure. *Computation Stat* 23:111–129.
- Cote, J., E. Bestion, S. Jacob, J. Travis, D. Legrand, and M. Baguette. 2017. Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes. *Ecography* 40:56–73.
- Cote, J., T. Brodin, S. Fogarty, and A. Sih. 2017. Non-random dispersal mediates invader impacts on the invertebrate community. *J. Anim. Ecol.* 86:1298–1307.
- Delgado, M. D., V. Penteriani, E. Revilla, and V. O. Nams. 2010. The effect of phenotypic traits and external cues on natal dispersal movements. *J. Anim. Ecol.* 79:620–632.
- Des Roches, S., D. M. Post, N. E. Turley, J. K. Bailey, A. P. Hendry, M. T. Kinnison, et al. 2018. The ecological importance of intraspecific variation. *Nat. Ecol. Evol.* 2:57.
- DiLeo, M. F., J. C. Siu, M. K. Rhodes, A. Lopez-Villalobos, A. Redwine, K. Ksiazek, et al. 2014. The gravity of pollination: integrating at-site features into spatial analysis of contemporary pollen movement. *Mol. Ecol.* 23:3973–3982.
- DiLeo, M. F., J. D. Rouse, J. A. Davila, and S. C. Lougheed. 2013. The influence of landscape on gene flow in the eastern massasauga rattlesnake (*Sistrurus c. catenatus*): insight from computer simulations. *Mol. Ecol.* 22:4483–4498.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46.
- Duckworth, R. A., and A. V. Badyaev. 2007. Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proc. Natl. Acad. Sci. USA.* 104:15017–15022.
- Dudaniec, R. Y., S. F. Spear, J. S. Richardson, and A. Storfer. 2012. Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. *PLoS One* 7:e36769.
- Dudaniec, R. Y., C. J. Yong, L. T. Lancaster, E. I. Svensson, and B. Hansson. 2018. Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*). *Mol. Ecol.* 27:2576–93.
- Duploux, A., S. C. Wong, J. Corander, R. Lehtonen, and I. Hanski. 2017. Genetic effects on life-history traits in the Glanville fritillary butterfly. *PeerJ* 5:e3371.
- Edelaar, P., and D. I. Bolnick. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. *Trends Ecol. Evol.* 27:659–665.
- Elliott, E. C., and S. J. Cornell. 2012. Dispersal Polymorphism and the Speed of Biological Invasions. *PLoS One* 7:0040496.
- Fountain, T., M. Nieminen, J. Siren, S. C. Wong, and I. Hanski. 2016. Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proc. Natl. Acad. Sci. USA.* 113:2678–2683.
- Fountain, T., A. Husby, E. Nonaka, M. F. DiLeo, J. H. Korhonen, P. Rastas, T. Schulz, M. Saastamoinen, and I. Hanski. 2018. Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly. *Evol. Appl.* 11:287–97.
- Francois, O., H. Martins, K. Caye, and S. D. Schoville. 2016. Controlling false discoveries in genome scans for selection. *Mol. Ecol.* 25:454–469.
- Frichot, E., and O. François. 2015. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* 6:925–929.
- Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* 30:1687–1699.
- Galipaud, M., M. A. Gillingham, M. David, and F. X. Dechaume-Moncharmont. 2014. Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. *Methods Ecol. Evol.* 5:983–91.
- Haag, C. R., M. Saastamoinen, J. H. Marden, and I. Hanski. 2005. A candidate locus for variation in dispersal rate in a butterfly metapopulation. *Proc. R. Soc. B* 272:2449–2456.
- Haikola, S., W. Fortelius, R. B. O'hara, M. Kuussaari, N. Wahlberg, I. J. Saccheri, et al. 2001. Inbreeding depression and the maintenance of genetic load in *Melitaea cinxia* metapopulations. *Conserv. Genetics* 2: 325–35.
- Hanski, I. 1994. A practical model of metapopulation dynamics. *J. Anim. Ecol.* 63:151–162.
- Hanski, I., and I. Saccheri. 2006. Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biol.* 4:719–726.
- Hanski, I., T. Schulz, S. C. Wong, V. Ahola, A. Ruokolainen, and S. P. Ojanen. 2017. Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nat. Commun.* 8.
- Hanski, I., T. Mononen, and O. Ovaskainen. 2011. Eco-evolutionary metapopulation dynamics and the spatial scale of adaptation. *Am. Nat.* 177: 29–43.
- Henriques-Silva, R., F. Boivin, V. Calcagno, M. C. Urban, and P. R. Peres-Neto. 2015. On the evolution of dispersal via heterogeneity in spatial connectivity. *Proc R. Soc B* 282:20142879.
- Holderegger, R., and H. H. Wagner. 2008. Landscape genetics. *Bioscience* 58:199–207.
- Kuussaari, M., M. Nieminen and I. Hanski. 1996. An experimental study of migration in the Glanville fritillary butterfly *Melitaea cinxia*. *J. Anim. Ecol.* 65:791–801.
- Kvist, J., C. W. Wheat, E. Kallioniemi, M. Saastamoinen, I. Hanski, and M. J. Frilander. 2013. Temperature treatments during larval development reveal extensive heritable and plastic variation in gene expression and life history traits. *Mol. Ecol.* 22:602–619.
- Kvist, J., A. L. Mattila, P. Somervuo, V. Ahola, P. Koskinen, L. Paulin, et al. 2015. Flight-induced changes in gene expression in the Glanville fritillary butterfly. *Mol. Ecol.* 24:4886–900.
- Legrand, D., J. Cote, E. A. Fronhofer, R. D. Holt, O. Ronce, N. Schtickzelle, et al. 2017. Eco-evolutionary dynamics in fragmented landscapes. *Ecography* 40:9–25.
- Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18:189–197.
- McDevitt, A. D., M. K. Oliver, S. B. Piertney, P. A. Szafranska, M. Konarzewski, and K. Zub. 2013. Individual variation in dispersal associated with phenotype influences fine-scale genetic structure in weasels. *Conserv. Genet.* 14:499–509.
- McRae, B. H. 2006. Isolation by resistance. *Evolution.* 60:1551–1561.
- Merckx, T., and H. Van Dyck. 2007. Habitat fragmentation affects habitat-finding ability of the speckled wood butterfly, *Pararge aegeria* L. *Anim. Behav.* 74:1029–1037.
- Moran, E. V., F. Hartig, and D. M. Bell. 2016. Intraspecific trait variation across scales: implications for understanding global change responses. *Glob. Change Biol.* 22:137–150.
- Morin, P. A., K. M. Karen, and B. L. Taylor. 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Resour.* 9: 66–73.

- Nair, A., T. Fountain, S. Ikonen, S. P. Ojanen, and S. van Nouhuys. 2016. Spatial and temporal genetic structure at the fourth trophic level in a fragmented landscape. *Proc. R. Soc. B.* 283:20160668.
- Nieminen, M., M. C. Singer, W. Fortelius, K. Schöps, and I. Hanski. 2001. Experimental confirmation that inbreeding depression increases extinction risk in butterfly populations. *Am. Nat.* 157:237–44.
- Niitepõld, K., A. D. Smith, J. L. Osborne, D. R. Reynolds, N. L. Carreck, A. P. Martin, et al. 2009. Flight metabolic rate and Pgi genotype influence butterfly dispersal rate in the field. *Ecology* 90:2223–32.
- Niitepõld, K. 2010. Genotype by temperature interactions in the metabolic rate of the Glanville fritillary butterfly. *J. Exp. Biol.* 213:1042–1048.
- Niitepõld, K., and M. Saastamoinen. 2017. A candidate gene in an ecological model species: Phosphoglucose isomerase (Pgi) in the Glanville fritillary butterfly (*Melitaea cinxia*). *Ann. Zool. Fenn* 54:259–273.
- Ochocki, B. M., and T. E. X. Miller. 2017. Rapid evolution of dispersal ability makes biological invasions faster and more variable. *Nat. Commun.* 8:14315.
- Ojanen, S. P., M. Nieminen, E. Meyke, J. Pöyry, and I. Hanski. 2013. Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): survey methods, data management, and long-term population trends. *Ecol. Evol.*, 3:3713–37.
- Orsini, L., C. W. Wheat, C. R. Haag, J. Kvist, M. J. Frilander, and I. Hanski. 2009. Fitness differences associated with Pgi SNP genotypes in the Glanville fritillary butterfly (*Melitaea cinxia*). *J. Evol. Biol.* 22:367–375.
- Palmer, S. C. F., A. Coulon, and J. M. J. Travis. 2014. Inter-individual variability in dispersal behaviours impacts connectivity estimates. *Oikos* 123:923–932.
- Paquette, S. R., B. Talbot, D. Garant, J. Mainguy, and F. Pelletier. 2014. Modelling the dispersal of the two main hosts of the raccoon rabies variant in heterogeneous environments with landscape genetics. *Evol. Appl.* 7:734–749.
- Peterman, W. E. 2014. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol. Evol.* 9:1638–1647.
- Peterman, W. E., T. L. Anderson, B. H. Ousterhout, D. L. Drake, R. D. Semlitsch, and L. S. Eggert. 2015. Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conserv. Genet.* 16:59–69.
- Peterson, M. A., and R. F. Denno. 1997. The influence of intraspecific variation in dispersal strategies on the genetic structure of planthopper populations. *Evolution*. 51:1189–1206.
- Pflueger, F. J., and N. Balkenhol. 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Mol. Ecol.* 23:2146–2156.
- Phillips, B. L., G. P. Brown, J. K. Webb, and R. Shine. 2006. Invasion and the evolution of speed in toads. *Nature* 439:803–803.
- Prunier, J. G., M. Colyn, X. Legendre, K. F. Nimon, M. C. Flamand. 2015. Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. *Mol. Ecol.* 24:263–83.
- Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24:4348–4370.
- Rius, M., and J. A. Darling. 2014. How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol. Evol.* 29:233–242.
- Rosseel, Y. 2012. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* 48:1–36.
- Saastamoinen, M. 2008. Heritability of dispersal rate and other life history traits in the Glanville fritillary butterfly. *Heredity* 100:39.
- Saastamoinen, M., G. Bocedi, J. Cote, D. Legrand, F. Guillaume, C. W. Wheat, et al. 2018. Genetics of dispersal. *Biol. Rev.* 93:574–599.
- Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature*. 392:491.
- Shirk, A. J., E. L. Landguth, and S. A. Cushman. 2018. A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Mol. Ecol. Res.* 18: 55–67.
- Smith, A. C., N. Koper, C. M. Francis, and L. Fahrig. 2009. Confronting collinearity: comparing methods for disentangling the effects of habitat loss and fragmentation. *Land Ecol.* 24:1271.
- Steele, C. A., J. Baumsteiger, and A. Storfer. 2009. Influence of life-history variation on the genetic structure of two sympatric salamander taxa. *Mol. Ecol.* 18:1629–1639.
- Swaegers, J., J. Mergeay, A. Van Geystelen, L. Therry, M. H. Larmuseau, and R. Stoks. 2015. Neutral and adaptive genomic signatures of rapid poleward range expansion. *Mol. Ecol.* 24:6163–76.
- Szucs, M., B. A. Melbourne, T. Tuff, C. Weiss-Lehman, and R. A. Hufbauer. 2017. Genetic and demographic founder effects have long-term fitness consequences for colonising populations. *Ecol. Lett.* 20:436–444.
- Wagner, N. K., B. M. Ochocki, K. M. Crawford, A. Compagnoni, and T. E. X. Miller. 2017. Genetic mixture of multiple source populations accelerates invasive range expansion. *J. Anim. Ecol.* 86:21–34.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating f-statistics for the analysis of population-structure. *Evolution*. 38:1358–1370.
- Wheat, C. W., C. R. Haag, J. H. Marden, I. Hanski, and M. J. Frilander. 2009. Nucleotide polymorphism at a gene (Pgi) under balancing selection in a butterfly metapopulation. *Mol. Biol. Evol.* 27:267–81.
- Wheat, C. W., H. W. Fescemyer, J. Kvist, E. V. Tas, J. C. Vera, M. J. Frilander, et al. 2011. Functional genomics of life history variation in a butterfly metapopulation. *Mol. Ecol.* 20:1813–28.
- Willing, E.-M., C. Dreyer, and C. Van Oosterhout. 2012. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One* 7:e42649.
- Wong, S. C., A. Oksanen, A. L. Mattila, R. Lehtonen, K. Niitepõld, and I. Hanski. 2016. Effects of ambient and preceding temperatures and metabolic genes on flight metabolism in the Glanville fritillary butterfly. *J. Insect. Physiol.* 85:23–31.
- Zheng, C. Z., O. Ovaskainen, and I. Hanski. 2009. Modelling single nucleotide effects in phosphoglucose isomerase on dispersal in the Glanville fritillary butterfly: coupling of ecological and evolutionary dynamics. *Philos. Trans. R. Soc. B* 364:1519–1532.

Associate Editor: Z. Gompert

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Cluster membership estimated from BAPS for 2011 (A) and 2012 (B). Each point is a patch and the colour indicates membership to a particular cluster.

Figure S2. Associations between the frequency of the *Pgi-c* allele and two indices of connectivity ($S_{i_{\text{water+forest}}}$ and log-transformed $S_{i_{\text{metapop}}}$) for each genetic cluster in 2011 (A) and 2012 (B).

Figure S3. Q-Q plots showing observed and expected $-\log_{10}$ p-values for latent factor mixed models testing for associations between 39 loci and connectivity metrics in new populations in 2011 (A-E) and Supporting Figures – DiLeo et al. *Evolution Letters* 2012 (F-G).

Figure S4. Scatterplots showing the relationship between the frequency of the *Pgi-c* allele and a patch connectivity metric incorporating roads as a facilitator to dispersal (A) and water as a barrier to dispersal (B) for the year 2012.

Figure S5. Scatterplots showing relationships between population-specific genetic differentiation (F_{st}) and the frequency of the *Pgi-c* allele (A), F_{st} and $S_{i_{\text{forest}}}$ (B) and F_{st} and $S_{i_{\text{water}}}$ (C) for newly colonized populations in 2011.

Figure S6. Sources of variation in patch connectivity per year. Yearly differences in connectivity for newly colonized populations, $S_{i_{\text{metapop}}}$, (A), geographic distance between source and target patches for newly colonized populations, d_{ij} (B), number of nests in source patches, N_j (C), and total number of source patches (D).

Figure S7. Relationships between connectivity and minor allele frequencies (MAF) of the four other candidate loci in newly colonized and old populations.

Table S1. Descriptions of landscape connectivity hypotheses.

Table S2. Pearson correlation coefficients among the five landscape connectivity variables.

Table S3. Posterior mean estimates, standard deviation, and quantiles of Bayesian nonspatial and spatial INLA models.

Table S4. Results of model selection on linear mixed effect models testing for associations between the frequency of the *Pgi-c* allele, population age (old or new), and competing metrics of patch connectivity (S_i) for the year 2011.

Table S5. Results of model selection on linear mixed effect models testing for associations between the frequency of the *Pgi-c* allele, population age (old or new), and competing metrics of patch connectivity (S_i) for the year 2012. Only the top 15 models are shown.

Table S6. Genomic inflation factors of for each value of k and each predictor tested in latent factor mixed models.

Table S7. Results of model selection on linear mixed effect models testing for associations between population genetic differentiation (F_{st}), population age (old or new), and competing metrics of patch connectivity (S_i) for the year 2011.

Table S8. Results of model selection on linear mixed effect models testing for associations between population genetic differentiation (F_{st}), population age (old or new), and competing metrics of patch connectivity (S_i) for the year 2011 with a singly outlier with a high F_{st} value removed. Only the top 15 models are shown.

Table S9. Results of model selection on linear mixed effect models testing for associations between population genetic differentiation (F_{st}), population age (old or new), and competing metrics of patch connectivity (S_i) for the year 2012.

Appendix A – Supplementary methods and results

Appendix B: Information about the 5 candidate and 40 neutral SNPs used.