





# Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly

Toby Fountain<sup>1,2</sup>  | Arild Husby<sup>1</sup> | Etsuko Nonaka<sup>1,3</sup>  | Michelle F. DiLeo<sup>1</sup>  |  
 Janne H. Korhonen<sup>1,4</sup> | Pasi Rastas<sup>1,5</sup> | Torsti Schulz<sup>1</sup>  | Marjo Saastamoinen<sup>1</sup>  |  
 Ilkka Hanski<sup>1†</sup>

<sup>1</sup>Department of Biosciences, University of Helsinki, Helsinki, Finland

<sup>2</sup>Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

<sup>3</sup>Department of Biology, Lund University, Lund, Sweden

<sup>4</sup>Department of Computer Science, Aalto University, Aalto, Finland

<sup>5</sup>Department of Zoology, University of Cambridge, Cambridge, UK

## Correspondence

Toby Fountain, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.  
 Email: toby.fountain@ebc.uu.se

## Funding information

Biotieteiden ja Ympäristön Tutkimuksen Toimikunta, Grant/Award Number: 283108 and 284601

## Abstract

Dispersal is important for determining both species ecological processes, such as population viability, and its evolutionary processes, like gene flow and local adaptation. Yet obtaining accurate estimates in the wild through direct observation can be challenging or even impossible, particularly over large spatial and temporal scales. Genotyping many individuals from wild populations can provide detailed inferences about dispersal. We therefore utilized genomewide marker data to estimate dispersal in the classic metapopulation of the Glanville fritillary butterfly (*Melitaea cinxia* L.), in the Åland Islands in SW Finland. This is an ideal system to test the effectiveness of this approach due to the wealth of information already available covering dispersal across small spatial and temporal scales, but lack of information at larger spatial and temporal scales. We sampled three larvae per larval family group from 3732 groups over a six-year period and genotyped for 272 SNPs across the genome. We used this empirical data set to reconstruct cases where full-sibs were detected in different local populations to infer female effective dispersal distance, that is, dispersal events directly contributing to gene flow. On average this was one kilometre, closely matching previous dispersal estimates made using direct observation. To evaluate our power to detect full-sib families, we performed forward simulations using an individual-based model constructed and parameterized for the Glanville fritillary metapopulation. Using these simulations, 100% of predicted full-sibs were correct and over 98% of all true full-sib pairs were detected. We therefore demonstrate that even in a highly dynamic system with a relatively small number of markers, we can accurately reconstruct full-sib families and for the first time make inferences on female effective dispersal. This highlights the utility of this approach in systems where it has previously been impossible to obtain accurate estimates of dispersal over both ecological and evolutionary scales.

## KEYWORDS

Colonization, COLONY, conservation genetics, dispersal, *Melitaea cinxia*, sibship reconstruction

<sup>†</sup>Deceased

## 1 | INTRODUCTION

The increasing availability of genomic information continues to transform the way we study natural populations. It is now possible to accurately and efficiently measure a wide range of important parameters that directly influence the fitness and survival of wild populations such as effective population size (Gilbert & Whitlock, 2015; Palstra & Fraser, 2012), effective number of breeders (Ackerman et al., 2017), extra pair paternity (Firth, Hadfield, Santure, Slate, & Sheldon, 2015; Griffith, Owens, & Thuman, 2002), heterozygosity (Fountain et al., 2016; Saccheri et al., 1998), inbreeding depression (Huisman, Kruuk, Ellis, Clutton-Brock, & Pemberton, 2016) and reproductive success (Coltman et al., 1999). Another key ecological parameter is dispersal, the ecological and evolutionary causes and consequences of which have been studied for decades. While for some species, including many birds and mammals, very detailed information on individual dispersal events can be obtained by tracking individuals in their natural habitat (e.g., Telfer et al., 2003; Tøttrup et al., 2012), for many others measuring dispersal in the field is difficult or even impossible (Clobert, Danchin, Dhondt, & Nichols, 2001). The use of molecular data makes it possible to examine the role of dispersal and gene flow in influencing the long-term and large-scale spatial genetic structure of populations in such species (Slatkin, 1985), providing an opportunity to investigate dispersal at previously unobtainable ecological and evolutionary scales, as well as in organisms where no dispersal data has previously been available. Additionally, this approach allows the estimation of effective dispersal, that is, dispersal directly associated with reproductive fitness, often a more biologically relevant parameter than general dispersal.

A large number of different estimators have therefore been proposed to infer dispersal from marker data. Initially,  $F$  statistics were used to indirectly infer spatial distribution of genetic variation using assumptions from Wright's island model (e.g., Dobzhansky & Wright, 1943). However, it became apparent that this approach has several limitations (reviewed in Whitlock & McCauley, 1999), including the inability to disentangle contemporary versus historical gene flow and dispersal, as well as poor performance in nonequilibrium populations where assumptions of the Wright's island model are violated. Assignment methods, where individuals are probabilistically assigned to populations based on their multilocus genotypes, were seen as an improved approach, providing contemporary estimates of gene flow and dispersal (Pritchard, Stephens, & Donnelly, 2000; Rannala & Mountain, 1997; Waser & Strobeck, 1998). However, power to detect dispersal with these methods relies on high levels of genetic structure between populations (Berry, Tocher, & Sarre, 2004; Paetkau, Slade, Burden, & Estoup, 2004; although see Hall et al., 2009), reducing their applicability in demographically dynamic populations (Lowe & Allendorf, 2010).

An alternative approach has been to infer dispersal events through first reconstructing relatedness between individuals (e.g., Lepais et al., 2010; Schunter, Pascual, Garza, Raventos, & Macpherson, 2014). For example, if two closely related individuals (e.g., parent-offspring or siblings), whose relationship is established

through molecular markers, are sampled in different populations, a dispersal event can be inferred. The utility of this approach has previously been limited by large uncertainties in relatedness estimates resulting from the limited number of available genetic markers and hence difficulties in statistically discriminating even closely related individuals from unrelated. For example, the central idea of relatedness estimation using genetic data is to infer genealogical relationship by comparing similarities in their multilocus genotypes given the Mendelian laws of inheritance. Therefore, the performance of these estimators depends on the number of loci (and alleles at a particular locus), allele frequencies and variance in relatedness within the populations from which the estimates are derived (Csilléry et al., 2006). With the advent of high-throughput sequencing (NGS) technology, it is now both relatively easy and economically feasible to genotype many individuals at hundreds or even thousands of loci, something that will considerably reduce uncertainty surrounding the relatedness estimates.

This uncertainty has been further reduced with significant methodological developments, including moving from looking at pairwise comparisons between individuals (e.g., Jones & Ardren, 2003), to considering multiple individuals simultaneously to reconstruct family relationships (Wang, 2007). This not only increases power by taking into account the genotypes of both of the parents, but can also reduce errors when inferring relationships (Wang, 2007), allowing the accurate reconstruction of even pedigrees in wild populations. A major improvement to pedigree reconstruction methods has been to consider the joint likelihood of relatedness among all individuals in the sample (Wang, 2004, 2007; Wang & Santure, 2009). This method is implemented in the software COLONY (Jones & Wang, 2010), which also considers missing genotypes and genotyping errors (Wang & Santure, 2009), something that can have large influence on the relatedness inferences if not modelled correctly (Wang, 2004). Incorporating error rates is particularly important given that current applications of pedigree reconstruction are likely to use NGS data that can have high error rates and missing genotypes (Glenn, 2011). It is therefore an ideal time to apply these methods to systems where it has previously been challenging or even impossible to have accurate estimates of dispersal across relevant biological spatial scales (Broquet & Petit, 2009).

Here, we report a study on the Glanville fritillary butterfly (*Melitaea cinxia*), in which molecular markers (SNPs) were used to infer dispersal across a highly dynamic natural metapopulation system over a six-year period. The Glanville fritillary has been studied for more than two decades in the Åland Islands in Finland, in a large network of about 4,000 small dry meadows (Hanski, 2011). While mark-release-recapture studies of individual butterflies (Hanski, Kuussaari, & Nieminen, 1994; Kuussaari, Nieminen, & Hanski, 1996), tracking studies of individuals with harmonic radar (Ovaskainen et al., 2008) and studies of recolonization of currently unoccupied but suitable habitat patches (van Nouhuys & Hanski, 2002) have provided much information about the extent and spatial scale of dispersal (Hanski, 2011), it has not been possible to examine dispersal at larger temporal and spatial scales. For example, previous work has been unable to directly estimate the levels

of gene flow among populations, a parameter of key importance for long-term persistence (Saccheri et al., 1998).

To address this, we sampled around 10,000 individuals (three larvae sampled from each larval nest) from three different regions of the metapopulation over a six-year period and genotyped them for 272 SNPs across the genome. By reconstructing family groups of full-sibs with known geographical position, it is possible to reconstruct the distance an individual female moved between egg laying, thereby inferring breeding dispersal distance. We additionally gained insight into another ecological parameter of importance to the population survival and dynamics of the species, namely the estimated number of surviving larval clutches produced by a single female during her lifetime. The large geographical and temporal scale of this study provides an unprecedented insight into dispersal movements and performance of individuals in a natural insect system.

## 2 | MATERIALS AND METHODS

### 2.1 | Study species and sampling

The Glanville fritillary butterfly is common in the Åland Islands in SW Finland, where it inhabits a large network of approximately 4,000 dry meadows that have been fully surveyed since 1993 (Hanski, 2011). Females lay their eggs in large clutches of 150–200 eggs at intervals of one to several days depending on weather conditions (Boggs & Nieminen, 2004). The larvae live gregariously until the last larval instar in the following spring, having spent the winter in a silken “winter nest” spun by the larvae at the base of the host plant. The offspring of two or more females may become mixed either because several females oviposit on the same host plant (M. Saastamoinen, pers. obs.) or because the larval groups merge following short movements while the larvae switch from one host plant individual to another (Kuussaari, van Nouhuys, Hellmann, & Singer, 2004). For further details of the life history, see Ehrlich and Hanski (2004) and for a description of the study area and field methods (Ojanen, Nieminen, Meyke, Pöyry, & Hanski, 2013).

In late August to early September in each year, the entire 4,000 meadows in the patch network are surveyed for the number of larval groups. We estimate that the probability of missing an existing population (false negatives) is around 10%, and these populations are mostly very small, consisting of just one or a few larval groups (Ojanen et al., 2013). The probability of detecting a larval group in a population is estimated to be around 50%–60% (Ojanen et al., 2013).

Since 2007, we have sampled three larvae per larval group for experiments and for DNA sampling. In this study, we analysed data collected from three separate regions in the years from 2007 until 2012 (Fig. S1). Our primary region of interest was the mainland area of Saltvik, where we have sampled a network of 235 patches. For comparison, we also sampled from the islands of Föglö and Sottunga, which have patch networks with 125 and 49 patches, respectively. Sottunga was unoccupied in 1991, at which point 62 larval groups were translocated there from the Finström region of Åland, which neighbours Saltvik, and the metapopulation has persisted ever since.

### 2.2 | DNA extraction and genotyping

Larval tissue was homogenized prior to extraction using TissueLyser (Qiagen) at 30/s for 1.5 mins with Tungsten Carbide Beads, 3 mm (Qiagen). DNA was extracted using the NucleoSpin 96 Tissue Core Kit (Macherey-Nagel). Where DNA yield was low, extracted DNA underwent two rounds of Whole Genome Amplification (WGA) (LGC Genomics). Genotyping was performed using a panel of 272 SNP markers on the KASP platform. Three separate criteria were used to select the panel of SNP markers. Markers from candidate genes relating to flight or dispersal traits, or genes that were differentially expressed after an experimental flight treatment were selected on the basis of previous studies ( $n = 184$ ; (Saastamoinen, Ikonen, Wong, Lehtonen, & Hanski, 2013; Somervuo et al., 2014; de Jong, Wong, Lehtonen, & Hanski, 2014; Kvist et al., 2015). Putatively neutral SNPs ( $n = 40$ ) from noncoding regions of the genome were also selected, with the remaining 44 SNPs chosen to ensure that all chromosomes were represented, based on linkage map information (Rastas, Paulin, Hanski, Lehtonen, & Auvinen, 2013). Initial SNP calling was performed on RNA-seq data using the “mpileup” function from “SAMtools” (Li et al., 2009) using default parameter values. This data set included 40 unrelated individuals sampled from across the Åland islands. Only SNPs which had a minor allele frequency (MAF)  $> 0.2$ , call rate  $> 0.9$  and SNP quality score  $> 100$  were retained. Neutral SNPs were selected using SOLiD matepair-1 genome sequences (Ahola et al., 2014) and called using an in-house method (Kvist et al., 2015; Rastas et al., 2013). Only heterozygous SNPs spanning all 31 chromosomes from noncoding regions were selected, as the genomic sequences originated from a single male individual. More detailed information on SNP validation and genotyping is reported in (Fountain et al., 2016). Any SNP or individual with a call rate of  $< 0.95$  was removed from further analyses.

### 2.3 | Reconstruction of families

We used COLONY (version 2.0.58 available from <https://www.zsl.org/science/software/colony>) software (Wang, 2004; Wang & Santure, 2009) to infer relationships (both full- and half-sib) between pairs of sampled individuals using the marker data described above. COLONY uses a full-pedigree method for sibship inference and parentage assignment (as opposed to pairwise inferences of relatedness) whereby sampled offspring are assigned to hypothetical maternal and paternal families (Wang, 2004; Wang & Santure, 2009). To find a good partition, COLONY uses simulating annealing for optimization, scoring each putative partition using a likelihood-based score (Wang, 2004; Wang & Santure, 2009). Although multiple mating is rare (approximately 6%–8% of the females mate more than once in the wild (Boggs & Nieminen, 2004)), we allowed this in COLONY by assuming polygamous mating in both sexes. We also allowed for inbreeding to occur (Saccheri et al., 1998) and a genotyping error rate of 1.1% per allele (S.C. Wong, pers. comm).

The analyses were carried out separately for each of the six years and the three areas, resulting in 18 different runs. Allele frequencies

were estimated on a per year and population basis. Uncertainty in family assignments as a result of the simulated annealing method was reduced by repeating each analysis five times for every data set (i.e., year and population) and retaining only those family relationships that were present in the results from all runs.

## 2.4 | Validating the performance of COLONY using simulated data

The accuracy of sibship reconstruction depends on factors such as sample size, number of sampled individuals for each family and the number and variability of the genetic markers. COLONY assumes that individuals in a sample are taken from a large, randomly mating population (Jones & Wang, 2010). Although our data sets violate this assumption, the accuracy of sibship assignments is not highly sensitive to population genetic structure (Wang & Santure, 2009). To assess the quality of the inferred sibships, we generated simulated data using a forward individual-based model constructed and parameterized for the Glanville fritillary metapopulation in the Åland Islands (see Appendix S1 for the description of the simulation model). We used single- and two-patch models to capture the general population genetic structure in the study areas and test the sensitivity of the COLONY predictions to the level of population genetic structure seen in the Glanville fritillary metapopulation. Although the abundance of larvae widely varied across years, within regions, as well as also across regions (Table S1), for the purpose of this analyses we set the population size near 500 (close to the mean sample size of 462), as computation in COLONY is demanding for large populations. We included all individuals from one generation in both one-patch and two-patch models (492 and 1,014, respectively), as the simulation model effectively randomly sampled at most a handful of offspring from each mother by retaining only offspring that were to survive to adulthood. We finally checked that the simulated data resembled the empirical data in key aspects such as family size distribution, minimum allele frequency, and the number and variation of marker loci. COLONY was then run on these data sets using the same settings as above. The accuracy of inferred sibships from COLONY could then be compared to the simulated data to assess error rates.

COLONY also inferred half-sib families, but these were not consistent across the five replicate runs, with some runs having as many as 48% additional predicted half-sib pairs compared to the consensus results. This result was verified with the simulated data, which demonstrated that 47% of the inferred half-sibs by COLONY were incorrect in the single population data and 44% in the two-patch data set. Therefore, in the remaining analyses, we focused on the full-sib families only.

## 2.5 | Dispersal distances

The yearly spatial scale of dispersal was estimated by fitting dispersal distance distributions to the distances of reconstructed dispersal events from the region of Saltvik ( $n = 230$ ) assuming either a linear model movement (lm) or a diffusion approximation of a random walk

movement model (rw) with constant settling rates (Nathan, Klein, Robledo-Arnuncio, & Revilla, 2012; Turchin, 1998). The parameters were estimated using the probabilistic modelling language Stan (Stan Development Team 2016). The annual parameters determining the scale of dispersal were modelled as lognormally distributed with Half-Cauchy priors of mean 0 and scale 1 on the mean and standard deviation of the lognormal distribution for both models. The probability density functions and the associated mean dispersal distances are given in equations (X1-X4).

$$f_{lm}(r) = \frac{1}{2\pi a^2} e^{-\frac{r}{a}}$$

$$f_{rw}(r) = \frac{1}{2\pi a} K_0 \left( \sqrt{\frac{1}{a}} r \right)$$

$$E_{lm}(r) = 2a$$

$$E_{rw}(r) = \frac{\pi}{2} \sqrt{a}$$

Connectivity of individual patches was calculated as

$$S_i = \sum_{j \neq i} A_j^{0.44} \frac{1}{2\pi a^2} e^{-\frac{d_{ij}}{a}} N_j$$

where  $A$  is the area of the target patch in hectares,  $2a$  is the mean dispersal distance,  $d$  is the distance between patches in kilometres, and  $N$  is the number of winter nests in the source patch.

## 2.6 | Spatial genetic structure

As a comparison with the direct estimates of dispersal distances from full-sib reconstruction, we additionally quantified the spatial scale of genetic structure by testing for associations between interindividual relatedness and geographical distance. The rate of decline of interindividual relatedness with distance is expected to reflect the scale of dispersal (Epperson, 2005; Hardy & Vekemans, 1999). We used SPAGeDi 1.5 (Hardy & Vekemans, 2002) to calculate pairwise kinship among individuals (Loiselle, Sork, Nason, & Graham, 1995) in the primary study region of Saltvik for each year. Mean kinship among individuals was calculated within patches, and among individuals within 12 distance classes that were predefined in one-kilometre intervals up to a maximum of 12 km. We used jackknifing over all loci to obtain mean estimates and standard errors of kinship coefficients, and tested the significance of correlations between kinship and geographical distance by permuting spatial locations of individuals 999 times. We randomly subsampled a single individual per larval triplet to attempt to control for family effects. Due to large sample size in 2012, we down sampled the analysis in this year to 120 randomly chosen patches to speed up computations. We repeated this analysis using all patches for all years with only putatively neutral SNPs and found the same pattern.

We additionally calculated  $F_{ST}$  for each region and year to assess overall genetic differentiation. Only a single individual per

larval triplet was included per patch, and patches with fewer than four families were excluded to prevent bias in estimates due to low sample size (Willing, Dreyer, & van Oosterhout, 2012). Calculations were made using the *adegenet* package in R (Jombart, 2008; R Core Team 2016).

### 3 | RESULTS

#### 3.1 | Population sizes and structure of empirical data

The number of larval groups in Åland is highly dynamic with large yearly differences in each of the three sampled regions. All three regions saw an initial decline in the number of larval groups, followed by a recovery towards the end of the sampled period (Table S1, Fig. S2). However, the recovery of Sottunga was far less pronounced than that of Föglö or Saltvik (in 2012, there were 44% of the number of larval groups collected in 2007, compared to 87% of larval groups in Föglö and 420% in Saltvik). By contrast, Saltvik had a rapid expansion in the number of larval groups between 2010 and 2012.

#### 3.2 | Full-sib family structure across patch networks

On average, over 71% of inferred full-sibs originated from a single larval group, but this varied between populations and years, ranging from 56% in Föglö in 2009 to 90% in Sottunga in 2011 (Table 1).

Out of the larval groups containing at least two individuals (80% of all larval groups, as some sampled individuals either had unsuccessful DNA extractions or were excluded in the SNP filtering process), 76% consisted of full-sibs only, while the remaining larval groups included less related individuals. Table 1 shows the distribution of full-sib families among three categories: those in which all members were located in a single larval group, those in which full-sibs occurred in two or more larval groups but within a single population, and those families in which family members were found in two or more different populations, implying that the female had dispersed between the populations. Temporal differences in the percentage of full-sib families found in different populations were much greater than spatial differences (Table 1). The percentage of full-sib families being found in multiple patches ranged from a minimum of 2% in 2010 to a maximum of 13% in 2009. By contrast, the averages for each of the regions were 5% for Sottunga, 7% for Föglö and 9% for Saltvik.

#### 3.3 | Full-sib family inference consistency and verification using simulations

Predicted full-sib family sizes varied between 1 and 20 (Fig. S3). The five replicate COLONY runs for full-sib inference gave highly consistent results for 17 of the 18 data sets (6 years times 3 regions), with less than 3.3% additional full-sib pairs inferred in a replicate run when compared to the consensus result obtained by accepting only

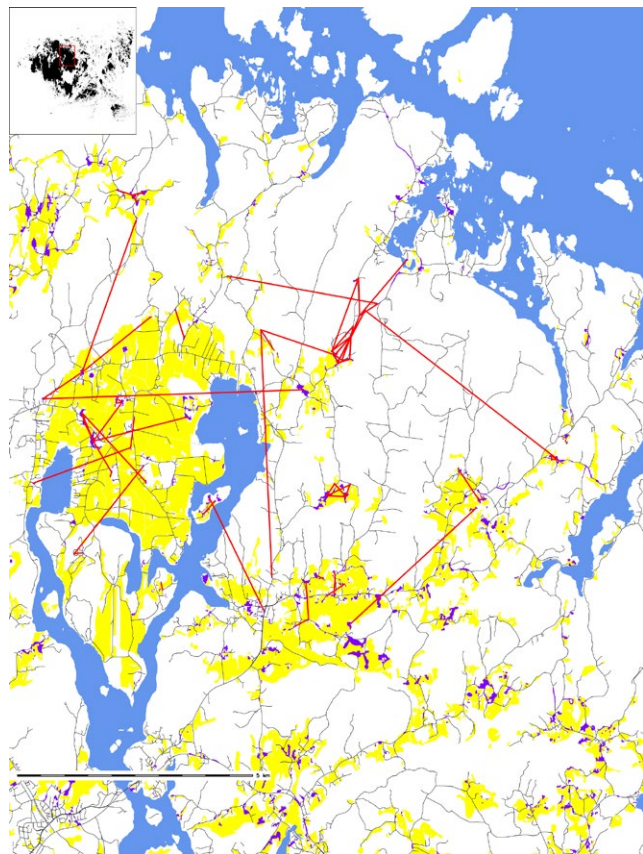
**TABLE 1** The number of larval groups containing two or more individuals (i.e., those included for the subsequent calculations), the predicted number of full-sib families and the proportion of inferred full-sib families on a single plant, within a patch and in multiple patches for each region and year

Region	Year	No of larval groups ( $\geq 2$ ind)	Predicted full-sib families	% larval groups with $\geq 2$ FS	% in single larval group	% in single patch	% in multiple patches
Sottunga	2007	52	56	80.77%	78.57%	17.86%	3.57%
	2008	13	22	84.62%	81.82%	18.18%	0.00%
	2009	18	23	77.78%	78.26%	13.04%	8.70%
	2010	7	6	85.71%	66.67%	33.33%	0.00%
	2011	11	10	90.91%	90.00%	0.00%	10.00%
	2012	25	26	80.00%	80.77%	11.54%	7.69%
Föglö	2007	95	174	56.84%	85.06%	9.77%	5.17%
	2008	25	46	72.00%	86.96%	10.87%	2.17%
	2009	25	16	92.00%	56.25%	25.00%	18.75%
	2010	7	9	71.43%	77.78%	22.22%	0.00%
	2011	37	37	81.08%	70.27%	24.32%	5.41%
	2012	93	116	82.80%	70.69%	19.83%	9.48%
Saltvik	2007	206	254	74.76%	75.98%	18.90%	5.12%
	2008	61	99	75.41%	77.78%	13.13%	9.09%
	2009	297	270	82.49%	66.30%	22.22%	11.48%
	2010	151	181	77.48%	69.06%	25.41%	5.52%
	2011	717	644	78.80%	62.58%	25.16%	12.27%
	2012	1,094	1,113	71.85%	72.51%	19.50%	8.00%
Total		2,934	3,102	75.77%	71.24%	20.25%	8.51%

the full-sib pairs inferred in all five runs. In the case where a single population was modelled, 100% of the predicted full-sibs were correct and 98% of all true full-sib pairs were detected by COLONY. In the simulated data based on a two-patch island model with weak genetic differentiation ( $F_{ST} = 0.03$ , see Appendix S1), we also found that 100% of the predicted full-sibs were correct and 99% of all true full-sib pairs were detected, demonstrating that COLONY predictions with our marker data are robust to the population genetic structure observed in the study metapopulation.

### 3.4 | Movement patterns

The distance between larval groups belonging to the same full-sib family can be used as a measure of female dispersal distance, as the female must have dispersed between laying the two egg clutches. Figure 1 shows one example of inferred movements between populations, from Saltvik in 2012, while Figure 2 shows the distribution of movement distances. 75% of movement distances are less than 1 km, and 85% are less than 2 km. There are three very long movement distances, more than 6 km. To further validate that these were real dispersal events, we checked the relatedness among the pairs of individuals in the respective families to ascertain that they all supported full-sib relatedness. The estimated scale of mean dispersal distances



**FIGURE 1** Example of breeding dispersal movements between populations in Saltvik in 2012. Highlighted purple areas are sampled patches, with red lines indicating a reconstructed dispersal event. Blue represents water and yellow agricultural fields

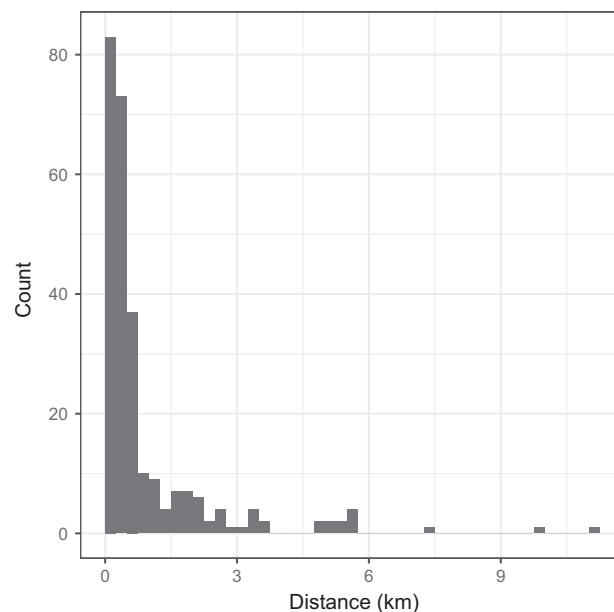
ranges from 0.66 km to 1.50 km for the negative exponential kernel and 0.74 to 1.48 km for the diffusion kernel (Table 2).

### 3.5 | Spatial genetic structure

We found significant negative relationships between kinship and distance for all years of study, and this was fairly consistent across the years (Figure 3). For 2007, 2008 and 2011, we found significant associations among individuals in distance classes up to two kilometres. In the remaining years, associations were significant up to three kilometres, although mean relatedness among individuals in the three kilometre distance class was quite low (Figure 3). Mean kinship within patches (Figure 3) and  $F_{ST}$  across patches in Saltvik (Table S2) was highest in 2007 and 2008, and lowest in 2011 and 2012. In general,  $F_{ST}$  was lower in both Sottunga and Föglö compared to Saltvik, but small sample sizes precluded calculations in most years (Table S2).

## 4 | DISCUSSION

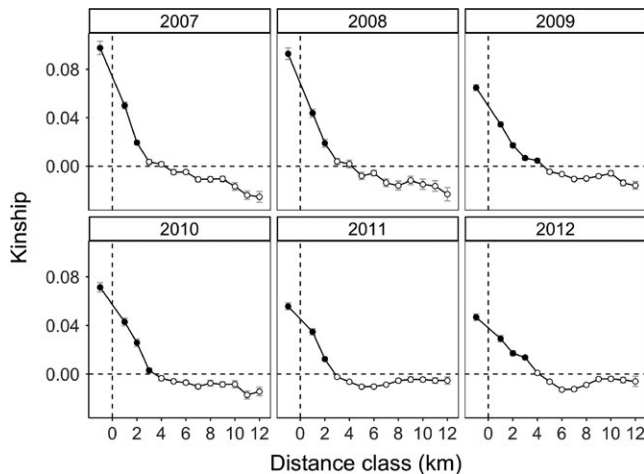
The ability to reconstruct relatedness of individuals is a powerful tool in the study of ecology and evolution in natural populations. Here we demonstrate that even in a complex dynamic system, like the Glanville fritillary butterfly metapopulation in the Åland Islands, it is possible to use a relatively small number of genetic markers to reconstruct individual relationships and provide estimates of ecological parameters that have previously been unavailable at such large spatial and temporal scales. Specifically, using a combination of empirical data and simulations we accurately reconstruct full-sib families with 272 SNP markers, and infer dispersal over all relevant spatial scales for a butterfly. We show that median dispersal



**FIGURE 2** Distribution of predicted breeding dispersal distances estimated from empirical SNP data for the Saltvik region

**TABLE 2** Observed and estimated annual scale of dispersal in the region of Saltvik

Year	Observations		Linear movement/negative exponential		Diffusion model for random walk with settling	
	<i>n</i>	Mean dispersal distance (km)	Mean dispersal distance (km)	<i>SD</i>	Mean dispersal distance (km)	<i>SD</i>
2007	13	1.02	1.10	0.22	1.11	0.24
2008	9	0.57	0.66	0.17	0.74	0.21
2009	31	0.89	0.92	0.12	0.94	0.14
2010	10	1.39	1.50	0.33	1.48	0.36
2011	78	1.26	1.27	0.10	1.26	0.12
2012	89	0.81	0.82	0.06	0.82	0.07

**FIGURE 3** Mean kinship and associated standard errors among individuals in Saltvik across distance classes for each year of study. The first distance class ( $x = -1$ ) shows mean kinship coefficients for within patch comparisons. Filled points above zero indicate individuals were significantly more similar than would be expected due to random sampling of the population

distance is 406 m consistent with previous work using tracking in the field with harmonic radar (Ovaskainen et al., 2008). Additionally, we also, for the first time in this species, shed light on the reproductive performance of females in the wild, providing estimates of how often females successfully oviposit clutches on one or multiple habitat patches during their lives. Both these variables are difficult or even impossible to observe in wild insect populations, yet can have wide implications on the persistence of the (meta)population across time.

#### 4.1 | Evaluation of relatedness estimates

As the performance of relatedness estimators depend on many factors, testing them with data generated from simulations with realistic assumptions and a known pedigree can increase the confidence level of the predictions. In addition, the flexibility of simulations allows researchers to test hypotheses that are difficult to do empirically. Despite this advantage, we are aware of only a limited

number of studies that analysed empirical sibship data in combination with simulated data (Charman, Sears, Green, & Bourke, 2010; Lepais et al., 2010). Therefore, to assess the reliability of COLONY's predictions for our SNP data sets, we ran COLONY on simulated SNP data sets with known genealogical relationships among individuals. We were primarily interested in knowing whether the number of SNPs (272) was sufficient, given the genetic architecture (i.e., linkage), minimum allele frequency, the distribution of family sizes and the mating system of the species. Although modest numbers of SNP markers (60–120) are reported to be sufficient for accurate reconstruction of sibships (Ackerman et al., 2017; Wang & Santure, 2009), the accuracy can depend on the complexity of genetic structure and other intrinsic population characteristics in a particular data set (Kopps, Kang, Sherwin, & Palsbøll, 2015). Furthermore, our study species violates a few assumptions in COLONY analyses (e.g., no genetic linkage among loci and random mating), although the algorithm is known to be robust to nonrandom mating and linkage of markers (Wang & Santure, 2009).

For the relatively small and well-connected networks of Sottunga and Föglo, a single-patch model was deemed adequate, conforming better to COLONY's assumptions, as moderate gene flow among subpopulations is expected in these networks. For the large network of Saltvik, some spatial genetic structure could be expected especially in years with low abundance (Nair, Fountain, Ikonen, Ojanen, & van Nouhuys, 2016; Orsini, Corander, Alasentie, & Hanski, 2008); hence, we also included a two-patch model with a typical level of genetic differentiation among the subpopulations (semi-independent networks, SInS) ( $F_{ST} = 0.03$ , Nair et al., 2016). The results suggest COLONY predictions are not sensitive to the observed level of population genetic differentiation for detecting full- or half-sibs, although the software performed poorly for half-sibs in both the single and two-patch models.

In these analyses, we included all the simulated individuals from one generation for simplicity. We were not concerned with the effects of sampling because sample sizes were relatively large (ranges from 21 to 3,200 larvae) and because three individuals were randomly sampled exhaustively from all the nests. COLONY uses allele frequencies estimated from the data to compute the likelihood, and therefore, more accurate estimates of allele frequencies from larger samples usually

lead to more accurate sibship predictions (Wang, 2012). We also did not simulate genotyping errors as they were thought to be fairly low (1.1%; S.C Wong pers. comm., but note that we allowed for genotyping errors in the empirical analyses when inferring full-sibs, see methods). These simplifications are limitations of our simulated data, but the purpose was to test the power and ability of COLONY to infer sibships using our empirical data set (i.e., with the same number of SNP loci, sample size, population structure).

The results with the simulated data indicated that the predictions from COLONY were highly accurate for full-sib pairs (100% of full-sib families were correctly predicted), while the performance was much poorer for half-sib pairs. This corroborates the results from the empirical SNP data, where predictions were highly consistent for full-sibs but inconsistent for half-sibs among separate COLONY runs. While it was unfortunate not to be able to accurately detect half-sibs, it has been shown that accurately detecting half-sibs is inherently difficult (Ackerman et al., 2017; Wang, 2012; Wang & Santure, 2009).

While it may be tempting to give a recommendation on the minimum number of SNP markers required to accurately estimate full-sib groups, previous work has shown this to be highly dependent on the focal study system. Wang (2012) showed as few as 60 SNPs could provide accurate family assignments, but estimates were sensitive to allele frequency misspecifications. For example, when a sample was dominated by a small number of large families, the large families tended to be split, due to an overestimation of their allele frequencies. Family structure, along with other ecological characteristics of the system, can therefore have significant implications on the number of markers needed for accurate sibship reconstruction. Kopps et al. (2015) further demonstrated the influence of intrinsic population characteristics on the power to accurately make kinship inferences, along with characteristics of the genetic markers used (e.g., marker type, MAF, typing error). Differences in mating system and the number of overlapping generations both led to a different minimum number of genetic markers needed for accurate relatedness classifications. Species-specific power calculations, as performed in the present study, are therefore essential to ensure accurate and reliable sibship estimation is possible when deciding the number of SNP markers to use.

## 4.2 | Inferring dispersal parameters from molecular data

In cases where larval groups were found in two different populations, it was possible to infer female effective dispersal distances, with the median dispersal distance for females being 406 m (mean: 1,001 m, range 67–11,004 m). The estimated average dispersal distance is consistent with previous estimates from mark-release-recapture (MRR) data, which range from 280 m to 544 m (Kuussaari et al., 1996; Niitepöld, Mattila, Harrison, & Hanski, 2011). Our results highlight significant annual variation in the scale of dispersal (Table 2), something not possible to detect with traditional MRR approaches, implying a large variation in the scale of colonization and extinction dynamics. Both linear movement and random walk models lead to similar

estimates of the scale of yearly movement, which is not surprising given that while biologically implausible, the linear model of movement can lead to similar distribution of dispersal distances compared to the random walk model (Hawkes, 2009; Turchin, 1998). For the first time, we were also able to estimate the frequency and scale of long-distance dispersal events in this system. Our estimate of the longest dispersal event is somewhat higher than those estimated based on MRR studies. Ovaskainen et al. (2008), however, used harmonic radar to track the movement of females butterflies in an area of 30 ha during a two hour period, and the maximum distance moved by an individual was 5,490 m. This shows that while unlikely, the very few long-distance dispersal events estimated in this study, are at least biologically plausible. This is particularly the case as we are potentially observing total dispersal across the entire breeding season with the genetic data, rather than that observed over a two hour period.

Previous estimates of female dispersal rate based on MRR data range from 8% to 40% with great variation among the studies (Hanski et al., 1994; Kuussaari et al., 1996; Niitepöld et al., 2011). Potential reasons explaining the variation in dispersal propensity include variation in weather conditions and habitat configuration among the studies. Factors such as size of a patch, butterfly density, habitat quality and even allele frequency in *Pgi* genotype have shown to influence dispersal propensity in the Glanville fritillary butterfly (Hanski et al., 1994; Kuussaari et al., 1996; Niitepöld et al., 2011). Similarly, models inferring dispersal from the long-term data on the population dynamics have further estimated that the number of habitat patches visited by a female during her life varies from one to ten (Zheng, Ovaskainen, & Hanski, 2009). Our results on the number of full-sib families found in two or more patches fall on the low side of the estimates on dispersal events, but we are only measuring successful colonization events. Our analysis cannot detect dispersal where the female lays eggs in only a single patch or whether she has moved before she laid her first clutch. However, our data show that female effective dispersal, or colonization distance (estimated here), is similar to the dispersal distance found in MRR data. It is this colonization distance that is the evolutionarily important feature of dispersal, as we are directly capturing gene flow, something not possible with observational studies.

We also used an alternative approach to our dispersal estimates based on family reconstruction, by assessing the spatial decay of kinship coefficients. By contrast to dispersal distances inferred from sibship reconstruction, we detected significant structure up to 3 km. However, this is expected because spatial genetic structure contains the signal of dispersal and mating that has occurred over many generations. It is also important to note that because kinship within each distance class is compared to mean kinship of the total sample, the measured extent of spatial genetic structure will vary depending on the sample taken. Thus, to truly estimate dispersal from patterns of spatial genetic structure one must control for effective population density, which is particularly difficult to estimate for metapopulations (Hardy et al., 2006; Vekemans & Hardy, 2004). Previous work has also demonstrated a lag time between major demographic events and genetic structure in Åland (Orsini et al., 2008), further highlighting



the added value of directly reconstructing dispersal through sibship analysis, which allowed us to track gene flow within single generations. While using family inference to infer the rate and scale of dispersal is not new and has been used to infer dispersal in plants (Dow & Ashley, 1996; Ellstrand & Marshall, 1985), mammals (Peacock & Smith, 1997; Telfer et al., 2003), fish (Jones, Planes, & Thorrold, 2005), birds (Woltmann, Sherry, & Kreiser, 2012) and other insects (Charman et al., 2010; Lepais et al., 2010), in this study we have demonstrated the ability to infer female dispersal even on the large scale of the Åland islands, where the direct observation of so many individuals would be impossible.

### 4.3 | Yearly and regional differences in estimated dispersal

We found yearly differences in the number of recorded dispersal events (i.e., full-sibs in multiple patches), with 2009 and 2011 showing the highest percentage of between-patch dispersal in both Sottunga and Saltvik, and 2009 and 2012 showing the highest percentages in Föglö (Table 1). This may be explained by the proportion of patch occupancy across years. For example, in both Sottunga and Saltvik, the years preceding the high dispersal years had the lowest proportion of occupied patches recorded across the study period (Table S1, Fig. S2). Following years where patch occupancy is low, the number of colonizations tends to be highest following the reciprocal metapopulation dynamics observed over the lifetime of the annual survey (Hanski, 2011; Ojanen et al., 2013). Negative density-dependent emigration has also been observed in *M. cinxia*, suggesting that butterflies tend to stay rather than disperse away from high-quality patches, and disperse when local density is low (Kuussaari et al., 1996). This could further explain patterns in Saltvik, where increased dispersal was observed in years following population bottlenecks (i.e., low total number of nests; Table S1, Fig. S2). However, an individual's decision to emigrate will also depend on a number of other factors such as local patch quality and patch boundary effects (Moilanen & Hanski, 1998), and weather during flight season (Kuussaari, Rytteri, Heikkinen, Heliölä, & von Bagh, 2016). Further work is required to determine the relative importance of these factors in explaining regional and yearly variation in dispersal.

### 4.4 | Estimation of other demographic parameters

From the field collected data, the vast majority of collected larval triplets included only full-sibs, while the remainder were most likely mixtures of the offspring of two or more females. This conclusion is drawn based on the following. Firstly, based on previous knowledge only 6%–8% of the females in the wild have been shown to mate more than once (Boggs & Nieminen, 2004). Secondly, of those females that have mated more than once, paternity analyses have indicated that around 80% of the clutches are still sired by only one male, in most cases the first one (Sarhan & Kokko, 2007). Finally, low levels of larval mixing have been reported from previous field studies (Kuussaari et al., 2004). Larval mixing may occur either through two or more females laying eggs on the same host plant

or by the larval groups mixing during the summer when they move from the foliated host plant to the next. In the field, approximately 10% of the host plants that contained eggs were observed to have more than one clutch (Kuussaari et al., 2004), but it is unknown whether these clutches were laid by one or more females.

In the majority (71%) of cases, full-sibs occurred in a single larval group. Previous studies, under semi-natural conditions, have shown that some females are able to lay up to ten clutches in their lifetime, with the average number of clutches laid by females being three (Saastamoinen, 2007). In 20% of the cases, the full-sibs occurred in more than two larval groups within a single population and only 9% of the full-sibs occurred in two or more larval groups in more than two local populations, indicating a successful dispersal event (i.e., establishment). Our results here indicate that of those females that successfully reproduce, most are producing only one clutch that survives over the summer. Considering this is crucial when making predictions about a populations survival and persistence.

## 5 | CONCLUSIONS

This study provides an example of the power of using relatively small numbers of genetic markers to uncover previously unknown ecological parameters. Caution should still be applied when designing such studies as the number of markers required will be highly dependent on the traits studied, the number of individuals sampled, and the population characteristics (e.g., level of population structure). This study also highlights the continuing value of combining genomic resources with long-term longitudinal field studies, without which many inferences of ecological and evolutionary dynamics in natural populations would be impossible.

### ACKNOWLEDGEMENTS

We thank Annukka Ruokolainen, Toshka Nyman and Laura Häkkinen for sample preparation and DNA extraction. This study was funded by the Academy of Finland (Grants 283108 to I.H. and 284601 to I.H. and M.S.). I.H. sadly passed away during the preparation of the manuscript.

### DATA ARCHIVING STATEMENT

Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.d461s>

### ORCID

Toby Fountain  <http://orcid.org/0000-0002-5501-4691>

Michelle F. DiLeo  <http://orcid.org/0000-0003-0101-5274>

Etsuko Nonaka  <http://orcid.org/0000-0001-6563-2840>

Marjo Saastamoinen  <http://orcid.org/0000-0001-7009-2527>

Torsti Schulz  <http://orcid.org/0000-0001-8940-9483>

## REFERENCES

- Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., ... Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: Empirical evaluations using hatchery steelhead. *Evolutionary Applications*, 10, 146–160.
- Ahola, V., Lehtonen, R., Somervuo, P., Salmela, L., Koskinen, P., Rastas, P., ... Hanski, I. (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nature Communications*, 5, 1–9.
- Berry, O., Tocher, M. D., & Sarre, S. D. (2004). Can assignment tests measure dispersal? *Molecular Ecology*, 13, 551–561.
- Boggs, C. L., & Nieminen, M. (2004). Checkerspot Reproductive Biology. In P. R. Ehrlich, & I. Hanski (Eds.), *On the Wings of Checkerspots* (pp. 92–111). Oxford, UK: Oxford University Press.
- Broquet, T., & Petit, E. J. (2009). Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology Evolution and Systematics*, 40, 193–216.
- Charman, T. G., Sears, J., Green, R. E., & Bourke, A. F. (2010). Conservation genetics, foraging distance and nest density of the scarce Great Yellow Bumblebee (*Bombus distinguendus*). *Molecular Ecology*, 19, 2661–2674.
- Clobert, J., Danchin, E., Dhondt, A. A., & Nichols, J. D. (2001). *Dispersal*. Oxford, UK: Oxford University Press.
- Coltman, D. W., Bancroft, D. R., Robertson, A., Smith, J. A., Clutton-Brock, T. H., & Pemberton, J. M. (1999). Male reproductive success in a promiscuous mammal: Behavioural estimates compared with genetic paternity. *Molecular Ecology*, 8, 1199–1209.
- Csilléry, K., Johnson, T., Beraldi, D., Clutton-Brock, T., Coltman, D., Hansson, B., ... Pemberton, J. M. (2006). Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, 173, 2091–2101.
- Dobzhansky, T., & Wright, S. (1943). Genetics of Natural Populations. X. Dispersion Rates in *Drosophila Pseudoobscura*. *Genetics*, 28, 304–340.
- Dow, B. D., & Ashley, M. V. (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Molecular Ecology*, 5, 615–627.
- Ehrlich, P. R., & Hanski, I. (2004). *On the Wings of Checkerspots*. Oxford, UK: Oxford University Press.
- Ellstrand, N. C., & Marshall, D. L. (1985). Interpopulation gene flow by pollen in wild radish, *Raphanus sativus*. *American Naturalist*, 126, 606–616.
- Epperson, B. K. (2005). Estimating dispersal from short distance spatial autocorrelation. *Heredity*, 95, 7–15.
- Firth, J. A., Hadfield, J. D., Santure, A. W., Slate, J., & Sheldon, B. C. (2015). The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution*, 69, 1336–1344.
- Fountain, T., Nieminen, M., Sirén, J., Wong, S. C., Lehtonen, R., & Hanski, I. (2016). Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences*, 113, 2678–2683.
- Gilbert, K. J., & Whitlock, M. C. (2015). Evaluating methods for estimating local effective population size with and without migration. *Evolution*, 69, 2154–2166.
- Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759–769.
- Griffith, S. C., Owens, I. P. F., & Thuman, K. A. (2002). Extra pair paternity in birds: A review of interspecific variation and adaptive function. *Molecular Ecology*, 11, 2195–2212.
- Hall, L. A., Palsbøll, P. J., Beissinger, S. R., Harvey, J. T., Berube, M., Raphael, M. G., ... Peery, M. Z. (2009). Characterizing dispersal patterns in a threatened seabird with limited genetic structure. *Molecular Ecology*, 18, 5074–5085.
- Hanski, I. (2011). Eco-evolutionary spatial dynamics in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences*, 108, 14397–14404.
- Hanski, I., Kuussaari, M., & Nieminen, M. (1994). Metapopulation structure and migration in the butterfly *Melitaea cinxia*. *Ecology*, 75, 747–762.
- Hardy, O. J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M. H., ... Troisoux, V. (2006). Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, 15, 559–571.
- Hardy, O. J., & Vekemans, X. (1999). Isolation by distance in a continuous population: Reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, 83, 145–154.
- Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Resources*, 2, 618–620.
- Hawkes, C. (2009). Linking movement behaviour, dispersal and population processes: Is individual variation a key? *Journal of Animal Ecology*, 78, 894–906.
- Huisman, J., Kruuk, L. E., Ellis, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, 113, 3585–3590.
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Jones, A. G., & Ardren, W. R. (2003). Methods of parentage analysis in natural populations. *Molecular Ecology*, 12, 2511–2523.
- Jones, G. P., Planes, S., & Thorrold, S. R. (2005). Coral reef fish larvae settle close to home. *Current Biology*, 15, 1314–1318.
- Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10, 551–555.
- de Jong, M. A., Wong, S. C., Lehtonen, R., & Hanski, I. (2014). Cytochrome P450 gene CYP337 and heritability of fitness traits in the Glanville fritillary butterfly. *Molecular Ecology*, 23, 1994–2005.
- Kopps, A. M., Kang, J., Sherwin, W. B., & Palsbøll, P. J. (2015). How well do molecular and pedigree relatedness correspond, in populations with variable mating systems, and types and quantities of molecular and demographic data?. *G3: Genes, Genomics, Genetics*, 5, 1815–1826.
- Kuussaari, M., Nieminen, M., & Hanski, I. (1996). An experimental study of migration in the Glanville fritillary butterfly *Melitaea cinxia*. *Journal of Animal Ecology*, 65, 791–801.
- Kuussaari, M., Rytteri, S., Heikkinen, R. K., Heliölä, J., & von Bagh, P. (2016). Weather explains high annual variation in butterfly dispersal. *Proceedings of the Royal Society of London B: Biological Sciences*, 283, 20160413.
- Kuussaari, M., van Nouhuys, S., Hellmann, J. J., & Singer, M. C. (2004). Larval biology of checkerspots. In P. R. Ehrlich, & I. Hanski (Eds.), *On the Wings of Checkerspots* (pp. 138–160). Oxford, UK: Oxford University Press.
- Kvist, J., Mattila, A. L., Somervuo, P., Ahola, V., Koskinen, P., Paulin, L., ... Hanski, I. (2015). Flight-induced changes in gene expression in the Glanville fritillary butterfly. *Molecular Ecology*, 24, 4886–4900.
- Lepais, O., Darvill, B., O'Connor, S., Osborne, J. L., Sanderson, R. A., Cussans, J., ... Goulson, D. (2010). Estimation of bumblebee queen dispersal distances using sibship reconstruction method. *Molecular Ecology*, 19, 819–831.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, 82, 1420–1425.
- Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, 19, 3038–3051.
- Moilanen, A., & Hanski, I. (1998). Metapopulation Dynamics: Effects Of Habitat Quality And Landscape Structure. *Ecology*, 79, 2503–2515.
- Nair, A., Fountain, T., Ikonen, S., Ojanen, S. P., & van Nouhuys, S. (2016). Spatial and temporal genetic structure at the fourth trophic level in

- a fragmented landscape. *Proceedings of the Royal Society of London B: Biological Sciences*, 283, 20160668.
- Nathan, R., Klein, E. K., Robledo-Arnuncio, J. J., & Revilla, E. (2012). Dispersal kernels. In J. Clobert, M. Baguette, T. G. Benton, & J. M. Bullock (Eds.), *Dispersal Ecology and Evolution* (pp. 187–210). Oxford, UK: Oxford University Press.
- Niitepöld, K., Mattila, A. L., Harrison, P. J., & Hanski, I. (2011). Flight metabolic rate has contrasting effects on dispersal in the two sexes of the Glanville fritillary butterfly. *Oecologia*, 165, 847–854.
- van Nouhuys, S., & Hanski, I. (2002). Colonization rates and distances of a host butterfly and two specific parasitoids in a fragmented landscape. *Journal of Animal Ecology*, 71, 639–650.
- Ojanen, S. P., Nieminen, M., Meyke, E., Pöyry, J., & Hanski, I. (2013). Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): Survey methods, data management, and long-term population trends. *Ecology and Evolution*, 3, 3713–3737.
- Orsini, L., Corander, J., Alasentie, A., & Hanski, I. (2008). Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology*, 17, 2629–2642.
- Ovaskainen, O., Smith, A. D., Osborne, J. L., Reynolds, D. R., Carreck, N. L., Martin, A. P., ... Hanski, I. (2008). Tracking butterfly movements with harmonic radar reveals an effect of population age on movement distance. *Proceedings of the National Academy of Sciences*, 105, 19090–19095.
- Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. *Molecular Ecology*, 13, 55–65.
- Palstra, F. P., & Fraser, D. J. (2012). Effective/census population size ratio estimation: A compendium and appraisal. *Ecology and Evolution*, 2, 2357–2365.
- Peacock, M. M., & Smith, A. T. (1997). The effect of habitat fragmentation on dispersal patterns, mating behavior, and genetic variation in a pika (*Ochotona princeps*) metapopulation. *Oecologia*, 112, 524–533.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: Austria.
- Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences*, 94, 9197–9201.
- Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., & Auvinen, P. (2013). LepMAP: Fast and accurate linkage map construction for large SNP datasets. *Bioinformatics*, 29, 3128–3134.
- Saastamoinen, M. (2007). Life-history, genotypic, and environmental correlates of clutch size in the Glanville fritillary butterfly. *Ecological Entomology*, 32, 235–242.
- Saastamoinen, M., Ikonen, S., Wong, S. C., Lehtonen, R., & Hanski, I. (2013). Plastic larval development in a butterfly has complex environmental and genetic causes and consequences for population dynamics. *Journal of Animal Ecology*, 82, 529–539.
- Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., & Hanski, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392, 491–494.
- Sarhan, A., & Kokko, H. (2007). Multiple mating in the Glanville fritillary butterfly: A case of within-generation bet hedging? *Evolution*, 61, 606–616.
- Schunter, C., Pascual, M., Garza, J. C., Raventos, N., & Macpherson, E. (2014). Kinship analyses identify fish dispersal events on a temperate coastline. *Proceedings of the Royal Society of London B: Biological Sciences*, 281, 20140556.
- Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, 39, 53–65.
- Somervuo, P., Kvist, J., Ikonen, S., Auvinen, P., Paulin, L., Koskinen, P., ... Hanski, I. (2014). Transcriptome analysis reveals signature of adaptation to landscape fragmentation. *PLoS ONE*, 9, e101467.
- Stan Development Team. (2016). Stan: A C++ library for probability and sampling. <http://mc-stan.org>.
- Telfer, S., Piertney, S. B., Dallas, J. F., Stewart, W. A., Marshall, F., Gow, J. L., & Lambin, X. (2003). Parentage assignment detects frequent and large-scale dispersal in water voles. *Molecular Ecology*, 12, 1939–1949.
- Tøttrup, A. P., Klaassen, R. H., Strandberg, R., Thorup, K., Kristensen, M. W., Jørgensen, P. S., ... Alerstam, T. (2012). The annual cycle of a trans-equatorial Eurasian-African passerine migrant: Different spatio-temporal strategies for autumn and spring migration. *Proceedings of the Royal Society of London B: Biological Sciences*, 279, 1008–1016.
- Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*, Vol. 1. Sunderland: Sinauer Associates.
- Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, 13, 921–935.
- Wang, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics*, 166, 1963–1979.
- Wang, J. (2007). Parentage and sibship exclusions: Higher statistical power with more family members. *Heredity*, 99, 205–217.
- Wang, J. (2012). Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics*, 191, 183–194.
- Wang, J., & Santure, A. W. (2009). Parentage and Sibship Inference From Multilocus Genotype Data Under Polygamy. *Genetics*, 181, 1579–1594.
- Waser, P. M., & Strobeck, C. (1998). Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, 13, 43–44.
- Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration:  $F_{ST} \neq 1/(4Nm + 1)$ . *Heredity*, 82, 117–125.
- Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by  $F_{ST}$  do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, 7, e42649.
- Woltmann, S., Sherry, T. W., & Kreiser, B. R. (2012). A genetic approach to estimating natal dispersal distances and self-recruitment in resident rainforest birds. *Journal of Avian Biology*, 43, 33–42.
- Zheng, C., Ovaskainen, O., & Hanski, I. (2009). Modelling single nucleotide effects in phosphoglucose isomerase on dispersal in the Glanville fritillary butterfly: Coupling of ecological and evolutionary dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364, 1519–1532.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Fountain T, Husby A, Nonaka E, et al. Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly. *Evol Appl*. 2018;11:287–297. <https://doi.org/10.1111/eva.12552>