

ORIGINAL ARTICLE

Inferring contemporary dispersal processes in plant metapopulations: comparison of direct and indirect estimates of dispersal for the annual species *Crepis sancta*

A Dornier and P-O Cheptou

Analyzing population dynamics in changing habitats is a prerequisite for population dynamics forecasting. The recent development of metapopulation modeling allows the estimation of dispersal kernels based on the colonization pattern but the accuracy of these estimates compared with direct estimates of the seed dispersal kernel has rarely been assessed. In this study, we used recent genetic methods based on parentage analysis (spatially explicit mating models) to estimate seed and pollen dispersal kernels as well as seed and pollen immigration in fragmented urban populations of the plant species *Crepis sancta* with contrasting patch dynamics. Using two independent networks, we documented substantial seed immigration and a highly restricted dispersal kernel. Moreover, immigration heterogeneity among networks was consistent with previously reported metapopulation dynamics, showing that colonization was mainly due to external colonization in the first network (propagule rain) and local colonization in the second network. We concluded that the differences in urban patch dynamics are mainly due to seed immigration heterogeneity, highlighting the importance of external population source in the spatio-temporal dynamics of plants in a fragmented landscape. The results show that indirect and direct methods were qualitatively consistent, providing a proper interpretation of indirect estimates. This study provides attempts to link genetic and demographic methods and show that patch occupancy models may provide simple methods for analyzing population dynamics in heterogeneous landscapes in the context of global change.

Heredity (2013) **111**, 1–7; doi:10.1038/hdy.2013.3; published online 27 February 2013

Keywords: colonization; seed dispersal; parentage analysis; gene flow; SEMM; metapopulation

INTRODUCTION

Since 1970s, plant population dynamics modeling has become a major area of research in plant biology (Jeltsch *et al.*, 2008). While first investigations were based on the paradigm of stable populations (Harper, 1977), the rise of spatial population ecology (MacArthur and Wilson, 1967; Gadgil, 1971) has laid the foundations for analyzing plant dynamics in heterogeneous environments better adapted to plant dynamics in natural habitats. Taking habitat heterogeneity into account is also central to forecast population dynamics in the context of global change. In line with this, the metapopulation concept considers that species are maintained at the regional scale through a balance between colonization and extinction of local populations (Levins, 1969). Over the last three decades, metapopulation dynamics has been influential in spatially structured habitats with major influences on the demographic, genetic and evolutionary processes of species (Hanski and Gaggiotti, 2004). In this context, the Levins model provides the conditions required for the maintenance of a viable metapopulation and the occupancy rate at equilibrium as a function of colonization and extinction (Levins, 1969). Analyzing metapopulation dynamics in natural populations requires studying patterns of extinct and patterns of colonization. Historically, genetic methods using population structure have been studied to infer colonization. This method allowed contrasting colonization patterns

(migrant pool vs propagule pool; Slatkin, 1977) to be sorted out, thanks to the link between the fixation index (FST) and the age of populations (see Giles and Goudet, 1997 for an empirical study). Such methods considers migration-drift equilibrium and are restricted to simple colonization scenarios. More recently, the development of stochastic patch occupancy models (hereafter SPOMs) coupled with an appropriate statistical framework provides powerful and flexible tools to estimate colonization from simple patch dynamic data and the geographical localization of suitable habitat (Etienne *et al.*, 2004). These models allow joint estimation of colonization and extinction from simple presence/absence of data. In the spatially realistic Levins model (Hanski and Ovaskainen, 2000), Patch colonization hinges on patch connectivity, that is, the contribution of all occupied patches weighted by the dispersal kernel, supposedly decreasing with distance (Moilanen and Nieminen, 2002). Some studies, mainly in animals, have reported that colonization is constrained by limited dispersal (Hanski, 1994; see Verheyen *et al.*, 2004 in plants) and have confirmed that SPOMs provide powerful tools to infer the dispersal of kernel in spatially structured habitats.

However, phenomenological dispersal kernels estimated in metapopulation models is based on occupancy patterns at equilibrium and does not necessarily represent the mechanistic dispersal of propagules inferred from direct measurements. Firstly, the inference of dispersal

kernels in SPOMs requires that patches be not too isolated to allow the possibility of stable balance between colonization and extinction nor too connected to prevent full metapopulation occupancy. Secondly, dispersal kernels in SPOMs can depart from dispersal processes when massive colonization from large source populations outside the focal metapopulation occurs (Bacles *et al.*, 2006). Inferring dispersal in such a situation would not allow the detection of limited dispersal even if limited dispersal occurs, a situation referred to as the 'propagule rain model' (Gotelli, 1991; Dornier *et al.*, 2011). Thirdly, colonization from dormant individuals generates false colonization or pseudo-colonization, which potentially biases dispersal estimates. The problem of pseudo-colonization has been considered a major limitation in the application of the metapopulation approach to plants (Freckleton and Watkinson, 2002).

Finally, the statistical methods used to estimate SPOM parameters require the metapopulation to be at equilibrium, and deviation from this assumption may generate bias in parameter estimates (Moilanen, 2000). This suggests that indirect colonization estimated from metapopulation models does not necessarily match the dispersal process. To what extent colonization estimated in SPOMs are robust estimates of dispersal processes or not remains an open question in plant ecology. Confronting SPOMs estimates to direct dispersal process in a metapopulation should help to clarify this point.

Direct methods of estimating dispersal allow the dispersal kernel to be estimated by relaxing the metapopulation assumptions (for example, equilibrium) mentioned above. In animals, the capture-mark-recapture framework allows dispersal to be estimated in a metapopulation context (Bennetts *et al.*, 2001). This approach involves marking many individuals, which means that the technique is often restricted to specific animal species (for example, birds or mammals). In plants, seed traps located at various distances from an identified source provide direct seed dispersal estimates (Bullock *et al.*, 2006) but this method is restricted to a single or a few sources only. Moreover, effective colonization encompasses various processes, from seed production to seed germination and seedling survival (Nathan and Muller-Landau, 2000), which cannot be captured by seed movement only. In this respect, parentage analysis based on highly polymorphic genetic markers (for example, microsatellites) provides a powerful method for direct dispersal estimates. Such analysis allows the joint estimate of seed and pollen dispersal kernels when the geographical position of potential parents and offspring is recorded (for example, Bacles *et al.*, 2006; see also Ashley, (2010) for a review). While previous methods performed parentage analysis and dispersal kernel estimates sequentially, recent methods allow joint parentage analysis and dispersal kernel estimation, which significantly improves statistical performance (Hadfield *et al.*, 2006). The spatially explicit mating model (hereafter, SEMM) allows seed and pollen immigration rates from an unknown source to be estimated within a unique statistical framework and seed and pollen kernel dispersal (including self-fertilization) from viable established seedlings and potential parents to be fitted (Burczyk *et al.*, 2006; Oddou-Muratorio and Klein, 2008). This method based on viable seedlings provides effective dispersal estimates that include all processes acting between dispersal and recruitment. The method requires substantial sampling effort of potential parents coupled with an elevated recapture rate of seedlings/offspring. To our knowledge, it has mainly been used for trees (Burczyk *et al.*, 2006; Oddou-Muratorio and Klein, 2008). Yet, inferring metapopulation processes requires rapid turnover of populations within a few years, which can be found in short-lived species. However, large population sizes (that is, high number of individual

genotyping) in such species is a major methodological limit to parentage analysis.

In this study, we used SEMM to estimate dispersal for the annual plant *Crepis sancta* in urban patches distributed regularly at the feet of ornamental trees in an urban environment. In this highly fragmented habitat, the species *Crepis sancta* (Asteraceae) survives, thanks to a colonization-extinction balance, in patches in which the local population rarely exceeds 40 individuals. The small population sizes allowed us to exhaustively sample adult individuals in the metapopulation over two generations. The study metapopulation comprised two set of several hundred, for which contrasting metapopulation dynamics have been documented (Dornier *et al.*, 2011). More specifically, we used SEMM (1) to estimate pollen and seed dispersal kernels in plant metapopulation (2) to estimate gene flow external to the study metapopulation systems (3) to compare direct dispersal kernel estimates (SEMM) to indirect colonization kernel estimates (SPOM) in the two contrasting metapopulation dynamics. The results demonstrate that direct methods are partly consistent with SPOM inference, and we discuss the discrepancy between the two approaches.

MATERIALS AND METHODS

The urban metapopulation system

The metapopulation system is composed of a set of favorable habitats (hereafter called 'patch') of equal area (1–2 m²) around ornamental trees along a street pavement in a matrix of unsuitable habitat (concrete matrix) in Montpellier city (southern France). Due to the lack of management by city authorities, these patches harbor diverse plant communities of about 100 wild species (frequent species include *Poa annua* L., *Stellaria media* L., *Sonchus oleraceus* L., *Senecio vulgaris* L., *Conyza spp.* and *Crepis sancta* L.). The annual plant *Crepis sancta* (Asteraceae) forms small transient populations in these patches (rarely more than 40 individuals), contrasting with the large populations found in the countryside around Montpellier. *C. sancta* is an outcrossing species pollinated mainly by bees, and reproduction begins in the early spring (beginning of March) and lasts about 2 months. Fruits (achenes) are dimorphic; achenes with a pappus (at the center of the capitulum) that are able to disperse and achenes without a pappus (at the periphery of the capitulum), which drop at the feet of plants (Imbert, 1999; Cheptou *et al.*, 2008). Germination occurs with autumn rainfall and the seedlings grow during winter as a rosette. Imbert, (1999) demonstrated that neither of the seed types exhibit dormancy in controlled conditions, and (Dornier *et al.*, 2011) found no evidence of a seed bank in *C. sancta*.

Patch networks

We chose two independent networks separated by 4 km within Montpellier, in which contrasting metapopulation dynamics have previously been documented. The HM network (Henri Marès street) is a one-dimensional stretch of 96 patches of about 1–2 m² spaced 5–7 m apart at the periphery of the city (03°51'21"E, 43°37'14"N, 60 m a.s.l.). The JC network (Jacques Cartier street) is a four-column matrix of 81 patches of small area (1 m²) spaced 5 or 10 m apart (within or between different lines), which occupies the city center (03°53'35"E, 43°36'25"N, 60 m a.s.l.). Within each network, all patches were numbered and GPS-located. In a previous analysis using SPOM methodology, Dornier *et al.*, (2011) found that patch dynamics was governed mainly by external processes (propagule rain model) in HM whereas a Levins-type metapopulation model (with rescue effect) with limited colonization distance was reported in the JC network. In the latter network, the negative exponential kernel of the Levins model documented an average distance of colonization of about 2 m. On the basis of these contrasting scenarios, we hypothesized that gene flow with external sources was higher in HM than in JC.

Genetic analysis

A fraction of leaf tissue was sampled at the end of the life cycle after seed dispersal (on dead individuals) over 2 consecutive years from all individuals in

both networks for genetic analyses (see Supplementary Figure A1). In 2006, we sampled 725 individuals in HM and 103 in JC. In 2007, 271 and 180 individuals were sampled in the HM and JC networks, respectively. We extracted DNA using the DNA 96 Plant Kit (Qiagen, Hilden, Germany) and genotyped all individuals for seven microsatellites developed for this species (Dubois *et al.*, 2007). We scored genotypes using the sequencer ABI PRISM 3100 Genetic Analyzer and Genemapper software 4.0 (Applied Biosystem, Waltham, MA, USA) for sizing alleles and checked the accuracy of genotyping by hand.

Spatially explicit mating model

In order to analyze our data, we used a slightly modified model developed by Burczyk *et al.* (2006). In our analysis all individuals within a network can be encapsulated in a unique neighborhood (Oddou-Muratorio and Klein, 2008). The neighborhood represents a hypothetical circle area, in which the focal individual is at the center and individuals outside this area are considered as emigrants belonging to the background population. In the infinite neighborhood size used here, all individuals belonging to the studied metapopulation are considered inside the neighborhood whereas all emigrants and unsampled/missed individuals are considered outside the neighborhood (background population). We consider that each juvenile is mothered either by an external mother due to seed immigration with the probability m_s or by a local female with the probability $(1 - m_s)$. When the mother is local, we consider three sources for pollen: self-fertilization occurs with the probability s , or the pollen belongs to an external father with the probability m_p or to a local father with the complementary probability $(1 - s - m_p)$. Thus, the probability of observing a seedling i with the genotype G is:

$$P(G_i) = m_s P(G_i | B_s) + (1 - m_s) \sum_j \Psi_{ij} [s P(G_i | M_{ij}, M_{ij}) + m_p P(G_i | M_{ij}, B_p) + (1 - s - m_p) \sum_k \Phi_{jk} P(G_i | M_{ij}, F_{ijk})],$$

Where M_{ij} and F_{ijk} are the genotypes of the j th putative mother and the k th putative father (crossed with the j th putative mother) of the offspring i , respectively. $P(G_i | B_s)$ represents the transition probability that an offspring i comes from an external mother belonging to the background population. $P(G_i | M_{ij}, M_{ij})$, $P(G_i | M_{ij}, B_p)$ and $P(G_i | M_{ij}, F_{ijk})$ represent the transient probability that an offspring i with a M_j mother will receive pollen from (i) an event of self-pollination, (ii) an external father belonging to the background population or (iii) the k th father of the population. B_s and B_p represent the genotypic composition of the background population and were considered as equal (see Burczyk and Chybicki, 2004). Ψ_{ij} and Φ_{jk} represent the relative reproductive success of the j th mother and the k th father, respectively, in which we included the seed and pollen dispersal kernel.

Dispersal kernel

We estimated the seed dispersal kernel by considering the effect of the distance between the mother and offspring with Ψ_{ij} and the pollen dispersal kernel with Φ_{jk} which includes the effect of the distance between the potential father and potential mother. We estimated pollen dispersal and seed dispersal kernels jointly. Thus, if r_{ij} represents the distance between offspring i and its mother j , $\Psi_{ij} = f(r_{ij}) / \sum_{i=1}^{s_h} f(r_{ih})$ with $f(r_{--})$ representing the seed dispersal kernel (see below for details) and s_h the number of mothers. Similarly, $\Phi_{jk} = g(r_{jk}) / \sum_{i=1}^{z_i} g(r_{ji})$ where r_{jk} is the distance between the k th father and the j th mother, $g(r_{--})$ is the pollen dispersal kernel and z_i is the number of fathers of the metapopulation.

The dispersal kernel describes the probability that a seed (or pollen) will arrive at a distance r from the source (Austerlitz and Smouse, 2001). We chose the family of exponential power functions and seed or pollen dispersal kernel, written as:

$$f(a, b, r) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp - \left((r/a)^b \right)$$

where Γ represents the gamma function and b is the shape parameter affecting the tail of the dispersal distribution (Austerlitz and Smouse, 2001; Oddou-

Muratorio and Klein, 2008). When $b = 1$, the exponential power function converges toward the negative exponential function whereas curves with $b < 1$ are classified as fat-tailed and those with $b > 1$ as thin-tailed. a is the scale parameter.

The average radial distance can be obtained from $\delta = a\Gamma(3/b)/\Gamma(2/b)$ (Oddou-Muratorio and Klein, 2008) and the negative exponential function from $\delta = 2a$ (Austerlitz and Smouse, 2001).

Kernel distribution is a decreasing function of distance satisfying $\int_0^\infty f(r)2\pi r dr = 1$. With a neighborhood size of N , we can consider that a fraction $\alpha = \int_0^N f(r)2\pi r dr$ of dispersal events occurs inside the neighborhood whereas the complementary fraction $1 - \alpha = \int_N^\infty f(r)2\pi r dr$ occurs outside the neighborhood (Goto *et al.*, 2006). With an infinite neighborhood size, we consider that N is half of the maximum distance between potential parents in a patch network in order to ensure that all individuals are included in the neighborhood. The fraction $1 - \alpha$ can be considered as the expected immigration rate—of seed or pollen—if the dispersal kernel is the only dispersal source (Goto *et al.*, 2006). Comparison of these expected and estimated rates (m_s and m_p) allows an estimate of all random processes (for example, long-distance dispersal) that are not captured by the kernels (Goto *et al.*, 2006).

Parameter estimate

We obtained maximum likelihood estimates for seed and pollen immigration rates (m_s and m_p), selfing rate (s) and kernel parameters (a_s and b_s for seed dispersal kernel; a_p and b_p for pollen dispersal kernel) by maximizing the log-likelihood function:

$$L(m_s, m_p, s, a_s, b_s, a_p, b_p) = \prod_{i=1} P(G_i)$$

We used the maximum likelihood framework and fitted parameters with a quasi-Newton-algorithm implemented with Mathematica (see Oddou-Muratorio and Klein, 2008 for details). We constructed three different dispersal models with a combination of negative exponential and exponential power dispersal curves for seed and pollen dispersal. The simplest model considers a negative exponential function for seed and pollen kernels (one parameter per kernel) whereas the more complex model integrates the exponential power function for each dispersal kernel (two parameters per kernel). We also considered a mixed scenario with a negative exponential function for the pollen and an exponential power function for the seed dispersal kernel. We subsequently chose the best model using the Likelihood Ratio Test. As SPOMs estimated from SPOMSIM (Moilanen, 2004) only estimate negative exponential kernels, we used negative exponential kernels for seed dispersal in order to compare direct genetic estimates from SEMM and indirect estimates from SPOMs. This analysis was performed for both metapopulations.

Data quality diagnostic

In neutral alleles, null alleles correspond to the absence of expression of specific allelic forms. Because null alleles induce errors in individual genotype identification, they are commonly recognized as an important bias in parentage analysis. We thus estimated the frequency of null alleles using the EM algorithm (Dempster *et al.*, 1977) implemented by default in genepop (Rousset, 2008). Then, in order to test the robustness of our analysis to the presence of the null allele, we performed a new analysis by changing systematically homozygous genotypes (for example, XX) into heterozygous genotypes using a common null allele (XN, N being a null allele) for loci experiencing the substantial occurrence of null alleles (> 0.1). This method allowed us to consider the potential bias due to null alleles without drastically reducing the accuracy of the estimate in our data set (Oddou-Muratorio *et al.*, 2009). We compared estimates in the initial and modified data set (XN data set hereafter) in order to control for the effect of null alleles. We also removed markers with a high frequency of null alleles to obtain two data sets of six loci (that is, by removing either one or the other locus with high null allele frequency) and one data set of five loci (the two alleles presenting the highest null alleles frequency were removed from the subset). A comparison between the global data set (seven loci) and the ‘quality diagnostic’ data sets (XN, 6- and 5-loci data sets) allowed us to test the robustness of our results against the quality of genetic markers.

RESULTS

Molecular markers

Microsatellite markers presented high levels of polymorphism (from 4 to 49 alleles per locus) and provided a powerful situation for parentage analysis with a probability of exclusion (Chakraborty *et al.*, 1988) close to 1 (prob of exclusion = 0.99). Null alleles were present at substantial frequencies for different markers (see Supplementary Table A1). For the JC network, comparison of results from the global data set and ‘quality diagnostic’ data sets (as well as the XN data set with six or five loci) showed that immigration rates were overestimated when using the global data set (see below for details). For the HM network, we failed to fit the model with the data set controlling for null alleles (XN data set) and the five-loci data set and our ability to check the influence of data quality was restricted to the use of the two data sets with six loci. For these data sets, we found relatively good congruence with the global data set (see Supplementary Table A2). As our aim was to compare the two patch networks, we decided to use the global data set. However, we should keep in mind that the global data set may upwardly bias the immigration rate but that such bias is likely to be similar for the two patch networks. Concerning the dispersal kernel, we observed good congruence in the fitted values between the global and other data sets with the exception of the five-loci dataset, for which the results showed substantial incongruence (see Supplementary Table A2) for the JC network. The seed dispersal kernel showed good congruence between the different data sets for the HM network (see Supplementary Table A2).

SEMM parameters

Using the global data set, the best model that we selected ($p_{\text{chisq}} < 0.05$) in the JC network combined an exponential power function for seed dispersal ($b < 1$) and an exponential function (b fixed to 1) for pollen (Table 1). However, due to the divergence of algorithms for models integrating at least one exponential power function, we succeeded in obtaining results only for the negative exponential model in the HM network (Table 1). The estimated seed immigration rates (0.59 for JC and 0.71 for HM, see Table 1) and pollen immigration rates (0.71 and 0.84 for JC and HM, respectively, Table 1) were high. As previously noted, these rates were overestimated because the mean seed immigration in the ‘quality diagnostic’ data sets (XN, six and five loci) was 0.43 (min–max: 0.38–0.51) for the JC network. For the HM network, estimates were obtained for the two six loci data sets only thus limiting the diagnostic. This comparison revealed good congruence between the global data set and the two six-loci data sets (mean of 0.693, with

min–max: 0.692–0.694, see Supplementary Table A2). In the global data set, these substantial overestimated immigration rates decreased the ability to fit the pollen dispersal curve because local mating was scarce.

Using the global data set, the best models showed that seed dispersal was highly restricted in the two networks (Table 1). We fitted a leptokurtic function ($b = 0.53$) with an average radial dispersal distance of 12.6 m for the JC network whereas the exponential function fitted in the HM network led to an average radial dispersal distance of 1.53 m (see Table 1 and Figure 1). The choice of data set used to estimate parameters slightly affected the seed dispersal kernels with congruent average dispersal distance between different data sets (see Supplementary Table A2). Using kernel dispersal in order to theoretically estimate the expected immigration rate if the seed dispersal kernel was the only source of immigrants (Goto *et al.*, 2006), we estimated that a very small fraction of seeds was expected to come from outside the neighborhood (1–2%) in the two networks. The pollen dispersal kernels that we estimated indicated a low mean dispersal distance (< 10 m) and very low selfing rate in the two networks (Table 1).

Direct genetic vs indirect demographic dispersal curve estimates

As the SPOMs estimated from SPOMSIM (Moilanen, 2004) only reported negative exponential kernel estimates, we reported negative exponential kernels although this model did not provide the best fit for the genetic data set. For the JC network, SEMM gave a mean dispersal distance of 0.367 m under a negative dispersal kernel whereas SPOM gave a mean dispersal distance of 2.2 m (Dornier *et al.*, 2011; see Table 2). We synthesized all dispersal curves for the JC network in Figure 1. For the HM network, SPOM inference found no evidence of restricted dispersal and showed that the propagule rain model was the best model for explaining patch dynamics. While SPOM inference could be interpreted as a flat kernel with a constant probability of colonization for all patches, SEMM revealed a highly restricted dispersal kernel, although this was slightly higher than the JC estimates (negative exponential).

DISCUSSION

In this study, we used recent SEMM to estimate seed and pollen immigration rates as well as dispersal curves in two different patch networks of the annual plant *C. sancta* in an urban landscape. While a previous study documented contrasting patch dynamics among the two patch networks (Dornier *et al.*, 2011), this study aimed to describe the underlying mechanisms of these contrasting metapopulation dynamics. The two patch networks revealed limited dispersal.

Table 1 Parameter estimates of the best-fitted seedling neighborhood model

JC	Seed dispersal				Pollen dispersal					LnL	P_{chisq}
	m_s	a_s	b_s	$dist_s(m)$	m_p	a_p	b_p	$dist_p(m)$	s		
EXP _{seed} /EXP _{pollen}	0.606	0.188	1 ^a	0.376	0.698	0.640	1 ^a	1.28	0.071	−3551.45	-
POW _{seed} /EXP _{pollen}	0.581	0.859	0.532	12.60	0.718	4.186	1 ^a	8.37	0.066	−3548.65	0.0179
POW _{seed} /POW _{pollen}	0.527	0.863	0.527	13.14	0.721	0.680	1.084	1.18	0.065	−3548.18	0.332
HM											
EXP _{seed} /EXP _{pollen}	0.715	0.764	1 ^a	1.528	0.840	0.711	1 ^a	1.422	0.061	−6071.21	-

Dispersal models: EXP_{seed}/EXP_{pollen}: Negative exponential on seed and pollen. POW_{seed}/EXP_{pollen}: Exponential power on seed and negative exponential on pollen. POW_{seed}/POW_{pollen}: Exponential power on seed and pollen.

^aFixed parameter. m_s and m_p are the seed and pollen immigration rate, respectively, a_s and b_s represent the scale and the shape parameter of the seed dispersal kernel and a_p and b_p the scale and the shape parameter of the pollen dispersal kernel. $dist_s(m)$ and $dist_p(m)$ are the average radial seed and pollen dispersal distance (in m). s is the self-fertilization rate. LnL is the log-likelihood of the model and p_{chisq} is the result of Log Likelihood Ratio Test between two models of two successive lines. The significant value ($P < 0.05$) is indicated in bold.

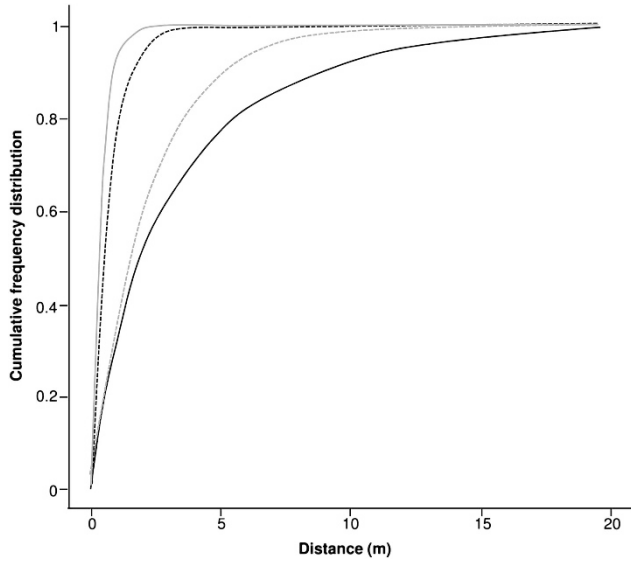


Figure 1 Cumulative frequency distribution of seed dispersal curve as a function of distance. Continuous and dotted black lines represent seed dispersal curve fitted with the best model in the JC and HM networks, respectively. Continuous gray line is the seed dispersal kernel for a negative exponential model for the JC network whereas the dotted gray line is the curve predicted by metapopulation modeling in the same network (Dornier *et al.*, 2011).

Table 2 Comparison of mean dispersal kernel between demographic and genetic methods

Dispersal function	SPOM estimated Exponential	SEMM estimated	
		Exponential	Exponential power
Estimated parameter	$a=0.45$	$a=0.180$	$a=0.859, b=0.532$
Average radial distance (m)	2.22	0.360	12.60

Abbreviations: SPOM, stochastic patch occupancy models; SEMM, spatially explicit mating model.
 SPOM estimated: dispersal kernel estimated using metapopulation modeling with $f(a, r) = \exp(-ar)$ (Dornier *et al.*, 2011) with a , the scale parameter. SEEM estimated: dispersal kernel estimated in the current study with $f(a, b, r) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp\left(-\left(\frac{r}{a}\right)^b\right)$.
 Estimated parameter: a is the scale of the kernel and b the shape of the parameter (b is fixed to 1 in exponential function). Average radial dispersal distance (in m) estimated from seed dispersal kernel.

Nevertheless, the analysis revealed substantial external migration, which was higher in the HM network than in the JC network. Our combined approach allowed us to link genetically estimated dispersal with patch dynamics and to assess the congruence between demographic and genetic methods of dispersal estimation.

Immigration rates in plant metapopulations

Immigration rates estimated from SEMM were surprisingly high despite a highly restricted kernel from which less than 2% of long-distance dispersal is theoretically expected (Goto *et al.*, 2006). The difference between theoretical predictions from the dispersal kernel and fitted values suggests that random processes and/or methodological bias have generated these substantial immigration rates (Goto *et al.*, 2006). Methodological problems include human error, problems during PCR procedures and errors during genotyping that are commonly reported in genetic studies and are known to lead to overestimates of immigration rate (Pompanon *et al.*, 2005). In our study, we controlled for the presence of null alleles using the XN data

set for the JC network and our results suggested that immigration rate was indeed inflated by the presence of null alleles. Although such results were not obtained for the HM network, it is likely that the presence of null alleles biased estimates in the same way for the two networks, given the absence of significant genetic structure among the networks (Dornier and Cheptou, 2012). Although changing the genetic data set for JC was found to affect immigration, the immigration rate remained relatively high and was unlikely to be due to human error or marker quality.

Assumptions regarding the genetic frequencies of the background population in the SEMM can induce a slight overestimation of immigration (Burczyk and Chybicki, 2004) whereas missing individuals are known to increase immigration when the actual parents are not sampled (Oddou-Muratorio and Klein, 2008). Although sampling effort was *a priori* exhaustive in this study, we cannot rule out the possibility that potential parents were missed in 2006 or that missing data in our data set led to an overestimation of immigration rate. However, although these different biases could combine to generate substantial false immigration, these methodological problems are not expected to produce bias on the fit of dispersal curves but only decrease the statistical power of the fit of the dispersal kernel. Importantly, the difference between the two networks cannot be explained only by methodological bias and such a difference suggests contrasting ecological processes.

From an ecological point of view, immigration rates are influenced by two contrasting random factors: long-distance dispersal and germination of the seed bank. Dornier *et al.* (2011) previously reported the absence of a substantial seed bank in *C. sancta* after analyzing soil samples from the same urban patches. Thus, pseudo-colonization can be excluded as a bias in the immigration estimates. The immigration rate may reflect spatial dispersal processes and in particular those from unknown sources. *C. sancta* produces many seeds with a pappus that provides high potential dispersal ability. Substantial immigration from an unknown source was previously considered as rare but recent studies using parentage analysis have found such seed immigration in trees (Ashley, 2010), with a record of 50–70% in fragmented populations of *Fraxinus excelsior* (Bacles *et al.*, 2006). Such data remain rare in annual plants but our study documented high seed immigration rates that differed between the two networks.

Seed and pollen dispersal kernels

Dispersal kernels were found to be limited with a maximum mean distance of a few meters in both networks. Shorter-distance seed dispersal was found in the HM network but this finding could reflect the different kernel function used in the selected models. Indeed, we were only able to fit a negative exponential function for HM whereas an exponential power function provided the best fit for the JC network. Indeed, comparison of the negative exponential functions showed more congruent estimates between the two networks. This is consistent with a previous study that documented similar phenotypic dispersal traits (estimated using the ratio of non dispersing to dispersing seeds) in the two networks (Cheptou *et al.*, 2008). Thus, although specific geographical processes are likely to affect seed dispersal differently, we suggest that the low statistical performance for the HM network is the main reason why the best model differed among the two metapopulations.

The substantial seed and pollen immigration rates estimated in the two patch networks limited the ability to fit a pollen dispersal curve and further investigations are required to obtain more robust results. Thus, the dispersal kernels for pollen must be considered with

caution. Finally, the low selfing rates found in the two networks are congruent with those obtained in the same populations based on multilocus heterozygosity (Dornier and Cheptou, 2012).

Genetic vs demographic dispersal in plant metapopulations

The combined analysis of patch dynamics estimated using the SPOMSIM software (Dornier *et al.*, 2011) and SEMM allows us to analyze links between the dispersal kernel and colonization pattern in fragmented plant populations. The central question is whether heterogeneous external immigration among networks coupled with relatively homogeneous seed dispersal estimated by SEMM is consistent with the metapopulation dynamics inferred in SPOMs.

In HM, the absence of limited colonization inferred from presence/absence data (SPOM) is consistent with the high seed immigration rates inferred from SEMM even in the presence of the restricted seed dispersal kernel found in our analysis. The results suggest that substantial seed immigration from an unknown source population is the main driver of the patch dynamics in this network and thus cancels the limited colonization in SPOM. However, the seed immigration rates obtained using SEMM (0.71) may be overestimated compared with the colonization rate fitted using SPOMS (Dornier *et al.*, 2011). It is important to note that a large population was found very close to the HM network and may be the potential source population of seed immigration that influenced the patch dynamics. Further investigations are required to investigate the actual influence of this potential source on metapopulation dynamics.

In the JC network, SEMM reported a substantial, although lower than in the HM network, immigration rate (38 or 59% according to the data set) coupled with a restricted dispersal kernel whereas SPOM found the Spatially Realistic Levins Models to be the best model. The two approaches are congruent and indicate that lower external migration coupled with a restricted seed dispersal kernel generate local colonization as inferred from SPOMS. In this network, the seed dispersal kernel appears to be the main driver of patch dynamics. Again, due to the relatively high seed immigration rate estimated using the XN data set (38%), it is likely that this network is driven by a combination of metapopulation scenarios whereby colonization is due to both seed immigration and local seed dispersal. The development of metapopulation models integrating these two colonization components would increase our understanding and prediction of patch dynamics in a highly fragmented landscape.

The presence of substantial immigration in a Levins metapopulation model opens interesting perspectives on the influence of dispersal and connectivity on demographic and genetic processes (Lowe and Allendorf, 2010). Genetic connectivity is classically estimated using FST (Lowe and Allendorf, 2010) but in the JC network, FST could reflect a mix of immigration and limited dispersal, which would lead to overestimating genetic connectivity (see also Waples and Gaggiotti, 2006). In this respect, a theoretical prediction of genetic structure and isolation by distance in a metapopulation confronted with local restricted dispersal and substantial immigration is required in order to better understand how connectivity influences genetic and demographic processes (Lowe and Allendorf, 2010).

CONCLUSION

By confronting genetic estimates of dispersal from SEMM and estimates of colonization from SPOMs in plant metapopulation, our study allows to better understand how congruent both estimates are. Although direct estimates (SEMM) allow quantifying dispersal kernels and external immigration, we showed that colonization estimates from indirect methods (SPOM) were, at least qualitatively,

consistent with direct estimates. Importantly, we clearly showed that information from indirect colonization encompasses various processes and must not be interpreted as a dispersal process only. Because patch occupancy models do not integrate local demography, such methods may provide manageable tools for analyzing population dynamics in spatially and temporally heterogeneous landscapes in the context of global change.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.b911r.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their helpful comments on the manuscript. The TrameVerte program (ANR-0.8-VILL-0003) and the chercheurs d'avenir -LR, France program (POC) provided financial support for this study.

- Ashley MV (2010). Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. *Cr Rev Plant Sci* **29**: 148–161.
- Austerlitz F, Smouse PE (2001). Two-generation analysis of pollen flow across a landscape. ii. Relation between ϕ_i (ft), pollen dispersal and interfemale distance. *Genetics* **157**: 851–857.
- Bacles CFE, Lowe AJ, Ennos RA (2006). Effective seed dispersal across a fragmented landscape. *Science* **311**: 628–628.
- Bennetts RE, Nichols JD, Lebreton J-D, Pradel R, Hines JE, Kitchens WK (2001). Methods for estimating dispersal probabilities and related parameters using marked animals. In: Clobert J, Danchin E, Dhondt AA, Nichols JD (eds) *Dispersal*. Oxford University Press: New York, pp 3–17.
- Bullock JM, Shea K, Skarpaas O (2006). Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecol* **186**: 217–234.
- Burczyk J, Adams WT, Birkes DS, Chybicki IJ (2006). Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics* **173**: 363–372.
- Burczyk J, Chybicki IJ (2004). Cautions on direct gene flow estimation in plant populations. *Evolution* **58**: 956–963.
- Chakraborty R, Meagher TR, Smouse PE (1988). Parentage analysis with genetic-markers in natural populations. 1. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**: 527–536.
- Cheptou PO, Carrue O, Rouifed S, Cantarel A (2008). Rapid evolution of seed dispersal in an urban environment in the weed *Crepis sancta*. *Proc Natl Acad Sci USA* **105**: 3796–3799.
- Dempster A, Laird N, Rubin D (1977). Maximum likelihood from incomplete data via the em algorithm. *Roy Stat Soc B* **39**: 1–38.
- Dornier A, Cheptou P-O (2012). Determinants of extinction in plant fragmented populations: *Crepis sancta* (Asteraceae) in urban environment. *Oecologia* **169**: 703–712.
- Dornier A, Pons V, Cheptou P-O (2011). Colonization and extinction dynamics of an annual plant metapopulation in an urban environment. *Oikos* **120**: 1240–1246.
- Dubois MP, Dornier A, Jarne P, Cheptou PO (2007). Nine polymorphic microsatellite markers in *Crepis sancta* (Asteraceae). *Mol Ecol Notes* **7**: 681–683.
- Etienne R, ter Braak C, Vos C (2004). Applications of stochastic patch occupancy models to real metapopulations. In: Hanski I, Gaggiotti OE (eds) *Ecology, Genetics and Evolution of Metapopulation*. Academic Press: Amsterdam, pp 105–132.
- Freckleton RP, Watkinson AR (2002). Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *J Ecol* **90**: 419–434.
- Gadgil M (1971). Dispersal: population consequences and evolution ecology. *Ecology* **52**: 253–261.
- Giles BE, Goudet J (1997). Genetic differentiation in *Silene dioica* metapopulations: Estimation of spatiotemporal effects in a successional plant species. *Am Nat* **149**: 507–526.
- Gotelli NJ (1991). Metapopulation models – the rescue effect, the propagule rain, and the core-satellite hypothesis. *Am Nat* **138**: 768–776.
- Goto S, Shimatani K, Yoshimaru H, Takahashi Y (2006). Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. *japonica* revealed by microsatellites. *Mol Ecol* **15**: 2985–2996.
- Hadfield JD, Richardson DS, Burke T (2006). Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol Ecol* **15**: 3715–3730.
- Hanski I (1994). A practical model of metapopulation dynamics. *J Animal Ecol* **63**: 151–162.

- Hanski I, Gaggiotti O (2004). Metapopulation biology: past, present, and future. In: Hanski I, Gaggiotti OE (eds) *Ecology, Genetics and Evolution of Metapopulation*. Academic Press: Amsterdam, pp 3–22.
- Hanski I, Ovaskainen O (2000). The metapopulation capacity of a fragmented landscape. *Nature* **404**: 755–758.
- Harper JL (1977). *Population Biology of Plants*. Academic Press: New York.
- Imbert E (1999). The effects of achene dimorphism on the dispersal in time and space in *Crepis sancta* (Asteraceae). *Can J Bot* **77**: 508–513.
- Jeltsch F, Moloney KA, Schurr FM, Köchy M, Schwager M (2008). The state of plant population modelling in light of environmental change. *Perspect Plant Ecol, Evol Syst* **9**: 171–189.
- Levins R (1969). Some demographic and genetic consequences of environment heterogeneity for biological control. *Bull Entomol Soc Am* **15**: 237–240.
- Lowe WH, Allendorf FW (2010). What can genetics tell us about population connectivity? *Mol Ecol* **19**: 5320–5320.
- MacArthur RH, Wilson EO (1967). *The Theory of Island Biogeography*. Princeton University Press: Princeton, NJ.
- Moilanen A (2000). The equilibrium assumption in estimating the parameters of metapopulation models. *J Animal Ecol* **69**: 143–153.
- Moilanen A (2004). Spomsim: software for stochastic patch occupancy models of metapopulation dynamics. *Ecol Model* **179**: 533–550.
- Moilanen A, Nieminen M (2002). Simple connectivity measures in spatial ecology. *Ecology* **83**: 1131–1145.
- Nathan R, Muller-Landau HC (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends Ecol Evol* **15**: 278–285.
- Oddou-Muratorio S, Klein EK (2008). Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Mol Ecol* **17**: 2743–2754.
- Oddou-Muratorio S, Vendramin GG, Buiteveld JB, Fady B (2009). Population estimators or progeny tests: what is the best method to assess null allele frequencies at *ssr* loci? *Conservation Genetics* **10**: 1343–1347.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005). Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* **6**: 847–859.
- Rousset F (2008). Genepop 007: a complete re-implementation of the genepop software for windows and linux. *Mol Ecol Resour* **8**: 103–106.
- Slatkin M (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol* **12**: 253–262.
- Verheyen K, Vellend M, Van Calster H, Peterken G, Hermy M (2004). Metapopulation dynamics in changing landscapes: A new spatially realistic model for forest plants. *Ecology* **85**: 3302–3312.
- Waples RS, Gaggiotti OE (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* **15**: 1419–1439.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)