Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review

Hugo Cayuela¹ | Quentin Rougemont¹ | Jérôme G. Prunier² | Jean-Sébastien Moore¹ | Jean Clobert² | Aurélien Besnard³ | Louis Bernatchez¹

¹Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec City, Québec, Canada
²Station d’Ecologie Théorique et Expérimentale, Unité Mixte de Recherche (UMR) 5321, Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS), Moulis, France
³CNRS, PSL Research University, EPHE, UM, SupAgro, IRD, INRA, UMR 5175 CEFE, Montpellier, France

Abstract
Dispersal is a central process in ecology and evolution. At the individual level, the three stages of the dispersal process (i.e., emigration, transience and immigration) are affected by complex interactions between phenotypes and environmental factors. Condition- and context-dependent dispersal have far-reaching consequences, both for the demography and the genetic structuring of natural populations and for adaptive processes. From an applied point of view, dispersal also deeply affects the spatial dynamics of populations and their ability to respond to land-use changes, habitat degradation and climate change. For these reasons, dispersal has received considerable attention from ecologists and evolutionary biologists. Demographic and genetic methods allow quantifying non-effective (i.e., followed or not by a successful reproduction) and effective (i.e., with a successful reproduction) dispersal and to investigate how individual and environmental factors affect the different stages of the dispersal process. Over the past decade, demographic and genetic methods designed to quantify dispersal have rapidly evolved but interactions between researchers from the two fields are limited. We here review recent developments in both demographic and genetic methods to study dispersal in wild animal populations. We present their strengths and limits, as well as their applicability depending on study objectives and population characteristics. We propose a unified framework allowing researchers to combine methods and select the more suitable tools to address a broad range of important topics about the ecology and evolution of dispersal and its consequences on animal population dynamics and genetics.

KEYWORDS
capture–recapture models, dispersal, dispersal kernel, gene flow, migration

1 INTRODUCTION

Dispersal is a central process in ecology and evolution as it deeply affects the demography (Benton & Bowler, 2012; Hanski & Gilpin, 1991; Hansson, 1991) and the genetic structuring of natural populations (Baguette, Blanchet, Legrand, Stevens, & Turlure, 2013; Olivieri, Michalakis, & Gouyon, 1995; Ronce, 2007), as well as adaptive processes (Hanski & Gaggiotti, 2004; Legrand et al., 2017; Ronce, 2007). From an applied point of view, dispersal influences the spatial dynamics of populations and their ability to respond to land-use changes, habitat degradation and climate change (Baguette et al., 2013; Caplat et al., 2016; Travis et al., 2013). Measuring dispersal is therefore of crucial importance, and methods to do so are continuously evolving. Broadly speaking, animal dispersal can be measured using demographic methods that mainly rely on capture–recapture approaches or through the use of molecular markers. While both are
rapidly evolving, interactions among researchers in these two specialized fields are often limited. We here review recent developments in both demographic and genetic methods to study animal dispersal to foster their combined use under a unified framework.

1.1 What is dispersal?

Dispersal designates the movement of an individual between its site of birth and its first breeding site (i.e., natal dispersal), or among successive breeding sites (i.e., breeding dispersal; Baguette & Van Dyck, 2007; Clobert, Galliard, Cote, Meylan, & Massot, 2009; Matthysen, 2012). Dispersal can be passive or active. In passive dispersers, movement is mainly driven by extrinsic factors such as wind, ocean currents or dispersal agents as animals (Bohonak & Jenkins, 2003; Burgess, Baskett, Grosberg, Morgan, & Strathmann, 2016; Nathan et al., 2008). In active dispersers, dispersal often implies specialized large-scale one-way movements potentially resulting in gene flow (Cote, Bestion et al., 2017; Ronce, 2007; Van Dyck & Baguette, 2005). Therefore, it is distinguished from migration, which implicates recurrent, two-way out and back movements, and from foraging movements implying frequent, short-distance movements to locate resources (Cote, Bocedi et al., 2017). Synonyms of dispersal have sometimes been used in the specialized literature dedicated to several taxonomic groups. For example, the term “straying” is the dispersal of mature fishes to spawn in a stream other than the one where they originated (Quinn, 1993). Furthermore, although the terms “dispersal” and “migration” are often considered synonyms in the context of population genetics (Broquet & Petit, 2009), formally they should be considered as two distinct ecological processes (Cote, Bocedi et al., 2017). In a population genetics context, we use the term “dispersal rate” to refer to the quantity \( \sigma^2 \), the mean squared axial parent-offspring distance (Rousset, 1997), while the term “migration” rate refers to the quantity \( m \) (Box 1), the proportion of genes in a subpopulation that originate from new immigrants at each generation. This conceptual point remains a source of ambiguity in many population genetics studies (also discussed in Broquet & Petit, 2009; Lowe & Allendorf, 2010).

Dispersal is usually thought as a three-stage process (Baguette & Van Dyck, 2007; Clobert et al., 2009; Matthysen, 2012) including: (a) emigration, which corresponds to the departure of an individual from its site of birth or its current breeding site, (b) transience that determines the movement of an individual in the landscape matrix, and (c) immigration, which designates the settlement in a new breeding site. Theory predicts that the evolution of dispersal depends on the balance between costs and benefits at each step of the process (Bonte et al., 2012) and that this balance is potentially affected by individual, social and environmental factors: that is, context- and condition-dependent dispersal (Bowler & Benton, 2005; Matthysen, 2012; Ronce & Clobert, 2012). At the individual level, morphological (e.g., body size and condition), behavioural (e.g., boldness and aggressiveness) and physiological traits (e.g., testosterone and corticosterone titres) can all influence the propensity to emigrate, the locomotor capacities mobilized during transience, and habitat selection associated with the immigration phase (Cote, Clobert, Brodin, Fogarty, & Sih, 2010; Davis & Stamps, 2004; Ronce & Clobert, 2012; Stamps, 2001). The covariation patterns between dispersal and phenotypic traits have been coined “dispersal syndromes” (Clobert et al., 2009; Ronce & Clobert, 2012). Individuals are expected to adjust their emigration and immigration decisions according to environmental and social cues that reflect the fitness prospects in a given site (i.e. informed dispersal Clobert et al., 2009). Site-specific environmental factors such as the quantity of food supplies, the density of heterospecifics and predation risks, may have a broad influence on emigration and immigration decisions (Bowler & Benton, 2005; Clobert, Ims, & Rousset, 2004; Matthysen, 2012). Emigration and immigration may also be profoundly affected by social factors including kin competition/selection and inbreeding risks (Bowler & Benton, 2005; Matthysen, 2012). Moreover, landscape composition and configuration have a strong influence on dispersal (Baguette et al., 2013; Cote, Bestion et al., 2017; Fahrig, 2003). Namely, individual movement during the transience phase is affected by the availability of sites and their level of geographic isolation, as well as the permeability of the landscape matrix (Baguette & Van Dyck, 2007; Baguette et al., 2013; Pflüger & Balkenhol, 2014).

1.2 Demographic and genetic consequences of dispersal

The processes occurring at the individual level have far-reaching consequences for the dynamics of spatially structured animal populations (Figure 1; Hanski, 1991; Hanski & Gaggiotti, 2004; Gilpin, 2012). Spatially structured populations are composed of a set of populations (or “subpopulations” in several demographic studies) occupying distinct sites (or “patches,” “demes”) that are linked by dispersing individuals (Revilla & Wiegand, 2008; Thomas & Kunin, 1999); they encompass all the population categories classically considered in the general “metapopulation” framework (i.e., “Levins metapopulation,” Hanski, 1998; “patchy populations,” Hastings & Harrison, 1994; “source-sink” and “pseudo-sinks” systems, Pulliam, 1988; “mainland-island” systems, Hanski & Gilpin, 1991). Condition- and context-dependent dispersal decisions deeply affect dispersal rates and distances (Clobert et al., 2009; Cote, Bestion et al., 2017), which in turn influences the dynamics and the long-term viability of spatially structured populations. Theory predicts that dispersal has a strong influence on the dynamics of populations by affecting the local population growth rate through net immigration (= immigration – emigration; Hastings, 1993). The contribution of dispersal to the rate of population growth is called demographic connectivity (Lowe & Allendorf, 2010). Dispersal increases the level of demographic similarity (i.e., absolute values of vital rates) and synchrony (i.e., relative change of vital rates) among populations over time (Abbott, 2011; Bjørnstad, Ims, & Lambin, 1999; Hastings, 1993; Ranta, Kaitala, Lindstrom, & Linden, 1995). It also reduces the risk of population extinction when the local population sizes are small and/or the local population growth rates are low (i.e., “rescue effect,” Gotelli, 1991; Harrison, 1991; Lowe & Allendorf, 2010). In parallel,
dispersal also increases the colonization rate of empty sites, which therefore decreases the extinction chances of the whole spatially structured population (Ebenhard, 1991; Gilpin, 2012).

As dispersal implies the movement of individuals that may contribute to reproduction, it can result in gene flow. Dispersal is called “effective” when the disperser (an animal or a gamete) successfully transmits its genes to the next generation, which leads to gene flow. “Non-effective” dispersal refers to cases where the disperser moves into another patch regardless of whether it successfully reproduces or not (Broquet & Petit, 2009). As dispersal is a costly process (Bonte et al., 2012), dispersers may pay acute survival and reproductive costs after immigrating in a new patch, which therefore affects lifetime reproductive success and gene flow (Ronce, 2007). Gene flow has profound influences on the
evolutionary trajectories of populations through modification of allele frequencies within populations (Figure 1). In the absence of gene flow, populations readily evolve through changes in allele frequencies due to the other four major evolutionary forces: mutation, random genetic drift (random fluctuation of allele frequencies in populations of finite size), recombination and selection. If gene flow occurs between populations from different demes, allele frequencies are expected to be homogenized thus reducing genetic differentiation. As a result, Wright (1951) showed that one migrant per generation may be sufficient to reduce allele frequency differences between populations and avoid the harmful effects of local inbreeding (i.e., inbreeding connectivity, see Lowe & Allendorf, 2010). Therefore, gene flow stemming from effective dispersal can play a major role in reducing population inbreeding and the fixation of deleterious mutations (Keller & Waller, 2002) and thus maintain fitness by counteracting the random loss of genetic diversity (Frankham, 2015). If populations experience a deleterious mutations load, then hybrid offspring between residents and immigrants may display heterosis and have higher fitness than their parents. Gene flow ultimately leads to the spread of immigrant alleles among nearby populations, thus increasing effective dispersal rates (Ingvarsson & Whitlock, 2000). More generally, dispersal may increase fitness within populations by introducing new adaptive variants (Frankham, 2015). Furthermore, the absence of gene flow is generally believed to be an important condition for speciation, although modest amounts of gene flow during secondary contacts can favour reinforcement (i.e., increase in prezygotic isolation due to selection against interspecific mating) and lead to complete reproductive isolation (Barton & Hewitt, 1985; Coyne & Orr, 2004; Servedio & Noor, 2003). The absence of gene flow may also favour local adaptation as gene flow may swamp locally adapted alleles and limit local adaptation (Morjan & Rieseberg, 2004). For instance, at migration–selection equilibrium, in a two-deme model, local adaptation can be maintained only when the effective migration rate $m$ is lower than selection $s$ favouring the locally adapted allele so that $m/s < 1$ (Bulmer, 1971; Lenormand, 2002; Yeaman & Otto, 2011). On the contrary, it is now acknowledged that dispersal may not always be random (Clobert et al., 2009; Edelaar, Siepielski, & Clobert, 2008; Garant, Kruuk, Wilkin, McCleery, & Sheldon, 2005) and may, for instance, favour local adaptation when immigrants select
their recipient patch according to their own phenotype, thus increasing assortative mating (Jacob et al., 2017). Gene flow between already diverged populations or species may also favour admixture (Kuhlwilm et al., 2016) and even subsequently adaptive introgression (Arnold & Kunte, 2017). Estimating the level and nature (e.g., random or not) of gene flow is hence of paramount importance to understand whether populations will be able to cope with global change, especially for low mobility species (e.g., Aitken & Whitlock, 2013; Corlett & Westcott, 2013; Kremer et al., 2012).

1.3 Demographic and genetic tools to estimate dispersal

A broad range of demographic and genetic methods have been used to quantify and study non-effective dispersal in free-ranging animal populations. Among the demographic methods, telemetry and capture–recapture (CR) surveys are the ones most commonly used to study animal movements (Hooten, Johnson, McClintock, & Morales, 2017; Hussey et al., 2015; Lebreton, Nichols, Barker, Pradel, & Speddelow, 2009; Lebreton & Pradel, 2002; Royle, Fuller, & Sutherland, 2018; Shafer, Northrup, Wikelski, Wittemyer, & Wolf, 2016). Telemetry methods pose a series of technical and conceptual challenges for the study of dispersal. First, despite remarkable advances in transmitter miniaturization, many species have body sizes too small to carry these devices without potential deleterious effects. Second, while telemetry methods are very efficient for tracking routine and cyclic movements related to foraging and migration (e.g., Bestley, Jonsen, Hindell, Harcourt, & Gales, 2015; Cumming, Henry, & Reynolds, 2017; Doherty et al., 2017; Hoenner, Whiting, Hindell, & McMahon, 2012; Moore et al., 2017), they are far less effective at detecting occasional dispersal events. Individuals are often surveyed over short time periods (rarely more than 3 years) due to limited battery capacities, which are not long enough to detect dispersal events, in particular for medium- to long-lived species. This limitation was recently exemplified by a study that revealed a mismatch between patterns of gene flow (resulting from dispersal) and migratory movements detected using telemetry data (Moore et al., 2017). Finally, the relatively high cost of telemetry devices generally allows surveying a small number of individuals (usually <30) and does not permit to quantify population-level dispersal rates and distances. For these reasons, we will focus on CR methods rather than telemetry for the remainder of this review.

In contrast to telemetry, CR methods allow: (a) examining dispersal in a broad range of taxa including small-sized organisms while accounting for non-exhaustive observation of individuals (e.g., Beirinckx, Van Gossum, Lajeunesse, & Forbes, 2006; Plășcu, Ozgul, Schmidt, & Băncilă, 2017; Vlasanek, Sam, & Novotny, 2013), (b) surveying the individuals throughout their entire lifetime to track punctual events of both natal and breeding dispersal (e.g., Balkiz et al., 2010; Blums, Nichols, Hines, Lindberg, & Mednis, 2003; Devillard & Bray, 2009); and surveying large samples of populations and estimate population-level dispersal and vital rates (e.g., Cayuela et al., 2016; Lebreton, Hines, Pradel, Nichols, & Speddelow, 2003; Serrano, Oro, Ursua, & Tella, 2005). Note that, while CR methods can be potentially suitable to examine dispersal in passive dispersers, these methods have been generally used in studies focusing on organisms with an active dispersal. Thus, the demographic methods considered in this study will therefore be more appropriate for examining dispersal in actively dispersing animals than in passive dispersers. During the last four decades, a broad range of CR models have been developed to quantify dispersal rates (Arnason, 1972; Lebreton et al., 2003; Schwarz, Schweigert, & Arnason, 1993) and distances (Ergon & Gardner, 2014; Fujiwara, Anderson, Neubert, & Caswell, 2006) and to test hypotheses about the effects of individual and

**FIGURE 2** Decision tree showing the demographic (i.e., CR modelling) and genetic methods that can be used to estimate non-effective and effective dispersal rate and distance. The description of all these methods is provided in section How to quantify non-effective and effective dispersal rates and distances using demographic and genetic approaches? [Colour figure can be viewed at wileyonlinelibrary.com]
environmental factors on each step of the dispersal process (i.e., emigration, transience and immigration; Grosbois & Tavecchia, 2003; Ovaskainen, 2004; Cayuela, Pradel, Joly, & Besnard, 2017; Cayuela, Pradel, Joly, Bonnaire, & Besnard, 2018).

Simultaneously, genetic approaches dedicated to quantifying both non-effective and effective dispersal also received considerable attention (Broquet & Petit, 2009). Recent decades in particular have seen increasing development of analytical tools to perform demographic inferences and estimating effective dispersal using either the allele frequency spectrum (Gutenkunst, Hernandez, Williams, & Bustamante, 2009), summary statistics relying on coalescent theory such as approximate Bayesian computation (Beaumont, 2010; Beaumont, Zhang, & Balding, 2002), or inferences based on blocks of identity-by-descent (Browning & Browning, 2011) making it possible to study the movement of genes between populations with increased precision. Concomitantly, decreasing sequencing costs make it increasingly easy to genotype or sequence large numbers of individuals using a range of methods from RADseq (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016) to whole genome sequencing (Ellegren, 2014; Fuentes-Pardo & Ruzzante, 2017), including whole genome pool sequencing data (Schlötterer, Tobler, Kofler, & Nolte, 2014). Careful use of these next-generation sequencing (NGS) data can provide further insights into levels of connectivity, even in species for which inferring genetic connectivity can be difficult due to large population size or high dispersal rates (Gagnaire et al., 2015).

1.4 | The goals of the review

In this review, we aim to propose a unified framework allowing demographers and population geneticists to select the most suitable tools according to their biological questions and the characteristics of studied populations. We first review the demographic and genetic methods available to estimate non-effective and effective dispersal rates and distances in animals (Figure 2). Next, we review the methods allowing investigation of the effects of environmental and individual factors on the three stages of dispersal (emigration/immigration and transience), in their non-effective and effective dimensions (Figure 3). Finally, we conclude this synthesis by giving a number of recommendations to help the reader select accurate methods, and eventually combine approaches, to address a set of important issues about the dynamics and the genetics of wild animal populations.

2 | HOW TO QUANTIFY NON-EFFECTIVE AND EFFECTIVE DISPERSAL RATES AND DISTANCES USING DEMOGRAPHIC AND GENETIC APPROACHES?

Demographic and genetic methods used to quantify effective and non-effective dispersal rates and distances are summarized in Figure 2.

2.1 | Estimating non-effective dispersal rates and distances using demographic approaches

2.1.1 | Estimating dispersal rates using multistate models

The first CR models developed to quantify dispersal rates among discrete sites (or patches) date from the 1970s. Arnason (1972, 1973) proposed the first multisite (or “multi-strata” in its general formulation) models with time-varying recruitment and survival, in
which individuals can be captured at three distinct dates, in different sites. This model can be viewed as a generalization of the single-site Cormack–Jolly–Seber model (Cllobert, Lebreton, & Allaine, 1987; Cormack, 1964; Jolly, 1965; Seber, 1965), allowing individuals to disperse between two sites across successive capture occasions. Twenty years later, Schwarz et al. (1993) proposed a generalization of this model (called “Arnason–Schwarz model”) by considering more than three capture occasions and a large number of recapture sites. The Arnason–Schwarz model paved the way for the development of multistate models (Lebreton & Pradel, 2002; Lebreton et al., 2009; Nichols & Kendall, 1995), in which one considers that individuals may move within a finite set of states that reflect individual variables such as body size (small vs. large), body condition (poor vs. good) or life history stages (juvenile vs. adult), rather than simple geographic states (sites). In such models, the transitions between states are modelled as first-order Markovian processes (i.e., in which the state at time $t$ only depends on the state at $t−1$). The basic parameters of the Arnason–Schwarz model are as follows:

\[
\begin{align*}
\phi_R^t &= \text{the probability that an individual alive in site } R \text{ at occasion } t−1 \text{ is still alive at occasion } t \text{ (i.e., survival probability).} \\
\psi_{RT}^t &= \text{the probability that an individual in site } R \text{ at occasion } t−1 \text{ disperses to site } T \text{ at occasion } t \text{ provided it survives (i.e., dispersal probability).} \\
\rho_R^t &= \text{the probability that an individual alive in site } R \text{ is recaptured at occasion } t \text{ (i.e., recapture probability).}
\end{align*}
\]

In 1993, Brownie, Hines, Nichols, Pollock, and Hestbeck (1993) developed the first software (MSSURVIV) specifically dedicated to the construction of multistate models. Other user-friendly programs (MARK, White & Burnham, 1999; White, Kendall, & Barker, 2006; MSURGE, Choquet, Reboulet, Pradel, Gimenez, & Lebreton, 2004) spurred a rapid and straightforward implementation of these models, resulting in an extensive use of multistate models to quantify dispersal rates in a broad range of taxa including insects (Chaput-Bardy, Grégoire, Baguette, Pagano, & Secondi, 2010), other arthropods (Mills, Gardner, & Oliver, 2005), birds (Cam, Oro, Pradel, & Jimenez, 2004; Doligez et al., 2002; Dugger, Ainley, Lyver, Barton, & Ballard, 2010), mammals (Sanderlin, Waser, Hines, & Nichols, 2012; Skvarla, Nichols, Hines, & Waser, 2004), fishes (Frank, Gimenez, & Baret, 2012; Haugen et al., 2007), reptiles (Dodd, Ozgul, & Oli, 2006; Roe, Brinton, & Georges, 2009) and amphibians (Funk, Greene, Corn, & Allendorf, 2005; Grant, Nichols, Lowe, & Fagan, 2010).

In summary, the Arnason–Schwarz model allows quantifying dispersal rates between pairs of sites while accounting for survival and recapture probabilities, using data collected across multiple surveys (e.g., years) and sites. It is a suitable approach to estimate dispersal rates (and possibly distances, see Fernández-Chacón et al., 2013) in species occupying discrete habitat patches or breeding sites. The Arnason–Schwarz model is not used to quantify dispersal in species occurring in relatively continuous and homogeneous environments. Yet, one of the most important issues with this model is that the number of parameters rapidly increases with the number of monitored sites (or more generally of states), which may result in problems of stability and precision of estimates, and identifiability of parameters (Lebreton & Pradel, 2002; Lebreton et al., 2009). Indeed, for $n$ states, the number of transitions among states to be estimated is $n(n−1)$ and is thus a function of $n^2$.

### 2.1.2 Estimating dispersal rates using multievent models

To circumvent the computational issues resulting from the exponential increase in parameter number of the Arnason–Schwarz model as states are added, Lagrange, Pradel, Bélisle, and Gimenez (2014) proposed a multievent model to estimate survival, dispersal and recapture probabilities while omitting the identity of sites. In multievent models, a distinction is made between events and states (Pradel, 2005). An event is what is observed in the field and thus coded in the individual capture history. This observation is related to the latent state (non-observable) of the individuals. Yet, observations can come with some uncertainty regarding the latent state. Multievent models aim at modelling this uncertainty in the observation process using hidden Markov chains. In their model, Lagrange et al. (2014) categorized the state of an individual in a given capture occasion as being in the same location as at $t−1$ or in a different location as at $t−1$. The states also include information about the capture status (captured or not) of the individual at $t−1$ and $t$. To date, this kind of multievent model has been used to quantify dispersal rates among numerous sites in birds (Lagrange et al., 2014, 2017) and amphibians (Cayuela et al., 2016; Denoël, Dalleur, Langrand, Besnard, & Cayuela, 2018).

This model provides accurate estimates of dispersal rates when the number of sites is large; it provides mean dispersal rates between all pairs of sites, contrary to the Arnason–Schwarz model that only provides pairwise dispersal rates. Its implementation in the E-SURGE program (Choquet, Rouan, & Pradel, 2009) also allows many possible model refinements (e.g., robust-design, trap-dependence) that have been developed for multistate models. Similar to the Arnason–Schwarz model, the Lagrange model is dedicated to estimating dispersal rate (and possibly distances, see Cayuela, Bonnare, & Besnard, 2018) in species occurring in discrete habitat patches or breeding sites and cannot be used to investigate dispersal in organisms occupying spatially continuous environments. Another limitation of this model is that it assumes that site characteristics and suitability do not vary over space and time, which appears to be an unrealistic assumption in many natural systems.

### 2.1.3 Estimating dispersal kernels using Fujiwara’s model

Dispersal kernels, the statistical distribution of dispersal distances in a spatially structured population, have been extensively used to
study dispersal (Nathan, Klein, Robledo-Arnuncio, & Revilla, 2012). They are probability functions (e.g., Gaussian, negative exponential, logistic) that describe the distribution of post-dispersal locations relative to the source point (Nathan et al., 2012). Fujiwara et al. (2006) first introduced a maximum-likelihood method to estimate dispersal kernels from CR data. Fujiwara’s model integrates three basic processes: dispersal, survival and sampling. Individuals are allowed to move freely in a one-dimensional space without any boundary. Dispersal is modelled as a density function \( k_d \) where \( d \) is the shape (Gaussian or Laplace) of the kernel. Survival is modelled as the probability \( \phi \) that an individual alive at time \( t \) is still alive at time \( t+1 \) and is independent of the location. The sampling process is modelled with a capture probability function \( \pi_i(x_t) \) giving the probability of capturing an individual conditional to its location \( x_t \) at time \( t \).

Fujiwara’s model was the first approach to estimate dispersal kernels assuming: The distribution of displacements is not always normally distributed, individuals can temporarily leave the study area and individuals may die during the study period. As with multistate and multievent models, Fujiwara’s model allows the examination of dispersal among discrete habitats patches or breeding sites. Contrary to later-developed spatially explicit CR models, this model does not allow quantifying dispersal kernels in spatially continuous environments. To our knowledge, Fujiwara’s model has not been used in further empirical studies, which might be due to the fact that this model has never been implemented in a user-friendly software.

### 2.1.4 Estimating dispersal kernels using spatially explicit CR models

Spatially explicit CR models are an extension of Cormack–Jolly–Seber CR models (Royle et al., 2018) and represent alternative approaches to fit dispersal kernels. Spatial CR models couple a spatiotemporal point process (Illian, Penttinen, Stoyan, & Stoyan, 2008) with a spatially explicit observation model. These models allow investigators to examine spatially explicit biological processes including density variation (Efford, 2011; Efford, Borchers, & Byrom, 2009), resource selection (Proffitt et al., 2015; Royle, Chandler, Sun, & Fuller, 2013) and dispersal (Ergon & Gardner, 2014; Royle, Fuller, & Sutherland, 2016) using encounter history data. Basically, these models assume that a population, composed of \( N \) individuals, is sampled and that each individual is associated with a spatial location that represents its activity centre expressed in \( x- \) and \( y- \)coordinates. The entire set of activity centres can be thought as the realization of point processes (Illian et al., 2008), a class of probability models for characterizing the spatial pattern and distribution of points. The activity centres are regarded as latent variables and are explicitly estimated along with other parameters of interest (e.g., probabilities of recapture, survival and dispersal) from the underlying point processes using marginal likelihood (Borchers & Efford, 2008) or Bayesian approaches using Markov chain Monte Carlo (Royle & Young, 2008).

Ergon and Gardner (2014) proposed a spatially explicit CR model that allows estimation of recapture and survival probability and fitting of dispersal kernels. The model integrates two basic parameters:

- \( \pi_i(x_t) \) = the capture probability of individual \( i \) in secondary session \( j \) within a primary session \( k \), which may depend on the latent activity centre of the individuals and the location of traps.
- \( \phi_{ik} \) = the probability that an individual \( i \) in the primary sampling period \( k \) survives to sampling period \( k + 1 \).

Dispersal is modelled as a shift in an individual’s activity centre. The dispersal process is described by individual dispersal direction \( (\theta_k) \) and distance \( (d_k) \) such that the change in the \( x- \) and \( y- \)coordinates of the activity centre is given by trigonometric functions. For \( d_k \), exponential, gamma and log-normal distributions, with zero-inflated versions for each of these distributions, can be considered in the model. The models can be fitted in the JAGS program using the R (R Core Team 2014) package rjags (Plummer, 2003).

To summarize, spatially explicit CR models are promising new tools to study dispersal distances. They allow fitting dispersal kernels using a great variety of distributions and are implemented in user-friendly \( R \) programs. Contrary to other capture-recapture models that estimate dispersal among discrete habitat locations, spatially explicit CR models allow quantification of dispersal in spatially continuous environments. They permit the use of individual detection data recorded using a variety of sampling methods including camera traps, acoustic sampling, non-invasive genetic sampling or direct physical capture.

### 2.2 Estimating non-effective dispersal rates and distances using genetic approaches

#### 2.2.1 Clustering and assignment approach

Genetic clustering analysis allows delineation of population boundaries by assigning individuals to discrete panmictic genetic clusters (Corander, Waldmann, & Sillanpää, 2003; Pritchard, Stephens, & Donnelly, 2000), sometimes with the use of geographic information (Caye, Jay, Michel, & Francois, 2017; Guillot, Estoup, Mortier, & Cosson, 2005; Guillot, Renaud, Ledevin, Michaux, & Claude, 2012), that can help to identify barriers to gene flow. These methods are valid when the species is effectively subdivided into discrete populations. In theory, \( F_0 \) migrants can be identified when individuals are well assigned to genetically differentiated groups. Non-effective dispersal rate is obtained by dividing the number of \( F_0 \)'s by the sample size (Broquet & Petit, 2009). While theoretically straightforward, several complications can arise (see also Broquet & Petit, 2009; Gagnaire et al., 2015). First, one needs to identify the number of discrete clusters present in the data, a task known to be difficult as statistically inferred clusters may be different from real populations and can be confounded by unsampled source populations (Falush, van Dorp, & Lawson, 2016; Pritchard et al., 2000). Second, isolation-by-distance (IBD) patterns are known to result in inflated signals of population clustering as most clustering methods assign individuals to discrete groups, assuming constant allele frequencies within each group and fail to take into account spatial autocorrelation in allele frequencies (Frantz, Cellina, Krier, Schley, & Burke, 2009; Meirmans, 2012; see
with NGS data, important concerns may arise as researchers tend to filter their data in ways that may not meet model assumption such as independence among loci. The use of minor allele frequency threshold filters can also introduce bias as rare variants contain information regarding population structure (Gravel et al., 2011; Mathieson & McVean, 2012). Other methods such as BayesAss (Wilson & Rannala, 2003), GeneClass2 (Piry et al., 2004), or BiMR (Faubet & Gaggiotti, 2008), are designed to estimate recent migration rate using MCMC and genotype data. The main limitations of these methods have already been reviewed elsewhere and include problems related to MCMC convergence, reduced accuracy with high numbers of populations and the need for moderate genetic differentiation ( \( F_{ST} \approx 0.05 \); Berry, Tocher, & Sarre, 2004; Paetkau, Slade, Burden, & Estoup, 2004; Hall et al., 2009; Faubet, Waples, & Gaggiotti, 2007; Broquet & Petit, 2009; Meirmans, 2014; Samarasing, Shuter, Wright, & Rodd, 2017). Given these limitations, such methods are not relevant for large population sizes and highly mobile species showing weak population structure (Gagnaire et al., 2015; Lowe 2003; Jones, Small, Paczolt, & Ratterman, 2010) and might not be appropriate for large-scale genomewide data sets. In contrast, the \( r \) package Assigner (Gosselin, Anderson, & Ferchaud, 2016) implements the methods of Anderson, Waples, and Kalinowski (2008) and Anderson (2010) and allows circumventing some of these limitations, such as low population differentiation (e.g., \( F_{ST} < 0.05 \)) while dealing with thousands of markers, for instance, using RADseq data. However, whether assignment results can be accurately translated to estimates of dispersal (m) remains to be investigated in more detail.

2.2.2 Parentage analysis and sibship reconstruction

Parentage analysis uses the genotypes of many individuals to identify parent–offspring relationships. It can be performed using exclusion methods where allelic mismatches are used to exclude individuals as possible parents of an offspring (Jones & Ardren, 2003; Jones, Small, Paczolt, & Ratterman, 2010; Marshall, Slate, Kruuk, & Pemberton, 1998). In most natural populations, it is impossible to sample all potential parents making exclusion-based approaches unreliable so maximum-likelihood or Bayesian methods are more commonly used to perform parentage analysis (Huisman, 2017; Jones & Ardren, 2003; Jones et al., 2010). Accurate assignments can be obtained from a small number of molecular markers: In general, optimal performances will be obtained with at least 15–20 polymorphic microsatellite markers (Jones et al., 2010) or at least 50–100 SNPs (Huisman, 2017). Nevertheless, it requires extensive sampling of all possible offspring (reviewed in Broquet & Petit, 2009; see Kamm et al., 2009 for an example). Parentage assignments, together with sibship reconstruction methods (reviewed in Wang & Santure, 2009; Wang, 2004, 2012; Städele & Vigilant, 2016; Blouin, 2003), allow estimating natal dispersal distances by measuring the geographic distance between parent and offspring spatial positions. These methods have also been used across numerous animal species to measure non-effective dispersal including insects (Fountain et al., 2018; Lepais et al., 2010), mammals (Burland, Barratt, Nichols, & Racey, 2001; Telfer et al., 2003; Waser, Busch, McCormick, & DeWoody, 2006), birds (Aguillon et al., 2017; Woltmann, Sherry, & Kreiser, 2012) and fishes (Almany, Berumen, Thorrold, Planes, & Jones, 2007; Almany et al., 2013, 2017; Jones, Planes, & Thorrold, 2005). Yet, despite their usefulness, parentage and sibship reconstruction analyses have several limitations. Specifically, they often require extensive sampling of offspring and potential parents to obtain accurate estimates, which can be infeasible when population size is large.

2.3 Estimating effective dispersal rates and distances using genetic approaches

2.3.1 Estimating migration rate \( m \)

\( F_{ST} \) as a biased estimator of migration rate

Traditionally, estimates of gene flow have been obtained under the island model introduced by Wright (1931). In this model, each population is made of the same, constant number of individuals \( N \) and receives and provides the same number of immigrants at a rate \( m \) per generation. Migration rates are thus symmetric and do not depend on geographic distance among populations (no IBD). The model also assumes that there is neither selection, nor mutation, and that migration–drift equilibrium is attained. Wright \( F \)-statistics allow measuring correlation of allele frequencies within and among such populations. In particular, Wright (1943) \( F_{ST} \) measures the variance of allele frequencies among populations (see review in Holsinger & Weir, 2009 and Alcaza & Rosenberg, 2017). Wright showed that, if all conditions of the islands models are met, then, \( F_{ST} \approx \frac{1}{4Nm} \) where \( N \) is the population size and \( m \) is the migration rate. Therefore, many researchers have used this relationship to estimate the product \( Nm \) as follows: \( Nm = \frac{1}{4} \left( \frac{1}{F_{ST}} - 1 \right) \). However, in reality, these conditions are rarely met. For instance, \( N_0/N \) ratios are known to be very far from one in nature (Frankham, 1995) and what is really measured is \( N_r m \) with \( N_r \) being the effective population size. However, obtaining accurate estimate of \( N_r \) is notoriously difficult (Charlesworth, 2009). In addition, a population may display very low \( F_{ST} \) due to large \( N_0 \) while being demographically independent (low \( m \)) from the other populations (Waples & Gaggiotti, 2006). The relationship between \( F_{ST} \) and \( N_r m \) is also affected by the mutation rate (\( \mu \)) and applies only when \( \mu \ll m \). While this could be a concern when the mutation rate is high, this should not be a problematic with SNPs data in which the mutation rate is lower. Details of the limitation this method have been reviewed in Whitlock and McCauley (1999) and Marko and Hart (2011). Given the many processes unrelated to gene flow that can result in high or low \( F_{ST} \) values, estimates of population connectivity based on \( F_{ST} \) alone are unlikely to be meaningful.

Coalescent approaches

The development of the coalescent theory (Kingman, 1982) has favoured the emergence of likelihood-based methods for inference
of population parameters. These methods can be exploited to directly assess the effective migration rate \( m \) or the product \( N_m \). It is noteworthy that in coalescence, \( m \) is scaled by the mutation rate, a parameter that is difficult to estimate (Ségurel, Wyman, & Przeworski, 2014) and reflects historical migration patterns over long time scales. Therefore, interpreting \( m \) might not reflect current levels of connectivity well. Earlier methods rely on coalescent theory and use MCMC to explore the space of genealogy (Beerli & Felsenstein, 1999, 2011). Then, isolation with migration (IM) models were developed (Wakeley, 1996) and implemented in the software IM (Nielsen & Wakeley, 2001) with various improvement to account for multiple loci, multiple demes, or to solve efficiently Felsenstein’s equation (Hey, 2010; Hey & Nielsen, 2004, 2007). These methods and their limitations have been reviewed elsewhere (e.g., Strasburg & Rieseberg, 2011). In general, they assume independence among loci, selective neutrality, free inter-locus recombination, no intra-locus recombination and migration–drift equilibrium. Violations of these assumptions have been shown to bias estimates of gene flow (Becquet & Przeworski, 2009; Strasburg & Rieseberg, 2011). False-positive rates were found when testing for the presence of migration using likelihood ratio tests in small data set (\( \sim 5\)–50 loci of 2,500 bp) and low divergence time (Cruickshank & Hahn, 2014) or small number of sample sites (Quinlin, Mayer, Elvinger, & Mardulyn, 2015). Two other important limitations of IM model are the assumption of a constant effective deme size and the inability to fit more complex and realistic models, including those with secondary contacts.

Recent model development has relaxed some of the previous assumptions using different variants of the IM model. For instance, it is now possible to include both asymmetric migration and variable population size (Costa & Wilkinson-Herbots, 2017). It is also now possible to infer complex histories using joint information from the blockwise site frequency spectrum and linkage disequilibrium (Beeravolu Reddy, Hickerson, Frantz, & Lohse, 2016) or to perform exact calculation of the joint allele frequency spectrum under an IM model using Markov chain representation of the coalescence (Kern & Hey, 2017).

Finally, approximate Bayesian computation (ABC) can be used to estimate the direction, symmetry and intensity of effective migration rate \( 4N_m m \) (Aeschbacher, Futschik, & Beaumont, 2013; Joseph, Hickerson, & Alvarado-Serrano, 2016; Moore et al., 2017; Rougemont & Bernatchez, 2018). ABC can be seen as a less rigorous, but very flexible, framework which potentially allows relaxing many assumptions made by the methods presented above (Beaumont et al., 2002; Ciliéry, Blum, Gaggiotti, & François, 2010). For example, Aeschbacher et al. (2013) used a two-step approach for inferring migration rate in Alpine ibex (Capra ibex). First, they estimated general population parameters (ancestral mutation rate and other more specific parameters). Second, they estimated migration between pairs of demes and showed how the accuracy of the pairwise approach increases with the number of parameters.

These methods mostly rely on the Kingman coalescent. Although this coalescent has been shown to be robust to departure from its major assumptions, it is not well suited if (a) the distribution of number of offspring among individuals is skewed (Eldon & Wakeley, 2006), (b) there are recurrent selective sweeps (Durrett & Schweinsberg, 2004), (c) sample sizes are larger than the effective population size (Wakeley & Takahashi, 2003), or (d) strong positive selection occurs (Spence, Kamm, & Song, 2016). Extensive efforts are currently being employed to develop more general classes of coalescent models (e.g., Spence et al., 2016) that relax some major assumptions of Kingman’s coalescent. In the light of these findings and other recent studies on the limit of demographic inferences based on the site frequency spectrum (Baharian & Gravel, 2018; Lapierre, Lambert, & Achar, 2017; Myers, Fefferman, & Patterson, 2008; Terhorst & Song, 2015), it is worth keeping in mind that low complexity models should be investigated first before testing more complex models with many parameters. With regard to these assumptions, more general classes of coalescent models appear very promising (Tellier & Lemaire 2014). In particular the spatial A-lambda-Fleming-Viot model (Barton, Etheridge, & Véber, 2010; Barton, Etheridge, & Véber, 2013; Etheridge, 2008; Etheridge & Véber, 2012), as pertaining to the multiple merger coalescent model (the A-Coalescent here), allows for the coalescence of more than two (multiple) lineages at a given generation. This model separately estimates dispersal distance \( \sigma^2 \) and the local population density \( \lambda \) and is not restricted to the study of the neighbourhood size \( 4\pi D \sigma^2 \) as most methods are (see section on IB above). This coalescent model was applied to a broad range of taxa including FLU virus (Guindon, Guo, & Welch, 2016), plants (Joseph et al., 2016) and humans (Ringbauer, Coop, & Barton, 2017) using blocks of identity-by-descent (see below).

### 2.3.2 Estimating effective dispersal distance

**IBD approaches to infer effective dispersal distance**

A widespread pattern observed in nature is the close genetic relatedness of individuals that are physically close to one another, and therefore, genetically distinct from geographically distant individuals (Vekemans & Hardy, 2004). This spatial autocorrelation generates a pattern of IBD in which individuals’ relatedness decreases with increasing geographic distance due to spatially limited dispersal (Malécot, 1948; Wright, 1943). Classically, IBD is tested by regressing linearized pairwise genetic distances (see link-based methods section for details), computed, for instance, as \( \Sigma_{i<j} \) against geographic distances (log-transformed in a two-dimensional habitat or untransformed in a one-dimensional habitat; Rousset, 1997, 2000). In a two-dimensional habitat, the slope of the regression is \( b = 1/(2D \pi) \), with \( 4\pi D \sigma^2 \) describing the “neighbourhood” size, where \( D \) represents the density of reproducing individuals and \( \sigma \) is the mean axial parent–offspring dispersal distance (Rousset, 1997, 2000; Sumner, Rousset, Estoup, & Moritz, 2001). When a direct estimate of population density is available, a non-trivial issue in structured metapopulations (Vekemans & Hardy, 2004), it becomes possible to infer \( \sigma \). This relationship holds when dispersal is homogeneous and spatially limited, when population density is homogeneous and when migration–drift equilibrium is reached (Rousset, 1997). Rousset (2000) then extended this approach at the individual level. Similar methods were
BOX 2 Quantifying dispersal and distance using pedigrees: the Florida Scrub-Jay case study

Recently, Aguillon et al. (2017) took advantage of an extensive data set for a single population of Florida Scrub-Jay including natal dispersal distance, sex, pedigree data and genotype data of almost all individuals in the population at more than 15,000 SNPs on autosomes and z-chromosomes. The geographic scale of the study was limited to ~10 km, providing an ideal setting to study the effects of recent dispersal at demographic equilibrium. Aguillon et al. (2017) first demonstrated limited and sex-biased dispersal of the Florida Scrub-Jay where half of the males only dispersed 488 m away from their parent's territories (territories are shown in Figure B) and half of the females dispersed less than 1,150 m (Figure A). Second, the authors estimated relatedness of individuals using identity-by-descent measures and clearly demonstrated sex-biased declines in identity-by-descent with distance, resulting in isolation by distance (Figure C). They then computed the distance (δ) where identity-by-descent diminishes halfway from its maximum value and found again greater isolation by distance (IBD) in males (δ = 620 m) than in females (δ = 903 m; Figure C). Thanks to the detailed pedigree information available, the authors then decomposed the effect of family relationship on IBD. For instance, in male–male comparisons, the highest IBD signal was apparently driven by short geographic distances between individuals from the highest pedigree classes, namely, parent–offspring, full-siblings, grandparent–grandchild, half-siblings and aunt/uncle-nice or nephew (Figures 2 and 3 in Aguillon et al., 2017). They also sequentially removed pedigree relationship classes and plotted the new IBD curves. While pattern of IBD softly decreases as classes were removed, the signal remained statistically significant even after removing all pairs with r ≥ 0.0625, indicating that if the strength of IBD was indeed driven by highly related individuals, the signal is also generated by dispersal events occurring at longer time scales. Interestingly, the authors showed similar IBD patterns in Z-linked markers. They then used their estimates of dispersal, population density and immigration rate to reconstruct sex-specific IBD patterns.

Although ideal for understanding the local process that generate dispersal, such studies will be hard to reproduce for other species given the amount of data needed and strict conditions required to observe local dispersal of individuals. Studies of this kind over larger spatial scales are nevertheless needed for conservation purposes and to gain insight into levels of connectivity between populations. For instance, the studied population of Florida Scrub-Jay is undergoing inbreeding depression due to decreased immigration rates (Aguillon et al., 2017; Chen, Cosgrove, Bowman, Fitzpatrick, & Clark, 2016).
also developed using kinship or autocorrelation statistics (Hardy & Vekemans, 1999; Loiselle, Sork, Nason, & Graham, 1995; Rousset, 2000; Vekemans & Hardy, 2004). While this method was used successfully in some of these studies, it relies on demographic equilibrium, which is often unrealistic (Leblois, Estoup, & Rousset, 2003; Leblois, Rousset, & Estoup, 2004) and can be confounded by ancestral structure (Meirmans, 2012). Therefore, the link between these estimates of dispersal and demographic connectivity is far from straightforward (Lowe & Allendorf, 2010). Importantly, inferences from IBD will perform best when sampling populations along regular grids, or regular networks, and when distances among samples are in accordance with the species dispersal ability, that is, in the order of σ (Leblois et al., 2003; Rousset, 2000; Vekemans & Hardy, 2004; Watts et al., 2007).

In addition to the direct inference of the mean axial squared parent-offspring dispersal rate, directional and non-directional Mantel correlograms (Borcard & Legendre, 2012; Oden & Sokal, 1986) may also be considered to assess the distance threshold at which the Mantel correlation becomes null, that is, the distance threshold below which allelic frequencies are positively autocorrelated and thus pairwise measures of genetic differentiation are smaller than expected by chance. It is a common mistake to interpret this distance threshold as an absolute estimate of the scale of gene flow (or as an upper estimate for effective dispersal distances) as it is primarily dependent on the considered sampling scheme, and more precisely, on the lag distance between sampling sites (Vekemans & Hardy, 2004). Nevertheless, Mantel correlograms may provide valuable information as to the relative spatial extent of gene flow across distinct genetic data sets (e.g., temporal or spatial replicates and age cohorts), provided they were gathered following a similar sampling scheme, and notably similar lag distances between sample sites. For instance, in a genetic study of the Florida Scrub-Jay (Aphelocoma coerulescens), Aguillon et al. (2017) found significant Mantel correlations at more distance classes in male-male pairs than in female-female pairs, a pattern consistent with the observed female-biased dispersal behaviour in this species. In Box 2, we detailed this study case, which combines both demographic and genetic data to refine our understanding of how restricted dispersed generates isolation by distance over very short distance. Although it was applied at a local scale, such approach could be deployed at larger spatial scales.

Cline analysis to measure effective dispersal distance
Cline theory provides an accurate framework for inferring dispersal distances (Barton, 1983; Lenormand, Guillemaud, Bourguet, & Raymond, 1998; Rieu, Lenormand, Carlier, de Lapeyre de Bellaire, & Ravigné, 2013; Sotka & Palumbi, 2006). At demographic equilibrium, if clines coincide and are more or less symmetric, then selection can be ignored and it is possible to infer dispersal σ such as σ = ωv√4R where ω is the cline width, R is the level of linkage disequilibrium and r is the recombination rate among loci (Barton, 1983). In most cases, however, selection coefficient must be estimated, which is not a trivial issue and different formulas will apply (but see Gagnaire et al., 2015). Numerous studies have described geographic clines of allelic frequencies, either falling along environmental gradients or along habitat boundaries (i.e., local adaptation clines; Sotka & Palumbi, 2006; Hare, Guenter, & Fagan, 2005; Galindo et al., 2010; Fabian et al., 2012; Bergland, Tobler, González, Schmidt, & Petrov, 2016; Van Woygaarden et al., 2017). These clines are often formed in secondary contact zones (Szymura & Barton, 1986). Recently, Gagnaire et al. (2015) suggested taking advantage of large genomewide data sets to identify selected and hitchhiker loci. They proposed using cline theory, either in the form of local adaptation clines, hitchhiking cline, hybrid clines or introgression tails, to infer patterns of connectivity in marine populations characterized by large effective population sizes and strong larval dispersal. In these populations, shallow levels of genetic differentiation make most traditional methods relying on neutral model inefficient. Importantly, identification of relevant outlier loci can be confounded by demographic factors (e.g., bottleneck, expansion, admixture), variation in local recombination rate, shared ancestral polymorphism and polygenic selection making their identification a non-trivial issue (Gagnaire et al., 2015; Hoban et al., 2016; Vitti, Grossman, & Sabeti, 2013). Nevertheless, the method advocated by Gagnaire et al. (2015) is highly promising and could be extended to species with sufficiently large effective population sizes and where natural selection is expected to be strong.

Using identity-by-descent blocks to infer recent demography
Another promising approach with the increased availability of whole genome sequencing data or other very dense polymorphism data (e.g., RADseq or high-density SNP chip data) is the analysis of the length of haplotype blocks (Gravel, 2012; Pool & Nielsen, 2009). Individuals immigrating into a new (genetically differentiated) population will transmit chromosomes that are broken down by recombination, with block size being gradually reduced with each generation of hybridization. Therefore, these long admixture tracts can provide information regarding recent migration rates (Liag & Nielsen, 2014). In the same vein, identity-by-descent segments, which are blocks of haplotypes inherited from a common ancestor by pairs of individuals (reviewed in Browning & Browning, 2012), have been used to infer recent migration rates (Palamara & Pe’er, 2013). Again, these segments are delimited by recombination history and the longer the segment, the more recent the migration event. In particular, Harris and Nielsen (2013) and Palamara and Pe’er (2013) developed theoretical expectations of the distribution of identity-by-descent under different demographic scenarios in the presence of migration. More recently, Ringbauer et al. (2017) derived a promising approach relying on diffusion approximation to infer patterns of isolation by distance of long identity-by-descent blocks. The model allows identifying population effective density (D) and dispersal rate σ² separately, thus overcoming the limitation of classical FST-based measures of isolation by distance. This scheme can account for changes in population density and the geographic spread of ancestry.
TABLE 1  Demographic and genetic methods to investigate environmental and individual variables on the three stages of non-effective dispersal

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Approach</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emigration/immigration</td>
<td>Individual state</td>
<td>Demography</td>
<td>Temporally fixed individual variables can be introduced as external covariates in multistate and multievent models. Temporally varying individual variables can be coded as states in multievent models; they are therefore introduced as discrete variables in the models. Genetic Parentage analysis and assignment method may permit to link emigration/immigration and individual variables. Environmental/social Demography Multistate models allow to compare groups of sites (e.g., small vs. large) by constraining model parameters. Spatiotemporally variable site characteristics or variation in the social context can be modelled using multievent models. Environmental and social information is coded as states in the model. Genetic Parentage analysis and assignment methods may permit to link emigration/immigration and environmental and social information. Individual genotyping (and relatedness measurement) of dispersers and residents can provide information about inbreeding avoidance, kin competition and individual fitness.</td>
</tr>
<tr>
<td>Transience</td>
<td>Individual state</td>
<td>Demography</td>
<td>Spatially explicit CR models and Ovaskainen’s diffusion models can incorporate individual covariates (fixed or time-specific). Genetic Parentage analysis can be used to assess relationships between parent–offspring dispersal distance and individual variables. Assignment methods may also be used to examine correlations between dispersal distances and individual factors. Environmental Demography Multistate model estimates can be used in ad hoc analyses to examine the effect of Euclidean and environmental distances on between-site dispersal probability. Multievent models can be used to examine the effect of physical barriers in the landscape on immigration probability. Ovaskainen’s diffusion models can be used to investigate landscape composition and configuration on movement path. Genetic Parentage analysis, assignment methods and link-based methods can be used to assess relationships between dispersal distance and landscape characteristics.</td>
</tr>
</tbody>
</table>

(assuming uniform diffusion). While all these methods rely on a good reference genome and high-quality genotype data for identity-by-descent segment inferences, they are promising in that they allow inference of very recent demographic events relevant to analyse contemporary genetic connectivity.

3 | HOW TO INFER ENVIRONMENTAL AND INDIVIDUAL EFFECTS ON NON-EFFECTIVE AND EFFECTIVE EMIGRATION AND IMMIGRATION?

Demographic and genetic methods used to study the influence of individual and environmental factors on effective and non-effective emigration and immigration are presented in Figure 3 and Table 1.

3.1 | Examining non-effective emigration and immigration using demographic approaches

3.1.1 | Disentangling emigration and immigration using multistate and multievent models

In the classical version of the Arnason–Schwarz model, the dispersal parameter $\psi_{RT}$ includes both emigration from site $R$ and immigration to site $T$. This formulation remained a limiting factor for dispersal studies for a long time as variables influencing emigration and immigration may differ or can differently affect both processes. For this reason, Grosbois and Tavecchia (2003) introduced a new parameterization of the Arnason–Schwarz model where $\psi_{RT}$ is decomposed into two distinct parameters:

\[
\pi_T = \text{the probability that an individual that was in site R at capture occasion t−1 emigrates at occasion t provided it survives.}
\]

\[
\mu_T^{RT} = \text{the probability that an individual that was in site R at capture occasion t−1 immigrates to site T at capture occasion t provided it survives and emigrates.}
\]

This parameterization was subsequently used in many studies of birds (Fernández-Chacón et al., 2013; Lok, Overdijk, Tinbergen, & Piersma, 2011; Péron, Crochet, Doherty, & Lebreton, 2010; Péron, Lebreton, & Crochet, 2010) and mammals (Devillard & Bray, 2009).

In the context of multievent models, Tournier, Besnard, Tournier, and Cayuela (2017) recently modified the structure of Lagrange models to separately estimate emigration and immigration probabilities. These extensions in the framework of both multievents and multistate models were important methodological developments allowing the study of the different steps of the dispersal process.
In a recent study, Denoël et al. (2018) investigated how the interplay between individual and environmental factors may lead to alternative dispersal strategies that, in turn, lead to the coexistence of contrasting site fidelity phenotypes. They addressed this issue in a pond-breeding amphibian, the great crested newt (Triturus cristatus, Figure C). They used a modified version of the Lagrange multievent CR model that includes heterogeneity mixtures. By doing so, they were able to assess if alternative breeding site fidelity phenotypes (i.e., lowly site faithful (LSF) vs. highly site faithful (HSF) individuals) could coexist within the studied spatially structured populations. In a first analysis, they showed that the probability of staying in the same breeding site between each time step depended on individual site fidelity status at \( t-1 \). The probability of remaining in the same breeding site was higher in individuals that were already site faithful at \( t-1 \).

In a second analysis, they highlighted that two distinct site fidelity strategies occurred in the population and that individuals belonging to each strategy differed in terms of phenotypic and life history traits. At both intra- and inter-annual scales, the site fidelity probability was always 1 in the HSF phenotype, while this probability fluctuated greatly over time in the LSF phenotype at both intra-annual (Figure A) and inter-annual levels (Figure B). In both HSF and LSF phenotypes, survival increased with body size. Yet, the HSF phenotype was characterized by a lower survival probability than the LSF phenotype (odd ratio HSF/LSF: 0.63; 95% CI: 0.43–0.94). The study also demonstrated that the probability of being assigned to the LSF phenotype depended on both intrinsic and extrinsic factors (i.e., sex, body size and pond surface). Males had a higher probability of belonging to the LSF phenotype than females (odd ratio: 2.62; 95% CI: 1.55–4.42), and, in both sexes, this probability increased with body size (Figure D and E). Furthermore, the probability of being in the LSF phenotype slightly increased with the mean surface of the pond occupied by the individuals during the three-year study period (Figure F and G).

The study of Denoël et al. (2018) illustrates the usefulness of multievent CR models to investigate dispersal syndromes in natural populations.
3.1.2 | Detecting inter-individual variation and dispersal strategies within populations

Multistate and multievent models examine temporal autocorrelation of the dispersal behaviour (memory models) as state–state transitions modelled as first-order Markovian processes. These models assess the repeatability of the individual’s dispersal behaviour, a central component in the framework of animal personalities (Bell, Hankison, & Laskowski, 2009; Stamps & Groothuis, 2010), by testing if the probability of dispersal at time $t$ depends on the individual dispersal state at $t−1$ (see, e.g., Denoëll et al., 2018 and Péron, Lebreton et al., 2010). Moreover, individual heterogeneity of unknown source in dispersal behaviour can be detected using capture-recapture mixture models with discrete classes of individuals (Pledger, Pollock, & Norris, 2003). Using such models, Denoëll et al. (2018) highlighted a dispersal syndrome implicating a covariation pattern between dispersal tendency, survival and body size (Box 3). Individual heterogeneity of dispersal probability can also be accommodated using random effects in multistate and multievent models (Gimenez & Choquet, 2010).

Multistate models have been used extensively to test the effects of individual variation on dispersal probabilities. In different studies, authors have extended the parameters’ space of the Arnason–Schwarz model to test the effects of age on dispersal probabilities (e.g., Blums, Nichols, Hines et al., 2003; Breton, Diamond, & Kress, 2006; Lebreton et al., 2003). For instance, Blums, Nichols, Hines et al. (2003) proposed a multistate model to quantify natal (i.e., pre-breeding) dispersal, breeding dispersal and age-dependent survival. Age classes were coded as states in the model. At the same time, Lebreton et al. (2003) introduced a class of age-dependent multistate CR models for the simultaneous estimation of natal dispersal, breeding dispersal and age-dependent recruitment. The effects of individual variation on dispersal have also been tested by including time-constant individual covariates (e.g., body condition, body mass and tarsus length; Barbraud, Johnson, & Bertault, 2003) and group effects (e.g., sex; Tavecchia, Pradel, Lebreton, Biddau, & Mingozzi, 2002). Other studies used the parameterization proposed by Grosbois and Tavecchia (2003) to examine the effect of individual age on emigration probabilities (Lok et al., 2011; Péron, Lebreton et al., 2010), again by coding age classes as states in the model. Moreover, a recent study used multievent models to deal with high number of sites (the parameterization of Tournier et al., 2017) to examine age-specific emigration rates by considering three age classes (two pre-breeding and one breeding states) in the model (Cayuela, Bonnaire et al., 2018).

3.1.3 | Detecting the influence of environmental factors on emigration and immigration

The effect of patch characteristics (e.g., past reproductive success, colony size, habitat management, predation risk) on dispersal probability has also been extensively studied using Arnason–Schwarz model (Blums, Nichols, Lindberg, Hines, & Mednis, 2003; Blums, Nichols, Hines et al., 2003; Cam et al., 2004; Dodd et al., 2006; Grant et al., 2010; Péron, Crochet et al., 2010; Spendelow et al., 2016). Other studies used the parameterization proposed by Grosbois and Tavecchia (2003) to examine the environmental effects on emigration probability (Fernández-Chacón et al., 2013; Péron, Lebreton et al., 2010; Péron, Crochet et al., 2010). In both cases, site-specific variation is coded as states in the models (e.g., small vs. large colonies). Moreover, time-specific factors uniformly affecting all studied sites (e.g., weather condition, sea level) are incorporated as continuous covariates in the models. Note that the existing multistate models do not allow examining the effects of spatiotemporally variable factors (e.g., variation in conspecific or heterospecific densities).

Recently, Cayuela et al. (2017) circumvented this issue by extending the parameters’ space of Lagrange’s model, which assumes that all sites have spatially and temporally homogenous proprieties. First, Cayuela et al. (2017) proposed a generalized version of the Lagrange model that allows survival and movement probabilities to differ according to temporarily fixed site characteristics. The states of the model include those of Lagrange model (captured or not at $t−1$ and $t$; move or stay between $t−1$ and $t$), and additional states correspond to site proprieties (A and B in the case of a two-habitat model). In subsequent work, Cayuela, Pradel et al. (2018) generalized their two-habitat models by considering situations where the state of sites is not constant over time. They developed two new multievent CR models that allow the estimation of emigration, immigration, survival and recapture probabilities when a site may appear or disappear over time or when the characteristics of sites change over space and time.

3.2 | Examining non-effective emigration and immigration using genetic approaches

Genetic approaches that identify both the origin and the destination of dispersing individuals, such as parentage analyses or assignment methods, can be used to infer the influence of individual phenotypes and local environmental characteristics on relative emigration and immigration rates. Parentage analyses, in particular, allow direct measurement of dispersal. For instance, parentage analysis and sibship reconstruction provide information regarding sex-biased dispersal (e.g., Aguillon et al., 2017; Biek et al., 2006; Prugnolle & de Meeus, 2002; van Dijk, Covas, Doutrelant, Spottiswoode, & Hatchwell, 2015; van Hooft, Keet, Brebner, & Bastos, 2018), the fitness of the dispersers (van Hooft et al., 2018), inbreeding avoidance (Nelson-Flower, Hockey, O’Ryan, & Ridley, 2012) and kin competition (Perrin & Mazalov, 2000). Regarding the inference of sex-biased dispersal, when phenotypic or fitness data are measured for all individuals for which dispersal data have been collected, it is then possible to draw correlation between emigration/immigration and these traits.
3.3 Examining effective emigration and immigration using genetic approaches

3.3.1 Detecting the influence of environmental factors on emigration and immigration

Emigration and immigration rates are directly impacted by the way dispersal individuals perceive and interpret environmental characteristics of local patches in the light of their own phenotypic and genotypic characteristics (Edelaar & Bolnick, 2012). For instance, Saint-Pé et al. (2018) showed that, in a brown trout (Salmo trutta trutta) population from a small French watershed stocked for decades, both the emigration propensity and the dispersal destination of individuals were influenced by their individual level of genetic admixture with the domestic strain. Nevertheless, when contemporary individual dispersal events cannot be identified, genetic approaches relying on the spatial distribution of allelic frequencies (e.g., measures of genetic differentiation) must carefully dissociate the influence of between-site landscape processes (geographic distance and/or landscape resistance) impacting the transience phase from the influence of at-site landscape processes (local carrying capacity, patch quality, etc.) impacting the emigration and immigration phases. To that aim, Murphy, Dezzani, Pilliod, and Storfer (2010) introduced gravity models (Fotheringham & O’Kelly, 1989) in which gene flow (modelled along a spatial network as a genetic distance) is explained by both at-site data (characteristics of network nodes in terms of production/attraction of propagules) and among-sites data (characteristics of network edges in terms of geographic distance and landscape resistance). Applied to a spatially structured population of Columbia spotted frogs (Rana luteiventris), they identified a negative effect of predation (locally decreasing immigration) and a positive effect of site productivity (locally increasing emigration) on gene flow along with a more classical negative effect of geographic distance, affecting the transience phase. Contrary to gravity models, other analytical procedures using location-specific landscape data (landscape genetics node-based methods; see Balkenhol, Waits, & Dezzani, 2009 for a review) do not explicitly consider the possible influence of between-patch processes on the distribution of neutral genetic variation (apart from spatial autocorrelation) and are thus only meaningful when studying adaptive genetic variation (Wagner & Fortin, 2013, 2015).

3.3.2 Emigration, immigration and isolation by environment

One may also consider the use of metrics of isolation by environment (IBE) that can be compared to pairwise measures of genetic differentiation in addition to classical metrics of IBD, IBB (isolation by barriers) and/or IBR (isolation by resistance) using various linked-based statistical tools (e.g., McRae, 2006; Bradburd, Ralph, & Coop, 2013; see next section for a review of other available statistical tools). IBE is defined as a pattern in which genetic differentiation increases with environmental differences, independent of geographic distance (Wang & Bradburd, 2014). IBE may arise from local (natural or sexual) selection against immigrants, but also from non-random dispersal resulting from individual habitat choice or local adaptation (Bolnick & Otto, 2013). For instance, Wang, Glor, and Losos (2013) showed that IBE explained 17.9% of variance in genetic divergence in 17 species of Anolis lizards from Greater Antillean islands, whereas 36.3% of variance was explained by classical IBD. In a meta-analysis by Sexton, Hangartner, and Hoffmann (2014), 74.3% of surveyed studies showed significant IBE.

4 HOW TO INFER THE ENVIRONMENTAL AND INDIVIDUAL EFFECTS ON NON-EFFECTIVE AND EFFECTIVE TRANSIENCE?

Demographic and genetic methods used to study the influence of individual and environmental factors on effective and non-effective transience are summarized in Figure 3 and Table 1.

4.1 Examining non-effective transience using demographic approaches

4.1.1 Studying transience using multistate and multievent models

Transience cannot directly be studied using multistate and multi-event CR models because the pathway used by the individual within the landscape matrix is ignored; these models allow estimating transitions between discrete sites. Yet, several studies attempted to examine transience in an indirect way, by modelling immigration probabilities according to the Euclidean distances between sites or the presence of physical barriers in the landscape matrix. For instance, Péron, Lebreton et al. (2010) as well as Fernández-Chacón et al. (2013) investigated how Euclidean distances among sites affect dispersal probabilities. In both studies, a post hoc analysis retrieved immigration estimates from the selected model and assessed the effect of between-sites distance on immigration using a generalized least squares approach. Moreover, in a recent study, Cayuela, Bonnaire et al. (2018) examined the effect of physical barriers (i.e., different kinds of roads) between sites on immigration probability by extending the parameters’ state-space of the multievent model proposed by Tournier et al. (2017). In this model, individuals may immigrate into sites given they crossed a physical barrier or not. To assess the significance of the physical barrier effect on transience, Cayuela, Bonnaire et al. (2018) compared the conditional immigration probability extracted from the best-supported models to the probability of reaching a site given individuals crossed an obstacle or not using a random dispersal hypothesis (i.e., the mean probability of arriving in a patch calculated from all the individuals occurring in all patches of the study area).
4.1.2 | Ovaskainen diffusion model

Ovaskainen (2004) developed a model that quantified movement behaviour of an individual in a heterogeneous landscape. In this model, the landscape is classified into a finite number of habitat types that differ from each other in terms of the individual’s movement behaviour. Three components are considered in the model: movement within a habitat type, behaviour at edges between habitat types and mortality. Movement within a habitat type is assumed to follow a random walk or a correlated random walk, which can be approximated by diffusion (Patlak, 1953). When deriving the diffusion approximation, the distributions specifying the random walk model aggregate into a single variable $D_i(t)$. As an individual’s response to boundaries between habitat types may be a major determinant of dispersal patterns (Crone & Schultz, 2008; Schtickzelle & Baguette, 2003; Schultz & Crone, 2001), the Ovaskainen model also accounts for edge-mediated behaviour. It assumes that an individual may bias its direction towards either of the habitat types when it is close to a boundary. In the diffusion approximation, such bias leads to a discontinuity in the probability density for the individual’s location (Ovaskainen & Cornell, 2003). In the model, this bias is quantified by habitat-specific multiplier $k_i$. Moreover, the model also accounts for individual mortality by assuming that an individual located in habitat type $i$ has a mortality probability $\mu_i(t)$ at time $t$.

Ovaskainen et al. (2008) extended the maximum-likelihood estimation scheme of their previous model to a more flexible Bayesian framework, using a MCMC method. They implemented the estimation schemes in two software programs: Mapper, a GIS-based interface for the triangulation of the layers representing the landscape and the set of mark-recapture sites; and Disperse, a program including the adaptive MCMC methods for parameter estimation. Overall, Ovaskainen’s model is of great interest because it separates the mortality occurring during the transience phase of the dispersal process and the mortality resulting from other activities (e.g., breeding; Ovaskainen et al., 2008). The model can also be used to examine the effect of landscape composition and configuration on the transience phase/process. Despite its flexibility and robustness, Ovaskainen’s model has been rarely used in empirical studies (Arelano, León-Cortés, & Ovaskainen, 2008; Ovaskainen, 2008; Ovaskainen et al., 2008; Wang et al., 2011), likely because it remains computationally intensive.

4.2 | Examining non-effective transience using genetic approaches

The study of transience based on indirect estimates of dispersal is rarely performed although landscape genetic approaches, as detailed below, could in principle be applied to estimates of dispersal obtained from assignments tests or parentage analysis (Castilla et al., 2017; Kamm, Gugerli, Rotach, Edwards, & Holderegger, 2010). Other approaches to measure connectivity rely on the combination of parentage analysis and dispersal kernels, and therefore estimate the distribution of dispersal distance. Such approaches, initially developed in plant seed dispersal (Nathan & Muller-Landau, 2000), are now used for animals with complex life cycles and dispersal patterns (e.g., Almany et al., 2017; Buston, Jones, Planes, & Thorrold, 2012; D’Aloia, Bogdanowicz, Majoris, Harrison, & Buston, 2013; Ismail et al., 2017). In general, these methods require very extensive sampling (e.g., Almany et al., 2017). Although these models do not directly infer the effect of landscape configuration on transience, future development could integrate GIS data accounting for the landscape configuration when fitting the kernel distribution and modelling dispersal. At least, information on the fragmentation of the landscape or connectivity levels in the sea (measured, e.g., by current forces and directionality) could be correlated to the estimated distribution kernels.

4.3 | Examining effective transience using genetic approaches

Both boundary-based and link-based landscape (see below) genetic procedures can infer the influence of between-patch processes (e.g., fragmentation, land cover conversion and climate change) on effective transience phase of dispersal. In most cases, boundary-based and link-based landscape genetic methods rely on the assumptions that both genetic drift and gene flow are spatially random: All patches (when relevant) are assumed to be of similar size and quality (Prunier, Dubut, Chikhi, & Blanchet, 2017) and all genotypes are considered equally likely to immigrate, disperse and settle (Edelaar & Bolnick, 2012).

4.3.1 | Boundary-based methods

Boundary-based methods include edge detection techniques (e.g., Cercueil, François, & Manel, 2007; House & Hahn, 2017; Jombart, Devillard, & Balloux, 2010; Jombart, Devillard, Dufour, & Pontier, 2008; Monmonier, 1973; Piry et al., 2016), and Bayesian clustering algorithms (reviewed in François & Durand, 2010) aim to delineate discrete or admixed populations in space (Wagner & Fortin, 2013). They allow identifying spatial genetic boundaries that are visually (e.g., Frantz et al., 2012; Prunier et al., 2014) or statistically (e.g., Balkenhol et al., 2014; Jay et al., 2012; Murphy, Evans, Cushman, & Storfer, 2008) compared to landscape patterns. Although sensitive to sampling design (e.g., Puechmaille, 2016) or IBD (e.g., Safner, Miller, McRae, Fortin, & Manel, 2011), boundary-based methods are an effective way of exploring the influence of landscape configuration on genetic structure, and ultimately, on spatial patterns of gene flow among patches.

4.3.2 | Link-based methods

Link-based landscape genetic methods address the question of how likely gene flow is between two patches considering the
spatiotemporal heterogeneity of the environment (Wagner & Fortin, 2013). They are based on the statistical comparison between pairwise measures of inter-individual or inter-population genetic differentiation and pairwise landscape distances quantifying the connectivity among sampling locations.

Pairwise landscape distances, considered as predictors in subsequent statistical approaches, quantify the connectivity among sampling locations under the non-exclusive assumptions of IBD (e.g., geographic of riparian distances), isolation by barriers (e.g., presence or number of barriers) and more generally isolation by resistance (IBR). Many procedures have been proposed to compute IBR pairwise landscape distances. Most of them rely on the initial design of one or several resistance surfaces, representing the extent to which the conditions at each grid cell are expected to constrain movement.

**BOX 4**  Identifying the spatial scales at which processes underlying genetic structures should be investigated: the Fenno-Scandinavian brown bear case study

The genetic structure of a population may be shaped by multi-level processes (e.g., sex-biased dispersal, social behaviours, territoriality) each acting at a given spatial scale. In a recent study, Schregel et al. (2018) proposed a simple analytical framework based on semivariogram analysis to identify the possible multiple spatial scales at which processes underlying the genetic structure are to be considered. A semivariogram depicts the spatial autocorrelation of measures at sample points (Figure A). The distance at which the semivariance flattens out is the range. Measures at sample locations separated by distances smaller than the range are spatially autocorrelated, whereas measures at locations separated by distances larger than the range are not. The semivariance at the range is called the total sill, while the semivariance at distance 0 is called the nugget and is due to background noise in the data (e.g., measurement errors). The partial sill is defined as the total sill minus the nugget. The relative partial sill (partial sill/total sill) provides the strength of the spatial autocorrelation in the data. The first step in the framework by Schregel et al. consists in getting the residuals from an IBD model, depicting the relationship between genetic and spatial distances. Then, a series of non-overlapping distance classes are defined (e.g., from 0 to the maximal recorded distance). For each distance class, the corresponding subset of residuals is used in a semivariogram analysis and the corresponding relative partial sill value, providing the local strength of the spatial autocorrelation in residuals at that distance class, is recorded. When values of relative partial sill are plotted against distance classes, local peaks indicate the spatial scales at which the population structure exhibits maximal strength. Applying this framework to an empirical brown bear (Ursus arctos) data set consisting in 1530 geo-referenced genotypes from Norway and Sweden, Schregel et al. identified two main peaks at different spatial scales, suggesting that two independent biological processes may shape the Scandinavian brown bear’s population genetic structure (Figure B): (a) a local process, to be analysed at the scale of the home range (<35 km) and affecting genetic patterns on a short-term time, (b) a large-scale process associated with sex-biased gene flow over multiple generations and responsible for population subdivision at scales >98 km. From these insights, they could then perform an informed scale-explicit analysis of the genetic structure in brown bears, finding that one of the four genetic clusters identified in central Scandinavia by STRUCTURE was actually caused by IBD.

![Figures adapted from Schregel et al. (2018). Figure A: Main characteristics of a semivariogram. Figure B: Strength of IBD-based population structure over geographic scales from 1 to 350 km, for each sex separately, and for combined data.](image)
or gene flow (Spear, Balkenhol, Fortin, Mcrae, & Scribner, 2010; Spear, Cushman, & McRae, 2015; but see Emaresi, Pellet, Dubey, Hirzel, & Fumagalli, 2011). The best approach to parameterize resistance surfaces is still debated (Milanesi et al., 2016; Zeller, McGarigal, & Whiteley, 2012). Parameterization may rely on expert opinion and/or on empirical evidence ensuing from experimentation (e.g., Stevens, Verkenne, Vandewoestijne, Wesselingh, & Baguette, 2006), movement pathway data (e.g., Reding, Cushman, Gosselink, & Clark, 2013) or species distribution models (e.g., Shafer et al., 2012; Wang, Yang, Bridgman, & Lin, 2008). The parameterization procedure may also incorporate an optimization step, notably allowing the calibration of non-linear responses of resistance values to original map pixel values (e.g., Peterman, Connette, Semlitsch, & Eggert, 2014; Wasserman, Cushman, Schwartz, & Wallin, 2010). Three major modelling frameworks can then be used to convert resistance surfaces (or original landscape maps) into pairwise measures of landscape connectivity: least-cost path modelling, circuit-based modelling and transect-based approaches. Least-cost path modelling (Adriaensen et al., 2003) assumes that organisms have enough knowledge of the whole landscape to follow the ideal path, whereas circuit-based modelling (McRae, 2006) incorporates many possible paths into the final measure of resistance, but both rely on resistance surfaces and thus incorporate the influence of landscape configuration on transience trajectories to provide pairwise matrices of landscape distances (Spear et al., 2015). Transect-based approaches use buffers of various widths and drawn along straight-lines (e.g., Emaresi et al., 2011) or least-cost paths (e.g., Van Strien, Keller, & Holderegger, 2012) to calculate the abundance of each original landscape feature. Although both approaches are primarily designed to quantify landscape composition between pairwise locations, the advantage of least-cost transects over straight-line transects is that they also allow taking landscape configuration into consideration (Van Strien et al., 2012).

Pairwise genetic distances (e.g., $F_{ST}$, $D_A$), considered as the quantity of interest in subsequent statistical approaches, may be computed at the population or at the individual level from neutral genetic data (Balkenhol & Fortin, 2015). Pairwise measures of genetic differentiation are usually computed from allele frequencies at the population-level and from genotypes at the individual level. In both cases, numerous metrics may be considered (Waits & Storfer, 2015). Additionally, individual-based genetic distances may also be calculated from transformed genetic data, such as individual ancestry values computed from clustering algorithms (e.g., Balkenhol et al., 2014; Murphy et al., 2008). Using populations as the sampling unit is a classical approach in landscape genetics, although it involves a challenging a priori decision about the putative delineation of demes, because the real spatial distribution of individuals may not be strictly discrete (Manel, Schwartz, Luikart, & Taberlet, 2003). Furthermore and as previously stated, the commonly used F-statistics are primarily measures of the balance between genetic drift and gene flow (not to mention mutation). Thus, they may only be considered as proxies for gene flow under the strict assumption of equal effective population sizes, demographic equilibrium (an assumption rarely met in postglacially established populations) or after having ruled out the possible influence of spatial heterogeneity on effective population sizes (Prunier, Dubut et al., 2017). Alternatively, using individuals as the sampling unit is adequate in the case of continuously distributed organisms and, in the case of patchily distributed organisms, it may provide more flexibility in the design of the sampling scheme (extent of the study area, lag distance between samples, etc.; Anderson et al., 2010) without specific loss of power for detecting isolation patterns (Luximon, Petit, & Broquet, 2014; Prunier et al., 2013).

### 4.3.3 | Link-based statistical models

Numerous linked-based statistical models have been proposed to relate pairwise genetic distances to pairwise landscape distances, while dealing with the inherent non-independence of pairwise data. They notably include linked-based linear models, such as simple and partial Mantel tests (Legendre, 2000), multiple regressions on distance matrices (Smouse, Long, & Sokal, 1986) and maximum-likelihood population-effects models (MLPE; Clarke, Rotbery, & Raybould, 2002; Selkoe et al., 2010; Van Strien et al., 2012), as well as linked-based non-linear models based on Bayesian approaches (e.g., Bradburd et al., 2013; Faubet & Gaggiotti, 2008). Note that the validity of the Mantel statistical test has been questioned and that it should be used with caution (see Diniz-Filho et al., 2013 for a review). Because these statistical approaches are in essence correlative, these linked-based linear models may be deeply flawed by multicollinearity among predictors, a major issue in landscape genetics (Prunier, Colyn, Legendre, Nimon, & Flamand, 2015; Wagner & Fortin, 2015). Multicollinearity, and notably statistical suppression situations (Paulhus, Robins, Trzesniewski, & Tracy, 2004), may obscure the interpretation of multivariate regressions through artefactual increases in regression coefficients and possible sign reversal (Prunier et al., 2015). The removal of redundant predictors, the creation of orthogonal synthetic predictors and/or the use of regression commonality analyses are different strategies that may be deployed to correct for multicollinearity (Dormann et al., 2013; Prunier, Colyn, Legendre, & Flamand, 2017; Prunier, Dubut, Loot, Tudesque, & Blanchet, 2018). Linked-based causal models (e.g., Cushman, McKelvey, Hayden, & Schwartz, 2006; Fourtune et al., 2018; Wang et al., 2013) are a promising alternative to the previously described models as they allow inferring causal relationships among the genetic response and landscape predictors beyond simple correlations, although they may also be sensitive to collinearity. Finally, most linked-based statistical models may be subject to model selection procedures based on the comparison of model fit parameters (e.g., Keller, Holderegger, & van Strien, 2013) or Akaike information criterion (Burnham & Anderson, 2002; but see Prunier et al., 2015; Franckowiak et al., 2017).

It is also crucial to consider the temporal and spatial scales of the ecological processes under study when performing linked-based statistical analyses, as it has been shown that the drivers of ecological processes may each act at unique scales in space and time (Anderson et al., 2010; Wiens, 1989). For instance, most statistical analyses may fail to detect significant relationships between genetic and landscape data when the rate of landscape change is faster than
can be resolved by common molecular markers (Anderson et al., 2010). A solution may consist in confronting genetic data with both historical and contemporary landscapes (e.g., Pavlacky, Goldizen, Prentis, Nicholls, & Lowe, 2009). Computing genetic distances from individual genotypes rather than from population-based allelic frequencies may also be worth considering, as metrics based on allelic frequencies may suffer from a loss of resolution due to the averaging of genetic information over individuals (Landguth et al., 2010; Prunier et al., 2013). Additionally, one may couple these approaches with genetic simulations to calculate the theoretical temporal lag in the expected genetic response to landscape change, given the data at hand (e.g., Prunier et al., 2014). Similarly, most statistical analyses may fail to detect significant relationships between genetic and landscape data when the grain size, the spatial extent of the study area or the lag distance between sampling points do not match the spatial scale of the considered ecological process (Anderson et al., 2010). For instance, restricting landscape genetic analysis to smaller scales (0–3 km) and neighbouring populations as defined by a population network provided the highest model fit in a landscape genetic study of the wetland grasshopper Stethophyma grossum (Keller et al., 2013). Similarly, Schregel et al. (2018) developed a simple statistical framework based on variogram analysis allowing the identification of the multiple spatial scales at which various biological processes have the highest influence on patterns of genetic structure (Box 4).

5 | COMPLEMENTARITY OF DEMOGRAPHIC AND GENETIC APPROACHES TO STUDY DISPERSAL

5.1 | Estimating non-effective dispersal rates and distances using demographic and genetic methods

Demographic and genetic methods can be used to estimate non-effective dispersal rates and distances, but they differ in terms of information gathered, sampling effort and technical constraints. Demographic surveys provide time-specific (i.e., monthly, yearly) dispersal distances and rates, while genetic approaches (i.e., assignment methods, sibship reconstruction and parentage analyses) provide a snapshot in time of dispersal estimates. In this context, demographic data are therefore more informative than genetic data by providing insights about temporal variation of non-effective dispersal estimates. Yet, demographic approaches involve important technical constraints. First, the sampling effort dedicated to collect demographic data is often heavy, requiring successive capture sessions during which a significant part of the population is surveyed. Hence, demographic surveys are usually avoided in species displaying very large population sizes, for example, millions of individuals as in many invertebrates and marine fishes. Demographic methods are also usually poorly suited to study dispersal in species having large dispersal

**FIGURE 4** Decision tree for selecting demographic or genetic methods to study non-effective dispersal rates and distances according to characteristics (i.e., population size, dispersal distance and dispersal rate) of the spatially structured populations. Overall, demographic methods are unsuitable to estimate non-effective dispersal when the population size is large and the dispersal distance is long due to technical and financial constraints. In these situations, genetic methods can be used to estimate non-effective dispersal rates and distances except when dispersal rates are high. High dispersal rates result in low genetic variation between patches (or demes), which decreases the inference accuracy of assignment analyses, parentage approaches and sibship reconstruction. Our decision tree also highlights the situations (high dispersal rates, long dispersal distance and large population size) where both demographic and genetic approaches are usually unsuitable to estimate non-effective dispersal [Colour figure can be viewed at wileyonlinelibrary.com]
abilities. Surveying many individuals over large areas is often infeasible, leading to an underestimation of the frequency of long-distance dispersal events (Reid, Thiel, Palsbøll, & Peery, 2016; Watts et al., 2007). Their use is also limited by our ability to mark individuals, especially in the absence of natural individual marking (e.g., apomictic coloration) and when a small body size does not allow the use of pit-tags. By contrast, assignment and parentage analyses efficiently detect non-effective dispersal with a reduced sampling effort (i.e., one sampling event per patch or deme) directed to a smaller subset of the population. These approaches are therefore suitable to estimate non-effective dispersal in large populations that cannot be surveyed using demographic approaches. They are also very useful when individuals cannot be surveyed using visual or internal marking, or when the survey is based on both direct and indirect (e.g., faeces) observations. Yet, these methods also have their own limitations. Particularly, the inferential accuracy of assignment and parentage analyses declines when genetic variation between patches decreases (Anderson et al., 2010), which precludes any dispersal assessment in spatially structured populations with high dispersal between patches (e.g., patchy populations).

To summarize, we recommend investigators choose demographic and genetic methods depending on a priori knowledge about the population system (high or low expected dispersal, population size), the sampling effort that can be allocated (1 year or multiple years) and the technical constraints induced by individual marking and surveying. We propose a decision tree (Figure 4) to help in the selection of the methods that are suitable to estimate non-effective dispersal according to the characteristics of the spatially structured populations. Moreover, we recommend the use of both demographic and genetic approaches when feasible. Their combination can be used to cross-validate dispersal estimates. It also allows investigating the dispersal process at different spatial and temporal scales: Demographic approaches permit the examination of multi-annual patterns of short- and medium-distance dispersal, while genetic methods detect long-distance dispersal events through short-term studies.

5.2 Estimating effective dispersal rates and distances using genetic methods

The use of genetic data to estimate dispersal rate should be guided by the question being tackled, namely, does the study aim to characterize direct dispersal? Is the question focused on recent dispersal (over a few generations) or the long-term patterns of migration rate (m) among differentiated populations? Indirect methods that rely on a population genetic models often reflect an average of demographic parameters over long-term historical periods. For instance, data from the site frequency spectrum SFS (e.g., single SNP) provide information about migration and changes in population size over tens to thousands of generations (Gutenkunst et al., 2009). This is true for summaries of the SFS, such as \( F_{ST} \) or associated statistics. Similarly, coalescent simulations, regardless of the kind of data (SNPs, microsatellite) or the use of diffusion to compute the SFS (Gutenkunst et al., 2009; Kimura, 1964), will provide information on long-term historical demography. Alternatively, long blocks of identity-by-descent can provide information about very recent demographic processes (Browning & Browning, 2012), and it is now becoming possible to infer migration rate or dispersal parameters and the beginning of migration from these data (Baharian et al., 2016; Gravel, 2012; Harris & Nielsen, 2013; Palamara & Pe’er, 2013; Ringbauer et al., 2017). From an ecological perspective, inferring patterns of recent dispersal can be relevant to draw inference about population connectivity and such methods might be of high relevance in the near future. Accurate estimates of dispersal and very recent effective population size obtained from these methods might serve as priors for the calibration of various models in subsequent demographic studies of dispersal in organisms where prior data are unavailable. For now, while these methods appear very promising, they are still restricted to a few species and have only been applied to human data (see in particular Baharian et al., 2016 and Ringbauer et al., 2017 for dispersal rates and Palamara and Pe’er (2013) for recent migration rates). They rely on heavy population haplotypic data with many individuals sampled and require a genetic map to determine blocks of identity-by-descent. However, as pointed out by Ringbauer et al. (2017), the promising results of such methods to infer dispersal and population density should justify further developments, even in non-model species. Similarly, these blocks of identity-by-descent could be used to infer relatedness that can then be directly exploited to study dispersal patterns.

5.3 Combining demographic and genetic methods to compare effective and non-effective dispersal rates and distances

The level and nature of correspondence between non-effective dispersal estimated from CR data and effective dispersal from genetic estimates has long been debated (Koenig, Van Vuren, & Hooge, 1996; Watts et al., 2007; Lowe & Allendorf, 2010; Yu, Nason, Ge, & Zeng, 2010; Wang & Shaffer, 2017). Theoretically, effective dispersal rates and distances are expected to be equal or less than non-effective dispersal rates and distances (Broquet & Petit, 2009; Slatkin, 1987). Empirical studies reported that effective dispersal exceeded non-effective dispersal (Fedy, Martin, Ritland, & Young, 2008; Slatkin, 1987; Watts et al., 2007; Yu et al., 2010), while others found the opposite (De Meester, Gómez, Okamura, & Schwenk, 2002; Favre, Balloux, Goudet, & Perrin, 1997; Lachish, Miller, Storfer, Goldizen, & Jones, 2011) or congruent patterns (Funk et al., 2005; Vandesoestjine & Baguette, 2004).

Different explanations have been proposed to explain these discrepancies. First, they may result from sampling design misconceptions. They can be due to temporal mismatch between demographic and molecular estimates. CR studies based on temporally restricted survey may lead to incorrect inferences when dispersal rates broadly vary in time (e.g., due to meteorological fluctuation). Moreover, molecular approaches strongly differ from one another in terms of lag time for detection of the genetic effects of landscape change (Landguth et al., 2010; Murphy et al., 2008). In particular, individual-
based methods achieve an equivalent effect size roughly 10 times faster than $F_{ST}$ (Landguth et al., 2010). Accordingly, incongruences between demographic and genetic estimates of dispersal may result from a lag time between contemporary dispersal estimates from CR data and historical gene flow (for an illustrative example, see Howeth, McGaugh, & Hendrickson, 2008). Second, CR surveys with spatially limited sampling effort can also entail mismatches between both types of dispersal estimates. Non-exhaustively sampling of all the patches occupied by a population usually leads to biased demographic inferences. When individuals immigrate into unsurveyed patches, dispersers become virtually dead, as apparent survival and permanent emigration are always confounded in the CR framework (Lebreton et al., 2009). This leads to underestimating both survival and dispersal rates. Moreover, spatially restricted CR surveys often do not allow detection of rare long-distance dispersal events (Reid et al., 2016; Watts et al., 2007). These methodological biases may result in effective dispersal rates and distances exceeding non-effective dispersal ones, a phenomenon called “Slatkin’s Paradox” (Slatkin, 1987; Yu et al., 2010).

Incongruent patterns between effective and non-effective dispersal may also result from the low accuracy of the statistical methods used. Concerning genetic approaches, it is important to keep in mind, as explained above, that $F_{ST}$ and other genetic divergence metrics cannot be considered as direct estimates of effective dispersal (Marko & Hart, 2011; Whitlock & McCauley, 1999). Gene flow and genetic drift interact as opposing forces, the former decreasing and the latter increasing genetic variability among populations (Hutchison & Templeton, 1999; Slatkin, 1985). Hence, mismatches between demographic and genetic estimates of dispersal could, in part, be due to the contribution of genetic drift to genetic differentiation, especially when the spatial extent of the genetic studies exceeds the maximal dispersal distance of surveyed species. In particular, Broquet and Petit (2009) suggested that genetic estimates of dispersal at scales larger than the species dispersal ability should not be interpreted at face value. At such large scales, mutation and selection are also likely to act as confounding factors and populations at opposite ends of the sampling distribution may not be at demographic equilibrium. Taking into account these factors is most important with dense genomewide data comprising hundreds of thousands of markers where different processes unrelated to gene flow can generate patterns of differentiation. Concerning CR methods, many studies that have compared non-effective and effective dispersal provided naïve demographic rates of dispersal without taking into account imperfect detection of individuals (but see Callens et al., 2011; Reid et al., 2016). Hence, one can legitimately ask whether the differences (or incongruence) between demographic and genetic estimates of dispersal found by these studies could result from statistical artefacts.

Beyond methodological artefacts, mismatches between non-effective and effective dispersal estimates may result from their respective biological meaning. As previously stated, demographic approaches focus on dispersal events that are followed, or not, by an effective reproduction. By contrast, genetic approaches used to analyse population structuring and gene flow quantify effective dispersal resulting from dispersal events necessarily followed by a successful reproduction. Accordingly, discrepancies between non-effective and effective dispersal may result from dispersal costs (Bonte et al., 2012; Clobert et al., 2009). Various forms of dispersal cost exist including energetic costs, resulting from the movement itself during the transience phase or the development of special dispersal organs and tissues (e.g., muscles and wings) necessary to initiate dispersal (e.g., Zera & Denno, 1997). Dispersal may also trigger time costs associated with searching for a new suitable patch in which to settle (Hinsley, 2000) and risk costs related to mortality risks (e.g., due to increased predation or settlement in unsuitable habitat) and accumulated damages or physiological changes (Soulbury, Baker, Iossa, & Harris, 2008; Srygley & Ellington, 1999). Moreover, opportunity costs can also occur due to loss of familiarity with the environment, the loss of benefits from familial nepotism or the loss of the social rank after settling into a new social context (Dickinson, Euaparadorn, Greenwald, Mitra, & Shizuka, 2009; Hansson, Bensch, & Hasselquist, 2004). All these costs, immediate or delayed, may lead to negative effects on survival and reproductive outputs, potentially resulting in a loss of lifetime reproductive success in dispersers (Bonte et al., 2012) and therefore mismatches between non-effective and effective dispersal estimates.

To summarize, combining well-designed and well-executed demographic and genetic studies allows for accurate comparisons of non-effective and effective dispersal rates and distances. Such integrative approaches permit investigation of dispersal costs in wild populations, especially when the reproductive success of dispersers and residents cannot be directly observed and/or are not precisely estimated using parentage and sibship analyses.

5.4 Combining demographic and genetic methods to study effective and non-effective emigration and immigration

Demographic methods and several genetic approaches (e.g., assignment and parentage analyses) allow studying fine-scale relationships between non-effective emigration/immigration and individual (i.e., phenotype-dependent dispersal) and environmental (condition-dependent dispersal) factors (Table 1). Demographic surveys usually collect precise phenotypic information during an individual’s lifetime. The demographic models presented above (i.e., multistate and multievent models) permit in-depth analysis of dispersal syndromes, by quantifying the covariation between individual dispersal probabilities (or distances), phenotypic traits (morphology, physiology or behaviour) and life history traits (survival and fecundity). Moreover, parentage analyses and sibship reconstruction methods provide valuable inferences about genealogical relationships, reproductive outputs and fitness in wild populations. Combining both approaches could be a suitable way to examine the influence of kin competition and kin selection on dispersal and the contribution of dispersal in the evolution of sociality (Platt & Bever, 2009). It would also allow
the examination of the role of dispersal in the evolution of inbreeding avoidance strategies by associating emigration and immigration decisions with the social context within the patches of departure and arrival.

Despite their great interest to model asymmetric gene flow, methods quantifying effective immigration (i.e., gravity models and other approaches dedicated to examining IBE) remain rarely used, perhaps due to their apparent complexity. Two distinct mechanisms may generate IBE: First, natural selection against immigrants may lead to a decrease of disperser’s fitness when their genotype is ill-adapted to the environmental conditions prevailing in the patch of arrival (Hendry, 2004; Nosil, Vines, & Funk, 2005). Alternatively, the concept of habitat matching choice and more generally of context- and condition-dependent dispersal assumes that individuals do not disperse randomly but rather base their emigration/immigration decisions according to their capacity to use the environment (Edelaar et al., 2008; Jacob et al., 2017). One of the main limits of IBE approaches is that they do not tease apart the relative contribution of the two mechanisms in directed gene flow. In this context, using demographic approaches could help determine whether non-effective immigration is followed by an erosion of fitness components (i.e., survival, reproductive success) or whether emigration/immigration decisions are based on environmental factors (i.e., similitudes of the characteristics of the patch of arrival and departure).

5.5 | Combining demographic and genetic methods to study non-effective and effective transience

To date, few demographic studies have analysed the effects of landscape composition and configuration on non-effective transience (Arellano et al., 2008; Cayuela, Bonnaire et al., 2018; Ovaskainen, 2004; Ovaskainen et al., 2008). This is mainly due to the scarcity of CR data collected in large geographic areas including a large number of patches. This is also potentially caused by the complexity of modelling systems, for example, Ovaskainen’s diffusion model (Ovaskainen, 2004) being computationally intensive. As well, few studies have analysed the landscape effects on non-effective transience using genetic approaches except that of Kamml et al. (2010) already mentioned above. Recently, Norman et al. (2017) developed a method that incorporates landscape information to perform spatial interpolation of relatedness. Many studies of connectivity patterns use parentage analysis or dispersal kernels (e.g., Almany et al., 2017; Buston et al., 2012; D’Alloia et al., 2013; Ismail et al., 2017), but these studies would be improved by the incorporation of landscape information.

Landscape genetic methods provide a suitable framework to study the effect of landscape composition and configuration on effective transience. They may be applied at relatively large spatial scales, which allow analysing landscape effects on transience in species with long dispersal distances (for which the use of demographic methods is technically complicated). They do not require an exhaustive sample of the individuals of a population, which allows studying landscape effects on transience in species displaying large population sizes (which is hardly possible with demographic, assignment and parentage analyses). However, they also display several limitations. First, they do not allow investigating the influence of phenotypic variation on transience (except for fixed traits, such as sex). Most of landscape genetic studies are temporally restrictive (i.e., single temporal point studies) due to financial constraints, while population genetic structuring may vary over time because of demographic or environmental causes (but see, e.g., McCairns & Bernatchez, 2008). Moreover, although several methods allow taking into account some spatial heterogeneity in genetic drift (e.g., Prunier, Dubut et al., 2017), it remains a difficult challenge to disentangle the relative contribution of other processes, especially if the landscape features limiting dispersal also constrain population sizes. On this specific point, combining demographic and landscape genetic approaches seems appropriate, as CR models usually provide robust census population size estimates (Nichols, 1992; Pollock, Nichols, Brownie, & Hines, 1990). Although census population size and effective population size are two very different metrics and their ratio is typically far from one (Frankham, 1995; Hedrick, 2005), a thorough estimation of census population size can be helpful to understand the factor influencing the variance in the ratio of effective population size to census size. Combining demographic and genetic approaches could also alleviate the problem of friction map parameterization that is often based on expert opinion. Demographic models could be used to examine the effects of landscape surface on non-effective dispersal, which could therefore inform landscape geneticists about how each surface is resistant to organisms’ movement.

6 | CONCLUSION

In this review, we proposed a unified framework allowing ecologists to select the most accurate tools to use with respect to their biological questions and population characteristics. We believe that an appropriate use of demographic and genetic methods, as well as their combination, allows investigators to address central issues in the field of ecology. Our review highlights that different demographic and genetic approaches can be used depending on the type of dispersal to infer (i.e., non-effective vs. effective), the age of targeted dispersal flow (i.e., historic vs. recent) and the characteristics of the populations (i.e., population size, intensity and extent of dispersal). Nevertheless, the integration of demographic and genetic methods remains hampered by the fact that the specialized knowledge required to apply these methods currently resides in two fields that have limited interactions. One potential solution would be the development of integrated analysis platforms to estimate both non-effective and effective dispersal rates and distance for the same spatially structured population. Such approaches would also account for population dynamics (i.e., survival, natality, census population size) in the gene flow inferences and landscape genetic analyses. To date, long-term capture–recapture surveys including multiple monitored sites are available to conduct such analyses on a broad range of
animal taxa. We believe that integrated approaches would open new avenues in the analyses of dispersal pattern and gene diffusion in wild populations.

AUTHOR CONTRIBUTIONS

The review was conceptualized by H.C. The MS was written by H.C., Q.R. and J.G.P. The redaction of the manuscript was contributed by J.S.M., J.C., A.B. and L.B.

ORCID

Hugo Cayuela http://orcid.org/0000-0002-3529-0736
Jérôme G. Prunier http://orcid.org/0000-0002-3353-3730

REFERENCES


CAYUELA ET AL.
sion commonality analyses on hierarchical genetic distances. Ecology,
Reid, B. N., Thiel, R. P., Palsboll, P. J., & Peery, M. Z. (2016). Linking genetic kinship and demographic analyses to characterize dispersal-


