Ann. Zool. Fennici 54: 97–112 ISSN 0003-455X (print), ISSN 1797-2450 (online)
Helsinki 15 May 2017

Ilkka Hanski: The legacy of a multifaceted ecologist

Metapopulations of marine species with larval dispersal: A counterpoint to Ilkka's Glanville fritillary metapopulations

Oscar E. Gaggiotti

Scottish Oceans Institute, East Sands, University of St. Andrews, St. Andrews, Fife KY16 8LB, UK

Received 29 Nov. 2016, final version received 17 Jan. 2017, accepted 18 Jan. 2017

Gaggiotti, O. E. 2017: Metapopulations of marine species with larval dispersal: A counterpoint to Ilkka's Glanville fritillary metapopulations. — *Ann. Zool. Fennici* 54: 97–112.

Marine ecologists have been slow at adopting the metapopulation paradigm in their research. As explained in the landmark book *Marine Metapopulations*, marine ecology was focused on local processes and used mainly experimental approaches that neglected the potential role of demographic connections among local populations, but this is no longer the case. Metapopulation theory is now integrated into marine ecology research and is being used as a helpful framework for describing spatial population processes, for asking new research questions, and for planning the management and conservation of marine species. Nevertheless, the use of the metapopulation paradigm in marine sciences is challenging because of the complex life histories of marine species and the peculiarities of the marine environment. Here, I highlight the many challenges faced by ecologists who study marine metapopulations and mention some of the approaches that can be used to overcome them. In particular, I focus on the problem of estimating connectivity, an important attribute of metapopulations that represents a fundamental knowledge gap in marine ecology. Following the repeated calls for integrative approaches that combine all sources of information I propose the use of the hierarchical Bayesian framework and provide an example that considers the joint analysis of the three leading approaches in connectivity research, genetics, microchemistry and biophysical modelling.

Introduction

Many, if not most, of the countless and important contributions of Ilkka Hanski to the field of population biology where based on his in-depth study of the Glanville fritillary (*Melitaea cinxia*) metapopulation network in the Åland Islands. As he explains in his insightful *Metapopulation Ecology* book (Hanski 1999), this is an ideal model system for conducting ecological research across large geographic areas. It has several practical advantages, which include the butterfly's life cycle and spatial structure of the habitat where it is found. The suitable habitat occurs in discrete patches, which simplifies the delimitation of local populations, and their limited size and the gregarious larval habits of M. *cinxia* greatly facilitate the census of local populations. Although sporadic long-distance dispersal may occur, most of the dispersal events involve nearby local populations. This and the geography of the Åland Islands lead to a hierarchical population structure that enables the study of ecological processes at a wide range of spatial scales.

Finally, there is a more prosaic but nevertheless important feature; the dense network of small roads that connect the many scattered farmhouses on the Åland Islands, which facilitate access to local populations of *M. cinxia*. Having said all of this, it is still true that planning and carrying out ecological research at the scale of the Glanville fritillary project is extremely challenging and requires an inquisitive mind, great determination and careful organisation; all attributes that Ilkka possessed in great abundance. It also requires the participation of a large number of devoted and hard-working students and young researchers. But this did not represent a problem to Ilkka because of his eagerness to foster the career of young scientists.

Although Ilkka focused his research on terrestrial metapopulations, the influence of his work extended to the marine sciences. Here, however, we are in a completely different realm with metapopulations of marine species representing a counterpoint to the Glanville fritillary metapopulations. Indeed, most marine species spend part of their life in the plankton at the mercy of strong oceanic currents and were initially seen as living in demographically open populations (Caley *et al*. 1996). On the other hand, butterflies were initially seen as living in closed populations because of their limited dispersal abilities and the patchy distribution of their host plants (Ehrlich *et al*. 1975). In both cases, however, the truth lies somewhere between these two extremes (cf. Hanski 1999, Hellberg 2009). Nevertheless, the differences between marine and terrestrial metapopulations are substantial and in some cases, extreme; compare for example insects with mobile adults but immobile larval stages and benthic species with fixed adults but free ranging planktonic larvae. The larval stages of many marine taxa can spend more than a month in the water column at the mercy of ocean currents so their potential for dispersal is enormous. Moreover, the strong asymmetric flows and dynamics of the marine environment and their temporal variability can lead to very large variance in reproductive success with most larvae being lost and only few individuals contributing recruits to local populations [reviewed by Hedgecock and Pudovkin (2011)]. Additionally, local population sizes of most marine species with larval dispersal are several orders of magnitude larger than that of most terrestrial species and they are logistically difficult to sample because of the peculiarities of the marine environment (Riginos *et al*. 2016). All these features make direct observation and study of individuals and local populations very difficult. Instead, marine biologists rely heavily on both remotely and locally deployed instruments with their own measurement errors.

In this article, I will describe the main obstacles faced by the study of marine metapopulations and explain how they may be overcome by combining tools and data from different research fields using truly integrative statistical frameworks. I will focus attention on the study of spatial structure of metapopulations, which is driven to a large extent by connectivity patterns, a metapopulation attribute that pervades Ilkka's research. Note, however, that the purpose is not to review the marine connectivity literature but rather to highlight the great potential of integrative Bayesian approaches as tools to study metapopulation processes in marine species.

What's so special about marine metapopulations?

It is clear that few if any marine metapopulation can play the important role of ecological model system or, as Ilkka put it, "ecological field facility" (Hanski 1999) played by the Glainville fritillary metapopulation. However, they still need to be studied in detail to answer a number of important questions in population biology. These questions are wide-ranging, from the basic description of metapopulation structure and dynamics to the more complex question of how local adaptation and speciation can take place despite the potential for very long-distance dispersal.

As stated before, marine metapopulations of species with pelagic larval stages possess important attributes that make them very different from most terrestrial metapopulations. In particular, the life history of marine species includes at least a subset of the following attributes: potential for very long-distance dispersal, huge fecundity and very large variance in reproductive

success, spatial distributions covering large geographic areas with very large local-population sizes and, importantly, several life stages showing sequential ecological adaptations to different habitats (Palumbi & Hedgecock 2005). These life-history attributes and the physical properties of the marine environment, which include strong and asymmetric physical flows over a very extended and heterogeneous habitat matrix, exacerbate the importance of spatial processes (Riginos *et al*. 2016).

In principle one could expect that the spatial population structure will be strongly determined by ocean circulation, in which case species sharing the same habitat should exhibit very similar spatial population structures. Surprisingly, this is not the case as shown by Selkoe *et al*. (2014) who uncovered extreme interspecific variation in spatial genetic patterns across 35 coral reef species inhabiting the Hawai'ian archipelago despite its almost perfect stepping-stone configuration, which in principle should lead to an isolation-by-distance structure. Thus, the question arises as to what are the physical and biological mechanisms that can lead to such large variability in spatial patterns. Ocean circulation can generate invisible barriers to dispersal (e.g. eddies) leading to negligible connectivity between neighbouring local populations (Gilg & Hilbish 2003) but can also facilitate dispersal between geographically distant populations (Mitarai *et al*. 2009). This can explain the lack of an isolation-by-distance pattern but not the large differences in spatial structure among species sharing the same habitat.

Another important characteristic of the marine environment is the temporal variability in ocean circulation patterns, which can include flow reversals even along straight coastlines (Gaylord & Gaines 2000, Selkoe *et al*. 2006). Temporal variability in ocean currents can arise through temporal variations in wind forcing (e.g. Mitarai *et al*. 2008). But larval connectivity is inherently an unpredictable and heterogeneous process on annual time scales even in the absence of spatial and temporal variability (Siegel *et al*. 2008). This stochasticity is due to advection of pelagic larvae by chaotic coastal circulation (Siegel *et al*. 2008). Thus, equilibrium assumptions underlying many statistical methods are likely to be violated leading to strong biases in estimates of population parameters.

Making inferences about metapopulation processes based on the structuring of genetic diversity

Marine population biologists have relied on the use of indirect methods based on natural tags as a means to overcome the challenge of directly observing pelagic species over large geographic areas. The most popular such approach uses molecular markers and population genetics tools and models. The underlying principle is that by studying the genetic structure of populations we gain a better understanding of the determinants of species ranges and factors that may facilitate the exchange of individuals between populations. Indeed, as any introductory population genetics book makes it clear, a species' population genetic structure is influenced by genetic (mutation and recombination) and ecological/demographic processes (demographic stochasticity, dispersal and selection). These two types of processes interact and leave complex genetic signatures that can be deciphered using the appropriate statistical methods.

Species ranges are strongly influenced by genetic drift, selection and dispersal. These processes, in turn are themselves influenced by factors such as demography (population dynamics, demographic history), geographic distance, topology, and many other environmental factors such as ocean currents, salinity, temperature, etc. (Fig. 1). Thus, it is possible to make inferences about the demographic history and ecology of species using the genetic signatures that the above-mentioned processes have left in the gene pool of a species. In principle, the only prerequisite is the availability of population genetic samples covering the species range or part thereof. Although this may be difficult, it is much easier than carrying out direct observations or mark-recapture studies.

There are several statistical genetics methods for making inferences about demographic processes. These include testing for changes in population sizes (e.g. Luikart & Cornuet 1998,

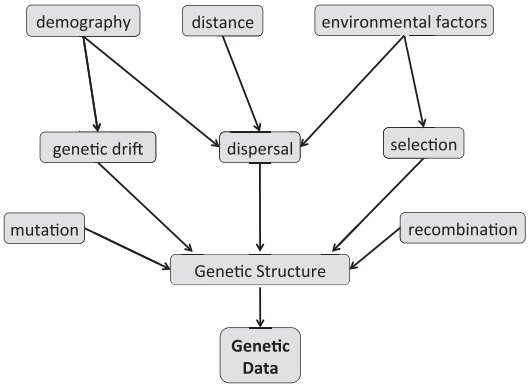

Fig. 1. Hierarchical structure of drivers of genetic structuring. Genetic data is obtained from a metapopulation subject to demographic, ecological, and genetic processes. The demographic and ecological processes are themselves driven by environmental factors. Thus, the integration of genetic and ecological data will enable testing of the factors that drive ecological processes.

Garza & Williamson 2001), estimating migration rates (Beerli & Felsenstein 2001, Wilson & Rannala 2003), and estimating population growth rates (Kuhner *et al*. 1998, Beaumont 1999). These approaches allow us to estimate demographic parameters but they do not make inferences about the drivers of processes such as changes in population size, migration or local adaptation. This inference is carried out *a posteriori* and without taking into account the uncertainty underlying parameter estimates. In the case of marine species, a typical example is the evaluation of the role of ocean circulation on larval dispersal (e.g. Galindo *et al*. 2006, Selkoe *et al*. 2010, White *et al*. 2010). The typical approach first obtains predicted oceanographic connectivity (as defined in Table 1) based on larval transport models (e.g. Cowen & Sponaugle 2009) and point estimates of genetic connectivity (F_{ST} or some related measure; *see* Table 1) using molecular markers and standard population genetics approaches (e.g. Selkoe *et al*. 2010, White *et al*. 2010). Using these point estimates, it then tests for correlations between oceanographic and genetic connectivity. However, ignoring uncertainty around point estimates precludes a rigorous evaluation of the validity of our conclusions. An alternative strategy is to use individual-based simulation models to obtain the genetic structuring (as measured by F_{ST} or related statistics) expected for a connectivity matrix generated by an oceanographic model (cf. Galindo *et al*. 2006). This predicted genetic differentiation is then compared with estimates obtained from real data in an *ad-hoc* fashion. Although this approach better integrates oceanographic and genetic models, the use of individual-based simulations to implement genetic models has several limitations (*see* Hoban *et al*. 2012).

Incorporating other sources of information to test hypotheses about drivers

The rigorous testing of hypotheses about drivers of demographic processes requires the joint analyses of genetic and non-genetic data. This can be accomplished using the Bayesian framework (e.g. Kittlein & Gaggiotti 2008, Gaggiotti *et al*. 2009), which allows us to use the prior distribution to introduce non-genetic data with which we can generate alternative priors, each representing an alternative hypothesis. These hypotheses are tested using genetic data introduced through the likelihood function. This approach is visualised in Fig. 2 and has been used to develop methods

Table 1. Definitions of the connectivity types mentioned in the article.

Type	Definition
Demographic Connectivity	Dispersal of individuals among subpopulations that survive until completion of the settlement process
Genetic Connectivity ↔ Reproductive connectivity	Dispersal of individuals among subpopulations that survive to reproduce
Oceanographic Connectivity	Dispersal probability between locations as predicted by oceanographic circulation patterns and larval traits

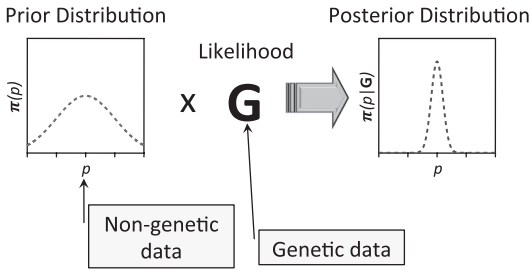

Fig. 2. Schematic representation of the Bayesian approach as implemented in e.g., Gaggiotti *et al.* (2002, 2004). In this example, there is only one parameter to estimate, which could be the allele frequency, *p*, at a particular locus. The prior distribution, $\pi(p)$, is used to introduce the non-genetic data and the likelihood function allows the incorporation of genetic data, **G**, with which we estimate the posterior distribution of *p*, $\pi(p|\mathbf{G})$.

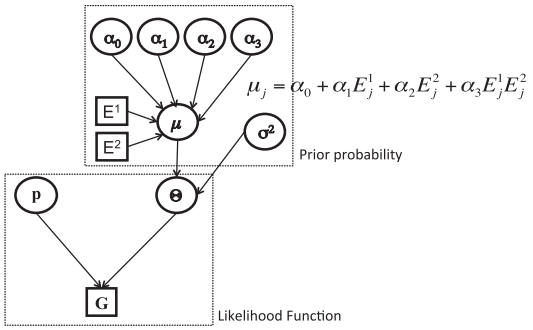

Fig. 3. Simplified and stylised Directed Acyclic Graph (DAG) of the statistical method implemented in GESTE (Foll & Gaggiotti 2006). The DAG describes the stochastic relationship between parameters (circles) and data (squares) of the Bayesian model. The effect of environmental factors E^1 and E^2 on the degree of genetic differentiation, Θ, are quantified by α_1 and α_2, respectively while the effect of the interaction is measured by α_3 and α_0 is the intercept. The prior for $\boldsymbol{\Theta}$ is lognormal with mean μ and variance σ^2. The allele frequency distributions at each locus and local population are described by the multidimensional matrix **p** and the genetic data (sample allele counts for each locus and population) by the multidimensional matrix **G**.

aimed at making inferences about drivers of various population processes. These include factors driving the colonisation of new habitat (Gaggiotti *et al*. 2002, 2004, Okuyama & Bolker 2005), drivers of spatial genetic differentiation patterns (Foll & Gaggiotti 2006), drivers of migration (Faubet & Gaggiotti 2008), and drivers of local adaptation (de Villemereuil & Gaggiotti 2015).

Figure 3 presents a stylised and simplified Directed Acyclic Graph (DAG) describing the stochastic relationships between the parameters (circles) and data (squares) of the Bayesian model implemented in GESTE (Foll & Gaggiotti 2006). This method allows testing hypotheses about the drivers of genetic differentiation in a metapopulation by focusing on population-specific F^j_{ST}s, which quantify the genetic differentiation between each local population *j* and the whole metapopulation. The likelihood function describes the probability of observing allele frequency counts given the allele frequencies $\mathbf{p} = (p_{ij})$ at each locus *i* and population *j*, and given genetic differentiation $\boldsymbol{\Theta} = (\theta_j)$, where $\theta_j = (1/F^j_{ST}) - 1$. The prior of θ_j is a log-normal distribution with mean μ_j and variance σ^2. The environmental data are introduced through the means of the log-normal distributions using a linear model. In this example, we consider two environmental variables (e.g. connectivity and salinity) described by vectors $\boldsymbol{E}^1 = (e_1^1, \ldots, e_j^1, \ldots, e_J^1)$ and $\boldsymbol{E}^2 = (e_1^2, \ldots, e_j^2, \ldots, e_J^2)$, where e_j^i is the observed value of environmental variable *i* for population *j*. If we consider an interaction term, then we can generate nine alternative models depending on which environmental variables are used to define the prior. The posterior distribution of each model is estimated using a Reversible Jump MCMC algorithm (Green 1995).

Challenges posed by the peculiarities of marine metapopulations and approaches to overcome them: The example of marine connectivity

Although a well-developed body of theory and statistical methods exist to analyse population genetic data (e.g. Balding *et al*. 2007), the peculiarities of marine species and the environment they inhabit limits the power of existing methods to carry out demographic inference. There are several problems but the most important one is that large local-population sizes and complex life histories allowing long-distance dispersal lead to very weak genetic signatures. For example, a very sharp decrease in population size may still result in effective sizes that are unlikely to leave

a strong bottleneck signature (e.g. Gaggiotti & Vetter 1999) or genetic differentiation among local populations can be very weak, which limits inferences about migration (Waples 1998). Two other important problems are the large spatial scale of marine metapopulations, which precludes sampling of all extant populations, and the extreme temporal variability of the marine environment that leads to non-equilibrium dynamics. This latter challenge arises because most statistical genetic methods to estimate population parameters (e.g. growth rates, migration rates, population sizes) assume equilibrium conditions (e.g. migration-drift or mutation-drift equilibrium). In this regard, it is important to note that the extreme temporal variability of the marine environment can lead to stochastic migration patterns, which in turn have important consequences on the genetic structure of metapopulations (e.g. Gaggiotti 1996, Gaggiotti & Smouse 1996) and, therefore, on estimates of demographic parameters from genetic data.

In what follows I will address these three challenges to the use of population genetics methods and the approaches that can allow us to overcome them. Throughout I will focus attention on the estimation of connectivity in marine species, an important knowledge gap in marine sciences that needs to be filled in order to advance our understanding of ecological and evolutionary processes in the marine realm (e.g. Thorrold *et al*. 2002, Lipcius *et al*. 2008, Gaines *et al*. 2010).

Connectivity, broadly defined as the exchange of individuals among local populations, is a fundamental spatial attribute of species and metapopulations and has been the focus of much research in the fields of ecology and evolution. At the ecological level, connectivity is one of the main drivers of the persistence and resilience of populations (Hastings & Botsford 2006), while at the evolutionary level it influences the extent to which species can adapt to local conditions (e.g. Hellberg 2009). The study of connectivity is also of fundamental importance in conservation biology where it is essential for effective spatial management and design of protected areas (Gaines *et al*. 2010).

The definition of connectivity given above is very general and, as such, it can cover many different aspects including physical transport, demographic exchanges and gene flow. Table 1 lists definitions for the different types of connectivity that are mentioned in this review. In what follows I will be referring to demographic connectivity, which is the result of reproduction, dispersal and recruitment (Botsford *et al*. 2009) and is measured shortly after larval settlement but before reproduction. Thus, demographic connectivity cannot be equated to gene flow because new recruits may not survive long enough to reproduce. Note, however, that it still is an important driver of this evolutionary force.

Although very informative, studies of connectivity so far have been largely descriptive (cf. Botsford *et al*. 2009) and only provide *qualitative* estimates of this fundamental parameter (Christie *et al*. 2017). The final objective of any empirical approach to quantifying demographic connectivity is to estimate a dispersal matrix that can be integrated and related to features of connectivity identified by models as driving metapopulation dynamics (Botsford *et al*. 2009). Thus, below I will only address population genetic methods that can estimate migration rates, m_{ij}, between pairs of populations i and j. Since the focus is on demographic connectivity, m_{ij} represents the proportion of new recruits in local population i that were contributed by local population j. It should be noted that there are many other population genetics methods that have been used in connectivity studies but, as explained below, they are not appropriate for the quantitative estimation of connectivity.

When using genetic data, it is important to note that the temporal scale of the estimates depends on the population genetics framework being used by the statistical method (Gaggiotti 2004). Methods based on coalescent theory (Kingman 1982) such as Migrate (Beerli & Felsenstein 2001), Lamarc (Kuhner 2006) or iMa (Hey & Nielsen 2007) provide estimates of connectivity over evolutionary time scales, while methods based on multi-locus genotypes approaches (cf. Gaggiotti 2004) such as BayesAss (Wilson & Rannala 2003), BIMr (Faubet & Gaggiotti 2008) or the method described by Broquet *et al*. (2009) provide estimates of migration rates over ecological time (i.e. during the last one or two generations). Naturally, marine ecolo-

gists are interested in ecological estimates and, therefore, in what follows I will mostly focus on multi-locus genotype approaches. I will start with the challenges posed by spatial scale and temporal variability, for which potential solutions exist, and then address the more difficult problem posed by the low genetic differentiation observed in many marine species, which require the development of a new statistical method.

Spatial scale

Although population genetics approaches are easier to implement than mark–recapture studies, the spatial scale of marine metapopulations can be vast and, therefore, we can only sample a small fraction of all the local-populations they comprise. This incomplete sampling can lead to the so-called "ghost population" effect whereby unsampled populations that exchange migrants with some of the sampled populations can upwardly bias migration rate estimates between sampled populations (Beerli 2004, Slatkin 2005).

It is impossible to completely overcome this problem but there are spatial statistic techniques that can help minimise it by estimating allele frequencies in the unsampled populations. Some of them have already been applied in the field of population genetics. For example, Vounastou *et al*. (2003) present a method based on a Multivariate Conditional Autoregressive (MCAR) approach and use it to model spatial variation in HLA-B allele frequencies. Wasser *et al*. (2004) used a Gaussian Process model to implement a continuous assignment method that estimates allele frequencies of unobserved locations and then assigns individuals to any location regardless of whether or not they were sampled. This method was recently extended by Rundel *et al*. (2013), who use MCMC to sample from the posterior distribution of allele frequencies conditional on the allele counts at observed locations. This allows them to introduce uncertainty in the spatial covariance parameters. A similar approach can be used to estimate migration rates. Moreover, it is possible to incorporate the output of larval transport models in order to help improve the estimation of allele frequencies in unsampled locations. Such a method can be implemented using the Stochastic Partial Differential Equation approach (SPDE; Lindgren *et al*. 2011), which can easily accommodate all kinds of geographically referenced data.

Temporally stochastic circulation patterns

As mentioned before, temporal variability in ocean circulation patterns is very large and can even include flow reversals. The time scale on which these changes operate is very short and, therefore, the long-generation time of many marine species can cover several years during which ocean circulation patterns exhibited extreme variability. In this regard, it is important to note that population genetics inference methods typically assume discrete generations and, therefore, each time step correspond to one generation. The direct consequence of this is that the estimates of per-generation migration rates provided by genotype-based methods represent population averages over several years. This is a real problem because a thorough understanding of connectivity and its drivers cannot be accomplished without taking into consideration temporal variability in ocean circulation patterns (Cowen & Sponaugle 2009). An important question in this regard is the extent to which stochastic circulation patterns translate into temporally stochastic connectivity patterns. Although the larval stages of marine species may not have the swimming abilities necessary to overcome advection and diffusion transport processes, their life history may provide mechanisms that can lead to dispersal patterns that differ from those of inert particles in a fluid environment (cf. Cowen & Sponaugle 2009). For example, reproduction of many marine species takes place at specific times of the year and this can decrease to some extent the effects of the physical environment. Also, the duration of the pelagic larval stage can mediate the effects of physical processes (Shanks *et al*. 2003, Selkoe & Toonen 2011). Moreover, some larvae exhibit complex behaviour such as vertical migration and oriented horizontal swimming, which allows them to have some control over the direction and extent of their dispersal (Shanks 2009).

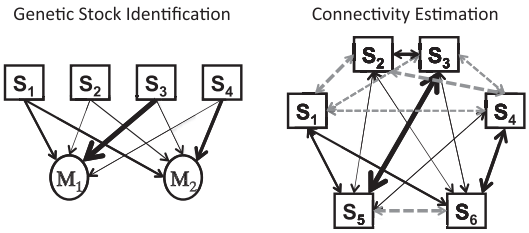

Fig. 4. Differences between the scenario considered by the Genetic Stock Identification method and the scenario that needs to be considered for the estimation of connectivity. GSI approach assumes two different types of demographic units, the potential source populations or stocks where reproduction takes place (S_i) and the genetic mixtures (M_i) that do not contribute migrants. The scenario that we need to consider for the estimation of connectivity does not make a distinction between local populations (S_i); reproduction takes place in all of them and all can contribute migrants to other local populations. In both cases the width of the arrows represent the relative magnitudes of the contribution of the different populations. For simplicity, the connectivity scenario considers symmetric migration but this is not necessary.

There is only one way in which the effect of temporal stochastic circulation patterns can be evaluated and this is by collecting spatio-temporal genetic samples. More precisely, we need tissue samples from adults and recently settled recruits in consecutive years (e.g. as in Christie *et al*. 2010). Instead of estimating local-population connectivity, the focus is on estimating larval connectivity, also referred to as "demographic connectivity" (Burgess *et al*. 2014), which requires the inference of the origins and destination of individual larvae. This may sound like an unachievable objective but there are several examples of studies that have done precisely this using assignment tests or parentage analyses (e.g. *see* references in Burgess *et al*. 2014). As Christie *et al*. (2017) demonstrate in a recent simulation study, neither of these methods is an ideal solution to the problem of estimating a demographic connectivity matrix as they can only provide qualitative evidence of connectivity between populations. Additionally, even under ideal conditions they cannot incorporate the uncertainty in the larvae origins into an estimate of demographic connectivity, being able only to provide very rough point estimates of migration between populations.

A more rigorous approach to estimate demographic connectivity is the Genetic Stock Identification method (e.g. Pella & Masuda 2001, Gaggiotti *et al*. 2002, Gaggiotti *et al*. 2004). This method was originally developed to estimate the proportionate contribution of different freshwater salmon stocks to the genetic mixture of harvested fish (e.g. Smouse *et al*. 1990) from a genetic sample of the harvest and baseline samples from all candidate stocks. The GSI method was later extended to identify the drivers of recent colonization events in a metapopulation (Gaggiotti *et al*. 2002, 2004), a scenario resembling that considered for the estimation of demographic connectivity. There are, however, some important differences between the scenario assumed by GSI methods and that envisaged for the estimation of connectivity (*see* Fig. 4). More precisely, the GSI approach assumes two different types of demographic units, the potential source populations or stocks where reproduction takes place and the genetic mixtures, which could be found for example in foraging grounds. The standard GSI approach considers a single genetic mixture but the method has been extended to allow for the simultaneous analysis of several genetic mixtures (Bolker *et al*. 2007). On the other hand, the scenario that we need to consider for the estimation of connectivity does not make a distinction between local populations; reproduction takes place in all of them and all can contribute migrants to other local populations. This is the scenario considered by BayesAss (Wilson & Rannala 2003) and BIMr (Faubet & Gaggiotti 2008) but these methods do not distinguish between adults and recently settled recruits and, therefore, provide estimates of migration rates averaged over several years that may encompass very different circulation patterns (*see* above). The method of Broquet *et al*. (2009) is based on samples taken at two different times, before and after dispersal and, therefore, is in principle better suited to estimate demographic connectivity. Nevertheless, as opposed to GSI-based methods, which use allele frequency counts in the source populations and individual genotypes in the mixture as input data, Broquet *et al*. (2009) use genotype counts before and after dispersal. This entails the estimation of genotype frequencies instead of allele frequencies; which requires

much larger sample sizes per population. This is illustrated by the results of the simulation study they present; in a scenario that only considers three local populations they observed very strong bias and low precision even when sampling 25% and 100% of pre- and post-dispersal individuals respectively. Thus, their method cannot be applied to species with very large population sizes, in which case it is only possible to sample a very small fraction of each local population. The limitation of Broquet *et al*.'s (2009) method can be overcome if we focus on allele frequencies among pre-dispersal individuals. Below I present a formulation that extends a previous GSI-based method (Gaggiotti *et al*. 2004) to allow for the estimation of the full migration matrix. But first I would like to address the third challenge posed by marine metapopulations.

Weak genetic differentiation

The very large local-population sizes of marine species with larval dispersal limit the effect of genetic drift and result in very low genetic differentiation among local populations (Waples 1998). This greatly limits the power of methods to estimate migration rates using genetic data (e.g. BayesAss or BIMr; *see* above). New developments in sequencing technologies (NGS) may help overcome this limitation but population genetic data is not the only source of information that can be used to estimate connectivity.

An alternative approach is to use microchemical fingerprinting to assign individuals to source populations. Calcified structures in marine invertebrates and fish (e.g. otoliths, shells, statoliths) can be used as natural tags of natal origins. These geochemical tags are the result of environmental conditions (temperature, salinity, seawater chemistry) that are recorded by the elemental composition of the calcified structure (Thorrold *et al*. 2007). Thus, larvae developing in areas that have different seawater characteristics will develop calcified structures that differ in their elemental composition (Zacherl *et al*. 2003). Hence, the basic prerequisite of this approach is substantial environmental variation across habitat patches or breeding sites (Thorrold *et al*. 2007), which parallels the requisite of genetic differentiation when using molecular markers. Another limitation of microchemistry is the need to characterise the elemental composition of young individuals across large geographic areas. However, this problem can be minimised using the same geostatistical techniques described for genetic data (*see* above).

Overcoming or at least minimising the limitations of genetic and microchemistry approaches require spatially and temporally extensive sampling and thorough molecular or chemical analyses. Thus, marine ecologists have resorted to using high-resolution biophysical models consisting of an underlying ocean circulation model to describe motion and an overlying particle-tracking model that describes the physics and biology of the larvae. Although these can be highly sophisticated incorporating some biological reality, they do not cover the full larval life history or they do not encompass the spatial extent of the dispersal process (Cowen & Sponaugle 2009). Doing this is still extremely time consuming and computationally expensive so most biophysical models make several simplifying assumptions that need to be verified. Thus, they do not constitute an inferential method. Rather, they provide predictions of connectivity that need to be validated with empirical data.

It is clear that no single approach can deal with the challenges faced by the studies of marine connectivity. They all have strengths and weaknesses and the best approach is to use an integrative framework that harnesses the information provided by these three independent data types to carry out statistical inference.

An integrative Bayesian framework

The need for combining different sources of information to estimate connectivity has been highlighted several years ago (e.g. Thorrold *et al*. 2002, Hedgecock *et al*. 2007) and there have been several mentions of the need to validate predictions made by biophysical models using real data (e.g. Thorrold *et al*. 2007, Cowen & Sponaugle 2009). Thus, over the last few years, there have been several attempts to integrate two different types of data; for example, output

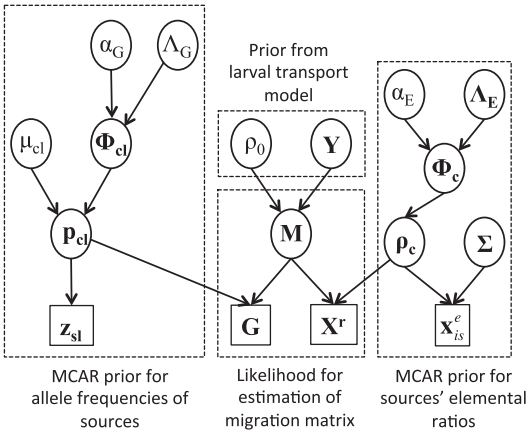

Fig. 5. Direct Acrylic Graph (DAG) of the Bayesian formulation for the estimation of the connectivity matrix. Square nodes denote known quantities (data) and the circles represent parameters to be estimated.

of biophysical models and genetic data (Selkoe *et al*. 2010, White *et al*. 2010, Alberto *et al*. 2011, Rivera *et al*. 2011) or microchemistry data (Lopez-Duarte *et al*. 2012). Similarly, there have been studies that combine genetic and microchemistry data (e.g. Perrier *et al*. 2011, Tanner *et al*. 2014). All these attempts, however, simply apply the two methods in tandem and then compare inferences or predictions obtained with the different approaches.

Although not specifically aimed at the estimation of demographic connectivity, statistically rigorous approaches have been proposed to carry out joint statistical inference using genetic and microchemistry data. Smith and Campana (2010) couple a genetic stock identification formulation based on allele frequencies with microchemistry data in order to estimate the proportionate contribution of different fish stocks to a genetic mixture. One disadvantage of this method is that instead of using individual genotypes in the mixture as input, it uses allele frequencies and, therefore, does not fully exploit all the information available in genetic data. A more recent study by Rundel *et al*. (2013) combines genetic and stable isotope data in a Bayesian framework for assigning wintering songbirds to breeding grounds. While not designed to generate a connectivity matrix, the study clearly shows how integration of data can increase spatial precision and accuracy of assignment beyond levels achievable from either the genetic or isotopic assignment techniques alone.

Below, I will introduce a Bayesian framework to integrate data from the three leading connectivity techniques in marine science: genetics, microchemistry, and biophysical modelling. Full details of the method and accompanying software will be published elsewhere. Here I will simply describe how the hierarchical Bayesian approach can be used to implement such a method. The underlying rationale is that genetics and microchemistry provide raw data for estimating parameters of a probabilistic model of migration (connectivity) between sources and recruit sampling sites. Biophysical-population models, on the other hand, provide predictions of connectivity patterns from existing knowledge about the species' life history and an ocean circulation model. Thus, the Bayesian framework incorporates genetic and microchemistry data through the likelihood function, while the connectivity predictions from the biophysical model are incorporated through the prior used for the migration rates. Figure 5 presents a Directed Acyclic Graph describing the approach.

Description of the Bayesian framework

We extend the Genetic Stock Identification approach of Gaggiotti *et al*. (2002, 2004) to allow for the estimation of the full connectivity matrix. Assume that n_k^r recruits (k = 1, …, K) were sampled from each of K sites and that each one was genotyped for L loci and assayed for t trace elements. Further assume that the same was done with n_s^e (s = 1 ,…, S) pre-dispersal individuals from each of S potential source populations (typically these will be the same as the pre-recruits sampling sites). Let $\mathbf{G} = (\mathbf{G}_{hk})$ be the observed multilocus genotypes of n^r recruits at L scored loci, where $\mathbf{G}_{hk} = (G_{hkl})$ and G_{hkl} denotes the genotype of recruit h at site k and locus l. The genetic data from the S source populations consists of allele counts $\mathbf{Z} = (\mathbf{z}_{sl})$, where $\mathbf{z}_{sl} = (z_{sla})$ and z_{sla} is the count of allele a at locus l for source population s. Similarly, let $\mathbf{X}^r = (\mathbf{X}^r_{hk})$ denote the elemental ratios of n^r recruits, where $\mathbf{X}^r_{hk} = (X^r_{hkj})$ denotes the vector of T (= t – 1 because data is expressed as the ratio

of each of T elements to a reference element, often calcium) elemental ratios for recruit h from site k. Microchemistry data from the S sources consists of elemental ratios $\mathbf{X}^e = (\mathbf{X}^e_{is})$ of pre-dispersal individuals, where $\mathbf{X}^e_{is} = (\mathbf{X}^e_{isj})$ denotes the vector of T elemental ratios for pre-dispersal egg i from source population s.

To minimise the ghost population effect we divide the study area into C grid cells and estimate maps of allele frequencies, $\mathbf{P}^{\text{ref}} = (p_{cl})$ (where $c = 1, \ldots, C$ and $s \in C$), and elemental ratios, $\mathbf{E}^{\text{ref}} = (e_c)$ from $\mathbf{Z}$ and $\mathbf{X}^e$, respectively, using spatial smoothing techniques (*see* Appendix).

The likelihood function for the estimation of connectivity is an extension of the Genetic Stock Identification approach presented by Gaggiotti *et al*. (2002, 2004) that includes microchemistry data. More precisely, the probability of observing an individual with genotype G_{hk} and profile $\mathbf{X}^r_{hk}$ is $P(\mathbf{G}_{hk}, \mathbf{X}^r_{hk}|\mathbf{p}_c, \mathbf{e}_\mathbf{c}, \mathbf{m}_k) = \sum_{c=1}^{C} P(\mathbf{G}_{hk}|\mathbf{p}_c) P(\mathbf{X}^r_{hk}|\mathbf{e}_\mathbf{c}) m_{kc}$, where $\mathbf{m}_k = (m_{kc})$ and m_{kc} is the probability that a recruit sampled at site k originated from grid cell c. $P(\mathbf{G}_{hk}|\mathbf{p}_c)$ is given in Gaggiotti *et al*. (2004: 813–814), while $P(\mathbf{X}^r_{hk}|\mathbf{e}_\mathbf{c})$ is multivariate normal. Thus, the likelihood function with both genetic and microchemistry data are $L(\mathbf{G}, \mathbf{X}^r|\mathbf{p}_c, \mathbf{e}_\mathbf{c}, \mathbf{m}_k) = \prod_{k=1}^{K} \prod_{h=1}^{n'_k} P(\mathbf{G}_{hk}, \mathbf{X}^r_{hk}|\mathbf{p}_c, \mathbf{e}_\mathbf{c}, \mathbf{m}_k)$, where $\mathbf{M} = (\mathbf{m}_k)$. The DAG describing this likelihood function is depicted in Fig. 5 (lower middle panel). Computation of the likelihood integrates over unobserved allele frequency and microchemistry surfaces using the Monte Carlo approximation described by Rundel *et al*. (2013).

The prior for $\mathbf{M}$ is defined using the results of a coupled biophysical-population model that produces point estimates of the probabilities of the sources of virtual recruits at each site (*see* below). Let the output of the model be noted $\mathbf{Y} = (y_{kc})$, where y_{kc} is the probability that a randomly-selected recruit at sampling site k is from grid cell source c. These results are incorporated by assuming that the vector $\mathbf{m} = (m_{kc})$ follows a Dirichlet distribution with parameters ρ_{kc} given by $\rho_{kc} = \rho_0 y_{kc}$, where ρ_0 determines the variance of the ρ_{kc} values and has to be estimated. We assume that the priors for ρ_{k0} are uniform between 0 and 100 for all k so as to obtain a weak prior. Thus, the prior for the vector of migration rates between all potential source populations and a given site k, $\mathbf{m}_k = (m_{kc})$, is given by

$$p(\mathbf{m}_k|\rho_{k1}, \rho_{k2}, \ldots, \rho_{kC}) = \left[\Gamma\left(\sum_{c=1}^{C} \rho_{kc}\right) \Big/ \prod_{c=1}^{C} \Gamma(\rho_{kc})\right] \prod_{c=1}^{C} m_{kc}^{\rho_{kc}-1}.$$

Having observed the data, our knowledge about the parameters is given by the posterior distribution:

$$P(\mathbf{M}, s_0|\mathbf{G}, \mathbf{X}^r, \mathbf{X}^{\text{ref}}, \mathbf{E}^{\text{ref}}, \mathbf{Y}) = L(\mathbf{G}, \mathbf{X}^r|\, \mathbf{P}^{\text{ref}}, \mathbf{E}^{\text{ref}}, \mathbf{M}) p(\mathbf{M}|\rho_0, \mathbf{Y}) p(\rho_0). \quad (1)$$

This posterior is estimated using standard MCMC approaches (e.g. Brooks 1998). The number of parameters that need to be estimated is very large, which can lead to convergence problems. In this regard, it is important to note that related methods we developed the past (Gaggiotti *et al*. 2002, 2004, Faubet & Gaggiotti 2008) require the estimation of similar number of parameters and involve manageable convergence times (a couple of days of computation time). Moreover, the method described above incorporates additional non-genetic data directly in the likelihood function, which should help avoid convergence issues. Finally, risks associated with computation time of MCMC approaches can be minimised by code parallelisation and use of techniques to improve mixing (*see* Brooks 1998), which we have implemented in other Bayesian methods we have developed (Faubet & Gaggiotti 2008, Foll & Gaggiotti 2008, Foll *et al*. 2014).

Concluding remarks

The adoption of the metapopulation paradigm among marine ecologists has lagged well behind that of terrestrial ecologists (Sale *et al*. 2006). According to Roughgarden (2006), the reason for this was an initial rejection due to long entrenched views among the community of marine ecologists, which was focused on small-scale manipulative experiments aimed at explaining intertidal community structure as driven by local dynamics and species interactions. However, this has changed substantially during the last decade. Indeed, metapopulation theory has now been integrated into marine ecology research and is being actively used as a helpful framework

for describing spatial population processes, for asking new research questions, and for planning the management and conservation of marine species (cf. Sale *et al*. 2006). Although Ilkka focused on terrestrial systems, marine ecologists are well aware of the great importance of his work for their discipline, as attested by the fact that he was invited to co-author the introductory chapter of the landmark book *Marine Metapopulations* (Kritzer & Sale 2006).

In this article, I have highlighted the many challenges faced by marine ecologists interested in the study of marine metapopulations and have mentioned some of the approaches that can be used to overcome these difficulties. In particular, I have focused on connectivity, an important attribute of metapopulations that represents an important knowledge gap in marine ecology (Lipcius *et al*. 2008, Gaines *et al*. 2010). There is an urgent need to develop approaches aimed at estimating connectivity in marine metapopulations because it is essential for the design of management and conservation plans for marine species. Indeed, marine ecosystems around the world are being subject to ever increasing demands for their important human benefits, usually described as ecosystem services. This increased demand comes from well-established sectors such as fishing and transportation that want to expand their activities as well as from emerging sectors such as renewable energy and offshore aquaculture (White *et al*. 2012). Minimising the combined impact of these activities requires the use of spatial management as an evolving paradigm for marine conservation policy. In particular, there has been a call for the use of networks of marine protected areas that can safeguard single species, as well as whole communities and even ecosystems, against anthropogenic perturbations while at the same time benefit fisheries by serving as a reservoir of recruits into fishing areas (Gaines *et al*. 2010, Harrison *et al*. 2012). In this spatial context, it is essential to be able to quantify exchanges of individuals among populations.

Besides estimating connectivity, the integrative approach can be used to validate the predictions of biophysical models. This can be done by comparing the fit of the Bayesian formulation 1, which incorporates predictions of the biophysical model, with that of a formulation where the prior for the migration rate is based on an island model or a stepping stone model. It is then possible to use model selection criteria such as DIC (Spiegelhalter *et al*. 2002) to determine the model that best fit the data. The framework can be further used to refine biophysical models in an iterative fashion (cf. Cowen & Sponaugle 2009). The fit of the Bayesian formulation 1 using a basic biophysical model can be compared with that of a formulation that uses a refined version of the basic model. If the modified version is a better fit, further refinements can be attempted until increased model complexity no longer improves the fit to the genetic and microchemistry data.

Besides the above-mentioned practical applications, the development and application of new integrative approaches will facilitate the study of the patterns, causes and consequences of spatial structuring in marine species and, therefore, will greatly contribute towards the advancement of marine ecology research.

Marine metapopulations will always be more difficult to study than the classic terrestrial metapopulations studied by Ilkka. For example, it would be very difficult and costly to carry out field experiments equivalent to those used to study spatial dynamics of colonisation in the Glainville fritillary and its two larval primary parasitoids (van Nouhuys & Hanski 2002). These field experiments required the manipulation of larvae and repeated direct observation of several experimental local populations, something that would be difficult to accomplish with the much smaller pelagic larvae that characterise the marine metapopulations considered here. Nevertheless, the methods described above can be modified to make them applicable in experimental settings, which would allow marine ecologists to carry out large-scale field experiments without the need of direct observations. This in turn will facilitate the application of metapopulation theory to marine systems.

Acknowledgements

First and foremost, I want to acknowledge the many things I learnt from Ilkka during the two and a half years I spent

in his Metapopulation Research Group as a senior research fellow. Secondly, I thank the editors of this special issue for giving me the opportunity to participate in this celebration of Ilkka's contributions to ecology and population biology. I would also like to acknowledge the many very useful comments made by an anonymous reviewer. Finally, I thank the support of the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland). MASTS is funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions.

References

Alberto, F., Raimondi, P. T., Reed, D. C., Watson, J. R., Siegel, D. A., Mitarai, S., Coelho, N. & Serrao, E. A. 2011: Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara Channel. — *Molecular Ecology* 20: 2543–2554.

Balding, D. J., Bishop, M. J. & Cannings, C. 2007: *Handbook of statistical genetics*. — John Wiley & Sons, Hoboken, NJ.

Beaumont, M. A. 1999: Detecting population expansion and decline using microsatellites. — *Genetics* 153: 2013–2029.

Beerli, P. 2004: Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. — *Molecular Ecology* 13: 827–836.

Beerli, P. & Felsenstein, J. 2001: Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. — *Proceedings of the National Academy of Sciences of the United States of America* 98: 4563–4568.

Bolker, B. M., Okuyama, T., Bjorndal, K. A. & Bolten, A. B. 2007: Incorporating multiple mixed stocks in mixed stock analysis: 'many-to-many' analyses. — *Molecular Ecology* 16: 685–695.

Botsford, L. W., Brumbaugh, D. R., Grimes, C., Kellner, J. B., Largier, J., O'Farrell, M. R., Ralston, S., Soulanille, E. & Wespestad, V. 2009: Connectivity, sustainability, and yield: bridging the gap between conventional fisheries management and marine protected areas. — *Reviews in Fish Biology and Fisheries* 19: 69–95.

Brooks, S. P. 1998: Markov chain Monte Carlo method and its application. — *Journal of the Royal Statistical Society D* 47: 69–100.

Broquet, T., Yearsley, J., Hirzel, A. H., Goudet, J. & Perrin, N. 2009: Inferring recent migration rates from individual genotypes. — *Molecular Ecology* 18: 1048–1060.

Burgess, S. C., Nickols, K. J., Griesemer, C. D., Barnett, L. A. K., Dedrick, A. G., Satterthwaite, E. V., Yamane, L., Morgan, S. G., White, J. W. & Botsford, L. W. 2014: Beyond connectivity: how empirical methods can quantify population persistence to improve marine protected-area design. — *Ecological Applications* 24: 257–270.

Caley, M. J., Carr, M. H., Hixon, M. A., Hughes, T. P., Jones, G. P. & Menge, B. A. 1996: Recruitment and the local dynamics of open marine populations. — *Annual Review of Ecology and Systematics* 27: 477–500.

Christie, M. R., Johnson, D. W., Stallings, C. D. & Hixon, M. A. 2010: Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. — *Molecular Ecology* 19: 1042–1057.

Christie, M. R., Meirmans, P. G., Gaggiotti, O. E., Toonen, R. J. & White, C. 2017: Disentangling the relative merits and disadvantages of parentage analysis and assignment tests for inferring population connectivity. — *ICES Journal of Marine Science* 74, doi:10.1093/icesjms/fsx044.

Cowen, R. K. & Sponaugle, S. 2009: Larval dispersal and marine population connectivity. — *Annual Review of Marine Science* 1: 443–466.

de Villemereuil, P. & Gaggiotti, O. E. 2015: A new FST-based method to uncover local adaptation using environmental variables. — *Methods in Ecology and Evolution* 6: 1248–1258.

Ehrlich, P. R., White, R. R., Singer, M. C., McKechnie, S. W. & Gilbert, L. E. 1975: Checkerspot butterflies — historical perspective. — *Science* 188: 221–228.

Faubet, P. & Gaggiotti, O. E. 2008: A new Bayesian method to identify the environmental factors that influence recent migration. — *Genetics* 178: 1491–1504.

Foll, M. & Gaggiotti, O. 2006: Identifying the environmental factors that determine the genetic structure of populations. — *Genetics* 174: 875–891.

Foll, M. & Gaggiotti, O. 2008: A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. — *Genetics* 180: 977–993.

Foll, M., Gaggiotti, O. E., Daub, J. T., Vatsiou, A. & Excoffier, L. 2014: Widespread signals of convergent adaptation to high altitude in Asia and America. — *American Journal of Human Genetics* 95: 394–407.

Gaggiotti, O. E. 1996: Population genetic models of source-sink metapopulations. — *Theoretical Population Biology* 50: 178–208.

Gaggiotti, O. E. 2004: Multilocus genotype methods for the study of colonisation processes. — In: Hanski, I. & Gaggiotti, O. E. (eds.), *Ecology, genetics and evolution of metapopulations*: 367–386. Elsevier-Academic Press, San Diego, CA.

Gaggiotti, O. E. & Smouse, P. E. 1996: Stochastic migration and maintenance of genetic variation in sink populations. — *American Naturalist* 147: 919–945.

Gaggiotti, O. E. & Vetter, R. D. 1999: Effect of life history strategy, environmental variability, and overexploitation on the genetic diversity of pelagic fish populations. — *Canadian Journal of Fisheries and Aquatic Sciences* 56: 1376–1388.

Gaggiotti, O. E., Brooks, S. P., Amos, W. & Harwood, J. 2004: Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. — *Molecular Ecology* 13: 811–825.

Gaggiotti, O. E., Jones, F., Lee, W. M., Amos, W., Harwood, J. & Nichols, R. A. 2002: Patterns of colonization in a metapopulation of grey seals. — *Nature* 416: 424–427.

Gaggiotti, O. E., Bekkevold, D., Jorgensen, H. B. H., Foll, M., Carvalho, G. R., Andre, C. & Ruzzante, D. E. 2009: Disentangling the effects of evolutionary, demographic, and environmental factors influencing the genetic struc-

ture of natural populations: Atlantic herring as a case study. — *Evolution* 63: 2939–2951.

Gaines, S. D., White, C., Carr, M. H. & Palumbi, S. R. 2010: Designing marine reserve networks for both conservation and fisheries management. — *Proceedings of the National Academy of Sciences of the United States of America* 107: 18286–18293.

Galindo, H. M., Olson, D. B. & Palumbi, S. R. 2006: Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. — *Current Biology* 16: 1622–1626.

Garza, J. C. & Williamson, E. G. 2001: Detection of reduction in population size using data from microsatellite loci. — *Molecular Ecology* 10: 305–318.

Gaylord, B. & Gaines, S. D. 2000: Temperature or transport? Range limits in marine species mediated solely by flow. — *American Naturalist* 155: 769–789.

Gelfand, A. E. & Vounatsou, P. 2003: Proper multivariate conditional autoregressive models for spatial data analysis. — *Biostatistics* 4: 11–25.

Gilg, M. R. & Hilbish, T. J. 2003: The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. — *Ecology* 84: 2989–2998.

Goovaerts, P. 1999: Geostatistics in soil science: state-of-the-art and perspectives. — *Geoderma* 89: 1–45.

Green, P. 1995: Reversible jump MCMC computation and Bayesian model determination. — *Biometrika* 82: 711–732.

Hanski, I. 1999: *Metapopulation ecology*. — Oxford University Press, Oxford.

Harrison, H. B., Williamson, D. H., Evans, R. D., Almany, G. R., Thorrold, S. R., Russ, G. R., Feldheim, K. A., van Herwerden, L., Planes, S., Srinivasan, M., Berumen, M. L. & Jones, G. P. 2012: Larval export from marine reserves and the recruitment benefit for fish and fisheries. — *Current Biology* 22: 1023–1028.

Hastings, A. & Botsford, L. W. 2006: Persistence of spatial populations depends on returning home. — *Proceedings of the National Academy of Sciences of the United States of America* 103: 6067–6072.

Hedgecock, D. & Pudovkin, A. I. 2011: Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. — *Bulletin of Marine Science* 87: 971–1002.

Hedgecock, D., Barber, P. H. & Edmands, S. 2007: Genetic approaches to measuring connectivity. — *Oceanography* 20: 70–79.

Hellberg, M. E. 2009: Gene flow and isolation among populations of marine animals. — *Annual Review of Ecology, Evolution, and Systematics* 40: 291–310.

Hey, J. & Nielsen, R. 2007: Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. — *Proceedings of the National Academy of Sciences of the United States of America* 104: 2785–2790.

Hoban, S., Bertorelle, G. & Gaggiotti, O. E. 2012: Computer simulations: tools for population and evolutionary genetics. — *Nature Reviews Genetics* 13: 110–122.

Kingman, J. F. C. 1982: On the genealogy of large populations. — *Journal of Applied Probability* 19A: 27–43.

Kittlein, M. J. & Gaggiotti, O. E. 2008: Interactions between environmental factors can hide isolation by distance patterns: a case study of Ctenomys rionegrensis in Uruguay. — *Proceedings of the Royal Society B* 275: 2633–2638.

Kritzer, J. P. & Sale, P. F. 2006: *Marine metapopulations*. — Elsevier Academic Press, San Diego, CA.

Kuhner, M. K., Yamato, J. & Felsenstein, J. 1998: Maximum likelihood estimation of population growth rates based on the coalescent. — *Genetics* 149: 429–434.

Kuhner, M. K. 2006: LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. — *Bioinformatics* 22: 768–770.

Lindgren, F., Rue, H. & Lindstrom, J. 2011: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. — *Journal of the Royal Statistical Society B* 73: 423–498.

Lipcius, R. N., Eggleston, D. B., Schreiber, S. J., Seitz, R. D., Shen, J., Sisson, M., Stockhausen, W. T. & Wang, H. V. 2008: Importance of metapopulation connectivity to restocking and restoration of marine species. — *Reviews in Fisheries Science* 16: 101–110.

Lopez-Duarte, P. C., Carson, H. S., Cook, G. S., Fodrie, F. J., Becker, B. J., DiBacco, C. & Levin, L. A. 2012: What controls connectivity? An empirical, multi-species approach. — *Integrative and Comparative Biology* 52: 511–524.

Luikart, G. & Cornuet, J. M. 1998: Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. — *Conservation Biology* 12: 228–237.

Mitarai, S., Siegel, D. A. & Winters, K. B. 2008: A numerical study of stochastic larval settlement in the California Current system. — *Journal of Marine Systems* 69: 295–309.

Mitarai, S., Siegel, D. A., Watson, J. R., Dong, C. & McWilliams, J. C. 2009: Quantifying connectivity in the coastal ocean with application to the Southern California Bight. — *Journal of Geophysical Research, Oceans* 114:

Okuyama, T. & Bolker, B. M. 2005: Combining genetic and ecological data to estimate sea turtle origins. — *Ecological Applications* 15: 315–325.

Palumbi, S. & Hedgecock, D. 2005: The life of the sea: implications of marine population biology to conservation policy. — In: Norse, E. A. & Crowder, L. B. (eds.), *Marine conservation biology: the science of maintaining the sea's biodiversity*: 33–46. Island Press, Washington.

Pella, J. & Masuda, M. 2001: Bayesian methods for analysis of stock mixtures from genetic characters. — *Fishery Bulletin* 99: 151–167.

Perrier, C., Daverat, F., Evanno, G., Pécheyran, C., Baglinière, J.-L., Roussel, J.-M. & Campana, S. 2011: Coupling genetic and otolith trace element analyses to identify river-born fish with hatchery pedigrees in stocked Atlantic salmon (*Salmo salar*) populations. — *Canadian Journal of Fisheries and Aquatic Sciences* 68: 977–987.

Riginos, C., Crandall, E., Liggins, L., Bongaerts, P. & Treml, E. 2016: Navigating the currents of seascape genomics: how spatial analyses can augment marine population genomic studies. — *Current Zoology* 62: 581–601.

Rivera, M. A. J., Andrews, K. R., Kobayashi, D. R., Wren, J. L. K., Kelley, C., Roderick, G. K. & Toonen, R. J. 2011: Genetic analyses and simulations of larval dispersal reveal distinct populations and directional connectivity across the range of the Hawaiian grouper (*Epinephelus quernus*). — *Journal of Marine Biology* 2011: 1–11.

Roughgarden, J. 2006: Foreword. — In: Kritzer, J. P. & Sale, P. F. (eds.), *Marine metapopulations*: xvii–xix. Elsevier Academic Press, San Diego, CA.

Rundel, C. W., Wunder, M. B., Alvarado, A. H., Ruegg, K. C., Harrigan, R., Schuh, A., Kelly, J. F., Siegel, R. B., DeSante, D. F., Smith, T. B. & Novembre, J. 2013: Novel statistical methods for integrating genetic and stable isotope data to infer individual-level migratory connectivity. — *Molecular Ecology* 22: 4163–4176.

Sale, P. F., Hanski, I. & Kritzer, J. P. 2006: The merging of metapopulation theory and marine ecology: Establishing the historical context. — In: Kritzer, J. P. & Sale, P. F. (eds.), *Marine metapopulations*: 3–28. Elsevier Academic Press, San Diego, CA.

Selkoe, K. A. & Toonen, R. J. 2011: Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. — *Marine Ecology Progress Series* 436: 291–305.

Selkoe, K. A., Gaines, S. D., Caselle, J. E. & Warner, R. R. 2006: Current shifts and kin aggregation explain genetic patchiness in fish recruits. — *Ecology* 87: 3082–3094.

Selkoe, K. A., Gaggiotti, O. E., Bowen, B. W., Toonen, R. J. & Lab, T. 2014: Emergent patterns of population genetic structure for a coral reef community. — *Molecular Ecology* 23: 3064–3079.

Selkoe, K. A., Watson, J. R., White, C., Ben Horin, T., Iacchei, M., Mitarai, S., Siegel, D. A., Gaines, S. D. & Toonen, R. J. 2010: Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. — *Molecular Ecology* 19: 3708–3726.

Shanks, A. L. 2009: Pelagic larval duration and dispersal distance revisited. — *Biological Bulletin* 216: 373–385.

Shanks, A. L., Grantham, B. A. & Carr, M. H. 2003: Propagule dispersal distance and the size and spacing of marine reserves. — *Ecological Applications* 13: S159–S169.

Siegel, D. A., Mitarai, S., Costello, C. J., Gaines, S. D., Kendall, B. E., Warner, R. R. & Winters, K. B. 2008: The stochastic nature of larval connectivity among nearshore marine populations. — *Proceedings of the National Academy of Sciences of the United States of America* 105: 8974–8979.

Slatkin, M. 2005: Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. — *Molecular Ecology* 14: 67–73.

Smith, S. J. & Campana, S. E. 2010: Integrated stock mixture analysis for continous and categorical data, with application to genetic–otolith combinations. — *Canadian Journal of Fisheries and Aquatic Sciences* 67: 1533–1548.

Smouse, P. E., Waples, R. S. & Tworek, J. A. 1990: A genetic mixture analysis for use with incomplete source population-data. — *Canadian Journal of Fisheries and Aquatic Sciences* 47: 620–634.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R. & van der Linde, A. 2002: Bayesian measures of model complexity and fit. — *Journal of the Royal Statistical Society B* 64: 583–616.

Tanner, S. E., Pérez, M., Presa, P., Thorrold, S. R. & Cabral, H. N. 2014: Integrating microsatellite DNA markers and otolith geochemistry to assess population structure of European hake (*Merluccius merluccius*). — *Estuarine, Coastal and Shelf Science* 142: 68–75.

Thorrold, S. R., Zacherl, D. C. & Levin, L. A. 2007: Population connectivity and larval dispersal using geochemical signatures in calcified structures. — *Oceanography* 20: 80–89.

Thorrold, S. R., Jones, G. P., Hellberg, M. E., Burton, R. S., Swearer, S. E., Neigel, J. E., Morgan, S. G. & Warner, R. R. 2002: Quantifying larval retention and connectivity in marine populations with artificial and natural markers. — *Bulletin of Marine Science* 70: 291–308.

van Nouhuys, S. & Hanski, I. 2002: Colonization rates and distances of a host butterfly and two specific parasitoids in a fragmented landscape. — *Journal of Animal Ecology* 71: 639–650.

Vounatsou, P., Smith, T. & Gelfand, A. E. 2003: Spatial modelling of gene frequencies in the presence of undetectable alleles. — *Journal of Applied Statistics* 30: 49–62.

Waples, R. S. 1998: Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. — *Journal of Heredity* 89: 438–450.

Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B. & Stephens, M. 2004: Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. — *Proceedings of the National Academy of Sciences of the United States of America* 101: 14847–14852.

White, C., Halpern, B. S. & Kappel, C. V. 2012: Ecosystem service tradeoff analysis reveals the value of marine spatial planning for multiple ocean uses. — *Proceedings of the National Academy of Sciences of the United States of America* 109: 4696–4701.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C. & Toonen, R. J. 2010: Ocean currents help explain population genetic structure. — *Proceedings of the Royal Society B* 277: 1685–1694.

Wilson, G. A. & Rannala, B. 2003: Bayesian inference of recent migration rates using multilocus genotypes. — *Genetics* 163: 1177–1191.

Zacherl, D. C., Manriquez, P. H., Paradis, G., Day, R. W., Castilla, J. C., Warner, R. R., Lea, D. W. & Gaines, S. D. 2003: Trace elemental fingerprinting of gastropod statoliths to study larval dispersal trajectories. — *Marine Ecology Progress Series* 248: 297–303.

Appendix

There are several approaches for the estimation of maps of allele frequencies and elemental rations. These include kriging (e.g. Goovaerts 1999), Multivariate Conditional Autoregressive models (MCAR; Gelfand & Vounatsou 2003), and Stochastic Partial Differential Equation approach (SPDE, Lindgren *et al*. 2011). Here I use a hierarchical Bayesian approach with a Multivariate Conditional Autoregressive prior, [MCAR(α, Λ); Gelfand & Vounatsou 2003] for the spatial trend. Allele frequency counts $\mathbf{Z}$ are modelled using a multinomial(n_s^e, $\mathbf{p}_{cl}$) likelihood with $\log(p_{cla}/p_{clK_l}) = \pi_{la} + \phi_{cla}$. The observed elemental ratios $\mathbf{X}^e$ are modelled with a multivariate normal $N(\mu_c, \Sigma)$ with $\mu_c = \rho_c + \Phi_c$. The spatially structured random effects ϕ_{cla} and Φ_c are modelled as conditional autoregressive priors (MCAR(α,Λ)). Briefly, $\Phi \sim N(0,[(\mathbf{D} - \alpha\mathbf{W}) \otimes \mathbf{\Lambda}]^{-1})$, where $\alpha \in (-1,1)$ is the smoothing parameter, $\mathbf{\Lambda}$ is a $K_l \times K_l$ (for allele frequencies) or $T \times T$ (for elemental ratios) positive definite and symmetric matrix, $\mathbf{D}$ = Diag(m_i) with m_i being the number of neighbours of region i, and $\mathbf{W}$ denotes the adjacency matrix of the map (i.e., $w_{ii} = 0$; $w_{ij} = 1$ if i is adjacent to j, and 0 otherwise).The symbol $\otimes$ is the Kronecker product. The DAGs of this Bayesian model are shown in Fig. 5 (left- and right-hand-side panels for genetic and microchemistry data, respectively).