Journal of Theoretical Biology ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Population genetics on islands connected by an arbitrary network: An analytic approach

Q1 George W.A. Constable [1], Alan J. McKane

Q2 *Theoretical Physics Division, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK*

HIGHLIGHTS

- Migration is modelled for organisms on demes connected by an arbitrary network.
- This individual based model is simplified via the diffusion approximation.
- A method of obtaining a one-dimensional effective theory is presented.
- The fixation time and probability are calculated from the effective system.
- Migration–selection balance and a specific hub topology are investigated.

ARTICLE INFO

Article history:
Received 13 February 2014
Received in revised form
9 May 2014
Accepted 21 May 2014

Keywords:
Metapopulation
Migration
Selection
Moran model
Fast-mode reduction

ABSTRACT

We analyse a model consisting of a population of individuals which is subdivided into a finite set of demes, each of which has a fixed but differing number of individuals. The individuals can reproduce, die and migrate between the demes according to an arbitrary migration network. They are haploid, with two alleles present in the population; frequency-independent selection is also incorporated, where the strength and direction of selection can vary from deme to deme. The system is formulated as an individual-based model and the diffusion approximation systematically applied to express it as a set of nonlinear coupled stochastic differential equations. These can be made amenable to analysis through the elimination of fast-time variables. The resulting reduced model is analysed in a number of situations, including migration–selection balance leading to a polymorphic equilibrium of the two alleles and an illustration of how the subdivision of the population can lead to non-trivial behaviour in the case where the network is a simple hub. The method we develop is systematic, may be applied to any network, and agrees well with the results of simulations in all cases studied and across a wide range of parameter values.

1. Introduction

The founders of population genetics, who reconciled Mendelian genetics and Darwinian evolution, did so through the use of mathematical models which were frequently deterministic and which involved the key processes of mutation, migration and Q3 selection (Hartl and Clark, 2007). Crucially, Fisher and Wright added genetic drift to this list by considering simple stochastic processes in systems where the population size, N, was finite (Fisher, 1930; Wright, 1931). Subsequent work tended to follow their original approach, assuming discrete generations and discrete state variables (corresponding to the number of individuals in the population carrying one particular type of allele) (Ewens, 1969). However adding more complexity to the models in the form of selection and migration makes this approach, based as it is on Markov chains, very unwieldy, and mathematical progress can be difficult (Ewens, 2004).

The solution to this dilemma is to take a mesoscopic perspective. That is, one uses a diffusion approximation in which the (discrete) number of individuals carrying a particular allele, n, is replaced by the (continuous) fraction $x=n/N$ carrying that allele (Crow and Kimura, 2009). Assuming x is continuous and is usually a good approximation for reasonably large N. This approximation, although originally suggested by Fisher (1922), was popularised by Crow and Kimura (1956); Kimura (1994), and proved to be a powerful tool and the starting point for many studies of more complex processes in population genetics (Kimura, 1955, 1964; Kimura and Weiss, 1964). Nevertheless, some quite straightforward models give rise to rather complicated equations even within

---

E-mail addresses: george.constable@postgrad.manchester.ac.uk (G.W.A. Constable), alan.mckane@manchester.ac.uk (A.J. McKane).

[1] Tel.: +44 1612754192; fax: +44 1612750480.

http://dx.doi.org/10.1016/j.jtbi.2014.05.033
0022-5193/© 2014 Published by Elsevier Ltd.

Please cite this article as: Constable, G.W.A., McKane, A.J., Population genetics on islands connected by an arbitrary network: An analytic approach. J. Theor. Biol. (2014), http://dx.doi.org/10.1016/j.jtbi.2014.05.033

this approximation. For instance, a model with migration between $\mathcal{D}$ subpopulations (demes) which includes selection, leads to a nonlinear partial differential equation in $\mathcal{D}$ variables for the probability distribution function (pdf) (Blythe and McKane, 2007) which seems quite intractable.

In this paper we show that equations such as these are in fact not as intractable as they seem, and in many cases can be reduced to a differential equation for a single variable, which can be straightforwardly analysed. The methodology which allows this reduction is based on the elimination of fast variables and relies on a number of factors. First, rather than formulating the diffusion approximation in terms of a partial differential equation for the pdf, it is more useful to work in terms of an equivalent stochastic differential equation (SDE) (McKane, 2009). This is a direct generalisation of the equation describing the deterministic dynamics (McKane et al., 2014). Second, this formulation of the system dynamics allows us to use much of the same intuition that is used to understand the deterministic process. In particular, we will see that in many of the cases of interest the dynamics can be divided into 'fast' variables and 'slow' variables (Serra et al., 1986). After a short time, the dynamics of the fast variables may be ignored, since they have decayed to their stationary values; all the dynamics is contained in the few (in our case, frequently only one) slow mode. Third, the method is systematic and intuitive, and also applies to more general systems, such as those with many alleles and those involving other processes.

Our intention here is to apply this methodology to the case of a simple two-allele, haploid model with migration, even though it may be extended to deal with more complicated systems. Population genetics models featuring migration were first considered by Wright (1931), who looked at what is now often referred to as the standard island model (Rousset, 2004). Instead of the well-mixed population of size N which had been previously studied, he considered a set of $\mathcal{D}$ well-mixed subpopulations. With migratory individuals being chosen from the global population, there was no spatial structure assumed, only interactions between the various subpopulations. These are called demes in the genetic context (Hartl and Clark, 2007), although we will also use the term islands in this paper to refer to areas in which there is no spatial structure, but between which interactions can occur. Within the modern nomenclature of ecology, this collection of demes effectively describes a metapopulation (Levins, 1969), or 'population of populations'. The case of one deme therefore should reduce to the well-mixed case.

Subsequently, the formulation of the stepping stone model (Kimura and Weiss, 1964) introduced what was a very simple topology into the description of migration; the islands were ordered, with migration from island i only allowed onto islands $i-1$ and $i+1$. Maruyama compared selection in the stepping stone model and in the island model (Maruyama, 1969). He concluded that if selection was additive (i.e. frequency-independent) and local, with the same selection pressure in all demes, then deme population structure played no role. In other words, the population behaved approximately as a well-mixed and spatially homogeneous population, with a size equivalent to the sum of the deme population sizes.

The books by Ewens (2004) and Moran (1962) describe variants of these models, analyses and conclusions, but for our purposes the next result of note is the work of Nagylaki (1980) who studied what would be in modern terminology an arbitrary network of demes. He constructed a migratory model with discrete generations (of the Wright-Fisher type) in the limit of strong migration, that is where the probability of a migration event is of the same order as that of a birth or death event. The effect of this assumption was to create a separation of timescales in the Markov chain. Nagylaki then employed his earlier results on Markov chains with timescale separation (Ethier and Nagylaki, 1980) to derive an equation in the diffusion limit. Starting with a neutral model, it was concluded that in the long-time limit the population behaved as if it were well-mixed, but with an effective population size less than or equal to the total unstructured population. Equality was shown to be achieved only if the migration matrix was symmetric.

The analysis which is used to reach these conclusions seems, to us at least, difficult to follow, with some parts of the proof relying on results from the theory of Markov chains and others relying on the nature of the diffusion approximation. Nevertheless, the results of the analysis are widely quoted and utilised. The work was extended (Nagylaki, 1980) to the case of different selection strengths on different islands, showing that once again the population was well-mixed with an effective population size, but now also with an effective selection coefficient. The situation where the selection on different islands operates in different directions was not discussed. In this case within certain parameter ranges a stable fixed point emerges, allowing coexistence of alleles deleterious in some demes, but advantageous in others. The deterministic implications of this have been discussed in Moran (1962), Eyland (1971), and Nagylaki and Lou (2008).

In the wake of this work a number of studies were carried out and a plethora of results obtained, all with a variety of different approximations and objectives. Several of these were concerned with an effort to determine the effective population size, which amounts to a rescaling of time for the structured population. Here we will avoid the temptation to describe the results that we obtain in terms of an effective population size, due to its amorphous definition, and at times misleading designation. We refer the reader to Charlesworth (2009) for a review of such work. Nagylaki's results on the diffusion limit of Markov chains with a separation of timescale have also been employed in Lessard (2009), where they were used to extend the results to systems where fitness is not just additive, but frequency dependent. In this case, however, it becomes important to specify carefully between whom individuals compete (on their own island, on islands connected to their own, or the whole population) as these can sometimes lead to different results. In turn, other work has focussed instead on the effect of migration on local deme properties (Blanquart et al., 2012).

The approach that we adopt in this paper will be to carefully define the model in terms of individuals (i.e. at the microscale). We will work within the context of continuous time Markov chains, that is, in terms of master equations (Gardiner, 2009). We will therefore not assume non-overlapping generations as in the Wright-Fisher model, but instead work with the continuous time Moran process (Ewens, 2004). As is well known, these two processes are essentially identical at medium to long times, up to a redefinition of time scales. The master equations for the Moran process involving birth, death, migration, mutation, and so on can be written down in a systematic way (Blythe and McKane, 2007), although it is too complicated to allow analytic progress to be made. As we have indicated the key to further progress is to write down a mesoscopic description which is achieved through a diffusion approximation itself derived by expanding the master equation in inverse deme size.

A related set of questions to those that we ask here have been studied in a model of language evolution (Baxter et al., 2006, 2008, 2012), in which each island is mapped on to a speaker having two different linguemes (different ways of saying the same thing) whose concentrations are modified through interaction events (analogous to migration events). While this model has similar features to the one we discuss here, it is distinct, and the methods of analysis and the final results are also different. We have already mentioned the work of Nagylaki (1980). Once again our model, analysis and conclusions differ. Throughout this paper we will

Please cite this article as: Constable, G.W.A., McKane, A.J., Population genetics on islands connected by an arbitrary network: An analytic approach. J. Theor. Biol. (2014), http://dx.doi.org/10.1016/j.jtbi.2014.05.033