The fixation probability of a beneficial mutation in a geographically structured population

Bahram Houchmandzadeh, Marcel Vallade

The fixation probability of a beneficial mutation in a geographically structured population.

Bahram Houchmandzadeh & Marcel Vallade
CNRS & Grenoble Universités, Laboratoire Interdiciplinaire de Physique, BP87, 38402 St-Martin d'Hères Cedex.

One of the most fundamental concepts of evolutionary dynamics is the "fixation" probability, *i.e.* the probability that a gene spreads through the whole population. Most natural communities are geographically structured into habitats exchanging individuals among themselves. The topology of the migration patterns is believed to influence the spread of a new mutant, but no general analytical results were known for its fixation probability. We show how for large populations, the fixation probability of a beneficial mutation can be evaluated for any migration pattern between local communities. Specifically, we demonstrate that for large populations, in the framework of the Voter Model of the Moran model, the fixation probability is always smaller or at best equal to the fixation probability of a non-structured population. In the "Invasion Processes" version of the Moran model, the fixation probability can exceed that of a non-structured population ; our method allows migration patterns to be classified according to their amplification effect. The theoretical tool we have developed to perform these computations uses the fixed points of the probability generating function which are obtained by a system of second order *algebraic* equations.

I. INTRODUCTION.

One of the main concepts of population genetics is the fixation probability[1] : a beneficial gene does not always spread, but has a probability π_f to take over the whole population. The fixation probability can be computed for a non-structured population[2, 3], where the progeny of an individual can replace any other one in the community with equal probability, and depends monotonously on the fitness of the new allele and the size of the community N. Natural communities however are geographically extended and subdivided into colonies that exchange individuals: the progeny of an individual in a given patch compete with the local descendants of other individuals in the same patch, and also with migrants from other nearby patches. Evaluating the fixation probability of these subdivided communities is of fundamental importance (see below), but no general results analogous to that of Moran were known for them.

Evolutionary dynamics is a competition between deterministic selection pressure and stochastic events due to random sampling from one generation to the other. For small populations, stochastic events drive the evolutionary dynamics : a large beneficial mutation with fitness say 1.1, has a probability of only 0.29 to take over a community of size $N = 4$, barely superior to a neutral mutation. For large non-structured population, selection is the main force : the same beneficial allele in a population of size $N = 10^6$ has a fixation probability of 0.09, orders of magnitude higher than the fixation probability of a neutral allele. It is then of fundamental importance to determine what are the driving forces in a large *subdivided* population and how the fixation probability is affected by this division and the pattern of migration routes.

Evolutionary problems related to geograhical subdivision have been investigated from the begining of the field of population genetics [4–7]. Maruyama [8, 9] was the first author to cast the problem of fixation probability into a rigorous generalized Moran model (see fig. I.1) to which he applied standard techniques of branching processes and Markov chains. Under the assumption that the progeny in a patch descend from only local individuals, he obtained an astonishing result: the migration pattern between the patches has *no influence* on the fixation probability of a gene and, in this respect, the population can be considered as non-structured. The simplifying assumptions of Maruyama have been criticized by many authors (see for example [1, 10, 11]), but the Maruyama result on fixation probability was largely held to be true.

Recently, Lieberman *et al.*[12] were able to show that, for patches of size 1 (where all the progeny descend from neighboring patches), some peculiar migration patterns can considerably amplify the effective fitness of a beneficial allele for the "invasion" version of the Moran model . Antal et al.[13, 14] analyzed a similar problem and showed that the location where a new mutant is introduced can play an important role in its fixation probability. They also pointed out the fundamental difference between the two versions of the generalized Moran model, called invasion process and voter model, in their amplifying capabilities. These new findings have spurred much interest in the field of evolutionary dynamics on graphs[11] and have applications beyond population genetics in domains such as the spread of epidemics[15] or cancer[16, 17] or the appearance of collective behavior[18].

In this article, we give the general solution of the Moran model on graph (see Fig. I.1) introduced by Maruyama without his simplifying assumptions. The mathematical tool we develop enables us to classify migration patterns according to their effect on the fixation probability and to make clear the differences between the two versions of the Moran model : the Voter Model (die first, then be replaced) and the Invasion Process (duplicate, then replace another). We demonstrate that when the selection pressure is not too small, in the Voter

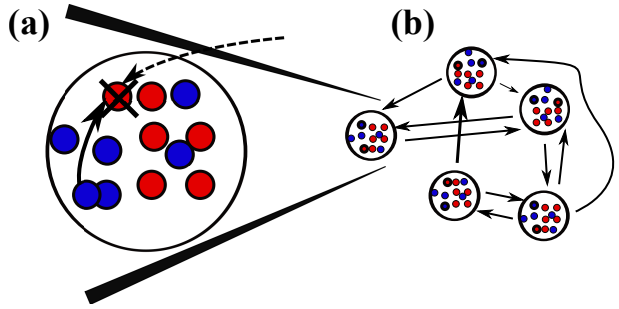

Figure I.1. (a) The (Voter version of) Moran model : in a community of fixed size N, death events occur continuously. When an individual dies, it is immediately replaced by the duplicate of another one, chosen at random. (b) The Maruyama islands, or the Moran model on graph : the meta community is composed of M islands, each of size N_i. When an individual dies in island i, it is replaced either by the progeny of a local individual, or by one from an other connected island. The connectivity between islands is given by the matrix m_{ki} , the probability of migration from island k to island i. In the Invasion process (IP) version of this model, one individual is first selected for duplication and then replaces another one.

Model framework, spatial subdivision always diminishes the chances of success of a beneficial allele : the fixation probability of a beneficial mutation in a subdivided population is always smaller or at best equal to that of a non-structured population of the same size. The equality happens only for "balanced" migration pattern, *i.e.* when the number of migrants send and received are balanced for *each* patch.

For the Invasion Process, the fixation probability of a new mutant in a subdivided population can exceed the chances of success of the same mutant in a non-structured populations. Our method allows migration patterns to be partitioned according to their effect on fixation probability. For large populations there is no absolute upper bound for the fixation probability, which can approach unity. However, for a given value of imbalance between migrants sent and received by each island, there exists an upper bound to the fixation probability which is higher than those of a non-structured population ; among the patterns with the same imbalance, some topologies can amplify the fixation probability to this higher value.

The mathematical method we develop is an extension of the tool we recently developed [19] for non-structured populations and is based on the dynamics of probability generating function (dPGF). The power of the method we present resides in its simplicity. The fixation probability is uniquely related to the fixed points of a partial differential equation and these fixed points are the roots of a system of algebraic second order equations. In this framework, the fixation probability acquires a particularly simple geometrical interpretation. We stress that our results are not limited to migration between neighboring patches, but encompass any migration pattern.

This article is organized as follow. We derive first the general outline of the dPGF method. The next section is dedicated to the application of the dPGF method to the Voter Model version of the Moran process. The following section treats the IP process following the same lines of argument and highlights fundamental differences between these two models. The main points are summarized in the conclusion.

II. THE DPGF EQUATION AND ITS FIXED POINTS.

We have recently shown[19] that the differential equation governing the probability generating function (PGF) can be used to compute efficiently the behavior of the Moran process for a non-structured population. We recall this method, which we will generalize to the Moran process on graph. Consider a community of fixed size N, where n haploid individuals carry the allele A with the fitness $1+s$ (compared to 1 for the others). Death events occur continuously and when an individual dies, it is immediately replaced by the duplicate of another one, chosen at random among the others, the randomness of this choice being weighted by the parent's fitness. The transition probabilities for the A allele to decrease or increase its population number by one unit during a short time dt read

$$W^-(n) = \mu n(N-n)/N \tag{II.1}$$
$$W^+(n) = \mu(1+s)(N-n)n/N \tag{II.2}$$

where μ is the death rate. Let us call $P(n,t)$ the probability density of observing n $A-$individuals at time t, and set $\phi(z,t) = \sum_n P(n,t)z^n$ as its PGF. Then, in units of generation time ($\mu/N = 1$), ϕ obeys the following equation :

$$c\frac{\partial \phi}{\partial t} = (1-z)(c-z)\frac{\partial}{\partial z}\left(N\phi - z\frac{\partial \phi}{\partial z}\right) \tag{II.3}$$

where $c = 1/(1+s)$. Note that for beneficial mutation, which we consider in this article, $s > 0$ and $c < 1$. The solution of this equation is given elsewhere [19]. The stationary solution $\phi_s(z) = \phi(z, t \to \infty)$ however can easily be used to compute the fixation probability. We observe from expression (II.3) that $\phi_s(z)$ obeys the first order ODE $N\phi_s - d\phi_s/dz = K$, the solution of which is

$$\phi_s(z) = \pi_0 + \pi_f z^N \tag{II.4}$$

Note that as $\phi(1,t) = 1$, $\pi_0 + \pi_f = 1$; moreover, as $\pi_0 = \phi_s(0)$ and $\pi_f = (1/N!)\phi_s^{(N)}(0)$, they represent respectively the loss of allele and the fixation probability.

The heart of our method lies in the fact that $z = c$ is a fixed point of eq.(II.3) ; if at time $t = 0$, there were exactly n_0 $A-$alleles present, for all subsequent times $\phi(c,t) = \phi(c,0) = c^{n_0}$. Inserting this relation into equation (II.4) leads to

$$\pi_f = \frac{1-c^{n_0}}{1-c^N} \tag{II.5}$$

which is the celebrated Moran result[2], usually obtained by more complicated means.

Consider now the Moran process on a graph : the community is subdivided into M colonies (or islands in Maruyama's terms), the size of $i-$th island being N_i. When an individual dies on the island i, it has a probability m_{ki} of being replaced by the progeny of an individual on the island k. The matrix of m_{ki}, the migration probability from island k to island i, uniquely specifies the migration pattern between islands. Let n_i be the number of A on island i. The transition probabilities for A to decrease or increase in number by one on this island is a small modification of expressions (II.1,II.2) :

$$W_i^-(n_i) = \mu n_i \sum_{k=1}^{M} m_{ki}(N_k - n_k)/N_k \tag{II.6}$$

$$W_i^+(n_i) = \mu(1+s)\,(N_i - n_i) \sum_{k=1}^{M} m_{ki} n_k/N_k \tag{II.7}$$

Let us make two remarks before going farther. Firstly, there are two possibilities for the Moran model on a graph. In the voter model (VM), first an individual dies and then is replaced by the progeny of an another individual. This is the case for example in a tropical forest where space is the limiting resource and a seed can grow only when an adult tree dies[20]. In the invasion process (IP), used for example for cancer propagation [16] first an individual duplicates and then the progeny replaces another one. The N_k in the denominator of eqs. (II.6-II.7) must be replaced by N_i ; moreover, the normalization constraint is different for these two cases:

$$\sum_{k=1}^{M} m_{ki} = 1 \;\; i = 1, ..., M \text{ (VM)} \tag{II.8}$$

$$\sum_{i=1}^{M} m_{ki} = 1 \;\; k = 1, ..., M \text{ (IP)} \tag{II.9}$$

The difference between the two versions becomes important for unbalanced graphs (see below). Secondly, in order to lighten the notation we set all islands to the same population size $N_i = N$. All the results obtained here can be trivially generalized to variable island size (see Mathematical details VI D).

We call $P(n_1, n_2, ..., n_M; t)$ the probability of observing n_i individuals on island i at time t and let $\phi(z_1, ..., z_M; t)$ be its PGF. It can be shown (see Mathematical Details VI A) that ϕ obeys the following equation for the VM :

$$\frac{\partial \phi}{\partial t} = \sum_i (z_i - 1) \left\{ N \sum_k m_{ki} z_k \partial_k \; - Nc\partial_i \right.$$
$$\left. - \; (z_i - c)\partial_i \sum_k m_{ki} z_k \partial_k \right\} \phi \tag{II.10}$$

where $\partial_i = \partial/\partial z_i$. The above equation can be treated along the same lines as before. We first note that the stationary solution for a connected graph reads

$$\phi_s(z_1, ..., z_M) = \pi_0 + \pi_f \left(z_1 z_2 ... z_M\right)^N \tag{II.11}$$

where again, π_0 and π_f stand for the loss and fixation probabilities. If we could find a fixed point $\zeta = (\zeta_1, ...\zeta_M)$ of eq. (II.10), then we would immediately deduce the fixation probability

$$\pi_f = \frac{1 - \phi(\zeta_1, ..., \zeta_M; t = 0)}{1 - \left(\zeta_1 ... \zeta_M\right)^N} \tag{II.12}$$

We see here how the general problem of finding the fixation probability is reduced to finding the fixed point of a PDE. A priori, the existence of such a fixed point ζ cannot be taken for granted. We shall see in the next section that a true fixed point actually exists for certain classes of connectivity and that a "quasi-fixed point" exists in the general case, when the global population is large enough. In this article, we are interested in the fixation probability of one mutant appearing on an island chosen at random. Therefore the initial condition is

$$P(n_1, ..., n_M; t = 0) = (1/M) \sum_{i=1}^{M} \delta_{n_i, 1} \tag{II.13}$$

and $\phi(z_1, ...z_M; t = 0) = (1/M) \sum_i z_i$. The fixation probability is then given by

$$\pi_f = \frac{1 - (1/M) \sum_i \zeta_i}{1 - \left(\zeta_1 ... \zeta_M\right)^N} \tag{II.14}$$

We call attention to the versatility of this method. Appearance of a mutant on a specific island, for example, can be treated with the same ease using (II.12).

III. FIXED POINTS OF THE VOTER MODEL.

a. Exact results. We first concentrate on the VM ; the IP will be discussed in the next section and treated as a natural extension. We first investigate for which kind of connectivity matrix the fixation probability is the same as for a non-structured population. Obviously $\zeta_i = c$ yields the non-structured fixation probability. Inserting this specific value into eq.(II.10) and arranging the summation orders, we get

$$\frac{\partial \phi(c, ...c)}{\partial t} = Nc(c-1) \left[\sum_i (T_i - 1)) \partial_i \right] \phi \Big|_{z_i = c} \tag{III.1}$$

where

$$T_i = \sum_k m_{ik} \tag{III.2}$$

is called the temperature of node i by Lieberman[12]. $T_i > 1$ represents a node which sends out more migrants

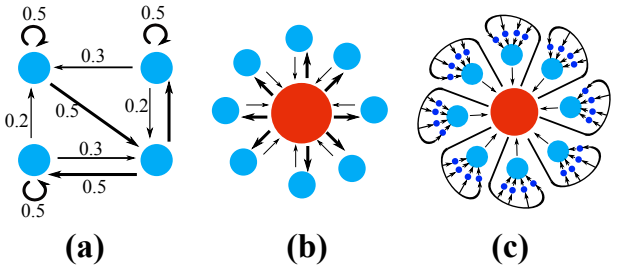

Figure III.1. Some particular special cases with simple fixed point. (a) Isothermal graph, where the number of migrants sent and received by each graph is equal (number on arrows denotes migration probabilities). (b) Star : one central island communicates exclusively with a group of P peripheral islands (see Mathematical details. (c) super star, where each secondary island possesses Q tertiary ones. migration routes are $1 \to 3 \to 2 \to 1$ (see VI B)

than it receives. If all the nodes are balanced, *i.e.* $T_i = 1\ \forall i$, then $\zeta_i = c$ is indeed a fixed point, and, according to eq.(II.14)

$$\pi_f = (1-c)/(1-c^{NM}) = \pi_f^{\text{n.s.}} \tag{III.3}$$

This generalizes the Maruyama finding to balanced islands. This result was already obtained with a different method by Lieberman et al. for the IP process, who call graphs obeying this condition " isothermal ".

Other particular solutions for simple symmetries such as star connectivity (see fig. III.1 and Mathematical details VI B) can be readily obtained.

b. Approximate solutions. Our aim, however, is to evaluate the fixation probability for any given connectivity, *i.e.* to determine the function $\pi_f(m_{ij})$. We have already showed that[19], for the classical Moran Model, and for not too small selection pressure $Ns \gtrsim 1$, second order terms not containing the factor N in the dPGF (II.3) can be negelected. Following the same approximation here, we find that there exist quasi-fixed points $\zeta = (\zeta_1,...,\zeta_M)$ which obey the following system of second order *algebraic* equations :

$$\sum_i m_{ki}(\zeta_i - 1) = c\left(1 - \frac{1}{\zeta_k}\right)\ ,\ k = 1,...,M \tag{III.4}$$

The above equations can be directly solved by standard numerical means. They are open however to a geometrical interpretation. Summing the above expression over the index k leads to:

$$\sum_i (\zeta_i + c/\zeta_i) = M(1+c) \tag{III.5}$$

This shows that the quasi-fixed points corresponding to the various connectivities lie on a particular hyper surface. Note that the exact fixed points found previously naturally belong to this hyper surface.

An important consequence of the above equation is that we can now sort migration patterns according to their effect on the fixation probability. This also allows us to demonstrate that π_f has an upper bound corresponding to $\pi_f^{\text{n.s.}}$. To illustrate these concepts and the geometrical meaning of these equations, we first consider the case of two islands, which readily generalizes to arbitrary M. There are only two independent migration parameters in this case: $m_{21} = 1 - m_{11}$ and $m_{12} = 1 - m_{22}$. Rearranging the terms, the above set of equations can be written as

$$\begin{aligned}(\zeta_1 - 1)m_{11} - (\zeta_2 - 1)m_{22} &= -\zeta_2 + (1+c) - c/\zeta_1 \\ (\zeta_1 - 1)m_{11} - (\zeta_2 - 1)m_{22} &= \zeta_1 - (1+c) + c/\zeta_2\end{aligned} \tag{III.6}$$

These are the equations of two conics crossing at $\zeta_1 = \zeta_2 = 1$ and therefore have one more solution. We can also regard the above equation as a linear system in migration

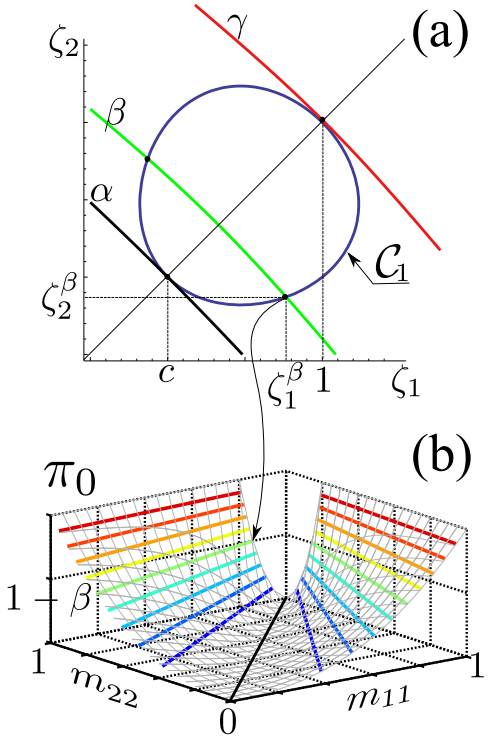

Figure III.2. Fixation probability as a function of connectivities for 2 islands. (a) A fixed point $\zeta = (\zeta_1, \zeta_2)$ belongs to the curve $\mathcal{C}_1$ (eq. III.7). Each point of this curve uniquely determines a fixation probability through eq.(II.14). Alternatively, choosing a value $\pi_f = \beta$ determines the fixed point : this is the point $(\zeta_1^\beta, \zeta_2^\beta)$ where the curves $\mathcal{C}_1$ and $\mathcal{C}_2^\beta$: $(1 - (\zeta_1 + \zeta_2)/2) \,/\, \left(1 - (\zeta_1\zeta_2)^N\right) = \beta$ cross. The crossing point between the implicit curve $\pi_f = \beta$ and the diagonal is a monotonic function of β. Therefore, possible values of π_f are bounded between two extremum values $\alpha = \pi_f^{\text{n.s.}}$ associated with the fixed point (c, c) and $\gamma = 1/N$ associated with the fixed point $(1, 1)$. (b) Surface and contour plot of $\pi_f(m_{11}, m_{22})$. Once the fixed point $(\zeta_1^\beta, \zeta_2^\beta)$ is chosen, in the connectivity plane (m_{11}, m_{22}), the straight line Δ^β : $(\zeta_1^\beta - 1)m_{11} - (\zeta_2^\beta - 1)m_{22} = -\zeta_2^\beta + (1+c) - c/\zeta_1^\beta$ is the level curve of $\pi_f(m_{11}, m_{22})$ for $\pi_f = \beta$. Here, for better visibility, $\pi_0 = 1 - \pi_f$ is represented.

probabilities m_{ij}, which has a solution only if $\zeta = (\zeta_1, \zeta_2)$ belongs to the curve

$$\mathcal{C}_1 \; : \; \zeta_1 + \zeta_2 + c\,(1/\zeta_1 + 1/\zeta_2) = 2(1+c) \qquad \text{(III.7)}$$

Choosing a point along this curve uniquely determines the fixation probability π_f (eq. II.14). On the other hand, once the point ζ is chosen, equation (III.6) determines the level curves of the function $\pi_f(m_{11}, m_{22})$, which are straight lines. This construction is detailed in figure III.2.

The important point we wish to stress is that π_f is bounded between two extremum values, the *maximum* corresponding to $\pi_f^{\text{n.s.}}$ and the minimum to the neutral case $1/MN$. This can be observed directly on the geometrical construction of figure III.2.a. Alternatively, this maximum principle can be demonstrated by searching for the constrained extremum of the function $\pi_f(\zeta_1, \zeta_2)$ (eq.II.14) under the constraint of eq.(III.7), using Lagrange multipliers (see Mathematical details VI C).

Figure III.3 shows the comparison between the above theoretical computations and direct numerical resolution of the Moran stochastic process for two islands by a Gillespie algorithm[21] (see Mathematical Details VI E). As can be observed, there is an excellent agreement between the theoretical results and "true" fixation probabilities obtained by numerical simulations.

Except for the ease of visualization when $M = 2$, nothing is specific to this case, and all of the above arguments can be extended to an arbitrary number of islands. Every quasi-fixed point $\zeta = (\zeta_1, ..., \zeta_M)$ belonging to the hyper surface (III.5) uniquely determines a fixation probability. In the $M(M-1)$ space of connectivity coefficients, the level curves of $\pi_f(m_{ij})$ are hyper-planes given by eq.(III.4).

The reverse procedure to explicitly compute the function $\pi_f(m_{ij})$ is simple : once the m_{ij} are specified, the fixed point ζ is found by solving the system of M second order *algebraic* equations. Finding ζ then allows for the computation of π_f through the algebraic expression (II.14). Figure III.4 shows the comparison between theoretical and numerical fixation probability, for five islands of population size $N = 50$, for various selection pressures ; for each selection pressure, 1200 random graphs have been generated. The relative deviation between theory and simulation is $\sim 10^{-3}$, which is due to the precision of numerical simulation and the number of stochastic trajectories used to evaluate the fixation probability.

The deviation of π_f from its maximum $\pi_f^{\text{n.s.}}$ is related to the deviation from "isothermy". The "non isothermy" of a graph can be measured by $\Delta = \text{Var}(T_i - 1)$. Figure III.4.c shows that even though Δ cannot be used as a single parameter to evaluate the fixation probability (or equivalently, the deviation of π_f from the maximum value), these two quantities are highly correlated.

The fixed point approximation developed above was derived for large selection pressure $Ns \gtrsim 1$. The domain of validity of our method is however much broader. In fact, it is not the population size of an island N, but the

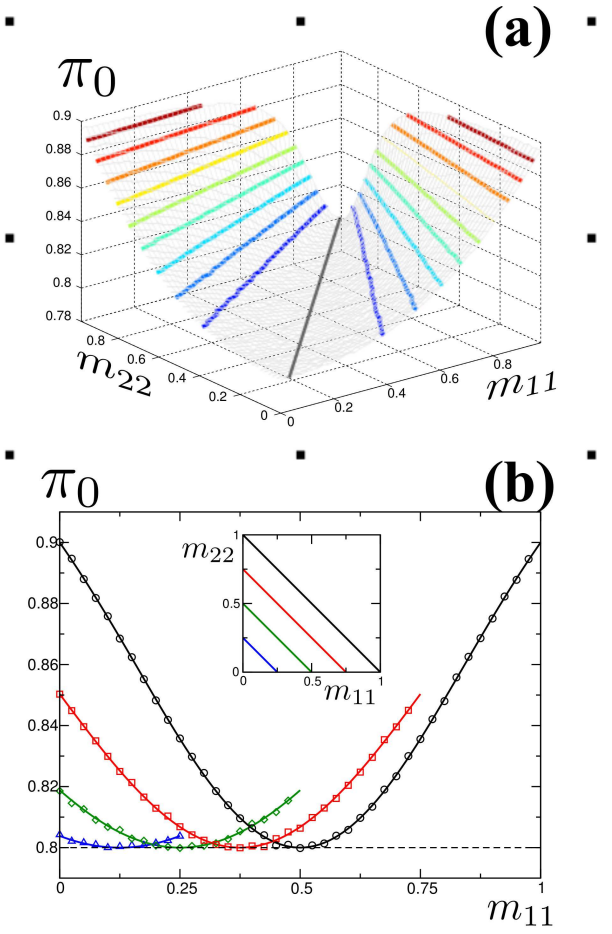

Figure III.3. Direct numerical resolution of the Moran stochastic process on graph for two islands ($M = 2$) and its comparison with theoretical curves. $c = 0.8$, $N = 50$. (a) surface and contour plot of the loss probability $\pi_0(m_{11}, m_{22}) = 1 - \pi_f(m_{11}, m_{22})$ of the numerical result obtained by a Gillespie algorithm. For each point (m_{11}, m_{22}), $R = 10^6$ trajectories have been stochastically generated. (b) Numerical data (symbols) and their corresponding theoretical curves (solid lines) along anti-diagonal slices (inset) : $\pi_0(m_{11}, K - m_{11})$ for $K = 1$(black, circles), 0.75(red, rectangles), 0.50(green, diamonds) and 0.25 (blue, triangles).

population size of the whole community MN which sets the limit of the validity : $MNs \gtrsim 1$, as can be seen in figure III.5. It is not the fixation probability, but the *limit of validity* of our approximation which seems to follow the Maruyama conjecture that " certain quantities are independent of the geographical structure of population" [9] . The limit $MNs \gtrsim 1$ is in fact what makes our method relevant : natural populations are always large and spatially subdived. Therefore even small additive fitnesses are amenable to our analytical treatment.

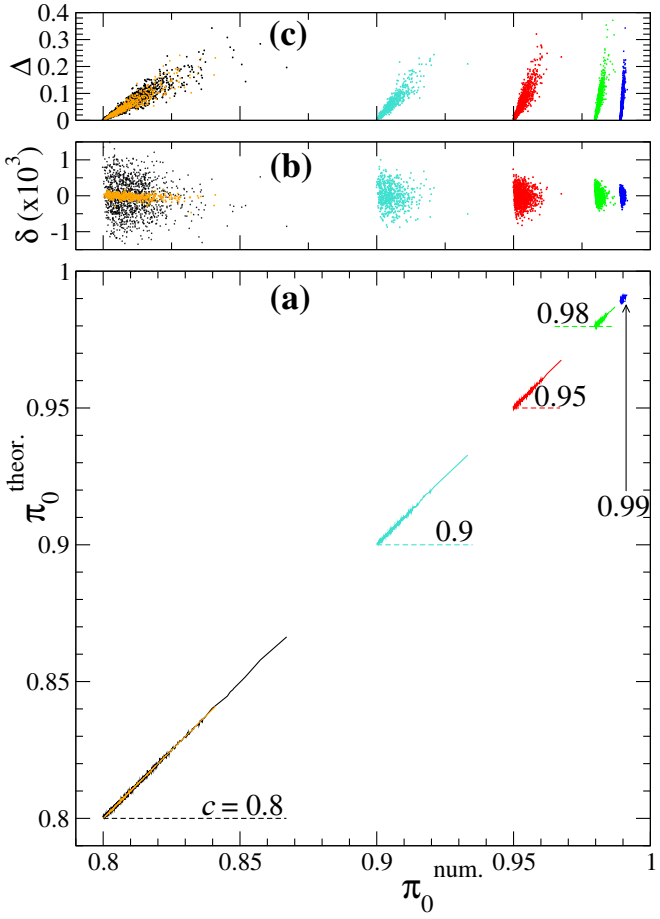

Figure III.4. Comparison between theoretical and numerical loss probability ($\pi_0 = 1 - \pi_f$) for $M = 5$, $N = 50$ and different fitness $c = 1/(1+s) = 0.99$(blue), 0.98 (green), 0.95(red), 0.9(cyan), 0.8(black and orange). For each value of c, 1200 random graph (only 480 for the high precision orange set) are generated ; For each graph, the loss probability is computed both numerically ($\pi_0^{\text{num.}}$) and theoretically ($\pi_0^{\text{theor.}}$). Numerical probabilities are computed by a Gillespie algorithm using $R = 10^6$ stochastic trajectories for each graph, except for for the $c = 0.8$, orange set, where $R = 10^8$ trajectories are used. (a) $\pi_0^{\text{theor.}}$ vs $\pi_0^{\text{num.}}$, where the closeness to the diagonal is a visual clue for the goodness of the theoretical result. Solid dotted lines indicate the corresponding $\pi_0^{\text{n.s.}}$. (b) relative deviation between numerical and theoretical results where $\delta = 2(\pi_0^{\text{num.}} - \pi_0^{\text{theor.}})/\left(\pi_0^{\text{num.}} + \pi_0^{\text{theor.}}\right)$. The relative deviation is $\sim 1/\sqrt{R} = 10^{-3}$; for $c = 0.8$ (orange set) the deviation is $\sim 1/\sqrt{R} = 10^{-4}$ (c) The deviation from isothermal $\Delta = \text{Var}(T_i - 1)$ versus the fixation probability π_f^{num} of graphs. There is a high positive correlation ($C \approx 0.9$) between Δ and $\pi_f - \pi_f^{\text{n.s.}}$

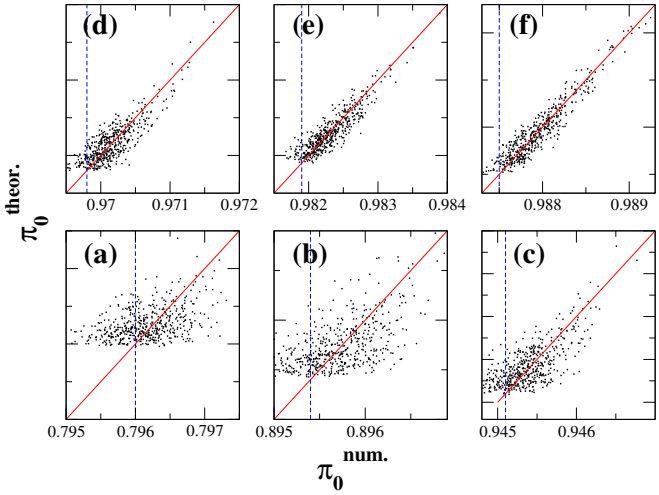

Figure III.5. Comparison between theoretical and numerical loss probability ($\pi_0^{\text{theor.}}$ vs $\pi_0^{\text{num.}}$) for $M = 5$, $c = 0.99$ ($s = 0.01$) and variable island population size : (a) $N = 1$, ($MNs = 0.05$) ; (b) 2, (0.1) ; (c) 4, (0.2) ; (d) 8, (0.4) ; (e) 16, (0.8) ; (f) 32, (1.6). For each value of c, 500 random connectivity matrix are generated. In each graph, horizontal and vertical axis scale are identical. We observe that for very small selection pressure $MNs \lesssim 0.5$, the deviation from the diagonal becomes large. To quantify this effect, we compare the standard deviation σ_1 and σ_2 of the two variables $\pi_0^{\text{num}} - \pi_0^{\text{theor.}}$ and $\pi_0^{\text{num}} - \pi_0^{\text{n.s.}}$; on each graph, $\pi_f^{\text{n.s.}}$ is represented by the dotted vertical line: ($\times 10^{-4}$) (a) $\sigma_1 = 3.9$, $\sigma_2 = 4.3$; (b) (3.1, 3.8) ; (c) (2.3, 3.3) ; (d) (1.7, 3.1) ; (e) (1.4, 3.1) ; (f) (1.1, 3.3). The two values become significantly different for $MNs \gtrsim 0.5$. As a comparison, in Figure III.4, the amplitude σ_2/σ_1 varies from 3.7 (for $c = 0.99$) to 20 (for $c = 0.8$).

IV. FIXED POINTS OF THE INVASION PROCESS.

For a non structured population, there is no difference between VM (die first then be replaced) and IP (duplicate first, then replace) version of the Moran model. For a subdivided population however, this difference has profound consequences when the islands are not balanced. We can understand this intuitively by considering an island that sends more migrants than it receives. In the IP version, when one A (beneficial) allele duplicates on *this* island, it has a high probability of spreading to other islands. In the VM version, as death occurs first, this advantage disappears. It is however less obvious why the IP advantage for one island could persist upon averaging over all islands.

Mathematically, the treatment of the IP is similar to VM, except that migration coefficients should obey row normalization $\sum_i m_{ki} = 1$ instead of column normalization (eq.II.8-II.9). The dPGF equation is therefore slightly different from expression (II.10) :

$$\frac{\partial \phi}{\partial t} = \sum_i (z_i - 1) \sum_k m_{ki} \{ N z_k \partial_k - N c \partial_i - (z_i - c) \partial_i z_k \partial_k \} \phi \qquad \text{(IV.1)}$$

but as before, we can find the fixation probability of a graph by the fixed point of the dPGF through the same eqs. (II.12,II.14).

For example, the isothermal theorem holds for IP and is demonstrated along the same lines as for the VM : for

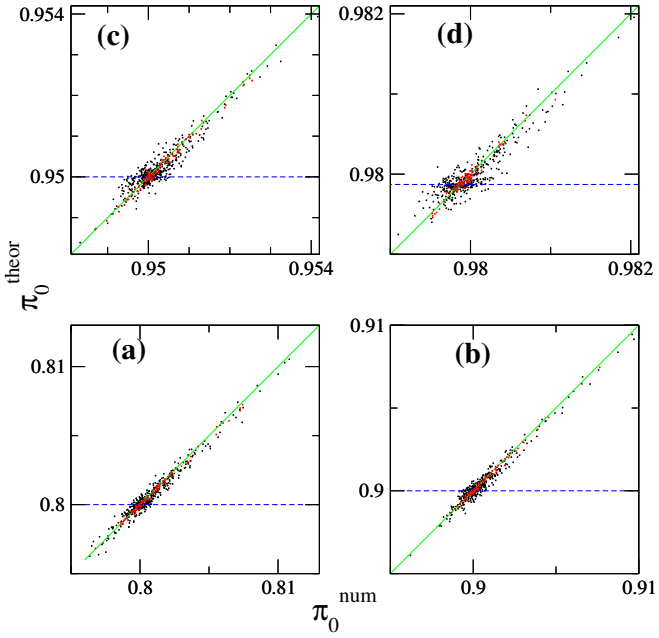

Figure IV.1. Comparison between theoretical and numerical loss probability ($\pi_0 = 1 - \pi_f$) of the Invasion Process for $M = 5$, $N = 50$ and different fitness $c = 1/(1+s) = 0.8$(a), 0.9 (b), 0.95(c), 0.98(d). Each graph contains 500 points : 400 points are computed numerically by generating 10^6 stochastic trajectories (black) and 100 by generating 10^7 ones(red). The solid (green) line designates the diagonal. The dotted (blue) horizontal line refers to the loss probability value of a non-structured population. In each graph, the mean relative deviation σ between numerical and theoretical values for the two data sets are ($\times 10^{-4}$) (a) 5.0, 2.0 ; (b) 3.4, 1.2 ; (c) 2.3, 1.1 ; (d) 1.5, 0.5.

the point $\zeta = (c, c, ..., c)$, the second order terms of the dPGF (II.10) automatically vanish ; the first order terms also vanish if

$$T_i = 1 \;\; i = 1, ..., M$$

where the temperature is now $T_i = \sum_k m_{ki}$. $(T_i - 1)$ has the meaning of balance between migrants received and sent by island i. All other exact results such as the "star" graph treated in preceding section have an analog for the IP (see Mathematical details VI B).

In the limit $MNs \gtrsim 1$, second order terms of the dPGF can be neglected and quasi-fixed points $\zeta = (\zeta_1, ..., \zeta_M)$ now obey the following system of second order *algebraic* equations :

$$\sum_i m_{ki}(\zeta_i - 1) = cT_k \left(1 - \frac{1}{\zeta_k}\right) \;\; , \;\; k = 1, ..., M \quad \text{(IV.2)}$$

The process of estimating the fixation probability is similar to the previous discussion of the VM version. Figure IV.1 shows, for 500 random graphs and various selection pressures, the comparison between the fixation probability found from expression (IV.2) and by direct numerical simulations. The same high degree of precision is observed as in the previous case.

It can be seen in figure IV.1 that many migration patterns show fixation probabilities higher than the non-structured value. The difference in the evaluation of the fixed point of invasion process is the presence of the term T_k on the right hand side of eqs.(IV.2) compared to eqs.(III.4). This modification excludes a constraint independent of the graph topology similar to eq.(III.5), and the fixed points no longer belong to a universal hypersurface. An important consequence is that for the Invasion Process, the fixation probability does not obey a gobal "maximum principle". We can however give a restricted version of this principle. Each migration pattern $\{m_{ki}\}$ can be considered as a point in a $M(M-1)$ space. We can partition this space into subspaces $\mathcal{E}_{\mathbf{T}}$ of dimension $(M-1)$, each subspace labeled by the vector $\mathbf{T} = (T_1, ..., T_M)$: the subspace $\mathcal{E}_{\mathbf{T}}$ contains all migration patterns which have temperature distribution $\mathbf{T}$. Using the same arguments as in the previous section, we can show that in each subspace $\mathcal{E}_{\mathbf{T}}$, the fixed points belong to the $(M-1)$ hypersurface

$$\sum_{k=1}^{M} T_k(\zeta_k + c/\zeta_k) = M(1+c) \quad \text{(IV.3)}$$

On each hypersurface, the fixation probability has a maximum value $\pi_f^{\max}(\mathbf{T}) > \pi_f^{\text{n.s.}}$. For a given temperature distribution $\mathbf{T}$, the set of all graphs with fixation probability $\in [\pi_f^{\max}(\mathbf{T}), \pi_f^{\text{n.s.}}]$ are fitness amplifiers. For some particular classes of graph, π_f can greatly exceed the non-structured value and approach unity, as noted by other authors[12, 15] (see Mathematical details VI B).

V. CONCLUSION.

Evolutionary dynamics of a non structured population is an interplay between deterministic selection pressure and stochastic sampling between generations. In a geographically subdivided populations, migration between colonies is the third ingredient to be taken into account. A migration pattern is a more complexe quantity than a number such as the fitness, and it has to be specified by an $M \times M$ matrix, where the coefficient m_{ki} weights migration importance from patch k to patch i. In the preceding sections, in the framework of the Moran model, we have developed a mathematical method that allows for the computation of the fixation probability of a beneficial allele in a subdivided population for arbitrary migration patterns.

This mathematical method allows us to partition the migration patterns according to their effect on the fixation probability and gives rise to important fundamental results. For example, the celebrated result of Maruyama[8, 9] was that the fixation probability does not depend on population subdivision and is equal to that of a non-strucured population $\pi_f^{\text{n.s.}}$. In order to

compute the fixation probability, Maruyama had used a severe approximation. Our investigation shows (in the VM version of the Moran model) that the Maruyama solution corresponds to the *upper* bound of the fixation probabilities and the effective fitness of a beneficial allele is degraded by the population subdivision.

The IP version of the Moran model is more often used for epidemic or cancer propagation. In this framework, Lieberman *et al.*[12] were able to build particular migration patterns capable of amplifying the fixation probability of a mutant in a subdivided population compared to a non-structured one. The method we develop in this article gives the general conditions for which a pattern can amplify the fitness of a mutant. We show that the important parameter in this case is the (vector) of migration imbalance $\mathbf{T}$, which sets the upper bound for the fixation probability $\pi_f^{\max}(\mathbf{T})$. One can then find all topologies correponding to this imbalance which have a fixation probability in the range of $[\pi_f^{\text{n.s.}}, \pi_f^{\max}(\mathbf{T})]$ and hence are fitness amplifiers.

The mathematcial tool we have developed is based on finding the fixed points of a partial differential equation ; for large selection pressure $MNs \gtrsim 1$, the fixed points are solutions of a system of second degree algebraic equation. The problem of finding the fixation probability is therefore mapped into a simple problem of geometry and of conics crossing in a M dimensional space. Until now, no general analytical tool was available to estimate the fixation probability of a subdivided population. Some solutions have been found for simple symmetries, but it was necessary to resort to numerical simulations in order to investigate a particular migration pattern[11]. Although faster algorithms have been developed[22], the numerical simulations remain extremely costly for large populations. The method we have developed in this article requires negligible computational time and allows precise investigation of wide classes of topologies and problems. For example, one could envisage problems in which the migration patterns themselves, *i.e.* the coefficients m_{ki}, are subject to selection in order to prevent epidemic propagation or help the emergence of cooperative behavior.

All species are distributed through space and are therefore subdivided into colonies. Understanding the evolutionary process for these populations is therefore of fundamental importance to our comprehension of evolution in general. We believe that the dPGF method we propose here is a valuable tool in advancing our comprehension of this process.

c. Acknowledgments. We are grateful to O.Rivoire, A. Halperin and E. Geissler for the careful reading of the manuscript and fruitful discussions. This work was partly funded by Agence Nationale de la Recherche Française (ANR) grant "Evo-Div."

VI. MATHEMATICAL DETAILS.

A. Deriving the dPGF equation on graph.

We will follow the same derivation as in [19], generalized to M islands. To shorten the notation, let us represent the state of the system by the vector $\mathbf{n} = (n_1, ..., n_M)$, its conjugated variable by $\mathbf{z} = (z_1, ..., z_M)$ and $\mathbf{z}^{\mathbf{n}} = z_1^{n_1} ... z_M^{n_M}$. More over, let the vector $a_k^{\pm}\mathbf{n}$ represent the state $\mathbf{n}$ where we have added or removed one individual to or from island k. Then the master equation governing the probabilities $P(\mathbf{n}; t)$ then reads

$$\begin{aligned}\frac{\partial P(\mathbf{n},t)}{\partial t} = \sum_{i=1}^{M} & W_i^+(a_i^-\mathbf{n})P(a_i^-\mathbf{n},t) - W_i^+(\mathbf{n})P(\mathbf{n},t) \\ & + W_i^-(a_i^+\mathbf{n})P(a_i^+\mathbf{n},t) - W_i^-(\mathbf{n})P(\mathbf{n},t)\end{aligned} \tag{VI.1}$$

where the transition probabilities $W_i^{\pm}$ are given by expressions (II.6,II.7). Multiplying both sides of equation (VI.1) by $\mathbf{z}^{\mathbf{n}}$ and summing over $\mathbf{n}$, we get :

$$\frac{\partial \phi}{\partial t} = \sum_i (z_i - 1)\left\langle \mathbf{z}^{\mathbf{n}} W_i^+(\mathbf{n})\right\rangle + \left(\frac{1}{z_i} - 1\right)\left\langle \mathbf{z}^{\mathbf{n}} W_i^-(\mathbf{n})\right\rangle \tag{VI.2}$$

where the symbol $\langle ... \rangle$ stands for summation over states $\mathbf{n}$. The computation rules are simple enough :

$$\begin{aligned}\left\langle n_k \mathbf{z}^{\mathbf{n}}\right\rangle &= z_k \frac{\partial \phi}{\partial z_k} \\ \left\langle n_k^2 \mathbf{z}^{\mathbf{n}}\right\rangle &= z_k \frac{\partial}{\partial z_k}\left(z_k \frac{\partial \phi}{\partial z_k}\right) \\ \left\langle n_k n_j \mathbf{z}^{\mathbf{n}}\right\rangle &= z_k z_j \frac{\partial^2 \phi}{\partial z_k \partial z_j}\end{aligned}$$

which leads to equations (II.10,IV.1) depending on the choice of row or column normalization of m_{ki}. Note that the stationary solution (II.11) of the dPGF is valid for connected graphs.

B. Some particular solutions of the Moran process on graph.

The article contains the general solution of fixation probability for the Moran process. Some graphs have simple symmetries and their π_f can be given in closed form. Here we explore a few such cases. These graphs, which have been already solved by other authors (at least for IP) are selected to illustrate the simplicity of the fixed point method.

d. Star connectivity. Consider fig. III.1b, where one central island, labeled 0, is connected to P peripheral islands and where self migration is prohibited $m_{ii} = 0$. Peripheral islands do not communicate among each other $m_{ij} = 0$ if $i, j > 0$. For the VM, this configuration imposes $m_{i0} = 1/P$ and $m_{0i} = 1$. The symmetric disposition of the islands imposes a fixed point of the form

$\zeta = (\zeta_0, \zeta_1, \zeta_1, ...\zeta_1)$. Plugging this solution into the VM dPGF (II.10), we get

$$\zeta_0 = c\,(Pc+1)\,/\,(P+c)$$
$$\zeta_1 = c^2/\zeta_0$$

For the IP, migrations are $m_{i0} = 1$ and $m_{0i} = 1/P$ $(i > 0)$. Using the dPGF (IV.1), we obtain

$$\zeta_0 = c\,(P+c)\,/\,(Pc+1)$$
$$\zeta_1 = c^2/\zeta_0$$

We emphasize that these are exact solutions of eq. (II.10,IV.1) respectively. In both cases, the fixation probability is

$$\pi_f^{\text{star}} = \frac{1-(\zeta_0+P\zeta_1)/(P+1)}{1-(\zeta_0\zeta_1^P)^N}$$

When $P \gg 1$ for IP, this reduces to

$$\pi_f^{IP} \approx \frac{1-c^2}{1-c^{2PN}} \tag{VI.3}$$

e. Super star. This configuration, (figIII.1c) is an extension of the star configuration, where a central island A is connected to P secondary islands B ; each secondary island is connected to Q tertiary islands C. For the VM, migration coefficients are $m_{AC} = 1$, $m_{CB} = 1/Q$ and $m_{BA} = 1/P$, and all other coefficients are zero. There is no longer an exact fixed point for this case ; the quasi-fixed points are

$$\zeta_A = c\frac{c^2PQ+cQ+1}{c^2Q+c+PQ}$$
$$\zeta_B = c\frac{c^2Q+c+PQ}{c^2+cPQ+Q}$$
$$\zeta_C = c^3/\zeta_A\zeta_B$$

For the IP, $m_{AC} = 1/PQ$, $m_{CB} = 1$, $m_{BA} = 1$, the ζ_i are the same as the above expressions, up to a circular permutation of indices A, B, C. In both cases, the fixation probability is

$$\pi_f^{\text{star}} = \frac{1-(\zeta_A+P\zeta_B+PQ\zeta_C)/(1+P+PQ)}{1-(\zeta_A\zeta_B^P\zeta_C^{PQ})^N} \tag{VI.4}$$

For IP, in the limit $P \gg 1$, $Q \gg 1$, this reduces to

$$\pi_f^{IP} \approx \frac{1-c^3}{1-c^{3PQN}}$$

Note that the expressions (VI.3,VI.4) for star and super-star are identical to those found by Lieberman[12].

C. The maximum Principle.

For VM the fixation probability

$$\pi_f = \frac{1-(1/M)\sum_i \zeta_i}{1-(\zeta_1...\zeta_M)^N}$$

subjected to the constraint

$$\sum_{i=1}^{M}(\zeta_i + c/\zeta_i) = M(1+c) \tag{VI.5}$$

possesses a maximum for the non-structured point $\zeta = (c, ...c)$. This result has a simple geometrical interpretation shown in figure III.2. It can be demonstrated as follows. Let us set

$$a = (1/M)\sum_i \zeta_i$$

$$g = \prod_i \zeta_i^{1/M}$$

the fixation probability can be written

$$\pi_f = \frac{1-a}{1-g^{NM}}$$

The geometrical mean g of a set of positive numbers is never larger than their arithmetical mean a, therefore

$$\pi_f \leq \frac{1-a}{1-a^{NM}} \tag{VI.6}$$

The range of variation of a constrained by the condition (VI.5) can be obtained using the Lagrange multiplier method:

$$\partial\left[a - \lambda\sum_{i=1}^{M}(\zeta_i + c/\zeta_i)\right]/\partial\zeta_i = 0$$

$$1/M - \lambda\left(1 - c/\zeta_i^2\right) = 0 \;\; i = 1, ..., M$$

which, for $\zeta_i > 0$, can be satisfied only if

$$\zeta_1 = \zeta_2 = ... = \zeta_M$$

The hypersurface (VI.5) has only two points along the diagonal,

$$\zeta_i = 1 \;\; i = 1, ..., M$$
$$\zeta_i = c$$

These two solutions correspond to the extremum values of a, so that the range of variation of this arithmetical mean is $c \leq a \leq 1$.

The right hand side of the inequality (VI.6) is the reciprocal of a geometrical sum: it is therefore a monotonically decreasing function of a and it takes its maximum value for $a = c$:

$$\pi_f \leq \frac{1-c}{1-c^{NM}}$$

This "maximum principle" holds as long as our approximate expression for the fixation probability is valid, that is to say for $NMs > 1$.

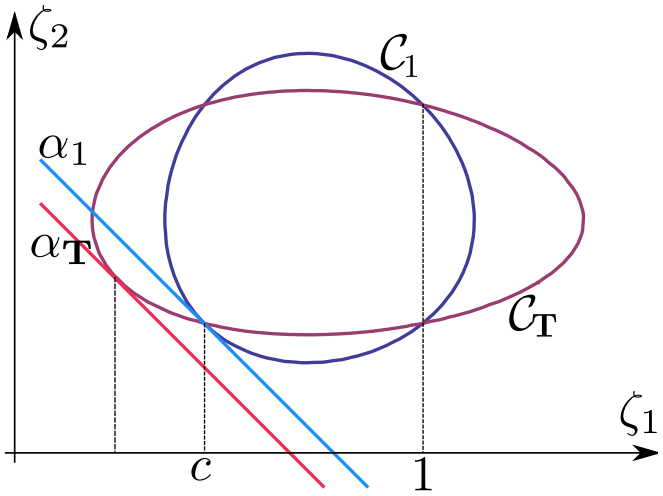

Figure VI.1. The restricted maximum principle for IP, illustrated for two islands. Two closed curves corresponding to temperature distributions $\mathbf{1} = \{1, 1\}$ and $\mathbf{T} = \{T, 2 - T\}$ are shown. The quasi fixed points are found as in figure III.2 by finding the point at which $\mathcal{C}_{\mathbf{T}}$ crosses the curve $\pi_f =$ constant. There is a maximum value α of the fixation probability. For isothermal distributions, $\alpha_1 = \pi_f^{\text{n.s.}}$. For non isothermal distributions, $\alpha_{\mathbf{T}} > \alpha_1$. Points of $\mathcal{C}_{\mathbf{T}}$ between the two curves $\pi_f = \alpha_1$ and $\pi_f = \alpha_{\mathbf{T}}$ correspond to fitness amplifiers graphs.

For the IP case, the maximum principle has a restricted version. Consider all graphs corresponding to a given (IP) temperature distribution $\mathbf{T} = \{T_1, ...T_M\}$. Then the quasi fixed point belongs to the hypersurface

$$\mathcal{C}_{\mathbf{T}} : \sum_{k=1}^{M} T_k(\zeta_k + c/\zeta_k) = M(1 + c) \tag{VI.7}$$

Given $\mathbf{T}$, all the above arguments for the maximum can be repeated, except that the extrema of π_f are now given by the crossing of $\mathcal{C}_{\mathbf{T}}$ and the curve

$$T_k \left(1 - c/\zeta_k^2\right) = \text{Constant} \;\; k = 1, ..., M$$

As a consequence, the maximum of the fixation probability $\pi_f^{\max}(\mathbf{T}) > \pi_f^{\text{n.s.}}$. Figure VI.1(analogous to figure III.2) illustrates this construction for the case of two islands. Note that $\sup_{\mathbf{T}} \left(\pi_f^{\max}(\mathbf{T})\right) = (1 - c/M)$, which corresponds to the limiting case $T_1 = M - \epsilon$, $T_{i>1} = \epsilon K_i$, $\epsilon \to 0$, $K_i = \mathcal{O}(1)$.

D. Generalization to variable island size.

The generalization to islands of variable sizes is straightforward. We describe this generalization for VM, and a similar generalization can be provided for the invasion process. The dPGF reads:

$$\frac{\partial \phi}{\partial t} = \sum_i (z_i - 1) \left\{ N_i \sum_k \frac{m_{ki}}{N_k} z_k \partial_k - N_i c \partial_i \right.$$
$$\left. - (z_i - c)\partial_i \sum_k \frac{m_{ki}}{N_k} z_k \partial_k \right\} \phi$$

The temperature of an island is defined as:

$$T_k = \sum_i \frac{N_i m_{ki}}{N_k}$$

The isothermal condition always corresponds to $T_k = 1$ for all the islands and it corresponds to the fixed point $\zeta_k = c \; \forall k$.

In the case of appearance of a mutant at a random site, the fixation probability is related to the coordinates ζ_i of the fixed point by:

$$\pi_f = \frac{1 - \sum_i (N_i / N_{\text{Tot}})\zeta_i}{1 - \prod_i \zeta_i^{N_i}} \tag{VI.8}$$

where $N_{\text{Tot}} = \sum_i N_i$ is the total number of individuals. The set of equations allowing the determination of the (quasi) fixed point reads:

$$\sum_i N_i m_{ki}(\zeta_i - 1) = N_k c(1 - \frac{1}{\zeta_k}) \;, \;\; k = 1, ..., M$$

and the fixed points lie on a hypersurface defined by

$$\sum_i N_i(\zeta_i + c/\zeta_i) = N_{\text{Tot}}(1 + c) \tag{VI.9}$$

One can readily show that the "maximum principle" for VM also holds in this more general case. we set

$$\hat{a} = \sum_i N_i \zeta_i / N_{Tot}$$

$$\hat{g} = \prod_i \zeta_i^{N_i / N_{Tot}}$$

Using the Lagrange multipliers method it is easy to show that $\hat{g}$ is never larger than $\hat{a}$ and that the range of variation of $\hat{a}$ under the constraint (VI.9) is $c \leq \hat{a} \leq 1$. Following the same reasoning as in section 6.4, one then deduces

$$\pi_f = \frac{1 - \hat{a}}{1 - \hat{g}^{N_{Tot}}} \leq \frac{1 - \hat{a}}{1 - \hat{a}^{N_{Tot}}} \leq \frac{1 - c}{1 - c^{N_{Tot}}}$$

The fixation probability given by (VI.8) is maximum for a non-structured population.

E. Numerical resolution.

f. Numerical simulation. The stochastic equations given by the rates (II.6,II.7) can be seen as $2M$ chemical reactions (two for each islands) for the species A:

$$A_i \xrightarrow{k^+} 2A_i \ ; \ A_i \xrightarrow{k^-} \emptyset$$

which we solve by the classical Gillespie algorithm ([21]) written in C++. We are interested here only in the fixation probability and not in the fixation time ; the program can therefore be accelerated by computing only the nature of the event that occurs at each turn (and not its time of occurrence). In general, to solve for the fixation probability of a given graph, $R = 10^6$ stochastic trajectories are generated. For $M = 5$ and $N = 50$, computing the fixation probability of 1000 islands takes about ~6 hours on a standard PC. The high precision orange set of Figure III.4 for 500 graph takes about 1 week on two computers.

g. Analytical computations. Equations (III.4) constitute a system of M second order algebraic equations. They can be efficiently solved by standard methods such as the "conjugated gradients". Mathematical software such as "Scilab" and "Matlab" have implemented these methods with the "fsolve" subroutine. In Mathematica (Wolfram Research), this routine is implemented by first setting the equation (III.4) (for VM)

```
f[m_,z_,c_]:= (m . (z-1) - c)*z + c
```

where m is the $M \times M$ connectivity matrix, and then by solving the equation

```
FindRoot[f[m, z, c] == 0, {z, z0}]
```

where z_0 is the $M-$dimensional vector $\{c, ...c\}$. For 5 islands, solving 1000 different graphs takes about 1 second .

[1] Z. Patwa and L. M. Wahl. The fixation probability of beneficial mutations. *J R Soc Interface*, 5(28):1279–1289, 2008.

[2] P.A.P. Moran. *The Statistical processes of of evolutionary theory.* Oxford University Press, 1962.

[3] W. J. Ewens. *Mathematical Population Genetics.* Springer-Verlag, 2004.

[4] S. Wright. Isolation by distance. *Genetics*, 28(2):114–138, 1943.

[5] G. Malécot. *Les mathématiques de l'hérédité.* Masson, Paris, 1948.

[6] M. Kimura and G. H. Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49:561–576, 1964.

[7] J. Felsenstein. A pain in the torus : some difficulties with models of isolation by distance. *American Naturalist*, 109:359–368, 1975.

[8] T. Maruyama. A markov process of gene frequency change in a geographically structured population. *Genetics*, 76(2):367–377, 1974.

[9] T. Maruyama. A simple proof that certain quantities are independent of the geographical structure of population. *Theor Popul Biol*, 5(2):148–154, 1974.

[10] Michael C Whitlock. Fixation probability and time in subdivided populations. *Genetics*, 164(2):767–779, 2003.

[11] P.A. Whigham and G. Dick. Evolutionary dynamics for the spatial moran process. *Genet. Program. Evolvable Mach.*, 9:157–170, 2008.

[12] Erez Lieberman, Christoph Hauert, and Martin A Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.

[13] T. Antal, S. Redner, and V. Sood. Evolutionary dynamics on degree-heterogeneous graphs. *Phys Rev Lett*, 96(18):188104, 2006.

[14] V. Sood, Tibor Antal, and S. Redner. Voter models on heterogeneous networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 77:041121, 2008.

[15] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys Rev Lett*, 92(17):178701, 2004.

[16] Natalia L Komarova, Anirvan Sengupta, and Martin A Nowak. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J Theor Biol*, 223(4):433–450, 2003.

[17] Natalia L Komarova. Viral reproductive strategies: How can lytic viruses be evolutionarily competitive? *J Theor Biol*, 249(4):766–784, 2007.

[18] Duncan J Watts. A simple model of global cascades on random networks. *Proc Natl Acad Sci U S A*, 99(9):5766–5771, 2002.

[19] B. Houchmandzadeh and M. Vallade. Alternative to the diffusion equation in population genetics. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(5 Pt 1):051913, 2010.

[20] S. P. Hubbel. *The unified neutral theory of Biodiversity and Biogeography.* Princeton University Press, 2001.

[21] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[22] Valmir C Barbosa, Raul Donangelo, and Sergio R Souza. Early appraisal of the fixation probability in directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(4 Pt 2):046114, 2010.