Multilocus Migration-Selection Models

REINHARD BÜRGER

Department of Mathematics, University of Vienna

KITP UCSB, 1 Oct 2008

How important is population subdivision

in maintaining genetic diversity?

General Model

- We consider a diploid population with discrete, nonoverlapping generations and indistinguishable sexes.
- The population is subdivided into finitely many colonies (demes) that exchange adult migrants.
- Mating occurs within each deme and is random.
- The number of loci and alleles per locus is arbitrary.
- There is constant, frequency-independent viability selection in each deme.
- Mutation and random genetic drift are ignored (among all other 'complications').

Mathematical questions addressed:

- What are the dynamical properties of multilocus systems subject to migration and selection?
- More precisely, when is the dynamics simple in the sense that convergence to equilibrium occurs?
- When does convergence to (quasi-) linkage equilibrium occur?
- Simple, positive answers may be expected for 'simple' genotype-fitness maps and for 'special' migration patterns.

Simple genotype-fitness maps:

- No or weak epistasis and no or intermediate dominance.

Special migration patterns:

- Migration is strong relative to selection (we shall also need that recombination is strong).
- Migration is weak relative to selection and recombination.
- Migration is independent of the deme of origin (Levene model).

The Central Biological Question

Under which conditions is multilocus polymorphism maintained in a subdivided population when it is not in a panmictic one?

More specifically:

- If there is no or weak epistasis and intermediate dominance at every locus and in every deme, how many loci can be maintained polymorphic at a (stable) equilibrium?
- If there is no or weak epistasis and intermediate dominance at every locus and in every deme, how many alleles per locus can be maintained at a (stable) equilibrium for a given number of loci? Is there a generic upper bound?
- In the absence of epistasis and of dominance, i.e., for purely additive selection, is the number of demes a generic upper bound for the number of alleles at a (stable) equilibrium?

Notation

There are

$L \geq 1$ loci	denoted by letters k, n
$I_n \geq 2$ alleles at locus n	denoted by letters i_n
$I = \prod_n I_n$ gametes	denoted by letters i, j
$\Gamma \geq 1$ demes	denoted by letters α, β
Δ_I	is the simplex of gamete frequencies in a single deme
Δ_I^Γ	is the state space for the full dynamics.

Alleles are denoted by $A_{i_k}^{(k)}$, i.e., $A^{(\text{number of locus})}_{\text{number of allele}}$

Further,

$p_{i,\alpha}$	is the frequency of gamete i in deme α,
$p^{(k)}_{i_k,\alpha}$	is the frequency of allele i_k at locus k in deme α,
$D_{i,\alpha}$	is the disequilibrium in gamete i in deme α,
$w_{ij,\alpha}$	is the fitness of genotype ij in deme α,
$m_{\alpha\beta}$	is the probability that an adult in deme α migrated from deme β.

We assume that the *backward migration matrix*, $M = (m_{\alpha\beta})$, is ergodic (clearly, it is stochastic). Its principal left eigenvector is denoted by μ, i.e.,

$$\mu^T M = \mu^T .$$

Averaging

We shall need to average $p_{i,\alpha}$ with respect to μ:

$$P_i = \mu^T p_i \,, \quad P = (P_1, \ldots, P_I)^T \in \Delta_I \,,$$

where $p_i = (p_{i,1}, \ldots, p_{i,\Gamma})^T \in \mathbb{R}^\Gamma$.

The deviation from the average gamete frequency P is denoted by

$$q = p - P \,.$$

Thus, $q_{i,\alpha} = p_{i,\alpha} - P_i$.

The average frequency of $A_{i_k}^{(k)}$ is denoted by $P_{i_k}^{(k)}$.

Linkage disequilibria

Let $\{\mathsf{K}, \mathsf{N}\}$ denote a nontrivial decomposition of the set of loci L, and let c_{K} designate the probability of reassociation of the genes at the loci in K, inherited from one parent, with the genes at the loci in N, inherited from the other.

Then, the linkage disequilibrium in gamete i in deme α is defined as (Nagylaki 1993)

$$D_{i,\alpha} = \frac{1}{\bar{w}_\alpha} \sum_j \sum_{\mathsf{K}} c_{\mathsf{K}} \left(w_{ij,\alpha}\, p_{i,\alpha} p_{j,\alpha} - w_{i_{\mathsf{K}} j_{\mathsf{N}}\, j_{\mathsf{K}} i_{\mathsf{N}},\alpha}\, p_{i_{\mathsf{K}} j_{\mathsf{N}},\alpha}\, p_{j_{\mathsf{K}} i_{\mathsf{N}},\alpha} \right) ,$$

where $D_i = (D_{i,1}, \ldots, D_{i,\Gamma})^T \in \mathbb{R}^\Gamma$. We shall write $D = (D_1, \ldots, D_I)^T$.

Dynamics

The life cycle is selection, recombination, migration.

Then, the *selection-recombination-migration dynamics* is given by

$$p'_{i,\alpha} = \sum_{\beta} m_{\alpha\beta} p^{\#}_{i,\beta}\,, \tag{SRM1}$$

where

$$p^{\#}_{i,\alpha} = p_{i,\alpha} \frac{w_{i,\alpha}}{\bar{w}_{\alpha}} - D_{i,\alpha} \tag{SRM2}$$

describes selection and recombination in deme α. We view these equations as a dynamical system on Δ_I^{Γ}.

MAIN RESULTS, PART I:

DYNAMICAL PROPERTIES

1. No selection, only migration and recombination

2. Selection is weak relative to migration and recombination

3. Migration is weak relative to selection and recombination; also epistasis is weak.

1. Migration and Recombination (No Selection)

Then, the dynamics of the gametic frequencies simplifies to

$$p_i' = M(p_i - D_i)\,,$$

where $D_{i,\alpha}$ now simply is

$$D_{i,\alpha} = \sum_{\mathsf{K}} c_{\mathsf{K}} \left(p_{i,\alpha} - p^{(\mathsf{K})}_{i_{\mathsf{K}},\alpha}\, p^{(N)}_{i_{\mathsf{N}},\alpha} \right) .$$

For the following, we recall that $q = p - P$ measures the spatial heterogeneity of the population.

Theorem 1. *For the migration-recombination dynamics, the manifold*

$$\Psi_0 = \left\{ p \in \Delta_I^{\Gamma} : p_{i,\alpha} = \prod_k P_{i_k}^{(k)} \ \ \forall i \, \forall \alpha \right\}$$
$$= \left\{ p \in \Delta_I^{\Gamma} : D = 0 \text{ and } q = 0 \right\}$$

is invariant and globally attracting at a uniform geometric rate. Every point on Ψ_0 is an equilibrium point.

Generically, the rate of convergence is

$$\lambda = \max(\kappa, 1 - c_{\min}) \,,$$

where $c_{\min} > 0$ is the smallest two-locus recombination rate and (generically) κ can be taken to be the modulus of the largest non-unit eigenvalue.

Essentially, the *proof* uses the method of Nagylaki (1993):

(i) Recursion relations for the linkage equilibria by passing to the much simpler set $d_{i,\alpha} = p_{i,\alpha} - \prod_k p^{(k)}_{i_k,\alpha}$;

(ii) An asymptotic invariance property of migration;

(iii) Decay of the linkage disequilibria for two embedded loci;

(iv) Decay of the linkage disequilibria for multiple loci;

(v) Decay of the spatial heterogeneity measure q.

2. Strong Migration, Strong Recombination, and Weak Selection

Set

$$w_{ij,\alpha} = 1 + \epsilon r_{ij,\alpha}\,,$$

where $\epsilon \geq 0$ is sufficiently small and $|r_{ij,\alpha}| \leq 1$.

Suppose that the (ergodic) backward migration matrix M and all recombination rates c_{K} are fixed.

We call the differential equation

$$\frac{dP_{i_n}^{(n)}}{d\tau} = P_{i_n}^{(n)} \left[\omega_{i_n}^{(n)}(\pi) - \bar{\omega}(\pi) \right] ,$$
$$D = 0 \, , \; q = 0$$

on Δ_I^{Γ} the *weak-selection limit* of (SRM). Here,

$$\omega_{i_n}^{(n)}(\pi) = \sum_{\alpha} \mu_{\alpha} r_{i_n,\alpha}^{(n)}(\pi) \;\; \text{and} \;\; \bar{\omega}(\pi) = \sum_{n} \sum_{i_n} \omega_{i_n}^{(n)} P_{i_n}^{(n)} \, ,$$

are the average selection coefficients of the allele $A_{i_n}^{(n)}$ and of the entire population, and $\pi = \left(P_1^{(1)}, \ldots, P_{I_1}^{(1)}, \ldots, P_1^{(L)}, \ldots, P_{I_L}^{(L)} \right)^T$.

The weak-selection limit 'lives' on the manifold Ψ_0 on which $D = q = 0$.

Its dynamics is well understood and simple. Essentially, it is a cartesian product of single-locus systems with multiple alleles. In particular, it is a generalized gradient system with potential $\bar{\omega}$:

$$\frac{d\bar{\omega}}{d\tau} = 2 \sum_n \sum_{i_n} P_{i_n}^{(n)} \left[\omega_{i_n}^{(n)}(\pi) - \bar{\omega}(\pi) \right]^2 \geq 0 \,.$$

Equality occurs exactly at the equilibria.

The fact that the weak-selection limit is a gradient is essential for the following theorem, which is based on a singular perturbation argument.

Theorem 2. *Assume that all equilibria of the weak-selection limit are hyperbolic (this holds generically!) and ϵ is sufficiently small.*

(a) The set of equilibria $\Xi_0 \subset \Delta_I^\Gamma$ of the weak-selection limit contains only isolated points, as does the set of equilibria $\Xi_\epsilon \subset \Delta_I^\Gamma$ of (SRM). As $\epsilon \to 0$, each equilibrium in Ξ_ϵ converges to the corresponding equilibrium in Ξ_0.

(b) In the neighborhood of each equilibrium in Ξ_0, there exists exactly one equilibrium point in Ξ_ϵ. The stability of each equilibrium in Ξ_ϵ is the same as that of the corresponding equilibrium in Ξ_0; i.e., each pair is either asymptotically stable or unstable.

(c) Every solution $p(t)$ of (SRM) converges to one of the equilibrium points in Ξ_ϵ.

This theorem (and its proof) shows that after a sufficiently long (but evolutionary short) time, the full SRM dynamics can be interpreted as a perturbation of the weak-selection limit.

In particular, a manifold Ψ_ϵ is approached on which quasi-linkage equilibrium prevails and solutions are spatially quasi-homogeneous.

PROOF. Follows Nagylaki, Hofbauer, and Brunovský (1999).

(i) Existence of and convergence to the manifold Ψ_ϵ, which is a perturbation of Ψ_0 and characterized by an equation of the form

$$(D, q) = \epsilon\psi(\pi, \epsilon)\,,$$

where ψ is a smooth function of π.

(ii) Derivation of the estimate

$$P_{i_n}^{(n)\prime} = P_{i_n}^{(n)} + \epsilon P_{i_n}^{(n)} \left[\omega_{i_n}^{(n)}(\pi) - \bar{\omega}(\pi)\right] + O(\epsilon^2)\,.$$

Rescaling time t in generations as $\tau = \epsilon t$ and letting $\epsilon \to 0$, we obtain the weak-selection limit.

(iii) The proof of Theorem 3.1 in NHB yields all statements in the above theorem except the first in (b).

(iv) Proof that (unstable) boundary equilibria remain on the boundary 'after the perturbation'. (This is quite obvious because at a boundary equilibrium of the weak-selection limit, a subset of alleles is absent from the population.)

For a single multiallelic locus, this result was proved by Nagylaki and Lou (TPB, 2007).

Consequences

1. The time $\tilde{t}$ required to approach Ψ_ϵ is evolutionary short and given by

$$\tilde{t} = \frac{\ln \epsilon}{\ln \lambda} + O(\epsilon)\,.$$

2. Suppose the assumptions of the above theorem. If $D(t) = O(\epsilon)$ and $q(t) = O(\epsilon)$, π is bounded away from the equilibria of the weak-selection limit, and p is within $O(\epsilon^2)$ of Ψ_ϵ, then

$$\Delta\bar{w}(p) > 0\,.$$

   Thus, the mean fitness is increasing for an evolutionary long time, of order $O(1/\epsilon)$.

3. Weak Migration, Weak Epistasis, and Strong Recombination

In the absence of migration, (SRM) reduces to

$$p'_{i,\alpha} = p_{i,\alpha}\frac{w_{i,\alpha}}{\bar{w}_\alpha} - D_{i,\alpha} \qquad \text{(NoMig)}$$

for every gamete i and deme α. Therefore, we have Γ decoupled multilocus selection dynamics, one for each deme. For a single deme, this dynamics is well known but, in general, may be complicated.

We write

$$m_{\alpha\beta} = \delta_{\alpha\beta} + a_{\alpha\beta}\,.$$

Because $M = (m_{\alpha\beta})$ is a stochastic matrix, we have

$$a_{\alpha\beta} \geq 0 \text{ for every } \beta \neq \alpha, \text{ and } \sum_{\beta} a_{\alpha\beta} = 0 \text{ for every } \alpha.$$

With $A = (a_{\alpha\beta})$, we have $M = I + A$.

We assume

$$M = I + A\,, \tag{1}$$

where $|a_{\alpha\beta}| \leq \epsilon$, and

$$w_{ij,\alpha} = \sum_n u^{(n)}_{i_n j_n,\alpha} + r_{ij,\alpha}\,, \tag{2}$$

where $u^{(n)}_{i_n j_n,\alpha} > 0$, $|r_{ij,\alpha}| \leq \eta$, and

$$\eta = \eta(\epsilon)\,. \tag{3}$$

Here, $\eta : [0,1) \to [0,\infty)$ is C^1 and satisfies $\eta(0) = 0$. Therefore, migration and epistasis need not be 'equally' weak. In particular, the case $\eta \equiv 0$, i.e., no epistasis, is included.

We write

$$U = (u^{(n)}_{i_n j_n, \alpha}) \qquad \text{and} \qquad W = (w_{ij,\alpha}) \text{ with } w_{ij,\alpha} \text{ as in (2)}.$$

Theorem 3. *Suppose that U is such that in the absence of epistasis ($\epsilon = 0$), every equilibrium of (NoMig) is hyperbolic (this holds generically!), all recombination rates c_{K} are fixed, and $\epsilon > 0$ is sufficiently small. Then, for every parameter combination (W, M) satisfying (1), (2), and (3), the following holds:*

(a) The set of equilibria $\Sigma_0 \subset \Delta_I^\Gamma$ of (NoMig) with $\epsilon = 0$ contains only isolated points, as does the set of equilibria $\Sigma_{(W,M)} \subset \Delta_I^\Gamma$ of (SRM). As $\epsilon \to 0$ in (2), each equilibrium in $\Sigma_{(W,M)}$ converges to the corresponding equilibrium in Σ_0.

(b) *In the neighborhood of each asymptotically stable equilibrium in Σ_0, there exists exactly one equilibrium point in $\Sigma_{(W,M)}$, and it is asymptotically stable. In the neighborhood of each unstable internal equilibrium in Σ_0, there exists exactly one equilibrium point in $\Sigma_{(W,M)}$, and it is unstable. In the neighborhood of each unstable boundary equilibrium in Σ_0, there exists at most one equilibrium point in $\Sigma_{(W,M)}$, and if it exists, it is unstable.*

(c) *Every solution $p(t)$ of (SRM) converges to one of the equilibrium points in $\Sigma_{(W,M)}$.*

PROOF. Essentially, it follows that of Theorem 2.1 in Nagylaki, Hofbauer, and Brunovský (1999) for weak epistasis and strong recombination (no migration). In fact, the above is a 'uniform' version and thus (slightly) stronger than theirs, even in the absence of migration.

For a single locus, the above theorem was proved by Nagylaki and Lou (TPB, 2007).

Other limiting cases:

- Strong migration, weak selection and recombination: convergence to spatial quasi-homogeneity.
- Strong migration and selection, weak recombination: the limiting dynamics is essentially a single-locus migration-selection model.
- Weak migration, weak selection, strong recombination: the limiting dynamics is a cartesian product of the 'slow evolution limits' for single loci (NL07).
- All evolutionary forces are weak: then, selection, recombination, and migration become decoupled.

In any of these cases, limit cycles can occur. Local perturbation results can be proved.

MAIN RESULTS, PART II:

MAINTENANCE OF MULTILOCUS POLYMORPHISM

1. Strong migration
2. Weak migration
3. Levene model

1. Polymorphism under Strong Migration

We prove that for two or more demes, arbitrarily many alleles can be maintained at arbitrarily many loci at a stable equilibrium.

This does neither require epistasis nor overdominance or underdominance. It is impossible, however, in the absence of epistasis AND dominance.

Theorem 4. *Let $L \geq 1$, $\Gamma = 2$, $I_n \geq 2$ for every $n \in \mathsf{L}$, let all recombination rates c_{K} be positive and fixed, and suppose (SRM).*

(a) There exists an open set Q of migration and selection parameters, such that for every parameter combination in Q, there is a unique, internal, asymptotically stable equilibrium point. This equilibrium is in quasi-linkage equilibrium, is spatially quasi-homogeneous, and attracts all trajectories with internal initial condition. Furthermore, every trajectory converges to an equilibrium point as $t \to \infty$.

(b) Such an open set, Q', also exists if the set of all fitnesses is restricted to be nonepistatic and to display partial dominance at every locus and in every deme.

PROOF. First, such an open set, with no epistasis and with partial dominance, is *constructed* for the weak-selection limit by extending the procedure in Remark 4.15 of Nagylaki and Lou (2007).

Then, the perturbation theorem (Theorem 2), together with various arguments in its proof, is invoked to extend this to the full SRM dynamics and to weak epistasis.

Remark. The above theorem also holds if migration is restricted to the Levene model.

Theorem 4 does not hold in the absence of dominance. Instead, we have:

Proposition 5. *If selection is nonepistatic and sufficiently weak relative to migration and recombination, then, generically, there exists a globally asymptotically stable* monomorphic *state if single-locus fitnesses are*

(a) additive, i.e., there is no dominance, or

(b) multiplicative.

In particular, no polymorphism is maintained.

Remark. Statement (a) can be generalized to DIDID.

2. Polymorphism under Weak Migration

For weak or no epistasis and intermediate dominance, we prove that as many alleles can be maintained at any locus as there are demes. The number of polymorphic loci is arbitrary.

Generically, it is impossible to maintain more alleles than there are demes.

To formulate the main theorem, we recall from above:

$$M = I + A\,, \tag{1}$$

where $|a_{\alpha\beta}| \leq \epsilon$, and

$$w_{ij,\alpha} = \sum_n u^{(n)}_{in jn,\alpha} + r_{ij,\alpha}\,, \tag{2}$$

where $u^{(n)}_{in jn,\alpha} > 0$, $|r_{ij,\alpha}| \leq \eta$, and

$$\eta = \eta(\epsilon)\,. \tag{3}$$

We denote by M_ϵ the set of all backward migration matrices of the form (1), where $|a_{\alpha\beta}| \leq \epsilon$.

Theorem 6. *Let $L \geq 1$, $\Gamma \geq \max_n I_n$, let all recombination rates c_{K} be positive and fixed, and suppose (SRM).*

(a) There exists an open set P_ϵ of migration and selection parameters (with $M \in \mathsf{M}_\epsilon$), such that a unique, fully polymorphic, asymptotically stable equilibrium exists for every $(W, M) \in \mathsf{P}_\epsilon$. This equilibrium exhibits weak linkage disequilibrium within every deme and attracts all trajectories with internal initial condition. Furthermore, every trajectory converges to an equilibrium point as $t \to \infty$.

(b) The set P_ϵ can be chosen such that either $w_{ii,\alpha} < w_{ij,\alpha} < w_{jj,\alpha}$ or $w_{ii,\alpha} > w_{ij,\alpha} > w_{jj,\alpha}$ holds $\forall i$, $\forall j \neq i$, and $\forall \alpha$.

(c) Such an open set, P'_ϵ, also exists if the set of all fitnesses are restricted to be nonepistatic and to display partial dominance.

Remarks

1. In contrast to the case of strong migration, here the number of alleles that is maintained at a locus cannot exceed the number of demes.

2. Also in contrast to strong migration, absence of dominance at every locus and in every deme is admitted. Indeed, any intermediate level of dominance is admitted.

3. Initially, it is counterintuitive that under strong migration more alleles can be maintained than under weak migration because it is generally assumed that a spatially heterogeneous environment is more efficient in maintaining genetic variation.

   However, the proofs indicate that the conditions for maintaining all alleles at a locus under strong migration are much more stringent than the conditions for maintaining Γ alleles in the weak migration case.

Theorem 7. *For arbitrary number of loci, sufficiently weak migration and epistasis, and partial dominance, the number of demes is the generic maximum for the number of alleles that can be maintained at any locus at any equilibrium (stable or not) of (SRM).*

Remark. For a single multiallelic locus, these results were proved by Nagylaki and Lou (2007).

Remark. In the Levene model, there are situations where the number of loci that can be maintained polymorphic depends on the number of demes!

Many thanks to:

Thomas Nagylaki

Josef Hofbauer

The Vienna Science and Technology Fund (WWTF)

The Program for Evolutionary Dynamics at Harvard University

3. Polymorphism in the Levene Model

The Levene model is characterized by the following migration scheme:

$$m_{\alpha\beta} = c_\beta \text{ for every } \alpha, \beta\,,$$

where $c_\beta > 0$ and $\sum_\beta c_\beta = 1$. Equivalently, individuals are randomly distributed across demes, in proportion to the size of each deme.

After one round of migration, allele and gamete frequencies are uniform among demes.

In the following, we assume soft selection and no epistasis.

General Results by Nagylaki (unpubl.)

- Let

$$\tilde{w}(\rho) = \prod_{\alpha} \bar{w}_{\alpha}^{c_\alpha} \, .$$

  We have $\Delta\tilde{w} \geq 0$, and $\Delta\tilde{w} = 0$ only when the gene frequencies are at equilibrium.

- The internal gene-frequency equilibria are the stationary points of $\tilde{w}(\rho)$, and these satisfy $(\forall i_n, \forall n)$

$$\sum_{\alpha} \frac{c_\alpha}{\bar{w}_\alpha(\rho)} [u_{i_n,\alpha}^{(n)}(p^{(n)}) - \bar{u}_{\alpha}^{(n)}(p^{(n)})] = 0 \, .$$

Nagylaki's results, continued

- For additive genes (no epistasis *and* no dominance), there exists exactly one stable gene-frequency equilibrium (point or manifold) and it is globally attracting. If there exists an internal gene-frequency equilibrium, it is the global attractor.

Interestingly, this is not true for multiplicative fitnesses!

- Generically, linkage disequilibria convergence to zero at a geometric rate. In particular, the linkage-equilibrium manifold is invariant.

Nagylaki's results, continued

- Suppose there are Γ demes and L diallelic loci with DIDID. If $L \geq \Gamma$, generically there exists no internal equilibrium.

In contrast to the previous results for weak or strong migration, here the number of loci that can be maintained polymorphic depends on the number of demes, i.e., at most $\Gamma - 1$ loci can be polymorphic.

We obtain complementary results and study why this is so.

In fact, we already know from the remark following Theorem 4 that in the Levene model arbitrarily many loci can be maintained polymorphic if there is some dominance.

Theorem 8. *Suppose there are $L \geq 1$ diallelic loci with DIDID. If $\Gamma \geq L + 1$, then for an open subset of parameters, a unique internal equilibrium point of the dynamics of gamete frequencies exists, which is in linkage equilibrium and globally attracting.*

Thus, less than Γ loci can be maintained polymorphic on an open set.

One can show that in the nongeneric case, where an internal equilibrium exists with DIDID and $L \geq \Gamma$, there exists in fact a manifold of equilibria.

Theorem 9. *Suppose there are $L \geq \Gamma \geq 1$ diallelic loci with DIDID. If there exists an internal equilibrium, then there exists a manifold of equilibria containing that equilibrium. Generically, this manifold has dimension $L - \Gamma + 1$.*