The Evolutionary Reduction Principle for Linear Variation in Genetic Transmission

Lee Altenberg [*]
University of Hawai'i at Manoa [†]

Abstract

The evolution of genetic systems has been analyzed through the use of modifier gene models, in which a neutral gene is posited to control the transmission of other genes under selection. Analysis of modifier gene models has found the manifestations of an "evolutionary reduction principle": in a population near equilibrium, a new modifier allele that scales equally all transition probabilities between different genotypes under selection can invade if and only if it reduces the transition probabilities. Analytical results on the reduction principle have always required some set of constraints for tractability: limitations to one or two selected loci, two alleles per locus, specific selection regimes or weak selection, specific genetic processes being modified, extreme or infinitesimal effects of the modifier allele, or tight linkage between modifier and selected loci. Here, I prove the reduction principle in the absence of any of these constraints, confirming a twenty-year-old conjecture. The proof is obtained by a wider application of Karlin's Theorem 5.2 (1982) and its extension to ML-matrices, substochastic matrices, and reducible matrices.

Keywords: evolution; evolutionary theory; modifier gene; recombination rate; mutation rate; spectral analysis; reduction principle; Karlin's theorem; ML-matrix; essentially non-negative matrix.

1 Introduction

Darwinian evolution occurs through the interaction of two fundamental processes: (1) natural selection, i.e. differential survival and reproduction; and (2) genetic transformation, i.e. change of genetic *content* during reproduction, which provides the variation upon which selection can act. The principal genetic transformations are sexual reproduction, recombination, and mutation, while a growing list of other transformations

1

includes gene conversion, methylation, deletions, duplications, insertions, transpositions, and other chromosomal alterations.

These two processes, augmented by a third — the randomness of sampling in finite populations — provide the basis for our causal explanations of the characteristics of organisms. In its simplest version: Transformation processes create new genetic states among offspring, and differential survival and reproduction of alternate genetic states result in the prevalence of states with the highest levels of survival and reproduction. In a more sophisticated version: evolutionary dynamics are the result of the joint action of selection, transformation, and random sampling processes that move populations to distribute over certain regions of genotype space.

The simpler version of the explanation runs into a quandary when trying to explain traits that in themselves do not cause different levels of survival or reproduction, in particular, the traits that make up the molecular, cellular, and organismal machinery of the genetic transformation processes themselves. Variation in the transformation machinery can produce differing distributions of offspring genotypes, without necessarily affecting the parent's survival or their quantity of offspring. To understand the fate of variation in the genetic transformation machinery requires the more sophisticated version of evolutionary causation, the evolutionary dynamics of the joint action of selection, transformation, and sampling processes.

The earliest mathematical treatments of evolutionary dynamics in the 1920s 'Modern Synthesis' of Darwinism and Mendelian genetics (Fisher, 1922; Haldane, 1924; Wright, 1931) dealt straightaway with many complex issues; however, the first analysis of genetic variation in the genetic transformation processes themselves waited another thirty years for Kimura's (1956) analysis of a model of recombination modification. The model examines the fate of a chromosomal alteration that eliminates recombination between two loci with a stable polymorphism that exhibits linkage disequilibrium. When the alteration occurs in the chromosome that has above-average fitness, it increases in frequency. This result affirmed Fisher's (1930, p. 130) assertion that "the presence of pairs of factors in the same chromosome, the selective advantage of each of which reverses that of the other, will always tend to diminish recombination, and therefore to increase the intensity of linkage in the chromosomes of that species." This result was the first instance of what was later to be called the "reduction principle" for the evolution of genetic transformations (Feldman, 1972; Feldman et al., 1980; Altenberg, 1984; Liberman and Feldman, 1986b; Altenberg and Feldman, 1987).

Nei (1967) introduced, and partially analyzed, a model for the evolution of recombination in which recombination rates between two loci are modified by a third, neutral locus. Feldman (1972) gave a complete linear stability analysis of the model under the assumption of additive, multiplicative, and symmetric viability selection regimes. He found that for populations near a polymorphic equilibrium with linkage disequilibrium under selection and recombination, genetic variation for recombination would survive if and only if it reduced the rate of recombination between the loci under selection.

Other transformation processes were analyzed with modifier gene models, in particular mutation and migration rates, and the reduction principle was found to emerge again (Karlin and McGregor, 1972; Feldman and Balkau, 1973; Balkau and Feldman, 1973; Karlin and McGregor, 1974; Feldman and Krakauer, 1976; Feldman et al., 1980).

Note that with migration modification, it is an organism's *location* that is being

2

transmitted and transformed. The idea that spatial location of organisms, and other environmental or acquired conditions, may function formally as heritable traits subject to transformation processes (Cavalli-Sforza and Feldman 1973; Karlin and McGregor 1974; Altenberg 1984, pp. 15-16, p. 178), is gaining recent acceptance, e.g. Schauber et al. (2007); Odling-Smee (2007). This concept follows quite naturally from the concept of generalized transmission that appears in models of cultural transmission and modifier genes.

These first reduction results were derived under narrow constraints on the selection regime, number of modifier alleles, or number of alleles under selection. Subsequent studies of modifiers of these three processes — recombination, mutation, and migration — have extended the result to cases of modifier polymorphisms and arbitrary selection coefficients (Liberman and Feldman, 1986a,b; Feldman and Liberman, 1986; Liberman and Feldman, 1989), but are still restricted to two alleles per selected locus, or two demes in the case of migration modification.

The repeated emergence of the reduction result in modifier models of different processes led to a study of mathematical underpinnings that might be common to all of them (Altenberg, 1984). The approach taken was to represent all possible transmission processes — in which recombination and mutation comprise special cases — using a general bi-parental transmission matrix, consisting of probabilities $T(i \leftarrow j, k)$ that parental haplotypes j and k produce a gamete haplotype i. Haplotype here refers to a gamete's genotype, or a gamete's contribution to a diploid genotype. (The addition of spatial subdivisions in the case of migration modification leads to a slightly changed representation (Altenberg, 1984, pp. 178-199).)

This approach allows the particulars of the processes to be abstracted out of the model, and reveals that what all the models have in common is the nature of the *variation* in transmission produced by variation at the modifier locus. All of the models fit the form

$$T_\alpha(i \leftarrow j, k) = \alpha \, P(i \leftarrow j, k) \text{ for } j, k \neq i,$$

where α is the modified parameter that represents an overall rate of transformation of haplotypes, over all haplotypes. This form is referred to as *linear variation* (Altenberg, 1984; Altenberg and Feldman, 1987) because the modifier gene scales all transmission probabilities between different haplotypes equally.

In the perturbation analysis of the evolutionary dynamics of the modifier locus (to be explained in detail in the next section), the stability matrix under linear variation has the form:

$$\mathbf{M}(\alpha, r)\mathbf{D} = \left\{ (1 - \alpha)[(1 - r)\mathbf{I} + r\mathbf{Q}] + \alpha[(1 - r)\mathbf{S} + r\mathbf{S}^{\Join}] \right\} \mathbf{D}, \qquad (1)$$

where α is the transmission parameter produced by the new modifier allele, $\mathbf{Q}$, $\mathbf{S}$, and $\mathbf{S}^{\Join}$ are stochastic matrices, $\mathbf{D}$ is a positive diagonal matrix, and r is the rate of recombination between the modifier and the nearest locus under selection.

The rare modifier allele will increase when rare at a geometric rate if and only if the spectral radius $\rho(\mathbf{M}(\alpha, r)\mathbf{D})$ exceeds 1. This formulation is extremely general, accommodating modifier models with arbitrary numbers of modifier alleles, numbers of loci and alleles per locus under selection, selection regimes, and transmission processes.

Although the model is extremely general, it is tractable thanks to a theorem of Karlin (1982):

Theorem 1 (Karlin Theorem 5.2, (1982, pp. 194–196))**.** *Let $\mathbf{M}$ be an arbitrary non-negative irreducible stochastic matrix. Consider the family of matrices*

$$\mathbf{M}(\alpha) = (1 - \alpha)\mathbf{I} + \alpha\mathbf{M}$$

Then for any diagonal matrix $\mathbf{D}$ with positive terms on the diagonal, the spectral radius

$$\rho(\alpha) = \rho(\mathbf{M}(\alpha)\mathbf{D})$$

is decreasing as α increases (strictly provided $\mathbf{D} \neq d\mathbf{I}$).

We see that $\mathbf{M}(\alpha, r)\mathbf{D}$ fits the form in Karlin's theorem only if $r = 0$, in which case the stability matrix becomes $\left\{(1 - \alpha)\mathbf{I} + \alpha\mathbf{S}\right\}\mathbf{D}$, and we immediately see that a new modifier can invade a population near equilibrium if and only if it reduces α so as to produce $\rho(\mathbf{M}(\alpha)\mathbf{D}) > 1$ (Altenberg 1984, Theorem 3.9, pp. 126-128; Altenberg and Feldman 1987, Result 3). If $r > 0$, one is left to evaluate

$$\frac{d}{d\alpha}\rho([(1 - \alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2]\mathbf{D}), \tag{2}$$

where $\mathbf{M}_1, \mathbf{M}_2 \neq \mathbf{I}$. A closed-form characterization of stochastic matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ that produce $d\rho/d\alpha < 0$ is not readily obtained.

The case of general $r > 0$ would seem unlikely to reverse the reduction result, since the only action of recombination with the modifier locus is to blend the equilibrium distribution of selected haplotypes with the distribution created by the new modifier allele. This blending should not alter the tendency of modifier alleles that reduce the transformation rate to become associated with haplotypes of above-average fitness (and the converse), which is the essence of the dynamics. The blending would merely lessen the association. Even free recombination cannot completely eliminate this emergent association: a modifier allele that produces perfect transmission ($\alpha = 0$) always invades when introduced, even if unlinked to the selected loci, i.e. $\rho(\mathbf{M}(0, 1/2)\,\mathbf{D}) > 1$ (Altenberg 1984, Theorem 3.5, p.118; Altenberg and Feldman 1987, Result 2).

Thus it was conjectured in Altenberg and Feldman (1987) that for arbitrary $r \in [0, 1/2]$, a new modifier allele can invade if and only if it decreases α. This paper proves the conjecture through an extension and expanded use of Karlin's theorem. The proof requires the extension of Karlin's theorem to essentially non-negative matrices (ML matrices), and to reducible matrices. This last of the constraints needed to prove the general reduction principle is thus removed.

I preface the result with a self-contained review of the general modifier model developed in Altenberg (1984); Altenberg and Feldman (1987) and also used in Zhivotovsky et al. (1994).

2 The Model

The evolutionary model examined here fits the general form

$$\overline{w}\, z_i' = \sum_{jk} \mathcal{T}(i \leftarrow j, k)\, w_{jk}\, z_j\, z_k, \tag{3}$$
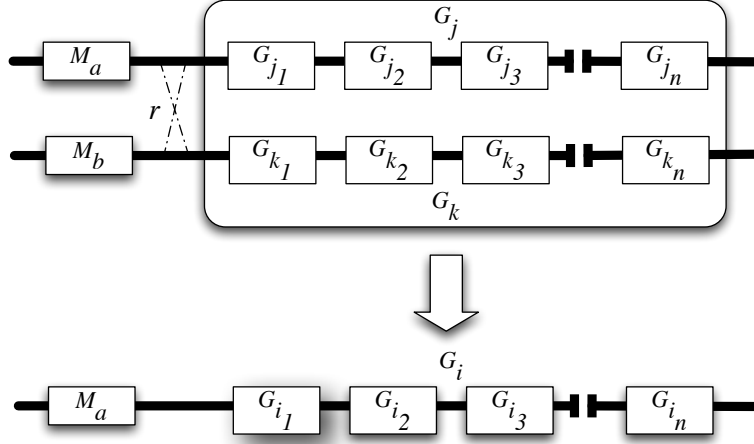
4

Figure 1: The configuration of loci in the modifier gene model considered here. M_a and M_b are the modifier alleles, G_j and G_k are the haplotypes undergoing viability selection, consisting of genes $G_{j_1} \ldots G_{j_n}$ and $G_{k_j} \ldots G_{k_n}$. With probabilities under the control of the modifier genotype M_a/M_b, haplotype $M_a G_i$ is produced.

where z_i is the frequency of haplotype i in the population, z_i' is the frequency in the next generation, $\mathcal{T}(i \leftarrow j, k)$ is the probability that parental haplotypes j and k produce an offspring haplotype i, $w_{jk} = w_{kj}$ is the fitness of diploid genotype jk, and $\overline{w} = \sum_{jk} w_{jk} z_j z_k$ is the mean fitness of the population. The model includes the general assumptions of an infinite population, frequency-independent viability selection, random mating, sex symmetry, no sex-linkage, and non-overlapping generations.

2.1   The Modifier Gene Model

The modifier gene model is a special case of (3) in which the genome is structured to contain a group of loci under selection, and a neutral locus, external to the group, that modifies their genetic transmission probabilities. The structure is illustrated in Fig. 1. Haplotypes will now have two indices, one for the allele at the modifier locus (a, b, c etc.), and one for the haplotype of the selected loci (i, j, k etc.). The modifier allele is assumed to be transmitted perfectly (no mutation nor segregation distortion), so that the only force acting upon it arises from its association with the selected loci. Recombination between the modifier locus and the nearest selected locus occurs at rate r_{ab}.

The situation where the modifier locus is internal to the set of selected loci is not considered because it is no longer possible to separate recombination events with the modifier locus from transformations acting on the selected haplotype. In order for the variation in transmission to be linear in this configuration, there would need to be complete interference between recombination events on either side of the modifier

locus and any other transformation processes being modeled.

Under this genetic structure, the transmission probabilities take on the form:

$$T_{(r)}(ai \leftarrow aj|bk) := (1 - r_{ab})T(ai \leftarrow aj|bk) + r_{ab}T^{\bowtie}(ai \leftarrow ak|bj)$$

where the probability that parental genotype aj, bk produces gamete haplotype ai, conditioned on the offspring containing modifier allele a, is:

$T(ai \leftarrow aj|bk)$, when no recombination occurs between the modifier and nearest selected locus, and

$T^{\bowtie}(ai \leftarrow ak|bj)$, when recombination occurs between the modifier and nearest selected locus (hence $aj|bk$ becomes $ak|bj$).

So, $1 = \sum_i T(ai \leftarrow aj|bk) = \sum_i T^{\bowtie}(ai \leftarrow ak|bj)$, $\forall a, b, j, k$.

2.2 Dynamical Recursion for the Modifier Model

With this genetic structure, recursion (3) becomes:

$$\overline{w} \, z'_{ai} = \sum_{bjk} T_{(r)}(ai \leftarrow aj|bk) \, w_{jk} \, z_{aj} \, z_{bk} \tag{4}$$

where

z_{ai} is the frequency of the haplotype with allele a at the modifier locus, and haplotype i at the selected loci, and z'_{ai} is for the next generation;

w_{jk} is the fitness coefficient for diploid genotype jk at the loci under selection;

$\overline{w} := \sum_{abjk} w_{jk} \, z_{aj} \, z_{bk}$ is the mean fitness of the population.

The evolutionary analysis consists of asking how the values of $T_{(r)}(ai \leftarrow aj|bk)$ determine whether a modifier allele a can invade a population and be protected from extinction.

3 Equilibria and their Stability

A population at equilibrium under (4) must satisfy the constraint:

$$\widehat{\overline{w}} \, \hat{z}_{bi} = \sum_{cjk} T_{(r)}(bi \leftarrow bj|ck) \, w_{jk} \, \hat{z}_{bj} \, \hat{z}_{ck}. \tag{5}$$

A perturbation of the equilibrium to $z_{bi} = \hat{z}_{bi} + \epsilon_{bi}$ produces:

$$\left(\widehat{\overline{w}} + 2 \sum_{bjck} \epsilon_{bj} w_{jk} \hat{z}_{ck} + \sum_{bjck} \epsilon_{bj} \epsilon_{ck} \right) (\hat{z}_{bi} + \epsilon'_{bi})$$
$$= \sum_{cjk} T_{(r)}(bi \leftarrow bj|ck) \, w_{jk} \, (\hat{z}_{bj} + \epsilon_{bj}) \, (\hat{z}_{ck} + \epsilon_{ck}).$$

6

The long-term evolution of genetic transmission depends on the properties that allow a new modifier allele to invade a population and be protected from extinction. Hence the analysis focuses on perturbations of the equilibrium by rare modifier alleles, entailing $\hat{z}_{ai} = 0$ for all i for new modifier allele a. Making this substitution, and ignoring all second and higher order terms in the perturbation, the linear recursion on a new modifier allele, a, that perturbs (5) can be represented in vector form as:

$$\boldsymbol{\epsilon}'_a = \mathbf{M} \, \mathbf{D} \, \boldsymbol{\epsilon}_a \tag{6}$$

where $\mathbf{M}$ is a (column) stochastic matrix, $\mathbf{D}$ a positive diagonal matrix, and

$$\boldsymbol{\epsilon}_a := [\epsilon_{ai}]_{i=1}^n, \quad \mathbf{D} := \mathbf{diag}\left[\hat{w}_i/\widehat{\overline{w}}\right]_{i,j=1}^n, \quad \hat{w}_i := \sum_{bj} w_{ij} \hat{z}_{bj}, \text{ and}$$

$$\mathbf{M} := \left[\sum_{bk} T_{(r)}(ai \leftarrow aj|bk) \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n .$$

Modifier allele a will increase at a geometric rate when rare if and only if the spectral radius $\rho(\mathbf{MD})$ exceeds 1. Clearly, if $\mathbf{D} = \mathbf{I}$, then $\rho(\mathbf{MD}) = \rho(\mathbf{M}) = 1$, so geometric rates of change in modifier allele frequencies require $\mathbf{D} \neq \mathbf{I}$, a situation described by saying there is a positive *selection potential* (Altenberg 1984, "fitness load" p. 63; Altenberg and Feldman 1987):

$$V = \frac{\max_i \hat{w}_i}{\widehat{\overline{w}}} - 1 > 0. \tag{7}$$

The analysis consists of evaluating the relationship between $T_{(r)}(ai \leftarrow aj|bk)$ and $\rho(\mathbf{MD})$.

4   Variation in Transmission

The reduction principle emerges in models where the modifier gene scales all the transition probabilities between different genotypes equally. Variation in transmission falls along a line that intersects the matrix for perfect transmission (where parental haplotypes are transmitted unchanged to their gametes, and in equal proportions), which functions as an "origin" in this space of matrices. For this reason it is called "linear variation" (Altenberg, 1984; Altenberg and Feldman, 1987).

A mechanistic derivation of linear variation is that each selected haplotype j has a certain probability, α, of being "hit" by some transforming process, and given that it is hit, it is transformed into various other selected haplotypes with different probabilities. These probabilities, $T(i \leftarrow j|k)$, include possible dependence on both parental haplotypes j and k. When the effect of the modifier gene is to scale the "hit" rate, α, up or down equally for all haplotypes, it produces linear variation in transmission.

4.1 Perfect Transmission

A genetic system that perfectly transmits parental selected haplotypes to gametes can be represented by:

$$T_{\mathbf{id}}(ai \leftarrow aj|bk) = T_{\mathbf{id}}^{\bowtie}(ai \leftarrow aj|bk) = \delta_{ij}, \tag{8}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Any genetic system can be characterized by the probability with which perfect transmission occurs. For a given modifier genotype ab, a global characterization of the extent of perfect transmission can be captured by a lower bound, $1 - \alpha_{ab}$, on the probability of perfect transmission over all selected haplotypes, where

$$1 - \alpha_{ab} := \min_{i,k} \left\{ T(ai \leftarrow ai|bk), T^{\bowtie}(ai \leftarrow ai|bk) \right\} \in [0, 1]. \tag{9}$$

This lower bound can be used to parameterize the transmission probabilities:

$$T(ai \leftarrow aj|bk) = (1 - \alpha_{ab})\, \delta_{ij} + \alpha_{ab}\, P(ai \leftarrow aj|bk) \tag{10}$$

and

$$T^{\bowtie}(ai \leftarrow aj|bk) = (1 - \alpha_{ab})\, \delta_{ij} + \alpha_{ab}\, P^{\bowtie}(ai \leftarrow aj|bk), \tag{11}$$

Thus $P(ai \leftarrow aj|bk) \geq 0$ and $P^{\bowtie}(ai \leftarrow aj|bk) \geq 0$ for all i, j, k, and for some i, k, $P(ai \leftarrow ai|bk) = 0$ or $P^{\bowtie}(ai \leftarrow ai|bk) = 0$.

Linear variation is defined to be where the sole effect of the modifier locus is to determine α:

$$T(ai \leftarrow aj|bk) = (1 - \alpha_{ab})\, \delta_{ij} + \alpha_{ab}\, P(i \leftarrow j|k) \tag{12}$$

and

$$T^{\bowtie}(ai \leftarrow ak|bj) = (1 - \alpha_{ab})\, \delta_{ik} + \alpha_{ab}\, P^{\bowtie}(i \leftarrow k|j). \tag{13}$$

4.2 External Stability of a Modifier Monomorphism

For the purposes of brevity, it becomes useful at this point to restrict the analysis to populations that are initially fixed on a single modifier allele. The steps that follow apply to modifier polymorphisms, but require additional characterizations of the polymorphisms; modifier polymorphisms under generalized transmission are analyzed in Altenberg (1984), and for particular cases of transmission processes, in Liberman and Feldman (1986a,b); Feldman and Liberman (1986); Liberman and Feldman (1989).

The population will initially be at equilibrium, fixed on modifier allele b where $\alpha_{bb} > 0$. Haplotypes will be listed only if they occur with positive frequency, so $\hat{z}_{bi} > 0$ for all i at equilibrium. To this population at equilibrium, modifier allele a is introduced. Linear variation and $\alpha_{bb} > 0$ preclude the possibility that the new modifier allele will cause the production of selected haplotypes not already present in the population.

For readability, let $\hat{\alpha}$ represent α_{bb}, α represent α_{ab}, and r represent r_{ab}. For linear variation (12)(13), the stability matrix $\mathbf{M}$ in (6) can be expressed as a function of α and

r:

$$\mathbf{M}(\alpha, r) = \left[\sum_k T_{(r)}(ai \leftarrow aj|bk) \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n \tag{14}$$
$$= (1 - \alpha)[(1-r)\mathbf{I} + r\mathbf{Q}] + \alpha[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}],$$

where the matrices are derived as follows:

$$\left[\sum_k \delta_{ij} \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n = \mathbf{I},$$

$$\mathbf{Q} := \left[\sum_k \delta_{ik} \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n = \left[\frac{w_{ij}}{\hat{w}_j} \hat{z}_{bi} \right]_{i,j=1}^n,$$

$$\mathbf{S} := \left[\sum_k P(i \leftarrow j|k) \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n, \text{ and}$$

$$\mathbf{S}^{\bowtie} := \left[\sum_k P^{\bowtie}(i \leftarrow k|j) \frac{w_{jk}}{\hat{w}_j} \hat{z}_{bk} \right]_{i,j=1}^n.$$

Letting $\mathbf{e}$ be the vector of ones, and $\mathbf{e}^\top$ its transpose, note that

$$\mathbf{e}^\top \mathbf{Q} = \mathbf{e}^\top \mathbf{S} = \mathbf{e}^\top \mathbf{S}^{\bowtie} = \mathbf{e}^\top, \tag{15}$$

since $\sum_i P(i \leftarrow j|k) = \sum_i P^{\bowtie}(i \leftarrow k|j) = \sum_k w_{jk} \hat{z}_{bk}/\hat{w}_j = 1$. $\mathbf{M}(\hat{\alpha}, r)$ can be used to express the equilibrium constraint (5):

$$\hat{\mathbf{z}}_b = \mathbf{M}(\hat{\alpha}, r)\mathbf{D}\hat{\mathbf{z}}_b \tag{16}$$
$$= \left\{ (1 - \hat{\alpha})[(1-r)\mathbf{I} + r\mathbf{Q}] + \hat{\alpha}[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}] \right\} \mathbf{D}\hat{\mathbf{z}}_b.$$

5   Results

A direct approach to obtain the reduction result would be to show that $d\rho(\mathbf{M}(\alpha, r)\mathbf{D})/d\alpha < 0$ for any r, but the proof is elusive. An alternative approach is to try to apply Karlin's Theorem 5.2 by extending a line from $\mathbf{I}$ through $\mathbf{M}(\alpha, r)$, and see if it intersects any matrix $\mathcal{M}$ where the spectral radius $\rho(\mathcal{M}\mathbf{D})$ is known. Since the spectral radius decreases along this line of matrices (as long as they are non-negative), this promises an inequality between $\rho(\mathbf{M}(\alpha, r)\mathbf{D})$ and $\rho(\mathcal{M}\mathbf{D})$, depending on their order along the line.

Thus we try to represent $\mathbf{M}(\alpha, r)$ as $\mathbf{M}(\alpha, r) = (1 - \beta)\mathbf{I} + \beta\mathcal{M}$, for a choice of $\mathcal{M}$ where we know $\rho(\mathcal{M}\mathbf{D})$. Finding an irreducible stochastic such $\mathcal{M}$ would give

$$\rho(\mathbf{M}(\alpha, r)\mathbf{D}) > \rho(\mathcal{M}\mathbf{D}) \text{ for } 0 < \beta < 1,$$

and

$$\rho(\mathbf{M}(\alpha, r)\mathbf{D}) < \rho(\mathcal{M}\mathbf{D}) \text{ for } \beta > 1,$$

which is the reduction result.

For our choice of $\boldsymbol{\mathcal{M}}$, we possess the crucial fact from (16) that $\rho(\mathbf{M}(\hat\alpha, r)\mathbf{D}) = 1$, so we might hope to find β and r to give $\mathbf{M}(\alpha, r) = (1-\beta)\mathbf{I} + \beta\mathbf{M}(\hat\alpha, r)$, but no such solution exists.

Fortunately, there exists a matrix solution $\mathbf{M}(\alpha, r) = (1 - \beta)\mathbf{I} + \beta\boldsymbol{\mathcal{M}}^\star$, but $\boldsymbol{\mathcal{M}}^\star$ may have negative elements along the diagonal for certain combinations of α, $\hat\alpha$, and r, and thus does not meet the conditions of Karlin's Theorem 5.2. Such a matrix is known as a 'Metzler–Leontief' (Seneta, 1981) or 'essentially nonnegative' (Cohen, 1981) matrix. Fortunately, Karlin's Theorem 5.2 can be extended to ML-matrices, which allows proof for all α, $\hat\alpha$, and r. We proceed in stages.

5.1 Metzler–Leontief (ML) Matrices

An ML-matrix is a square real matrix where all off-diagonal elements are non-negative. The *spectral abscissa* of a matrix $\mathbf{A}$ is defined as $\pi(\mathbf{A}) = \max_i\{\mathrm{Re}(\lambda_i)\}$, where $\{\lambda_i\}$ are the eigenvalues of $\mathbf{A}$. Let $\mathbf{B}$ refer to an ML-matrix, where $[\mathbf{B}]_{ij} \geq 0$ for all $i \neq j$. We utilize the following properties of ML-matrices:

Lemma 1 (Spectral Abscissa of ML Matrices)**.**

1. $\pi(\mathbf{B})$ *is an eigenvalue of* $\mathbf{B}$, *referred to as the Perron root as it is for positive matrices;*

2. $\pi(\mathbf{B}) = \rho(\mathbf{B} + g\mathbf{I}) - g$, *for any* $g \geq -\min_i\{[\mathbf{B}]_{ii}\}$;

3. *If* $\mathbf{Bz} = \lambda\mathbf{B}$ *and* $\mathbf{z} > 0$, *then* $\lambda = \pi(\mathbf{B})$. *If in addition* $\mathbf{B}$ *is irreducible, then there also exists* $\mathbf{v} > 0$ *such that* $\mathbf{v}^\top\mathbf{B} = \pi(\mathbf{B})\mathbf{v}^\top$, *and* $\mathbf{z}$ *and* $\mathbf{v}$ *are unique up to constant multiples.*

Proof. These are found in, or follow directly from, Theorem 2.6 in (Seneta, 1981, pp. 45–46) and Theorem 3 in (Gantmacher, 1959, p. 66). $\qquad\square$

5.2 Solving for $\mathcal{M}^*$

Lemma 2. *Define*

$$\boldsymbol{\mathcal{M}}(\alpha, c_1, c_2) := (1 - \alpha)\left[(1 - c_1)\,\mathbf{I} + c_1\mathbf{Q}\right] + \alpha[(1 - c_2)\mathbf{S} + c_2\mathbf{S}^{\bowtie}].$$

Then for $\alpha, \hat\alpha \in (0, 1)$, *and* $r \in [0, 1]$:

1. $\mathbf{M}(\alpha, r) = (1-\beta)\mathbf{I} + \beta\boldsymbol{\mathcal{M}}(\hat\alpha, c_1, c_2)$ *is solved by*

$$\beta = \alpha/\hat\alpha, \ c_1 = r\hat\alpha(1 - \alpha)/[\alpha(1 - \hat\alpha)], \ and \ c_2 = r,$$

to give

$$\mathbf{M}(\alpha, r) = \left(1 - \frac{\alpha}{\hat\alpha}\right)\mathbf{I} + \frac{\alpha}{\hat\alpha}\boldsymbol{\mathcal{M}}^\star, \tag{17}$$

where

$$\boldsymbol{\mathcal{M}}^{\star} := \boldsymbol{\mathcal{M}}\left(\hat{\alpha}, r\frac{\hat{\alpha}(1-\alpha)}{\alpha(1-\hat{\alpha})},\, r\right) \tag{18}$$

$$= \left(1-\hat{\alpha}-r\frac{\hat{\alpha}}{\alpha}+r\hat{\alpha}\right)\mathbf{I} + r\hat{\alpha}\left(\frac{1}{\alpha}-1\right)\mathbf{Q} + \hat{\alpha}[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}].$$

2. $\mathbf{e}^{\top}\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, r) = \mathbf{e}^{\top}$ *and* $\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, r)\,\mathbf{D}\,\hat{\mathbf{z}}_b = \hat{\mathbf{z}}_b$ *for any* $c_1 \in \Re$.

3. $\boldsymbol{\mathcal{M}}^{\star}$ *is an ML-matrix and, if* $(1-\hat{\alpha})r\mathbf{Q} + \hat{\alpha}[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}]$ *is irreducible,* $\pi\left(\boldsymbol{\mathcal{M}}^{\star}\,\mathbf{D}\right) = 1$.

4. *If, in addition,* $\alpha \geq \hat{\alpha}$ *or* $\alpha \geq \hat{\alpha}\,r/[1-\hat{\alpha}(1-r)]$, *then* $\boldsymbol{\mathcal{M}}^{\star}$ *is non-negative and* $\pi\left(\boldsymbol{\mathcal{M}}^{\star}\,\mathbf{D}\right) = \rho\left(\boldsymbol{\mathcal{M}}^{\star}\,\mathbf{D}\right) = 1$.

Proof. Straightforward evaluation verifies (17) and (18) from:

$$\boldsymbol{\mathcal{M}}^{\star} = (1-\hat{\alpha})\left[\left(1-r\frac{\hat{\alpha}(1-\alpha)}{\alpha(1-\hat{\alpha})}\right)\mathbf{I} + r\frac{\hat{\alpha}(1-\alpha)}{\alpha(1-\hat{\alpha})}\mathbf{Q}\right]$$
$$+ \hat{\alpha}\ [(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}].$$

By (15), $\mathbf{e}^{\top}\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, c_2) = [(1-\hat{\alpha})(1-c_1+c_1) + \hat{\alpha}(1-c_2+c_2)]\mathbf{e}^{\top} = \mathbf{e}^{\top}$. Observe that $\mathbf{I}\mathbf{D}\hat{\mathbf{z}}_b = \mathbf{Q}\mathbf{D}\hat{\mathbf{z}}_b = \mathbf{D}\hat{\mathbf{z}}_b$:

$$\mathbf{Q}\mathbf{D}\hat{\mathbf{z}}_b = \left[\frac{w_{ij}}{\hat{w}_j}\hat{z}_{bi}\right]_{i,j=1}^{n}\mathbf{diag}\left[\frac{\hat{w}_i}{\widehat{\overline{w}}}\right]_{i,j=1}^{n}\hat{\mathbf{z}}_b$$

$$= \left[\sum_j \frac{w_{ij}}{\widehat{\overline{w}}}\hat{z}_{bi}\hat{z}_{bj}\right]_{i,j=1}^{n} = \left[\frac{\hat{w}_i}{\widehat{\overline{w}}}\hat{z}_{bi}\right]_{i,j=1}^{n} = \mathbf{D}\hat{\mathbf{z}}_b.$$

Hence $\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, r)\,\mathbf{D}\,\hat{\mathbf{z}}_b =$

$$(1-\hat{\alpha})\left[(1-c_1)\,\mathbf{I} + c_1\mathbf{Q}\right]\mathbf{D}\hat{\mathbf{z}}_b + \hat{\alpha}[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}]\mathbf{D}\hat{\mathbf{z}}_b$$

is invariant for all $c_1 \in \Re$. Since $\boldsymbol{\mathcal{M}}(\hat{\alpha}, r, r) := \mathbf{M}(\hat{\alpha}, r)$, then

$$\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, r)\mathbf{D}\hat{\mathbf{z}}_b = \boldsymbol{\mathcal{M}}(\hat{\alpha}, r, r)\mathbf{D}\hat{\mathbf{z}}_b = \mathbf{M}(\hat{\alpha}, r)\mathbf{D}\hat{\mathbf{z}}_b = \hat{\mathbf{z}}_b. \tag{19}$$

The off-diagonal elements in (18) are non-negative as they all derive from non-negative stochastic matrices — $\mathbf{Q}$, $\mathbf{S}$, and $\mathbf{S}^{\bowtie}$ — multiplied by non-negative coefficients under conditions $\hat{\alpha} \in (0, 1)$, $c_1 \geq 0$, and $r \in [0, 1]$.

If $(1-\hat{\alpha})r\mathbf{Q} + \hat{\alpha}[(1-r)\mathbf{S} + r\mathbf{S}^{\bowtie}]$ is irreducible, then since $\hat{\mathbf{z}}_b$ is strictly positive, $\hat{\mathbf{z}}_b$ is the right Perron eigenvector of $\boldsymbol{\mathcal{M}}(\hat{\alpha}, c_1, r)\mathbf{D}$ with Perron root 1, by Lemma 1 (3). The coefficient on $\mathbf{I}$, $1 - \hat{\alpha} - r\hat{\alpha}/\alpha + r\hat{\alpha}$, is non-negative when $\alpha > \hat{\alpha}$, for then $1 - \hat{\alpha} - r\hat{\alpha}/\alpha + r\hat{\alpha} \geq (1-\hat{\alpha})(1-r) \geq 0$, but is negative if $\alpha < \hat{\alpha}\,r/[1 - \hat{\alpha}(1-r)]$ (found by simple rearrangement). Thence $\boldsymbol{\mathcal{M}}^{\star}$ may have negative diagonal elements. When $\boldsymbol{\mathcal{M}}^{\star}$ is non-negative, then $\pi(\boldsymbol{\mathcal{M}}^{\star}\,\mathbf{D}) = \rho(\boldsymbol{\mathcal{M}}^{\star}\,\mathbf{D})$. $\qquad\square$

5.3 Extending Karlin's Theorem 5.2 to ML-Matrices

Friedland's (1981) Donsker–Varadhan related variational formula for the spectral radius is applied to $\mathbf{A} \geq 0$ in order to show that it applies also to the Perron root of $\mathbf{B} = \mathbf{A} - g\,\mathbf{I}$:

Lemma 3 (Variational Formula for ML-Matrices). *For any irreducible ML-matrix $\mathbf{B}$, the spectral abscissa is*

$$\pi(\mathbf{B}) = \sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} \sum_{i=1}^{n} p_i \frac{[\mathbf{Bx}]_i}{x_i}, \tag{20}$$

where $\mathcal{P}_n = \{\mathbf{p} \colon \mathbf{p} \geq 0, \mathbf{e}^\top \mathbf{p} = 1\}$.

Proof. Define for any matrix $\mathbf{A}$,

$$f(\mathbf{A}, \mathbf{p}, \mathbf{x}) := \sum_{i=1}^{n} p_i \frac{[\mathbf{Ax}]_i}{x_i}. \tag{21}$$

The Donsker–Varadhan related variational formula (Friedland, 1981, Corollary 3.1) for irreducible $\mathbf{A} \geq 0$ gives us $\rho(\mathbf{A}) = \sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} f(\mathbf{A}, \mathbf{p}, \mathbf{x})$. Substituting $\mathbf{A} = \mathbf{B} + g\,\mathbf{I}$:

$$
\begin{aligned}
\rho(\mathbf{A}) &= \rho(\mathbf{B} + g\,\mathbf{I}) = \sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} \sum_{i=1}^{n} p_i \frac{[(\mathbf{B} + g\,\mathbf{I})\mathbf{x}]_i}{x_i} \\
&= \sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} \sum_{i=1}^{n} p_i \frac{[\mathbf{Bx}]_i}{x_i} + g.
\end{aligned}
$$

Knowing from Lemma 1 (2) that $\rho(\mathbf{A}) = \pi(\mathbf{B}) + g$ proves the lemma. Moreover,

$$\sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} f(\mathbf{A}, \mathbf{p}, \mathbf{x})$$

and

$$\sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > \mathbf{0}} f(\mathbf{B}, \mathbf{p}, \mathbf{x})$$

are both attained for the same $\mathbf{p}$ and $\mathbf{x}$. $\qquad\square$

Lemma 4 (Derivative of the Spectral Abscissa). *Let $\mathbf{B}(\beta)$ be a function of $\beta \in \Re$, in continuity class C^2, such that $\mathbf{B}(\beta)$ is an ML-matrix for $\beta \geq 0$. Let $\mathbf{p}(\beta)$ and $\mathbf{x}(\beta)$ ($\mathbf{x}$ normalized so $\mathbf{e}^\top \mathbf{x}(\beta) = 1$) be the vectors at which the supremum and the infimum in (20), are attained, respectively. Then*

$$\frac{d}{d\beta}\pi(\mathbf{B}(\beta)) = \sum_{i=1}^{n} p_i(\beta) \frac{[\frac{d\mathbf{B}(\beta)}{d\beta}\,\mathbf{x}(\beta)]_i}{x_i(\beta)}.$$

12

Proof. Let $\mathbf{A}(\beta) = \mathbf{B}(\beta) + g\,\mathbf{I} \geq 0$ be a non-negative matrix associated with $\mathbf{B}(\beta)$. So $\frac{d}{d\beta}\mathbf{A}(\beta) = \frac{d}{d\beta}\mathbf{B}(\beta)$. By Lemma 1 (2), $\frac{d}{d\beta}\rho(\mathbf{A}(\beta)) = \frac{d}{d\beta}\pi(\mathbf{B}(\beta))$. Differentiating (21):

$$\frac{d}{d\beta}f(\mathbf{A}, \mathbf{p}, \mathbf{x}) = \frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{A}}\frac{d\mathbf{A}}{d\beta} + \frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{p}}\frac{d\mathbf{p}}{d\beta} + \frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{x}}\frac{d\mathbf{x}}{d\beta}$$

Since $\mathbf{p}(\beta)$ and $\mathbf{x}(\beta)$ are unique critical points of $f(\mathbf{A}(\beta), \mathbf{p}(\beta), \mathbf{x}(\beta))$ (Friedland and Karlin, 1975; Friedland, 1981; Karlin, 1982),

$$\left.\frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{p}}\frac{d\mathbf{p}}{d\beta}\right|_{\substack{\mathbf{A}(\beta),\\\mathbf{p}(\beta),\mathbf{x}(\beta)}} = \left.\frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{x}}\frac{d\mathbf{x}}{d\beta}\right|_{\substack{\mathbf{A}(\beta),\\\mathbf{p}(\beta),\mathbf{x}(\beta)}} = 0,$$

hence

$$\frac{d}{d\beta}\rho(\mathbf{A}(\beta)) = \frac{d}{d\beta}\pi(\mathbf{B}(\beta)) = \left.\frac{\partial f(\mathbf{A}, \mathbf{p}, \mathbf{x})}{\partial \mathbf{A}}\frac{d\mathbf{A}}{d\beta}\right|_{\substack{\mathbf{A}(\beta),\\\mathbf{p}(\beta),\mathbf{x}(\beta)}}$$

$$= \sum_{i=1}^{n}p_i(\beta)\frac{[\frac{d\mathbf{A}(\beta)}{d\beta}\mathbf{x}(\beta)]_i}{x_i(\beta)} = \sum_{i=1}^{n}p_i(\beta)\frac{[\frac{d\mathbf{B}(\beta)}{d\beta}\mathbf{x}(\beta)]_i}{x_i(\beta)}. \quad \square$$

Theorem 2 (Extension to ML-Matrices). *Let $\mathbf{L}$ be an irreducible ML-matrix such that $\mathbf{e}^{\top}\mathbf{L} = \mathbf{e}^{\top}$. Consider the family of matrices*

$$\mathbf{C}(\beta) = (1 - \beta)\mathbf{I} + \beta\,\mathbf{L}$$

with $\beta > 0$. Then for any diagonal matrix $\mathbf{D}$ not a multiple of $\mathbf{I}$, with positive terms on the diagonal, the spectral abscissa (and Perron root),

$$\pi(\mathbf{C}(\beta)\mathbf{D}),$$

is strictly decreasing as β increases.

Proof. Let $\mathbf{B}(\beta) = \mathbf{C}(\beta)\mathbf{D}$. Applying Lemma 4, we have:

$$\frac{d}{d\beta}\pi(\mathbf{C}(\beta)\mathbf{D}) = \sum_{i=1}^{n}p_i(\beta)\frac{[(\mathbf{L} - \mathbf{I})\mathbf{D}\mathbf{x}(\beta)]_i}{x_i(\beta)},$$

since $\frac{d}{d\beta}\mathbf{B}(\beta) = \frac{d}{d\beta}\mathbf{C}(\beta)\,\mathbf{D} = (\mathbf{L} - \mathbf{I})\mathbf{D}$.

Following Karlin's proof in Karlin (1982), we note that for $\beta > 0$,

$$\mathbf{L} - \mathbf{I} = \frac{1}{\beta}\left([(1 - \beta)\mathbf{I} + \beta\mathbf{L}] - \mathbf{I}\right) = \frac{1}{\beta}[\mathbf{C}(\beta) - \mathbf{I}].$$

Hence

$$\frac{d\pi(\mathbf{C}(\beta)\mathbf{D})}{d\beta} = \frac{1}{\beta}\sum_{i=1}^{n}p_i(\beta)\frac{[(\mathbf{C}(\beta) - \mathbf{I})\mathbf{D}\mathbf{x}(\beta)]_i}{x_i(\beta)} = \frac{1}{\beta}\left[\pi(\mathbf{C}(\beta)\mathbf{D}) - \sum_{i=1}^{n}p_i(\beta)\frac{[\mathbf{D}]_{ii}x_i(\beta)}{x_i(\beta)}\right]$$

$$= \frac{1}{\beta}\left[\pi(\mathbf{C}(\beta)\mathbf{D}) - \sum_{i=1}^{n}p_i(\beta)[\mathbf{D}]_{ii}\right]. \tag{22}$$

As $\mathbf{x}(\beta)$ is unique and produces the infimum, for $\mathbf{0} < \mathbf{x} \neq \mathbf{x}(\beta)$:

$$\sum_{i=1}^{n} p_i(\beta) \frac{[\mathbf{C}(\beta)\mathbf{D}\mathbf{x}(\beta)]_i}{x_i(\beta)} < \sum_{i=1}^{n} p_i(\beta) \frac{[\mathbf{C}(\beta)\mathbf{D}\mathbf{x}]_i}{x_i}.$$

Let $\mathbf{y} > \mathbf{0}$ be the right Perron eigenvector of $\mathbf{C}(\beta)$. Since $\mathbf{e}^{\top}\mathbf{L} = \mathbf{e}^{\top}\mathbf{C}(\beta) = \mathbf{e}^{\top}$, we know that $\pi(\mathbf{C}(\beta)) = 1$, hence $\mathbf{C}(\beta)\,\mathbf{y} = \mathbf{y}$. Now, set $\mathbf{x} = \mathbf{D}^{-1}\mathbf{y}$, where $\mathbf{y}$ is scaled so that $\mathbf{e}^{\top}\mathbf{x} = 1$. Assuming $\mathbf{D}^{-1}\mathbf{y} \neq \mathbf{x}(\beta)$, we get:

$$\pi(\mathbf{C}(\beta)\mathbf{D}) = \sum_{i=1}^{n} p_i(\beta) \frac{[\mathbf{C}(\beta)\mathbf{D}\mathbf{x}(\beta)]_i}{x_i(\beta)} \tag{23}$$

$$< \sum_{i=1}^{n} p_i(\beta) \frac{[\mathbf{C}(\beta)\mathbf{D}\mathbf{D}^{-1}\mathbf{y}]_i}{(\mathbf{D}^{-1}\mathbf{y})_i} = \sum_{i=1}^{n} p_i(\beta) \frac{y_i}{y_i/[\mathbf{D}]_{ii}} = \sum_{i=1}^{n} p_i(\beta)[\mathbf{D}]_{ii}$$

We verify that $\mathbf{D}^{-1}\mathbf{y} \neq \mathbf{x}(\beta)$: should $\mathbf{D}^{-1}\mathbf{y} = \mathbf{x}(\beta)$, then (23) is an equality, thence $\pi(\mathbf{C}(\beta)\mathbf{D}) = \sup_{\mathbf{p}\in\mathcal{P}_n} p_i[\mathbf{D}]_{ii} = \max_i[\mathbf{D}]_{ii}$. The supremum requires $p_i(\beta) = 0$ for all $\{i\colon\ [\mathbf{D}]_{ii} < \max_j[\mathbf{D}]_{jj}\}$, which is nonempty because $\mathbf{D} \neq d\,\mathbf{I}$ for any d. In contradiction, irreducible $\mathbf{L}$ implies $\mathbf{p} > \mathbf{0}$ (Friedland, 1981, eq. (3.5), Theorem 3.2, and Corollary 3.1).

Application to (22) gives:

$$\frac{d}{d\beta}\pi(\mathbf{C}(\beta)\mathbf{D}) = \frac{1}{\beta}\left(\pi(\mathbf{C}(\beta)\mathbf{D}) - \sum_{i=1}^{n} p_i(\beta)[\mathbf{D}]_{ii}\right) < 0. \qquad \square$$

Corollary 1 (Extension to Substochastic-ML Matrices). *Let $\mathbf{L}$ be an irreducible ML-matrix such that $\mathbf{e}^{\top}\mathbf{L} \leq \mathbf{e}^{\top}$ and $\mathbf{e}^{\top}\mathbf{L} \neq \mathbf{e}^{\top}$; hence, by extension from non-negative matrices, it will be called a substochastic ML-matrix. Consider the family of matrices $\mathbf{C}(\beta) = (1-\beta)\mathbf{I} + \beta\,\mathbf{L}$ with $\beta > 0$. Then for any diagonal matrix $\mathbf{D}$ with positive terms on the diagonal, the spectral abscissa (and Perron root) $\pi(\mathbf{C}(\beta)\mathbf{D})$ is strictly decreasing as β increases.*

Proof. The steps are identical to those in the proof of Theorem 2 except that $\mathbf{e}^{\top}\mathbf{L} \leq\neq \mathbf{e}^{\top}$ means $\mathbf{e}^{\top}\mathbf{C}(\beta) \leq\neq \mathbf{e}^{\top}$, hence $\pi(\mathbf{C}(\beta)) < 1$ (Seneta 1981, Corollary 3, p. 30; Corollary 3, p. 52 and Lemma 1 (2)). No assumption that $\mathbf{D}^{-1}\mathbf{y} \neq \mathbf{x}(\beta)$ is needed, and (23) becomes:

$$\pi(\mathbf{C}(\beta)\mathbf{D}) \quad \leq \quad \sum_{i=1}^{n} p_i(\beta) \frac{\pi(\mathbf{C}(\beta))\,y_i}{(\mathbf{D}^{-1}\mathbf{y})_i} = \pi(\mathbf{C}(\beta))\sum_{i=1}^{n} p_i(\beta)[\mathbf{D}]_{ii} < \sum_{i=1}^{n} p_i(\beta)[\mathbf{D}]_{ii}.$$

Hence from (22), $\frac{d}{d\beta}\pi(\mathbf{C}(\beta)\mathbf{D}) < 0$. No assumption $\mathbf{D} \neq d\,\mathbf{I}$ is used. $\qquad \square$

5.4   Main Result

Theorem 3 (The Reduction Principle for Linear Variation). *Suppose a new allele at a modifier locus is introduced into a population fixed at the modifier locus near an*

equilibrium with a positive selection potential among the selected haplotypes, and $\hat{\alpha} < 1$. The new modifier allele produces linear variation in the transmission of the selected loci, and is linked to the nearest selected locus with recombination rate r.

If $\mathbf{M}(\alpha, r)$ is irreducible, the new modifier allele will increase (decrease) in frequency at a geometric rate if it brings the transmission closer to (further away from) perfect transmission, i.e., $\alpha < \hat{\alpha}$ ($\alpha > \hat{\alpha}$).

If $\mathbf{M}(\alpha, r)$ is reducible, then:

1. *the new modifier allele will increase in frequency at a geometric rate if $\alpha < \hat{\alpha}$, provided it occurs in at least one haplotype whose marginal fitness differs from the mean fitness of the population; and if $\alpha > \hat{\alpha}$, it will either:*

2. *decrease in frequency at a geometric rate; or*

3. *not increase in the case where it occurs in a haplotype whose block in the Frobenius normal form of $\mathbf{M}(\alpha, r)$ is an isolated block with all marginal fitnesses equal.*

Proof. For $\alpha = 0$, the theorem reduces to Result 2 in Altenberg and Feldman (1987). Henceforth only $\alpha > 0$ is considered. Lemma 2 shows that $\mathcal{M}^\star$ is an ML-matrix for all $\alpha, \hat{\alpha} \in (0, 1)$, and $r \in [0, 1]$. We substitute $\mathbf{C}(\beta) = (1-\beta)\mathbf{I} + \beta\mathcal{M}^\star$ and $\beta = \alpha/\hat{\alpha}$, thus $\mathbf{C}(\alpha/\hat{\alpha}) = \mathbf{M}(\alpha, r)$.

Case 1: Irreducible $\mathbf{M}(\alpha, r)\mathbf{D}$.

If $\mathbf{M}(\alpha, r)$ is irreducible and $\mathbf{D}$ positive on the diagonal, $\mathbf{M}(\alpha, r)\mathbf{D}$ is irreducible, and Theorem 2 applies. Hence for $\beta > 0$, $\frac{d}{d\beta}\pi(\mathbf{C}(\beta)\mathbf{D}) < 0$. Therefore,

$$\pi(\mathbf{C}(\alpha/\hat{\alpha})\mathbf{D}) > \pi(\mathbf{C}(1)\mathbf{D}) = 1, \text{ for } \alpha < \hat{\alpha},$$

and

$$\pi(\mathbf{C}(\alpha/\hat{\alpha})\mathbf{D}) < \pi(\mathbf{C}(1)\mathbf{D}) = 1, \text{ for } \alpha > \hat{\alpha}.$$

Since $\mathbf{M}(\alpha, r)$ is always non-negative, the Perron root is in fact the spectral radius:

$$\pi(\mathbf{C}(\alpha/\hat{\alpha})\mathbf{D}) = \pi(\mathbf{M}(\alpha, r)\mathbf{D}) = \rho(\mathbf{M}(\alpha, r)\mathbf{D}).$$

For $\hat{\alpha} = 1 > \alpha$, $\mathcal{M}^\star$ is unsolvable, but as a limit for $\hat{\alpha} < 1$, $\rho(\mathbf{M}(\alpha, r)\mathbf{D}) \geq 1$ is assured.

Case 2: Reducible $\mathrm{M}(\alpha, r)\mathbf{D}$.

If $\mathbf{M}(\alpha, r)$ is reducible or $\mathbf{D}$ is not strictly positive on the diagonal, $\mathbf{M}(\alpha, r)\mathbf{D}$ is reducible, and Theorem 2 must be extended. To analyze the dependence of the spectral abscissa on α in the case of reducible $\mathbf{C}(\beta)$ or where some $[\mathbf{D}]_{ii} = 0$, we utilize the Frobenius normal form of $\mathbf{C}(\beta)\mathbf{D}$. The Frobenius normal form permutes the indices of $\mathbf{C}(\beta)\mathbf{D}$ so that along the diagonal are irreducible square block matrices $\mathbf{C}_h(\beta)\mathbf{D}_h$, where h indexes the diagonal blocks.

The Frobenius normal form, $\mathbf{A}$, of a reducible matrix has the structure (Gantmacher, 1959, p. 75):

$$\mathbf{A} = \left(\begin{array}{cccc|cccc} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} & & & & \\ \mathbf{0} & \mathbf{A}_2 & \ddots & \vdots & & & \mathbf{0} & \\ \vdots & & \ddots & \mathbf{0} & & & & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_t & & & & \\ \hline \mathbf{A}_{t+1,1} & \mathbf{A}_{t+1,2} & \cdots & \mathbf{A}_{t+1,t} & \mathbf{A}_{t+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \cdots & \cdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{A}_{t+s,1} & \mathbf{A}_{t+s,2} & \cdots & \mathbf{A}_{t+s,t} & \mathbf{A}_{t+s,t+1} & \cdots & \cdots & \mathbf{A}_{t+s} \end{array} \right) \tag{24}$$

Blocks $\mathbf{A}_1, \ldots, \mathbf{A}_t$ are referred to as *isolated* blocks, while $\mathbf{A}_{t+1}, \ldots, \mathbf{A}_{t+s}$ are referred to as *non-isolated* blocks. Isolation is defined in terms of the rows, so in a matrix of the form

$$\left(\begin{array}{cc} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_3 & \mathbf{A}_2 \end{array} \right)$$

the block $\mathbf{A}_1$ is defined as isolated, whereas $\mathbf{A}_2$ is non-isolated if $\mathbf{A}_3 \neq \mathbf{0}$. Thus, by definition, block $\mathbf{A}_1$ in Frobenius normal form is always isolated.

At this point we require the following lemma:

Lemma 5. *The eigenvalues of $\mathbf{A}$ are the eigenvalues of the irreducible diagonal block matrices in the Frobenius normal form of $\mathbf{A}$.*

Proof. Let $h \in \mathcal{I}$ refer to isolated blocks, and $h \in \mathcal{N}$ refer to non-isolated blocks. The non-isolated blocks are contained in a principal submatrix, $\mathbf{A}_{\mathcal{N}}$, of $\mathbf{A}$. In (24) this would be

$$\mathbf{A}_{\mathcal{N}} = \left(\begin{array}{cccc} \mathbf{A}_{t+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{A}_{t+s,t+1} & \cdots & \cdots & \mathbf{A}_{t+s} \end{array} \right). \tag{25}$$

The eigenvalues of $\mathbf{A}$ must be eigenvalues either of an isolated block $\mathbf{A}_h$, $h \in \mathcal{I}$, or of the submatrix of non-isolated blocks, $\mathbf{A}_{\mathcal{N}}$. This is readily seen, because, letting $\mathbf{x}$ be an eigenvector of $\mathbf{A}$, if $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$, then for isolated block h, $\mathbf{A}_h \mathbf{x}_h = \lambda \mathbf{x}_h$, where $\mathbf{x}_h$ refers to the subvector of $\mathbf{x}$ with indices in block h. Hence either

1. λ is an eigenvalue of $\mathbf{A}_h$; or

2. $\mathbf{x}_h = \mathbf{0}$. If $\mathbf{x}_h = \mathbf{0}$ for all $h \in \mathcal{I}$, then $\mathbf{A}_{\mathcal{N}} \mathbf{x}_{\mathcal{N}} = \lambda \mathbf{x}_{\mathcal{N}}$, hence λ is an eigenvalue of $\mathbf{A}_{\mathcal{N}}$.

In submatrix $\mathbf{A}_{\mathcal{N}}$, we see from (25) that $\mathbf{A}_{t+1}$ is an isolated block. Hence, repeating the argument above, the eigenvalues of $\mathbf{A}_{\mathcal{N}}$ must be eigenvalues of either $\mathbf{A}_{t+1}$ or of the principal submatrix containing blocks $\mathbf{A}_{t+2}$ through $\mathbf{A}_{t+s}$. Continuing recursively in this manner, the lemma is proved. $\qquad\square$

Thus the eigenvalues of $\mathbf{C}(\beta)\mathbf{D}$ — including the spectral abscissa — must be eigenvalues of the diagonal blocks $\mathbf{C}_h(\beta)\mathbf{D}_h$. Therefore, the spectral abscissa for $\mathbf{C}(\beta)\mathbf{D}$ is the maximum of the spectral abscissas:

$$\pi(\mathbf{C}(\beta)\mathbf{D}) = \max_h \pi(\mathbf{C}_h(\beta)\mathbf{D}_h).$$

The spectral abscissas of each block are either constant, or strictly decreasing:

1. constant if $\mathbf{D}_h = d_h \, \mathbf{I}_h$ and $\mathbf{e}_h^\top \mathbf{C}_h(\beta) = \mathbf{e}_h^\top$;

2. strictly decreasing if either

    (a) $\mathbf{D}_h \neq d \, \mathbf{I}_h$ (by Theorem 2), or

    (b) $\mathbf{C}_h(\beta)$ is substochastic-ML, i.e. $\mathbf{e}_h^\top \mathbf{C}_h(\beta) \leq \mathbf{e}_h^\top, \neq \mathbf{e}_h^\top$ (by Corollary 1).

In the case where all blocks are isolated, it is not possible for all blocks to be constant, as this would entail that $\mathbf{D} = \mathbf{I}$, which is contrary to hypothesis (7). In the case where there is at least one non-isolated block, then (referring to the indices in (24)) there is at least one non-zero matrix $\mathbf{C}_{t+k,j}$, for some $t \in \{1, \ldots, s\}$, and $j \in \{1, \ldots, t\}$. This entails that

$$\mathbf{e}_{t+k}^\top \mathbf{C}_{t+k,j}(\beta) + \mathbf{e}_j^\top \mathbf{C}_j(\beta) \leq \mathbf{e}_j^\top, \neq \mathbf{e}_j^\top$$

hence $\mathbf{e}_j^\top \mathbf{C}_j(\beta) \leq \mathbf{e}_j^\top, \neq \mathbf{e}_j^\top$, thus $\mathbf{C}_j(\beta)$ is a substochastic ML-matrix. Hence if $\mathcal{N}$ is nonempty, at least one isolated block will be an irreducible substochastic ML-matrix, which by Corollary 1 is a decreasing block. Therefore, in all cases, there will be at least one isolated block h for which $\pi(\mathbf{C}_h(\beta)\mathbf{D}_h)$ is strictly decreasing in β.

We have as a reference point (16), where for $\beta = 1$, $\mathbf{C}(1)\mathbf{D}\hat{\mathbf{z}}_b = \hat{\mathbf{z}}_b > \mathbf{0}$. Hence, by Lemma 1 (3), $\pi(\mathbf{C}(1)\mathbf{D}) = 1$. We know by Theorem 6 in (Gantmacher, 1959, pp. 77–78), extended to ML-matrices by Lemma 1 (3), that $\pi(\mathbf{C}_h(1)\mathbf{D}_h) = 1$ for all $h \in \mathcal{I}$, and $\pi(\mathbf{C}_h(1)\mathbf{D}_h) < 1$ for all $h \in \mathcal{N}$.

For the parameter range $0 < \beta = \alpha/\hat{\alpha} < 1$, i.e. $\alpha < \hat{\alpha}$, at least one isolated block has spectral abscissa strictly decreasing in β, thus $\pi(\mathbf{C}_h(\beta)\mathbf{D}_h) > 1$ for $\beta < 1$. Hence,

$$\pi(\mathbf{C}(\alpha/\hat{\alpha})\mathbf{D}) = \rho(\mathbf{M}(\alpha, r)\mathbf{D}) = \max_h \pi(\mathbf{C}_h(\alpha/\hat{\alpha})\mathbf{D}_h) > 1, \text{ for } \alpha < \hat{\alpha}. \quad (26)$$

As long as the new modifier allele a is introduced into haplotypes that are part of the isolated block for which $\pi(\mathbf{C}_h(\alpha/\hat{\alpha})\mathbf{D}_h) > 1$, i.e $\epsilon_{a_h} \neq \mathbf{0}$, then ϵ_a will grow at a geometric rate. This proves conclusion (1) in the theorem.

For the parameter range $\beta = \alpha/\hat{\alpha} > 1$, i.e. $\alpha > \hat{\alpha}$, if there are no isolated blocks for which $\mathbf{D}_h = d_h \, \mathbf{I}_h$, then $\pi(\mathbf{C}_h(\beta)\mathbf{D}_h) < 1$, for all $h \in \mathcal{I}$ when $\beta < 1$. And recalling that $\pi(\mathbf{C}_h(1)\mathbf{D}_h) < 1$ for all $h \in \mathcal{N}$, then $\pi(\mathbf{C}_h(\beta)\mathbf{D}_h) < 1$ for all $h \in \mathcal{N}$ when $\beta > 1$ since all blocks are either constant or decreasing. Therefore,

$$\pi(\mathbf{C}(\alpha/\hat{\alpha})\mathbf{D}) = \rho(\mathbf{M}(\alpha, r)\mathbf{D}) = \max_h \pi(\mathbf{C}_h(\beta)\mathbf{D}_h) < 1, \text{ for } \beta > 1,$$

yielding conclusion (2) in the theorem. If there is a constant isolated block, which requires $\mathbf{D}_h = d_h \, \mathbf{I}_h$, then $\pi(\mathbf{C}(\beta)\mathbf{D}) = \rho(\mathbf{M}(\alpha,r)\mathbf{D}) = 1$ for $\beta > 1$. If the new modifier allele is introduced associated with haplotypes within the constant block, then conclusion (3) in the theorem holds. $\qquad\square$

Conditions for $\mathbf{M}(\alpha,r)$ to be reducible are limited but important. For $r = 0$, $\mathbf{S}$ must be reducible. This is the case in general when recombination is the only transformation acting on the selected haplotypes, and the Frobenius normal form of $\mathbf{S}$ consists of isolated irreducible blocks. If no block has all marginal fitnesses the same, then conclusion (2) applies. For $\mathbf{M}(\alpha,r)$ to be reducible when $r > 0$ and $\alpha < 1$, $\mathbf{Q} + \mathbf{S} + \mathbf{S}^{\bowtie}$ must be reducible. This is precluded if all fitnesses w_{ij} are positive, making $\mathbf{Q} > \mathbf{0}$. For $\mathbf{Q} + \mathbf{S} + \mathbf{S}^{\bowtie}$ to be reducible requires either that some haplotypes be lethal, or that $\mathbf{S} + \mathbf{S}^{\bowtie}$ be reducible and that specific zeros in $\mathbf{S} + \mathbf{S}^{\bowtie}$ be matched by $w_{ij} = 0$, a highly non-generic selection regime.

6 Discussion

6.1 Mathematical Issues

Theorem 3 establishes the full generality of the reduction principle for modifier genes that produce linear variation in genetic transmission, assuring that no exceptions to the evolutionary reduction of genetic transformation rates can come about due to selection regime, allele or locus multiplicity, different genetic processes, or recombination with the modifier locus, for modifiers that produce linear variation in transmission.

In previous studies, analytical tractability required tradeoffs between various simplifying assumptions: Liberman and Feldman (1986a,b); Feldman and Liberman (1986); Liberman and Feldman (1989) allowed no more than two alleles at each selected locus, or two demes in the case of migration modification, in order to obtain results for all $r > 0$. Altenberg (1984); Altenberg and Feldman (1987) required $r = 0$ or $\alpha = 0$ in order to obtain results for arbitrary allele and locus number, genetic processes, and selection regimes. These tradeoffs are now removed.

In the process of proving Theorem 3, I have had to extend Karlin's Theorem 5.2 to ML-matrices, substochastic matrices, and reducible matrices, which may have application outside the current context of modifier gene theory. Moreover, the treatment of reducible matrices here produces much stronger and more detailed results than in Altenberg (1984); Altenberg and Feldman (1987), where the spectral radius $\rho(\mathbf{M}(\alpha,0)\mathbf{D})$ was only proven to be non-increasing in α for reducible $\mathbf{M}(\alpha,0)$. Here, it is proven that $\rho(\mathbf{M}(\alpha,0)\mathbf{D})$ is strictly decreasing in α on $\alpha \in (0,\hat{\alpha})$, and strictly decreasing in α on $\alpha \in (\hat{\alpha},0)$ except when the marginal fitnesses are all equal within an isolated block in the Frobenius normal form of $\mathbf{M}(\alpha,0)\mathbf{D}$, in which case $\rho(\mathbf{M}(\alpha,0)\mathbf{D}) = 1$.

It should be noted that the method of proof utilized here does not allow us to say that $\frac{d}{d\alpha}\rho(\mathbf{M}(\alpha,r)\mathbf{D}) < 0$ for $r > 0$. That would be a much more direct means of proof for Theorem 3 and would obviate the need to extend Karlin's theorem to ML-matrices. A proof, however, runs up against impediments at step (23) that have not been surmounted. Such a result would be needed in order to say, as was done in Result

2 in Altenberg and Feldman (1987), that the asymptotic strength of selection either for or against a new modifier allele increases with $|\alpha - \hat{\alpha}|$. Theorem 3 alone cannot rule out the possibility that the function $\rho(\mathbf{M}(\alpha, r)\mathbf{D})$ meanders non-monotonically over $(0, \hat{\alpha})$ or $(\hat{\alpha}, 1)$, only that it never crosses the $\rho = 1$ line on those intervals. This perverse possibility is not consistent, however, with the behavior of the spectral radius for other examples of simple variation in matrix elements, where it is convex for variation in diagonal elements (Cohen, 1978, 1979; Friedland, 1981), and concave for variation in off-diagonal elements (Deutsch and Neumann, 1984, 1985). Out of these considerations, I conjecture that for any $r \in (0, 1/2)$, $\frac{d}{d\alpha}\rho(\mathbf{M}(\alpha, r)\mathbf{D}) < 0$.

6.2 Linear Variation

Linear variation is in a sense 'impartial' in that all transitions — from advantageous to deleterious haplotypes, and vice versa — are scaled exactly the same. So the production of both advantageous as well as deleterious haplotypes is reduced by a modifier that reduces transformation rates. But Theorem 3 shows that a reducing modifier allele always generates its own linkage disequilibrium with advantageous haplotypes and thus invades the population, while a modifier allele that increases transformation rates generates linkage disequilibrium with disadvantageous haplotypes and thus goes extinct. This is the essence of the reduction principle.

Given this generality of the reduction principle for linear variation in transmission, we must ask why natural genetic systems in fact depart from the reduction principle and do not evolve to perfect transmission. The reduction principle guides us to the list of assumptions that must be violated in order for departures from reduction to occur.

Removal of the near-equilibrium assumption of the model allows several classes of departure from reduction. Systems may be kept far from equilibrium by fluctuating selection, finite population size that produces drift, and *de novo* mutations. Recombination has been found to evolve under a special form of cyclic selection (Charlesworth, 1976; Hamilton, 1980), and under newly changed directional selection (Maynard Smith, 1988; Charlesworth, 1993; Barton, 1995), and in finite populations (Felsenstein, 1974; Felsenstein and Yokoyama, 1976). Removal of the assumption of Mendelian segregation of the modifier locus, e.g. meiotic drive, allows the evolution of recombination (Feldman and Otto, 1991).

Removal of the assumption of linear variation has produced many examples of departure from reduction. Mechanistically, linear variation requires that the modifier controls the 'hit' rate of a transforming process acting on the selected haplotypes, and that this process be the only transforming process, otherwise transmission probabilities between different genotypes will not be scaled equally. Furthermore, for α to enter linearly in the recursion, the transformation process cannot allow multiple hits.

When spelled out thus, we see that the mechanistic requirements for linear variation are not realized biologically. Multiple hits are the norm in chiasma formation and in point mutation. Multiple transformation processes are also the norm, as recombination, mutation, gene conversion, deletions, duplications, transpositions, etc. all happen in gamete formation.

The case of multiple hits has been explored in a model of recombination modification for a multi-locus system (Zhivotovsky et al., 1994). There, a refined reduction

principle again holds.

The case of multiple simultaneous transformation processes acting during transmission has received a good deal of theoretical attention, and it is here that numerous departures from the reduction principle have been found. A detailed review can be found in Feldman et al. (1997). The cases exemplify the "principle of partial control" (Altenberg, 1984, p. 149): when a modifier gene has only partial control over the transformations occurring at selected loci, then it may be possible for this part of the transformations to evolve an increase in rates.

The problem of characterizing exactly the conditions on variation in transmission that will evolve under modifier gene control remains an open question. Its solution will require a deeper understanding of the differential properties of the spectra of non-negative matrices.

7   Acknowledgments

References

Altenberg, L., 1984. A Generalization of Theory on the Evolution of Modifier Genes. Ph.D. thesis, Stanford University. Searchable online and available from University Microfilms, Ann Arbor, MI.

Altenberg, L. and Feldman, M. W. 1987. Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle. Genetics 117:559–572.

Balkau, B. and Feldman, M. W. 1973. Selection for migration modification. Genetics 74:171–174.

Barton, N. H. 1995. A general model for the evolution of recombination. Genetical Research (Cambridge) 65:123144.

Cavalli-Sforza, L. and Feldman, M. W. 1973. Models for cultural inheritance. I. Group mean and within group variation. Theoretical Population Biology 4:42–55.

Charlesworth, B. 1976. Recombination modification in a fluctuating environment. Genetics 83:181195.

Charlesworth, B. 1993. Directional selection and the evolution of sex and recombination. Genetical Research (Cambridge) 61:205224.

Cohen, J. E. 1978. Derivatives of the spectral radius as a function of non -negative matrix elements. Math. Proc. Camb. Phil. Soc. 83:183–190.

Cohen, J. E. 1979. Random evolutions and the spectral radius of a non-negative matrix. Math. Proc. Camb. Phil. Soc. 86:345–350.

Cohen, J. E. 1981. Convexity of the dominant eigenvalue of an essentially nonnegative matrix. Proceedings of the American Mathematical Society 81:657–658.

Deutsch, E. and Neumann, M. 1984. Derivatives of the perron root at an essentially nonnegative matrix and the group inverse of an m–matrix. Journal of Anal. Appl. 102:1–29.

Deutsch, E. and Neumann, M. 1985. On the first and second order derivatives of the perron vector. Lin. Alg. Appl. 71:57–76.

Feldman, M., Christiansen, F., and Brooks, L. D. 1980. Evolution of recombination in a constant environment. Proc Natl Acad Sci USA 77:4838–4841.

Feldman, M. and Liberman, U. 1986. An evolutionary reduction principle for genetic modifiers. Proc Natl Acad Sci USA 83:4824–4827.

Feldman, M. W. 1972. Selection for linkage modification: I. random mating populations. Theoretical Population Biology 3:324–346.

Feldman, M. W. and Balkau, B. 1973. Selection for linkage modification II. a recombination balance for neutral modifiers. Genetics 74:713–726.

Feldman, M. W. and Krakauer, J., 1976. Genetic modification and modifier polymorphisms. Pages 547–583 *in* S. Karlin and E. Nevo, eds. Population Genetics and Ecology. Academic Press, New York.

Feldman, M. W. and Otto, S. P. 1991. A comparative approach to the population genetic theory of segregation distortion. American Naturalist 137:443456.

Feldman, M. W., Otto, S. P., and Christiansen, F. B. 1997. Population genetic perspectives on the evolution of recombination. Annual Review of Genetics 20:261–295.

Felsenstein, J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.

Felsenstein, J. and Yokoyama, S. 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. Genetics 83:845–859.

Fisher, R. A. 1922. On the dominance ratio. Proceedings of the Royal Society of Edinburgh 42:321–341.

Fisher, R. A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Friedland, S. 1981. Convex spectral functions. Linear and Multilinear Algebra 9:299–316.

Friedland, S. and Karlin, S. 1975. Some inequalities for the spectral radius of non-negative matrices and applications. Duke Mathematical Journal 42:459–490.

Gantmacher, F. R., 1959. The Theory of Matrices, volume 2. Chelsea Publishing Company, New York.

Haldane, J. B. S. 1924. A mathematical theory of natural and artificial selection. Part I. Transactions of the Cambridge Philosophical Society 23:19–41.

Hamilton, W. D. 1980. Sex versus non-sex versus parasite. Oikos 35:282290.

Karlin, S., 1982. Classification of selection-migration structures and conditions for a protected polymorphism. Pages 61–204 *in* M. K. Hecht, B. Wallace, and G. T. Prance, eds. Evolutionary Biology, volume 14. Plenum Publishing Corporation.

Karlin, S. and McGregor, J. 1972. The evolutionary development of modifier genes. Proc Natl Acad Sci USA 69:3611–3614.

Karlin, S. and McGregor, J. 1974. Towards a theory of the evolution of modifier genes. Theoretical Population Biology 5:59–103.

Kimura, M. 1956. A model of a genetic system which leads to closer linkage by natural selection. Evolution 10:278–287.

Liberman, U. and Feldman, M. W. 1986a. Modifiers of mutation rate: A general reduction principle. Theoretical Population Biology 30:125–142.

Liberman, U. and Feldman, M. W. 1986b. A general reduction principle for genetic modifiers of recombination. Theoretical Population Biology 30:341–371.

Liberman, U. and Feldman, M. W., 1989. The reduction principle for genetic modifiers of the migration rate. Pages 111–137 *in* M. W. Feldman, ed. Mathematical Evolutionary Theory. Princeton University Press.

Maynard Smith, J. 1988. Selection for recombination in a polygenic modelthe mechanism. Genetical Research (Cambridge) 51:5963.

Nei, M. 1967. Modification of linkage intensity by natural selection. Genetics 57:625–641.

Odling-Smee, J. 2007. Niche inheritance: A possible basis for classifying multiple inheritance systems in evolution. Biological Theory 2:276–289.

Schauber, E., Goodwin, B., Jones, C., and Ostfeld, R. 2007. Spatial selection and inheritance: applying evolutionary concepts to population dynamics in heterogeneous space. Ecology 88:1112–1118.

Seneta, E., 1981. Non-negative Matrices and Markov Chains. Springer, New York.

Wright, S. 1931. Evolution in mendelian populations. Genetics 16:97–159.

Zhivotovsky, L. A., Feldman, M. W., and Christiansen, F. B. 1994. Evolution of recombination among multiple selected loci: A generalized reduction principle. Proc Natl Acad Sci USA 91:1079–1083.