J. Math. Biol. (1997) 35: 391–408

Journal of
Mathematical
Biology

Multilocus dynamics under haploid selection

Valery Kirzhner[1], **Yuri Lyubich**[2]

[1] Institute of Evolution, The University of Haifa, Mount Carmel, Haifa, 31905, Israel
[2] Department of Mathematics, Technion, Haifa 32000, Israel

Received 27 November 1995; received in revised form 17 January 1996

Abstract. A general haploid selection model with arbitrary number of multiallelic loci and arbitrary linkage distribution is considered. The population is supposed to be panmictic. A dynamically *equivalent* diploid selection model is introduced. There is a *position effect* in this model if the original haploid selection is not multiplicative. If haploid selection is additive then the fundamental theorem is established even with an estimate for the change in the mean fitness. On this basis exponential convergence to an equilibrium is proved. As rule, the limit states are single-gamete ones. If, moreover, linkage is *tight*, then the single-gamete state with *maximal* fitness attracts the population for almost all initial states.

Key words: Haploid selection – Position effect – Additive selection – Fundamental theorem – Convergence to equilibrium – Tight linkage

1 Introduction

There are a few works devoted to dynamics of populations under haploid one- or two-locus selection (Moran 1958; Feldman 1971; Nagylaki 1992; Rutschman 1994). In the present paper we consider haploid n-locus selection in a panmictic population with any number n and any number of alleles at each locus. The linkage distribution is arbitrary. We use the concepts and notations introduced by Reiersöl 1962 in multilocus population dynamics. (For a presentation and development of Reiersöl's method with relevant references see Lyubich 1992.) An equation for haploid selection has a compact and convenient form (15). In this way one can construct a diploid selection model (see (18)) which has the dynamics on the gamete level same as the given haploid selection model. In this sense these models are equivalent. In particular, if the fundamental theorem of natural selection is valid for one of them, then it is automatically valid for other (see formula (19)). The fundamental theorem is actually valid if the haploid selection is additive. We derive this from explicit formula (34) for the mean fitness in next generation in terms of initial one. Moreover, we establish a quantitative formulation (see (53), (49)) of the fundamental theorem in the spirit of Lyubich et al. (1976) (see Lyubich 1992, Sect. 9.2 for details) and Nagylaki 1989. As a result, the population with any initial state exponentially tends to an equilibrium. We describe all equilibria; in particular, they are single-gamete if the selective contributions of alleles are pairwise distinctive.

In absence of crossing-overs the model becomes formally one-locus. Its evolution is well-known (see Nagylaki 1992). Since this dynamical system is hyperbolic in the above mentioned distinctive case, its small perturbations admit the same dynamics. Therefore in this case the population with tight linkage tends to the single-gamete state whose fitness is maximal except for a thin set of initial states.

If a haploid model is multiplicative then the equivalent diploid model is so. However, it is not true for additive haploid models except for some exotic ones. Moveover, every nonmultiplicative haploid selection model generates a diploid selection model with a position effect.

2 The basic equation of haploid selection

We consider a set $L = \{1, \ldots, n\}$ of autosomal loci. For every locus $i \in L$ we denote by Γ_i the set of alternative alleles,

$$\Gamma_i = \{a_{i1}, \ldots, a_{im_i}\}, \quad 1 \leqq i \leqq n , \tag{1}$$

so that m_i is the number of alleles at the locus i. We assume $m_i \geqq 2$. Then there are $N = m_1 \ldots m_l$ genotypes of gametes

$$g = a_{1j_1}a_{2j_2} \ldots a_{nj_n} = \prod_{1 \leqq i \leqq n} a_{ij_i} . \tag{2}$$

We denote by Γ the set of all gametes g. As usual, *a state* of the population is an arbitrary probability distribution on the set Γ, $\{p(g) \mid g \in \Gamma\}$, so

$$p(g) \geqq 0, \qquad \sum_g p(g) = 1 \tag{3}$$

(g runs over Γ from now on.) It is convenient to consider the states as formal linear combinations of gametes, namely,

$$G = \sum_g p(g)g . \tag{4}$$

According to (3) the states are the *convex* combinations of the *single-gamete states* g: $p(g) = 1, p(h) = 0$ for $h \neq g$. We say that a gamete g *is contained* in a state G if $p(g) \neq 0$.

We will also deal with arbitrary *real* linear combinations

$$X = \sum_g \xi(g)g . \tag{5}$$

They form a vector space of dimension N. There is a special linear form

$$s(X) = \sum_g \xi(g) \tag{6}$$

on this space. The states are just vectors X with $s(X) = 1$ and coordinates $\xi(g) \geqq 0$, so the set of all states is the basis simplex.

Haploid selection with fitnesses $\lambda(g) > 0$, $g \in \Gamma$, transforms any initial state G into the state

$$\hat{G} = \frac{\Lambda G}{w(G)} = \frac{\sum_g \lambda(g)\, p(g) g}{w(G)}\,, \tag{7}$$

where

$$w(G) = \sum_g \lambda(g)\, p(g) \tag{8}$$

is the mean fitness. Obviously,

$$w(G) = s(\Lambda G) > 0\,, \tag{9}$$

where Λ is a linear operator, namely,

$$\Lambda X = \sum_g \lambda(g) \xi(g) g \tag{10}$$

for X given by (5). Note that $w(G) \neq 1$ in general.

The next stage is the fertilization. We assume that the zygotes (g, h) form at random by independent association of the gametes g and h.

Finally, crossing-overs between gametes g and h in a zygote (g, h) occur with some probabilities $r(U|V)$, where $U|V$ runs over all partitions of the set L of loci into two classes U,V: $U \cup V = L$, $U \cap V = \emptyset$. Therefore

$$r(U|V) \geqq 0, \qquad \sum_{U|V} r(U|V) = 1\,. \tag{11}$$

This *linkage distribution* determines a recombination process.

If a partition $U|V$ occurs in meiosis then the zygote (g, h) produces the gametes $g_U h_V$ and $g_V h_U$ where g_U, g_V, etc. are *subgametes*, for instance

$$g_U = \prod_{i \in U} a_{ij_i}, \qquad g_V = \prod_{i \in V} a_{ij_i}\,. \tag{12}$$

(Note that $g_U g_V = g$.) Respectively, we set

$$X_U = \sum_g \xi(g) g_U\,. \tag{13}$$

If the recombination starts with a state G then in the next generation

$$G' = \sum_{U|V} r(U|V) G_U G_V \tag{14}$$

(see Reiersöl 1962; Lyubich 1992, Sect. 6.2.) However, in our model $\hat{G}$ plays the role of G. Therefore

$$G' = \sum_{U|V} r(U|V) \hat{G}_U \hat{G}_V$$

and, by substitution from (7)

$$G' = \frac{\sum_{U|V} r(U|V)\ (\Lambda G)_U (\Lambda G)_V}{w(G)^2}\ . \tag{15}$$

This is the *basic equation* of the evolution under haploid selection. The states G such that $G' = G$ are *equilibria.* In particular, all single-gamete states are equilibria. In general, some other equilibria are possible even internal ones (see, for example, Feldman 1971).

3 The equivalent diploid model

Let us consider a diploid selection model with the same set L of loci and the same linkage distribution $\{r(U|V)\}$. Let $\zeta(g, h)$ be the fitness of the zygote (g, h). (As usual, $\zeta(g, h) = \zeta(h, g)$ and $\zeta(g, g) > 0$.) The corresponding evolution equation on the gamete level is

$$G' = \frac{1}{W(G)} \sum_{U|V} r(U|V) \sum_{g,h} \zeta(g_U h_V, h_U g_V) p(g_U h_V) p(h_U g_V) g \tag{16}$$

where

$$W(G) = \sum_{g,h} \zeta(g, h) p(g) p(h) \tag{17}$$

is the mean fitness on the zygote level (see Lyubich 1992, Sect. 9.5). Below we show that equation (16) turns into equation (15) under the relation

$$\zeta(g, h) = \lambda(g)\lambda(h) \tag{18}$$

between zygote and gamete fitness. Thus, the evolution in our haploid model is the same as on the gamete level of the diploid model with fitness function (18). In this sense these models are *equivalent.*

At first, it follows from (18) that

$$W(G) = \left[\sum_g \lambda(g) p(g)\right]^2 = [w(G)]^2\ . \tag{19}$$

This is the denominator in (15). By the way, (19) shows that the fundamental theorem for a haploid model is equivalent to a similar theorem for the equivalent diploid model.

Now by (18) we transform the numerator in (16) into the expression

$$\sum_{U|V} r(U|V) \sum_{g,h} \lambda(g_U h_V) \lambda(h_U g_V) p(g_U h_V) p(h_U g_V) g\ . \tag{20}$$

In this sum we can make substitution

$$g_U h_V = \gamma, \qquad h_U g_V = \chi\ . \tag{21}$$

It is invertible and its inverse is the same,

$$g = \gamma_U \chi_V , \qquad h = \chi_U \gamma_V \; . \tag{22}$$

Therefore (20) takes the form

$$\sum_{U|V} r(U|V) \sum_{g,h} \lambda(g)\lambda(h)p(g)p(h)g_U h_V \; , \tag{23}$$

(after (21) we change the notation γ, χ for g, h). Now we obtain

$$\sum_{U|V} r(U|V) \sum_{g} \lambda(g)p(g)g_U \sum_{h} \lambda(h)p(h)h_V = \sum_{U|V} r(U|V)(\Lambda G)_U(\Lambda G)_V \; . \tag{24}$$

This is just the numerator in (15).

Let us discuss some important types of selection on both of levels, zygotes and gametes. The following definitions are usual.

Diploid selection with fitnesses $\zeta(g, h)$ is called *additive* if

$$\zeta(g, h) = \sum_{i=1}^{n} \zeta_i(g_i, h_i) \tag{25}$$

where g_i and h_i are genes at i-th locus in gametes g and h and ζ_i is a positive symmetric function on the set of all pairs of genes (a_{ij}, a_{ik}), $1 \leqq j \leqq m_i$, $1 \leqq k \leqq m_i$.

Similarly, haploid selection with fitnesses $\lambda(g)$ is called *additive* if

$$\lambda(g) = \sum_{i=1}^{n} \alpha_i(g_i) \tag{26}$$

where now α_i is a positive function on the set of all genes a_{ij}, $1 \leqq j \leqq m_i$.

The values $\zeta_i(g_i, h_i)$ and $\alpha_i(g_i)$ are the *contributions* of the pair (g_i, h_i) or the single g_i into the fitness $\zeta(g, h)$ or $\lambda(g)$.

If we change the sums in (25) and (26) into products then the above definitions turn into the definitions of *multiplicative* selection. So, now we have

$$\zeta(g, h) = \prod_{1 \leqq i \leqq n} \theta_i(g_i, h_i) \tag{27}$$

in diploid case and

$$\lambda(g) = \prod_{1 \leqq i \leqq n} \beta_i(g_i) \tag{28}$$

in haploid one.

It follows from (18) that for multiplicative haploid selection the equivalent diploid selection is also multiplicative with

$$\theta_i(g_i, h_i) = \beta_i(g_i)\beta_i(h_i) \; . \tag{29}$$

Diploid selection is called *recombination-invariant* (Lyubich 1992, Sect. 9.7) if

$$\zeta(g_U h_V, h_U g_V) = \zeta(g, h) \tag{30}$$

for all pairs of gametes g, h and all crossing-overs $U|V$. Any violation of this relation is a *position effect*.

Obviously, the additive and the multiplicative diploid selection models are both recombination-invariant, so there is no a position effect in these models.

In contrast, any haploid selection model except for the multiplicative one causes a position effect in the equivalent diploid model. Let us formulate this result as a theorem (see Appendix 1 for a proof).

Theorem 1. *If for a haploid selection model the equivalent diploid selection model is recombination-invariant, then this haploid selection is multiplicative.*

We see that no dynamical information can be transferred from recombination-invariant nonmultiplicative diploid selection models to haploid ones. Note that some standard diploid selection models are recombination-invariant (see, for example, Lewontin and Kojima 1960). The general equation (16) allows position effect but, certainly, this may be also applicable to any recombination-invariant model. However, in this case the nonmultiplicative diploid dynamics does not correspond any haploid selection. In particular, this concerns diploid additive models whose dynamics was investigated rather completely (Karlin and Feldman 1970; Karlin 1978; Kun and Lyubich 1980; Lyubich 1992, Sects. 9.5 and 9.6). Indeed, as rule, additive diploid selection cannot be of haploid origin except for an exotic case when the haploid selection is additive and multiplicative at the same time (see Appendix 1).

However, the dynamical picture of weak multilocus diploid selection (Nagylaki 1993) is compatible with a position effect. By (18) the dynamics under weak haploid selection,

$$\lambda(g) = 1 + O(\varepsilon), \qquad \varepsilon \to 0 , \tag{31}$$

is directly reduced to the diploid case.

Concluding this section we note that every diploid selection model of haploid origin satisfies the equation

$$\zeta(g, h) = \sqrt{\zeta(g, g)}\,\sqrt{\zeta(h, h)} , \tag{32}$$

since $\zeta(g, g) = \lambda(g)^2$ follows from (18) for $h = g$. Conversely, (32) implies (18) with $\lambda(g) = \sqrt{\zeta(g, g)}$.

As a consequence of (32), if diploid selection is of haploid origin then

$$\min_{g} \zeta(g, g) \leqq \min_{g \neq h} \zeta(g, h) \leqq \max_{g \neq h} \zeta(g, h) \leqq \max_{g} \zeta(g, g) . \tag{33}$$

In this sense *an advantage of heterozygotes is impossible under haploid selection.*

Note that the estimates (33) are strict if the homozygote fitnesses $\zeta(g, g)$ are pairwise distinct.

4 The fundamental theorem under additive haploid selection

Let haploid selection be additive. In this case the mean fitness changes to the next generation by the following relation

$$w(G') = \frac{s(\Lambda^2 G)}{w(G)} . \tag{34}$$

This is a consequence of the formula

$$s(\Lambda(X_U X_V) = s(\Lambda X)s(X) \tag{35}$$

which is proved in Appendix 2. In turn, to derive (34) from (35) we take $X = \Lambda G$ and note that (15) implies

$$\Lambda G' = \frac{\sum_{U|V} r(U|V)\Lambda(X_U X_V)}{w(G)^2}\,. \tag{36}$$

Since $w(G') = s(\Lambda G')$ by (9), we obtain from (36)

$$w(G') = \frac{\sum_{U|V} r(U|V)s(\Lambda(X_U X_V))}{w(G)^2}\,. \tag{37}$$

Applying (35) we get

$$w(G') = \frac{\sum_{U|V} r(U|V)s(\Lambda X)s(X)}{w(G)^2} = \frac{s(\Lambda X)s(X)}{w(G)^2}\,. \tag{38}$$

which yields (34) because of $s(X) = s(\Lambda G) = w(G)$ and $s(\Lambda X) = s(\Lambda^2 G)$.

By the way, (34) can be rewritten as

$$w(G') = w(\hat{G})\,, \tag{39}$$

where G is defined by (7). Equality (39) is true given additivity, so *there is no recombination effect on the mean fitness created on the selection stage.* This conclusion fails in general for nonadditive selection.

Formula (34) implies the following fundamental theorem of Fisher-Ewens type:

Theorem 2. *If haploid selection is additive, then*

$$w(G') \geqq w(G)\,, \tag{40}$$

for every state

$$G = \sum_g p(g)g\,. \tag{41}$$

Equality is attained if and only if all gametes contained in G are of the same fitness.

Proof. By the Cauchy inequality

$$\sum_g \lambda(g)p(g) \leqq \sqrt{\sum_g \lambda^2(g)p(g)}\,\sqrt{\sum_g p(g)} = \sqrt{\sum_g \lambda^2(g)p(g)}\,. \tag{42}$$

Thus,

$$s(\Lambda G) \leqq \sqrt{s(\Lambda^2 G)}\,. \tag{43}$$

Now (34) yields

$$w(G') \geqq \frac{s(\Lambda G)^2}{w(G)} = w(G)\,, \tag{44}$$

since $s(\Lambda G) = w(G)$ by (8). The theorem is proved.

It follows from a well-known fact regarding the Cauchy inequality that (42) is an equality if and only if the vectors $\{\lambda(g)\sqrt{p(q)}\}$ and $\{\sqrt{p(q)}\}$ are proportional. This means that $w(G') = w(G)$ if and only if $\lambda(g)$ is the same for all of g contained in the state G. So, we have

Corollary. *If a state G is an equilibrium then the fitnesses of all gametes contained in G are the same.*

Thus, *there is no polymorphic equilibrium containing all gametes except for the trivial case of constant fitness function.*

Let us say that two gametes g and h are *equally* fit if $\lambda(g) = \lambda(h)$. In this case we will write $g \equiv h$. As a result, we get a partition of the set Γ of all gametes into some classes $E_1, \dots, E_m$, $m \leqq n$. All gametes from the same class are equally fit but no two gametes from different classes are not so.

For any state G one can introduce the probabilities of the classes,

$$p(E_k) = \sum_{g \in E_k} p(g), \quad 1 \leqq k \leqq m \,. \tag{45}$$

Obviously,

$$\sum_{k=1}^{m} p(E_k) = 1 \,. \tag{46}$$

In these terms

$$w(G) = \sum_{k=1}^{m} \lambda_k p(E_k) \,. \tag{47}$$

where λ_k is the common fitness of all gametes g from E_k.

Now we can quantitatively sharpen our fundamental theorem.

Theorem 3. *Let*

$$\delta = \min_{j \neq k} |\lambda_j - \lambda_k| \,. \tag{48}$$

Then

$$\Delta w(G) \equiv w(G') - w(G) \geqq \frac{\delta^2}{2w(G)} \sum_{k=1}^{m} p(E_k) q(E_k) \tag{49}$$

where $q(E_k) = 1 - p(E_k)$, $1 \leqq k \leqq m$.

Proof. We start with the Lagrange identity

$$\sum_g a_g^2 \sum_h b_h^2 - \left(\sum_g a_g b_g\right)^2 = \frac{1}{2} \sum_{g \neq h} (a_g b_h - a_h b_g)^2 \tag{50}$$

where a_g and b_h are arbitrary numbers. Setting $a_g = \lambda(g)\sqrt{p(g)}$ and $b_h = \sqrt{p(h)}$ we obtain

$$\sum_g \lambda^2(g) p(g) - \left(\sum_g \lambda(g) p(g)\right)^2 = \frac{1}{2} \sum_{g \neq h} [\lambda(g) - \lambda(h)]^2 p(g) p(h) \tag{51}$$

or

$$s(\Lambda^2(G) - s(\Lambda G)^2 = \frac{1}{2} \sum_{g \neq h} [\lambda(g) - \lambda(h)]^2 p(g) p(h) \,. \tag{52}$$

It follows from (52), (9) and (34) that

$$\Delta w(G) = \frac{1}{2w(G)} \sum_{g \not\equiv h} [\lambda(g) - \lambda(h)]^2 p(g) p(h) \,. \tag{53}$$

This exact formula implies not only the inequality $w(G') \geqq w(G)$ but also the following estimate:

$$\Delta w(G) \geqq \frac{\delta^2}{2w(G)} \sum_{g \not\equiv h} p(g) p(h) \,. \tag{54}$$

The summation (54) can be done as

$$\begin{aligned} \sum_{g \not\equiv h} p(g)p(h) &= \sum_{g \in \Gamma} p(g) \sum_{h:\, h \not\equiv g} p(h) = \sum_{g \in \Gamma} p(g) \left[1 - \sum_{h:\, h \equiv g} p(h) \right] \\ &= \sum_{k=1}^{m} (1 - p(E_k)) \sum_{g \in E_k} p(g) = \sum_{k=1}^{m} p(E_k) q(E_k) \,. \end{aligned} \tag{55}$$

Now (49) directly follows from (54) and (55).

For our purposes it is useful to change (49) to the inequality

$$\Delta w(G) \geqq \frac{\delta^2}{w(G)} p(E_j) q(E_j), \quad 1 \leqq j \leqq m \,. \tag{56}$$

In order to do it, we transform the last summation in (55) in such a way:

$$\begin{aligned} \sum_{k=1}^{m} p(E_k) q(E_k) &= \sum_{k=1}^{m} p(E_k)(1 - p(E_k)) = 1 - \sum_{k=1}^{m} p^2(E_k) \\ &= 1 - p^2(E_j) - \sum_{k \neq j}^{m} p^2(E_k) \,. \end{aligned} \tag{57}$$

Therefore

$$\begin{aligned} \sum_{k=1}^{m} p(E_k) q(E_k) \geqq 1 - p^2(E_j) - \left(\sum_{k \neq j}^{m} p(E_k) \right)^2 &= 1 - p^2(E_j) - (1 - p(E_j))^2 \\ &= 2p(E_j) q(E_j) \,. \end{aligned} \tag{58}$$

5 The convergence to an equilibrium

Let G_0 be an initial state of the population and let

$$G_t = \sum_{g} p_t(g) g, \quad t = 0, 1, 2, \ldots \tag{59}$$

be the corresponding state in t-th generation. We investigate the asymptotic behavior of G_t as $t \to \infty$.

Theorem 4. *Under haploid additive selection, every trajectory* $\{G_t\}_0^\infty$ *tends to an equilibrium at an exponential rate.*

Proof. By compactness of the basis simplex, every subsequence of $\{G_t\}_0^\infty$ contains a convergent subsequence. Let Ω be the ω-limit set of the trajectory, i.e. Ω is the set of the limits of all convergent subsequences. This set is evolutionarily invariant, i.e. if $G \in \Omega$ then $G' \in \Omega$.

The fundamental theorem implies the existence of $w_\infty = \lim w(G_t)$ as $t \to \infty$. Therefore, $w(G_t) \to w_\infty$ for all $G \in \Omega$. Hence, $w(G_{t+1}) \to w_\infty$ as well, so $w(G') = w(G)$. In this case all gametes contained in G have the same fitnesses, say $\lambda(g) = \lambda_1$ hence,

$$G = \sum_{g \in E_1} p(g) g \,. \tag{60}$$

We see that $p(E_1) = 1$ for all $G \in \Omega$ hence, $p_t(E_t) \to 1$, thus, $p_t(E_i) \to 0$ for $i = 2, \ldots, m$.

To estimate the rate of this convergence, we turn to (56) for $G = G_t, j = 1$. Then

$$w(G_{t+1}) - w(G_t) \geqq \frac{\delta^2}{w(G_t)} p_t(E_1)\,[1 - p_t(E_1)] \,. \tag{61}$$

Further, it is convenient to pass to the deviation of the mean fitness from its limit value

$$\Delta_t = w_\infty - w(G_t) = \lambda_1 - w(G_t) \,. \tag{62}$$

Obviously, $\Delta_t \geqq 0$ and

$$\Delta_t = \lambda_1 - \sum_{k=1}^{m} \lambda_k p_t(E_k) \leqq \lambda_1(1 - p_t(E_1)) \,. \tag{63}$$

In these terms (61) and (63) imply

$$\Delta_t - \Delta_{t+1} \geqq \frac{\delta^2 p_t(E_1)}{\lambda_1 w(G_t)} \Delta_t \,, \tag{64}$$

or

$$\Delta_{t+1} \leqq \left[1 - \frac{\delta^2 p_t(E_1)}{\lambda_1 w(G_t)}\right] \Delta_t \,. \tag{65}$$

Since

$$\lim_{t \to \infty} \left[1 - \frac{\delta^2 p_t(E_1)}{\lambda_1 w(G_t)}\right] = 1 - \frac{\delta^2}{\lambda_1^2} < 1 \,, \tag{66}$$

(65) yields the exponential estimate

$$\Delta_t = O(\chi^t) \tag{67}$$

with any χ such that

$$1 - \frac{\delta^2}{\lambda_1^2} < \chi < 1 \,. \tag{68}$$

Thus, from (61)

$$1 - p_t(E_1) \leqq \frac{w(G_t)}{\delta^2 p_t(E_1)} \Delta_t = O(\chi^t) \,, \tag{69}$$

i.e. $p_t(E_1) \to 1$ at an exponential rate. (It is just the rate of elimination of the gametes whose fitnesses are different from the limit fitness.)

If evolution started from a state with equally fit gametes, then it would be reduced to a pure recombination process. Because this process is also exponentially convergent (see Lyubich 1992, Sect. 6.6), we should still have the same property in our case, when recombination and selection are both present. Let us specify this idea.

From the above analysis we get

$$G_t = p_t(E_1)H_t + O(\chi^t) \tag{70}$$

with

$$H_t = \frac{\sum_{g \in E_1} p_t(g)g}{p_t(E_1)} \,. \tag{71}$$

Obviously, $s(H_t) = 1$. Now

$$\Lambda G_t = p_t(E_1)\lambda_1 H_t + O(\chi^t) \tag{72}$$

and then

$$\Lambda G_t = \lambda_1 H_t + O(\chi^t), \qquad w(G_t) = s(\Lambda G_t) = \lambda_1 + O(\chi^t) \tag{73}$$

because of (69). The evolutionary equation (15) written as

$$G_{t+1} = \frac{\sum_{U|V} r(U|V)(\Lambda G_t)_U(\Lambda G_t)_V}{w(G_t)^2} \tag{74}$$

can be reduced to an asymptotic form,

$$H_{t+1} = \sum_{U|V} r(U|V)\,(H_t)_U(H_t)_V + O(\chi^t)\,. \tag{75}$$

It is an evolutionary equation of a recombination process up to an additional term $O(\chi^t)$.

In the case of the number of loci $n = 1$ we have

$$H_{t+1} = H_t + O(\chi^t)\,. \tag{76}$$

Thus, in this case H_t tends to a limit H_∞ and

$$H_{t+1} = H_\infty + O(\chi^t)\,. \tag{77}$$

In the general case $n > 1$ one can write

$$H_{t+1} = \rho H_t + \sum_{U|V \neq L|\emptyset} r(U|V)(H_t)_U(H_t)_V + O(\chi^t) \tag{78}$$

where $\rho = r(L|\emptyset)$ is the probability that no recombination occurs. One can assume that $\rho < 1$ because in the case $\rho = 1$ (78) coincides with (76). Moreover, by induction on n, one can assume that, with nonempty U and V, $(H_t)_U$ and $(H_t)_V$ exponentially tend to some limits. Indeed, their evolutionary equations are natural projections of (75) on subsystems of loci $U \subset L$ and $V \subset L$. These projections are of same form (75) with some induced linkage distributions (cf. Reiersöl 1962; Lyubich 1992, p. 248). Thus,

$$H_{t+1} = \rho H_t + F + O(\varepsilon^t)\,. \tag{79}$$

with some constant F and some ε, $\chi \leqq \varepsilon < 1$. This yields

$$H_t = \frac{F}{1-\rho} + O(\max(\varepsilon^t, \rho^t)) \,. \tag{80}$$

if $\varepsilon \neq \rho$. The last condition can always be provided due to a variation of ε. The theorem is proved.

What are the equilibria in the additive situation? We know that all gametes contained in such a state G are of the same fitness. Moreover,

$$G = \prod_{1 \leqq i \leqq n} H_i \tag{81}$$

where H_i are some single-locus states, $1 \leqq i \leqq n$. Let $a_{i1}, \ldots, a_{ik_i}$ be genes contained in the state H_i. Then G contains only the gametes $g = a_{1j_1}, \ldots, a_{nj_n}$ with $1 \leqq j_1 \leqq k_1, \ldots, 1 \leqq j_n \leqq k_n$. By additivity,

$$\lambda(g) = \sum_{i=1}^{n} \alpha(a_{ij_i}) \,. \tag{82}$$

This sum does not depend on the choice of the chain $j_1, \ldots, j_n$ if and only if $\alpha(a_{ij})$ does not depend on j, $j \leqq k_i$. Thus, a *state G is an equilibrium if and only if it is a product of single-locus states such that the contributions to the fitness function are the same for all present alleles.*

Suppose that there exists a pair of alleles whose contributions are the same. Then the variety of equilibria is infinite. However, this variety is finite if the contributions of all alleles are pairwise distinct. Under the last condition, we call the haploid additive selection *distinctive*. It is just the case in which the variety of equilibria is the set of single-gamete states. Note that this situation is compatible with the existence of gametes g, h with are equally fit. However, if $\lambda(g) \neq \lambda(h)$ for all pairs g, $h (g \neq h)$ then selection is distinctive.

Example. Consider a system two loci with two alleles at each one, say

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$$

If $\alpha(a_{11}) = \alpha(a_{22}) \neq \alpha(a_{12}) = \alpha(a_{21})$ then selection is distinctive but $\lambda(a_{11}a_{21}) = \lambda(a_{12}a_{22})$.

Now by Theorem 4 we get

Corollary. *Under haploid additive distinctive selection, every trajectory tends to a single-gamete state γ at an exponential rate.*

Our estimate of this rate is (69), where χ is an arbitrary real number satisfying (68). In other words,

$$\lim \sqrt[t]{1 - p_t(\gamma)} = 1 - \frac{\delta^2}{\lambda(\gamma)^2} \,, \tag{69'}$$

where δ is the minimal difference between the fitnesses of gametes (see (48)).

6 Tight linkage

This is the case when $r(L|\emptyset) = \rho$ is close to 1, i.e., the probability of at least one crossing-over,

$$\sum_{U|V \neq L|\emptyset} r(U|V) = 1 - \rho \,, \tag{83}$$

is small enough. Assume that *there are no equally fit gametes*: $\lambda(g) \neq \lambda(h)$ if $g \neq h$. Then we prove that *the population tends to the single-gamete state with maximal fitness for almost all of initial states.*

At first, we consider the extreme case $\rho = 1$, i.e. $r(U|V) = 0$ except for $U|V = L|\emptyset$. The evolutionary equation (15) is reduced to the following one:

$$G' = \frac{\Lambda G}{w(G)} \,. \tag{84}$$

This equation describes the evolution under haploid selection without crossing-overs. Actually, it is a 1-locus model with gametes in the role of genes. It is well-known that such a population tends to a single-gamete state whose fitness is maximal among gametes which are contained in the initial state (Nagylaki 1992). To pass to our general case, we investigate the population dynamics near equilibria.

By differentiation of (84)

$$d(G') = \frac{w(G)\Lambda dG - w(dG)\Lambda G}{w(G)^2} \tag{85}$$

because ΛG and $w(G) = s(\Lambda G)$ are linear. In particular, at a single-gamete state $G = h$ we have $\Lambda G = \lambda(h)h$, $w(G) = \lambda(h)$,

$$Y = \frac{\Lambda X - s(\Lambda X)h}{\lambda(h)} \tag{86}$$

where $X = dG$, $Y = d(G')$. Using the expansion

$$X = \sum_g \xi(g)g \tag{87}$$

we obtain

$$Y = \frac{\sum_{g \neq h} \lambda(g)\xi(g)(g - h)}{\lambda(h)} \,. \tag{88}$$

Note that

$$\sum_g \xi(g) = s(X) = 0 \tag{89}$$

because $s(G) = 1$ for all of states G. Therefore one can write (87) as

$$X = \sum_{g \neq h} \xi(g)(g - h) \,. \tag{90}$$

Comparing (88) with (90), we see that the eigenvalues of the differential (86) are $\lambda(g)/\lambda(h)$, $g \neq h$. By our assumption they are different from 1. Some of them are more than 1 if $\lambda(h)$ is not maximal, others are less than 1. On the other hand, the evolution under the equation (84) is invertible,

$$G = \frac{\Lambda^{-1}G'}{s(\Lambda^{-1}G')} \,. \tag{91}$$

Such a hyperbolic invertible structure is invariant under small variation of the parameter ρ. In this situation every state γ with $\lambda(\gamma) < \max\{\lambda(g) | g \neq \gamma\}$ only attracts the sequences $\{G_t\}$ whose initial states G_t lie on the stable variety at γ. Being convergent, these sequences have to be attracted by the single-gamete state with maximal fitness.

Appendix 1

Proof of Theorem 1. Let

$$\zeta(g, h) = \lambda(g)\lambda(h) \tag{A1.1}$$

and

$$\zeta(g_U h_V, h_U g_V) = \zeta(g, h) \tag{A1.2}$$

for all $U|V$. By the substitutions

$$\omega(g, h) = \ln[\zeta(g, h)], \qquad \alpha(g) = \ln[\lambda(g)] \tag{A1.3}$$

we get

$$\omega(g, h) = \alpha(g) + \alpha(h) \tag{A1.4}$$

and

$$\omega(g_U h_V, h_U g_V) = \omega(g, h) \,. \tag{A1.5}$$

Hence,

$$\alpha(g_U h_V) + \alpha(h_U g_V) = \alpha(g) + \alpha(h) \,. \tag{A1.6}$$

It follows from (A1.4) and (A1.6) that for every gamete k

$$\omega(k, g_U h_V) + \omega(k, h_U g_V) = \omega(k, g) + \omega(k, h) \,. \tag{A1.7}$$

The system of equations (A1.5) and (A1.7) imply that $\omega(g, h)$ is additive,

$$\omega(g, h) = \sum_{i=1}^{n} \omega_i(g_i, h_i) \tag{A1.8}$$

(Lyubich 1992, Sect. 9.7). Thus,

$$\alpha(g) + \alpha(h) = \sum_{i=1}^{n} \omega_i(g_i, h_i) \,. \tag{A1.9}$$

In particular,

$$\alpha(g) = \frac{1}{2}\sum_{i=1}^{n} \omega_i(g_i, g_i) \,. \tag{A1.10}$$

Coming back to (A1.3) we obtain

$$\lambda(g) = \prod_{1 \leqq i \leqq n} \beta_i(g_i) \tag{A1.11}$$

where

$$\beta_i(g_i) = \exp[\tfrac{1}{2}\,\omega_i(g_i, g_i)]\ . \tag{A1.12}$$

The Theorem 1 is proved.

Corollary. *If haploid selection with fitnesses $\lambda(g)$ satisfies the equation*

$$\lambda(g_U h_V) + \lambda(h_U g_V) = \lambda(g) + \lambda(h) \tag{A1.13}$$

for all $U|V$, then it is additive.

Proof. Let

$$\mu(g) = \exp[\lambda(g)] \tag{A1.14}$$

and

$$\zeta(g, h) = \mu(g)\mu(h)\ , \tag{A1.15}$$

so that we have new haploid selection fitnesses $\mu(g)$ and the equivalent diploid selection fitnesses $\zeta(g, h)$. The latter is recombination-invariant by (A1.13). Indeed,

$$\zeta(g_U h_V, h_U g_V) = \mu(g_U h_V)\mu(h_U g_V) =$$

$$\exp[\lambda(g_U h_V) + \lambda(h_U g_V)] = \exp[\lambda(g) + \lambda(h)] = \zeta(h, g)\ . \tag{A1.16}$$

By Theorem 1, $\mu(g)$ is multiplicative, and hence $\lambda(g) = \ln\,\mu(g)$ is additive,

$$\lambda(g) = \sum_{i=1}^{n} \alpha_i(g_i)\ . \tag{A1.17}$$

Now we consider *additive* diploid selection $\zeta(g, h)$ of haploid origin. By Theorem 1 equivalent haploid selection $\lambda(g)$ is of form (A1.11). On the other hand,

$$\zeta(k, g_U h_V) + \zeta(k, h_U g_V) = \zeta(k, g) + \zeta(k, h) \tag{A1.18}$$

by additivity. By substitution $\zeta(k, g) = \lambda(k)\lambda(g)$ etc., we obtain (A1.13) for the haploid selection. By Corollary this one is of form (A1.17). It remains to discuss the compatibility of (A1.11) and (A1.17).

If $n = 1$ then (A1.11) coincides with (A1.17) taking $\alpha_1(g_1) = \beta_1(g_1)$. Starting with $n = 2$ we prove that *the additive formula* (A1.17) *can be transformed to a multiplicative formula* (A1.11) *if and only if at every locus i, except, maybe, for one of them, all alleles contribute the same value α_i so, $\alpha_i(a_{ij})$ does not depend on j.*

Let $n = 2$ and let $\alpha_1(a_{11}) \neq \alpha_1(a_{12})$. Taking the genes a_{21} and a_{22} for instance, we get

$$\alpha_1(a_{11}) + \alpha_2(a_{21}) = \beta_1(a_{11})\beta_2(a_{21})\ , \tag{A1.19}$$

$$\alpha_1(a_{11}) + \alpha_2(a_{22}) = \beta_1(a_{11})\beta_2(a_{22})\ , \tag{A1.20}$$

$$\alpha_1(a_{12}) + \alpha_2(a_{21}) = \beta_1(a_{12})\beta_2(a_{21})\ , \tag{A1.21}$$

$$\alpha_1(a_{12}) + \alpha_2(a_{22}) = \beta_1(a_{12})\beta_2(a_{22})\ . \tag{A1.22}$$

Hence,

$$\frac{\alpha_1(a_{11}) + \alpha_2(a_{21})}{\alpha_1(a_{11}) + \alpha_2(a_{22})} = \frac{\alpha_1(a_{12}) + \alpha_2(a_{21})}{\alpha_1(a_{12}) + \alpha_2(a_{22})} \tag{A1.23}$$

and then

$$[\alpha_1(a_{11}) - \alpha_1(a_{12})][\alpha_2(a_{21}) - \alpha_2(a_{22})] = 0 \ . \tag{A1.24}$$

Since $\alpha_1(a_{11}) - \alpha_1(a_{12}) \neq 0$ we obtain $\alpha_2(a_{21}) = \alpha_2(a_{22})$.

The general case $n > 2$ can be reduced to the previous one by interpretation of subgametes corresponding to the loci $2, \ldots, n$ as alleles of a 'superlocus'. Obviously, the selection remains to be additive and multiplicative after this procedure. If, as before, $\alpha_1(a_{11}) \neq \alpha_1(a_{12})$ we conclude that for all subgametes

$$g = \prod_{2 \leqq i \leqq n} a_{ij_i} \tag{A1.25}$$

the sums $\sum_{i=2}^{n} \alpha_i(a_{ij_i})$ are the same. Hence, $\alpha_i(a_{ij})$ does not depend on j, $1 \leqq j \leqq m_i$. Conversely, if this property takes place then

$$\lambda(g) = \alpha_1(g_1) + \delta, \qquad \delta = \text{const} \ . \tag{A1.26}$$

Therefore

$$\lambda(g) = \prod_{1 \leqq i \leqq n} \beta_i(g_i) \tag{A1.27}$$

where

$$\beta_1(g_1) = \alpha_1(g_1) + \delta, \qquad \beta_i(g_i) = 1 \quad (i > 1) \ . \tag{A1.28}$$

Remark. The equation (A1.18) appeared as a sufficient condition for the validity of a fundamental theorem for diploids (Lyubich 1992, Sect. 9.5). This result generalizes the fundamental theorem for additive diploid selection (Ewens 1969). Since the fundamental theorem for haploid selection is equivalent to such a theorem for the equivalent diploid selection, we have another proof of our fundamental theorem for additive haploid selection. Indeed, (A1.18) is true in this case. So, we have a nonadditive case when the fundamental theorem takes place.

Appendix 2

Here we prove formula (35). At first, we define the function α on subgametes. For every nonempty subset $U \subset L$ the possible subgametes are of a form

$$\gamma = \prod_{i \in U} a_{ij_i} \ . \tag{A2.1}$$

We set

$$\alpha(\gamma) = \sum_{i \in U} \alpha(a_{ij_i}) \ . \tag{A2.2}$$

In particular, one can take $\gamma = g_U$ where g is from (2). It is clear from (26) that

$$\lambda(g) = \alpha(g_U) + \alpha(g_V) \tag{A2.3}$$

for every partition $U|V$.

Let us denote by Γ_U the set of all subgametes γ of form (A2.1). Then (13) can be written as

$$X_U = \sum_{\gamma \in \Gamma_U} \xi_U(\gamma)\gamma \tag{A2.4}$$

where

$$\xi_U(\gamma) = \sum_{g : g_U = \gamma} \xi(g) = \sum_{\delta \in \Gamma_V} \xi(\gamma\delta) \; . \tag{A2.5}$$

Hence,

$$X_U X_V = \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} \xi_U(\gamma)\xi_V(\delta)\,(\gamma\delta) \tag{A2.6}$$

and

$$\Lambda[X_U X_V] = \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} \lambda(\gamma\delta)\xi_U(\gamma)\xi_V(\delta)\,(\gamma\delta) \; . \tag{A2.7}$$

As a result

$$s(\Lambda[X_U X_V]) = \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} \lambda(\gamma\delta)\xi_U(\gamma)\xi_V(\delta) \; . \tag{A2.8}$$

If follows from (A2.3) that

$$\lambda(\gamma\delta) = \alpha(\gamma) + \alpha(\delta) \; . \tag{A2.9}$$

Then (A2.8) takes the form

$$\begin{aligned} s(\Lambda[X_U X_V]) &= \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} [\alpha(\gamma) + \alpha(\delta)]\xi_U(\gamma)\xi_V(\delta) \\ &= \left\{ \sum_{\gamma \in \Gamma_U} \alpha(\gamma)\xi_U(\gamma) + \sum_{\delta \in \Gamma_V} \alpha(\delta)\xi_V(\delta) \right\} s(X) \; , \end{aligned} \tag{A2.10}$$

because, by (A2.5)

$$\sum_{\gamma \in \Gamma_U} \xi_U(\gamma) = \sum_{\gamma \in \Gamma_U} \sum_{g : g_U = \gamma} \xi(g) = \sum_g \xi(g) = s(X) \tag{A2.11}$$

and similarly,

$$\sum_{\delta \in \Gamma_V} \xi_V(\delta) = s(X) \; . \tag{A2.12}$$

On the other hand,

$$\begin{aligned} s(\Lambda X) = \sum_g \lambda(g)\xi(g) &= \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} \lambda(\gamma\delta)\xi(\gamma\delta) \\ &= \sum_{\substack{\gamma \in \Gamma_U \\ \delta \in \Gamma_V}} [\alpha(\gamma) + \alpha(\delta)]\xi(\gamma\delta) \\ &= \sum_{\gamma \in \Gamma_U} \alpha(\gamma) \sum_{\delta \in \Gamma^V} \xi(\gamma\delta) + \sum_{\delta \in \Gamma_V} \alpha(\delta) \sum_{\gamma \in \Gamma_U} \xi(\gamma\delta) \; . \end{aligned} \tag{A2.13}$$

By virtue of (A2.5)

$$s(\Lambda X) = \sum_{\gamma \in \Gamma_U} \alpha(\gamma)\xi_U(\gamma) + \sum_{\delta \in \Gamma_V} \alpha(\delta)\xi_V(\delta) \; . \tag{A2.14}$$

Comparing (A2.10) and (A2.14) we obtain (35).

This work was supported by the Israeli Ministry of Sciences.

References

Ewens, W. J.: Mean fitness increases when fitnesses are additive. Nature **221**, 1076 (1969)

Feldman, M. W.: Equilibrium studies of two locus haploid populations with recombination. Theor. Popul. Biol. **2**, 229–318 (1971)

Karlin, S.: Theoretical aspects of multi-locus selection balance. I. pp 503–587 in Studies in Mathematics, Vol. 16: Studies in Mathematical Biology, edited by S. A. Levin. Mathematical Association of America, Washington, D.C. (1978)

Karlin, S., Feldman M.: Convergence to equilibria of the two locus additive viability model. J. Appl. Prob. **7**, 262–271 (1970)

Kun, L. A., Lyubich, Yu. I.: Convergence to equilibrium in a polylocus polyallele population with additive selection. Probl. Inform. Transmiss. **16**, 152–161 (1980)

Lewontin, R. C., Kojima, K.: The evolutionary dynamics of complex polymorphisms. Evolution **14**, 458–472 (1960)

Lyubich, Yu. I., Maistrovskii, G. D., Ol'khovskii, Yu. G.: Convergence to equilibrium under the action of selection in a single locus population. Soviet Math. Dokl. **17**, 55–58 (1976)

Lyubich, Yu. I.: Mathematical Structures in Population Genetics. Springer, Berlin (1992) (Original Russian edition: "Matematicheskie structury v populyatsionnoi genetike", Naukova Dumka, Kiev, 1983)

Moran, P. A. P.: The effect of selection in a haploid genetic population. Proc. Camb. Phil. Soc. **54**, 463–467 (1958)

Nagylaki, T.: Rate of evolution of a character without epistasis. Proc. Natl. Acad. Sci. USA **86**, 1910–1913 (1989)

Nagylaki, T.: Introduction to Theoretical Population Genetics. Springer, Berlin (1992)

Nagylaki, T.: The evolution of multilocus systems under weak selection. Genetics **134**, 627–647 (1993)

Reiersöl, O.: Genetic algebras studied recursively and by means of differential operators. Math. Scand. **10**, 25–44 (1962)

Rutschman, D. H.: Dynamics of the two-locus haploid model. Theor. Popul. Biol. **45**, 167–176 (1994)