Loss of genetic variation in the two-locus multiallelic haploid model

March 5, 2020

Martin Pontz[a,b,*] and Marcus W. Feldman[c]

[a] Department of Mathematics, University of Vienna, Austria

[b] Vienna Graduate School of Population Genetics

[c] Department of Biology, Stanford University, Stanford, CA

[*] Corresponding author

Address for correspondence:

Martin Pontz
Institut für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
1090 Wien
Austria
E-mail: martin.pontz@univie.ac.at
Phone: +43 1 4277 50784

1

Abstract

In the evolutionary biology literature, it is generally assumed that in deterministic haploid selection models, in the absence of variation-generating mechanisms such as mutation, no polymorphic equilibrium can be stable. However, results corroborating this claim are scarce and almost always depend upon additional assumptions. Using ideas from game theory, we establish a condition on the fitness parameters of haplotypes formed by two loci such that a monomorphism is a global attractor. Further, we show that no isolated equilibrium exists, at which an unequal number of alleles from two loci is present. Under the assumption of convergence of trajectories to equilirium points, we settle the two-locus three-allele case for a fitness scheme formally equivalent to the classical symmetric viability model.

Key words: selection, recombination, haploid population, unstable equilibria, genetic variation, global stability

Introduction

A recent paper by Novak and Barton (2017) raises one of the main questions of population genetics right in the title: "When does frequency-independent selection maintain genetic variation?" They note that, while the answer is generally assumed to be "never" for constant selection acting on an idealized haploid population, basically only the cases of no recombination and of no selection have been solved and corroborate this claim. Well known results from perturbation theory, of course, expand these results to small parameter values of the respective force. In fact, Novak and Barton (2017) give a new, more standard, proof for the case of weak selection. The other extreme, tight linkage (low recombination rate), was solved by Kirzhner and Lyubich (1997), who also arrive at the same conclusion for additive fitnesses and arbitrary linkage. These three results hold for any number of loci and any number of alleles. They all incorporate convergence of the solutions to equilibrium points, via the powerful method of identifying a Lyapunov function.

Besides the general additive case and the trivial one-locus case, only the two-locus two-allele case has recieved attention for intermediate values of recombination and selection. Feldman (1971) was one of the first to rigorously analyze existence and stability of polymorphic equilibria in a two-locus two-allele haploid system with a simple fitness

scheme. He showed that whenever a polymorphism exists, it is unique and unstable. A general fitness scheme was considered by Rutschman (1994). He showed convergence of the trajectories to equilibrium points in most parameter regimes. However, parameter combinations in which an internal (polymorphic) equilibrium was possible, couldn't be treated in the same way. The final answer to the question of loss of genetic variation in the two-locus two-allele case is the paper by Bank, Bürger and Hermisson (2012). They showed that for the fitness parameter combinations not covered by Rutschman, an equilibrium exists, but it is always unstable.

We consider a well mixed haploid population with constant selection on two loci, each with an arbitrary number of alleles. Our fitness scheme is general without any restriction on the epistatic interaction between alleles. For convenience, the dynamics are stated in continuous time. First, we apply a method used on a game theoretic problem by Hofbauer and Su (2016). With its help we can show that if one allele dominates all the others from the same locus, then this allele becomes fixed. An allele dominating another here means that the fitness of an haplotype containing the *dominating* allele is greater than the fitness of an haplotype containing the other, *dominated*, allele for every choice of background allele. Further, we state and prove that no isolated equilibrium exists if the numbers of alleles at the two loci are unequal. This is done by finding a system of linear equations, whose solution corresponds to an internal equilibrium. For unequal numbers of alleles at the two loci, this system is overdetermined and thus, by basic linear algebra has, in general, no solution.

We also use the above-mentioned system of linear equations to give a different proof of the result by Bank, Bürger and Hermisson (2012). Finally, we show that there is a unique unstable polymorphic equilibrium in the two-locus three-alleles model with centrosymmetric fitnesses.

Model

In the two-locus haploid model considered here, we assume that at one locus the alleles are $A_1, ..., A_m$, while at the other locus the alleles are $B_1, ..., B_n$. Let p_{ij} and s_{ij} be the frequency and the fitness, respectively, of haplotype $A_i B_j$ and define the matrices $S = (s_{ij})_{m \times n}$ an $P = (p_{ij})_{m \times n}$. In the following, we will mainly use the vector p, which is defined as the vector of all rows of P, i.e., $p = (p_{11}, \ldots, p_{1n}, p_{21}, \ldots, p_{m1}, \ldots, p_{mn})^T$. Following Nagylaki (1992, pp. 189–195) and Novak and Barton (2017), we can write

the change in frequency over time, $\dot{p}_{ij} = \frac{dp_{ij}}{dt}$, as

$$\dot{p}_{ij} = r(p_i q_j - p_{ij}) + p_{ij}(s_{ij} - \bar{s}), \tag{1}$$

where $\bar{s} = \sum_{ij} s_{ij} p_{ij}$ is the mean fitness, $p_i = \sum_{j=1}^{n} p_{ij}$ and $q_j = \sum_{i=1}^{m} p_{ij}$ are the marginal frequencies of the alleles. As always, the sum of all haplotype frequencies is one, $\sum_{ij} p_{ij} = 1$. The quantities $p_i p_j - p_{ij}$ are measures of linkage disequilibrium (LD), and $r > 0$ denotes the recombination rate.

We investigate stability properties of the monomorphic equilibria and existence and stability properties of polymorphic equilibria. At equilibrium, the coordinates are denoted by aˆ and satisfy

$$0 = r(\hat{p}_i \hat{q}_j - \hat{p}_{ij}) + \hat{p}_{ij}(s_{ij} - \hat{\bar{s}}) \quad \text{for } 1 \le i \le m \text{ and } 1 \le j \le n. \tag{2}$$

These are mn quadratic equations in mn variables.

The state space of (1) is the $mn-$dimensional simplex Δ_{mn} as defined by

$$\Delta_{mn} = \{x \in \mathbb{R}^{mn} : x_{ij} \ge 0, \sum_{ij} x_{ij} = 1\}. \tag{3}$$

Remark 1. *We note that system* (1) *is invariant with respect to adding a constant c to S. If $s'_{ij} = s_{ij} + c$, then $\bar{s}' = c + \bar{s}$ and thus $s'_{ij} - \bar{s}' = s_{ij} - \bar{s}$ for every pair (i, j). It will be especially useful to take $c = -\hat{\bar{s}}_E$, where $\hat{\bar{s}}_E$ denotes mean fitness at an equilibrium E. At the same equilibrium in the scaled fitness scheme the scaled mean fitness is $\hat{\bar{s}}'_E = \hat{\bar{s}}_E - \hat{\bar{s}}_E = 0$.*

Results

Stability of monomorphic equilibria

First, we determine the conditions under which monomorphism are stable, which is in turn used to give an upper bound for the number of stable hyperbolic monomorphic equilibria.

It is easy to see that each monomorphism is an equilibrium for (1).

To determine the local stability of equilibria, we compute the Jacobian J of $\dot{p}$, given

4

by (1). For every pair (i, j) the following holds:

$$\frac{\partial \dot{p}_{ij}}{\partial p_{ij}} = r(p_i + q_j - 1) + s_{ij}(1 - p_{ij}) - \bar{s}, \tag{4a}$$

$$\frac{\partial \dot{p}_{ij}}{\partial p_{il}} = rq_j - p_{ij}s_{il} \quad \text{if } l \neq j, \tag{4b}$$

$$\frac{\partial \dot{p}_{ij}}{\partial p_{kj}} = rp_i - p_{ij}s_{kj} \quad \text{if } k \neq i, \tag{4c}$$

$$\frac{\partial \dot{p}_{ij}}{\partial p_{kl}} = -p_{ij}s_{kl} \quad \text{if } l \neq j \text{ and } k \neq i. \tag{4d}$$

If we fix (u, v) and sum over the corresponding column of J, we get:

$$\sum_{i,j} \frac{\partial \dot{p}_{ij}}{\partial p_{uv}} = \frac{\partial \dot{p}_{uv}}{\partial p_{uv}} + \sum_{i \neq u} \frac{\partial \dot{p}_{iv}}{\partial p_{uv}} + \sum_{j \neq v} \frac{\partial \dot{p}_{uj}}{\partial p_{uv}} + \sum_{i \neq u, j \neq v} \frac{\partial \dot{p}_{ij}}{\partial p_{uv}}$$

$$= r(p_u + q_v - 1) + s_{uv}(1 - p_{uv}) - \bar{s} + \sum_{i \neq u}(rp_i - p_{iv}s_{uv})$$

$$+ \sum_{j \neq v}(rq_j - p_{ul}s_{uv}) - \sum_{i \neq u, j \neq v} p_{ij}s_{uv}$$

$$= r(\sum_i p_i + \sum_j q_j - 1) + s_{uv}(1 - \sum_{ij} p_{ij}) - \bar{s}$$

$$= r - \bar{s}.$$

Since this holds for every column of the Jacobian, $(1, \ldots, 1)$ is a left eigenvector with the eigenvalue $r - \bar{s}$. Because this eigenvector is normal to the simplex, the corresponding eigenvalue carries no information about the stability of an equilibrium.

Keeping this in mind, we can compute the Jacobian at the monomorphism with $p_{11} = 1$. At this equilibrium, $\hat{\bar{s}} = s_{11}$. Applying this to (4a) yields $\frac{\partial \dot{p}_{11}}{\partial p_{11}} = r - \hat{\bar{s}}$.

The other diagonal entries are given by

$$\frac{\partial \dot{p}_{1j}}{\partial p_{1j}} = s_{1j} - s_{11} \text{ if } j \neq 1, \quad \frac{\partial \dot{p}_{i1}}{\partial p_{i1}} = s_{i1} - s_{11} \text{ if } i \neq 1, \tag{5a}$$

and

$$\frac{\partial \dot{p}_{ij}}{\partial p_{ij}} = -r + s_{ij} - s_{11} \text{ if } i \neq 1 \text{ and } j \neq 1. \tag{5b}$$

The remaining non-zero entries are given by:

$$\frac{\partial \dot{p}_{1j}}{\partial p_{ij}} = r \text{ if } i \neq 1 \text{ and } j \neq 1, \quad \frac{\partial \dot{p}_{i1}}{\partial p_{ij}} = r \text{ if } i \neq 1 \text{ and } j \neq 1, \tag{6a}$$

$$\frac{\partial \dot{p}_{11}}{\partial p_{1j}} = r - s_{1j} \text{ if } j \neq 1, \quad \frac{\partial \dot{p}_{11}}{\partial p_{i1}} = r - s_{i1} \text{ if } i \neq 1, \tag{6b}$$

5

and

$$\frac{\partial \dot{p}_{11}}{\partial p_{ij}} = -s_{ij} \ \text{ if } i \neq 1 \text{ and } j \neq 1. \tag{6c}$$

All other entries are zero.

After recalling the lexicographical order of the double indices in the vector p, inspection of the non-zero entries shows that $J|_{p_{11}=1}$ is an upper right triangular matrix and therefore, the eigenvalues are the diagonal entries given by (5).

We can, in general, relabel alleles and loci, such that the Jacobian of any monorphism is an upper right triangular matrix. This means that for the monomorphism $A_{i_0}B_{j_0}$, the $mn - 1$ eigenvalues that determine stability are given by

$$s_{i_0 j} - s_{i_0 j_0} \ \text{ if } j \neq j_0, \quad s_{i j_0} - s_{i_0 j_0} \ \text{ if } i \neq i_0, \tag{7a}$$

and

$$-r + s_{ij} - s_{i_0 j_0} \ \text{ if } i \neq i_0 \text{ and } j \neq j_0. \tag{7b}$$

However, for certain choices of the fitness values, some of the monomorphisms are not isolated and have an eigenvalue equal to zero as the following result shows.

Lemma 1. *Every point on the edge connecting $A_{i_0}B_{k_1}$ with $A_{i_0}B_{k_2}$, given by $p_{i_0 k_1} + p_{i_0 k_2} = 1$, is an equilibrium if and only if $s_{i_0 k_1} = s_{i_0 k_2}$.*

Proof. On the edge connecting $A_{i_0}B_{k_1}$ with $A_{i_0}B_{k_2}$ $p_{i_0 k_1} + p_{i_0 k_2} = 1$ and $p_{ij} = 0$ for all i and j that do not form the pairs $(i_0 k_1)$ or $(i_0 k_2)$. Hence, $p_i q_j = 0$ and thus, after plugging this into (1), $\dot{p}_{ij} = 0$ if $i \neq i_0$ and $j \neq k_1, k_2$. Therefore, the only equations of (1) with non-zero right-hand side are

$$\dot{p}_{i_0 k_1} = r((p_{i_0 k_1} + p_{i_0 k_2})p_{i_0 k_1} - p_{i_0 k_1}) - p_{i_0 k_1}(s_{i_0 k_1} - s_{i_0 k_1}p_{i_0 k_1} - s_{i_0 k_2}p_{i_0 k_2})$$
$$= -p_{i_0 k_1}(s_{i_0 k_1} - s_{i_0 k_2})(1 - p_{i_0 k_1}) = -\dot{p}_{i_0 k_2}.$$

This is zero for all $0 \leq p_{i_0 k_1} \leq 1$ if and only if $s_{i_0 k_1} = s_{i_0 k_2}$. $\qquad\square$

Remark 2. *An analoguos result holds for the edge connecting $A_{k_1}B_{j_0}$ with $A_{k_2}B_{j_0}$*

To avoid the complications of degenerate cases, we assume for this and the next section that

$$\text{All monomorphisms of (1) are hyperbolic equilibria.} \tag{$\mathcal{H}$}$$

This means that no eigenvalue of a monomorphic equilibrium is zero. In particular, this implies that no two fitness values of haplotypes that share an allele are the same.

Proposition 1. *The monomorphism $A_{i_0}B_{j_0}$ is asymptotically stable if and only if $r > \max_{i,j}(s_{ij}) - s_{i_0 j_0}$. In this case (i_0, j_0) is the unique pair such that $\max_j(s_{i_0 j}) = s_{i_0 j_0} = \max_i(s_{i j_0})$.*

Proof. The eigenvalues of the monomorphism $A_{i_0}B_{j_0}$ given by (7) are negative if and only if $(i_0 j_0)$ is the unique pair such that $\max_j(s_{i_0 j}) = s_{i_0 j_0} = \max_i(s_{i j_0})$ and $r > \max_{i,j}(s_{ij}) - s_{i_0 j_0}$ holds. $\qquad\square$

The number of asymptotically stable monomorphisms is bounded by the following

Corollary 1. *In a system with n alleles present at one locus and m at the other, there are between one and $\min(m, n)$ asymptotically stable monomorphisms.*

Proof. By Proposition 1, the monomorphism corresponding to the largest entry of the fitness matrix is always locally asymptotically stable because $r > 0$.

Let $m < n$ and let m be the number of rows of the fitness matrix. Assumption $(\mathcal{H})$ implies that each row has a unique maximum. There are exactly m such maxima. If each of those is also the maximum in its respective column then the corresponding monomorphisms are locally asymptotically stable for sufficiently large r by Theorem 1. There cannot be more. $\qquad\square$

Corollary 2. *Let $s_{i_0 j_0} = \max_{i,j}(s_{ij})$, then for every $r > 0$ the monomorphism $A_{i_0}B_{j_0}$ is the only locally asymptotically stable monomorphism if and only if for all pairs (k, l) $s_{kl} \leq \max\left(\max_i(s_{il}), \max_j(s_{kj})\right)$ with equality only if $(k, l) = (i_0, j_0)$, i.e., $s_{i_0 j_0}$ is the only value that is maximal in both its row and column.*

Proof. For each pair $(k, l) \neq (i_0, j_0)$ assume that $s_{kl} < \max\left(\max_i(s_{il}), \max_j(s_{kj})\right)$. Then by Proposition 1 and Corollary 1 it is clear that for every $r > 0$, $A_{i_0}B_{j_0}$ is the only locally asymptotically stable monomorphism.

Conversely, suppose there is a monomorphism $A_k B_l$ with $k \neq i_0$ and $l \neq j_0$ such that $\max_j s_{kj} = s_{kl} = \max_i s_{il}$. There exists $r_0 > 0$ such that $r_0 = s_{i_0 j_0} - s_{kl}$. Then (7) implies that $A_k B_l$ is stable for $r > r_0$. This contradicts the assumption that for every $r > 0$, $A_{i_0}B_{j_0}$ is the only stable monomorphism. $\qquad\square$

If we assume that genetic variation is never maintained in a haploid population under selection and recombination, then the monomorphism described in Corollary 2 would be the natural candidate for a global attractor. However, no proof could be found to verify this.

Dominating alleles

In the following, we apply ideas from game theory, in particular from a paper by Hofbauer and Su (2016) about dominating strategies, to alleles. For a special class of fitness schemes, this allows us to prove global stability of a monomorphism.

Theorem 1. *If there exist alleles A_{i_0} and A_{i_1} such that $s_{i_0 j} > s_{i_1 j}$ holds for every j, then $p_{i_1} \to 0$.*

Proof. Without loss of generality, suppose that $s_{1j} > s_{2j}$ holds for all j. We show that the minimal quotient $Q(t) = \frac{p_{1l}(t)}{p_{2l}(t)} = \min_j \left(\frac{p_{1j}(t)}{p_{2j}(t)} \right)$ is increasing along trajectories in t. First, note that $l = l(t)$ can assume different values in $\{1, \ldots, m\}$ at different times. Further, if p_1, $p_2 > 0$, then for every given t

$$\frac{p_{1l}}{p_{2l}} = \frac{p_{1l}p_2}{p_{2l}p_2} = \frac{1}{p_2} \sum_j \frac{p_{1l}p_{2j}}{p_{2l}} \leq \frac{1}{p_2} \sum_j \frac{p_{1j}p_{2j}}{p_{2j}} = \frac{1}{p_2} \sum_j p_{1j} = \frac{p_1}{p_2} \tag{8}$$

holds, which is equivalent to

$$p_1 p_{2l} - p_2 p_{1l} \geq 0. \tag{9}$$

Then,

$$\dot{Q} = \frac{1}{p_{2l}^2} \left(p_{2l} \left[r(p_1 q_l - p_{1l}) + p_{1l}(s_{1l} - \bar{s}) \right] - p_{1l} \left[r(p_2 q_l - p_{2l}) + p_{2l}(s_{2l} - \bar{s}) \right] \right) \tag{10a}$$

$$= \frac{1}{p_{2l}^2} \left(r q_l (p_1 p_{2l} - p_2 p_{1l}) + p_{1l} p_{2l}(s_{1l} - s_{2l}) \right) \geq Q(s_{1l} - s_{2l}) > 0. \tag{10b}$$

Define $\delta = \min_l (s_{1l} - s_{2l}) > 0$. Then (10) implies

$$\dot{Q} \geq Q\delta \tag{11}$$

and subsequently

$$Q(t) \geq Q(0)e^{\delta t} \to \infty \tag{12}$$

as $t \to \infty$.

This implies that $\frac{p_{1j}}{p_{2j}}$ goes to infinity as $t \to \infty$ for all j. Since the numerator is bounded, $p_{2j} \to 0$ for all j and thus $p_2 \to 0$ as $t \to \infty$. $\square$

Remark 3. *A similar theorem holds for the other locus.*

This can be interpreted such that allele A_{i_0} dominates A_{i_1}, since for every background allele B_J the haplotype containing allele A_{i_0} is fitter than that containing A_{i_1}. Almost intuitively, this leads to the loss of A_{i_1}. Ultimately, if A_{i_0} dominates all other alleles at the same locus, these should all go extinct. This is formalized in the following

Theorem 2. *If there is an allele A_{i_0} with $s_{i_0 j} > s_{ij}$ for all $i \neq i_0$ and all j, then the monomorphism $A_{i_0} B_{j_0}$ with $s_{i_0 j_0} > s_{i_0 j}$ for all $j \neq j_0$ is globally asymptotically stable.*

Proof. For each $i \neq i_0$ we apply Theorem 1. Then allele A_{i_0} becomes fixed in the population. Because of $(\mathcal{H})$, there exists a unique j_0 such that $s_{i_0 j_0} > s_{i_0 j}$ for all $j \neq j_0$. By (7) the monomorphism $A_{i_0 j_0}$ is the only asymptotically stable monomorphism $\qquad \square$

Polymorphic equilibria

Monomorphisms are equilibria where exactly one allele at each locus is present. Next, we look at equilibria where at least three alleles are involved.

While system (1) with state space Δ_{mn}, with $m < n$, can always be imbedded in the system with state space Δ_{nn} where both loci have the same number of alleles, it simplifies derivations if we think of the equilibria in subsystems as being fully polymorphic.

From now on, we are thus mainly interested in the existence of fully polymorphic equilibria, i.e., states, where all haplotypes are present. Therefore, the inequalities

$$0 < \hat{p}_{ij} < 1 \quad \forall i, j \tag{13}$$

hold, which is equivalent to $p \in \Delta_{mn}^0$. Here, ^{0} denotes the interior of the simplex.

Our main result is

Theorem 3. *If $m \neq n$, then (2) has either no or infinitely many solutions for which (13) holds. Thus, there are no isolated equilibria with all mn haplotypes present.*

In order to prove the theorem, we first need

Lemma 2. *Define the matrix $\tilde{S}(\sigma)$ by*

$$\tilde{s}_{ij}(\sigma) = \frac{s_{ij} - \sigma}{r + \sigma - s_{ij}}, \tag{14}$$

where $\sigma \in \mathbb{R}$. For a given fitness matrix S and $r > 0$, a solution of (2) that fulfills (13) exists if and only if $\exists \hat{\sigma} \in \mathbb{R}$, $\mathbf{x} \in \Delta_m$ and $\mathbf{y} \in \Delta_n$ such that

$$r + \hat{\sigma} - s_{ij} > 0 \quad \forall i, j, \tag{15a}$$

$$\tilde{S}(\hat{\sigma})\mathbf{y} = 0, \tag{15b}$$

$$\mathbf{x}^T \tilde{S}(\hat{\sigma}) = 0. \tag{15c}$$

If a solution exists, it is given by

$$\hat{p}_{ij} = \frac{r x_i y_j}{r + \hat{\sigma} - s_{ij}} \quad \forall i, j, \tag{16}$$

9

and then

$$\hat{p}_i = x_i, \ \hat{q}_j = y_j, \ and \ \hat{\tilde{s}} = \hat{\sigma} \tag{17}$$

holds.

Here, Δ_m denotes the $m-$dimensional simplex.

Proof. ($\Rightarrow$) From (2), we can compute the following identities:

$$\hat{p}_{ij} = \frac{r\hat{p}_i\hat{q}_j}{r + \hat{\tilde{s}} - s_{ij}} \quad \forall i, \, j, \tag{18a}$$

$$\dot{p}_i = \sum_j \dot{p}_{ij} = \sum_j \hat{p}_{ij}(s_{ij} - \hat{\tilde{s}}) = 0 \quad \forall i, \tag{18b}$$

$$\dot{q}_j = \sum_i \dot{p}_{ij} = \sum_i \hat{p}_{ij}(s_{ij} - \hat{\tilde{s}}) = 0 \quad \forall j. \tag{18c}$$

We plug (18a) into (18b) and (18c) to get:

$$\dot{p}_i = r\hat{p}_i \sum_j \frac{s_{ij} - \hat{\tilde{s}}}{r + \hat{\tilde{s}} - s_{ij}}\hat{q}_j = 0 \quad \forall i, \tag{19a}$$

$$\dot{q}_j = r\hat{q}_j \sum_i \frac{s_{ij} - \hat{\tilde{s}}}{r + \hat{\tilde{s}} - s_{ij}}\hat{p}_i = 0 \quad \forall j. \tag{19b}$$

Then (13) entails $\hat{p}_i, \hat{q}_j > 0 \quad \forall i, \, j$. Thus, we can write (19) as

$$\sum_j \frac{s_{ij} - \hat{\tilde{s}}}{r + \hat{\tilde{s}} - s_{ij}}\hat{q}_j = 0 \quad \forall i, \tag{20a}$$

$$\sum_i \frac{s_{ij} - \hat{\tilde{s}}}{r + \hat{\tilde{s}} - s_{ij}}\hat{p}_i = 0 \quad \forall j, \tag{20b}$$

or in matrix terms, after we set $\hat{p}_i = x_i$ and $\hat{q}_j = y_j$:

$$\tilde{S}(\hat{\sigma})\mathbf{x} = 0, \tag{20c}$$

$$\mathbf{y}^T \tilde{S}(\hat{\sigma}) = 0. \tag{20d}$$

The remaining condition (15a) is implicit in (18a) because of (13).

($\Leftarrow$) We define

$$\hat{p}_{ij} = \frac{rx_iy_j}{r + \hat{\sigma} - s_{ij}}. \tag{21}$$

We have to show that

$$x_i = \sum_j \hat{p}_{ij} = \hat{p}_i, \tag{22a}$$

$$y_j = \sum_i \hat{p}_{ij} = \hat{q}_j \tag{22b}$$

10

and

$$\hat{\sigma} = \sum_{ij} s_{ij}\hat{p}_{ij} = \hat{\bar{s}}. \tag{22c}$$

We can write

$$\hat{q}_j = \sum_i \hat{p}_{ij} = \sum_i \frac{rx_iy_j}{r + \hat{\sigma} - s_{ij}} = y_j \sum_i \frac{(r + \hat{\sigma} - s_{ij})x_i}{r + \hat{\sigma} - s_{ij}} + y_j \sum_i \frac{(s_{ij} - \hat{\sigma})x_i}{r + \hat{\sigma} - s_{ij}} = y_j,$$

since $\mathbf{x} \in \Delta_m$ and $\sum_i \frac{(s_{ij} - \hat{\sigma})x_i}{r + \hat{\sigma} - s_{ij}} = 0$ by (15c).

An analogous computation shows that $\hat{p}_i = \sum_j \hat{p}_{ij} = x_i$. This implies

$$\sum_{ij} \hat{p}_{ij} = 1, \tag{23}$$

because $\mathbf{x} \in \Delta_m$.

We can rewrite (21) as

$$rx_iy_j - \hat{p}_{ij}(r + \hat{\sigma} - s_{ij}) = 0. \tag{24}$$

Summing (24) over all i and j, yields

$$r\left(1 - \sum_{ij} \hat{p}_{ij}\right) + \sum_{ij} \hat{p}_{ij}s_{ij} - \hat{\sigma}\sum_{ij} \hat{p}_{ij} = 0, \tag{25a}$$

which implies

$$\sum_{ij} \hat{p}_{ij}s_{ij} = \hat{\sigma} = \hat{\bar{s}}. \tag{25b}$$

Now, we can write (24) as

$$r\hat{p}_i\hat{q}_j - \hat{p}_{ij}(r + \hat{\bar{s}} - s_{ij}) = 0, \tag{26}$$

which is (2). $\qquad\square$

With this characterization of the polymorphism at hand, we can prove Theorem 3.

Proof of Theorem 3. If $\tilde{S}$ is such that there exist no vectors x, y and values r and $\hat{\sigma}$ such that all conditions in (15) are fulfilled, then no equilibrium exists. The following argument shows that if there is a solution that satisfies (15), then there are infinitely many provided $m \neq n$.

Assume $m < n$ and suppose there exists $\mathbf{x} \in \Delta_m$, $\mathbf{y} \in \Delta_n$ and $\hat{\sigma} \in \mathbb{R}$ such that (15) holds. Then $\text{rank}(\tilde{S}) \leq m - 1$ which, because of the rank-nullity theorem, implies

11

that $\dim(\ker(\tilde{S})) \geq n - (m-1) \geq 2$. This means that at least one additional linearly independent vector $\mathbf{y}'$ exists in the kernel of $\tilde{S}$. This solution vector does not necessarily lie in the simplex. However, $\frac{\mathbf{y}+\epsilon\mathbf{y}'}{1+\epsilon\sum_j \mathbf{y}'_j}$ defines a one dimensional manifold that lies in the simplex for $0 < \epsilon < \epsilon^*$ with $\epsilon^* > 0$ sufficiently small and is a solution of (15b). Therefore, $\hat{p}_{ij}$ is at least one-dimensional. $\qquad\square$

The characterization of internal equilibria given by Lemma 2 yields necessary conditions for the existence of equilibria with an equal number of alleles present.

Proposition 2. *If $n = m$ and an isolated equilibrium $(\hat{p}_{ij})$ of (2) satisfying (13) exists, then the following holds:*

(a) No row or column of $\tilde{S}(\hat{\hat{s}})$ consists only of entries of the same sign.

(b) $\mathrm{rank}(\tilde{S}(\hat{\hat{s}})) = n - 1$.

(c) $\det(\tilde{S}(\hat{\hat{s}})) = 0$.

(d) There is no other equilibrium with the same $\hat{\hat{s}}$.

Proof. Let $\mathbf{y}$ be the vector for which (15b) holds. First, assume that there exists a row i of $\tilde{S}$ with all entries positive. This implies $\sum_j \tilde{s}_{ij} y_j > 0$, a contradiction, because (15b) holds for $\mathbf{y}$. An analogous argument works for column j and $\mathbf{x}$. This yields statement (a).

The rank of $\tilde{S}$ has to be smaller than n, because (15b) can only have a nontrivial solution if $\tilde{S}$ is singular.

Now, assume that $\mathrm{rank}(\tilde{S}) < n-1$. Then by the rank-nullity theorem, $\dim(\ker(\tilde{S})) > n - (n-1) = 1$. Therefore, at least one additional linearly independent vector $\mathbf{y}'$ exists in the kernel of $\tilde{S}$. This solution vector does not necessarily lie in the simplex. However, $\frac{\mathbf{y}+\epsilon\mathbf{y}'}{1+\epsilon\sum_j \mathbf{y}'_j}$ defines a one dimensional manifold that lies in the simplex for $0 < \epsilon < \epsilon^*$ with $\epsilon^* > 0$ sufficiently small and is a solution of (15b). This contradicts the assumption of an isolated equilibrium. Thus statement (b) is true. Statement (c) immediately follows.

If $\mathrm{rank}(\tilde{S}) = n - 1$, then $\mathbf{y}$ spans the kernel of $\tilde{S}$. Hence no other equilibrium with the same $\hat{\hat{s}}$ is possible. Hence, statement (d) follows. $\qquad\square$

Note, that statements (a)-(d) are not sufficient conditions for the existence of an internal equilibrium. In particular, statement (a) of Proposition 2 does not imply that a positive solution vector exists. It is also not clear if there always exists a $\hat{\sigma}$ that simultaneously fulfills (15a) and statements (a) and (c) of Proposition 2.

In simple situations Lemma 2 and Proposition 2 allow us to get further results.

Statements (c) and (d) combined yield an upper bound for the number of internal equilibria, since each zero of $\det(\tilde{S}(\sigma))$ gives rise to at most one equilibrium. Since $\det(\tilde{S}(\sigma))$ is a rational function of σ, the degree of the numerator polynomial determines the maximal possible number of internal equilibria. For small numbers of alleles this argument gives rise to a feasible method to determine all admissible internal equilibria. In fact, for two alleles we show in the following that there is at most one internal equilibrium, which is also true for three alleles with a centrosymmetric fitness scheme (see SI).

If S' is the scaled fitness scheme with respect to the equilibrium E, then

$$\tilde{S}(\hat{\tilde{s}}_E) = \tilde{S}'(0) \tag{27}$$

holds, by Remark 1. Therefore, statement (c) of Proposition 2 implies $\det \tilde{S}'(0) = 0$. This yields an additional identity that the scaled fitnesses s'_{ij} have to satisfy. For small numbers of alleles, this can be solved explicitly and implicit coordinates for E can be obtained.

Two explicit cases with an equal number of alleles at both loci

Two alleles

In the two allele case the following Theorem is already known from Bank, Bürger and Hermisson (2012), but we present a different proof.

Theorem 4. *System* (1) *restricted to two alleles at each locus has a unique internal equilibrium if and only if there exists $\hat{\tilde{s}} \in \mathbb{R}$ such that $r > s_{i_0 j_0} - \max_{(i,j) \neq (i_0,j_0)} s_{ij}$ and $\min(s_{11}, s_{22}) > \hat{\tilde{s}} > \max(s_{12}, s_{21})$ or the reverse order holds. Here, $s_{i_0 j_0} = \max_{ij} s_{ij}$. If the equilibrium exists, it is unstable.*

Proof. ($\Rightarrow$) Without loss of generality, we assume $s_{11} = s_{i_0 j_0} = \max_{ij} s_{ij}$. Since the equilibrium is unique, statement (a) of Proposition 2 yields

$$0 < \text{sign}(s_{11} - \hat{\tilde{s}}) = -\text{sign}(s_{12} - \hat{\tilde{s}}) = \text{sign}(s_{22} - \hat{\tilde{s}}) = -\text{sign}(s_{21} - \hat{\tilde{s}})$$

This implies

$$s_{11} > s_{22} > \hat{\tilde{s}} > \max(s_{12}, s_{21}). \tag{28}$$

Additionaly, Lemma 2 tells us that (15a) has to hold, which we combine with (27) to conclude

$$r > s_{11} - \hat{\tilde{s}} > s_{11} - s_{22}. \tag{29}$$

13

($\Longleftarrow$)   Without loss of generality, we assume $s_{11} > s_{22} > \hat{\tilde{s}} > s_{12} > s_{21}$. By applying Proposition 1, we know that the monomorphism $A_1 B_1$ is asymptotically stable for all r, while the monomorphisms $A_1 B_2$ and $A_2 B_1$ are unstable for all r. The monomorphism $A_2 B_2$ is asymptotically stable if $r > s_{11} - s_{22}$. This is exactly the condition $r > s_{i_0 j_0} - \max_{(i,j) \neq (i_0, j_0)} s_{ij}$ with $(i_0, j_0) = (1, 1)$. There are no other boundary equilibria, since by Theorem 3 there is no equilibrium on the edges where one allele is fixed. Thus, the sum of all indices on the boundary is 2. Since the index theorem by Hofbauer (1990) implies that the sum over the indices of all saturated equilibria is 1, the sum of all indices of internal equilibria has to equal -1 (they are saturated by definition). This entails an odd number of internal equilibria, because the index of a hyperbolic equilibrium is either $+1$ or -1.

However, the degree of the numerator polynomial (in σ) of the determinant of the corresponding matrix $\tilde{S}(\sigma)$ is two (see SI for the exact expression). Hence, there are at most two values of σ such that $\tilde{S}$ is singular and this is a necessary condition by statement (b) of Proposition 2. Thus there are up to two internal equilibria. By the index argument above, only an odd number of internal equilibria is possible and therefore, there is exactly one internal equilibrium. It has index -1, which implies an odd number of positive eigenvalues and subsequently its instability. $\square$

With an additive scaling of the fitness matrix, detailed in Remark 1, we can easily express the equilibrium coordinates of the internal equilibrium.

Corollary 3. *Let S be the scaled fitness scheme such that $\hat{\tilde{s}} = 0$ at the internal equilibrium and $s_{22} = \check{s}_{22}$ holds, with*

$$\check{s}_{22} = \frac{s_{12} s_{21}(r - s_{11})}{s_{12}(s_{21} - s_{11}) + s_{11}(r - s_{21})}, \tag{30}$$

then the equilibrium is given by

$$p_{11} = \frac{s_{12} s_{21}(r - s_{11})}{r(s_{11} - s_{12})(s_{11} - s_{21})}, \quad p_{12} = -\frac{s_{11} s_{21}(r - s_{11})}{r(s_{11} - s_{12})(s_{11} - s_{21})} \tag{31a}$$

$$p_{21} = -\frac{s_{11} s_{12}(r - s_{11})}{r(s_{11} - s_{12})(s_{11} - s_{21})}, \quad p_{22} = 1 - p_{11} - p_{12} - p_{21}. \tag{31b}$$

It is admissible if $\operatorname{sign}(s_{11} s_{12}) = \operatorname{sign}(s_{21} \check{s}_{22}) = -\operatorname{sign}(s_{11} \check{s}_{22}) = -1$ *and*

$$r > \begin{cases} s_{11} & \text{if } s_{11} > 0 \\ s_{12} + s_{21} - \dfrac{s_{12} s_{21}}{s_{11}} & \text{if } s_{11} < 0 \end{cases} \tag{32}$$

14

Proof. According to Remark 1, we can formally set $\hat{\check{s}} = 0$ in $\tilde{S}$, restricted to two alleles per locus, which simplifies further analysis. $\tilde{S}$ is a singular matrix if $s_{22} = \check{s}_{22}$. Subsequently we solve (15b) and (15c) and use formula (16) to derive (30). Admissibility conditions follow from Theorem 4, after one resolves the dependence of $\check{s}_{22}$ on r with repect to $\text{sign}(\check{s}_{22})$ and $r > \check{s}_{22}$. See also SI. $\qquad\square$

Three alleles

We now investigate the case of three alleles at each locus. Additionally, we assume a centrosymmetric fitness scheme:

$$
S_c = \begin{array}{c} \\ \end{array}\begin{array}{ccc} B_1 & B_2 & B_3 \\ \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{21} \\ s_{13} & s_{12} & s_{11} \end{pmatrix} & & \begin{array}{c} A_1 \\ A_2 \\ A_3 \end{array} \end{array} \tag{33}
$$

In the sense that the matrix is the same but the entries correspond to haplotypes formed by single alleles rather than by diploid genotypes, S_c is formally equivalent to the well-studied symmetric viability model. Similarly, it implies that the fitnesses stay the same under a simultaneous exchange of alleles A_1 with A_3 and B_1 with B_3. Unfortunately, it violates assumption $(\mathcal{H})$, since $s_{32} = s_{12}$ and $s_{23} = s_{21}$ and thus every point on the edges connecting the monomorphisms A_1B_2 with A_3B_2 and A_2B_1 with A_2B_3 is an equilibrium by Lemma 1.

Since we are mainly interested in the stability of a potential internal equilibrium, from now on, we assume that an internal equilibrium exists with mean fitness at this equilibrium given by $\hat{\check{s}} = 0$ (see Remark 1).

This simplifies (14) considerably, because

$$
\tilde{S}_c = \begin{pmatrix} \frac{s_{11}}{r-s_{11}} & \frac{s_{12}}{r-s_{12}} & \frac{s_{13}}{r-s_{13}} \\ \frac{s_{21}}{r-s_{21}} & \frac{s_{22}}{r-s_{22}} & \frac{s_{21}}{r-s_{21}} \\ \frac{s_{13}}{r-s_{13}} & \frac{s_{12}}{r-s_{12}} & \frac{s_{11}}{r-s_{11}} \end{pmatrix}. \tag{34}
$$

The coordinates of the equilibrium for which S is scaled, are described by the following

Proposition 3. *Let $\tilde{S}_c$ be as in (33) and $s_{11} \neq s_{13}$. In addition, define*

$$
\check{s}_{22} = \frac{2s_{12}s_{21}(r-s_{11})(r-s_{13})}{(r-s_{12})E - s_{21}(E+s_{12}F)}, \tag{35}
$$

with $E = s_{11}(r-s_{13}) + s_{13}(r-s_{11})$ and $F = s_{11} + s_{13} - 2r$.

Then, an equilibrium, given by (16) satisfying (13), exists if and only if

$$s_{22} = \check{s}_{22}, \tag{36a}$$

$$r > \max_{ij} s_{ij}, \tag{36b}$$

$$\text{sign}(s_{22}s_{12}) = \text{sign}(s_{22}s_{21}) = -1 \tag{36c}$$

and

$$\text{sign}(s_{22}s_{11}) = +1 \ \text{ or } \ \text{sign}(s_{22}s_{13}) = +1. \tag{36d}$$

Its coordinates are

$$\hat{p}_1 = \hat{p}_3 = -\frac{(r - s_{11})(r - s_{13})s_{21}}{r(E + s_{21}F)}, \tag{37a}$$

$$\hat{p}_2 = 1 - 2\hat{p}_1 = \frac{(r - s_{21})E}{r(E + s_{21}F)} \tag{37b}$$

and

$$\hat{q}_1 = \hat{q}_3 = -\frac{(r - s_{11})(r - s_{13})s_{12}}{r(E + s_{12}F)}, \tag{37c}$$

$$\hat{q}_2 = 1 - 2\hat{q}_1 = \frac{(r - s_{12})E}{r(E + s_{12}F)}. \tag{37d}$$

Proof. ($\Rightarrow$) If the equilibrium exists, we can scale S such that $\tilde{S}$ is given by (33) and $\hat{\tilde{s}} = 0$. Then, condition (15a) simplifies to

$$r - s_{ij} > 0 \quad \forall i, j, \tag{38}$$

which implies (35b). The equation in statement (c) of Proposition 2 for (33) is linear in s_{22} and thus has the unique solution $\check{s}_{22}$. Statemant (a) implies (35c) and (35d), because otherwise, at least one row or column of (33) consists only of entries of the same sign.

($\Leftarrow$) The determinant of (33) is linear in s_{22} and is zero if $s_{22} = \check{s}_{22}$. Thus, $\tilde{S}_c$ is a singular matrix and (15b) can have a admissible solution. Equating the first and third row of (15b) yields

$$\tilde{s}_{11}y_1 + \tilde{s}_{12}y_2 + \tilde{s}_{13}y_3 = \tilde{s}_{13}y_1 + \tilde{s}_{12}y_2 + \tilde{s}_{11}y_3 \tag{39a}$$

which simplifies to

$$(\tilde{s}_{11} - \tilde{s}_{13})(y_1 - y_3) = 0, \tag{39b}$$

16

where $\tilde{s}_{ij}$ denote the entries of $\tilde{S}_c$. Since $s_{11} \neq s_{13}$, (38b) holds if and only if $y_1 = y_3$. By applying $\sum_j y_j = 1$, we can express (15b) in terms of y_1 and solve for it. An analogous argument yields x_1 and we get (36) (at equilibrium $\hat{p}_i = x_i$ and $\hat{q}_j = y_j$).

By Lemma 2 these coordinates determine an equilibrium for (1). However, it remains to show that $\hat{p}_i$ and $\hat{q}_j$ given by (36) are positive, given the conditions (35). Without loss of generality, $s_{22} = \check{s}_{22} > 0$. This implies that $s_{12}, s_{21} < 0$. Let h_1 denote the numerator and h_2 the denominator of $\check{s}_{22}$ given in (34). Since h_1 is clearly positive, the same has to hold for h_2. Condition (37) also holds for $\check{s}_{22}$ and simplifies to

$$0 < r - \check{s}_{22} = \frac{(r - s_{12})(r - s_{21})E}{h_2}. \tag{40}$$

Therefore, we conclude that $\text{sign}(E) = \text{sign}(h_2) = \text{sign}(s_{22}) = 1$, since the first two factors of the numerator are positive because of (37). This is possible only if (35d) is true, since otherwise, $s_{11}, s_{13} < 0$ imply $E < 0$, provided (37) holds. Because of (37), $F < 0$ and thus the denominators of (36) are all positive. A simple check reveals that this also holds for each numerator of (36).

For any fitness scheme that can be achieved by adding a constant to s_{ij}, an equilibrium exists and its coordinates are given by (36) where a constant is added to each s_{ij}. $\qquad\square$

If $s_{11} = s_{13}$, then it is clear that $\tilde{S}_c$ given by (33) is singular. However, if $\text{rank}(\tilde{S}_c) = 2$, then one can easily check that the equilibrium is not admissible. If $\text{rank}(\tilde{S}_c) = 1$, then the equilibrium is not isolated by Proposition 2.

Corollary 4. *The equilibrium given by* (36) *is centrosymmetric, i.e.,*

$$\hat{p}_{11} = \hat{p}_{33}, \quad \hat{p}_{12} = \hat{p}_{32}, \tag{41a}$$

$$\hat{p}_{13} = \hat{p}_{31}, \quad \hat{p}_{21} = \hat{p}_{23}. \tag{41b}$$

Proof. Using (16) one can derive the exact expressions for $\hat{p}_{ij}$ given (36), they are rather lengthy and thus only shown in the SI. Then it is easy to check that they satisfy (40). $\qquad\square$

Before we prove the instability of an internal poylmorphism, if it exists, we note that Corollary 4 gives rise to the coordinate transformation $(\mathcal{U})$ inspired by Feldman and Karlin (1970), which provides the starting point for the proof.

We define:

$$u_1 = p_{11} - p_{33}, \quad u_2 = p_{12} - p_{32}, \quad u_3 = p_{13} - p_{31}, \quad u_4 = p_{21} - p_{23},$$

$$u_5 = p_{11} + p_{33} - p_{12} - p_{32}, \quad\quad u_6 = p_{13} + p_{31} - p_{21} - p_{23}, \quad\quad (\mathcal{U})$$

$$u_7 = p_{11} + p_{12} + p_{32} + p_{33} - p_{13} - p_{21} - p_{23} - p_{31}, \quad u_8 = p_{22}.$$

From $(\mathcal{U})$ and the simplex condition $\sum_{ij} p_{ij} = 1$, we derive the reverse transformation:

$$p_{11} = \frac{1}{8}(1 + 4u_1 + 2u_5 + u_7 - u_8), \quad p_{12} = \frac{1}{8}(1 + 4u_2 - 2u_5 + u_7 - u_8),$$

$$p_{13} = \frac{1}{8}(1 + 4u_3 + 2u_6 - u_7 - u_8), \quad p_{21} = \frac{1}{8}(1 + 4u_4 - 2u_6 - u_7 - u_8),$$

$$p_{22} = u_8,$$

$$p_{23} = \frac{1}{8}(1 - 4u_4 - 2u_6 - u_7 - u_8), \quad p_{31} = \frac{1}{8}(1 - 4u_3 + 2u_6 - u_7 - u_8),$$

$$p_{32} = \frac{1}{8}(1 - 4u_2 - 2u_5 + u_7 - u_8), \quad p_{33} = \frac{1}{8}(1 - 4u_1 + 2u_5 + u_7 - u_8).$$

With this, we now derive the transformed system of equations from (1). However, this rather lengthy system of ODEs is only shown in the SI. The equilibrium coordinates $\hat{p}_{ij}$ are also transformed into the equilibrium $\hat{u}_i$ (see SI). Because of Corollary 4 and $(\mathcal{U})$, $\hat{u}_i = 0$ for $i = 1, 2, 3$ and 4.

With these prerequisites, we can prove the main theorem of this section.

Theorem 5. *If the polymorphic equilibrium (36) of (1) under the fitness scheme S_c (eq. 32) exists, then it is unstable.*

Proof. For the new system in (u_i), we compute the Jacobian J_u (see SI for the derivation). We evaluate it at the hyperplane H given by $\hat{\sigma} = 0$ and $\hat{u}_i = 0$ for $i = 1, 2, 3$ and 4. Clearly, H contains the equilibrium.

The resulting 8×8 matrix is then in block diagonal form

$$J_u|_H = \begin{pmatrix} C_1 & \mathbf{0} \\ \mathbf{0} & C_2 \end{pmatrix}, \tag{42}$$

where each block is a square matrix of dimension 4. As the determinant of the full matrix is the product of the determinants of C_1 and C_2, we can analyze them separately.

The resulting expression for $\det(C_1)$ (see SI for the expression) is then evaluated at the coordinates $\hat{u}_i$, $i = 5, 6, 7$ and 8. After some simplification, it can be written as:

$$\det(C_1) = \frac{-s_{12}s_{21}(s_{11} - s_{13})^2(r - s_{11})(r - s_{12})(r - s_{13})(r - s_{21})}{(E + s_{12}F)(E + s_{21}F)} \tag{43a}$$

$$= -\frac{r^2(r - s_{12})(s_{11} - s_{13})^2(r - s_{21})}{(r - s_{11})(r - s_{13})}\hat{p}_1\hat{q}_1 < 0. \tag{43b}$$

18

This inequality holds, because condition (37) ensures that the ratio in inequality (42b) is positive and $-\hat{p}_1\hat{q}_1 < 0$, since (36) is an admissible equilibrium.

The determinant of C_1 is the product of the four eigenvalues and we can thus conclude that at least one of them has to be positive. This implies that the equilibrium is unstable. $\qquad\square$

Remark 4. *Neither Proposition 3 nor Theorem 5 states that the equilibrium it concerns is the only internal equilibrium. If there were two isolated equilibria E_1 and E_2 with $\hat{\hat{s}}_{E_1}$ and $\hat{\hat{s}}_{E_2}$ respectively then both Proposition 3 and Theorem 5 would remain valid if we set once $\hat{\hat{s}}_{E_1} = 0$ and once $\hat{\hat{s}}_{E_2} = 0$. However, in the SI we prove uniqueness of the internal equilibrium, if conditions (35c) and (35d) of Proposition 3 hold.*

Discussion

We have conducted a rather general mathematical analysis of haploid two-locus multiallele dynamics with constant selection and recombination. The model we use is the standard continuous-time model for selection on haploids with recombination, see eq. (1).

In the first section, we provide conditions for stability of hyperbolic monomorphisms and show that at least one of them is always stable. If both loci exhibit the same number of alleles, then there are at most n stable monomorphisms. If the number of alleles is unequal, then the smaller of them is the upper bound for the number of stable monomorphism. We also characterize the fitness matrices such that the only stable monomorphism is the same for every $r > 0$ and claim that it is also globally asymptotically stable for all r. But this remains unproven.

However, we use ideas about dominating strategies from game theory to prove global stability in section 2. If for a fixed background allele the haplotype formed with one allele is greater than with another allele and this holds for every background allele, than the fitter allele is the dominating allele. The dominated allele goes extinct. Subsequently, if one allele dominates all other alleles on the same locus, then they all go extinct and the dominating allele is fixed. Thus the two-locus multi-allele problem is reduced to a one-locus multi-allele problem. There, it is known that the allele with the highest fitness gets fixed. Similar to the game theoretic problem, we apply a quasi-concave Lyapunov function to prove global convergence. Speaking informally, this approach helps to get control over the terms of linkage disequilibria. These terms are introduced when we consider population genetic dynamics with more than one locus

and they are the reason why the usual Lyapunov methods that work for one-locus models break down. Potentially, other multi-locus convergence problems could be treated by means of quasi-concave Lyapunov functions.

Lemma 2 represents a very useful and intuitive characterization of polymorphisms in terms of two linear homogeneous systems of equations. Solvability of both systems in (15) is necessary and sufficient for the existence of internal equilibria. If the numbers of alleles at the two loci are different, then one of the systems is overdetermined and has, in general, no solution. However, in the degenerate case, where a solution exists, we showed that there is a manifold of solutions. This means for an unequal number of alleles at the loci, there is either no internal equilibrium or there are infinitely many. This immediate consequence is formalized in Theorem 3. If the two loci have the same number of alleles, we state necessary conditions for the existence of an isolated internal equilibrium in Proposition 2.

If there are either two or three alleles (with centrosymmetric fitnesses) at both loci, these general results on the existence of internal equilibria are used to establish uniqueness of the polymorphism and its instability if it exists.

With two alleles at each locus it is rather straightforward to prove uniqueness of the internal equilibrium by combining an index theorem by Hofbauer (1990) with Proposition 1, Theorem 3 and Proposition 2. The index theorem also entails that the equilibrium is a saddle point. This approach shows the uniqueness and instability of the polymorphism simultaeously and is not as technical and computationally difficult as that of Bank, Bürger and Hermisson (2012).

For three alleles at both loci, we need additional assumptions to establish an analogous result. We assume a centrosymmetric fitness scheme, generalizing that of the classical symmetric viability two-locus two-allele diploid model.

After scaling S_c such that $\hat{\hat{s}} = 0$, we determine the exact equilibrium coordinates, which also exhibit centrosymmetry. This allows us to apply a coordinate transformation that exploits this symmetry. With its help, instability of the internal equilibrium is shown. Note that in this proof we do not need the uniqueness of the equilibrium. However, in the SI, we show that the internal equilibrium for this centrosymmetric three-allele model is in fact unique.

Conditions (35) together with Proposition 1 imply that three monomorphisms are locally asymptotically stable. This in turn implies, by Theorem 4, that in each of the three two-allele subsystems spanned by these three monomorphisms a unique biallelic unstable polymorphism exists. Generalizing this argument in a rather speculative fash-

ion, we claim that $n^2 + \sum_{k \geq 2}^{n} \binom{n}{k} = n^2 + 2^n - (n+1)$ is the maximum number of isolated equilibria for system (1) with n alleles at both loci. There, we assume that for $k > 1$ alleles at both loci $\binom{n}{k}$ isolated equilibria exist. This is only proven for $n = 2$, because the centrosymmetry assumption in our treatment of $n = 3$ entails that four monomorphisms are not isolated, since Lemma 1 ensures the existence of edges where every point is an equilibrium. However, if we set $s_{23} = s_{21} + \epsilon$ and $s_{32} = s_{12} + \epsilon$, then local perturbation theory implies that there are $3^2 + \binom{3}{2} + \binom{3}{3} = 9 + 3 + 1 = 13$ isolated equilibria for ϵ sufficiently small. According to the claim we made above, this should hold for the general three allele case. The claim also implies uniqueness of the equilibrium with all n alleles present.

With the results presented here and the assumption that the trajectories converge to equilibrium points, it is clear that genetic variation, if it is maintained at two loci through haploid selection and recombination, only occurs with the same number (larger or equal 3) of alleles at both loci. If at one locus exactly two alleles occur or exactly three, which in addition are centrosymmetric, then genetic variation is always lost regardless of the number of alleles at the other locus. Ultimately, the population is fixed for one allele at each locus. Additionally, variation vanishes if the fitness scheme is of the form given in Theorem 2. The genotype with the maximal fitness becomes fixed.

Acknowledgements

References

Bank C., Bürger R., Hermisson J. 2012. Limits to parapatric speciation: Dobzhansky-Muller incompatibilities in a continent-island model. Genetics, 191, 845-863.

21

Feldman M.W. 1971. Equilibrium studies of two locus haploid populations with recombination. Theor. Popul. Biol. 2, 299-318.

Feldman M.W., Karlin S. 1970. Linkage and selection: Two locus symmetric viability model. Theor. Popul. Biol. 1, 39-71.

Hofbauer J. and Su J.J. 2016. Global stability of spatially homogeneous equilibria in migration-selection models. SIAM J. Appl. Math. 76, 578-597.

Hofbauer J. 1990. An index theorem for dissipative semiflows. Rocky Mountain Journal of Mathematics. 20, 1017-1031.

Kirzhner V.M. and Lyubich Y. 1997. Multilocus dynamics under haploid selection. J. Math. Biol. 35: 391-408

Nagylaki T. 1992. Introduction to Theoretical Population Genetics. Springer-Verlag

Novak S., Barton N.H. 2017. When does frequency-independent selection maintain genetic variation? Genetics 207: 653-668

Rutschman D. 1994. Dynamics of the two-locus haploid model. Theor. Popul. Biol. 45, 167-176.