A MUTATION-SELECTION-RECOMBINATION MODEL
IN POPULATION GENETICS

L. HATVANI, F. TOÓKOS, AND G. TUSNÁDY

Bolyai Institute, University of Szeged, Aradi vértanúk tere 1
H-6720 Szeged, Hungary
Helmholtz Zentrum München, German Research Center for Environmental Health
Institute for Biomathematics and Biometry Ingolstaedter Landstrasse 1
D-85764 Neuherberg, Germany
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences
P.O.Box 127, H-1364 Budapest, Hungary

ABSTRACT. We construct a new continuous time selection-mutation-recombination model for population dynamics, which describes the development of the distribution of the different gametes in the population. We show that cyclic mutation rates can result in stable and unstable limit cycles due to Hopf bifurcation. In addition, we give a qualitative characterization of the whole dynamics in the simplex, which is the phase space of the system. If only selection acts, then Fisher's Fundamental Law is valid: the mean fitness is a Lyapunov function and every orbit converges to some rest point.

1. INTRODUCTION

There are several models in the literature describing the change of gene frequencies in a randomly mating population (see, e.g., [1], [11], [14], [15], [18], [22], [26]). The classical selection and mutation model for 1 locus and n alleles, due to Fisher, Wright, and Haldane, is as follows. Denote the alleles by $A_1, \ldots, A_n$, and their relative frequencies by $x_1, \ldots, x_n$, respectively. Let w_{ij} be the fitness of the genotype $A_i A_j$ (the probability that an individual with genotype $A_i A_j$ survives into procreative age), $W = (w_{ij})_{i,j=1}^n$ the fitness matrix and ε_{ij} the mutation rate of $A_j \to A_i$ (the probability that the allele A_j mutates into A_i). Then

$$\varepsilon_{ij} \geq 0 \quad \text{and} \quad \sum_{i=1}^n \varepsilon_{ij} = 1 \qquad (j = 1, \ldots, n).$$

Introduce the notations $\mathbf{x} := \{x_1, x_2, \ldots, x_n\}^T$, $(W\mathbf{x})_i = \sum_{j=1}^n w_{ij} x_j$; the latter can be interpreted as the fitness of A_i. The gene distribution $x'_1, \ldots, x'_n$ of the next

generation is given by

$$x'_i = \frac{\sum_{j=1}^n \varepsilon_{ij} x_j (W\mathbf{x})_j}{W(\mathbf{x})} = \frac{\sum_{j,k=1}^n \varepsilon_{ij} x_j w_{jk} x_k}{W(\mathbf{x})} \qquad (i = 1, \ldots, n),$$

where $W(\mathbf{x}) = \mathbf{x} \cdot W\mathbf{x} = \sum_{r,s=1}^n w_{rs} x_r x_s$ is the mean fitness, the normalizing factor. From this discrete system we can obtain the following continuous model:

$$\dot{x}_i = \frac{\sum_{j,k=1}^n \varepsilon_{ij} x_j w_{jk} x_k}{W(\mathbf{x})} - x_i \qquad (i = 1, \ldots, n).$$

This equation was studied by Hadeler [16], [17] and Hofbauer [21] among others.

As Hofbauer and Sigmund [22, p. 266] write, "a biologically more satisfactory way" to derive a continuous time selection-mutation model might be the following. Let $M = (m_{ij})_{i,j=1}^n$ be the matrix of Malthusian fitness parameters, i.e., m_{ij} denotes the difference between the birth rate and the death rate of zygote $A_i A_j$. Assuming that mutation effects are small and they change the gene frequencies in a linear way, and the effects of selection and mutation are independent, one arrives at the model

$$(1.1) \qquad \dot{x}_i = x_i \Big((M\mathbf{x})_i - \mathbf{x} \cdot M\mathbf{x} \Big) + \sum_{j=1}^n (\varepsilon_{ij} x_j - \varepsilon_{ji} x_i) \qquad (i = 1, \ldots, n)$$

studied by Crow, Kimura [11] and Akin [1]. These two models are not independent of each other. As Hofbauer [21] showed, with proper parametrization, a connection can be established between them. The two systems are essentially equivalent provided that there is not much difference between the fitnesses of the genotypes and that the mutation rates are small.

Other selection-mutation equations include the haploid and diploid sequence-space model [3], [6], probability and stochastic models (see, e.g., [27], [4], [5]). Bürger [8], [9] gives a good overview of the various existing models and a detailed discussion of equation (1.1). For a discussion of the multi-locus models see, e.g., [10].

The purpose of this paper is to create a deterministic continuous time multi-locus selection - mutation - recombination model which coincides with the classical selection (mutation) model in the case when only selection (mutation) acts. Proceeding with the study of Hofbauer [21], in Section 3 we will show that circular mutation can result in limit cycles due to Hopf bifurcation. In the case of 3 alleles we present an example for the phenomenon that the circular mutation can make the system nonpersistent. In the case of 4 alleles there are 2 zygotes not involved in the cyclization in the sense that mutation cannot happen to gametes issuing from them. We show by an example that stability properties of the limit cycle depend on the relation of production coefficients h of zygotes. Namely, if production coefficients of the zygotes in the cycle are larger than those of the zygotes out of the cycle, then the limit cycle is stable, otherwise it is unstable. Besides we give a complete qualitative description of the dynamics

of the system around the equilibrium. We discuss the pure selection, no mutation case in Section 4 and show that in that case our model agrees with the Crow-Kimura model (1.1). In Section 5 we briefly discuss the case of several loci. For the reader's convenience and to make the paper self-contained, in the Appendix we reformulate some concepts and results from bifurcation theory used in the paper.

2. **THE MODEL**

Suppose that there are d loci with l_1, l_2, ..., l_d alleles on them. A gamete is a d-dimensional vector whose coordinates correspond to the alleles. Thus, there are $n = \prod_{i=1}^{d} l_i$ different types of gametes: $A_1, \ldots, A_n$. Denote the numbers of these gametes in the gene pool at time t by $m_1(t), \ldots, m_n(t)$, respectively. Let $m(t)$ be the total number of gametes in the gene pool, i.e., $m(t) := \sum_{i=1}^{n} m_i(t)$. Moreover, let $g(i)$ denote the death rate of the gamete of type A_i, and $h(i, j, k)$ denote the number of gametes of type A_i produced by the genotype (zygote) $A_j A_k$ per capita and per unit time. For example, $h(i, j, k)$ may be the product $h(i, j, k) = M_{jk}(i) \cdot w(j, k)$, where $w(j, k)$ is the Wrightian fitness of genotype $A_j A_k$ and $M_{jk}(i)$ is the probability that genotype $A_j A_k$ produces a gamete of type A_i. Thus $M_{jk}(i)$ includes the mutation and expresses also the linkage disequilibrium due to the recombination for $d > 1$.

Mating is random, so the number of individuals with genotype $A_j A_k$ at time t is equal to $\dfrac{m_j(t) \cdot m_k(t)}{m(t)}$. Hence the increment in the number of gametes of type A_i in a time interval Δt is

$$m_i(t + \Delta t) - m_i(t) = -\Delta t \cdot g(i) \cdot m_i(t) + \Delta t \cdot \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k) \cdot \frac{m_j(t) \cdot m_k(t)}{m(t)}.$$

Letting $\Delta t \to 0$ we obtain

$$(2.1) \qquad \dot{m}_i = -g(i)m_i + \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k) \cdot \frac{m_j \cdot m_k}{m} \quad (i = 1, \ldots, n).$$

Now, let us turn from numbers of gametes to their distribution. Let $x_i(t) = \dfrac{m_i(t)}{m(t)}$ be the relative frequency of A_i at time t (so $\sum_{i=1}^{n} x_i(t) \equiv 1$). Then

$$(2.2) \qquad \dot{x}_i(t) = \frac{\dot{m}_i(t)}{m(t)} - x_i(t) \cdot \frac{\dot{m}(t)}{m(t)}.$$

From equation (2.1) we obtain

$$\frac{\dot{m}_i(t)}{m(t)} = -g(i)x_i(t) + \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k)x_j(t)x_k(t).$$

Introducing the notation

$$S(\mathbf{x}) := \frac{\dot{m}(t)}{m(t)} = \frac{\sum_{i=1}^{n} \dot{m}_i(t)}{m(t)} = \sum_{i=1}^{n} \left(-g(i)x_i + \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k)x_j x_k \right),$$

from (2.2) we get

$$(2.3) \qquad \dot{x}_i = -g(i)x_i + \sum_{j=1}^{n} \sum_{k=1}^{n} h(i,j,k)x_j x_k - S(\mathbf{x})x_i \qquad (i = 1, \ldots, n).$$

This system of nonlinear equations is our most general model in this paper. We shall assume that $g(i) \geq 0$, $h(i,j,k) \geq 0$ for all $i, j, k = 1, \ldots, n$.

There are two basic requirements towards the system: for any initial distribution $x_1(0), x_2(0), \ldots, x_n(0)$ $(x_i(0) \geq 0,\ i = 1, 2, \ldots, n)$ the corresponding solution has to satisfy $x_1(t) + x_2(t) + \cdots + x_n(t) \equiv 1$, and $x_i(t) \geq 0$ for all i and for all $t > 0$. In other words, the phase space of (2.3) may be restricted to the simplex

$$\Delta^{n-1} := \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0\ (i = 1, 2, \ldots, n)\}.$$

The first part of the requirement can be easily verified. Indeed, set $U(\mathbf{x}) := \sum_{k=1}^{n} x_k$. Adding up the equations in (2.3) we get

$$\dot{U}(\mathbf{x}) = S(\mathbf{x}) - S(\mathbf{x})U(\mathbf{x}) = S(\mathbf{x}) \cdot (1 - U(\mathbf{x})).$$

$U(\mathbf{x}) \equiv 1$ is an equilibrium of this system, which implies that the hyperplane $x_1 + x_2 + \cdots + x_n = 1$ is invariant. This makes it possible to decrease the dimension of the system by 1 by considering the restriction of the system on the hyperplane.

The second part of the requirement, the nonnegativity of the coordinates, is also satisfied, as shown by the following lemma.

Lemma 2.1. *The set $B = \{\mathbf{x} \in \mathbb{R}^n :\ x_k > 0,\ k = 1, 2, \ldots, n\}$ is positively invariant with respect to the dynamics defined by (2.3).*

Proof. Consider an arbitrary solution $\mathbf{x}$ with $\mathbf{x}(0) \in B$. Since the negative members of the right-hand side of (2.3) contain a factor x_i, for every i $(1 \leq i \leq n)$ there exists a constant λ_i such that $\dot{x}_i(t)/x_i(t) > -\lambda_i$, i.e., the function $t \mapsto x_i(t)e^{\lambda_i t}$ is increasing. Therefore, this function cannot vanish, hence the solution cannot reach the boundary of B. $\qquad \square$

Remark 2.2. Due to continuous dependence on initial conditions the set $\{\mathbf{x} \in \mathbb{R}^n : x_k \geq 0,\ k = 1, 2, \ldots, n\}$ is also positively invariant.

Remark 2.3. In genetics certain gametes cannot be distinguishable from the point of view of selection and mutation. Due to this fact we can often reduce n, the dimension of the phase space of our basic system (2.3) introducing equivalence classes in the set of gametes. Let us denote the different classes by the natural numbers $1, 2, \ldots, f$; f can be called the "phenotype dimension" of the system $(f \leq n)$. For this reason we may consider the dimension n of the phase space as an arbitrary natural number not only in the one-locus cases, but also in the multi-locus ones.

In (2.3) $g(i)$'s are the death rates of gametes. One may say that the selection between gametes are involved into the model through the differences between the coefficients $h(i, j, k)$ $(j, k = 1, \ldots, n)$, so it is reasonable to investigate the case

$$g(i) = g = \text{const.} \quad (i = 1, \ldots, n).$$

Then (2.3) takes the following form:

$$S(\mathbf{x}) = -g + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k) x_j x_k,$$

$$\dot{x}_i = \sum_{j=1}^{n} \sum_{k=1}^{n} h(i, j, k) x_j x_k - \left(\sum_{p=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} h(p, j, k) x_j x_k \right) x_i;$$

g drops out of the model. To get nearer to standard forms of model equations in population genetics (see, e.g., (1.1)), introduce the following notations:

$$(2.4) \qquad H(i) := h(i, j, k)_{j=1,\ldots,n}^{k=1,\ldots,n} \in \mathbb{R}^{n \times n} \quad (i = 1, 2, \ldots, n), \quad H := \sum_{p=1}^{n} H(p);$$

($H(i)$ is an $n \times n$ matrix for each i). Then system (2.3) has the form

$$(2.5) \qquad\qquad \dot{x}_i = \mathbf{x} \cdot H(i)\mathbf{x} - x_i \{\mathbf{x} \cdot H\mathbf{x}\} \qquad (i = 1, \ldots, n).$$

Remark 2.4. Coefficients $h(i, j, k)$ $(i = 1, \ldots, n)$ involve the fitness of genotype $A_j A_k$. In certain cases one needs to separate the death rate of genotypes as a factor of fitness. Let g_{jk} denote the death rate of genotype $A_j A_k$. It is worth noticing that if all these death rates are equal to a constant g, and there is no selection *between gametes*, then $g(i) = g$ $(i = 1, \ldots, n)$ is also satisfied. In fact, the number of $A_j A_k$ zygotes dying in unit time is

$$g_{jk} \frac{m_j m_k}{m}.$$

Therefore, the number of gametes of type i dying in unit time is

$$\frac{m_i}{m} \left(\sum_{k=1}^{n} g_{ik} m_k \right),$$

so the death rate of the gamete of type i is

$$g(i) = \frac{1}{m} \left(\sum_{k=1}^{n} g_{ik} m_k \right).$$

In the case when all the zygotes have the same death rate, i.e., $g_{jk} = g$ $(j, k = 1, \ldots, n)$, then we have $g(i) = g$ $(i = 1, \ldots, n)$ and we get back system (2.5).

Let us formulate our equation (2.5) for the case when there is no recombination (e.g., there is only one locus). Denote by $b(j, k)$ the number of gametes produced

by a zygote of type $A_j A_k$ per unit time, and denote by $\gamma(i, j)$ the probability that a gamete j mutates into a gamete i. Obviously,

$$b(j, k) = b(k, j); \quad \sum_{i=1}^{n} \gamma(i, j) = 1 \qquad (j, k = 1, \ldots, n).$$

Then the number of gametes of type A_i produced from zygotes of type $A_j A_k$ per unit time is

$$\frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) \frac{m_j m_k}{m},$$

and the total number of gametes A_i produced per unit time is

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) \frac{m_j m_k}{m}.$$

It means that in (2.1) and, consequently, in (2.5) we have

$$
\begin{aligned}
h(i, j, k) &= b(j, k) \frac{\gamma(i, j) + \gamma(i, k)}{2}, \\
(H(i)\mathbf{x})_j &= \sum_{k=1}^{n} h(i, j, k) x_k = \sum_{k=1}^{n} \frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) x_k, \\
\mathbf{x} \cdot H(i)\mathbf{x} &= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) x_k \right) x_j \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) x_j x_k, \\
\sum_{p=1}^{n} \mathbf{x} \cdot H(p)\mathbf{x} &= \sum_{p=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\gamma(p, j) + \gamma(p, k)}{2} b(j, k) x_j x_k \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} b(j, k) x_j x_k.
\end{aligned}
$$

Therefore, the system is:

$$(2.6) \qquad \dot{x}_i = \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\gamma(i, j) + \gamma(i, k)}{2} b(j, k) x_j x_k - (\mathbf{x} \cdot B\mathbf{x}) x_i \quad (i = 1, \ldots, n).$$

This is a continuous time selection-mutation model. In contrast to (1.1), we did not make any restriction on selection and mutation.

Remark 2.5. Equation (2.6) gives, as a special case, the continuous time mutation equation studied by Hofbauer and Sigmund [22, Section 20.2] in the case of one locus. In fact, if there is no selection, only mutation, so $b(j, k) = b = \text{const.}$, then

$$b \sum_{j,k=1}^{n} \frac{\gamma(i, j)}{2} x_j x_k = \frac{b}{2} \sum_{j=1}^{n} \gamma(i, j) x_j,$$

and the model reads as follows:

$$\dot{x}_i = b \sum_{j=1}^n \gamma(i,j)x_j - bx_i \quad (i = 1, \ldots, n),$$

which is the same as the continuous time mutation equation in [22].

Remark 2.6. Equation (2.6) gives the classical continuous time selection equation for the case of one locus as a special case. In fact, if there is no mutation, only selection, so $\gamma(i,j) = \delta(i,j)$ (Kronecker delta), then

$$\sum_{j,k=1}^n \frac{\gamma(i,j)}{2}b(j,k)x_jx_k = \frac{1}{2}\left(\sum_{k=1}^n b(i,k)x_k\right)x_i,$$

and (2.6) has the form

(2.7) $$\dot{x}_i = x_i\{(B\mathbf{x})_i - \mathbf{x} \cdot B\mathbf{x}\} \qquad (i = 1, \ldots, n),$$

which is the well-known selection equation (see [22, p. 250]; B is the matrix of Malthusian fitness parameters).

3. MUTATION RESULTING IN PERIODIC ORBITS

Fisher's Fundamental Theorem of Natural Selection says: if only selection acts, then the mean fitness is increasing and the system tends to an equilibrium. It means that system (2.7) has no non-trivial periodic solution.

Hofbauer [21] generalized Fisher's theorem to selection mutation model (1.1) assuming that mutation rates depend only on the target genes and the effect of mutation is small in some sense. He also showed that (1.1) with more general mutation rates can have non-trivial periodic solutions. E. and M. Baake gave a necessary and sufficient condition for the mean fitness to be a Lyapunov function in the stochastic model [5]. Limit cycles appear in the diploid sequence-space model [6] due to Hopf bifurcation.

Our main goal in this section is to show that, in the one-locus case, with a special choice of the parameters we can achieve cyclic mutation in system (2.5) and obtain periodic solutions. The periodic solutions appear due to Hopf bifurcation, and in higher dimensions pitchfork bifurcations occur as well. We start our investigation with 3 alleles, then consider the 4-allele and, briefly, the 5-allele cases. We show by an example of 4 alleles that the stability properties of the periodic solution depends on the proportion of the mutational effect in the cycle to one outside of the cycle. As it was mentioned in Remark 2.3, these cases can occur in multi-locus models, too.

3 alleles. Consider the following choice of parameters in system (2.5):

$$h(1, 1, 1) = 1 \quad h(1, 2, 1) = a \quad h(1, 3, 1) = 1$$
$$h(2, 1, 2) = 1 \quad h(2, 2, 2) = 1 \quad h(2, 3, 2) = a$$
(3.1)
$$h(3, 1, 3) = a \quad h(3, 2, 3) = 1 \quad h(3, 3, 3) = 1$$
$$h(i, j, k) = a \text{ otherwise.}$$

Here $a > 0$ is the mutation parameter which determines the strength of the cycliza-tion. If a is small, then the cyclization is strong, if it is near 1, then the cyclization is weak. According to this choice the heterozygote diploid cells always prefer one of the alleles to the other one to inherit into the gamete. $A_1 A_2$ prefers A_2, $A_2 A_3$ prefers A_3 and $A_3 A_1$ prefers A_1. The homozygotes obviously produce their own type with the highest probability.

System (3.1) is seemingly not symmetric in the last two variables of h, how-ever this is just a cosmetic problem. One can introduce the coefficients $\tilde{h}(i, j, k) = (h(i, j, k) + h(i, k, j))/2$. The new coefficients $\tilde{h}(i, j, k)$ are symmetric in j and k, and the corresponding system is identical to the original. Moreover, since the sum $\sum_i h(i, j, k)$ is independent of j and k, there is no selectional difference between the genotypes.

We are mainly interested in the dynamics of the system on the biologically im-portant invariant $x_1 + x_2 + x_3 = 1$ plane.
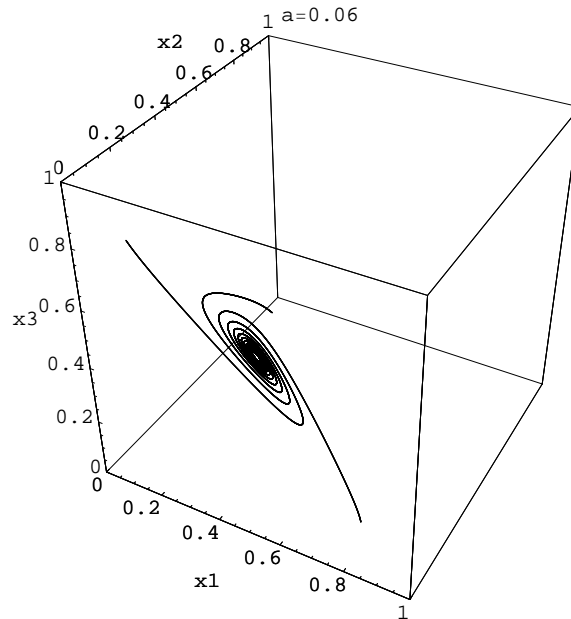

FIGURE 1. Asymptotically stable equilibrium of the system with (3.1)

Theorem 3.1. 1. If $a > \frac{1}{19}$, i.e, the cyclization is weak, then the equilibrium point $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ in system (2.5) with coefficients (3.1) is asymptotically stable with re-spect to the plane $\sum_{i=1}^{3} x_i = 1$ (see Figure 1).
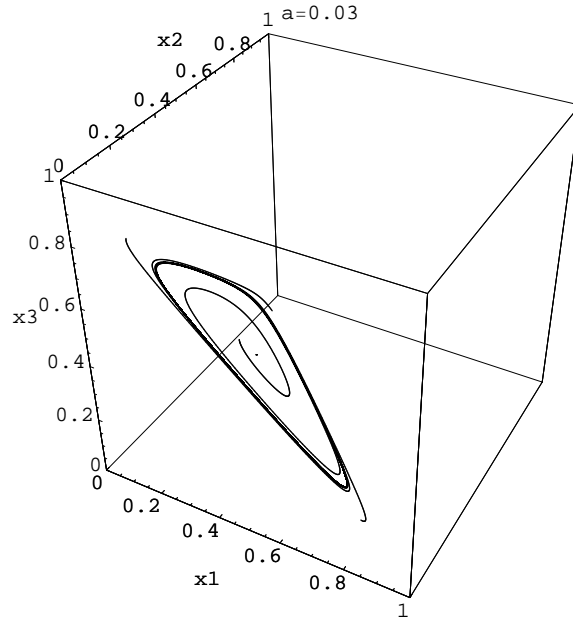
FIGURE 2. Limit cycle of the system with (3.1)

2. *If $0 < a < \frac{1}{19}$, i.e., the cyclization is strong, then system (2.5) with coefficients (3.1) has a nontrivial periodic solution on the plane $\sum_{i=1}^{3} x_i = 1$ in a neighborhood of the equilibrium point $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, and this periodic solution is asymptotically stable on the plane (see Figure 2).*

Proof. Perform the following linear transformation on system (2.3):

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = T \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \qquad T := \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix}.$$

Then the equilibrium point $(1/3, 1/3, 1/3)$ of the original system moves into $x = y = z = 0$, and the invariant plane into $z = 0$. After substituting $z = 0$ we obtain an equation of the form

$$(3.2) \qquad \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = F(x, y), \qquad (F(0, 0) = 0).$$

We can study the stability of the equilibrium position $x = y = 0$ by Lyapunov's first method [19]. The Jacobi matrix of F at $(0, 0)$ has the eigenvalues

$$(3.3) \qquad \lambda_{1,2} = \frac{1}{6}(1 - 19a) \pm i \frac{\sqrt{3}}{6}(a - 1).$$

Their real part is positive if $a \in I_1 := [0, 1/19)$ and negative when $a \in I_2 := (1/19, \infty)$. Therefore, according to Lyapunov's theorem, if $a \in I_2$, then the equilibrium position $x = y = 0$ is asymptotically stable, which proves the first statement of the theorem.

If $a \in I_1$ then the equilibrium point is unstable. We present two different ways to prove the second part of the theorem.

A. The triangle $A := \{x_1 + x_2 + x_3 = 1, \quad x_i \geq 0, \ i = 1, 2, 3\}$ is positively invariant in the original coordinates, so its transformed image, A', containing the origin, is also positively invariant in x and y. Therefore any solution starting from A' has a nonempty, compact ω-limit set, which must be different from the equilibrium point $x = y = 0$, because this equilibrium is unstable. It is easy to check with MAPLE that the equation $F(x, y) = 0$ has no other solution besides $x = y = 0$, hence system (3.2) has no other equilibrium. By the Poincaré-Bendixson theorem the ω-limit set must be an asymptotically stable limit cycle.

B. Formula (3.3) shows that the real part of the complex conjugate eigenvalues becomes 0 at the critical value $a = 1/19$, and the derivative of the real part with respect to a is not zero. This implies (see Theorem B in Appendix), that at this critical value a Hopf bifurcation occurs with the bifurcation parameter a. It means that, for sufficiently small ε and for $a \in (1/19 - \varepsilon, 1/19)$, system (3.2) has a limit cycle around the origin, and locally this is the only closed orbit.

To see the stability of the periodic solution we need to calculate its Floquet-exponent. By MAPLE we obtain the number $-189/152$, which means that the periodic solution is orbitally asymptotically stable with asymptotic phase. (Proof B works only for values of a near $1/19$.)

This completes the proof of the theorem. $\qquad\square$

Remark 3.2. This is only a local result. It states the existence of the periodic solution and the attraction only in a certain neighborhood. According to our computer tests the periodic solution seems to be unique and globally attractive, and the equilibrium point also seems to be globally attractive for $a > 1/19$. It would be interesting to find rigorous mathematical proofs for these results.

If we modify the coefficients in (3.1) a little bit, we can get an even richer structure. Consider the following choice of parameters:

(3.4)
$$
\begin{array}{ccc}
h(1,1,1) = 1 & h(1,2,1) = a & h(1,3,1) = 1 \\
h(2,1,2) = 1 & h(2,2,2) = 1 & h(2,3,2) = a \\
h(3,1,3) = a & h(3,2,3) = 1 & h(3,3,3) = 1 \\
h(2,1,1) = 0 & h(1,2,2) = 0 & h(1,3,3) = 0 \\
h(3,1,1) = 0 & h(3,2,2) = 0 & h(2,3,3) = 0 \\
\multicolumn{3}{c}{h(i,j,k) = a \text{ otherwise.}}
\end{array}
$$

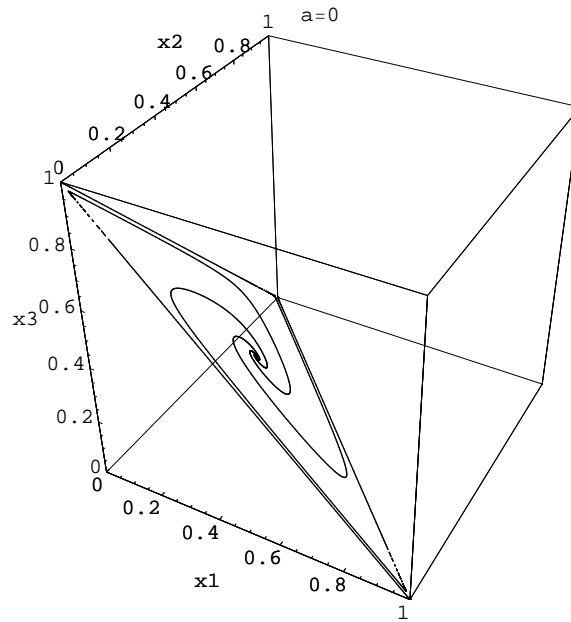For these values we have the following theorem which can be proved similarly to Theorem 3.1.

FIGURE 3. The boundary of the simplex is an attractor in case (3.4)

Theorem 3.3.   1. *If $a > 1$, then the equilibrium $(1/3, 1/3, 1/3)$ is an asymptotically stable node.*

2. *If $1/11 < a < 1$, then the equilibrium $(1/3, 1/3, 1/3)$ is an asymptotically stable focus.*

3. *If $0 < a < 1/11$, i.e., the cyclization is strong enough, then the system has a limit cycle.*

4. *If $a = 0$, then the extreme distributions*

$$
\begin{aligned}
P_1: \quad & x_1 = 1 \quad x_2 = 0 \quad x_3 = 0 \\
P_2: \quad & x_1 = 0 \quad x_2 = 1 \quad x_3 = 0 \\
P_3: \quad & x_1 = 0 \quad x_2 = 0 \quad x_3 = 1
\end{aligned}
$$

are equilibria, they are clockwise connected by the sides of the positively invariant triangle, and all trajectories starting from inside the triangle are cycling to the boundary (see Figure 3).

Remark 3.4. Part 4 of Theorem 3.3 expresses the possibility of the lack of the persistence of the alleles in the gene pool. In the real life, alike in computer simulations (see Figure 3), solutions with the properties in 4 cannot exist, all the trajectories end up in one of the vertices. It would be interesting to find conditions for system (2.5) (or for the most general system (2.3)) guaranteeing persistence for all alleles.

Remark 3.5. The solution curves on Figures 2–3 were created numerically by the use of MATHEMATICA.

4 alleles. We show that in this case both stable and unstable periodic solutions can appear. The easiest way to generalize system (3.1) is the following:

$$
\begin{aligned}
&h(1,1,1) = 1 \quad h(1,2,1) = a \quad h(1,3,1) = 1 \quad h(1,4,1) = 1 \\
&h(2,1,2) = 1 \quad h(2,2,2) = 1 \quad h(2,3,2) = a \quad h(2,4,2) = 1 \\
&h(3,1,3) = 1 \quad h(3,2,3) = 1 \quad h(3,3,3) = 1 \quad h(3,4,3) = a \\
&h(4,1,4) = a \quad h(4,2,4) = 1 \quad h(4,3,4) = 1 \quad h(4,4,4) = 1 \\
&\qquad\qquad\qquad h(i,j,k) = a \text{ otherwise.}
\end{aligned}
$$

(3.5)

This choice preserves the cyclic mutation with parameter a. Again, the heterozygote diploid cell $A_1 A_2$ prefers A_2, $A_2 A_3$ prefers A_3, $A_3 A_4$ prefers A_4, and $A_4 A_1$ prefers A_1. We also have two other heterozygotes, $A_1 A_3$ and $A_2 A_4$, which are not in this cycle. According to (3.5), they have no preference for either of the two alleles; e.g., $A_1 A_3$ produces both A_1 and A_3 with weight the same weight 1, i.e., with the weight of the preferred directions of the mutations in the cycle.

Just like in the 3-allele case, let us restrict ourselves to the $x_1 + x_2 + x_3 + x_4 = 1$ invariant hyperplane by performing the transformation

(3.6)
$$
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = U \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix} \qquad U = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -3 & 1 \end{pmatrix}
$$

This maps the hyperplane into the subspace $y_4 = 0$, and the equilibrium point $(1/4, 1/4, 1/4, 1/4)$ into the origin. We are interested in the dynamics inside the transformed image of the set $\{\mathbf{x} \in \mathbb{R}^4 : \sum x_i = 1, \ x_i \geq 0\}$. This is a positively invariant tetrahedron in the phase space (y_1, y_2, y_3). The vertices of the tetrahedron correspond to the marginal distributions when only one allele is present in the population. Inside the tetrahedron all the alleles are present, at the origin their proportions are equal.

Theorem 3.6. *Consider system* (2.5) *with the choice of coefficients* (3.5) *after transformation* (3.6).

1. *If $a > \frac{1}{17}$, then the origin is an asymptotically stable equilibrium position.*
2. *At $a = \frac{1}{17}$ a pitchfork bifurcation occurs.*
3. *If $0 < a < \frac{1}{17}$, then the origin is an unstable equilibrium position. There appear two further equilibria in the tetrahedron; they are asymptotically stable.*
4. *At $a = 0$ Hopf bifurcation occurs with periodic solutions appearing for positive values of a.*

Proof. The eigenvalues of the Jacobi-matrix at the origin are

$$
\frac{1}{4} - \frac{17}{4} a, \quad -4a \pm \frac{1}{4} i(-1 + a).
$$

This immediately implies that for $a > 1/17$ the origin is asymptotically stable. Using MAPLE we can see that in this case there is no other equilibrium point in the tetrahedron. At $a = 1/17$ the real eigenvalue becomes 0, while the real part of the complex conjugate eigenvalues is still negative. Since conditions (6.2) are satisfied, Theorem A in the Appendix applies with $k = 2$. It means that for $a < 1/17$ there are two new equilibrium positions besides the origin, the origin becomes unstable and the new equilibria are asymptotically stable. When a is between 0 and $1/17$, MAPLE gives that there is no other equilibrium in the tetrahedron; and as a tends to 0, the two new equilibria approach the edges $P_1 - P_3$ and $P_2 - P_4$ (where P_i denotes the vertex corresponding to the marginal distribution with only A_i present in the population).

At $a = 0$ the real part of the complex conjugate eigenvalue-pair becomes 0, while the imaginary part is non-vanishing. Since the derivative of the real part with respect to a is $-4 \neq 0$, hypothesis H3 from the Appendix is also satisfied. According to Theorem B, Hopf bifurcation occurs at $a = 0$. To determine the direction of the bifurcation (whether the periodic solutions appear for positive or negative a's) we need to calculate the second derivative of μ^* as described after Theorem B in the Appendix. MAPLE gives that the second derivative of μ^* with respect to ε is $\frac{2}{5} > 0$. Therefore, periodic solutions appear for positive a's and their periods are close to $2\pi/(1/4) = 8\pi$. $\qquad\qquad\square$
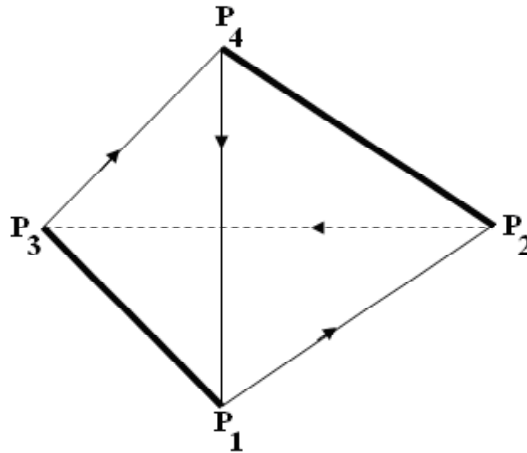

FIGURE 4. Dynamics on the edges at $a = 0$ with coefficients (3.5). The thick lines denote the equilibrium points.

The case $a = 0$ is particularly interesting. The vertices of the tetrahedron are equilibria, all edges and sides are invariant. This can be expected because if the cyclization is absolutely strong and some alleles are missing, these alleles cannot be produced by the others. The origin is unstable, the two stable equilibria reach the edges $P_1 - P_3$ and $P_2 - P_4$, all the points on these edges are equilibria. In this way every side of the tetrahedron contains an edge consisting of only equilibrium points.

The dynamics on the edges of the tetrahedron is shown on Figure 4 and the dynamics on each of the sides can be seen on Figure 5.
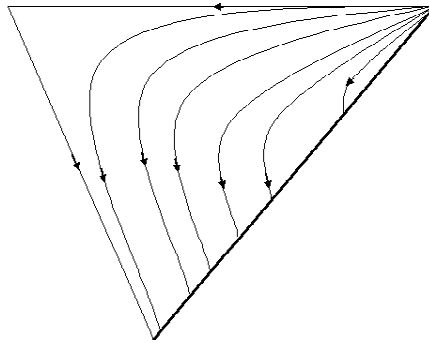

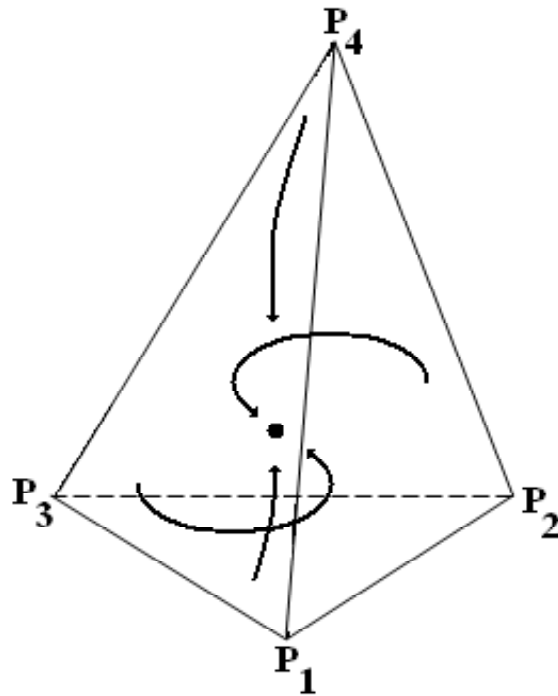FIGURE 5. Dynamics on each of the sides of the tetrahedron at $a = 0$ with coefficients (3.5)

FIGURE 6. The behavior of the system with coefficients (3.5) for $a > \frac{1}{17}$

We have the following conjecture based on our computer simulations:

Conjecture 3.7.   1. *For $a > \frac{1}{17}$ the origin is globally asymptotically stable in the tetrahedron (see Figure 6).*

   2. *For $0 < a < \frac{1}{17}$ the solutions tend to one of the two new equilibria with probability 1; the basins of these equilibria are separated by the two-dimensional stable manifold of the origin, which also contains the periodic orbit (see Figure 7).*

3. *At $a = 0$ the solutions tend to one of the equilibria on the edges $P_1 - P_3$ and $P_2 - P_4$, the basins of the edges are separated by the central manifold of the origin.*
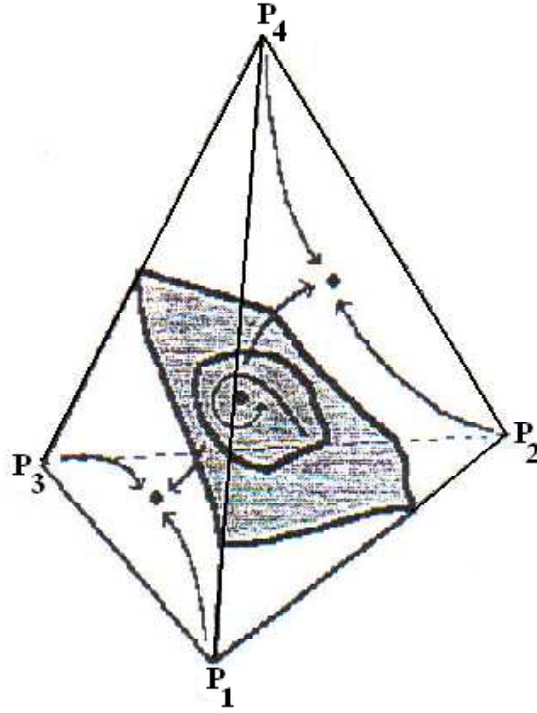

FIGURE 7. The behavior of the system with coefficients (3.5) for $0 < a < \frac{1}{17}$ based on numerical results

The appearing periodic solutions are unstable. However, we can modify the coefficients in such a way that we obtain stable limit cycles. Consider the following choice:

$$
\begin{array}{llll}
h(1,1,1) = 1 & h(1,2,1) = a & h(1,3,1) = a & h(1,4,1) = 1 \\
h(2,1,2) = 1 & h(2,2,2) = 1 & h(2,3,2) = a & h(2,4,2) = a \\
h(3,1,3) = a & h(3,2,3) = 1 & h(3,3,3) = 1 & h(3,4,3) = a \\
h(4,1,4) = a & h(4,2,4) = a & h(4,3,4) = 1 & h(4,4,4) = 1
\end{array}
$$

(3.7)

As opposed to the previous cases we extend the coefficients symmetrically, i.e., so that $h(i,k,j) = h(i,j,k)$. This way the cyclic mutational effect gets stronger and the heterozygotes A_1A_3 and A_2A_4 play lesser role (we can regard it as selectional disadvantage).

In this system the eigenvalues at the origin after the transformation are:
$$
-\frac{1}{4} - \frac{15}{4}a, \quad \frac{1}{4} - \frac{17}{4}a \pm \frac{1}{2}i(a-1).
$$

Theorem 3.8. *Consider system* (2.5) *with the choice of coefficients* (3.7) *after transformation* (3.6).

1. If $a > \frac{1}{17}$, then the origin is an asymptotically stable equilibrium position.
2. At $a = \frac{1}{17}$ Hopf bifurcation occurs. The periodic solutions are asymptotically stable and appear for $a < \frac{1}{17}$ on the unstable manifold of the origin.

The main difference is that now the two bifurcations switch places, so that when the periodic solutions appear the origin has a one-dimensional stable and a two-dimensional unstable manifold.

Proof. The proof is similar to that of Theorem 3.6. Now the second derivative of μ^* is $-4608/289 < 0$, so the periodic solutions appear for $a < \frac{1}{17}$. For the asymptotic stability, according to the Appendix, we have to calculate $V'''(0)$, where V is the displacement function of the Poincaré map on the central manifold. Performing the algorithm described in [23, Section 4A] we obtain that $V'''(0) = -144\pi < 0$, therefore the origin is a "vague attractor" and the periodic orbits are asymptotically stable.

An alternative, more direct proof of the stability can be given using Lyapunov's method (see the Appendix). □

Actually, we can say even more about the behavior of solutions of the system undergoing the Hopf bifurcation. In this case Chafee's theorem [23, Theorem 3A.1] applies (with one periodic orbit), so we can give a good description of the flow of the vector field near the bifurcation point.
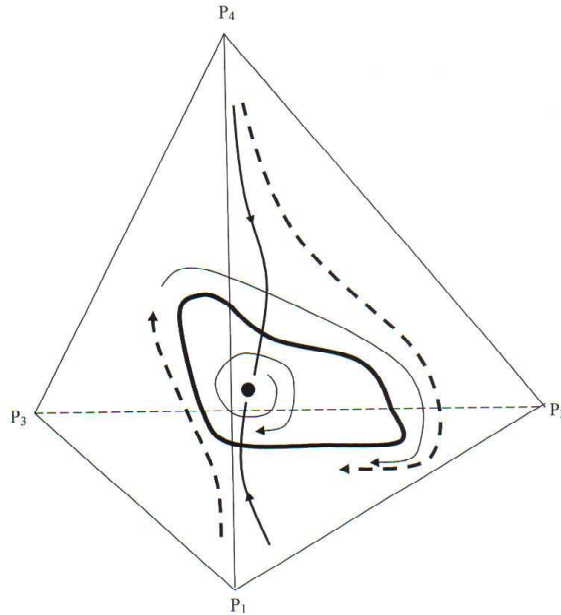

FIGURE 8. Dynamics of the system with coefficients (3.7) for $-\frac{1}{15} < a < \frac{1}{17}$

Theorem 3.9. *Consider system* (2.5) *with the choice of coefficients* (3.7) *after transformation* (3.6). *Let $B(r)$ denote the open ball centered at the origin with radius r.*

There exist numbers r_1, r_2 and ε such that $0 < r_2 \leq r_1$, $\varepsilon > 0$, and such that the following assertions are true (see Figure 8).

1. *For each $a \in \left[\frac{1}{17} - \varepsilon, \frac{1}{17}\right)$ there exists a closed orbit $\gamma(a)$ which lies inside a neighborhood $B(r(a))$, where $0 < r(a) \leq r_2$ and $r(a) \to 0$ as $a \to \frac{1}{17}^{-}$. Moreover, $\gamma(a)$ lies on a local two-dimensional integral manifold $M^2(a)$ homeomorphic to an open disc in $\mathbb{R}^2$ and containing the origin.*

2. *For each $a \in \left[\frac{1}{17} - \varepsilon, \frac{1}{17}\right)$ that part of $M^2(a)$ which lies inside $\gamma(a)$ is filled by solutions of (2.3) which approach the origin as $t \to -\infty$ and which, except for the equilibrium point at the origin, approach $\gamma(a)$ as $t \to +\infty$. No other solutions of (2.3) remain in $B(r_1)$ for all $t < 0$.*

3. *For each $a \in \left[\frac{1}{17} - \varepsilon, \frac{1}{17}\right)$ that part of $M^2(a)$ lying outside $\gamma(a)$ but contained in $B(r_2)$ is filled by solutions of (2.3) which remain in $M^2(a) \cap B(r_1)$ for all $t > 0$ and which approach $\gamma(a)$ as $t \to +\infty$.*

4. *For each $a \in \left[\frac{1}{17} - \varepsilon, \frac{1}{17}\right)$ there exists a one-dimensional local integral manifold $M^1(a)$ homeomorphic to an open interval and containing the origin, which consists of the origin and two different solutions which approach the origin as $t \to +\infty$. No other solutions in $B(r_2)$ approach the origin as $t \to +\infty$.*

5. *If for a given $a \in \left[\frac{1}{17} - \varepsilon, \frac{1}{17}\right)$, $\mathbf{y}(t)$ is a solution of the system with initial value $\mathbf{y_0} \in B(r_2)$, then $\mathbf{y}(t)$ remains in $B(r_1)$ for all $t > 0$. Moreover, if $\mathbf{y}(t) \not\to 0$ as $t \to +\infty$ (see item 4 above) then as $t \to +\infty$, $\mathbf{y}(t)$ approaches $\gamma(a)$, more precisely, its ω-limit set is $\gamma(a)$.*

Remark 3.10. From Theorem B in the Appendix and Theorem 3.1 in [23] we can also conclude that $\gamma(a)$ has period $\approx \frac{2\pi}{8/17} = \frac{17\pi}{4}$ and the radius grows like $\sqrt{\frac{1}{17} - a}$.
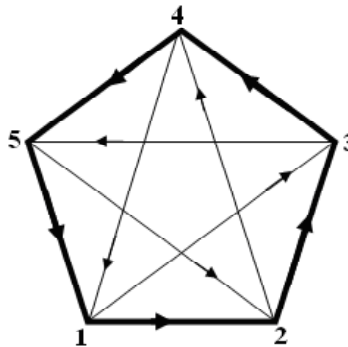

FIGURE 9. Cyclization with 5 alleles. First cycle: 1-2-3-4-5. Second cycle: 1-3-5-2-4.

5 alleles. Periodic orbits can be naturally produced by cyclic mutation for more alleles as well. In this case the picture may be even more complex, since with more alleles we can afford even more cycles. For example, if we have 5 alleles, it is possible

to break up the heterozygotes into two cycles: $A_1 - A_2 - A_3 - A_4 - A_5 - A_1$ and $A_1 - A_3 - A_5 - A_2 - A_4 - A_1$ (see Figure 9). Let a and b denote the mutation parameters of the two cycles, respectively. Thus, e. g., the heterozygote $A_1 A_2$ produces A_1 with probability a and A_2 with probability $1 - a$, $A_3 A_5$ produces A_3 with probability b and A_5 with $1 - b$. If we also introduce a general mutation rate c for the other mutations (to allow e. g., for $A_1 A_3$ to produce A_2), then we see that we can obtain Hopf bifurcation at $c = 0$. For certain values of the parameters the eigenvalues at the origin are two complex conjugate pairs, with the real parts changing sign at $c = 0$ simultaneously. Therefore we obtain two different families of cycles.

4. SELECTION

In this section we investigate the purely selectional case of the general model (2.3). First we rewrite the model into a matrix form similar to (2.5). We introduce, besides (2.4), the matrix notations

$$
G(i) := \begin{pmatrix} g(i) & \cdots & g(i) \\ & \vdots & \\ g(i) & \cdots & g(i) \end{pmatrix}, \quad G := \begin{pmatrix} g(1) & \cdots & g(1) \\ & \vdots & \\ g(n) & \cdots & g(n) \end{pmatrix}.
$$

Let us express $S(\mathbf{x})$ by these notations:

$$
\begin{aligned}
S(\mathbf{x}) &= -\sum_{p=1}^{n} g(p) x_p + \sum_{p=1}^{n} \sum_{j,k=1}^{n} h(p, j, k) x_j x_k \\
&= -\mathbf{x} \cdot G\mathbf{x} + \mathbf{x} \cdot H\mathbf{x} = \mathbf{x} \cdot (H - G)\mathbf{x}.
\end{aligned}
$$

Now the right-hand side of (2.3) reads as follows:

$$
\begin{aligned}
-g(i) x_i + \sum_{j,k=1}^{n} h(i, j, k) x_j x_k - S(\mathbf{x}) x_i \\
&= -(\mathbf{x} \cdot G(i)\mathbf{x}) x_i + \mathbf{x} \cdot H(i)\mathbf{x} - S(\mathbf{x}) x_i \\
&= \mathbf{x} \cdot H(i)\mathbf{x} - x_i \{\mathbf{x} \cdot (H - (G - G(i)))\mathbf{x}\},
\end{aligned}
$$

where the expression

$$
\vec{x} \cdot (G - G(i))\mathbf{x} = \sum_{p=1}^{n} g(p) x_p - g(i) = \overline{g}(\mathbf{x}) - g(i)
$$

is the deviation of the death rate $g(i)$ of the i-th gamete A_i from the *mean death rate* $\overline{g}(\mathbf{x}) := \sum_{p=1}^{n} g(p) x_p$. Therefore, equation (2.3) has the form

(4.1) $\dot{x}_i = \mathbf{x} \cdot H(i)\mathbf{x} - x_i \{\mathbf{x} \cdot (H - (G - G(i)))\mathbf{x}\}$ $(i = 1, \ldots, n)$.

Now let us turn to the purely selectional case. Then

(4.2) $h(i, j, k) = 0$ if $j \neq i$ and $k \neq i$,

i.e., matrix $H(i)$ introduced in (2.4) has nonzero elements only in its i-th row and i-th column. Therefore

$$\mathbf{x} \cdot H(i)\mathbf{x} = x_i \left\{ \sum_{k=1}^{n} h(i,i,k)x_k + \sum_{\substack{j=1 \\ j \neq i}}^{n} h(i,j,i)x_j \right\} = x_i(M\mathbf{x})_i,$$

where the matrix M of the fitnesses is defined by

$$m(i,j) := \begin{cases} 2h(i,i,j) & \text{if } i \neq j \\ h(i,i,i) & \text{if } i = j. \end{cases}$$

The mean fitness of the system is

$$\mathbf{x} \cdot M\mathbf{x} = \sum_{i=1}^{n} x_i(M\mathbf{x})_i = \sum_{i=1}^{n} \mathbf{x} \cdot H(i)\mathbf{x} = \mathbf{x} \cdot \left(\sum_{i=1}^{n} H(i)\mathbf{x} \right) = \mathbf{x} \cdot H\mathbf{x},$$

and

$$\mathbf{x} \cdot (H - (G - G(i)))\mathbf{x} = \mathbf{x} \cdot (M - G)\mathbf{x} + \mathbf{x} \cdot G(i)\mathbf{x} = \mathbf{x} \cdot (M - G)\mathbf{x} + (G\mathbf{x})_i.$$

Finally, equation (4.1) has the form

(4.3) $$\dot{x}_i = x_i\{((M - G)\mathbf{x})_i - \mathbf{x} \cdot (M - G)\mathbf{x}\} \qquad (i = 1, \ldots, n).$$

This is a replicator equation in the sense of [22, Section 7] (see also [7, Section 4.10]). The "payoff matrix" $M - G$ (the selectional matrix) is symmetric, i.e., (4.3) is a "partnership game" if and only if $g(1) = \ldots = g(n)$. In this case G is dropped from equation (4.3) and the equation gives the classical selection model (see (2.7) and [22, p. 250]); M is the matrix of Malthusian fitness coefficients, and Fisher's law holds.

In the other extreme case, when $m(i,j) = \text{const.} \ (i,j = 1, \ldots, n)$ and there are at least two different death rates, system (4.3) has the form

(4.4) $$\dot{x}_i = x_i \{\overline{g}(\mathbf{x}) - g(i)\} \qquad (i = 1, \ldots, n).$$

The derivative of the mean death rate $\overline{g}(\mathbf{x})$ with respect to (4.4) reads as follows:

$$\begin{aligned} (\overline{g})^{\cdot}(\mathbf{x}) &= \sum_{i=1}^{n} g(i)\dot{x}_i = \sum_{i=1}^{n} g(i)x_i \left(\sum_{j=1}^{n} g(j)x_j - g(i) \right) \\ &= \left(\sum_{i=1}^{n} g(i)x_i \right)^2 - \sum_{i=1}^{n} g(i)^2 x_i. \end{aligned}$$

By Jensen's inequality we have $(\overline{g})^{\cdot}(\mathbf{x}) \leq 0$, and the inequality is strong if $x_i > 0$ $(i = 1, \ldots, n)$. This means that $\overline{g}$ is a Lyapunov function to (4.4) and the set

$$M := \{\mathbf{x} \in \Delta^{n-1} : (\overline{g})^{\cdot}(\mathbf{x}) = 0\}$$

does not contain any interior point of the $(n-1)$-dimensional simplex $\Delta^{n-1} \subset \mathbb{R}^n$. By LaSalle's Invariance Principle [22, Theorem 2.6.1] the positive limit set of any solution of (4.4) is contained in M, so the boundary of Δ^{n-1} attracts all solutions.

To be more precise, we show that every solution converges to some rest point located on the boundary of Δ^{n-1}, and the limit rest point can be computed from the initial values of the solution.

Let $t \mapsto \mathbf{x}(t)$ be a solution of (4.4) and suppose that $x_p(0) > 0$ $(1 \le p \le n)$. By equation (4.4) we have

$$\frac{d}{dt}\left(\frac{x_i(t)}{x_p(t)}\right) = (g(p) - g(i))\frac{x_i(t)}{x_p(t)} \qquad (i = 1, \ldots, n).$$

Denoting by $\mathbf{x}^*$ a positive limit point of $\mathbf{x}$ we can conclude:

- If $g(i) = g(p)$, then

$$\frac{x_i^*}{x_p^*} = \frac{x_i(0)}{x_p(0)}.$$

- If $g(i) < g(p)$ and $x_i(0) > 0$, then $x_p^* = 0$.
- If $g(i) > g(p)$, then $x_i^* = 0$.

So we proved the following theorem, which is a generalization of a result of Akin and Hofbauer [2] (see also [22, Theorem 19.2.1]) and shows the extreme cases of viability and gametic selection (see, e.g., [20, Chapter 6]).

Theorem 4.1.   1. *If $g(1) = \ldots = g(n)$, i.e., there is no selection amongst gametes (there is only viability selection), then the mean fitness $\mathbf{x} \cdot M\mathbf{x}$ is increasing along solutions of the selection equation (4.3), and each orbit of this equation converges to some rest point as $t \to \infty$.*

2. *If $m(i, j) = 1$ $(i, j = 1, \ldots, n)$, i.e., there is no selection amongst zygotes (there is only gametic selection), but there are at least two different death rates amongst $g(1), \ldots, g(n)$, then the mean death rate $\sum_{i=1}^{n} g(i)x_i$ is decreasing along solutions of the selection equation (4.3), and each solution $\mathbf{x}$ of this equation converges to some rest point located on the boundary of the $n - 1$-dimensional simplex. Namely, if $x_p(0) > 0$ for some p $(1 \le p \le n)$, then gametes having death rates greater than $g(p)$ either are missing in the population, or die out asymptotically.*

Remark 4.2. If selection operates amongst neither gametes nor zygotes, then (4.3) has the form

$$\dot{x}_i = 0 \quad (i = 1, \ldots, n).$$

This is the Hardy–Weinberg Law in biology (see, e.g., [9], [20]) saying that the population without selection, mutation, and recombination is in a state where the gene probabilities remain unchanged from generation to generation, and gamete frequencies are the products of the corresponding gene frequencies provided that mating is random. In the cases treated in Theorem 4.1 this law holds but asymptotically.

5. RECOMBINATION

The multi-locus case is significantly more complicated due to the recombination. In the following we would only like to demonstrate the possibility of the recombination in the easiest case: two loci, two alleles on each locus, no mutation, no selection.

Let a_1, a_2 be the alleles on the first locus and b_1, b_2 on the second one. Let us denote the gametes as follows:

$$A_1 : \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \quad A_2 : \begin{pmatrix} a_1 \\ b_2 \end{pmatrix} \quad A_3 : \begin{pmatrix} a_2 \\ b_1 \end{pmatrix} \quad A_4 : \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}.$$

We shall assume that there is neither selection nor mutation, so the coefficients $h(i, j, k)$ are determined by Mendel's probability and the recombination. Let the probability of crossover between the two loci be r, i.e., for example, a diploid cell with genotype $A_1 A_4$ produces gametes A_2, A_3 with probability r and A_1, A_4 with probability $1 - r$. The case $r = 1/2$ corresponds to independent loci.

Theorem 5.1. *In the "two loci-two alleles" case with no selection or mutation the set of equilibrium points is the two-dimensional manifold defined by $D := x_1 x_4 - x_2 x_3 = 0$ (Wright manifold). This manifold is globally attractive in the phase space $\{\mathbf{x} \in \mathbb{R}^4 : \sum_{i=1}^4 x_i = 1, \ x_i \geq 0 \ (i = 1, 2, 3, 4)\}$: each orbit converges to some rest point in D as $t \to \infty$.*

Proof. A simple computation shows that system (2.5) has the form

$$(5.1) \qquad \begin{aligned} \dot{x}_i &= -rD \qquad (i = 1, 4), \\ \dot{x}_i &= rD \qquad (i = 2, 3), \end{aligned}$$

and the derivative of the "linkage disequilibrium coefficient" D with respect to (5.1) is

$$\dot{D} = \dot{x}_1 x_4 + x_1 \dot{x}_4 - \dot{x}_2 x_3 - x_2 \dot{x}_3 = -rD.$$

It immediately implies that the manifold $D = 0$ is invariant and all orbits tend to this manifold.

Obviously,

$$\Psi_1(\mathbf{x}) := x_1 + x_2, \quad \Psi_2(\mathbf{x}) := x_1 + x_3, \quad \Psi_3(\mathbf{x}) := x_1 - x_4$$

are first integrals of (5.1). It is easy to check that if $\mathbf{x}^* \in D$ is a point of the positive limit set of an arbitrary solution $\mathbf{x} = \widetilde{\mathbf{x}}(t)$ of (5.1) then

$$x_1^* = c_1 c_2, \quad x_2^* = c_1(1 - c_2), \quad x_3^* = c_2(1 - c_1),$$

where

$$c_1 := \Psi_1(\widetilde{\mathbf{x}}(0)), \quad c_2 = \Psi_2(\widetilde{\mathbf{x}}(0)), \quad c_3 = \Psi_3(\widetilde{\mathbf{x}}(0)).$$

This completes the proof. $\qquad \qquad \square$

Remark 5.2. The "linkage disequilibrium coefficient" D is a measure of the statistical dependence between the two loci. In fact, since the frequencies of the alleles a_1, a_2, b_1, b_2 are given by

$$x_1 + x_2, \quad x_3 + x_4, \quad x_1 + x_3, \quad x_2 + x_4,$$

respectively, it follows, e.g., that

$$D = x_1 x_4 - x_2 x_3 = x_1 - (x_1 + x_3)(x_1 + x_2) = \Pr\binom{a_1}{b_1} - \Pr(a_1)\Pr(b_1),$$

where $\Pr(\cdot)$ denotes the frequency in the gene pool. Consequently, $D = 0$ if and only if

$$\Pr\binom{a_i}{b_j} = \Pr(a_i)\Pr(b_j) \quad (i, j = 1, 2),$$

which says that the two loci are independent. In this case the population is said to be in linkage equilibrium. Accordingly, Theorem 5.1 can be interpreted in the following way. A population in linkage equilibrium remains in linkage equilibrium for ever. If $r > 0$, the linkage disequilibrium coefficient converges to zero as $t \to \infty$, i.e., recombination drives asymptotically the population into linkage equilibrium. This is a common phenomenon in multi-locus models (cf. [9]).

There are a lot of open questions related to the model and system (2.3). It would be interesting to investigate how large the dimension of the attractor with cyclic mutation in the one-locus case can be, and to see whether chaotic behavior can occur. In the other direction: under what condition does (2.3) have a single globally asymptotically stable equilibrium? In the multi-locus case with selection it would be important to find a Lyapunov function at least for a class of the coefficients. Last, but not least, the model also has its discrete and stochastic version, which are also worth studying.

6. APPENDIX

To make the paper self-contained, in this section we cite the theorems and algorithms used in Section 3.

Bifurcation theorems. Consider the equation

$$(6.1) \qquad \qquad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mu),$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mu \in \mathbb{R}$, and the bifurcation value of μ is μ_0. Assume that $\mathbf{f} \in C^1(E \times J, \mathbb{R}^n)$ and $\mathbf{f}(\cdot, \mu) \in C^3(E, \mathbb{R}^n)$, where E is an open subset in $\mathbb{R}^n$ and $J \subset \mathbb{R}$ is an interval containing μ_0.

Theorem A (Pitchfork Bifurcation). *Suppose that $\mathbf{f}(\mathbf{x_0}, \mu_0) = \mathbf{0}$ and that the $n \times n$ matrix $A = D_1\mathbf{f}(\mathbf{x_0}, \mu_0)$ (D_1 is the differential operator with respect to $\mathbf{x}$) has the simple eigenvalue $\lambda = 0$. Let $\mathbf{v}$ and $\mathbf{w}$ denote eigenvectors of A and A^T, respectively, belonging to the eigenvalue $\lambda = 0$. Furthermore, suppose that A has k eigenvalues with negative real part and $(n - k - 1)$ eigenvalues with positive real part and that the following conditions are satisfied:*

$$(6.2) \qquad \begin{aligned} \mathbf{w}^T\mathbf{f}_\mu(\mathbf{x}_0, \mu_0) &= 0, & \mathbf{w}^T[D_1\mathbf{f}_\mu(\mathbf{x}_0, \mu_0)\mathbf{v}] &\neq 0, \\ \mathbf{w}^T[D_1^2\mathbf{f}(\mathbf{x}_0, \mu_0)(\mathbf{v}, \mathbf{v})] &= 0, & \mathbf{w}^T[D_1^3\mathbf{f}(\mathbf{x}_0, \mu_0)(\mathbf{v}, \mathbf{v}, \mathbf{v})] &\neq 0, \end{aligned}$$

where $\mathbf{f}_\mu$ denotes the derivative of $\mathbf{f}$ with respect to μ. Then there is a smooth curve of equilibrium points of (6.1) in $\mathbb{R}^n \times \mathbb{R}$ passing through $(\mathbf{x}_0, \mu_0)$ and tangent to the hyperplane $\mathbb{R}^n \times \{\mu_0\}$. Depending on the signs of the expressions in (6.2), there is one equilibrium point of (6.1) near $\mathbf{x}_0$ when $\mu < \mu_0$ (or when $\mu > \mu_0$) and there are three equilibrium points of (6.1) near $\mathbf{x}_0$ when $\mu > \mu_0$ (or when $\mu < \mu_0$). The two new equilibrium points have stable manifolds of dimension $k + 1$, while the dimension of the stable manifold of the third (old) equilibrium point decreases from $k + 1$ to k during the bifurcation. [25, Section 4.2]

To formulate the Hopf Bifurcation Theorem on the appearance of periodic solutions we need some hypotheses on system (6.1).

- **(H1):** $\mathbf{f} \in C^k(E \times J, \mathbb{R}^n)$, $(k \geq 2)$, $\mathbf{f}(\mathbf{0}, \mu) = \mathbf{0}$);
- **(H2):** $A = D_1\mathbf{f}(\mathbf{0}, \mu_0)$ has simple (i.e., of algebraic multiplicity one) eigenvalues at $\pm i\omega_0$ ($\omega_0 > 0$), and no other eigenvalue of A belongs to $i\omega_0\mathbb{Z}$;
- **(H3):** $\text{Re}(D\sigma(\mu_0)) \neq 0$, where $\sigma(\mu)$ is the branch of eigenvalues of $D_1\mathbf{f}(\mathbf{0}, \mu)$ through $i\omega_0$ at $\mu = \mu_0$.

Theorem B (Hopf Bifurcation). *Let hypotheses (H1)–(H3) be satisfied. Then there exist C^{k-1}-functions $\varepsilon \mapsto \mu^*(\varepsilon)$, $\varepsilon \mapsto \omega^*(\varepsilon)$, and $\varepsilon \mapsto \mathbf{x}^*(\varepsilon)$, defined for ε sufficiently small, taking values in $\mathbb{R}$, $\mathbb{R}$ and $C(\mathbb{R}, \mathbb{R}^n)$, respectively, such that $\mathbf{x}^*(\varepsilon)$ is a $\frac{2\pi}{\omega^*(\varepsilon)}$ periodic solution of equation (6.1) with $\mu = \mu^*(\varepsilon)$. Moreover, μ^* and ω^* are even, $\mu^*(0) = \mu_0$, $\omega^*(0) = \omega_0$, $\mathbf{x}^*(-\varepsilon)(t) = \mathbf{x}^*(\varepsilon)(t + \frac{\pi}{\omega^*(\varepsilon)})$, and $\mathbf{x}^*(\varepsilon)(t) = \varepsilon\,Re(e^{i\omega_0 t}\mathbf{p}) + o(\varepsilon)$ for $\varepsilon \searrow 0$, where $\mathbf{p}$ is an eigenvector of A belonging to the eigenvalue $i\omega_0$.* [13, Theorem X.2.1]

We will also need to determine the direction of the bifurcation, i.e., whether the periodic solutions appear for $\mu > \mu_0$ or $\mu < \mu_0$. Since μ^* is even, it suffices to find the second derivative of μ^* with respect to ε at 0 (provided the derivative is not 0): if it is positive (negative), then the periodic solutions appear for $\mu > \mu_0$ ($\mu < \mu_0$). Of course we do not know μ^* explicitly, however, we can find the second derivative at 0 as follows (see [13, Appendix VIII]).

Choose $\mathbf{q}$ in $\mathbb{C}^n$ such that $A^T\mathbf{q} = i\omega_0\mathbf{q}$ and $\mathbf{q} \cdot \mathbf{p} = \sum_{i=1}^n q_i p_i = 1$. Introduce the complex number

$$\begin{aligned}
c = &\frac{1}{2}\mathbf{q} \cdot D_1^3\mathbf{f}(\mathbf{0}, \mu_0)(\mathbf{p}^2, \overline{\mathbf{p}}) \\
&+ \mathbf{q} \cdot D_1^2\mathbf{f}(\mathbf{0}, \mu_0)(-A^{-1}D_1^2\mathbf{f}(\mathbf{0}, \mu_0)(\mathbf{p}, \overline{\mathbf{p}}), \mathbf{p}) \\
&+ \frac{1}{2}\mathbf{q} \cdot D_1^2\mathbf{f}(\mathbf{0}, \mu_0)((2i\omega_0 - A)^{-1}(D_1^2\mathbf{f}(\mathbf{0}, \mu_0)(\mathbf{p}, \mathbf{p}), \overline{\mathbf{p}})),
\end{aligned}$$

where $\overline{\mathbf{p}}$ denotes the conjugate of $\mathbf{p}$. Then the second derivative of μ^* is

$$(6.3) \qquad\qquad -\frac{\mathrm{Re}(c)}{\mathrm{Re}\,\mathbf{q} \cdot D_{1,2}\mathbf{f}(\mathbf{0}, \mu_0)\mathbf{p}}.$$

Stability of the appearing periodic solutions. There are different ways to guarantee stability for the appearing cycles. One of them is to determine the sign of the displacement function of the Poincaré mapping defined on the 2-dimensional central manifold for $\mu = \mu_0$. This can be done, for example, following [23, Section 4A]. If the third derivative of the displacement function V is negative, then the origin is a so-called "vague attractor" and the periodic orbits are attractive.

A more direct proof of stability is based on the following

Theorem C. *Suppose that hypotheses (H1)–(H3) are satisfied and that the eigenvalues of A are contained in the left half plane (with the exception of $\pm i\omega_0$). Furthermore, let the origin be Lyapunov attractive for $\mu = \mu_0$. Then the periodic orbits obtained from Theorem B are attractive.* [23, Theorem 3B.4]

According to this theorem, in order to prove the asymptotic stability of the appearing periodic solutions, it is enough to show that the equilibrium of the system is asymptotically stable at the bifurcation value, when the linear approximation of the system has a pair of purely imaginary complex conjugate eigenvalues. Lyapunov developed a method to determine stability properties in this critical case (see [12]). Although it is simple and effective, the algorithm is not very wide-spread. Now we present this algorithm for the general case of an $n+2$-dimensional system, where $\mathbf{0}$ is the equilibrium point with n negative real eigenvalues and a pair of purely imaginary eigenvalues $\pm i\lambda$, $\lambda > 0$, then we apply it to complete the proof of Theorem 3.8.

Let the system be given in the following form:

$$(6.4) \qquad \begin{cases} \frac{dx}{dt} = -\lambda y + X \qquad \frac{dy}{dt} = \lambda x + Y \\ \frac{dx_s}{dt} = p_{s1}x_1 + \cdots + p_{sn}x_n + \alpha_s x + \beta_s y + X_s \quad (s = 1, \ldots, n), \end{cases}$$

where $x, y, x_1, \ldots, x_n$ are the state variables; $X, Y, X_1, \ldots, X_s$ are analytic functions whose Taylor expansion have no constant and linear terms; p_{sr}, α_s, and β_s $(s, r = 1, \ldots, n)$ are constants; the eigenvalues of the $n \times n$ matrix $P = (p_{i,j})_{i,j=1}^n$ are negative real numbers, and both X and Y vanish at $x = y = 0$.

The first step of the algorithm is to introduce polar coordinates r and ϑ instead of x, y:

$$x = r \cos \vartheta, \ \ y = r \sin \vartheta.$$

In the new variables the equations have the form

$$\frac{dr}{dt} = X \cos \vartheta + Y \sin \vartheta, \qquad \frac{d\vartheta}{dt} = \lambda + \Theta,$$

where Θ denotes an analytic function of variables $r, \vartheta, x_1, \ldots, x_n$ vanishing at $x_1 = \cdots = x_n = 0$ and having in its expansion with respect to $x_1, \ldots, x_n$ coefficients which are rational functions of $\sin \vartheta$ and $\cos \vartheta$. It follows from the properties of Θ that as long as r and x are sufficiently small, ϑ is a strictly increasing continuous function of t, thus we can change the independent variable t for ϑ, which yields the form

$$(6.5) \quad \begin{cases} \frac{dr}{d\vartheta} = rR \\ \frac{dx_s}{d\vartheta} = q_{s1}x_1 + \cdots + q_{sn}x_n + (a_s \cos \vartheta + b_s \sin \vartheta)r + Q_s \quad (s = 1, \ldots, n). \end{cases}$$

Here R and Q are the same type of functions as Θ; the functions Q_s will not contain in their expansions terms of lower order than 2 with respect to the variables r and $x_1, \ldots, x_n$, and $q_{sr} = \frac{p_{sr}}{\lambda}$, $a_s = \frac{\alpha_s}{\lambda}$, $b_s = \frac{\beta_s}{\lambda}$.

Since the right-hand sides of (6.5) are analytic, every solution

$$(r(\cdot\,; c), x_1(\cdot\,; c), \ldots, x_n(\cdot\,; c))$$

of the system with the initial condition $r(0; c) = c$ can be searched for sufficiently small c in the form

$$(6.6) \quad \begin{cases} r = c + u^{(2)}(\vartheta)c^2 + u^{(3)}(\vartheta)c^3 + \cdots \\ x_s = u_s^{(1)}(\vartheta)c + u_s^{(2)}(\vartheta)c^2 + u_s^{(3)}(\vartheta)c^3 + \cdots \quad (s = 1, \ldots, n), \end{cases}$$

where $u^{(k)}(0) = 0$, $u_s^{(k)}(0) = 0$ for all s and k.

In order to determine $u^{(k)}$ and $u_s^{(k)}$ we substitute (6.6) into (6.5) and compare the corresponding coefficients of c. We obtain:

$$\frac{du^{(k)}}{d\vartheta} = U^{(k)}$$

$$\frac{du_s^{(k)}}{d\vartheta} = q_{s1}u_1^{(k)} + \cdots + q_{sn}u_n^{(k)} + (a_s \cos \vartheta + b_s \sin \vartheta)u^{(k)} + U_s^{(k)},$$

where $U^{(k)}$ and $U_s^{(k)}$ are rational functions of $u^{(1)}, \ldots, u^{(k-1)}$ and $u_s^{(1)}, \ldots, u_s^{(k-1)}$, whose coefficients are rational functions of $\sin \vartheta$ and $\cos \vartheta$.

We determine $u^{(k)}$ and $u_s^{(k)}$ successively. Suppose there exists an m such that for $i < m$, $u^{(i)}$ is a periodic function of ϑ, but $u^{(m)}$ is not. Then $u^{(m)} = g\vartheta + v$, where g is a constant different from 0, and v is a finite sum of sines and cosines of multiples of the angle ϑ. The stability depends on this constant g.

Theorem D (Lyapunov). *If the number g produced by the algorithm above is negative, then the origin is an asymptotically stable equilibrium of* (6.4). *If g is positive, then the equilibrium is unstable.* [12, Chapter 8]

Based on the above the proof of the asymptotic stability of the periodic solutions in Theorem 3.8 goes as follows.

Proof. According to Theorem C, it is sufficient to show that the origin is asymptotically stable at $a = \frac{1}{17}$. Performing Lyapunov's algorithm described above, we obtain the following.

In our case $n = 1$ and $\lambda = \frac{8}{17}$. In system (6.4) $p_{11} = -\frac{8}{17}$, $\alpha_1 = \beta_1 = 0$. After changing for polar coordinates we have $q_{11} = -1$, $a_1 = b_1 = 0$. If we perform the following step of the algorithm, we get the following values:

$$U^{(1)} = 0 \quad U_1^{(1)} = 0$$
$$U^{(2)} = 0 \quad U_1^{(2)} = 3 - 12\sin\vartheta\cos\vartheta - 6\sin^2\vartheta$$
$$U^{(3)} = -72\sin\vartheta\cos\vartheta\cos(2\vartheta) + 36\sin^4\vartheta$$
$$-72\sin^2\vartheta - 36\cos^4\vartheta$$
$$u_1^{(1)} = 0$$
$$u^{(2)} = 0 \quad u_1^{(2)} = 3\cos(2\vartheta)$$
$$u^{(3)} = -36\vartheta + \tfrac{9}{2}\cos(4\vartheta)$$

Therefore $m = 3$, $v = \frac{9}{2}\cos(4\vartheta)$, $g = -36$, and Theorem D implies that the origin is asymptotically stable. $\qquad\square$

REFERENCES

[1] Akin E.: Hopf bifurcation in the two locus genetic model. Mem. Amer. Math. Soc., **44**, no. 284, AMS, Providence R.I. (1983)

[2] Akin E., Hofbauer J.: Recurrence of the unfit. Math. Biosci. **61**, 51–63 (1982)

[3] Baake E.: Diploid models on sequence space. J. Biol. Syst. **3**, 343–349 (1995)

[4] Baake E.: Mutation and recombination with tight linkage. J. Math. Biol. **42**, 455–488 (2001)

[5] Baake E., Baake M.: An exactly solved model for mutation, recombination and selection. Canad. J. Math. **55**, 3–41 (2003)

[6] Baake E., Wiehe, T.: Bifurcations in haploid and diploid sequence space models. J. Math. Biol. **35**, 321–343 (1997)

[7] Britton N. F.: Essential mathematical biology. Springer-Verlag, London (2003)

[8] Bürger R.: Mathematical properties of mutation-selection models. Genetica **102-103**, 279–298 (1998)

[9] Bürger R.: The mathematical theory of selection, recombination, and mutation. Wiley, Chichester (2000)

[10] Christiansen F. B.: Population genetics of multiple loci. Wiley, Chichester (1999)

[11] Crow J. F., Kimura M.: An introduction to population genetics theory. Harper and Row Publishers (1970)

[12] Chetayev N. G.: The stability of motion. Pergamon Press (1961)

[13] Diekmann O., Van Gils S. A., Verduyn Lunel S. M., Walther H.-O.: Delay equations. Functional-, complex- and nonlinear analysis. Springer-Verlag, New York (1995)

[14] Ewens W. J.: Population genetics. Methuen and Co., Ltd., London (1969)

[15] Farkas M.: Dynamical models in biology. Academic Press, London (2001)

[16] Hadeler K. P.: Mathematik für Biologen (in German). Springer-Verlag, Berlin (1974)

[17] Hadeler K. P.: Stable polymorphisms in a selection model with mutation. SIAM J. Appl. Math. **41**, 1–7 (1981)

[18] Haldane J. B. S.: A mathematical theory of natural and artifical selection. V. Proc. Camb. Philos. Soc. **23**, 838–844 (1927)

[19] Hale J. K.: Ordinary differential equations. Wiley-Interscience, New York-London-Sydney-Toronto (1969)

[20] Hartl D., Clark A. G.: Principles of population genetics. Sinauer Associates, Inc., Sunderland (1997)

[21] Hofbauer J.: The selection mutation equation. J. Math. Biol. **23**, 41–53 (1985)

[22] Hofbauer J., Sigmund K.: Evolutionary games and replicator dynamics. Cambridge University Press, Cambridge (1998)

[23] Marsden J. E., McCracken M.: The Hopf bifurcation and its applications. Applied Mathematical Sciences, vol. 19, Springer-Verlag, New York (1976)

[24] Moran P. A. P.: The statistical processes of evolutionary theory. Clarendon Press, Oxford (1962)

[25] Perko L.: Differential equations and dynamical systems. Texts in Applied Mathrematics, vol. 7, Springer-Verlag, New York (1996)

[26] Scheuer P. A. G., Mandel S. P. H.: An inequality in population genetics. Heredity **13**, 519–524 (1959)

[27] Steinsaltz D., Evans S. N., Wachter K. W.: A generalized model of mutation-selection balance with application to aging. Advances in Applied Mathematics **35**(4), 16–33 (2005)