Math. Model. Nat. Phenom.
Vol. 10, No. 5, 2015, pp. 84–99
DOI: 10.1051/mmnp/201510506

Genetic Recombination as a Chemical Reaction Network

S. Müller[1]*, J. Hofbauer[2]

[1]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, 4040 Linz, Austria

[2]Department of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria

Dedicated to the memory of the Viennese chemists and mathematicians Rudolf Wegscheider (1859–1935), Hilda Geiringer (1893–1973), and Friedrich J. M. Horn (1927–1978)

Abstract. The process of genetic recombination can be seen as a chemical reaction network with mass-action kinetics. We review the known results on existence, uniqueness, and global stability of an equilibrium (for all marginal frequencies and all recombination rate constants), from both the population genetics and the reaction networks point of view.

Keywords and phrases: population genetics, chromosomal crossover, detailed balance, Lyapunov function

Mathematics Subject Classification: 34D23, 37N25, 92D10, 92E20

1. Introduction

Population genetics deals with the change of gene frequencies in a population subject to the evolutionary forces of selection, recombination, and mutation [3]. Interestingly, all these processes can be viewed as chemical reactions with mass-action kinetics; in particular, mutation as first order, recombination as second order, and selection as third order reactions.

Recombination, or chromosomal crossover, is the exchange of genetic material between a pair of homologous chromosomes. It occurs when matching regions on matching chromosomes break and then reconnect to the other chromosome. Recombination is the major evolutionary force to produce and maintain variation in a (sexual) population.

In the simplest case, the process of recombination

$$\left|\begin{matrix}A\\B\end{matrix}\right| + \left|\begin{matrix}a\\b\end{matrix}\right| \overset{k}{\leftrightsquigarrow} \left|\begin{matrix}A\\b\end{matrix}\right| + \left|\begin{matrix}a\\B\end{matrix}\right|$$

*Corresponding author. E-mail: stefan.mueller@ricam.oeaw.ac.at

involves two loci, two alleles per locus (alleles A, a at the first and B, b at the second locus), and hence four gametes (AB, ab, Ab, aB). It can be written as a reversible chemical reaction

$$g_1 + g_4 \overset{k}{\Leftrightarrow} g_2 + g_3.$$

This very simple reaction network involves four species g_1, g_2, g_3, g_4, two complexes $g_1 + g_4$, $g_2 + g_3$, and has deficiency zero. The recombination $g_1 + g_4 \to g_2 + g_3$ occurs at the rate $k\, p_1 p_4$ determined by the rate constant times the frequencies of the reacting gametes. In the chemical setting, this corresponds to the assumption of mass-action kinetics. The dynamical system for the gamete frequencies amounts to

$$\dot{p}_1 = \dot{p}_4 = k\,(-p_1 p_4 + p_2 p_3) = -\dot{p}_2 = -\dot{p}_3.$$

There are conservation laws

$$(p_1 + p_2)^{\cdot} = (p_1 + p_3)^{\cdot} = (p_4 + p_2)^{\cdot} = (p_4 + p_3)^{\cdot} = 0$$

of which three are linearly independent. The equilibrium manifold is given by the conic

$$p_1 p_4 = p_2 p_3.$$

In each stoichiometric compatibility class, solutions converge to a unique equilibrium. This is usually proved by considering the so-called linkage disequilibrium function $D = p_1 p_4 - p_2 p_3$, which satisfies

$$\dot{D} = -k\, D\, (p_1 + p_2 + p_3 + p_4)\,.$$

Hence, D converges to 0.

Alternatively, one can try an ansatz for a Lyapunov function in the form $V(p) = \sum_{i=1}^{4} F(p_i)$. Then,

$$\dot{V}(p) = \sum_{i=1}^{4} F'(p_i)\, \dot{p}_i = -k\, D\, (F'(p_1) + F'(p_4) - F'(p_2) - F'(p_3))\,.$$

For the choice $F'(p) = \ln p$, we obtain

$$\begin{aligned} \dot{V}(p) &= -k\, D\, (\ln p_1 + \ln p_4 - \ln p_2 - \ln p_3) \\ &= -k\, (p_1 p_4 - p_2 p_3)\, (\ln(p_1 p_4) - \ln(p_2 p_3)) \le 0 \end{aligned}$$

due to the monotonicity of the logarithm. This shows that $V(p)$, with the convex function $F(p) = p \ln p - p$, is a global Lyapunov function.

The main object of this paper is to study the general recombination model (in continuous time) with an arbitrary number of genetic loci and arbitrary numbers of alleles at each locus. We will see that this leads to a chemical reaction network which is reversible and satisfies the Wegscheider conditions [23], since the rate constants of a reaction and its reverse coincide. Using results about complex-balancing equilibria by Horn and Jackson [11], and the graph-theoretical concept of reachability [22], we show the existence of a unique, positive, detailed-balancing equilibrium (in every stoichiometric compatibility class and for all rate constants). A generalization of the above entropy-like Lyapunov function allows us to prove global stability.

In population genetics, this general recombination model was studied by Geiringer [7] (for discrete time) and further by [3,17–19], thereby using linkage disequilibrium functions, induction on the number of loci, or cumulants. Proofs for global stability based on the entropy as Lyapunov function were independently given in [1] (for continuous time) and [13] (for discrete time). We present a proof using entropy and a proof by induction on the number of loci adapted from [13].

As the literature cited above, we follow a deterministic approach, although a stochastic approach is possible and required for small populations. A comparison of deterministic and stochastic aspects of recombination in the special case of single crossover can be found in [2].

2. Mathematical model

Notation: We denote the positive real numbers by $\mathbb{R}_{>0}$ and the non-negative real numbers by $\mathbb{R}_{\geq 0}$. For a finite index set I, we write $\mathbb{R}^I$ for the real vector space of formal sums $x = \sum_{i\in I} x_i\, i$ with $x_i \in \mathbb{R}$. Viewing the elements of I as indicator functions, a vector $x \in \mathbb{R}^I$ can be seen as a function $x\colon I \to \mathbb{R}$, and $x(i) = x_i$. For $x, y \in \mathbb{R}^I_{\geq 0}$, we define $x^y \in \mathbb{R}_{\geq 0}$ as $\prod_{i\in I} x_i^{y_i}$, where we set $0^0 = 1$. Given a matrix $Y \in \mathbb{R}^{I\times J}$, we denote by $Y^j \in \mathbb{R}^I$ the column vector indexed by $j \in J$. For $x \in \mathbb{R}^I_{\geq 0}$ and $Y \in \mathbb{R}^{I\times J}_{\geq 0}$, we define $x^Y \in \mathbb{R}^J_{\geq 0}$ as $(x^Y)_j = x^{Y^j} = \prod_{i\in I} x_i^{Y_{ij}}$ for $j \in J$.

2.1. Genetic recombination

We consider a finite set of loci $\mathcal{L}$ with $L = |\mathcal{L}| \geq 1$, finite sets of alleles $\mathcal{A}_i$ with $A_i = |\mathcal{A}_i| \geq 2$ for $i \in \mathcal{L}$, the resulting set of gametes

$$\mathcal{G} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_L,$$

the set of recombination patterns

$$\mathcal{P} = \{\{I, J\} \mid I \subseteq \mathcal{L},\, J = \mathcal{L} \setminus I\},$$

and the distribution of recombination rate constants

$$c\colon \mathcal{P} \to \mathbb{R}_{\geq 0}.$$

Clearly, there are $|\mathcal{G}| = \prod_{i\in\mathcal{L}} A_i$ gametes and $|\mathcal{P}| = 2^{L-1}$ recombination patterns, including the trivial recombination $\{\emptyset, \mathcal{L}\}$.

In a recombination following pattern $\{I, J\}$, alleles at loci I "stay together", but "get separated" from alleles at loci J. More precisely, for gametes $g, h \in \mathcal{G}$ and a recombination pattern $\{I, J\} \in \mathcal{P}$, we define $g_I h_J \in \mathcal{G}$ as

$$(g_I h_J)_i = \begin{cases} g_i, & \text{if } i \in I, \\ h_i, & \text{if } i \in J \end{cases}$$

and the resulting recombination as

$$\{g, h\} \overset{c(\{I,J\})}{\rightsquigarrow} \{g_I h_J, g_J h_I\} \tag{2.1}$$

with rate constant $c(\{I, J\})$. For $g_I h_J$, $g_J h_I \in \mathcal{G}$ and $\{I, J\} \in \mathcal{P}$, we find

$$(g_I h_J)_I (g_J h_I)_J = g \quad \text{and} \quad (g_I h_J)_J (g_J h_I)_I = h$$

and obtain the reverse recombination

$$\{g_I h_J, g_J h_I\} \overset{c(\{I,J\})}{\rightsquigarrow} \{g, h\}$$

which occurs with the same rate constant.

2.2. Dynamics

We introduce the distribution of gamete frequencies

$$p\colon \mathcal{G} \to \mathbb{R}_{\geq 0}$$

with $\sum_{g\in\mathcal{G}} p(g) = 1$. In fact, we identify the function $p\colon \mathcal{G} \to \mathbb{R}_{\geq 0}$ with the vector $p \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$ and write $p = \sum_{g\in\mathcal{G}} p(g)\, g$. In other words, we consider elements of the simplex

$$S_{\mathcal{G}} = \{p \in \mathbb{R}^{\mathcal{G}}_{\geq 0} \mid \textstyle\sum_{g\in\mathcal{G}} p(g) = 1\}.$$

Recombination (2.1) causes a change in gamete frequencies proportional to $g_I h_J + g_J h_I - g - h$ at the rate $c(\{I,J\})\, p(g)\, p(h)$ determined by the recombination rate constant times the frequencies of the recombining gametes. We formulate the dynamical system for the vector $p \in S_{\mathcal{G}}$ of all gamete frequencies, that is, for $p = \sum_{g\in\mathcal{G}} p(g)\, g$, by summing over all recombination partners and patterns:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{1}{2} \sum_{g,h\in\mathcal{G}} \sum_{\{I,J\}\in\mathcal{P}} c(\{I,J\})\, p(g)\, p(h)\, (g_I h_J + g_J h_I - g - h)\,. \tag{2.2}$$

The factor $\frac{1}{2}$ is needed since $\{g,h\} \rightsquigarrow \{g_I h_J, g_J h_I\}$ and $\{h,g\} \rightsquigarrow \{h_I g_J, h_J g_I\}$ represent the same recombination.

Clearly, recombination (2.1) causes a change in gamete frequencies only if $\{g,h\} \neq \{g_I h_J, g_J h_I\}$ and $c(\{I,J\}) > 0$. Further, different recombination patterns may give rise to the same recombination (with different rate constants, in general). In order to view recombination as a chemical reaction, we have to ensure inequality of left- and right-hand sides and positivity of rate constants. Moreover, we have to sum over the rate constants of all contributing recombination patterns which can be seen as reaction mechanisms.

2.3. Chemical reactions

In order to determine the set of chemical reactions arising from a process of genetic recombination, we start with the following observation: Given two gametes which differ on a subset of loci, all recombination patterns which agree on this subset give rise to the same recombination (with different rate constants, in general).

Let $K \subseteq \mathcal{L}$ be a subset of loci. The recombination pattern $\{I,J\} \in \mathcal{P}$ induces the subpattern $\{I,J\}_K \in \mathcal{P}_K$ where $\{I,J\}_K = \{I\cap K, J\cap K\}$ and

$$\mathcal{P}_K = \{\{I,J\} \mid I \subseteq K,\, J = K \setminus I\}.$$

We write $\{I,J\} \geq \{I,J\}_K$ and, for simplicity, $\mathcal{P}_K^* = \mathcal{P}_K \setminus \{\{\emptyset, K\}\}$. The set of all recombination subpatterns amounts to

$$\breve{\mathcal{P}} = \bigcup_{K\subseteq\mathcal{L}} \mathcal{P}_K,$$

and we introduce the distribution of cumulative recombination rate constants

$$\begin{aligned} \breve{c}\colon \breve{\mathcal{P}} &\to \mathbb{R}_{\geq 0}, \\ \{I,J\} &\mapsto \sum_{\substack{\{I',J'\}\in\mathcal{P}:\\ \{I',J'\}\geq\{I,J\}}} c(\{I',J'\}). \end{aligned}$$

For a subpattern $\{I,J\}$ with $J = K\setminus I$ and $K \subseteq \mathcal{L}$, the cumulative rate constant $\breve{c}(\{I,J\})$ sums over all patterns which agree with $\{I,J\}$ on K.

An important parameter is the cumulative rate constant for the recombination subpattern $\{\{i\},\{j\}\}$, that is, for the case that an allele at locus i gets separated from an allele at locus j. We assume that $\breve{c}(\{\{i\},\{j\}\}) > 0$ for all pairs of loci i, j. Otherwise, the two loci can be identified.

To explicitly state a chemical reaction arising from a recombination pattern and a pair of gametes, we introduce the set $\Delta(g,h) = \{i \in \mathcal{L} \mid g_i \neq h_i\}$ for gametes $g, h \in \mathcal{G}$. In genetic terms, g and h are heterozygous at the subset of loci $\Delta(g,h)$ and homozygous otherwise.

Now, gametes $g,h \in \mathcal{G}$ and a recombination pattern $\{I,J\} \in \mathcal{P}$ give rise to a reaction mechanism, only if $|\Delta(g,h)| \geq 2$, $\{I,J\}_{\Delta(g,h)} \neq \{\emptyset, \Delta(g,h)\}$, and $c(\{I,J\}) > 0$. In other words, only if the gametes are heterozygous at two or more loci, if the subpattern induced on these loci is non-trivial, and if the recombination rate constant is non-zero. Every pattern $\{I',J'\} \in \mathcal{P}$ with $\{I',J'\} \geq \{I,J\}_{\Delta(g,h)}$ and

$c(\{I', J'\}) > 0$ gives rise to a mechanism for the same reaction, that is, to the same recombination (with different rate constant, in general). The effect of all such patterns is summarized in the chemical reaction

$$g + h \xrightarrow{k} g_I h_J + g_J h_I \tag{2.3}$$

with rate constant $k \equiv k(g + h \to g_I h_J + g_J h_I) = \breve{c}(\{I, J\}_{\Delta(g,h)}) > 0$. Note that $g + h$ stands for $\{g, h\}$ such that $g + h$ equals $h + g$. The reverse reaction

$$g_I h_J + g_J h_I \xrightarrow{k} g + h$$

occurs with the same rate constant.

2.4. Reaction networks

A chemical reaction network $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ consists of three finite sets: a set $\mathcal{S}$ of species, a set $\mathcal{C} \subset \mathbb{R}^{\mathcal{S}}_{\geq 0}$ of complexes, and a set $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$ of reactions. Complexes are the left- and right-hand sides of reactions. A complex $y \in \mathcal{C}$ can be seen as a formal sum of species $y = \sum_{s \in \mathcal{S}} y_s\, s$, where y_s is the stoichiometric coefficient of species s. For a reaction $(y, y') \in \mathcal{R}$, we write $y \to y'$. It is required that each complex appears in at least one reaction and that there are no reactions of the form $y \to y$.

A chemical reaction network $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ together with a vector of rate constants $k \in \mathbb{R}^{\mathcal{R}}_{>0}$ gives rise to a weighted directed graph with complexes as nodes, reactions as edges, and rate constants as labels. The connected components of this graph are called linkage classes. (Note that linkage classes have nothing to do with genetic linkage.) A network is called weakly reversible if every component is strongly connected, that is, if there exists a directed path from each node to every other node in the component.

In the process of genetic recombination, the reacting species are the gametes, that is, $\mathcal{S} = \mathcal{G}$. Every complex $g + h$ is a formal sum (with stoichiometric coefficients equal to one) of two gametes g and h, which differ at two or more loci, and every reaction $g + h \to g_I h_J + g_J h_I$ arises from a pair of gametes and a recombination pattern $\{I, J\}$, under the conditions specified in the previous subsection. The set of all chemical reactions (with corresponding rate constants) amounts to

$$\begin{aligned} \mathcal{R} = \Big\{ g + h \xrightarrow{k} g_I h_J + g_J h_I \;\Big|\; & g, h \in \mathcal{G},\ \{I, J\} \in \mathcal{P} \text{ with } |\Delta(g, h)| \geq 2, \\ & \{I, J\}_{\Delta(g,h)} \neq \{\emptyset, \Delta(g, h)\}, \text{ and} \\ & k \equiv \breve{c}(\{I, J\}_{\Delta(g,h)}) > 0 \Big\}. \end{aligned} \tag{2.4}$$

For each reaction $y \to y'$ we have its reverse $y' \to y$, and both occur with the same rate constant. Hence, we can combine them in the reversible reaction $y \Leftrightarrow y'$, which we identify with $y' \Leftrightarrow y$, and write $k(y \Leftrightarrow y')$ for $k(y \to y') = k(y' \to y)$. From (2.4), we obtain the set of all reversible reactions

$$\mathcal{R}_{\Leftrightarrow} = \{ y \overset{k}{\Leftrightarrow} y' \mid (y \xrightarrow{k} y') \in \mathcal{R} \} \tag{2.5}$$

and the set of all complexes

$$\mathcal{C} = \{ y \mid (y \to y') \in \mathcal{R} \}. \tag{2.6}$$

In the examples and schemes below, we determine the set of all (reversible) reactions in another way. We first iterate over subsets of two or more loci and then over non-trivial subpatterns on these loci:

$$\begin{aligned} \mathcal{R}_{\Leftrightarrow} = \Big\{ g + h \overset{k}{\Leftrightarrow} g_I h_J + g_J h_I \;\Big|\; & K \subseteq \mathcal{L} \text{ with } |K| \geq 2, \\ & g, h \in \mathcal{G} \text{ with } |\Delta(g, h)| = K, \\ & \{I, J\} \in \mathcal{P}^*_{\Delta(g,h)} \text{ with } k \equiv \breve{c}(\{I, J\}) > 0 \Big\}. \end{aligned}$$

Thereby, we extend the definition of $g_I h_J$ to the subpattern $\{I, J\} \in \mathcal{P}_{\Delta(g,h)}$ in the obvious way: $(g_I h_J)_i = g_i$ for $i \in I$, $(g_I h_J)_i = h_i$ for $i \in J$, and $(g_I h_J)_i = g_i = h_i$ for $i \in \mathcal{L} \setminus (I \cup J)$.

Finally, we consider the graph arising from the reaction network $(\mathcal{S}, \mathcal{C}, \mathcal{R})$, in particular, its linkage classes. We observe that species (gametes) consist of alleles and complexes (pairs of gametes) contain two alleles at each locus. Since reactions separate alleles, but do not consume or produce them, only complexes which contain the same alleles are connected by a reaction. Moreover, if complexes $g + h$ and $g' + h'$ are connected by a reaction then $\Delta(g, h) = \Delta(g', h')$, and every subpattern $\{I, J\} \in \mathcal{P}^*_{\Delta(g,h)}$ which gives rise to a reaction involving $g + h$ gives rise to a reaction involving $g' + h'$, and vice versa. Hence every linkage class is a symmetric graph. If no reaction is precluded by a zero rate constant, then every linkage class is a complete graph, characterized by two (possibly identical) alleles at each locus.

2.5. Examples and schemes

We consider examples of genetic recombination for small numbers of loci and alleles and depict the corresponding chemical reaction networks as graphs. Further, we present schemes for arbitrary numbers of loci and compute the resulting numbers of linkage classes, complexes, and reversible reactions. (The numbers of linkage classes and complexes determine the deficiency of a network, cf. Section 4.) For simplicity, we assume that no reaction is precluded by a zero rate constant. In this case, all linkage classes are complete graphs.

For each locus $i \in \mathcal{L} = \{1, \dots, L\}$, we consider the set of alleles $\mathcal{A}_i = \{0, 1, \dots, A_i - 1\}$. Instead of $c(\{I, J\})$ we write $c(I)$ and further omit the set brackets, e.g., $c(\{\{1\}, \mathcal{L} \setminus \{1\}\}) \equiv c(\{1\}) \equiv c(1)$.

Example 2.1 ($L = 2$ loci with $A_1 = A_2 = 2$ alleles)**.**

$$\begin{vmatrix}0\\0\end{vmatrix} + \begin{vmatrix}1\\1\end{vmatrix} \overset{c(1)}{\Longleftrightarrow} \begin{vmatrix}1\\0\end{vmatrix} + \begin{vmatrix}0\\1\end{vmatrix}$$

The graph has $l = 1$ linkage class, $m = 2$ complexes, and $r = 1$ reversible reaction.

Example 2.2 ($L = 3$ loci with $A_1 = A_2 = A_3 = 2$ alleles)**.**

$$\begin{vmatrix}0\\0\\0\end{vmatrix} + \begin{vmatrix}1\\1\\0\end{vmatrix} \overset{c(1)+c(2)}{\Longleftrightarrow} \begin{vmatrix}1\\0\\0\end{vmatrix} + \begin{vmatrix}0\\1\\0\end{vmatrix} \quad , \quad \begin{vmatrix}0\\0\\1\end{vmatrix} + \begin{vmatrix}1\\1\\1\end{vmatrix} \overset{c(1)+c(2)}{\Longleftrightarrow} \begin{vmatrix}1\\0\\1\end{vmatrix} + \begin{vmatrix}0\\1\\1\end{vmatrix}$$

$$\begin{vmatrix}0\\0\\0\end{vmatrix} + \begin{vmatrix}1\\0\\1\end{vmatrix} \overset{c(1)+c(3)}{\Longleftrightarrow} \begin{vmatrix}1\\0\\0\end{vmatrix} + \begin{vmatrix}0\\0\\1\end{vmatrix} \quad , \quad \begin{vmatrix}0\\1\\0\end{vmatrix} + \begin{vmatrix}1\\1\\1\end{vmatrix} \overset{c(1)+c(3)}{\Longleftrightarrow} \begin{vmatrix}1\\1\\0\end{vmatrix} + \begin{vmatrix}0\\1\\1\end{vmatrix}$$

$$\begin{vmatrix}0\\0\\0\end{vmatrix} + \begin{vmatrix}0\\1\\1\end{vmatrix} \overset{c(2)+c(3)}{\Longleftrightarrow} \begin{vmatrix}0\\1\\0\end{vmatrix} + \begin{vmatrix}0\\0\\1\end{vmatrix} \quad , \quad \begin{vmatrix}1\\0\\0\end{vmatrix} + \begin{vmatrix}1\\1\\1\end{vmatrix} \overset{c(2)+c(3)}{\Longleftrightarrow} \begin{vmatrix}1\\0\\1\end{vmatrix} + \begin{vmatrix}1\\1\\0\end{vmatrix}$$

$$\begin{array}{ccc}
\begin{vmatrix}0\\1\\0\end{vmatrix} + \begin{vmatrix}1\\0\\1\end{vmatrix} & \overset{c(1)}{\Longleftrightarrow} & \begin{vmatrix}1\\1\\0\end{vmatrix} + \begin{vmatrix}0\\0\\1\end{vmatrix} \\
\Big\Updownarrow{\scriptstyle c(2)} \quad \overset{c(3)}{\Longleftrightarrow} & & \Big\Updownarrow{\scriptstyle c(2)} \\
\begin{vmatrix}0\\0\\0\end{vmatrix} + \begin{vmatrix}1\\1\\1\end{vmatrix} & \underset{c(1)}{\Longleftrightarrow} & \begin{vmatrix}1\\0\\0\end{vmatrix} + \begin{vmatrix}0\\1\\1\end{vmatrix}
\end{array}$$

The graph has $l = 7$ linkage classes, $m = 16$ complexes, and $r = 12$ reversible reactions. The last class has $2^{L-1} = 4$ complexes and $\binom{2^{L-1}}{2} = \binom{4}{2} = 6$ reactions.

Scheme 2.3 ($L \geq 2$ loci with $A_i = 2$ alleles, $i = 1, \ldots, L$)**.** .

As already mentioned, every linkage class is determined by two (possibly identical) alleles at each locus. We sum over subsets of two or more loci having different alleles. Given a subset of cardinality k, alleles can be chosen in $1^k\, 2^{L-k}$ ways. Hence,

$$\begin{aligned} l &= \sum_{k=2}^{L} \binom{L}{k} 2^{L-k} \\ &= 3^L - (2^L + L\, 2^{L-1}) \\ &= 3^L - 2^{L-1}(2 + L). \end{aligned}$$

Further, every linkage class with k loci having different alleles contains 2^{k-1} complexes and $\binom{2^{k-1}}{2}$ reversible reactions. Hence,

$$\begin{aligned} m &= \sum_{k=2}^{L} \binom{L}{k} 2^{L-k}\, 2^{k-1} \\ &= 2^{L-1} \textstyle\sum_{k=2}^{L} \binom{L}{k} \\ &= 2^{L-1}(2^L - (1 + L)) \end{aligned}$$

and

$$\begin{aligned} r &= \sum_{k=2}^{L} \binom{L}{k} 2^{L-k} \binom{2^{k-1}}{2} \\ &= \textstyle\sum_{k=2}^{L} \binom{L}{k}\, 2^{L-k}\, 2^{k-1}(2^{k-1} - 1)\, 2^{-1} \\ &= \textstyle\sum_{k=2}^{L} \binom{L}{k}\, (2^{L-3}\, 2^k - 2^{L-2}) \\ &= 2^{L-3}(3^L - (1 + L\, 2)) - 2^{L-2}(2^L - (1 + L)) \\ &= 2^{L-3}(3^L - 1) - 2^{L-2}(2^L - 1) \\ &= 2^{L-3}(3^L - 2^{L+1} + 1). \end{aligned}$$

Example 2.4 ($L = 2$ loci with $A_1 = 2$ and $A_2 = 3$ alleles, respectively)**.**

$$\left|\begin{matrix}0\\0\end{matrix}\right| + \left|\begin{matrix}1\\1\end{matrix}\right| \overset{c(1)}{\Longleftrightarrow} \left|\begin{matrix}1\\0\end{matrix}\right| + \left|\begin{matrix}0\\1\end{matrix}\right|$$

$$\left|\begin{matrix}0\\0\end{matrix}\right| + \left|\begin{matrix}1\\2\end{matrix}\right| \overset{c(1)}{\Longleftrightarrow} \left|\begin{matrix}1\\0\end{matrix}\right| + \left|\begin{matrix}0\\2\end{matrix}\right|$$

$$\left|\begin{matrix}0\\1\end{matrix}\right| + \left|\begin{matrix}1\\2\end{matrix}\right| \overset{c(1)}{\Longleftrightarrow} \left|\begin{matrix}1\\1\end{matrix}\right| + \left|\begin{matrix}0\\2\end{matrix}\right|$$

The graph has $l = 3$ linkage classes, $m = 6$ complexes, and $r = 3$ reversible reactions.

Scheme 2.5 ($L \geq 2$ loci with $A_i \geq 2$ alleles, $i = 1, \ldots, L$)**.**

$$
\begin{aligned}
l &= \sum_{K \subseteq \mathcal{L}: |K| \geq 2} \prod_{i \in K} \binom{A_i}{2} \prod_{i \in \mathcal{L} \setminus K} A_i \,, \\
m &= \sum_{K \subseteq \mathcal{L}: |K| \geq 2} \prod_{i \in K} \binom{A_i}{2} \prod_{i \in \mathcal{L} \setminus K} A_i \; 2^{|K|-1} \,, \\
r &= \sum_{K \subseteq \mathcal{L}: |K| \geq 2} \prod_{i \in K} \binom{A_i}{2} \prod_{i \in \mathcal{L} \setminus K} A_i \binom{2^{|K|-1}}{2} \,.
\end{aligned}
$$

2.6. Mass-action kinetics

Let $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ be a chemical reaction network and $k \in \mathbb{R}^{\mathcal{R}}_{>0}$ a vector of rate constants. Under the assumption of mass-action kinetics, the rate of a reaction $(y \to y') \in \mathcal{R}$, which depends on the species concentrations $x \in \mathbb{R}^{\mathcal{S}}_{\geq 0}$, is given by $k(y \to y')\, x^y$, that is, by a monomial in the reactant concentrations with the corresponding stoichiometric coefficients as exponents.

As detailed in Subsection 2.3, the effect of all patterns causing recombination (2.1) can be summarized in the chemical reaction (2.3), provided that the recombination is effective and the cumulative rate constant is positive. Then, reaction (2.3) occurs at the rate $\breve{c}(\{I, J\}_{\Delta(g,h)})\, p(g)\, p(h)$. In chemical terms, it follows mass-action-kinetics.

Hence, we obtain a dynamical system equivalent to (2.2), by summing over all reactions (2.4) and assuming mass-action kinetics:

$$
\frac{\mathrm{d}p}{\mathrm{d}t} = \sum_{(g+h \to g'+h') \in \mathcal{R}} k(g + h \to g' + h')\, p(g)\, p(h)\, (g' + h' - g - h) \,. \tag{2.7}
$$

The right-hand side of (2.7) can be written as a product of the stoichiometric matrix $N \in \mathbb{R}^{\mathcal{G} \times \mathcal{R}}$ and the rate vector $v_k(p) \in \mathbb{R}^{\mathcal{R}}_{\geq 0}$. Thereby, the column vector of N indexed by $(g + h \to g' + h') \in \mathcal{R}$ is given by $(g' + h' - g - h) \in \mathbb{R}^{\mathcal{G}}$ and the component of $v_k(p)$ indexed by $g + h \to g' + h'$ is given by $k(g + h \to g' + h')\, p(g)\, p(h)$. Hence,

$$
\frac{\mathrm{d}p}{\mathrm{d}t} = N v_k(p). \tag{2.8}
$$

Complex balancing

The right-hand side of the dynamical system (2.7) can also be written as a product of the complex matrix $Y \in \mathbb{R}^{\mathcal{G} \times \mathcal{C}}$, the Laplacian matrix $A_k \in \mathbb{R}^{\mathcal{C} \times \mathcal{C}}$ of the weighted directed graph, and the vector of monomials $p^Y \in \mathbb{R}^{\mathcal{C}}$. The column vector of Y indexed by $y \in \mathcal{C}$ is given by $y \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$ itself, that is, $Y^y = y$, and A_k is defined as follows: $(A_k)_{y'y} = k_{y \to y'}$ if $(y \to y') \in \mathcal{R}$, $(A_k)_{yy} = -\sum_{(y \to y') \in \mathcal{R}} k_{y \to y'}$, and $(A_k)_{y'y} = 0$ otherwise. We obtain

$$
\frac{\mathrm{d}p}{\mathrm{d}t} = Y A_k\, p^Y \tag{2.9}
$$

Recall that $(p^Y)_y = p^{Y^y} = p^y$ for $y \in \mathcal{C}$. For a particular complex $y = g + h$, we have $p^y = p^{g+h} = p(g)\, p(h)$.

An equilibrium of (2.9) is called complex-balancing if $A_k\, p^Y = 0$. That is, if at each complex the rates of all reactions sum up to zero.

Detailed balancing

In the process of genetic recombination, all reactions are reversible. Moreover, the rate constants of a reaction and its reverse coincide. Hence, we obtain a dynamical system equivalent to (2.7), by summing over all reversible reactions (2.5):

$$
\frac{\mathrm{d}p}{\mathrm{d}t} = \sum_{(g+h \Leftrightarrow g'+h') \in \mathcal{R}_{\Leftrightarrow}} k(g + h \Leftrightarrow g' + h') \left(p(g)\, p(h) - p(g')\, p(h')\right) \left(g' + h' - g - h\right) . \tag{2.10}
$$

An equilibrium of (2.10) is called detailed-balancing if $p(g)\,p(h) = p(g')\,p(h')$ for all $(g+h \Leftrightarrow g'+h') \in \mathcal{R}_{\Leftrightarrow}$. In general, an equilibrium of a reversible reaction network is called detailed-balancing if the rates of each reaction and its reverse coincide. Clearly, every detailed-balancing equilibrium is complex-balancing.

2.7. Conserved quantities

The change over time (2.8) lies in a subspace of $\mathbb{R}^{\mathcal{G}}$, and every trajectory in $\mathbb{R}^{\mathcal{G}}_{\geq 0}$ lies in a coset of this subspace. We define the stoichiometric subspace

$$S = \operatorname{im} N$$

and the stoichiometric compatibility classes

$$S(p) = (p + S) \cap \mathbb{R}^{\mathcal{G}}_{\geq 0}$$

for $p \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$. Every stoichiometric class is characterized by its orthogonal projection on $S^{\perp} = (\operatorname{im} N)^{\perp} = \ker N^T$, that is, by a vector of conserved quantities. For $u \in S^{\perp}$, that is, $u^T N = 0$, we have

$$\frac{\mathrm{d}(u^T p)}{\mathrm{d}t} = 0,$$

that is, $u^T p = const.$

We observe that the vector $1 \equiv 1^{\mathcal{G}} = \sum_{g\in\mathcal{G}} g$ is orthogonal to all columns of N: $1^T(g'+h'-g-h) = 0$ for all $(g+h \to g'+h') \in \mathcal{R}$, that is, $1^T N = 0$. Since $1^T p = \sum_{g\in\mathcal{G}} p(g)$, we have

$$\frac{\mathrm{d}(\sum_{g\in\mathcal{G}} p(g))}{\mathrm{d}t} = 0,$$

and, as one consequence, the simplex $S_{\mathcal{G}}$ is invariant.

Further, we consider for each locus and each allele at this locus the subset of gametes which contain this allele and define the corresponding formal sum of gametes

$$u_i(a) = \sum_{g\in\mathcal{G}: g_i = a} g \quad \text{for } i \in \mathcal{L} \text{ and } a \in \mathcal{A}_i,$$

where $u_i(a) \in \{0,1\}^{\mathcal{G}}$. As already mentioned, only complexes which contain the same alleles are connected by a reaction. Hence, $u_i(a)^T(g'+h'-g-h) = 0$ for all $(g+h \to g'+h') \in \mathcal{R}$, that is, $u_i(a)^T N = 0$, and the marginal frequencies

$$p_i(a) = u_i(a)^T p = \sum_{g\in\mathcal{G}: g_i = a} p(g)$$

are conserved quantities, that is,

$$\frac{\mathrm{d}p_i(a)}{\mathrm{d}t} = 0.$$

For each $i \in \mathcal{L}$, we have

$$\sum_{a\in\mathcal{A}_i} u_i(a) = \sum_{g\in\mathcal{G}} g$$

and

$$\sum_{a\in\mathcal{A}_i} p_i(a) = \sum_{g\in\mathcal{G}} p(g).$$

Hence, there are at least $1 + \sum_{i\in\mathcal{L}}(A_i - 1)$ linearly independent vectors in $S^{\perp}$ and as many independent marginals.

We define the marginal compatibility classes

$$M(p) = \{p' \in \mathbb{R}^{\mathcal{G}}_{\geq 0} \mid p'_i(a) = p_i(a) \text{ for } i \in \mathcal{L} \text{ and } a \in \mathcal{A}_i\}$$

for $p \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$. They are determined by the marginal frequencies of the alleles, which are conserved quantitites of the dynamics.

Clearly, $S(p) \subseteq M(p)$. In Propositions 3.2 and 3.3, we will show that every marginal and every stoichiometric compatibility class contain a unique equilibrium. Hence, the two classes coincide, cf. Corollary 3.5.

3. Results

We determine the equilibria for the process of genetic recombination and prove convergence to a unique equilibrium.

First, we rewrite the dynamical system (2.2). Using the symmetry in the double sum over recombination partners, we obtain

$$\begin{aligned}\frac{\mathrm{d}p}{\mathrm{d}t} &= \sum_{g,h\in\mathcal{G}} \sum_{\{I,J\}\in\mathcal{P}} c(\{I,J\})\, p(g)\, p(h)\, (g_I h_J - g) \\ &= \sum_{\{I,J\}\in\mathcal{P}} c(\{I,J\}) \left(\sum_{g,h\in\mathcal{G}} p(g)\, p(h)\, g_I h_J - p \right).\end{aligned}$$

Thereby, we assumed $\sum_{g\in\mathcal{G}} p(g) = 1$, that is, $p \in S_{\mathcal{G}}$.

For $K \subseteq \mathcal{L}$, we define the set of subgametes $\mathcal{G}_K = \prod_{i\in K} \mathcal{A}_i$, the projection $\mathcal{G} \to \mathcal{G}_K$, $g \mapsto g_K$, where $(g_K)_i = g_i$ for $i \in K$, and its linear extension to the corresponding vector spaces: $\mathbb{R}^{\mathcal{G}}_{\geq 0} \to \mathbb{R}^{\mathcal{G}_K}_{\geq 0}$, $p \mapsto p_K$, where $p_K = \sum_{g\in\mathcal{G}} p(g)\, g_K$, that is, $p_K(g_K) = \sum_{h\in\mathcal{G}:h_K=g_K} p(h)$. If $K = \{i\}$ with $i \in \mathcal{L}$, we recover the marginal frequencies $p_i = \sum_{g\in\mathcal{G}} p(g)\, g_i$, that is, $p_i(g_i) = \sum_{h\in\mathcal{G}:h_i=g_i} p(h)$.

Let $\{I,J\} \in \mathcal{P}$. For $g \in \mathcal{G}_I$ and $h \in \mathcal{G}_J$, we define $gh \in \mathcal{G}$ as $(gh)_i = g_i$ for $i \in I$ and $(gh)_i = h_i$ for $i \in J$ and extend the multiplication $\mathcal{G}_I \times \mathcal{G}_J \to \mathcal{G}$ linearly to $\mathbb{R}^{\mathcal{G}_I}_{\geq 0} \times \mathbb{R}^{\mathcal{G}_J}_{\geq 0} \to \mathbb{R}^{\mathcal{G}}_{\geq 0}$. Hence, we write

$$\begin{aligned}\frac{\mathrm{d}p}{\mathrm{d}t} &= \sum_{\{I,J\}\in\mathcal{P}} c(\{I,J\}) \left(\sum_{g\in\mathcal{G}} p(g)\, g_I \sum_{h\in\mathcal{G}} p(h)\, h_J - p \right) \\ &= \sum_{\{I,J\}\in\mathcal{P}} c(\{I,J\})\, (p_I\, p_J - p)\,.\end{aligned} \tag{3.1}$$

In fact, we may sum over $\{I,J\} \in \mathcal{P}^*$ since the contribution of $\{\emptyset, \mathcal{L}\}$ is identically zero.

The projection of a trajectory of the dynamical system is the trajectory of a projected dynamical system with the same structure: For $K \subseteq \mathcal{L}$ and $\{I',J'\} = \{I,J\}_K \in \mathcal{P}_K$, we find

$$(p_I\, p_J)_K = \sum_{g,h\in\mathcal{G}} p(g)\, p(g)\, (g_I h_J)_K = \sum_{g,h\in\mathcal{G}} p(g)\, p(g)\, g_{I'} h_{J'} = p_{I'}\, p_{J'}$$

and hence

$$\begin{aligned}\frac{\mathrm{d}p_K}{\mathrm{d}t} &= \sum_{\{I',J'\}\in\mathcal{P}_K} \sum_{\substack{\{I,J\}\in\mathcal{P}: \\ \{I,J\}\geq\{I',J'\}}} c(\{I,J\})\, ((p_I\, p_J)_K - p_K) \\ &= \sum_{\{I',J'\}\in\mathcal{P}_K} \check{c}(\{I',J'\})\, (p_{I'}\, p_{J'} - p_K)\,.\end{aligned}$$

3.1. Equilibria

Now, we are in a position to characterize the equilibria on the simplex.

Lemma 3.1. *For all recombination rate constants, $p \in S_{\mathcal{G}}$ is an equilibrium of the dynamical system* (2.2) *if and only if*

$$p = \prod_{i\in\mathcal{L}} p_i, \quad \textit{that is,} \quad p(g) = \prod_{i\in\mathcal{L}} p_i(g_i). \tag{3.2}$$

Proof. We show that, if (3.2), then

$$p_I = \prod_{i\in I} p_i.$$

Indeed, for $\{I, J\} \in \mathcal{P}$ and $J = \{1, \dots, |J|\}$, we find

$$\begin{aligned} p_I(g_I) &= \sum_{h_J \in \mathcal{G}_J} p(g_I h_J) \\ &= \sum_{h_1 \in \mathcal{G}_1} \dots \sum_{h_{|J|} \in \mathcal{G}_{|J|}} \prod_{i \in I} p_i(g_i) \prod_{i \in J} p_i(h_i) \\ &= \prod_{i \in I} p_i(g_i) \sum_{h_1 \in \mathcal{G}_1} p_1(h_1) \dots \sum_{h_{|J|} \in \mathcal{G}_{|J|}} p_{|J|}(h_{|J|}) \\ &= \prod_{i \in I} p_i(g_i). \end{aligned}$$

Hence,

$$p_I \, p_J = \prod_{i \in I} p_i \prod_{i \in J} p_i = \prod_{i \in \mathcal{L}} p_i = p$$

for all $\{I, J\} \in \mathcal{P}$, and $p \in S_{\mathcal{G}}$ is an equilibrium of the dynamical system (3.1) equivalent to (2.2).

It remains to show that $p \in S_{\mathcal{G}}$ is an equilibrium only if (3.2). We proceed by induction on the number of loci:

For $\mathcal{L} = \{1\}$, there is no non-trivial recombination. Every $p \in S_{\mathcal{G}}$ is an equilibrium which coincides with its marginals: $p(g) = p_1(g)$ for $g \in \mathcal{G}$, that is, $p = p_1$.

For $L \geq 2$, we consider subsets of loci $K \subset \mathcal{L}$ with $|K| < L$. The projection of an equilibrium $p \in S_{\mathcal{G}}$ of the dynamical system (with loci $\mathcal{L}$) is an equilibrium of the projected dynamical system (with loci K). By the induction hypothesis, $p_K = \prod_{i \in K} p_i$ and hence

$$p_I \, p_J = \prod_{i \in I} p_i \prod_{i \in J} p_i = \prod_{i \in \mathcal{L}} p_i$$

for all $\{I, J\} \in \mathcal{P}^*$. Summing over $\{I, J\} \in \mathcal{P}^*$ in (3.1), we obtain

$$0 = \sum_{\{I,J\} \in \mathcal{P}^*} c(\{I, J\}) \left(\prod_{i \in \mathcal{L}} p_i - p \right)$$

and hence $p = \prod_{i \in \mathcal{L}} p_i$. (There always exists a non-zero rate constant for some non-trivial recombination pattern.) □

If the dynamics is not restricted to the simplex, then $p \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$ is an equilibrium if and only if

$$p = \left(\textstyle\sum_{g \in \mathcal{G}} p(g) \right)^{1-L} \displaystyle\prod_{i \in \mathcal{L}} p_i.$$

In any case, an equilibrium does not depend on the recombination rate constants.

Without loss of generality, we assume

$$p_i(a) > 0 \quad \text{for all } i \in \mathcal{L} \text{ and } a \in \mathcal{A}_i$$

in the following. In other words, all marginal frequencies are positive, and hence boundary equilibria are excluded.

Clearly, every equilibrium is determined by its marginals, and we have the following result.

Proposition 3.2. *Every marginal compatibility class contains a unique, positive equilibrium.*

In chemical terms, every equilibrium is detailed-balancing since

$$p(g)\,p(h) = \prod_{i\in\mathcal{L}} p_i(g_i)\,p_i(h_i) = \prod_{i\in\mathcal{L}} p_i(g_i')\,p_i(h_i') = p(g')\,p(h')$$

for all $(g + h \Leftrightarrow g' + h') \in \mathcal{R}_{\Leftrightarrow}$, cf. Equation (2.10). Recall that only complexes which contain the same alleles are connected by a reaction. In fact, we can derive the following result entirely in the chemical setting.

Proposition 3.3. *Every stoichiometric compatibility class contains a unique, positive, detailed-balancing equilibrium.*

Proof. First, we determine the set of positive detailed-balancing equilibria. Using positivity, we write the condition for detailed balancing,

$$p(g)\,p(h) = p(g')\,p(h') \quad \text{for all } (g + h \Leftrightarrow g' + h') \in \mathcal{R}_{\Leftrightarrow},$$

as

$$p^{g'+h'-g-h} = 1 \quad \text{for all } (g + h \to g' + h') \in \mathcal{R}$$

and even more abstractly as

$$p^N = 1,$$

where $N \in \mathbb{R}^{\mathcal{G}\times\mathcal{R}}$ is the stoichiometric matrix and $1 \equiv 1^{\mathcal{R}}$. Clearly, the trivial solution is given by $p^* = 1 \equiv 1^{\mathcal{G}}$. To determine all solutions, we take the logarithm,

$$N^T \ln p = 0,$$

and note that $\ker N^T = (\operatorname{im} N)^\perp = S^\perp$ and $\dim(S^\perp) \geq 1$. Hence, we can write the set of positive detailed-balancing equilibria as

$$Z = \{p \in \mathbb{R}^{\mathcal{G}}_{>0} \mid \ln p - \ln p^* \in S^\perp\}.$$

Clearly, every detailed-balancing equilibrium is complex-balancing. Now, if there exists a positive complex-balancing equilibrium p^*, then the set of positive complex-balancing equilibria is given by Z and there are no other positive equilibria [11, Theorem 6A]. Moreover, every stoichiometric compatibility class contains a unique positive equilibrium [11, Lemma 4B]. Hence, the sets of positive detailed- and complex-balancing equilibria coincide, and every stoichiometric compatibility class contains a unique positive equilibrium, which is detailed-balancing.

It remains to preclude boundary equilibria. We consider an arbitrary initial value $p \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$ on the boundary, that is, $p(g) = 0$ for some $g \in \mathcal{G}$, and define the set of gametes $\mathcal{G}_0 = \{g \in \mathcal{G} \mid p(g) > 0\}$. Note that for each locus and each allele at this locus there is a gamete which contains this allele and occurs with a positive frequency. By Lemma 3.4 below, the set of all gametes $\mathcal{G}$ is reachable from $\mathcal{G}_0$. Now, let $p(t)$ be the solution of the dynamical system with $p(0) = p$. By [22, Theorem 2, p. 618], $p(t) \in \mathbb{R}^{\mathcal{G}}_{>0}$ for $t > 0$, and hence p is not an equilibrium. □

In the proof, we have used fundamental results about complex balancing by Horn and Jackson [11]. Necessary and sufficient conditions for complex balancing are given by Horn [10] and for detailed balancing by Vol'pert and Hudjaev [22] and Feinberg [5]. For the relation between complex and detailed balancing, see [4].

In the proof of Proposition 3.3, we have also used the purely graph-theoretical concept of reachability. Let $\mathcal{S}$ and $\mathcal{R}$ be the species and reactions of a chemical reaction network, and let $\mathcal{S}_0 \subseteq \mathcal{S}$ be a set of species. Iteratively, we define

$$\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{g' \mid g, h \in \mathcal{S}_{i-1} \text{ and } (g + h \to g' + h') \in \mathcal{R}\}$$

for $i \geq 1$. Since the graph is finite, we find $\mathcal{S}_i = \mathcal{S}_{i^*}$ for some $i^* \geq 0$ and all $i \geq i^*$, and the set of species reachable from $\mathcal{S}_0$ is given by $\mathcal{S}_{i^*}$.

Recall that every pair of loci gets separated by some recombination pattern with positive rate constant and hence every pair of loci gets separated by some reaction in the resulting network. We have the following result.

Lemma 3.4. *In a chemical reaction network arising from a process of genetic recombination, every gamete is reachable from a given set of gametes, provided that every allele is contained in some gamete in this set.*

Proof. We use induction on the number of loci:

For $L = 1$, the gametes coincide with the alleles.

For $L \geq 2$, we consider subsets of loci $K \subset \mathcal{L}$ with $|K| < L$ and project the network and the given set of gametes on the loci K. In the projected network, every pair of loci gets separated by some pattern, and in the projected set, every allele (at loci K) is contained in some gamete. By the induction hypothesis, every gamete $g_K \in \mathcal{G}_K$ is reachable in the projected network, and hence some gamete $h \in \mathcal{G}$ with $h_K = g_K$ is reachable.

It remains to show that every gamete $g' \in \mathcal{G}$ is reachable. Let $\mathcal{L} = \{1, \dots, L\}$ and $G = \mathcal{L} \setminus \{1\}$, $H = \mathcal{L} \setminus \{2\}$. By the argument above, some gametes $g, h \in \mathcal{G}$ with $g_G = g'_G$, $h_H = g'_H$ are reachable. If g' equals either g or h, then it is reachable. Otherwise, since loci 1 and 2 get separated by some pattern, we find the reaction $g + h \to g' + h'$, and hence g' is reachable. □

On the one hand, by Proposition 3.2, every marginal compatibility class contains a unique equilibrium. Hence, the set of all equilibria can be parametrized by $1 + \sum_{i \in \mathcal{L}} (A_i - 1)$ independent marginals. On the other hand, by Proposition 3.3, every stoichiometric compatibility class contains a unique equilibrium. Since stoichiometric classes are contained in marginal classes, we have the following result.

Corollary 3.5. *The stoichiometric compatibility classes coincide with the marginal compatibility classes, and* $\dim(S^\perp) = 1 + \sum_{i \in \mathcal{L}} (A_i - 1)$.

3.2. Convergence

Our main results concern the convergence of the dynamics to a unique equilibrium. For the first theorem, we provide two proofs: one by induction (as in the original literature) and one using the entropy as a Lyapunov function. For the second theorem, formulated in the chemical setting, we rely on results from chemical reaction network theory which are based on the same Lyapunov function.

Theorem 3.6. *In every marginal compatibility class and for all recombination rate constants, a process of genetic recombination converges to the unique equilibrium given by* (3.2).

First Proof (Induction). Every marginal compatibility class is characterized by a unique equilibrium. Given an equilibrium $p \in S_{\mathcal{G}}$, that is, $p = \prod_{i \in \mathcal{L}} p_i$, we consider trajectories in the class $M(p)$. We proceed by induction on the number of loci:

For $L = 1$, we have $M(p) = p$.

For $L \geq 2$, we consider subsets of loci $K \subset \mathcal{L}$ with $|K| < L$. The projection of a trajectory $\phi : \mathbb{R}_{\geq 0} \to M(p)$ of the dynamical system (with loci $\mathcal{L}$) is a trajectory of the projected dynamical system (with loci K). By the induction hypothesis, $\phi(t)_K \to \prod_{i \in K} p_i$ as $t \to \infty$ and hence

$$\phi(t)_I \, \phi(t)_J \to \prod_{i \in I} p_i \prod_{i \in J} p_i = \prod_{i \in \mathcal{L}} p_i = p$$

for all $\{I, J\} \in \mathcal{P}^*$. Summing over $\{I, J\} \in \mathcal{P}^*$ in the dynamical system (3.1) equivalent to (2.2), we obtain the non-autonomous differential equation

$$\begin{aligned} \frac{\mathrm{d}\phi}{\mathrm{d}t} &= \sum_{\{I,J\} \in \mathcal{P}^*} c(\{I, J\}) \, (\phi_I \, \phi_J - \phi) \\ &= f(t) - \sum_{\{I,J\} \in \mathcal{P}^*} c(\{I, J\}) \, \phi \end{aligned}$$

with

$$f(t) = \sum_{\{I,J\} \in \mathcal{P}^*} c(\{I, J\}) \, \phi_I \, \phi_J$$

and

$$f(t) \to \sum_{\{I,J\}\in\mathcal{P}^*} c(\{I,J\})\, p$$

as $t \to \infty$. In other words, the differential equation is asymptotically autonomous. The limiting equation

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \sum_{\{I,J\}\in\mathcal{P}^*} c(\{I,J\})\,(p-\phi)$$

is linear, and hence $\phi(t) \to p$ in the limiting equation. Moreover, $\{p\}$ is the maximal compact invariant set in the limiting system, and therefore $\phi(t) \to p$ holds also for all solutions of the original dynamical system, see e.g. [14, 15]. □

Second Proof (Lyapunov function). We consider the classical entropy function

$$H(p) = -\sum_{g\in\mathcal{G}} p(g)\ln p(g) = -\,p^T \ln p \geq 0$$

which defines a continuous function on the simplex $S_{\mathcal{G}}$. If $p(g) > 0$ for all $g \in \mathcal{G}$, then H is smooth and

$$\dot{H}(p) = -\sum_{g\in\mathcal{G}} \dot{p}(g)\ln p(g) - \sum_{g\in\mathcal{G}} \dot{p}(g) = -\,\dot{p}^T \ln p,$$

since $\sum_{g\in\mathcal{G}} p(g) = 1$. Using the dynamical system (2.10) equivalent to (2.2), we obtain

$$\dot{H}(p) = \sum_{(g+h\Leftrightarrow g'+h')\in\mathcal{R}_\Leftrightarrow} k(g+h \Leftrightarrow g'+h')\,(p(g)\,p(h) - p(g')\,p(h'))\cdot \\ \cdot\,(\ln(p(g)\,p(h)) - \ln(p(g')\,p(h'))) \geq 0.$$

Equality $\dot{H}(p) = 0$ holds if and only if p is a detailed-balancing equilibrium, that is, if and only if (3.2) holds.

Given an initial point $p(0) \in \mathbb{R}^{\mathcal{G}}_{>0}$ in the interior, the entropy $H(p(t))$ increases strictly towards its maximum on $M(p(0))$, and $p(t)$ converges to the unique equilibrium p in the class $M(p(0))$.

Given an initial point $p(0) \in \mathbb{R}^{\mathcal{G}}_{\geq 0}$ on the boundary, we have $p(t) \in \mathbb{R}^{\mathcal{G}}_{>0}$ for $t > 0$, cf. the proof of Proposition 3.3. In genetic terms, recombination immediately produces all gametes, as long as all alleles are present in the population. □

The entropy as Lyapunov function was used by Akin [1] and Lyubich [13] (referring to a paper by Kun and Lyubich [12]) to prove global stability for recombination. For chemical reaction networks with detailed balancing, see Vol'pert and Hudjaev [21, 22] (who acknowledge previous use of the entropy function by Zel'dovich [24]). For complex balancing, see [8, 9, 11, 20].

Theorem 3.7. *A process of genetic recombination gives rise to a reversible chemical reaction network with mass-action kinetics. In every stoichiometric compatibility class and for all reaction rate constants, the dynamics converges to a unique, positive, detailed-balancing equilibrium.*

Proof. By [22, Theorem, pp. 642–643], the ω-limit set of every solution consists either of a unique positive detailed-balancing equilibrium or boundary detailed-balancing equilibria. By Proposition 3.3, there are no boundary equilibria, and every solution converges to a unique, positive, detailed-balancing equilibrium. □

4. Final remarks

Note that we have not used a central concept of chemical reaction network theory, the deficiency

$$\delta = m - l - s,$$

where m is the number of complexes, l the number of linkage classes, and s the dimension of the stoichiometric subspace.

The deficiency zero and one theorems state that there exists a unique (asymptotically stable) positive complex-balancing equilibrium, for all reaction rate constants and all stoichiometric compatibility classes, if the network is weakly reversible and either (0) its deficiency is zero or (1a) the deficiencies of the individual linkage classes are zero or one and (1b) the individual deficiencies add up to the total deficiency, see [6].

In Example 2.1 ($L = 2$, $A_1 = A_2 = 2$), we find $\delta = 2 - 1 - 1 = 0$. However, already in Example 2.2 ($L = 3$, $A_1 = A_2 = A_3 = 2$), the deficiencies of the individual linkage classes are zero, but $\delta = 16-7-4 = 5$.

In fact, the individual deficiencies are zero in the entire Scheme 2.3 ($L \geq 2$, $A_i = 2$ for $i = 1, \dots, L$): Every linkage class is characterized by two different alleles at some loci $K \in \mathcal{L}$ with $|K| \geq 2$ and two identical alleles at other loci. Hence $2^{|K|-1} - 1 - (2^{|K|-1} - 1) = 0$, whereas

$$\begin{aligned}\delta &= m - l - s \\ &= 2^{L-1}(2^L - (1+L)) - (3^L - 2^{L-1}(2+L)) - (2^L - (1+L)) \\ &= 2^{L-1}(2^L - 1) - 3^L + 1 + L,\end{aligned}$$

using $s = \dim(S) = |\mathcal{G}| - \dim(S^\perp) = 2^L - (1+L)$. For $L = 3, 4, 5, \dots$, we find $\delta = 5, 44, 259, \dots$, and the deficiency zero and one theorems do not apply.

More importantly, there exist δ necessary and sufficient conditions on the rate constants for the existence of positive complex-balancing equilibria, see Horn [10]. The conditions involve the Laplacian matrix of the weighted directed graph of complexes and reactions, in particular, the quotients of so-called tree constants, see [16]. For the existence of positive detailed-balancing equilibria, additionally the Wegscheider conditions have to be fulfilled, that is, the products of rate constants in a cycle and its reverse must coincide, see [5, 22, 23]. In the process of genetic recombination, the rate constants of a reaction and its reverse coincide, and all conditions for the existence of positive complex- and detailed-balancing equilibria are fulfilled.

Acknowledgements. SM was supported by the project "Local Team and International Consortium for Computational Modeling of a Cyanobacterial Cell", Reg. No. CZ.1.07/ 2.3.00/20.0256, financed by the European Union and supported by the Ministry of Education of the Czech Republic.

We thank an anonymous referee for helpful comments.

References

[1] E. Akin. *The Geometry of Population Genetics*, vol. 31 of Lect. Notes in Biomath., Springer, New York, 1979.

[2] E. Baake. *Deterministic and stochastic aspects of single-crossover recombination*, in Proceedings of the International Congress of Mathematicians. Volume IV, Hindustan Book Agency, New Delhi, 2010, pp. 3037–3053.

[3] R. Bürger. *The Mathematical Theory of Selection, Recombination, and Mutation*, John Wiley & Sons, 2000.

[4] A. Dickenstein, M. Pérez Millán. *How far is complex balancing from detailed balancing?*, Bull. Math. Biol., 73 (2011), pp. 811–828.

[5] M. Feinberg. *Necessary and sufficient conditions for detailed balancing in mass action systems of arbitrary complexity*, Chemical Engineering Science, 44 (1989), pp. 1819 – 1827.

[6] M. Feinberg. *The existence and uniqueness of steady states for a class of chemical reaction networks*, Arch. Rational Mech. Anal., 132 (1995), pp. 311–370.

[7] H. Geiringer. *On the probability theory of linkage in Mendelian heredity*, Annals Math. Statist., 15 (1944), pp. 25–57.

[8] A. N. Gorban. *General H-theorem and entropies that violate the second law*, Entropy, 16 (2014), pp. 2408–2432.

[9] J. Higgins. *Some remarks on Shear's Liapunov function for systems of chemical reactions*, Journal of Theoretical Biology, 21 (1968), pp. 293–304.

[10] F. Horn. *Necessary and sufficient conditions for complex balancing in chemical kinetics*, Arch. Rational Mech. Anal., 49 (1972), pp. 172–186.

[11] F. Horn, R. Jackson. *General mass action kinetics*, Arch. Ration. Mech. Anal., 47 (1972), pp. 81–116.

[12] L. A. Kun, Y. I. Lyubich. *The H-theorem and convergence to equilibrium for free multi-locus populations*, Kibernetika, (1980), p. 150.

[13] Y. I. Lyubich. *Mathematical Structures in Population Genetics*, vol. 22 of Biomathematics, Springer-Verlag, Berlin, 1992. Translated from the 1983 Russian original by D. Vulis and A. Karpov.

[14] L. Markus. *Asymptotically autonomous differential systems*, in Contributions to the theory of nonlinear oscillations, vol. 3, vol. 36 of Annals of Mathematics Studies, Princeton University Press, 1956, pp. 17–29.

[15] K. Mischaikow, H. Smith, H. R. Thieme. *Asymptotically autonomous semiflows: chain recurrence and Lyapunov functions*, Trans. Amer. Math. Soc., 347 (1995), pp. 1669–1685.

[16] S. Müller, G. Regensburger. *Generalized mass-action systems and positive solutions of polynomial equations with real and symbolic exponents (invited talk)*, in Computer Algebra in Scientific Computing, V. P. Gerdt, W. Koepf, W. Seiler, and E. V. Vorozhtsov, eds., vol. 8660 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 302–323.

[17] T. Nagylaki. *The evolution of multilocus systems under weak selection*, Genetics, 134 (1993), pp. 627–47.

[18] T. Nagylaki, J. Hofbauer, P. Brunovský. *Convergence of multilocus systems under weak epistasis or weak selection*, Journal of Mathematical Biology, 38 (1999), pp. 103–133.

[19] S. Shahshahani. *A new mathematical framework for the study of linkage and selection*, vol. 211 of Memoirs of the AMS, Amer. Math. Soc., 1979.

[20] D. Siegel, D. MacLean. *Global stability of complex balanced mechanisms*, J. Math. Chemistry, 27 (2000), pp. 89–110.

[21] V. M. Vasil'ev, A. I. Vol'pert, S. I. Hudjaev. *The method of quasi-stationary concentrations for the equations of chemical kinetics*, Comput. Math. Math. Phys., 13 (1974), pp. 187–206.

[22] A. I. Vol'pert, S. I. Hudjaev. *Analysis in classes of discontinuous functions and equations of mathematical physics*, vol. 8 of Mechanics: Analysis, Martinus Nijhoff Publishers, Dordrecht, 1985.

[23] R. Wegscheider. *Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reactionskinetik homogener Systeme*, Monatshefte für Chemie und verwandte Teile anderer Wissenschaften, 22 (1901), pp. 849–906.

[24] Y. B. Zel'dovich. *The proof of uniqueness of the solution of mass law equations*, Zhurnal fizicheskoi khimii, 11 (1938), pp. 685–687.