Coevolution of the mitotic and meiotic modes of eukaryotic cellular division

Valmir C. Barbosa,[1,*] Raul Donangelo,[2,3] and Sergio R. Souza[2,4,5]

[1]*Programa de Engenharia de Sistemas e Computação, COPPE, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, 21941-972 Rio de Janeiro-RJ, Brazil*

[2]*Instituto de Física, Universidade Federal do Rio de Janeiro, Caixa Postal 68528, 21941-972 Rio de Janeiro-RJ, Brazil*

[3]*Instituto de Física, Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, 11.300 Montevideo, Uruguay*

[4]*Instituto de Física, Universidade Federal da Bahia, Campus Universitário de Ondina, 40210-340 Salvador-BA, Brazil*

[5]*Departamento de Física, ICEx, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte-MG, Brazil*

(Received 29 April 2018; revised manuscript received 9 August 2018; published 13 September 2018)

The genetic material of a eukaryotic cell (one whose nucleus and other organelles, including mitochondria, are enclosed within membranes) comprises both nuclear DNA (ncDNA) and mitochondrial DNA (mtDNA). These differ markedly in several aspects but nevertheless must encode proteins that are compatible with one another for the proper functioning of the organism. Here, we introduce a network model of the hypothetical coevolution of the two most common modes of cellular division for reproduction: by mitosis (supporting asexual reproduction) and by meiosis (supporting sexual reproduction). Our model is based on a random hypergraph, with two nodes for each possible genotype, each encompassing both ncDNA and mtDNA. One of the nodes is necessarily generated by mitosis occurring at a parent genotype, the other by meiosis occurring at two parent genotypes. A genotype's fitness depends on the compatibility of its ncDNA and mtDNA. The model has two probability parameters, p and r, the former accounting for the diversification of ncDNA during meiosis, the latter for the diversification of mtDNA accompanying both meiosis and mitosis. Another parameter, λ, is used to regulate the relative rate at which mitosis- and meiosis-generated genotypes are produced. We have found that, even though p and r do affect the existence of evolutionary pathways in the network, the crucial parameter regulating the coexistence of the two modes of cellular division is λ. Depending on genotype size, λ can be valued so that either mode of cellular division prevails. Our study is closely related to a recent hypothesis that brings mitochondria to the center stage and views the appearance of cellular division by meiosis, as opposed to division by mitosis, as an evolutionary strategy for boosting ncDNA diversification to keep up with that of mtDNA. Our results indicate that this may well have been the case, thus lending support to the first hypothesis in the field to take into account the role of such ubiquitous and essential organelles as mitochondria.

DOI: 10.1103/PhysRevE.98.032409

I. INTRODUCTION

A cell is said to be eukaryotic if it has a nucleus as well as other organelles enclosed in membranes providing separation from the cellular medium. With one single exception known to date [1], these other organelles include mitochondria, the cell's powerhouses. Every multicellular organism is a eukaryote, and so are numerous unicellular organisms as well, such as unicellular algae and fungi. A eukaryote's genotype comprises the genetic material found in both its nuclear DNA (ncDNA) and its mitochondrial DNA (mtDNA). Both types of DNA are essential for the proper functioning of the cell, so despite the fundamental differences between ncDNA and mtDNA (such as shape, size, multiplicity, and inheritance patterns), the proteins their genes synthesize must be compatible with one another. In fact, it is thought that such compatibility is key to an organism's fitness in evolutionary terms [2–4], as well as to regulating metabolic functions and supporting healthy aging [5]. Here, we consider eukaryotic cells exclusively.

A cell's ncDNA is organized as pairs of chromosomes. Its mtDNA, in turn, is part of a single chromosome in each mitochondrion. This form of organization is deeply entwined with how the organism reproduces since it supports cellular division (the central reproductive event at the cellular level) by either of the two most common modalities, mitosis and meiosis. The mitotic mechanism of division is used by an organism both for its somatic cells to multiply and for asexual (or clonal) reproduction if such is the case. During mitosis, the cell gets divided into two identical cells, each inheriting from the original cell an exact copy of its ncDNA and possibly mutated copies of its mtDNA. This is not to say that ncDNA never incurs mutations, which in fact constitute the prevailing cause of abnormalities such as cancer [6], only that such mutations are so rare as to be negligible in normal mitosis.

The other common mechanism of cellular division, that of meiosis, is central to the sexual reproduction of organisms. When a cell undergoes meiosis, its ncDNA is first "shuffled" through recombination and mutation of the genetic material in the paired chromosomes. Each chromosome in the resulting pair gets inherited by one of the cells produced by the division, called a gamete. Each gamete inherits a mutated version of the parent cell's mtDNA, just as in the case of mitosis. The

*valmir@cos.ufrj.br

encounter of two gametes, one from each parent during sexual reproduction, gives rise to a somatic cell of the resulting offspring, now with the paired chromosomes restored (one chromosome from each parent). This cell's mtDNA is in general inherited from only one of the parents (the mother). Cell division by meiosis is a much more complex process than division by mitosis, and as such requires substantially more time to complete.

Curiously, some organisms reproduce both asexually and sexually, depending on environmental and other factors [7,8], which hints at the possibility of a deep evolutionary past in which the modes of cellular division were much less well defined and coexisted much more freely. If such a past really existed, then the events that took place in it must have lied at the very roots of the evolution of sex and of meiosis as the currently prevalent mode of cellular division for reproduction. However, in spite of the evidence we find today in the form of organisms adopting asexual as well as sexual reproductive strategies, a widely accepted theory of how sex evolved is still lacking, even though proposals ranging from the purely biological [9,10] to the algorithmic and game theoretic [11] have been put forward.

In this paper we aim to explore, via mathematical modeling and computer simulations, what seems to be the most recent proposal as to why sex evolved in the first place and moreover has endured ever since [12,13]. This proposal is based on two core assumptions (cf. [12] and references therein). Assumption 1 is that the mutation rates in mtDNA transmission during cellular division, known to be much higher than that in ncDNA during recombination, have been consistently high since ancient times. Assumption 2 is that the inheritance of mtDNA from only one of the two gametes produced by meiosis, which is the rule for all eukaryotes with very few exceptions, has all along been constrained by natural selection and as such has also been the rule since the beginning. What the new theory posits is that sex evolved in response to the evolutionary advantage afforded by ncDNA mutations during recombination since these mutations could then stand up to those of mtDNA and thus help maintain compatibility between the two forms of DNA. This is backed by Assumption 1. As for Assumption 2, it is needed to prevent mtDNA from acquiring even more variability through the recombination that could take place if mtDNA material were inherited from both gametes. The fundamental nature of both recombination and mutation has been expressed mathematically in a number of occasions (cf., e.g., [14–18] and references therein), but the new proposal calls for their roles to be explored during the coevolution of two very distinct types of DNA. Such exploration lies at the core of this study, which has targeted both the purported centrality of mitochondria in the evolution of sex and the relevance of Assumptions 1 and 2 to such centrality, if indeed they can be said to have held.

Our model is essentially a network of genotypes with accompanying dynamical equations, each genotype accounting for both ncDNA and mtDNA, being therefore represented by three sequences, two standing for an ncDNA pair and one for mtDNA. Each genotype occupies two nodes of the network and correspondingly has two abundances associated with it (how many of it there are in each of the two nodes). These nodes differ in how the genotype they both represent relates to other genotypes, the chief distinction being the mode of cellular division through which genotype production comes about. One of them represents genotypes produced exclusively by mitosis, via the cloning of any other genotypes, regardless of how those genotypes are themselves produced. Genotypes represented by the other node necessarily result from the process of meiosis on the parent genotypes' sides, again with no restrictions on the mode of cellular division leading to those parents. Clearly, then, our model allows for the free coexistence and intermixing of both modes of cellular division.

The model also includes provisions to incorporate the Darwinian principles of random mutations and natural selection, in the form of two probability parameters, p and r, and of a measure for a genotype's fitness quantifying the compatibility of its ncDNA and mtDNA. The parameters p and r are meant to reflect the inherent randomness of ncDNA recombination and mtDNA mutation. Their role in the model is to regulate the network's density by affecting the interconnectedness of the genotypes, and also the pace of the model's dynamics. A third and final parameter λ helps account for the different durations of mitotic and meiotic cellular division.

Our model's dynamical equations are reminiscent of the well-known quasispecies equations [19–21], originally introduced to model the evolution of prebiotic molecules and RNA viruses [22–24], and of our own modifications thereof to incorporate network structure [25–27]. They are presented in Sec. II, along with all other details of the model. We then proceed with the presentation of results in Sec. III, discussion in Sec. IV, and conclusions in Sec. V.

II. MODEL

We represent each genotype as a triplet of sequences, each one comprising L binary digits (0's or 1's) that stand for the sequence's alleles (gene variants). For genotype i, we denote these sequences by i_1, i_2, and $i_{\rm mt}$, where i_1 and i_2 are the two chromosomes of ncDNA and $i_{\rm mt}$ is the single chromosome of mtDNA. This particular choice for the representation of a genotype incurs two strong simplifications, viz., that each allele gets reduced to only two possibilities and that the length of the mitochondrial chromosome is the same as that of the two nuclear chromosomes (whereas, in fact, it should be orders of magnitude shorter). As will become apparent, these simplifications have to do with practical limitations regarding the number of distinct genotypes given L, henceforth denoted by G, as well as the size of a network capable of accommodating twice as many genotypes. The value of G can be calculated by first noting that, for each of the 2^L possibilities for $i_{\rm mt}$, there are $\binom{2^L}{2}$ possibilities for the (unordered) pair i_1, i_2 if $i_1 \neq i_2$ and 2^L possibilities if $i_1 = i_2$. That is,

$$G = 2^L\left[\binom{2^L}{2} + 2^L\right] = 2^{3L-1} + 2^{2L-1}. \tag{1}$$

A schematic representation of mitosis and meiosis as they act on such sequences is given in Fig. 1.

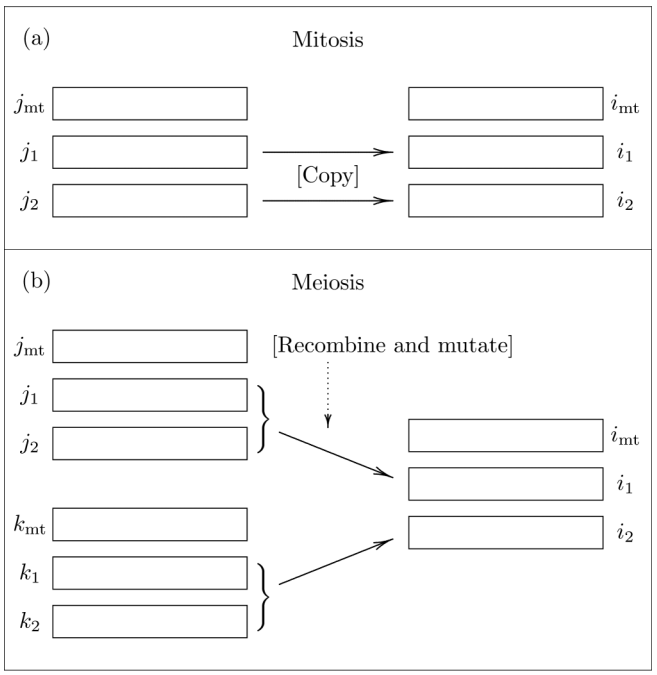

FIG. 1. (a) Generation of genotype i by mitosis from genotype j. Sequences j_1 and j_2 are copied to i_1 and i_2, respectively, while i_{mt} is a possibly mutated copy of j_{mt}. (b) Generation of genotype i by meiosis at parents j and k. Sequence i_1 originates from recombination and mutation between j_1 and j_2, sequence i_2 from k_1 and k_2. Sequence i_{mt} is a possibly mutated copy of either j_{mt} or k_{mt}.

A. Fitness of a genotype

Such simplifications also facilitate the definition of a genotype's fitness in terms of how compatible its ncDNA and mtDNA are. For genotype i, its fitness, denoted by f_i, is given by $f_i = 2^{-d_i}$, where d_i is the number of loci at which i_{mt} differs from both i_1 and i_2. Clearly, fitnesses range from 2^{-L} (when i_1 and i_2 are identical and i_{mt} differs from them at all loci) to 1 (when i_1 and i_2 do not necessarily equal each other at all loci but i_{mt} coincides with them wherever equality happens). Fitness f_i, therefore, grows exponentially with the number of alleles in i_{mt} that are identical to their counterparts in at least one of i_1, i_2. We remark that defining the fitness of a genotype in a way that focuses entirely on how compatible its ncDNA and mtDNA are is fully consistent with the central purpose of this study, noted in Sec. I, which is to explore the coevolution of two fundamentally different types of DNA that nevertheless are part of the same genotype and as such must function harmoniously if that genotype is to have any evolutionary advantage. This differs markedly from how fitness is defined in simpler studies in evolutionary dynamics, particularly those that target the replicative ability of single-sequence genotypes, in which case the fittest genotype can be equated with a so-called wild-type sequence (cf., e.g., one of our own studies on quasispecies [25] and references therein).

Later in the sequel it will become handy for us to know how many distinct genotypes there exist for which fitness equals some fixed value 2^{-k}. This number, n_k, can be easily calculated, as follows. Let α be the number of loci at which i_1 and i_2 differ and $\beta \geqslant k$ be the number of loci at which i_1 and i_2 are equal. Clearly, $\alpha + \beta = L$. If the partition between the α and the β loci were fixed, then the value of n_k, which depends on whether $\alpha = 0$ or $\alpha > 0$, would be given as follows. For $\alpha = 0$, n_k would equal simply $\binom{L}{k}2^L$. For $\alpha > 0$, it would equal $\binom{\beta}{k}2^{2\alpha+\beta-1}$, where 2^α possibilities for i_1 (hence for i_2) and 2^α for i_{mt} are accounted for on the $i_1 \neq i_2$ side of the partition, as well as 2^β possibilities for i_1 (hence for i_2) on the $i_1 = i_2$ side, and moreover such total number of possibilities gets divided by 2 to account for the fact that swapping i_1 and i_2 does not change genotype i. Letting the partition vary yields

$$\begin{aligned} n_k &= \binom{L}{k}2^L + \sum_{\alpha=1}^{L-k}\binom{L}{\alpha}\binom{L-\alpha}{k}2^{\alpha+L-1} \\ &= 2^{L-1}\binom{L}{k}\left[1 + \sum_{\alpha=0}^{L-k}\binom{L-k}{\alpha}2^\alpha\right] \\ &= 2^{L-1}\binom{L}{k}(1+3^{L-k}). \end{aligned} \tag{2}$$

As expected, $\sum_{k=0}^{L} n_k = G$. Moreover, the sum total of all genotypes' fitnesses, henceforth denoted by F, is given by

$$F = \sum_{k=0}^{L} n_k 2^{-k} = \frac{3^L + 7^L}{2}. \tag{3}$$

B. A note on hypergraphs

Most network models rely on graphs as the natural means of representation. A graph is defined simply as a set N of nodes and a set of edges that can be any subset of $N \times N$, so clearly an edge is defined by the pair of nodes it interconnects. Such a pair is unordered for undirected graphs, ordered for directed graphs. Given the ordered pair (i, j), the edge it stands for is said to be directed from node i to node j.

As it turns out, however, this representational scheme is not entirely adequate in the present case, owing mainly to the need to represent not only the generation of genotypes by mitosis, but also the generation that results from meiosis on the parents' side. To this end, we resort to the generalization of graphs known as hypergraphs [28]. A hypergraph shares with a graph the fact of being defined on a set N of nodes, but differs from a graph in that the means it employs to represent an interconnection, now called a hyperedge, is more general than an edge. At the level of generality that we require in this paper, a hyperedge is a nonempty multiset with elements from N.

Just as in the case of graphs, directed hypergraphs can also be considered [29]. In a directed hypergraph, a hyperedge is partitioned into node multisets S and D, called the hyperedge's source and destination node multisets, respectively. Our use of hypergraphs will include directed hyperedges like $\{i, j\}$ with $S = \{j\}$ and $D = \{i\}$ (really the same as an edge directed from j to i in a graph), possibly with $i = j$, and directed hyperedges like $\{i, j, k\}$ with $S = \{j, k\}$ and $D = \{i\}$, possibly with any of the node repetitions $i = j$, $i = k$, $j = k$, or $i = j = k$.

For $i \in N$, we denote the set of all source multisets S for which the hyperedge directed from S to $D = \{i\}$ exists by I_i (the input set to node i). We write either $j \in I_i$ or

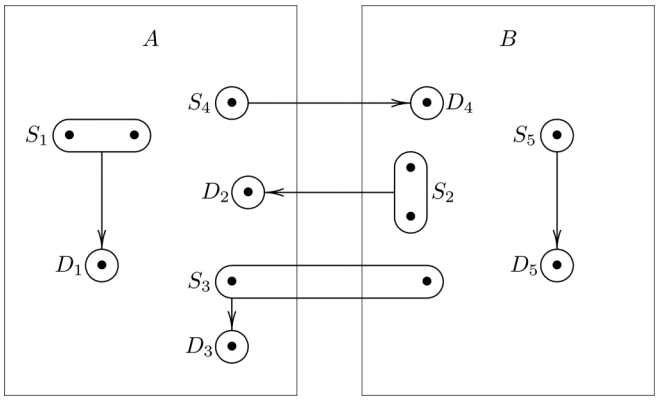

FIG. 2. A fragment of hypergraph H, showing some of its hyperedges and only those nodes that participate in them. The complete node sets A and B are identical (each contains nodes representing all genotypes), but a hyperedge leading to a node in A is of a different nature than a hyperedge leading to a node in B since nodes in set A represent genotypes generated by meiosis while those in B represent genotypes generated by mitosis. In this fragment, three hyperedges $S_1 \to D_1$, $S_2 \to D_2$, and $S_3 \to D_3$ indicate genotype generation by meiosis, two from same-set parents, one from mixed parents. Two other hyperedges, $S_4 \to D_4$ and $S_5 \to D_5$, indicate genotype generation by mitosis, the former from a genotype in A, the latter from another genotype in B.

$jk \in I_i$ for the two possible types of hyperedge in our case. Correspondingly, we denote the output set of node j by O_j and that of the node pair j, k by O_{jk}, writing $i \in O_j$ and $i \in O_{jk}$, respectively. In these notations, the use of jk refers to an unordered pair of genotypes.

C. Network structure

Our network of interacting genotypes is represented by a directed hypergraph H whose set of nodes N has two nodes for each of the G possible genotypes. We partition N into sets A and B, each with a full set of distinct genotypes. That is, any possible genotype is represented in H by a node in A and another in B. What distinguishes these two nodes is the set of hyperedges directed toward them: nodes in B are meant to represent genotypes generated by mitosis, so I_i for $i \in B$ contains single-node sets only; nodes in A are for genotypes generated by meiosis of their parents, so I_i for $i \in A$ contains two-node multisets only. In either case, the nodes that go into the multisets of I_i originate from the entirety of N, regardless of the partition into A and B. That is to say, it is possible for a genotype in B to originate by mitosis from either a genotype in A or one in B. In the same vein, each of the two genotypes that undergo meiosis to give rise to a genotype in A can be a member of A or B. This is illustrated in Fig. 2.

The description of hypergraph H is completed by specifying its hyperedges. We do this based on two probability parameters p and r, which regulate the occurrence of mutation that accompanies ncDNA recombination and that which mtDNA undergoes, respectively. Probability p, therefore, affects the expected number of hyperedges incoming to nodes in A, whereas probability r has a similar effect on nodes in both A and B. Clearly, then, the resulting H is to be regarded as a sample of the random hypergraph defined by the probabilities p and r on set N. By proceeding in this way, we allow for much greater variation regarding genotype interaction.

Henceforth, we let $h(s, t)$ denote the Hamming distance between sequences s and t and $R(s, t)$ denote the set of all sequences that can result from recombination between s and t. Each sequence in $R(s, t)$, therefore, equals both s and t at $L - h(s, t)$ loci and equals either one or the other at the remaining $h(s, t)$ loci. Therefore, $R(s, t)$ contains $2^{h(s,t)}$ sequences.

D. Mitotic hyperedges

A mitotic hyperedge is directed from $S = \{j\}$ to $D = \{i\}$ with $j \in A \cup B$ and $i \in B$, possibly with $i = j$. This hyperedge can only exist if genotypes i and j have the same ncDNA (i.e., both $i_1 = j_1$ and $i_2 = j_2$) since normal division by mitosis does not alter the nuclear genetic material. If this holds, then the hyperedge exists with probability $p_{j,i}$, given by

$$p_{j,i} = r^{h(i_{\rm mt}, j_{\rm mt})}. \tag{4}$$

That is, the existence of the hyperedge depends on how likely it is for the mtDNA of genotype j to mutate into that of genotype i. If the two are identical (hence $i = j$), then $p_{j,i} = 1$. Consequently, every genotype in B has at least two incoming hyperedges, one originating from itself and another originating from its identical counterpart in A.

E. Meiotic hyperedges

A meiotic hyperedge has $S = \{j, k\}$ and $D = \{i\}$ with $j, k \in A \cup B$ and $i \in A$, allowing for $i = j$, $i = k$, $j = k$, or $i = j = k$. Unlike its mitotic counterpart, a meiotic hyperedge is in no way constrained by how the genetic material of the genotypes involved relate to one another. Its existence is, therefore, unconditionally random and occurs with probability $p_{jk,i}$, whose calculation must take into account the recombination of genetic material of both j and k, possibly affected by mutation, and the inheritance by genotype i of a mutated version of the mtDNA of either j or k. Given all the independences involved, $p_{jk,i}$ can be expressed as

$$p_{jk,i} = \pi_{jk,i}\rho_{jk,i}, \tag{5}$$

where $\pi_{jk,i}$ is the probability of the ncDNA-related sample-space events and $\rho_{jk,i}$ is the probability of the mtDNA-related ones.

In order to calculate probability $\pi_{jk,i}$, we must take into account the fact that, even though sequences i_1 and i_2 are interchangeable (swapping them does not affect genotype i), there are two fundamentally distinct ways they can originate from genotypes j and k, depending on whether i_1 is paired with j and i_2 with k, or conversely. Thus, $\pi_{jk,i}$ is given by

$$\pi_{jk,i} = \pi_{j\to i_1}\pi_{k\to i_2} + \pi_{j\to i_2}\pi_{k\to i_1} - \pi_{\rm mixed}, \tag{6}$$

where $\pi_{j\to i_1}$ is the probability that i_1 results from recombination at j with the intervening effect of mutation, and similarly for $\pi_{j\to i_2}$, $\pi_{k\to i_1}$, and $\pi_{k\to i_2}$. As for $\pi_{\rm mixed}$, it is the probability that, in a population of genotype-i individuals, some have inherited i_1 from genotype j and i_2 from genotype k (let E_1 be the event in sample space corresponding to this

pattern of inheritance) while others have inherited i_1 from genotype k and i_2 from genotype j (event E_2). That is, π_{mixed} is the probability of event $E_1 \cap E_2$. This probability is counted twice as the first two terms in Eq. (6) are added up since the first of them is the probability of event E_1 and the second is the probability of event E_2. Subtracting π_{mixed} off the total is meant to correct for this.

The probabilities appearing in Eq. (6) are made explicit by resorting to the set $R(s,t)$ of all sequences that can result from recombination between sequences s and t. We do this by assuming that all recombinations between ncDNA sequences occur without any bias toward any of the two sequences [so, every sequence in $R(s,t)$ is equiprobable, following Mendel's first law, or law of segregation, which seems reasonable given how much variation there is from one organism to another [30] and that we focus on no specific organism] and that, in a way similar to that of the mitotic case, the recombination-related mutation that accompanies the transformation of sequence s into sequence t occurs with probability $p^{h(s,t)}$. We then have

$$\pi_{j\to i_1} = 2^{-h(j_1,j_2)} \sum_{\ell\in R(j_1,j_2)} p^{h(\ell,i_1)}, \tag{7}$$

$$\pi_{j\to i_2} = 2^{-h(j_1,j_2)} \sum_{\ell\in R(j_1,j_2)} p^{h(\ell,i_2)}, \tag{8}$$

$$\pi_{k\to i_1} = 2^{-h(k_1,k_2)} \sum_{\ell\in R(k_1,k_2)} p^{h(\ell,i_1)}, \tag{9}$$

$$\pi_{k\to i_2} = 2^{-h(k_1,k_2)} \sum_{\ell\in R(k_1,k_2)} p^{h(\ell,i_2)}, \tag{10}$$

and

$$\begin{aligned}\pi_{\text{mixed}} = 2^{-1}[&\pi_{j\to i_1}\pi_{k\to i_2}(\pi_{j\to i_2}+\pi_{k\to i_1}-\pi_{j\to i_2}\pi_{k\to i_1})\\ &+\pi_{j\to i_2}\pi_{k\to i_1}(\pi_{j\to i_1}+\pi_{k\to i_2}-\pi_{j\to i_1}\pi_{k\to i_2})].\end{aligned} \tag{11}$$

Equation (11), in particular, can be understood by analyzing the hypothetical case of a completely uniform population of genotype-i individuals, that is, one in which every i_1 comes from j and every i_2 from k, or conversely. The probability that this happens is $\pi_{j\to i_1}(1-\pi_{k\to i_1})\pi_{k\to i_2}(1-\pi_{j\to i_2})+\pi_{j\to i_2}(1-\pi_{k\to i_2})\pi_{k\to i_1}(1-\pi_{j\to i_1})$, which can be rewritten as $\pi_{j\to i_1}\pi_{k\to i_2}+\pi_{j\to i_2}\pi_{k\to i_1}-2\pi_{\text{mixed}}$. This expression is entirely consistent with that on the right-hand side of Eq. (6) since the amount to be subtracted off $\pi_{j\to i_1}\pi_{k\to i_2}+\pi_{j\to i_2}\pi_{k\to i_1}$ to ensure that the probability of a mixed population is counted only once is exactly half the amount to be subtracted to ensure uniformity.

As for probability $\rho_{jk,i}$, its expression is simply

$$\rho_{jk,i} = r^{h(i_{\text{mt}},j_{\text{mt}})} + r^{h(i_{\text{mt}},k_{\text{mt}})} - r^{h(i_{\text{mt}},j_{\text{mt}})+h(i_{\text{mt}},k_{\text{mt}})}, \tag{12}$$

where the term subtracted at the end is the probability that, in a population whose individuals all share genotype i, some have inherited their mtDNA from genotype j, some from genotype k. As in the case of Eq. (6), the probability that this happens is counted twice as the first two terms are added up. The subtraction fixes this.

F. Interpretation of the base probabilities

As we consider Eqs. (4), (7)–(10), and (12), it is important to note that the base probabilities used (p, r, and 2^{-1}) can be interpreted as point probabilities affecting each one of the loci in question equally and independently but not the others (whose probability of being affected is 0). Thus, if b is one of the three base probabilities and h is the number of loci at which two sequences differ, then the probability that all h loci get affected is indeed the b^h used in various forms in those equations. This formulation is reminiscent of the uniform-susceptibility model of quasispecies mutation, introduced elsewhere in an attempt to circumvent the implausibility of certain common assumptions (cf. [25] and references therein).

G. Network dynamics

Given hypergraph H, an instance of the random-hypergraph model described so far, the resulting network dynamics is expressed by a set of $2G$ coupled differential equations, one for each of the genotypes (nodes) in the network. These equations give the rates at which the genotypes' abundances vary with time and depend on the same probabilities $p_{j,i}$ [Eq. (4)] and $p_{jk,i}$ [Eq. (5)] used to obtain hypergraph H in the first place. Readily, through these probabilities the parameters p and r affect network dynamics as much as network structure. Mutation rates in mtDNA are usually much higher than those in ncDNA (cf., e.g., [31] and references therein), suggesting that we use $p \ll r$.

For use in the dynamics, such probabilities must be normalized so that summing them up for all $i \in O_j$ with j fixed yields 1, and so does summing them up for all $i \in O_{jk}$ with j,k fixed. The probabilities' normalized versions are, respectively,

$$q_{j,i} = \frac{p_{j,i}}{\sum_{\ell\in O_j} p_{j,\ell}} \tag{13}$$

and

$$q_{jk,i} = \frac{p_{jk,i}}{\sum_{\ell\in O_{jk}} p_{jk,\ell}}. \tag{14}$$

We first give the differential equations for the absolute abundances of the genotypes. For genotype i, this absolute abundance is denoted by X_i. If $i \in A$ (that is, genotype i is the result of meiosis for parents $jk \in I_i$), we have

$$\dot{X}_i = \sum_{\substack{jk\in I_i\\ j\neq k}} f_j f_k q_{jk,i} \min\{X_j, X_k\} + \sum_{\substack{jk\in I_i\\ j=k}} f_j^2 q_{jj,i} \frac{X_j}{2}. \tag{15}$$

In this equation, the influence exerted on $\dot{X}_i$ by each $jk \in I_i$ depends both on the fitnesses f_j and f_k of the two parents and on probability $q_{jk,i}$. It also depends on how abundant the pairs they form can be, which in turn depends on whether $j \neq k$ or $j = k$. In the former case, the number of possible pairs is given by $\min\{X_j, X_k\}$. In the latter, the number is $X_j/2$.

If $i \in B$ (that is, genotype i results from mitosis for $j \in I_i$), then a simpler equation ensues,

$$\dot{X}_i = \lambda \sum_{j\in I_i} f_j q_{j,i} X_j, \tag{16}$$

TABLE I. Summary of the hyperedges of H in relation to the model's parameters and dynamics.

Hyperedge's source multiset S	Hyperedge's destination multiset D	Properties
Genotype set $\{j\}$ for $j \in A \cup B$	Genotype set $\{i\}$ for $i \in B$	The existence of a hyperedge from S to D depends on probability r. The dynamics, which represents the generation of genotype i from mitosis and mitochondrial mutation at genotype j, depends on r, on the fitness f_j of genotype j, and on the rate λ.
Genotype multiset $\{j,k\}$ for $j,k \in A \cup B$	Genotype set $\{i\}$ for $i \in A$	The existence of a hyperedge from S to D depends on probabilities p and r. The dynamics, which represents the generation of genotype i from meiosis with recombination at genotypes j,k, as well as on mitochondrial mutation at either j or k, depends on p and r, and on the fitnesses f_j and f_k of genotypes j and k, respectively.

where the role played by each $j \in I_i$ in altering $\dot{X}_i$ is again dependent on the fitness f_j, the probability $q_{j,i}$, and the abundance X_j. The parameter λ allows us to tune the entire dynamics so that the speed with which mitosis and meiosis occur relative to each other can be experimented with. Normally, mitosis is a substantially faster process than meiosis (cf., e.g., [32] and references therein), which to some extent is already accounted for in Eqs. (15) and (16), owing to the presence of squared fitnesses in the former equation. Tuning through the value of λ is then expected to work in conjunction with this. In particular, decreasing λ while all else remains constant leads the rate at which meiosis acts relative to mitosis to increase.

Equations (15) and (16) imply the unbounded growth of every genotype's absolute abundance. A better approach is then to turn to the genotypes' relative abundances instead since these are constrained to lie in the interval [0,1]. The relative abundance of genotype i, denoted by x_i, is

$$x_i = \frac{X_i}{\sum_{\ell \in A\cup B} X_\ell}, \tag{17}$$

hence we have

$$\dot{x}_i = \frac{\dot{X}_i}{\sum_{\ell\in A\cup B} X_\ell} - x_i \frac{\sum_{\ell\in A\cup B} \dot{X}_\ell}{\sum_{\ell\in A\cup B} X_\ell} = \frac{\dot{X}_i}{\sum_{\ell\in A\cup B} X_\ell} - x_i\phi, \tag{18}$$

where

$$\phi = \sum_{\substack{jk\in A\cup B\\ j\neq k}} f_j f_k \min\{x_j, x_k\} + \sum_{\substack{jk\in A\cup B\\ j=k}} f_j^2 \frac{x_j}{2} + \lambda \sum_{j\in A\cup B} f_j x_j. \tag{19}$$

From Eq. (18) we can easily derive the counterparts of Eqs. (15) and (16) for relative abundances:

$$\dot{x}_i = \sum_{\substack{jk\in I_i\\ j\neq k}} f_j f_k q_{jk,i} \min\{x_j, x_k\} + \sum_{\substack{jk\in I_i\\ j=k}} f_j^2 q_{jj,i} \frac{x_j}{2} - x_i\phi \tag{20}$$

for $i \in A$ and

$$\dot{x}_i = \lambda \sum_{j\in I_i} f_j q_{j,i} x_j - x_i\phi \tag{21}$$

for $i \in B$.

H. Summary

A summary of the topological and dynamical properties of hypergraph H is given in Table I. In this table, emphasis is placed on the structure of each type of hyperedge, as well as on how the existence of each hyperedge and the associated dynamics depend on the model's parameters.

I. A special case

By Eq. (5), for $p = r = 1$ we have $p_{jk,i} = 1$ for all $i \in A$ and all $j,k \in A \cup B$, meaning that every genotype $i \in A$ has the same input set I_i. By Eq. (4), for $r = 1$ we similarly have $p_{j,i} = 1$ for all $i \in B$ and all $j \in A \cup B$, provided the ncDNA of genotype j is identical to that of i. In general, then, genotypes in B can have distinct input sets. In specifying the special case of this section, we aim to study the setup in which not only every possible connection is present but also the value of X_i is the same for all $i \in A$ and likewise for all $i \in B$. This can be imposed for $t = 0$, but having it hold subsequently requires moreover that every genotype keep receiving the same input as all others in the same set (A or B). This is already true of genotypes in A (by virtue of their shared input set), but making it true of genotypes in B as well requires that we circumvent the fact that, inside a group of same-ncDNA genotypes, fitness can be distributed differently than inside another group. We can circumvent this by constraining the nature of the genotypes that constitute sets A and B so that fitness distribution is the same for every occurring ncDNA.

Our criterion for including genotypes in A and B is the simplest possible: genotype i is to be included if and only if both i_1 and i_2 are sequences of 0's, so now all genotypes in B have identical input sets as well. The number of genotypes

in A or B is therefore no longer the G of Eq. (1), but instead $G_0 = 2^L$ (the number of possibilities for mtDNA). Likewise, the number of genotypes having fitness 2^{-k}, previously given by the n_k of Eq. (2), now amounts to $n_{k,0} = \binom{L}{k}$ (since k is now the number of loci at which mtDNA is 1). Consequently, the sum total of fitnesses in set A or B, previously given by F as in Eq. (3), is now denoted by F_0 and given by

$$F_0 = \sum_{k=0}^{L} n_{k,0} 2^{-k} = \left(\frac{3}{2}\right)^L. \tag{22}$$

By Eqs. (13) and (14), $q_{k,i} = q_{jk,i} = 1/G_0$. Letting a stand for any $i \in A$ and b for any $i \in B$, we obtain

$$\dot{X}_a = \frac{F_0^2}{2G_0} X_a + \frac{F_0^2}{G_0} \min\{X_a, X_b\} + \frac{F_0^2}{2G_0} X_b \tag{23}$$

and

$$\dot{X}_b = \frac{\lambda F_0}{G_0} X_a + \frac{\lambda F_0}{G_0} X_b. \tag{24}$$

Rescaling time by the factor $F_0^2/2G_0$ and letting $\mu = \lambda/F_0$ allow us to rewrite these equations as

$$\dot{X}_a = X_a + 2\min\{X_a, X_b\} + X_b \tag{25}$$

and

$$\dot{X}_b = 2\mu X_a + 2\mu X_b. \tag{26}$$

Except for the trivial case of $X_i = 0$ all over $A \cup B$ initially, these two differential equations entail an unbounded exponential growth of both X_a and X_b.

There are two regimes to be considered. The first one, valid while $X_a \leqslant X_b$, leads to System 1,

$$\dot{X}_a = 3X_a + X_b, \tag{27}$$

$$\dot{X}_b = 2\mu X_a + 2\mu X_b, \tag{28}$$

whose eigenvalues are $u_\pm = \mu + \frac{3}{2} \pm \sqrt{(\mu + \frac{3}{2})^2 - 4\mu}$. Assuming $X_a(0) = 0$ yields

$$\frac{X_a}{X_b} = \frac{1 - e^{(u_- - u_+)t}}{u_+ - 3 - (u_- - 3)e^{(u_- - u_+)t}} \tag{29}$$

and, therefore,

$$\lim_{t\to\infty} \frac{X_a}{X_b} = \frac{1}{u_+ - 3}. \tag{30}$$

This steady state can be reached only if it happens while $X_a \leqslant X_b$, hence for $\mu \geqslant 1$ ($\lambda \geqslant F_0$), with $\mu = 1$ implying $X_a = X_b$. For $\mu < 1$ ($\lambda < F_0$) the value of X_a catches up with that of X_b before the steady state can be reached. In this case, the second regime takes over.

This second regime is valid for $X_a \geqslant X_b$ and leads to System 2,

$$\dot{X}_a = X_a + 3X_b, \tag{31}$$

$$\dot{X}_b = 2\mu X_a + 2\mu X_b, \tag{32}$$

of eigenvalues $u_\pm = \mu + \frac{1}{2} \pm \sqrt{(\mu + \frac{1}{2})^2 + 4\mu}$. For $X_a(0) = X_b(0)$, we obtain

$$\frac{X_a}{X_b} = \frac{3[(u_- - 4)e^{u_+ t} + (4 - u_+)e^{u_- t}]}{(u_+ - 1)(u_- - 4)e^{u_+ t} + (4 - u_+)e^{u_- t}} \tag{33}$$

and

$$\lim_{t\to\infty} \frac{X_a}{X_b} = \frac{3}{u_+ - 1}. \tag{34}$$

Similarly to the first regime, this steady state can be reached only if it happens while $X_a \geqslant X_b$, hence for $\mu \leqslant 1$ ($\lambda \leqslant F_0$), again with $\mu = 1$ implying $X_a = X_b$.

This brief analysis of System 1 [Eqs. (27) and (28)] and System 2 [Eqs. (31) and (32)] reveals how $X_a > X_b$ can be reached in the long run, having started at $X_a(0) = 0$. The general picture is that the genotypes are initially subject to the solution to System 1, which remains the case for as long as $X_a \leqslant X_b$. This can either endure indefinitely, with the genotypes eventually reaching the steady state of Eq. (30), or end when $X_a = X_b$ occurs before that steady state is reached. The outcome depends on the value of λ, with $\lambda \geqslant F_0$ implying the former, $\lambda < F_0$ implying the latter. It follows that, in order for the genotypes to go beyond $X_a = X_b$ and eventually reach a steady state in which $X_a > X_b$, we must have $\lambda < F_0$. In this case, the genotypes become subject to the solution to System 2 as soon as $X_a = X_b$ happens, and from then on converge to the steady state of Eq. (34), necessarily with $X_a > X_b$. Note that the key to this development from $X_a(0) = 0$ is assigning a sufficiently small value to λ. As we show in Sec. III, this continues to hold when we leave the special case and return to the general model.

III. RESULTS

Henceforth, we use x_A to denote the total relative abundance of genotypes generated by meiosis, and likewise x_B for those generated by mitosis. That is, $x_A = \sum_{i\in A} x_i$ and $x_B = 1 - x_A$. All our results come from time stepping Eqs. (20) and (21) for a fixed hypergraph H (obtained by Monte Carlo sampling as described in Secs. II D and II E), assuming $x_i(0)$ to be selected uniformly at random for $i \in B$ and $x_i(0) = 0$ for $i \in A$ [hence $x_A(0) = 0$ and $x_B(0) = 1$], then averaging or binning over hypergraphs and initial conditions. Assuming total dominance of mitosis at $t = 0$ allows us to use our model to probe the hypothetical evolutionary past in which cellular division by mitosis was the rule but through mutation and natural selection gave rise to the appearance of meiosis.

Unlike our previous work based on similar dynamical equations on random networks [25–27], in which a few thousand nodes could be handled within reasonable bounds on computational resources, the situation is now significantly more severe. The reason behind this is twofold, since not only does it involve a much faster growth of the number of nodes with sequence length L ($2G = 2^{3L} + 2^{2L}$ now versus 2^L or 2^{L+1} before), but also it requires hyperedges (rather than edges) to be handled. As a consequence, in this paper we use $L = 3, 4$, whereas in those previous publications we reached $L = 10$ or even a little higher.

These limitations notwithstanding, we have been able to elicit richly varied behavior from suitable valuations of the three parameters p, r, and λ. Our results are presented in Figs. 3–7. In Fig. 3 we explore the parameter landscape as each possibility leads to a form of coexistence between

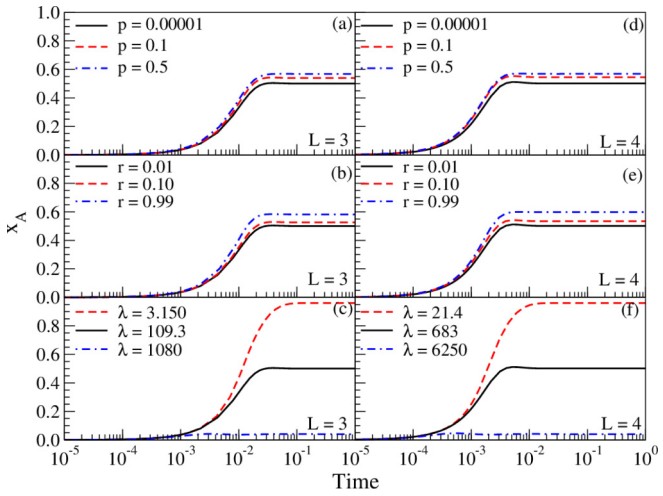

FIG. 3. Evolution of the relative abundance of genotypes generated by meiosis (x_A) for $L = 3$ (a)–(c) and $L = 4$ (d)–(f). Parameter values omitted from each panel default to $p = 10^{-5}$, $r = 0.01$, and $\lambda = 109.3$ (for $L = 3$) or $\lambda = 683$ (for $L = 4$).

mitosis- and meiosis-generated genotypes. We do this for both $L = 3$ and 4.

Figures 4–6, in turn, focus on how the relative abundances of genotypes become distributed in the steady state, given a scenario in which mitosis dominates ($x_A \approx 0.04$, in Fig. 4), another in which mitosis and meiosis coexist at roughly the same proportion ($x_A \approx 0.5$, in Fig. 5), and another in which meiosis dominates ($x_A \approx 0.96$, in Fig. 6). All three figures share the same value of p and r, with the value of λ being

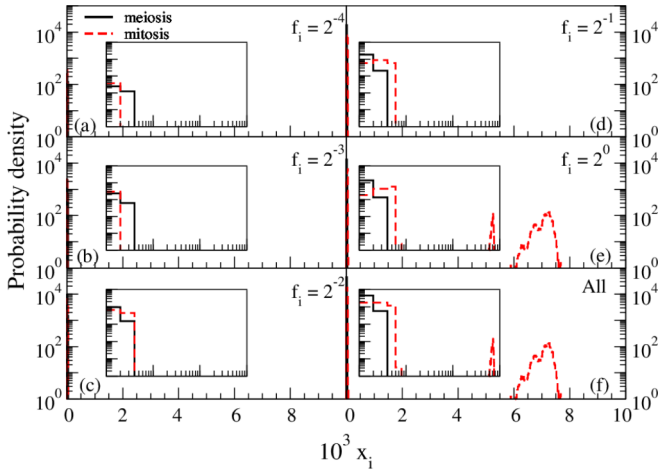

FIG. 4. Probability density of the stationary-state relative abundance of genotype $i \in A \cup B$ for $L = 4$ and $x_A \approx 0.04$ (i.e., mitosis dominates), with $p = 10^{-5}$, $r = 0.99$, and $\lambda = 2.84 \times 10^3$. Each of panels (a)–(e) refers to genotypes having the same number (k) of loci at which the mtDNA sequence differs from both ncDNA sequences, therefore having the same fitness $f_i = 2^{-k}$, as indicated. Panel (f) refers to all genotypes. All panels contain density spikes very near $x_i = 0$ for both mitosis- and meiosis-generated genotypes. These are highlighted in the logarithmic-scale insets, with x_i in the range 10^{-5}–10^{-2} and densities ranging as in the containing panels.

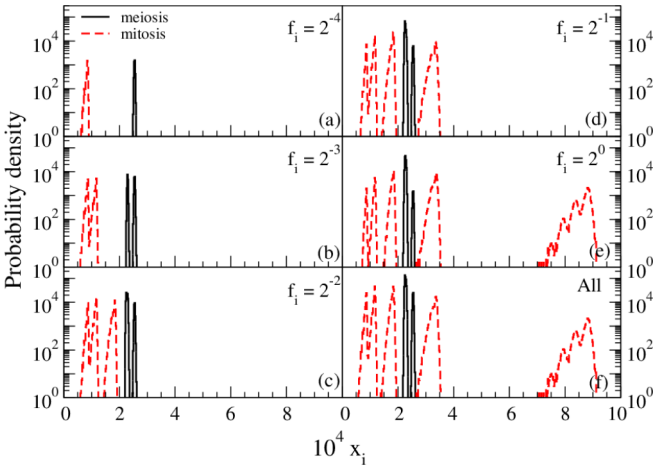

FIG. 5. Probability density of the stationary-state relative abundance of genotype $i \in A \cup B$ for $L = 4$ and $x_A \approx 0.5$ (i.e., mitosis and meiosis coexist at about the same proportion), with $p = 10^{-5}$, $r = 0.99$, and $\lambda = 938$. Each of panels (a)–(e) refers to genotypes having the same number (k) of loci at which the mtDNA sequence differs from both ncDNA sequences, therefore having the same fitness $f_i = 2^{-k}$, as indicated. Panel (f) refers to all genotypes.

set so that the desired steady-state value of x_A is reached. These figures are relative to $L = 4$ only, but contain a lot of further detail in that they include separate histograms for each of the $L + 1$ possible values of a genotype's fitness (2^{-L} through 2^0).

Figure 7 is given in the same spirit of the previous three, now containing plots for both $L = 3$ and 4. All its panels refer to the eventual preponderance of meiosis over mitosis ($x_A \approx 0.96$) for fixed p, with λ adjusted to support the desired steady-state value of x_A while r is varied.

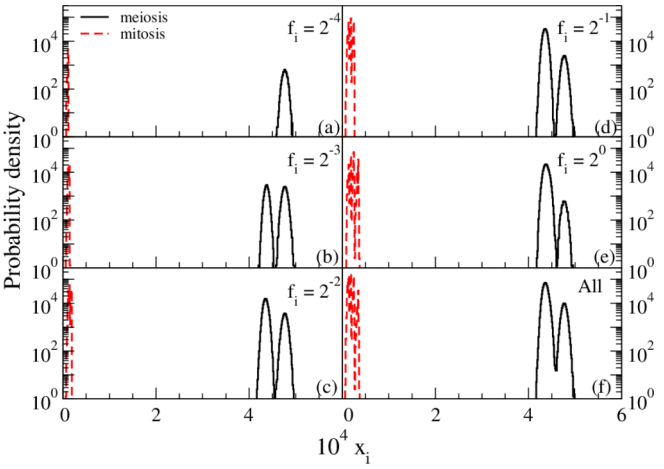

FIG. 6. Probability density of the stationary-state relative abundance of genotype $i \in A \cup B$ for $L = 4$ and $x_A \approx 0.96$ (i.e., meiosis dominates), with $p = 10^{-5}$, $r = 0.99$, and $\lambda = 27.7$. Each of panels (a)–(e) refers to genotypes having the same number (k) of loci at which the mtDNA sequence differs from both ncDNA sequences, therefore having the same fitness $f_i = 2^{-k}$, as indicated. Panel (f) refers to all genotypes. Panels (a)–(c) contain density spikes very near $x_i = 0$ for mitosis-generated genotypes.

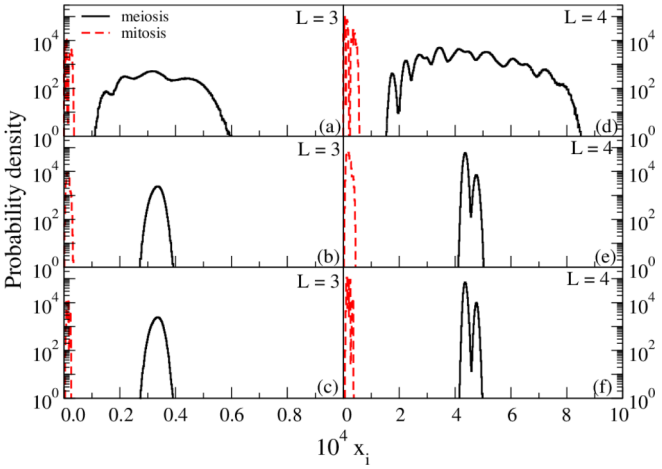

FIG. 7. Probability density of the stationary-state relative abundance of genotype $i \in A \cup B$ for $L = 3$ (a)–(c) and $L = 4$ (d)–(f), in both cases for $x_A \approx 0.96$. Probability p is fixed at $p = 10^{-5}$; probability r varies from $r = 0.01$ (a), (d), to $r = 0.9$ (b), (e), to $r = 0.99$ (c), (f); the value of λ for each panel is as follows: $\lambda = 3.15$ (a), $\lambda = 4.1$ (b), $\lambda = 4.05$ (c), $\lambda = 21.4$ (d), $\lambda = 27.7$ (e), $\lambda = 27.7$ (f). Panel (f) is identical to Fig. 6(f). Panels (a)–(c) contain density spikes very near $x_i = 0$ for mitosis-generated genotypes.

IV. DISCUSSION

Probabilities p and r influence network structure by affecting the probabilities $p_{j,i}$ and $p_{jk,i}$ that each hyperedge exists: decreasing p leads to sparser meiotic hyperedges, decreasing r leads to both sparser mitotic hyperedges and sparser meiotic hyperedges. They also influence network dynamics on a fixed hypergraph, both through the sparseness of the hyperedges incoming to any given genotype and by relativizing these hyperedges' importance (since now the normalized versions of $p_{j,i}$ and $p_{jk,i}$, respectively $q_{j,i}$ and $q_{jk,i}$, are the ones in charge). Even though we might expect the eventual steady state to depend on how p and r relate to each other, the plots in Fig. 3 indicate that, in fact, it depends much more strongly on the value of λ.

As explained in Sec. II G, λ is intended to complement the effect of squared fitnesses in lowering the rate of genotype production by meiosis when compared to that of mitosis (to which fitnesses contribute linearly). We see in Fig. 3 that, for $p = 10^{-5}$ and $r = 0.01$, using $\lambda = 3.15$ (for $L = 3$) or $\lambda = 21.4$ (for $L = 4$) ensures the eventual prevalence of meiosis over mitosis, while increasing these values slowly reverses the trend and eventually leads to the prevalence of mitosis in the steady state [Figs. 3(c) and 3(f)]. In each of Figs. 3(c) and 3(f), which are the only two inside which the value of λ varies from plot to plot, decreasing the value of λ not only contributes to the survival of meiosis-generated genotypes at ever higher levels in the steady state, it also leads such a steady state to be reached at a pace that is ever increasing as well. This is as expected, as per our comments in Sec. II G following the introduction of λ in Eq. (16).

Still regarding Fig. 3, for the values of λ leading mitosis and meiosis to coexist approximately in the same proportion (these are the default values for λ in the figure), varying r while p remains fixed at 10^{-5} [Figs. 3(b) and 3(e)] or varying p while r remains fixed at 0.01 [Figs. 3(a) and 3(d)] is practically innocuous. This remains true if the other values of λ in Figs. 3(c) and 3(f) are used instead (data not shown), suggesting that, even though $p \ll r$ is known to hold currently, it need not have held all along the evolutionary process. This obviates the need for Assumption 1 (see Sec. I), at least as far as predicting the eventual situation of mitosis- versus meiosis-generated genotypes is concerned. In fact, this is true of Assumption 2 as well since pushing the value of r very near 1 ($r = 0.99$) is practically tantamount to eliminating r from both Eqs. (4) and (12) and yet produces no relevant effect. The fact that the two assumptions are unnecessary in the face of our results by no means invalidates the spirit of the hypothesis we set out to explore in the first place, namely, that by evolving in combination with mtDNA, the inheritance of ncDNA through cellular division by mitosis can give rise to meiosis as the prevalent mechanism. In this case, the mitochondria are seen to act not as predicated by Assumptions 1 and 2, but by strongly influencing a genotype's fitness as well as the multitude of pathways through which evolution can work.

Of all the p/r ratios used in Fig. 3, we use the lowest ($p = 10^{-5}$ to $r = 0.99$) in Figs. 4–6, which in sequence refer to a progression in the eventual relative abundance of genotypes generated by meiosis (from very low in Fig. 4 to very high in Fig. 6, with λ adjusted accordingly all along). These figures refer to $L = 4$ and show the probability density of relative genotype abundances for each possible fitness value. By Eq. (2), 16 of the mitosis-generated genotypes have the least possible fitness (2^{-4}), followed by 128, 480, 896, and 656 genotypes as we move higher up. The same holds for the meiosis-generated genotypes.

What we see in Fig. 4 (which corresponds to the meiosis-generated genotypes reaching only a small relative abundance in the steady state, $x_A \approx 0.04$) is that only the very fittest of the mitosis-generated genotypes survive, that is, only about 30% of them. This indicates that coevolution with genotypes generated by meiosis exerts considerable pressure on those generated by mitosis even in conditions that are highly favorable to the latter. As the value of λ gets decreased to support $x_A \approx 0.5$ (Fig. 5), we start to see nonzero relative abundances at all fitness levels for both mitosis- and meiosis-generated genotypes. Nevertheless, some of the fittest mitosis-generated genotypes continue to dominate, occurring with the highest relative abundances in isolation, though at one full order of magnitude lower than previously. What supports $x_A \approx 0.5$ in the steady state even in these conditions is the fact that meiosis-generated genotypes are already winning at the three lowest fitness levels (roughly 29% of such genotypes), as well as the clash between mitosis- and meiosis-generated genotypes at the following fitness level (about 41% of the genotypes in each group). Decreasing λ considerably further leads to the steady-state scenario shown in Fig. 6, which corresponds to the rise of the meiosis-generated genotypes to dominance ($x_A \approx 0.96$). Readily, genotypes at all fitness levels contribute, which contrasts sharply with Fig. 4.

A further view of the probability densities associated with the relative abundance of genotypes is given in Fig. 7, aiming to complement Figs. 4–6 by including some $L = 3$ cases and by varying the value of r. The plots in Fig. 7 are all given

for $x_A \approx 0.96$ in the steady state, a setting similar to that of Fig. 6, now letting r vary from 0.01 [Figs. 7(a) and 7(d)], to 0.9 [Figs. 7(b) and 7(e)], to 0.99 [Figs. 7(c) and 7(f)]. As before, the value of λ has been calibrated to yield the desired x_A. As mentioned earlier, increasing the value of r leads to denser networks in relation to both the hyperedges leading to mitosis-generated genotypes and to those leading to meiosis-generated genotypes. As shown in Fig. 7, the effect of this is to progressively narrow the interval where nonzero densities are found, particularly so for the winning variety, that of meiosis-generated genotypes. This seems to be indicating that denser networks tend to homogenize relative abundances across genotypes generated in the same manner, at least to a certain extent.

A salient feature that is common to practically all panels in all of Figs. 4–7, though at varying degrees, is the presence of density peaks clustered together wherever nonzero densities occur. This is true almost exclusively of the $L = 4$ cases [in Figs. 4–6 and in Figs. 7(d)–7(f)], which seems to indicate something to be expected as L grows. Whether this is so, maybe in conjunction with further structural properties of the networks, has remained an elusive question and is the subject of further research.

To finalize, we return to the special case of Sec. II I, where we found F_0 to be a strict upper bound on the value of λ in order for meiosis-generated genotypes to prevail over mitosis-generated ones. This indication that lowering the value of λ in the special case is crucial to the prevalence of meiosis-generated genotypes is true of the general case as well, as demonstrated in Fig. 3. In fact, the analog of $\lambda < F_0$ in the general case is $\lambda < F$, which as we see in Figs. 3(c), 3(f), 6, and 7 (those in which the prevalence of meiosis over mitosis occurs), holds as well: by Eq. (3), we have $F = 185$ for $L = 3$, $F = 1\,241$ for $L = 4$, therefore substantially higher than the values of λ used in those figures. We find it remarkable that such constraint on the value of λ should remain essentially valid all the way up from the drastically simplified case of Sec. II I.

V. CONCLUSION

Our network model of how the mitotic and meiotic modes of cellular division may have coevolved, eventually giving rise to meiosis as the prevalent mechanism underlying reproduction in eukaryotes, includes an oversimplified representation of genotypes and only three parameters. Of these, two (probabilities p and r) aim to account for the stochastic variability in DNA during cellular division and the third (λ) aims to adjust the growth rate of genotype populations generated by mitosis to that of genotype populations generated by meiosis. Probability p is related to the recombination that ncDNA undergoes during meiosis, while probability r is related to mutations that mtDNA undergoes when transmitted from parent cell to offspring and is therefore present in division by both mitosis and meiosis.

In spite of its simplicity, we have found the model to yield surprisingly complex behavior and to contemplate a variety of scenarios regarding the steady-state coexistence of the two modes of cellular division. In particular, we have identified values of λ for which the genotypes generated by meiosis rise to dominance from initial conditions in which those generated by mitosis dominate absolutely, thus providing theoretical support to the most recent hypothesis as to why eukaryote reproduction has come to be dominated by meiosis as the most common mode of cellular division. Recall from Sec. I that at the core of this hypothesis is the random diversification of ncDNA afforded by meiosis as a means to catch up with that of mtDNA. Our three parameters have been meant precisely to capture this mismatch. Together with the various model details, they provide a mathematically sound view into what may have happened along evolutionary time.

However, our model's support to the said hypothesis is not without its caveats. Even though the hypothesis rests on two main assumptions (Assumptions 1 and 2), given in Sec. I, that place heavy weight on $p \ll r$ and on the inheritance of mtDNA solely from one of the two parents when cellular division happens by meiosis, our results indicate that neither of the two assumptions is essential. Instead, they suggest that, in association with mitochondria, cellular division by meiosis is naturally poised to navigate a fitness landscape heavily influenced by how the two forms of DNA interact much more efficiently than cellular division by mitosis. The crucial parameter is then λ since it provides a direct means to adjust the relative rate at which the two processes unfold.

Our conclusions still lack in robustness, though. Whether they will stand further stressing depends on our future ability to extend the analysis in Sec. II I to a less restrictive case or our computational results to substantially lengthier genotypes. Exactly how to achieve either end will be the subject of further research, but at this point we believe the most promising of the avenues to be the latter (compute on substantially lengthier genotypes). As indicated in Sec. III, the central difficulty to be addressed is the storage of a data structure representing a hypergraph whose size grows exponentially with sequence length L. This difficulty is shared by many other problems in data science and engineering, and in the case at hand it might be tempting to resort to Monte Carlo sampling not only to obtain hypergraph H (as indicated at the beginning of Sec. III) but also to reduce its number of nodes and/or hyperedges. Such a naïve approach would fall short of solving the problem since essentially it lacks some fundamental guarantees regarding the type and likelihood of the errors that would ensue. Fortunately, though, the said difficulty may be addressable in the near future through the use of probabilistic data structures that, at the cost of a modestly bounded probability of error, may be capable of handling fast increases in hypergraph size with only occasional failures. Initial progress has been made for the simpler case of graphs [33].

Of course, the ability to handle larger values of L will expose even more strongly the implausibility of our same-length assumption for ncDNA and mtDNA sequences. Handling this will require that we consider the nature of the so-called mitonuclear interactions, that is, the combined effect of mitochondrial- and nuclear-gene expression, more carefully. Doing this may make it possible for us to generalize the definition of fitness given in Sec II A to allow ncDNA sequences to be lengthier than mtDNA ones in our model. Recent studies have identified the convergence of multiple ncDNA and mtDNA genes (many more of the former than of the latter) onto regulating some essential mitochondrial

functions [34]. These discoveries are bound to provide the basis for a more suitable definition, for example, by pairing (in the sense of our current fitness definition) the shorter mtDNA sequence with several subsequences of the lengthier ncDNA sequences.

ACKNOWLEDGMENTS

We acknowledge partial financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and a BBP grant from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). We also acknowledge financial support from Agencia Nacional de Investigación e Innovación (ANII) and Programa de Desarrollo de las Ciencias Básicas (PEDECIBA). We thank Núcleo Avançado de Computação de Alto Desempenho (NACAD), Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE), Universidade Federal do Rio de Janeiro (UFRJ), for the use of supercomputer Lobo Carneiro, where part of the calculations were carried out.

[1] A. Karnkowska, V. Vacek, Z. Zubáčová, S. C. Treitli, R. Petrželková, L. Eme, L. Novák, V. Žárský, L. D. Barlow, E. K. Herman *et al.*, Curr. Biol. **26**, 1274 (2016).
[2] N. Osada and H. Akashi, Mol. Biol. Evol. **29**, 337 (2012).
[3] F. S. Barreto and R. S. Burton, Mol. Biol. Evol. **30**, 310 (2012).
[4] D. B. Sloan, D. A. Triant, M. Wu, and D. R. Taylor, Mol. Biol. Evol. **31**, 673 (2013).
[5] A. Latorre-Pellicer, R. Moreno-Loshuertos, A. V. Lechuga-Vieco, F. Sánchez-Cabo, C. Torroja, R. Acín-Pérez, E. Calvo, E. Aix, A. González-Guerra, A. Logan *et al.*, Nature (London) **535**, 561 (2016).
[6] C. Tomasetti, L. Li, and B. Vogelstein, Science **355**, 1330 (2017).
[7] S. Billiard, M. López-Villavicencio, M. E. Hood, and T. Giraud, J. Evol. Biol. **25**, 1020 (2012).
[8] Y. Y. Yang and J. G. Kim, J. Ecol. Environ. **40**, 12 (2016).
[9] S. P. Otto and T. Lenormand, Nat. Rev. Genet. **3**, 252 (2002).
[10] M. Hartfield and P. D. Keightley, Integr. Zool. **7**, 192 (2012).
[11] A. Livnat and C. Papadimitriou, Commun. ACM **59**, 84 (2016).
[12] J. C. Havird, M. D. Hall, and D. K. Dowling, Bioessays **37**, 951 (2015).
[13] E. Pennisi, Science **353**, 334 (2016).
[14] E. Cohen, D. A. Kessler, and H. Levine, Phys. Rev. Lett. **94**, 098102 (2005).
[15] J.-M. Park and M. W. Deem, Phys. Rev. Lett. **98**, 058101 (2007).
[16] D. B. Saakian, J. Stat. Phys. **128**, 781 (2007).
[17] D. B. Saakian and C.-K. Hu, Phys. Rev. E **88**, 052717 (2013).
[18] D. B. Saakian, Phys. Rev. E **97**, 012409 (2018).
[19] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[20] M. Eigen and P. Schuster, Naturwissenschaften **64**, 541 (1977).
[21] C. K. Biebricher and M. Eigen, in *Quasispecies: Concept and Implications for Virology*, Vol. 299 of Current Topics in Microbiology and Immunology, edited by E. Domingo (Springer, Berlin, 2006), pp. 1–31.
[22] E. Domingo, Contrib. Sci. **5**, 161 (2009).
[23] A. S. Lauring and R. Andino, PLoS Pathog. **6**, e1001005 (2010).
[24] A. Más, C. López-Galíndez, I. Cacho, J. Gómez, and M. A. Martínez, J. Mol. Biol. **397**, 865 (2010).
[25] V. C. Barbosa, R. Donangelo, and S. R. Souza, J. Theor. Biol. **312**, 114 (2012).
[26] V. C. Barbosa, R. Donangelo, and S. R. Souza, J. Stat. Mech. (2015) P01022.
[27] V. C. Barbosa, R. Donangelo, and S. R. Souza, J. Stat. Mech. (2016) 063501.
[28] C. Berge, *Hypergraphs* (North-Holland, Amsterdam, 1989).
[29] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, Discrete Appl. Math. **42**, 177 (1993).
[30] J. L. Gerton and R. S. Hawley, Nat. Rev. Genet. **6**, 477 (2005).
[31] C. Haag-Liautard, N. Coffey, D. Houle, M. Lynch, B. Charlesworth, and P. D. Keightley, PLoS Biol. **6**, e204 (2008).
[32] R. S. Cha, B. M. Weiner, S. Keeney, J. Dekker, and N. Kleckner, Genes Dev. **14**, 493 (2000).
[33] A. McGregor, SIGMOD Rec. **43**, 9 (2014).
[34] J. N. Wolff, E. D. Ladoukakis, J. A. Enríquez, and D. K. Dowling, Philos. Trans. R. Soc. London, B **369**, 20130443 (2014).