International Journal of Modern Physics C
Vol. 20, No. 8 (2009) 1299–1309

World Scientific
www.worldscientific.com

PHASE TRANSITION IN THE GENOME EVOLUTION FAVORS NONRANDOM DISTRIBUTION OF GENES ON CHROMOSOMES

JAKUB KOWALSKI, WOJCIECH WAGA, MARTA ZAWIERTA
and STANISŁAW CEBRAT*

Department of Genomics, Faculty of Biotechnology
Wrocław University, ul. Przybyszewskiego 63/77 51-148, Wrocław, Poland
**cebrat@smorfland.uni.wroc.pl*

Received 8 January 2009
Accepted 27 April 2009

We have used the Monte Carlo-based computer models to show that selection pressure could affect the distribution of recombination hotspots along the chromosome. Close to the critical crossover rate, where genomes may switch between the Darwinian purifying selection or complementation of haplotypes, the distribution of recombination events and the force of selection exerted on genes affect the structure of chromosomes. The order of expression of genes and their location on chromosome may decide about the extinction or survival of competing populations.

Keywords: Monte Carlo simulation; chromosome structure; evolution; complementation; recombination; competition.

1. Introduction

According to neo-Darwinism there are three forces driving the evolution of biological systems: mutations, recombinations, and selection, of which the only directional force is selection favoring the fittest individuals.[1] Nevertheless, there is a very sophisticated interplay between these forces. Selection favors some structures where recombination events as well as mutations are not totally random. For example, in nature there are known genetic elements which can transfer specific blocks of information inside the genome or between genomes. It is also known, that the direction of gene on the DNA molecules decides about its mutation probability (for review see Ref. 2). In fact, recombination and mutations could be highly biased in both frequency and location.

The Monte Carlo (MC) modeling of the genome evolution in the sexually reproducing populations has revealed that depending on the intragenomic recombination rate and effective population size (or more precisely inbreeding coefficient), genomes can switch between two different strategies of evolution — the Darwinian purifying selection and complementation of haplotypes.[3,4] The strategy of purifying selection is advantageous in large, randomly mating, panmictic populations with high

intragenomic recombination rate, where the genetic relations between sexual partners are very low (so-called Mendelian populations). In such situations selection eliminates effectively individuals with homozygous defective loci and the fraction of defective, recessive alleles in the genetic pool of population is relatively low. In small populations, on the other hand, the probability for an individual of inheriting the long strings of genes from both parents originated from the common ancestor is higher. Under such condition the other strategy is more advantageous — complementing the whole strings of defective alleles. The situation is illustrated in Figs. 1 and 2. The complementing strategy enables the accumulation of a higher fraction of defective alleles in the genetic pool of the population. The second possibility is very often overlooked and even recently some models of biological evolution are built on the base of so-called Wright–Fisher (W–F) model[5–7] of Mendelian population, assuming that genetic loci are not linked and alleles in different loci inherit independently.[8,9]

It is obvious that switching between the two strategies depends on the evolutionary costs and reproduction potential of populations. It has been found that such a switching has a character of phase transition.[10] For a given set of parameters of evolving populations — population size, length of chromosomes, mutation rate — the critical intragenomic recombination rate can be found. Below this recombination rate the complementing strategy is more advantageous while above this critical recombination rate the purifying strategy ensures a higher reproduction potential of the population. Population size influences the value of critical recombination

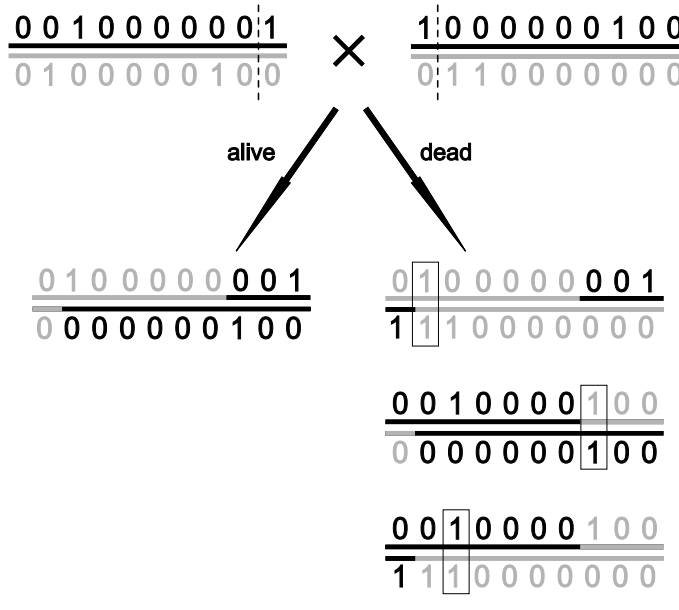

Fig. 1. The purifying selection strategy. After recombination, the surviving offspring may have less defects than parental genomes.

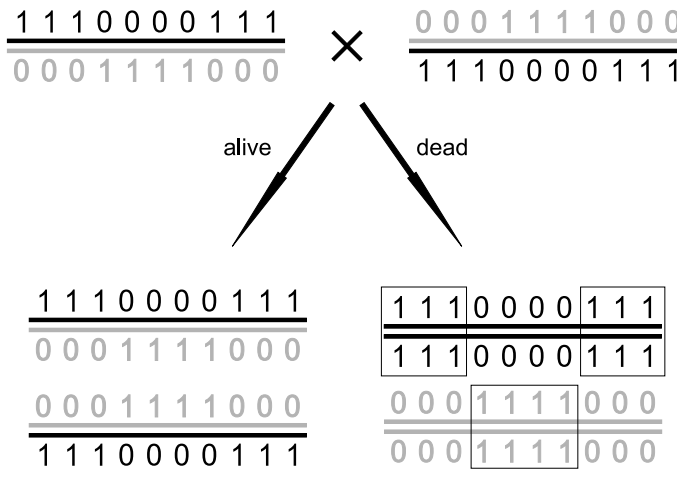

Fig. 2. The complementing strategy. In this example fraction of defective genes reaches 0.5, but the probability of the offspring survival is still relatively high. Please note that there is no recombination.

rate only if the panmixia (random selection of mates) is assumed. In such populations some power law relations between the value of critical recombination rate and population sizes have been found.[10] If some constraints are imposed on the totally random selection of mates in the panmictic populations (i.e. distance limit between mates in the spacially distributed populations), the value of critical intragenomic recombination rate depends on the inbreeding coefficient. Modeling the evolution of populations on the lattice, when the distances for looking for the mates and for placing the offspring are the constant parameters of the model, the value of the critical intragenomic recombination rate does not depend on the population size.[4] Furthermore, simulations on lattices revealed that under some conditions a restriction imposed on recombination rate increases the population expansion and enables the sympatric speciation.[4]

In this paper we are studying the functional properties of chromosomes in the Penna model under recombination rate close to critical values. The Penna model has been used in many simulations of age structured populations.[11] It is very convenient for such studies because selection values of genes differ in the model depending on the period of life when they are expressed. In particular we now study the role of distribution of the recombination hotspots along the chromosomes.

2. Model

Detailed descriptions of the standard diploid Penna model can be found in many original papers[12] or reviews.[13,14] We have used the modified, diploid version of the Penna model with sexual reproduction. Panmictic populations are composed of N individuals represented by their diploid genomes. Each genome is composed of one pair of haplotypes (bitstrings) $L = 128$ bits long. Bits correspond to genes and, if

they are located at the same positions in both bitstrings, they correspond to alleles. Bits set to 0 correspond to the wild — correct alleles, bits set to 1 correspond to the defective alleles. A defective phenotype is determined only if both bits at the corresponding positions in both bitstrings are set to 1. Bits (genes) are switched on (expressed) chronologically. In the standard Penna model, during each MC step one bit in each bitstring is switched on; in the first step the first bit, in the second step the second bit and so on. In our version the sequence of switching on the bits can be modified. It will be described in detail in the next section. Anyway, the number of switched-on bits determines the age of an individual. When the individual reaches the age $R = 80$ it can start to reproduce. In one MC step, each female at or above this minimum reproduction age can give birth to B offsprings. She copies her genome introducing a new mutation into each haplotype in the random position with probability M. If the bit chosen for mutation is 0 it is changed for 1, if it is 1 it stays 1 (there are no reversions). Two copied and mutated bitstrings are paired and recombined with probability C in a random position, mimicking crossover. One product of recombination of each pair of chromosomes is randomly assorted to the female gamete. Then, the female looks for a male with age $a \geq 80$, who produces a male gamete in the same way. Male and female gametes fuse forming a diploid genome of a newborn. Its first bit pair will be switched on in the next MC step. The male individual, after reproduction with one female, returns to the pool of males at the reproduction age and it can produce an offspring with other female(s) in the same MC step. The scheme of the process is shown in Fig. 3. As mentioned above, if both bits at corresponding positions are set to 1 (are defective) they determine the phenotype to be deleterious when switched on. If the number of

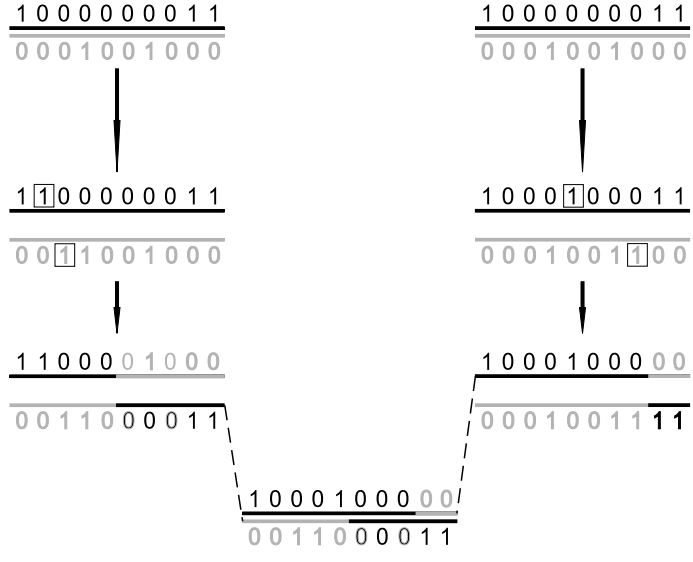

Fig. 3. The scheme of reproduction process.

switched-on defective phenotypes reaches the declared threshold T — the individual dies because of genetic death. The size of panmictic populations is controlled by a Verhulst factor operating in our version of the model only at birth, as proposed in Ref. 15. The Verhulst survival probability is $V = 1 - N_t/N_{\max}$, where: V — survival probability of the newborn, N_t — the actual size of the population, and $N_{\max}$ is called sometimes the maximum capacity of the environment.

3. Genome Evolution Far from the Critical Point of Recombination

In Figs. 4 and 5 the age and the genetic structure of populations simulated with the standard diploid Penna model are shown. Parameters of simulations were: one pair of bitstrings, $L = 128$, $N_{\max} = 10\,000$, $R = 80$, $M = 1$, $T = 1$, $B = 2$, and $C = 1$ or 0. The results show that populations evolving under recombination rate 1 are larger, they have a lower fraction of defective alleles in the part of genomes expressed before the reproduction age (up to 80), a larger fraction of individuals at reproduction ages and a higher maximum lifespan. Nevertheless, populations evolving without crossover survived, though they had very high fractions (0.5) of defective genes expressed before the minimum reproduction age. If we assume that the defective genes are distributed randomly in the pool of haplotypes, the average survival probability until the minimum reproduction age is of the order of 4×10^{-8}. Under such conditions, populations should be extinct. In fact, in the whole genetic pool only two different and complementing haplotypes exist apart from minor mutations. The effect of accumulation of defects in the diploid version of the Penna model was observed previously.[16] Furthermore, it has also been found that

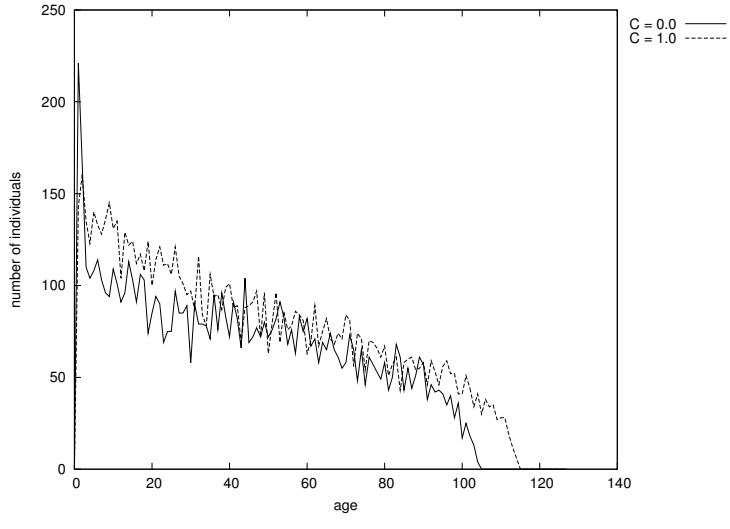

Fig. 4. The age structure of populations.

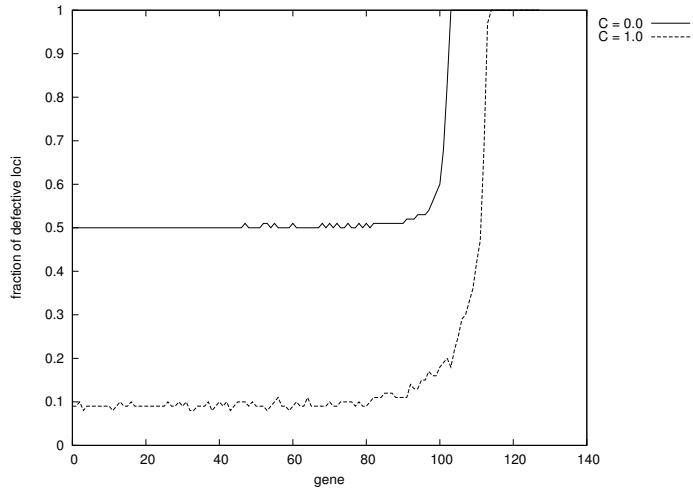

Fig. 5. The genetic structure of populations.

gamete recognition — the process in which complementing gametes preferentially form zygotes — increases the reproduction potential of populations.[17]

Intuitively, it seems that under fixed parameters of simulations the only condition which influences the fraction of defective alleles is the period of life when the genes are expressed. The fraction of defects should not depend on the position of a particular gene in the chromosome. To check this we have performed simulations with an inverted second part of chromosomes (bits from 42 to 128). Now, the bit expressed in the 42nd MC step is located at the end of chromosome. We have simulated the evolution of populations with the standard order of bits (not inverted) and with the inverted second part of bitstrings under the same conditions (the same values of the rest of the parameters). The parameters describing the obtained populations (size of populations, fraction of defective genes, maximum lifespan) suggest that the structure of chromosomes — the co-linearity or the lack of co-linearity between the physical sequence of genes on chromosomes and order of their expression do not influence the evolutionary value of population at least under the studied crossover rates, $C = 1$ and 0. Nevertheless, in the next sections it will be shown that the fitness of the two populations with different structures of chromosomes are equivalent only far from the critical value of intragenomic recombination rate.

4. The Role of Chromosome Structure Close to the Critical Point of Recombination

The critical crossover rate for population without inversion of bits was found. All parameters of simulations but the crossover rate were constant. Populations switch their strategy of genomes evolution from the complementing haplotypes to the purifying selection at a recombination rate of about 0.075. There is a characteristic

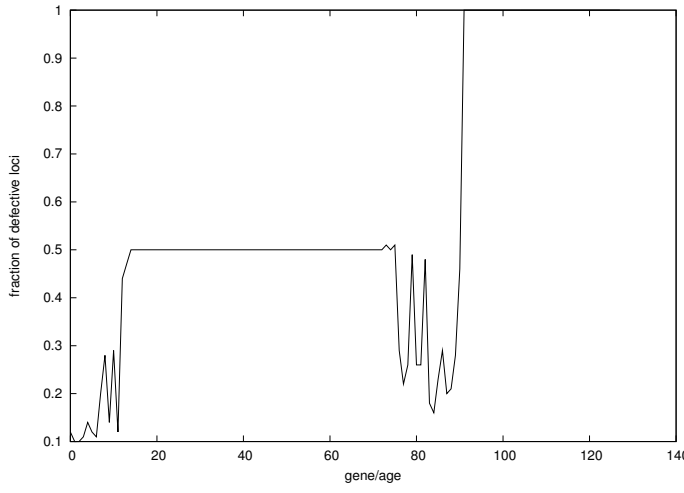

Fig. 6. The population's genetic structure. $L = 128$, $R = 80$, $T = 1$, and $C = 0.075$.

distribution of defective genes along the chromosome under critical crossover rate — the first bits of the bitstrings and the bits expressed at the beginning of the reproduction period have lower fraction of defects, it seems that these bits are at least partially under the purifying selection (Fig. 6).

Notice that results shown in Fig. 5 suggest that all genes expressed before the reproduction age are under the same selection pressure. Nevertheless, the selection pressure weakens gradually for genes expressed after the minimum reproduction age because the defective genes may be transferred to the offspring before the expression of the defective phenotype. If a sequence of expression of bits is like in the standard Penna model (no inversions), then the chromosome is asymmetric — at one end (beginning of the bitstring) the bits expressed early during the life span and very strongly watched by selection are located, while at the other end the "empty space" can be found. "Empty" means that all genes located at this end are defective. There is no genetic information in this space which could help the individual to survive. Can we find any role for this space?

In real genomes, if two markers (genes) are physically very close, they are usually inherited together because recombination has to happen just between them to separate them and only in such a rare case they can be transferred to separate gametes. It is said that they are linked. If two genes are far away from each other, the probability of recombination between them is higher and they can be inherited independently with higher probability. That is why we have repeated the simulations with inverted parts of chromosomes, close to the critical recombination rate. In the standard structure of Penna chromosomes (not inverted) the empty space is at the end of chromosome and any recombination in this space has no effect on the probability of disruption of any linked, functional group of genes. In the inverted chromosome, the "empty space" is in its central part. Recombination in this part

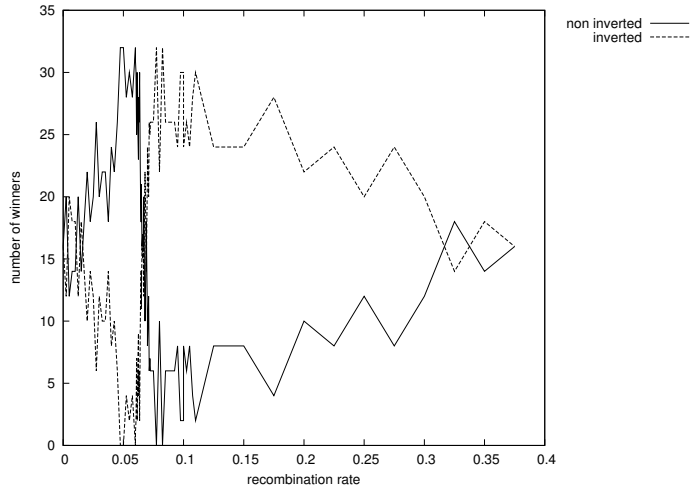

Fig. 7. Number of winning and losing competitions for populations with standard genome structure and populations with "inverted" regions of genomes.

increases the probability of independent transfer of genes located at the two ends of the chromosome to the gametes and consequently, to the zygotes.

To show that the location of this empty space affects the evolutionary value of populations we have simulated the competition of pairs of populations with different structures of chromosomes, that is, with and without inversion of their second parts. For the first 400 000 MC steps, populations evolved independently, each in its own environment under independently operating Verhulst factor. After 400 000 MC steps two populations (one with inverted chromosome and the other one without inversion) were put into one environment with doubled $N_{\max}$ but with one Verhulst factor operating for both populations, though the populations cannot crossbreed, they behave like a different species competing in the same ecological niche. We have performed such competitions for 32 pairs of populations, each evolving under the same crossover rate. The results are shown in Fig. 7. For low recombination rates — close to 0, as well as for high recombination rates — close to 1, the inversions of chromosomes have no significant influence on the fitness of the populations, sometimes the populations with the standard chromosomes are winning and sometimes the populations with the inverted parts of chromosomes are winning. In contrast, the structure of chromosomes decides about winning/losing close to the critical recombination rates. When the recombination rate increases from small C approaching its critical value, the populations with inverted chromosomes are losing. Then, the role is suddenly switched and populations with inverted chromosomes become the winners. It can be interpreted as: The effective values of recombinations in the inverted chromosomes are higher and they reach their critical value earlier. Since the reproductive potential at the critical point is the lowest, the inverted population loses. Under only a slightly higher recombination rate the population with inverted

chromosomes is already "above" the critical point under purifying selection while the one without inversions just enters the critical range and it is losing now.

To support this interpretation we have performed simplified simulations without chronologically switching on the bits. Individual genomes were composed of two bitstrings 128 bits long. Only 100 of them were under selection, mutations in the rest of them (28) were neutral. These 28 bits were located at the end of chromosome (corresponds to the standard Penna chromosome without inversion) or in the middle of chromosome (corresponds to the chromosome with inversion). The populations evolved for the first 100 000 MC steps with sexual reproduction. At each MC step 5% of individuals were killed randomly and the gap was filled up by offspring generated by the surviving individuals. The newborns survived if their genomes had no loci with both alleles defective. For such populations, we have checked how the probability of generating the surviving offspring depends on the recombination rate. Results are presented in Fig. 8. The critical value of recombination rate was lower for populations with the neutral region located in the middle of chromosomes. Under the critical recombination rate the probability of generating the surviving offspring is the lowest. Nevertheless, only a slight increase of the recombination rate shifts this population into the purifying strategy with much higher probability of generating the surviving offspring. The population with the neutral region located at the end of chromosomes reaches the critical point of recombination (with the lowest probability of the offspring survival) when the population with the neutral part in the middle of chromosome is already under purifying selection. Thus, the neutral regions at the end give a higher critical value than neutral regions in the middle of chromosomes. The plots in Fig. 8 show the regions where two populations

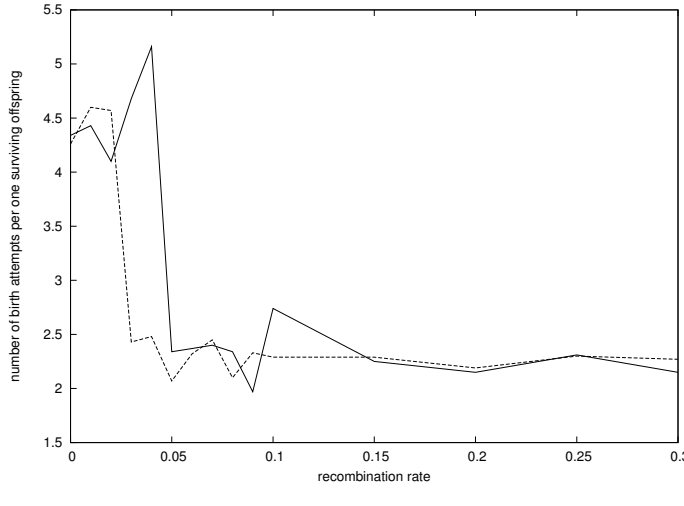

Fig. 8. Results of simplified simulations without chronological switching on bits. Dashed line for the genomes with neutral region in the middle of the bitstrings, solid line for the neutral region at the end of the bitstrings.

significantly differ in the reproduction potential and the difference depends exclusively on the recombinational structure of chromosomes.

Along the natural chromosomes, the recombination events are not evenly distributed. There are some regions with very high recombination frequency (called recombinational hotspots) and some regions with very low frequency of recombination (called recombinational deserts).[18–21] The location of the "empty space" in the middle of the inverted Penna chromosome disrupts the genetic linkage between genes expressed before the minimum reproduction age. This "empty space" may be interpreted as a recombination hotspot. In the natural chromosomes hotspots do not need to be physically long, they could be just places promoting the recombination. Understanding the distribution of hotspots and its relations to the selection pressure on adjacent genes may be very important for successful genetic manipulations with eukaryotic chromosomes and genomes.

5. Conclusions

The results of computer simulations have shown that the distribution of recombination events along the chromosomes plays an important role in the evolution of genomes and populations and it depends on the selection pressure exerted on the genes separated by recombination. Since the Penna model produces a gradient of selection pressures, it seems that it can be used for simulating the self-organization of the chromosomes' structures.

Acknowledgments

We thank D. Stauffer for comments and discussions. The work was done in the frame of European programs: COST Action MP0801, FP6 NEST — GIACS and UNESCO Chair of Interdisciplinary Studies, University of Wrocław. Calculations have been carried out in Wrocław Center for Networking and Supercomputing (http://www.wcss.wroc.pl), Grant #102.

References

1. F. J. Ayala and J. A. Kiger, *Modern Genetics* (The Benjamin/Cummings Publishing Company, Inc., 1980).
2. D. Mackiewicz and S. Cebrat, *To Understand Nature, From Genetics to Mathematics*, eds. J. Miekisz and M. Lachowicz (World Scientific, 2009).
3. M. Zawierta, P. Biecek, W. Waga and S. Cebrat, *Theory Biosci.* **125**, 123 (2007).
4. W. Waga, D. Mackiewicz, M. Zawierta and S. Cebrat, *Theory Biosci.* **126**, 53 (2007).
5. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
6. S. Wright, *Genetics* **16**, 97 (1931).
7. S. Wright, in *Proceedings of the 6th International Congress of Genetics*, Vol. 1 (1932), p. 356.
8. C. Som, H. C. Bagheri and H.-U. Reyer, *BMC Evol. Biol.* **7**, 80 (2007).
9. R. J. Redfield, *Nature* **369**, 145 (1994).

10. M. Zawierta, W. Waga, D. Mackiewicz, P. Biecek and S. Cebrat, *Int. J. Mod. Phys. C* **19**, 917 (2008).
11. T. J. P. Penna, *J. Stat. Phys.* **78**, 1629 (1995).
12. A. Łaszkiewicz, S. Cebrat and D. Stauffer, *Adv. Complex Syst.* **8**, 7 (2005).
13. S. M. de Oliveira, P. M. C. de Oliveira and D. Stauffer, *Evolution, Money, War and Computers* (Stuttgard-Leipzig, Teubner, 1999).
14. D. Stauffer, S. M. de Oliveira, P. M. C. de Oliveira and J. S. S'a Martins, *Biology, Sociology, Geology by Computational Physicists* (Amsterdam, Elsevier, 2006).
15. J. S. S'a Martins and S. Cebrat, *Theory Biosci.* **119**, 156 (2000).
16. A. Łaszkiewicz, S. Szymczak and S. Cebrat, *Int. J. Mod. Phys. C* **14**, 765 (2003).
17. S. Cebrat and D. Stauffer, *Int. J. Mod. Phys. C* **19**, 259 (2008).
18. A. Yu *et al.*, *Nature* **409**, 951 (2001).
19. M. J. Daly, J. D. Rioux, S. F. Schaffner, T. J. Hudson and E. S. Lander, *Nature Genet.* **29**, 229 (2001).
20. A. J. Jeffreys, J. Kauppi and R. Neumann, *Nature Genet.* **29**, 217 (2001).
21. N. Arnheim, P. Calabrese and M. Nordborg, *Am. J. Hum. Genet.* **7**, 5 (2003).