Available online at www.sciencedirect.com
ScienceDirect
Theory in Biosciences 125 (2007) 123–132
Theory in Biosciences
www.elsevier.de/thbio

The role of intragenomic recombination rate in the evolution of population's genetic pool

Marta Zawierta, Przemysław Biecek, Wojciech Waga, Stanisław Cebrat*

Department of Genomics, Wrocław University, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland

Abstract

In a simple computer model of population evolution, we have shown that frequency of recombination between haplotypes during the gamete production influences the effectiveness of the reproduction strategy. High recombination rates keeps the fraction of defective alleles low while low recombination rate or uneven distributed recombination spots change the strategy of genomes' evolution and result in the accumulation of heterozygous loci in the genomes. Even short fragment of chromosome with restricted recombination influences the genetic structure of neighboring regions.

Keywords: Monte Carlo simulation; Population evolution; Recombination; Mutation; Recessiveness

1. Introduction

There is an assumption, widely accepted by neo-Darwinists, that the biological evolution is driven by two kinds of forces: random-mutational pressure and recombination processes producing raw biological material, and directional-selection

*Corresponding author. Tel.: +48 71 3756303; fax: +48 71 3252151.
E-mail address: cebrat@microb.uni.wroc.pl (S. Cebrat).

1431-7613/$ - see front matter © 2007 Elsevier GmbH. All rights reserved.
doi:10.1016/j.thbio.2007.02.002

choosing the best fitted individuals among products of those random processes (Ayala and Kiger, 1980). In fact, the mutational pressure and recombinations are highly biased in both the frequency and localization in the genomic sequences. Mutations very rarely are of advantage; instead, usually they are neutral or deleterious. Elimination of all mutational processes (fortunately impossible) would freeze the genetic information and then, even recombinations would not help to avoid the loss of biological diversity. On the other hand, a too high frequency of mutations leads to the genetic meltdown of populations when the elimination of deleterious mutations is too costly or cannot be compensated by high reproduction rates (de Oliveira, 2006). In the computer models of evolution of coding sequences, when the biodiversity or divergence rate is a measure of the evolution rate, the order of one mutation per genome per generation could be observed, which the optimum of mutational pressure is Dudkiewicz et al. (2005). Similar results have been obtained by the phenomenological analysis of the effects of mutational pressure (Azbel, 1999).

A possible fate of any asexually reproducing (no recombination), finite size population, even with moderate mutation rate is, the genetic meltdown which could be considered as a Muller ratchet effect (Coe and Mao, 2005). Recombinations could help the populations to avoid the genetic meltdown. It is thought that recombination speeds up the evolution, because it decouples the deleterious genes from the advantageous ones. The mutant genes pooled in one haplotype could be eliminated together by one genetic death, lowering the evolutionary costs. Without recombination, the Hill-Robertson effect of genetic hitchhiking (Kliman and Rey, 1993) would neutralize beneficial mutations by closely linked harmful genes. Simply, recombination can produce a recombinant, which is better than both parental forms.

When analyzing the role of recombination, it is very important to consider its frequency and distribution in the genomes. Recombination—mainly crossing-over in the diploid genomes—is a complicated, multi-step process, which could be genetically risky, and, to some extent, it is associated with mutations (Gorlov and Gorlova, 2001). Thus its frequency cannot be freely increased (Stauffer and Cebrat, 2006). Cytological studies have shown that the frequency of recombination in the human genomes varies for males and females and for different chromosomes and 50–100 recombination events occur per meiosis. The probability of recombination between two points in the chromosome is used by geneticists as a measure of "genetic distance" between the loci. If the probability of recombination between two loci equals 1% it corresponds to the distance called 1 centiMorgan (cM). The length of the human genome corresponds to a few thousands cM, which means that on average there are about ten genes per cM. But in fact, genetic distances measured in centiMorgans do not correspond to the physical distances measured in nucleotides. Considering the number of cross-over events and the total length of the human genome one can calculate that the expected average physical size of 1 cM should correspond to 1 million base pairs (1 Mbp). The comparative analyses of genomic sequences confirm the results of the cytological studies concerning the frequency of recombination, but reveal highly uneven distribution of recombination events—very

conserved regions, where recombination seems to be forbidden, and some other regions where cross-over occurs more often than expected from simple calculation (Yu et al., 2001). Thus, some regions are called recombination "deserts", while some other regions are called the recombination hot spots.

In this paper, we have analyzed the role of such uneven distribution of recombination along the chromosome on the processes of whole populations' evolution. In our context, we discuss only the reciprocal recombination, when homologous regions of chromosomes are exchanged.

2. Model

In the standard version of our model, the population is composed of $N = 1000$ individuals, each represented by its diploid "genome" corresponding to two bitstrings (haplotypes) $L = 100$ bits long. Bits set to 0 or 1 correspond to the correct (wild) or defective genes, respectively. Only, if both genes are set for 1 at the same position on the bitstrings, the phenotype of the "locus" is defective (all defects are recessive) and the individual dies because of a genetic death. Thus, one single active mutation causes death. Additionally, during each Monte Carlo step 2% of population dies because of random death. The emptied space after individuals' death could be occupied by the newborns. To give birth, the female genome is replicated and during the replication a new mutation is introduced into each copy of bitstring with probability M into randomly chosen locus. If the value of the locus is 0 it is replaced by 1, if it is 1 it stays 1 (there are no reversions). The two copies of bitstrings recombine with probability C at the randomly chosen point just by exchanging the corresponding arms. After these processes, each of the two new bitstrings corresponds to the gamete. Randomly chosen gamete is joined with another one produced in the same way from a randomly chosen male individual. The pair of gametes corresponds to the newborn's diploid genome. The newborn's sex is established with equal probability to be male or female.

Any modifications of the model have been noted in the results section.

3. Results and discussion

3.1. Positive selection for heterozygosity

At the first stage of studies we have analyzed the effect of selection for heterozygosity on the accumulation of defective genes. We have assumed that, like in case of sickle cell anemia or cystic fibrosis, heterozygous loci are advantageous for carriers while defective homozygous loci are lethal (individuals with both defective alleles in the same locus die, while a single sickle-cell allele helps against malaria). In this modification of the model, the reproduction probability of individuals grew

linearly with the number of heterozygous loci in their genomes. First, we have tried three different strategies of reproduction without recombination:

1. Cloning—the randomly chosen parental genome is replicated and mutations are introduced into each new haplotype.
2. Two parental individuals (male and female) are randomly chosen, from the first, the first bitstring is taken for replication and from the second, the second bitstring is replicated. The two copies, after mutation, form the newborn's genome. Note that this strategy divides the whole pool of bitstrings into two subsets: the "first" one and the "second" one. Thus, each genome in the population has one bitstring from the first set and the second bitstring from the second set. One can say that this strategy is possible if the haplotype recognition is assumed.
3. Two parental individuals (female and male) are randomly chosen, from each genome one bitstring is randomly chosen for replication and both bitstrings after mutation are joined to form a newborn's genome. Note, that there are not two separate sets of bitstrings.

The evolution of heterozygosity in the genomes of populations is shown in Fig. 1. The results of cloning are trivial—the equilibrium, where all genomes were heterozygous in all loci is established quickly. The second strategy leads to the same equilibrium, two sets of complementary haplotypes evolve and combination of two haplotypes—one from each set forms diploid genome with all heterozygous loci. The results of the third strategy do not seem to be so intuitive, but in equilibrium, like in the second strategy, the fraction of defective loci in the genetic pool of the population reaches 0.5. Since even 1 locus with both alleles

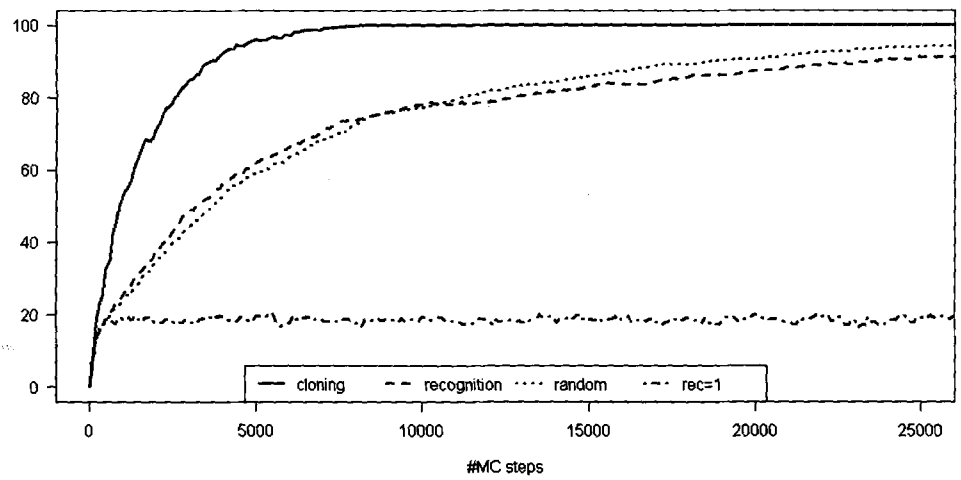

Fig. 1. Dynamics of heterozygosity in simulations with positive selection for heterozygous loci. Strategies of reproduction: without recombination; cloning, haplotype recognition (recognition), randomly chosen haplotypes (random) and, recombination rate 1 (rec = 1) per gamete production: The y-axis represents the number of heterozygous loci and the x-axis time of simulations in MC steps. See Section 3.1 for more explanation.

defective kills the individual, the probability of killing the individual because of the genetic structure of single locus is 0.25. The survival probability of a newborn with 100 loci is of the order of 10^{-13}. Thus, we have checked the Hamming distances between the bitstrings in the whole population. In this case, Hamming distance is the sum of differences between the values of bits in all corresponding loci in a given pair of bitstrings. We have counted Hamming distances for all possible pairs of bitstrings in the population. The final distribution shows only two values, 0 and 100, which means that the third strategy also leads to the selection of two complementing haplotypes. We will call them H1 and H2. There are two possible lethal combinations of haplotypes: (H1/H1 or H2/H2) and two surviving combinations (H1/H2, H2/H1).

In this strategy, the survival probability of a newborn is 0.5, even though with that size of the genome and 50% of defective genes (with random distribution of defects along the bitstrings), the surviving probability would be of the order of 10^{-13}. Selection drives the genome structure of population away from the random distribution of defective genes to the highly correlated structures where the probability of reproduction is much higher. Obviously, this is possible without recombination between haplotypes. That is why we have checked the fourth strategy—with recombination.

4. Before copying the bitstring from the parental genome as in the third strategy, one recombination event happens between parental bitstrings at a randomly chosen point.

In this reproduction strategy the fraction of defective genes in population drops (Fig. 1) and we have not observed complementing sets of haplotypes. The Hamming distance distribution is Gaussian.

3.2. No positive selection for heterozygosity

One can argue that the emergence of sets of complementing haplotypes in simulations described in the above section is a result of positive selection for heterozygous loci. Thus, we have simulated the evolution of populations without the assumption of higher reproduction potential of heterozygotes. Simulations were performed according to the standard model. In this version, it is indifferent for selection if there are two wild alleles (0/0) in the locus or the locus is heterozygous (1/0). In Fig. 2, we have presented the relation between the fraction of defective alleles in the genome and the recombination rate. Decreasing the recombination rate from 1 to 0, we have observed the transition from relatively low frequency of defective genes, to the frequency 0.5 at the recombination rate about 0.15.

Like in cases with positive selection for heterozygosity, we have observed the emergence of complementing haplotypes—the transition from low frequency of defective genes is connected with emergence of negative correlations between two haplotypes in the genomes. Nevertheless, now, these phenomena are not connected with any preferences for heterozygous states. In Fig. 3 we have presented the

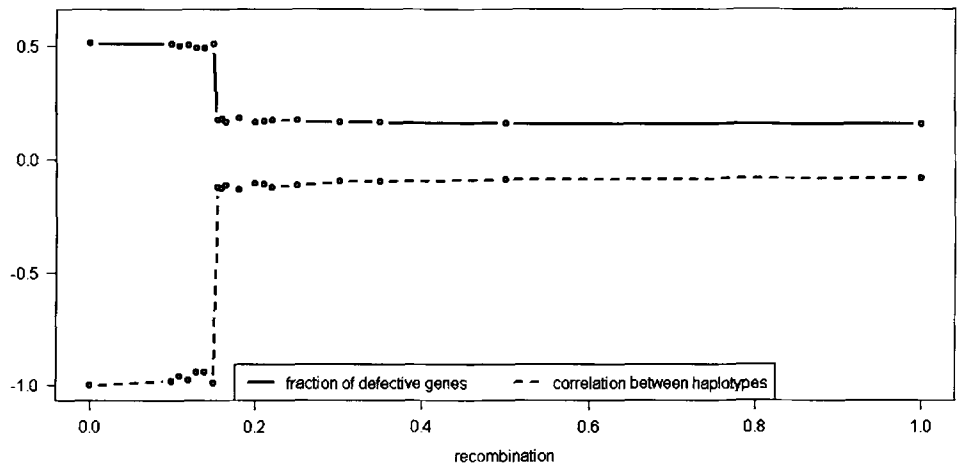

Fig. 2. Relation between recombination rate (x-axis) and fraction of defective genes in the haplotypes (higher diagram) and correlation in defects' distribution between two haplotypes in single genomes (lower diagram).

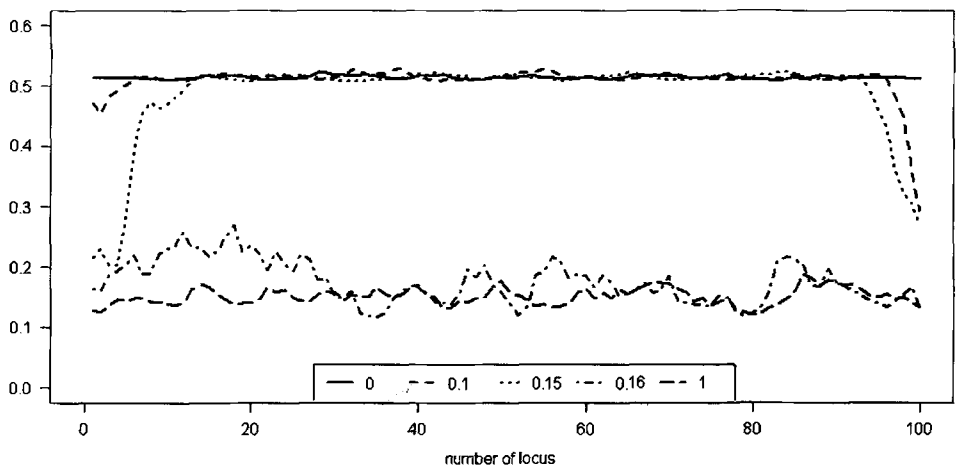

Fig. 3. Distribution of defective loci along the chromosome after simulations with different recombination rate after 200,000 MCs.

distribution of defective genes along the haplotypes. Simulations without recombination gave the evenly distributed defective genes in the entire chromosome. The frequency reaches 0.5 (in fact it is slightly higher, because the defects are counted before killing the individuals with homozygous defective loci). For intermediate recombination rate (of the order of 0.1) the fraction of defective genes at both ends of chromosomes are lower. In Fig. 4 it is shown that the higher density of defects spread from the middle of chromosomes to their ends with time.

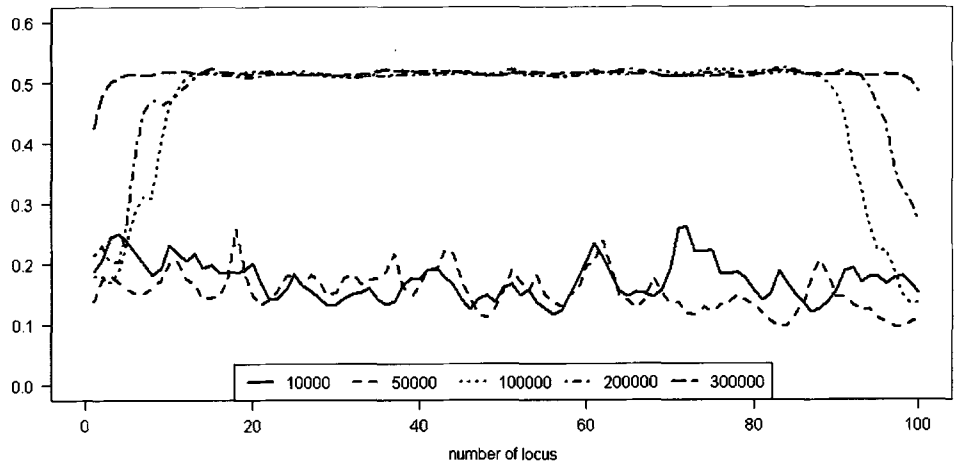

Fig. 4. Distribution of defective loci along the chromosome after different time of simulations with recombination rate 0.15 per gamete production.

There are three important conclusions from these results:

1. In the equilibrium with low recombination rate, only two complementary haplotypes exist in the genetic pool of population. All other haplotypes generated by rare recombination events or mutations are eliminated by selection because there is a very high probability of forming the homozygous defective loci (lethal) in the diploid offspring.
2. The distribution of defective genes in haplotypes is random, in such a sense that the haplotypes from independent simulations are not complementary and in fact, two independent simulations should produce two different "species" with 0 probability of the hybrids survival (the probability equals 0.75^L, where L is the number of bits in the bitstring). This phenomenon was already observed in the Penna model simulation of ageing (Laszkiewicz et al., 2003). These authors have noticed that very low recombination rate between haplotypes leads to the speciation with very low hybrid survival.
3. If we imagine a region of chromosome where the recombination is forbidden, then in the flanking regions, even with higher recombination rate, some linkage and conservation of structure should be observed. That means that "recombination deserts", where no recombination happens, should affect the genetic structure of the neighboring fragments of chromosomes. That is why we call these region—"driver".

3.3. The role of driver in the generation of linkage disequilibrium

In the further simulations, we have stuck the driver to one end of chromosome—to locus number 1. The recombination inside driver has been forbidden, while the

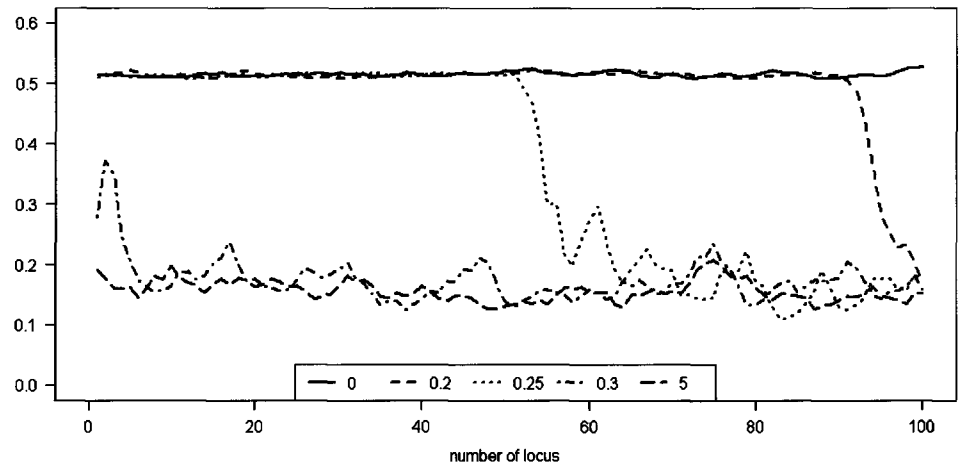

Fig. 5. Distribution of defective loci along the chromosome after the simulations with different recombination rate if "driver" is spliced to the first locus of chromosome.

recombination in the rest of chromosome has been performed like in the standard model. The results are shown in Fig. 5. In this series of simulations, loci linked to the driver have a tendency to be negatively correlated, even under higher recombination rate than if the chromosome was free of a driver. Two effects have been observed:

The first one—the stabilization of the chromosome structure (negative correlation between two haplotypes in the genome) increases with the decrease of recombination rate. Almost half of the chromosomes are stabilized after 300 000 MCs under a recombination rate 0.25 and shorter fragments of chromosome under higher recombination rate. Without driver, the stabilization of chromosome was observed for lower recombination rate—about 0.15 and had a character of sharp transition (cf. Figs. 3 and 5).

The second observation is that the length of stabilized fragments is growing with time of evolution (cf. Figs. 4 and 6). The effect has a tendency for spreading with time.

In the panmictic populations, like in our model, each female can choose randomly the male independently for each reproduction. Thus, the inbreeding coefficient which is a measure of how close, on average, two organisms in the population are genetically related depends on the size of the population. The inbreeding coefficient describes also the probability that an individual received both alleles in one locus from one ancestor.

In the fully Mendelian processes, all alleles are supposed to be inherited independently, but it is not true for real life because some physically linked alleles are co-transferred. In extreme conditions, if there are no recombinations between homologous chromosomes, the whole chromosomes (haplotypes) can originate from one ancestor. In such a case it is important if both homologous chromosomes in a given genome are identical or it is a pair of chromosomes inherited from one

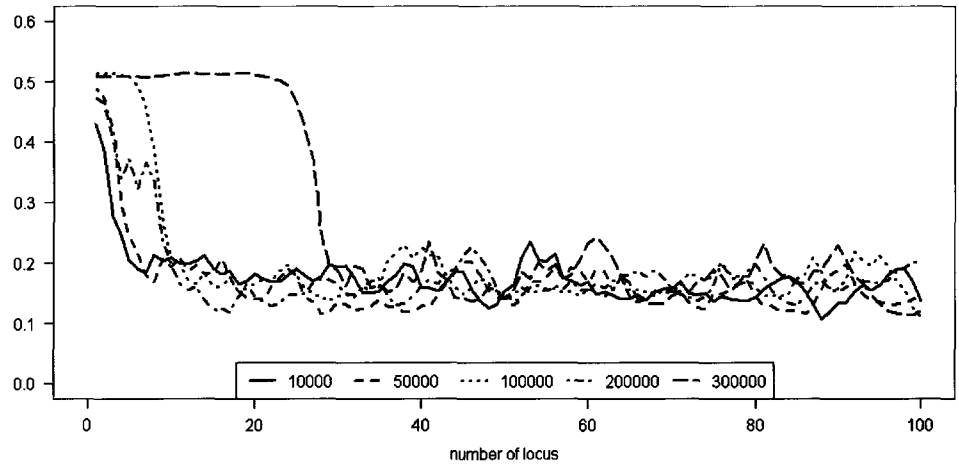

Fig. 6. Distribution of defective loci after different time of simulation with recombination rate 0.28. Driver is spliced to the first locus.

ancestor. If they are identical—the probability of homozygosity of some loci in it is higher. Our observations show that driver can be used as indicator for identity or complementarity of chromosomes (haplotypes). In nature, the Major Histocompatibility Complex can play this role, at least at two levels (Penn and Potts, 1999). At the first level, it helps to avoid incestuous sex contacts eliminating (or decreasing) the probability of reproduction involving a pair of too closely related partners (works for decreasing the inbreeding coefficient) (Garver-Apgar et al., 2006), and at the second level, it leads to the spontaneous abortion if MHC antigens of fetus resemble those of mother (Hedrick and Black, 1997). Gamete (or haplotype) preselection has also been suggested in the computer model analyzing the evolution costs of the inside intron recombination (Cebrat et al., 2006).

This model can also be useful for simulation and analysis of the genetic evolution of hybridogenesis (Milinski, 1994). In the hybridogenetic populations, individuals produce gametes containing haplotypes of only one parent; the whole genetic information of the other parent is discarded before the gametes' production. This strategy of reproduction could be considered as a prototype of haplotype recognition.

4. Conclusions

The intragenomic recombination rate or uneven distribution of recombination spots play a very important role in shaping the structure of chromosomes and a possibility of sexual reproduction between random partners. The results of simulations suggest that forming the complementing structures of chromosomes can increase the robustness of small populations; such populations can survive under higher mutational pressure and with higher genetic load. On the other hand, it can lead to the fast speciation effect because "drivers", evolving independently, after

some time lost the complementarity and an outbreeding depression instead of heterosis is observed (heterosis means hybrid vigor, when the fitness of hybrid offspring is higher than of both parents). This outbreeding depression can be considered as the first step toward the speciation. It seems reasonable to assume that some barriers emerge during the evolution which enhance the probability of forming the offspring with higher survival probability; therefore paying some costs in biodiversity of the population. If the barriers are formed at the level of gamete preselection, the procedures of fertilization in vitro and embryo transfer (FIVET) may break them. If it is true, FIVET results should carefully be statistically analyzed and, eventually, the FIVET procedure, itself, should be reconsidered.

Acknowledgments

Authors thank D. Stauffer, S. Moss de Oliveira and P.M.C. de Oliveira for discussions. This work was supported by the Poland's State Committee for Scientific Research, grant # 105/E-344/SPB and Polish Foundation for Science. It was done in the frame of European programs: COST Action P10 and NEST—GIACS.

References

Ayala, F.J., Kiger, J.A., 1980. Modern Genetics. The Benjamin/Cumings Pub. Comp. Inc., California.

Azbel, M., 1999. Phenomenological theory of mortality evolution: its singularities, universality, and superuniversality. Proc. Natl. Acad. Sci. USA 96, 3303–3307.

Cebrat, S., Pękalski, A., Scharf, F., 2006. Monte Carlo simulations of the inside intron recombination. Int. J. Mod. Phys. C 17, 305–315.

Coe, J.B., Mao, Y., 2005. Gompertz mortality law and scaling behavior of the Penna model. Phys. Rev. E 72, 051925.

Dudkiewicz, M., Mackiewicz, P., Nowicka, A., Kowalczuk, M., Mackiewicz, D., Polak, N., Smolarczyk, K., Banaszak, J., Dudek, M.R., Cebrat, S., 2005. Correspondence between mutation and selection pressure and the genetic code degeneracy in the gene evolution. Future Generat. Comput. Systems 21 (7), 1033–1139.

Garver-Apgar, C.E., Gangestad, S.W., Thorhill, R., Miller, R.D., Olp, J.J., 2006. Major histocompatibility complex alleles, sexual responsivity, and unfaithfulness in romantic couples. Psych. Sci. 17, 830–835.

Gorlov, I.P., Gorlova, O.Y., 2001. Cost-benefit analysis of Recombination and its application for understanding of chiasma interference. J. Theor. Biol. 12, 12–45.

Hedrick, P.W., Black, F.L., 1997. Random mating selection within families against homozygotes for HLA in South Ameridians. Hereditas 127, 51–58.

Kliman, R.M.K., Rey, J., 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. 10, 1239–1258.

Laszkiewicz, A., Szymczak, Sz., Cebrat, S., 2003. Speciation effect in the Penna aging model. Int. J. Mod. Phys. C 14, 765–774.

Milinski, M., 1994. Hybridogenetic frogs on an evolutionary dead end raw. Trends Ecol. Evol. 9, 62.

de Oliveira, P.M.C., 2006. Chromosome length scaling in haploid asexual reproduction. J. Phys.: Condens. Matter 18, 1–9.

Penn, D.J., Potts, W.K., 1999. The evolution of mating preferences and major histocompatibility complex genes. Am. Naturalist 153, 145–164.

Stauffer, D., Cebrat, S., 2006. Extinction in genetic bit-string model with sexual recombination. Adv. Compl. Syst. 9, 147–156.

Yu, A., et al., 2001. Nature Comparison of human genetic and sequence-based physical maps. 409, 951–953.