

Common Ancestor and Genetic Diversity in Penna Model

Qiuhui Pan¹, Zhirui Wang¹, Zhen Wang¹ and Mingfeng He^{1,2,*}

¹ School of Innovation Experiment, and School of Mathematical Science, Dalian University of Technology, Dalian 116024, China.

² College of Advanced Science and Technology, and School of Mathematical Science, Dalian University of Technology, Dalian 116024, China.

Received 16 April 2009; Accepted (in revised version) 29 May 2009

Communicated by Dietrich Stauffer

Available online 12 August 2009

Abstract. The characteristic of individual is described by the Penna model. Based on information entropy and the Penna model we define the entropy in the Penna model to discuss common ancestors and genetic diversity for two reproduction modes with and without mutation. About the problem of common ancestor, we find that all living individuals at any time step have a set of common ancestors which belongs to different times. They have the most recent common ancestor and earlier common ancestors. The analysis of genetic diversity shows that the complexity of ecosystems is caused by mutations and there is no contribution of sexual reproduction to conserving genetic diversity at long time scales. Moreover, in stable environment genetic diversity in asexual reproduction mode is higher than that in the sexual case.

PACS: 65.40.Gr, 87.15.Aa, 05.10.Ln

Key words: Penna model, entropy, common ancestor, genetic diversity, mutation.

1 Introduction

The research of common ancestor has always been the focus of the academia. In the Archaeology field, fossils are an important tool to study common ancestor for scientists [1]. With the development of gene technology, a new method to reveal the secrets of biological evolution is DNA sequence analysis [2–4]. Some other theoretical studies about common ancestor are based on mathematics and simulation [5–11]. For example, Rohde et al. suggest that all present-day human have one most recent common ancestor and earlier exactly the same set of common ancestors [5].

*Corresponding author. *Email addresses:* qhpan@dlut.edu.cn (Q. Pan), wzrwork@gmail.com (Z. Wang), kinglilin@gmail.com (Z. Wang), mfhe@dlut.edu.cn (M. He)

Recently, there are many researches about biodiversity including species diversity [12–14] and genetic diversity [14–16]. Tilman et al. discuss the diversity and stability in plant communities [12]. They use data from a long-term biodiversity experiment with plant communities to show that diverse systems can be both stable and unstable. Emerson and Kolm show that species diversity can drive speciation [13]. Global efforts to conserve species have been strongly influenced by the heterogeneous distribution of species diversity across the Earth. This is manifest in conservation efforts focused on diversity hotspots [16–18]. The conservation of genetic diversity within an individual species [19,20] is an important factor in its survival in the face of environmental changes and disease. Therefore, genetic diversity within species is an important subject in the research of biodiversity. Rauch and Bar-Yam show that diversity within species is distributed unevenly by simulation [15], which implies that diversity loss owing to severe extinction events is high, and focusing conservation efforts on highly distinctive groups can save much of the diversity. In this paper, the biodiversity under consideration is also about the genetic diversity within species.

Penna presents a simple model for biological aging based on bit strings [21]. The model works under the effect of the Verhulst factor, mutations, death by genetic diseases or age and a minimum reproduction age. The sexual Penna model introduced by Stauffer et al. [22,23] corresponds to a reproductive regime of diploid organisms, the population being divided into males and females. Makowiec et al. discuss the “Eve effect” in the Penna model [24], which is also studied in a few other papers [25–27]. Moreover, information entropy proposed by Shannon is used to discuss the uncertainty of events and the amount of information [28].

Based on the Penna model and information entropy we define the entropy in the Penna model to discuss common ancestor so-called “Eve effect” and genetic diversity for asexual and sexual reproduction. From the results we find that all present-day individuals within a species have exactly the same set of ancestors which belongs to different times and mutation is the basic condition of conserving genetic diversity.

The paper is organized in the following way. In Section 2, we give the model of entropy in the Penna model to discuss common ancestor and genetic diversity for asexual and sexual reproduction. In Section 3, we draw some conclusions from this study.

2 The entropy in Penna model

2.1 Penna model [21–23]

2.1.1 Asexual model

Each individual of the population is described by a bit string of zeros and ones, where each bit position corresponds to a year. The size of this bit string is A_{\max} that means each individual will die when its age becomes A_{\max} . Genetic diseases are represented by a bit “1” in the strings. Each “year” one more bit position is looked at. There is a

limit number T of diseases each individual can accumulate: if at some age an individual has already acquired T diseases, it dies at that time step. An individual dies not only when the number of accumulated diseases reaches the threshold T and the age becomes A_{\max} , but also as a result of competition for food and space against other individuals, represented by the logistic Verhulst factor:

$$V = \frac{N(t)}{N_{\max}},$$

where N_{\max} is the maximum population size the environment can support and $N(t)$ is the current population size. At each time step and for each individual a random number between zero and one is generated and compared with V : if it is smaller than V , the individual dies independently of its age or genome.

At every time step, each individual with age equal to or greater than the minimum reproductive age R generates B offspring. This reproducing process is repeated every year until death. The offspring genome is inherited from the father except for M additional mutations placed at randomly chosen positions. The age for a newborn is set to one.

All these processes of testing the survival of each individual and the process of reproduction, both applied over the whole population, represents a time step, i.e., one year in the simulation.

2.1.2 Sexual model

The population is divided into males and females. Each individual is represented by two bit strings of size A_{\max} read in parallel. One string is from its father, the other come from its mother. For an individual, six random dominant positions will be selected. Genetic diseases are represented by "1" bits in the bit string. If an individual has two bits equal to 1 in the same position (homozygote), it will start to suffer the effects of an inherited disease from that year on until its death. For the dominant positions, i.e., positions chosen at the beginning of the simulation, the presence of a bit 1 in one of the strings (heterozygote) represents a disease for the individual. Otherwise the sexual individual dies according to the same rule as the asexual one.

At every time step, each female with age equal to or greater than the minimum reproductive age R randomly chooses a male with age also equal to or greater than R to mate, and generates B offspring. This mating process is repeated every year until death. The offspring genome is constructed in the following way: the father genome is cut in a random position, generating four bit string pieces. Two complementary pieces, each one coming from one of the original strings, are joined to form the offspring string which contains the genetic charge to be inherited from the father. After this, M random mutations are included. The same procedure is repeated with the mother genome, to produce the other string of the baby. The sex of the newborn is randomly chosen.

All these processes of testing the survival of each individual and the process of reproduction, both applied over the whole population, represent a time step, i.e., one year in the simulation.

Table 1: Values used in the simulations.

Quantity	Value
Maximum population size	$N_{\max} = 10^5$
Initial population size	$N_0 = 1000$
Size of the bit strings	$A_{\max} = 32$
Maximum number of deleterious mutations	$T = 4$
Minimum reproduction age	$R = 8$
Mutation rate of bit string	$M = 2$
Birth rate	$B = 1$

2.2 The entropy of kinship

2.2.1 The model

At the beginning, each individual has a different family name expressed as i . The family name i of offspring will be inherited from its father whether sexual or asexual reproduction. In other words, different individuals from the same ancestor have the same family name i . There is a set $Q_1 = \{1, 2, \dots, N_0\}$ concluding all the family names. That means $i \in Q_1$, where N_0 is the initial population size.

We use $N_1(t, i)$ to express the population size of individual with family name i at time step t . $N_1(t, i)$ and $N(t)$ will be obtained by simulation in Penna model.

We define the entropy of kinship in the following:

$$S_1(t) = - \sum_{i \in Q_1, N_1(t, i) \neq 0} \left(\frac{N_1(t, i)}{N(t)} \cdot \ln \frac{N_1(t, i)}{N(t)} \right),$$

where $Q_1 = \{1, 2, \dots, N_0\}$.

2.2.2 The results

The values used for all the simulations in this paper are shown in Table 1.

In the following we will discuss the population size and the entropy of kinship for different reproduction modes with or without mutation and the results are shown in Figs. 1 and 2.

Fig. 1 shows that after some time steps stability is reached and the population self-organizes. And there is a small difference among different reproduction modes.

The entropy of kinship shows the information of family name. From Fig. 2 we can see that the entropy of kinship decreases sharply and trends to zero. The decrease of entropy means the individuals with some family name die out. At last the entropy goes to zero. It shows that all living individuals have the same family name. In a word, all present-day individuals have the same ancestor.

However, in the process of approaching zero the entropy is not monotonically decreasing. The value of entropy has the characteristic of fluctuation. At the beginning, the entropy of kinship decreases sharply as a result of decreasing family name. And then

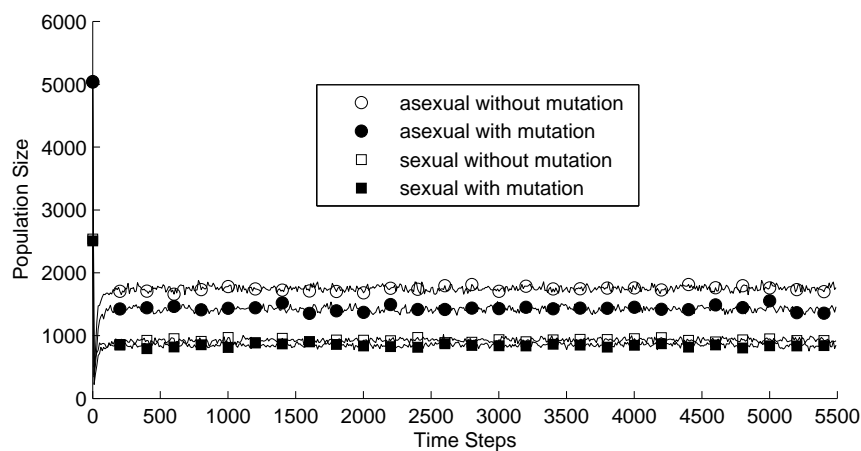


Figure 1: The population size as the function of time step for different reproduction modes with or without mutation.

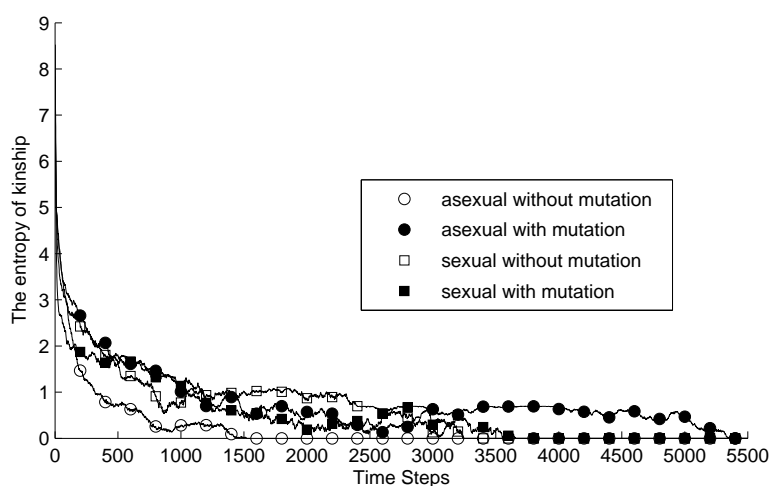


Figure 2: The entropy of kinship as the function of time step for different reproduction modes with or without mutation.

the entropy fluctuates in a small range. At this time, the amount of ancestors is small. Maybe there are only two different family names. This fluctuation is caused by the number of offspring from these two different ancestors. This fluctuation process is long term. When the number of ancestor is smaller, the competition between individuals from two ancestors becomes more and more difficult. But ultimately the common ancestor will be chosen from the winner.

From Fig. 2 we can see all living individuals of a species have the same ancestor at long time scales called common ancestor. We now ask whether or not the common

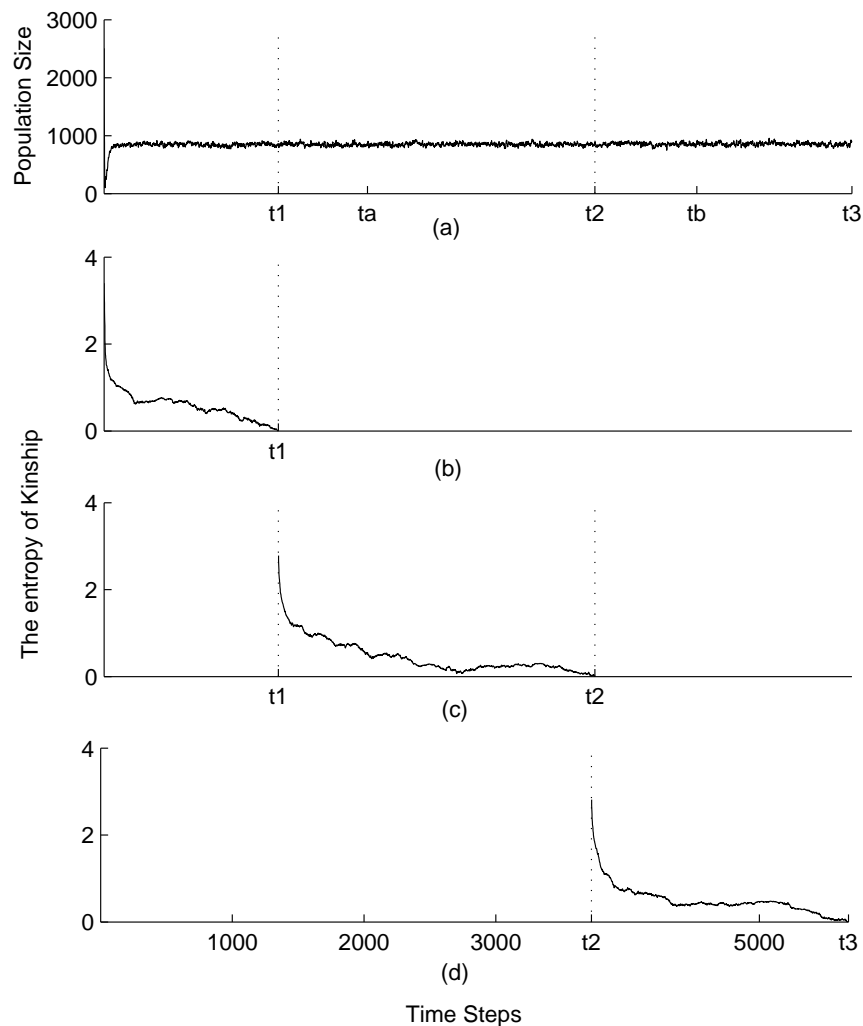


Figure 3: A set of common ancestors.

ancestor is unique? That drives us to do the following simulation. Take the case of sexual with mutation as an example and the parameter is same as Fig. 2, the results are obtained in Fig. 3. The entropy defined in this model is relative to family name. At time step $t=0$, each individual has a different family name i_0 , where $i_0 \in \{1, 2, \dots, N_0\}$. The entropy of kinship relative to i_0 is shown in Fig. 3(b). When the entropy is zero at time step t_1 , all the living individuals $N(t_1)$ are assigned a new different family name i_1 , where $i_1 \in \{1, 2, \dots, N(t_1)\}$. The same as before, the entropy relative to i_1 reaches zero when $t=t_2$ shown in Fig. 3(c). And then the same simulation as before will be performed. The entropy relative to i_2 is obtained in Fig. 3(d), where $i_2 \in \{1, 2, \dots, N(t_2)\}$.

From Fig. 3(a), we can see that the equilibrium state of the population is reached. Fig. 3(b) shows that the entropy reaches zero at t_1 time step. That means all the living individuals at time step t_1 have a common ancestor called C_0 which is an individual from the set $\{1, 2, \dots, N_0\}$. Same as Fig. 3(b), Fig. 3(c) shows that all the living individuals at t_2 time step also have common ancestor C_1 from the set $\{1, 2, \dots, N(t_1)\}$. At this time, all the living individuals have two common ancestors called C_0 and C_1 who belong to different times, where C_1 is the most recent common ancestor. Same as before, all the living individuals at t_3 time step have three common ancestors called C_0 , C_1 and C_2 , where C_2 is an individual from the set $\{1, 2, \dots, N(t_2)\}$. Thus, all the living individuals at any time step have a set of common ancestors which belongs to different times. Furthermore, they have the most recent common ancestor. But the most recent common ancestor will be different for individuals living at different time step. For example, the most recent common ancestor for individuals living at t_a time step is C_0 . And the most recent common ancestor for individuals living at t_b time step is C_1 , while C_0 is also the common ancestor for these individuals. Therefore, if we know the common ancestor of a species from a certain time by some technology (e.g. archaeology), we can not ensure that there is no common ancestor before or after this time.

2.3 The entropy of genetic diversity

2.3.1 The model

In Penna model each individual is described as a bit string of zeros and ones. The individual will be classified by bit string, viz. the individual with the same bit string will have the same characteristic. The different characteristic will be understood as genetic diversity. Each individual has a characteristic code i . There is a decimal numeral conversion from each binary bit string. This decimal numeral expresses the characteristic code i for corresponding individuals. In other words, different individuals with the same bit string have the same characteristic code. The genome of each individual is represented by two bit strings of size A_{\max} in the case of sexual reproduction mode. These two genotype strings will be represented by one phenotype string of size A_{\max} . At this time the individual will be classified by its phenotype bit string. There is a set $Q_2 = \{0, 1, 2, \dots, 2^{A_{\max}} - 1\}$ concluding all the characteristic codes. That means $i \in Q_2$.

On the process of reproduction, the characteristic code of offspring will be calculated by its bit string.

We use $N_2(t, i)$ to express the population size of individual with i characteristic code at t time step. $N_2(t, i)$ and $N(t)$ will be obtained by simulation in Penna model.

We define the entropy of genetic diversity in the following:

$$S_2(t) = - \sum_{i \in Q_2, N_2(t, i) \neq 0} \left(\frac{N_2(t, i)}{N(t)} \cdot \ln \frac{N_2(t, i)}{N(t)} \right),$$

where $Q_2 = \{0, 1, 2, \dots, 2^{A_{\max}} - 1\}$.

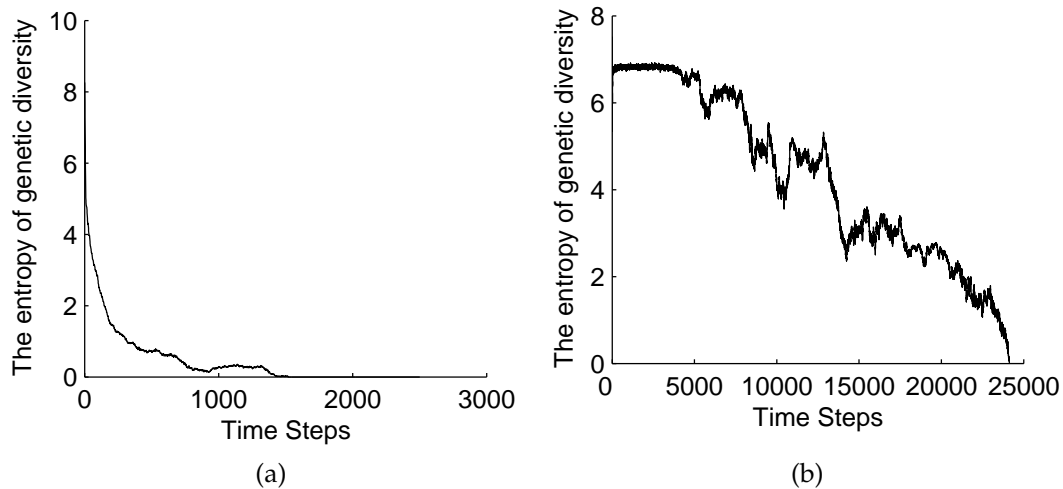


Figure 4: (a) The entropy of genetic diversity as the function of time step for sexual reproduction without mutation. (b) The entropy of genetic diversity as the function of time step for asexual reproduction without mutation.

2.3.2 The results

The values used for all the simulations in this paper are shown in Table 1. In the following we will discuss the entropy of genetic diversity for different reproduction modes with or without mutation (see Fig. 4).

Fig. 4 shows that the entropy of genetic diversity will get to zero no matter what reproduction modes. It means that the characteristic of individual will trend to be the same. Furthermore, this phenomenon is more obvious for asexual reproduction. In asexual reproduction model, the entropy of genetic diversity reaches zero quickly. In sexual reproduction model, genetic diversity is steady incipiently, but after some time steps the entropy of genetic diversity tends to reach zero as in the asexual reproduction model. From this phenomenon we can see that the genetic diversity will be conserved in the case of sexual reproduction without mutation just for a short time. At long time scales, there is no contribution of sexual reproduction to conserving genetic diversity. This is a phenomenon worthy of attention.

From Fig. 5 we can see that the entropy of genetic diversity will reach an equilibrium value for both different reproduction modes with mutation. It shows that the ecosystem has connatural complexity which is caused by mutation. In other words, the basic condition of conserving genetic diversity is mutation. This dynamic equilibrium shows that this system with diversity will be formed by different individuals, but the degree of diversity is the same for sexual and asexual reproduction. In addition, the equilibrium in asexual reproduction mode is higher than that in sexual reproduction. This result is surprising to us. We want to know if this result always appears. So the following simulation is made. We discuss the entropy as the function of time steps with different bit string size

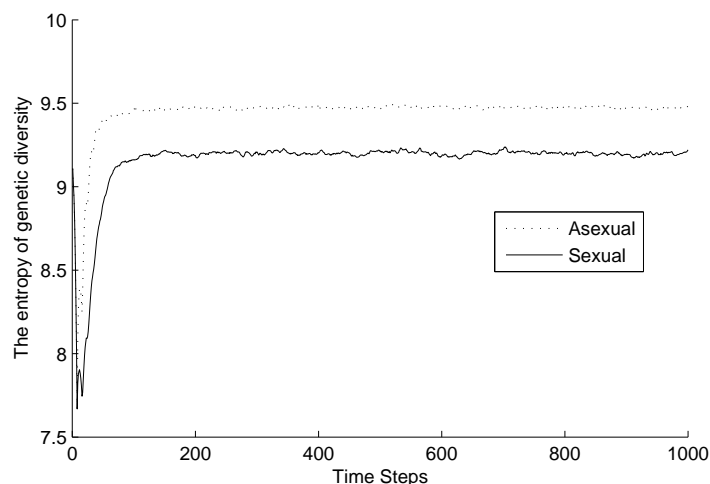


Figure 5: The entropy of genetic diversity as the function of time step for two different reproduction modes with mutation.

A_{\max} . The parameter of A_{\max} is taken as $A_{\max} = 8, 16, 24$ and other parameters (T , R and M) are taken in proportion (see from Fig. 6).

From Fig. 6 we can see that the result that the equilibrium in asexual reproduction mode is higher than that in sexual reproduction always appears, no matter what value A_{\max} is. This can be interpreted as the diversity being higher for asexual reproduction in the stable environment. It is pointed out that Sa Martins and de Oliveira found that sexual reproduction survives a catastrophe better than asexual reproduction [29] and He et al. stated so [30]. The advantage of sexual reproduction was thought to be a higher genetic diversity in the sexual case. Both papers find that the individual of sexual reproduction has better acclimatization ability in harsh or changing environment. This is not inconsistent with the results shown in Fig. 6. We can see that the two reproduction modes have merits and demerits respectively. In nature, some species have two reproduction modes, such as hydra, spirogyra and so on [31]. The asexual reproduction mode is taken in stable environment but sexual reproduction is taken in harsh environment. Furthermore, other simulations are made with different N_{\max} . The same result as Fig. 6 is found, i.e., the equilibrium in asexual reproduction mode is higher than that in sexual reproduction.

3 Conclusions

The characteristic of individual is described by the Penna model. Based on the concept of information entropy we define the entropy in the Penna model to discuss the problem of common ancestor and genetic diversity for asexual and sexual reproduction with or without mutation. We can make the following conclusions:

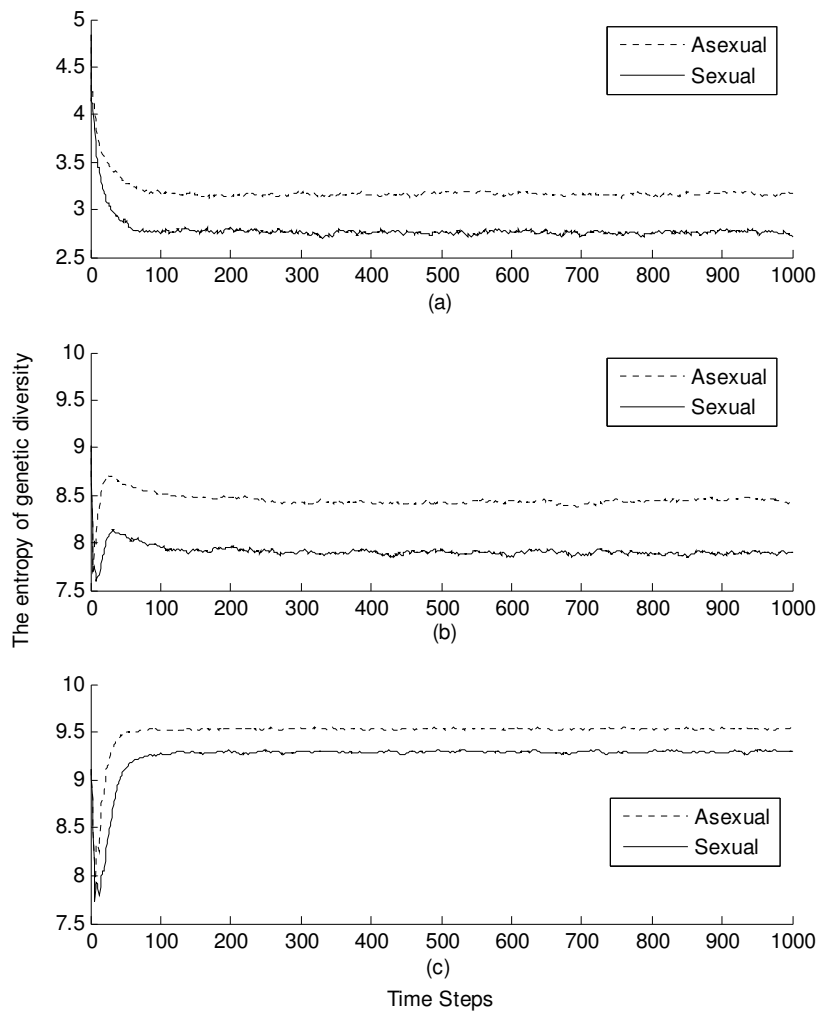


Figure 6: The entropy of genetics diversity as the function of time steps with different bit string size A_{\max} , where (a) $A_{\max} = 8$, (b) $A_{\max} = 16$ and (c) $A_{\max} = 24$.

(a) For a species, we find that all living individuals at any time step have a set of common ancestors which belongs to different times. In other words, they have the most recent common ancestor and earlier common ancestors in different years. It is possible that a lucky man from us may become the common ancestor of all the humans in the future.

(b) The genetic diversity will be conserved in the case of sexual reproduction without mutation just for a short time. The genetic diversity will be lost at last for both two reproduction modes without mutation. In large time scale, there is no contribution of sexual reproduction to conserving genetic diversity. The genetic diversity will be conserved for both two reproduction modes with mutation all the time. Therefore, there is no contri-

bution of sexual reproduction to conserving genetic diversity at long time scales and the complexity of ecosystem is caused by mutation. Moreover, in stable environment genetic diversity in asexual reproduction mode is higher than in the sexual case. It is possible that the individual in sexual reproduction has better acclimatization ability in harsh or changing environment.

References

- [1] S. Tavaré, C.R. Marshall, O. Will, C. Soligo and R.D. Martin, *Nature* 416 (2002) 726-729.
- [2] R. Thomson, J.K. Pritchard, P. Shen, P. J.Oefner, and M.W. Feldman, *Proc. Natl. Acad. Sci. USA* 97 (2000) 7360-7365.
- [3] H. Kaessmann, F. Heissig, A. von Haeseler and S. Pääbo, *Nature Genetics* 22 (1999) 78-81.
- [4] A.E. Barry, A. Leliwa, M. Choi, K.M. Nielsen, D.L. Hartl and K.P. Day, *Molecular and Biochemical Parasitology* 130 (2003) 143-147.
- [5] D.L.T. Rohde, S. Olson and J.T. Chang, *Nature* 431 (2004) 562-566.
- [6] G. Coop and R.C. Griffiths, *Theoretical Population Biology* 66 (2004) 219-232.
- [7] O. Zhaxybayeva and J.P. Gogarten, *Trends in Genetics* 20 (2004) 182-187.
- [8] J. Wakeley and N. Aliacar, *Genetics* 159 (2001) 893-905.
- [9] H.C. Harpending, M.A. Batzer, M. Gurven, L.B. Jorde, A.R. Rogers, and S.T. Sherry, *Proc. Natl. Acad. Sci. USA* 95 (1998) 1961-1967.
- [10] P.F. Slade, *Theoretical Population Biology* 57 (2000) 35-49.
- [11] J.T. Chang, *Advances in Applied Probability* 31 (1999) 1002-1026, 1027-1038.
- [12] D. Tilman, P. B. Reich and J. M. H. Knops, *Nature* 441 (2006) 629-632.
- [13] B.C. Emerson and N. Kolm, *Nature* 434 (2005) 1015-1017.
- [14] M. Vellend, *The American Naturalist* 166 (2005) 199-215.
- [15] E.M. Rauch and Y. Bar-Yam, *Nature* 431 (2004) 449-452.
- [16] J.R. Prendergast, R.M. Quinn, J.H. Lawton, B.C. Eversham and D.W. Gibbons, *Nature* 365 (1993) 335-337.
- [17] N. Myers, R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca and J. Kent, *Nature* 403 (2000) 853-858.
- [18] K.J. Gaston, *Nature* 405 (2000) 220-227.
- [19] P.R. Ehrlich and E.O. Wilson, *Science* 253 (1991) 758-761.
- [20] D.P. Faith, *Biological Conservation* 68 (1994) 69-74.
- [21] T.J.P. Penna, *J. Stat. Phys.* 78 (1995) 1629-1633.
- [22] S. Moss de Oliveira, P.M.C. de Oliveira and D. Stauffer, *Braz. J. Phys.* 26 (1996) 626-630.
- [23] D. Stauffer, P.M.C. de Oliveira, S. Moss de Oliveira, T.J.P. Penna and J.S.S. Martins, *An. Acad. Bras. Cienc.* 73 (2001) 15-32.
- [24] D. Makowiec, J. Dabkowski and M. Groth, *Physica A* 273 (1999) 169-181.
- [25] P.M.C. de Oliveira, S. Moss de Oliveira and D. Stauffer, *Theory in Bioscience* 116 (1997) 3.
- [26] P.M.C. de Oliveira, S. Moss de Oliveira and Jan P. Radomski, *Theory in Bioscience* 120 (2001) 77.
- [27] T. Preece and Y. Mao, *Phys. Rev. E* 74 (2006) 051915.
- [28] C.E. Shannon, *The Bell System Technical Journal* 27 (1948) 379-423, 623-656.
- [29] J.S. Sá Martins and S.M. de Oliveira, *Int. J. Mod. Phys. C* 9 (1998) 421.
- [30] M.He, H.Ruan, C.Yu and L.Yao. *Int. J. Mod. Phys. C* 15 (2004) 289-299.
- [31] A. G. Desnitski, *Protistology* 1 (2000) 195-198.