Cellular automata-based artificial life system of horizontal gene transfer

Ji-xin Liu

Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China
E-mail: jessonlew@hotmail.com

Published in *The Journal of Engineering*; Received on 19th June 2015; Accepted on 15th July 2015

Abstract: Mutation and natural selection is the core of Darwin's idea about evolution. Many algorithms and models are based on this idea. However, in the evolution of prokaryotes, more and more researches have indicated that horizontal gene transfer (HGT) would be much more important and universal than the authors had imagined. Owing to this mechanism, the prokaryotes not only become adaptable in nearly any environment on Earth, but also form a global genetic bank and a super communication network with all the genes of the prokaryotic world. Under this background, they present a novel cellular automata model general gene transfer to simulate and study the vertical gene transfer and HGT in the prokaryotes. At the same time, they use Schrodinger's life theory to formulate some evaluation indices and to discuss the intelligence and cognition of prokaryotes which is derived from HGT.

1 Introduction

In biology, the trait – anatomical, biochemical or behavioural – is the result of gene–environment interaction. Moreover, this interaction is affected by two factors which are mutation and natural selection. They will introduce some heritable genetic changes and may give rise to alternative traits in organisms. From a genetic viewpoint, evolution is a generation-to-generation change in the frequencies of alleles within a population that shares a common gene pool [1]. This thinking is generally accepted so influenced that many algorithms and models are inspired from it such as ant colony optimisation, evolutionary programming, genetic algorithm (GA) etc.

However, the above idea of evolution may be prevailing in eukaryotes, but not prokaryotes. The research of Mathieu and Sonea [2] has shown that there is another more important and universal mechanism – horizontal gene transfer (HGT) – in the evolution of prokaryotes. HGT is any process in which an organism incorporates genetic material from another organism without being the offspring of that organism. The opposite conception to HGT is vertical gene transfer (VGT) which is based on mutation and it usually occurs when an organism receives genetic material from its ancestor. The terms VGT and HGT can be conflated as general gene transfer (GGT) in this paper.

It has been proved that antibiotic resistance of prokaryotes is contributed by HGT [3]. Moreover, HGT has been observed between prokaryotes and eukaryotes [4]. Moreover, there also have been a few examples that some eukaryotes can receive genes from bacteria, fungi and plants by HGT [5]. Although the highly significant HGT has attracted more and more concern in biology, the achievement of research or application with HGT is shortage in non-biological field. Consequently, in this paper, we will present a novel cellular automata (CA) model to simulate VGT and HGT in prokaryotes, and study the performance of this simulation system with Schrodinger's life theory.

This paper is organised as follows. Section 2 describes the principle of HGT in prokaryotes, and the theory of life entropy and complexity. In Section 3, we will present the GGT-CA model and establish its simulation system. Section 4 details two experiments and analyses the results. In Section 5, the significance and application of HGT will be discussed. Section 6 provides conclusions.

2 Materials and methods

2.1 Horizontal gene transfer

A biological classification is introduced by Woese *et al.* [6] that divides cellular life forms into archaea, bacteria and eukaryote domains. In fact, both archaea (e.g. acidophiles and methanogens) and bacteria (e.g. *Escherichia coli* and streptococcus) are separated from prokaryotes. The key distinguish between prokaryotes and eukaryotes are that eukaryotes have nuclear envelope containing their deoxyribonucleic acid (DNA), whereas the genetic material in prokaryotes is not membrane-bound. This distinction is so important that it may be the main reason of why HGT is critical to prokaryotes as VGT to eukaryotes.

With the containing of nuclear envelope, eukaryotes' DNA is isolated and stable. Since DNA is a kind of molecule, only two major methods can change it in eukaryotes: one is some external effect at molecular level such as radiation, viruses, transposons and mutagenic chemicals; another is autologous errors that occur during meiosis or DNA replication. Both ways can cause mutation (i.e. VGT). By contrast, the genomes of prokaryotes are held within an irregular DNA/protein complex in the cytosol called the nucleoid, which lacks a nuclear envelope [7]. In simple terms, this means prokaryotes' DNA is naked with no protective membrane, and can be changed more easily by HGT. Moreover, it implies that the class and trait of a prokaryotes' cell also can be easily changed with the environmental change by HGT. That is why the prokaryotes are so adaptable. Mathieu and Sonea [2] list three main modes of HGT: (i) transformation (A and A′): a prokaryote cell can take up exogenous DNA from some adjacent lysed prokaryotes; (ii) transduction (B): a few phages will inject fragment of foreign donor's chromosome replacing homologous nucleotide sequence of the receiver's chromosome; (iii) conjugation (B′): genes can be transferred through adjacent direct cell contact. All three are so-called HGT. Fig. 1 is an explanation of HGT.

On the one hand, eukaryotes are easier and earlier to be observed than prokaryotes. Almost all species of large organisms are eukaryotes, including animals, plants and fungi, but all prokaryotes are microorganisms. On the other hand, the trait of eukaryotes is varied and colourful in comparison with prokaryotes' simple and monotonous. The obvious evidence is that most of prokaryotes have one of only three basic shapes (see Fig. 2). These are why most thinking in genetics and its extension has focused on VGT, but not HGT. Let us take some examples. In bionic mathematical modelling, Nowak [8] has summarised the mainstream about evolutionary dynamics, and VGT is the central idea of most works. In the research of CA, Sirakoulis *et al.* [9] utilised CA to study DNA sequence evolution; Gerlee and Anderson [10] presented a CA model of clonal evolution in cancer; and Mizas *et al.* [11] tried to predict the mutation of DNA sequences by using GAs and CA. Although

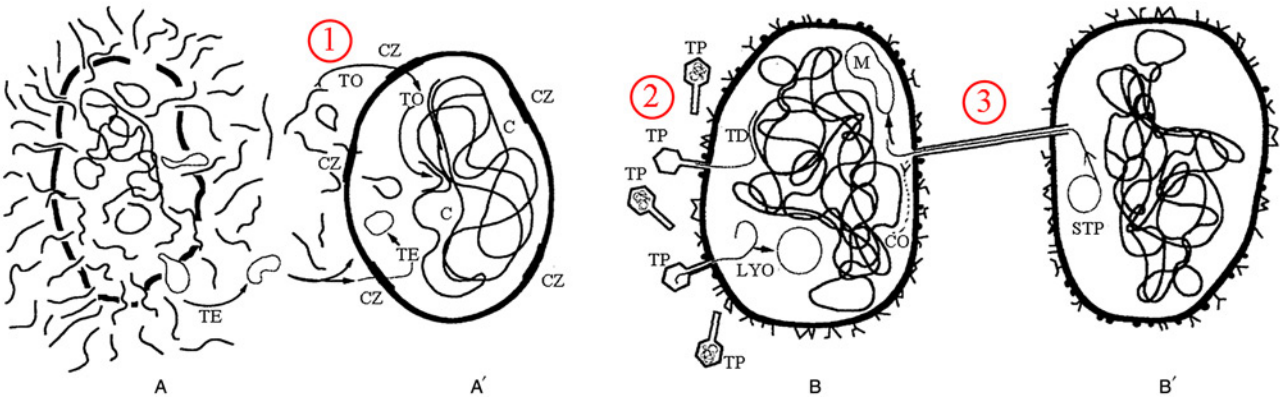

Fig. 1 *Three main mechanisms of HGT in prokaryotes*

these works also have been involved in prokaryotes, VGT would be still the focal point of their discussion.

Fortunately, there is a growing awareness that HGT is a vital important phenomenon. Rivera and Lake [12] indicate that considerable HGT has occurred between prokaryotes; Ochman *et al.* [13] explain the principle and significance of HGT; additional evidence suggests that HGT might also be an important evolutionary mechanism in protist evolution [14]; Richardson and Plamer [15] state: '*HGT has played a major role in bacterial evolution and is fairly common in certain unicellular eukaryotes. However, the prevalence and importance of HGT in the evolution of multicellular eukaryotes remain unclear*'; and HGT also may be a strong support for the theory that all living organisms on Earth are descended from a common ancestor [16, 17].

In a word, HGT is too weighty to be ignored by some other research fields such as mathematical modelling and evolutionary computation. Thus, this paper will try to integrate VGT with HGT as GGT by using CA in order to simulate the behaviour of prokaryotes.

2.2 Life entropy and complexity

Before we establish the CA model of GGT, there are two problems need to be solved. In the first place, why choose CA? CA is a kind of dynamical system modelling tool which is discrete and finite in time, space and state [18, 19]. The essential features of CA are (i) at a given time, each cell is in one of a number of states; (ii) the cells are organised according to a fixed geometry; (iii) each cell communicates only with other cells in its neighbourhood; and (iv) there is a universal clock. Each cell may change to a new state at each tick of the clock depending on its present state, and the present states of its neighbours. The rules for changing state are called the CA transition rules. At each clock tick (or 'generation') the behaviour of each cell depends only on the states of its neighbours and its own state. To conclude, the character of CA is that simple local neighbour rules can lead to complex, dynamic results of global large-scale system. This is very advantageous in solving non-linear modelling problems. Therefore a few CA models of prokaryotes have been studied. Sugiura *et al.* [20] advance a CA model for population densities of bacteria; Indekeu and Giuraniuc [21] use CA to simulate bacterial towers; Abdul Karim *et al.* [22] build a CA model of hantarvirus infection from macroscopic scale. However, there is little CA model that focus on HGT. The goal of this paper is to spread the idea of HGT.

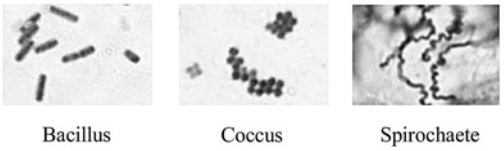

Fig. 2 *Three basic shapes of prokaryotes*

Second, how to actualise the quantitative analysis about GGT-CA model? From Schrodinger's book [23], we can gain enlightenment that entropy is a great quantitative index to GGT-CA. It is that the feature of entropy in life system is as different as non-living. The tendency of life entropy is towards far away from maximum entropy (the state of equilibrium or death). The entropy can be measured from any process of exchange. In fact, GGT is a typical process of exchange which is based on gene information; besides, GGT-CA evidently belongs to artificial life system [24] for simulating the behaviour of prokaryotes. However, a life system is more complex than any inorganic matter, and entropy may be imperfect. In a dynamical system, there are three main factors that should be considered, including 'order', 'information' and 'equilibrium'. In view of this, 'simple measure for complexity' – as a more reasonable measure index based on probability distribution – is presented from the discussion of L-Ruiz *et al.* [25], Feldman and Crutchfield [26] and Shiner *et al.* [27]. Hence, life entropy and complexity will be two critical evaluation indices about GGT-CA system.

3 GGT-CA model and simulation system

In this section, we will illustrate the operation mechanism of GGT-CA system, and the computational method of all evaluation indices. Fig. 3 is the flowchart of GGT-CA. The main steps are as follows:

Step 1: The first task is initialising CA-space and CA-state. We select two-dimensional rectangular as geometry of CA-space, and its $M \times N$ equivalent matrix can be formed as

$$\boldsymbol{S}(t) = [S(m,\ n)]_{M\times N} \tag{1}$$

The CA-state of each prokaryotic cell can be described as a vector (see (2)), including age, gene quantity and class. In this vector, age and class must be finite positive integer, and gene quantity must be finite positive real number. With the increasing of time step, the system has to judge whether the age of current cell is up to the maximum and decide the cell is alive or dead. If the state is not dead, the age can add one; or this cell must be vanishing. The range of gene quantity is a non-negative real number closed interval $[g_{\min},\ g_{\max}]$; and it will be classified as several sub-ranges which is one-to-one correspondence with the classification of prokaryotes such as $[c_{i\min},\ c_{i\max}]$, $i = 1,\ 2,\ \ldots,\ k$

$$P_{m,n}(t) = \left[a_{m,n}(t),\ g_{m,n}(t),\ c_{m,n}(t)\right]^{\mathrm{T}} \tag{2}$$

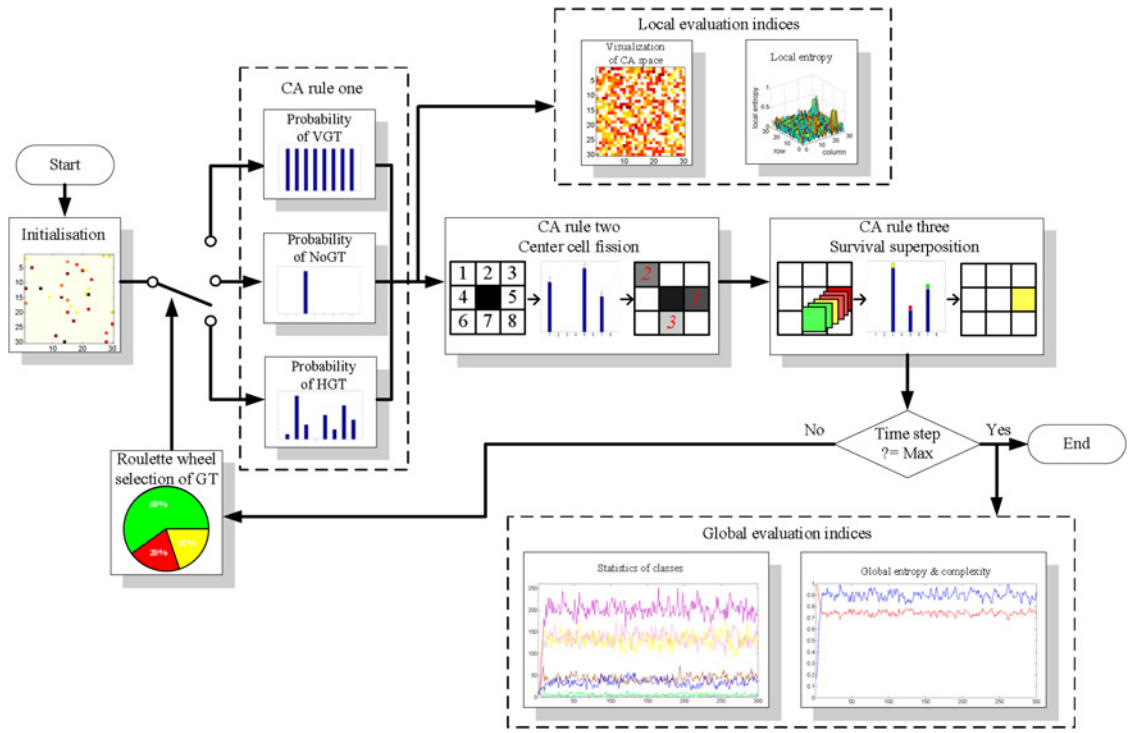

Fig. 3 *Flowchart of GGT-CA simulation system*

Step 2: Then the CA rule one – selection of GT – will be executed. There are three GT strategies which are VGT, HGT and no gener transfer (NoGT). VGT and HGT have been introduced in Section 1; and NoGT means that the current cell gene class remains unchanged. The method of selection adopts roulette wheel, and its probability distribution can be dependent on subjective experience or feedback parameter of system

$$p_{j\mathrm{GT}} = \begin{cases} 1/k, & \text{if(selection} = \text{VGT)} \\ 1 \text{ or } 0, & \text{if(selection} = \text{NoGT)} \\ \dfrac{(n_{cj} + b \times \varepsilon)}{\sum_{i=1}^{k} (n_{ci} + b \times \varepsilon)}, & \text{if(selection} = \text{HGT)} \end{cases}, \quad \sum_{j=1}^{k} p_{j\mathrm{GT}} = 1 \tag{3}$$

Equation (3) shows the probability distribution of the classes of current centre cell will be varied in each neighbour region, when GT strategy is decided:

(1) If selection is VGT, the feature of mutation will imply the uniform distribution. Since the variation trend is random, each class has the same probability.
(2) If selection is NoGT, the class of current centre cell will stay the same. Therefore the probability of only one class can be 1, and all the other will be 0.
(3) If selection is HGT, the number (n_{ci}, $i = 1, 2, \ldots, k$) of each class in the neighbour region will be made statistics. Then the number of blank unit (b) can be acquired by subtracting the number of all classes in the current region from amount of one neighbour region (it is 9 in a Moore mode). According to the introduction in Section 2.1, the proportion of n_{cj} to the number of all non-blank units in one neighbour region can be used to represent the probability of which the class j will be selected as the variation trend. The purpose of this method is to simulate the process of transformation and conjugation which is from the neighbours. On the other hand, we adopt some other means to embody the effect of transduction by using $b \times \varepsilon$ ($\varepsilon \geq 0$) to change the probability of that every class may be selected as the variation trend. In fact, transduction usually is small probability and it is depended on phages that might bring some strange chromosomes from ancient or afar.

When CA rule one is done by every cell in the whole CA-space, two local evaluation indices can be calculated and recorded. One is visualisation of CA-space which shows the space distribution of each class; another is local entropy ((4)) which means the level of disorder in each neighbour region

$$E_{m,n} = -\sum_{j=1}^{k} \left[p_{j\mathrm{GT}} \times \log_2(p_{j\mathrm{GT}}) \right] \tag{4}$$

Step 3: If the class c_z is chosen in one neighbour region, the vector value of current centre cell will be updated as follows and stored in the register

$$\boldsymbol{R}_{mn}(t+1) = [a_{mn}(t) + 1,\ (g_{mn}(t) + (c_{z\,\min} + c_{z\,\max})/2)/2,\ c_z]^{\mathrm{T}} \tag{5}$$

After that, the CA rule two – centre cell fission – is starting. This aims to simulate the binary fission of prokaryotes, a form of asexual reproduction. In short, we premise that each cell will have a copy after the process of their GT. The probability of one cell in the current neighbour region being selected is that the proportion of the probability value's reciprocal of this cell's class in (6) (W_{r}) to the sum of W_{r}. The reason is that there will be a rapid deterioration or fewer resources in a local area when the number of one class is more and more. If the offspring were to born in there, they would evolve into the local excessive class and fall into a hobble. Therefore we design that it has a negative relationship with the probability ($p_{j\mathrm{GT}}$ in (6)) of one neighbour's class in the current neighbour region that the probability of one cell of the neighbour region is selected to occupy for an offspring from the binary fission of centre cell. Moreover if a unit is blank, it will

have more opportunities to be selected. The probability distribution can be seen in (6)

$$p_{\mathrm{r}} = \frac{W_{\mathrm{r}}}{\sum_{i=1}^{9} W_i}, \quad \left(W_{\mathrm{r}} = \frac{1}{p_{j\mathrm{GT}}}, \quad j = 1,\ 2,\ \ldots,\ k \right) \tag{6}$$

Step 4: There will be a new problem when the CA rule two is done. In CA-space, some units have been selected by not only one cell, even from different classes. This means that some offsprings of prokaryotes will be superposition in the same unit. However, every unit in CA-space only can be occupied by not more than one cell. Actually, the survival pressure would be increasing with growth of cell population because of limited resources. Thereby the CA rule three – survival superposition – need to be carried out. The number (n'_{cs}) of each class in one unit with superposition will be made statistics first; then the surviving probability of each class in this unit can be taken as the proportion of the reciprocal of n'_{ci} to the sum of all n'_{ci} (i from 1 to k'). The probability distribution of this stage is as follows

$$p_{\mathrm{s}} = \frac{(1/n'_{\mathrm{cs}})}{\sum_{i=1}^{k'} 1/n'_{ci}} \tag{7}$$

If the chosen class is c_{s}, the vector value of current centre cell needs to be changed

$$\boldsymbol{P}_{mn}(t+1) = \left[\frac{\sum_{i=1}^{n'_{\mathrm{cs}}} a_i(t)}{n'_{\mathrm{cs}}}, \ \frac{\sum_{i=1}^{n'_{\mathrm{cs}}} g_i(t)}{n'_{\mathrm{cs}}}, \ c_{\mathrm{s}} \right]^{\mathrm{T}} \tag{8}$$

Step 5: The whole cell in CA-space can be updated by the data from (5) and (8). Thus, the GGT-CA completes once system running. If the number of time step is not maximum, the system will loop again; otherwise, the entire simulation system could end. Finally, we need to compute three global evaluation indices. The first one is the statistics of each class in every time step for calculating the next two indices; the second is the global entropy ((10)) which shows the variation of disorder level in the whole running of GGT-CA – an artificial life system; and the third is global complexity ((11)) which shows the variation of relationship about 'order', 'information' and 'equilibrium' during the system running

$$p_{cj}(t) = \frac{n_{cj}(t)}{\sum_{i=1}^{k} n_{ci}(t)} \tag{9}$$

$$E_{\mathrm{global}}(t) = -\sum_{i=1}^{k} \left[p_{cj}(t) \times \log_2(p_{cj}(t)) \right] \tag{10}$$

$$C_{\mathrm{global}}(t) = \left[E_{\mathrm{global}}(t) \times \left(\log_2 k - E_{\mathrm{global}}(t) \right) \right] / \left(\log_2 k \right)^2 \tag{11}$$

In the end, the description of GGT-CA system is done. In the next section, we will have two experiments with GGT-CA for analysing the system performance and significance of simulating HGT.

4 Results

4.1 Experiment one

According to the flowchart in Fig. 3, we select MATLAB as soft platform to achieve the GGT-CA system. The purpose of experiment one is to study the function of different GTs. There are three parts in this experiment: in the first part, only VGT would be adopted; second, only HGT would be adopted; third can be called hybrid GTs which means all of VGT, HGT and NoGT would be adopted and their probability distributions of roulette wheel are 5, 5 and 90%.

The parameters of GGT-CA are initialised as follows: Time step = 100 $M = N = 30$ $g_{\min} = 0$, $g_{\max} = 63$ $k = 8$ $\varepsilon = 0.1$.

Fig. 4 is the total evaluation indices in experiment one. Moreover, their analyses are as follows:

(1) V_{CS} means visualisation of CA-space. Since time step is too much, we only select six samples which are $t = 0$, 2, 6, 12, 40, 100. In VGT, the space distribution of cell keeps random; in HGT, the space distribution shows that order is prevailing; in Hybrid GTs, the trend of space distribution is from random to order.
(2) E_{L} means local entropy. Six time steps are the same as they were in V_{CS}. In VGT, the level of disorder is the highest; in HGT, it is the lowest; and in hybrid GTs, it is in the middle.
(3) S_{PC} means the statistics of each class in every time step. In VGT, the number of each class is in the approximate degree; in HGT, each class has a tendency of convergence and the difference is obvious; and in hybrid GTs, the distribution is very intricate and indefinite.
(4) E_{G} (light line) and C_{G} (darker line) mean the global entropy and complexity. For the rationality of comparison, both of them are normalised. In VGT, the entropy of system reach maximum and its complexity quickly stay at a very low level; in HGT, the entropy tends to far away from the state of equilibrium and the complexity of system is very high; in hybrid GTs, the tendency of entropy is avoiding maximum with a tempered speed and the complexity tends to high level with a complex change.

In conclusion, the system only with VGT looks like a molecular diffusion system which is satisfied with the theory of statistical physics; only with HGT have the feature of living organism; and with both VGT and HGT could be a good artificial life system which not only achieve the requirement of Schrodinger's life theory, but also simulate the behaviour of prokaryotes. The result shows that we have realised the simulation of prokaryotes as GGT-CA on the macroscopic and microscopic scales.

4.2 Experiment two

The purpose of Section 4.2 is to analyse VGT and HGT which is more important or universal. We only select the hybrid GTs. The parameters of system would be initialised as same as Section 4.1. The probability distributions of roulette wheel about VGT, HGT and NoGT also are 0.05, 0.05 and 0.9 at the beginning. In the process of system iteration, the probability of roulette wheel will be adjusted with the statistics of each GT by selecting. Only the time of VGT and HGT need to be recorded. In the end, we would make differences of the sum of VGT's selected times subtracting HGT's in each system run. The system will run independently 20 times. Fig. 5 is the result about this experiment. We can see that HGT becomes prevailing in spite of the initial probabilities of HGT and VGT are same at 5%. This also proves the viewpoint of Mathieu and Sonea [2], Jain *et al.* [28] and Ochman *et al.* [13]. They all believe that HGT is more important and universal in the evolution of prokaryotes.

Finally, the above two experiments show that GGT-CA is a valid simulation method for prokaryotes and HGT plays a key role in this system. On the basis of GGT-CA, we will discuss the important influence of HGT to the intelligence and cognition of prokaryotes in the next section.

5 Discussion

5.1 Diversity based on HGT

An example from Eigen and Winkler's book [29] could be appropriate to explain the difference between VGT and HGT. It said that, if you want to visit a friend in one city, you can have two methods. One is that you can go all round this city, and you will

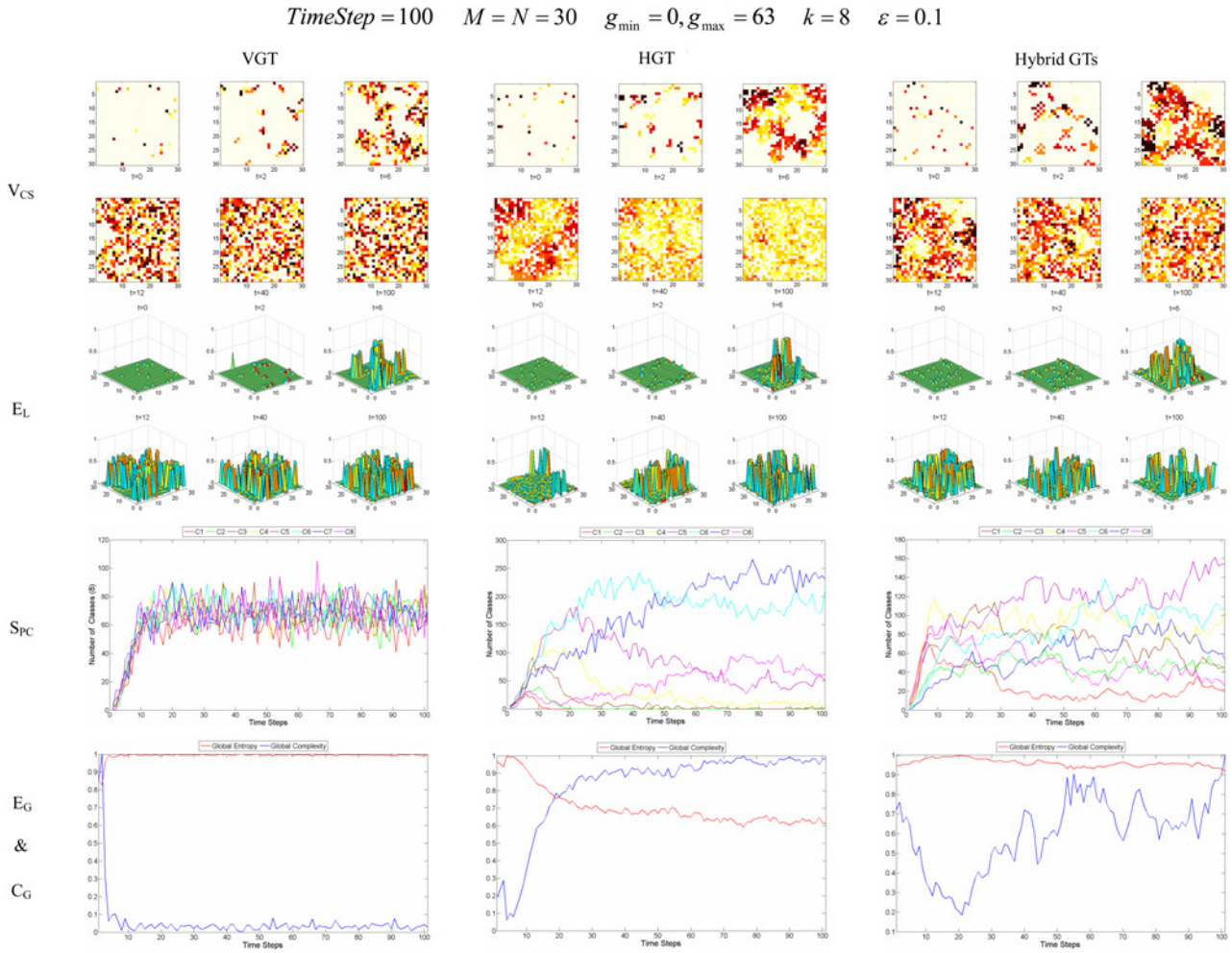

Fig. 4 *Evaluation indices of three times running in GGT-CA system*

meet your friend with a small probability in theory. Another is that you can find his address from telephone book. Even he is not at home, you may obtain some information about him from his neighbour. The latter is wise, and almost any people can select it easily. However, you would surprise that the prokaryotes may make the same decision by HGT.

The first key problem in a biological system is diversity. If a species that has a large degree of genetic diversity among its population, it will have more chances to adapt, survive and evolve. Diversity is so important, and it directly relates to the mode of variation. As we know in Section 1, two mechanisms of variation – VGT and HGT – are in prokaryotes. Simulating extinction event will be a good method to test the function of two GTs to diversity. Extinction event or mass extinction could be caused by the catastrophes such as asteroid collision, volcanic eruption and so on. It was studied that the worst extinction event can kill off over about 80 or 90% of species on Earth. When we simulate the extinction event in GGT-CA, S_{PC} could be a good index towards diversity.

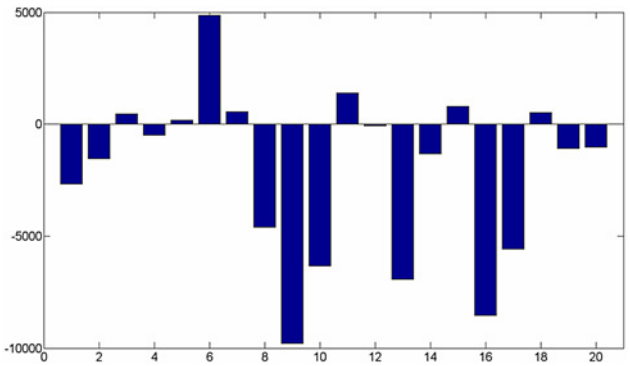

Fig. 5 *Differences of the sum of VGT's selected times subtracting HGT's in each system run*

The parameters in GGT-CA are the same as Section 4.2. When $t = 39, 79$, we will delete the cell of half dominant classes all in CA-space. This will simulate the extinction event. Before and after extinction event in Fig. 6, the diversity in VGT is uniform distribution by randomness all the time, but in HGT it can vary with changes of population. The trend of variation with VGT is random (mutation), and natural selection decides which is the fittest. This method refers to GA, and it also means mutation and natural selection are not correlated in VGT. On the contrary, the evolution with HGT can be guided from current class distribution in environment which is the result of natural selection. A feedback is built between HGT and natural selection. This means, although both VGT and HGT can keep the diversity in system, HGT contribute more efficient and smart genetic variation to cope with any environmental change.

5.2 *Intelligence and cognition of prokaryotes*

To some extent, the prokaryotes look like having intelligence and cognition in their evolution and behaviour. Therefore the second key problem is confirming that the intelligence and cognition of prokaryotes is on account of HGT or not. E_G and C_G will be two

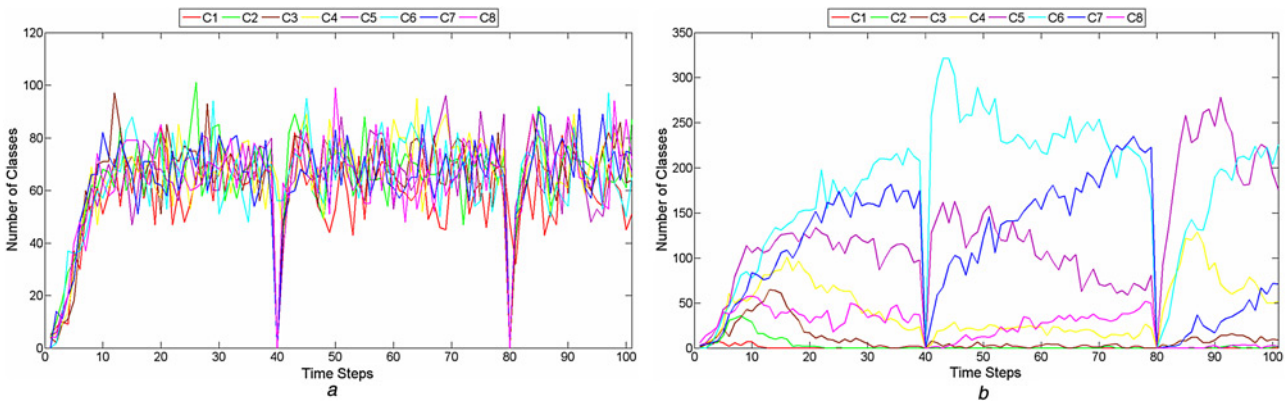

Fig. 6 *Comparison of diversity*
a VGT
b HGT

good indices in this problem. According to the extinction event experiment in Section 5.1, Fig. 7 is the result of E_G and C_G. Before and after extinction event, E_G and C_G of VGT stay at high entropy and low complexity as a non-living system; however, E_G and C_G of HGT tend to low entropy and high complexity even when the environment have been changed greatly.

As Schrodinger's theory [23], the feature of a life system is tending to go far away from maximum entropy (or the state of equilibrium). In his opinion, any process would create an increase of the entropy. However, it is generally known that information exchange will effectively improve the order of a system (avoiding the increase of the entropy). The foundation of information exchange is on expression and communication. If a system can spontaneously express information, it can be considered as having intelligence; if a system can spontaneously communicate information, it can be considered as having cognitive performance. Examples of this abound. Human can exchange information by the speech and writing; and the waggle dance is a mechanism whereby honey bees send the information to other bees in their colony about the direction and distance to patches of flowers yielding nectar and pollen, to water sources, or to new housing locations [30, 31]. In fact, gene is the carrier of genetic information, and HGT substantially is a process that the prokaryotes exchange gene each other. These are expressions and communications in prokaryotes. Therefore the result in Fig. 7 shows GGT-CA will become an intelligent life simulation system with cognitive performance when HGT is enabled. Antibiotic resistance also has been proved that is derived from HGT, and it is the embodiment of intelligence and cognition of prokaryotes to solving the extinction event.

Our opinion also can be supported by the research of Mathieu and Sonea [2]. They thought HGT would cause the amazing attributes and fateful achievements of prokaryotes: (i) a non-Darwinian evolution, based on co-operation and unity; (ii) the discovery and extensive use of entirely original mechanisms for gene exchanges involving all the genes of the prokaryotic world; (iii) a global communication network through genes, by and for all prokaryotes; (iv) a computer-like generalised system of solving problems; (v) generalised capacity for genetic engineering; (vi) a system of mixed communities favourably stabilising their environment and sustaining our entire biosphere; and (vii) a major and still active role in the origin and evolution of eukaryotes.

Our future work will be interested in the application of HGT in two fields – evolution computation and pattern recognition. In evolution computation, GA is a search heuristic that mimics the mutation and natural evolution. One drawback of GA is that it easily tends to converge towards local optima when solving some complex problems. The main reason is that the fitness function of GA is single and changeless, but a complex problem usually could be high dimensional and multimodal. Therefore we consider that HGT can be introduced into GA for designing a kind of adaptive fitness functions. On the other hand, feature extraction is a critical technique in pattern recognition. The cognitive performance of GGT-CA could be derived from the self-organising class distribution (S_{PC}) of HGT. In other words, the class distribution of HGT

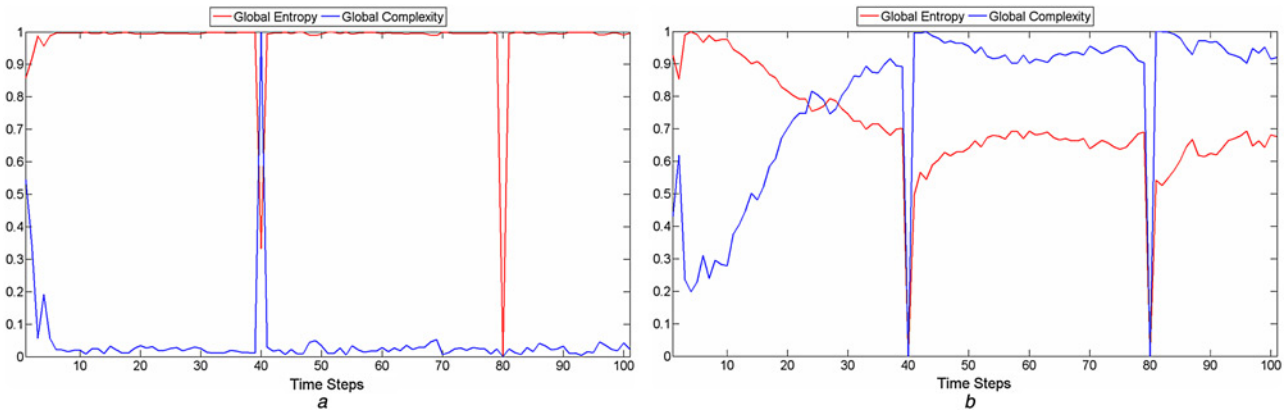

Fig. 7 *Comparison of the feature of life system*
a VGT
b HGT

will be a good representation to current environment, which might lead to a new feature extraction method.

6 Conclusions

This paper has presented a novel CA model GGT-CA to simulate the behaviour of prokaryotes as an artificial life system. In the first place, we have explained the fundamental of HGT and Schrodinger' life theory as the theoretical basis of GGT-CA. Second, we have introduced each component in GGT-CA and algorithms of some key parts. Third, we have completed the whole system running and using five evaluation indices to analyse the performance of GGT-CA by two experiments. In conclusion, we have successfully actualised the simulation of VGT and HGT, and the evaluation indices show that GGT-CA is satisfied with the requirement of artificial life system.

After system running and analysing, we have studied the diversity based on HGT and discuss its significance. On the basis of these, we have utilised GGT-CA to demonstrate the function of HGT in the intelligence and cognition of prokaryotes. Finally, we have suggested the application of HGT in evolution computation and pattern recognition.

7 Acknowledgments

This work was supported by the China National Natural Science Funds (grant no. 61401220), the Provincial Natural Science Foundation of Science and Technology Bureau of Jiangsu province (grant no. BK20140884), the University Natural Science Research Project of Jiangsu province (grant no. 14KJB510022) and the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (grant no. NY213109).

8 References

[1] Stoltzfus A.: 'Mutation and the dual causation of evolutionary change', *Evol. Dev.*, 2006, **8**, pp. 304–317
[2] Mathieu L.G., Sonea S.: 'A powerful bacterial world', *Endeavor*, **19**, (3), pp. 112–117
[3] Walsh T.R.: 'Combinatorial genetic evolution of multiresistance', *Curr. Opin. Microbiol.*, 2006, **9**, pp. 476–482
[4] Kondo N., Nikoh N., Ijichi N., Shimada M., Fukatsu T.: 'Genome fragment of Wolbachia endosymbiont transferred to X chromosome of host insect', *Proc. Natl. Acad. Sci., USA*, 2002, **99**, pp. 14280–14285
[5] Gladyshev E.A., Meselson M., Arkhipova I.R.: 'Massive horizontal gene transfer in bdelloid rotifers', *Science*, 2008, **320**, pp. 1210–1213
[6] Woese C.R., Kandler O., Wheelis M.K.: 'Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya', *Proc. Natl. Acad. Sci., USA*, 1990, **87**, pp. 4576–4579
[7] Thanbichler M., Wang S., Shapiro L.: 'The bacterial nucleoid: a highly organized and dynamic structure', *J. Cell. Biochem.*, 2005, **96**, pp. 506–521
[8] Nowak M.A.: 'Evolutionary dynamics: exploring the equations of life' (Belknap Press of Harvard University Press, Cambridge, MA, 2006)
[9] Sirakoulis G.Ch., Karafyllidis I., Mizas Ch., Mardiris V., Thanailakis A., Tsalides P.: 'A cellular automaton model for the study of DNA sequence evolution', *Comput. Biol. Med.*, 2003, **33**, pp. 439–453
[10] Gerlee P., Anderson A.R.A.: 'A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype', *J. Theor. Biol.*, 2008, **250**, pp. 705–722
[11] Mizas Ch., Sirakoulis G.Ch., Mardiris V., Karafyllidis I., Glykos N., Sandaltzopoulos R.: 'Reconstruction of DNA sequences using genetic algorithms and cellular automata: towards mutation prediction?', *BioSystems*, 2008, **92**, pp. 61–68
[12] Rivera M.C., Lake J.A.: 'The ring of life provides evidence for a genome fusion origin of eukaryotes', *Nature*, 2004, **431**, pp. 152–155
[13] Ochman H., Lawrence J.G., Groisman E.A.: 'Lateral gene transfer and the nature of bacterial innovation', *Nature*, 2000, **405**, pp. 299–304
[14] Bapteste E., Susko E., Leigh J., MacLeod D., Charlebois R.L., Doolittle W.F.: 'Do orthologous gene phylogenies really support tree-thinking?', *BMC Evol. Biol.*, 2005, **5**, pp. 33–42
[15] Richardson A.O., Plamer J.D.: 'Horizontal gene transfer in plants', *J. Exp. Bot.*, 2007, **58**, pp. 1–9
[16] Steel M., Penny D.: 'Common ancestry put to the test', *Nature*, 2010, **465**, pp. 168–169
[17] Theobald D.L.: 'A formal test of the theory of universal common ancestry', *Nature*, 2010, **465**, pp. 219–222
[18] Corne D.W., Frisco P.: 'Dynamics of HIV infection studied with cellular automata and conformon-P systems', *BioSystems*, 2008, **91**, pp. 531–544
[19] Jones D.H., McWilliam R., Purvis A.: 'Designing convergent cellular automata', *BioSystems*, 2009, **96**, pp. 80–85
[20] Sugiura K., Kawasaki Y., Kinoshita M., Murakami A., Yoshida H., Ishikawa Y.: 'A mathematical model for microcosms: formation of the colonies and coupled oscillation in population densities of bacteria', *Ecol. Model.*, 2003, **168**, pp. 173–201
[21] Indekeu J.O., Giuraniuc C.V.: 'Cellular automaton for bacterial towers', *Physica A*, 2004, **336**, pp. 14–26
[22] Abdul Karim M.F., Md Ismail A.I., Ching H.B.: 'Cellular automata modelling of hantarvirus infection', *Chaos Solitons Fractals*, 2009, **41**, pp. 2847–2853
[23] Schrodinger E.: 'What is life?' (Cambridge University Press, Cambridge, 1944)
[24] Bedau M.A., McCaskill J.S., Packard N.H., *ET AL.*: 'Open problems in artificial life', *Artif. Life*, 2000, **6**, pp. 363–376
[25] L-Ruiz R., Mancini H.L., Calbet X.: 'A statistical measure of complexity', *Phys. Lett. A*, 1995, **209**, pp. 321–326
[26] Feldman D.P., Crutchfield J.P.: 'Measures of statistical complexity: why?', *Phys. Lett. A*, 1998, **238**, pp. 244–252
[27] Shiner J.S., Davison M., Landsberg P.T.: 'Simple measure for complexity', *Phys. Rev. E*, 1999, **59**, pp. 1459–1464
[28] Jain R., Rivera M.C., Lake J.A.: 'Horizontal gene transfer among genomes: the complexity hypothesis', *Proc. Natl. Acad. Sci., USA*, 1999, **96**, pp. 3801–3806
[29] Eigen M., Winkler R.: 'Laws of the game: how the principles of nature govern chance' (Princeton University Press, Princeton, 1993)
[30] Riley J., Greggers U., Smith A., Reynolds D., Menzel R.: 'The flight paths of honeybees recruited by the waggle dance', *Nature*, 2005, **435**, pp. 205–207
[31] Seeley T.D., Visscher P.K., Passino K.M.: 'Group decision making in honey bee swarms', *Am. Sci.*, 2006, **94**, pp. 220–229