

Genetic Algorithm Inspired by Mimetic Octopus RNA Editing*

Xuecheng Zhang and Qiuxuan Wu

*the Institute of Automation
Hangzhou Dianzi University
Hangzhou, Zhejiang Province, China
wuqx@hdu.edu.cn*

Lifeng Zhang and Botao-Zhang

*the Institute of Automation
Hangzhou Dianzi University
Hangzhou, Zhejiang Province, China
wuqx@hdu.edu.cn*

Abstract - *The mimetic Octopus is known for its ability to perform complex behaviors, and its intelligence has gone beyond human recognition. Researches have shown that mimetic Octopus can sacrifice the evolution of its own genomic DNA in exchange for extensive RNA editing. Inspired by the ability of self-editing RNA to modify its own genetic information, by encoding the four bases A, C, G, U, writes RNA genetic algorithm, and calculate the different test functions. After many comparative experiments, it is concluded that this algorithm has certain effectiveness in optimization performance, which can be found quickly jump out of the local loop, and then find the global optimal value. The research of this algorithm is helpful to further explore the influence of RNA editing mechanism on the evolution of mimetic Octopus.*

Index Terms - RNA-editing RNA algorithm nitrogen base

I. INTRODUCTION

The mimetic Octopus is one of the most unique type of cephalopod, which has a high intelligent organism [1], with the largest and most complex nervous system [2], it can imitate the morphology of different creatures up to eighteen. And using imitation skills to protect themselves, to avoid harm and so on. Studies [3] have shown that, the mimetic Octopus gene is very unique, whom can edit its own RNA, and then change the characteristics and functions of the protein in the body at a much faster rate than DNA mutation. In fact, Vertebrate can also edit RNA, but rarely used. There are about 20,000 genes in human and Octopus, but only a few dozen RNA editing sites may encode functional proteins in human, however, there are at least 60,000 active RNA editing sites in Octopus. The compiled RNA can allow Octopus have a diverse nervous system to meet a variety of challenges.

RNA editing refers to the genetic information transmitted by RNA through modification and processing of post-transcriptional mature RNA molecules. For mimetic Octopus, although scientists still do not know the true purpose of RNA editing, it is undeniable that the mimetic Octopus's nervous system is more diverse after RNA editing, whom can learn more new skills, such as opening the aquarium's cover to escape, hiding themselves to avoid natural enemies and so on.

Genetic algorithm lacks complete convergence, low search efficiency and premature convergence. Inspired by the ability of RNA self-editing in mimetic Octopus, this paper

compiles a genetic algorithm for RNA. Different from the standard genetic algorithm (SGA) and other RNA algorithm [4-5], this algorithm uses a digital coding method based on RNA basic groups to overcome the limitations of genetic algorithms. For example, it uses 0, 1, 2, 3 instead of bases A, G, C, U; At the same time, in order to reflect randomness, the algorithm does not adopt the traditional roulette method, but choose the best and worst to form a new population, then begin to cross-computation; In order to reflect self-editing, different mutation probabilities are set, and the mutation probability of base A is set to be larger to complete the mutation to G. Finally, the algorithm is used to optimize several different test functions, then compared with standard genetic algorithm (SGA) in many experiments, it concluded that the algorithm is effectiveness.

II. RNA EDITING AND BIOLOGICAL SIGNIFICANCE

A. RNA editing phenomenon

Generally speaking, RNA editing is the insertion, deletion or substitution of the transcribed RNA bases in the coding region, which makes the transcribed products not faithfully reflect the DNA sequence, and produces the product of polymorphic gene phenomena. There are two main types of RNA editing phenomena: one is the insertion or deletion of bases, such as the insertion or deletion of bases U or C; the other is editing substitution, such as C to U, A to C, A to G[6-8], and so on, the most common way of editing in higher plants is U to C, while in mammals, is A to I.

At present, various forms of RNA editing have been discovered, and RNA editing can occur in several important aspects, such as translation on the pathway of gene expression, creation and deletion of splicing sites, and so on.

B. Editing mechanism of RNA mimetic Octopus

New research [9] shows that mimetic octopus often does not strictly follow the genetic instructions in DNA, but sacrifices the evolution of genomic DNA in exchange for a large number of RNA editing. Scientists have found that in the RNA coding region of octopus, especially in the cells of the nervous system, an abnormal editing ratio appears, and the purpose of editing is to cause protein changes.

* This work is partially supported by Key Projects of Science and Technology Plan of Zhejiang Province (Grant no. 2019C04018).

Caroline et al. [10] have conducted an in-depth study of the octopus genome and RNA editing mechanisms. After sequence the genome of the double-spotted octopus, they have found a wide range of messenger RNA editing in transcripts are produced in genes, involved in neural excitability, and lead to protein diversity. In order to identify polymorphic positions in the genome, they used the GATK method to predict gene patterns, and classified SNPs based on the type of nucleotide changes. They [11-12] have found that ADAR enzyme plays an important role in the editing process of RNA and leads to Mutations in base A to G, especially in the process of neural tissue and development.

C. Biological significance of RNA editing

RNA editing has a dual role in expanding the diversity of proteins and maintaining the conservation of ethnic development [13-15]. Because the edited site can accumulate more mutations than the unedited site, it also increases protein diversity.

In general, RNA editing has the following important meanings: (1) RNA editing can change genetic information, often resulting in changes in the encoded amino acid species; (2) RNA editing causes polymorphisms in gene expression products, which can be edited to express a variety of homologous proteins; (3) RNA editing can repair gene function, and a variety of gene frame-shift mutations usually occur in organisms, however, such mutations are often harmful, but after editing the RNA, it can add or remove genes to repair the functions of the gene[16].

III. MAIN IMPLEMENTATION STEPS OF RNA ALGORITHM

RNA is short for ribonucleic acid, each RNA molecule is made up of a long chain of nucleotide units, including a nitrogen base, a riboside and a phosphate group. The single-stranded RNA molecule shown in Figure 1. There are four nitrogen bases: A (adenine), G (guanine), C (cytosine), U (uracil). It is precisely the different combinations of these four bases that make up a complex function of RNA molecules.

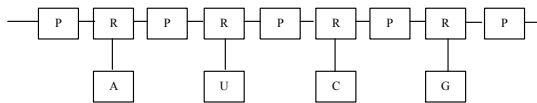


Figure 1. Part of a single-stranded RNA structure

A. Individual coding

Individual coding is primary to be solved when applying genetic algorithm, it is also an important step to design genetic algorithm. The standard genetic algorithm encoding includes binary coding, gray coding, and so on, even binary code solves the operation of decimal coding mode, but it is easy to encounter the problem of Hamming Cliff. However, this paper adopts a new coding method for individual coding, that is, the digital coding of bases. The digital coding of bases alleviates the problem of Hamming cliff to a certain extent. For RNA molecules, 0 (00), 1 (01), 2 (10), 3 (11) digital encoding can be combined with the four bases cytosine (C), uracil (U),

adenine (A) and guanine (G), which can well simulate the evolution of biology.

B. Individual selection

In the process of algorithm searching, on the one hand, we should choose the optimal individual with stronger adaptability, that is, the search can converge to the global optimal solution; on the other hand, the diversity of individuals should be fully maintained in the process of searching. In this paper, in order to maintain the diversity of population, it is decided to select the optimal top ten individuals and the worst twenty individuals to form a new population. This method guarantees the diversity of population.

C. Individual crossover

Crossover is an important method to generate new individuals. New individuals are produced by cross-recombination between two chromosomes. Common crossover strategies include a little exchange, k-point exchanges. In this paper, by analyzing the characteristics of RNA molecules, a substitution operation based on the equivalent length of k-points is adopted, in which one subsequence of a chromosome is replaced by another subsequence of the same length. It is assumed that the original individual in this paper is $K_1K_2K_3K_4K_5K_6K_7$, when the subsequence of K_3 is replaced by a sequence of the same length K_3' , the new sequence becomes $K_1K_2K_3'K_4K_5K_6K_7$, as shown in Figure 2.

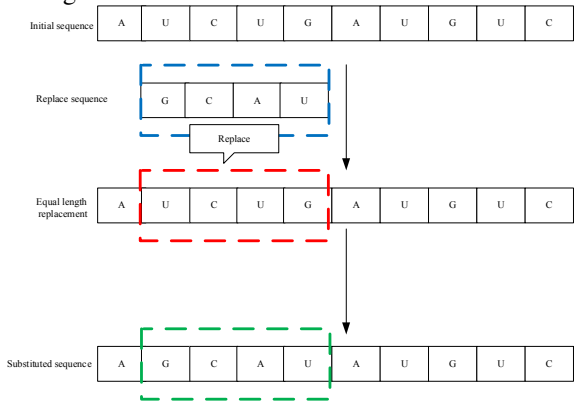


Figure 2. The part of initial sequence is replaced by the equal length

D. Individual mutation

The purpose of mutation operation is not only make the algorithm have global search ability, but also maintain the diversity of the population. When the genetic algorithm approaches the optimal solution through selection operation, the local optimization ability of the mutation operator can make them closer to the optimal solution. Generally, the mutation probability should be smaller, otherwise it may deviate from the optimal solution with a large probability, which will lead to the poor search ability. In this paper, individual mutation is the most important part of the algorithm, and it is also an important embodiment of RNA self-editing. In this paper two different mutation probabilities

are set, that is high mutation probability and low mutation probability, as shown in Figure 3.

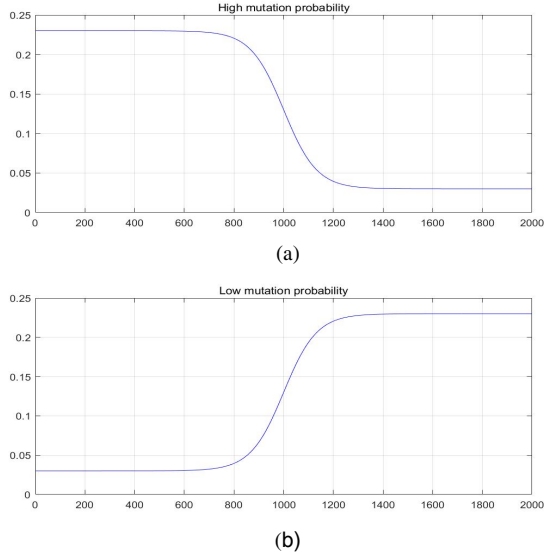


Figure 3. Different mutation probabilities: (a) high mutation probability; (b) low mutation probability

IV. ALGORITHM PERFORMANCE ANALYSIS

In order to verify the effectiveness of the proposed algorithm, several representative functions are selected for the optimization test. The function expressions are shown in equations.

$$\min f_1(x) = \sum_{i=1}^2 x_i \sin(\sqrt{|x_i|}), x_1, x_2 \in [-512, 512] \quad (1)$$

$$\min f_2(x) = \sin(x_1) + \cos(x_2) + 0.1(x_2 + x_1), \quad (2)$$

$$x_1, x_2 \in [-12, 12]$$

$$\max f_3(x) = \left(\frac{a_0}{b_0 + (x_1^2 + x_2^2)} \right)^2 + (x_1^2 + x_2^2)^2$$

$$a_0 = 3, b_0 = 0.05 \quad (3)$$

$$x_i \in [-5.12, 5.12]$$

$$\min f_4(x) = x_1 \sin(\sqrt{|x_2 + 1 - x_1|}) \cos(\sqrt{|x_2 + 1 + x_1|})$$

$$+ (x_2 + 1) \cos(\sqrt{|x_2 + 1 - x_1|}) \sin(\sqrt{|x_2 + 1 + x_1|}), \quad (4)$$

$$x_i \in [-512, 512]$$

This functions are all two-dimensional with special forms, of which $f_1(x)$ is a Schwefel function with symmetry, it is easy to fall into local optimal value; $f_2(x)$ is a multi-peak function, with multiple peaks, which is inseparable; $f_3(x)$ is a Needle in haystack function (NiH), it is uneasy to search for the optimal value; $f_4(x)$ is a Rana function, the optimal solution is located on the boundary of the search area, with multiple minimum points.

A. Experimental results and Analysis

In order to verify the effectiveness and the influence of experimental parameters, I have tried many experiments.

(1) Setting the iterations is 2000, and the crossover probability is 0.8, the average fitness and optimal fitness of functions as shown in the Figures.

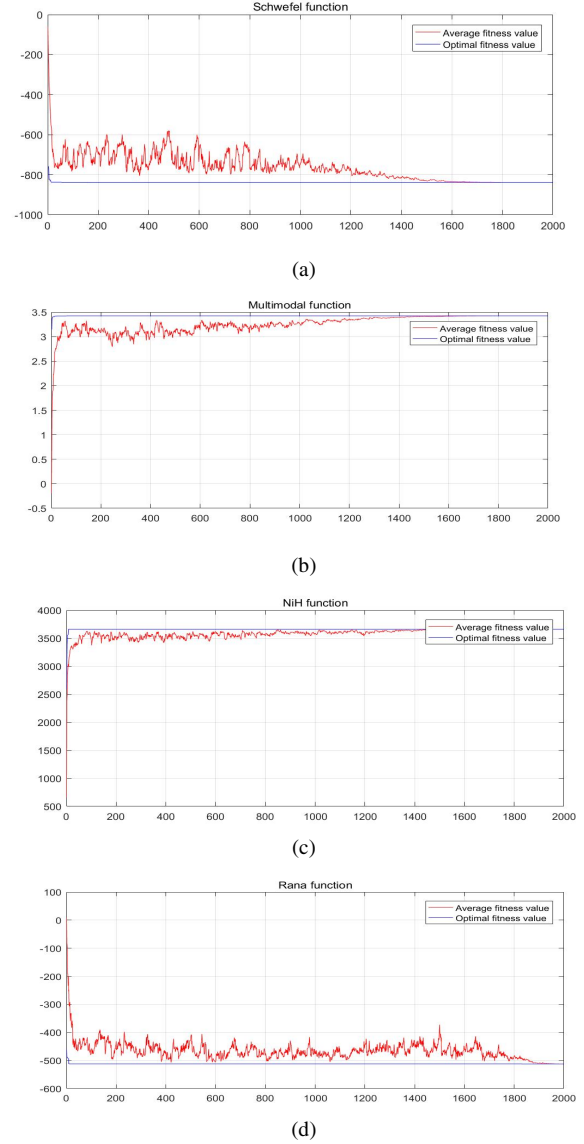
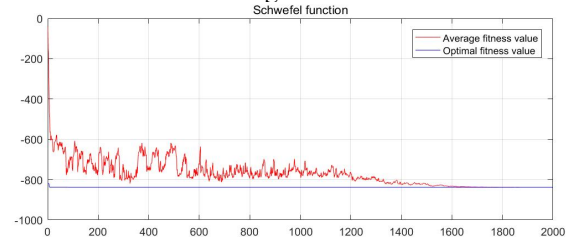


Figure 4. Setting the crossover is 0.8, iterations is 2000, the figures obtained by different functions using this RNA algorithm: (a) Schwefel function; (b) Multimodal function; (c) NiH function; (d) Rana function

(2) Setting the iterations is 2000, and the crossover probability is 0.4, the average fitness and optimal fitness of functions as shown in the Figures.



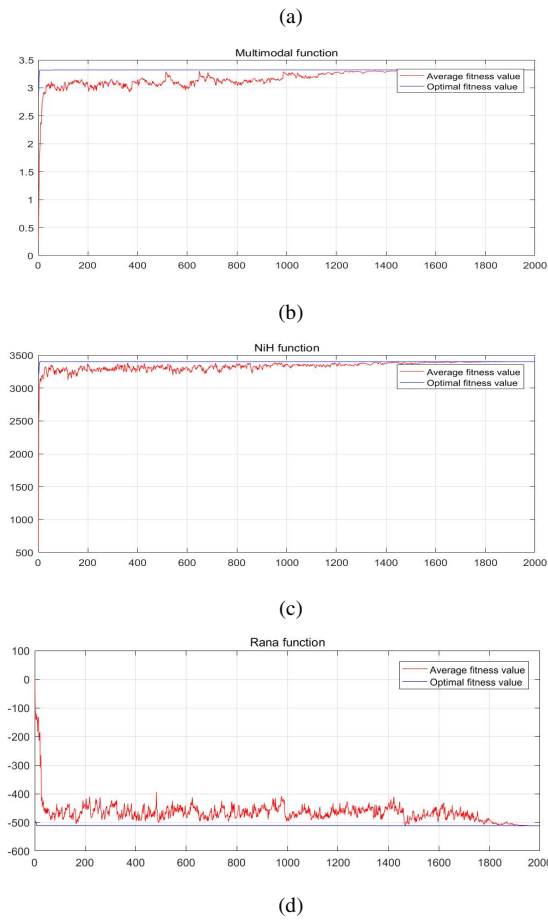


Figure 5. Setting the crossover is 0.4, iterations is 2000, other parameters are the same, different results after using this RNA algorithm: (a) Schwefel function; (b) Multimodal function; (c) NiH function; (d) Rana function

(3) In order to further verify the effectiveness of the algorithm, the proposed algorithm is compared with the standard genetic algorithm (SGA). Setting the SGA population size is also 100, the crossover probability is 0.8, the mutation probability is 0.03, and the chromosome coding length is 10, the iterations is 2000, as shown in the Figures.

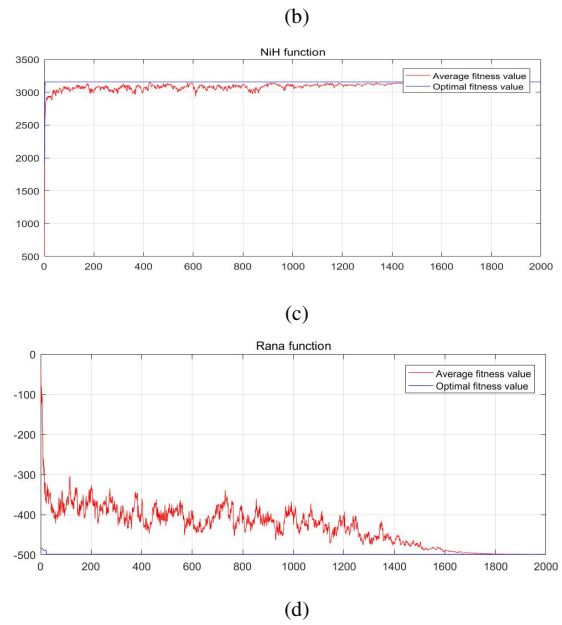
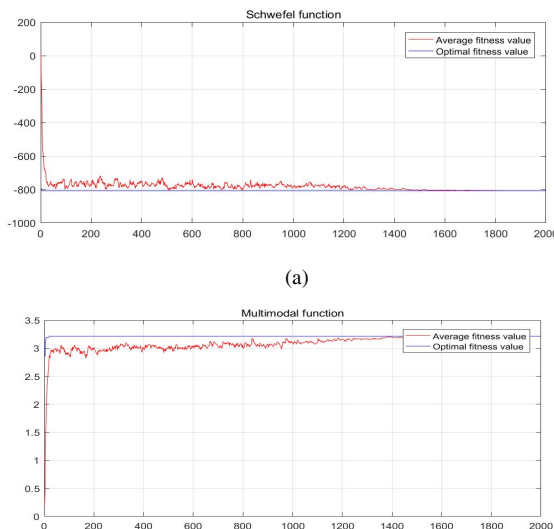


Figure 6. Setting the crossover is 0.8, iterations is 2000, other parameters are the same, different results after using this SGA algorithm: (a) Schwefel function; (b) Multimodal function; (c) NiH function; (d) Rana function

B. Analysis

Comparing the standard genetic algorithm with the RNA genetic algorithm of this paper, the following optimization results are obtained. It can be seen in Table 1, the algorithm in this paper can show a high success rate when performing different function optimization.

TABLE I
COMPARISON OF OPTIMIZATION RESULTS

Method	Functions	Optimal solution	Optimal solution coordinate	Success rate
SGA	$f_1(x)$	-809.56	(-413.97, -419.96)	68%
	$f_2(x)$	3.25	(7.5, 6.38)	76.4%
	$f_3(x)$	3156.78	(3.25e-9, 4.53e-9)	34.5%
	$f_4(x)$	-500.25	(-512, -493)	52.3%
RNA	$f_1(x)$	-837.96	(-420.86, -420.9)	100%
	$f_2(x)$	3.42	(7.95, 6.38)	99.5%
	$f_3(x)$	3600	(0, 0)	99.2%
	$f_4(x)$	-511.7	(-488.61, 512)	100%





For the problem that the optimization accuracy of the two test functions is not high for the positive RNA genetic algorithm [17], the RNA genetic algorithm in this paper is used for test comparison.

$$\begin{aligned}
 \min f_2 &= \frac{(x_1 - 100)^2 + (x_2 - 100)^2}{4000} \\
 &- \cos(x_1 - 100) \cos\left(\frac{x_1 - 100}{\sqrt{2}}\right) + 1, \\
 x_1, x_2 &\in [-600, 600]
 \end{aligned} \quad (5)$$

The functional equation is expressed by an equation the following optimization results are obtained in Table 2, the

algorithm in this paper can show a high success rate when performing different function optimization.

TABLE II
COMPARISON OF OPTIMIZATION RESULTS

Method	Functions	Optimal solution	Average evolutionary algebra	Success rate
plus-strand RNA		-511.73	500	95%
		0	349	70%
RNA		-511.73	167	100%
		0	274	75%

V. CONCLUSION

Inspired by the RNA editing of Octopus, an RNA genetic algorithm is proposed in this paper. Firstly, the genetic algorithm adopts the 0, 1, 2, 3 digital coding method corresponding to four RNA bases A, G, C, U in the initial coding. On one hand, the use of this digital coding method can avoid the problem of Hamming cliff; On the other hand, it can represent various characteristics of the bases, such as the complementarity of bases. Secondly, in order to reflect the self-editing characteristics of RNA, the specific base A is mutated into G in the process of mutation, and in the program expressed as 2 into 3, and the probability of 2 into 3 is higher, so that the mutation ability can be guaranteed on the basis of the general mutation probability. Thirdly, in order to verify the effectiveness of the algorithm, four different functions, such as Schwefel function, multi-peak function, NiH function and Rana function are tested. By optimizing the four functions, it is found that the algorithm can quickly jump out of the local optimal value and find the global optimal value with a higher probability; At the same time, set different parameters when testing the functions, such as chang the crossover probability and the mutation probability, and so on. When the crossover probability is large, the algorithm loses its optimization performance and becomes a random algorithm, while when it is low, the algorithm converges slowly. Finally, comparing with SGA, it is found that the SGA is easy to fall into the local optimum value and converge slowly when searching for these complicated functions, and it is difficult to jump out of the optimum value after many experiments, so the algorithm is effectiveness in this paper.

In the future, we will investigate in methods which can be useful for further explore about the influence of RNA editing mechanism on the evolution of mimetic Octopus.

REFERENCES

[1] Christinel H, Norah S, Healy H, et al. The evolution of conspicuous facultative mimicry in octopuses: an example of secondary adaptation[J]. Biological Journal of the Linnean Society, 2010, 101(1):68-77.
[2] Zullo L, Hochner B. A new perspective on the organization of an invertebrate brain[J] Communicative & Integrative Biology, 2011, 4(1):26-29.

[3] Shomrat T, Turchetti-Maia A L, Stern-Mentch N, et al. The vertical lobe of cephalopods: an attractive brain structure for understanding the evolution of advanced learning and memory systems[J]. J Comp Physiol a Neuroethol Sens Neural Behav Physiol, 2015, 201(9):947-956.
[4] Tao J, Wang N. DNA computing-based RNA genetic algorithm with applications in parameter estimation of chemical engineering processes[J]. Computers & Chemical Engineering, 2007, 31(12):1602-1618.
[5] Chang L, Ho C, Chen Y. Applying Multiobjective Genetic Algorithm to Analyze[J]. Journal of Water Resources Planning & Management, 2010, 136(5):539-546.
[6] Smith M, Leung D W, Gillam S, et al. Sequence of the gene for iso-l-cytochrome c in saccharomyces cerevisiae[J]. Cell, 1979, 16(4):753-761.
[7] Alon S, Garrett S C, Levanon E Y, et al. The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. [J]. 2015, 4.
[8] Wang L, Kimble J, Wickens M. Tissue-specific modification of gld-2 mRNA in C. elegans: likely C-to-U editing[J]. RNA-a Publication of the Rna Society, 2009, 10(9):1444-1448.
[9] Garrett S, Rosenthal J J. RNA editing underlies temperature adaptation in K+ channels from polar octopuses. [J]. Science, 2012, 335(6070):848-51.
[10] Caroline B. Albertin, Oleg Simakov, Therese Mitros. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature, 2015, 524, 220–224.
[11] Albertin C B, Simakov O, Mitros T, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties[J]. Nature, 2015, 524(7564):220-224.
[12] Nishikura K. Functions and Regulation of RNA Editing by ADAR Deaminases[J]. Annual Review of Biochemistry, 2010, 79(1):321-349.
[13] Yang Y, Lv J, Gui B, et al. A-to-I RNA editing alters less-conserved residues of highly conserved coding regions: Implications for dual functions in evolution[J]. Rna-a Publication of the Rna Society, 2008, 14(8):1516-1525.
[14] Gott J M, Visomirski L M, Hunter J L. Substitutional and insertional RNA editing of the cytochrome c oxidase subunit 1 mRNA of Physarum polycephalum.[J]. Journal of Biological Chemistry, 1993, 268(34):25483-6.
[15] Valleriani A, Zhang G, Nagar A, et al. Length-dependent translation of messenger RNA by ribosomes[J]. Physical Review E, 2011, 83(4 Pt 1):042903.
[16] Mai Te-Lun,Chuang Trees-Juen. A-to-I RNA editing contributes to the persistence of predicted damaging mutations in populations. [J]. Genome research,2019,29(11).
[17] Mingze Xu. Research on Plus-strand RNA genetic algorithms[D]. Zhejiang University,2011.