

Extended Abstract

A non-linear analogy procedure for gene repair

Xiaoyu Deng¹, Jun Meng^{2,*}

¹ Center for Data Mining and Systems Biology, College of Electrical Engineering, Zhejiang University; 2583588104@qq.com

² Center for Data Mining and Systems Biology, College of Electrical Engineering, Zhejiang University; junmeng@zju.edu.cn

* Correspondence: junmeng@zju.edu.cn; Tel.: +86-186-6810-5681

† Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12-16 June 2017.

Published: 9 June 2017

Abstract: Interactions between sub-systems and their environment are everywhere. In this paper, logistic map is applied, which felicitously stimulates the chaotic behavior of gene, to make analogy of a health DNA sequence and a mutation DNA sequence. By successfully repairing mutational part of a DNA sequence segment with help of neighbor-compressed method, we propose that compressing the neighbor healthy part to replay the mutational part is an effective way for gene repair. Since information ecosystem also possesses chaotic characteristics, the concept that extracting information from the neighboring environment can amend or complete information of corresponding elements also apply to information ecosystems.

Keywords: gene repair; logistic map; neighbor-compressed method; information ecosystem; neighboring environment.

1. Introduction

Universe is a nonlinear system[1] and chaos is everywhere[2]. With no doubt, DNA sequences also possess chaotic characteristics[3-5]. So it makes sense that using a chaotic model to simulate some biological process or status.

The logistic map is a polynomial mapping of degree 2, which is often cited as an archetypal example of how complex, chaotic behavior can arise from very simple nonlinear dynamical equations[6]. Taking advantage of logistic map for making analogies to living matters may help with further exploration about mysteries of life. In order to study the bioinformatic feature of DNA sequence and conduct an analogy procedure of repairing mutational part in DNA sequence, logistic map was used to generate a health DNA sequence and a mutation DNA sequence.

After conducting an effective gene repair, we found that compressing the neighbor healthy part to replay the mutational part is an effective way for gene repair. As DNA sequence and information ecosystems[7, 8] have chaotic characteristics, as well as behaviors inside body are similar to information ecosystems' interactions, it can extend to information ecosystems that extracting information from the neighboring environment can amend or complete information of subsystem.

2. Materials and Methods

2.1. Logistic map for DNA sequence

- Two different chaotic statuses were set, which respectively were health status with logistic map's coefficient $R=3.9$, and mutation status and with logistic map's coefficient $R=3.6$;
- Two number sets were generated by logistic map. one number set of 500 numbers all in health status was generated with $n=500$, $R=3.9$, and $x_n=0.5$. Another number set of 500 numbers was also generated with $n=500$ and $x_n=0.5$, but four fifth of this set was in health status with $R=3.9$ and one fifth of it was in mutation status with $R=3.7$.
- The two number sets were transformed to be nucleotide sequences, health DNA sequence and mutation DNA sequence. The rule transforming number set into 'A', 'G','C','T' sequence is presented in table 1.

Table 1. The rule transforming the number set into 'A' 'G' 'C' 'T'

Value of x_n ($n=1, 2, 3, \dots, 500$)	Corresponding letter
$0 \leq x_n < 0.28$	A
$0.28 \leq x_n < 0.60$	G
$0.60 \leq x_n < 0.84$	T
$0.84 \leq x_n \leq 1.00$	C

2.2. Neighbor-compressed method for DNA sequences repair

- Suppose that DNA sequence and $\{1, 2, 3, 4\}$ have a first mapping relationship as follows,

Where

$$\mathcal{A}_1 = \{A, G, C, T\},$$

and

$$\mathcal{B}_1 = \{1, 2, 3, 4\},$$

the first mapping relationship is

$$\mathcal{A}_1 \rightarrow \mathcal{B}_1.$$

- Because length of the healthy part in mutation DNA sequence is 3 times longer than the length of the mutational part, the healthy part should be compressed to be 1/4 of the original length so that the new sequence we get from compression could take place the mutational part. The rule that every four letters in sequences should be taken as a whole, which independently corresponds to one single number. Following is the fourth mapping relationship between all kinds of tetrad-letter and numbers (from 1 to 256) in 1/4 compression method:

where

$$\mathcal{A}_3 = \{AAAA, AAAG, AAAC, AAAT, AAGA, AAGG, AAGC, \dots, TTTA, TTTG, TTTC, TTTT\}$$

And

$$\mathcal{B}_3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots, 252, 253, 254, 255, 256\},$$

the fourth mapping relationship is

$$\mathcal{A}_3 \rightarrow \mathcal{B}_3.$$

For the sake of unifying the Y-axis,

$$C_3 = B_3/64.$$

For $\forall u \in C_3$,

set $v = 1$, when $u \in (0, 1]$,

set $v = 2$, when $u \in (1, 2]$,

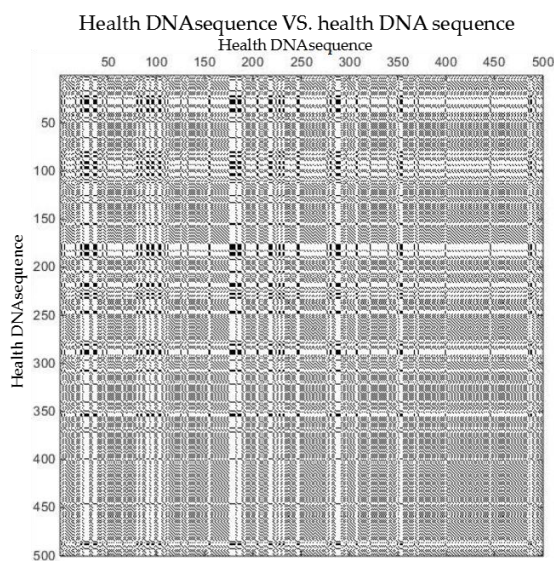
set $v = 3$, when $u \in (2, 3]$,

set $v = 4$, when $u \in (3, 4]$.

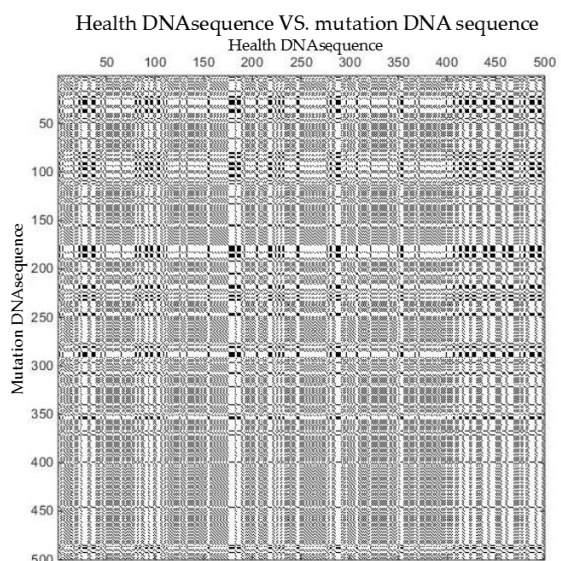
- After compressing the healthy part of the mutation DNA sequence and obtaining the new sequence, we displaced the mutational part by the new sequence. So the new built-up sequence was the repaired DNA sequence.

3. Result

- Function 'seqdotplot' of bioinformatics toolbox in matlab was used to create dot plots of four pairs, visualizing the matches between every pair, which made one sequence the X-axis and another sequence the Y-axis.



(a)



(b)

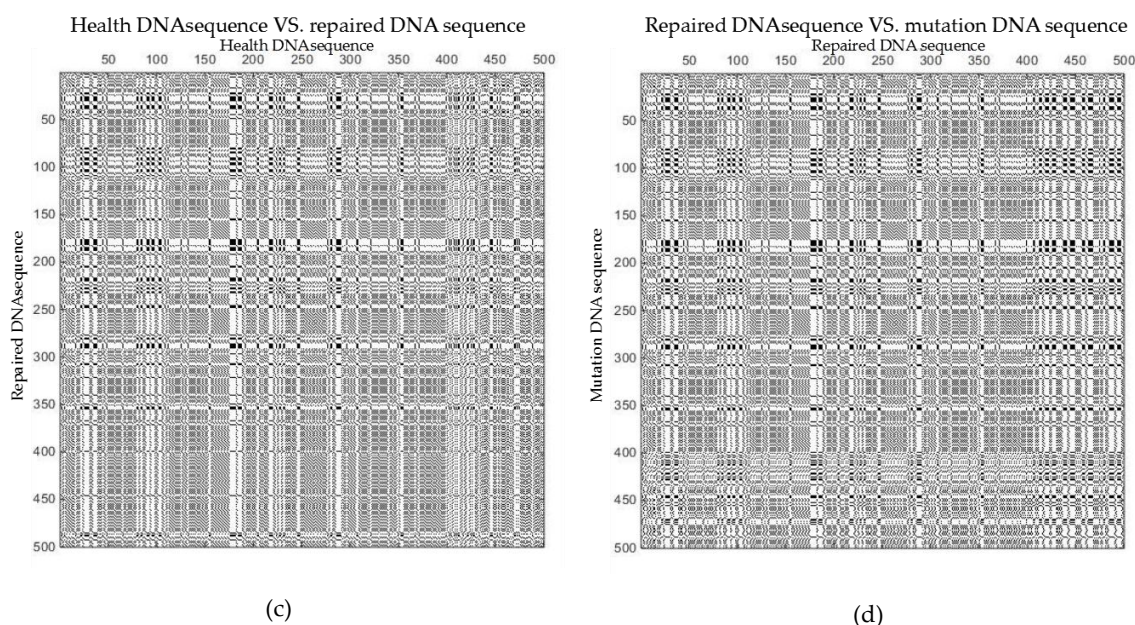


Figure 1. Four groups of pair comparison. (a) It is the dot plot of health DNA sequence and health DNA sequence, health DNA sequence as both X-axis and Y-axis. This group is a control group to compare with other groups. (b) It is the dot plot of health DNA sequence and mutation DNA sequence, health DNA sequence as X-axis and mutation DNA sequence as Y-axis. Comparing group (a) and (b), it is clear that mutational part of mutation DNA sequence is far different from corresponding part of healthy DNA sequence. (c) It is the dot plot of health DNA sequence and repaired DNA sequence, health DNA sequence as X-axis and repaired DNA sequence as Y-axis. Comparing group (a), (b) and (c), it is evident that repaired DNA sequence is much more similar to healthy DNA sequence than mutation DNA sequence. (d) It is the dot plot of repaired DNA sequence and mutation DNA sequence, repaired DNA sequence as X-axis and mutation DNA sequence as Y-axis. Comparing group (b) and group (d), it is obvious that these two plots is extremely familiar, which indicates that repaired DNA sequence and healthy DA sequence share great similarity.

- Rescaled range analysis (R/S) was used to obtain the Hurst exponent for quantitative analysis of similarities among health DNA sequence, mutation DNA sequence, and repaired DNA sequence.

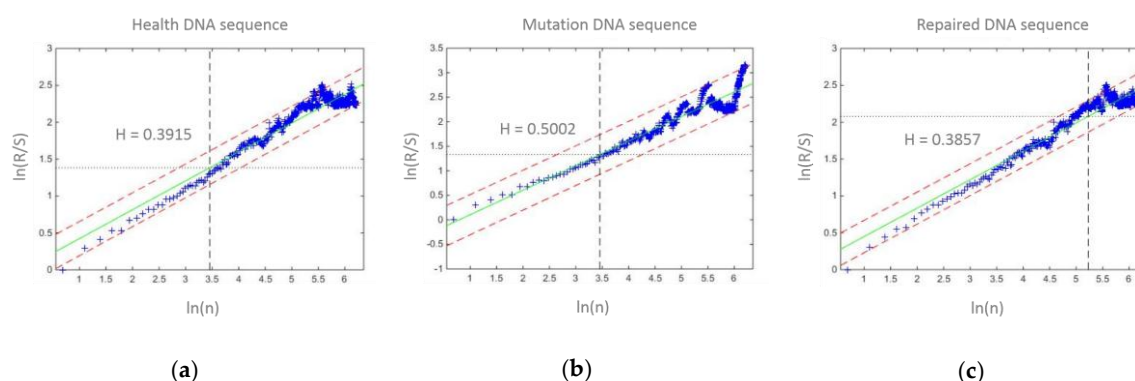


Figure 5. Obtaining the Hurst exponent by rescaled range analysis. Three sequences were analyzed by R/S and their $\ln(n)$ - $\ln(R/S)$ graph were drawn. The Hurst exponent is the slope of the $\ln(n)$ - $\ln(R/S)$. (a) It is $\ln(n)$ - $\ln(R/S)$ graph of health DNA sequence and its Hurst exponent is 0.3915. (b) It is $\ln(n)$ - $\ln(R/S)$ graph of mutation DNA sequence and its Hurst exponent is 0.5002. (c) It is $\ln(n)$ - $\ln(R/S)$ graph of repaired DNA sequence and its Hurst exponent is 0.3857.

That indicates health DNA sequence and repaired DNA sequence share high similarity in nonlinearity. On the contrary, mutation DNA sequence differs from them in nonlinearity.

As a result, visually or quantitatively, health DNA sequence and repaired DNA sequence proved to be much more similar to each other than mutation DNA sequence. So it shows this analogy procedure for gene repair is effective.

4. Discussion

- This kind of repair method is essentially to extract health information from mutational part's healthy neighbor, and to amend the mutational part by the health information. It exhibits that extracting information from the neighboring environment can amend or complete information of sub-system. This concept may also be appropriate for other systems possessing chaotic characteristics, such as information ecosystems.
- Organism can be regarded as a huge information ecosystem, where organs, tissues, cells or even DNA sequences exchange information to their neighboring environment. In the general information ecosystem, regardless of the actual boundary between an element and the environment, the former passes information, matter and energy through its internal structure. The actions and interactions of system components eventually create outflow from the system into the environment, so that the environment contain information and feature of element.
- That is to say, in the information ecosystem, environment can not only make influence on elements, further, it can reserve elements' information. Thus, if there is any element that is amiss or lost, its environment may provide help for amending or completing.

Acknowledgments: We are particularly grateful to Xiao Chen and Yiyang Jiang for valuable advice on simulation work.

Author Contributions: J.M. and X.D. designed and performed the simulation; J.M. and X.D. analyzed the; X.D. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scott, A. C., *The Nonlinear Universe: Chaos, Emergence, Life*. Springer Publishing Company, Incorporated: 2007.
2. Barnsley, M. F., - Fractals Everywhere (Second Edition). *Fractals Everywhere* **1993**, 412–415.
3. Peng, C. K.; Buldyrev, S. V.; Goldberger, A. L.; Havlin, S.; Mantegna, R. N.; Simons, M.; Stanley, H. E., Statistical properties of DNA sequences. *Physica A Statistical Mechanics & Its Applications* **1995**, 221, (1–3), 180-192.
4. Chatzidimitrioudeismann, C. A.; Larhammar, D., Long-range correlations in DNA. *Nature* **1993**, 361, (6409), 212-3.
5. Gutierrez, J.; Iglesias, A.; Rodriguez, M.; Burgos, J.; Moreno, P., Analyzing the multifractal structure of DNA nucleotide sequences. *Chaos and Noise in Biology and Medicine* **1998**, 7, 315-319.
6. Boeing, G., Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction. *Social Science Electronic Publishing* **2016**.
7. Manson, S. M., Simplifying complexity: a review of complexity theory. *Geoforum* **2001**, 32, (3), 405-414.
8. Chen, M., Game Behavior Analysis of Information Resources Allocation in Information Ecosystem. *Information Studies Theory & Application* **2010**, 54, (4), 15-16.



© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)