ISSN 2231-346X (Print)
JUSPS-A Vol. 30(12), 409-412 (2018). Periodicity-Monthly
Section A
ISSN 2319-8044 (Online)

JOURNAL OF ULTRA SCIENTIST OF PHYSICAL SCIENCES
An International Open Free Access Peer Reviewed Research Journal of Mathematics
website:- www.ultrascientist.org
Estd. 1989

Modelling of Friedreich's Ataxia and other Genetic Disorders as Defective / Noisy Genetic Information Processing

SWASTI WAGH[1] and D.K. WAGH[2]

[1]M.Sc.. Applied Mathematics from G.S. Institute of Tech. and Sc., Indore, M.P. (India)
[2]Ph.D., Retired Prof. of Mathematics from G.S. Institute of Tech. and Sc., Indore, M.P., (India)
Corresponding Author Email ; swastiwagh@gmail.com : waghdiwakar42@gmail.com
http://dx.doi.org/10.22147/jusps-A/301201

Acceptance Date 19th December, 2018, Online Publication Date 27th December, 2018

Abstract

Genetic disorders are due to mutation of some gene. For example, Friedreich's Ataxia is due to mutation of frataxin gene. The mutation causes transmission error (noise) in the transmission channel. In this paper we show how this transmission error can be calculated using information theory approach.

Key words : Genetic disorder, Freiderich's ataxia, Information theory, Transmission error.

1. Introduction

Information theory was founded by Claude Shannon in 1948 to construct Mathematical model of communication system[1,2,3]. Although the information theory was first developed to study transmission of messages through communication channels, it was later applied to other research fields. Now a day's information theory is not a part of communication theory only but finds application in physics, computer science and biotechnology. Living organisms process and transmit information at many levels including the mechanism of inheritance. They can be considered as open thermodynamic system which constantly exchange matter, energy, entropy and information. DNA double helix in the cells is the physical unit which plays the basic role in coding and transmission of the information necessary to maintain life. This understanding motivated the application of information theory to DNA sequence analysis somewhere in 1970-77 [4]. The researches during this period were devoted to define significant parameters and their estimation. The aim of these studies was to obtain quantitative measures of the complexities of DNA sequences. However, the real success of the information theory application to DNA sequence analysis could be possible after 1987 when long range DNA sequence data become

available.The references of recent work on the application of information theory to DNA sequence analysis are given in[5,6].

Friedreich Ataxia[7,8,9] is a rare genetic disorder and is incurable till now. It is autosomal recessive and is caused when both copies of FXN gene are mal functioning. Because of this adequate quantity of frataxin protein, which is required for Fe-S cluster biogenesis, is not produced. This leads to loss of energy production in mitochondria and deposition iron on the cell causing its death. DNA sequence of Friedreich's ataxia patients show increase in GAA repeats. This gene mutation corrupts the message and creates noise in communication channel. Hence the adequate quantity of frataxin protein is not produced and normal functioning of the gene is affected.

Other genetic disorders are due to mutation of some other gene. Due to these mutations the error in biological transmission increases. This increase in error is reflected as symptoms of genetic disorder. By now we know many types of genetic disorders such as FRDA, SCA, and Parkinson.These disorders make the life of patients very uncomfortable and dependent on others.

The objective of this paper is to calculate the transmission error in Friedreich's ataxia due to mutation of FXN gene. The method is illustrated by taking a hypothetical small DNA sequence. This sequence is then mutated by removing some codon and introducing some GAA repeats. Then we calculate the transmission error for both the sequences. The difference in the two errors gives the error due to mutation. The procedure can be applied to actual DNA sequences of normal and diseased person and the transmission error due to GAA repeats can be calculated .The procedure of calculation of transmission error in Friedreich's ataxia due to mutation of FXN gene can be used for other genetic disorders also.

2. Modelling of Biological Information Processing As Communication Channel :

The biological information processing can be modelled as communication channel, where input is the sequence of four nucleotides a = adinine, t = thimine, c = cytosine, g =guanine and output is the amino acid chain in the protein. The information source generates sequence of symbols (messages) chosen from a finite alphabet set 'X' given by,

$$X = \{a, t, c, g\} \tag{2.1}$$

The elements of set X are called bases and represent nucleotides .In general, symbols of the set X are not emitted with equal probability. The sum of the probabilities with which a, t, c, g are emitted is one,

$$p(a) + p(t) + p(c) + p(g) = 1 \tag{2.2}$$

The output sequence is the sequence of codons (triplets of bases).The information is coded in terms of codons. These codons represent amino acids required to synthesise particular protein. The synthesis starts from the start codon and ends at the stop codon.

To illustrate the mathematical procedure, we consider a small hypothetical sequence 'S' of codons as given below,

$$S = (\text{atg}, \text{ttt}, \text{atc}, \text{atc}, \text{act}, \text{tag}) \tag{2.3}$$

Let us suppose that the above sequence gives a message to synthesise frataxin protein in a normal person. In FRDA mutation of frataxin gene occurs and the sequence S is mutated by unwanted GAA repeats. Let the mutated sequence of FRDA patient be,

$$S' = (\text{atg},\text{ttt},\text{atc},\text{gaa},\text{gaa},\text{gaa},\text{act},\text{tag}) \tag{2.4}$$

Let Y = (1, 2, 3) denote the position of bases in codon.

We then calculate the probability matrices p(X) and p'(X) for the sequence S and muted sequence S' and obtain,

$$p(X) = (5/18 , 4/9 , 1/6 , 1/9) \tag{2.5}$$

$$p(X) = (5/12 , 7/24 , 1/12 , 5/24) \tag{2.6}$$

The elements of these matrices give the probabilities with which the bases a,t,c,g are generated in the sequence S and S' respectively

Then we calculate the conditional probability matrices P(Y/X) and P'(Y/X)for the sequence S and S'.The rows of the conditional probability matrix give the probabilities with which the bases a,t,c,g occupy the position 1,2,3 in the codon. This gives,

$$P'(Y/X) = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

$$P'(Y/X) = \begin{bmatrix} \frac{3}{10} & \frac{4}{10} & \frac{3}{10} \\ \frac{2}{7} & \frac{3}{7} & \frac{2}{7} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{5} & 0 & \frac{2}{5} \end{bmatrix} \tag{2.8}$$

The joint probability matrix P(X, Y) is obtained by multiplying the rows of P (Y/X) by P (a), P (t), P(c) and P (g) respectively. Thus for the sequence S, we get,

$$P\ (XY) = \begin{bmatrix} \frac{2}{9} & \frac{1}{18} & 0 \\ \frac{1}{9} & \frac{2}{9} & \frac{1}{9} \\ 0 & \frac{1}{18} & \frac{1}{9} \\ 0 & 0 & \frac{1}{9} \end{bmatrix} \tag{2.9}$$

The rows of this matrix give the probabilities with which the bases a,t,c,g occupy the positions 1,2,3 in the codon.

Simillarly, for the mutated sequence S', we obtain the joint probability matrix P'(X Y) as,

$$P'(XY) = \begin{bmatrix} \frac{1}{8} & \frac{1}{6} & \frac{1}{8} \\ \frac{1}{12} & \frac{1}{8} & \frac{1}{12} \\ 0 & \frac{1}{24} & \frac{1}{24} \\ \frac{1}{8} & 0 & \frac{1}{12} \end{bmatrix} \tag{2.10}$$

The entropy H (Y/X) is the measure of error (noise) in transmission system and is given by,

$$H(Y/X) = -\sum_{J=1}^{4} \sum_{K=1}^{3} P\ (x_j, y_k) \log_2 P\ (y_k/x_j) \tag{2.11}$$

Using this formula, we calculate the entropies H(Y/X) and H'(Y/X) for the sequence S and S' and obtain H(Y/X) = 1.02 bits, H'(Y/X) = 1.392 bits.

As mentioned above H(Y/X) is the error (noise) in the communication system. Hence the difference

$$H'(Y/X) - H(Y/X) = 0.372 \text{ bits} \tag{2.12}$$

gives the error (noise) due to mutation of the sequence S.

3. **Discussion and Result**

In the previous section we have given the method to calculate transmission error due to mutation of a small hypothetical DNA sequence. The same procedure can be applied to calculate transmission error of actual DNA sequences which are quite large.

In the problem considered above, the sequence S was supposed to convey message for preparing frataxin protein. This sequence was then mutated by introducing repeats of GAA codons.Because, FRDA is a genetic disorder in which DNA sequence of frataxin gene is mutated and is found to contain large number of GAA repeats.

FRDA is a progressive disease. Number of GAA repeats increase with the progress of the disease. Calculating the transmission error at different stages we can determine the average rate at which transmission error is increasing. It may be then possible to correlate clinical symptoms with rate at which error is increasing.

4. **Concluding Remarks**

The scope of modelling of genetic disorders as biological information transmission through a noisy channel is not limited to estimate redundancy or divergence in DNA sequences of genetic disorders and to estimate error in biological information transfer. But it has far reaching implications. Information theory approach will add a new dimension to ataxia cure research. Because, Shannon has shown that it is possible to have almost error free communication in a noisy channel by maintaining the transmission rate below the channel capacity. So, cure of genetic disorders is possible if we can develop medicines, radiations or anything else which can control rate of biological information transmission.

References

1. Claude E Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, Vol. *27*, pp. 379-423, 623-656, (1948).
2. Claude E Shannon and Warren Weaver, The Mathematical Theory of Communication,The University of Illinois Press, Urbana, Illinois, 1949, ISBN 0-252.
3. Thomas M Cover, Joy A,Thomas, Elements of Information Theory, Wiley Interscience, A Jhon Wiley & Sons Inc., Second edition July (2006).
4. Ramon Roman – Rolda N Pedro Bernaola-Galvan and Jose L. Oliver, Application of information Theory to DNA sequence Analysis-A Review, Pattern Recognition,Vol. *29*, pp. 1187-1194, (1996) Elsevier Science Ltd.
5. Ahmad A. Rushdi A Mathematical Model of DNA Replication, IMACST:Vol.*1* Number 1 Decemer (2010).
6. Susana Vinga, Information Theory Applications for Biological Sequence Analysis ,Briefing in Bioinformatics Advancs Access Published Sept (2013).
7. Frequently asked questions about ataxia. National Ataxia foundation (NAF) publication (2014).
8. Friedreich Ataxia Booklet published by Ataxia UK (2015).
9. Role of frataxin Wikipedia, the free encyclopaedia, (https://en.wikipedia.org/wiki/frataxin), 11 Dec (2018).