The Language Of Genes

Sambhav Sharma, Nitish Rawat, Bairagi Nath Behera.
Department Of Computer Science and Engineering,
National Institute Of Technology, Calicut

Abstract

Just in a short time after the DNA structure became known to us and with that we came across knowing various gene expressions, we came across Noam Chomsky work about the expression and representation of languages, its grammar, syntaxes, computation and ambiguity[1].

Now as we know that the DNA is a richly-expressive language that determines the various structures and functions of life and it further determines the structure of RNA and Proteins too. Hence the interest started in this field, developing gene regulation , gene structure and expression,and other forms of mutation and rearrangement ,conformation of macromolecules and computational of the biological data we have. So the trend increased in developing grammars for the biological sequesnce. This is done in order to make it easy to perform various operations like searching, pattern recognisation etc[1].

Here we intend to present a formal language based linguistic heriustics of biological sequences from basic mathematical linguistic and its applications and then extending those all to DNA and Protein sequences. Just by using the Formal Language concept and realting the biological sequence with that we bring on the computational results and further interweaving of these fields will be instrumental in extending our understanding of the language of life[2].

References

[1] The language of genes - David B. Searls, Bioinformatics Division, Genetics Research, GlaxoSmithKline Pharmaceuticals.

[2] The Computational Linguistics of Biological Sequences - David B. Searls.