

Sažetak:

Genetic code holds information that is written in the linear structure of nucleotides, which are marked with four letters (A, G, C, T/U) and translated to amino acids with higher structure. Since its deciphering, it has been viewed as a language of life and many parallels between DNA and natural language have been developed, such as linearity of signs, coding, the randomness of the code, its reading and decoding, transcription, translation, information, editing (splicing), etc. Concurrently, possibilities of analysing DNA – genetic text – with linguistic methods have been explored.

This paper aims to shortly introduce main theoretical concepts based on the analogy between natural language and DNA – so-called DNA grammars – and present the most often applied linguistic methods for DNA analysis (e.g. Damerau-Levenshtein distance, Bag-of-Words model, Zipf's law, Menzerath-Altmann law).