

基于非线性方法的DNA序列分析

赵小杰¹、刘勇¹、姚力¹、陈惟昌²

1 北京师范大学信息学院

2 中日友好临床医学研究所生物物理研究室

以人类DNA序列为原始数据, 对其进行数字编码。在利用功率谱分析编码DNA序列的基础上, 利用Hurst指数进一步分析序列的自相似性。从分析结果看出, DNA序列中的确存在长程相关现象, 而且这种长程相关现象与DNA的组成结构有关, 表现为它的结构基团的Hurst指数大于其功能基团的Hurst指数。同时, 从内含子含量的角度分析序列的长程相关程度, 结果表明, 同一序列中, 内含子含量与不同的化学基团具有不同的关系。这表明不同的化学结构对DNA序列特征具有不同的贡献。这些结论从非线性方法的角度对DNA序列的分析提供了新的思路。

ANALYSIS OF DNA SEQUENCE CONSTRUCTURE BASED ON NONLINEAR METHOD

The primary data based on the human DNA sequences were digitally coded. After the coded DNA sequence being analyzed by power spectrum, its self-comparability was further analyzed by using the Hurst index. The results show that the long range correlation does exist in the DNA sequences, which relates to the DNA composition constructions. It denotes that the Hurst index of structural group is larger than the Hurst index of functional group. At the same time, the long range correlation with intron content was also analyzed. The results show that the intron content is correlated differently with different chemical group in the same sequence. It indicates that the different chemical construction contribute to the DNA sequence characteristic differently. These conclusions suggest that it would be a new method to analyze the DNA sequence construction by nonlinear method.

关键词

长程相关(Long range correlation); 功率谱(Power spectrum); Hurst指数(Hurst index); M编码(M coding)