

DNA序列的“双符三阶”图形编码

陈志华^{1,2}、陈惟昌²、陈志义³、王自强²、房青²、李红艳²

1 中国医学科学院 中国协和医科大学

2 中日友好临床医学研究所生物化学与分子生物学研究室

3 中国科学院自动化研究所国家模式识别实验室

DNA的图形编码是在几何意义下,在不同位置,用不同的标记符号及不同的方向线段,对DNA的序列进行编码。DNA图形编码相对于DNA的字符编码而言,具有直观、简明、形象和便于比较局部DNA序列的相似性等特点。在分析已知各类DNA的图形表示模式的基础上,提出一种DNA序列的“双符三阶”图形编码,并以此对一些特异DNA编码序列进行分析。DNA图形编码与DNA字符编码呈一一对应关系,具有简便易行、编译方便、形象丰富、便于比较等优点。适用于DNA短序列的相似性检测与分析,在生物信息学上有一定的应用前景。

BINOTE AND TRISTAVE GRAPHIC CODE OF DNA SEQUENCE

The graphic coding of DNA sequence is the geometric representation of DNA sequences by means of different notes and line segments in different positions. The letter coding (C, T, A, G) of DNA is the most fundamental representation of DNA sequence for information storage, comparison and analyzing of the gene data. But as compared to the DNA letter coding, the DNA graphic coding is more intuitive, direct visualizing and more convenient for searching and comparing the similarity of DNA segments. After analyzing different formats of DNA graphic representations, a new graphic coding pattern for DNA sequence with binote and tristave method was suggested. The binote-tristave graphic code of DNA sequence is originated from the horizontal displacement of the digital coding plane of DNA nucleotides. It is one to one corresponding to the DNA letter coding, plotable and readable by computer. The binote-tristave DNA graphic code is especially suitable for searching and comparing the unknown DNA short sequences and is of significant value for the bioinformatic analysis of gene data.

关键词

DNA字符编码(DNA letter coding); DNA数字编码(DNA digital coding); DNA双符三阶图形编码(DNA binote-tristave graphic coding)