生物物理学报(2002) 18(2), 206-212

PDF文档

遗传密码格式的组合编码数分析

陈惟昌[1]、陈志义[2]、陈志华[3]、王自强[1]
1 中日友好临床医学研究所生物物理研究室
2 中国科学院自动化研究所国家模式识别实验室
3 中日友好临床医学研究所生物化学及分子生物学研究室

用N个密码子对m个编码对象进行编码的编码格式是m元N维空间中的一个顶点。64个密码子对20种氨基酸和终止密码子进行编码格式的组合编码数是一个十分巨大的数字。对多元高维编码空间的拓扑特性进行了分析和研究,并由此推导出m-N空间的特性三角的排列方式以及给出特性三角公式的数学证明。指出,目前的遗传密码的编码格式是21元64维编码空间的一个顶点。应用组合数学分析的方法,计算了遗传密码格式的最大组合编码数C_M=4.19×10^{84},基因组遗传密码的组合编码数C_G=1.13×10^{80}以及线粒体遗传密码的组合编码数C_T=1.38×10^{79}等。分析结果表明,遗传密码的指定是一个小概率事件,可能来源于λ简并后的偶数三联密码配对的组合编码的对称破缺。

ANALYSIS OF COMBINATORIAL CODING NUMBERS OF THE GENETIC CODE PATTERNS

The coding pattern which uses N codons to encode m objects is a vertex in N dimension space of m elements. The combinatorial coding number of 64 codons to encode 20 amino acids and the terminate code is very huge. The topological properties of the polynomial high dimension spaces (the m-N spaces) were first analyzed and the characteristic triangles (Chen Weichang Triangles) of the m-N spaces were obtained. A mathematical proof of the characteristic triangles was also given. Obviously, the coding pattern of the genetic code is a vertex in a 64 dimension space of 21 elements. Using the combinatorial mathematical method, the following combinatorial numbers of genetic coding patterns had been calculated: the maximum combinatorial number of genetic coding patterns C_m(Cm=4.19×10^{84}); the combinatorial number of genomic coding pattens CG(CG=1.13×10^{80}, the combinatorial number of mitochondrial coding patterns CT(CT=1.38×10^{79}. It is suggested that the determination of the genetic code is an event of extremely small probability. The origin of the genomic genetic codes might be from the symmetry breaking of the triplets pairs of the mitochondrial codes.

关键词

遗传密码(Genetic code); 编码对象(码象)和编码元(码子)(Coded object (codim) and coding element (codon)); 组合编码数(Combinatorial coding numbers); 多元高维空间(高维栅格空间)(Polynomial high dimension space (hypergrid space)); 多项式定理(Polynomial theorem); 特性三角(广义贾宪与帕斯卡三角)(Chen Weichang triangle (extended Jia Xian and Pascal triangle))