生物物理学报(2002) 18(1), 71-75

PDF文档

大肠杆菌、酵母和果蝇基因保守位点的信息熵分析

吕军[1,2]、李宏[1]、马克健[1]
1　内蒙古大学理论物理和理论生物物理研究室
2　内蒙古工业大学物理教研室

对大量的大肠杆菌(Escheri chiacoli)、酵母(Yeast)和果蝇(Drosophila melanogaster)已知基因起始密码子和终止密码子上、下游各30个碱基序列,用重新定义单碱基信息冗余(记为D_1(l),l是位点)和紧邻碱基的信息冗余(记为D_2(l)),统计计算每个位点的D_1(l)和D_2(l)值。从结果看,双碱基比单碱基携带更多的信息;酵母和果蝇基因起始密码子上游-3位点D_1(-3)和D_2(-3)有一明显峰值;大肠杆菌基因起始密码子上游SD区域D_1(l)和D_2(l)有明显峰值,与他人结论相同。发现酵母基因起始密码子下游的+4位点与+5位点的紧邻碱基的D_2(l)有一峰值,其关联模式为TC(联合概率为0.211).这说明用重新定义的信息冗余去确认DNA序列中存在的保守位点是完全可行的。

AN INFORMATION ENTROPY ANALYSIS OF CONSERVATIVE SITES OF E.coli、 YEAST AND Drosophila GENES

The formulation of the single base information redundancy D_1(l) and the adjacent base related information redundancy D_2(l) are revised. For the sequences of upstream and down-stream the start codon and the terminal codons of E.coli, yeast and Drosophila genes, the D_1(l) and D_2(l) for each site l (l=-30, -29, …, +32, +33) are calculated. The results shown that D_2(l) have more information than D_1(l). In site -3 of coding start sequences, D_1(-3) and D_2(-3) have a distinct peak value for yeast and Drosophila. In the SD region of E.coli gene sequences, D_1(l) and D_2(l) have obvious peak value distribution, which is consistent with the others' results. D_2(l) in site +4 of coding start sequences in yeast also have a peak value, whose related mode is TC (the combined probability is 0.211). Therefore, the revised information redundancies applied in this thesis are feasible to confirm the conservative sites in DNA sequence.

关键词

信息熵(Information entropy); 关联(Correlation); 保守位点(Conservative sites)