

大肠杆菌Sigma70启动子的识别

张颖、贾芸、吕军*
内蒙古工业大学理学院

应用多样性增量结合二次判别分析 (Increment of Diversity with Quadratic Discriminant analysis, IDQD) 方法, 对大肠杆菌 σ 70启动子进行识别。使用受试者操作特性 (receiver operating characteristic, ROC) 曲线和精度召回率曲线 (Precision Recall Curves, PRC) 进行性能评估。10-fold交叉检验给出, 在正负集之比为1:1时, ROC曲线下面积和PRC曲线下面积均为95%。结果表明, IDQD算法有能力应用于原核启动子的识别。识别精度高于现有算法。

The Recognition of Sigma70 Promoters in Escherichia coli K-12

The method of Increment of Diversity with Quadratic Discriminant analysis (IDQD) was used to predict the σ 70 promoters in Escherichia coli K-12. The performance evaluations by using Receiver Operator Characteristics (ROC) and Precision Recall Curves (PRC) were carried out. 10-fold cross-validation test gave out the results that area under ROC (auROC) and area under PRC (auPRC) were all 95% when positives/negatives ratio was 1:1. Results show the IDQD method is capable at recognition of prokaryotic promoter regions, and the accuracy is better than other top softwares currently published.

关键词

大肠杆菌 (Escherichia coli K-12); 启动子 (promoter); 多样性增量 (Increment of Diversity); 二次判别分析 (Quadratic Discriminant analysis)