

细菌 ncRNA 基因预测工具的评估

金城

(《微生物学通报》编委会 北京 100101)

细菌非编码 RNA (ncRNA) 在 RNA 的转录调节、染色体复制、RNA 加工与修饰、mRNA 翻译与稳定性、蛋白质降解与转运以及细菌感染等许多过程中都发挥着重要作用^[1]。目前, 基于实验鉴定的细菌 ncRNA, 已开发出不同的生物信息学预测方法^[3-6], 其中比较基因组学和转录单元预测方法是目前最常用的方法。

比较基因组学和转录单元预测方法的结合使用, 具有较高的特异性, 但大多数方法的普适性较低, 不能够预测菌株特异的 ncRNA 基因, 而基于机器学习方法的预测工具受阳性训练集以及特征向量参数的制约, 假阳性很高, 同时阳性检出率很低。本期发表了刘林梦、欧竑宇等的论文“利用不同 G+C 含量细菌基因组评估细菌 ncRNA 基因预测工具”^[7], 作者针对细菌 ncRNA 基因预测工具普适性低的问题, 比较了不同 G+C 含量细菌基因组中 ncRNA 基因转录起始区域和终止区域的序列特征, 发现了随基因组 G+C 含量不同, ncRNA 基因转录单元特征存在明显差异。进而对 3 个代表性工具 sRNAPredict、PORTRAIT 和 sRNAscanner 进行了预测准确性的评价, 发现 3 种 ncRNA 基因工具预测效果随基因组 G+C 含量变化而不同, 不同 G+C 含量基因组中 ncRNA 基因的转录起始和终止区特征可作为 ncRNA 基因预测的重要参数之一。这些结果将能够为不同细菌的 ncRNA 基因预测提供一些有益的参考。

由于已知的实验验证细菌 ncRNA 基因数量上的限制, 本文研究工作的样本量还较小, 相信通过系统整合 RNA-seq 转录组测序信息, 扩大训练集中的细菌 ncRNA 基因数量, 该论文的结果将对细菌 ncRNA 基因预测工具的开发有更大帮助。

关键词: 细菌 ncRNA 基因, G+C 含量, 预测

参考文献

- [1] Li W, Ying X, Lu Q, et al. Predicting sRNAs and their targets in bacteria[J]. *Genomics, Proteomics & Bioinformatics*, 2012, 10(5): 276-284.
- [2] Pichon C, Felden B. Small RNA gene identification and mRNA target predictions in bacteria[J]. *Bioinformatics*, 2008, 24(24): 2807-2813.
- [3] Rivas E, Klein RJ, Jones TA, et al. Computational identification of noncoding RNAs in *E. coli* by comparative genomics[J]. *Current Biology*, 2001, 11(17): 1369-1373.
- [4] Sridhar J, Sambaturu N, Sabarinathan R, et al. sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes[J]. *PLoS One*, 2010, 5(8): e11970.
- [5] Gruber AR, Neuböck R, Hofacker IL, et al. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures[J]. *Nucleic Acids Research*, 2007, 35: W335-W338.
- [6] Arrial RT, Togawa RC, Brigido MM. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*[J]. *BMC Bioinformatics*, 2009, 10(1): 239.
- [7] 刘林梦, 温权, 欧竑宇. 利用不同 G+C 含量细菌基因组评估细菌 ncRNA 基因预测工具[J]. *微生物学通报*, 2014, 41(12): 2583-2592.

Assessment of bacterial ncRNA gene prediction tools

JIN Cheng

(The Editorial Board of Microbiology China, Beijing 100101, China)

Keywords: Bacterial ncRNA gene, G+C content, Prediction