

实验研究

日本血吸虫乙醛脱氢酶全长基因编码序列的预测、验证与分析

王玮^{1, 2}, 刘德立², 胡薇¹, 冯正¹, 杨忠^{3, 4} *

1 中国疾病预防控制中心寄生虫病预防控制所, 世界卫生组织疟疾、血吸虫病和丝虫病合作中心, 卫生部寄生虫病原与媒介生物学重点实验室, 上海200025; 2 华中师范大学生命科学院, 武汉430079; 3 上海生物信息技术研究中心, 上海200235; 4 华东理工大学生物反应器工程国家重点实验室, 上海200237

收稿日期 修回日期 网络版发布日期 接受日期

摘要

目的 利用生物信息学和实验相结合的方法, 补平已知拼接序列中的缺失片段, 获得日本血吸虫乙醛脱氢酶全长基因编码序列。方法 通过生物信息学方法从已公开发表的日本血吸虫转录组数据库中提取乙醛脱氢酶表达序列标签(EST)序列数据, 与其他物种的同源基因进行多序列联配, 寻找分别与同一蛋白氨基酸序列的N端和C端配对的序列; 设计引物, 用RT-PCR扩增得到全长基因中间的缺失片段并测序。最终获得全长基因序列, 并分析该蛋白的理化性质。结果 找到日本血吸虫乙醛脱氢酶基因的可能EST序列片段8条, 对其中的1对EST序列(AAW27891和AAW27047对应的氨基酸序列, 中间缺少约80个氨基酸)进行blastn比对结果, 预测为同一基因的2条片段。根据这两对EST序列设计正、反向引物, 通过PCR扩增、对扩增产物进行测序及序列的生物信息学鉴定, 找回了缺少的核酸序列, 并与预测序列大小大致吻合(430 bp)。组合成1条完整的日本血吸虫乙醛脱氢酶基因的编码序列(提交GenBank, 其登录号为EF503564)。ORF全长为1 596 bp, 编码531个氨基酸, 编码的蛋白相对分子质量理论值为Mr 573 30.7, pI值为7.94, 此序列的290~297位氨基酸与乙醛脱氢酶的模式序列 [LIVMFGA] -E- [LIMSTAC] - [GS] -G- [KNLM] - [SADN] - [TAPFV] 相匹配。结论 利用生物信息学和实验相结合的方法, 可以补平已知拼接序列中的缺失片段, 获得日本血吸虫乙醛脱氢酶的全长基因编码序列。

关键词 [日本血吸虫](#) [乙醛脱氢酶](#) [生物信息学](#) [序列拼接](#)

分类号

Putation, Identification and Bioinformatics Analysis of *Schistosoma japonicum* Aldehyde Dehydrogenase Full Coding Sequence

WANG Wei 1, 2, LIU De-li 2, HU Wei 1, FENG Zheng 1, YANG Zhong 3, 4 *

1 National Institute of Parasite Diseases, Chinese Centre for Disease Control and Prevention, WHO Collaborating Centre for Malaria, Schistosomiasis and Filariasis, Shanghai 200025, China; 2 College of Life Sciences, Central China Normal University, Wuhan 430079, China; 3 Shanghai Center of Bioinformatics and Technology, Shanghai 200235, China; 4 State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China

Abstract

Objective To acquire the full coding sequence of *Schistosoma japonicum* aldehyde dehydrogenase, and fill the gaps of the partial aldehyde dehydrogenase sequences. Method Putative sequence fragments of the *S. japonicum* aldehyde dehydrogenase were extracted from the transcriptome database by use of bioinformatics tools, through the multiple sequences alignment with homologous sequences of other species. Primers were designed according to the EST sequences matching the N terminal and C terminal respectively, and the gap sequence fragment was amplified by RT-PCR and sequenced. The full gene sequence was obtained finally by combining the old 2 EST sequences with the amplified sequence. The physico-chemical parameters of the new sequence were analyzed by using bioinformatics software. Result Eight EST sequences of *S. japonicum* were predicted as partial sequences of aldehyde dehydrogenase. Two of which (AAW27891, AAW27047) were predicted to represent the N terminal and C terminal of one protein, respectively. The gap between them was deduced as about 80 amino acids according to the result of multiple sequences alignment. Primers located on the flanking of the gap were designed according to the known EST sequences of AAW27891 and AAW27047. The gap between the AAW27891 and AAW27047 were obtained by RT-PCR and then sequenced, as well as confirmed by bioinformatics software. The full sequence of aldehyde dehydrogenase was reassembled by filling of the gap sequence. The reassembled gene coding sequence was submitted to GenBank with an accession number of EF503564. The coding sequence contains an intact ORF of

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF \(276KB\)](#)

▶ [\[HTML全文\]\(OKB\)](#)

▶ [参考文献\[PDF\]](#)

▶ [参考文献](#)

服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“日本血吸虫”的 相关文章](#)

▶ 本文作者相关文章

· [王玮](#)

·

· [刘德立](#)

· [胡薇](#)

· [冯正](#)

· [杨忠](#)

·

1 596 bps with deduced 531 amino acids. Bioinformatic analysis of new amino acids sequence was performed as deduced molecular weight of 57 330.7 and PI value of 7.94. The aldehyde dehydrogenase pattern of [LIVMFGA] -E- [LIMSTAC] - [GS] -G- [KNLM] - [SADN] - [TAPFV] was found located in the position 290-297 of the new sequence. Conclusion The gap between two partial nucleotide sequences is filled and the full coding sequence of aldehyde dehydrogenase gene has been obtained by the method combining bioinformatics tools and experiments together.

Key words [Schistosoma japonicum](#) [Aldehyde dehydrogenase](#) [Bioinformatics](#) [Sequence splicing](#)

DOI :

通讯作者 杨忠 yangzzh@ecust.edu.cn

作者个人主页 王玮1; 2; 刘德立2; 胡薇1; 冯正1; 杨忠3; 4 *