

学术探讨

单体型组装MEC问题的参数化算法研究

谢民主^{1,2}, 王建新², 陈建二²

1. 湖南师范大学 物理与信息科学学院, 长沙 410081

2. 中南大学 信息科学与工程学院, 长沙 410083

收稿日期 修回日期 网络版发布日期 2007-11-29 接受日期

摘要 单体型组装MEC问题指如何利用个体的DNA测序片段数据, 翻转最少的SNP位点值以确定该个体单体型的计算问题。根据片段数据的特点提出了一个时间复杂度为 $O(nk_22^{k_2+m\log m+mk_1})$ 的参数化算法, 其中 m 为片段数, n 为单体型的SNP位点数, k_1 为一个片断覆盖的最大SNP位点数 (通常小于10), k_2 为覆盖同一SNP位点的片断的最大数 (通常不大于10)。对于实际DNA测序中的片段数据, 即使 m 和 n 都相当大, 该算法也可以在较短的时间得到MEC问题的精确解, 具有良好的可扩展性和较高的实用价值。

关键词 [生物信息学](#) [单体型检测](#) [参数化算法](#) [单核苷酸多态性](#)

分类号

Research on parameterized algorithm of Haplotypes assembly MEC problem

XIE Min-zhu^{1,2}, WANG Jian-xin², CHEN Jian-er²

1. College of Physics and Information Science, Hunan Normal University, Changsha 410081, China

2. School of Information Science and Engineering, Central South University, Changsha 410083, China

Abstract

The haplotype assembly MEC problem is the computational problem of inducing a pair of haplotypes from an individual's DNA fragments sequencing data by correcting minimum SNPs. Based on the characters of DNA fragments, the paper introduces a parameterized algorithm of time complexity $O(nk_22^{k_2+m\log m+mk_1})$ with m fragments, n SNPs, the maximum number of SNP sites that a fragment covers k_1 (usually smaller than 10) and the maximum number of the fragments covering a SNP site k_2 (usually no more than 10). For the practical fragment data, the algorithm can solve the MEC problem efficiently even if m and n are larger and it is scalable and applicable in practice.

Key words [bioinformatics](#) [haplotyping](#) [parameterized algorithm](#) [SNPs \(Single-Nucleotide Polymorphisms\)](#)

DOI:

通讯作者 谢民主 xieminzhu@sina.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(659KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“生物信息学”的相关文章](#)

▶ 本文作者相关文章

· [谢民主](#)

·

· [王建新](#)

· [陈建二](#)