

Поиск праймеров с применением алгоритма Ахо-Корасика

Научный руководитель – Губайдуллин Ирек Марсович

Давлеткулов Т.М.¹, Ахметзянова Л.У.²

1 - Уфимский государственный нефтяной технический университет, Уфа, Россия, *E-mail*: *russianlolik@gmail.com*; 2 - Институт нефтехимии и катализа РАН, Уфа, Россия, *E-mail*: *www.lianab@mail.ru*

В настоящее время молекулярная диагностика интенсивно развивается. Одним из методов диагностики является петлевая изотермическая амплификация LAMP (loop-mediated isothermal amplification) [1]. Ее применяют для детекции различных видов инфекций (в том числе и новой коронавирусной инфекции SARS-CoV-2). Для проведения успешной LAMP необходимо правильно подобрать или найти праймеры в целевом участке или полном геноме. Праймеры - это короткие фрагменты нуклеотидной последовательности, обычно от 10 до 30 нуклеотидов. Для проведения LAMP часто используют наборы из шести праймеров, и поиск этих последовательностей в геноме весьма трудоемкая задача. В данной работе поиск праймеров в геномах представлен как поиск подстроки (образца) в строке. Поиск проводился на геномах разной длины и осуществлялся с помощью алгоритма Ахо-Корасика [2], который позволяет искать сразу несколько вхождений, что немаловажно для проведения LAMP. Программа поиска праймеров реализована на языке программирования Python. Программа получает на вход 6 праймеров в виде строк (F3, F2, F1c, B1c, B2, B3) и геном в формате FASTA. Затем с помощью библиотеки BioPython происходит считывание генома. С помощью алгоритма Ахо-Корасика, библиотеки *ruahocorasick* и библиотеки *multiprocessing*, создается два отдельных процесса: на первом происходит поиск праймеров F3, F2, B1 на 5'-3' цепи, а на другом процессе поиск праймеров F1c, B2, B3 на комплементарной 3'-5' цепи. Затем, рассчитывается расстояние между праймерами. Рекомендуемое расстояние: «F3 / F2»: 0-60 нуклеотидов, «F2 / F1»: 20-40 нуклеотидов, «F1 / B1c»: 0-30 нуклеотидов. Если праймеры удовлетворяют данному условию и могут объединиться в наборы, то программа выводит индексы вхождений праймеров на выход пользователю. На рисунке 1 представлены графики зависимости времени поиска от длины нуклеотидной последовательности для последовательного поиска и многоядерного поиска. Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 20-37-90091).

Источники и литература

- 1) Notomi T., Okayama H., Masubuchi H., Yonekawa T., Watanabe K., Amino N., Hase T. Loop-mediated isothermal amplification of DNA // *Nucleic Acids Res.* 2000. V.28(12). E63. DOI: 10.1093/nar/28.12.e63.
- 2) Aho A. V., Corasick M. J. Efficient string matching: An aid to bibliographic search // *Commun. ACM.* 1975. Vol. 18 № 6. P. 333–340.

Иллюстрации

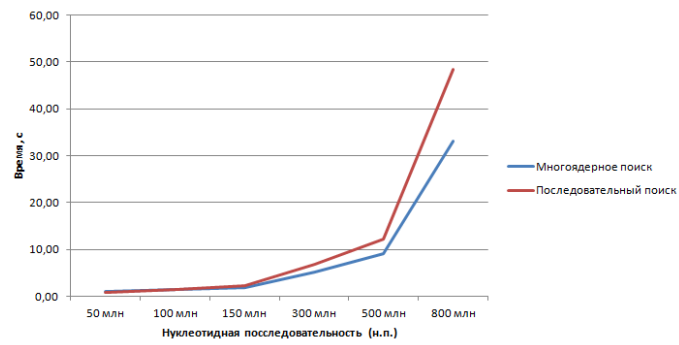


Рис. 1. Рисунок 1 – График зависимости времени от длины нуклеотидной последовательности с применением алгоритма Ахо-Корасика.