

ОЦЕНКА РАБОТЫ КЛАССИФИКАТОРОВ НА ОСНОВЕ ФИЗИЧЕСКИХ СВОЙСТВ ДНК НА ПРИМЕРЕ *E.COLI* И *B.SUBTILIS*

Рясик А.А., Орлов М.А., Зыкова Е.А., Ермак Т.В.¹, Сорокин А.А.

Институт биофизики клетки РАН, 142290, Московская область, Пущино, ул.
Институтская 3

¹Институт цитологии и генетики СО РАН, 630090, Новосибирск, пр. Академика
Лаврентьева 10

Анализ распределений физических свойств и топологических характеристик молекулы ДНК все чаще используется для поиска регуляторных участков в бактериальных геномах. Так, например, возможности применения электростатического потенциала для определения положения промоторов на хромосоме *E.coli* были продемонстрированы в работе [1]. При этом, в большинстве исследований рассмотрены распределения статических свойств молекулы ДНК, таких как электростатический потенциал, склонность ДНК к изгибам, дестабилизация двойной спирали ДНК под действием скручивания, термостабильность. Динамические характеристики молекулы используются значительно реже, а их особенности распределения в регуляторных областях мало изучены. Совместное использование статических и динамических физических и топологических свойств молекулы ДНК представляется перспективным направлением в поиске и анализе регуляторных областей прокариотических геномов, так как позволило бы более полно описать ДНК-белковое взаимодействие, учитывая сложность этого многостадийного процесса.

В данной работе исследовано поведение разработанных ранее классификационных моделей [2] на полногеномных последовательностях *E.coli* и *B.subtilis*. Указанные классификаторы, позволяющие на основании распределений статических и динамических физических свойств ДНК и нуклеотидной последовательности предсказывать один из заданных функциональных участков генома (промоторы, гены, промотороподобные участки из “промоторных островков”, участки генома без описанных функций). Обучение классификаторов проводилось на последовательностях *E.coli* с помощью двух алгоритмов: Naive Bayes и Random Forest, и показало хорошие результаты на тестовых выборках (специфичность, чувствительность и точность составляли около 90%). В представленной работе была проведена оценка предсказательной силы классификаторов применительно к *E.coli* и *B.subtilis*, ошибочные предсказания исследованы и классифицированы.

Работа поддержана грантом РФФИ №16-37-00303 мол_а.

1. Темлякова Е.А. “Роль электростатического потенциала ДНК в формировании промоторной функции в геноме *E. coli*” дисс. канд. физ.-мат. наук: 03.01.02, 12.05.2016, Пущино

2. Ryasik A.A., Orlov M.A., Zykova E.A., Ermak T.V., Sorokin A.A., 2017, Classifiers of DNA sequences based on physical characteristics profiles of DNA, FEBS Journal, p. 308.