

УДК 51-76

СРАВНЕНИЕ МЕТОДОВ ВЫВОДА ДЕМОГРАФИЧЕСКОЙ ИСТОРИИ ПОПУЛЯЦИЙ С ПОМОЩЬЮ GADMA2 И STDPOPSIM

Носкова Е.Э. (Университет ИТМО)

Научный руководитель – к.т.н., доцент ФИТиП Ульяновцев В.И.
(Университет ИТМО)

Введение. Демографическая история популяций — эволюционная история этих популяций, которая включает в себя такие параметры как время разделения, численности популяций, темпы миграции и отбора. Вывод демографической истории популяций — это процесс реконструкции этой истории на основе генетических данных. Демографическая история дает представление о силах, формирующих генетическое разнообразие, а также имеет важное значение для исследований в области биологии и медицины.

Существует множество методов и программных решений для вывода демографической истории популяций, каждый из которых использует различные математические модели и статистики данных. Например, `dadi`, `moments` и `mom2` используют статистику, основанную на частоте аллелей, но отличаются методом вычисления правдоподобия. Другим примером является программное обеспечение `momentsLD`, которое использует статистику, основанную на неравновесном сцеплении генов, для вывода демографических историй. Каждый из этих инструментов имеет свои ограничения, поэтому важно сравнивать их эффективность и точность.

Основная часть. Существующие программные решения для вывода демографических историй требуют параметризованной модели на вход и используют различные алгоритмы оптимизации для поиска параметров, которые дают максимальное значение правдоподобия. Однако наличие целого множества программных обеспечений с различными интерфейсами приводит к ошибкам в спецификации этих моделей. Кроме того, алгоритмы оптимизации иногда оказываются неэффективными. `GADMA2` решает эти проблемы, объединяя несколько существующих методов для оценки значения правдоподобия под общим интерфейсом и предоставляя эффективный алгоритм глобальной оптимизации. Это делает `GADMA2` идеальной основой для сравнения и оценки эффективности различных методов вывода демографической истории популяций.

Использование симулированных генетических данных является наиболее точным способом проведения сравнений и оценки эффективности методов. Библиотека `stdpopsim` упрощает процесс симулирования, она предоставляет каталог биологических видов и демографических сценариев, позволяя исследователям генерировать симулированные генетические данные для тестирования и проверки методов вывода демографических историй.

Выводы. Мы сравнили четыре метода вывода демографических историй популяций (`dadi`, `moments`, `mom2` и `momentsLD`), которые доступны в программном обеспечении `GADMA2`, на двух наборах данных, симулированных с помощью библиотеки `stdpopsim`. Мы проверили точность методов для различных моделей демографической истории, включая неправильно специфицированные, которые по построению не способны отразить истинную историю, используемую для моделирования данных.

Носкова Е.Э. (автор)

Ульянцев В.И. (научный руководитель)