

Немного новостей по биоинформатике за 2024

Автор

Мазяр Алексей Николаевич, группа 21213

1. Разработан питоновский пакет для филогенетического анализа - Kssdtree

а. Ссылка:

<https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btae566/7762101>

В отличие от других инструментов для поиска и анализа информации об ортологичных генах Kssdtree использует алгоритм по разбиению генома на к-меры. Это позволяет намного быстрее проводить крупные филогенетические анализы, особенно по сравнению с классическими инструментами для выравнивания (BLAST, CLUSTAL, MUSCLE). При этом используемый алгоритм более точен, чем иные подобные (Mash, Sourmash, Bindash), так как не сжимает пространство рассматриваемых к-меров, ведь не использует хэш-функции (которые из-за коллизий иногда выдают ошибочный положительный результат).

Было проведено сравнение точности и скорости алгоритма Kssdtree с упомянутыми Mash, Sourmash, Bindash на реальных и искусственных датасетах, специально отобранных для проверки таких инструментов. Kssdtree показал наибольшую точность и скорость (хотя Bindash был близко, но всё же позади).

2. Предложен инструмент для автоматизации создания предикторов биоактивности пептидов - AutoPeptideML

а. Ссылка:

<https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btae555/7760207>

Разработка прогностических моделей биологической активности пептидов это сложный и ресурсоемкий процесс. Традиционные методы создания таких моделей могут занимать много времени, а также сопровождаться утомительным подбором полезных примеров биоактивности пептидов.

AutoPeptideML решает эти проблемы, автоматизируя многие шаги этого процесса: подбор данных, их разбиение на группы для обучения и тестирования, выбор модели и подбор гиперпараметров, оценка точности модели. Также этот инструмент предлагает пользователям удобный интерфейс, что позволяет экспертам без технического опыта работать с ним.

3. Описан и протестирован метод представления связей лекарств и бактерий в человеческом организме - DHDMР

а. Ссылка:

<https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btae562/7760205>

Как известно, бактерии играют большую роль в жизнедеятельности человеческого организма, поэтому прогнозирование взаимосвязей между лекарствами и микробами очень важно для понимания влияния микрофлоры на эффективность и безопасность препаратов. В отличие от традиционных методов построения графов этих связей (GCNMDA, EGATMDA, GSAMDA и т.д.) алгоритм DHDMР подразумевает использование более далёких и многочисленных связей, а также гомогенных (лекарства с лекарствами, бактерии с бактериями) и гетерогенных (лекарства с бактериями) **гипер**графов вместе. При тестировании модели, построенной с использованием этого алгоритма, выяснилось, что в сравнении с моделями других методов она имела наибольшую точность и распознавала наибольшую часть истинно существующих взаимодействий препаратов с микрофлорой.