

Пайплайн для отбора минимального паспортизирующего набора однонуклеотидных полиморфизмов (SNP): опыт с соей культурной (*Glycine max* L.)

И.В. Стрембовский*, Д.С. Ульянов, М.В. Климушина, М.А. Самарина, А.С. Ермолаев, П.Ю. Крупин, Г.И. Карлов, М.Г. Дивашук

Всероссийский научно исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия

* ilya.strembovskiy@mail.ru

Цель: В сентябре 2025 года в РФ на юридическом уровне вводится понятие генетической паспортизации растений. Методика паспортизации основана на генетическом скрининге выборки растений сорта по набору полиморфных участков ДНК и последующем определении характерной структуры сорта по этим полиморфизмам – построении генетического паспорта сорта. Скрининг проводится с использованием специальных наборов генетических маркеров. Подобные наборы должны обладать высокой дифференцирующей способностью, то есть однозначно различать сорта и при этом иметь минимальный размер, чтобы не усложнять скрининг. Для поддержки генетической паспортизации растений в РФ мы предлагаем пайплайн, адаптированный для отбора минимального паспортизирующего набора однонуклеотидных полиморфизмов (SNP) и апробированный на коллекции сортов и линий сои культурной (*Glycine max* L.).

Материалы и методы: Предлагаемый пайплайн состоит из ряда последовательных этапов:

- 1) секвенирование коллекции образцов короткими рядами Illumina, качественная оценка ридов и их тримминг,
- 2) картирование подготовленных массивов ридов на референсную сборку *Glycine_max_v4.0* в Bowtie2,
- 3) выявление однонуклеотидных полиморфизмов в FreeBayes,
- 4) фильтрация выявленных SNP по глубине покрытия ≥ 10 , качеству ≥ 30 и отсутствию мультялельных сайтов в bcftools и отбор пригодных для целей паспортизации полиморфизмов,
- 5) фильтрация набора SNP в plink2 (LD pruning, $r^2 < 0,2$),
- 6) максимизация информативности набора через рекурсивное исключение признаков,
- 7) маскирование полученного минимизированного паспортизирующего набора SNP и его искусственное расширение за счет повторного отбора,
- 8) визуализация результатов (ggplot2, филогенетические деревья).

Оценка эффективности предлагаемого пайплайна проводилась на коллекции из 295 сортов и линий сои культурной (*Glycine max* L.).

Результаты: Полученный на первом этапе минимизированный набор из 20 SNP позволил достоверно различить все 295 сортов и линий сои, однако для 235 сортов и линий паспорта различались только на 1 SNP набора. Поэтому для повышения устойчивости системы итоговый набор был расширен до 72 SNP, покрывающих большую часть хромосом сои за исключением хромосом 5 и 10, где полиморфизмы не соответствовали критериям фильтрации. В итоговом паспортизирующем наборе каждый сорт и линия имели уникальный SNP-профиль. Для 15 сортов при построении паспорта фиксировались гетерозиготные сигналы SNP или их отсутствие, однако это не повлияло на идентификацию благодаря избыточности набора. При этом 280 сортов и линий различались больше чем по одному SNP, что подтверждает высокую дискриминационную мощность метода.

Выводы: Тестирование предлагаемого нами пайплайна на 295 сортах и линиях сои подтвердило его применимость для разработки стандартизированных SNP-наборов генетической паспортизации растений.

Финансирование: Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, государственное задание FGUM-2023-0005.