

На правах рукописи

Савиченко Виолетта Георгиевна

РАЗРАБОТКА ТЕХНОЛОГИИ МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКОЙ
ИДЕНТИФИКАЦИИ СОРТОВ СОИ (*GLYCINE MAX* (L.) MERR.) НА ОСНОВЕ
МУЛЬТИПЛЕКСНОГО МИКРОСАТЕЛЛИТНОГО АНАЛИЗА

Специальность: 4.1.2. Селекция, семеноводство и биотехнология растений

Автореферат диссертации на соискание ученой степени

Кандидата биологических наук

Москва – 2026

Работа выполнена в лаборатории молекулярно-генетических исследований отдела биологических исследований Федерального государственного бюджетного научного учреждения «Федеральный научный центр «Всероссийский научно-исследовательский институт масличных культур имени В.С. Пустовойта» (ФГБНУ ФНЦ ВНИИМК).

Научный руководитель Рамазанова Светлана Алексеевна,
кандидат биологических наук, ведущий научный
сотрудник лаборатории молекулярно-генетических
исследований отдела биологических исследований
ФГБНУ ФНЦ ВНИИМК

Официальные оппоненты: Шилов Илья Александрович,
доктор биологических наук, профессор РАН,
заведующий лабораторией анализа геномов
ФГБНУ «Всероссийский научно-
исследовательский институт сельскохозяйственной
биотехнологии»

Ущাপовский Игорь Валентинович,
кандидат биологических наук, заместитель
директора по научной работе ФГБНУ
«Федеральный научный центр лубяных культур»

Ведущая организация ФГБНУ «Северо-Кавказский федеральный
научный центр садоводства, виноградарства,
виноделия»

Защита состоится «21» декабря 2026 г. в 15:00 ч. на заседании диссертационного совета 35.2.030.08 на базе ФГБОУ ВО «Российский государственный аграрный университет–МСХА имени К.А. Тимирязева», 127434, г. Москва, ул. Прянишникова, д. 19, тел: 8 (499) 976-17-14.

Юридический адрес для отправки почтовой корреспонденции (отзывов): 127434, г. Москва, ул. Тимирязевская, д. 49.

С диссертацией можно ознакомиться в Центральной научной библиотеке имени Н.И. Железнова ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева» и на сайте Университета www.timacad.ru.

Автореферат разослан «___» _____ 20__ г.

Ученый секретарь
диссертационного совета

Е.А. Вертикова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность и степень разработанности темы. Соя (*Glycine max* (L.) Merr.) является одной из важнейших сельскохозяйственных культур. Расширение ее сортимента определяет необходимость внедрения надежного инструмента подтверждения сортовой принадлежности, контроля генетической однородности семенного материала, выявления фальсификации на рынке семян, а также защиты прав селекционеров. В настоящее время в качестве такого инструмента рассматривается молекулярно-генетическая (ДНК) паспортизация и идентификация. Актуальность ДНК-паспортизации и идентификации сортов возрастает в связи с совершенствованием нормативного регулирования семеноводства («О семеноводстве» № 454-ФЗ). Однако утвержденная методика, регламентирующая проведение ДНК-паспортизации сортов сои, на сегодняшний день отсутствует.

Целесообразность использования микросателлитных (SSR) маркеров для идентификации сортов и гибридов сельскохозяйственных культур, в том числе и сои, была показана в работах отечественных и зарубежных исследователей. Данный тип ДНК-маркеров сочетает высокий уровень полиморфизма, кодоминантный тип наследования и относительную простоту проведения анализа. При этом объединение нескольких SSR-маркеров в мультиплексные системы, обеспечивающие одновременную амплификацию их фрагментов, в сочетании с высокоразрешающими методами разделения и детекции позволяет повысить пропускную способность, точность и воспроизводимость генотипирования, что имеет особое значение при проведении массовых анализов.

Вместе с тем остаются нерешенными вопросы формирования оптимального набора SSR-маркеров, обладающего достаточной дискриминационной способностью, и стандартизации условий получения результатов генотипирования, обеспечивающих информативность и воспроизводимость ДНК-профилей. Недостаточно изучен внутрисортовой SSR-полиморфизм сои и его влияние на межсортовую дифференциацию. Кроме того, идентификация сортов предполагает сравнение их ДНК-профилей, поэтому при увеличении объема анализируемых данных возникает необходимость автоматизации этой процедуры. Перспективным для этих целей является применение методов машинного обучения. Решение указанных вопросов требует комплексного подхода к молекулярно-генетической идентификации сортов данной культуры.

Цель исследования – изучение внутри- и межсортового SSR-полиморфизма сои для молекулярно-генетической идентификации сортов с использованием машинного обучения.

Задачи исследования:

1. Разработать мультиплексные системы SSR-маркеров, оптимизировать и стандартизировать условия их применения для получения воспроизводимых результатов генотипирования сои.
2. Изучить внутрисортную генетическую структуру сортов сои на основе ДНК-профилей, сформированных с помощью мультиплексных систем SSR-маркеров.
3. Оценить межсортную дифференциацию сои с учетом внутрисортного генетического SSR-полиморфизма.
4. Изучить возможность применения машинного обучения для сопоставления ДНК-профилей сортов сои и частичной автоматизации процесса их идентификации.
5. Оценить эффективность использования мультиплексных систем SSR-маркеров для контроля прохождения гибридизации сои.

Научная новизна исследования. Впервые при изучении внутрисортной генетической структуры исследованного набора сортов сои установлено, что сорта представлены различным числом биотипов с неодинаковой частотой встречаемости, выявлены специфичные и общие биотипы для разных сортов сои.

Впервые установлено, что межсортная дифференциация сои определяется биотипным составом выборок индивидуальных растений, при этом наличие общих биотипов ограничивает идентификацию сорта по единичному растению, а специфичные биотипы имеют дифференцирующее значение.

Впервые показана возможность применения алгоритма машинного обучения Random Forest для частичной автоматизации сопоставления ДНК-профилей при молекулярно-генетической идентификации сортов сои и установлено, что результат идентификации зависит от представленности биотипов в обучающей выборке.

Теоретическая и практическая значимость. Результаты исследования вносят вклад в понимание внутрисортной генетической структуры сои и ее влияния на дифференциацию сортов. Полученные данные расширяют представления о сорте сои как объекте молекулярно-генетической идентификации и формируют теоретическую основу его характеристики по совокупности ДНК-профилей индивидуальных растений.

Разработанные мультиплексные системы в оптимизированных условиях позволяют проводить воспроизводимый анализ 11 SSR-маркеров в каждой, сокращая продолжительность исследования по сравнению с моноплексным. Получаемые ДНК-профили могут использоваться для идентификации сортов и

гибридов сои, выявления и отбора биотипов, а в сочетании с моделью машинного обучения для частично автоматизированного определения сортовой принадлежности семенного материала. Результаты исследования также могут применяться при формировании генетических паспортов новых сортов сои и в качестве дополнительного инструмента для оценки их отличимости, однородности и стабильности.

Методология и методы исследования. Теоретическая часть исследования основана на систематическом анализе отечественной и зарубежной литературы. Экспериментальная часть основана на применении методов молекулярно-генетического, биоинформатического и статистического анализов, подробно изложенных в главе «Материалы и методы исследования» диссертации.

Основные положения, выносимые на защиту:

1. Разработанные мультиплексные системы SSR-маркеров обеспечивают получение воспроизводимых ДНК-профилей растений сои.
2. Сорт сои представляет собой совокупность различающихся SSR-профилями биотипов.
3. Межсортная дифференциация сои определяется составом и частотой встречаемости биотипов.
4. Алгоритм Random Forest обеспечивает частичную автоматизацию молекулярно-генетической идентификации сортов сои.

Степень достоверности и апробации результатов исследования. Достоверность результатов подтверждается достаточным объемом экспериментальных данных, полученных в ходе лабораторных исследований при использовании современных подходов молекулярно-генетического анализа. Обработка результатов проводилась с использованием программных средств (MS Excel и языка программирования R). Основные результаты диссертационной работы представлены на 9 конференциях всероссийского и международного уровней: «Актуальные вопросы биологии, селекции, технологии возделывания и переработки сельскохозяйственных культур» (Краснодар, 2021); «Научное обеспечение устойчивого развития агропромышленного комплекса в условиях аридизации климата», посвященной 35-летию ФГБНУ РосНИИСК «Россорго» (Саратов, 2021); «Актуальные вопросы биологии, селекции, технологии и переработки сельскохозяйственных культур, экологии и экономики в сельском хозяйстве» (Краснодар, 2022; 2023; 2025); «Генетические ресурсы растений для генетических технологий: к 100-летию Пушкинских лабораторий ВИР» (Санкт-Петербург, 2022); «Совершенствование способов управления технологическими процессами в садоводстве и виноградарстве» (Краснодар, 2023); «Рядом с Н.И. Вавиловым –

научные школы России по обеспечению продовольственной и экологической безопасности страны» (Москва, 2024); «Современное состояние, проблемы и перспективы развития аграрной науки» (Симферополь, 2026).

Публикации. По материалам диссертации опубликовано 17 работ, в том числе 5 – статьи в рецензируемых научных изданиях, рекомендованных ВАК Минобрнауки РФ, патент, две базы данных, методические рекомендации.

Личный вклад автора. Автором самостоятельно осуществлены обзор литературных источников, оптимизация методических условий микросателлитного анализа и разработка мультиплексных систем SSR-маркеров. Массовые анализы по генотипированию выполнены совместно с сотрудниками лаборатории по разработанной автором технологии. Обработка и интерпретация результатов проведены при непосредственном участии автора.

Структура диссертации. Диссертация изложена на 179 страницах, состоит из введения, основной части, содержащей 24 рисунка, 12 таблиц, заключения, предложений для селекционной практики, списка сокращений, списка литературы, включающей 211 наименований, в том числе 78 – на иностранном языке, и 5 приложений.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1 Обзор литературы

В главе приведен структурированный анализ отечественной и зарубежной научной литературы, посвященной вопросам сортовой идентификации сои. Рассмотрены ограничения традиционных методов идентификации, приведено обоснование выбора ДНК-маркеров для исследования, проанализированы методические аспекты SSR-анализа. Обобщены результаты ДНК-идентификации генотипов сои и других культур с использованием SSR-маркеров, а также современное состояние и проблемы внедрения ДНК-паспортизации в РФ.

Глава 2 Материалы и методы исследования

Всего в исследовании 85 сортов сои: 73 – из РФ, 12 – из Канады, США, Украины и Беларуси. Состав и объем исследуемого материала определялись задачами исследования. При оценке эффективности выделения ДНК использовали 5 типов растительного материала (части зародышей семян, корни, гипокотили и семядоли проростков, листья) сорта Пума по 12 индивидуальных растений каждого типа. Материалом для первичной оценки SSR-маркеров служили смеси по 10–15 растений каждого из 20 сортов; для разработки мультиплексных систем – по одному индивидуальному растению 15 сортов. Положительный контрольный образец (ПКО) для сопоставления результатов

фрагментного анализа в разных сериях – сорт сои Лира (ФГБНУ ФНЦ ВНИИМК).

Молекулярно-генетическую идентификацию по разработанной технологии проводили для 60 сортов сои, представленных выборками индивидуальных растений. В обучающую выборку модели Random Forest вошли 53 сорта, в т.ч. 37 сортов селекции ФГБНУ ФНЦ ВНИИМК (по 47 семян категории ОС) и 16 сортов других учреждений (по 23–30 семян неизвестной категории). Независимая тестовая выборка включала 7 сортов, отсутствующих в обучении (по 8 семян), а также сорта, участвующие в обучении, но представленные другими партиями.

При идентификации гибридов исследовали 16 предполагаемых гибридных растений от 8 комбинаций скрещиваний, а также их родительские растения.

Выделение ДНК осуществляли преимущественно на автоматической станции Auto-Pure 96 (Allsheng, КНР) набором реагентов МагноПрайм ФИТО (ООО «НекстБио», РФ) из частей зародышей или листьев. Для оценки выделенной ДНК использовали спектрофотометрический и флуориметрический методы, а также электрофорез в 1%-ном агарозном геле.

Разработка мультиплексных систем SSR-маркеров включала: 1) Поиск SSR-локусов в базе данных SoyBase и публикациях с последующей *in silico* проверкой праймеров на эталонной сборке генома сои Glycine_max_v4.0 в Primer-BLAST. 2) Оценка информативности (N_a , N_e , PIC) и специфичности. 3) Разработка новых пар праймеров в Primer-BLAST на основе последовательностей ДНК эталонной сборки генома сои с учетом размера ПЦР-продукта, температуры плавления, отсутствия димеров и их экспериментальная проверка. 4) Формирование мультиплексных наборов SSR-маркеров. 5) Оптимизация условий амплификации (подбор температурно-временных режимов, концентрации праймеров, флуорофоров) и интерпретации результатов фрагментного анализа (секвенирование по Сэнгеру фрагментов ДНК ПКО, разработка панелей аллельных бинов в программе GeneMarker).

При оптимизации условий мультиплексных ПЦР сравнивали два температурно-временных режима: протокол А – 3-стадийная ПЦР со ступенчатым снижением температуры отжига в течение первых 10 циклов (touchdown); протокол Б – 2-шаговая ПЦР с отжигом и элонгацией на одной стадии (подробные условия описаны в главе «Материалы и методы исследования» диссертации). Оптимальный протокол выбирался по результатам сравнительного анализа электрофореграмм. Концентрации праймеров подбирали путем варьирования объема исходного раствора праймера (100 пкМ/мкл) при внесении, оценивая относительную интенсивность

сигнала. Итоговые концентрации устанавливали после распределения флуорофоров между SSR-маркерами.

В качестве основного в исследовании применялся метод капиллярного электрофореза (Нанофор-05, ИАП РАН, РФ), также в работе использовали: электрофорез в 2%-ном агарозном геле, в 8%-ном полиакриламидном геле (ПААГ) в нативных условиях и в 8%-ном ПААГ в денатурирующих условиях.

Генетическое разнообразие сортов оценивали по доли полиморфных локусов (P), количеству выявленных и эффективных аллелей (N_a и N_e), наблюдаемой и ожидаемой гетерозиготности (H_o и H_e), рассчитанными в GenALEx. Растения с идентичными по 22 SSR-маркерам ДНК-профилями относили к одному биотипу, при наличии хотя бы одного отличающегося аллеля выделяли новый биотип. Внутрисортную структуру характеризовали числом и частотой встречаемости биотипов. Межсортную дифференцию оценивали на основе данных генетических дистанций Нея с последующей кластеризацией сортов методом UPGMA в программной среде R.

Классификация ДНК-профилей растений проводилась по алгоритму машинного обучения Random Forest (далее RF) (Breiman, 2001), реализованного в пакете «randomForest». При обучении модели RF использовали параметры: число деревьев в ансамбле (Ntree) – 2000, число случайно отбираемых признаков для разделения каждого узла дерева (Mtry) – 9. Класс растения определяли по большинству голосов деревьев в ансамбле. Для исключения отнесения растений с низкой оценкой предсказаний к сортам обучающей выборки установлен порог 0,7.

Глава 3 Экспериментальное обоснование методических условий SSR-анализа генотипов сои

Для ДНК-идентификации сортов сои необходимы методы, обеспечивающие точность, воспроизводимость и достаточную пропускную способность анализа. В связи с этим осуществлен выбор способов выделения ДНК и детекции ПЦР-продуктов, как этапов SSR-анализа, влияющих на стабильность получаемых ДНК-профилей.

3.1 Сравнительная оценка эффективности выделения ДНК из разных частей растений сои

По результатам спектрофотометрических измерений наибольшая концентрация нуклеиновых кислот получена при выделении из частей зародыша – 160,84–265,83 нг/мкл. Значения соотношения A260/A280 для большинства образцов находились в пределах 1,8–2,2, за исключением некоторых образцов, выделенных из корней и гипокотилей. Наибольшие значения A260/A230 получены в образцах, выделенных из зародышей (от 1,47 до 1,72), самыми загрязненными оказались образцы, полученные из корней

(0,31–0,72) и гипокотилей (0,41–0,61). Флуориметрическая оценка содержания дцДНК и анализ в 1%-ном агарозном геле показали наличие фрагментированной ДНК во всех образцах, а образцы из зародышей и листьев содержали также высокомолекулярную ДНК. Результаты ПЦР показали, что имеющиеся примеси (на основании A260/A230) и деградированная ДНК, не оказали негативного влияния. Таким образом, ДНК, выделенная из любой части растения сои, пригодна для ПЦР. В качестве основного материала для экстракции выбраны части зародышей семян, это позволяет получить препарат ДНК, пригодный для SSR-анализа, без этапа проращивания семян.

3.2 Сравнительная оценка способов разделения и детекции фрагментов ДНК SSR-маркеров

Проведены ПЦР, одни и те же продукты которых были разделены различными способами. При разделении в агарозном геле и ПААГ нативных условиях близкие по размеру фрагменты в большинстве случаев не дифференцировались. Денатурирующий ПААГ показал более высокую разрешающую способность – дифференцированы фрагменты с разницей в 3 и более п.н. При этом определение размеров фрагментов проводили только по результатам капиллярного электрофореза, поскольку при других методах на точность определения размеров влияло искривление электрофоретического фронта и другие факторы. Фрагменты, отличающиеся на 2 п.н. (Soyhsp176), удалось дифференцировать только с помощью капиллярного электрофореза. Для ДНК-идентификации генотипов сои был выбран метод капиллярного электрофореза, обеспечивающий наиболее высокую разрешающую способность, точность и воспроизводимость результатов при высокой пропускной способности анализа.

Глава 4 Подбор информативных SSR-маркеров для ДНК-идентификации генотипов сои

4.1 Поиск SSR-локусов сои и *in silico* проверка специфичности праймеров

На первом этапе отобрано 138 ди-, три- и тетрануклеотидных SSR-локусов с 20 хромосом генома сои, включая часть ранее апробированных во ВНИИМК (Satt141, Satt681, Satt181, Soyprp1 и Soyhsp176). В результате *in silico* проверки праймеров на эталонной сборке генома сои из 138 отобрано 47, обеспечивающих специфичное связывание с ДНК-матрицей, с температурой плавления от 50 до 65 °С и ожидаемыми фрагментами 100–600 п.н.

4.2 Оптимизация температуры отжига праймеров

Экспериментально определены температуры отжига праймеров, варьирующие от 47 до 64 °С. Для объединения маркеров в мультиплексной ПЦР помимо прочего необходимо учитывать, что праймеры должны

гибридизоваться при одинаковой или близкой температуре для сохранения специфичности амплификации.

4.3 Оценка информативности и специфичности SSR-маркеров

Оценку SSR-маркеров проводили по результатам разделения продуктов в денатурирующем ПААГ. Satt631, Sat374, Satt311, Satt398, Satt257 и Satt713 исключены из-за выявления неспецифичных продуктов или нулевых аллелей. Для остальных рассчитывали показатели информативности (таблица 1).

Таблица 1 – Основные показатели информативности SSR-маркеров

№	SSR-маркер	N _a	N _e	PIC	№	SSR-маркер	N _a	N _e	PIC
1	Satt532	3	2,27	0,56	22	Satt333	4	2,17	0,49
2	Sat413	3	2,13	0,53	23	Satt500	3	2,33	0,48
3	Satt307	3	2,27	0,56	24	Satt633	2	2,00	0,38
4	Satt286	4	2,13	0,53	25	Satt638	2	1,35	0,22
5	Satt149	3	1,89	0,47	26	Satt359	3	2,17	0,47
6	Satt309	3	1,72	0,42	27	Satt635	4	3,23	0,63
7	Sat420	3	1,69	0,41	28	Satt353	3	2,22	0,44
8	Sct189	4	3,23	0,69	29	Satt304	2	1,47	0,27
9	Satt141	6	2,75	0,64	30	Satt651	4	2,90	0,66
10	Satt681	6	4,76	0,79	31	Satt263	3	2,78	0,56
11	Soyprp1	3	2,55	0,61	32	Satt287	4	2,50	0,55
12	Soyhsp176	2	1,22	0,18	33	Sat222	3	2,50	0,53
13	Satt181	3	2,80	0,64	34	Satt292	4	4,00	0,70
14	Satt168	5	4,00	0,71	35	Satt614	5	4,17	0,72
15	Satt549	3	2,38	0,49	36	Satt507	4	3,17	0,69
16	AW277661	4	2,63	0,54	37	Satt604	1	1,00	0,00
17	Satt684	4	2,78	0,58	38	Satt506	1	1,00	0,00
18	Satt471	2	1,72	0,33	39	Satt197	4	2,13	0,53
19	CSSR531	4	2,40	0,56	40	Satt442	4	2,61	0,62
20	Satt636	4	2,45	0,50	41	Satt596	7	5,63	0,82
21	Satt207	4	2,86	0,56	-	-	-	-	-
Среднее по 41 маркеру							3,49	2,54	0,51

Количество выявленных аллелей (N_a) варьировало от 1 (Satt604, Satt506) до 7 (Satt596), эффективных (N_e) – от 1,00 до 5,63, PIC – от 0 до 0,82. 25 SSR-маркеров характеризуются как высокоинформативные (PIC > 0,5), 13 – информативные (0,25 < PIC < 0,5). Исключены SSR-маркеры, по которым выявлено 2 и менее аллелей. Для всех остальных с помощью капиллярного электрофореза уточнены размеры фрагментов и аллельный состав. Количество выявленных фрагментов для большинства маркеров соответствовало полученным при разделении в ПААГ. Однако для части SSR-маркеров по результатам капиллярного электрофореза выявлено меньше аллельных вариантов, поэтому они были исключены. Также исключены динуклеотидные Sct189, Sat413 и Sat420 из-за наличия статтеров с более высокой интенсивностью сигнала как менее пригодные для надежной интерпретации.

Глава 5 Разработка мультиплексных систем SSR-маркеров для ДНК-идентификации генотипов сои

5.1 Разработка праймеров для формирования мультиплексных систем

Для решения проблем несовместимости при формировании мультиплексных систем к 19 SSR-локусам (Satt141, Satt549, AW277661, Satt681, Satt286, CSSR531, Soyprp1, Satt197, Satt359, Satt353, Satt181, Satt442, Satt149, Satt263, Satt287, Satt596, Satt309, Satt292, Satt614) сконструированы новые последовательности праймеров (Савиченко, 2026). Проверка разработанных праймеров по результатам амплификации в моноплексных ПЦР в идентичных условиях на выборке индивидуальных растений сои показала их высокую специфичность – получены целевые фрагменты ДНК, размеры которых согласуются с рассчитанным в Primer-BLAST по эталонной сборке генома сои. Установлены диапазоны размеров амплифицированных фрагментов для каждого маркера для их учета при формировании мультиплексных наборов.

5.2 Формирование мультиплексных систем SSR-маркеров и оптимизация условий их применения

В результате работы по формированию мультиплексных наборов было создано две системы для совместной амплификации фрагментов 11 SSR-маркеров в каждой: Satt507, Satt681, Satt149, Satt532, Satt141, Satt614, Satt442, Satt263, Satt309, Satt359, Satt286 в первой и CSSR531, Satt596, Satt287, Satt197, AW277661, Satt181, Satt168, Satt549, Satt353, Satt292, Soyprp1 во второй.

При оптимизации температурно-временных условий было выявлено, что проведение мультиплексных ПЦР по протоколу *Б* (2-шаговой ПЦР) обеспечивает стабильную амплификацию только целевых фрагментов ДНК по всем SSR-маркерам обеих мультиплексных систем по сравнению с протоколом *А*, где наблюдалось наличие неспецифичных фрагментов, а в мультиплексной системе 2 – амплификация была не стабильна для части маркеров (рисунок 1).

Для отдельных SSR-маркеров сохранялась относительно низкая интенсивность сигнала, что особенно заметно на каналах детекции TAMRA и ROX, на данном этапе все праймеры добавляли в одинаковой концентрации (по 4 пкМ/мкл). Кроме того, существовала потенциальная проблема перекрытия фрагментов ДНК разных SSR-маркеров, имеющих близкие диапазоны размеров (Satt681 и Satt359, Satt596 и Satt549). Для выравнивания интенсивности сигналов и обеспечения надежного разделения фрагментов проведено перераспределение флуорофоров между SSR-маркерами с учетом их флуоресценции и итоговый подбор концентраций праймеров (таблица 2).

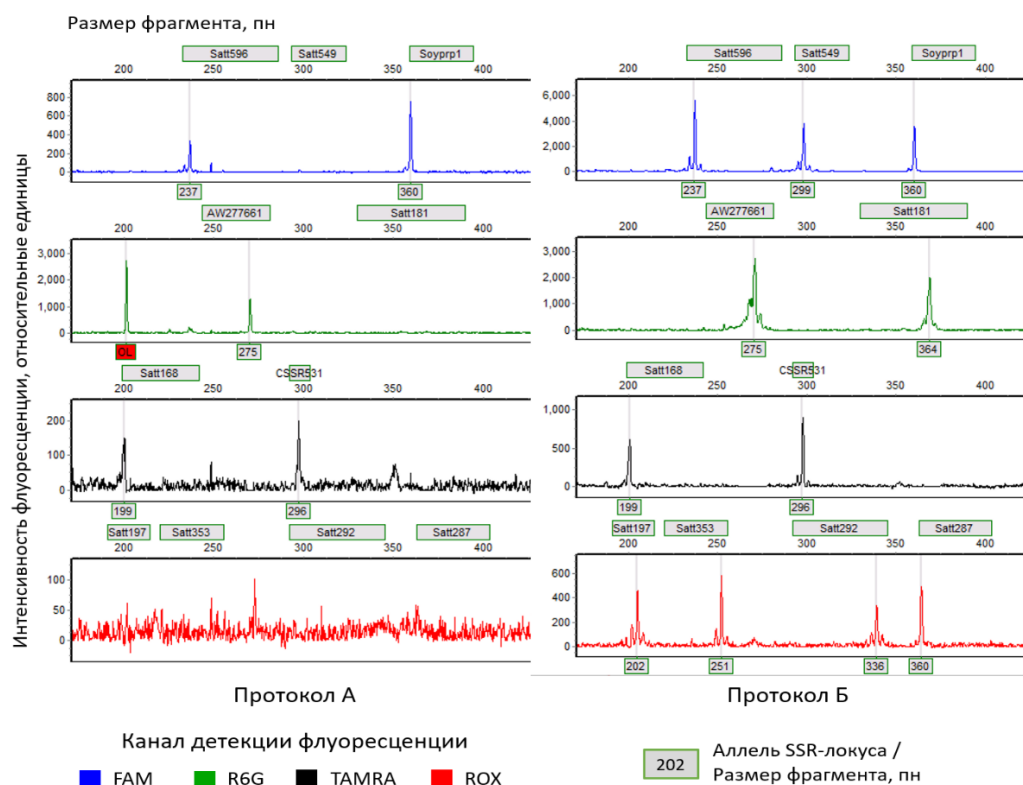


Рисунок 1 – Мультиплексная система 2 при разных протоколах амплификации, ПКО

Таблица 2 – Характеристика разработанных мультиплексных систем SSR-маркеров сои

Флуорофор ¹	SSR-маркер	Концентрация праймеров, пкМ/мкл ¹	Количество выявленных фрагментов ДНК ²	Диапазон размеров фрагментов ДНК ²	Мотив ³	Размер фрагмента ПКО, п.н. ³
Мультиплексная система 1						
FAM	Satt507	0,75	5	211–223	(ATT)22	217
	Satt681	1,75	8	330–351	(TTA)20	333
	Satt149	0,50	4	398–425	(AAT)11	407
R6G	Satt532	0,75	5	162–174	(AAT)14	165
	Satt141	0,75	6	277–331	(ATA)21	319
	Satt614	4,50	17	409–547	(TTA)25	427
TAMRA	Satt442	2,25	10	174–210	(TAA)30	192
	Satt263	4,00	8	308–353	(TTA)22	338
ROX	Satt309	2,00	3	255–276	(ATT)8	261
	Satt359	4,00	5	348–378	(TAT)15	348
	Satt286	4,75	5	451–469	(ATT)18	466
Мультиплексная система 2						
FAM	CSSR531	0,25	4	113–123	(TAA)12	117
	Satt596	0,75	14	240–288	(AAT)13	240
	Satt287	0,75	6	360–399	(TTA)8	360
R6G	Satt197	0,75	9	151–211	(ATT)24	202
	AW277661	0,50	6	257–284	(TAT)23	275
	Satt181	0,75	8	331–385	(ATT)16	364

TAMRA	Satt168	2,75	5	199–232	(ATA)8	199
	Satt549	2,75	6	296–320	(TAT)16	299
ROX	Satt353	1,75	5	218–251	(TTA)17	248
	Satt292	3,00	8	294–339	(ATA)21	336
	Soyprp1	3,25	5	360–393	(TAT)11	360

Примечание: ¹ – по результатам: оптимизации условий мультиплексных ПЦР; ² – генотипирования, описанным в главе 6; ³ – секвенирования ПКО

5.3 Анализ нуклеотидных последовательностей фрагментов ДНК

Проведено секвенирование фрагментов ДНК ПКО с последующим выравниванием нуклеотидных последовательностей на эталонную сборку генома сои. Для всех изученных последовательностей подтверждено соответствие целевым SSR-локусам. Установлено, что различия между аллелями обусловлены вариабельностью количества повторяющихся мотивов при сохранении консервативности фланкирующих областей. На основании результатов секвенирования для ПКО установлены количество повторов и размеры фрагментов ДНК (таблица 2).

5.4 Формирование панелей аллелей SSR-маркеров

Для воспроизводимого определения аллелей в программе GeneMarker сформированы панели аллелей с указанием их размеров и диапазонов обнаружения (Свидетельство № 2026623222). При интерпретации результатов фрагментного анализа положение панелей контролируется относительно ПКО и установленным для него значениям (таблица 2), при необходимости корректируется. Это обеспечивает сопоставимость результатов между разными сериями анализов.

Разработанные мультиплексные системы в совокупности со всеми подобранными и оптимизированными условиями анализа ДНК, представляют собой технологию молекулярно-генетической идентификации сортов сои.

Глава 6 Молекулярно-генетическая идентификация генотипов сои

6.1 Формирование ДНК-профилей сортов сои с помощью разработанных мультиплексных систем SSR-маркеров

С помощью разработанной технологии проведено генотипирование индивидуальных растений 60 сортов сои, выявлен внутрисортной полиморфизм. Учитывая распространенный в литературе подход к анализу смешанных образцов ДНК, оценена полнота представления аллельного состава сорта при использовании данного подхода. Для этого был исследован смешанный образец ДНК сорта Дуар, сформированный из 47 проб ДНК, ранее проанализированных индивидуально. Часть аллелей, выявленных при индивидуальном анализе, в смешанном образце не детектировалась или не могла быть однозначно интерпретирована. Наложение сигналов близких по размеру фрагментов и различия в их интенсивности, обусловленные

неодинаковым вкладом отдельных аллелей в суммарный сигнал, затрудняли разграничение целевых и артефактных пиков. На рисунке 2 пример интерпретации фрагментов одного из маркеров.

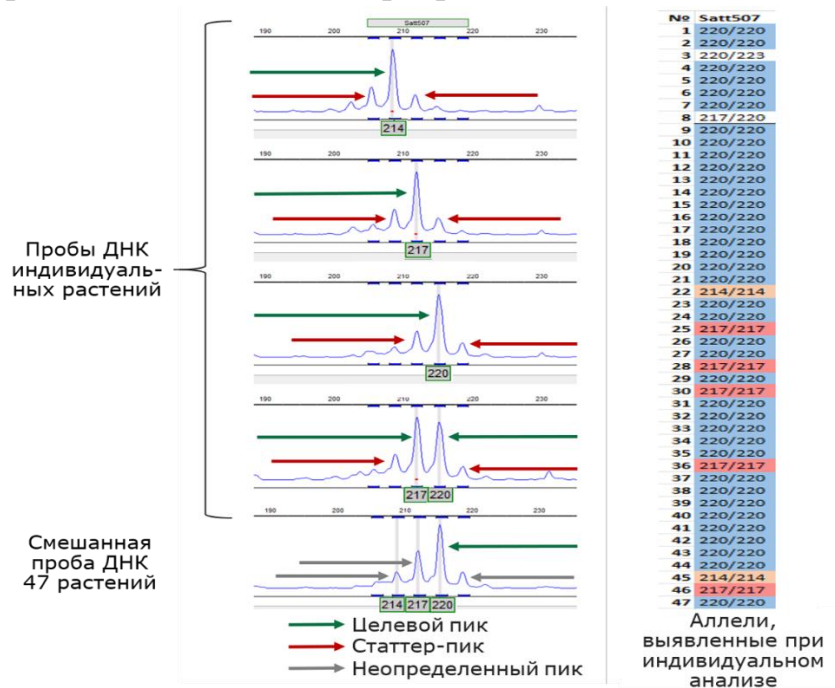


Рисунок 2 – Сравнение результатов фрагментного анализа по SSR-маркеру Satt507, сорт Дуар

Результаты свидетельствуют об ограниченной информативности смешанного ДНК-профиля вследствие неполноты аллельного состава и невозможности оценки генетической структуры сорта. Для идентификации сортов целесообразно анализировать выборку индивидуальных растений и формировать на ее основе полный ДНК-профиль.

6.2 Анализ генетической структуры и дифференциации сортов сои

Установлены различия по уровню внутрисортного генетического разнообразия. Наименьшим доля полиморфных локусов (4,6 %) выявлена в сортах Бинго, Лира и Юг-30, наибольшая (100 %) – сорт Дуар. N_a варьировало от 1,045 до 2,909, N_e – от 1,004 до 1,934, со средними значениями 1,625 и 1,155 соответственно. Несмотря на относительно высокую долю полиморфных локусов, значения H_o оставались низкими и в среднем составляли 0,008, что обусловлено биологическими особенностями сои как самоопыляющейся культуры и преобладанием в сортах гомозиготных генотипов. При этом среднее значение H_e составило 0,091.

В анализируемых сортах выявлено от 2 до 34 биотипов (рисунок 3), при этом количество обнаруженных биотипов следует рассматривать с учетом различного объема выборок индивидуальных растений.

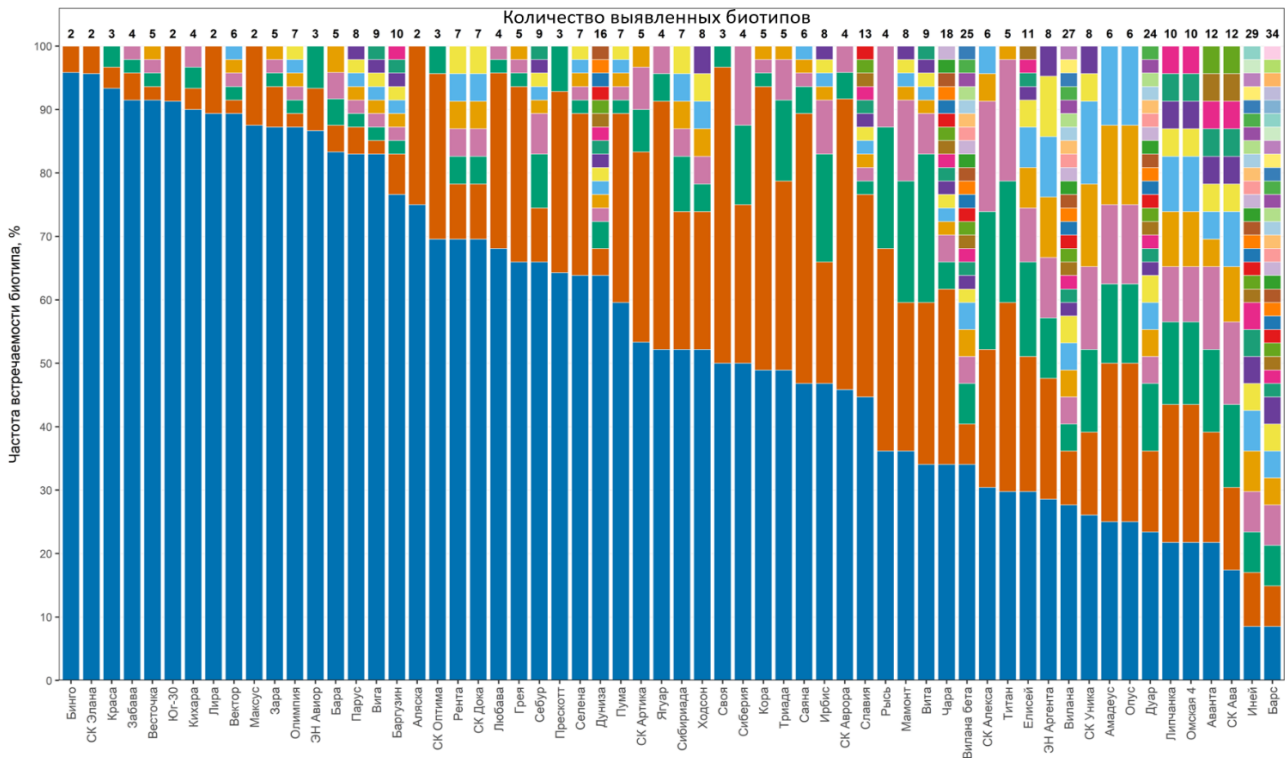


Рисунок 3 – Биотипный состав сортов сои по данным SSR-анализа

У части сортов наблюдалось выраженное преобладание одного биотипа, по два основных биотипа обнаружено в сортах Своя (50 и 47 %), Кора (49 и 45 %), Саяна (47 и 43 %) и СК Аврора (по 46 %), при наличии других малочисленных. Наиболее не однородными являлись сорта Барс (34 биотипа), Иней (29), Вилана (27), Вилана бета (25) и Дуар (24).

У ряда сортов выявлены идентичные биотипы. Так, Зара и Себур имели идентичный биотип, встречающийся с частотой 87,2 и 66 % соответственно. Выявлены сорта, имеющие несколько общих биотипов: Краса и Забава, Баргузин и Саяна по 2 биотипа, Парус и Славия – 3, с разной частотой встречаемости. Для дифференциации таких сортов необходимо учитывать как частоту встречаемости общих, так и наличие специфичных биотипов. Важно отметить, что идентичные биотипы со средней и высокой частотой встречаемости обнаружены у сортов селекции одного учреждения, что может быть связано с их общим происхождением. Единичные совпадения биотипов могут быть обусловлены механическим загрязнением семенного материала.

Кластеризация по UPGMA позволила дифференцировать большинство анализируемых сортов и выявить группы генетически близких (рисунок 4). Сопоставление дендрограммы с индивидуальными ДНК-профилями показало, что близость определялась наиболее распространенными аллельными вариантами. Сорта, не дифференцированные при кластеризации (Зара и Себур, Парус и Славия) различаются по составу малочисленных биотипов. Это

подтверждает необходимость формирования полного ДНК-профиля на основании анализа выборки индивидуальных растений, позволяющего учитывать его внутрисортную структуру и аллельное разнообразие.



Рисунок 4 – Дендрограмма генетических различий сортов сои с графическим представлением ДНК-профилей индивидуальных растений

6.3 Идентификация сортов сои по ДНК-профилям с помощью алгоритма классификации Random Forest

В настоящей работе для оценки возможности автоматической классификации индивидуальных растений по их сортовой принадлежности применен алгоритм Random Forest (RF). Для анализа распределения правильных и ошибочных классификаций модель RF была применена к обезличенным ДНК-профилям растений обучающей выборки. Общая точность классификации составила 92 %, коэффициент Каппа – 0,918, ошибка Out-of-bag – 7,98 %, что свидетельствует о высокой классификационной способности модели. Несмотря на генетическую неоднородность изученных сортов, большинство растений классифицированы в соответствии со своей сортовой принадлежностью. Ошибочная классификация наблюдалась в основном для сортов, имеющих общие биотипы (Саяна – Баргузин, Зара – Себур, Парус – Славия). Биотипы, специфичные для сортов, классифицировались верно, общие – относились к сортам, в которых имеют большую распространенность (рисунок 5). Единичные растения, отнесенные к другим сортам с высокой вероятностью, имеют идентичные или близкие по составу SSR-профили, с

низкой вероятностью – отсутствующие или мало представленные в обучающей выборке биотипы.

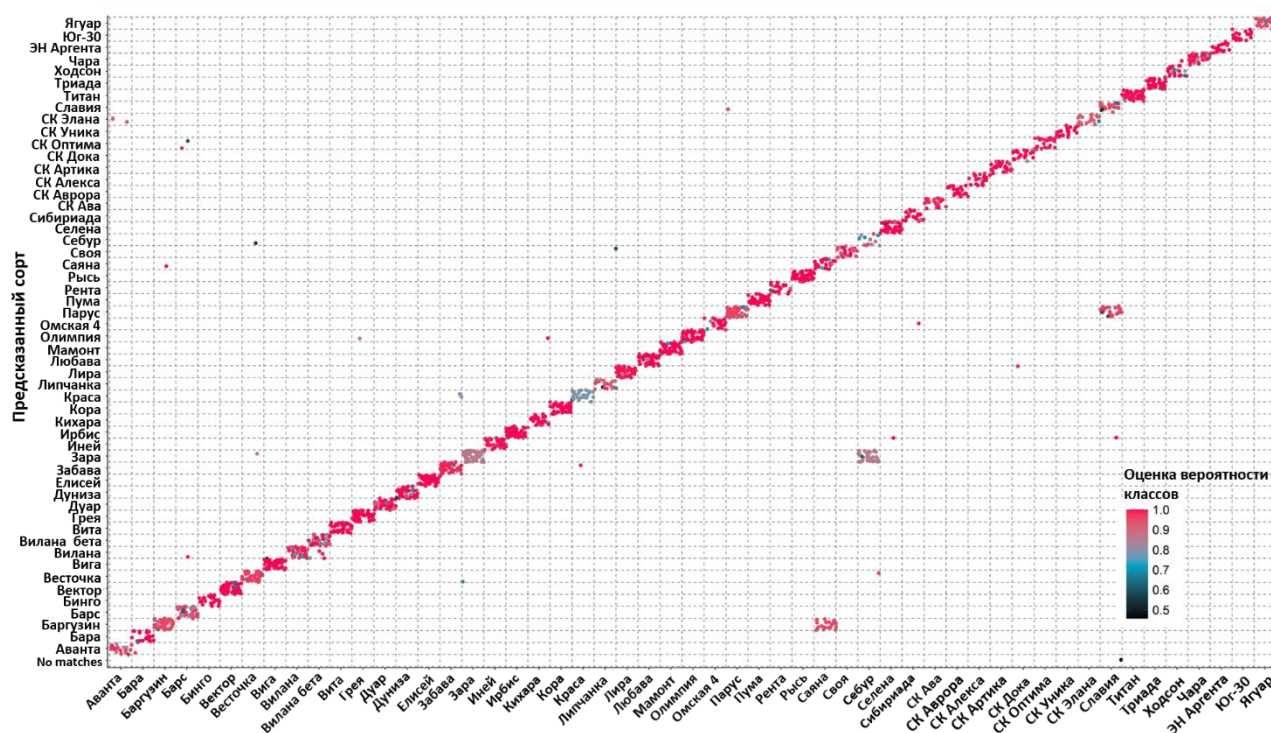


Рисунок 5 – Классификация растений сои из обучающей выборки моделью RF

Таким образом, классификация отдельного растения не всегда позволяет однозначно установить сортовую принадлежность, необходимо учитывать структуру выборки, частоты и наличие общих и специфичных биотипов.

Далее проведена независимая проверка классификационной способности модели RF на тестовой выборке, включавшей растения сортов, представленных в обучающей выборке другими партиями семян, а также сортов, отсутствующих при обучении. То есть оценена возможность идентификации неизвестного материала по ДНК-профилям с использованием обученной модели RF (рисунок 6).

Большинство растений сортов, представленных в обучающей выборке, классифицированы верно. Ошибки наиболее выражены для растений сортов Барс, Вилана бета и Вилана. Это объясняется наличием в анализируемых партиях биотипов мало представленных или отсутствующих в обучающей выборке. Для растений сортов, не включенных в обучающую выборку, ошибочного отнесения к известным модели сортам не выявлено.

Независимая проверка показала возможность использования модели RF для сопоставления индивидуальных ДНК-профилей с базой данных сортов. Точность классификации зависела от представленности биотипов в обучающей выборке и наличия идентичных у разных сортов. Поэтому идентификацию

сорта необходимо проводить по совокупности результатов анализа выборки индивидуальных растений с учетом внутрисортовой генетической структуры.

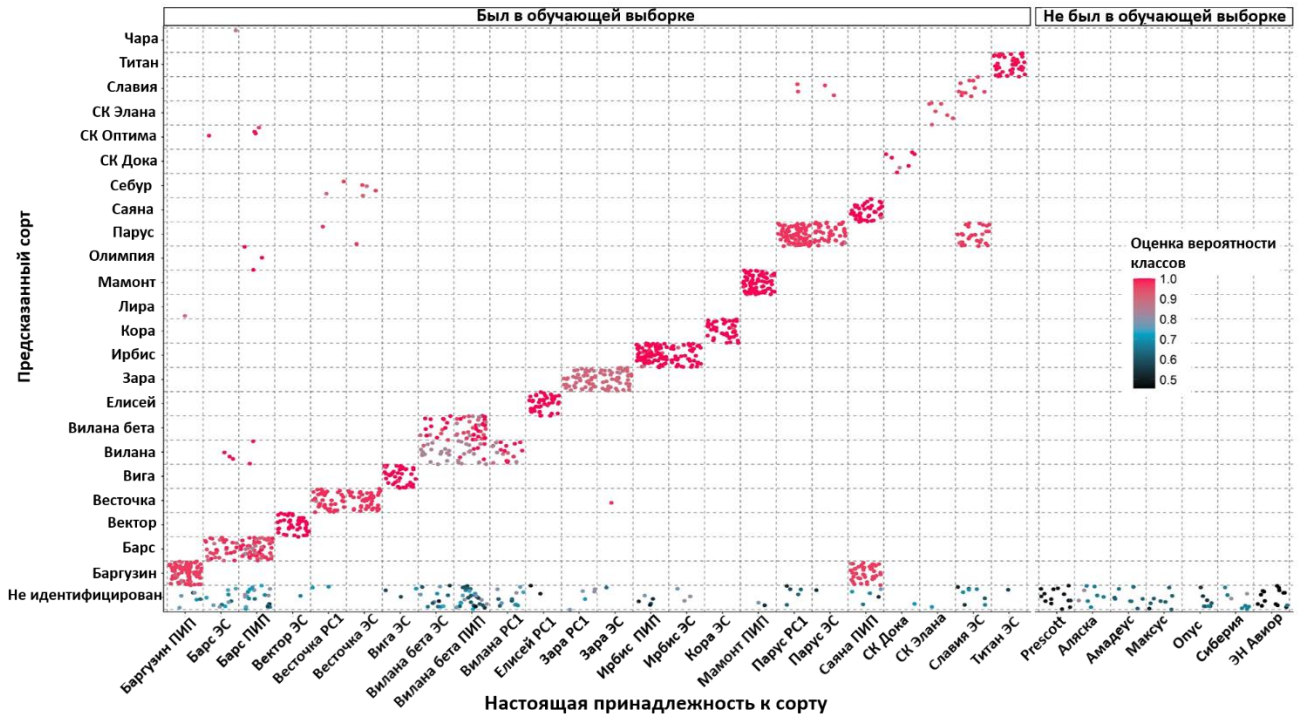


Рисунок 6 – Определение сортовой принадлежности с помощью технологии молекулярно-генетической идентификации сои и модели RF

6.4 Идентификация гибридов F_1 сои с помощью разработанных мультиплексных систем SSR-маркеров

Гибридное происхождение анализируемых растений устанавливалось по наличию в их SSR-профилях аллелей обеих родительских форм и отсутствию аллелей, несоответствующих родителям. Сравнение профилей предполагаемых гибридных растений и родителей подтвердило гибридное происхождение 10 из 16 проанализированных растений. Среди остальных обнаружены 3 растения, сформировавшиеся в результате самоопыления, что подтверждалось наличием только материнских аллелей при отсутствии отцовских. Выявлено два растения, не принадлежащие анализируемым комбинациям – в их SSR-профилях обнаружены аллели, отсутствующие у обеих родительских форм.

Для каждой гибридной комбинации был характерен индивидуальный набор информативных SSR-маркеров, поэтому для идентификации гибридов целесообразно применять полный набор маркеров разработанных систем. Кроме того, внутрисортовой полиморфизм определяет необходимость генотипирования растений, непосредственно участвовавших в скрещиваниях, для достоверной идентификации гибридов.

ЗАКЛЮЧЕНИЕ

1. Установлено, что надежность молекулярно-генетической идентификации генотипов сои определяется информативностью используемой системы SSR-маркеров, стандартизацией условий получения и интерпретации ДНК-профилей. Выделение ДНК из частей зародыша семени позволяет исключить этап проращивания семян и получить препараты ДНК, пригодные для ПЦР. Капиллярный электрофорез обеспечивает необходимую разрешающую способность при разделении амплифицированных фрагментов. Разработанные две мультиплексные системы позволяют одновременно амплифицировать по 11 SSR-маркеров в каждой. В совокупности эти этапы обеспечивают точность и воспроизводимость результатов генотипирования, а также повышают пропускную способность анализа.

2. Изученные сорта сои характеризуются различной степенью внутрисортowego полиморфизма. Установлена ограниченная информативность смешанного ДНК-профиля, поскольку он не отражает полноту аллельного состава сорта и не позволяет выявлять биотипы и определять их частоту встречаемости. Сорт сои как объект молекулярно-генетической идентификации следует рассматривать как структурированную совокупность генотипических вариантов (биотипов).

3. Межсортная дифференциация сои определяется не только преобладающими ДНК-профилями, но и биотипным составом всей выборки. Наличие общих биотипов у разных сортов ограничивает возможность идентификации по единичному растению, тогда как малочисленные специфичные биотипы могут иметь существенное значение для сортовой идентификации. Поэтому идентификация сорта должна быть основана на сопоставлении выборок индивидуальных растений.

4. Показана возможность частичной автоматизации сопоставления ДНК-профилей при молекулярно-генетической идентификации сортов сои с помощью алгоритма Random Forest. Однако результат классификации отдельного растения зависит от представленности соответствующего биотипа в обучающей выборке и не является окончательным заключением о сортовой принадлежности исследуемого материала. Идентификация сорта должна быть основана на совокупности результатов анализа выборки индивидуальных растений с учетом соотношения правильно классифицированных, неидентифицированных и нетипичных генотипов, а также состава общих и специфичных биотипов.

5. Ввиду наличия внутрисортowego полиморфизма у сои при идентификации гибридов F_1 необходимо анализировать родительские растения, непосредственно использованные в скрещивании. Информативность SSR-

маркеров определяется различием в аллельном составе конкретных родительских форм, применение полного набора маркеров из разработанных мультиплексных систем позволяет выявлять истинные и ложные гибриды без предварительного подбора маркеров для каждой отдельной комбинации.

ПРЕДЛОЖЕНИЯ ДЛЯ СЕЛЕКЦИОННОЙ ПРАКТИКИ

1. Внедрить разработанную технологию молекулярно-генетической идентификации сортов сои в практику селекционно-семеноводческих центров и испытательных лабораторий для решения задач сортового контроля (патент на изобретение № 2871306).

2. Использовать алгоритм классификации Random Forest для обеспечения частичной автоматизации процедуры сопоставления ДНК-профилей сортов сои и повышения воспроизводимости результатов идентификации.

3. Применять разработанные системы SSR-маркеров в селекционном процессе для идентификации исходного материала, включая отбор и выявление биотипов, подтверждение гибридности.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

Публикации в рецензируемых научных изданиях, рекомендованных ВАК РФ:

1. Рамазанова, С. А. Поиск новых SSR-локусов ДНК для создания эффективной технологии генотипирования сои / С. А. Рамазанова, **В. Г. Савиченко**, Э. Г. Устарханова [и др.] // Масличные культуры. – 2021. – № 4 (188). – С. 18–24. DOI: 10.25230/2412-608X-2021-4-188-18-24

2. Рамазанова, С. А. Оценка эффективности выделения ДНК из разных частей растений сои для создания генетических паспортов / С. А. Рамазанова, **В. Г. Савиченко**, Д. Л. Савиченко [и др.] // Кормопроизводство. – 2023. – № 10. – С. 21–24. DOI: 10.25685/krm.2023.10.2023.002

3. **Савиченко, В. Г.** Поиск информативных SSR-маркеров для паспортизации сортов сои / В. Г. Савиченко, С. А. Рамазанова, С. З. Гучетль // Масличные культуры. – 2024. – № 2 (198). – С. 10–15. DOI: 10.25230/2412-608X-2024-2-198-10-15

4. **Савиченко, В. Г.** Генетический полиморфизм сортов сои селекции ВНИИМК / В. Г. Савиченко, Д. Л. Савиченко, С. А. Рамазанова [и др.] // Достижения науки и техники АПК. – 2024. – Т. 38. – № 10. – С. 5–10. DOI: 10.53859/02352451_2024_38_10_5

5. **Савиченко, В. Г.** Разработка мультиплексных систем SSR-маркеров сои (*Glycine max* (L.) Merr.) для паспортизации и идентификации сортов / В. Г. Савиченко // Труды по прикладной ботанике, генетике и селекции. – 2026. – Т. 187. – № 2. – С. – 193–206. DOI: 10.30901/2227-8834-2026-2-013

Публикации в рецензируемых журналах и материалах конференций:

6. Ramazanova, S. A. Genetic analysis of the homogeneity of soybean varieties by microsatellite loci / S. A. Ramazanova, **V. G. Savichenko**, S. Z. Guchetl // IOP Conference Series: Earth and Environmental Science. IOP Publishing Ltd, 2021. – V. 699. – № 1. – P. 012007. DOI: 10.1088/1755-1315/699/1/012007

7. **Савиченко, В. Г.** Определение внутрисортного полиморфизма сортов сои по микросателлитным (SSR) локусам / В. Г. Савиченко, С. А. Рамазанова // Научное обеспечение устойчивого развития агропромышленного комплекса в условиях аридизации климата. Сборник материалов международной научно-практической конференции, посвященной 35-летию ФГБНУ РосНИИСК «Россорго». Саратов, 2021. – С. 232–238.

8. **Савиченко, В. Г.** Идентификация сортов сои селекции ВНИИМК методом микросателлитного анализа / В. Г. Савиченко, С. А. Рамазанова // Актуальные вопросы биологии, селекции, технологии возделывания и переработки сельскохозяйственных культур. Сборник материалов 11-й Всероссийской конференции молодых ученых и специалистов. Краснодар, 2021. – С. 97–101. DOI: 10.25230/conf11-2021-97-101

9. Рамазанова, С. А. Анализ полиморфизма микросателлитных локусов ДНК для генотипирования сои методом капиллярного электрофореза / С. А. Рамазанова, **В. Г. Савиченко** // Генетические ресурсы растений для генетических технологий : к 100-летию Пушкинских лабораторий ВИР : материалы Всероссийской научно-практической конференции, г. Санкт-Петербург, 22–23 июня 2022 г. : тезисы докладов : научное электронное издание / под редакцией Ю. В. Ухатовой, Е. А. Соколовой ; Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова. – Санкт-Петербург : ВИР, 2022. – С. 124–125.

10. **Savichenko, V.** Fragment analysis of microsatellite DNA loci for genotyping soybean varieties / V. Savichenko, S. Ramazanova, S. Guchetl // International Scientific and Practical Conference "Current Issues of Biology, Breeding, Technology and Processing of Agricultural Crops" (CIBTA2022). Conference Proceedings (To the 110th anniversary of V.S. Pustovoit All-Russian Research Institute of Oil Crops). United States, 2023. – V. 2777. – № 1. – P. 020028-1-020028-6. DOI: 10.1063/5.0140280

11. **Савиченко, В. Г.** Детекция фрагментов ДНК при проведении молекулярно-генетической паспортизации сортов сои / В. Г. Савиченко, С. А. Рамазанова // Актуальные вопросы биологии, селекции, технологии возделывания и переработки сельскохозяйственных культур : Сборник материалов 12-й Международной конференции молодых учёных и специалистов, Краснодар, 01–03 марта 2023 года. – Краснодар: Федеральный научный центр "Всероссийский научно-исследовательский институт

масличных культур имени В.С. Пустовойта", 2023. – С. 242–246.

DOI: 10.25230/conf12-2023-242-246

12. **Савиченко, В. Г.** Идентификация сортов сои на основе полиморфизма SSR-маркеров / В. Г. Савиченко // Научные труды Северо-Кавказского федерального научного центра садоводства, виноградарства, виноделия. – 2023. – Т. 37. – С. 68–72. DOI: 10.30679/2587-9847-2023-37-68-72

13. Рамазанова, С. А. Изучение генетического разнообразия сортов сои разного происхождения с использованием микросателлитных локусов / С. А. Рамазанова, **В. Г. Савиченко**, С. З. Гучетль // Агронаука. – 2023. – Т. 1. – № 1. – С. 96–103.

Патенты и базы данных:

1. Патент № 2871306 С1 Российская Федерация МПК С12Q 1/68. Способ молекулярно-генетической идентификации сортов сои на основе мультиплексного микросателлитного анализа : заявл. 26.12.2025 : опубл. 06.10.2026 / **В. Г. Савиченко**, С. А. Рамазанова, С. З. Гучетль ; заявитель Федеральное государственное бюджетное научное учреждение "Федеральный научный центр "Всероссийский научно-исследовательский институт масличных культур имени В.С. Пустовойта".

2. Свидетельство о государственной регистрации базы данных № 2026622727 Российская Федерация. База данных ДНК-профилей сортов сои по микросателлитным маркерам : заявл. 25.05.2026 : опубл. 10.06.2026 / **В. Г. Савиченко**, С. З. Гучетль, С. В. Иванов [и др.] ; заявитель Федеральное государственное бюджетное научное учреждение "Федеральный научный центр "Всероссийский научно-исследовательский институт масличных культур имени В.С. Пустовойта".

3. Свидетельство о государственной регистрации базы данных № 2026623478 Российская Федерация. База данных панелей бинов m1 и m2 для определения аллелей микросателлитных локусов ДНК сои : заявл. 07.07.2026 : опубл. 17.07.2026 / **В. Г. Савиченко**, С. З. Гучетль, С. А. Рамазанова ; заявитель Федеральное государственное бюджетное научное учреждение "Федеральный научный центр "Всероссийский научно-исследовательский институт масличных культур имени В.С. Пустовойта".

Методические рекомендации:

1. Молекулярно-генетическая паспортизация сортов сои: методические рекомендации / **В. Г. Савиченко**, С. З. Гучетль, С. А. Рамазанова. – Краснодар: ООО «Просвещение-Юг», 2026. – 36 с.