

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ
УЧРЕЖДЕНИЕ “ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК”**

На правах рукописи

АФОННИКОВ ДМИТРИЙ АРКАДЬЕВИЧ

**КОМПЬЮТЕРНЫЕ МЕТОДЫ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО
ФЕНОТИПИРОВАНИЯ РАСТЕНИЙ**

1.5.8 – математическая биология, биоинформатика

Диссертация на соискание ученой степени

доктора биологических наук

Научный консультант

академик РАН, д.б.н., проф. Колчанов Н.А.

Новосибирск

2023

ОГЛАВЛЕНИЕ

Оглавление	1
Введение	5
Глава 1. Обзор литературы	15
1.1. Задачи феномики растений	15
1.2. Получение и анализ изображений	22
1.3. Фенотипирование растений	42
1.4. Интеграция данных в феномике растений	50
1.5. Заключение по обзору литературы и постановка задачи исследования	52
Глава 2. Высокопроизводительное фенотипирование опушения листа у растений	55
2.1. Опушение листа у растений	55
2.2. Протокол получения микроизображений сгибов листа пшеницы	64
2.2.1. Растительный материал и условия выращивания	64
2.2.2. Протокол пробоподготовки и получения изображения сгибов листа	65
2.3. Оценка количественных характеристик опушения листа на основе анализа двумерных микроизображений сгиба	73
2.3.1. Алгоритм автоматической оценки количественных характеристик опушения листа LHDetect	73
2.3.2. Алгоритм автоматической оценки количественных характеристик опушения листа LHDetect2	76
2.3.3. Компьютерная программа LHDetect2 для оценки количественных характеристик опушения листа пшеницы	89
2.3.4. Статистические методы анализа количественных характеристик опушения листа	95
2.4. Разнообразие опушения листьев у мягкой пшеницы и ее сородичей	97
2.4.1. Растительный материал и условия выращивания	97
2.4.2. Фенотипирование растений и обработка данных	99
2.4.3. Количественные характеристики опушения листа пшеницы и ее сородичей	101

2.4.4. Разнообразие опушения листа у мягких пшениц и их сородичей и его связь с геномным составом	112
2.5. Изучение фенотипического проявления генов опушения листа пшеницы <i>Triticum aestivum</i> L.	117
2.5.1. Растительный материал и условия выращивания	117
2.5.2. Фенотипирование растений и обработка данных	119
2.5.3. Анализ опушения листа	120
2.6. Анализ растений замещенных линий пшеницы (<i>Triticum aestivum</i> L.) в условиях нормального и ограниченного водообеспечения	130
2.6.1. Растительный материал и условия выращивания	130
2.6.2. Связь количественных характеристик опушения листа у изогенных и замещенных линий пшеницы с физиологическими показателями растений в условиях нормального и ограниченного водообеспечения	134
2.7. Высокопроизводительное фенотипирование опушения листа картофеля	149
2.7.1. Условия, материалы и методы для анализа опушения листа у картофеля	149
2.7.2. Оценка количественных характеристик опушения листьев картофеля с использованием анализа цифровых микроизображений	151
2.8. Высокопроизводительное фенотипирование опушения листа табака	156
2.8.1. Генетические линии табака с измененным паттерном экспрессии гена пролиндегидрогеназы	158
2.8.2. Растительный материал и условия выращивания генетически модифицированных линий табака	159
2.8.3. Модификация параметров алгоритма для анализа опушения листа табака	161
2.8.4. Сравнение характеристик опушения у ГМ-линий и контрольного сорта табака	166
2.9. Заключение по главе 2	171
Глава 3. Высокопроизводительное фенотипирование зерен пшеницы	173
3.1. Зерна злаков и их характеристики важные для селекционно-генетических исследований	173
3.2. Приложение SeedCounter для морфометрии зерен пшеницы	178

3.2.1. Протоколы получения изображений зерен пшеницы для фенотипирования с помощью мобильных устройств	178
3.2.2. Алгоритм обработки изображений для морфометрии зерен	179
3.3. Поиск локусов количественных признаков для характеристик размера, формы и цвета зерен мягкой пшеницы	193
3.3.1. Растительный материал и условия выращивания	193
3.3.2. Протокол фенотипирования зерен	194
3.3.3. Оценка характеристик зерен	196
3.3.4. Анализ локусов количественных признаков и приоритизация генов в области QTL	200
3.3.5. Анализ характеристик зерен в популяции ITMI	204
3.3.6. Анализ QTL для характеристик зерен в популяции ITMI	209
3.3.7. Приоритизация генов из районов QTL для характеристик зерен	218
3.4. Изучение взаимосвязи между характеристиками размера, формы и цвета зерен мягкой пшеницы, временем их хранения и всхожестью	222
3.4.1. Биологический материал	223
3.4.2. Протокол съемки зерен и обработка изображений и статистический анализ	223
3.4.3. Анализ влияния генотипа и времени хранения на признаки зерен	226
3.4.4. Анализ связи между всхожестью и признаками зерен	234
3.5. Заключение по главе 3	241
Глава 4. Высокопроизводительное фенотипирование колоса пшеницы	245
4.1. Колос и его характеристики важные для селекционно-генетических исследований	245
4.2. Протоколы получения изображений колоса	247
4.3. Идентификация колоса и остей на изображении	250
4.3.1. Предварительная обработка изображения	250
4.3.2. Бинаризация изображения	251
4.3.3. Идентификация остей	251
4.4. Алгоритмы морфометрии колоса	255

4.5. Алгоритмы предсказания индекса плотности колоса и типа его формы	260
4.6. Оценка точности распознавания областей остей и колоса	263
4.7. Анализ характеристик колосьев и предсказание типа колоса	268
4.8. Заключение по главе 4	280
Глава 5. Базы данных для описания фенотипа пшеницы	283
5.1. База данных WheatPGE	283
5.2. База данных SpikeDroid DB	298
5.3. Заключение по главе 5	307
6. Заключение	309
Выводы	316
Список публикаций в рецензируемых журналах и авторские свидетельства по теме диссертации	319
Использованные сокращения	323
Список литературы	325
Приложение	377

ВВЕДЕНИЕ

Актуальность работы. Глобальные климатические и социальные процессы оказывают всё более существенное влияние на условия ведения сельского хозяйства в разных странах. С одной стороны, в связи с увеличением населения Земли стабильно возрастает потребность в продовольствии. По данным ФАО (FAO, Food and Agriculture Organization, <http://www.fao.org/>), численность и доходы населения должны вырасти вдвое к 2050 году. Это увеличивает конкуренцию за природные (водные, земельные) и сельскохозяйственные ресурсы. С другой стороны, мировое сельское хозяйство сталкивается с такими неблагоприятными климатическими изменениями как повышение сезонных температур во всех широтах, засуха, увеличение содержания углекислого газа в атмосфере (Long, Ort, 2010). Это требует пересмотра методов и подходов в селекции. Необходимо создавать сорта сельскохозяйственных культур с более высоким потенциалом урожайности и устойчивости к абиотическим стрессам (White et al., 2012). Последнее особенно важно, так как расширение посевных площадей возможно только за счет районов, где условия возделывания отличаются от оптимальных, и где урожайность имеющихся сортов может снижаться (Long, Ort, 2010).

Важной стратегией улучшения сельскохозяйственных культур является поиск генетических образцов, которые эволюционировали в суровых климатических условиях и поэтому адаптировались к ним. Они в огромном количестве представлены в генбанках всего мира, однако большинство из этих образцов требует изучения и детального описания (Cobb et al., 2013). У растений, адаптированных к суровым условиям, необходимо охарактеризовать множество различных признаков, представляющих ценность при создании

новых сортов. Только тогда они будут эффективно использоваться в селекционной работе.

Анализ данных большого масштаба актуален и при поиске генов, ответственных за важные хозяйственные признаки растений. Подавляющее большинство этих признаков, такие как урожайность, биомасса растения, сроки основных фаз роста, устойчивость к стрессу и заболеваниям контролируются полигенно. Основная идея поиска генов (локусов количественных признаков, QTL), контролирующих такие признаки, заключается в выявлении ДНК-маркеров, которые ассоциированы с уровнем экспрессии исследуемого признака (Kearsey, Farquhar, 1998; Würschum, 2012; Хлесткина, 2013). В последнее время технологии секвенирования ДНК позволяют недорого и быстро прочитывать последовательности как полных геномов индивидуальных организмов, так и отдельных его локусов (Mardis, 2008; Varshney et al., 2009). Это позволяет эффективно решать задачи идентификации геномных маркеров (Elshire et al., 2011; Miller et al., 2007; Paux et al., 2012; Нестеров и соавт, 2015) и выявления их вариаций для большого количества генотипов (Lijavetzky et al., 2007; Li et al., 2013; Cuesta-Marcos et al., 2010; Xu et al., 2010). Как результат, высокопроизводительные технологии секвенирования привели к созданию новых методов селекции, таких как маркер-контролируемый отбор, (Marker-assisted selection, MAS) (Collard, Mackill, 2008) и геномная селекция (Meuwissen et al., 2001).

Следует отметить, что в основе успешного поиска взаимосвязи между генотипом и фенотипом лежат статистические методы, эффективность и точность которых зависит от размера анализируемой выборки. Большой размер выборки позволяет уменьшить ошибку идентификации QTL, а также увеличить точность определения генетического расстояния между маркерами. Это демонстрируется аналитическими расчетами (Frisch et al., 1999),

численными экспериментами (Visscher et al., 1996), практическим опытом проведения QTL (Beavis, 1998). Фактор размера выборки также важен для методов поиска полногеномных ассоциаций (genome-wide association studies, GWAS) и геномной селекции (Meuwissen et al., 2001; Korte et al., 2012). Поэтому для идентификации локусов, отвечающих за количественные признаки, все чаще используют выборки из тысяч растений.

При проведении подобных масштабных экспериментов одним из узких мест в настоящее время является измерение фенотипических параметров растения (Furbank, Tester, 2011), которое до недавнего времени было трудоемким, неавтоматизированным и зачастую базировалось на субъективных оценках фенотипа экспертами. Все это препятствовало эффективной оценке фенотипа для тысяч растений. Попытки решить эту проблему привели к формированию новой области науки на стыке биологии, информатики и инженерии - феномики.

Цель работы. Целью работы является создание экспериментально-компьютерной платформы для высокопроизводительного компьютерного фенотипирования растений на основе анализа цифровых изображений для оценки характеристик опушения листа, размера, формы и цвета зерен, формы колоса и изучение с ее помощью взаимосвязи признаков фенотипа растений с генотипом и ответом на воздействие окружающей среды.

Задачи, поставленные для достижения указанной цели, включали:

1. Разработать экспериментально-компьютерную платформу ICGPhenoPlant для фенотипирования растений, включая базы данных для хранения полученной информации о характеристиках:

- (а) опушения листьев пшеницы, картофеля и табака;
- (б) размера, формы и цвета оболочки зерен пшеницы;
- (в) размера и формы колоса пшеницы.

2. С помощью платформы исследовать разнообразие опушения листа у сортов и линий мягкой пшеницы и ее сородичей, оценить влияние генов контроля опушения листьев пшеницы на формирование этого признака и его изменчивости у растений при воздействии засухи.

3. С помощью платформы оценить разнообразие опушения листа российских сортов картофеля и изменчивость опушения листа у линий табака с модификацией гена пролиндегидрогеназы.

4. С помощью платформы провести анализ размера, формы и цвета оболочки зерен пшеницы популяции ITMI и их связи с локусами количественных признаков, длительностью хранения в генбанке и всхожестью.

5. С помощью платформы оценить характеристики формы, размера, остистости колосьев пшеницы и на этой основе построить метод предсказания их типа и индекса плотности.

Научная новизна.

Экспериментально-компьютерная платформа для высокопроизводительного компьютерного фенотипирования растений ICGPhenoPlant включает впервые разработанный комплекс методов для фенотипирования опушения листьев пшеницы на основе анализа микроизображений сгибов листа программой LHDetect2. Он позволяет оценивать количественные характеристики опушения листа на изображении с высокой точностью: распределение трихомов по длинам, среднее число трихомов, среднюю длину трихомов, соотношение средней длины и количества трихомов. С помощью методов фенотипирования опушения листьев пшеницы:

(а) впервые для 47 образцов пшеницы и ее диких сородичей с помощью программы LHDetect2 проведен высокоточный анализ количественных

характеристик опушения, выявлены их основные особенности и связь с хромосомным набором генома пшениц;

(б) впервые количественно оценено влияние на число и длину трихомов генов, контролирующих опушение листа для ряда сортов, замещенных и почти изогенных линий мягких пшениц;

(в) впервые на основе точной оценки количественных характеристик опушения листа пшеницы программой LHDetect2 показаны их изменения в ответ на засушливые условия произрастания и их связь с изменением физиологических параметров растений;

С помощью экспериментально-компьютерной платформы ICGPhenoPlant было проведено фенотипирование опушение листьев двудольных растений, картофеля и табака, и показано что:

(а) растения сибирских сортов картофеля демонстрируют разнообразие количественных характеристик опушения листьев;

(б) растения генетически-модифицированных линий табака с измененным паттерном экспрессии гена пролиндегидрогеназы демонстрируют достоверно большее количество трихомов на листе и меньшую среднюю длину трихомов.

В рамках разработки экспериментально-компьютерной платформы ICGPhenoPlant впервые было разработано приложения SeedCounter для фенотипирования зерен злаков с помощью мобильных устройств. С его помощью впервые исследована связь количественных характеристик зерен пшеницы и сроком их хранения в генбанке и показано, что для большинства признаков цвета зерна существует значимая статистическая линейная связь между величиной признака и годом урожая.

В рамках разработки экспериментально-компьютерной платформы III CGPhenoPlant разработано приложение WERrecognizer для оценки

морфометрических характеристик колосьев пшеницы, которое впервые использует модель двух четырехугольников для описания формы колоса на основе анализа его цифровых изображений. Показано, что геометрические параметры колоса позволяют предсказывать его тип и плотность с высокой точностью.

Положения, выносимые на защиту:

1. Экспериментально-компьютерная платформа ICGPhenoPlant разработанная для фенотипирования растений, включает программные продукты LHDetect, WERecognizer и SeedCounter, которые позволяют с высокой точностью на основе двумерных изображений оценивать количественные параметры опушения листьев, формы, размеров колоса и цвета оболочки зерен растений, а также информационные ресурсы WheatPGE и SpikeDroid, обеспечивающие возможность накопления и сохранения информации о фенотипе, генотипе и месте произрастания растений в соответствующих базах данных.

2. Высокопроизводительное компьютерное фенотипирование опушения листьев пшеницы с помощью программы LHDetect2 позволяет установить достоверную статистическую связь между количественными признаками опушения листа, хромосомным составом генома, аллельным составом генов, контролирующим развитие опушения, и изменением физиологических параметров растений в процессе ответа на воддефицит.

3. Программа LHDetect2 позволяет достоверно оценить изменчивость количественных характеристик опушения листьев для генетически модифицированных линий табака с измененным паттерном экспрессии гена пролиндегидрогеназы и увеличенным содержанием пролина по сравнению с контрольным генотипом SR1.

4. Компьютерное фенотипирование зерен мягкой пшеницы урожаяв различных годов с помощью приложения SeedCounter демонстрирует достоверные изменения характера пигментации их оболочки в зависимости от срока хранения.

5. С помощью приложения SeedCounter установлено, что у мягкой пшеницы существует достоверная корреляция между характером пигментации оболочки зерновки, сроком хранения зерна и их всхожестью.

Теоретическая значимость работы.

Разработан комплекс экспериментально-компьютерных методов ICGPhenoPlant для фенотипирования важных признаков растений: опушения листа (у пшеницы, картофеля и табака), характеристик размера, формы и цвета оболочки зерен, морфометрических характеристик колоса (у пшеницы). Данный комплекс методов позволяет получать быстро и с высокой точностью оценивать количественные характеристики фенотипа растений, что позволяет использовать предложенные подходы для массового анализа при решении задач поиска взаимосвязи генотип–фенотип и окружающая среда–фенотип у растений.

Научно-практическая значимость работы.

Разработанный комплекс экспериментально-компьютерных методов и баз данных ICGPhenoPlant доступен для исследователей в области генетики и селекции сельскохозяйственных растений в Интернет для свободного использования. Это обеспечивает широкое использование предложенных методов для решения задач в области генетики и селекции, поиска взаимосвязи между генотипом, фенотипом и окружающей средой, и, в перспективе, создания новых высокопродуктивных сортов сельскохозяйственных растений.

Апробация работы.

Результаты диссертационной работы были представлены на конференциях БГРС (2012, 2014, 2016, 2018, 2020, 2022 гг.); PlantGen (2015, 2017, 2019, 2021 гг.); VII Московском международном конгрессе «Биотехнология: состояние и перспективы развития» (г. Москва, 2013); 1-м Международном семинаре «Генетика и геномика растений для продовольственной безопасности» (НГУ, 2016); конференции PhenoDays-2014 (Bon, France, 2014); XIII Новосибирском инновационно-инвестиционном форуме «Сельхозмашиностроение и переработка агропродукции» (2018 г.); Школе-конференции по количественной генетике (ВИР, 2018 г.); Конференции «125 лет прикладной ботаники в России» (ВИР, 2019 г.); на VII съезде Вавиловского общества генетиков и селекционеров, посвященному 100-летию кафедры генетики СПбГУ (ВИР, 2019); на II Европейской школе биотехнологов (Минск, 2019); на Международной научно-практической конференции «Развитие сельского хозяйства на основе современных научных достижений и интеллектуальных цифровых технологий «Сибирь – Агробиотехнологии» («САБИТ 2019») (СибНИИРС, 2019); форуме Технопром-2019 (г. Новосибирск, 2019); на Третьем Международном форуме «Передовые цифровые и производственные технологии» (СПбПУ, 2021); конференции "В генетику - через разные отрасли науки и практики" (Центр Сириус, 2022).

Публикации.

По теме диссертационной работы опубликовано 24 работы, в том числе из них 12 в зарубежных и российских журналах с импакт-фактором, 7 в российских журналах, индексируемых в системе WoS без импакт-фактора, 2 журнала, индексируемых в системе SCOPUS и три авторских свидетельства. Все работы опубликованы в соавторстве.

Личный вклад автора.

Данная работа была бы невозможна без участия большого количества исследователей. Автор внес решающий вклад в постановку цели и задач исследования, обоснование выбора методов для решения задач, обработку данных, интерпретацию и обобщение полученных результатов. В большинстве статей, опубликованных по тематике диссертации автор был автором для переписки.

Основные результаты главы 2 были получены в соавторстве с Пшеничниковой Т.А., Генаевым М.А., Дорошковым А.В., Ибрагимовой С.С., Осиповой С.В., Пермязковой М.Д., Пермязковым А.В., Колчановым Н.А. Основные результаты главы 3 получены в соавторстве с Комышевым Е.Г., Генаевым М.А., Ковалем В.С., Ефимовым В.М., Шмаковым Н.А., Арифом М.А.Р., Бёрнером А. Основные результаты главы 4 получены в соавторстве с Генаевым М.А., Комышевым Е.Г., Гончаровым Н.П., Кручининой Ю.В., Ковалем В.С. Основные результаты главы 5 получены в соавторстве с Комышевым Е.Г., Генаевым М.А., Ковалем В.С., Гончаровым Н.П.

Структура и объем работы.

Работа состоит из Введения, 5 глав, включая Обзор литературы и 4 главы с описанием результатов работы, Заключение, Выводов, Списка публикаций по теме диссертации, Списка использованных сокращений, Списка литературы и Приложения. Объем работы составляет 391 лист. В работе 87 рисунков и 29 таблиц. Список литературы включает 438 источников.

Благодарности.

Автор выражает глубокую признательность научному консультанту академику РАН Колчанову Н.А. Автор выражает благодарность сотрудникам ИЦиГ СО РАН Пшеничниковой Т.А., Генаеву М.А., Дорошкову А.В., Ибрагимовой С.С., Комышеву Е.Г., Ковалю В.С., Ефимову В.М., Кочетову

А.В, а также А. Бёрнеру и Арифуну М.А.Р. за участие в работе на разных ее этапах.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Задачи феномики растений

Проблема важности изучения фенома в его взаимосвязи с информацией о полном геноме и геномных вариациях была сформулирована уже на заре постгеномной эры (Freimer, Sabatti, 2003). В широком смысле под феномом подразумевают наблюдаемое проявление генотипа в масштабе всего организма (Houle et al., 2010). В свою очередь, область науки, возникшая на стыке биологии, информатики и инженерии, связанная с разработкой методов измерения параметров фенотипа, его описания, обработки фенотипических данных получила название феномика.

Интенсивное развитие получили методы феномики при изучении растений (Furbank, Tester, 2011). Особенностью феномики растений является необходимость учета широкого спектра условий внешней среды (температуры, влажности, освещенности, типа почвы, и др.), в которой произрастают растения, поскольку предполагается, что до 50% фенотипических вариаций могут быть обусловлены средовыми факторами (Kearsey, 1998).

Существующие подходы к решению задачи взаимосвязи между фенотипом, генотипом и окружающей средой у растений выдвигают ряд вполне естественных требований к основным направлениям развития феномики растений.

Прежде всего, технологии фенотипирования должны быть направлены на определение характеристик фенотипа для одного организма, поскольку именно реализация генетической информации на фенотипическом уровне у отдельных организмов позволяет адекватно оценить их популяционную изменчивость. В некоторых селекционно-генетических экспериментах фенотипические характеристики линии или сорта оценивают усреднением по группам растений, например, для растений из одной делянки (урожайность,

высота растения). Однако даже в одной делянке идентичность условий произрастания строго не выполняется (из-за конкуренции растений между собой, различной освещенности и т.п.). Разумеется, анализ индивидуальных фенотипов более трудоемок. Тем не менее, именно такой анализ находится в фокусе разработки современных фенотипических методов (Eberius, Lima-Guerra, 2009; Klukas et al., 2014; Hartmann et al., 2011).

Одним из важнейших свойств фенотипа является его многоуровневый характер, поскольку проявление генома можно описать на всех уровнях организации живых систем – от молекулярного, до целого организма (Houle et al., 2010). Методы анализа микроскопических изображений позволяют изучать особенности экспрессии генов в органах, ансамблях клеток, детально охарактеризовать микроскопическую структуру органов и тканей (Sapri P.G., Heisler, 2013). Однако область интересов селекционеров в большей степени связана с изучением фенотипических характеристик на уровне органов (таких как корень, лист, стебель, соцветие, колос, семена), физиологических свойств растений (скорость развития на отдельных этапах онтогенеза, показатели эффективности фотосинтеза и эффективность использования воды, устойчивость к стрессу) или их общих характеристик, таких как продуктивность, биомасса, устойчивость к заболеваниям. Поэтому фенотипические характеристики, связанные с этими уровнями организации или физиологическими параметрами, представляют для феномики растений наибольший интерес.

При анализе растения в целом или его отдельных органов для биолога, прежде всего, важны параметры, характеризующие форму, размер, массу. Поэтому одним из важных направлений современной феномики является разработка методов морфометрического анализа. Типичные примеры такого анализа – определение формы и размера листьев (Bylesjö et al., 2008; Backhaus

et al., 2010) , зерен (Herridge et al., 2011) или корней (Yazdanbakhsh, Fisahn, 2009). При решении этой задачи важно быстро и точно оценить различные характеристики формы и размеров органа растения, создать методы их сравнения, например, для дальнейшей классификации.

Еще одной часто встречающейся задачей феномики является подсчет объектов, например, числа зерен (Tanabata et al, 2012), числа трихомов на поверхности листа (Pomeranz et al., 2013), число корней (Yazdanbakhsh, Fisahn, 2009). Для этой задачи требуется быстрая идентификация биологических объектов на цифровых изображениях и их подсчет с минимальной ошибкой.

Ряд методов основан на анализе цветовых характеристик органов растений. Как правило, эти методы используются для определения спелости плода (Darrigues et al., 2008), степени поражения растения засухой или заболеванием (Camargo, Smith, 2009).

Разработка высокопроизводительных методов определения физиологических характеристик отдельных растений также является предметом феномики: массово оцениваются такие характеристики как скорость транспирации растения (Wallach et al., 2010), эффективность транспирации (Vadez et al., 2014), проводимость листового покрова (Vadez et al., 2015).

При измерении фенотипических характеристик одним из важных условий является возможность анализа фенотипа без повреждения растений. Такие неразрушающие технологии позволяют наблюдать изменение фенотипических характеристик для одного и того же растения в процессе его роста и развития. В результате можно оценивать скорость изменения фенотипа в зависимости от генотипа или условий среды. В некоторых случаях допустимо получать образцы для фенотипирования с частичным повреждением органов растений, но при сохранении их способности выжить и дать потомство.

Не всегда с помощью методов фенотипических измерений оказывается возможным напрямую измерять характеристики растений, традиционно используемых в селекции. Поэтому исследователям приходится оценивать их косвенными способами. Например, проективная площадь побега на двумерных изображениях демонстрирует корреляцию с суммарной площадью листа, сухой и сырой биомассой у растений *Arabidopsis thaliana* (Leister et al., 1999), пшеницы и ячменя (Golzarian et al., 2011). Если анализ изображения дает оценку лишь «косвенных» признаков, прибегают к перенормировке данных – на основе обработки достаточно большого числа образцов из тестовой выборки вычисляют специальную «подгоночную» функцию (не обязательно линейную), которая связывает «косвенную» и «прямую» фенотипические характеристики (Eberius, Lima-Guerra, 2009).

Отметим, что возможности феномики не ограничены анализом набора характеристик, традиционно используемых для описания фенотипа. Очень часто высокопроизводительные методы позволяют получать совершенно новую информацию о фенотипе растения, такую, которая раньше не была доступна для анализа. Например, при анализе устойчивости различных генотипов ячменя к засухе для оценки состояния растений использовали 388 фенотипических характеристик, полученных на основе анализа изображений в видимом, инфракрасном, ультрафиолетовом диапазонах, различных характеристик формы и размера побегов (Klukas et al., 2014).

Одной из важных составляющих селекционно-генетического эксперимента является необходимость учета условий окружающей среды. При анализе растений для решения этой задачи эксперимент можно проводить в полностью контролируемых условиях (при заданном уровне полива, освещенности, температуры). Варьируя контролируемые параметры среды (например, в теплице или специальном боксе) можно оценить их влияние на

изменчивость фенотипа растения (Chen et al., 2014). Однако искусственные условия не могут полностью воспроизвести разнообразие абиотических факторов, которые действуют на растения в естественных условиях. Поэтому одним из актуальных направлений феномики является изучение растений в полевых условиях (Araus, Cairns, 2014; Comar et al., 2012; Billau et al., 2012; Walter et al., 2012). Особенности такого анализа заключаются в том, что в полевых условиях объект исследования удален от лабораторного оборудования, что затрудняет анализ растений высокопроизводительными автоматическими методами. Для ускорения сбора информации о фенотипах растений в поле разрабатываются технологии с использованием сенсоров, установленных на движущейся платформе (Busemeier et al., 2013) или удаленного мониторинга с использованием БПЛА (беспилотных летательных аппаратов) (Honkavaara et al., 2013). Одним из перспективных направлений фенотипирования в полевых условиях может стать широкое внедрение мобильных устройств, которые даже при использовании низкопроизводительных технологий измерения фенотипа могут существенно ускорить процесс за счет прямого ввода информации в объединенную базу данных (Vankadavath et al., 2009), в том числе и через мобильный Интернет. Такие сетевые технологии были использованы для обеспечения контроля за распространением сорных растений в посевах на основе вовлечения в этот процесс добровольцев (технологии краудсорсинга) (Rahman et al., 2015).

Большое разнообразие фенотипических характеристик, проведение экспериментов в разных лабораториях и географических точках обуславливают актуальность стандартизации описания фенотипических параметров растения, условий среды и селекционно-генетического эксперимента (White et al., 2013; Krajewski et al., 2015). Такая стандартизация необходима для того, чтобы можно было сравнивать одни и те же наборы

признаков, измеренные исследователями в разных лабораториях и разных экспериментах. Одним из направлений решения данной проблемы является описание фенотипических признаков растений в виде онтологий – структурированных и контролируемых словарей биологических терминов (Shrestha et al., 2012), а также формирование стандартных протоколов по измерению и описанию фенотипических характеристик растений (Cornelissen et al., 2003).

Массовый анализ тысяч растений накладывает самое основное требование на методы фенотипирования: необходимость высокой степени автоматизации этого процесса (Kolukisaoglu, Thurgow, 2010). Основной особенностью автоматизации фенотипирования является исключение из этого процесса человека (работника), что значительно повышает производительность эксперимента. Одно из решений в этом направлении – создание автоматизированных боксов для выращивания растений (Arvidsson et al., 2011) или автоматизированных теплиц (Hairmansis et al., 2014; Klukas et al., 2014; Hartmann et al., 2011). Боксы обеспечены установками поддержания микроклимата и освещения, камерой и сенсорами для получения изображений. В автоматических теплицах растения высаживаются в вазоны, которые в процессе эксперимента постоянно перемещаются и автоматически подаются к регистрирующим биологические показатели устройствам, весь процесс роста растений происходит без участия человека, роль которого заключается в формировании программы эксперимента и вводе ее в компьютер. Теплицы обеспечены установками автоматического полива, камерами для измерения фенотипических параметров. Вазоны снабжены чипом с RFID-меткой (аналогично чипу на банковской карте), по которой в экспериментальной установке растение проходит идентификацию. Вокруг таких установок формируются исследовательские центры, например, Австралийский Центр

Феномики (<http://www.plantphenomics.org.au/>), сеть немецких центров по феномике

(http://www.dppn.de/dppn/EN/Infrastructures,%20Facilities/facilities_node.html).

Эти центры организованы в масштабную международную сеть по исследованиям в области феномики растений, International Plant Phenotyping Network (IPPN, <http://www.plant-phenotyping.org/>).

Автоматизация фенотипирования позволяет решить еще одну важную проблему. Оценка ряда параметров фенотипа человеком часто требует специальной квалификации, подготовки или большого опыта. К таким параметрам можно отнести качественные характеристики, такие как степень поврежденности листа растения заболеванием, поражения засухой или степень окраски, или опущения органов растения. Обычно эксперт на основе визуального анализа определяет одну из нескольких степеней, характеризующую признак: слабая, умеренная, сильная и т.д. Источником ошибок при этом может оказаться субъективизм эксперта. Использование компьютерного анализа при автоматическом фенотипировании позволяет этого избежать и оценивать характеристики с одинаковой точностью вне зависимости от того, кем и в какой лаборатории производится такая оценка. Автоматизация позволяет существенно снизить требования к квалификации экспериментатора: фенотипирование можно поручить менее квалифицированному специалисту. Все это позволяет существенно удешевить процесс фенотипирования вдобавок к увеличению его производительности. Еще одним преимуществом автоматизации процесса фенотипирования является возможность непосредственной записи информации о фенотипе от измерительного прибора в компьютерную базу данных, что актуально при анализе большого количества фенотипических признаков для тысяч растений.

1.2. Получение и анализ изображений

Получение изображений.

Методы фенотипирования на основе анализа изображений удовлетворяют большинству требований, сформулированных выше. Они могут быть неинвазивными, изображения могут быть получены быстро и в автоматическом режиме. Макроскопические изображения в видимом свете наиболее просты для получения и позволяют оценивать морфологию и ряд физиологических параметров органа, целого растения или популяции. Полезную информацию содержат изображения, полученные с помощью автофлуоресценции хлорофилла. Этот подход позволяет оценить активность фотосистемы II *in vivo* (Barbagallo et al., 2003). Съёмка в инфракрасном (ИК) диапазоне или при помощи мультиспектральных камер используется для оценки водного статуса растения (Zarco-Tejada et al., 2005). Томографические снимки также используются в задачах фенотипирования (Kaminuma et al., 2008). Наиболее ценным достоинством томографии является возможность неинвазивной реконструкции 3D-изображения не только видимых, но и скрытых от прямого наблюдения структур. Однако для томографии необходимо специальное дорогостоящее оборудование. Перспективным подходом является воссоздание объёмной модели органа или даже целого побега растения (Paprocki et al., 2012) а также использование специализированных систем активных дальнометров оптического диапазона (LIght Detection and Ranging) - LIDAR (Delagrangé and Rochon, 2011). Такие системы способны получать информацию об объёмной структуре объекта.

Обработка и анализ изображений.

Методы обработки и анализа изображений лежат в основе большинства технологий массового фенотипирования растений. Благодаря прогрессу в разработке эффективных алгоритмов, методов машинного обучения, созданию

библиотек и пакетов программ с удобным интерфейсом анализ изображений стал широко использоваться в современной биологии (Eliceiri et al., 2012).

В области обработки и анализа медицинских и биологических изображений используются шесть основных классов методов и алгоритмов: улучшение качества изображения (для уменьшения шумов, удаления артефактов), его сегментация (разделение на информативные области, определение их контуров и границ, выделение компактных объектов), количественный анализ формы и размера объектов на изображении, совмещение изображений (для изучения динамики их формы и размеров), сжатие и обеспечение хранения в базах данных (Eliceiri et al., 2012).

Список и основные характеристики библиотек и программных пакетов для анализа изображений приведены в табл. 1.1. Работа с библиотеками анализа изображений, такими как OpenCV требует знания навыков программирования. Однако такие пакеты как ImageJ, имеют удобный оконный пользовательский интерфейс, наборы базовых функций для решения основных задач анализа изображений, указанных выше, а также возможность для формирования вычислительных конвейеров из элементарных операций для обработки множества изображений однотипным способом. Развитие методов анализа изображений привело к созданию большого количества программ, направленных на определение фенотипических характеристик. Для систематизации этих программ была разработана база данных, доступная через Интернет, в которой хранится описание программ анализа изображений для растений, ссылки на них в Интернет, ссылки на статьи, в которых эти программы были описаны (Lobet et al., 2013).

Таблица 1.1. Программы для обработки изображений.

Название	Описание	Интерфейс	Коммерческий	Веб-сайт	Публикация
Matlab	Пакет прикладных программ широкого спектра	командная строка	да	http://www.mathworks.com	
ImageJ	Расширяемое программное обеспечение для анализа и обработки изображений	графический	нет	http://rsbweb.nih.gov/ij/	(Abràmoff et al., 2004)
Fiji	Программный пакет основанный на ImageJ	графический	нет	http://fiji.sc/Fiji	(Schindelin et al, 2012)
OpenCV	Библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения	программный	нет	http://opencv.org/	(Bradski, Kaehler, 2008)
FARSIGHT	Набор инструментов для обработки и анализа изображений	командная строка	нет	http://www.openmicroscopy.org/site/support/bio-formats5.1/users/farsight	
ProStack	Платформа для управления программными модулями обработки и анализа изображений	графический	нет	http://urchin.spbcas.ru/downloads/ProStack/ProStack.htm	(Kozlov et al., 2007)
CellProfiler	Программное обеспечение для решения биологических задач автоматизированного количественного фенотипирования на основе анализа изображений	графический	нет	http://www.cellprofiler.org	(Bray et al., 2014)

Растровое изображение представляет собой матрицу элементарных ячеек, содержащих информацию о цвете (для цветных изображений) или интенсивности свето-потока (для черно-белых) (Форсайт и Понс, 2004). Элементы матрицы называется пикселями изображения (pixel производное от “picture element”).

Среди обычных растровых изображений выделяют цветные, монохромные (градации серого, черно-белые) и бинарные. Монохромные изображения имеют всего один канал, хранящий значения интенсивности (градации) яркости пикселей в заданном диапазоне значений. Типичный формат представления монохромных изображений – это целочисленный массив размерности $N \times M$, содержащий значения яркости пикселей в диапазоне от 0 до 255 (один байт). Здесь N и M – размерность изображения в пикселях по горизонтали и вертикали, соответственно. Бинарные изображения являются частным случаем монохромных. Значение интенсивности свето-потока для них может принимать либо 0 или 1.

Цветное изображение состоит из нескольких целочисленных массивов. Один из популярных форматов представления цветных изображений является 24-битное RGB изображение (RGB это аббревиатура английских слов Red (красный), Green (зеленый), Blue (синий)). В этом случае изображение представлено массивом размерности $N \times M \times 3$ элементов. Каждый элемент хранит значение интенсивности одного из цветов: красного, зеленого и синего. Диапазон значений интенсивности связан с количеством выделенных бит на канал (цвет). Для 24-битного изображение интенсивность может принимать значение от 0 до 255. Такие изображения получаются в результате съёмки объектов цифровой камерой, наиболее часто используемый способ для фенотипирования растений.

Как правило, все изображения так или иначе содержат дефекты (помехи

и шум). Дефекты вносятся как фотосенсорами, в силу их конструкционных несовершенств и дискретизации, так и недостатками в протоколах получения изображений (неравномерность освещения, блики). Например, шум «соль и перец» представляет собой одиночные темные пиксели внутри светлых областей (перец), или одиночные светлые пиксели внутри темных областей (соль). Он может возникать в результате пороговой бинаризации изображения, а также после аналого-цифровых преобразований.

Существует множество различных методов для устранения шумов на изображении. Одним из них является применение двумерных линейных сглаживающих фильтров с конечной импульсной характеристикой фильтра. В качестве маски используется множество весовых коэффициентов, заданных во всех пикселях окрестности S , обычно симметрично окружающих текущий пиксель изображения. Примером такого сглаживающего фильтра является фильтр Гаусса. Преобразование для Гауссова фильтра с центром в окружности радиусом r считается по формуле:

$$y(m, n) = \frac{1}{2\pi r^2} \sum_{u,v} e^{-\frac{(u^2 + v^2)}{2r^2}} x(m + u, n + v) \quad (1.1)$$

где $x(m, n)$ – значение яркости пикселя в координате (m, n) изображения. Такое преобразование типа позволяет уменьшить “шум” изображения и сгладить некоторые “артефакты”, блики.

Сегментация изображений.

Биологические объекты на изображениях как правило характеризуются однородностью освещенности или цвета (по отношению к фону или остальному изображению). Примерами могут служить зерна на белом листе бумаги, лист растения на фоне почвы и т.п. Поэтому на изображении биологические объекты можно выделить как области одинаковые по

интенсивности цвета или каких-либо его компонент. Задача определения такой области базируется на понятии связности.

Связная область – это множество пикселей изображения, имеющих хотя бы одного соседа из данного множества (Гонсалес, Вудс, 2012). Таким образом, связная область состоит из множества пикселей изображения, в котором любые два пикселя можно соединить друг с другом через последовательность соседей, принадлежащих к этой же области. В зависимости от определения отношения соседства, т.е. какой из пикселей считается соседним к данному, различают 4-связность и 8-связность (рис. 1.1).

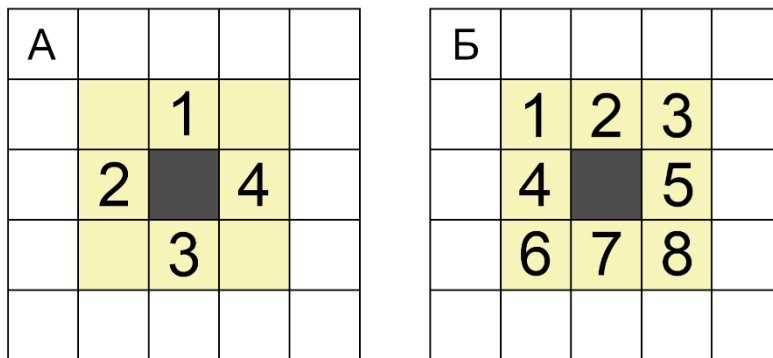


Рисунок 1.1. Примеры определения двух типов связности для центрального пикселя изображения (показан черным) и его соседних пикселей (показаны желтым). Пронумерованы пиксели, удовлетворяющие условию 4-связности (А) и 8-связности (Б).

Примеры связных и несвязных областей изображения приведены на рис.

1.2.

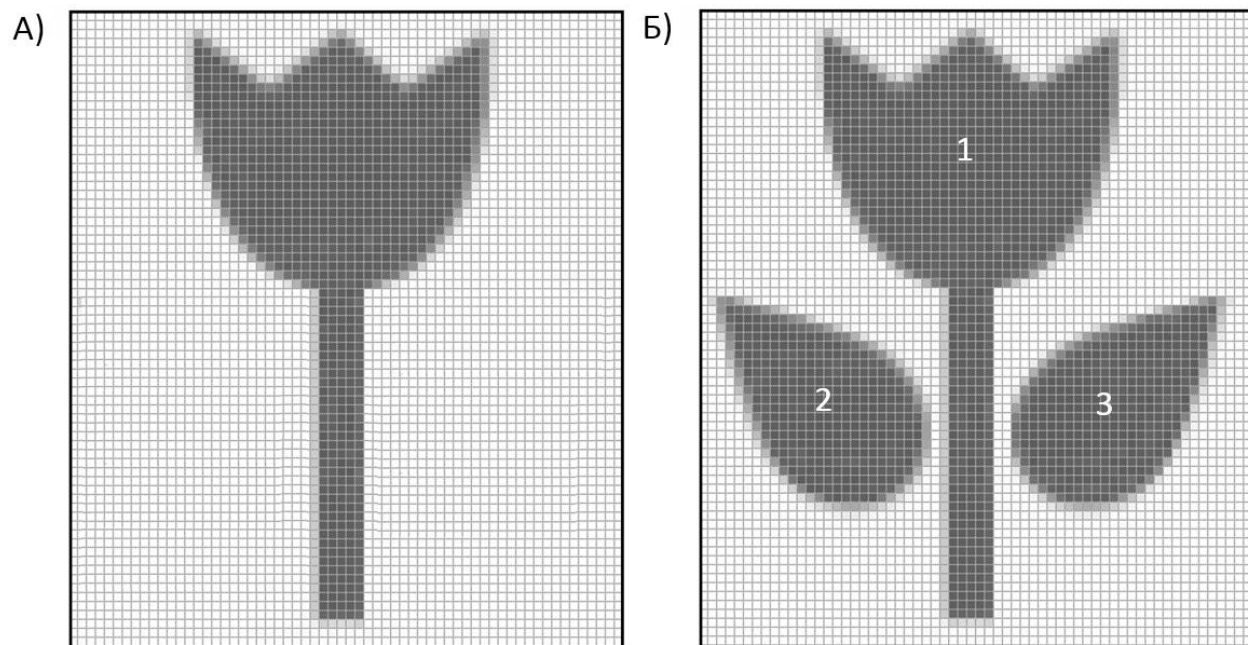


Рисунок 1.2. Примеры связных и несвязных областей на изображении. А) Темные пиксели на изображении образуют одну связную область; Б) темные пиксели не образуют единой связной области, а распадаются на три связных области (пронумерованных 1, 2 и 3).

Для каждой связной области можно определить контур, замкнутую или незамкнутую в пределах изображения линия (кривая), непрерывно соединяющую все пиксели вдоль границы связной области. Контур является полезным инструментом для анализа формы, а также для обнаружения и распознавания объектов.

Выделение связных областей на изображении – задача сегментации. Сегментация используется для выделения объектов или областей интереса на изображениях для дальнейшего анализа. Наиболее простой пример сегментации – пороговая бинаризация монохромного изображения, при которой все пиксели изображения на основе заранее выбранного порога интенсивности разделяются на два класса: пиксели фона и пиксели “интересующих объектов” (Гонсалес, Вудс, 2012). Например, если значение интенсивности пикселя превышает пороговое значение – он обозначается как

пиксель объекта, иначе – как пиксель фона, или наоборот, в зависимости от задачи. Порог при этом зачастую выбирается на основе анализа распределения пикселей изображений по яркости (гистограмм). Ниже приведены широко применяемые методы сегментации (Гонсалес, Вудс, 2012):

- Глобальная пороговая обработка (использование одного значения порога для всех пикселей изображения).
- Метод Оцу (Otsu, 1979) Этот метод максимизирует межклассовую дисперсию по интенсивности пикселей для областей.
- Метод с использованием нескольких порогов (аналог метода k -средних) (Fukunaga, 1972).
- Метод переменного порога (разбиение исходного изображения на непересекающиеся прямоугольные области, в каждой из которых для сегментации используется свое значение порога).
- Метод выращивания областей (группирует пиксели или подобласти в более крупные области по заранее заданным критериям укрупнения).
- Метод морфологических водоразделов, рассматривающих изображение как трехмерную поверхность, заданной двумя пространственными координатами и уровнем яркости в качестве высоты поверхности (рельефа) и основанном на поиске границ водоразделов для такой топографической области.

Следует отметить, что указанные алгоритмы можно обобщить для работы с различными цветовыми компонентами изображений для их сегментации по цвету.

Применение описанных методов зависит от поставленных задач, требований к особенностям программного обеспечения и вычислительных ресурсов. В некоторых задачах изображения получены в лабораторных

условиях, имеют высокое качество и низкий уровень шума и могут быть успешно обработаны достаточно простыми алгоритмами.

Важным вопросом в связи с анализом изображений является оценка точности работы алгоритмов сегментации. Для этого обычно используется тестовая выборка изображений, на которых области размечены вручную, т.е. для каждого пикселя изображения указана его принадлежность к одному из нескольких классов. Такие изображения считаются стандартом и в идеале автоматический анализ должен в точности воспроизвести классификацию пикселей на изображении. Однако автоматический анализ неизбежно содержит ошибки, поэтому необходимо оценить степень согласия между стандартом и автоматической классификацией.

Поэтому после обработки изображений происходит сравнение принадлежности пикселей к различным классам и в качестве меры точности используется коэффициент Жаккара (известен также аналог этой меры – IoU, Intersection over Union) (Tan, 2005):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (1.2)$$

здесь множество A – пиксели области изображения, полученной в результате сегментации с помощью разработанного алгоритма и заданных значений его параметров; множество B – пиксели области изображения, маркированные вручную. Если классификация пикселей автоматическая совпадает с ручной, то эта мера равна 1. Чем выше различия, тем ближе эта мера к 0 (рис. 1.3).

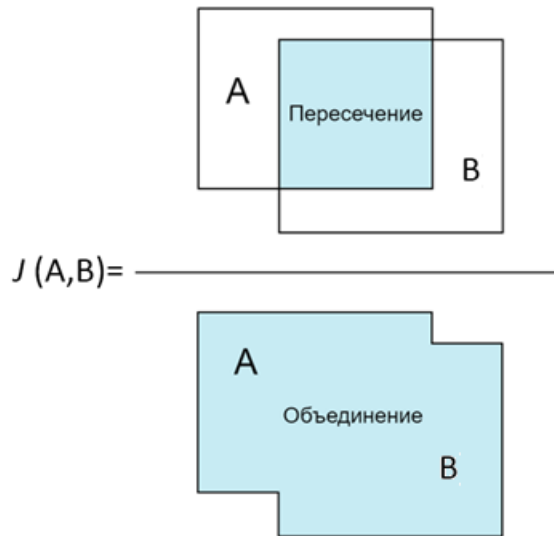


Рисунок 1.3. Иллюстрация оценки коэффициента Жаккара.

Другой задачей оценки точности алгоритмов распознавания объектов на изображении при их подсчете является сравнение их количества на изображении. Такая задача возникает, например, при подсчете зерен на изображении. Для оценки точности алгоритма в этом случае используют сравнение количества объектов на изображениях тестовой выборки с количеством объектов, оцененным автоматически. Удобно использовать две меры оценки точности для таких задач: среднюю абсолютную ошибку MAE (Mean Absolute Error) и среднюю абсолютную процентную ошибку (Mean Absolute Percentage Error), которые вычисляются по формулам:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{i=N} \frac{|n_i - n'_i|}{n_i}, \quad (1.3)$$

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{i=N} \left(\frac{|n_i - n'_i|}{n_i} \right), \quad (1.4)$$

где N – число изображений в анализируемой выборке, n_i – число объектов,

определенных на изображении i вручную, n'_i – число объектов, определенных автоматически. Чем больше значение параметров MAE и MARE, тем выше ошибка определения числа объектов на изображении. При точном совпадении количества распознанных объектов эти меры равны 0. MARE удобна тем, что позволяет оценить относительную ошибку метода распознавания (и не зависит от количества объектов на изображении).

Другим способом оценить точность количественного распознавания объектов на изображении служит коэффициент корреляции Пирсона между предсказанным и истинным количеством объектов для тестовых изображений.

Представление объектов и морфометрия

После выделения объектов на изображении встает вопрос об оценке и численном представлении их размера и формы. Эта задача является важной, так как позволяет отобразить свойства биологических объектов в числовом пространстве, что позволяет провести дальнейшее сравнение объектов, их классификацию.

Один из примеров представления объектов – метод прослеживания их границ (Гонсалес, Вудс, 2012). В этом случае задается направление обхода границы объекта, движение осуществляется попиксельно и на каждом шаге составляется упорядоченный список пикселей границы. Это движение можно закодировать, сопоставив каждому шагу число – направление движения из предыдущего пикселя в последующий согласно 4- или 8-связности. Такая последовательность называется цепным кодом Фримена. Такое представление объектов удобно для использования в последующих алгоритмах сравнения и классификации. Например, сходный алгоритм был использован, в частности,

для описания формы листьев и их классификации по видам растений в методе LeafSnap (Kumar et al., 2012).

Однако для биологов такое представление не всегда будет понятным. Обычно более предпочтительны интуитивно понятные характеристики размера и формы объекта, такие как площадь на изображении, вытянутость, изрезанность границы и т.п. Такие числовые характеристики объекта называют дескрипторами (следует иметь ввиду, что не все дескрипторы могут иметь какое-либо биологическое значение).

Одним из простейших дескрипторов границы объекта (контура) является ее длина (периметр объекта). Ее можно оценить, как количество пикселей, формирующих границу объекта.

Для связной области можно оценить ее площадь в пикселях. Другим примером дескриптора формы является коэффициент округлости, который определяется как отношение площади области к площади круга (максимально компактная фигура) с таким же периметром (Гонсалес, Вудс, 2012).

Геометрические дескрипторы объектов будут использованы в настоящей работе для описания формы зерен и колоса. Они будут детально описаны в соответствующих разделах диссертации.

Анализ цвета.

Цвет поверхности - характеристика ее спектральной отражательной способности, которая является результатом действия многих факторов, таких как поглощение излучения от источника света на разных длинах волн, его отражение и рассеяние (Форсайт, Понс, 2004). Наиболее полно характеристики отражения или поглощения излучения в разных интервалах длин волн могут быть оценены спектрометрами. В большинстве же обычных цифровых камер используются матрицы, реагирующие на отраженное излучение в диапазоне

видимых глазу длин волн (400-780 нм). Следует отметить, что восприятие цвета человеческим глазом имеет свои особенности, связанные со строением зрительного аппарата: между восприятием цвета поверхности глазом и ее спектральными характеристиками не существует однозначного соответствия. Например, один и тот же оттенок серого цвета может быть воспроизведен отраженным излучением с совершенно различными интенсивностями для разных длин волн.

При изучении человеческого восприятия цвета было выяснено, что достаточно трех основных цветов, красного, зеленого и синего, чтобы при их смешивании в различных пропорциях получить все остальные воспринимаемые человеком цвета (Форсайт, Понс, 2004). Данный вывод подтверждается строением человеческого глаза, т.к. человеческая сетчатка глаза имеет три вида рецепторов (колбочек), ответственных за цветное зрение.

Для цифрового представления цветов были разработаны различные цветовые модели (цветовые пространства). Цветовая модель задает систему координат, в которой однозначно определяются цвета. Существует целый ряд различных цветовых моделей, которые были разработаны для наилучшего способа описания цвета в таких системах как телевидение, фото и видео камеры, цветная печать. Среди из наиболее часто использующихся в анализе цифровых изображений растений необходимо отметить следующие.

Цветовая модель RGB - наиболее известное цветовое пространство, в котором цвет кодируется относительной интенсивностью трех его компонент, красной (R), зеленой (G) и синей (B) (рис. 1.4). Эти компоненты описываются целыми числами, чаще всего от 0 до 255. Чем выше значения, тем выше интенсивность цвета. Цвета с одинаковыми значениями компонент представляют собой оттенки серого цвета. Такое представление используется в основном в мониторах и цифровых камерах.

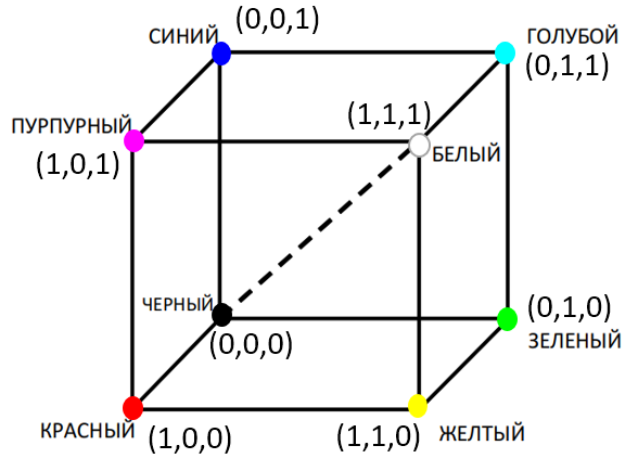


Рисунок 1.4. Иллюстрация представления цвета в пространстве (R,G,B).

Модель HSV (HSB) – цветовое пространство, предложенное в середине 1970-х и использующее три компоненты цвета (рис. 1.5 А). Цветовой тон (Hue) варьирует от 0 до 360 и определяет цветность (значения близкие к 0 или 360 соответствуют красному цвету, 60 – желтому, 120 - зеленому, 180 - голубому, 240 - синему, 300 – малиновому). Насыщенность (Saturation) имеет тем большие значения, чем насыщенней оттенок цвета, а малые значения параметра соответствуют оттенкам серого. Яркость (Value/Brightness) принимает меньшие значения для темных цветов, большие – для ярких. Один из недостатков цветовых моделей HSV и RGB заключается в том, что количество различимых глазом уровней насыщенности и оттенков в этих цветовых пространствах уменьшается при приближении яркости (V) к нулю.

Пространство CIE $L^*a^*b^*$ предложено в 1976 году Международной Комиссией по Освещению (International Commission on Illumination, CIE) для обеспечения более однородного различия в цвете в отношении восприятия человеком (рис. 1.5 Б). Как и в HSV, в CIE $L^*a^*b^*$ значение яркости (компонента L^*) отделено от хроматической составляющей цвета (Pathare et al.,

2013). Параметр L^* является приблизительной оценкой яркости. Параметр a^* принимает положительные значения для красноватых цветов и отрицательные значения для зеленоватых, тогда как параметр b^* принимает положительные значения для желтоватых цветов и отрицательные значения для голубоватых. Цветовая модель CIE $L^*a^*b^*$ нашла широкое применение в программном обеспечении для обработки изображений и цветокоррекции. Эта модель используется для результатов оценки характеристик цвета в спектрофотометрах.

Пространство YCbCr (рис. 1.5 В) применяется в компонентном видео и цифровой фотографии в различных алгоритмах встраивания информации в изображения, а также алгоритмах сжатия изображений. Канал Y характеризует интенсивность (яркость) света. Каналы Cb, Cr – цветоразностные каналы, несущие в себе информацию о цвете: Cr - хроматический красный; Cb - хроматический синий (Фисенко и Фисенко, 2008).

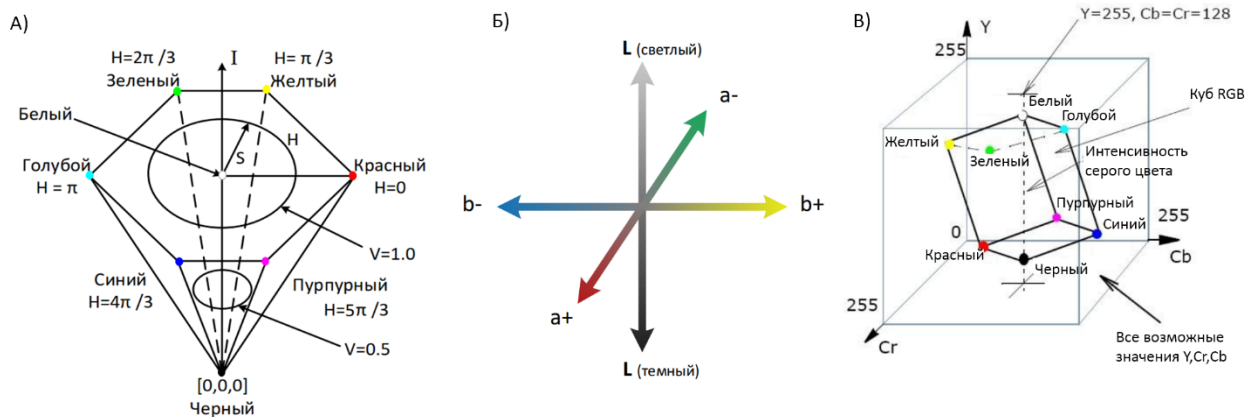


Рисунок 1.5. Иллюстрация представления цвета в различных цветовых пространствах: А) HSV; Б) $L^*a^*b^*$; В) YCrCb.

Существует также целый ряд других цветовых пространств, их описание и характеристики можно найти в специальной литературе по анализу изображений (Домасев и Гнатюк, 2009; Фисенко и Фисенко, 2008).

Компоненты для одного и того же цвета в различных системах связаны правилами преобразования, так что зная значения цветовых компонент цвета в одной системе всегда можно получить значения компонент этого цвета в других цветовых пространствах. Например, из значений компонент RGB можно получить значения цветовых компонент HSV и наоборот. Все это позволяет в зависимости от задачи выбрать то или иное представление цвета изображения.

В задачах компьютерного зрения и анализа цветовых характеристик наибольший интерес представляет собой такие цветовые модели как HSV и $L^*a^*b^*$, в которых цвета представляются с точки зрения таких же понятий, которыми оперирует человек при описании цвета: цветовой тон, насыщенность, яркость (светлота).

Распознавание образов и методы машинного обучения

Распознавание образов – это идентификация и классификация объектов, которые характеризуются конечным набором некоторых свойств и признаков. Задача распознавания образов является одной из основных в большинстве интеллектуальных систем, таких как машинное зрение (Гонсалес, Вудс, 2012). Цель таких систем состоит в получении изображения и составление его описания. В целом задачу распознавания можно разделить на два этапа. Первый это детектирование (обнаружение) объектов на изображении. Вторым этапом – это классификация.

Задача детектирования образа состоит в определении всех областей изображения, возможно содержащих детектируемый объект.

Задача классификации состоит в определении, к какому конкретному классу принадлежит обнаруженный объект. Роль исходной информации для классификации образа играет понятие вектор признаков. Признак – это

некоторое качественное или количественное измерение распознаваемого объекта произвольной природы. Вектор признаков – это множество признаков одного образа. В рамках задачи распознавания считается, что каждому образу ставится в соответствие единственное значение вектора признаков и наоборот: каждому значению вектора признаков соответствует единственный образ.

Классификация в системах распознавания основывается на образах, правильная классификация которых известна. Можно считать, что все объекты или явления разбиты на конечное число классов. Задача распознавания, на конечном этапе, состоит в том, чтобы отнести новый распознаваемый объект к какому-либо классу (Местецкий Л. М., 2008).

Классификатором или решающим правилом классификации называется правило отнесения образа к одному из классов на основании его вектора признаков. Практическая разработка системы классификации осуществляется по схеме решения следующих задач:

1. Задача генерации признаков – выбор признаков, описывающих образ наиболее полно.
2. Задача селекции признаков – отбор наиболее эффективных в классификации признаков.
3. Задача построения классификатора – выработка правила классификации, по которому на основании вектора признаков будет осуществляться классификация.
4. Задача оценки системы – выбор правила для количественной оценки правильности системы классификации.

Задача классификации нередко решается с помощью методов машинного обучения, так как число признаков и параметров влияния их на конечный результат зачастую достигает нескольких десятков или сотен (Sarker, 2021).

Алгоритмы машинного обучения обычно используют и обрабатывают данные для выявления и изучения паттернов связи между данными (объектами, изображениями, индивидуумами, бизнес-процессами, транзакциями, событиями и т.д.). Алгоритмы машинного обучения в основном делятся на четыре категории: обучение с учителем, обучение без учителя, обучение с частичным привлечением учителя и обучение с подкреплением.

Обучение с учителем, как правило, является задачей машинного обучения по определению функции, которая сопоставляет входные данные с выходными на основе выборочных пар ввода-вывода. Оно использует помеченные обучающие данные и набор обучающих примеров для вывода функции. Наиболее распространенными задачами этого класса являются “классификация”, которая разделяет данные, и “регрессия”, которая выявляет соответствие между данными.

Обучение без учителя сводится к анализу неразмеченных наборов данных без необходимости вмешательства человека. Этот подход широко используется для извлечения генеративных признаков, выявления значимых тенденций и структуры данных, группировки данных. Такими задачами являются кластеризация, оценка плотности, изучение признаков, уменьшение размерности, поиск правил ассоциации, обнаружение аномалий и т.д.

Обучение с частичным привлечением учителя можно определить, как гибридизацию вышеупомянутых методов, поскольку оно работает как с помеченными, так и с немаркированными данными. Некоторые области применения, в которых используется этот подход: машинный перевод, обнаружение мошенничества, маркировка данных и классификация текста.

Обучение с подкреплением - это тип алгоритма машинного обучения, который позволяет программному обеспечению в виде агентов и машин

автоматически оценивать режим оптимального поведения в конкретном контексте для повышения его эффективности.

В фенотипировании растений наиболее часто используются первый и второй подходы (распознавание/классификация/регрессия и кластеризация объектов). В задачах классификации требуется разбить все объекты на два (истина/ложь) или более классов. В некоторых случаях классификации на множество меток объектам может быть присвоено несколько классов одновременно.

Для решения этой задачи используются такие алгоритмы как наивный Байесовский метод (John, Langley, 1995), линейный дискриминантный анализ (McLachlan, 2004), метод решающих деревьев (Quinlan, 1986), логистическая регрессия (LeCessie, Van Houwelingen, 1992), метод k -ближайших соседей (Aha et al., 1991), машины опорных векторов (Keerthi et al., 2001), случайный лес (Breiman, 2001), а также нейронные сети (Горбань, Россиев, 1996).

При решении задачи классификации на обучающей выборке, для которой заранее известны отношения между классами объектов и их признаками, алгоритмы машинного обучения подбирают параметры на основе минимизации значения ошибки классификации. Для оценки точности алгоритма используется тестовая выборка объектов, для которых известна принадлежность к определенным классам. Для объектов из тестовой выборки с помощью алгоритма машинного обучения производится классификация, которая сравнивается с известной.

Наглядным представлением качества предсказания является матрица ошибок. Это матрица размером $N \times N$, используемая для оценки эффективности модели классификации, где N - количество целевых классов. Ячейки матрицы i, j , содержат количество объектов для которых алгоритм предсказал класс i , а истинным классом объекта является j . Эта матрица дает

представление о том, насколько хорошо работает классификационная модель и какие ошибки она допускает. Для бинарной классификации это матрица 2×2 , показанная на рисунке 1.6. Для нее оценивается 4 характеристики ошибок: *tp*- истинное положительное предсказание (true positive), *tn*- истинно отрицательное предсказание (true negative), *fp* – ложное положительное предсказание (false positive), *fn* – ложное отрицательное предсказание.

		Истинные значения класса	
		1	0
Предсказанные значения класса	1	tp	fp
	0	fn	tn

Рисунок 1.6. Матрица ошибок для бинарной классификации.

Зная указанные выше параметры можно оценить такие меры качества предсказания как точность (англ. термин *precision*), полноту (англ. термин *recall*) и общую меру точности, *F*-меру (van Rijsbergen, 1979).

$$pr = \frac{tp}{tp+fp}, \quad (1.5)$$

$$re = \frac{tp}{tp+fn}, \quad (1.6)$$

$$F = 2 \frac{pr \cdot re}{pr+re}. \quad (1.7)$$

Чем выше *F*, тем меньше ошибок в идентификации объектов на изображении. В случае если все объекты идентифицированы правильно и не идентифицировано ни одного лишнего объекта *F*-мера равна 1.

1.3. Фенотипирование растений

Фенотипирование корней

Корневая система высших растений - основной поставщик питательных веществ из почвы. При этом архитектура корневой системы крайне вариабельна как между видами, так и у генетических форм одного вида (Osmond et al., 2007). Особенности структуры корневой системы культурных растений влияют на их устойчивость к стрессовым условиям и продуктивность. Ключевые морфологические признаки можно разделить на макроскопические (размеры, топология корневой системы) и микроскопические (особенности корневых волосков, тканей корня).

К основным можно отнести следующие макроскопические признаки (Lynch, 2007): длина главного корня у стержневой системы; диаметр корня и его кончика; размер и количество боковых корней; разветвлённость; наличие, размеры и положение придаточных корней. К микроскопическим признакам относятся, прежде всего, количество, размеры и положение корневых волосков, характеристики эпидермальной ткани.

Особенностью корневой системы является трудность ее исследования неразрушающими методами (без удаления растения из почвы), тем более, в полевых условиях. В связи с этим множество физиологических, генетических и селекционных работ ведутся в контролируемых лабораторных условиях. При этом растения, как правило, выращиваются на гидропонике.

Типичным подходом для наблюдения за ростом корней является использование прозрачной агаризованной среды и выращивание в плоских чашках, расположенных вертикально. В таких условиях корни растут по поверхности среды и доступны для непрерывной неинвазивной двумерной визуализации. Например, для изучения влияния сахарозы на рост корней растений *A. thaliana* была разработана автоматизированная система PlaRoM

(Plant ROot Monitoring platform) (Yazdanbakhsh, Fisahn, 2009). Чашки снимаются камерой в ИК диапазоне, из серий двумерных изображений корней при помощи компьютерной программы извлекаются данные о скорости их роста. Этот подход показал высокую точность, позволяя оценивать временную динамику скорости роста корней в течение дня, а также топологию и геометрию корней, размеры зоны корневых волосков, их количество.

Существуют технологии эффективной системы выращивания растений в плоских ёмкостях наполненными землёй («ризотронах») с контролируемыми условиями среды. Визуализация производится в видимом или ИК-спектре сквозь прозрачную стенку ёмкости. Подобная автоматизированная система массового выращивания и анализа изображений GROWSCREEN-Rhizo была использована для риса *O. sativa* и арабидопсиса *A. thaliana*. Несмотря на то, что часть корней в такой системе оказывается скрытой в почве и недоступна для анализа, по видимой части корневой системы можно достаточно точно получить ее полные характеристики (Nagel et al., 2012). В этой работе было продемонстрировано, что увеличение плотности почв приводит к уменьшению длины основного корня и уменьшению длины латеральных корней. Возможности системы GROWSCREEN-Rhizo не ограничиваются оценкой характеристик корня, а позволяют также измерять размеры побегов и листьев. Это позволило количественно оценить уменьшение размеров листьев и скорости роста побега у растений ячменя, произрастающих на более плотной почве.

Для детальной оценки характеристик формы и размеров корневой системы разрабатываются методы получения и анализа объемных изображений. В связи с этим активно применяется выращивание растений в прозрачной среде с дальнейшим получением серии снимков под разными углами. Для описания формы и размеров корня можно использовать

характеристики проекции корня на нескольких двумерных изображениях (Iyer-Pascuzzi et al., 2010). Однако наиболее информативна реконструкция объёмной модели корневой системы на основании большой серии двумерных изображений (Torr et al., 2013). Такой метод позволяет более точно оценить топологию корней, форму, размеры и геометрию области, занятой корневой системой, точно рассчитать углы между корнями, толщину и объём отдельных корней.

При выращивании растений кукурузы, гороха, картофеля в полевых условиях для фенотипирования корневой системы растение извлекается из почвы, очищается от неё, затем проводится либо его визуальная оценка, измерение при помощи специальной подложки, размеченной размерными рисками и углами, либо съёмка при помощи цифровой камеры с последующей цифровой обработкой (Bucksch et al., 2014). Недостатком методов полевого фенотипирования корней является то, что они требуют извлечения растения из земли, его гибели и из-за этого не позволяют проводить эксперименты с отслеживанием изменения параметров корня в процессе роста растения.

Фенотипирование листьев

Листья являются основным фотосинтетическим органом растений, выполняют такие важные функции, как дыхание, транспирация и гуттация. Посредством листьев растения взаимодействуют с окружающей средой над поверхностью почвы. У цветковых растений листья демонстрируют большое разнообразие. При этом внешние характеристики листа (размер, форма, опушение и т. д.) являются одними из базовых для видовой идентификации растения. Морфологические характеристики листа служат параметрами для видовой классификации растений в системе Leaf Snap (Kumar et al., 2012). С помощью этой системы пользователь может проанализировать изображение

листа, полученного с помощью камеры мобильного телефона и определить, к какому виду относится растение.

Анализ формы и размеров листа по изображению можно оценивать полуавтоматически, с использованием таких пакетов как ImageJ. Так, с помощью скриптов пакета ImageJ в работе Читвуда и соавт. (Chitwood et al., 2014) было проанализировано более 1200 генотипов винограда, произрастающих в разных условиях с целью установить влияния генетических факторов, терруара (уникальных свойств климата, почв) на формирование морфологии листа. На основе геномных ассоциаций были выявлены полиморфизмы, влияющие на форму и размер листа винограда. Для анализа листьев разрабатываются и специальные программные пакеты. Например, программа LAMINA на основе изображения сорванного листа на светлом фоне позволяет оценить его длину, ширину, площадь, округлость, симметричность и т.п. (Bylesjö et al., 2008). Аналогичные задачи позволяют решить программы LEAFPROCESSOR (Backhaus et al., 2010) и LeafJ (Wang et al, 2013). В работе Ванга и соавт. (Wang et al, 2013) данные о форме листа (его контур) извлекаются на основе фотографий, сделанных в полевых условиях на природном фоне.

Важная задача при фенотипировании листа – выявление сегментов листовой пластинки, поврежденных различными заболеваниями или вредителями. В случае легко различимых видимых поражений используется анализ цветных изображений, например, при выявлении антракноза у табака, парши яблони (Wijekoorn et al., 2008), угловой пятнистости хлопка (Camargo, Smith, 2009). В случае слабо различимых человеческим глазом заболеваний используются мультиспектральные сенсоры. Они эффективны для определения фузариоза пшеницы (Bauriegel et al., 2011) или фитофтороза томатов (Wang et al., 2013).

Фенотипирование соцветий, семян и плодов

Одними из важнейших, непосредственно влияющих на продуктивность урожая культурных растений являются репродуктивные органы: соцветия, семязачатки, семена, плоды и др. Зачастую качественные и количественные признаки именно данных органов определяют потенциальную целесообразность выращивания того или иного сорта сельскохозяйственных культур.

В работе Аль-Тама и соавт. (Al-Tam et al., 2013) предложен метод P-TRAP (Panicle TRAIit Phenotyping) фенотипирования соцветия и зерен у риса на основе двумерных изображений, полученных на фоне белого листа бумаги. Соцветие риса представляет собой метелку из основной ветки, вторичных и третичных ветвей. Программа распознает узлы ветвления на соцветии, а также кончики метелки и на основе этих данных формирует описание соцветия в виде графа. Программа рассчитывает 15 параметров, характеризующих соцветие и 9 параметров, характеризующих зерна риса, имеет удобный пользовательский интерфейс, позволяющий анализировать соцветие в интерактивном режиме.

Кроуэлл и соавт. разработали пакет PANorama (Crowell et al., 2014), позволяющий оценивать характеристики органов растений, которые имеют ветвящуюся структуру. Помимо анализа соцветия риса в работе продемонстрировано применение этого пакета для изучения соцветия кукурузы, формы листа томата, структуры побега арабидопсиса.

Задача определения формы и размера семян у растений на основе анализа цифровых изображений привлекает к себе внимание еще с 80-х годов прошлого столетия (Sapirstein et al., 1987). Как правило для анализа используются изображения семян на светлом фоне, сделанные фотокамерой или сканером. Результаты компьютерной морфометрии используются для

определения формы и размеров зерен (Smykalova et al., 2013; Whan et al., 2014), предсказания на этой основе характеристик урожайности растений (Novaro et al., 2001), их классификации (Zapotoczny et al., 2008).

В работе Дуана и соавт. (Duan et al., 2011) представлена высокопроизводительная система для анализа признаков риса, связанных с урожайностью. Эта система автоматически определяет число зерновок у растения, извлекает зерна и определяет их длину и ширину, массу 1000 зерен и др. характеристики. За сутки работы система позволяет обработать ~1400 растений.

Методы анализа изображений позволяют достаточно эффективно оценивать спелость плодов растений. Например, система TATC (Tomato Analyzer Color Test) производит сканирование среза плодов фруктов и овощей и даёт точную количественную оценку цвета и его однородности (Darrigues et al., 2008). Подсчет плодов также можно осуществлять на основе цифровых изображений, полученных непосредственно в местах их произрастания (Lu, Sang, 2015).

Для фенотипирования плодов развиваются методы автоматизированного анализа изображений, полученных с помощью сенсоров. Например, метод реконструкции виноградных гроздей на основе данных лазерного дальномера (Schöler, Steinhage, 2015) позволяет воссоздавать 3D-модель грозди с точностью до 0.025 мм. После воссоздания 3D-модели она может быть проанализирована для получения необходимых данных о фенотипических признаках плодов винограда.

Анализ цвета оболочки злаков

Цвет оболочки зерен злаков является важным признаком, который характеризует содержащиеся в ней пигменты и метаболиты. Одна из

актуальных задач анализа изображений зерен - это классификация. Задачи классификации в отношении зерен могут быть разными. Например, необходимо классифицировать зерна по признакам цвета/текстуры на несколько различных генотипов (Pourreza et al., 2012; Olgun et al., 2016). Часто к параметрам цвета и текстуры при классификации добавляются характеристики размера и формы (Majumdar, Jayas, 2000; Chaugule, Mali, 2014; Sabanci et al., 2016).

С этой же задачей тесно связана задача сортировки зерен по цвету и текстуре (Pearson, 2010), в частности, для массового анализа большого количества зерен разрабатываются сортеры, позволяющие отделить здоровые зерна от мусора и поврежденных зерен. Работам в области оценки качества зерен на основе анализа изображений, включая и гиперспектральные, посвящен обзор Хуанга и соавт. (Huang et al., 2015). В обзоре Гонга и соавт. представлены, в том числе, и инженерные подходы для решения задачи определения качества зерна (Gong et al., 2015).

Иногда рассматривают классификацию зерен только по их типу окраски (красная или белая). Так, в работе Рам и соавт. с помощью спектрофотометра и спектрометра построена процедура для определения цвета оболочки зерна у краснозерных и белозерных сортов пшеницы (Ram et al., 2002). В работе МакКейга и соавт. была проведена классификация на краснозерные и белозерные сорта с помощью спектрометрических данных для 262 генотипов как мягкой, так и твердой пшеницы (McCaig et al., 1993).

Еще одна задача, которая решается на основе анализа цветовых характеристик: определение зерен, пораженных патогенами (Ahmad, 1999; Goriewa-Duba et al., 2018), или поврежденных механически (Delwiche et al., 2013).

Фенотипирование побегов

Биомасса растения - один из наиболее важных показателей его развития. Однако растение приходится удалять из почвы, чтобы определить его живую или сухую массу. Это ограничивало возможности оценки массы растения в процессе его роста. Методы высокопроизводительного фенотипирования позволили успешно решить задачу оценки биомассы растения на основе неинвазивной технологии анализа изображений побега.

Оценку биомассы арабидопсиса осуществляют по площади розетки на изображении, полученном сверху (De Vuylder et al., 2012). Помимо биомассы, методы анализа изображений позволяют также оценивать различные характеристики формы розетки (компактность, количество листьев), динамику роста биомассы и т.п. В некоторых работах информация о видимом изображении розетки комбинируется с другими фенотипическими параметрами, такими как содержание хлорофилла, оцениваемое по флюоресценции (Jansen et al., 2009) или вес растения вместе с почвой (Tisné et al., 2013). Все это дает возможность изучать комплексный ответ растения на стрессовые факторы, например, засуху.

У злаковых растений для оценки биомассы и размеров побега производится несколько цифровых снимков как сверху, так и сбоку. В работе Голзарян и соавт. (Golzarian et al., 2011) для оценки сухой биомассы растения ячменя и пшеницы вычислялась площадь проекции побега на изображениях сверху и сбоку. Авторы протестировали несколько функций аппроксимации (линейную, квадратичную, кубическую и экспоненциальную) для того чтобы связать значения площади проекции побега, возраста растения и его сухой массы. Наилучшую аппроксимацию сухого веса получили линейной функцией. Голзарян и соавт. также показали, что стрессовые условия (засоление) могут существенно смещать оценку биомассы на основе площади

проекции побега, однако учет возраста растения позволяет устранить такое смещение.

1.4. Интеграция данных в феномике растений

В процессе высокопроизводительного фенотипирования экспериментатор может получать огромное количество самых разнородных данных (характеристики фенотипа, генотипа, окружающей среды), часть из которых измеряется в реальном времени (например, погодные условия). Для формализации, хранения и быстрого доступа к фенотипической информации используются базы данных.

Одной из полезных функций баз данных является учет и систематизация генотипов (образцов) при проведении экспериментов у растений. Для решения подобных задач была разработана база PlantDB (Exner et al., 2008), которая содержит информацию о генотипах растений, их родословных, результатах генотипирования и рассчитана на использование для работы молекулярно-генетической лаборатории. Обширную информацию по образцам различных растений содержат генетические коллекции (<http://www.ars-grin.gov/>). Однако подобные базы данных не позволяют хранить информацию о фенотипах растений. С развитием технологий фенотипирования создание баз данных, интегрирующих информацию о генотипе и фенотипе становилось все более актуальным. Одной из первых таких баз данных стала GERMINATE (Lee et al, 2005). В этой базе информация содержала несколько функциональных модулей: паспорт генотипа (происхождение, родословную, таксономическую принадлежность), модуль данных (характеристики фенотипа, генотипа, наборы маркеров), информационный модуль (страна и организация пользователя, который внес информацию), модуль структуры базы данных, модуль интеграции (содержит описание доступных генетических коллекций). Работа базы данных была протестирована на таких видах как горох, пшеница,

ячмень, салат и использовалась для работы в нескольких научных институтах. Позднее структура базы данных GERMINATE была адаптирована для описания генотипов и фенотипов у ячменя в системе Hordeum Toolbox (Blake et al., 2012).

При проведении селекционно-генетического эксперимента необходимо также учитывать и условия внешней среды. Примером создания интегрированной системы для описания экспериментов с растениями в различных условиях (в растительных боксах для арабидопсиса) является Phenopsis DB (Fabre et al., 2011). В работах Кёл и соавт. описаны системы для контроля создания генетических линий растений в результате экспериментов по скрещиванию (Köhl et al., 2008) или трансформации (Köhl, Gremmels, 2010), которые учитывают также и условия проведения селекционных экспериментов. Интегрированная система на основе этих баз данных была разработана для контроля проведения экспериментов по селекции картофеля в рамках проекта TROST, включавшего 11 организаций и 11 мест произрастания (Billiau et al., 2012).

Для удобного ввода фенотипических данных в полевых условиях разрабатываются интерфейсы баз данных и мобильных устройств («электронные полевые журналы»). Так, система «PHENOME» для сбора информации по характеристикам томатов в полевых условиях использует карманные компьютеры, а затем синхронизирует данные в общей базе (Vankadavath et al., 2009). Система хранения и сбора фенотипической информации в полевых условиях через мобильные устройства представлена в работе Кёл и Греммельс (Köhl, Gremmels, 2010).

Все эти работы демонстрируют успешное использование информационных технологий (баз данных и систем распределенного сбора

информации) для решения таких комплексных задач, как контроль проведения селекционных и генетических экспериментов в полевых условиях.

1.5. Заключение по обзору литературы и постановка задачи исследования

Для успешного решения задач по идентификации связи фенотипических признаков растений с их генами современные методы геномики предоставляют большой арсенал средств для идентификации генотипа растения. Узким местом при решении этой задачи является определение фенотипических признаков растений, которые до недавнего времени не были автоматизированными и в силу этого, являлись трудоемкими, неточными и низкоэффективными.

Необходимость устранения препятствий для получения широкомасштабных данных для фенотипических признаков растений привело к развитию феномики, бурно развивающейся области науки. Для нее характерны широкая апробация новых методов фенотипирования для решения разнообразных задач, вовлечение новейших компьютерных и инженерных технологий в создание этих методов.

Применение методов феномики позволит существенно улучшить наши представления о фенотипе растений, сформировать новые важные количественные характеристики для его описания и детального исследования связи фенотип-генотип у растений.

Важной задачей является применение разработанных методов для решения конкретных селекционно-генетических задач, связанных с оценкой физиологической роли исследуемых признаков и изучению механизма их генетического контроля.

Дополнительной особенностью феномики растений является большой объем данных как результат применения современных методов фенотипирования. Хранение и обработка этих данных, обеспечение доступа к ним для биологов, обуславливает создание баз данных, интегрирующих информацию о генотипе, фенотипе и окружающей среде для растений. Поэтому одной из немаловажных задач является разработка таких баз данных для исследуемых объектов и признаков.

С учетом этого в рамках настоящей работы была сформулирована задача разработки экспериментально-компьютерной платформы ICGPhenoPlant, Основанной на едином подходе для фенотипирования растений:

1. Разработка протоколов получения изображений органов растений для их последующего использования в определении количественных признаков растений. Эти протоколы должны быть, по возможности, просты в реализации, опираться на недорогие и массовые средства получения изображений, доступны в условиях полевых измерений фенотипа.
2. Разработка алгоритмов анализа изображений для получения количественных оценок признаков растений.
3. Тестирование алгоритмов на независимых выборках изображений для оценки их точности. Оценка факторов протокола, которые влияют на точность алгоритмов распознавания изображений.
4. Создание баз данных для хранения информации о фенотипе растений, их генотипе и условиях окружающей среды, в которой они произрастают.
5. Применение платформы ICGPhenoPlant для решения реальных задач по исследованию связи между генотипом и фенотипом, а также для

оценки влияния факторов окружающей среды на фенотип растений.

Мы разрабатывали нашу платформу для изучения видов растений, которые исследуются в ФИЦ ИЦиГ СО РАН: пшеница, картофель и табак. К числу признаков, которые важны для генетиков, изучающих эти растения относились: опушение листа (играет роль в устойчивости растений к абиотическому и биотическому стрессу), размер, форма и цвет оболочки зерен злаков, морфометрические параметры колоса пшеницы.

Следующие главы диссертации будут посвящены описанию платформы ICGPhenoPlant и ее применению для исследования указанных признаков. Поскольку в работе для каждого из этих признаков был использован разный биологический материал, разные протоколы получения изображений, алгоритмы их анализа, то для целостности восприятия материала мы посвятили каждую из последующих глав анализу одного из признаков растений, указанных выше (опушение листа, характеристики зерен, колоса). Для каждой главы использовали сходную структуру изложения: (1) описание признака и важности его изучения для селекционеров и генетиков; (2) описание биологического материала; (3) описание протоколов получения изображений; (4) описание алгоритмов обработки изображений; (5) оценка точности методов анализа изображений; (6) анализ биологических данных и обсуждение его результатов. Общее резюме по разработанной платформе ICGPhenoPlant будет дано в разделе Заключение.

ГЛАВА 2. ВЫСОКОПРОИЗВОДИТЕЛЬНОЕ ФЕНОТИПИРОВАНИЕ ОПУШЕНИЯ ЛИСТА У РАСТЕНИЙ

2.1. Опушение листа у растений

Структура и функция опушения листьев растений

Выросты эпидермальной ткани - трихомы (ед. ч. трихóм, см. Приложение, Таблица П1) широко распространены среди растений разных таксономических групп (Solereder, 1908, цит. по Мирославов Е.А., 1974; Швачко и соавт., 2020). Они имеют различное строение и функции. По структуре трихомы разных видов растений бывают одноклеточные и многоклеточные, простой формы и весьма замысловатой (рис. 2.1, Швачко и соавт., 2020). Трихомы бывают не секретирующие и секретирующие (депонирующие и/или выделяющие секрет). Первые участвуют в формировании микроклимата у поверхности листа, участвуют в защите от неблагоприятных биотических и абиотических факторов среды.

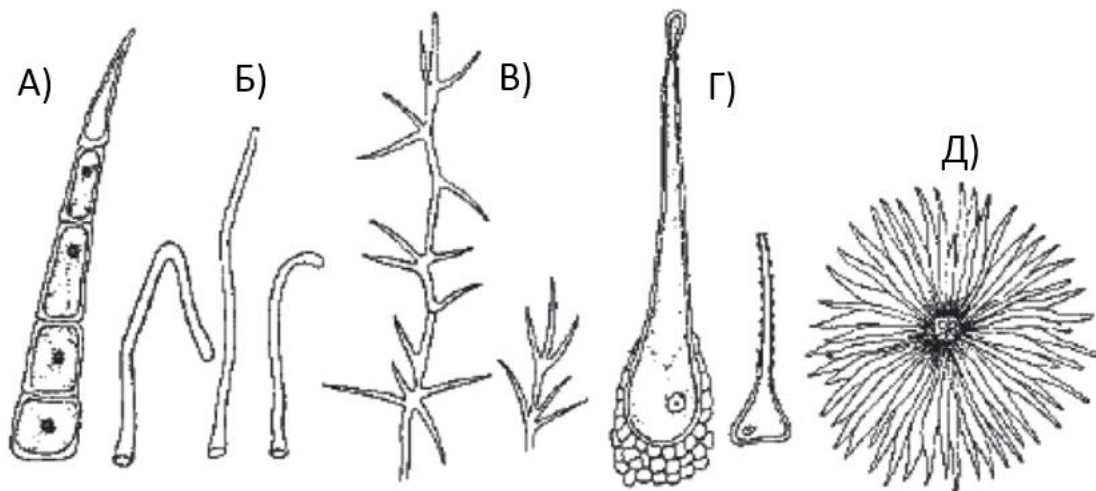


Рисунок 2.1. Типы трихомов: А), в – простые многоклеточные; Б) – простые одноклеточные; В,Г) – железистые; Д) – простые звездчатые. Рисунок взят из работы Швачко и соавт. (Швачко и соавт., 2020).

Исследования роли опушения листа экологической направленности были проведены у различных видов растений. Было обнаружено, что светопоглощение листа было значительно уменьшено у опушённого вида мучнистой энцелии *Encelia farinosa* по сравнению с неопушённым видом *E. californica* (Ehleringer et al. 1976) и коррелировало с количеством осадков в течение сезона. Увеличение степени опушения, тем не менее, не влияло на эффективность фотосинтетического процесса. У этого же вида было показано, что опушение регулирует температуру листа в зимний и в летний периоды, приближая её к оптимальной для фотосинтеза (Ehleringer, Mooney, 1978). Этому способствовало увеличение устойчивости пограничного слоя воздуха у поверхности листа. Развитие трихомов на листе провоцировало водный стресс в летний период роста (Ehleringer, 1982). Испытавшая этот стресс листовая меристема впоследствии порождала густо опушённые листья.

Зависимость эффективности фотосинтеза от опушения была показана на растущих листьях винограда. У опушённого сорта по сравнению с неопушённым наблюдалась более высокая эффективность фотосистемы II и более высокий уровень фотосинтеза (Liakopoulos et al., 2006). Наличие опушения рассеивало избыток световой энергии, и клетки меньше нуждались в клеточной фотозащите в виде выработки большого количества компонентов ксантофиллового цикла. Эти пигменты осуществляют нефотохимическое тушение избыточной энергии, вызывающей повреждения структур клетки. В листьях неопушённого вида дуба, находящихся на солнце, также больше синтезировалось ксантофиллов для обеспечения фотопротективного эффекта, который у опушённого вида обеспечивался наличием трихомов на листьях (Morales et al., 2002). Искусственное удаление опушения приводило к достоверному увеличению синтеза ксантофиллов и снижению активности фотосистемы II. Увеличение приспособленности через изменение морфологии

листа изучалось у растения *Piriqueta caroliniana*. Было обнаружено, что в засушливых условиях самой высокой была скорость роста у растений с более плотным опушением (Picotte et al., 2007). Этот признак прямо влиял на рост растений через повышение эффективности использования воды и накопление биомассы. Таким образом, многие исследования свидетельствуют о важном защитном значении опушения листа, особенно в условиях засухи, при тепловых нагрузках и повышенной солнечной радиации.

Трихомы злаковых участвуют в защите от неблагоприятных биотических и абиотических факторов среды (Scillinger et al., 1968; Roberts et al., 1979; Nameed et al., 2002; Лихенко и др., 2007). Среди злаковых культур, наиболее важных для мирового сельскохозяйственного производства, физиологический эффект опушения листа был наиболее изучен у риса. Линия культурного риса *Oriza sativa* L. с интрогрессией от дикого вида *O. nivata* унаследовала густое опушение (Namaoka et al., 2017). Линия показала ограниченный уровень транспирации и повышенный уровень эффективности использования воды благодаря эффекту увеличения устойчивости пограничного слоя воздуха у поверхности листа.

Особый интерес представляют секретирующие или железистые трихомы. Это специализированные волоски, которые обнаруживаются на поверхности около 30% всех сосудистых растений. Такие трихомы способны синтезировать, хранить и выделять большое количество специализированных метаболитов, многие из которых являются коммерчески важным в качестве лекарственных средств, ароматизаторов, пищевых добавок и природных пестицидов (Wagner, 1991; Duke et al., 2000; Wagner et al., 2004). Многие из железистых трихомов служат для защиты от патогенных насекомых и микроорганизмов (Glas et al., 2012). Трихомы, в том числе и несекретирующие, могут физически препятствовать движениям растительоядных членистоногих

и поеданию ими тканей растения (Cardoso, 2008; Pott C. et al., 2012).

Механизмы развития эпидермальных образований растений представляют интерес как с практической (адаптивный, хозяйственно ценный признак), так и с фундаментальных точек зрения. Трихомы являются удобным объектом для исследования механизмов морфогенеза цветковых растений. Отчасти это связано с их поверхностным расположением и удобством прижизненного наблюдения. Другим достоинством трихомов, в отличие от устьиц и корневых волосков, является их факультативность: мутанты, имеющие рудиментарные, искаженные трихомы или не образующие трихомов, в условиях искусственного выращивания, как правило, жизнеспособны (Hulskamp et al., 1998). Для ряда двудольных растений этот признак достаточно хорошо изучен: известен ряд молекулярных механизмов, обеспечивающих контроль формирования трихомов у двудольных. В существенно меньшей степени изучены молекулярные механизмы формирования трихомов у однодольных.

Опушение листа у пшеницы: морфология, функции и генетический контроль

Для мягкой пшеницы характерно формирование не ветвящихся волосков на поверхности листа. Следует отметить, что у этой культуры существует намного большее разнообразие типов опушения, чем у риса и ячменя (Вавилов, 1990). Эти различия выражаются в плотности расположения трихомов на поверхности листа и их длине. Пример микрофотографий адаксиальной поверхности листьев мягкой пшеницы с различным типом опушения, полученных с помощью сканирующей электронной микроскопии показаны на рис. 2.2.

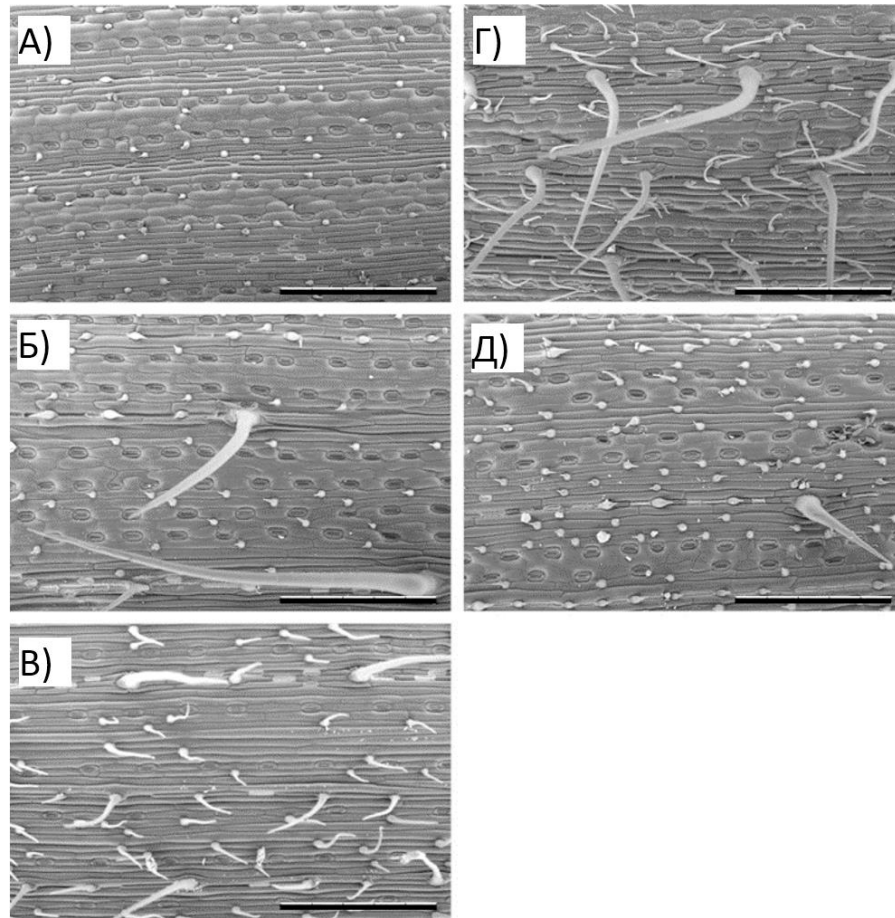


Рисунок 2.2. Микрофотографии адаксиальной поверхности предфлагового листа различных сортов и линий пшеницы, полученные с помощью сканирующего электронного микроскопа. (А) Сорт 'Родина', (Б) почти изогенная линия 102/00ⁱ, (В) сорт 'Саратовская-29', (Г) почти изогенная линия i:C29 *H12^{aesp}*, (Д) почти изогенная линия i:C29 *glabrous*. Масштабная линейка = 500 мкм. Рисунок адаптирован из работы (Doroshkov et al., 2016).

Опушение листа у пшеницы имеет большое биологическое значение в защите от вредителей и при адаптации к факторам среды. В частности, интенсивное, «войлочное», «мохнатое», опушение характерно для ряда засухоустойчивых сортов, относящихся к степной экологической группе; для сортов, произрастающих во влажном климате, напротив, характерно очень слабое, «мохнатое», опушение (Крупнов, Цапайкин, 1990).

В связи с большим разнообразием опушения листа мягкой пшеницы можно предположить, что для этого вида существует сложная генетическая система контроля формирования этого признака. Различные аллельные варианты генов могут влиять на основные процессы формирования опушения: гены дифференцировки клеток эпидермиса, ответственные за формирование паттерна распределения трихомов на поверхности листа; гены ответственные за рост трихомов и обуславливающие соотношение числа трихомов разного размера; гены, влияющие на эндорепликацию клеток трихомов и влияющие на их размер.

Ранее для мягкой пшеницы, был выявлен доминантный ген, определяющий опушение ряда сортов СССР, расположенный на хромосоме 4А (Майстренко, 1976). Этот ген был назван *H11* (Arbuzova et. al., 1996). В 2002 году Такета с соавторами, применив для описания опушения листа методы электронной микроскопии, провели генетический и моносомный анализ этого признака для опушённого китайского сорта ‘Хонг-Манг-май’ (Taketa et. al., 2002). Им удалось выявить доминантный ген, влияющий на формирование опушения листа сорта ‘Хонг-Манг-май’. Этот ген был назван *H12*. При помощи моносомного анализа ген *H12* был локализован в хромосоме 7В. Ген *H12* неаллелен гену *H11*. Для интрогрессивной линии 102/00ⁱ неопушённого сорта ‘Родина’ было показано наличие гена опушения листа в хромосоме 7В (Pshenichnikova et. al., 2007). Таким образом, к началу работы над диссертацией с использованием молекулярных маркеров было локализовано два гена опушения листа. Ген *H11* локализован на длинном плече хромосомы 4В, а ген *H12*^{aesp} картирован на длинном плече хромосомы 7В линии 102/00ⁱ сорта ‘Родина’ с интрогрессией от *Aegilops speltoides* (Dobrovolskaya et. al., 2007).

Влияние известных генов формирования трихомов на количественные характеристики опушения листа мягкой пшеницы и их взаимодействие ранее не было изучено.

Морфология опушения листа у картофеля

У паслёновых (*Solaneae*), к которым относится картофель *Solanum tuberosum* опушение листа представлено широким спектром многоклеточных обычно не ветвистых трихомов разного размера (Luckwill et. al., 1943; Tingey et al., 1982). Например, трихомы томата *S. lycopersicon* отличаются по размеру, морфологии и химическому составу, что свидетельствует о разной функциональной роли волосков. Они бывают как секретирующие, так и несекретирующие, имеющие как одноклеточные основания, так и многоклеточные (Luckwill, 1943).

Основные классы вторичных метаболитов, которые способны производить железистые трихомы пасленовых – это терпеноиды (Gershenzon, Dudareva 2007), фенилпропены (Gang et al., 2001), флавоноиды (Treutter, 2006), метилкетоны (Fridman et al., 2005), ацильные сахара (Kroumova, Wagner 2003). Помимо вторичных метаболитов, трихомы многих видов паслёновых также способны производить значительные количества белков с защитными функциями, такими как ингибиторы протеаз (Liu J. et al., 2006), полифенол-оксидазы (Yu et. al., 1992) и филлопанины (Shepherd et al., 2005).

Для картофеля (безвирусный, сорт ‘Russet Burbank’) отмечено наличие как секретирующих трихомов с головкой из четырех клеток, так и более крупных не секретирующих многоклеточных трихомов (McCauley, Evert, 1988).

При изучении ультраструктуры трихомов картофеля сорта ‘Bintje’ S.

tuberosum установлено наличие двух типов секретирующих трихомов у картофеля (Lyshede, 1980). Один тип имеет восьмиклеточную шаровидную головку на одноклеточной ножке. Другой тип имеет булавовидные многоклеточные головки также на одноклеточной ножке. Короткие железистые трихомы с круглой головкой в англоязычной литературе часто называют «тип А» (Gibson, 1971), показано, что их секрет содержит фенольные соединения (Ave, Tingey, 1986). Более длинные железистые трихомы с булавовидной головкой называют «тип В», их секрет содержит сложные эфиры сахарозы (Neal et al., 1990). Следует отметить, что для большинства культурных сортов картофеля, в отличие от диких видов, характерны трихомы типа А и крупные трихомы без булавовидной головки и не содержащие характерного для трихомов типа В секрета (рис. 2.3; Tingey et al., 1982; Tingey, Laubengayer, 1981; McCauley, Evert, 1988).

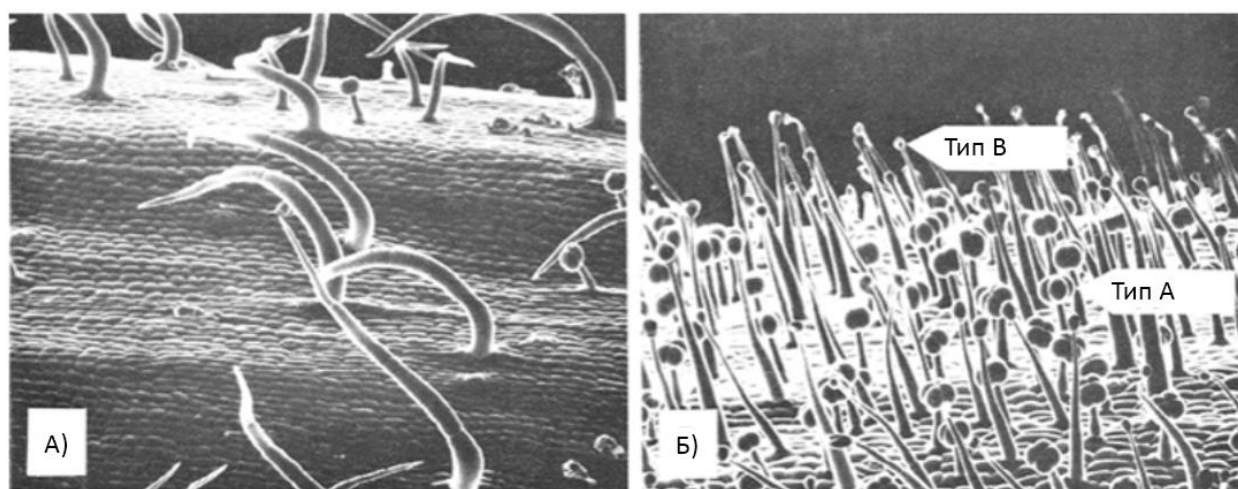


Рисунок 2.3. Опушение культурного картофеля *S. tuberosum* сорт 'Hudson' (А) и его дикого сородича *S. berthaultii* (Б), электронная микрофотография (Tingey et al., 1982). Стрелками указаны два типа трихомов.

Показано, что опушение у культурного картофеля является одним из

важных факторов сопротивления растения насекомым, в частности колорадскому жуку, картофельной блохе, и картофельной цикадке в том числе и в полевых условиях (Gregory et. al., 1986; Flanders et al., 1992; Maharijaya et. al., 2015).

Плотное опушение железистыми трихомами, эффективное в защите от вредителей распространено среди дикорастущих видов картофеля (рис. 2.3) и рассматривается как перспективный признак для интрогрессии в культурные формы (Фасулати и др., 2011).

Методы фенотипирования опушения листьев у растений

Выявление генетических механизмов, контролирующих опушение листа пшеницы затруднено из-за отсутствия эффективных методик количественных оценок его характеристик. Типичными методами оценки опушения листьев являются сканирующая электронная микроскопия (Perazza et. al., 1999; Nagata et. al., 1999; Luo et. al., 1999), визуальный подсчет при помощи оптико-увеличительных приборов (Lacape et. al., 2005; Dobrovolskaya et. al., 2007). К недостаткам этих методов можно отнести низкую точность качественных оценок и трудоемкость методов подсчета, что затрудняет их применение для анализа большого количества образцов. Перспективными в этой связи являются компьютерные методы, основанные на автоматическом анализе цифровых изображений. Так, в работе Bensch и соавт. (Bensch et. al., 2009) разработали алгоритм подсчета трихомов на листовых пластинках *A. thaliana*, который позволяет анализировать расположение трихомов на цифровых изображениях. Методы, основанные на анализе изображений, получают в последнее время все более интенсивное развитие (Failmezger et al., 2013; Pomeranz et al., 2013). Они позволяют существенно ускорить процесс оценки параметров фенотипа, проводить их количественные оценки.

2.2. Протокол получения микроизображений сгибов листа пшеницы

2.2.1. Растительный материал и условия выращивания

Для оценки точности предложенных алгоритмов автоматического анализа изображений листьев пшеницы использовались растения разных генотипов с различающимся типом опушения.

Сорт ‘Родина’ практически не имеет трихомов на поверхности листа. Редко наблюдаются короткие трихомы длиной до 50 мкм.

Сорт ‘Диамант 2’ имеет слабое опушение поверхности листа. Большинство трихомов короче 100 мкм. Некоторые трихомы достигают 300 мкм в длину. Средняя плотность трихомов - 4 шт. на 1 кв. мм.

У растений линии ‘Саратовская 29’ (С29) трихомы на адаксиальной поверхности листа имеют длины в пределах от 25 до 250 мкм, однако из-за наличия редких длинных трихомов средняя длина волосков у этого сорта составляет 321 мкм (Дорошков и соавт., 2011).

Линия 102/00ⁱ была получена в результате скрещивания сорта ‘Родина’ и дикого сородича мягких пшениц, *Aegilops speltoides* Tausch (Lapochkina et al., 2003). Эта линия несет интрогрессированный ген опушения листа (Pshenichnikova et al., 2007) и демонстрирует большое число коротких (не более 75 мкм) и ограниченное число гораздо более длинных трихомов, так, что средняя длина трихомов составляет 84 мкм (Дорошков и соавт., 2011).

Линия 821 получена в результате скрещивания ‘Саратовская 29’ и *Triticum timofeevii* (Leonova et al., 2001) и имеет опушение близкое по характеристикам к тетраплоидному родителю.

Растения выращивались в условиях гидропонной теплицы ЦКП «Искусственного выращивания растений» ИЦиГ СО РАН во время весенних и

осенних вегетаций 2008-2011 годов. Во время выращивания поддерживались условия, включающие температуру, водный режим и длину дня, необходимые для выращивания яровой пшеницы. В качестве субстрата был использован керамзит, в качестве питательного раствора - раствор Кнопа (Shive, 1915).

Подбор и выращивание биологического материала проводились сотрудниками Сектора генетики качества зерна под руководством Пшеничниковой Т.А.

2.2.2. Протокол пробоподготовки и получения изображения сгибов листа

Метод оценки количественных характеристик опущения листа пшеницы основан на анализе микроизображений поперечного сгиба листа. Основные процедуры для получения изображений опущения листа пшеницы заключаются в следующем.

Лист отделяется от растений. На поверхности листа выбирается область, в которой будет анализироваться опущение. Как правило, для оценки характеристик опущения листа пшеницы исследуются три области: у основания листа, в его середине и на конце листа (см. рис 2.4). Для изучения характеристик опущения адаксиальной (верхней) поверхности листа, он складывается поперек адаксиальной поверхностью наружу. Для изучения характеристик опущения абаксиальной (нижней) поверхности листа, он складывается наружу абаксиальной поверхностью. В результате формируется поперечный сгиб листа, который используется для оценки количественных характеристик опущения.

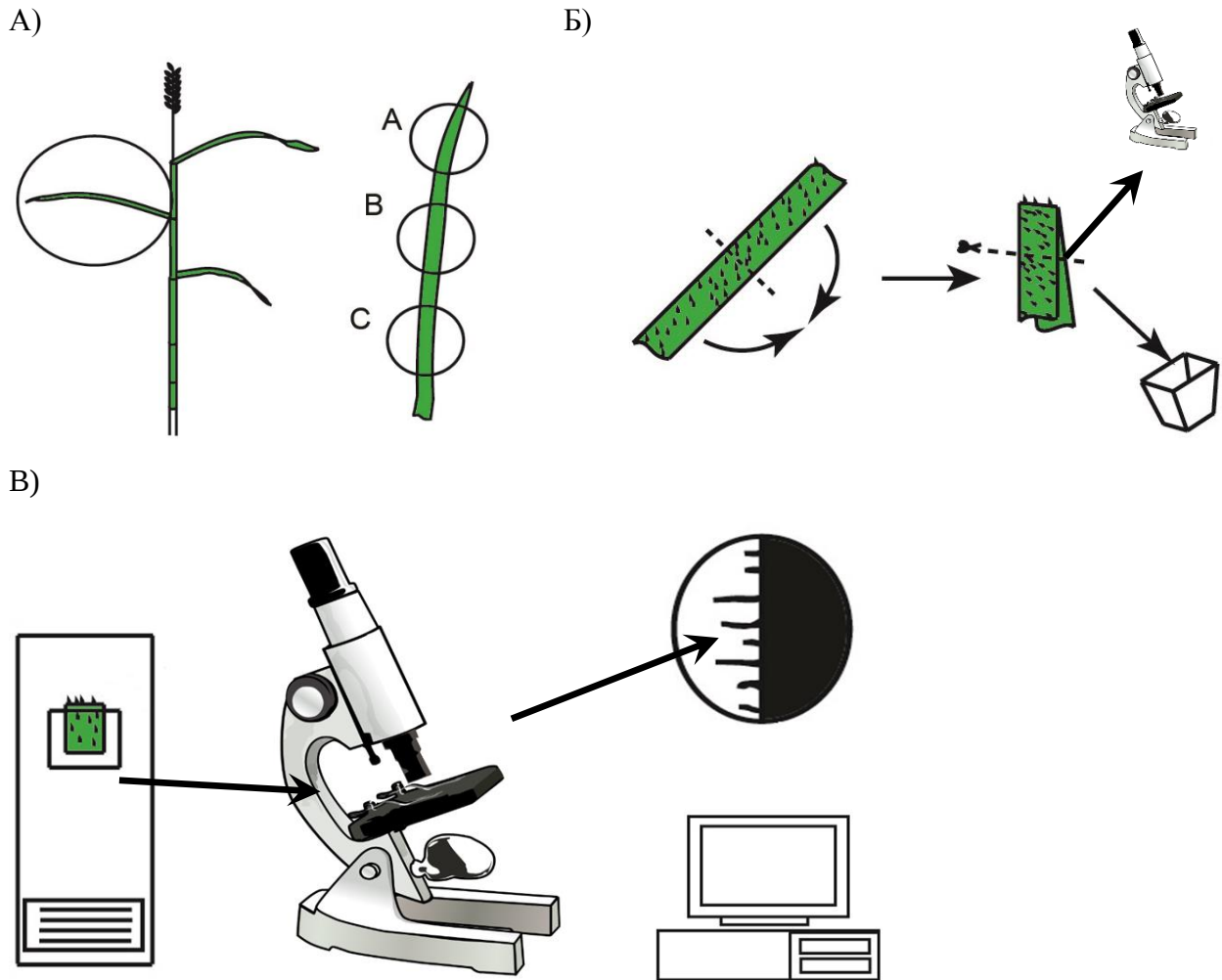


Рисунок 2.4. Протокол пробоподготовки образцов сгибов листа для получения микрофотографий. А) Выбор листа для фенотипирования и области поверхности листа для анализа; Б) сгиб листа и удаление остатков на концах образца; В) размещение согнутого фрагмента на предметном стекле, его фиксация липкой лентой, размещение на предметном стекле микроскопа и получение микрофотографий.

Фрагменты листа на расстоянии 1-1.5 см от сгиба удаляются с помощью ножниц, а оставшаяся часть помещается на предметное стекло микроскопа и закрепляется клейкой лентой с отступом 0.2-0.5 см от сгиба. Предметное стекло с образцом помещается под окуляр светового микроскопа.

Для получения изображений сгиба листа пшеницы использовали Микроскоп ZEISS Axioscop 2 plus с камерой Axioscam HRc (размер CCD

сенсора 2/3 дюйма) и адаптером с увеличением 0.63×. Объектив 5×. При таком масштабе пиксель изображения разрешением 1300 x 1030 соответствует размеру 2.1×2.1 мкм. Аппаратура была предоставлена ЦКП микроскопического анализа биологических объектов ИЦиГ СО РАН.

На полученных изображениях граница листа расположена вертикально, область листа (темная) располагается справа, а область фона – слева (рис. 2.5 Б). Трихомы идентифицируются как выросты, сужающиеся от оснований (место прикрепления к поверхности листа) к вершинам.

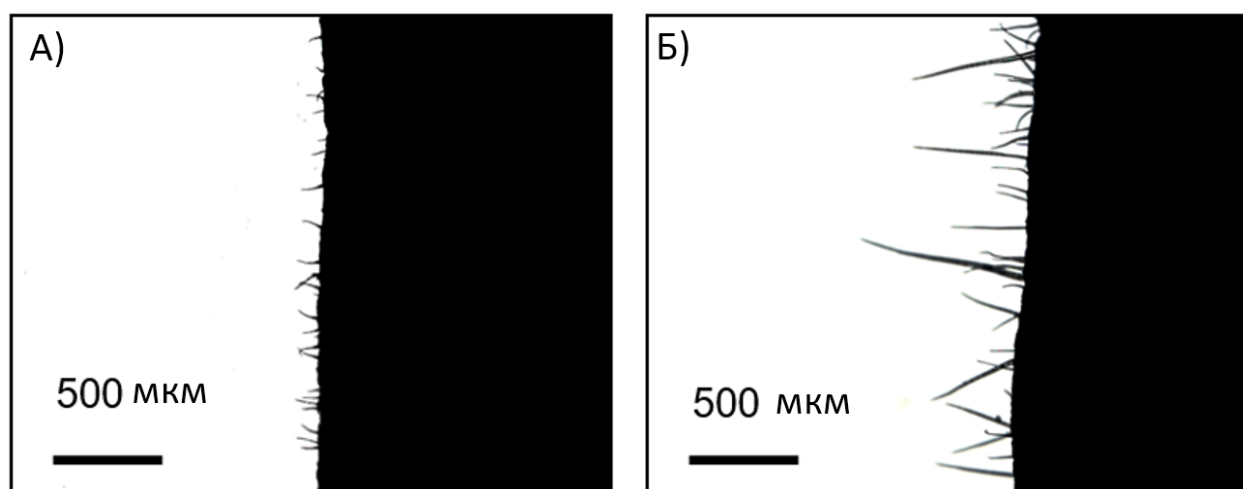


Рисунок 2.5. Пример типичных изображений сгиба листа, полученных в результате протокола, описанного выше. Образцы были взяты для сгибов листа слабо опушенного растения (А) и сильно опушенного растения (Б). Шкала соответствует 500 мкм.

Необходимо отметить, что трихомы, на изображении полученном таким образом, характеризуют опушение небольшого фрагмента листа, соответствующего его торцевой поверхности, обращенной наружу от сгиба. Это хорошо видно из схемы образца сгиба листа, представленного на рис. 2.6. В этом случае зависимость числа трихомов на сгибе листа и плотности

опущения (числа трихомов на единицу площади поверхности) можно описать на основе простой модели (рис. 2.6).

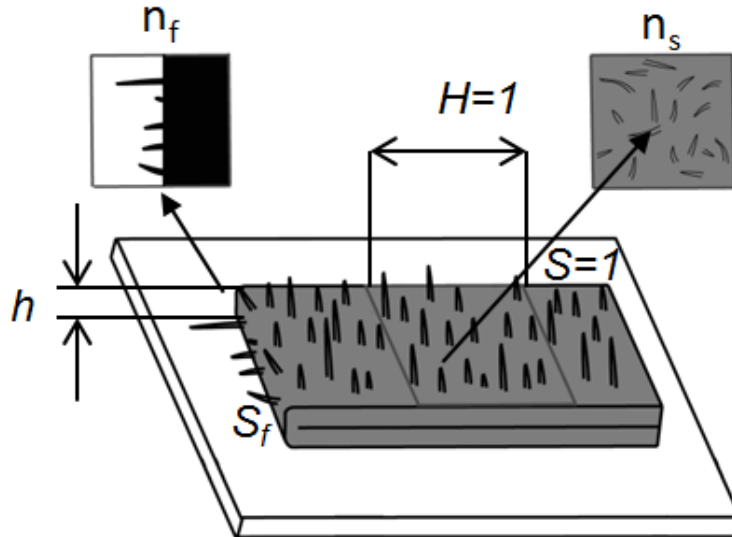


Рисунок 2.6. Получение изображения сгиба и поверхности листа для сравнения числа трихомов на поперечном сгибе листа и единичной площади поверхности. Изображены элементы площади поверхности единичного размера $S=1$, имеющего высоту $H=1$, поверхность сгиба площадью S_f с высотой сгиба h . Стрелки указывают на схематические изображения сгиба и поверхности и соответствующие им значения числа трихомов на единицу длины сгиба (n_f) и на единицу площади поверхности (n_s).

Изображение сгиба содержит трихомы, которые находятся на участке поверхности, соответствующем сгибу листа, и имеющему площадь S_f . Следовательно число трихомов n_s на фрагменте поверхности единичной площади относится к числу трихомов n_f на сгибе как $n_s/n_f = 1/S_f$, где n_f – число трихомов, нормированных на длину сгиба. При этом площадь сгиба листа S_f можно оценить приблизительно как площадь прямоугольника со стороной h (высотой сгиба листа, см. рис. 2.6). Тогда отношение числа трихомов на сгибе к числу трихомов на поверхности листа с единичной площадью можно оценить

по формуле $n_s/n_f=1/h$. Это означает, что между числом трихомов на сгибе листа n_f и на его поверхности единичной площади n_s должна наблюдаться линейная зависимость $n_s \sim n_f/h$.

Чтобы подтвердить предложенную модель и показать, что число трихомов на линии сгиба листа может служить адекватной оценкой плотности трихомов на его поверхности (то есть числа трихомов на единицу площади поверхности листа) мы сравнили среднее число трихомов на 1 мм линии сгиба листа (n_f) и среднее число трихомов на 1 мм² поверхности листа (n_s) вблизи линии сгиба. Оценки n_f и n_s были получены для 200 изображений поверхности листа на расстоянии от 1 см до 3 см от линии сгиба листа. Для получения образцов мы использовали сорта и гибриды пшеницы как со слабым, так и с интенсивным опушением адаксиальной поверхности листа: сорт 'Родина' (без опушения), 'Диамант 2' (слабое опушение), 'Саратовская 29' (интенсивное опушение), интрогрессионные линии 102/00ⁱ и 821 (Подробнее см. раздел 2.2 «Растительный материал и условия выращивания»). У растений были взяты фрагменты предфлагового листа в 4 местах на разных удалениях от основания. Таким образом были обработаны 50 растений разных генотипов.

Чтобы подсчитывать трихомы на поверхности листа, его фрагмент помещали на предметное стекло так, чтобы адаксиальная поверхность располагалась сверху. Изображения поверхности были получены с помощью микроскопа Carl Zeiss Axioskop 2 Plus. Трихомы были идентифицированы с помощью аутофлуоресценции эпидермиса листа. Примеры изображений трихомов на сгибе и поверхности листа из эксперимента приведены на рис. 2.7.

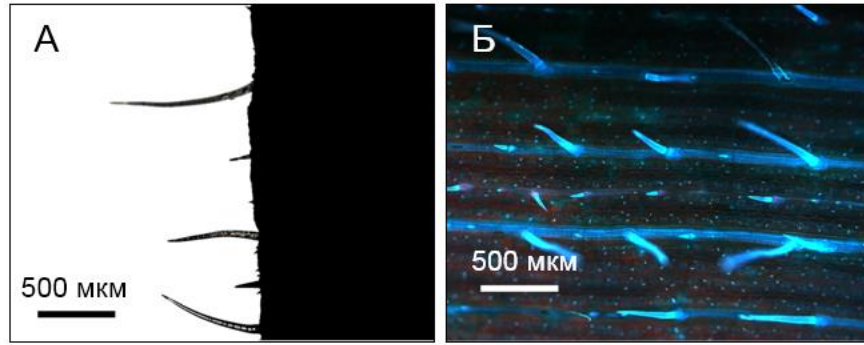


Рисунок 2.7. Микрофотография сгиба листа растения линии 102/00ⁱ (А) и флуоресцентная микрофотография поверхности листа (Б) вблизи линии сгиба. На линии сгиба листа было выявлено пять трихомов (1,83 трихома на 1 мм линии сгиба). На листовой поверхности был обнаружен двадцать один трихом (3,63 трихома на 1 мм² листовой поверхности). Шкала=500 мкм.

Площадь изображения для линии сгиба и поверхностных микрофотографий составила 2730 x 2163 мм (5776 мм²). Плотность трихомов на листовой поверхности растений варьировала от 0.4 до 110 трихомов на 1 мм². Трихомы, видимые на линии сгиба, и трихомы, видимые на поверхности листа, подсчитывались вручную. Данные были нормализованы следующим образом: количество линий сгиба было нормализовано к длине линии сгиба, видимой на микрофотографии, а количество трихомов на листовой поверхности было нормализовано к площади листовой поверхности, видимой на изображении.

Зависимость между оценками числа трихомов на единицу длины сгиба (n_f) и числа трихомов на единицу площади (n_s) в соседнем со сгибом участке листовой поверхности для 200 образцов сгиба листа показана на рис. 2.8. Из рисунка видно, что при значениях $n_f > 3$ зависимость $n_s(n_f)$ подчиняется линейному закону (коэффициент корреляции значим, $p < 0.01$).

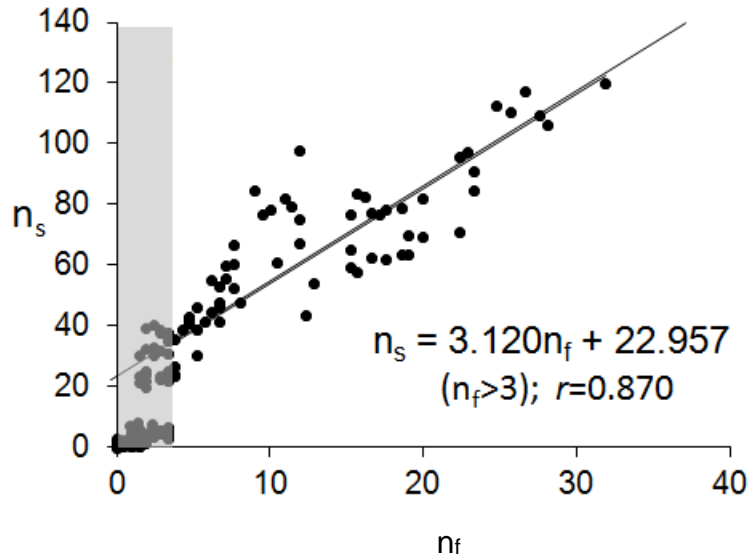


Рисунок 2.8. Зависимость n_s от n_f для 200 изображений сгибов и поверхности листа 50 растений различных генотипов. Приведены уравнение $n_s(n_f)$ для образцов с $n_f > 3$, коэффициент корреляции и линия регрессии. Серым цветом отмечена зона неопределенности ($n_f < 3$), для которой регрессионная зависимость $n_s(n_f)$ не выполняется.

Однако для образцов листьев с редкими трихомами ($n_f \leq 3$) часто встречаются случаи, когда близким значениям n_f соответствуют сильно различающиеся значения n_s , что характеризует данный интервал значений как область неопределенности (помечена на рис. 2.8 серым цветом). В этой области значений, оценки плотности опушения листа дают сильный разброс. Такой разброс можно объяснить в рамках предложенной модели сгиба листа: если опушение слабое (т.е. на единицу площади число трихомов мало), может возникнуть ситуация, когда на участок поверхности сгиба размером S_f не попадает ни одного трихома. Это происходит в силу случайного расположения трихомов на поверхности и того, что $S_f < 1$. При значениях n_s меньше определенного порога, значения n_f с большой вероятностью будут равны 0. Это означает, что линия регрессии $n_s(n_f)$ должна пересекать ось Y в точке $n_{s0} > 0$ (свободный параметр регрессии). Это означает, что из оценки свободного

параметра регрессии можно оценить значение h – толщины сгиба листа. В нашем случае значение свободного параметра составило 3.120. Поскольку n_s нормировалось на 1 мм^2 , оценка толщины сгиба на основе свободного параметра уравнения линейной регрессии в данной модели дает $h=1/3.120 \approx 0.320 \text{ мм}$, что оказывается близко к удвоенной толщине листа пшеницы.

Проведенный анализ показывает, что число трихомов на микроизображении сгиба листа является приемлемой оценкой количества трихомов на единицу площади его поверхности. Однако с учетом большого разброса оценок количества трихомов для слабого опушения листа оценку количественных характеристик опушения производили путем усреднения результатов по нескольким изображениям сгибов, полученных для разных участков поверхности листа.

Предложенный нами способ получения образцов сгиба листа пшеницы, их фотографирование под оптическим микроскопом для дальнейшего компьютерного анализа позволяет существенно ускорить процесс оценки количественных характеристик опушения листа пшеницы. Время получения изображения одного образца для настроенного микроскопа составляет менее 1 мин, время компьютерной обработки – несколько секунд. По нашему опыту работы один человек за 1 час может собрать образцы сгибов листа и обработать их согласно предложенным рекомендациям для 10 растений. При этом для каждого растения берется предфлаговый лист, образцы получаются в местах трех сгибов, для каждого сгиба делается три изображения (всего 30 изображений). Из этого времени около 10 мин. уходит на работу с микроскопом.

2.3. Оценка количественных характеристик опущения листа на основе анализа двумерных микроизображений сгиба

Нами были разработаны алгоритмы автоматического анализа изображений сгибов листа пшеницы для определения количественных характеристик опущения, которые являются важными для генетиков. Во-первых, это количество трихомов на изображении; во-вторых, это характеристики размеров трихомов (длина, толщина); также интерес могут представлять особенности взаимного распределения трихомов на поверхности листа. Нами были разработаны два подхода для анализа изображений, реализованные в программах LHDetect и LHDetect2. Ниже мы опишем эти алгоритмы, и проведем сравнение их точности.

2.3.1. Алгоритм автоматической оценки количественных характеристик опущения листа LHDetect

В алгоритме LHDetect (Дорошков и соавт., 2009) первый этап анализа заключается в определении принадлежности пикселя изображения листу или фону. Предварительный анализ изображений показал, что изображения листа и фона можно хорошо различить по цвету (в том числе и по его интенсивности). Данные в виде битовых карт цифровых микроизображений содержат информацию об интенсивности цветовых компонент (RGB – красной, зеленой и синей) каждого пикселя изображения. Поэтому для решения первой задачи мы использовали классификацию пикселей в пространстве компонент RGB. Для этого в правой и левой части изображения определялись вертикальные полосы, которые заведомо принадлежат листу и фону (рис. 2.9 А). Для пикселей каждой из полос определялись средние значения компонент RGB. Каждый пиксель рисунка относился к типу «фон», если в пространстве RGB его положение ближе к средним значениям компонент RGB пикселей полосы фона и к типу «лист», если в пространстве

RGB его положение ближе к средним значениям компонент RGB пикселей полосы листа.

После классификации пикселей на «фон» и «лист» определялась граница листа. Для этого изображение разбивалось на горизонтальные полосы фиксированной высоты, подобранной опытным путем (высота полосы была в 5 раз больше средней толщины трихома). В каждой из полос находился пиксель фона, являющийся крайне правым в своей полосе. Полученные в каждой полосе крайние правые пиксели являлись узлами границы. Соединив узлы ломаной, получим линию границы листа (рис. 2.9 Б).

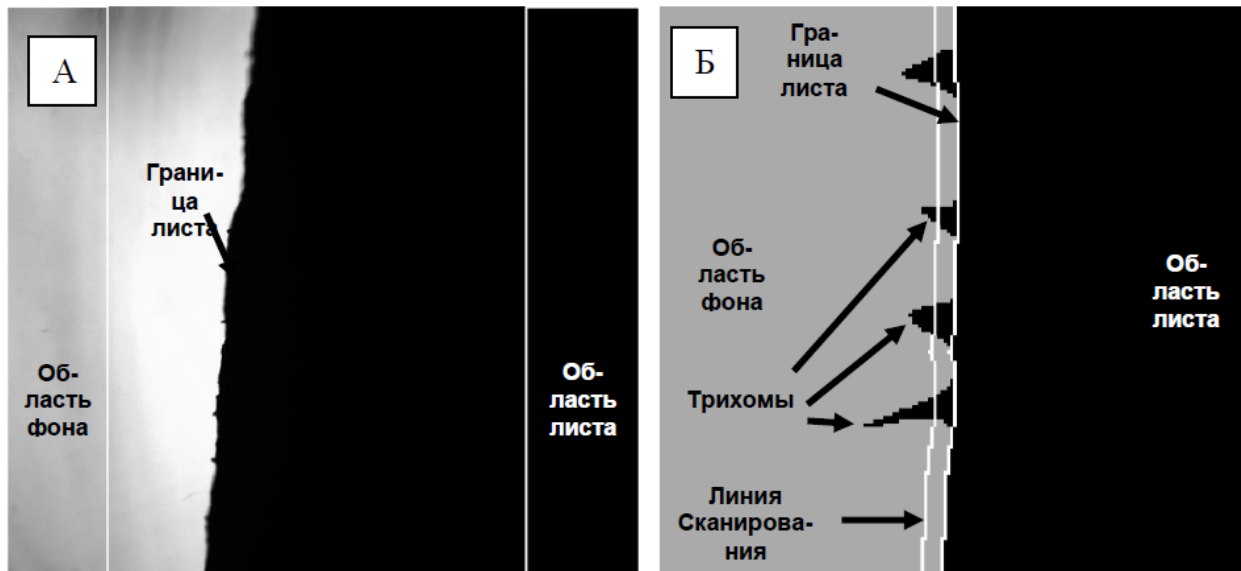


Рисунок 2.9. Схема работы алгоритма определения характеристик опушенности листа LHDetect. (А) Выбор полос для определения цветовых характеристик фона и листа. (Б) Граница листа и работа алгоритма подсчета числа трихомов.

Для подсчета числа трихомов на разном удалении от границы листа линия границы смещалась по горизонтали влево от поверхности листа на 1,2,3, ..., i , ..., 121 пикселей (всего анализировалось 120 линий). Для каждого i -го смещения подсчитывалось число пересечений линии с трихомами (т.е. число

непрерывных отрезков на такой линии, пиксели которых по цветовой классификации были отнесены к типу «лист»). Это число s_i является оценкой количества трихомов на удалении от поверхности листа, эквивалентном i пикселям (размер пикселя составлял 2.1 мкм). На (рис. 2.9 Б) представлен пример одной из линий проекции границы, которая пересекает 4 трихома. Число трихомов, имеющих на микрофотографии длину i , n_i , вычисляется как $n_i = s_i - s_{i+1}$. В настоящей работе мы использовали более грубую оценку распределения трихомов по длине: мы подсчитывали числа трихомов, длины которых находились в пределах $i, i+5$, т.е. рассчитывали n_i с шагом 5 пикселей (~10 мкм). Полное число трихомов n_f на микрофотографии вычислялось путем суммирования всех значений n_i . В результате такого анализа каждое изображение листа характеризовалось 25 значениями вектора n_f ($i=1,6,11, \dots, 121$ пикселей) и значением n_f , которые количественно описывают морфологию опушения листа. Полученные таким образом параметры позволяют описывать признак опушенности листа количественно (число трихомов на изображении и их распределение по длинам), что гораздо более информативно по сравнению с использованием качественных характеристик («без опушения», «слабое опушение», «сильное опушение»), которые определялись тактильно.

Предложенный алгоритм LHDetect отличается простотой и быстротой расчетов. Использование этого алгоритма впервые позволило определять количественные характеристики опушения листьев пшеницы для десятков растений за короткое время (Дорошков и соавт., 2011).

Однако, у алгоритма LHDetect есть целый ряд недостатков. Прежде всего, он не позволяет оценивать размер отдельного трихома. Во-вторых, он дает неверную оценку длин трихомов в случае, если они располагаются по отношению к поверхности листа наклонно (не под углом 90°). В-третьих, он никак не различал случаи пересечения трихомов между собой, что тоже могло

повлиять на точность метода. Поэтому нами был предложен более совершенный алгоритм, LHDetect2, основанный на выделении трихомов как отдельных объектов на изображении.

2.3.2. Алгоритм автоматической оценки количественных характеристик опущения листа LHDetect2

Модифицированный алгоритм основан на том, что на полученных изображениях сгибов листа можно выделить несколько областей (лист, фон), границу между листом и фоном и трихомы (рис. 2.10). Для каждого трихома можно определить вершину, основание и центральную линию, которая может аппроксимировать длину трихома. Для того, чтобы выделить указанные объекты на изображении мы проводили предварительную обработку изображения и его сегментацию.

Алгоритм предварительной обработки и анализа микроизображений сгибов листа пшеницы, включает подавление шума и сегментацию. Для подавления шума изображение преобразуется в оттенки серого и размывается Гауссовым фильтром. В результате на изображении устраняются мелкие пятна грязи в фоновой области и воздушные пузыри на изображении трихомов.

Сегментация служила для разделения изображения на фоновую область (более светлые пиксели) и листовую область, содержащую трихомы (более темные пиксели). Для сегментации мы использовали алгоритм, основанный на классификации пикселей по порогу интенсивности. Для оценки пороговой интенсивности были проанализированы значения интенсивности пикселей для 1352 изображениях сгибов листьев. Распределение пикселей по интенсивности представлено на рис. 2.11.

Распределение на рис. 2.11 имеет пять четко выраженных пиков: в интервале от 0 до 50 для пикселей области листа; в интервале от 50 до 205 для

мелких пятен в фоновой области; от 205 до 228 для трихомов и более крупных пятен, от 228 до 240 для пикселей листа вблизи границы листа/фона и трихомов, не находящихся в фокусе, и от 240 до 255 для фоновой области. В таблице 2.1 характеристики этих интервалов приведены более детально, а на рисунке 2.12 пиксели из этих категорий на изображении сгиба листа показаны различными цветами.

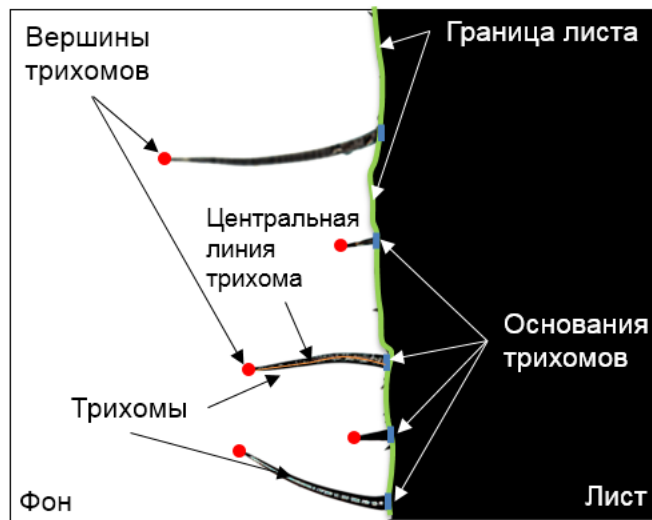


Рисунок 2.10. Основные объекты на изображении сгиба листа: фон (белый), лист (черный), граница между листом и фоном (зеленый), трихомы (черный/серый), вершины трихомов (красный) и их основания (голубой), центральная линия трихома (оранжевый).

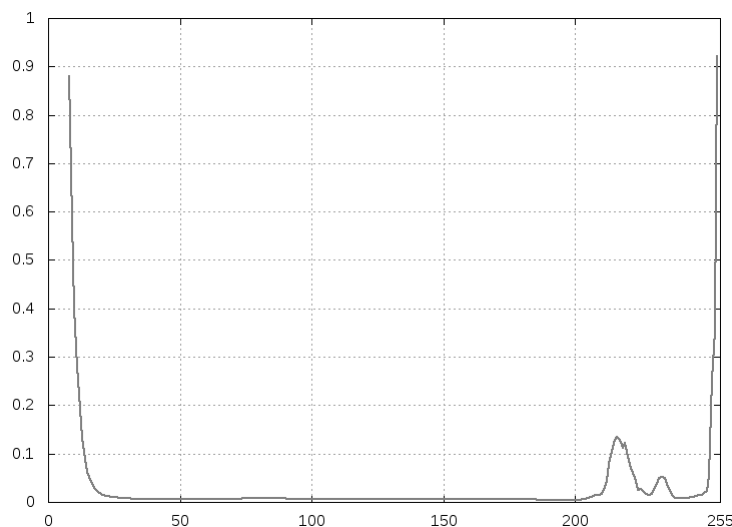


Рисунок 2.11. Распределение интенсивности пикселей в 1352 изображениях сгибов листа. По оси X-интенсивность серого уровня; по оси Y-доля пикселей относительно общего числа пикселей во всех изображениях.

Таблица 2.1. Сегментация изображения сгиба листа на основе интенсивности серого уровня. Каждый сегмент снабжен описанием цвета, иллюстрирующим результат сегментации на рис. 2.12.

Номер сегмента	Интенсивность	Цвет сегмента на рис. 2.12	Объекты, ассоциированные с сегментом
1	0-50	зеленый	Лист
2	50-205	красный	Небольшой шум и пятна
3	205-228	голубой	Трихомы, большие пятна
4	228-240	темно-синий	Граница листа и фона, небольшие пятна и шум
5	240-255	белый	Фон

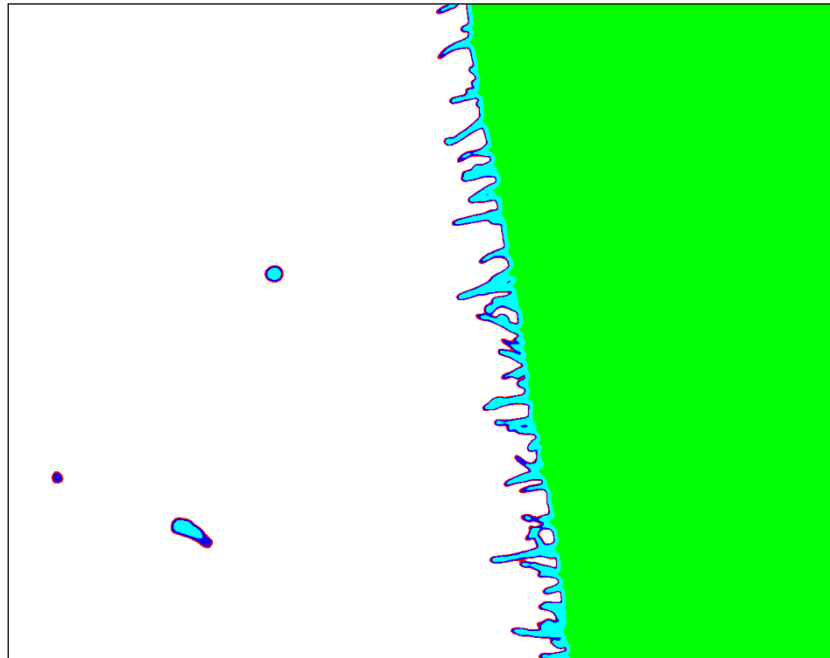


Рисунок 2.12. Пример сегментации изображения сгиба листа по интенсивности пикселей.

Данный анализ показал, что сегментация по порогу интенсивности пикселей позволяет выделить указанные выше объекты на изображении сгиба листа.

Мы использовали порог интенсивности, установленный на уровне 240. Предполагалось, что пиксели с интенсивностью серого уровня выше этого значения относятся к фоновой области, а пиксели с интенсивностью серого уровня ниже этого значения относятся к области листа.

После определения фона и листа на изображении необходимо было определить между ними границу, исключив из рассмотрения трихомы. Мы аппроксимировали границу линией, которая содержала пиксели листа, контактирующие с фоновыми пикселями и проходящие через основания

трихомов. С этой целью изображение было разделено на перекрывающиеся горизонтальные полосы, каждая из которых имеет высоту $ws = 70$ пикселей (~147 мкм). Полосы были смещены по вертикали друг относительно друга на 1 пиксель, с перекрыванием. В каждой из таких полос был идентифицирован пиксель фона, который имел максимальную координату X (крайний правый пиксель). Пиксель листа рядом с ним справа считался пограничной точкой между листом и фоном. После того, как все пограничные точки были идентифицированы, они были соединены ломаной линией, которая аппроксимировала линию границы между листом и фоном. Любому пикселю листа, расположенному в пределах *min_trichome* слева от границы, присваивалась категория «фон» для исключения из рассмотрения небольших выступов на поверхности листа, которые не являлись трихомами. После этого пиксели листа преобразовывались в категорию «фон» для того, чтобы в последующем использовать алгоритм бинарной классификации для определения трихомов.

После преобразования пикселей листа в пиксели фона мы применили бинаризацию с пороговым значением интенсивности 205. Все пиксели, имевшие интенсивность меньше или равную этому значению, считались пикселями трихомов. Использование этого порогового значения позволило нам избавиться от небольших пятен вблизи трихомов и отдельных трихомов, которые были не в фокусе. В результате бинаризации на изображении оставались только контуры, которые соответствовали трихомам. Для идентификации контуров и их границ использовали функцию библиотеки OpenCV 2.0 (Bradski, Kaehler, 2008).

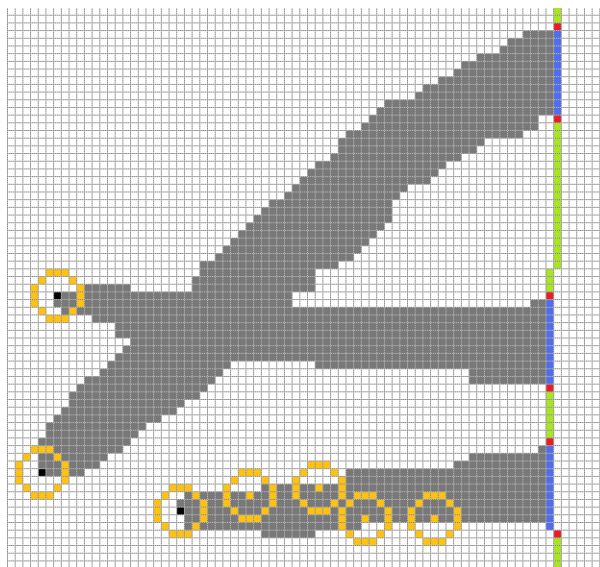


Рисунок 2.13. Идентификация вершин и оснований трихомов. Различные объекты на изображении приведены разными цветами: белый цвет - пиксели фона; серый цвет - пиксели трихомов; зеленый цвет - граница листа и фона; черный цвет - пиксели, идентифицированные как вершины трихомов; красный цвет - границы основания трихома; синий цвет - основание трихома; желтый цвет - окружности, с помощью которых можно определить, представляет ли его центральный пиксель вершину трихома. На рисунке изображены два объекта (контура). Верхний является составным, он состоит из двух пересекающихся трихомов, имеет две вершины и два основания. Нижний объект простой, он имеет одну вершину и одно основание (состоит из одного трихома).

Далее для каждого контура (объекта) были определены вершины и основания (см. рис. 2.13). Любой объект, состоящий из более чем одного трихома, может иметь более чем одну вершину и более чем одно основание, причем число вершин не обязательно совпадает с числом оснований.

Чтобы найти вершины трихомов в контуре, алгоритм обходит все пиксели границы контура. Для каждого пикселя границы контура i идентифицируются еще два граничных пикселя объекта в позициях $i \pm N$ и измеряется угол, образованный этими тремя пикселями, с пикселем i в качестве

его вершины. Если этот угол меньше α , то вершина угла считается предполагаемой вершиной трихома.

Поскольку ширина любого трихома вблизи вершины может быть намного меньше N , чтобы повысить точность определения значения угла α , мы использовали более надежный, хотя и более медленный алгоритм. Мы рассматривали окружность с радиусом R и центром в предполагаемом пикселе вершины трихома и подсчитывали, сколько пикселей трихома попадает внутрь этой окружности. Доля этих пикселей по отношению ко всей совокупности пикселей внутри окружности умноженная на 2π являлась оценкой значения угла изгиба границы контура. Например, если она близка к половине – то угол близок к π , если близка к четверти, то значение угла было близко к $\pi/2$ и т.п. Основания трихомов определялись как пиксели ранее определенной границы между листом и фоном, контактирующие с пикселями контура трихома. Любой пиксель, контактирующий с пикселями фоновой области, пикселем листа/границы и пикселем основания трихома одновременно, считался пикселем границы основания трихома. Вершины трихомов и их основания, идентифицированные с помощью этого алгоритма, показаны на рис. 2.13.

Все объекты на изображении, соответствующие описанным выше контурам могут быть классифицированы как простые (состоящие из одного трихома) или составные (состоящие из более чем одного трихома, см. рис. 2.13). Количество волосков в каждом объекте рассчитывается как $\max(H, B)$, где H -количество вершин в объекте и B -количество оснований. Если объект простой (один трихом), то $H=B=1$.

В нашей базе данных изображений было идентифицировано 20074 объекта, состоящих в общей сложности из 31037 трихомов. Составные объекты составляют 31.1% и занимают 15.4% от общей площади всех

идентифицированных трихомов. Трихомы в составных объектах составляют 45.2%. Среднее число трихомов в сложном объекте составляет 2.8. Следовательно, составные объекты достаточно информативны для анализа опушения.

В качестве оценки длины трихома мы используем длину его центральной линии. Для ее построения в простых объектах мы используем алгоритм, предложенный Лейфером и его коллегами (Leifer et al., 2011). Граница простого объекта делится на две части: верхнюю часть, от вершины трихома до верхнего пикселя границы основания трихома, и нижнюю часть, от вершины трихома до нижнего пикселя границы основания трихома. После деления алгоритм обходит по очереди каждый пиксель обеих частей границ от вершины к основанию; на каждой итерации пара пикселей, принадлежащих верхней и нижней частям границы, соединяется отрезком для которого определяется середина. Срединные точки соединяются ломаной линией, которая аппроксимирует центральную линию трихома. Длина центральной линии представляет собой оценку длины соответствующего трихома (рис. 2.14 А).

Для обработки составных объектов, включающих несколько трихомов, использован другой алгоритм. Для каждого пикселя объекта вычисляется расстояние до ближайшего пикселя границы. Полученные значения нормализуются. Центральная линия трихомов определяется как путь наименьшего веса в графе от пикселя, представляющего вершину трихома, до пикселя, представляющего середину основания трихома. Результаты работы алгоритма приведены на рис.2.14 А, Б.

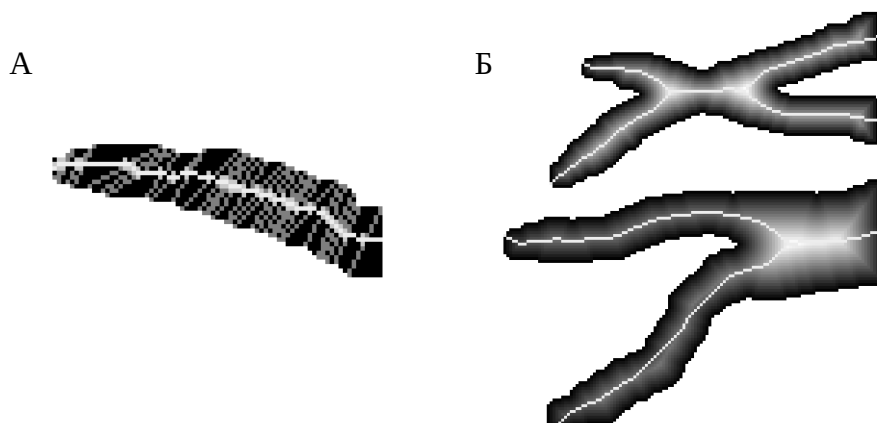


Рисунок 2.14. Идентификация центральных линий в простых и составных объектах. (А) в простом объекте центральная линия аппроксимируется ломаной линией, проходящей через середины отрезков, которые соединяют верхние и нижние граничные пиксели объекта. (Б) в составном объекте центральная линия формируется пикселями объекта, наиболее удаленными от границы объекта.

Работа алгоритма LHDetect2 зависит от ряда параметров:

- σ : параметр размытия Гауссовского фильтра при предварительной обработке изображения;
- *min_trichome*: наименьшая длина трихома;
- *alpha* (в градусах): наибольшая кривизна контура, при которой выбранный пиксель объекта будет помечен как предполагаемая вершина трихома;
- *N*: расстояние в пикселях между предполагаемым вершиной трихома и соседними пикселями границы трихома, используемыми для оценки угла *alpha*;
- *R*: радиус окружности с центром в предполагаемом пикселе вершины трихома для надежной оценки кривизны контура;

- *percent*: наибольший процент пикселей трихома в пределах тестовой окружности радиуса R , при котором пиксель считается вершиной трихома.

Для оценки оптимальных параметров алгоритма мы использовали набор из 76 изображений сгибов листа растений двух генотипов пшеницы с контрастным опушением листьев: сорта ‘Саратовская 29’ и линии 102/00ⁱ. На каждом изображении вручную круглым маркером были отмечены области вершины трихома, каждая из которых имела диаметр 15 пикселей. Количество вершин было равно количеству трихомов на изображении. С помощью предложенного алгоритма мы определяли положение вершины каждого трихома на изображении. Предполагалось, если программа определит вершину трихома, которая попала в размеченную область, то вершина трихома идентифицирована правильно.

Для каждого изображения в выборке 76 изображений мы рассчитали F -меру точности предсказания вершин (формула 1.7) (van Rijsbergen, 1979). Затем усредняли значение этого параметра по всей выборке. Оптимальные параметры подбирались автоматически методом перебора по многомерной сетке значений таким образом, чтобы максимизировать значение F -меры. Были оптимизированы следующие параметры:

- σ менялся от 0 до 2 с шагом 0,25.

- *min_trichome* варьировался от 4 до 10 пикселей с шагом в 2 пикселя, ему соответствовала и величина параметра $R = \text{min_trichome}/2$;

- *alpha* изменялся от 80° до 120° с шагом 1°;

- *percent* варьировался от 10 до 90 с приращением на 1-2 в диапазоне от 24 до 40; с приращением на 10 для другой области значений (10, 20, 24, 25, 26, 28, 30, 32, 34, 35, 36, 40, 50, 60, 70, 80, 90).

Параметр N был фиксирован: $N=2$.

В результате этого процесса были получены следующие значения параметров, максимизирующих F -меру на обучающей выборке: $\alpha=105^\circ$, $\min_trichome=6$ пикселей, $percent=30$, $\sigma=1.0$. Оценка F при этом составила 0.85, оценка доли ложно предсказанных вершин, $fp=0.07$, а оценка доли не обнаруженных вершин $fn=0.19$.

Независимая оценка точности подсчета количества трихомов на изображении осуществлялась на выборке из 70 дополнительных изображений тех же генотипов, что и в выборке для оценки оптимальных параметров. В качестве меры оценки точности мы использовали MAE, MAPE и коэффициент корреляции Пирсона r . Кроме того, мы оценили точность оценки длины трихомов. Для этого из выборки тестовых изображений были отобраны 20, на которых вручную с помощью модуля Segmented line tool из пакета ImageJ (rsbweb.nih.gov/ij/) были оценены длины трихомов. Эти значения были сопоставлены с оценками, полученными автоматическим методом, а точность оценивалась с помощью коэффициента корреляции Пирсона.

Сравнение точности работы предложенного алгоритма с анализом трихомов, проведенным вручную показано на рис. 2.15 А. Из рисунка видно, что между количеством трихомов на изображении, посчитанным вручную, и количеством трихомов, определенным автоматически, существует хорошее соответствие. Коэффициент корреляции Пирсона $r=0.986$. Оценка коэффициента регрессии, удовлетворяющей условию $y(0)=0$, дает зависимость $y=0.887x$. Точки на графике располагаются вблизи прямой $y=x$, однако для больших значений точки располагаются несколько ниже этой линии. Это свидетельствует о том, что программа несколько недооценивает количество трихомов (в среднем, на 3-4 штуки на изображение). Это происходит за счет

отбрасывания небольших выростов на поверхности листа (длина которых не превышает 12.6 мкм, 6 пикселей), часть которых тем не менее экспертом были определены как трихомы.

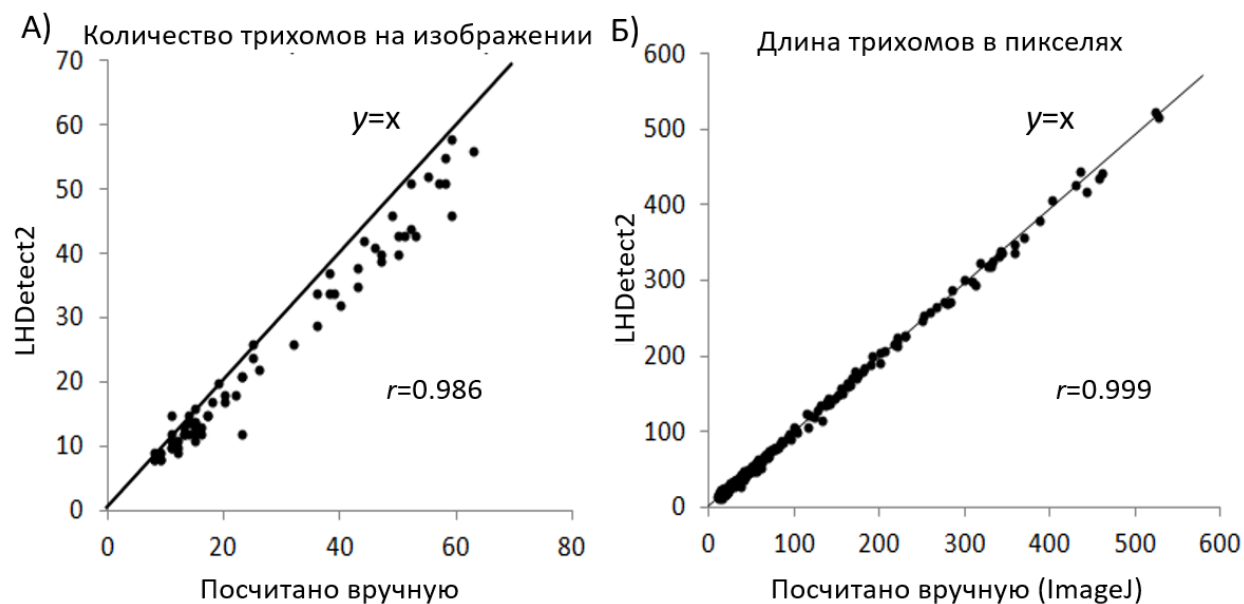


Рисунок 2.15. Сравнение оценок числа трихомов на изображении сгиба листа и их длины, полученных программой LHDetect2 и вручную. (А) Число трихомов, подсчитанных программой LHDetect2 (ось Y) и определенных вручную (ось X). Показана линия $y=x$ и коэффициент корреляции между двумя оценками. (В) Длина трихома, оцененная программой LHDetect2 (ось Y) и соответствующее значение, полученное вручную с использованием пакета ImageJ (ось X). Также показана линия $y=x$ и приведен коэффициент корреляции между двумя оценками.

Что касается определения длины трихомов, то и в этом случае можно наблюдать хорошее соответствие между длинами, полученными вручную и автоматически. Оценка коэффициента регрессии, удовлетворяющей условию $y(0)=0$, дает зависимость $y=0.980x$, коэффициент корреляции Пирсона $r=0.999$. Среднее отклонение длины трихомов, полученной программой LHDetect2 и

ручной оценки составляет 3 пикселя (6.3 мкм), что в среднем составляет 6% от длины трихома.

Сравнение точности программы LHDetect2 с программой, LHDetect, представлено в табл. 2.2. Эти методы мы сравнивали по двум параметрам характеризующих точность определения числа трихомов на изображении: средняя абсолютная ошибка (MAE, формула 1.3) и средняя абсолютная процентная ошибка (MAPE, формула 1.4). Кроме того, выборку из тестовых изображений мы визуально классифицировали на простые (содержат мало перекрывающихся трихомов) и сложные (содержат большое число перекрывающихся трихомов). Выборка содержала, таким образом, 44 простых и 26 сложных изображений. Среднее число трихомов было 15.64 на простом изображении и 48.42 на сложном. Результаты сравнения двух программ по параметрам MAE и MAPE показывают, что ошибка LHDetect примерно на 0.8 трихомов больше, чем у LHDetect2.

Таблица 2.2. Сравнение точности оценки числа трихомов на изображении сгиба листа, полученных методом LHDetect и LHDetect2. Приведены значения MAE и MAPE оценок числа трихомов, посчитанными автоматически в сравнении с ручной оценкой. Отдельно приведены оценки этих параметров для всей выборки из 70 тестовых изображений, а также для 44 простых и 26 сложных изображений.

Выборка изображений	LHDetect2		LHDetect	
	MAE, шт	MAPE, %	MAE, шт	MAPE, %
Полный набор	3.33	11.65	4.54	14.22
Простые изображения	1.82	11.23	1.64	11.60
Сложные изображения	5.88	12.38	9.46	18.64

Интересно, что для простых изображений ошибка у двух алгоритмов

примерно одинакова. Сильные различия наблюдаются для сложных изображений, которые содержат большое число трихомов с их сильным перекрытием. Ошибка MAE на этих изображениях для алгоритма LHDetect почти в два раза выше, чем для LHDetect2. Ошибка MAPE алгоритма LHDetect почти в полтора раза выше, чем для LHDetect2. Отметим, что программа LHDetect не позволяла оценивать длины отдельных трихомов.

2.3.3. Компьютерная программа LHDetect2 для оценки количественных характеристик опушения листа пшеницы

Описанный выше алгоритм был реализован в программе LHDetect2. Это консольное приложение, реализованное на языке C с использованием библиотеки анализа цифровых изображений OpenCV 2.0 (Bradsky, Kaehler, 2008). Эта программа свободно доступна по адресу <https://github.com/genaev/LHDetect2>. Кроме того, для удобства доступа пользователей был реализован Web сервис для анализа микрофотографий сгиба листа, который доступен по адресу <http://wheatdb.org/lhdetect2>. Необходимо отметить, что возможность выполнения анализа цифровых изображений через web-сервис для фенотипирования растений была реализована впервые. Вид основной страницы Web-сервиса LHDetect2 показан на рис. 2.16. Страница содержит поля для выполнения анализа изображений: (1) опцию выбора единиц измерения размеров пикселей; (2) единицы измерения размера для случая, если выбрана пользовательская шкала; (3) размер пикселя в пользовательских единицах измерения; (4) размер бина по длине трихомов, необходимый для построения гистограммы; (5) название файла с изображением сгиба листа, полученный по протоколу, описанному в разделе 2.1.

WheatPGE *beta*

[Login](#) | [Registration](#)
Language: [ENG](#) | [RUS](#)

[Main](#) [Plants](#) [Search](#) [Timeline](#) [Blog](#)

LHDetect2

Leaf image processing for wheat hairiness

Type of measurement units: pixels (Pixel) or other (User): (1)

Name of measurement units, if its type is user-defined (for example, mkm): (2)

Width of image in user units to determine measurement scale: (3)

Step of trichome length counting in units defined: (4)

Input image file (png,tif,bmp,jpeg and gif formats supported): (5)

(6)

Examples

(click on image for full size photo)

The program input is leaf image file in TIFF/PNG/JPG/BMP/GIF format prepared according the protocol ([LeafImageProtocol.pdf](#)).

The detailed description of the method could be found in the paper of Genaev et al "Extraction of quantitative characteristics describing wheat leaf pubescence with a novel image-processing technique" (2012) PLANTA, DOI: 10.1007/s00425-012-1751-6

© 2009-2019 wheatdb.org

Рисунок 2.16. Интерфейс Web-сервиса LHDetect2. Цифрами обозначены основные поля для ввода параметров выполнения программы: (1) выбор единиц измерения размеров пикселей; (2) ввод единицы измерения размера для случая, если выбрана пользовательская шкала; (3) ввод размер пикселя в пользовательских единицах измерения; (4) размер бина для построения гистограммы распределения трихомов по длине; (5) название файла с изображением сгиба листа; (6) кнопка запуска программы.

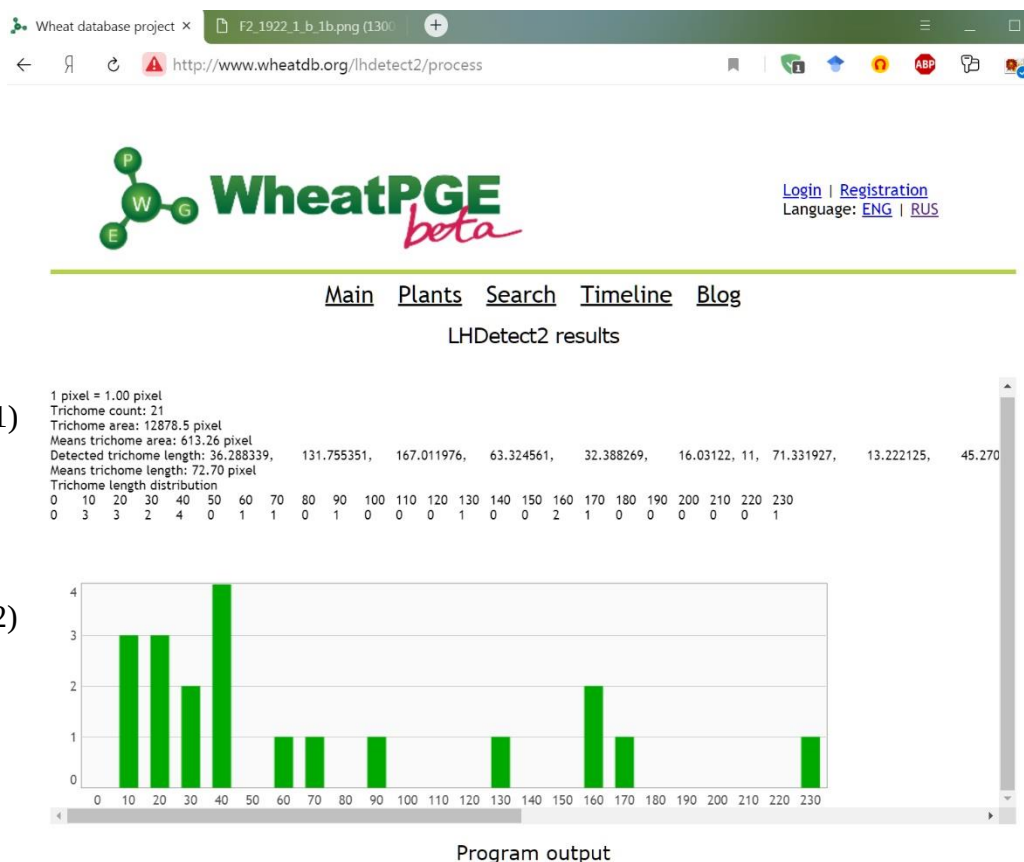


Рисунок 2.17. Визуализация результатов работы Web-сервиса LHDetect. Общая информация. Цифрами обозначены: (1) Вывод общих параметров, характеризующих трихомы на изображении (число трихомов, общая площадь, средняя площадь трихома, длины трихомов, средняя длина трихома, распределение длин трихомов в текстовом виде); (2) визуализация гистограммы распределения трихомов на изображении по длинам.

Результат работы Web-сервиса LHDetect показан на рис. 2.17 и 2.18. Он содержит общую статистику (число трихомов, общая площадь, средняя площадь трихома, длины трихомов, средняя длина трихома, распределение длин трихомов в текстовом виде), визуализацию гистограммы распределения трихомов на изображении по длинам (рис. 2.15).

```

Contour:1 area:190.00 complex:0 trichome count:1
  Head pixel 1
    Distance to the last trichome: -1
    Length: 36.29
    588 1015
  Foot pixel 2
    618 1002
    618 994

Contour:2 area:8.00 complex:0 trichome count:0
  Head pixel 0
  Foot pixel 2
    625 806
    625 804

Contour:3 area:923.50 complex:0 trichome count:1
  Head pixel 1
    Distance to the last trichome: 203.90
    Length: 131.76
    507 753
  Foot pixel 2
    625 802
    625 792

Contour:4 area:2292.00 complex:1 trichome count:2
  Head pixel 2
    Distance to the last trichome: 77.24
    Length: 167.01
    457 719
    Distance to the last trichome: 0.00
    Length: 63.32
    563 755
  Foot pixel 4
    625 747
    624 729
    624 726
    624 717

Contour:5 area:387.00 complex:1 trichome count:3
  Head pixel 3
    Distance to the last trichome: 70.38
    Length: 32.39
    595 634
    Distance to the last trichome: 6.00
    Length: 16.03
    614 650
    Distance to the last trichome: 10.41
    Length: 11.00
    617 634
  Foot pixel 5
    631 655
    630 643
    628 632
    629 633

```

Рисунок 2.18. Визуализация результатов работы Web-сервиса LHDetect. Вывод количественных характеристик идентифицированных объектов.

В дополнение к описанным выше характеристикам программа выдает количественные параметры для каждого из идентифицированных объектов (рис. 2.18): площадь, расстояние между основаниями до ближайшего трихома по вертикали, длину, координаты вершин и оснований.

Отметим, что гистограмма распределения трихомов по длинам, при условии одинаковых интервалов длин (бинов) является удобным способом описания количественных характеристик опушения листа. Эта информация может быть представлена в виде вектора n , который представляет гистограмму

в виде упорядоченного набора чисел. Каждый элемент этого вектора n_i является оценкой числа трихомов в интервале длины, соответствующему i -му бину гистограммы. Отметим, что пользователь может задавать интервалы длин не только в пикселях, но и в произвольных единицах измерения, например, микрометрах.

Такое представление удобно для сравнения опушения разных образцов как точек в многомерном пространстве, координаты которых равны элементам вектора, а размерность пространства равна числу бинов гистограммы. Сравнивать между собой можно вектора \mathbf{n} , полученные для идентичных интервалов длин.

В дальнейшем, при анализе опушения у пшеницы мы будем использовать следующие количественные характеристики опушения листьев: распределение трихомов по длинам (\mathbf{n}); число трихомов на изображении сгиба n_f ; среднюю длину трихомов l_m на изображении поперечного сгиба листа.

Последний параметр можно оценить, как

$$l_m = \frac{1}{n_f} \sum_1^{n_f} l_i, \quad (2.1)$$

где n_f – число трихомов, идентифицированных на сгибе листа, i – номер трихома, l_i – оценка его длины.

Индекс опушения, дополнительный параметр, который позволяет оценить его тип:

$$H_{L/N} = l_m/n_f, \quad (2.2)$$

описывающий отношение размера трихомов (средняя длина, l_m) к плотности опушения (число трихомов, n_f). Если среднее число трихомов на изображении меньше 3, то образец считается неопушенным и этот параметр не вычисляется. Высокие значения $H_{L/N}$ характерны для листьев с редкими длинными

трихомами, низкие значения (близкие к 1) характерны для плотного опушения, образованного короткими трихомами. Таким образом этот параметр характеризует тип опушения.

Дополнительно для оценки влияния отдельных генов на характер опушения мы рассчитывали параметр $K_{L/N}$ (Doroshkov et al., 2016). Этот параметр оценивает согласованность влияния гена на длину l_m и число n_f трихомов. Параметр рассчитывается как соотношение изменения длины трихомов для генотипов, содержащих различные аллели генов опушения к изменению числа трихомов и вычисляется по формуле

$$K_{L/N} = (l_{m2} - l_{m1}) / (n_{f2} - n_{f1}), \quad (2.3)$$

где n_{f2} и l_{m2} — число и длина трихомов генотипа, несущего исследуемый ген в аллельном состоянии 1, а n_{f1} и l_{m1} — число и длина трихомов генотипа с аллельным состоянием гена 2. Чем больше значение $K_{L/N}$, тем в большей степени ген влияет на длину трихомов по отношению к влиянию на число трихомов. Гены, имеющие сходные значения $K_{L/N}$, имеют сходный характер влияния на длину l_m и число n_f трихомов.

Очевидно, что указанные параметры (n , n_f , l_m , $H_{L/N}$) можно усреднять по набору изображений одного листа, растения, генотипа, и т.п. В таком случае они будут характеризовать опушение для соответствующей выборки изображений.

Микроизображения сгибов листа пшеницы в описанных далее работах были получены нами по единому протоколу (раздел 2.2.2 «Протокол пробоподготовки и получения изображения сгибов листа») и в одинаковом масштабе (размер пикселя 2.1 мкм), поэтому указанные характеристики являются сравнимыми для всех исследованных в этой главе образцов.

2.3.4 Статистические методы анализа количественных характеристик опушения листа

Основная цель анализа количественных характеристик – провести их сравнение между разными образцами, генотипами, растениями и т.п. Для этого можно использовать ряд статистических тестов. Один из наиболее известных подходов - дисперсионный анализ (как однофакторный, так и многофакторный) (Шеффе, 1980). При решении подобных задач можно также пользоваться и *t*-статистикой Стьюдента сравнения двух выборочных средних. Одна из решаемых задач: выявление генетических факторов, влияющих на опушение листа. При таком анализе независимыми переменными являются признаки генотипов растений (плоидность, аллельное состояние определенных генов и т.п.), зависимыми факторами являются признаки опушения (среднее число трихомов на изображении, средняя длина трихома, индекс опушения).

Второй тип задач, который возникает при анализе опушения: сравнение распределений длин трихомов для двух выборок (разных растений, генотипов, популяций и т.п.). В этом случае можно использовать непараметрические тесты сравнения двух распределений (хи-квадрат, Колмогорова-Смирнова и т.п.). В некоторых случаях различия можно оценить, зная стандартную ошибку среднего для количества трихомов в интервале заданной длины при 95% доверительном интервале. Если различие таково, что существенно превышает стандартную ошибку среднего, можно говорить о его значимости.

Третий тип задач, который возникает при анализе опушения – классификация образцов на однородные группы со сходными характеристиками опушения. Для этого оказывается удобно использовать методы кластерного анализа (Sneath, Sokal, 1973). Для того чтобы выделить в

популяции группы растений, имеющих сходный тип опушения, можно использовать иерархическую кластеризацию полученных для их векторов n . Для каждой пары растений k и j , вычисляется евклидово расстояние между векторами n_k и n_j ,

$$d_{kj} = \left(\sum_{i=1}^{i=n_b} (n_{ik} - n_{ij})^2 \right)^{1/2}, \quad (2.4)$$

где i это индекс интервала длины трихомов, n_b – количество таких интервалов в эмпирическом распределении. Это расстояние отражает степень различий опушения листа у двух растений (насколько различаются у них распределения трихомов по длине). Чем больше расстояние, тем сильнее различается опушение, в случае идентичного распределения трихомов по длинам у растений $d_{kj}=0$. В результате кластерного анализа с использованием различных аггломеративных стратегий (ближайшего соседа невзвешенного среднего и др.) получается бинарное дерево, конечные вершины которого соответствуют образцам в популяции, а длины ветвей отражают степень различий в опушении между ними. В настоящей работе мы использовали, прежде всего, иерархический кластерный анализ, который позволяет визуализировать относительные расстояния между группами объектов.

Для анализа количественных характеристик опушения листьев мы также использовали и другие методы многомерного статистического анализа (Ефимов, Ковалева, 2008): кластеризация методом k -средних, анализ главных компонент, многомерное шкалирование и др.

Следует отметить, что применение методов многомерного анализа к векторам распределения трихомов по длинам учитывает различия в количестве трихомов с разными длинами и, таким образом, является более информативным, чем сравнение усредненных параметров (числа трихомов или

их средней длины).

Для статистической обработки данных нами использовались пакеты программ Excel, PAST (Hammer et al., 2001).

2.4. Разнообразие опушения листьев у мягкой пшеницы и ее сородичей

Разработанный нами подход высокопроизводительной оценки количественных характеристик опушения листа пшеницы позволил впервые провести массовый анализ характеристик опушения у большого числа растений различных генотипов и тем самым описать разнообразие опушения листьев у пшеницы и ее сородичей.

2.4.1. Растительный материал и условия выращивания

Всего было изучено 47 образцов от 12 видов родов *Triticum* и *Aegilops*. Образцы были отобраны по 2 или более представителей каждой формы (за исключением *T. dicoccoides*). Таким образом, было изучено 24 диплоидных, 10 тетраплоидных и 13 гексаплоидных образцов. В целом в нашей выборке было представлено широкое внутри- и межвидовое разнообразие, характерное для видов родов *Triticum* и *Aegilops*.

Гексаплоидные генотипы включали 9 сортов *T. aestivum* L. с различной морфологией опушения листьев, 2 образца *T. spelta* и 2 образца *T. petropavlovskii*. Тетраплоидные генотипы включали набор образцов *T. durum* Desf., *T. dicoccoides* Körn., *T. dicoccon* Schrank ex Schübl. и *T. timopheevii* Zhuk. Представители диплоидных видов включали набор *T. monococcum* L., *T. boeoticum* Boiss., *T. urartu* Tumanian ex Gandilyan, *Ae. speltoides* Tausch. и *Ae. tauschii* Coss. Кроме того, изучались синтетические образцы тетраплоидные *T.*

palmovae G. Ivanov и гексаплоидный Synthetic 6x. Весь материал представлен яровой пшеницей. Происхождение, ботанические сведения для всех изученных образцов и используемые сокращения представлены в Приложении, табл. П2.

Анализируемая выборка образцов, представляет все возможные варианты хромосомных наборов пшениц: 16 представителей генома AA, 4 представителя геномов CC и DD, 6 представителей генома AABB, три представителя генома AAGG, 1 представитель генома AADD и 13 представителей генома гексаплоидной пшеницы (AABBDD). Это разнообразие геномов было представлено 848 растениями, имеющими также большую изменчивость в опушении: растения без трихомов, с длинными и редкими трихомами, с короткими и частыми трихомами, а также рядами промежуточных форм. Таким образом, изучение опушения в выборке образцов позволяет не только оценить его разнообразие у мягкой пшеницы и ее сородичей, но также и сравнить вклад таких факторов как хромосомный набор генома и уровень пloidности в формирование опушения листа.

Растения выращивали сотрудники Сектора генетики качества зерна ИЦиГ СО РАН в гидропонной теплице в ЦКП "Лаборатория искусственного выращивания растений" Института цитологии и генетики СО РАН с 14-часовым фотопериодом и дневной/ночной температурой воздуха 18/20 °C до и 20/22 °C после кущения. В качестве субстрата для питания растений использовался керамзит с раствором Кнопа (Shive, 1915). Подбор и выращивание биологического материала проводились сотрудниками Сектора генетики качества зерна под руководством Пшеничниковой Т.А.

2.4.2. Фенотипирование растений и обработка данных

Количественное определение адаксиального опушения поверхности листовой пластинки проводилось по описанной выше методике (раздел 2.3). Предфлаговый лист для каждого растения был отделен на стадии выхода в трубку. Он был сложен поперечно в трех равноудаленных положениях, так что верхняя сторона листа была снаружи. Изображения сгибов листьев были сделаны с помощью микроскопа Axioskop 2 Plus (Zeiss, Германия) с ПЗС-камерой AxioCam HRc. Все изображения были одинаковых размеров (1030 на 1300 пикселей, что соответствовало 2163 на 2730 мкм), поэтому полученные от них данные были сопоставимы для разных образцов. В результате фенотипирования оценивались:

- гистограмма распределения длины трихомов с интервалом длин 42 мкм;
- средняя длина трихомов на изображение l_m , в мкм;
- количество трихомов на изображении n_f , в шт.;
- индекс опушения, $H_{L/N}$ (формула 2.2), описывающий отношение размера трихомов (средняя длина, l_m) к плотности опушения (число трихомов, n_f). В наших образцах этот индекс изменялся от значений, близких к 1, до значений, превышающих 70.

Чтобы охарактеризовать опушение листьев у одного образца эти оценки мы усреднили по всем изображениям для выборки из 10-25 растений этого образца коллекции. Также были рассчитаны 95% доверительные интервалы для средней длины трихомов, их количества на изображении и количества в каждом интервале длин эмпирического распределения.

Для упрощенной классификации генотипов по характеристикам их опушения, мы проанализировали распределение средней длины трихомов, их

количества и индекса $H_{L/N}$ в исследованных генотипах растений и на основании этого отнесли значения этих параметров для каждого генотипа к одной из четырех категорий: очень низкие, низкие, умеренные и высокие. При определении категорий мы учитывали не только значение параметра, но и характер его распределения (например, мультимодальность).

Для более точной классификации образцов на основе характеристик опушения мы использовали кластерный анализ на основе сравнения векторов n , представляющих эмпирические распределения трихомов по длинам (см. раздел 2.3.4). При анализе оказалось, однако, что для образцов со слабым опушением многие интервалы гистограмм имели нулевые или близкие к нулю значения. Поэтому мы решили увеличить размер интервалов для гистограмм, снизив их количество до четырех: 0-50; 50-100; 100-1000; >1000 мкм. К преобразованным значениям вектора n мы применили иерархический кластерный анализ на основе использования Евклидова расстояния между векторами n_k и n_j , для образцов k и j (формула 2.4). На основе полученных таким образом значений расстояний была проведена кластеризация образцов методом невзвешенного среднего (UPGMA), реализованного в программе PAST (Hammer et al., 2001).

Как было указано выше, указанные образцы пшеницы представляют большой интерес не только для анализа разнообразия опушения, но также и для оценки влияния на опушение хромосомного набора (наличия в геноме трех субгеномов A, S(B,G), D) и уровня пloidности. Для оценки этого влияния на характеристики n_f , l_m , $H_{L/N}$, мы использовали метод дисперсионного анализа. Характеристики генома (наличие субгеномов A, B, D, уровень пloidности) использовались в качестве независимых переменных, а характеристики опушения в качестве зависимых. Для каждой пары зависимых и независимых

характеристик анализ проводился независимо. Всего было проведено 15 сравнений (5 независимых переменных и 3 зависимых). В случае геномной конституции мы рассматривали семь категорий этой переменной: AA, SS (BB, GG), DD, AABV, AAGG, AADD и AABVDD. Для уровня плоидности рассмотрено три категории значений: диплоид, тетраплоид, гексаплоид. Для типов геномов мы рассматривали две категории: наличие или отсутствие субгенома. Результаты анализа считались значимым при $p < 0.05$.

2.4.3. Количественные характеристики опушения листа пшеницы и ее сородичей

Результаты оценки количественных характеристик опушения для 47 образцов приведены в Приложении, табл. ПЗ. Из таблицы видно, что среднее число трихомов для образцов варьирует от 0 до 55, индекс опушения изменяется от 1 до 62, а средняя длина находится в диапазоне от 7 до 710 мкм.

Растения с хромосомным набором AA характеризовались количеством трихомов от малых (4) до самых высоких (54). Длина трихомов для большинства этих образцов варьировала умеренно от 30 до 70 мкм. Только два генотипа из 16 имеют относительно высокую среднюю длину трихомов. Это представители *T. boeoticum* (образцы №5 и 6) со значениями параметра $l_m = 185$ и 123 мкм соответственно. Индекс опушенности для большинства этих образцов не превышал 3. Таким образом, эти геномы характеризуются умеренной длиной трихомов с широкими вариациями их числа.

Количественные характеристики опушения растений с геномами SS и DD имели много общего. Это низкое значение числа трихомов (все образцы, кроме образца 18, имеют $n_f < 8$), длины трихомов (в основном более 100 мкм) и высокое значение индекса опушения (все значения более 16, кроме образца 18). Таким образом, опушение этой группы образцов можно охарактеризовать как

редкое с длинными трихомами.

Опушение у тетраплоидных растений было разделено на три группы. Образцы с хромосомным набором AABB включены в первую группу, которая характеризовалась очень редкими трихомами небольшой длины (все растения *T. durum*). Во вторую группу также вошли растения с геномом AABB, но с умеренным количеством трихомов средней длины (*T. dicoccoides*, *T. dicoccon*). В третью группу вошли все растения с хромосомными наборами AAGG и AADD, которые имели редкие, но очень длинные (более 400 мкм) трихомы и соответственно экстремальные значения $H_{L/N}$ (до ~60).

Опушение гексаплоидов характеризовалось умеренной длиной трихомов от 30 до 150 мкм, за исключением образца 39 (15 мкм). Количество трихомов варьировало от 0 до умеренных значений (40). Таким образом, несмотря на довольно широкие вариации, длина трихомов и их количество у растений этой группы не достигали экстремально больших значений. Индекс $H_{L/N}$ у гексаплоидов не превышал 11.

Рассмотрим распределение указанных параметров в выборке исследованных образцов. Они приведены на рис. 2.19. Степень проявления характеристик обозначена цветом кружков в оттенках серого (очень малые, малые, умеренные, высокие).

На гистограмме распределения количества трихомов на изображении (n_f , рис. 2.19 А) большинство изученных образцов имеют небольшое количество трихомов ($n_f < 15$). Оставшиеся образцы с количеством трихомов на изображении более 15 распределяются относительно равномерно по остальным интервалам длин. В интервалах, соответствующих очень малым, малым, умеренным и большим значениям попадает 16, 16, 6, и 9 образцов, соответственно (рис. 2.19 А).

Распределение образцов по средней длине трихомов характеризуется концентрацией практически 90% образцов в интервале длин 0-240 мкм (1/3 интервала значений). Оставшиеся 10% образцов распределены в области от ~400 до 720 мкм. Это растения с хромосомными наборами AAGG/AADD (Приложение, табл. ПЗ). Мы отнесли этот интервал к категории высоких значений длины трихомов. Интервал 0-240 мкм был разделен на три интервала одинаковой длины (очень малые, малые и умеренные значения). Таким образом, в интервалы с очень малыми, малыми, умеренными и высокими значениями длин трихомов попадают 26, 12, 4, и 5 образцов, соответственно (рис. 2.19 Б).

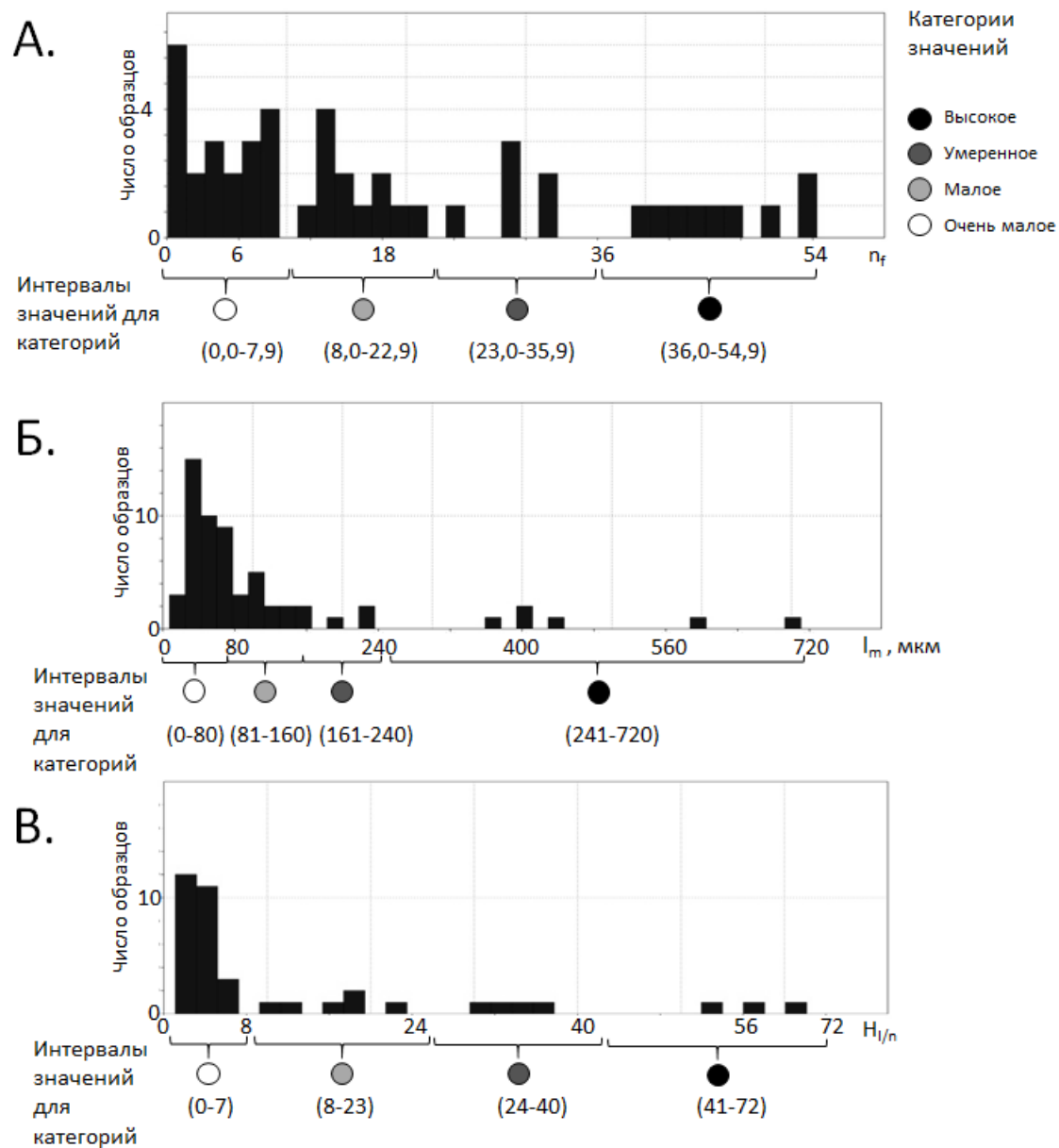


Рисунок 2.19. Гистограммы распределений параметров опушения n_f (А), l_m (Б) и $H_{L/N}$ (В) в 47 образцах пшениц. Оттенки серого для кружков внизу гистограмм отражают соответствующие интервалы для категорий параметров.

Величины индекса опушения H_{LN} (рис. 2.19 В) имеют пик распределения в интервалах низких значений (0-8). Небольшое количество генотипов имеют значения этого индекса в интервалах от 8 до 72 и разбивается на несколько кластеров. Таким образом, величины индекса опушения делятся на 4 интервала: 0-7, 8-23, 24-40, и 41-72. В эти интервалы попадают 26, 7, 4 и 3 генотипов, соответственно.

Полученная в результате иерархического кластерного анализа дендрограмма с выделенными на ней кластерами показана на рис. 2.20. Для каждого из кластеров справа показано характерное распределение длины трихомов. Было выявлено девять кластеров, 7 из которых включали по меньшей мере 4 образца; два кластера содержали 3 и 2 образца соответственно. Один генотип, № 46, *T. aestivum* сорт 'Саратовская 29', не был отнесен ни к одному из кластеров. Этот образец имеет достаточно высокое значение количества трихомов $n_f = 40$ шт. и высокое значение средней длины трихомов ($l_m = 147$ мкм). На рис. 2.20 показано, что качественные (соответственно категориям на рис. 2.19) и количественные (кластерный анализ) классификации образцов по характеристикам трихомов хорошо согласуются. Выявленные кластеры действительно представляют собой группы генотипов с однородным опушением.

Рассмотрим подробнее основные характеристики опушения листьев растений для выявленных нами кластеров. Как видно из рис. 2.20, 28 из 47 образцов относятся к категориям очень малых и малых (рис. 2.19) по числу трихомов и расположены в кластерах от С2 до С8. Среди этих кластеров особняком отстоит кластер С2, в который входили *T. timopheevii* и искусственно полученный образец *T. palmovae*. Он однороден по составу генома и содержит только представителей с тетраплоидными геномами ААGG/ААDD, которые не

встречаются в других кластерах. Характерной чертой этой группы являются редкие, но очень длинные трихомы.

Растения неопушенных образцов образовали кластер С5, который включает 7 тетраплоидных и гексаплоидных образцов, несущих геномы АА и ВВ. Среди них были все три образца *T. durum*. Четыре гексаплоида в этом кластере были представлены *T. spelta* (образцы 35 и 36) и *T. aestivum* (образцы 39 и 40). Интересно, что представители этого же геномного состава встречаются в других кластерах, характеризующихся заметным опушением.

Растения, с опушением представленным очень малым количеством трихомов находятся в кластерах С3, С4, С6 и С7. Кластер С7 содержит 4 диплоида с геномом АА, представленных *T. monococcum* (образцы 1, 2) и *T. boeoticum* (образцы 5, 7), а также два тетраплоида с геномами ААВВ (*T. dicoccoides*, *T. dicoccon*). Все растения в этом кластере, кроме образца 5 (*T. boeoticum*), имеют незначительное число и длину трихомов, а также индекс опушения. Такими же свойствами обладают растения кластера С4, включающие два образца: *T. monococcum* (№ 3) и *T. aestivum* (ААВВDD, № 41). Два других кластера С3 и С6 характеризуются редким опушением с умеренно длинными трихомами. Все они являются диплоидами с хромосомным набором SS/DD. Кластер С3 представлен видами *Ae. speltoides* (№№ 17, 19 и 20) и *Ae. tauschii* (№ 22), кластер 6 – только *Ae. tauschii* (№№ 21-23).

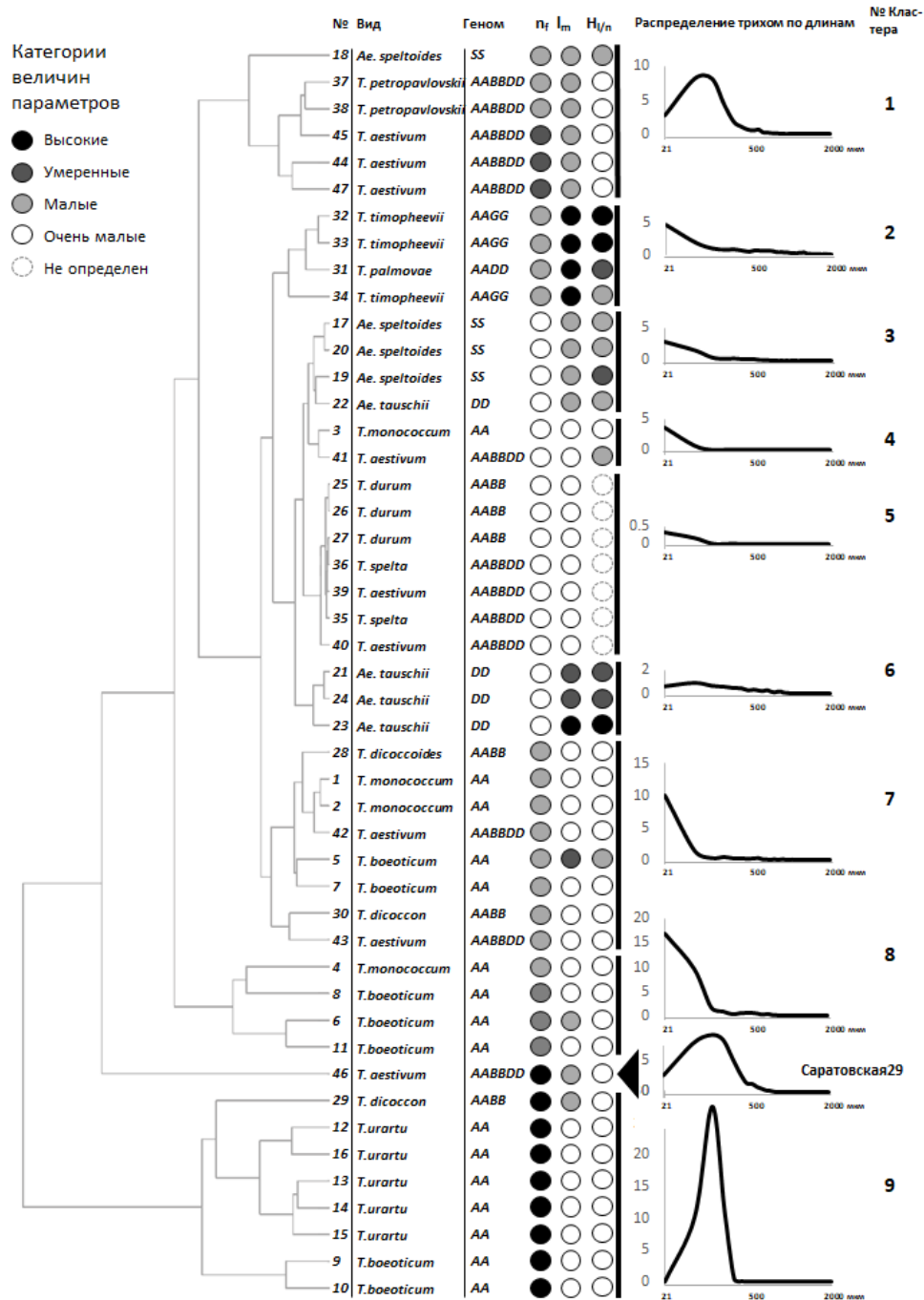


Рисунок 2.20. Результаты кластеризации образцов по распределению длин трихомов. Номера образцов (в соответствии с табл. ПЗ Приложения), названия видов, хромосомный набор, качественные характеристики опушения (согласно рис. 2.19) и распределение трихомов по длинам представлено для каждого кластера. Справа указаны номера кластеров (С1-С9).

Образцы, характеризующиеся интенсивным опушением, образовали кластер С1, включающий 6 членов. Из них все кроме одного (*Ae. speltoides*, геном SS, № 18), были гексаплоидными видами *T. aestivum* и *T. petropavlovskii*. Все они имеют умеренные значения параметра n_f и l_m ; значения $H_{L/N}$ были низкими. По числу трихомов растения кластера С1 были сходны с представителями кластера С8, но для последнего характерно отсутствие длинных трихомов. То же самое было характерной чертой кластера С9, у представителей которого было много коротких трихомов. За исключением образца № 29 (*T. dicoccon*, ААВВ), они были представлены геномами АА (*T. urartu* и *T. boeoticum*). Образец 46, *T. aestivum*, сорт ‘Саратовская 29’, не попал ни в какой из описанных выше кластеров. В то же время по плотности и длине опушения, а также по распределению длины трихомов он близок к представителям кластера С1.

Особенности распределения трихомов по длине, характерные для обнаруженных кластеров, отражают общие свойства опушения, описанные выше. Растения с интенсивным опушением имеют высокие значения числа трихомов (ось Y). При этом все графики можно разделить на 2 группы независимо от количества трихомов. К первой группе относятся растения со слабым опушением (кластеры С2-С7) и умеренным опушением (кластер С8), которые характеризуются наличием пика для распределения длины трихомов в области низких значений. Ко второй группе относятся растения с высокой плотностью опушения (кластеры С1, С9 и сорт ‘Саратовская 29’), которые характеризуются наличием максимума распределения в интервалах умеренной длины (50-100 мкм).

Мы построили диаграмму рассеяния проанализированных растений по величинам n_f и l_m . Она представлена на рис. 2.21.

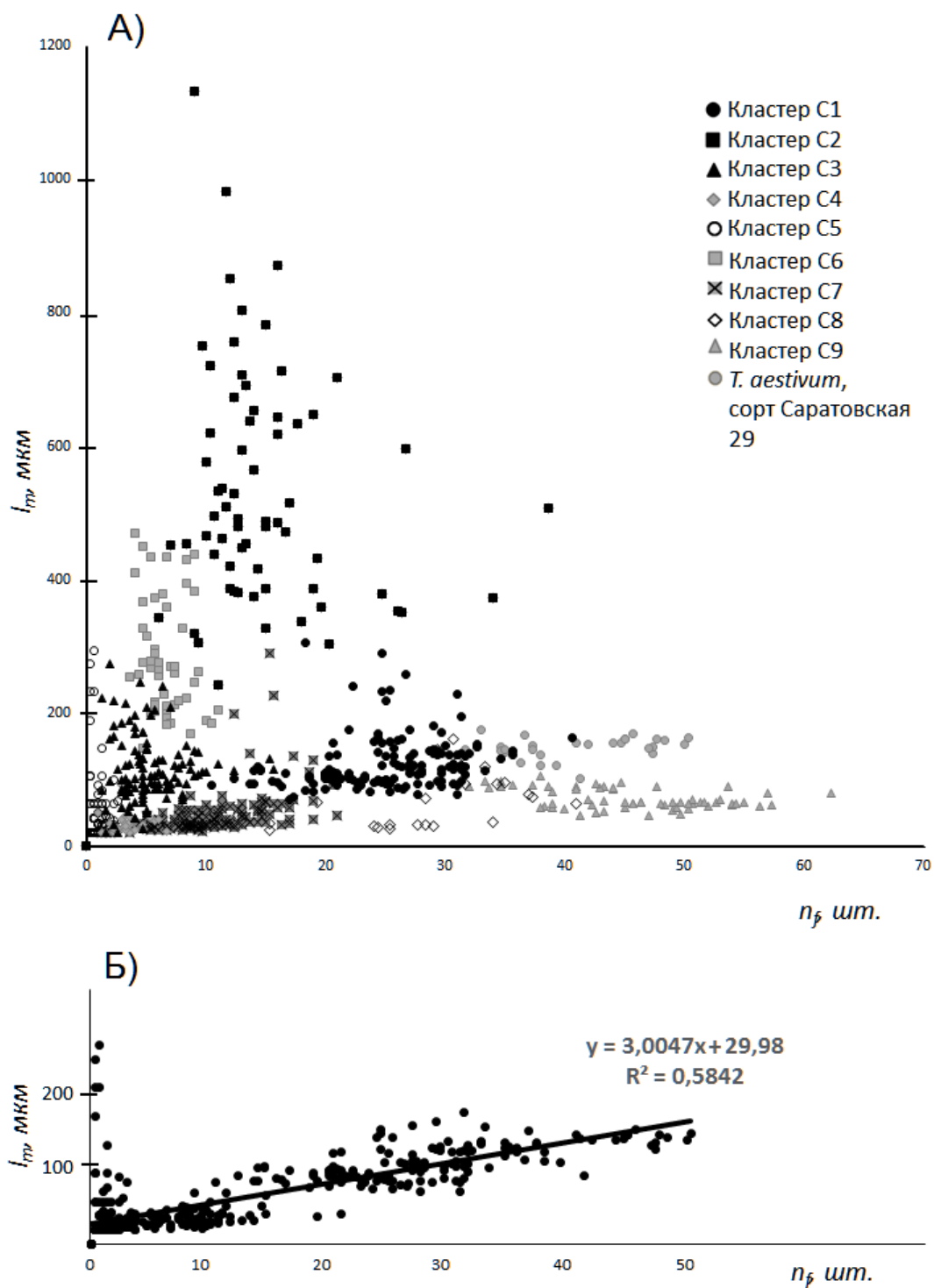


Рисунок 2.21. Диаграмма рассеяния значений средней длины (l_m) и среднего числа (n_f) трихомов для растений всех исследованных образцов (А) и для растений гексаплоидных видов (Б). Маркер определенной формы соответствует каждому кластеру из рис. 2.19.

На диаграмме рассеяния для средней длины (l_m) и среднего числа (n_f) трихомов всех исследуемых растений выявленные кластеры образовали компактную область (рис. 2.21 А). Например, растения, относящиеся к кластеру С5, расположены в области малых значений n_f и рассеяны практически вдоль оси Y. Растения, относящиеся к кластерам С2, С3 и С6, расположены в области малых значений n_f , но точки, которым они соответствуют, разбросаны также по вертикали. Растения, относящиеся к кластерам С8 и С9, расположены в области низких значений l_m и рассеяны вдоль горизонтального направления.

Растения, принадлежащие к гексаплоидным пшеницам (*T. aestivum*, *T. petropavlovskii* и *T. spelta*), расположены в основном вдоль линии, проходящей через начало координат (рис. 2.21 Б). Значения индекса опушения $H_{L/N}$ для них близки и колеблются от 3 до 5 (Приложение, табл. ПЗ). Точки, соответствующие *T. spelta*, расположены в области очень небольшого числа трихомов, но относительно высокой средней длины (300 мкм). Коэффициент корреляции Пирсона между n_f и l_m в группе гексаплоидных растений пшеницы составляет $r = 0.76$ ($p \ll 0.0001$). Если исключить из выборки гексаплоидов слабо опушенную *T. spelta*, коэффициент корреляции увеличивается до $r = 0.92$; ($p \ll 0.0001$). Во всей выборке растений такого высокого уровня значимости корреляции не наблюдалось ($r = 0.05$; $p > 0.1$). Растения родственных гексаплоидам видов из родов *Triticum* и *Aegilops* (расположенные преимущественно в кластерах С9, С3 и С6) распределены в большей степени вдоль осей X и Y и не представлены в областях диаграммы рис. 2.21 А, характерных для образцов опушенных гексаплоидных пшениц. Следует отметить, что для растений с длинным опушением ($l_m > 200$ мкм) высокой ($n_f > 20$) плотности трихомов не наблюдалось и наоборот, плотное опушение растений обычно характеризовалось средней длиной трихомов ($l_m < 200$ мкм).

На диаграмме нет точек, связанных с высокими значениями n_f (например, как для кластера 9) и, одновременно, высокими значениями l_m (как для кластера 2). Для растений мягкой пшеницы средняя длина трихомов наблюдалась в диапазоне 0-200 мкм, а среднее число трихомов - в диапазоне 0-50.

Для установления статистически значимых взаимосвязей между количественными характеристиками опушения листьев и структурными особенностями генома был проведен однофакторный дисперсионный анализ. Результаты приведены в табл. 2.3. В качестве фиксированных переменных использовались структура генома (табл. 2.3), уровень ploидности и наличие отдельно каждого из трех диплоидных геномов.

Таблица 2.3. Статистическая зависимость между характеристиками генома (хромосомным набором и уровнем ploидности), наличия/отсутствия субгеномов А, В (G,S), D и количественных характеристик опушения листьев (n_f , l_m , $H_{L/N}$).

Независимая переменная	Зависимая переменная		
	n_f	l_m	$H_{L/N}$
Хромосомный набор	<0,005**	<<0,005***	<0,05*
Уровень ploидности	0,1263	<0,05*	<0,01*
А субгеном	<0,05*	0,2982	0,9584
В (G,S) субгеном	<0,01*	0,5762	0,05631
D субгеном	<0,05*	<0,005**	<0,005**

Как видно из табл. 2.3, фактор хромосомного набора оказывал существенное влияние на все три характеристики опушения. Что касается уровня ploидности, то он был значимо связан с длиной трихомов. Плотность опушения (число трихомов на изображении) значимо зависела от типа генома

для диплоидных видов, тогда как длина опушения и индекс опушения $H_{L/N}$ были подвержены влиянию только наличия/отсутствия D-субгенома.

2.4.4. Разнообразие опушения листа у мягких пшениц и их сородичей и его связь со структурой генома

Формирование аллополиплоидных видов растений сопровождается значительными структурными и функциональными изменениями в геноме (Feldman, Levy, 2012; Feldman et al., 2012). Они заключаются не только в перестройках и элиминации некодирующих частей генома, но и сопровождаются потерей или подавлением генов, определяющих фенотипические признаки растений.

В данном разделе мы исследовали разнообразие важной морфологической характеристики растений – опушение листьев у ди-, тетра- и гексаплоидных видов, относящихся к родам *Triticum* и *Aegilops*. Мы показали, что на характеристики опушения статистически значимо влияет структура генома. Это означает, что виды с разными наборами субгеномов геномов различаются по типу опушения. Уровень плоидности влияет только на длину трихомов и индекс опушения $H_{L/N}$. Это отражается более длинным опушением у *T. timopheevii*, чем у донора одного из его геномов, *Ae. speltoides*, и слабым опушением у *T. durum* по сравнению с родственными ему видами. На изменение количества трихомов (плотность опушения) оказывает значимое влияние наличие отдельных субгеномов А и В, тогда как наличие/отсутствие генома D существенно влияет на все изученные параметры опушения. На рис. 2.22 суммированы полученные нами результаты по количественным характеристикам опушения у представителей родов *Triticum* и *Aegilops*. На диаграмме представлены хромосомный набор исследуемых образцов (набор субгеномов), их происхождение и характерные особенности опушения.

Очевидно, что диплоидные виды, носители генома АА, характеризуются большим разнообразием в плотности опушения.

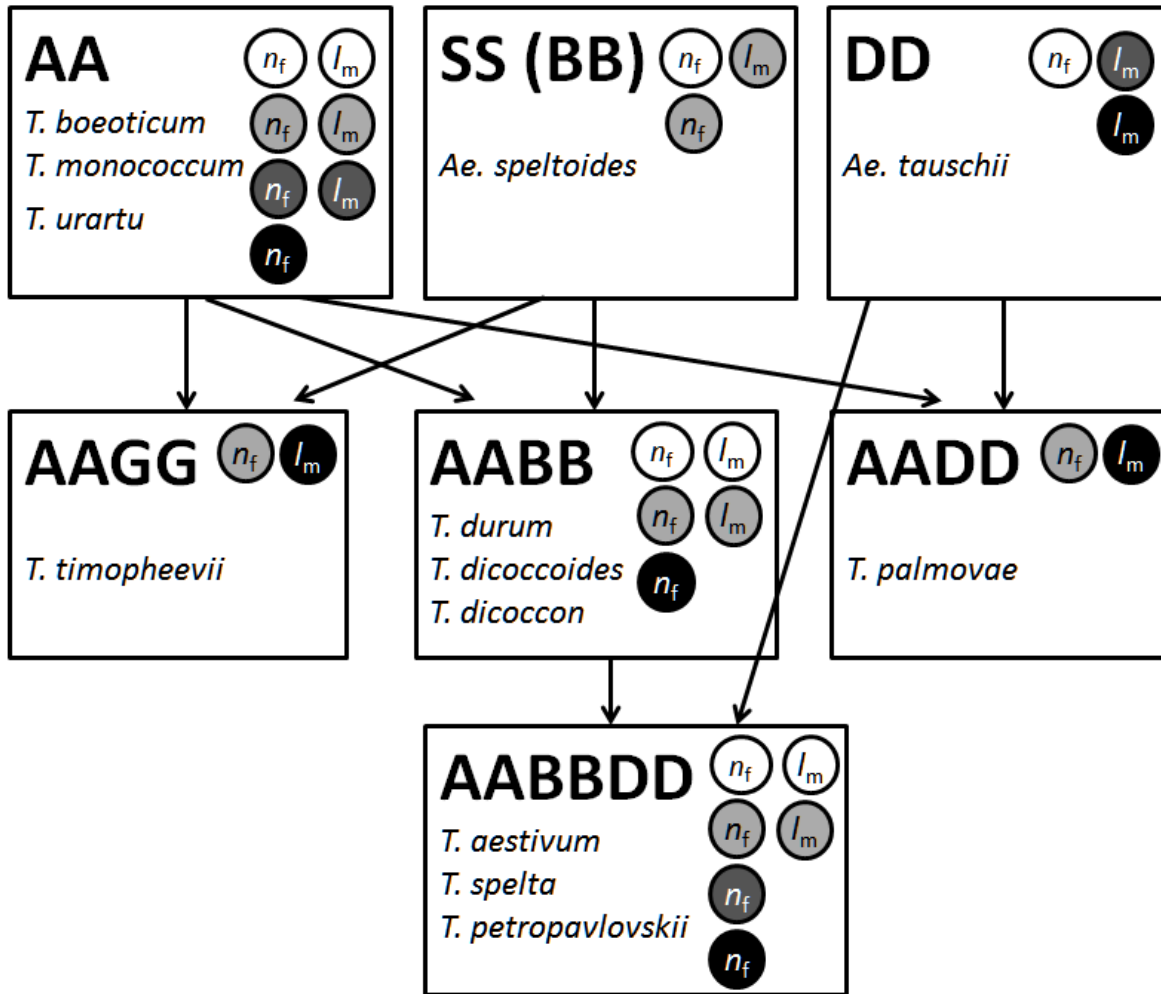


Рисунок 2.22. Распределение количественных особенностей опушения в изученных нами видов для родов *Aegilops* и *Triticum* различных хромосомных наборов. Происхождение тетраплоидных и гексаплоидных геномов показано стрелками. Шкала серого для категорий величин признаков соответствует кодировке на рис. 2.19, 2.20.

Виды *T. boeoticum* и *T. urartu* характеризуются плотным опушением (они образуют отдельный кластер С8, рис. 2.20). *T. monosocum* характеризуется низкой плотностью трихомов, при этом длинные трихомы отсутствуют, а короткие преобладают (см. также Приложение, табл. П3). Представители этого вида расположены в разных кластерах. Образцы рода *Aegilops*, носители геномов SS и DD, характеризуются умеренно длинным и длинным, но редким опушением и образуют независимые кластеры (С3 и С6, рис. 2.20). Вероятно, эти межгеномные и межвидовые различия являются результатом длительного периода около 3 миллионов лет независимой эволюции видов этих родов (Gill et al., 2004). Аллотетраплоидные геномы характеризуются различными морфологическими признаками опушения (рис. 2.20). Среди тетраплоидов, носителей генома AABB и прямых предков гексаплоидных видов, наблюдаются образцы с наиболее слабым опушением из всей выборки. Именно культурные формы *T. durum* попали в кластер С5 вместе со слабо опушенными представителями гексаплоидов.

Можно предположить, что гены, контролирующие опушение, были элиминированы или подавлены в процессе одомашнивания тетраплоидов (Matsuoka, 2011). У диких и культурных эммеров обнаружено более плотное и длительное опушение, и эта изменчивость привела к разделению этих образцов на несколько кластеров по типу опушения (рис. 2.20). Образцы-носители генома AAGG образовали отдельный кластер и демонстрируют самое длинное опушение с редкими трихомами. Так, тетраплоидные представители двух дивергентных линий, носители геномов AABB и AAGG (Matsuoka, 2011) резко различаются по типу опушения. Специфическое длинное опушение характерно для эндемичного вида *T. timopheevii*, обитающего в горах Кавказа. Его адаптивное значение еще предстоит установить.

Тип опушения *T. timopheevii* характерен для искусственно синтезированного тетраплоидного вида *T. palmovae*, имеющего общий с этим видом геном AA, но несущего геном DD вместо GG. Вероятно, длинные трихомы *T. palmovae* были унаследованы от *Ae. tauschii*. Согласно полученным в настоящем исследовании данным, носители генома DD имеют самые длинные трихомы. Следует отметить, что средняя длина трихомов у *T. timopheevii* и *T. palmovae* достоверно выше, чем у любого из исследованных образцов *Ae. speltoides* и *Ae. tauschii*. Это может указывать на некоторую модификацию экспрессии генов опушения в сложном геноме *T. palmovae* и *T. timopheevii*. Аналогичное явление мы наблюдали в изогенных линиях мягкой пшеницы сорта ‘Саратовская 29’, несущих дополнительный ген опушения *H12^{aesp}* от *Ae. speltoides* (Дорошков и др., 2016). Вновь полученные линии имели в среднем более длинные трихомы по сравнению с исходными родительскими формами.

Опушение гексаплоидов с геномной конституцией AABBDD характеризуется умеренной длиной трихомов (рис. 2.20). При этом, в отличие от ди- и тетраплоидных видов, плотность опушения показывает высокую изменчивость от очень слабого опушения у *T. spelta* до интенсивного опушения у сортов мягкой пшеницы ‘Саратовская 29’ и ‘Хонг-Манг-май’. Поэтому часть изученных образцов были сгруппированы в один кластер, остальные распределены между представителями гексаплоидов (рис. 2.19, 2.20). В то же время морфотип опушения гексаплоидных пшениц ($l_m > 100$ мкм при $n_f > 35$) не встречается у ди- и тетраплоидных родичей пшеницы в нашей выборке (рис. 2.20). Это говорит в пользу того, что гексаплоидная пшеница также испытала потерю части разнообразия, характерного для родительского вида на стадии формирования, а именно морфотипа опушения с длинными трихомами.

Мягкая пшеница демонстрирует наибольшую изменчивость среди изученных гексаплоидов по плотности опушения от $n_f = 0,54$ для сорта 'Родина' до $n_f = 40,5$ для сорта 'Саратовская 29'. В нашей работе мы описали опушение только у яровых сортов различного происхождения.

Уникальным среди всех изученных сортов является сорт 'Саратовская 29', который в иерархическом анализе образовал отдельный кластер (рис. 2.20). С помощью генетического анализа было показано, что этот генотип несет в себе два гена опушения листьев (Дорошков и др., 2011).

Результаты проведенного впервые массового сравнения количественных характеристик опушения более чем для 40 образцов ди-, тетра - и гексаплоидных видов пшеницы, а также видов рода *Aegilops* хорошо согласуются с качественными оценками изменчивости рода *Triticum* по этому признаку, описанными Н. Вавиловым (Вавилов, 1990), который, несомненно, изучил гораздо больше образцов обоих родов. Он отметил, что все диплоидные виды имеют либо бархатистое, либо длинное реснитчатое опушение, что соответствует нашим количественным описаниям опушения у носителей генома AA (рис. 2.20, *T. monococcum* или *T. boeoticum*). Тетраплоиды, по его словам, чаще всего неопушенные или имеют короткие бархатистые (*T. durum*, *T. dicoccon.*, *T. dicoccoides*, рис. 2.20) или длинное реснитчатое опушение (*T. timopheevii*, рис. 2.20). Гексаплоиды в его работах описываются прежде всего, как носители короткого бархатистого опушения, но не длинного-реснитчатого.

2.5. Изучение фенотипического проявления генов опушения листа пшеницы *Triticum aestivum* L.

К началу диссертационной работы у пшеницы было известно два гена, контролирующих опушение листа. С помощью генетического анализа было показано, что сорт ‘Саратовская 29’ несет в себе два гена опушения листьев (Дорошков и др., 2011). Один из них, *HL1*, был картирован на хромосоме 4В (Dobrovolskaya et al., 2007), хромосомная локализация второго гена, *HL3* была неизвестна. Для изучения влияния указанных генов на опушение листа в секторе генетики качества зерна были подобраны генотипы, которые несут различные аллели генов, контролирующих опушение у пшеницы. Это позволило впервые изучить проявление этих генов как в совокупности, так и поотдельности на характер опушения листа.

2.5.1. Растительный материал и условия выращивания

Генетические особенности изученного материала, его происхождение, а также количественные характеристики опушения листьев подробно описаны в работе Дорошкова и соавт. (Doroshkov et al., 2016). Выборка была представлена тремя неродственными яровыми сортами: ‘Янецкис Пробат’ (ЯП), ‘Родина’ и ‘Саратовская 29’ (С29), из которых последний сорт является густоопушённым носителем двух генов опушения *HL1* и *HL3* (Doroshkov et al., 2011). На его генетическом фоне созданы: межсортовая замещённая линия С29 (ЯП 4В), донором хромосомы 4В и рецессивного аллеля гена *HL1* которой является сорт ЯП, а реципиентом является линия С29, а также две изогенные линии *i:C29 glabrous* и линия *i:C29 HL2^{aesp}*, несущая дополнительный к двум упомянутым генам сорта-реципиента ген опушения листа. В выборку включена также интрогрессированная линия 102/00ⁱ из коллекции «Арсенал» (Larochkina, 2001) и *Ae. speltoides* Tausch (образец k-389 из коллекции ВИР, г. Санкт-

Петербург). Линия 102/00ⁱ несёт ген *HL2^{aesp}* на фоне генотипа сорта ‘Родина’, у которого практически полностью отсутствует опушение. Донором опушения листа для линии 102/00ⁱ явился упомянутый выше образец *Ae. speltoides* Tausch. Опушение листа этой линии контролируется геном *HL2^{aesp}*, расположенным в 7В хромосоме (Pshenichnikova et al., 2007; Dobrovolskaya et al., 2007).

Аллельный состав генов опушения и их качественное фенотипическое проявление представлены на рис. 2.23.

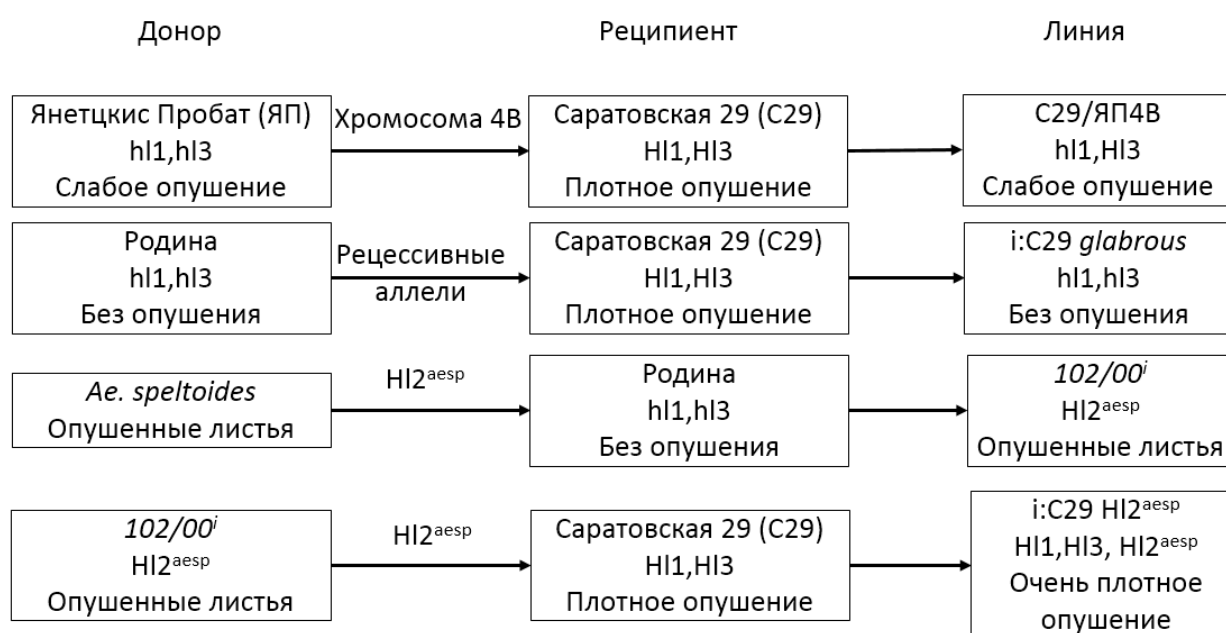


Рисунок 2.23. Генотипы пшеницы-реципиента и донора - носители различных аллелей опушения листьев и их фенотипов

Наблюдения и гибридизация проводились в условиях гидропонной теплицы ЦКП «Искусственного выращивания растений» Института цитологии и генетики СО РАН. Во время выращивания поддерживались температура, водный режим и режим освещения, необходимые для выращивания яровой пшеницы. В качестве субстрата был использован керамзит, в качестве питательного раствора - раствор Кнопа (Shive, 1915). Подбор и выращивание

биологического материала проводились сотрудниками Сектора генетики качества зерна под руководством Пшеничниковой Т.А.

2.5.2. Фенотипирование растений и обработка данных

Опушение листа анализировалось на адаксиальной поверхности предфлагового листа растения в фазу трубкования или колошения. Для определения количественных характеристик опушения использовалась прогпама LHDetect 2 (Genaev et al., 2012). С каждого анализируемого растения был взят один предфлаговый лист, сделано три препарата сгибов в середине, у основания и вершины листа согласно протоколу (Genaev et al., 2012). Снимки сгибов листа были сделаны при помощи микроскопа Axioskop 2 Plus (Zeiss, Germany) с CCD-камерой AxioCam HRc в Центре коллективного пользования микроскопического анализа биологических объектов СО РАН. В результате анализа были получены гистограммы распределения трихомов по длинам для каждого изображения. В настоящей работе использовались гистограммы распределения трихомов по длинам с шагом 42 мкм. В нашем масштабе это значение эквивалентно 20 пикселям изображения.

На основе гистограмм для каждого растения вычислялись средняя длина трихомов l_m и среднее количество трихомов на изображении сгиба листа n_f , которое характеризует плотность опушения (Genaev et al., 2012). Значения в каждом интервале гистограмм всех растений одного генотипа, представленных в виде численного массива, усреднялись и рассчитывались 95% доверительные интервалы. Полученные гистограммы распределения трихомов по длинам позволяют оценить соотношение трихомов разной длины, характерное для каждого исследованного генотипа. Для сравнения морфологии опушения листа использовалось 25-30 растений каждого сорта и линии.

2.5.3. Анализ опушения листа

В табл. 2.4 представлены оценки среднего количества и длины трихомов на листьях. С29 - сильно опушенный сорт; он несет как *H11* (Дорошков и др. 2011; Майстренко, 1976; Dobrovolskaya et al., 2007), так и второй ген, *H13*, с неизвестной хромосомной локализацией. Линия С29 моносомная по хромосоме 4В, несет только одну копию *H11* (табл. 2.4) и ее опушение менее выражено слабее, чем у С29. Количество трихомов на листьях у ЯП примерно в 11 раз меньше, чем у С29, а средняя длина трихомов на листьях более чем в три раза меньше (табл. 2.4). ЯП не несет ни *H11* ни *H13*. Таким образом, у межвидовой линии с заменой одной хромосомы С29/ЯП 4В отсутствует *H11*, а трихомы на ее листьях короче и менее многочисленны, чем у сорта С29. Моногенный контроль интрогрессированной опушенности листьев был продемонстрирован при скрещивании линии 102/00ⁱ со слабо опушенным сортом 'Диамант 2' (Pshenichnikova et al., 2007). Эта линия, как и *Ae. speltooides* несет гены *H12^{aesp}*, которые сами по себе дают менее плотное опушение, чем у С29, несущего как *H11*, так и *H13*. Плотность опушения листьев в линии 102/00ⁱ превышала таковую у линии *Ae. speltooides* k-389 на 40%, а средняя длина трихомов на листьях была примерно вдвое больше, чем на листьях k-389. Листья линии i:С29 *H12^{aesp}*, которая несет *H12^{aesp}* в дополнение к *H11* и *H13*, развили наиболее интенсивное опушение из всех образцов (табл. 2.4). Среднее количество трихомов на изображении превысило таковое у С29 на 9%, а у 102/00ⁱ - в четыре раза. Средняя длина трихомов линии i:С29 *H12^{aesp}* превышала таковую у сорта 'Саратовская 29' на 58%, а у линии 102/00ⁱ - на 13%. Листья сорта 'Родина' были лишь слегка опушены (табл. 2.4). Неопушенная почти изогенная линия i:С29 лишена как *H11*, так и *H13*, и характеристики ее опушения сопоставимы с таковыми для Янетцкис

Пробат (табл. 2.4).

Таблица 2.4. Сравнение количественных характеристик опушения на листьях линий пшеницы, несущих различные комбинации генов *HL*.

Генотип	Набор генов контроля опушения	n_f	l_m (мкм)
C29	<i>HL1 HL1. HL3 HL3</i>	40.5±2.2	147.1±5.9
C29/ЯП4В	<i>hl1 hl1; HL3 HL3</i>	16.5±0.8	65.5±3.1
ЯП	<i>hl1 hl1; hl3 hl3</i>	3.0±0.5	31.4±5.0
<i>Ae. speltoides</i>	<i>HL2^{aesp} HL2^{aesp}</i>	6.2±1.3	97.6±17.0
102/00 ⁱ	<i>hl1 hl1; hl3 hl3; HL2^{aesp} HL2^{aesp}</i>	10.2±0.8	205.9±13.8
<i>i</i> :C29 <i>HL2^{aesp}</i>	<i>HL1 HL1. HL3 HL3. HL2^{aesp} HL2^{aesp}</i>	44.2±1.2	231.9±5.0
‘Родина’	<i>hl1 hl1; hl3 hl3</i>	0.5±0.2	16.0±4.9
<i>i</i> :C29 <i>glabrous</i>	<i>hl1 hl1; hl3 hl3</i>	6.1±0.8	31.6±2.3

Таблица 2.5. Влияние генотипа на среднюю длину и количество волосков на листьях. Результаты многомерного дисперсионного анализа MANOVA/CVA. Группирующая переменная: генотип.

Лямбда Уилкса	8.008324	След Пиллаи:	1.75
df1:	16	df1:	16
df2:	678	df2:	680
<i>F</i> :	422.1	<i>F</i> :	298.1
<i>P</i>	0	<i>P</i>	2.09 10 ⁻²⁹⁴

Анализ показал положительную связь между длиной и количеством трихомов на поверхности листьев во наборе всех образцов ($r=+0.66$, $p<0.005$). Многомерный дисперсионный анализ выявил значительную зависимость средней длины и количества трихомов на листьях от генотипа растения (табл. 2.5).

Это также подтвердило значительные парные различия ($p<0.005$) для каждой пары образцов, за исключением ЯП/’Родина’ и ЯП/i:C29 *glabrous* (табл. 2.6).

Показатели средней длины трихомов на листьях и их среднего числа не могут полностью охарактеризовать опушение листьев, поскольку они не отражают изменчивость, присутствующей в длине трихомов на листьях. Поэтому, дополнительно, их совместное распределение было проанализировано для всех генотипов, и они были сгруппированы на диаграммах рассеяния.

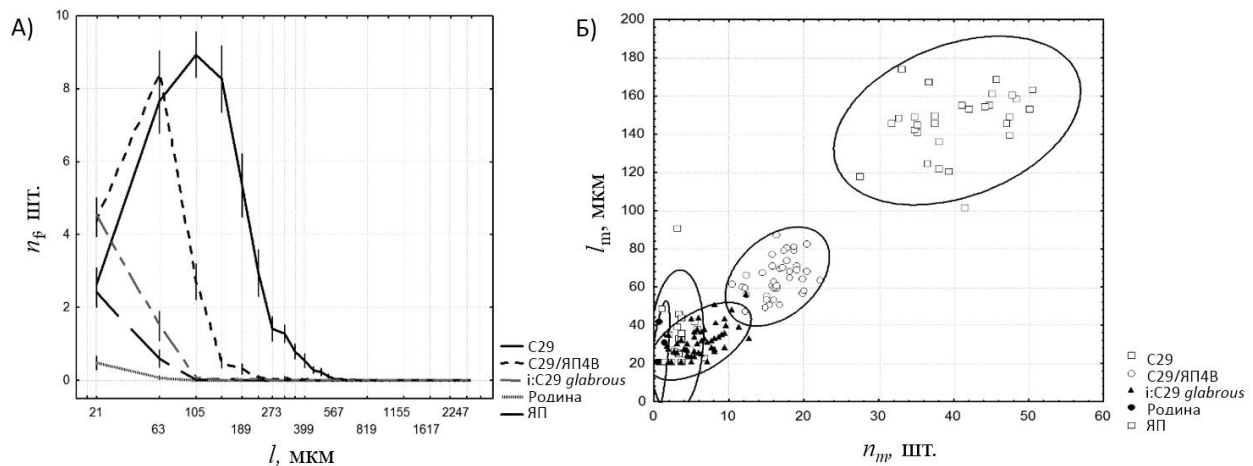


Рисунок 2.24. Количественная оценка опушения листьев у сортов ‘Родина’, ‘Янетцкис Пробат’ (ЯП), ‘Саратовская 29’ (С29), неопушенной линии *i:C29 glabrous* и линии, полученной на основе сорта ‘Саратовская 29’ с межвидовой заменой одной хромосомы С29/ЯП4В. (А) Распределение длины трихомов. Шкала является логарифмической. Вертикальные линии обозначают 95% доверительный интервал для каждого значения длины трихомов. (В) Диаграмма рассеяния среднего числа (ось X) и средней длины трихомов (ось Y) на изображениях. Эллипсы определяют области доверительного интервала 95%.

У сорта С29 модальная длина составляла 105 мкм, в то время как самые длинные трихомы листа имели длину 693 мкм (рис. 2.24 А). Отсутствие *Hll*, как и в С29/ЯП4В, увеличивает долю более коротких трихомов, так что модальная длина составила всего 63 мкм, а максимальная длина составляла всего 399 мкм. Таким образом, основной компонент изменения опушения листа у этой линии был обусловлен измененной пропорцией длинных трихомов.

Распределение длины трихомов листьев у растений с генотипами, в которых отсутствовали *H11* и *H13*, характеризовалось модой, близкой к нулю, и максимальной длиной менее 105 мкм. У сорта 'Родина' максимальная длина трихомов составила менее 63 мкм. Различия в высоте гистограмм соответствовали различиям в плотности опушения (рис. 2.24 А, табл. 2.7). Анализ диаграмм рассеяния (рис. 2.24 Б) показал, что растения сорта 'Родина' занимают компактные области, определяемые минимальной длиной трихомов листьев 0-60 мкм и средним количеством 0-6 трихомов на изображение. Опушение ЯП было несколько интенсивнее, чем у сорта 'Родина'. Точки, соответствующие неопушенной линии *i:C29 glabrous* располагаются в той же области, но немного смещены по числу трихомов и их количеству в большую сторону. Линия С29/ЯП4В несет *H13*, но не *H11*. На диаграмме рассеяния (число/длина трихомов) ее растения занимают область, не пересекающуюся с сортом 'Родина' и ЯП, но пересекающуюся с областью рассеяния линии *i:C29 glabrous*. Образцы С29 расположены в отдельной области и характеризуются длиной трихомов 100-180 мкм и средним числом 26-52 трихомов на изображение.

Аналогичный анализ для носителей *H12^{aes}* представлен на рис. 2.25. Линия 102/00ⁱ не отличается от линии *Ae. speltoides* k-389 по диапазону длины трихомов (0-500) мкм. Оба генотипа показали моду около 21 мкм, но они различаются по количеству длинных трихомов и их максимальной длине. Самые длинные трихомы на листьях k-389 составляли 525 мкм, тогда как линия 102/00ⁱ была способна формировать на листьях трихомы длиной до 1323 мкм (рис. 2.25 А).

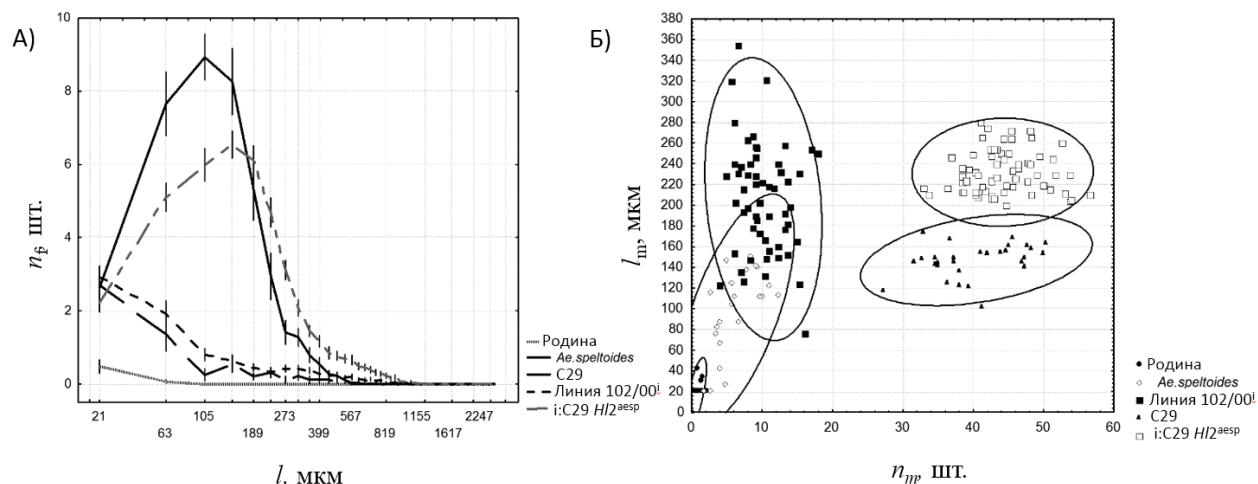


Рисунок 2.25. Количественная оценка опушения листьев в линии 102/00ⁱ линия, линии i:C29 *HL2*^{aesp} и их родительские формы сорт 'Родина', 'Саратовская 29' (C29) и *Ae. speltoides*, k-389. (А) Распределение длин трихомов. Шкала является логарифмической. Вертикальные линии обозначают 95% доверительный интервал для каждого значения длины трихомов. (Б) Диаграмма рассеяния среднего числа (ось X) и средней длины трихомов (ось Y) на изображениях. Эллипсы определяют области доверительного интервала 95%.

Анализ диаграммы рассеяния (рис. 2.25 Б) показал, что области диаграммы для генотипов 102/00ⁱ и *Ae. speltoides* k-389 перекрывались но демонстрировали значительные различия в длине трихомов, варьирующие от 0 до 360 мкм. Оба генотипа имели плотность опушения листьев в диапазоне от 0 до 18 трихомов на изображение. Сравнение гистограмм распределения длины трихомов у C29 и i:C29 *HL2*^{aesp} показало, что последний образец формировал большее количество листовых волосков длиной более 200 мкм. У него образовалось больше длинных трихомов на листьях, чем у линии 102/00ⁱ (рис. 2.25 Б). Максимальная длина волосков листьев для линии i:C29 *HL2*^{aesp} (1323 мкм), также была больше, чем у C29 (693 мкм). Точки, соответствующие образцам линии i:C29 *HL2*^{aesp} на диаграмме рассеяния (рис. 2.25 Б) находятся далеко от положения C29 из-за большей средней длины трихомов (области

рассеяния для этих двух генотипов практически не перекрываются).

Набор сравниваемых сортов и линий сильно различался в отношении опушения листьев. Плотное опушение листьев у С29 отражает присутствие *HL1* и *HL3*, в то время как эффект *HL2^{aesp}* заключался в стимулировании роста более редких, но более длинных трихомов. Модель, согласовывающая количественные различия в опушении с содержанием генов *HL*, представлена на рис. 2.25. Относительное влияние каждого из генов *HL* как на длину, так и на количество волосков в листьях было определено количественно путем вычисления параметра $K_{L/N}$ (формула 2.3), отражающего связь между длиной волосков в листьях и плотностью (табл. 2.7).

Таблица 2.7. Влияние различных комбинаций генов контроля опушения *HL* на отношение длины трихомов к их количеству ($K_{L/N}$).

Ген	Доминантный и рецессивный генотипы	$K_{L/N}$
<i>HL1</i>	С29 и С29/ЯП4В	3.4
<i>HL3</i>	С29/ЯП4В и 'Родина'	3.09
<i>HL2^{aesp}</i>	Линия 102/00 ⁱ и 'Родина'	19.6
<i>HL2^{aesp}</i>	i:С29 <i>HL2^{aesp}</i> и С29	22.9

Отсутствие *HL1*, но наличие *HL3*, как у линии С29/ЯП4В, уменьшает как длину, так и количество волосков на листьях (рис. 2.24, табл. 2.4). *HL1* оказывает дозированный эффект, поскольку опушение было несколько снижено, когда ген присутствовал только в однократной дозе (как в моносомной линии С29/ЯП4В). Отсутствие *HL3* и *HL1* (как у неопушенной линии i:С29 *glabrous*) привело к значительной потере как длины, так и количества трихомов. Распределения сорта 'Родина' и линии С29/ЯП4В на

диаграмме рассеяния, а также сравнение соответствующих значений $K_{L/N}$ позволяют предположить, что механизм действия $Hl3$, вероятно, такой же, как и у $Hl1$. Значения $K_{L/N}$, связанные с двумя генами Hl , находятся в пределах 10% друг от друга, что означает, что они вызывают сходный фенотипический эффект. Хотя генотипы ЯП и *i:C29 glabrous* являются носителями комбинаций рецессивных аллелей $hl1:hl1/hl3:hl3$, оба, тем не менее, демонстрируют определенную степень опущения листьев; это подразумевает существование дополнительных, пока неизвестных генов, лежащих в основе контроля признака.

$Hl2^{aesp}$ отвечает за опущение у *Ae. speltoides* k-389 и в линии 102/00ⁱ (табл. 2.4, рис. 2.25). Его взаимодействие с $Hl1$ и $Hl3$ было исследовано путем создания почти изогенной линии, несущей все три доминантных аллеля генов опущения Hl . Результатом действия всех трех доминантных аллелей генов Hl было увеличение доли длинных трихомов и, в меньшей степени, общего количества трихомов (рис. 2.24 Б). Значение $K_{L/N}$, связанное с $Hl2^{aesp}$, было более чем в пять раз выше, чем значение, связанное с $Hl3$ или $Hl1$ (табл. 2.7). Относительно небольшая разница между значениями $K_{L/N}$ $Hl2^{aesp}$, рассчитанными для линий 102/00ⁱ и *i:C29* $Hl2^{aesp}$, позволяет предположить независимую активность гена $Hl2^{aesp}$ в присутствии $Hl1$ и $Hl3$.

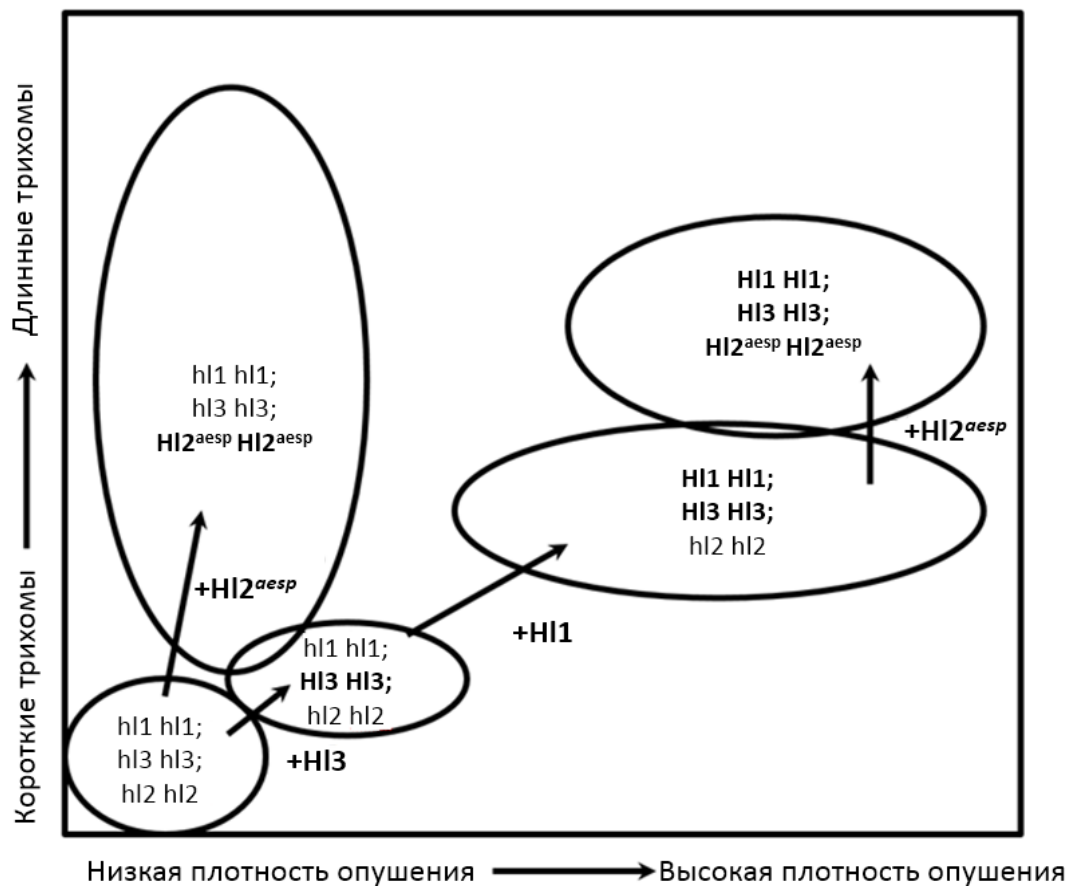


Рисунок 2.26. Модель влияния генов *H11*, *H13* и *H12^{aesp}* на длину и количество трихомов на листе пшеницы.

Общая схема влияния генов аллельных комбинаций трех генов пшеницы на опушение листа отражена на рис. 2.26: ген *H12^{aesp}* в большей степени оказывает влияние на длину трихомов. Возможно его функция связана с изменением размера клеток трихомов. Гены *H13* и *H11* влияют одновременно и на длину трихомов и на их количество. Вероятно, они затрагивают как пути инициации трихомов на листе, так и роста клеток отдельных трихомов.

Таким образом, разработанные методы фенотипирования позволили выявить комплексную природу действия трех генов на развитие опушения

листа у пшеницы.

2.6. Анализ растений замещенных линий пшеницы (*Triticum aestivum* L.) в условиях нормального и ограниченного водообеспечения.

Известно, что опушение листьев растений играет важную физиологическую роль, обеспечивая микроклимат листа у его поверхности, защищая лист от поражающего воздействия излучения. Поэтому большой интерес представляет оценка связи интенсивности и характера опушения листьев на физиологические характеристик растений.

2.6.1 Растительный материал, условия выращивания, фенотипирование растений, методы анализа данных

Растительный материал использованный в этом разделе описан выше (см раздел 2.5.1 «Растительный материал и условия выращивания»). Растения были выращены в условиях фитотрона СИФИБР СО РАН в двух независимых повторностях. Сроки посева повторностей различались на 1 неделю. Растения (по 6 – 10 зерен на линию) выращивали в сосудах Митчерлиха, заполненных 4 кг почвенного субстрата. Фотопериодический режим день/ночь был 18/6 часов, температурный режим поддерживался на уровне 10-15° С ночью и 20-23° С днем. Освещение в течение дня естественное, подсветка в вечерние и утренние часы 400-500 мкмоль/м²с, влажность воздуха 40-50%. В каждом независимом эксперименте растения выращивали на двух режимах водообеспечения, оптимальном и воддефицитном. Оптимальный режим соответствовал 60 %, а дефицитный – 30 % воды от полной влагоемкости почвы, которую определяли, как описано в руководстве (Журбицкий, 1968). Водный режим в период выращивания растений поддерживали весовым методом, взвешивание проводили дважды в неделю. Все физиологические показатели измеряли на

стадии стеблевания на флаговых и предфлаговых листьях растений. Также на этой стадии определяли биомассу главного побега. Ответ растения на водный стресс оценивался на основе индекса толерантности, рассчитанного по формуле $T_d/T_c \times 100$, где T_d это среднее значение веса сухой биомассы для растений под воздействием засухи, а T_c та же величина для растений в нормальных условиях.

С помощью портативной системы для изучения газообмена и флуоресценции хлорофилла GFS-3000 (Walz, Германия) были измерены следующие показатели: скорость транспирации, устьичная проводимость, скорость ассимиляции CO_2 (или нетто-фотосинтез), фоновая флуоресценция хлорофилла, потенциальная и реальная эффективность фотосинтеза, скорость транспорта электронов в фотосистеме II, нефотохимическое тушение флуоресценции. Эффективность использования воды (WUE) была рассчитана как отношение нетто фотосинтез/транспирация. Сокращенные название признаков и единицы измерений представлены в табл. 2.8. Параметры газообмена и флуоресценции хлорофилла были измерены в листьях шести растений каждого генотипа в каждой из повторностей и усреднены для каждого генотипа в каждой из повторностей.

Также для шести растений были измерены длина и ширина предфлагового листа на стадии начала колошения. Из этих листьев были приготовлены влажные микропрепараты средней части листа, на которых было посчитано число устьиц в поле зрения микроскопа на верхней и нижней стороне. Использовался объектив 16× и окуляры 10×. Диаметр поля зрения был 0.95 мм, что соответствует площади 0,71 мм². Эти данные были усреднены по всем растениям и нормированы на 1 мм² площади.

Таблица 2.8. Обозначения параметров, характеризующих физиологические характеристики растений и опушения листа у пшеницы.

Признак	Обозначение	Ед. Изм.
Скорость транспирации	<i>TR</i>	ммоль/м ² с
Устьичная проводимость	<i>SC</i>	ммоль/м ² с
Скорость нетто-фотосинтеза	<i>PhR</i>	ммоль/м ² с
Эффективность использования воды	<i>WUE</i>	безразм.
Биомасса побега	<i>BM</i>	г
Интенсивность флуоресценции хлорофилла	<i>F₀</i>	25000 ммоль (фотон)/м ² с
Потенциальная эффективность фотосинтеза	<i>F_v/F_m</i>	безразм.
Реальная эффективность фотосистемы	<i>Yld</i>	безразм.
Скорость транспорта электронов фотосистемы II	<i>ETR</i>	ммоль /м ² с
Коэффициент нефотохимического тушения флуоресценции	<i>NPQ</i>	безразм.
Число трихомов на изобраении сгиба верхней стороны листа	<i>TN_{up}</i>	шт.
Число трихомов на изобраении сгиба нижней стороны листа	<i>TN_{bot}</i>	шт.
Средняя длина трихомов на верхней стороне листа	<i>TL_{up}</i>	мкм
Средняя длина трихомов на нижней стороне листа	<i>TL_{bot}</i>	мкм
Число устьиц на верхней стороне листа	<i>SN_{up}</i>	шт.
Число устьиц на нижней стороне листа	<i>SN_{bot}</i>	шт.
Длина листа	<i>L</i>	мм
Ширина листа	<i>W</i>	мм

Также были приготовлены препараты сгиба листа в трёх местах (у основания, посередине и на кончике) для определения количественных

характеристик опущения программой LHDetect2 согласно протоколу (см. Раздел 2.2.2). На основе результатов работы программы LHDetect2 были рассчитаны средняя длина (в мкм) трихомов на нижней и верхней стороне листа и их число на изображениях сгиба, полученных с обеих сторон. Указанные параметры опущения измерялись для листьев шести растений каждого генотипа в каждой из повторностей, их средние значения характеризовали опущение генотипов. Все принятые сокращения также представлены в табл. 2.8.

Для оценки влияния генотипа на физиологические и фенотипические характеристики растений в ответ на контрастные условия орошения был проведен двухфакторный дисперсионный анализ. Линии (генотипы) и водные режимы рассматривались как независимые факторы, а повторности - как случайный фактор. *F*-критерий использовался для оценки достоверных различий между средними значениями физиологических, фотосинтетических характеристик и характеристик листовой поверхности сортов и линий-доноров и реципиентов. Пул данных был обработан с использованием метода неметрического многомерного масштабирования (NMDS) (Oksanen, 2011). NMDS и анализ главных компонент были проведены с использованием пакета PAST (Hammer et al., 2001). Для изучения специфического влияния засухи на изменение рассматриваемых характеристик растений мы также рассчитали нормированную разницу их значений в условиях полива и засухи по формуле $X_i = (X_{in} - X_{id}) / (X_{id})$, где X_{in} – значение *i*-го признака в контрольных условиях X_{id} – значение признака в условиях засухи. Мы использовали полученные нормированные значения для проведения многомерного статистического анализа (анализ главных компонент и кластерный анализ генотипов, основанный на расчете коэффициентов корреляции Пирсона).

2.6.2. Связь количественных характеристик опущения листа у изогенных и замещенных линий пшеницы с физиологическими показателями растений в условиях нормального и ограниченного водообеспечения

Средние значения изученных физиологических параметров, а также уровни значимости отличий между сортами-реципиентами и производными линиями представлены в табл. 2.9, 2.10, 2.11. Эти данные были использованы в многомерном статистическом анализе.

Пул данных по 18 физиологическим и морфологическим признакам, измеренным в двух водных режимах, был проанализирован с помощью NMDS с целью изучения изменчивости характеристик растений генотипов друг относительно друга на плоскости в зависимости от распределения векторов признаков в многомерном пространстве (Holland, 2008) (рис. 2.25). Из рисунка видно, что сорта и линии образуют две отдельные области, характеризующие проявление признаков в двух условиях полива. При водном дефиците площадь, ограниченная многоугольником с вершинами, соответствующими генотипам в пространстве физиологических и морфологических параметров, уменьшается. При этом относительное расположение генотипов в этом пространстве остается неизменным. Это говорит о сходной у большинства генотипов реакции на засуху. В пределах каждой области генотипы расположились относительно друг друга в зависимости от разнообразия проявления отдельных признаков.

Таблица 2.9. Сравнительные значения параметров газообмена сортов и их производных форм - изогенных и замещающих линий при контрастных водных режимах. Сравнения сделаны с получателем, 'Янетцкис Пробат' сравнивался с 'Саратовской 29'. Условия: К-контроль, З – засуха.

Генотипы	TR, ммоль м ⁻² с ⁻¹		SC, ммоль м ⁻² с ⁻¹		PhR, мкмоль м ⁻² с ⁻¹		WUE		ВМ, г		IT [#]
	К	З	К	З	К	З	К	З	К	З	
C29	0.38	0.40	26.9	28.5	2.36	2.40	7.1	8.3	3.9	1.3	33.4
<i>i:C29 glabrous</i>	0.92 ^{***}	0.50 [*]	67.4 ^{***}	35.2 [*]	3.25 ^{***}	2.36	3.8 ^{***}	5.4 ^{***}	3.4 [*]	0.98 ^{**}	30.4
<i>i:C29 H12^{aesp}</i>	0.27 [*]	0.29 [*]	19.2 [*]	20.8 [*]	1.54 ^{***}	1.93 ^{***}	6.1	8.6	4.1	1.1	27.1
C29/ЯП4В	0.63 ^{***}	0.53 ^{**}	45.4 ^{***}	38.2 ^{**}	2.4	2.08 [*]	5.5 ^{**}	4.3 ^{***}	3.4 [*]	1.2	36.5
'Родина'	0.75	0.47	55.8	34.1	2.9	2.82	5.5	7.1	4.5	1.1	24.8 ^{**}
102/00 ⁱ	0.70	0.38 [*]	52.0	26.3 [*]	2.8	1.85 ^{***}	5.0	6.1	4.9	1.1	23.5 ^{**}
ЯП	0.92 ^{***}	0.24 ^{***}	68.3 ^{***}	16.7 ^{***}	3.3 ^{***}	2.02 ^{***}	4.1 ^{***}	9.4	5.4 ^{***}	1.2	23.1 ^{**}

*** - P<0,001; ** - P<0,01; * - P<0,05; Обозначения характеристик приведены в табл. 2.8. # - индекс устойчивости к засухе; *курсив* – значимые различия генотипа C29 и других связанных с ним генотипов

Таблица 2.10. Сравнительные значения морфологических характеристик сортов и их производных форм - изогенных и замещающих линий при контрастных водных режимах. Сравнения сделаны с соответствующим генотипом-акцептором, 'Янетцкис Пробат' с 'Саратовской 29'. Условия: К-контроль, З – засуха.

Генотипы	Трихомы								Устьица				Размер листа			
	TNup		TNbot		TLup		TLbot		SNup		SNbot		L		W	
	К	З	К	З	К	З	К	З	К	З	К	З	К	З	К	З
C29	39.7	55.4	48.2	66.4	140.9	138.8	133.8	115.8	40.3	56.1	30.1	30.1	278.0	109.4	9.3	5.0
<i>i:C29 glabrous</i>	5.3 ***	9.4 ***	3.3 ***	18.1 ***	45.0 ***	37.5 ***	67.8 ***	42.7 ***	42.6	72.2 ***	31.9	30.0	293.4	93.8	9.4	4.3
<i>i:C29 HI2^{aesp}</i>	41.9	61.4 **	56.4***	72.2***	213.9 ***	191.9 ***	188.2 ***	165.3***	44.4	66.1 *	32.4	32.8	267.9	93.8	9.3	4.3
C29/ЯП4В	19.5 ***	33.4 ***	21.9***	35.4***	70.7 ***	62.8 ***	64.7 ***	61.7 ***	44.0	75.2 ***	29.7	33.0	231.1 ***	105.9	8.4	4.7
'Родина'	0.56	0.4	0.2	0.2	18.9	11.1	11.2	17.3	47.2	67.0	27.4	27.6	176.1	137.2	9.9	5.7
102/00 ⁱ	14.6 ***	23.8 ***	10.5***	14.4***	281.4 ***	141.4 ***	223.1***	125.0 ***	40.7 **	57.0 **	27.0	36.4 *	284.5 **	117.1	11.9	5.8
ЯП	4.0 ***	15.5 ***	2.7 ***	10.3 ***	31.4 **	38.8 ***	42.9 ***	41.8 ***	33.2 ***	52.0	24.2	32.3 *	381.0 ***	132.3	10.5	4.3

*** - $P < 0,001$; ** - $P < 0,01$; * - $P < 0,05$; Обозначения как указаны в табл. 2.8

Таблица 2.11. Сравнительные значения параметров флуоресценции хлорофилла у сортов и их производных форм - изогенных и замещающих линий при контрастных водных режимах. Сравнения сделаны с получателем, 'Янетцкис Пробат' с 'Саратовской 29'. Условия: К-контроль, З – засуха.

Сорта и линии	F ₀		Fv/Fm		Yield		ETR		NPQ	
	К	З	К	З	К	З	К	З	К	З
С29	117.3	70.8	0.785	0.760	0.452	0.525	30.3	35.1	0.38	0.41
i:C29 <i>glabrous</i>	107.4	61.4	0.752***	0.750	0.484*	0.513	32.5*	34.4	0.63***	0.45
i:C29 <i>HI2aesp</i>	110.9	58.7*	0.696***	0.749	0.477	0.538	31.9	36.0	0.26***	0.35**
С29/ЯП4В	112.3	77.3	0.722***	0.674***	0.478	0.476**	32.0	31.9**	0.39	0.32*
'Родина'	102.3	79.6	0.732	0.742	0.502	0.500	33.7	33.5	0.44	0.43
102/00 ⁱ	95.7	61.8***	0.776***	0.799***	0.464*	0.480*	29.4**	30.5***	0.63**	0.66***
ЯП	92.6***	65.7	0.779	0.751	0.537***	0.523	36.0***	35.1	0.36	0.32**

*** - P<0,001; ** - P<0,01; * - P<0,05; Обозначения как указаны в табл. 2.8

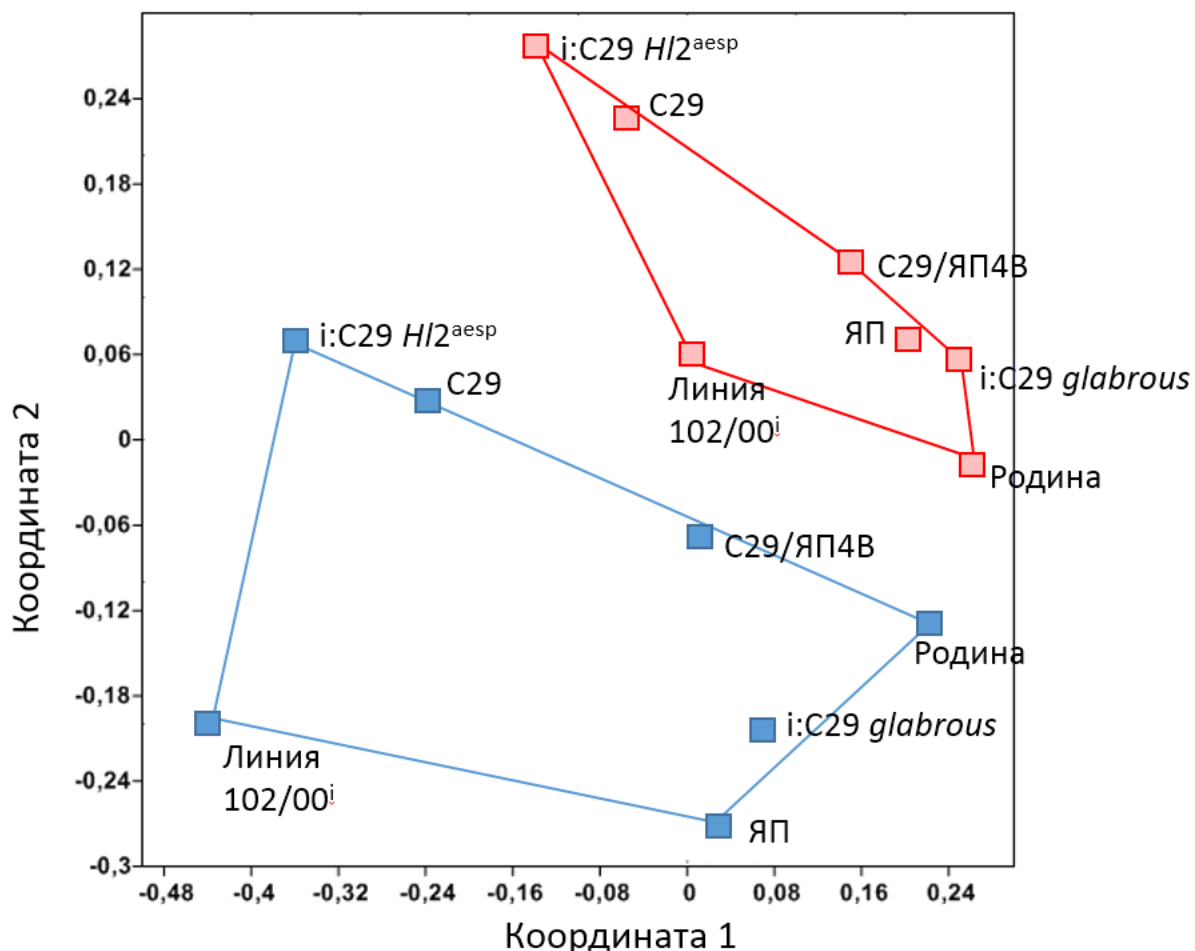


Рисунок 2.27. Неметрическое многомерное шкалирование сортов и линий пшеницы по 18 фотосинтетическим параметрам и морфологическим характеристикам листьев при контрастных водных режимах: красный многоугольник соответствует значениям параметров при водном дефиците, синий – при поливе.

Следует отметить, что густо опушённый сорт-реципиент C29 в обоих случаях на диаграмме располагается близко с линией *i:C29 Hl2*^{aesp}, несущей дополнительный ген опушения листа. Линии этого же реципиента, лишённые одного или двух генов опушения, а также неопушённые сорта ЯП и 'Родина' группировались вместе на противоположной стороне многоугольника. Такое расположение характерно для двух условий полива. Линия 102/00ⁱ, несущая интрогрессированное от *Ae. speltoids* опушение, располагалась особняком от

остальных генотипов. Это говорит о контрастном проявлении отдельных признаков у трёх групп генотипов. Действительно, согласно измеренным показателям газообмена (табл. 2.9), С29 и изогенная линия имели самые низкие значения параметров TR и SC на поливе и не снижали эти показатели на засухе. Также они сохраняли высокую эффективность использования воды, WUE, в обоих условиях. Одновременно оба генотипа характеризуются наиболее высокой удельной плотностью трихомов и их длиной на обеих сторон листа (табл. 2.10). Отметим, что экспрессия признаков опущения ещё более усиливается при водном дефиците. Обособленная линия 102/00ⁱ обладала самой большой изменчивостью по параметрам флюоресценции хлорофилла среди изученных генотипов (табл. 2.11). Линия достоверно увеличивала по сравнению с исходным сортом 'Родина' потенциальную эффективность фотосинтеза, уменьшала реальную эффективность фотосинтеза и уровень электронного транспорта. У неё был самый высокий коэффициент NPQ в обоих условиях полива.

Фенотипическое разнообразие физиологических реакций на разные водные режимы обеспечивали 2 главные компоненты (табл. 2.12). В обоих вариантах первая компонента (PC1) имела значительное преимущество. Вклад в PC1 отдельных признаков различался в двух вариантах опыта. На поливе SC и ETR вносили наибольший вклад, далее следовали длина трихомов на обеих сторонах листа и число устьиц на нижней стороне листа. На засухе на первое место выходят плотность трихомов на верхней стороне листа, затем следуют другие параметры опущения, ETR и SC. Во вторую компоненту (PC2) на поливе важный вклад вносила плотность трихомов на нижней стороне листа, менее значимы были SC и число трихомов на верхней стороне листа. В условиях стресса наибольший вклад в PC2 вновь вносила плотность опущения.

Таблица 2.12. Анализ основных компонентов различий в характеристиках признаков при двух водных режимах. Обозначения признаков указаны в табл. 2.8.

Главная компонента	Условия полива		Условия водного дефицита	
	Вклад в общую дисперсию, %	Вклад основных признаков в величину компоненты	Вклад в общую дисперсию, %	Вклад основных признаков в величину компоненты
PC1	76.7	SC (0.53); ETR (0.40); TLup (0.39); TLbot (0.37); SNbot (0.30)	81.3	TNbot (-0.40); TNup (-0.34); TLup (-0.34); ETR (0.33); SC (0.33); TLbot (-0.33); SNbot (0.29); SNup (0,23)
PC2	17.1	TNbot (0.60); SC (-0.48); TNup (0.46); SNbot (0.23)	11.8	TNbot (0.69); TNup (0.35); SC (0.34); SNup (0.33); TLbot (-0.25); TLup (-0.24)

Для изучения взаимодействия признаков и их изменчивости при изменении условий выращивания был проведён кластерный анализ, в котором в качестве данных была взята нормированная разница значений каждого признака в условиях контроля и засухи. Кластеризация производилась на основании изменения каждого признака в ответ на стресс. В результате анализа образовалось четыре кластера (рис. 2.28). Отдельный кластер образовали размеры листа (L , W) и биомасса побега (BM). Вместе кластеризовались признаки газообмена - устьичная проводимость (SC), скорость транспирации (TR) и скорость нетто фотосинтеза (PhR). Длина трихомов кластеризовалась с показателями флуоресценции хлорофилла (NPQ , F_v/F_m , Yld , ETR) и числом устьиц на нижней стороне листа. Число трихомов на нижней и верхней стороне ($TNup$) и ($TNbot$) и число устьиц на верхней стороне листа кластеризовались с эффективностью использования воды (WUE) (рис. 2.28).

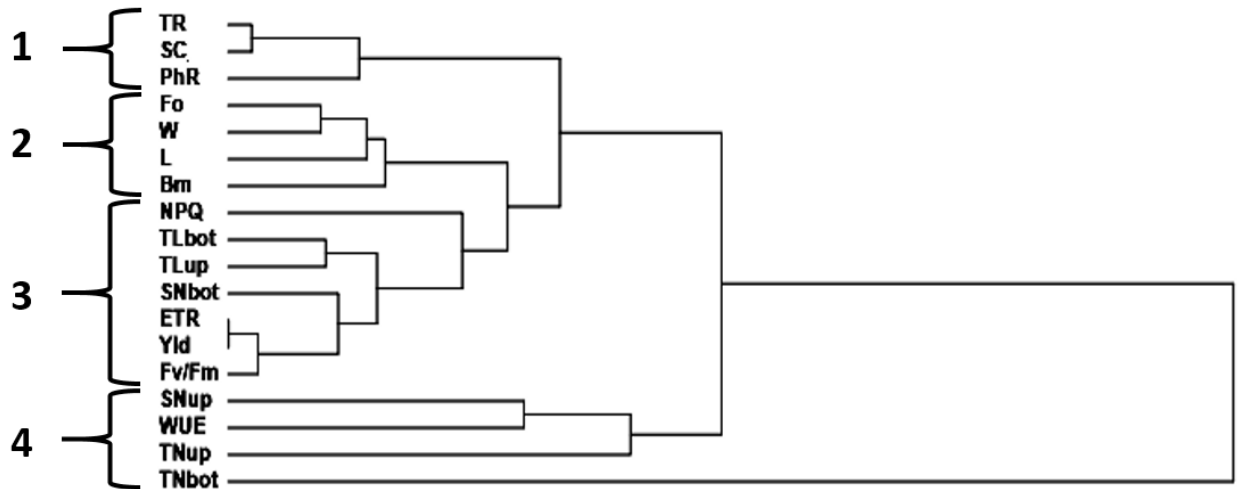


Рисунок 2.28. Кластерный анализ параметров газообмена, флуоресценции хлорофилла и морфологических признаков на основе нормализованных различий значений между нормой и засухой. Фигурные скобки обозначают кластеры, которые группируют сходную изменчивость признаков между генотипами.

Таким образом, длина и число трихомов листа пшеницы изменяются в ответ на водный стресс согласованно с фотосинтетическими показателями.

Мы провели кластеризацию изученных генотипов и анализ главных компонент на основании нормированной разницы значений каждого признака в условиях контроля и засухи (рис. 2.29). Линии и сорта на дендрограмме образовали два кластера. В первый кластер (рис. 2.29А) вошли слабо и очень слабо опушённые образцы: сорта ‘Родина’ и ЯП, изогенная линия *i:C29 glabrous (hl1 hl3)*, утратившая оба гена опушения листа исходного сорта, а также замещенная линия С29/ЯП4В, лишённая одного гена опушения. Второй кластер (рис. 2.29) был представлен хорошо опушённым сортом ‘Саратовская 29’, его дочерней линией *i:C29 Hl2^{aesp}*, несущей дополнительный ген опушения, и производная от неопушённого сорта ‘Родина’ линия 102/00ⁱ, несущая длинное

редкое опушение, интрогрессированное от *Ae. speltoides*. Данная кластеризация, основанная на нормированной разнице между средними значениями признаков в разных условиях полива, согласуется с результатами NMDS, основанными на абсолютных средних значениях.

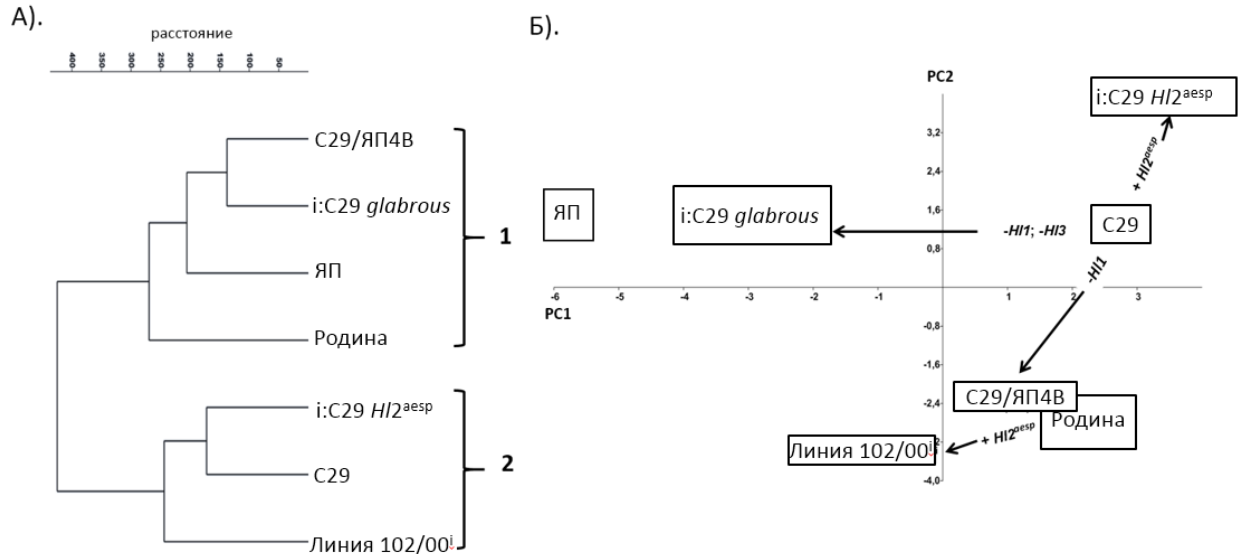


Рисунок 2.29. Кластеризация генотипов (А) и схема анализа основных компонентов (Б) на основе нормализованных различий между значениями в условиях контроля и засухи. Фигурные скобки обозначают кластеры генотипов со сходной реакцией на засуху.

Анализ главных компонент нормированных различий по изученным признакам показывает характер изменения ответа растения на засуху у различных генотипов (рис. 2.29 Б, показаны стрелками). Рассчитанные компоненты описывали 33% и 24% вариации, соответственно. Сильно опушенный сорт С29 и практически неопушенный сорт ЯП занимали противоположные положение в пространстве первой компоненты. Замещение доминантных аллелей генов *HI1* и *HI3* опушения листа на рецессивные в

генотипе сорта С29 вызывало смещение изучаемых характеристик в сторону неопушенного генотипа. Индексы устойчивости биомассы побега уменьшались аналогичным образом (табл. 2.9). Это говорит о значительной роли опушения в проявлении комплекса физиологических признаков в контрастных условиях полива. Изменение гена *HL2^{aesp}* на доминантный на фоне генотипа С29 изменило характеристики в направлении компоненты РС2. У линии i:С29 *HL2^{aesp}* было достоверно снижено фотохимическое тушение (*NPQ*) по сравнению с родительским сортом (табл. 2.11). Замещение хромосомы 4В с доминантным геном *Hl1* у сорта С29 на гомологичную от ЯП с рецессивным аллелем, также, как и замещение хромосомы 7В сорта 'Родина' на гомеологичную от *Ae. speltoides*, приводят к сдвигу генотипов на диаграмме в сторону одновременного уменьшения значений обоих главных компонент. Смещение линии С29/ЯП4В в сторону слабоопушенных генотипов не сопровождалось снижением биомассы побега (табл. 2.9). Однако у линии по сравнению с С29 снижалось большинство показателей первичной фотохимии особенно в условиях засухи, а также параметр *WUE* (табл. 2.10). Эту же область пространства главных компонент занимал неопушенный сорт 'Родина' и его производная интрогрессированная линия с геном длинного опушения *HL2^{aesp}*. Сдвиг линии в отрицательную сторону РС2 связан, вероятно, с увеличением нефотохимического тушения (*NPQ*) и снижением на засухе всех показателей первичной фотохимии.

Сорт-реципиент С29, на генетическом фоне которого были созданы линии с различным типом опушения, является одним из наиболее засухоустойчивых среди всех яровых сортов бывшего СССР и России (Кумаков, 1985; Ильина, 1989). Сорт характеризуется высокой экологической пластичностью и в прошлом занимал более 20 млн га посевов в различных экологических зонах (Мамонтова, 1980). Он обладает плотным опушением,

которое контролируется двумя доминантными генами, *H11* и *H13* (Дорошков и соавт., 2011). Параметры газообмена у C29 и производных от нее линий с разными вариантами генов опушения были обратно пропорциональны плотности и длине трихомов. Линия i:C29 *glabrous*, в генотипе которой остались лишь рецессивные гены опушения, а количество трихомов на верхней стороне листа сократилось в 9 раз, имела повышенную транспирацию, нетто-фотосинтез и сниженную *WUE* (табл. 2.9). Напротив, линия i:C29 *H12^{aesp}* с повышенной плотностью, и, важно отметить, длиной трихомов, имела сниженные показатели газообмена. Линия 102/00ⁱ с дополнительным геном опушения продемонстрировала менее выраженный, но тот же эффект. По сравнению с исходным сортом 'Родина' у нее увеличилось количество и длина трихомов и снизились все параметры газообмена в условиях дефицита воды. При кластеризации нормированных значений фотосинтеза число трихомов и *WUE* группировались вместе. Эти результаты согласуются с данными, полученными для других видов растений (Ehleringer, 1976; Liakopoulos et al., 2006; Picotte et al., 2007). Так, у изолиний сои, в 4 раза различающихся по степени опушения, у более опушенных линий транспирация была уменьшена при сохранении более благоприятной *WUE* (Baldochi et al., 1983).

Среди комплекса параметров флуоресценции хлорофилла существенную связь с опушением обнаружила только скорость электронного транспорта. Этот признак вносил существенный вклад в PC1 наравне с длиной трихомов и SC, а на засухе и с плотностью трихомов на поверхности. При анализе нормированной разницы значений признаков параметры флуоресценции хлорофилла и длины трихомов на обеих сторонах листа были объединены в один большой кластер (рис. 2.28), что отражает сходные тенденции изменения этих признаков в ответ на стресс. По-видимому, от длины трихомов зависит поглощение квантов света листовой поверхностью. Это отразилось в

существенной разнице по NPQ между изогенными линиями, созданными на основе С29, и родительским генотипом в условиях оптимального полива и одинаковой освещенности. Значения NPQ были обратно пропорциональны плотности опушения и длине трихомов (табл. 2.11.). В вододефицитных условиях NPQ также снижалось у линии i:С29 $HI2^{aesp}$ с дополнительным геном опушения. Эти факты согласуются с представлением о защитной функции опушения при избыточной освещенности листьев. В работе Галмес и соавт. (Galmés et al., 2007) было обнаружено, что наличие трихомов у опушенного и неопушенного подвидов вида *Digitalis minor* является важным механизмом уменьшения чувствительности к фотоповреждению. Наблюдавшиеся авторами различия по опушению были связаны с географическим распространением подвидов при большей экологической распространенности опушенного вида, включавшей полуаридные области.

Вместе с тем, линия 102/00ⁱ, несущая редкое, но длинное опушение листа, по-видимому, не получила преимуществ от введения дополнительного гена опушения. По сравнению с исходным сортом 'Родина', у линии 102/00ⁱ была существенно снижена скорость ассимиляции CO_2 в условиях водного стресса (табл. 2.9), реальная эффективность фотосинтеза, скорость транспорта электронов (табл. 2.11). У неё существенно, в 1.4 раза, было повышено NPQ по сравнению с исходным сортом, то есть энергия квантов света использовалась менее эффективно. Объяснить это можно тем, что по молекулярным данным (Dobrovolskaya et al., 2007), она несет целую хромосому 7S от дикого злака *Ae. speltoides*, в которой локализован ген $HI2^{aesp}$. Возможно, что данная хромосома несёт другие генетические факторы, контролирующие общую реакцию линии на засуху. В частности, у линии была снижена плотность устьиц (табл. 2.9), что на засухе могло оказаться критическим для данного генотипа, и привело к снижению транспирационного тока воды, скорости ассимиляции CO_2 и, как

следствие, - к снижению засухоустойчивости.

Большинство генотипов в условиях полива имело сопоставимое число устьиц на верхней стороне листа. Засуха вызывала увеличение их плотности (табл. 2.10), составлявшее 40-70%. Увеличение плотности устьиц у пшеницы как реакция на водный дефицит отмечалось ранее (Давыдов, 2007; Ху, Zhou, 2008). Увеличение плотности трихомов как реакции на засуху отмечали у опушенного сорта вида *Olea europaea* L. (Boughalleb, Hajlaoui, 2010). Однако, именно две линии сорта С29 с редуцированным по сравнению с исходным сортом числом трихомов, показали наиболее значительное, до 70%, увеличение числа устьиц. Возможно, что у засухоустойчивого сорта морфогенез трихомов и устьиц взаимосвязаны.

Биомасса побега является интегральным показателем адаптивности растения к различным условиям. Слабоопушенные сорта ЯП и 'Родина' показали существенные преимущества перед С29 по уровню транспирации, устьичной проводимости и уровню нетто-фотосинтеза. Даже при низких значениях параметра *WUE* в нелимитирующих условиях они образовывали большую, чем у С29 биомассу. Но в условиях засухи оба сорта значительно её снижали и показали низкое значение индекса устойчивости биомассы. При этом сорт ЯП, имевший самые высокие показатели фотосинтеза и самый высокий прирост биомассы на поливе, на засухе снижал нетто-фотосинтез в 1,5 раза и показал самую низкую устойчивость к засухе. В группе линий, родственных С29, биомасса побега в оптимальных условиях была прямо пропорциональна плотности опушения и длине трихомов (табл. 2.9.). В условиях стресса линия i:С29 *glabrous*, лишённая опушения, сформировала наименьшую массу побега. Несмотря на высокие фотосинтетические показатели, линии с редуцированным опушением имели сниженную или

сравнимую с родителем биомассу в благоприятных условиях. Видимо, сорт С29 не имеет генетически обусловленного физиологического потенциала для реализации возможностей интенсивного фотосинтеза. Возможно, также это связано с уменьшением WUE у линий, лишенных опушения листа. Замена доминантных аллелей генов $H11$ и $H13$ или только гена $H11$ в генотипе С29 сдвигает изогенную и замещенную линии в пространстве физиологических и фенотипических параметров в сторону незасухоустойчивого сорта ЯП (рис. 2.29). Аналогичный результат был обнаружен при изучении взаимоотношений между различными фенотипическими признаками и биомассой как отображением приспособленности у вида *Convolvulus chilensis* в условиях недостаточного водоснабжения (Gianoli et al. 2005). Оказалось, что листовые трихомы были единственным признаком, обнаружившим адаптивную пластичность на засухе, а их плотность положительно коррелировала с биомассой.

Добавление интрогрессированного от *Ae. speltoids* гена $H12^{aesp}$ в разные генотипы вызвало неоднозначный эффект. Линия 102/00ⁱ, несущая данный ген на генетическом фоне сорта 'Родина', снизила устойчивость к водodefициту. Линия i:С29 $H12^{aesp}$ с самым густым опушением, продемонстрировала снижение физиологических реакций в контрастных условиях полива. По сравнению с С29, она сохранила уровень WUE , сравнимую с засухоустойчивым сортом. Можно предполагать, что дополнительное опушение, образуемое геном $H12$, создает еще более плотный по сравнению с С29 пограничный слой, в котором замедлен газообмен. Это может вызывать снижение скорости нетто-фотосинтеза. Более точно эффект длинного опушения можно установить после создания изогенной линии на генетическом фоне сорта 'Родина'.

Следует отметить, что все изученные сорта (за исключением практически

неопушенного сорта 'Родина') на засухе увеличивали степень опушения (табл. 2.10). Это реакция направлена на увеличение пограничного слоя, замедление транспирации и, в конечном счёте, - на сохранение воды в тканях листа и поддержании постоянной температуры его поверхности. Это явление можно считать универсальным механизмом, поскольку его наблюдали и у других видов растений (Ehleringer, 1982; Morales et al., 2002; Picotte et al., 2007; Numaoka et al., 2017). Генетически обусловленное присутствие густого опушения делает растения пшеницы пре-адаптированными к засухе, так как наличие пограничного слоя, обеспечиваемого наличием трихомов, уменьшает транспирацию, устьичную проводимость и нетто-фотосинтез.

Опушение листа чрезвычайно распространено среди яровых сортов Сибири и Казахстана по сравнению, например, с сортами Германии. Это наблюдали как среди современных, так и среди стародавних сортов двух стран, хотя специального отбора на этот признак селекционерами не проводится (Генаев и др., 2012). Недавно было высказано предположение, что признак может быть полезен для повышения засухоустойчивости риса и снижении водозатрат при его выращивании (Numaoka et al., 2017). Опушённые листья нового генотипа риса с интрогрессией соответствующего гена имели низкий уровень транспирации и высокий уровень WUE при различных уровнях интенсивности света.

2.7. Высокопроизводительное фенотипирование опушения листа картофеля

2.7.1. Условия, материалы и методы для анализа опушения листа у картофеля

В работе было использовано 35 сортов картофеля Российского происхождения: ‘Брянский деликатес’, ‘Великан’, ‘Вымпел’, ‘Гала’, ‘Голубизна’, ‘Гусар’, ‘Евразия’, ‘Жуковский ранний’, ‘Ильинский’, ‘Импала’, ‘Колобок’, ‘Кортни’, ‘Красавчик’, ‘Крепыш’, ‘Ломоносовский’, ‘Метеор’, ‘Накра’, ‘Никулинский’, ‘Регги’, ‘Ред Скарлетт’, ‘Сударыня’, ‘Танго’, ‘Удача’, ‘Фаворит’, ‘Фрителла’, ‘Чароит’, ‘Любава’, ‘Кемеровчанин’, ‘Кузнечанка’, ‘Танай’, ‘Тулеевский’, ‘Златка’, ‘Сафо’, ‘Юна’ и ‘Невский’ (стандарт). Сорта были выращены в поле СИБ НИИРС (Новосибирская Обл.) в 2016 году. Сроки посадки – середина-конец мая. Образцы листьев для анализа собирались 3-5 августа 2016 г.

Для определения количественных характеристик опушения были использованы листья пятого порядка, начиная с верхушки. Листья были собраны в поле и помещены черешками в воду для сохранения тургора. В лабораторных условиях из них были вырезаны фрагменты, заключённые между жилок, по три для каждого листа, как показано на рис. 2.30.

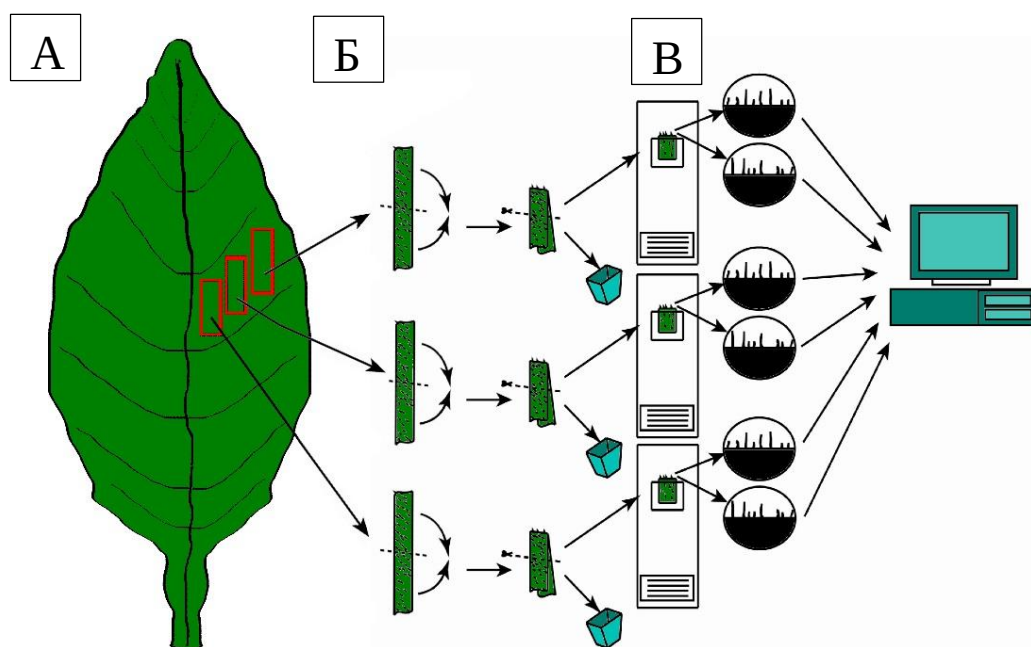


Рисунок 2.30. Основные этапы получения изображений трихомов картофеля для фенотипирования опушения листа. А) вырезание фрагментов; Б – приготовление препаратов; В – получение изображений и их компьютерная обработка.

Фрагмент сгибался пополам таким образом, чтобы верхняя поверхность листа оказалась снаружи, и помещался на предметное стекло. Фиксация фрагмента листа на предметном стекле производилась липкой лентой с отступом от сгиба на 3 мм. Такая методика позволяет получить в проходящем свете контрастное изображение трихомов на светлом фоне. Образцы согнутых фрагментов листа помещались под микроскоп ZEISS Axioscop 2 plus с камерой Axioscam HRC и адаптером с увеличением 0.63, при помощи которого были получены их цифровые микрофотографии. Полученные изображения имели размеры в пикселях 1300 x 1030, размеры одного пикселя составили 2.1 x 2.1 мкм. Для идентификации трихомов и оценки количественных характеристик опушения использовали программу LHDetect2 (Genaev et al., 2012).

2.7.2. Оценка количественных характеристик опушения листьев картофеля с использованием анализа цифровых микроизображений

Анализ изображений показывает наличие на сгибе листа картофеля трихомов двух типов: едва заметные короткие типа А и длинные типа В (рис. 2.31).

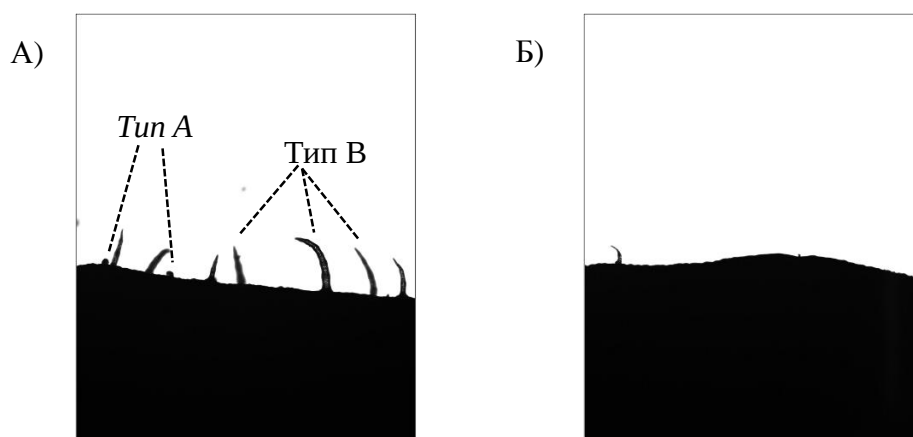


Рисунок 2.31. Примеры изображений сгиба листа картофеля. А – сорт ‘Колобок’, для которого указаны типы трихомов на изображении; Б – сорт ‘Красавчик’.

Отметим, что использованная версия программы LHDetect2 не позволяет точно оценить длину трихомов типа А (см. раздел 2.1.3 «Морфология и генетический контроль опушения листа у картофеля»), поскольку подавляющее большинство из них слишком короткие и недостаточно выступают за пределы поверхности листа на изображении. Трихомы типа В хорошо заметны на контрастном фоне и значительно выступают над поверхностью листа, что позволяет произвести их автоматический подсчет с помощью программы LHDetect2. Были рассчитаны распределения трихомов по длинам с шагом 42 мкм, была оценена их средняя длина в микрометрах (l_m) и число трихомов на изображении (n_f).

Частоты трихомов в бинах распределения по длинам были усреднены для

каждого генотипа и использованы для кластерного анализа и анализа главных компонент. Для статистической обработки данных была использована программа PAST (Hammer et al., 2001).

Мы оценили среднюю длину и среднее значение числа трихомов типа В на изображении для каждого из исследованных сортов картофеля. Гистограммы распределений значений этих параметров приведены на рис. 2.32 (панели А и Б, соответственно).

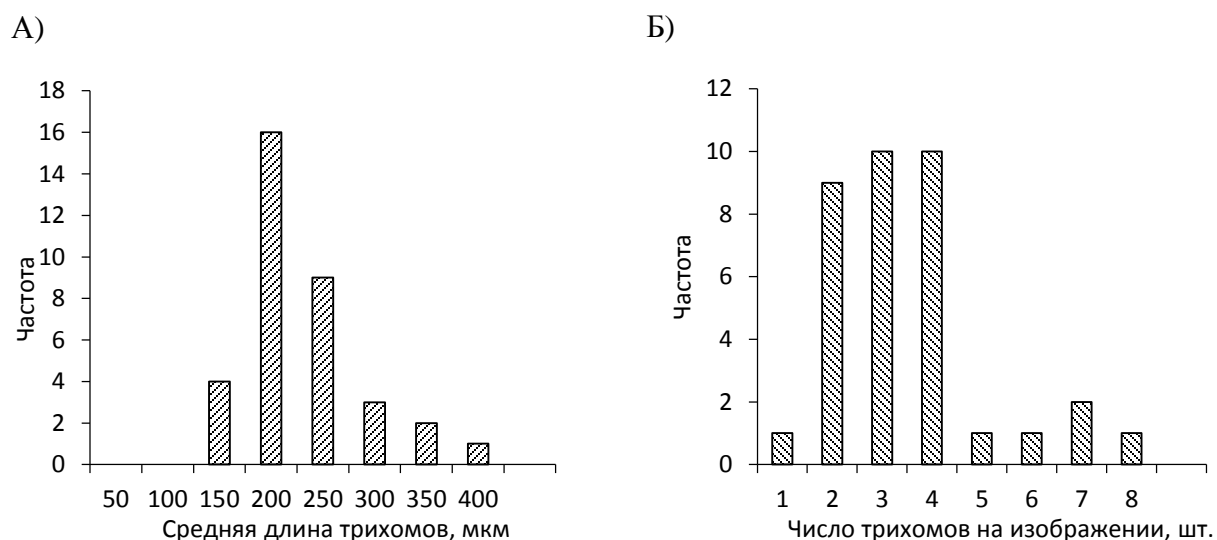


Рисунок 2.32. Гистограммы распределений средней длины и числа трихомов на изображении для исследованных сортов картофеля. (А) Распределение средней длины трихомов (ось X): минимум 124.17 мкм, максимум 382.67 мкм, медиана 191.61 мкм, среднее значение 207.39 мкм, дисперсия 3398.91 мкм², стандартное отклонение 58.30 мкм. (Б) Распределение среднего числа трихомов (ось X): минимум 1 шт., максимум 7.33 шт, среднее значение 2.94 шт., медиана 2.41 шт., дисперсия 2.40 шт², стандартное отклонение 1.55 шт.

Из гистограмм видно, что основное количество растений имеет размеры трихомов типа В в пределах 150-250 мкм, число трихомов на изображении – 2-4. В то же время можно выделить сорта с наиболее длинными трихомами (более 350 мкм, сорт ‘Кортни’) и наиболее короткими трихомами (менее 175 мкм,

сорта ‘Импала’, ‘Красавчик’, ‘Регги’, ‘Тулеевский’), а также сорта с интенсивным опушением (более 5,5 трихомов на изображении, сорта ‘Великан’, ‘Кемеровчанин’, ‘Танай’) и слабым опушением (менее 1,5 трихомов на изображении, сорт ‘Красавчик’).

На рис. 2.33 приведена диаграмма рассеяния средней длины и количества трихомов типа В. В основном исследованные сорта располагаются в области графика, соответствующем числу трихомов от 1 до 4-х и длине от 100 до 250 мкм. Отметим, что для сортов с интенсивным опушением средняя длина трихомов оказывается не более 200 мкм.

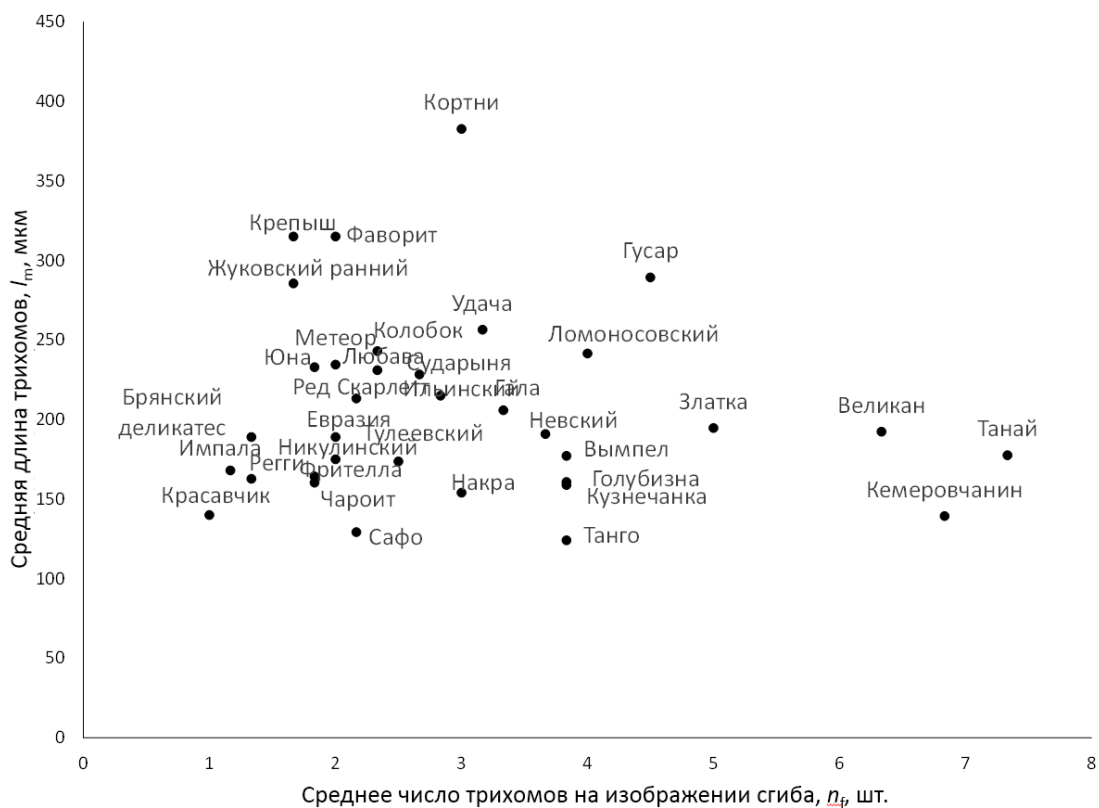


Рисунок 2.33. Диаграмма рассеяния средней длины и числа трихомов на изображении сгиба листа у 35 сортов картофеля.

Значимой линейной зависимости между числом трихомов и их средней

длиной обнаружено не было: коэффициент корреляции между этими величинами составил -0.147 ($p < 0.39$).

Мы провели анализ главных компонент для количественных характеристик опушения исследованных нами сортов в многомерном пространстве количества трихомов в различных интервалах длин. Результаты приведены на рис. 2.34. На две главные компоненты пришлось 58% дисперсии (43% на первую, PC1, и 15% на вторую, PC2).

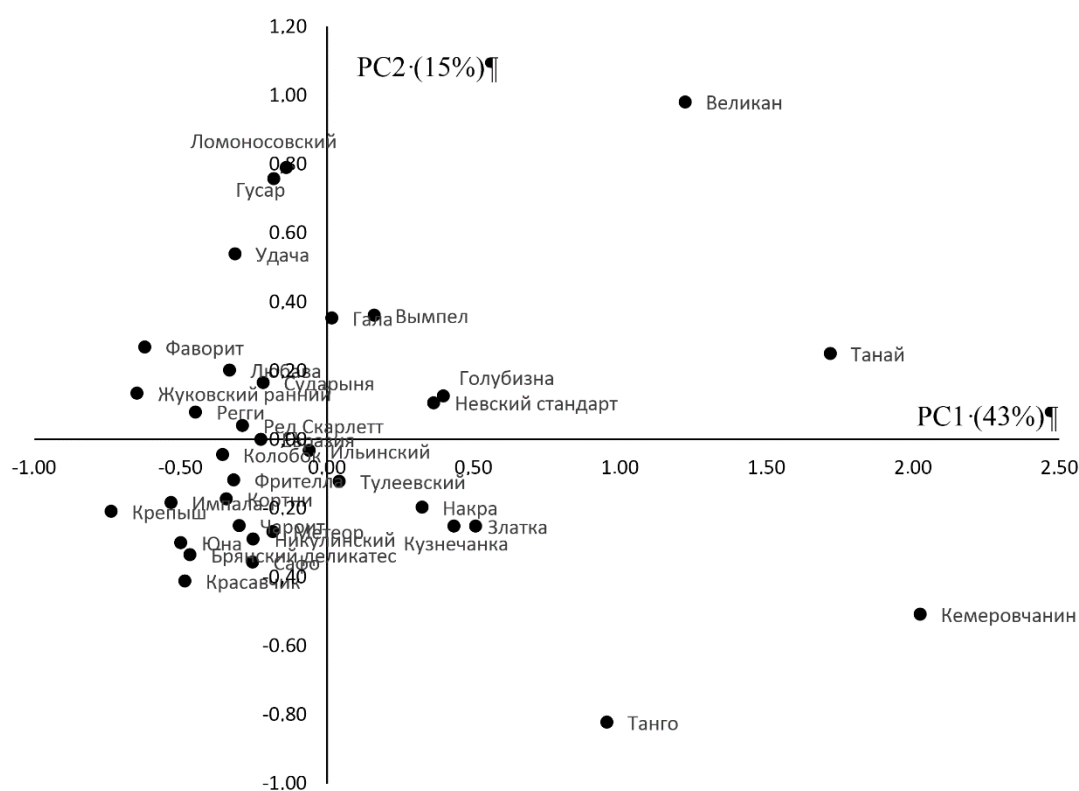


Рисунок 2.34. Диаграмма рассеяния 35 сортов картофеля в пространстве двух первых компонент, полученных в результате анализа главных компонент. В скобках рядом с наименованиями компонент указана доля общей дисперсии, которая приходится на компоненту.

Оказалось, что коэффициент корреляции Пирсона между весами первой компоненты и числом трихомов составил 0.9 , а длиной трихомов -0.43 . Таким образом, увеличение первой компоненты характеризует увеличение числа

трихомов при уменьшении их длины. Из рисунка видно, что наибольшее значение первой компоненты наблюдается для наиболее опушенных сортов ('Великан', 'Танай', 'Кемеровчанин'), которые, преимущественно, имеют умеренные и малые длины трихомов. Что касается второй компоненты, то оказалось, что ее веса отрицательны для числа трихомов в интервалах длин 42-200 мкм, и положительны для числа трихомов в интервалах длин 200-400 мкм. Таким образом вторая компонента отражает преобладание в опушении трихомов умеренной длины. Например, сорт 'Танго' характеризуется преобладанием коротких трихомов (~100 мкм), а сорт 'Великан', трихомов умеренной длины (~200 мкм). Сорта 'Великан', 'Танай', 'Кемеровчанин' характеризуются наибольшим отклонением от других в пространстве главных компонент, т.е. демонстрируют наибольшие отклонения по свойствам опушения листьев по отношению к основной массе исследованных растений.

Умеренная доля объясненной дисперсии для первых двух компонент (~60%) указывает, что изменчивость опушения листьев картофеля является комплексной и не может быть сведена к изменению лишь небольшого числа параметров, таких как число трихомов и их средняя длина.

Таким образом, предложенный метод оценки количественных характеристик опушения листьев картофеля показал высокую производительность и эффективность. С его помощью удалось охарактеризовать количественные характеристики опушения листьев для 35 сортов картофеля, выделить сорта, имеющие наиболее интенсивное опушение ('Великан', 'Танай', 'Кемеровчанин').

2.8. Высокопроизводительное фенотипирование опушения листа табака

Ранее, для изучения роли пролина в поддержании стрессоустойчивости у растений, в лаборатории генной инженерии ИЦиГ СО РАН были получены модельные трансгенные линии табака со сниженным уровнем активности гена деградации пролина – пролиндегидрогеназы (ПДГ) и увеличенным содержанием пролина (Кочетов и соавт., 2004). Было показано, что растения табака с модифицированной экспрессией генов метаболизма пролина проявляют большую устойчивость при увеличении содержания пролина в листьях - устойчивость к засолению, к засухе, низким температурам, к солям тяжелым металлов. (Колодяжная и соавт., 2006, Ибрагимова и соавт., 2012, Ибрагимова и соавт., 2015). Содержание пролина у полученных ГМ линий табака оказалось различным, что, вероятно, отражает активность введенной конструкции. При этом уровень содержания пролина не всегда напрямую коррелировал с уровнем устойчивости. Так, линия Pdh8, имея относительно меньший уровень содержания пролина в сравнении с линиями Pdh5 и Pdh6, проявляла большую устойчивость к неблагоприятным факторам окружающей среды (Ибрагимова и соавт., 2012).

Аминокислота пролин играет важную роль в процессе онтогенеза растений и в ответе на стресс (Trovato et al., 2019, Dar et al., 2016). Концентрация пролина в растительных клетках многократно возрастает в ответ на повышение или понижение температуры, засуху, засоление почвы, недостаток питательных веществ, повышенный уровень УФ -излучения или воздействие тяжелых металлов, вызывая у растений осмотический стресс (Кузнецов, Шевякова 1999). Действуя как совместимый осмолит, пролин, играет важную роль в оксидативном стрессе. Повышение уровня пролина в клетке приводит к модуляции тургора клетки, тем самым создавая

осмотический баланс, стабилизирует клеточные мембраны, структуры белков и ферментов, препятствуя утечке электролитов в клетке и окислительному стрессу. Таким образом, пролин выступает как сигнальная молекула стрессового ответа у растений (Nayat et al., 2012).

Изменения концентраций пролина являются лишь одним из множества ответов растений на стресс: физиологических, морфологических, анатомических (Nameed et al., 2010; Pyas et al., 2020). Все большее число данных свидетельствует об участии пролина в процессах роста и развития растений (Trovato et al., 2019).

Оказалось, что помимо проявления устойчивости растений к различным видам стресса, у полученных в лаборатории генной инженерии ИЦиГ СО РАН ГМ растений табака варьирует степень опущения листьев в обычных, не стрессовых условиях. Однако, достоверно оценить изменчивость опущения на основе визуального и тактильного анализа было затруднительно.

В настоящем разделе описаны результаты применения технологий высокопроизводительного компьютерного фенотипирования для оценки количественных характеристик опущения листа у растений ГМ линий табака со сниженной активностью гена пролиндегидрогеназы и увеличенным содержанием пролина, и контрольной линии SR1. Метод оценки количества трихомов и их длины на основе анализа изображений сгиба листа, разработанный нами ранее для пшеницы (Genaev et al., 2012), был адаптирован для табака.

2.8.1 Растительный материал и условия выращивания генетически модифицированных линий табака

ГМ линии табака с измененным паттерном экспрессии гена, контролирующего деградацию пролина - пролиндегидрогеназы (ПДГ) в клетке были получены ранее с использованием сорта табака 'Petit Havana' SR-1 (линия SR1, *Nicotiana tabacum* L.), методом агробактериальной трансформации листовых дисков. Искусственная супрессия гена ПДГ была достигнута путем введения конструкции в геном табака, содержащей фрагмент гена ПДГ арабидопсиса в антисмысловой ориентации, помещенный под управление 35S РНК вируса мозаики цветной капусты. В ГМ линиях табака с этой конструкции считывается мРНК, образующая гетеродуплекс с мРНК гена ПДГ растения-хозяина. Это приводит к индукции механизма РНК-интерференции, в результате чего происходит частичное разрушение мРНК гена ПДГ и снижение его активности (Кочетов и соавт., 2004). Полученные ГМ линии табака характеризовались увеличенным содержанием пролина по сравнению с исходной линией табака SR1 (табл. 2.13) (Колодяжная и соавт., 2006).

Таблица 2.13. Содержание пролина у ГМ линий табака.

Линии	Содержание пролина	
	Пролин сырой массы, мкг/г	Сравнение с контролем, %
Pdh5	1294 ± 127	200
Pdh6	1115 ± 297	172
Pdh8	864 ± 306	133
SR1(контроль)	648 ± 94	100

Растения ГМ линии табака (Pdh5, Pdh6, Pdh8) и контрольной линии SR1

выращивали в гидропонной теплице (ЦКП ЛИВР ИЦИГ СО РАН) в керамзитной смеси, с регулярным поливом раствором Кноппа. Для анализа брали правильно сформированные листья, без признаков поражения и деформации из среднего яруса, у растений в фазе цветения.

2.8.2. Подготовка образцов и получение изображений

Подготовку препаратов для определения плотности опушения листа проводили согласно протоколу, приведенному в работе Дорошкова с соавторами (Дорошков и соавт., 2016). С каждого анализируемого листа вырезали не менее трех фрагментов размером 10×20 мм, заключенных между радиальными жилками листа, сгибали его пополам, помещали на предметное стекло и фиксировали липкой лентой так, чтобы край сгиба листа оставался свободным. Препараты готовились только из адаксиальной стороны листа (рис. 2.35). Параметры опушения оценивали на препаратах листьев, собранных с 5 растений каждой из исследуемых линий. Снимали показания с 4-х сгибов каждого листа.

Съемка подготовленных препаратов производилась на микроскопе Zeiss Axioscop 2 plus, при увеличении 5 (объектив) × 10 (окуляр) (ЦКП микроскопии ИЦИГ СО РАН).

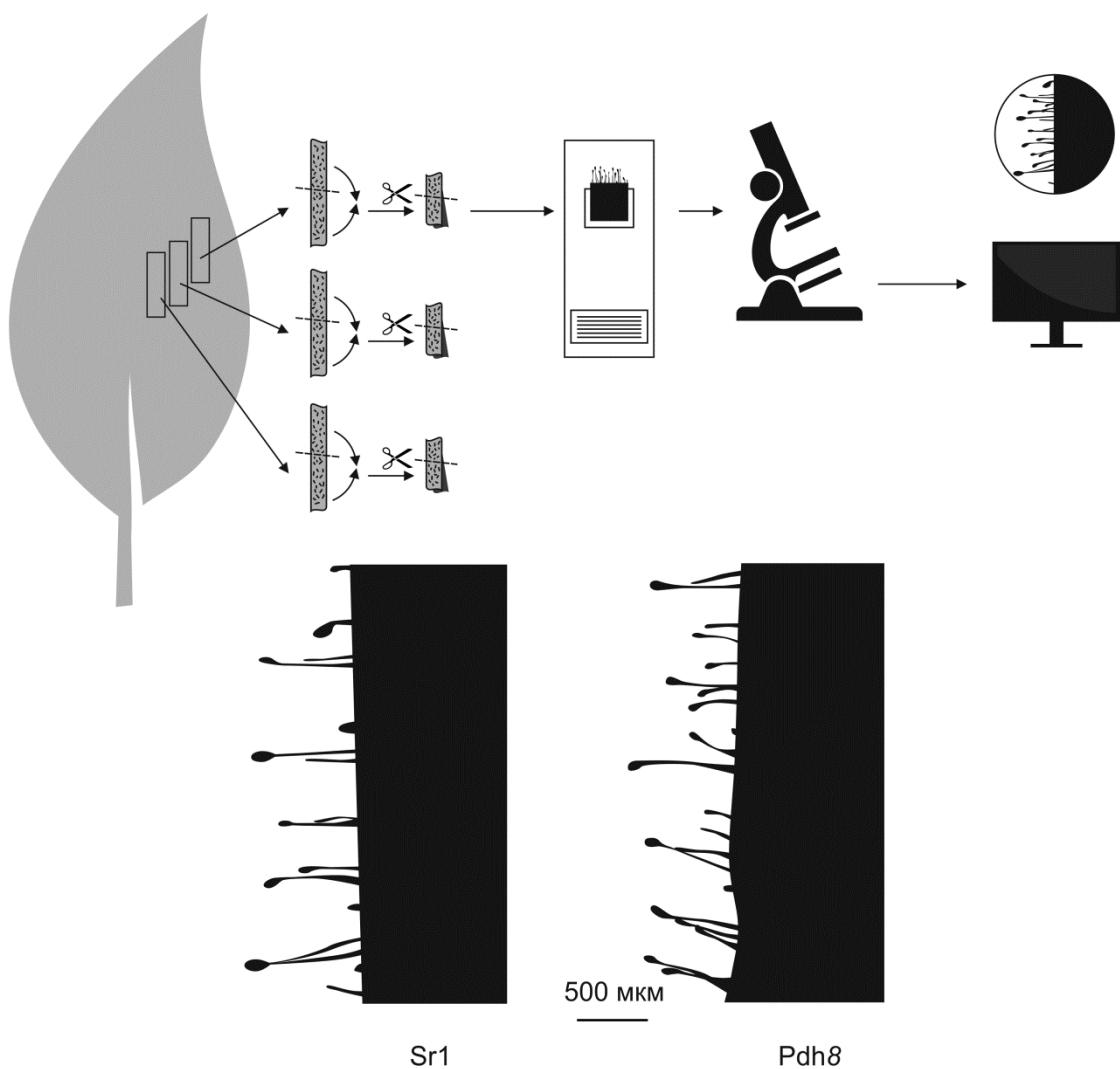


Рисунок. 2.35. Подготовка фрагментов листа для микрофотосъемки и изображение сгиба листа табака контрольной линии SR1 и ГМ линии Pdh8, используемое для извлечения данных.

2.8.3. Модификация параметров алгоритма для анализа опушения листа табака

В отличие от простых, несекретирующих трихомов пшеницы, трихомы листа табака характеризуются бóльшими размерами, высокой плотностью расположения и высокой метаболической активностью. У табака два вида железистых трихомов: короткие с одноклеточной головкой и длинные трихомы с многоклеточной или одноклеточной головкой (Akers et al., 1978). Поэтому в настоящей работе нам необходимо было модифицировать алгоритм LHDetect2, предложенный ранее для автоматического анализа опушения листьев пшеницы (Genaev et al., 2012).

В настоящей работе параметры модели алгоритма LHDetect2 определялись из условия максимизации меры F на тестовой выборке из 20 изображений сгибов листа табака, тестирование производилось на независимой выборке из 53 изображений, не входивших в обучающую выборку. Однако точность алгоритма, разработанного нами для пшеницы, на изображениях трихомов табака оказалась низкой ($F = 0.37$). Причина этому – секретизирующие головки сферической и овальной формы, не имеющие резкого изгиба линии границы в области вершины. Для большинства пикселей границы головки степень изгиба контура оказывается одинаковой, и по описанному выше критерию многие пиксели подходили на кандидатуру вершины.

Чтобы избежать этой неопределенности и выбрать вершину, удаленную от основания головки трихома, была проведена модификация алгоритма (рис. 2.36). В результате модификации, значения параметра W для каждого трихома в процессе поиска вершины менялись в пределах от W_{\min} до W_{\max} , где W_{\min} и W_{\max} – целые числа, параметры алгоритма. На этапе поиска вершины трихомов, как и в случае пшеницы, производится поиск пикселей, для которых кривизна

линии границы максимальна. Как и в предыдущей версии, для текущего пикселя границы определяется пара пикселей, удаленных от анализируемого на W_i пикселей (вперед и назад по направлению обхода). При этом значения W_i последовательно увеличиваются от W_{\min} до W_{\max} с шагом 1 пиксель. Для каждой такой пары пикселей измеряется величина угла α_i с вершиной в тестируемом пикселе. Если на определенном шаге α_i становится меньше порогового значения α , перебор останавливается и тестируемый пиксель считается потенциальной вершиной трихома (рис. 2.36).

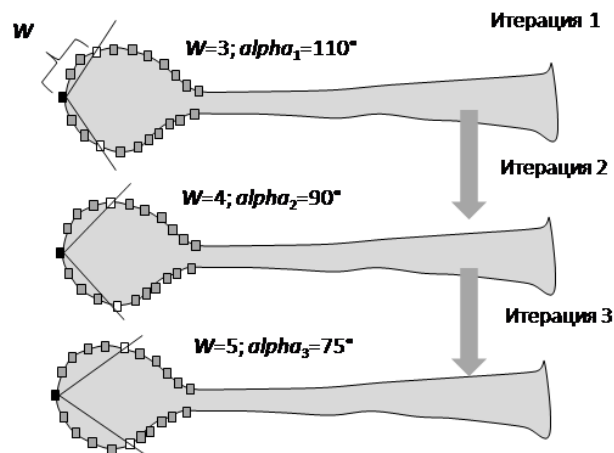


Рисунок 2.36. Модификация алгоритма поиска вершины для анализа трихомов табака. Серым контуром обозначен трихом на трех итерациях алгоритма. Серыми квадратами показаны пиксели границы контура трихома в области головки. Проиллюстрировано изменение угла α_i при различных значениях параметра W для пикселя контура секретирующей головки одного трихома, показанного черным цветом (кандидат на вершину). В данном примере пороговое значение $\alpha=70^\circ$. Белым цветом показаны пиксели границы, удаленные от текущего, на W пикселей. Линиями показан угол, образованный тремя пикселями. Критерий $\alpha_i < \alpha$ будет выполняться на четвертой итерации алгоритма. Таким образом, пиксель, показанный черным, является вершиной трихома.

Поскольку головка трихома табака имеет форму, вытянутую вдоль оси, проходящей через середину основания, данное условие будет выполняться для

пикселей границы, расположенных на стороне головки, противоположной основанию трихома. После модификации алгоритма был получен новый набор оптимальных параметров LHDetect2 на обучающей выборке изображений: $\alpha=90^\circ$, $min_trichome=10$ пикселей, $percent=30$, $\sigma=1.0$, $W_{min}=4$, $W_{max}=30$.

Точность оценки количества трихомов, определяли на основе коэффициента корреляции Пирсона - для истинного и предсказанного значений, средней абсолютной ошибки (MAE) и средней абсолютной процентной ошибки (MAPE) (формулы 1.2 и 1.4):

Для того чтобы определить ошибку автоматической оценки длин трихомов, мы вручную, с помощью инструмента “segmented line” из пакета программы ImageJ (rsbweb.nih.gov/ij/) на тестовом наборе изображений для каждого трихома проводили центральную линию и оценивали её длину. Затем для каждого трихома сравнивали это значение со значением, полученным автоматически (использовали коэффициент корреляции Пирсона, меры MAE и MAPE).

Необходимо отметить, что предложенный нами алгоритм не позволяет различить на изображении трихома с секретирующими головками и без них, он рассчитывает количественные характеристики общего пула всех трихомов, не разделяя их на два типа.

Модифицированная программа LHDetect2, как и прежняя ее версия для пшеницы, по итогам анализа каждого изображения выдает список числа трихомов на изображении и для каждого из них оценку длины в мкм. На основе этих данных мы оценивали параметры: n_f – число трихомов на изображении, l_m – среднюю длину трихомов на изображении. Для каждого генотипа нами было получено несколько десятков изображений для разных листьев и разных участков листа. Сравнивались распределения n_f и l_m для образцов разных

генотипов, а также распределения трихомов по длинам для этих образцов.

Результаты оценки точности обеих версий алгоритмов расчета количественных характеристик опушения показывают, что алгоритм с адаптивным выбором параметра W оказывается значительно более точным. Общая мера точности $F=0.90$, что почти в три раза выше, чем у алгоритма с постоянным значением W ($F=0.37$). Ошибки MAE и MAPE для алгоритма с адаптивным выбором параметра W составили, соответственно, 0.78 и 7. Эти параметры также в несколько раз меньше, чем для алгоритма с постоянным значением W (4.57 и 52, соответственно). Таким образом, алгоритм с адаптивным выбором параметра W оказывается значительно более точным.

Диаграммы рассеяния для числа трихомов, идентифицированных на изображении автоматическим методом и вручную, показаны на рис. 2.37. Из рисунка следует, что в случае использования модифицированного алгоритма, точки располагаются близко к прямой $y=x$, а коэффициент корреляции между двумя оценками близок к 1 ($r=0.927$). В случае использования версии LHDetect2 с постоянным значением параметра W , точки располагаются ниже прямой $y=x$, что свидетельствует о существенной недооценке числа трихомов. О более низкой точности этой версии алгоритма свидетельствует и значение коэффициента корреляции между двумя значениями, $r=0.535$, что почти в два раза меньше, чем в случае алгоритма с адаптивно меняющимся параметром W .

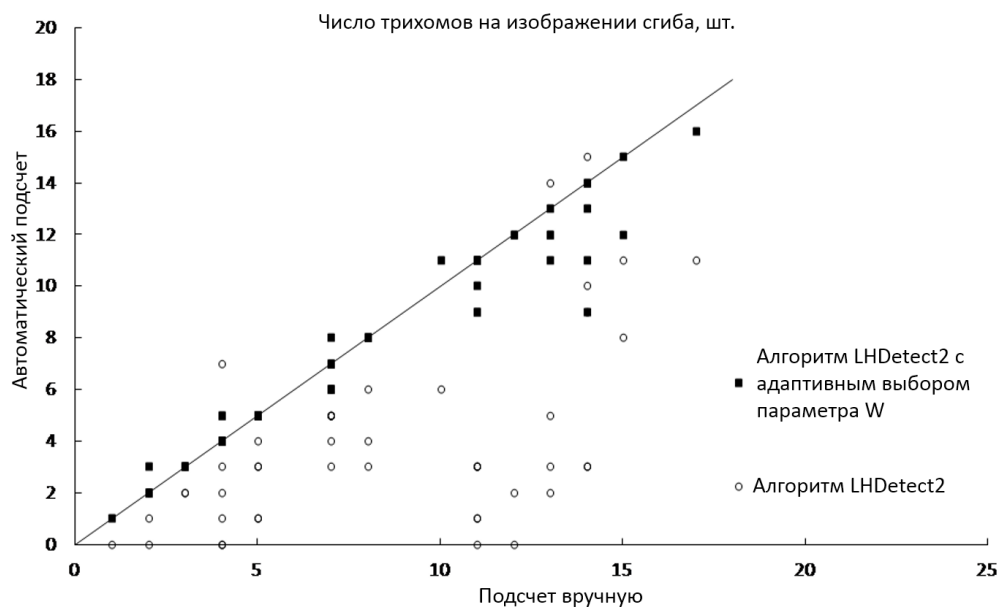


Рисунок 2.37. Зависимость между оценками количества трихомов на микроизображениях сгиба листа табака на тестовых изображениях, полученных автоматическим (оси Y) и ручным (оси X) подсчетом. Маркеры двух типов соответствуют различным вариантам алгоритма определения вершин трихомов: черные квадраты – использовалась версия LHDetect2 с изменяющимся параметром W ; белые кружки – использовалась версия LHDetect2 с фиксированным значением параметра W . Прямая линия соответствует зависимости $y=x$.

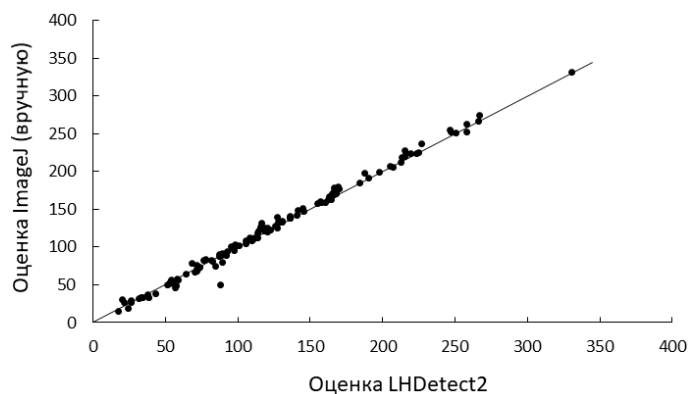


Рисунок 2.38. Сравнение оценок длины трихомов на изображении сгиба листа табака, полученных автоматически программой LHDetect2 с изменяющимся параметром W (ось X) и вручную, с помощью программы ImageJ (ось Y).

На рис. 2.38 представлена диаграмма рассеяния значений длин трихомов на изображении, определенной автоматически (алгоритм с изменяющимся параметром W) и вручную, при помощи пакета ImageJ для 20 тестовых изображений (125 трихомов). Коэффициент корреляции Пирсона между двумя оценками составил $r=0.996$. Средняя абсолютная ошибка (MAE) при этом составила 4.429 пикселя.

Полученные результаты показывают, что модификация алгоритма оценки количественных характеристик опушения для табака, основанная на использовании адаптивного изменения интервала W , позволила повысить точность и обеспечить возможность изучения количественных характеристик опушения. Предложенный метод позволяет получить оценки как количества трихомов, так и их длины с высокой степенью точности, при этом затрачивая гораздо меньше времени, чем в случае подсчета вручную. Это позволяет анализировать большее число изображений.

2.8.4. Сравнение характеристик опушения у ГМ-линий и контрольного сорта табака

Предварительный визуальный и тактильный анализ показал, что морфология опушения у ГМ линий табака Pdh5, Pdh6, Pdh8 в сравнении с контрольным сортом SR1 различаются, однако определить степень различий было невозможно. Автоматический анализ изображений сгибов листа позволил оценить количественные характеристики опушения листьев табака. Основные статистики для числа трихомов на изображении (n_t), средней длины трихомов (l_m), а также оценка коэффициента корреляции Пирсона для этих двух значений в выборке растений различных генотипов ($r(n_t, l_m)$) приведены в табл. 2.14.

Таблица 2.14. Количественные характеристики опушения ГМ-линий табака (Pdh) и контрольной линии SR1. Параметры, которые достоверно отличаются у ГМ линий от контрольной, отмечены символом * ($p < 0.01$).

Линии	Кол-во изображений N	Среднее число трихомов, (n_f) и 95% доверительный интервал	$Var(n_f)$	Длина трихомов, l_m , мкм, и 95% доверительный интервал	$Var(l_m)$, мкм ²	$r(n_f, l_m)$
SR1	53	4.78±0.35	1.67	308.39±19.82	5271.98	-0.06
Pdh5	15	7.10±0.75*	2.14	250.36±22.32*	1857.82	-0.44
Pdh6	16	7.75±0.78*	2.26	229.74±12.50*	586.89	-0.06
Pdh8	24	13.19±0.72*	3.02	273.57±16.98*	690.47	-0.13

Оказалось, что средние значения количества трихомов и их средняя длина для ГМ линий табака достоверно ($p < 0.01$) отличаются от таковых у растений контрольного генотипа SR1. Для генотипа Pdh5 уровень значимости отличий от линии SR1 составил по параметру n_f $p < 1.7e-5$, по параметру l_m $p < 0.0004$, для линии Pdh6 уровень значимости составил, соответственно, для n_f $p < 3.61e-7$ и для l_m $p < 4.6e-9$, а для линия Pdh8: для n_f $p < 1.5e-21$; для l_m $p < 0.002$. При этом оказалось, что у всех ГМ растений количество трихомов достоверно выше по сравнению с контрольной линией. Так, для линий Pdh5 и Pdh6 значение выше почти в полтора раза, а для линии Pdh8 – почти в 3 раза. Для средней длины трихомов наоборот, этот параметр для растений всех ГМ линий ниже, чем у контрольных растений. Наибольшие отличия по длине (в 1.3 раза) наблюдаются для растений линии Pdh6. Анализ коэффициентов корреляции Пирсона между количеством трихомов и их средней длиной у листьев табака показал слабую отрицательную зависимость: значения коэффициентов

корреляции близки по модулю к нулю и отрицательны как у контрольной линии SR1, так и у двух ГМ линий (Pdh6 и Pdh8). Для линии Pdh5 значение коэффициента корреляции $r(n_f, l_m)$ достаточно высоко по модулю (-0.44), но все же достоверно не отличается от 0 ($p < 0.09$). Таким образом, между числом трихомов и их средней длиной не выявлено статистической связи.

Из табл. 2.14 также следует, что в ряду генотипов Pdh5-Pdh6-Pdh8 идет увеличение количества трихомов на единицу площади, однако для длины каких-либо тенденций не наблюдается. Разработанный нами подход позволяет получить распределение трихомов по длинам в виде гистограмм, которые приведены на рис. 2.39. Гистограммы составлены для длин трихомов с интервалом шага 126 мкм. Из рис. 2.39 (А) следует, что наиболее сильные различия между количеством трихомов у линии SR1 и ГМ линий наблюдаются в интервалах длин от 0 до 380 мкм. Такие различия выражены как при сравнении растений ГМ линий и контрольной линии SR1, так и при сравнении ГМ линий между собой. При длине трихомов в пределах 400-600 мкм и более различия между количеством трихомов у SR1 и линий Pdh5 и Pdh6 практически отсутствуют, а для линии Pdh8 - достоверны. При длине трихомов более 700 мкм они исчезают для всех генотипов, однако и количество трихомов в этом интервале длин практически неотличимо от нуля. Отметим, что если в распределении трихомов по длинам у ГМ линий наблюдается пик в области 252 мкм, то для SR1 пик смещен в область большей длины (378 мкм), а для интервалов большей длины число трихомов у данной линии больше, чем у Pdh5 и Pdh6. Это, вероятно, обуславливает в итоге большее значение средней длины трихомов для линии SR1 (табл. 2.14).

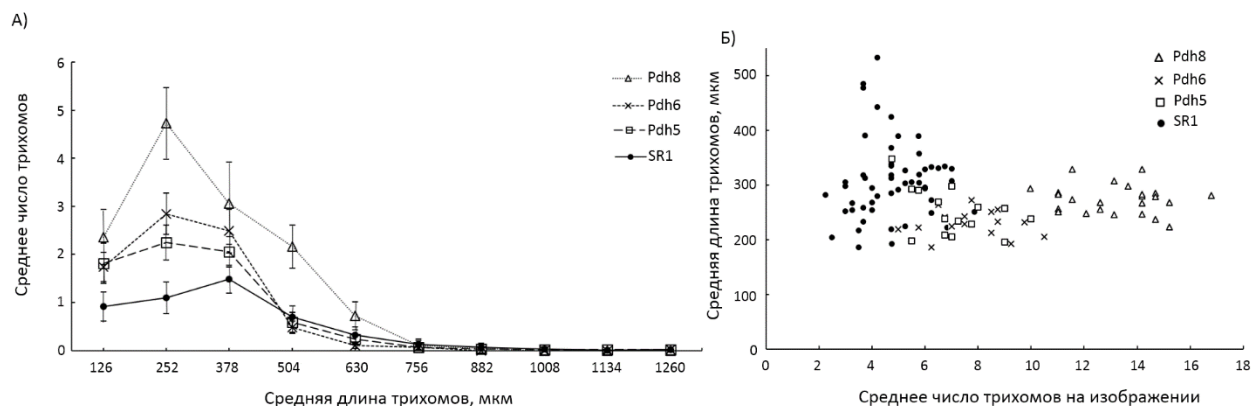


Рисунок 2.39. Сравнение количественных характеристик опушения листа контрольной (SR1) и трех ГМ линий табака.

А) гистограммы распределения трихомов по длинам на изображении. По оси Х отложены интервалы длин трихомов с шагом 126 мкм, по оси Y – количество трихомов интервале длин, усреднённое по всем изображениям. Вертикальными линиями показаны 95%-доверительные интервалы среднего значения количества трихомов; Б) диаграмма рассеяния образцов листьев табака для четырех генотипов по значениям среднего числа (ось X) и средней длины (ось Y) трихомов на изображении.

Была также построена диаграмма рассеяния для образцов листьев табака по параметрам n_f и l_m , (рис. 2.39 Б). Образцы контрольной линии SR1 располагаются в области низких значений количества трихомов (от 2 до 7). Разброс значений количества трихомов на изображении для линий Pdh5 и Pdh6 находится почти в одинаковых пределах (4-11 шт.), что и обуславливает близость средних значений этого параметра для двух групп растений (табл. 2.14). Далеко по значениям n_f отстоят от других генотипов образцы линии Pdh8: количество трихомов варьирует от 10 до 17, этот интервал практически не пересекается с интервалами n_f для других генотипов. Что касается средней длины трихомов, то по этому параметру растения ГМ линий близки (у всех разброс в пределах 200-300 мкм), а образцы генотипа SR1 демонстрируют более широкий разброс 180-530 мкм, с заметным числом образцов со средней

длиной трихомов выше 300. Все это обуславливает более высокое значение средней длины трихомов и ее дисперсию у растений SR1 (табл. 2.14). Также на этой диаграмме наблюдается отсутствие какой-либо заметной корреляции между n_f и l_m у рассмотренных образцов растений, что согласуется с данными, представленными в табл. 2.14.

Таким образом, полученные нами результаты демонстрируют, что ГМ линии табака с измененным паттерном экспрессии гена пролиндегидрогеназы и повышенным содержанием пролина по сравнению с контрольным генотипом имеет достоверно большее количество трихомов на адаксиальной стороне листа, и при этом длина трихомов достоверно меньше, чем у растений контрольной линии SR1.

Все большее число накопленных данных свидетельствует об участии пролина в онтогенезе растений в отсутствие стресса (Trovato et al. 2019). Он обеспечивает растительные клетки вегетативных тканей энергией, способствуя их более быстрому росту и развитию (Mattioli et al. 2009). Низкий уровень пролина приводит к нарушению клеточного деления и к формированию аномальных зародышей, что было показано на кукурузе (Zadebagheri et al. 2014). Известно, что пролин играет важную роль в делении клетки, участвует в регуляции генов клеточного цикла, в частности, циклинов, а также является компонентом структурных белков клеточной стенки, выполняющих функцию механической поддержки клеточной стенки (Biancucci et al. 2015).

В условиях стресса, изменение степени опушения листьев растений под воздействием высокой концентрации пролина, может вносить дополнительный вклад в устойчивость организма к стрессу. Увеличение плотности трихомов на адаксиальной поверхности листа и уменьшение их длины, по-видимому, является одной из универсальных и неспецифических реакций растений на

стресс. В настоящей работе мы наблюдаем аналогичный эффект при постоянно высоком уровне содержания пролина в ГМ линиях табака. Очевидно, что увеличенное содержание пролина в ГМ линиях табака, способствует более интенсивному делению клеток эпидермиса и формированию большего числа трихомов.

Выявленный факт усиления степени опущения листа у ГМ – линий с повышенным уровнем свободного пролина заслуживает внимания. Табак – одна экономически важных индустриальных культур. Практически все органы табака покрыты железистыми трихомами. Они накапливают и секретируют: терпены, сахара, никотин, производные жирных кислот и флавоноиды. Полученные данные свидетельствуют о влиянии пролина, как важного компонента метаболома, на фенотип растений, а ГМ линии табака представляют собой перспективную генетическую модель для изучения роли пролина в морфогенезе растений.

2.9. Заключение по главе 2

Исследования фенотипического проявления опущения листа растений являются очень трудоёмкими, в том числе и у мягких пшениц. Это визуальный и тактильный анализ, визуальный анализ микроскопических изображений. Это сильно затрудняло проведение генетических исследований по такому важному признаку, как опущение листа. С середины 70-х годов 20-го века до начала 2000-х годов у мягкой пшеницы было выявлено и картировано два гена, контролирующих опущение, *HL1* и *HL2*, на хромосомах 4В и 7В, соответственно. Однако оставались неизвестными характер их действия, распространённость среди сортов, наличие аллелей и т.д. Разработка метода цифровой количественной оценки характеристик опущения позволила существенно ускорить изучение этого признака. Алгоритм LHDetect2 для

автоматического определения количественных характеристик позволяет с высокой точностью определять количество трихомов на изображении сгиба листа, длины трихомов. Дополнительно, в результате работы он определяет: площадь трихомов на изображении, положение оснований трихомов (верхней и нижней границы основания). Все это позволяет с высокой точностью характеризовать опушение листьев пшеницы.

Применение данного метода для анализа опушения картофеля и его адаптация для изучения опушения табака показали, что этот метод можно обобщить для изучения и двудольных растений и изучить на основе такого метода разнообразие опушения у различных представителей одного вида или изменчивость под влиянием генетических изменений.

Таким образом, с помощью разработанных технологий высокопроизводительного фенотипирования данные о генетическом контроле и средовой регуляции опушения листа и его вкладе в физиологический ответ на засуху существенно пополнены по сравнению с исследованиями предыдущих десятилетий. Эта работа показала, что метод компьютерного фенотипирования обладает высокой разрешающей способностью в широком диапазоне характеристик опушения и позволяет быстро оценивать большое количество растений.

ГЛАВА 3. ВЫСОКОПРОИЗВОДИТЕЛЬНОЕ ФЕНОТИПИРОВАНИЕ ЗЕРЕН ПШЕНИЦЫ

3.1. Зерна злаков и их характеристики важные для селекционно-генетических исследований

Одно из важнейших направлений селекционно-генетических исследований мягкой пшеницы - поиск и изучение генов, контролирующих показатели урожайности. Некоторые из них - это характеристики семян пшеницы (размер и форма), которые, как было показано, связаны с массой семян (Brinton, Uauy, 2019; Huang et al., 2015; Zhang et al., 2014). Размер и форма семян связаны с качеством и составом муки: мелкие зерна могут способствовать повышению качества муки, оказывая при этом негативное влияние на ее выход (Baasandorj et al., 2015). Для поиска генов, контролирующих эти признаки у пшеницы, используется анализ локусов количественных признаков (Quantitative Trait Loci, QTL). Этот анализ позволяет идентифицировать наборы маркеров, которые связаны с признаками размера или формы семян. Исследования показали, что размер и форма семян пшеницы контролируются большим количеством локусов, расположенных почти на всех хромосомах (Li et al., 2021; Ali et al., 2020; Alemu et al., 2020; Kumari et al., 2018; Kumar et al., 2016; Williams and Sorrells, 2014; Gegas et al., 2010; Breseghello and Sorrells, 2007). Идентификация этих локусов в сочетании с молекулярным анализом может помочь в поиске генов, участвующих в контроле массы или размера семян (Huang et al., 2002; Ma et al., 2016; Ma et al., 2015; Ma et al., 2012; Su et al., 2011; Tikhenko, 2020). Основываясь на генетических и молекулярных исследованиях как модельного организма *A. thaliana*, так и злаков, в настоящее время установлено, что на массу семян влияют многочисленные факторы на генетическом и молекулярном уровнях, которые приводят к динамическим изменениям в делении клеток, увеличению

их размера и дифференцировке во время развития семени. Некоторые важные молекулярные механизмы такие как убиквитинирование, действие фитогормонов и G-белков, фотосинтез, эпигенетические модификации и микроРНК способствуют увеличению массы семян (Chen et al., 2021; Li et al., 2019). Знание путей, контролирующих развитие семян у хорошо изученных модельных организмов, позволяет определить приоритетные гены-кандидаты, контролирующие эти признаки у пшеницы, а также обеспечивает их дальнейшее изучение молекулярными методами (Ma et al., 2016).

Еще одной важной характеристикой семян пшеницы является цвет оболочки. Он характеризует содержащиеся в ней пигменты и метаболиты. Фиолетовая и синяя окраска семян определяется наличием антоцианов, желтоватый цвет может быть обусловлен наличием каротиноидов (Smolikova, Medvedev, 2015). Красновато-коричневая или темно-коричневая окраска семян обусловлена присутствием флавоноидов, таких как проантоцианидины и флорафены (Lachman et al., 2017). Генетический контроль формирования цвета как в семенах, так и в других органах растений осуществляется генами, кодирующими ферменты, участвующие в биосинтезе пигментов, а также регуляторными генами (Khlestkina, 2013). Для ряда пигментов эти гены хорошо изучены; однако для некоторых пигментов молекулярные механизмы биосинтеза недостаточно ясны (Glagoleva et al., 2020). Присутствие пигментов в оболочке семян влияет на различные технологические свойства семян (Ma et al., 2021) и связано с антиоксидантными свойствами (Li et al., 2015). Поэтому сорта и линии с разнообразной окраской семян представляют интерес в пищевой промышленности (Loskutov and Khlestkina, 2021; Khlestkina et al., 2016). Цвет оболочки семян пшеницы также связан с такими важными характеристиками, как всхожесть и предуборочное прорастание (Pre-Harvest Sprouting, PHS). Красные семена менее чувствительны к PHS (Groos et al.,

2002); поиск QTL, связанных с цветом оболочки семян и их устойчивости к PHS часто выполняется одновременно (Lin et al., 2016; Kumar et al., 2009).

Другой важной характеристикой семян является их жизнестойкость и всхожесть. Сохранность семян важна для фермеров, селекционеров и промышленных предприятий, заинтересованных в переработке семян и коммерческой торговле, а также для сохранения генетических ресурсов в генетических банках растений. Во время хранения семена накапливают повреждения на структурном и метаболическом уровнях, которые нарушают их функцию и снижают устойчивость к неблагоприятным внешним условиям (Priestley, 1986; Shvachko and Khlestkina, 2020). Внешние условия влияют на эти процессы: более высокие температуры и влажность негативно влияют на жизнеспособность семян и ускоряют их деградацию (Mira et al., 2015; Smolikova et al., 2021). Однако даже при хранении в специальных условиях генбанка семена постепенно теряют способность к прорастанию (Nagel, Börner, 2010; Guberac et al., 2003; Guzzon et al., 2021). В этой связи вопросы, связанные со старением семян, оказываются актуальными для поддержания жизнеспособности генетических коллекций.

Процессы, связанные со старением зерна, изучаются на биохимическом уровне (Rajjou, Debeaujon, 2008; Kurek et al., 2019; Pirredda et al., 2020), что позволяет определить их молекулярные механизмы и, по возможности, снизить воздействие повреждающих факторов при хранении (Нау, 2021). Также ведутся активные исследования по поиску генов, связанных с долговечностью семян (Arif et al., 2012; Clerckx et al., 2004; Börner et al., 2018; Smolikova et al., 2021). Это позволяет идентифицировать гены, способствующие более длительному хранению зерна, и определить механизмы генетического контроля длительной жизнеспособности.

Большую роль в этих исследованиях играют методы фенотипирования семян, не только их размера, формы, но и цвета оболочки. Для эффективной морфометрии зерна были предложены методы обработки цифровых изображений (Granitto et al., 2005; Pourreza et al., 2012; Tanabata et al., 2012). Большинство из этих подходов были реализованы с использованием программного обеспечения для настольных ПК. Типичные изображения для анализа – зерна на светлом фоне, полученные с помощью цифровой камеры или сканера (Herridge et al., 2011; Tanabata et al., 2012; Whan et al., 2014). Эти подходы позволяют пользователям оценивать большое количество морфометрических параметров зерна, описывающих форму и цвет (Bai et al., 2013). Компьютерные методы с использованием анализа изображений зерна также позволяют проводить автоматическую идентификацию сорта зерновых культур (Chen et al., 2010; Wiesnerová and Wiesner, 2008; Zapotoczny, 2011), определять содержание влаги в семенах и прогнозировать выход манной крупы у пшениц твердых сортов (Novaro et al., 2001; Tahir et al., 2007).

Методы цифрового анализа изображений также используются для изучения физиологических свойств зерен и их изменения под воздействием различных факторов на основе оценки морфологии и цвета семян (Farris et al., 2020; Huang et al., 2015; Goriewa-Duba et al., 2018; Martín-Gómez et al., 2019; Cervantes et al., 2019). Эти методы позволяют количественно оценить степень изменения параметров зерен и выявить среди них те, которые наиболее тесно связаны с физиологическим состоянием, включая старение (Dell'Aquila, 2009; Behtari et al., 2014). Цифровой анализ изображений был использован для выявления изменений цвета оболочки семян, известных как потемнение, при искусственном старении чечевицы (*Lens culinaris* Medick.), что позволило предположить, что прогрессирующий и непрерывный процесс ухудшения качества можно контролировать с помощью изменения цвета (Dell'Aquila,

2006). Эффект потемнения оболочки семян был также обнаружен у огурца (*Cucumis sativus* L.) после десятидневного хранения при пониженном содержании влаги и повышенной температуре (Dell'Aquila, 2009). Изменения цвета были также обнаружены после ускоренного старения семян *Medicago sativa* и *Onobrychis viciifolia* (Vehtari et al., 2014).

Технические средства для морфометрии зерна демонстрируют высокую производительность и точность, однако они установлены в ограниченном количестве лабораторий по исследованию растений. По-прежнему существует потребность в недорогих, высокопроизводительных методах анализа зерна (Whan et al., 2014).

Крупномасштабные эксперименты по селекции требуют обработки фенотипических данных значительного объема, часто в полевых условиях и, следовательно, без доступа к настольным компьютерам и сканерам. В этом случае цифровая камера является приемлемым вариантом, но изображения должны быть впоследствии скопированы на ноутбук или ПК.

Современные мобильные устройства (смартфоны и интернет-планшеты) содержат цифровые камеры с высоким разрешением. Мобильные устройства оснащены многоядерными процессорами с достаточной вычислительной мощностью для обработки и анализа изображений. Эти функции позволяют пользователям получать и обрабатывать изображения там, где это необходимо. Для морфометрии органов растений был разработан ряд приложений для мобильных устройств. Приложение Leafsnap (Kumar et al., 2012) способно идентифицировать виды растений в режиме реального времени на основе изображений их листьев: пользователь фотографирует лист растения с помощью мобильного устройства и отправляет изображения с камеры на удаленный сервер, где они обрабатываются. Leaf Doctor (Pethybridge and

Nelson, 2015) - это еще одно мобильное приложение, которое оценивает процент повреждения площади листа на основе изображений полуавтоматическим способом.

Данный раздел работы посвящен разработке и применению методов высокопроизводительного фенотипирования зерен пшеницы, как на основе мобильных устройств, так и с помощью цифровых фотокамер.

3.2. Приложение SeedCounter для морфометрии зерен пшеницы.

3.2.1. Протокол получения изображений зерен пшеницы для фенотипирования с помощью мобильных устройств

Для анализа изображений с помощью мобильных устройств необходимо получить цветное изображение зерен, произвольно размещенных на листе белой бумаги стандартного формата (A4, Letter, Legal, A3, A5, B4, B5 или B6). Рекомендуется свести к минимуму любой контакт между зернами. Чтобы уменьшить количество ошибок, пользователь должны обеспечить следующие условия для получения изображения: лист бумаги должен располагаться на темном фоне и следует избегать яркого бокового освещения.

Границы листа бумаги на изображении должны быть параллельны сторонам рамки кадра (рис. 3.1 А). Фиксированный размер бумаги позволяет рассчитать масштаб изображения и оценить размеры зерен в метрических единицах. Приложение SeedCounter получает изображения непосредственно с камеры мобильного устройства.

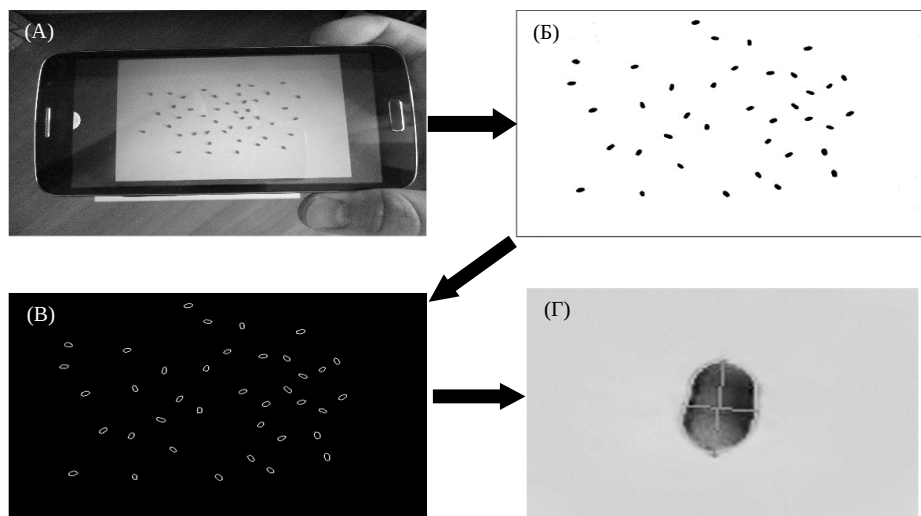


Рисунок 3.1. Основные этапы получения изображения зерен на листе бумаги. А) Получение изображения с помощью камеры мобильного устройства и распознавание бумаги. Б) Изображение после аффинного преобразования и бинаризации. В) Контуры зерен, идентифицированные на изображении. Д) Изображение зерна с основными осями, показанными пересеченными линиями.

3.2.2. Алгоритм обработки изображений для морфометрии зерен

Алгоритм реализован с использованием библиотеки обработки изображений OpenCV (Dawson-Howe, 2014;) и состоит из нескольких шагов.

Лист бумаги распознается как светлая область четырехугольной формы на темном фоне. Для распознавания исходное цветное изображение (рис. 3.1 А) преобразуется в оттенки серого с помощью функции `cvtColor()`. Для определения площади листа функция `adaptiveThreshold()` выполняет адаптивную бинаризацию всего изображения, а функция `canny()` используется для определения границ бумаги. Набор линий, близких к границам листа, генерируется функцией `HoughLinesP()` с параметром длины, изменяющимся от 20 до 80% от соответствующей стороны изображения. Из-за искажений на изображении не все эти линии для одной и той же стороны параллельны, а

линии на соседних сторонах не перпендикулярны. Поэтому, чтобы выбрать линии, приближающиеся к границам бумаги, мы группируем их относительно их взаимного угла и расстояния, получая четыре кластера линий, соответствующих сторонам бумаги. Для каждого кластера мы восстанавливаем линию границы листа с наименьшим расстоянием от пикселей линий кластера. Пересечения между линиями границ листа определяют вершины четырехугольного изображения на бумаге. Если форма листа бумаги на изображении отличается от прямоугольной, мы преобразуем изображение листа в прямоугольное с помощью аффинных преобразований. Этот шаг выполняется с использованием функции `getPerspectiveTransform()` для вычисления матрицы преобразования, а функция `Warperspective()` используется для преобразования изображения, делая противоположные края параллельными, а все углы равными 90° .

Зерна идентифицируются как контуры путем применения функции `findContours()` к фрагменту изображения, соответствующему листу бумаги. Мы производим дальнейшую корректировку границ зерен, используя локальную бинаризацию канала HSV для соседних областей исходного изображения. Локальная бинаризация уменьшает влияние затенения при определении границ зерен. Она включает преобразование локального сегмента изображения в цветовое пространство HSV и последующее преобразование в оттенки серого на основе параметров калибровки и цветовых гистограмм. Результирующий канал отражает степень соответствия пикселей изображения цвету зернистости. Локальная бинаризация дает более точное определение границ зерен.

Для определения границ зерен, которые на изображении соприкасаются друг с другом используется метод водораздела (Roerdink and Meijster, 2000),

реализованный в функции watershed(). Полученные контуры зерен аппроксимируются эллипсоидами, что позволяет оценить размер главной и второстепенной главных осей, соответствующих длине и ширине зерна (рис. 3.1 Г). Дополнительно программа определяет область изображения зерна и расстояние между геометрическим центром масс и точкой пересечения главных осей.

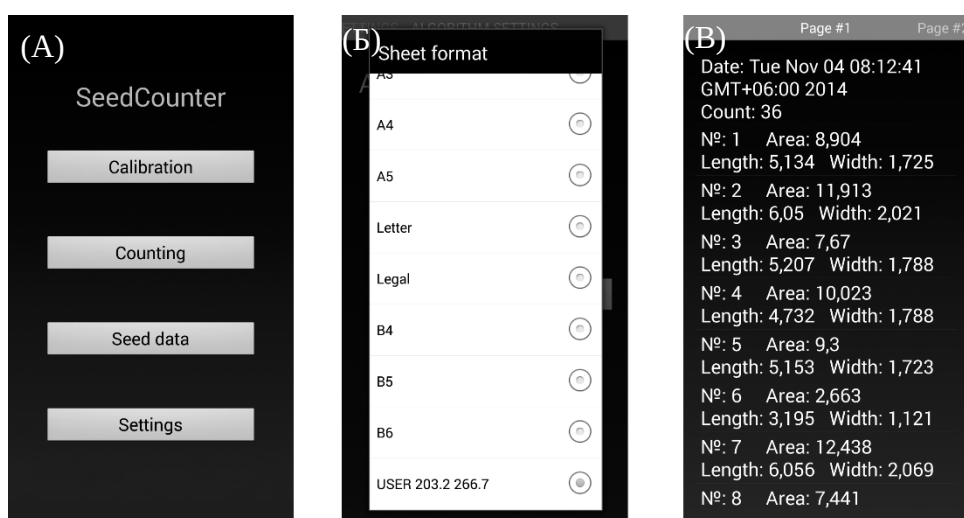


Рисунок 3.2. Интерфейс мобильного приложения SeedCounter. А) Основное меню. Б) Выбор параметра «Размер бумаги». В) Вид экрана с результатами работы программы (число зерен и длина, ширина, проективная площадь для каждого зерна).

Пользователь мобильного приложения может настроить параметры обработки изображений и распознавания семян, используя опцию "Calibration" в главном меню (рис. 3.2 А). Пользователь должен сделать снимок одного зерна на бумаге, обработать изображение и убедиться, что алгоритм правильно идентифицирует зерно и помечает его красной прямоугольной рамкой. Параметры алгоритма на этом этапе сохраняются автоматически. Пользователь также может использовать меню программы для определения размера листа

бумаги (включая определяемые пользователем размеры), а также разрешения камеры и изображения, чтобы включить алгоритм разделения семян прикосновением и бинаризацию в цветовом пространстве HSV.

Данные о количестве подсчитанных семян и параметрах формы семян для каждого семени хранятся в формате XML и могут быть отображены с помощью меню "Данные о семенах" (рис. 3.2 В). Пользователь может просмотреть данные, удалить их, экспортировать в формате tsv или отправить на веб-сервер SeedCounter. В последнем случае пользователь получает URL-адрес данных, который позволяет загружать данные в веб-браузере.

При оценке точности работы нашего метода мы рассмотрели два типа ошибок. Во-первых, мы оценили точность идентификации количества зерен. Для этого пятьдесят зерен пшеницы одного сорта высыпали на лист, и количество зерен оценивали программой. После этого одно зерно удаляли с листа, зерна перемешивали (без контроля контактов зерен) и снова оценивали количество зерен. Эта процедура была повторена 40 раз. Мы провели эту серию измерений, используя различные мобильные устройства, разрешения камер и условия освещения. Для каждой серии оценок количества зерен мы рассчитали среднюю абсолютную ошибку (MAE) и среднюю абсолютную процентную ошибку (MAPE) (формулы 1.3 и 1.4).

Во-вторых, мы оценили точность оценки длины и ширины зерна. Мы измерили длину и ширину 250 зерен пяти сортов пшеницы, причем каждое зерно располагалось в строгом порядке, используя микроскоп Carl Zeiss Aioscop 2 plus, оснащенный цифровой камерой с адаптером AhoCamHRc TV2/3c 0.63. Мы разместили зерна на листе бумаги в том же порядке и применили программное обеспечение SeedCounter для оценки их длины и ширины. Была проведена серия морфометрических измерений 250 зерен с

использованием различных мобильных устройств, разрешений камер и условий освещения. Для каждого эксперимента мы рассчитали MAE отдельно для длины и ширины и рассчитали средние значения. Та же процедура использовалась для расчета MAPE для ширины и длины. Коэффициенты корреляции Пирсона, r_w и r_l , были рассчитаны для этих параметров отдельно.

Мобильные устройства, использованные для эксперимента, включали смартфоны Samsung Galaxy Grand 2, Sony Ericsson XPERIA pro mini и интернет-планшет DNS AirTab m101w, работали под управлением ОС Android, использовалось максимальное разрешение камеры.

Мы использовали следующие три типа осветительных приборов: лампу дневного света мощностью 11 Вт (цветовая температура 4000 К, световой поток 900 лм), лампу дневного света мощностью 5 Вт (4000 К, 400 лм) и галогенную лампу мощностью 35 Вт (2700 К, 190 лм). Использовались четыре типа искусственного освещения: лампа дневного света мощностью 11 Вт (L1); лампа дневного света мощностью 11 Вт и две лампы дневного света мощностью 5 Вт (L2); лампа дневного света мощностью 11 Вт и четыре лампы дневного света мощностью 5 Вт (L3); и лампа дневного света мощностью 11 Вт, четыре лампы дневного света мощностью 5 Вт и галогенная лампа (L4). Лампы были установлены на высоте 60 см над листом бумаги. Лист помещали на стол с темной столешницей, и эксперименты проводили в темной комнате. Чтобы оценить точность измерений при дневном свете, мы также измерили зерна без использования искусственного освещения в пасмурную погоду в помещении и в ясный день на открытом воздухе. Подробная информация об условиях эксперимента приведена в табл. 3.1.

Таблица 3.1. Условия освещения для измерения точности морфометрии зерна пшеницы. Единицы измерения Люкс (лк) использовались для оценки интенсивности света в условиях естественного освещения.

Тип освещения	Источники освещения	Освещенность (лк)	Световая температура источника
L1	Лампа дневного света 11-Вт	900 лк	4000К
L2	Лампа дневного света 11-Вт, две лампы дневного света 5-Вт	1700 лк	4000К
L3	Лампа дневного света 11-Вт, четыре лампы дневного света 5-Вт	2500 лк	4000К
L4	Лампа дневного света 11-Вт, четыре лампы дневного света 5-Вт, галогенная лампа 35-Вт	2690 лк	4000К и 2700К
L5	Дневное освещение, пасмурный день, внутри здания	(1280 лк)	–
L6	Дневное освещение, солнечный день, снаружи здания	(656000 лк)	–

Мы использовали двухфакторный дисперсионный анализ для оценки влияния типа устройства и условий освещения на количество зерен и точность формы. Мы рассматривали тип устройства и освещение как независимые переменные, а оценки ошибок (MAE и MAPE) - как зависимые переменные.

Чтобы сравнить точность приложения SeedCounter с доступным программным обеспечением, мы сравнили наши результаты с измерениями, полученными с помощью приложения Smartgrain (Tanabata et al., 2016), работающего на персональном компьютере (Intel Core i7, 2400 МГц, 4 Гб оперативной памяти) и изображений со сканера HP Scanjet 3800 с разрешением

600 точек на дюйм.

Для проведения эксперимента по оценке точности метода мы использовали зерна следующих пяти сортов пшеницы из коллекции зерновых лаборатории хромосомной инженерии Института цитологии и генетики СО РАН: ‘Аленькая 1102 II-12’, 84/98w 99 II-13, ‘Синтетик 6x x-12’, ‘Purple Chance’ 4480 II-03 и ‘Альседо’ n-99. Растения были выращены в поле под Новосибирском в 2014 году. Эти сорта имеют зерна разной формы и размера. Сорт ‘Альседо’ имеет овальную форму и имеет среднюю длину ~ 7 мм и ширину ~ 3,6 мм. Сорт ‘Синтетик’ имеет удлиненную форму зерна и среднюю длину ~ 8 мм и ширину ~ 2,3 мм. Сорт ‘Аленькая’ имеет меньшие размеры, со средней длиной ~5 мм и средней шириной ~2,4 мм. Сорта 84/98 и ‘Purple Chance’ похожи по внешнему виду и имеют среднюю длину/ширину 6,5 мм/2,6 мм и 7 мм/2,9 мм соответственно.

Точность оценки количества зерен для различных серий экспериментов показана в табл. 3.2. В таблице показано, что МАЕ оценки количества зерен на листе близок к 1, а МАРЕ близок к 2%. Более детальный анализ показал, что наибольшие ошибки при подсчете количества зерен возникают, если два или более зерен на бумаге соприкасаются и что в условиях плохого освещения алгоритм не разделяет большую часть зерен. Если все зерна на листе разделены, ошибка подсчета семян исчезает.

Таблица 3.2. Оценка точности подсчета зерен мобильным приложением SeedCounter.

Эксперимент	MAE ^a (мм)	MAPE ^a (%)	$r(N_j, N_j')$ ^a
Sam_L1	1.425	0.035	0.996
Sam_L2	1.375	0.036	0.994
Sam_L3	0.65	0.015	0.998
Sam_L4	0.975	0.024	0.997
Sam_L5	1.15	0.029	0.992
Sam_L6	0.55	0.017	0.998
Sony_L1	1	0.024	0.995
Sony_L2	0.8	0.019	0.995
Sony_L3	0.675	0.017	0.996
Sony_L4	0.775	0.020	0.997
Sony_L5	0.75	0.018	0.996
Sony_L6	0.775	0.018	0.996
DNS_L1	1.2	0.031	0.997
DNS_L2	0.5	0.012	0.997
DNS_L3	0.125	0.003	0.999
DNS_L4	0.725	0.017	0.998
DNS_L5	1.175	0.030	0.996
DNS_L6	0.775	0.020	0.997

^a Средняя абсолютная ошибка (MAE), средняя абсолютная процентная ошибка (MAPE) и коэффициент корреляции Пирсона $r(N_j, N_j')$ между реальным числом зерен и оцененным с помощью программы.

Таблица 3.3. Точность оценок длины и ширины зерен пшеницы для приложения SeedCounter и программы SmartGrain.

Эксперимент	MAE ^a (мм)	MAPE ^a (%)	r_l^a	r_w^a
Sam_L1	0.284	7.453	0.974	0.903
Sam_L2	0.296	7.576	0.970	0.906
Sam_L3	0.283	7.339	0.972	0.906
Sam_L4	0.327	8.306	0.969	0.900
Sam_L5	0.398	9.081	0.918	0.867
Sam_L6	0.349	8.437	0.953	0.886
Sony_L1	0.313	8.277	0.972	0.877
Sony_L2	0.310	8.121	0.971	0.879
Sony_L3	0.298	7.787	0.974	0.886
Sony_L4	0.327	8.418	0.966	0.874
Sony_L5	0.301	7.727	0.967	0.877
Sony_L6	0.338	8.546	0.953	0.886
DNS_L1	0.295	7.852	0.976	0.884
DNS_L2	0.296	7.688	0.973	0.887
DNS_L3	0.287	7.730	0.979	0.889
DNS_L4	0.311	8.229	0.975	0.887
DNS_L5	0.351	8.798	0.957	0.843
DNS_L6	0.346	8.264	0.976	0.884
SmartGrain	0.305	6.973	0.980	0.942

^a Средняя абсолютная ошибка (MAE), средняя абсолютная процентная ошибка (MAPE) и коэффициент корреляции Пирсона для длины (r_l) и ширины (r_w) зерен между реальными и оцененными на основе анализа изображений характеристиками.

Точность оценки длины и ширины зерен различными устройствами в разных условиях показана в табл. 3.3. Таблица показывает, что точность оценки размера зерна составляла приблизительно 0.30 мм (среднее значение для всех серий: 0.31 мм), что составляет примерно 8% от линейных размеров зерна (среднее значение для всех серий: 8.03%). Коэффициенты корреляции между контрольной длиной и ее оценкой во всех экспериментах были не ниже 0.9. Для ширины зерна этот параметр был немного меньше, но все же больше 0.84. Интересно, что ошибки в оценках длины зерна для SeedCounter и Smartgrain близки друг к другу; однако для ширины зерна SmartGrain демонстрирует более высокую точность.

Мобильные устройства в среднем демонстрируют наилучшие показатели при оценке размера зерна в условиях освещения L3 (две лампы дневного света, световой поток 2500 лк). Наихудшая производительность была получена в условиях L5 (пасмурный день, в помещении).

Таблица 3.4. Оценка влияния типа мобильного устройства и условий освещения на точность морфометрии зерен пшеницы. Приведены p -значения для двухфакторного дисперсионного анализа. Жирный шрифт для значимых величин ($p < 0.05$).

Тип ошибки	Условия освещения	Тип устройства
Подсчет зерен, MAE	0.004	0.365
Подсчет зерен, MAPE	0.003	0.306
Оценка размера, MAE	0.036	0.771
Оценка размера, MAPE	0.094	0.890

Двухфакторный дисперсионный анализ показал, что условия освещения существенно влияют на оценку количества зерен, а также длины и ширины зерен ($p < 0.05$) (табл. 3.4). Интересно, что наибольшее среднее значение MAE

(формула 1.2) для подсчета зерна, 0.458, было получено для условий освещения с наименьшим световым потоком (только лампа L1 мощностью 11 Вт), тогда как другие условия освещения имели более низкие значения MAE: 0.058 для L2, 0.1 для L3, 0.058 для L4 и 0.275 для L5. Следует отметить, что погрешность подсчета семян в условиях без искусственного освещения меньше, чем при самом низком световом потоке, но больше, чем при всех других условиях контролируемого освещения. Результаты, приведенные в табл. 3.4, показывают, что тип устройства не оказывает существенного влияния на измерения количества зерен/размеров.

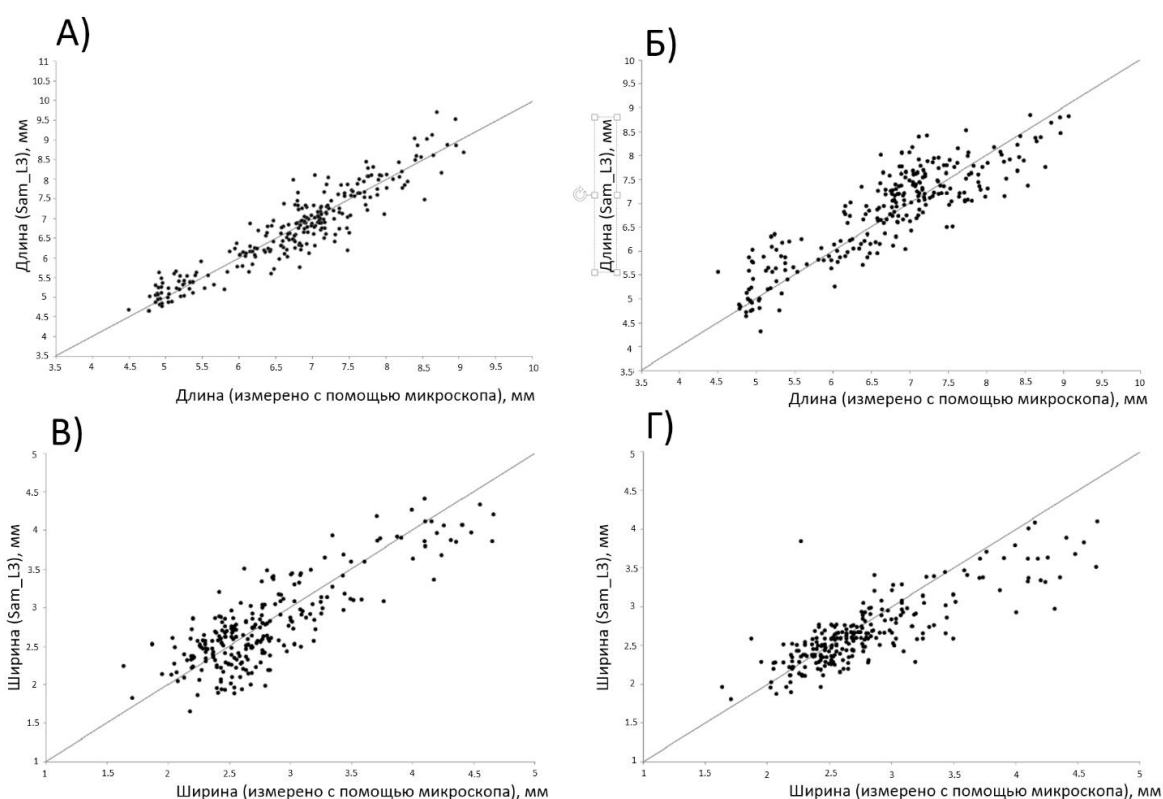


Рисунок 3.3. Диаграммы рассеяния для длины и ширины зерен оцененных с помощью Samsung mobile device (ось Y) по сравнению с оценками, сделанными с помощью микроскопа (ось X). А) Длина, измеренная в условиях освещения L3; Б) Длина, измеренная в условиях освещения L6; В) Ширина, измеренная в условиях освещения L3; Г) Ширина, измеренная в условиях освещения L6.

На рис. 3.3 показана диаграмма рассеяния измерений длины и ширины для 250 семян, полученных с помощью микроскопа и камеры Samsung с использованием ламп дневного света (L3) и солнечного света (L6) в условиях освещения. Этот рисунок демонстрирует, что при хороших условиях освещения оценки размера зерна, полученные мобильным устройством, согласуются с измерениями микроскопа. Однако в условиях солнечного света наше программное обеспечение имеет тенденцию недооценивать размеры зерен для более крупных зерен и переоценивать их для более мелких зерен. Этот эффект, вероятно, обусловлен эффектом тени, который вносит систематическое смещение в оценку размера зерна, когда изображение делается под прямыми яркими солнечными лучами.

Мы оценили время, затраченное на анализ одного изображения мобильными устройствами и программным обеспечением SmartGrain при разных разрешениях изображения. Результаты приведены в дополнительной таблице S3. Время обработки изображений с низким разрешением (2592×1944 пикселей) составляет примерно 30 секунд. Для камеры с более высоким разрешением (Samsung 3264×2448) это значение близко к 1 минуте. Интересно, что это сопоставимо со временем обработки изображений SmartGrain (при аналогичных разрешениях 3510×2550).

Таким образом, нами было разработано приложение для морфометрии зерна с использованием мобильных устройств. Протокол настройки анализа прост и использует лист белой бумаги стандартного размера в качестве фона для преобразования пикселей в метрическую шкалу. Чтобы проверить точность программы, мы провели серию экспериментов по анализу изображений с использованием трех типов мобильных устройств и шести условий освещения. В нашей работе средние абсолютные погрешности оценок длины/ширины

составляют примерно 7-10%, а коэффициенты корреляции длины и ширины между расчетными и фактическими значениями близки к 0.93 и 0.87 соответственно. Аналогичный анализ, проведенный в недавней работе Миллера и др. (Miller et al. 2016), продемонстрировал значение $R^2 = 0.996$ для оценки точности длины зерен кукурузы, на основе анализа цифровых изображений (планшетный сканер документов Epson V700, изображение 1200 точек на дюйм, 24-битное цветовое разрешение). Санкаран и др. (Sankaran et al. 2016) оценивали размер семян нута на основе изображений и для своего метода получили оценки коэффициентах корреляции Пирсона между ними и реальными значениями в диапазоне от 0.86 до 0.93 (цифровая зеркальная камера Canon 70D, установка штатива, зум-объектив 15-85 мм, разрешение изображения 2700x1800 пикселей). Уан и др. (Whan et al. 2014) проанализировали результаты измерений длины и ширины семян пшеницы с помощью следующих трех методов: GrainScan (разработанный авторами), Smartgrain (Tanabata et al., 2012) и SeedCount (Next Instruments, 2015). Они использовали сканер Epson Perfection V330 (Seiko Epson Corporation, Сува, Япония) для получения цветных изображений с разрешением 300 точек на дюйм. Уан и др. продемонстрировали, что среднее значение коэффициента корреляции Пирсона между истинными параметрами и оценками на основе изображений для GrainScan была очень высокой (0.981 – 0.996), аналогичной точности для SeedCount (0.991 – 0.994), в то время как средняя точность для Smartgrain была ниже (0.871 – 0.947). Наши результаты показывают, что точность и эффективность SeedCounter сопоставимы с результатами, полученными с помощью настольных ПК/сканеров/цифровых камер. Нужно иметь в виду, что для наших экспериментов мы использовали камеры с умеренным разрешением и слабо контролируруемыми условиями освещения.

Мы обнаружили, что освещение листа оказывает существенное влияние

на точность нашего метода, в отличие от типа смартфона (табл. 3.4). Мы использовали дисперсионный анализ с шестью различными типами освещения в качестве независимой переменной. Наши данные показали, что влияние самой яркости на точность не является однозначным: изображения, сделанные при высокой яркости под прямыми солнечными лучами, демонстрируют повышенную погрешность по сравнению с изображениями средней яркости и искусственным освещением. В условиях низкой освещенности (лампа дневного света мощностью 11 Вт или без искусственного освещения) точность оценки количества зерен снижается. Условия освещения галогенными лампами и лампами дневного света (экспериментальные условия SAM_L4, Sony_L4 и DNS_L4) вызвали небольшой эффект мерцания на изображениях. Этот эффект может усложнить процесс распознавания бумаги и привести к искажению результатов. Эффект мерцания также присутствовал в условиях SAM_L3 и DNS_L3, но его можно значительно подавить с помощью техники "ночной съемки". Расположение источников света и их угол наклона к поверхности бумаги могут исказить результаты измерений и ухудшить условия распознавания листа. Более яркий рассеянный свет устраняет искажения, связанные с появлением темных пятен на границах листа, которые могут быть неправильно распознаны как зерна, позволяет более эффективно отделять соприкасающиеся зерна и снижает вероятность того, что зерно на изображении сольется с фоном.

Мобильное приложение SeedCounter для устройств Android доступно бесплатно в магазине Android Play (<https://play.google.com/store/apps/details?id=org.wheatdb.seedcounter>). Для приложения SeedCounter требуется версия API Android 15, рекомендуется использовать Oracle / Sun JDK 6 или 7. SeedCounter использует библиотеку OpenCV для обработки изображений. SeedCounter распространяется по

лицензии BSD (Berkley Software Distribution).

Мобильные приложения могут значительно ускорить процесс подсчета количества зерен пшеницы в колосе. Для современных смартфонов с многоядерными процессорами время, необходимое для расчета примерно 50 зерен с помощью мобильного устройства, составляет несколько секунд, в зависимости от мобильного устройства и разрешения камеры. Время, необходимое для ручного подсчета того же количества зерен существенно больше. Кроме того, мобильные устройства позволяют обрабатывать серию изображений в фоновом режиме и автоматически сохранять и передавать данные на сервер. Увеличение количества зерен до 100 увеличивает время работы алгоритма на 5-10 секунд. Время, необходимое для оценки длины и ширины 50 зерен под микроскопом, составляет примерно 40-60 минут. Мобильное приложение выполняет этот анализ примерно за одну минуту.

Таким образом, мобильное приложение SeedCounter позволяет проводить крупномасштабные измерения фенотипических параметров зерен пшеницы, таких как длина, ширина, площадь и количество зерен в колосе, как в "полевых условиях", так и в лаборатории.

3.3. Поиск локусов количественных признаков для характеристик размера, формы и цвета зерен мягкой пшеницы.

3.3.1. Растительный материал и условия выращивания

Растительный материал был предоставлен проф. Андреасом Бернером, Институт генетики растений и исследований растениеводства Лейбница (IPK), Гатерслебен, Германия. Семена представляли 114 образцов популяции ITMI/MP мягкой пшеницы (*T. aestivum* L.). Картирующая популяция ITMI была получена путем опыления растений *T. aestivum* (сорт 'Opata 85') пыльцой

синтетической гексаплоидной яровой пшеницы ‘Synthetic’ W7984 (Arif et al., 2021). Растения каждого генотипа были выращены в сезоне 2014 года на экспериментальных полях ИРК в Гатерслебене, Германия.

3.3.2. Протокол фенотипирования зерен

Мы модифицировали протокол съемки зерен для фенотипирования для использования цифровой камеры в лабораторных условиях. Съемка зерен производилась на столе на белом фоне (лист бумаги формата А4). Зерна рассыпались небольшой группой, до 20 штук. В область каждого кадра помещалась калибровочная палитра ColorChecker (x-rite ColorChecker® Classic Mini, <https://xritephoto.com/camera>), которая использовалась при обработке изображений для цветовой коррекции и определения масштаба. Освещение было настроено таким образом, что позволяло избежать теней. RGB изображения колосьев были получены с помощью цифровой фотокамеры Canon EOS 600D, with Canon EF 100mm f/2.8 Macro USM lens в формате JPG или PNG. Пример полученного в результате такого протокола изображения представлен на рис. 3.4.



Рисунок 3.4. Изображение зерен линии ITMI_110, полученное по протоколу съемки цифровой камерой в лабораторных условиях.

Для нашего анализа мы использовали два изображения для каждого генотипа: 15 и 5 семян. Разделение первоначально использовалось для проверки воспроизводимости оцененных значений признаков. Однако между двумя повторностями никаких значимых различий между средними значениями признаков семян не наблюдалось в соответствии с F -тестом. Поэтому мы использовали средние значения характеристик для 20 изображений семян из двух повторов в качестве входных данных для анализа QTL.

3.3.3. Оценка характеристик зерен

Обработка изображений выполнялась с помощью приложения SeedCounter (Komyshev et al., 2017), версии для настольных ПК, дополненной модулем расчета цветовых характеристик, функции которого будут описаны ниже. Приложение реализовано с помощью библиотеки OpenCV (Dawson-Howe, 2014) на языке программирования Java.

Основные этапы обработки изображений включают:

1. распознавание цветовой шкалы ColorChecker, масштабирование и цветокоррекция;
2. цветовая сегментация изображения;
3. выделение замкнутых контуров на изображении;
4. фильтрация полученных контуров, распознавание зерен;
5. извлечение характеристик размера и формы зерен;
6. извлечение цветовых характеристик зерен.

Первый этап был реализован на основе алгоритма, описанного в работе (Genaev et al., 2019).

Цветовая шкала была идентифицирована с использованием ее эталонного изображения для идентификации нескольких дескрипторов путем поиска ключевых точек. Для каждого дескриптора эталонного изображения определялся дескриптор на изображении, ближайший по расстоянию Хэмминга. Области, соответствующие разным цветам палитры, были идентифицированы путем выравнивания калиброванного изображения шкалы с эталонным изображением с использованием алгоритма RANSAC (Kaehler and Bradski, 2016). Определив цветовую шкалу, мы корректируем цвет

изображения с помощью метода, используемого в эпилюминесцентной микроскопии (Quintana et al., 2011). Смещение цветов на изображении зерен относительно эталона оценивалось с использованием параметра D_{col} : чем выше его значение, тем значительнее искажение цвета. Если D_{col} близок к нулю, изображение практически не содержит цветовых искажений. Масштаб изображения (размер пикселя, мм) вычислялся из отношения площади известной цветовой шкалы к ее площади на изображении (с учетом поправки на ее ориентацию). На последующих этапах область цветовой шкалы была исключена из анализа.

Этапы анализа 2-4 описаны в разделе 3.2.2. «Алгоритм обработки изображений для морфометрии зерен». Детали этапов 5 и 6 будут описаны ниже.

Характеристики размера зерен включают в себя длину (sL), ширину (sW) и проецируемую на изображении площадь (sA). Метод оценки этих параметров описан в разделе 3.2.2. «Алгоритм обработки изображений для морфометрии зерен». Дополнительно оценивали округлость (Seed Circularity, sCi), закругленность (Seed Roundness, sRo), шероховатость (Seed Rugosity, sRg), цельность (Seed Solidity, sSo) (Cervantes et al., 2016; Zdilla et al., 2016). Индексы округлости (Circularity и Roundness) отражают насколько форма контура близка к форме окружности, но рассчитываются по-разному.

Округлость - это мера сходства плоской фигуры с окружностью, рассчитываемая по формуле (Cervantes et al., 2016)

$$sCi = \frac{4\pi \times area}{perimeter^2}, \quad (3.1)$$

где *area* и *perimeter* это площадь и периметр контура, соответственно. Т.к. круг, это фигура с наименьшим периметром при заданной площади, фигура отличная от окружности будет иметь индекс округлости меньше единицы.

Для контуров, имеющих много небольших выпуклостей, периметр увеличивается, а индекс округлости принимает более низкие значения. В этих случаях для описания близости формы фигуры к окружности целесообразно использовать индекс Roundness, поскольку эта величина не зависит от неровностей линии контура. Этот индекс рассчитывается как отношение площади фигуры (*area*) к квадрату длины главной оси вращения (Major Axis). Для окружности длина Major axis = $2r$ и индекс вычисляется как (Cervantes et al., 2016)

$$sRo = \frac{4 \times area}{\pi [Major\ axis]^2}. \quad (3.2)$$

Для фигуры отличной от круга знаменатель будет увеличиваться и индекс принимает значения меньше единицы.

Индекс шероховатости (Rugosity, *sRg*) определяется как отношение периметра контура к выпуклому периметру (Cervantes et al., 2016):

$$sRg = \frac{Ps}{Pc}, \quad (3.3)$$

где *Ps* - периметр контура, а *Pc* - выпуклый периметр контура, также известный как выпуклая оболочка, то есть наименьшая выпуклая фигура, которая содержит все точки изображения.

Индекс компактности/целостности (Solidity, *sSo*) - это отношение площади контура к площади его выпуклой оболочки (Zilla et al., 2016):

$$sSo = \frac{\textit{Contour Area}}{\textit{Convex Hull Area}}. \quad (3.4)$$

Для описания цветовых характеристик зерен мы использовали представление цвета в виде четырех моделей: RGB, HSV, Lab, YCrCb (Busin et al., 2008; Gowda and Yuan, 2019; Комышев и соавт., 2020). Каждая из них представляет цвет в виде трех компонент. Значения компонент одного пространства могут быть получены путем преобразований значений компонент другого. Диапазон значений интенсивности каналов составляет от 0 до 255, кроме канала Hue в цветовой модели HSV, диапазон значений которого составил от 0 до 180. Цветовые дескрипторы вычислялись для каждой из этих компонент для каждого зерна в отдельности.

Первый тип дескрипторов: средние значения интенсивности компонент для пикселей зерна. Эти характеристики вычислялись в два этапа. Во-первых, оценки средних значений и стандартных отклонений для компонент проводились по всем пикселям изображений зерен. После этого отбрасывались пиксели, с отклонением интенсивности компонент от среднего более чем на 3 стандартных отклонения. Для оставшихся пикселей вновь вычислялось средние значение компонент, которое использовалось далее. В дальнейшем дескрипторы средних значений компонент будут обозначены малой буквой *m*, например, для цветового пространства RGB это три параметра RGB_mR, RGB_mG, RGB_mB; для остальных пространств обозначения аналогичные.

Следующий набор характеристик - доминантные цвета зерен. Дескрипторы доминантных цветов обеспечивает представление репрезентативных цветов на изображении или его области (Cieplinski, 2001). Для вычисления этих дескрипторов, все пиксели, относящиеся к изображению зерна в цветовом пространстве RGB кластеризуются по интенсивности

компонент методом k -средних где k -число репрезентативных кластеров, которое может быть любым большим 2. В настоящей работе $k=3$. Полученные кластеры $i=1, \dots, k$ ранжируются по доле пикселей p_i , которые к ним относятся. Для каждого кластера определяется центроид c_i , представляющий доминантный цвет. Для каждого такого центроида определялись значения интенсивностей цветовых компонент (например, для пространства RGB это параметры RGB_dCj_i , где j – обозначение цветовой компоненты, i – номер доминантного кластера, например, RGB_dCR_1 обозначает компоненту R для первого доминантного цвета в пространстве RGB). Использование трех доминантных цветов позволяет более точно оценить оттенки окраски зерен.

В итоге, для каждого зерна определялось 3 характеристики размера, 4 характеристики формы и 48 характеристик цвета. Характеристики рассчитывались для каждого зерна 114-х генотипов пшеницы. Затем для каждого генотипа среднее значение использовалось в качестве входных данных для анализа QTL.

3.3.4. Анализ локусов количественных признаков и приоритизация генов в области QTL

Выделение ДНК, генотипирование и выявление QTL было проведено доктором Миан Абдур Рехман Арифом, Институт ядерных исследований для сельского хозяйства и биологии, Файзалабад, Пакистан, детали анализа приведены в работе (Arif et al., 2022). Для исследований использовался чип для пшеницы 20K Infinium SNP array. Для картирования маркеров использовался метод QTL Ici Mapping 4.2.53 (<http://www.isbreeding.net/> (версия от сентября 2019) один из наиболее недавних методов для анализа QTL (Arif et al., 2021). Значения $LOD > 2.0 \leq 3$ считались значимыми для идентификации QTL, значения > 3.0 считались высоко значимыми. С помощью данного метода также

были выявлены эпистатические парные QTL. Обозначение QTL приводится в соответствии с рекомендациями Каталога генных символов (McIntosh et al., 2008). Расположение QTL в геноме пшеницы было визуализировано с помощью пакета “circlize” R (Gu et al., 2014).

Чтобы идентифицировать возможные гены-кандидаты, связанные с признаками семян, мы провели приоритизацию на основе нескольких критериев с использованием аннотации генов, предоставленной в базе данных KEGG (Kanehisa et al., 2017). Анализ включал локусы, для которых значение LOD превышало три, и состоял из нескольких этапов.

На первом этапе анализа мы определили физическую локализацию маркеров, выровняв их последовательности с последовательностью сборки генома пшеницы IWGS 2.1 (Zhu et al., 2021). Данные о последовательности генома и аннотации были получены с сайта URGI (https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v2.1). Мы рассматривали только аннотации генов с высокой степенью достоверности. Последовательности маркеров были получены из работы (Wang et al., 2014) и базы данных маркеров Gramene (<https://archive.gramene.org/markers/>) (Tello-Ruiz et al., 2021). Последовательности маркеров были выровнены на референсный геном с помощью программы blastn пакета BLAST+ (Camacho et al., 2009) с использованием $e\text{-value}=1e-17$ (другие параметры были установлены по умолчанию). Это позволило искать сходные последовательности длиной от 50 нт и более и избежать лишних совпадений. Таким образом, для каждого из выбранных нами QTL мы получили список генов аннотации генома пшеницы IWGS 2.1. Нам не удалось определить физические границы QTL для нескольких локусов, поскольку для одной из последовательностей маркера на

хромосоме, соответствующей маркеру выравнивание не обнаруживалось.

Поскольку растительные пигменты могут синтезироваться в различных тканях и органах, при определении приоритетных генов для QTL цвета семян мы дополнительно фильтровали гены по уровню экспрессии в семенах. Для этой цели мы использовали данные экспрессии, представленные для пшеницы в базе данных expVIP (Borrill et al., 2016). Данные в текстовом формате были загружены с сайта URGI (https://urgi.versailles.inra.fr/download/iwggsc/IWGGSC_RefSeq_Annotations/v1.1/iwggsc_refseqv1.1_rnaseq_mapping_2017July20.zip). Мы использовали данные экспериментов RNA-seq, в которых столбец "High level tissue " содержит значение "seed". Мы считали, что ген экспрессируется в этих экспериментах, если уровень его экспрессии, измеренный в единицах TPM \geq 1. Для выполнения фильтрации мы разработали скрипты на Python, учитывающие преобразование идентификатора гена между аннотациями генома пшеницы версии 2.1 (геном) и 1.2 (транскриптом).

На втором этапе анализа мы сгенерировали список ортологичных групп белков из базы данных KEGG (Kanehisa et al., 2017), которые связаны с формированием признаков размера и цвета семян. Хорошо известно, что цвет семенной оболочки определяется наличием в ней специфических растительных пигментов (Lachman et al., 2017). Поэтому мы выбрали ортогруппы, участвующие в KEGG-путях биосинтеза этих пигментов и ряда их предшественников. Список таких путей включает метаболизм триптофана (map00380), основной путь биосинтеза терпеноидов (map00900), биосинтез каротеноидов (map00906), биосинтез фенилпропаноидов (map00940), биосинтез флавоноидов (map00941), биосинтез антоцианов (map00942), биосинтез изофлавоноидов (map00943), биосинтез флавонов и флавонолов

(map00944). Мы получили 307 ортогрупп KEGG для этих путей.

Размер семян зависит от множества биологических процессов, происходящих на молекулярном уровне, включая убиквитинирование белка, ответ на гормональные сигналы, биосинтез и транспорт белка и т.д. Поэтому было невозможно выделить пути, соответствующие этим процессам, основываясь только на их описании в базе данных KEGG. Однако гены, участвующие в развитии семян, были довольно хорошо изучены экспериментально у *A. thaliana* и риса (*Oryza sativa*). Поэтому мы использовали три недавних обзора литературы, описывающих молекулярные процессы развития семян у арабидопсиса и риса (Chen et al., 2021; Li et al., 2019; Li et al., 2018). Мы объединили набор генов из трех работ и удалили дублированные идентификаторы. Во время компиляции мы преобразовали идентификаторы генов из работы (Chen et al., 2021) из формата RAP в формат MSU с помощью инструмента "ID converter" на веб-сайте базы данных OryzaExpress (http://bioinf.mind.meiji.ac.jp/OryzaExpress/ID_converter.php). Мы идентифицировали ортогруппы KEGG для выбранных генов и использовали их для нашего анализа.

Присвоение ортологичных групп KEGG генам пшеницы по их последовательности было выполнено с использованием инструментов BlastKOALA (Kanehisa, et al., 2016) и KofamKOALA (Aramaki et al., 2020). Ортогруппы были присвоены генам по результатам по крайней мере одной из программ. Мы считали гены в локусах QTL приоритетными если ортогруппы к которым они принадлежали были в списках, связанных с признаками цвета и размера зерна.

3.3.5. Анализ характеристик зерен в популяции ITMI

На рис. 3.5 показано распределение шести из 55 характеристик семян в популяции ITMI. Три из них характеризуют размер (sL , sW , sA), три характеризуют цвет (Lab_mL , Lab_ma , Lab_mb). Распределения имеют колоколообразную форму, гипотеза о нормальности не была отвергнута для характеристик длины и площади семян (тест Шапиро-Уилкса, $p < 0.05$), но отвергнута для характеристик ширины и цвета (рис. 3.5). На основе этого теста для всех характеристик гипотеза о нормальности не была отвергнута для 22 из 55 признаков.

Чтобы визуализировать распределение генотипов в пространстве рассматриваемых признаков, мы провели анализ главных компонент для признаков формы/размера (все 7 признаков) (рис. 3.6), цвета (12 признаков со средними значениями цветовых компонент в четырех цветовых пространствах) семян независимо (рис. 3.7) и всех этих 19 признаков одновременно (рис. 3.8). Из диаграммы главных компонент в пространстве признаков размер/форма видно, что первая компонента характеризует форму семян и коррелирует с округлостью. Вторая компонента характеризует размер семян (в наибольшей степени связан с шириной и площадью). Наиболее характерными генотипами являются ITMI_082 (наиболее округлые семена), ITMI_075 (большая площадь), ITMI_048 (малая площадь), а также ITMI_111 и Synthetic W7984 (наиболее удлинённые). Второй родительский генотип расположен на этой диаграмме в крайнем правом углу диаграммы, близко к оси X, т.е. он имеет округлую форму зерен. На этой диаграмме трудно выделить какие-либо заметные кластеры, это облако с некоторыми отстоящими от него генотипами. Примечательно, что это облако более разрежено в верхней полуплоскости ($PC2 > 0$) и компактное в нижней полуплоскости ($PC2 < 0$).

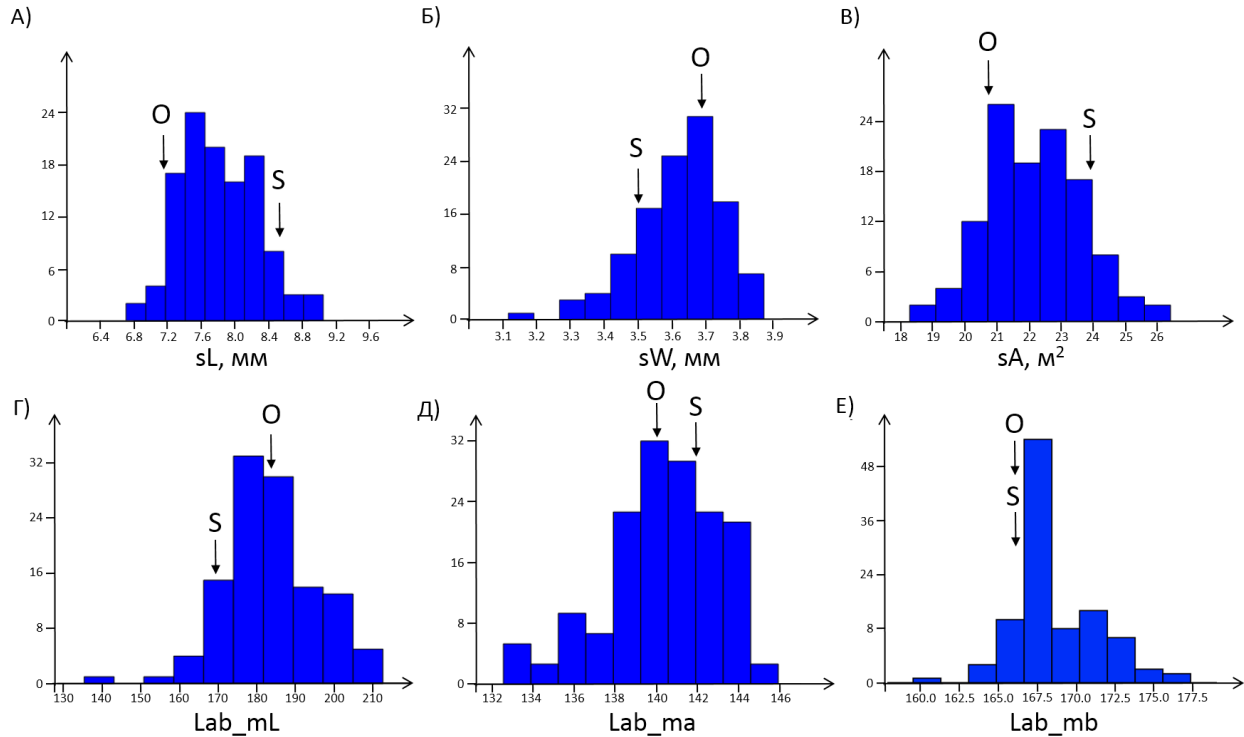


Рисунок 3.5. Распределение шести характеристик размера и цвета зерен пшеницы в популяции ITMI. Ось X – значение характеристики, ось Y – частота встречаемости в выборке. (А) длина зерна, sL; (Б) ширина зерна, sW; (В) площадь проекции на изображении, sA; (Г) интенсивность компоненты L цветового пространства Lab (светлота); (Д) интенсивность компоненты a цветового пространства Lab (краснота); (Е) интенсивность компоненты b цветового пространства Lab (желтизна). Стрелки указывают значение признаков для родительских генотипов Opata (O) и Synthetic (S).

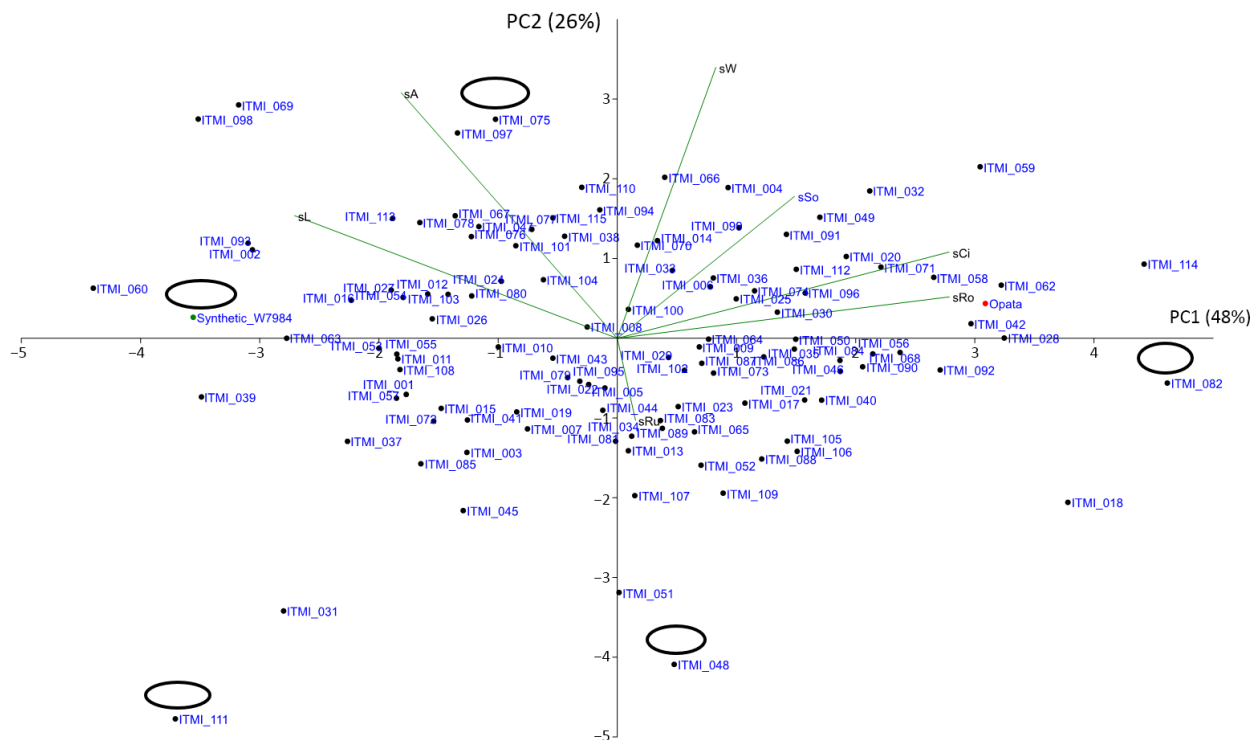


Рисунок 3.6. Диаграмма главных компонент (двойственный график) для характеристик размера и формы семян популяции ITMI. Эллипсы представляют размер и форму семян для некоторых контрастных генотипов в одинаковом масштабе. Родительские генотипы показаны зелеными (Synthetic W7984) и красными (Opata) точками. Оси PC1, PC2 обозначают основные компоненты 1 и 2, процент объясненной дисперсии показан в скобках. Зеленые линии обозначают направление признаков в пространстве главных компонент.

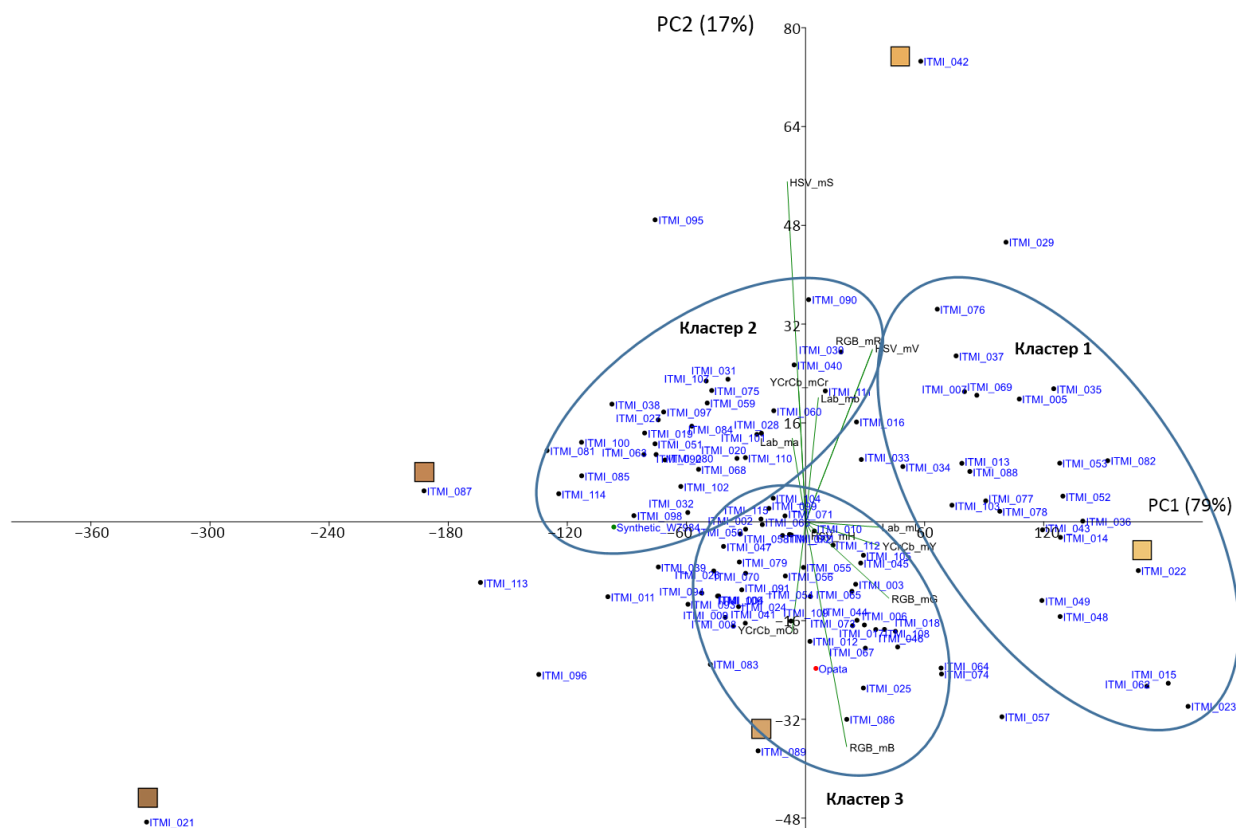


Рисунок 3.7. Диаграмма главных компонент (двойственный график) для характеристик цвета оболочки семян популяции ITMI. Прямоугольники представляют образцы цвета оболочки для нескольких контрастных генотипов. Родительские генотипы показаны зелеными (Synthetic W7984) и красными (Opata) точками. Оси PC1, PC2 обозначают основные компоненты 1 и 2, процент объясненной дисперсии показан в скобках. Зеленые линии обозначают направление признаков в пространстве главных компонент. Эллипсами обведены три кластера генотипов.

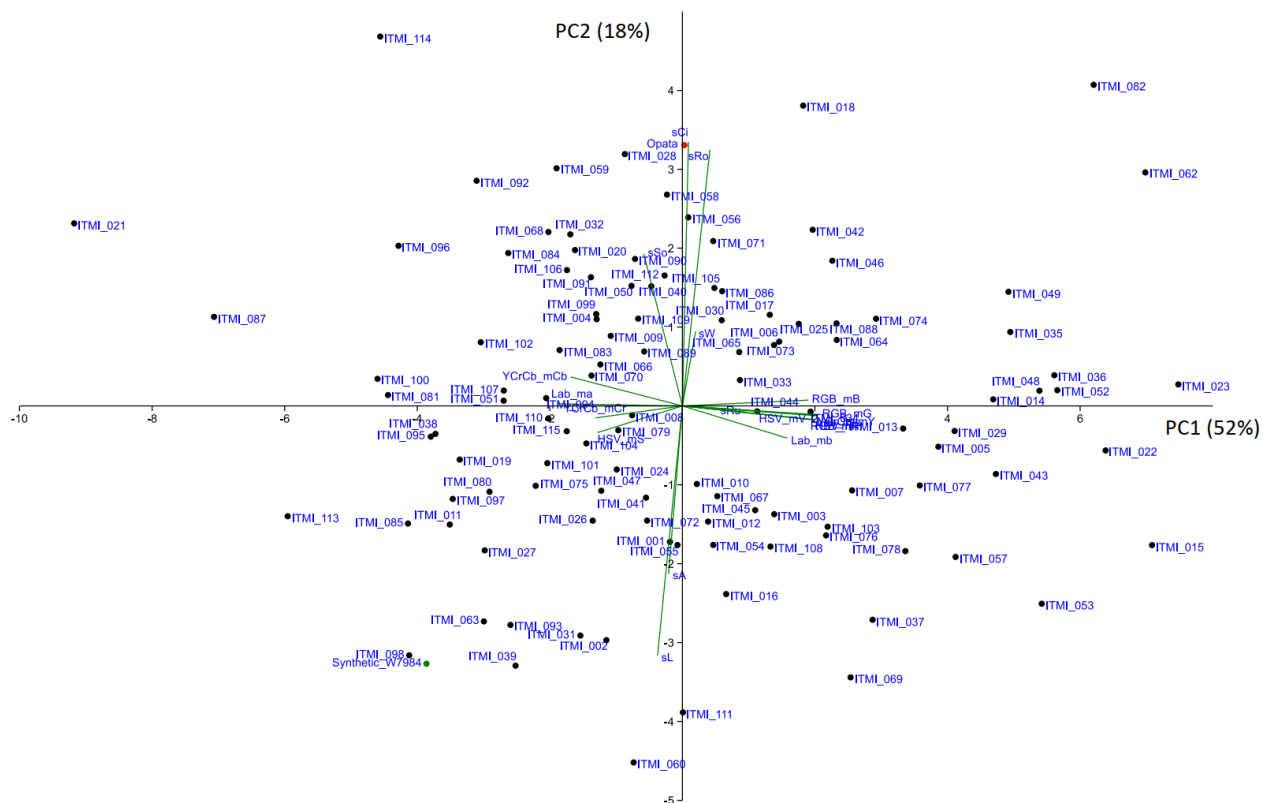


Рисунок 3.8. Диаграмма главных компонент (двойственный график) для 19 характеристик размера/формы и цвета оболочки семян популяции ITMI. Родительские генотипы показаны зелеными (Synthetic W7984) и красными (Opata) точками. Оси PC1, PC2 обозначают основные компоненты 1 и 2, процент объясненной дисперсии показан в скобках. Зеленые линии обозначают направление признаков в пространстве главных компонент.

Диаграмма главных компонент в пространстве цветовых характеристик показывает, что первая компонента в первую очередь характеризует светлоту оболочки (коррелирует с Lab_mL, YCrCb_mY). Вторая компонента характеризует насыщенность цвета семян и их красноватый оттенок (положительно коррелирует с характеристиками HSV_mS, RGB_mR и Lab_ma). Наиболее характерными генотипами являются ITMI_2 (самая светлая оболочка), ITMI_042 (самый насыщенный цвет), ITMI_088 (самый бледный оттенок оболочки) и ITMI_021 и ITMI_087 (самый темный оттенок оболочки).

На графике можно различить три кластера (рис. 3.7). Семена растений первой группы имеют более светлую окраску (большие значения PC1), большая их часть имеет более красноватую окраску (положительные значения PC2). Семена растений второго кластера имеют более темный (отрицательные значения PC1) и более красноватый цвет (положительные значения PC2). Семена растений третьего кластера имеют менее красноватую окраску (положительные значения значений PC2, PC1 разброс вблизи значений 0). Интересно, что родительские генотипы Synthetic_W7984 и Opata попадают в отдельные кластеры на графике (кластер 2 и 3 соответственно).

Диаграмма главных компонент в пространстве 7 признаков размера/формы и 12 признаков цвета (рис. 3.8) показывает, что первая компонента характеризует цвет оболочки семян: отрицательные значения являются характеристиками красноватого цвета (PC1 положительно коррелирует с Lab_mb, отрицательно коррелирует с Lab_ma). Вторая компонента характеризует размер/форму семян (положительно коррелирует с округлостью, sRo и сферичностью, sCi; отрицательно коррелирует с площадью, sA и длиной, sL). На этом графике не обнаружено четких кластеров для генотипов.

3.3.6. Анализ QTL для характеристик зерен в популяции ITMI

Генетический анализ трех характеристик размера семян (sL, sW и sA), четырех характеристик формы (sCi, sRo, sRu и sSo) и 48 характеристик цвета выявил в общей сложности 20, 22 и 170 QTL, соответственно (всего 212) (рис. 3.9), на всех хромосомах генома пшеницы, кроме хромосом 1A и 5D. Количество QTL для одного признака варьировало от одного (характеристики: HSV_mS, HSV_dCH_2, HSV_dCS_2 и HSV_dCV_2) до десяти (характеристика: sA). Для большинства признаков число QTL варьировало от трех (13

признаков) до четырех QTL (19 признаков).

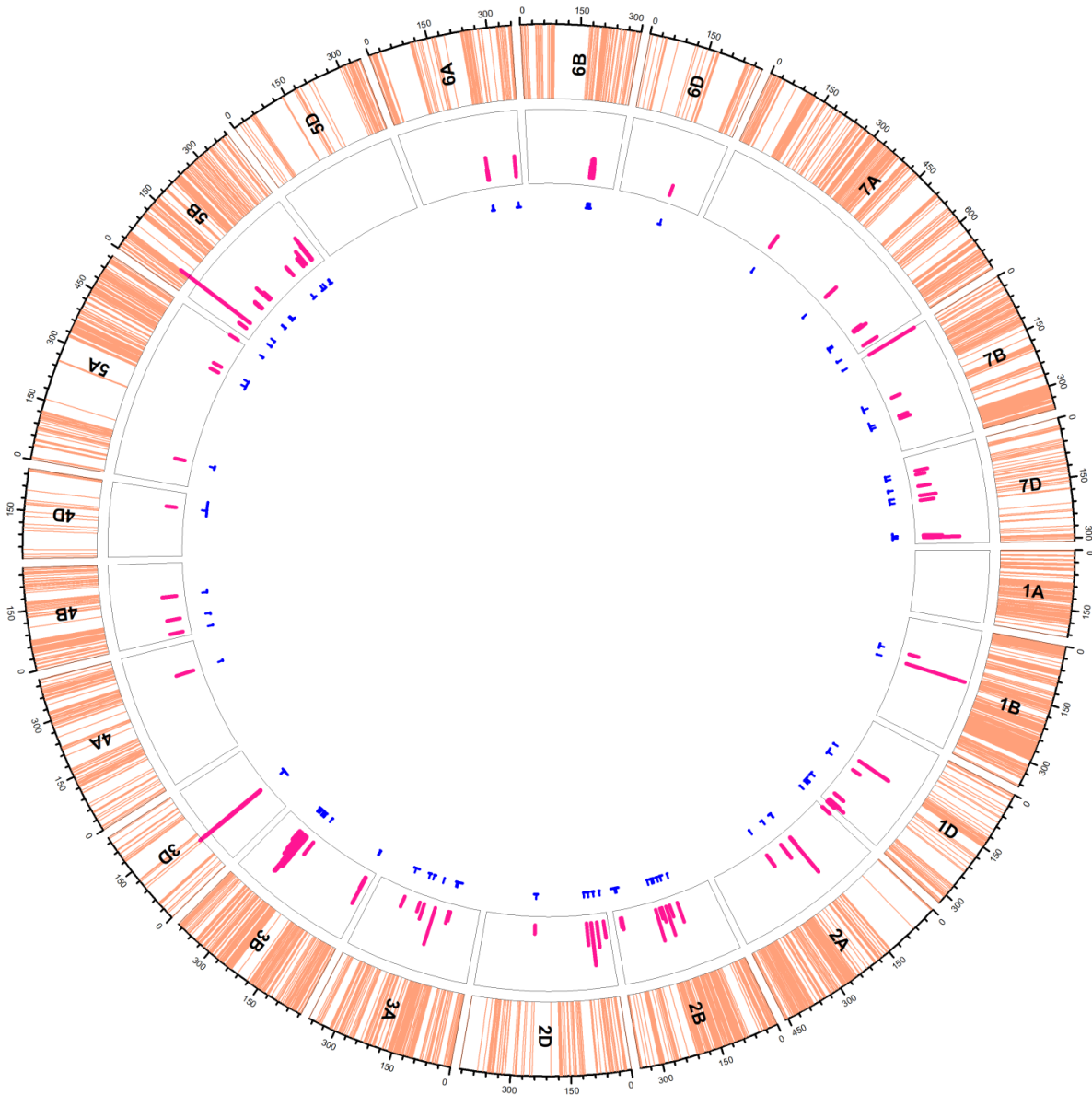


Рисунок 3.9. Распределение аддитивных QTL (синие линии во внутреннем круге) на хромосомах мягкой пшеницы. Светло-оранжевые линии на внешней дорожке указывают позиции SNP на каждой хромосоме; розовые полосы во втором круге указывают значения LOD QTLs. Координаты хромосом даны в сМ. Синие линии под окружностью трека указывают доверительный интервал QTL, а маленькие вертикальные линии указывают на положение пика QTL.

Среди хромосом наибольшее количество QTL наблюдалось на хромосоме 3В (46 QTL), за которой следуют хромосомы 3D и 6В с 34 и 27 QTL соответственно. Хромосома 5В несла 15 QTL, а хромосома 2В несла 14 QTL, тогда как хромосома 7А несла 12 QTL. За этим последовала хромосома 1D с 11 QTLs. Хромосомы 3А и 7D несли по девять QTL каждая, а хромосома 6А несла семь QTL. Хромосомы 6А и 2D обнаружили семь и шесть QTL соответственно. Пять QTL находились на каждой из хромосом 2А и 4В, тогда как четыре QTL находились на каждой из хромосом 5А и 7В. С другой стороны, два QTL были обнаружены на каждой из хромосом 1В и 6D. Наконец, хромосомы 4А и 4D несли по одному QTL каждая. Что касается групп, хромосомы 3-й группы несли наибольшее количество QTL (89), тогда как хромосомы 4-й группы несли наименьшее количество QTL (пять). Хромосомы 6-й группы несли 36 QTL, тогда как каждая из хромосом 2-й и 7-й групп несла по 25 QTL. С другой стороны, хромосомы 5-й группы несли 19 QTL, а хромосомы 1-й группы несли 13 QTL.

Кроме того, мы также смогли обнаружить 34 пары эпистатических QTL, контролирующих по меньшей мере 22 признака в линиях пшеницы. Пять признаков связаны с влиянием более чем одной пары эпистатических QTL (рис. 3.10). Эти QTL включали все хромосомы пшеницы, кроме хромосом 1А, 4А, 4В и 6А. Наиболее часто задействованной хромосомой была 3D (12 раз), за которой следовали хромосомы 3А (11 раз) и 3В (девять раз). Хромосома 2D была задействована шесть раз, в то время как хромосомы 5В и 5D были задействованы по четыре раза каждая. Хромосомы 1D, 2В и 7А были задействованы по три раза каждая. Двукратное вовлечение наблюдалось для хромосом 1В, 2А, 4D, 6В и 6D, тогда как хромосомы 5А, 7В и 7D были вовлечены только один раз.

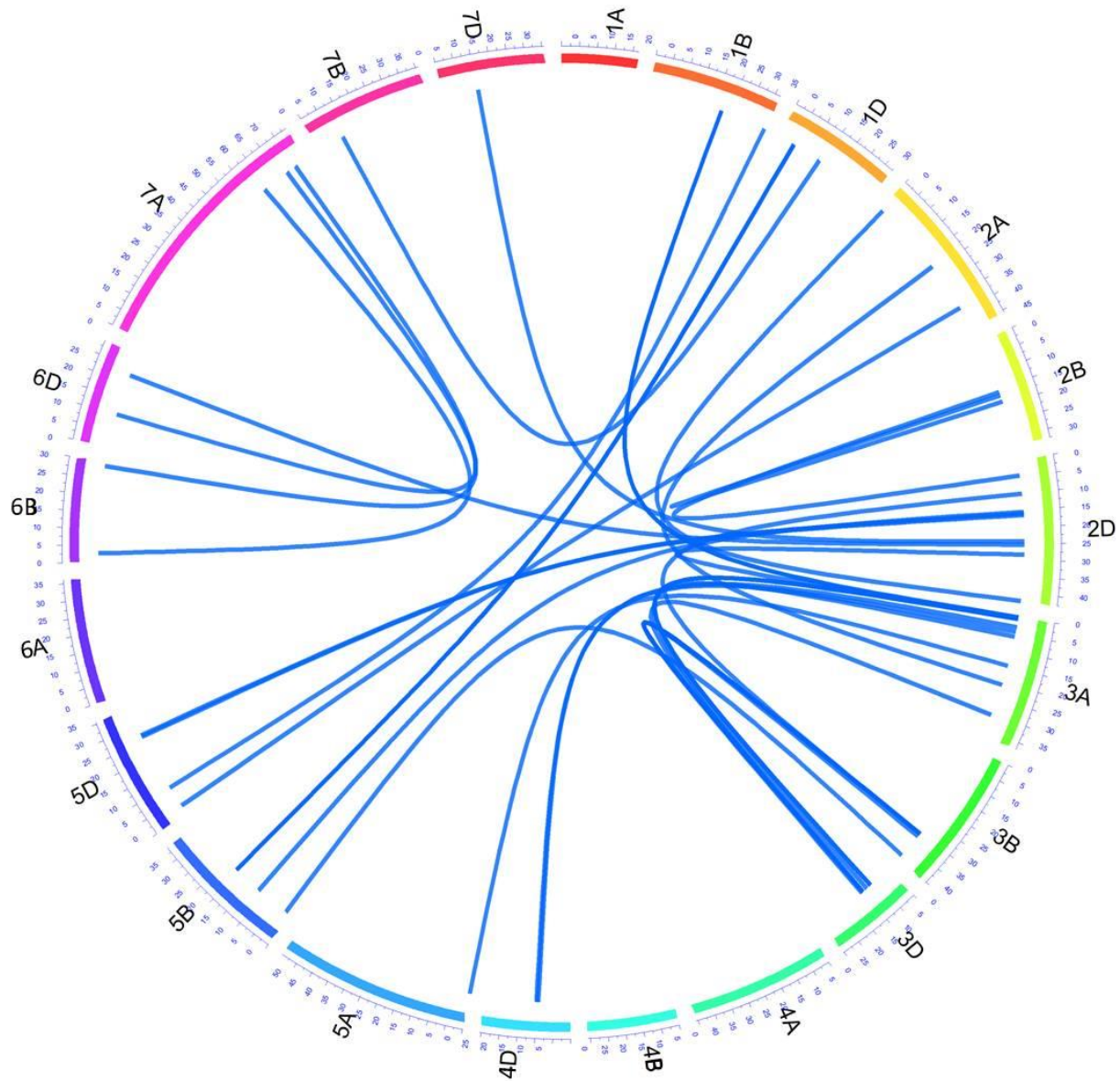


Рисунок 3.10. Набор эпистатических пар QTL в популяции ITMI. Внешний круговой график представляет координаты генома мягкой пшеницы, расположенные на хромосомах 1-21 (1A-7D) по часовой стрелке. Цифры на цветном внешнем круге представляют координаты в сМ соответствующих хромосомам. Синие дуги представляют пары эпистатических QTL, контролирующие различные признаки.

Мы наблюдали заметное совпадение между местоположениями QTL для

разных признаков. Например, два QTL, связанных с формой семени, *Q.sCi-2B^c* (округлость) и *Q.sSo-2B^c* (целостность), расположены в одной позиции 129 сМ хромосомы 2В. Хромосома 3В имеет локусы с множественными ассоциациями признаков: позиция 39.179 сМ (2 признака размера), позиция 298.179 сМ (7 признаков цвета), позиция 299.179 сМ (12 признаков цвета), позиция 300.179 сМ (2 признака цвета), позиция 306.179 сМ (10 признаков цвета), позиция 308.179 (два цветовых признака), позиция 311.179 сМ (2 цветовых признака и 1 признак формы, шероховатость), позиция 324.179 сМ (3 цветовых признака). Это неудивительно, потому что наши параметры, оцененные по изображениям, представляют собой различные количественные оценки одних и тех же биологических свойств семян (например, массы семян, концентрации пигмента в оболочке и т.д.). Это говорит о том, что набор наших характеристик на самом деле является вырожденным, и многие из них, по сути, контролируются одними и теми же генами.

Чтобы оценить сходство различных анализируемых признаков в рамках подобной вырожденности, когда несколько признаков контролируются одним локусом, мы сгруппировали признаки иерархически по степени перекрытия местоположений QTL (рис. 3.11). Древоподобная диаграмма демонстрирует несколько интересных особенностей. Во-первых, характеристики размера/формы (правая часть дерева) четко отделены от цветовых признаков (за исключением шероховатости, sRu). Во-вторых, некоторые признаки с небольшим числом QTL (от 1 до 2) также отделены от других признаков. В-третьих, значительное количество признаков, связанных с желтизной, образуют большую группу. Наконец, признаки, связанные со светлотой (Lab_mL, HSV_mV, YCrCb_mY), попадают в один и тот же кластер, и их QTL сильно перекрываются. Другие цветовые признаки неравномерно распределены по дереву, в частности образуя большой кластер цветовых

признаков.

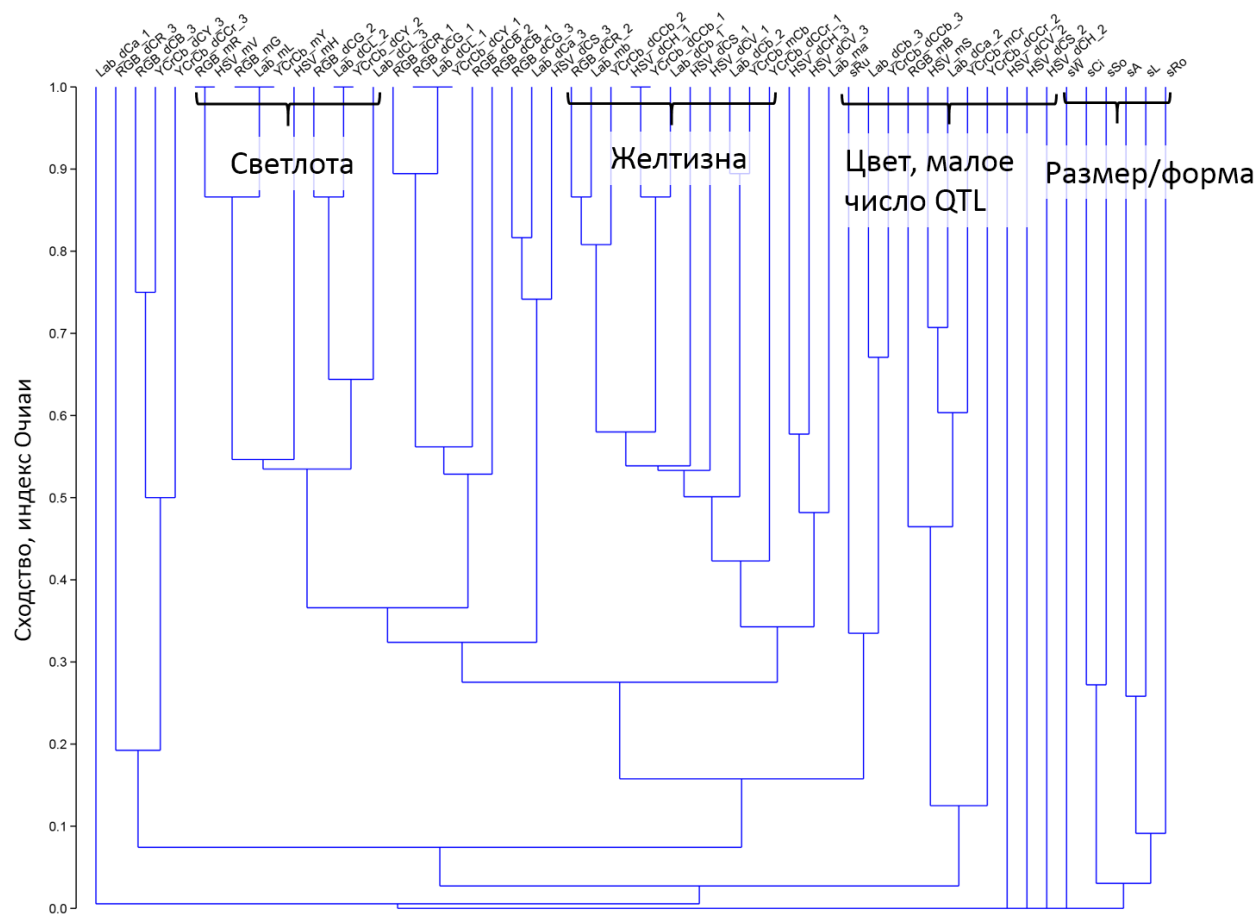


Рисунок 3.11. Дерево сходства для признаков семян, полученных по степени перекрытия между их местоположениями QTL. Вертикальная ось представляет меру сходства, основанную на индексе Очиаи (ось Y). Листья соответствуют признакам семян. Группы признаков с сильным перекрытием местоположений QTL показаны фигурными скобками.

Из наших результатов (рис. 3.11) видно, что многие локусы цветковых признаков перекрываются друг с другом, но не с QTL размера/формы. Исключением является шероховатость (Ru), которая отражает шероховатость оболочки и, вероятно, связана с цветовыми искажениями на границе зерен на фоне изображения. С одной стороны, эта диаграмма отражает вырожденность оцениваемых признаков, что указывает на их избыточность. С другой стороны,

расположение нескольких QTL, связанных с разными цветовыми характеристиками зерен, в одном и том же участке генома может указывать на более надежную идентификацию ассоциации данного участка генома с признаками цвета. Использование множества цифровых представлений одного и того же признака на первый взгляд выглядит излишним и сбивающим с толку. Разумным шагом было бы выбрать одну или несколько числовых характеристик, которые наиболее эффективны при идентификации QTLs. Однако, различное числовое представление одного и того же биологического признака полезно и позволяет оценить его детали. С нашей точки зрения, перекрывание множества QTL для различных признаков в одном и том же положении генома может свидетельствовать о важности этого участка генома.

Наш анализ позволил идентифицировать ряд QTL, связанных с характеристиками семян пшеницы: формой, размером и цветом оболочки. Подобные исследования давно и интенсивно проводятся на основе анализа как QTL, так и полногеномного анализа ассоциаций (Alemu et al., 2020; Williams and Sorrells, 2014; Ren et al., 2021; Arif et al., 2021). В работе (Williams and Sorrells, 2014;) исследовались трехмерные характеристики размера и формы семян на основе анализа изображений семян, полученных в двух проекциях, и использования дескрипторов на основе анализа Фурье с использованием двух популяций, для одной из которых, SynOpDH, родительские генотипы совпадают с популяцией ITMI. Авторы этой работы обнаружили QTL, который влияет на ряд характеристик формы. Для длины зерен было обнаружено восемь QTL на хромосомах 2A, 2D, 4B, 5A, 5B, 6A, 7A, 7D. В нашей работе мы обнаружили меньшее количество QTL для этого признака, расположенных на хромосомах 2A, 2D, 3B, 5B (2 QTL) и 7D. Для ширины зерен Виллиамс и Соррелс обнаружили три QTL на хромосомах 2A, 5A и 6A, тогда как мы обнаружили четыре QTL для этого признака, расположенных на хромосомах

1D, 2D, 3B, 4D. Эти различия могут быть связаны с использованием данных из нескольких разных локализаций выращивания растений за несколько лет.

Ариф и соавт. (Arif et al., 2021) ранее проанализировали 92 образца из популяции ITMI по большому количеству признаков, включая такие признаки, как вес тысячи зерен (TKW) и цвет зерна (КС) в разных регионах произрастания пшеницы и разных сезонах выращивания. Мы не обнаружили никакого совпадения QTL для веса зерен TKW из работы (Arif et al., 2021) с признаками, характеризующими размер и форму семян в нашей работе. Что касается цвета семян, то Ариф и соавт. обнаружили 15 QTL. Сравнение наших результатов с результатами этой работы показало, что из 15 QTL один точно соответствовал QTL, который был найден в нашей работе. Это *Q.KC_Pu07-3B* (Arif et al., 2021), ограниченный маркерами AX-94979462, IAAV6088 и расположенный в нашей работе на хромосоме 3B в позиции 306.179. В нашей работе несколько QTL, связанных с характеристиками цвета оболочки зерен, соответствуют этому локусу, например, QTL *Q.HSV_dCH_1-3B^m*. Также интересно отметить QTL *Q.KC_Mo07-3D* и *Q.KC_Mo08-3D* (Arif et al., 2021), ограниченные маркерами D_GDS7LZN02IJRXZ_309 слева и CAP12_c2615_128 справа, расположенными на хромосоме 3D в положении 76 cM. В нашей работе мы обнаружили серию QTL, связанных с цветом, локализованных на хромосоме 3D в положении 100-102, ограниченных маркерами CAP12_c2615_128 слева и BS00067163_51 справа. Таким образом, QTL из нашей работы и работы Ариф и соавт. (Arif et al., 2021) находятся в непосредственной близости. Для других QTL, связанных с цветом, мы не обнаружили совпадений. Например, Ариф и соавт. идентифицировали пять QTL, связанных с цветом оболочки на хромосоме 5A. Однако в нашей работе только два локуса в других положениях этой хромосомы были связаны с цветом.

Среди локусов, связанных с цветом, участок на хромосоме 3D (положение 100 сМ), ограниченный маркерами CAP12_c2615_128 и BS00067163_51, пожалуй, является наиболее интересным. Как указано выше, он расположен рядом с QTL, связанным с цветом оболочки, определенным в работе Ариф и соавт. (Arif et al., 2021). В нашей работе с этим положением на хромосоме связаны 34 различных признака, характеризующих цвет семян. Все они являются цветовыми параметрами в различных цветовых пространствах. Мы локализовали физические координаты этого участка на хромосоме 3D: 573.6-580.8 млн. п.н. в соответствии с аннотацией генома IWGS v2.0. Интересно, что Ланг и соавт. (Lang et al., 2021) недавно провели анализ QTL *PHS-3D*, связанного с устойчивостью семян к предуборочному прорастанию, для популяции синтетической гексаплоидной пшеницы. Оказалось, что на физической карте эта область расположена на хромосоме 3D в позициях 571.9-574.3 млн. п.н., что совпадает с физической локализацией QTL, которую мы идентифицировали. Авторы также показали, что генотипы растений, восприимчивые к предуборочному прорастанию, характеризуются делецией ~2,4 млн. п.н., включающей 20 генов в этой области генома. Оказалось, что в этой области расположен ген, кодирующий транскрипционный фактор *Myb10-D*, который придает устойчивость к предуборочному прорастанию за счет активации путей биосинтеза флавоноидов и абсцизовой кислоты. Интересно, что растения, которые не содержат делеций в этой области и устойчивы к предуборочному прорастанию, имеют красновато-коричневую окраску семенной оболочки (Lang et al., 2021).

3.3.7. Приоритизация генов из районов QTL для характеристик зерен

Поиск ортологичных групп для восьми путей биосинтеза пигментов и их предшественников выявил 307 ортологов KEGG, участвующих в этих процессах. Обзор литературы выявил 155 генов арабидопсиса и 42 гена риса, участвующих в молекулярных процессах развития семян. Было обнаружено, что 193 из этих генов имеют идентификаторы последовательностей в базе данных KEGG, и 109 из них были связаны с ортологичными группами KEGG.

Для определения приоритетных генов мы использовали 48 значимых QTL с $LOD > 3$, для которых были идентифицированы позиции маркеров в геномной последовательности пшеницы. На этой основе мы идентифицировали 2787 уникальных генов, локализованных в участках, ограниченных маркерами. Из них 1422 гена были связаны с размером/формой семян и 1365 генов связаны с цветом семян. После фильтрации по уровню экспрессии осталось 823 гена, связанных с цветом семян. Для этих наборов генов мы выполнили поиск их принадлежности к ортогруппам KEGG с использованием сервисов BlastKOALA и KofamKOALA. Для 464 генов, связанных с локусами признаков размера, и 321 гена, связанных с признаками цвета, были обнаружены такие ортогруппы.

Для 29 генов из локусов, связанных с размером семян, мы обнаружили совпадение в списке ортогрупп, полученном на основе анализа литературных данных (Приложение, табл. П4). 11 генов, идентифицированных таким образом, принадлежат регуляторным белкам (факторы транскрипции EREBP, HD-ZIP, MYBP; локусы на хромосомах 3A, 2B, 2D, 7D). Они могут быть вовлечены в регуляцию роста и развития семян. В частности, известно, что факторы транскрипции, связанные с реакцией на этилен (EREBP), участвуют в определении размера семян, массы семян и накопления жиров и белка у *A.*

thaliana (Jofuku et al., 2005). В работе (Li et al., 2016) идентифицировала два фактора транскрипции из семейства AP2/EREBP, TaPARG, расположенные на 2A и 2D хромосомах пшеницы, которые регулируют несколько признаков, связанных с урожайностью, включая массу семян.

6 генов относятся к факторам инициации трансляции (ELF2C; локусы на хромосомах 2B, 7D).

5 генов относятся к ферментам, связанным с процессами убиквитинирования (локусы на хромосомах 2D). Это убиквитин-протеинлигаза E3, убиквитинтиоэстераза белка OTUB1. Известно, что убиквитины и родственные им ферменты играют важную роль в развитии семян, контролируя пролиферацию клеток (Li and Li, 2014). Например, гены семейства лигаз E3 участвуют в биосинтезе амилозы в пшенице (Parveen et al., 2021); показано, что ген TaGW2-6A из этого семейства контролирует размер семян (Lv et al., 2022).

4 гена обладают хитиназной активностью (локус на хромосоме 7D). Это ферменты, участвующие в защите семян от патогенов, таких как бактерии или грибы (Gomez et al., 2002). Однако это не единственная их роль в формировании и функционировании семян. Из протеомных данных, полученных для риса было показано, что хитиназа 14 взаимодействует с геном GW2 (убиквитинлигазы кольцевого типа E3). Также было показано, что GW2 контролирует размер семян посредством регуляции уровней или активности хитиназы 14 и фосфолицераткиназы (Lee et al., 2018).

Также обнаружено 2 гена с цитокиндегидрогеназной активностью (хромосомы 3A и 7D) и один ген киназы, содержащий домен aarF (хромосома 7D). Они также могут быть связаны с развитием семян пшеницы (Chen et al., 2021).

Для генов из локусов, связанных с цветом семенной оболочки, 14 обнаружили совпадение с ортогруппами метаболических путей базы данных KEGG, связанных с биосинтезом пигментов (Приложение, табл. П5). Восемь генов были вовлечены в биосинтез фенилпропаноидов (локусы на хромосомах 3А, 3В, 6А и 6В), два гена были вовлечены в путь биосинтеза каротиноидов (локусы на хромосомах 2А и 6А), по одному гену каждый был вовлечен в биосинтез флавонов и флавонолов, биосинтез флавоноидов, метаболизм триптофана и биосинтез терпеноидов.

На хромосоме 2А мы обнаружили несколько генов, которые участвуют в биосинтезе растительных пигментов. Среди них один ген, обозначенный как *NCED*, участвует в пути биосинтеза каротиноидов. У риса мутации этого гена приводят к изменению окраски семян околоплодника (Amer Hamzah et al., 2021). Экспрессия этого фермента контролируется абсцизовой кислотой (Sano and Marion-Poll, 2021), а *NCED* также участвует в биосинтезе АВА (Lang et al., 2021). Функции этого гена в развитии семян хорошо известны (Matilla et al., 2015): он является важным регулятором развития семян, зиготического эмбриогенеза и покоя. Другим геном, связанным с биосинтезом каротиноидов, который мы обнаружили среди приоритетных генов, является *CYP707A* ((+) - 8'-гидроксилаза абсцизовой кислоты) (Приложение, табл. П5). Его функции тесно связаны с геном *NCED*; было показано его участие в тех же процессах, связанных с развитием семян (Matilla et al., 2015). Интересно, что два этих гена расположены вблизи локусов, связанных с покраснением семенной оболочки (*Q.YCrCb_dCCr_1-2A³* и *Q.HSV_dCH_3-6A⁴/Q.Lab_ma-6A⁴*).

Мы обнаружили два гена, которые участвуют в пути биосинтеза флавоноидов (Приложение, табл. П5), которые обеспечивают различную окраску семян злаков (Shoeva et al., 2016; Khlestkina, 2013). К ним относятся

гомолог FG2 (флавонол-3-О-глюкозид L-рамнозилтрансфераза), мутации которого приводят к фенотипу с изменением цвета семян у сои (Rojas Rodas et al., 2014). Было показано, что экспрессия шикимат-О-гидроксициннамоилтрансферазы повышена в растениях пшеницы с высокой антиоксидантной активностью семян (Ma et al., 2022).

В области QTL *Q.Lab_dCb_3-3A*ⁱ мы обнаружили два гена, участвующих в биосинтезе фенилпропаноидов. Они оба кодируют 4-кумарат-КоА-лигазу. Этот фермент катализирует превращение *p*-кумаровой кислоты в *p*-кумароил-КоА, который в дальнейшем служит источником биосинтеза как лигнина (структурного компонента оболочки семян), так и флавоноидов. В исследованиях транскриптомной ассоциации у *Brassica napus* было показано, что экспрессия 4CL во время развития семян положительно коррелирует с содержанием семенной оболочки (долей массы семян, приходящейся на оболочку) (Zhang et al., 2022). Интересно, что при анализе экспрессии генов у растений *B. napus* с коричневой окраской семян уровень экспрессии генов, кодирующих этот фермент, был выше, чем у растений с желтой окраской (Hong et al., 2017).

Таким образом анализ позволил нам идентифицировать ряд генов-кандидатов, связанных с размером/формой семян и их цветом, на основе биоинформатического анализа и аннотации генов в соответствии с литературными данными и базой данных KEGG. Мы идентифицировали восемь локусов, связанных в геноме с признаками размера/формы семян, для которых мы обнаружили 28 ортологичных генов, вовлеченных в генные сети, контролирующие эти признаки.

3.4. Изучение взаимосвязи между характеристиками размера, формы и цвета зерен мягкой пшеницы, временем их хранения и всхожестью

Большинство работ по изучению старения зерна основаны на искусственном кратковременном воздействии таких факторов, как высокая температура и влажность (Delouche and Baskin, 1973). Это позволяет быстро активировать метаболические процессы в зерне, аналогичные старению (Dell'Aquila, 2009; Dell'Aquila, 2006; Смоликова, 2014). Изучение старения семян в естественных условиях позволяет исследовать процессы, которые происходят в реальных условиях содержания, например, в генбанках, но эти исследования затруднены из-за необходимости длительного хранения зерна. Насколько нам известно, до сих пор не проводилось систематического исследования количественных характеристик семян, таких как размер/форма /цвет, для семян, хранящихся в течение нескольких лет в условиях генбанка.

В этом разделе диссертационной работы мы анализируем морфологические характеристики зерен и цвет их оболочки для 44 рекомбинантных инбредных линий мягкой пшеницы популяции ITMI, которые были выращены в четыре разных сезона: 2003, 2004, 2009 и 2014 гг. Для этих образцов мы исследуем связь между характеристиками зерен (размер, форма и цвет оболочки) с годом урожая (фактически, с длительностью хранения в генбанке). Кроме того, для 19 образцов, выращенных в те же сезоны, оценивали всхожесть и сравнили связь этого параметра с характеристиками зерен.

3.4.1. Биологический материал

Мы изучали 44 образца из популяции рекомбинантных инбредных линий (РИЛ) картирующей популяции мягкой пшеницы (*Triticum aestivum* L.) Международной Инициативы по картированию пшеницы (ITMI), описанной выше в разделе 3.3. Растения каждого генотипа были выращены в 2003, 2004, 2009 и 2014 гг.

Для 19 генотипов (ITMI_001, ITMI_006, ITMI_009, ITMI_010, ITMI_016, ITMI_020, ITMI_024, ITMI_025, ITMI_026, ITMI_030, ITMI_032, ITMI_045, ITMI_055, ITMI_082, ITMI_084, ITMI_088, ITMI_105, ITMI_112, ITMI_114) проф. А. Бернером были получены данные по всхожести зерен в 2003, 2004, 2009 и 2014 г. (Afonnikov et al., 2022).

3.4.2. Протокол съемки зерен и обработка изображений и статистический анализ

Зерна фенотипированы в 2020 г. Съемка зерен производилась как описано в разделе 3.3.2. «Протокол фенотипирования зерен», по 20 штук для каждого года и генотипа.

Обработка изображений выполнялась с помощью приложения SeedCounter (Komyshev et al., 2017), версии для настольных ПК, см. раздел 3.3.3. «Оценка характеристик зерен».

В итоге, для каждого зерна определялось 3 характеристики размера, 4 характеристики формы и 48 характеристик цвета. Характеристики рассчитывались для каждого зерна 44-х генотипов пшеницы, отдельно для каждого из 4-х лет урожая.

Для кластеризации признаков и анализа главных компонент мы

проводили нормализацию и стандартизацию данных по строкам (зерна) и столбцам (признаки). На основе преобразованных данных реконструировалось дерево близости признаков методом UPGMA с использованием Евклидова расстояния. Эти же данные были использованы для анализа главных компонент, который проводился отдельно для признаков размера/формы зерен и цвета. Анализ выполнялся с использованием языка программирования Python, библиотеки Scipy (модули `scipy.cluster.hierarchy` и `scipy.spatial.distance`) для кластеризации и Sklearn (модули `sklearn.preprocessing` и `sklearn.decomposition`) для анализа главных компонент. Визуализация выполнялась с помощью библиотеки Matplotlib.

Для проверки гипотезы о связи между годом урожая (длительностью хранения семян) и характеристиками размера/формы и цвета зерен мы использовали несколько статистических подходов, которые опишем ниже. Мы использовали в качестве независимой переменной год урожая, а в качестве зависимых переменных – характеристики зерен.

Во-первых, для оценки влияния длительности хранения зерен на их характеристики мы проводили дисперсионный анализ (независимая переменная – год урожая). Во-вторых, для выявления трендов между годом урожая и значением признака использовали коэффициент корреляции Пирсона. Однако, поскольку годы урожая (2003, 2004, 2009 и 2014) разделены неравномерными временными интервалами, мы применили три варианта кодировки независимой переменной. Первый вариант кодировки (Year01) приписывал для годов 2003 и 2004 значение 0, а для 2009 и 2014 года значение 1. Эта кодировка позволяет разделить годы урожая на ранние (2003, 2004) и более поздние (2009, 2014). Второй вариант кодировки (YearRank) приписывал разным годам ранг в порядке возрастания (соответственно, 1 для 2003, 2 для

2004, 3 для 2009 и 4 для 2014). Такой способ нивелировал разницу временных интервалов между годами сбора урожая. Третий вариант кодировки (Year) использовал значение года урожая. Поскольку дисперсионный анализ показал, что вклад генетической компоненты в дисперсию характеристик зерен оказывается значимым, чтобы устранить этот эффект мы из значений переменных вычитали среднее по соответствующему генотипу. Для оценки значимости коэффициента корреляции мы использовали бутстреп анализ и перестановочный тест, в каждом случае использовали по 2000 реплик и вычисляли коэффициент корреляции Пирсона с параметром YearRank. Для всех случайных выборок определяли максимальное значение коэффициента корреляции и минимальное значение. Связь между признаком и YearRank считалась значимой, если реальный коэффициент корреляции был либо меньше минимального, либо больше максимального значения в обоих тестах. Обработка проводилась в пакете JACOBI4 (Polunin et al., 2019).

Оценка статистической связи между всхожестью семян различных генотипов и их характеристиками проводилась на основе линейных коэффициентов корреляции Пирсона. Предварительно мы оценили вклад в дисперсию значений всхожести факторов года урожая и генотипа. Поскольку год урожая имел значимый вклад в дисперсию всхожести, мы из значений всхожести для каждого генотипа и года вычитали среднее по соответствующему году урожая. Анализ корреляций между признаками зерен и всхожестью в дальнейшем проводился аналогично тому, как это было сделано для года урожая.

3.4.3. Анализ влияния генотипа и времени хранения на признаки зерен

Результаты кластеризации 55 характеристик зерен приведены на рис. 3.12.

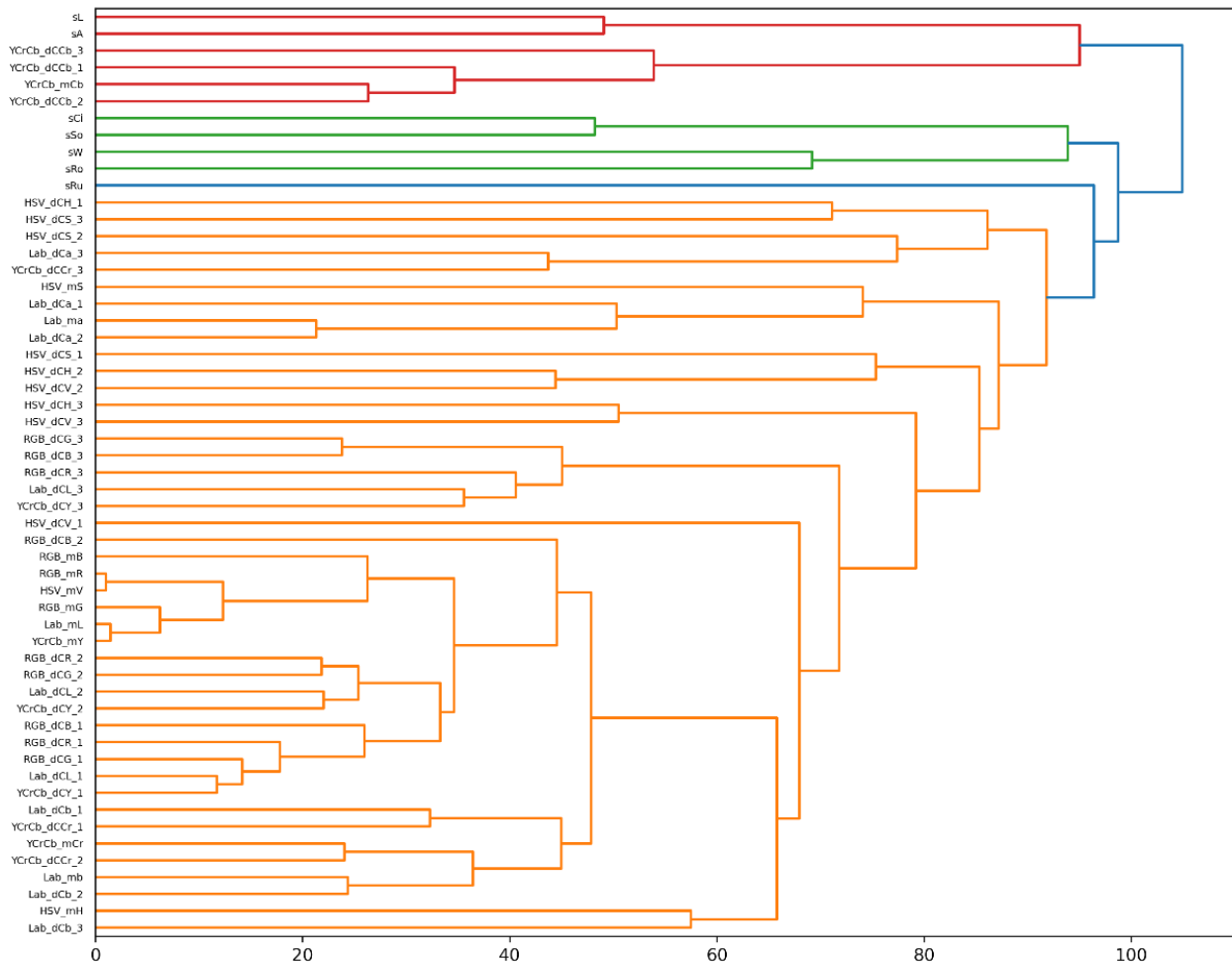


Рисунок 3.12. Кластеризация характеристик размера/формы и цвета зерен пшеницы в популяции ITMI. Разными цветами выделены 4 характерных кластера признаков.

На диаграмме рис. 3.12 можно выделить несколько характерных кластеров признаков. Первый, самый большой, показан оранжевым цветом, включает исключительно характеристики цвета зерен. При этом в этом

кластере можно выделить несколько подкластеров. Прежде всего это группа признаков связанная со средними значениями R,G,B компонент. К ней близки ряд характеристик других цветовых пространств, отражающих, преимущественно, средние значения яркости цвета или светлоты (YCrCb_mY, HSV_mV, Lab_mL). Близко к ним расположены и соответствующие компоненты первых и вторых доминантных цветов (RGB_dCB_2, RGB_dCR_2, RGB_dCG_2, Lab_dCL_2, YCrCb_dCY_2, RGB_dCB_1, RGB_dCR_1, RGB_dCG_1, Lab_dCL_1, YCrCb_dCY_1). Признаки этого кластера, следовательно, можно отнести к характеристикам яркости цвета зерен.

Рядом располагается группа параметров, характеризующих оттенки цвета b для шкалы Lab и Cr для шкалы YCrCb: Lab_mb, Lab_dCb_1, Lab_dCb_2, YCrCb_mCr, YCrCb_dCr_1, YCrCb_dCr_2. Эти параметры характеризуют изменение оттенков цвета от синего до желтого (b), либо от красного до зеленого (Cr). Рядом с этим кластером располагаются среднее значение Hue шкалы HSV (HSV_mH) и Lab_dCb_3.

Большой набор параметров не формирующих выраженной группировки включает компоненты для второго и третьего доминантных цветов (RGB_dCG_3, HSV_dCH_3 и т.п.). Анализ средних значений цветовых компонент доминантных цветов показал, что для пространств, в которых одна из компонент связана с яркостью (V для HSV, L для Lab, Y для YCrCb) три доминантных цвета зерен преимущественно отличаются именно по этой компоненте (различия составляют десятки единиц), в то время как для остальных двух компонент, характеризующих оттенки цвета, эти различия составляют, несколько единиц (до десятка). Таким образом можно сделать вывод, что три доминантных цвета отражают, в большей степени, различия в освещенности разных сторон зерен, в меньшей – неоднородность их окраски.

К неструктурированной группе признаков также относятся средние значения компонент а шкалы Lab и S шкалы HSV (Lab_ma, HSV_mS).

Группировка указанных выше параметров на дереве хорошо согласуется со статистической зависимостью между ними: коэффициенты корреляции между значениями компонент RGB и характеристиками яркости YCrCb_mY, HSV_mV, Lab_mL превышают 0.8. Для параметров Lab_ma, HSV_mH коэффициенты корреляции наоборот, меньше -0.8. Для этих значений коэффициентов корреляции p -value близка к 0.

В отдельную группу попадают характеристики формы (sCi, sSo, sW, sRo), характеризующие округлость/вытянутость семян. Значения шероховатости (sRu) попадают в большой кластер с характеристиками цвета зерен. По-видимому, это может быть связано с тем, что неровная граница зерен (высокое значение sRu) приводит к размыванию цвета близких к ней пикселей, что вносит свой вклад в значения цветовых компонент.

Интерес представляет кластер, включающий значения площади и длины зерен, характеризующих размер. Оба этих параметра тесно связаны (коэффициент корреляции Пирсона составляет 0.81, $p=0$). Однако в этот кластер попадают и значения цветовой компоненты Сb шкалы YCrCb, как среднее, YCrCb_mCb, так и значения трех доминантных цветов (YCrCb_dCb_1, YCrCb_dCb_2, YCrCb_dCb_3). Эту связь отражают и коэффициенты корреляции Пирсона: их значения между sL и указанными параметрами цвета составляют, соответственно, 0.37, 0.38, 0.37 и 0.27.

Полученные данные свидетельствуют, что цветовые характеристики зерен формируют одну группу признаков, характеристики размера/формы – другую. Тем не менее, между отдельными признаками из этих групп существуют значимые статистические связи (например, между размером и

компонентой С_b пространства YCrCb).

Для выявления значимости влияния генотипа и времени хранения зерен на характеристики размера/формы и цвета зерен был проведен дисперсионный анализ. Влияние двух факторов (генотип и год урожая) рассматривались независимо, для значений p -value мы делали поправку Бонферрони на число анализируемых признаков (55), так что $p_{\text{corr}} = 0.05/55=0.0009$. Анализ показал, что генотип оказывает значимое влияние на подавляющее большинство анализируемых признаков, для 53 из 55 признаков за исключением sRu (rugosity) и HSV_dCH_2. Для второго фактора, года урожая, число таких признаков составило 34, а для 21 из них $p > p_{\text{corr}}$. К признакам размера/форм зерен, для которых влияние года урожая не значимо, относятся sA (площадь), sRu (шероховатость), sW (ширина). К признакам цвета, для которых влияние года урожая не значимо, относятся компоненты разных цветовых шкал в основном для 2 и 3 доминантных цветов.

Чтобы определить наличие линейной корреляции между признаками зерен и годом урожая, мы оценили коэффициенты корреляции Пирсона между этими величинами. Мы предположили, что систематический тренд, связанный с годом урожая может отражать изменение характеристик зерен, связанных с длительностью их хранения. Мы использовали 4 вида кодировки года урожая как это описано в материалах и методах, а перед анализом корреляций мы вычли из значений признаков их средние по генотипам. Результаты приведены в Приложении, табл. Пб.

Прежде всего обращает на себя внимание то, что коэффициенты корреляции между признаками зерен и годом урожая, представленном в различных кодировках являются близкими, разница наблюдается, преимущественно, во втором знаке. Так, для признака Lab_ma (среднее

значение цветовой компоненты а шкалы Lab) и кодировки Year01 коэффициент корреляции равен -0.69, для кодировки Year = -0.66, для кодировки YearRank -0.64. Этот признак имеет одни из наибольших по модулю и отрицательных коэффициентов корреляции. Для признака sL (длина зерна) все оценки коэффициентов корреляции с годом урожая низки по абсолютному значению: для кодировки Year01 -0.00, для кодировки Year 0.06, для кодировки YearRank 0.02. Это свидетельствует об устойчивости наших оценок коэффициентов корреляции на основе трех разных кодировок по отношению к значению года урожая. При этом необходимо отметить, что наиболее низкие по модулю оказались коэффициенты корреляции с такой кодировкой года, как YearRank. Для 41 из 55 признаков значение корреляции с другими признаками у такой кодировки было меньше по модулю, чем у какой-либо другой кодировки (для кодировки Year01 это число составило 8, для кодировки Year – 0). Именно поэтому мы выбрали для рандомизационных тестов проверку корреляций признаков с кодировкой года урожая YearRank.

Результаты рандомизационных тестов показали, что для 45 из 55 признаков зависимость между YearRank и признаками зерен является значимой как для бутстреп, так и для пермутационного теста. Например, для указанного выше параметра Lab_ma коэффициент корреляции в реальной выборке составил -0.64, минимальные значения в экспериментах пермутации/бутстреп составили -0.07 и -0.06, соответственно, максимальные 0.06 и 0.06. Таким образом, для этого признака выявленные зависимости высоко значимы. Для 10 признаков это условие оказалось не выполнено (связь не значима). Например, для признака sL (реальное значение коэффициента корреляции 0.02) значения в экспериментах пермутации/бутстреп составили - -0.06 и - -0.05, соответственно, максимальные – 0.06 и 0.06.

Большинство характеристик цвета зерен, 44 из 48, демонстрируют значимую связь с годом урожая. Среди признаков размера/формы только один, шероховатость, демонстрирует значимую связь с годом урожая. Для R_u коэффициент корреляции с YearRank составил в реальной выборке 0.08, максимальные и минимальные значения для пермутаций составили, соответственно, -0.05 и 0.06, бутстрепа - -0.05 и 0.05. Мы видим, однако, что в сравнении с таким параметром, как Lab_{ma}, абсолютное значение реального коэффициента корреляции лишь ненамного превосходит максимальное значение в рандомизированных выборках. Это согласуется и с тем фактом, что дисперсионный анализ показал отсутствие значимой связи между годом урожая и этим признаком (p -value 0.21).

Следует отметить, что наибольшие по абсолютному значению отрицательные значения коэффициентов корреляции (все меньше или равные -0.6) наблюдаются для признаков, связанных с компонентой a пространства Lab, которая характеризует красный оттенок цвета. Это признаки Lab_{ma}, Lab_{dCa_1}, Lab_{dCa_2} (среднее значение и значение для доминантных компонент). Чем больше год, тем меньше значение компоненты a . Это означает, что зерна ранних годов урожая имеют более красный оттенок оболочки, по сравнению с более поздними годами урожая. Несколько меньшие по модулю отрицательные значения коэффициентов корреляции наблюдаются для признаков компоненты Cr шкалы YCrCb, которая является красной цветоразностной компонентой этого пространства и также связана с красным оттенком (Busin et al., 2008; Gowda and Yuan, 2019; Комышев и соавт., 2020): чем больше год, тем меньше значение компоненты Cr. Также следует отметить, отрицательные значения коэффициентов корреляции между годом урожая и компонентами яркости HSV_{mV} (для YearRank $r = -0.24$), YCrCb_{mY} (для YearRank $r = -0.15$), Lab_{mL} (для YearRank $r = -0.15$). Это свидетельствует, что

чем меньше год урожая, тем более светлым является цвет оболочки зерен. Всего мы наблюдаем 35 признаков (все относятся к характеристикам цвета), для которых связь с годом урожая является отрицательной.

Значимые положительные коэффициенты корреляции между годом урожая и признаками зерен были обнаружены для 10 признаков, лишь один из них – признак формы *rugosity* (как было отмечено выше). Остальные – цвета. Наибольший по модулю положительный коэффициент корреляции (0.49) наблюдается для признака *HSV_mH*, который отражает среднее значение цветовой компоненты *hue* шкалы *HSV*. Эта компонента связана с оттенком цвета. Для более ранних лет урожая она меньше, для поздних больше.

Для 10 признаков коэффициент корреляции с годом урожая близок к 0 и не значим, согласно рандомизационным тестам. Из них 6 признаков – характеристики размера/формы, 4 – цвета (различные компоненты доминантных цветов).

Интересно, что признаки с незначимой согласно ANOVA связью с годом урожая имеют и более низкие по абсолютному значению коэффициенты корреляции (Приложение, табл. Пб). Так для признаков, у которых коэффициент корреляции с *YearRank* менее -0.31 или более 0.18 ANOVA демонстрирует только значимую связь с годом урожая. В то же время, существуют признаки, для которых анализ ANOVA демонстрирует значимую связь с годом урожая, а коэффициенты корреляции при этом низкие. К ним относятся такие признаки размера/формы как *sCi*, *sSo*, *sL*, *sRo*. Можно предположить, что при наличии достоверной зависимости их значений от года урожая, между годом урожая и значением признака отсутствует тренд. Ряд признаков не демонстрируют достоверной связи с годом урожая согласно ANOVA с поправкой Бонферрони, однако для некоторых из них мы наблюдаем

значимые коэффициенты корреляции с YearRank. Если рассмотреть наибольшие из них по модулю, это признаки RGB_dCR_2 ($r=-0.31$), HSV_dCV_1 ($r=-0.28$), HSV_dCV_2 ($r=-0.22$), то для них результаты ANOVA являются значимыми без учета поправки Бонферрони (p -values, соответственно 0.001, 0.001 и 0.015). В целом, результаты ANOVA и проведенного корреляционного анализа находятся в хорошем согласии, а для признаков, для которых и результаты ANOVA и корреляционного анализа значимы можно предполагать наличие тренда, связанным с годом урожая.

На рис. 3.13 приведены интервальные диаграммы распределений значений признаков Lab_ma, HSV_mV (значимая положительная связь с годом урожая) HSV_mH (значимая отрицательная связь с годом урожая) и sL (нет значимых корреляций с годом урожая) для всех генотипов в зависимости от года урожая.

Из рис. 3.13 видно, что для признаков Lab_ma, HSV_mV более ранние годы урожая соответствуют более высоким значениям признаков, а для признака HSV_mH наоборот, более низким. В целом тренд систематического уменьшения (признаки Lab_ma, HSV_mV) или увеличения (признак HSV_mH) хорошо заметны на этих рисунках.

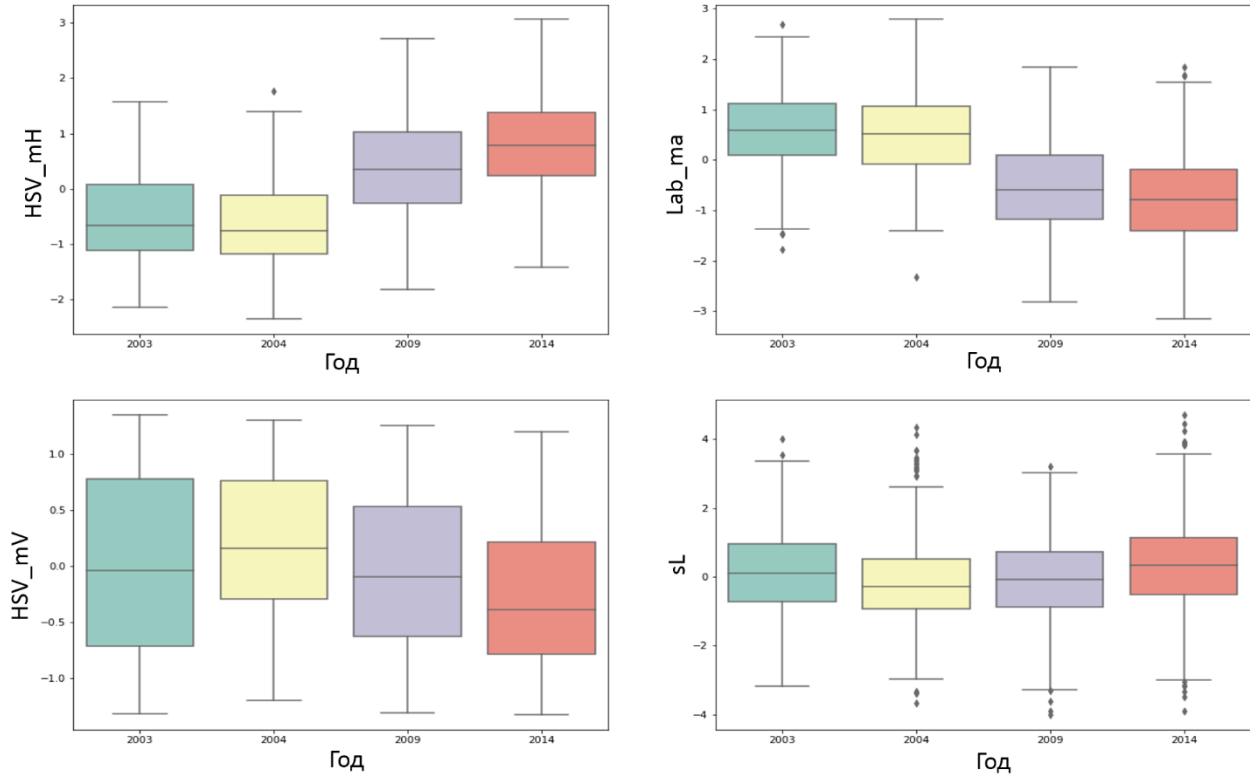


Рисунок 3.13. Интервальные диаграммы распределений нормализованных параметров HSV_mH, Lab_ma, HSV_mV и sL (оси Y) в образцах зерен, собранных в сезоны разных лет (ось X).

Для признака sL (длина зерна) отличия в длине зерен для разных годов заметны, однако какого-либо систематического тренда по годам не наблюдается. Это отражает тот факт, что согласно тесту ANOVA для этого признака наблюдается значимая зависимость от года урожая, однако коэффициент корреляции между ним и годом не является значимым.

3.4.4. Анализ связи между всхожестью и признаками зерен

Для того, чтобы исследовать зависимости между всхожестью зерен и их признаками мы вначале с помощью метода ANOVA оценили их зависимость от двух факторов: генотипа и года урожая. Оказалось, что для генотипа такая связь

оказалась не значимой (p -value=0.98), а для года урожая – значимой (p -value=1.2·10⁻¹⁵). Действительно, для 2003 и 2004 годов средние значения всхожести составили 92.0 и 98.6%, для 2009 и 2014 годов 64.9 и 88.2%, соответственно (см. рис. 3.14А). Особенно выделяются данные по всхожести различных образцов для 2009 года: они систематически ниже остальных и имеют существенно больший разброс. Для устранения эффекта зависимости от года урожая у всех значений всхожести мы вычли соответствующее среднее для года урожая.

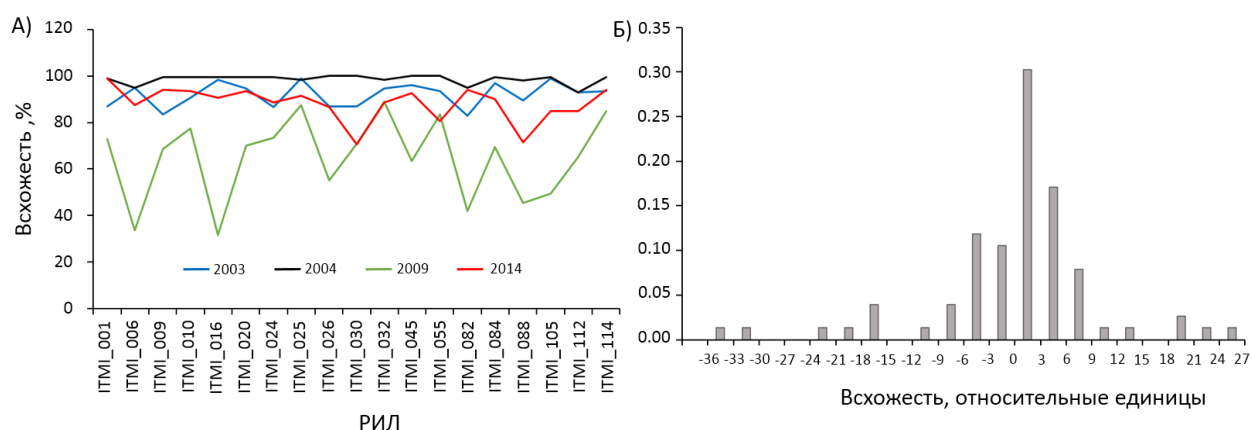


Рисунок 3.14. Всхожесть растений популяции ITMI для 19 генотипов, оцененные в 2003, 2004, 2009 и 2014 г. (А) Значения всхожести (ось Y) для генотипов (ось X) в разные годы показаны ломаными линиями разного цвета (см. легенду внизу графика); (Б) Распределение значений всхожести после вычитания среднегодовых значений.

Распределение значений всхожести после вычитания среднегодовых значений показано на рис. 3.14 Б. Это распределение разбивается на центральную часть (в пределах значений от -15 до 15) и отстоящие пики (за пределами этого интервала). Интересно, что далеко отстоящие пики соответствуют в основном данным 2009 года (ITMI_006, ITMI_016, ITMI_082, ITMI_088, ITMI_105 меньше -15 и ITMI_032, ITMI_055, ITMI_114 больше 15).

Два значения меньше -15 наблюдаются и для 2014 г (ITMI_030 и ITMI_088). Поскольку выбросы могут оказывать влияние на оценки коэффициентов корреляции, мы решили исключить данные по этим генотипам и годам и рассматривать только данные по всхожести, которые соответствуют центральной области распределения.

Связь между всхожестью и значением признака зерна оценивали на основе коэффициента корреляции Пирсона. Для оценки его значимости, как и для года урожая провели два рандомизационных теста: бутстреп и пермутации. Результаты для всех признаков приведены в Приложении, табл. П7.

Значимая положительная связь была обнаружена только между всхожестью и средним значением hue (шкала HSV). Отрицательные значимые коэффициенты корреляции были выявлены для ряда признаков, которые связаны с красным оттенком зерна. Их можно разделить на две группы: для первой группы значения коэффициентов корреляции слабо отличаются от пороговых. Это признаки, связанные с яркостью по шкале HSV (HSV_mV и HSV_dCV_1) и красной компонентой шкалы RGB (RGB_dCR_1, RGB_dCR_2). Вторая группа – признаки, коэффициенты корреляции которых с всхожестью более чем в 2 раза отличаются от порога. Это признаки, связанные с красной компонентой цвета зерен цветовых пространств YCrCb и Lab: YCrCb_dCCr_1, YCrCb_dCCr_2, YCrCb_mCr, Lab_dCa_1, Lab_dCa_2, Lab_ma.

Из таблицы видно, что связь между всхожестью зерен и цветовой компонентой Lab_ma их оболочки наиболее выражена и отрицательна. Это означает, что зерна с меньшими значениями всхожести имеют более красный оттенок оболочки. Это же подтверждается и негативной зависимостью всхожести от компоненты Cr шкалы YCrCb. учетом негативной зависимости всхожести от компоненты HSV_mV, можно сделать вывод, что они являются

еще и более темными.

Диаграммы рассеяния для всхожести и трех признаков, Lab_ma, YCrCb_mCr и HSV_mH представлены на рис. 3.15. Видно, что для меньших значений всхожести значения Lab_ma, YCrCb_mCr выше, а значения HSV_mH ниже, чем для больших значений всхожести.

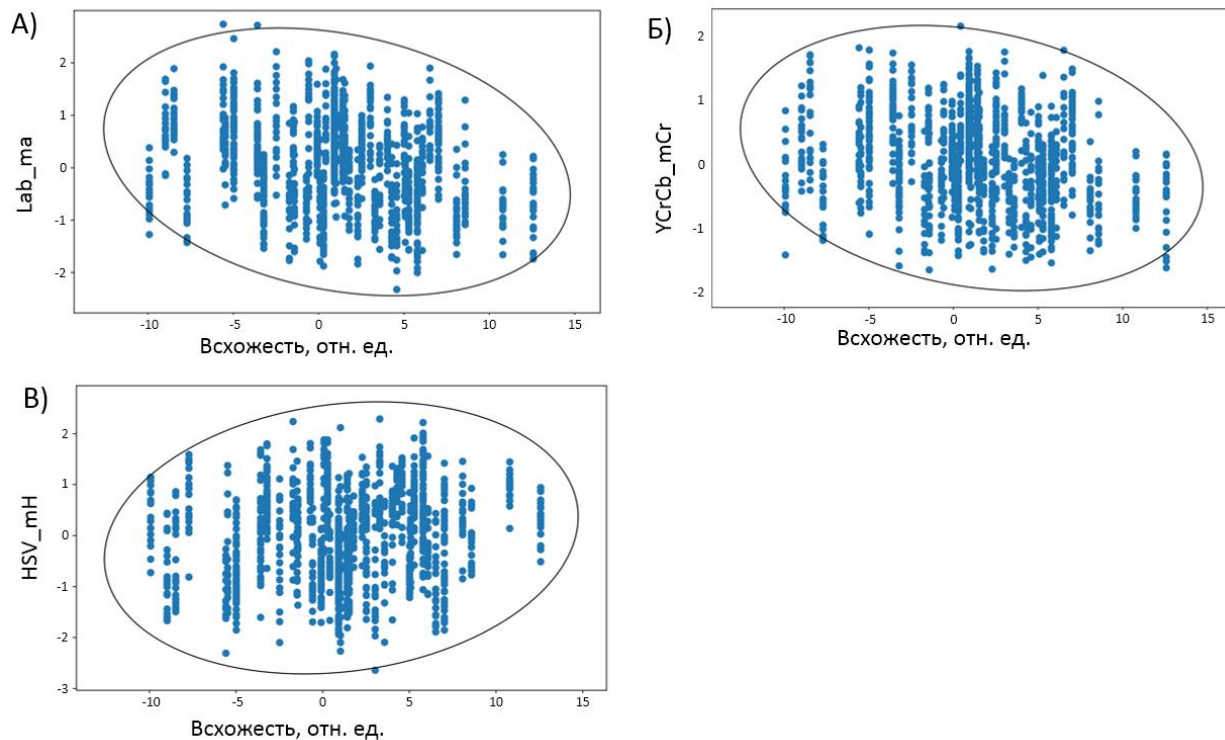


Рисунок 3.15. Диаграммы рассеяния для значений всхожести (оси X) и признаков цвета зерен (оси Y): (A) Lab_ma, (B) YCrCb_mCr, (C) HSV_mH. Показаны эллипсоиды рассеяния при уровне значимости 95%.

Отметим, что ни для одного из признаков формы/размера зерен значимой связи с всхожестью обнаружено не было (Приложение, табл. П7). Наибольший по абсолютному значению коэффициент корреляции наблюдается для площади

зерна sA (0.067), однако он оказывается незначимым.

Фактор года урожая является комплексным и может включать как изменчивость свойств зерен в результате воздействия природных факторов на растения (условия внешней среды в зависимости от года выращивания), так и результат длительности хранения зерен. Если первый вклад можно рассматривать как случайный фактор (условия внешней среды меняются год от года произвольным образом), то второй вклад, принимая во внимание идентичность условий хранения зерен, можно считать детерминированным: время хранения зерен постоянно увеличивается. Таким образом в результате хранения зерен можно ожидать систематическое изменение признаков, уменьшение, либо увеличение.

Для размера/формы зерен наблюдаем значимые результаты теста дисперсионного анализа, но не корреляций (рис. 3.15 Г). Эти результаты показывают, что свойства размера/формы зерен в большей степени подвержены воздействию условий выращивания растений, что хорошо согласуется с существующим представлениями о влиянии условий выращивания на характеристики урожайности (Fenner, 1992; Nasehzadeh and Ellis, 2017; Yang and Zhang, 2006). При этом процесс хранения не оказывает существенного воздействия на эти характеристики.

Для большинства же характеристик цвета мы наблюдаем одновременно значимые результаты для дисперсионного анализа и корреляционного анализа с годом урожая. Следовательно, эти результаты отражают зависимость цвета оболочки зерен прежде всего от времени хранения: чем дольше время хранения, тем зерна становятся более светлыми и более красноватыми.

Цвет зерен может изменяться под действием различных факторов. Воздействие высоких температур и высыхание в условиях микроволновой печи

(Zhao et al., 2017; Zhao et al., 2020; Bai et al., 2018) или нагревание (Chung et al., 2013) приводят к потемнению зерен и их покраснению. Этот эффект известен как *browning* (Wettlaufer and Leopold, 1991) и ассоциирован с поранением семян (Seo et al., 2020). *Browning effect* наблюдался в условиях контролируемой *deterioration* семян in lentil (Dell'Aquila, 2006) and cucumber (Dell'Aquila, A. 2009). Как и в указанных выше работах мы показали, что у пшеницы с увеличением времени хранения увеличивается краснота оболочки зерен. Однако, в отличие от *browning*, мы наблюдали значимый эффект осветления, хотя и достаточно слабый: с увеличением длительности хранения компоненты цвета, связанные с яркостью/светлотой оболочки зерен, увеличивались. Возможная причина этого – более мягкие условия хранения при нормальных температурах, в отличие от условий искусственного старения, с использованием экстремальных воздействий температуры/влажности. Полученные нами результаты демонстрируют важность анализа зерен, которые были подвергнуты именно естественному старению.

Наши данные находятся в хорошем согласии с результатами оценки изменения для морфологических свойств зерен пшеницы эйнкорн, которые хранились в естественных условиях при различных температурах в течение 360 дней (Kibar and Kılıç, 2020). В этой работе показано, что характеристики цвета зерен, включая светлоту и красноту, положительно коррелируют с временем хранения, что хорошо согласуется с нашими результатами. Мы не наблюдали значимых линейных корреляций между формой/размером зерен и длительностью их хранения, в то время как в работе Кибар (Kibar, 2016) для зерен пшеницы и кукурузы была выявленная отрицательная корреляция между физическим размером зерна и временем хранения в промежутке до 90 дней: с увеличением времени хранения размеры зерен уменьшались. Следует, однако отметить, что в эксперименте (Kibar, 2016) зерна пшеницы были собраны в

одно и то же время и относились к одному генотипу, в то время как в нашем эксперименте мы изучали зерна разного года урожая. Наши результаты можно объяснить тем, что фактор условий выращивания вносит более существенный вклад в характеристики размера зерен, чем длительность их хранения.

Наши данные также показали, что компонента красного цвета оболочки зерен демонстрирует отрицательную связь с всхожестью зерен. Ряд экспериментов по оценке влияния цвета семян на их всхожесть демонстрирует аналогичные результаты. Так, в работе (Janampa-Santome et al., 2020) показано что темно окрашенные семена культуры маки перуанской из высокогорных Анд *Lepidium meyenii* (Маса) имеют меньшую всхожесть по сравнению с более светлыми. Статистически значимый эффект отрицательной взаимосвязи между плотностью окраски семян без оболочки и процентом прорастания наблюдался в эксперименте по ускоренному старению семян *M. sativa* и *O. viciifolia* (Behtari et al., 2014).

Этот результат можно объяснить физиологическими свойствами семян и биохимическим составом их оболочки, что в итоге определяет их всхожесть (Mohamed-Yasseen et al., 1994; Debeaujon et al., 2000). Задержка прорастания была связана с присутствием конденсированных и окисленных дубильных веществ (коричневый пигмент), которые могут делать оболочку семян более плотной и менее водопроницаемой. Цвет семян также может быть связан с изменением других вторичных метаболитов, которые могут влиять на их способность к прорастанию (Kuchlan et al., 2010). Отметим, что аналогичные молекулярные механизмы играют важную роль и в устойчивости зерен к преждевременному прорастанию: хорошо известно, что красные зерна более устойчивы к этому, чем белые (Groos et al., 2002; Imtiaz et al., 2008).

3.5. Заключение по главе 3

Методы обработки изображений для морфометрии и классификации семян внедряются с 1980-х годов. Обновления этих методов появляются постоянно, в том числе в последние годы. Тем не менее, по-прежнему существует необходимость в фенотипировании семян с использованием простых и недорогих инструментов. Они могут быть эффективно реализованы с высокой пропускной способностью. Несмотря на простоту, они достаточно эффективны для идентификации QTL, связанных с морфологией и размером семян. Мобильные устройства являются ценными инструментами в этом отношении. Они предоставляют исследователю все необходимое для простого фенотипирования, включая цифровую камеру, мощный процессор и доступ в Интернет. Они могут применяться вдали от лаборатории, но при этом обеспечивают достаточную точность оценки фенотипических параметров. Мобильные устройства также удобны для нового типа фенотипирования растений на основе "краудсорсинга".

Мобильные приложения могут значительно ускорить процесс подсчета количества зерен пшеницы в колосе. Время, необходимое для вычисления примерно 50 зерен с помощью мобильного устройства, составляет примерно несколько секунд, в зависимости от мобильного устройства и разрешения камеры. Время, необходимое для ручного подсчета того же количества зерен, больше, кроме того, мобильные устройства позволяют обрабатывать серию изображений в фоновом режиме и автоматически сохранять и передавать данные на сервер. Увеличение количества зерен до 100 увеличивает время выполнения алгоритма на 5-10 секунд. Время, необходимое для оценки длины и ширины 50 зерен под микроскопом, составляет примерно 40-60 минут. Мобильное приложение выполняет этот анализ примерно за одну минуту.

В нашей работе мы создали метод высокопроизводительного фенотипирования зерен пшеницы. Оно реализовано как в виде мобильного приложения SeedCounter на платформе Android, так и для ПК. Для фенотипирования зерен с помощью анализа цифровых изображений разработаны протоколы, которые могут быть реализованы как в полевых (мобильные устройства), так и лабораторных (цифровые фотокамеры) условиях.

Мы разработали усовершенствование программы SeedCounter для ПК за счет фенотипирования цветковых характеристик оболочки зерен. В совокупности, разработанные технологии позволяют оценивать характеристики формы, размера и цвета оболочки зерен с высокой точностью.

С помощью этих технологий проведен поиск QTL по семи признакам размера/формы семян и 48 признакам окраски оболочки, оцененным на основе анализа цифровых изображений в картирующей популяции рекомбинантных инбредных линий мягкой пшеницы ITMI. Анализ выявил 212 аддитивных и 34 пары эпистатических QTL на всех хромосомах генома пшеницы, кроме хромосом 1A и 5D. Мы продемонстрировали, что один локус может контролировать десятки характеристик семян (преимущественно, связанных с окраской оболочки). Анализ перекрытия локусов показал, что это типично для цветковых признаков и редко встречается для признаков размера/формы семян. Для ряда наиболее значимых QTL мы определили физическое расположение их маркеров на хромосомах пшеницы. Используя литературу и данные KEGG, мы идентифицировали наборы генов у арабидопсиса и риса из сетей, контролирующих размер и цвет семян. Эта информация вместе с координатами маркеров в геноме пшеницы использовалась для определения приоритетов генов пшеницы в регионах QTL. Мы идентифицировали 29 генов-кандидатов

из локусов, связанных с размером семян, и 14 генов из локусов, связанных с цветом семенной оболочки. Гены, которые мы идентифицировали как возможные кандидаты, ассоциированные с размером/формой семян и формированием цвета в популяции ITMI, могут быть дополнительно исследованы более детально с использованием генетических и молекулярных методов для установления механизмов, контролирующих эти важные признаки. Наши результаты продемонстрировали сложный характер генетического контроля признаков семян пшеницы и эффективность методов анализа изображений для получения новых QTL для характеристик семян.

Анализ размера, формы и цвета зерен 44 RILs из популяции ITMI разных годов урожая позволил нам охарактеризовать эти образцы и показать, что генотип значимо влияет на изменчивость всех признаков зерен, а год урожая примерно на три четвертых из них (включая все признаки формы и размера). При этом оценка линейной корреляции между признаками размера/формы зерен показала отсутствие такой зависимости, что в совокупности с результатами ANOVA позволяет сделать вывод, что на эти признаки существенное влияние оказывают условия выращивания растений, а не длительность хранения семян.

Для большинства признаков цвета наблюдаются значимые зависимости с годом урожая, что демонстрирует тренд в их изменении, связанный с годом урожая и, соответственно, с длительностью хранения. Для большинства признаков цвета эти зависимости отрицательные. Наиболее сильная отрицательная связь наблюдается для признаков, связанных с красным оттенком оболочки: чем больше год урожая, тем зерна менее красные. В то же время, с увеличением длительности хранения значимо изменяется светлота оболочки зерен: чем больше длительность хранения, тем зерна более светлые.

Анализ линейных корреляций между всхожестью семян пшеницы разных генотипов и годов урожая и их признаками позволил выявить отрицательную линейную связь между красной компонентой цвета оболочки и всхожести: чем краснее зерна, тем ниже их всхожесть.

Полученные результаты демонстрируют прежде всего проявления метаболических изменений в оболочках зерен, связанные с длительным хранением и их связь с уменьшением жизнеспособности семян в результате этого.

Таким образом, предложенные нами технологии фенотипирования зерен пшеницы продемонстрировали высокую эффективность и производительность, точность оценки характеристик зерен и информативность. С их помощью были успешно решены задачи поиска взаимосвязи между фенотипическими характеристиками и генотипом, между свойствами зерен и их физиологическими свойствами: всхожестью и изменением содержания пигментов в процессе длительного хранения.

ГЛАВА 4. ВЫСОКОПРОИЗВОДИТЕЛЬНОЕ ФЕНОТИПИРОВАНИЕ КОЛОСА ПШЕНИЦЫ

4.1. Колос и его характеристики, важные для селекционно-генетических исследований

Одними из важных хозяйственно ценных признаков злаков являются характеристики колоса, которые тесно связаны с продуктивностью растения (Гончаров, 2012). Для селекционеров и генетиков важны такие параметры как число зерен в колосе, масса 1000 зерен и др. Форма зерновки также является полезным селекционным признаком, поскольку наравне с размером и выравненностью определяет рыночную стоимость зерна. Характеристики формы колоса также оказывают существенное влияние на урожайность растений (Dobrovolskaya et al., 2015). К ним относятся тип колоса, его длина и профиль, число колосков в колосе. Существенными для селекционеров также являются остистость-безостость колосьев, число плодородных и стерильных колосков (озерненность), ломкость колоса, свойства колосковой чешуи и др. (Kopopatskaya et al., 2016).

Характеристики зерновок и колоса контролируются множеством генов, тесно связанных с архитектурой соцветия злаков (Sreenivasulu and Schnurbusch, 2012). Идентификация многих из этих генов является сложной задачей (Dobrovolskaya et al., 2015). Часто для этого используется поиск локусов количественных признаков (Boden et al., 2015), требующий анализа большого количества растений при поиске ассоциаций между фенотипом и наборами генетических маркеров. Ранее были идентифицированы некоторые основные QTL и гены, ассоциированные с доместикацией и обуславливающие морфологию зерновок пшеницы, ячменя (*Hordeum vulgare* L.) и риса (*Oryza sativa* L.) (Jantasuriyarat et al., 2004; Nalam et al., 2007). Это локус Q (признак спельтоидность, “spelt factor”), локализованный на длинном плече хромосомы

5A, локус Tg (признак жесткая чешуя, “tenacious glume”), локализованный на коротком плече хромосомы 2D и локус Sog (признак мягкая чешуя, “soft glume”), локализованный на хромосоме 2A (Sood et al., 2009; Amagai et al., 2017). Многие вопросы, касающиеся генетики признаков, обуславливающих структуру колоса, таких как число, размер и форма зерен, включая их вариабельность и функциональное влияние на фенотип для большинства видов пшеницы в настоящее время не изучены.

Для идентификации генов, которые могут быть ассоциированы с признаками зерновки и колоса, важным направлением является анализ вариаций указанных признаков в специально собранных коллекциях растений (Гончаров, 2012; Youssef et al., 2017). Например, в работе Бодена и соавт. (Boden et al., 2015) для идентификации локусов, ассоциированных со структурой колоса, было проанализировано более 13 тыс. колосьев. Сложность подобных исследований обусловлена необходимостью получения больших массивов фенотипических данных, включающих промеры тысяч растений. Такой анализ обычно проводится экспертом визуально, что является трудоемким, в особенности, если требуется исследовать детали строения колоса или формы зерновок. Альтернативой такому трудоемкому анализу является использование новых технологий высокопроизводительного фенотипирования, основанных на методе компьютерного анализа цифровых изображений (Strange et al., 2015). Такие технологии обеспечивают высокую степень автоматизации сбора информации о фенотипе, ее хранение в базах данных, интеграцию с данными о генотипе и параметрах окружающей среды, создают основу для интеллектуального анализа полученной информации.

Это направление феномики особенно актуально при фенотипировании колоса, т.к. множество морфологических признаков колоса принято оценивать качественно, а не количественно. Например, по данным портала Онтологии

Сельскохозяйственных культур <http://www.cropontology.org/> большинство признаков колоса пшеницы не имеют количественной оценки. К таким признакам относятся форма, в том числе и плотность колоса, цвет колоса (колосковых чешуй), опушение колосковых чешуй, тип остистости, цвет остей, форма колоса, ломкоколосость и т.д. В связи с этим, применение подходов цифрового анализа изображений для описания формы зерна и колоса, а также их сопоставление с оценками признаков колоса, выполненных экспертами-селекционерами, представляется актуальной задачей, т.к. позволит существенно ускорить процесс получения данных за счет автоматизации и увеличит точность оценки фенотипических параметров зерна и колоса, устранив субъективизм и неточность измерений, присущих человеку.

В настоящем разделе диссертации мы предлагаем методы количественной оценки формы и размера колоса пшеницы на основе анализа цифровых изображений.

4.2. Протоколы получения изображений колоса

Нами были предложены два протокола для получения изображения колоса.

Протокол 1. Колос располагается вертикально перед синим фоном. Колос и цветовая шкала ColorChecker держатся на прищепках, которые помещаются на штатив. С помощью данного метода, вращая колос относительно его оси, возможно производить съемку колоса в четырех или более проекциях. При съемке использовались два импульсных источника света Falcon TE-300 (мощность вспышки выставлена в положение 2.5), в качестве модификаторов света использовали софтбоксы Falcon 60x60 см; камера Canon 350D, объектив EF-S 18-55mm f/3.5-5.6; параметры съемки выдержка 1/160, диафрагма 11, ISO 100, фокусное расстояние 55 мм, съемка в RAW-формат; баланс белого

устанавливался при обработке RAW-файла по белому сектору ColorChecker. Расстояние от объекта: до фона 60 см, до источников света 100 см, до фотоаппарата 120 см. Фотоаппарат, установленный на штативе, объект и источники света расположены на одной высоте, комфортной для работы фотографа. Схема размещения камеры и источников света относительно объекта приведена на рис. 4.1 А, результат изображения полученного в результате выполнения протокола на рис. 4.1 Б.

Протокол 2. Колос лежит на стеклянном столике, который располагается на столе с поверхностью синего цвета (фон). Фотокамера фиксируется на стойке типа "журавль" над стеклом. С помощью данного метода можно производить съемку лицевой проекции колоса. Для освещения использовали два импульсных источника света Falcon TE-300 (мощность вспышки выставлена в положение 1.0 и 1.4), в качестве модификаторов использовали софтбоксы Falcon 60x60 см; камера Canon 600D, объектив EF-S 28-135 мм f/3.5-5.6; параметры съемки выдержка 1/160, диафрагма 10, ISO 200, фокусное расстояние 112 мм, съемка в RAW-формат; баланс белого устанавливался при проявке RAW-файла по белому фону ColorChecker. Расстояние от фотоаппарата до объекта 70 см, от источников света до объекта 60 см, высота стола 60 см, высота просветного столика со стеклом (расстояние от объекта до синего фона) 20 см. Схема размещения камеры и источников света относительно объекта приведена на рис. 4.1 В, результат изображения полученного в результате выполнения протокола на рис. 4.1 Г.

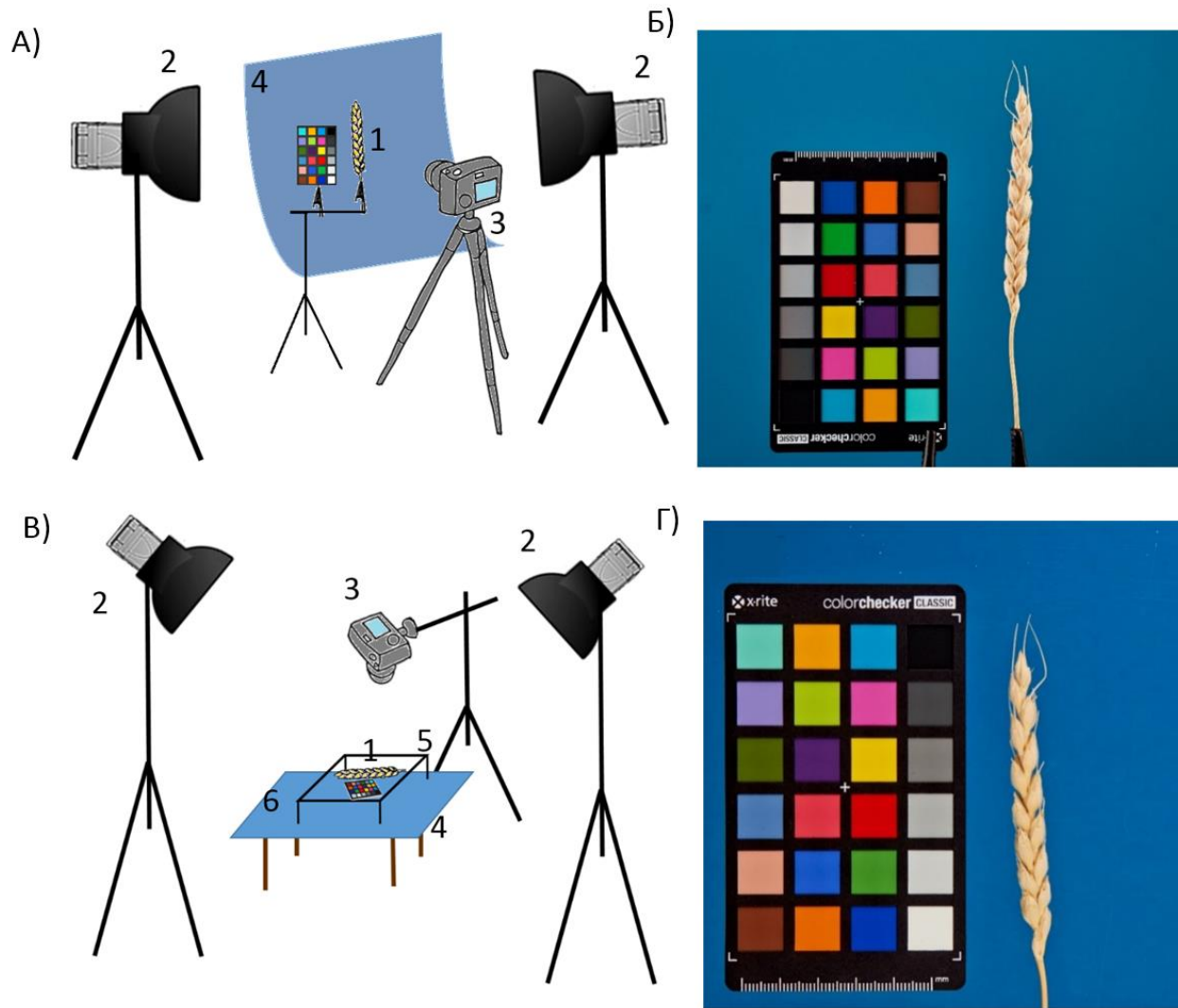


Рисунок 4.1. Протоколы съемки колосьев пшеницы. Протокол 1 («на прищепке»): А) - Схема размещения камеры и источников света относительно объекта съемки: 1 – колос, 2 – источники света, 3 – камера, 4 - бумажный фон; Б) – изображение, полученное в результате выполнения протокола. Протокол 2 («на столе»): В) Схема размещения камеры и источников света относительно объекта съемки (1- колос, 2 – источники света, 3 – камера, зафиксированная на стойке типа "журавль"; 4 – предметный стол; 5 – просветной стол, 6 – фон; Г) изображение, полученное в результате выполнения протокола.

Протокол 2, в отличие от протокола 1, позволяет избавиться от прищепок для фиксации колоса. В дополнение к этому расположение стекла на

определенной высоте над поверхностью фона позволяет избежать теней под объектом. Однако протокол 2 не позволяет фиксировать колос под произвольными углами вращения вокруг своей оси при съемке. В область каждого кадра помещается цветовая шкала ColorChecker Mini Classic target (<https://xritephoto.com/camera>), что позволяет произвести цветовую коррекцию изображения. Такая коррекция позволяет устранить искажения цвета на изображении, возникающие из-за разных условий освещения. Другим преимуществом использования цветовой шкалы является ее стандартный размер 63 x 108 мм, что позволяет оценить масштаб изображения.

3-канальные (3x: 0-255, “RGB”) изображения колосьев получаются с помощью цифровой фотокамеры в формате jpg или png.

4.3. Идентификация колоса и остей на изображении

4.3.1. Предварительная обработка изображения

Обработка изображений осуществлялась с использованием пакета OpenCV (Open Source Computer Vision Library, <https://opencv.org>) (Kaehler, A.; Bradski, G., 2016). На первом этапе анализа мы выполняли идентификацию цветовой шкалы на изображении. Она расположена в левом нижнем углу рядом с колосом. Цветовая шкала может быть наклонена по отношению к вертикальной оси изображения, а в случае протокола «на прищепке» ее плоскость не всегда перпендикулярна оптической оси объектива. Для идентификации шкалы использовалось ее эталонное изображение, на котором определялись ряд дескрипторов с помощью поиска ключевых точек, как это описано в разделах 3.3.2 и 3.3.3. Эта шкала также использовалась для определения масштаба.

На следующем этапе предварительной обработки мы сглаживали изображение при помощи фильтра Гаусса с ядром 3×3 для удаления шумов, удаляли элементы изображения штатива и прищепки (для протокола «на

прищепке»).

4.3.2. Бинаризация изображения

Бинаризацию изображения проводили после преобразования изображения в цветовое пространство HSV, в котором светлота выделена в отдельный канал, поэтому цвет при различном уровне освещения характеризуется более устойчиво. Для бинаризации изображения на области колоса и фона использовали интервалы значения каналов HSV, которые подбирались на обучающей выборке (см. ниже). В результате мы получали изображение, сегментированное на пиксели фона и остальные пиксели, которые принадлежали как области самого колоса, так и области остей. После удаления области ColorChecker и сегментации по цвету колоса на изображении выбирался самый большой по площади контур, который соответствовал колосу и остям. Меньшие по размеру контуры, которые соответствовали мусору и фрагментам чешуек исключались из дальнейшего анализа.

4.3.3. Идентификация остей

Цвет остей близок к цвету колоса. На некоторых изображениях ости интенсивно пересекаются или даже слипаются в пучки, что существенно затрудняет идентификацию отдельных остей, а в некоторых случаях делает это невозможным. С учетом этого использовался алгоритм, состоящий из двух этапов.

На первом этапе идентифицировали пиксели остова остей. Поскольку толщина остей существенно меньше, чем толщина тела колоса, для получения остова остей использовали алгоритм частичной скелетизации. Пиксели на границе колоса и фона удалялись итерационно при условии сохранения связности областей колоса и остова так, что удаление очередного пикселя на

границе колоса не увеличивало количество изолированных областей колоса на изображении. Для остановки алгоритма использовался порог на количество итераций, который подбирался в результате оптимизации алгоритма на обучающей выборке изображений (см. ниже). В процессе эрозии последовательность удаляемых пикселей контура запоминалась. На основе этого можно было определить, пиксели, удаление которых приводило к формированию остова. Эти пиксели отмечались как пиксели остей.

Работа алгоритма выделения областей колоса и остей зависит от следующих параметров:

1. Целевые значения каналов HSV и интервалы их допустимых отклонений для бинаризации изображения на области фона и колоса/остей.
2. n – количество итераций до остановки алгоритма скелетизации.

Для подбора этих параметров мы использовали выборку изображений 93 колосьев гибридов F2 австралийского сорта мягкой пшеницы ‘Triple Dirk’ с образцом KU506 китайской пшеницы *Triticum yunnanense*, полученных как по протоколу «на столе», так и «на прищепке». Для каждого колоса, на основании экспертной оценки, приведена классификация по типам остистости: безостые (рис. 4.2 А), с зачатками остей (рис. 4.2 Б), полуостистые (рис. 4.2 В) и короткоостые (рис. 4.2 Г) (Konopatskaia et al., 2016).



Рисунок 4.2. Типы остистости колосьев. (А) Безостые. (Б) С зачатками остей. (В) Полуостистые. (Г) Короткоостые.

Распределение изображений колосьев в этой выборке по типам протокола и остистости приведено в табл. 4.1. Для каждого из этих изображений вручную были маркированы пиксели остей и пиксели тела колоса. Изображения были произвольным образом разделены на тестовую (30 изображений) и обучающую (63 изображений) выборки, так что соотношение типов остистости колосьев и

протокола в тестовом и обучающем наборе было приблизительно одинаковым.

Таблица 4.1. Распределение колосьев из выборки изображений по типам остистости.

Тип остистости	Количество изображений	протокол «на прищепке»	протокол «на столе»
безостые	16	10	6
с зачатками остей	4	4	0
полу-остистые	14	14	0
короткоостые	59	36	23

Для оценки точности алгоритма сегментации изображения на фон и колос использовался индекс Жаккара J (Tan, 2005) (формула 1.2, раздел 1.2). Мы рассчитывали следующие индексы Жаккара: J_e – точность бинаризации всего колоса вместе с остями, J_b – точность распознавания пикселей тела колоса (область колоса за вычетом пикселей, принадлежащих остям), и J_a – точность распознавания остей. В процессе подбора параметров мы оптимизировали среднее значение J_e для тестовой выборки, после оптимизации независимо оценивали J_b и J_a на тестовой выборке. Для оптимизации использовали генетический алгоритм (Sivanandam, 2008). Блоки параметров (особи) состояли из наборов 7 целевых цветов HSV, и их диапазонов (dH, dS, dV). Блоки могли обмениваться (кроссинговер) целевыми цветами со связанными диапазонами. Размер популяций от 20 до 100 особей.

После выделения областей остей и колоса рассчитываются характеристики остистости колоса: площадь области остей S_a (мм²) рассчитывается как количество формирующих ее пикселей умноженное на

площадь одного пикселя; количество остей, N_a , которое определялось по количеству контактов между областями колоса и остей (количество оснований остей); суммарная длина остей L_a (мм), которая определялась как сумма длин пикселей, формирующих скелетизированный остов области остей; средняя длина остей $l_{a=} L_a/N_a$ (мм). В силу того, что часть изображений содержала сильное пересечение и/или наложение остей (рис. 4.2), выделить отдельные ости на них было невозможно, и мы не оценивали длины отдельных остей.

4.4. Алгоритмы морфометрии колоса

Идентификация контура колоса и его выпрямление

После удаления остей в местах сочленения с “телом” колоса иногда образовывались ломаные изгибы. Поэтому мы выполняли сглаживание контура колоса на основе алгоритма вычисления эллиптических дескрипторов Фурье (Kuhl, Giardina, 1982). Для распознанного контура колоса после этапа удаления областей остей вычисляются дескрипторы для 70-ти гармоник. Далее – на их основании вычисляются точки сглаженного контура.

Осевую линию колоса мы аппроксимировали ломаной линией, построенной на основе сегментов контура “тела” колоса. Построение этой линии выполнялось итеративно. На первом шаге для пикселей контура определялся центр масс и главные оси эллипсоида, аппроксимирующего контур. Большая ось соответствовала направлению стержня колоса в центре масс. Перпендикуляр к главной оси делит область колоса на две части. На втором шаге аналогичная процедура выполнялась для каждой из этих частей. В результате, каждая из них также делилась соответствующей осью на две части, для которых выполнялся следующий шаг итерации. Данная процедура выполнялась последовательно, пока количество сегментов не превышало 20. Центры масс каждого из полученных сегментов определяли ломанную,

аппроксимирующую осевую линию колоса. На последних этапах итерации в некоторых случаях размеры сегмента, поперек оси могли быть больше, чем вдоль оси. В этом случае в качестве оси сегмента мы строили ось, сонаправленную большей оси сегмента, полученного на предыдущем этапе, но проходящую через центр масс текущего сегмента.

После определения осевой линии колоса производится процедура выпрямления контура колоса так, что эта линия проходит вертикально, ее верхний конец соответствует вершине колоса, нижний – основанию колоса, а расстояния пикселей трансформированного контура до центральной линии равны расстояниям соответствующих пикселей на исходном контуре. Такое преобразование позволяет устранить деформации контура колоса, вызванные его изгибом. Определение размера и количественных характеристик формы колоса применяется к выпрямленным контурам колосьев.

Интегральные характеристики формы колоса

Характеристики формы колоса делятся на несколько категорий. К первой категории относятся интегральные характеристики формы. Длина колоса, L_e (мм), которая аппроксимируется длиной осевой линии колоса. P_e (мм), периметр контура тела колоса. S_e (мм²), площадь области колоса. Индекс квадрата, $SQI=S_e/(L_e^2)$ - отношение площади колоса к квадрату его длины.

Округлость (circularity, I): индекс отражает насколько форма контура близка к форме окружности (формула 3.1, раздел 3.3.3). Индекс варьирует от 0 до 1. Значение 1 индекс принимает для идеального круга. На контурах, имеющих много небольших выпуклостей на поверхности, периметр увеличивается, а индекс округлости принимает более низкие значения. В этих случаях целесообразно использовать индекс закругленности (roundness, R), поскольку эта величина не зависит от таких неровностей периметра (рассчитывается по формуле 3.2). Индекс шероховатости (rugosity, Rg)

определяется как отношение периметра контура к выпуклому периметру, рассчитывается по формуле 3.3. Индекс компактности/целостности (solidity, S) - это отношение площади контура к площади его выпуклой оболочки, рассчитывается по формуле 3.4.

Модель сечений

Первая модель для описания формы колоса представляла собой набор сечений, которые определялись перпендикулярами к его осевой линии, отложенными с шагом в $1/21$ длины колоса. Для каждого перпендикуляра определялось два расстояния до пикселей контура колоса (с обеих сторон от осевой линии). Таким образом, эта модель определялась 40-ка параметрами.

Модель четырехугольников

Контур колоса, расположенного горизонтально, можно представить в виде двух четырехугольников (рис. 4.3), верхнего и нижнего, с общим основанием – осевой линией колоса (длина L_e). Геометрия верхнего четырехугольника определяется четырьмя независимыми параметрами:

x_{u1} – расстояние от вершины колоса до проекции B' вершины B на основание AD ;

x_{u2} – расстояние от B' до проекции C' вершины C на основание AD ;

y_{u1} – расстояние от вершины B до ее проекции B' на основание AD ;

y_{u2} – расстояние вершины C до ее проекции C' на основание AD ;

Расстояние x_{u3} от проекции C' до основания колоса D можно вычислить, зная длину колоса $x_{u3} = L_e - x_{u2}$.

Для нижнего четырехугольника определяются аналогичные параметры x_{b1} , x_{b2} , x_{b3} , y_{b1} , y_{b2} (рис. 4.3).

Параметры верхнего и нижнего четырехугольника подбирались таким образом, чтобы минимизировать квадратичное отклонение по перпендикуляру от осевой линии между пикселями контура и ребрами четырехугольника.

Минимизация проводилась с помощью алгоритма Левенберга-Марквардта (Ranganathan, 2004), реализованного в библиотеке Apache (class org.apache.commons.math3.fitting.leastsquares.LevenbergMarquardtOptimizer); <http://commons.apache.org/proper/commons-math/javadocs/api-3.4/org/apache/commons/math3/optimization/general/LevenbergMarquardtOptimizer.html>).

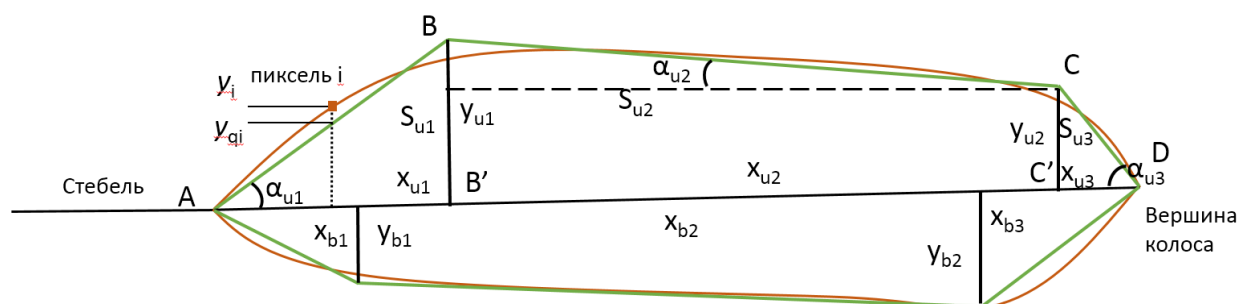


Рисунок 4.3. Представление формы колоса в виде двух четырехугольников. Черной горизонтальной линией показана осевая линия колоса. Контур колоса показан коричневой линией. Четырехугольники, аппроксимирующие контур колоса показаны зелеными линиями. На рисунке для верхнего четырехугольника показаны основные параметры, характеризующие его геометрию.

Мы также оценили ряд производных параметров для двух четырехугольников:

α_{u1} – величина наклона ребра AB относительно основания верхнего четырехугольника (градусы);

α_{u2} – величина наклона ребра BC относительно основания верхнего четырехугольника (градусы);

α_{u3} – величина наклона ребра BD относительно основания верхнего четырехугольника (градусы);

t_{1u} – тангенс угла α_{u1} ;

t_{2u} – тангенс угла α_{u2} ;

t_{3u} – тангенс угла α_{u3} ;

S_{u1} – площадь треугольника ABV' (мм^2);

S_{u2} – площадь трапеции $VB'V'C$ (мм^2);

S_{u3} – площадь треугольника ABV' (мм^2);

S_u – площадь верхнего четырехугольника (мм^2);

y_{um} – среднее значение высоты верхнего четырехугольника (мм);

Аналогично вычислялись параметры для нижнего четырехугольника.

Для двух четырехугольников также вычислялись:

$A_{Ix2} = (x_{u1} - x_{b1})^2 / x_{u1} + (x_{u2} - x_{b2})^2 / x_{u2} + (x_{u3} - x_{b2})^2 / x_{u3}$ – индекс асимметрии по длинам сегментов (безразм.);

$A_{Iy2} = (y_{u1} - y_{b1})^2 / y_{u1} + (y_{u2} - y_{b2})^2 / y_{u2}$ – индекс асимметрии по высотам сегментов (безразм.)

$A_{Ixy2} = A_{Ix2} + A_{Iy2}$ – общий индекс асимметрии (безразм.).

Радиальная модель

Параметры радиальной модели вычисляются следующим образом: от центра масс контура колоса проводятся 360 лучей, начиная от направления главной оси контура с шагом в 1 градус, до ближайшей точки линии контура к центру масс вдоль луча. Длины полученных таким образом 360 отрезков представляют радиальную модель.

Полный список параметров, характеризующих размер и форму колоса, приведен в Таблице П8 Приложения.

4.5. Алгоритмы предсказания индекса плотности колоса и типа его формы

Выборка изображений колосьев

Мы оценили работоспособность нашего подхода на задаче предсказания плотности колоса и его типа. Для этого из базы данных SpikeDroid (Генаев и соавт, 2019) нами были взяты изображения колосьев для 249 растений, которые были проаннотированы вручную. Мы использовали оцифрованные данные 1245 фотографий колосьев, полученных для 9 различных видов и гибридов гексаплоидных пшениц (табл. 4.2). Эти виды пшениц имеют контрастные формы колоса, которая контролируется хорошо изученными генами (Swaminathan, Rao, 1961). У каждого растения выбирался главный колос. Для каждого колоса было получена 5 фотографий: в одной проекции по протоколу «на столе» и четырех проекциях по протоколу «на прищепке».

Таблица 4.2. Виды и гибриды пшениц, использованные в задаче предсказания индекса плотности и типа формы колоса

Вид	Количество растений	Количество изображений
<i>T. compactum</i> Host	63	315
F ₂ Triple Dirk x <i>T. yunnanense</i>	52	260
<i>T. aestivum</i> L.	50	250
<i>T. antiquorum</i> Heer ex Udacz.	20	100
<i>T. sphaerococcum</i> Perc.	19	95
<i>T. spelta</i> L.	18	90
Speltiforme	9	45
<i>T. yunnanense</i>	9	45
<i>T. macha</i> Dekapr. et Menabde	9	45

Для каждого растения экспертом был определен тип колоса, измерена длина колоса и посчитано число колосков в колосе. По типу колоса 110 растений было отнесено к категории компактный, 72 к категории нормальный и 67 к категории спельта. Количественный признак, который имеет высокую корреляцию с формой колоса – это индекс плотности D (1), где $(A-1)$ – число колосков в колосе без верхушечного колоска; B – длина стержня колоса, см (Фляксбергер, 1935).

$$D = \frac{10(A-1)}{B}. \quad (4.1)$$

Методы предсказания характеристик колоса

Для предсказания индекса плотности мы использовали метод случайного леса, реализованный в классе `RandomForestRegressor` пакета `sklearn.ensemble` из библиотеки `scikit-learn` (Pedregosa et al., 2011). Для определения типа колоса мы использовали логистическую регрессию и случайный лес, реализованные в классах `LogisticRegression` и `RandomForestClassifier` из этой же библиотеки. Для предсказания мы использовали 44 признака модели четырехугольников, 7 индексов, которые описывают геометрические характеристики колоса (`Perimeter`, `Ear area`, `Awns area`, `Circularity index`, `Roundness`, `Solidity`, `Rugosity`). Для признаков модели сечений и радиальной модели число признаков было уменьшено до 10 с помощью анализа главных компонент (по данным выборки отбирались наибольшие компоненты так, что в итоге им соответствовало в сумме более 90% вариаций). В итоге мы анализировали для каждого изображения 71 признак.

Для изображений, полученных по протоколу «на прищепке», один и тот же колос представлен в четырех проекциях. Мы объединяли параметры всех четырех изображений, ранжировав их в порядке уменьшения ширины колоса. В результате получалось 284 признака (от 1 до 142 описывают лицевую сторону

колоса, от 143 до 284 - боковую сторону колоса).

Для предсказания характеристик формы колоса мы проводили отбор наиболее значимых признаков, по их вкладу в предсказание плотности колоса. Ранжирование признаков по значимости осуществлялось по усредненному значению результатов 8 различных подходов: (1) коэффициенты линейной регрессии (класс `LinearRegression`); (2) коэффициенты регрессии с регуляризацией лассо (Lasso); (3) коэффициенты регрессии с регуляризацией ridge (Ridge); (4) Отбор признаков по критерию стабильности (Stability) (Meinshausen, Bühlmann, 2010), использовалась реализация в классе `RandomizedLasso` библиотеки `sklearn`. (5) ранг признака, оцененный путем рекурсивного удаления признаков с наименьшей значимостью (RFE) (Guyon et al., 2002); (6) Ранжирование признаков на основе среднего уменьшения энтропии признака, которая вычисляется при построении деревьев решений в модели случайного леса (RF), использовался класс `RandomForestRegressor` в пакете `sklearn.ensemble`. (7) Ранжирование признаков на основе вычисления корреляции между регрессором и целевым признаком (Corr), использовался класс `f_regression` в пакете `sklearn.feature_selection`. (8) Ранжирование на основании значения максимального информационного коэффициента (MIC), который позволяет учитывать не линейные зависимости между признаками (Reshef et al., 2011), использовался класс `MINE` в пакете `mineru`. Оценка значимости признака для всех описанных стратегий может меняться в интервале от 0 до 1.

Оценка точности предсказания плотности колоса

Для оценки качества предсказания индекса плотности мы использовали значения средней абсолютной ошибки (MAE) и средней абсолютной процентной ошибки (MAPE). Они вычислялись по формулам 1.3 и 1.4. В этих формулах N – число колосьев; n_j – индекс плотности, посчитанный вручную;

n'_j – предсказанное значение индекса плотности. Дополнительно мы оценивали коэффициент корреляции Пирсона между предсказанными значениями и значениями определенными экспертом. Для оценки качества классификации колосьев по форме использовалась мера F , которая вычисляется как среднее гармоническое между точностью и полнотой (формулы 1.5, 1.6, 1.7). В этих формулах параметры tp , tn , fp , fn соответствуют числу истинно-положительных, истинно-отрицательных, ложно-положительных, ложно-отрицательных решений. Для оценки точности мы проводили процедуру кросс-валидации: данные были поделены на обучающую и тестовую выборки в соотношении 70% на 30%. Значения MAE, MARE, F усреднялись по 5 итерациям кросс-валидации как для обучающей, так и для тестовых выборок.

4.6. Оценка точности распознавания областей остей и колоса

На тестовой выборке изображений после подбора оптимальных параметров алгоритма было достигнуто среднее значение точности распознавания тела колоса $J_b = 0.925$ и остей $J_a = 0.660$. На обучающей точность распознавания тела колоса $J_b = 0.932$ и остей $J_a = 0.634$. Использование цветокоррекции не оказало существенного влияния на точность распознавания колоса и остей: среднее значение индекса Жаккара J_b сегментации “тела” колоса с учетом цветокоррекции составило 0.925, для остей с учетом цветокоррекции среднее значение J_b составило 0.679.

На рис. 4.4 приведены примеры результатов применения алгоритма к изображениям для колоса с идентификатором “6450” из БД SpikeDroidDB. Для данного изображения была получена оценка точности распознавания колоса $J_b=0.963$ и остей $J_a=0.796$. На изображении с выделенными остями заметно что основная доля неразмеченных пикселей остей концентрируется на их кончиках, а для большей части остей их пиксели правильно идентифицировались.

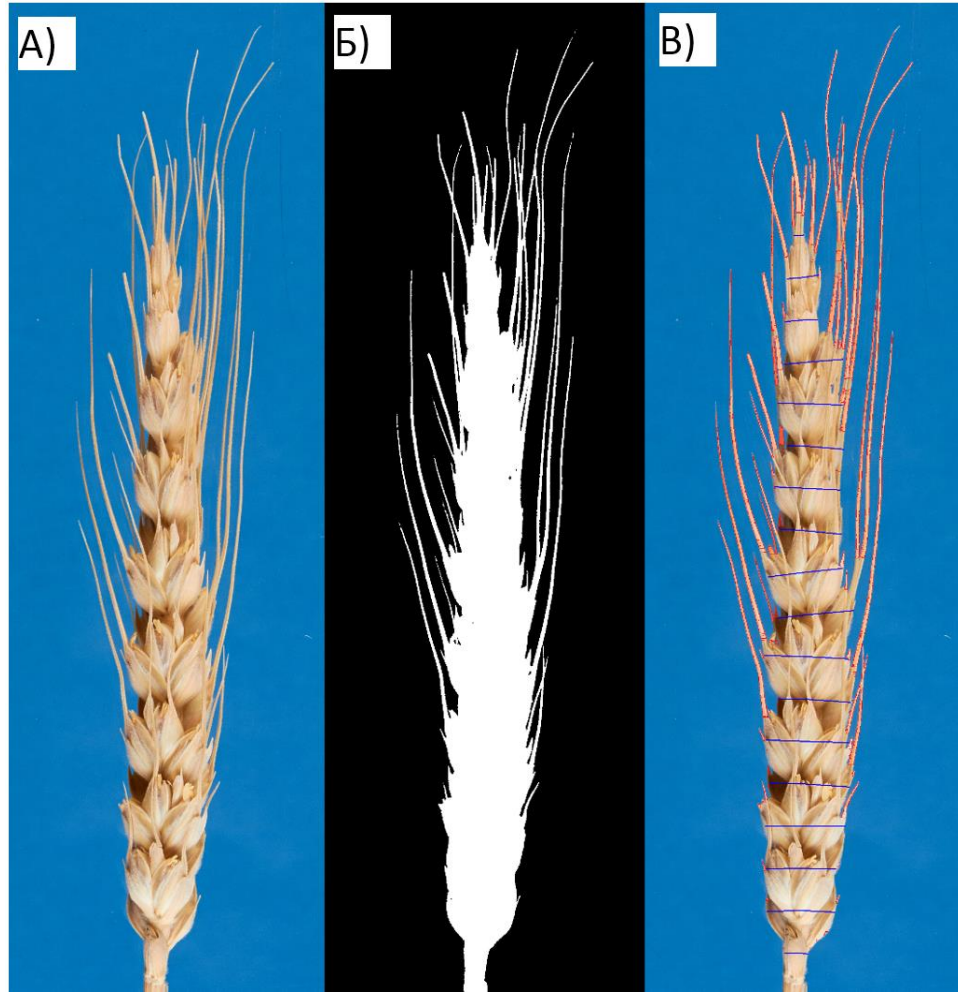


Рисунок 4.4. Этапы алгоритма определения областей остей на примере изображения колоса “6450”. (А) Изображение колоса на синем фоне. (Б) Бинаризованное изображение колоса. (В) Исходное изображение колоса с отмеченными (красным) распознанными областями остей и (синим) ребрами профиля.

Мы провели оценку влияния различных факторов, связанных с протоколом получения изображения и его обработки на точность определения остей: масштаб съемки (количество пикселей на единицу фотографируемой площади); тип протокола («на столе» или «на прищепке»); проекция колоса (лицевая или боковая сторона колоса) для протокола «на прищепке».

Прежде всего, мы исследовали влияние на точность идентификации остей масштаба изображения. В выборке изображений, использованных нами,

оказались изображения 4 масштабов (т.е. снятые на 4 разных удалениях объектива от колоса). Это хорошо заметно из графиков зависимости между площадью остей колоса, выраженной в пикселях и в мм² (рис. 4.5).

При этом изображения, полученные по протоколу «на столе» расположены вдоль одной линии (масштаб 4, оранжевые кружки). Остальные величины масштаба соответствуют изображениям, полученным по протоколу «на прищепке». Таким образом, при применении предложенных протоколов различия в масштабе изображений являются неизбежными и использование цветовой шкалы для определения масштаба изображения является оправданным.

Распределение коэффициента Жаккара J для идентификации тела колоса и остей для всех 93 изображений приведено на рис. 4.6.

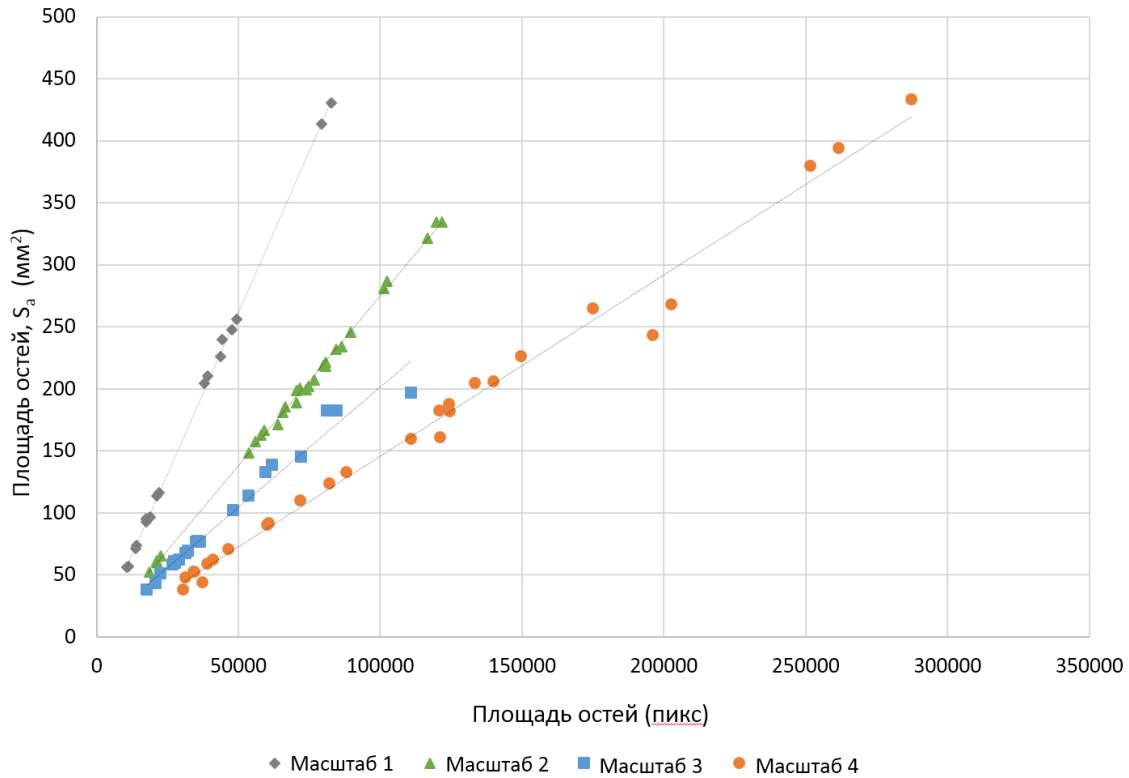


Рисунок 4.5. Отношение количества пикселей остей к площади остей (в мм²), сгруппированных по масштабам съемки. Значения линейной регрессии: масштаб 1, $\alpha = 0.00526$; масштаб 2, $\alpha = 0.00277$; масштаб 3, $\alpha = 0.00213$; масштаб 4, $\alpha = 0.00146$.

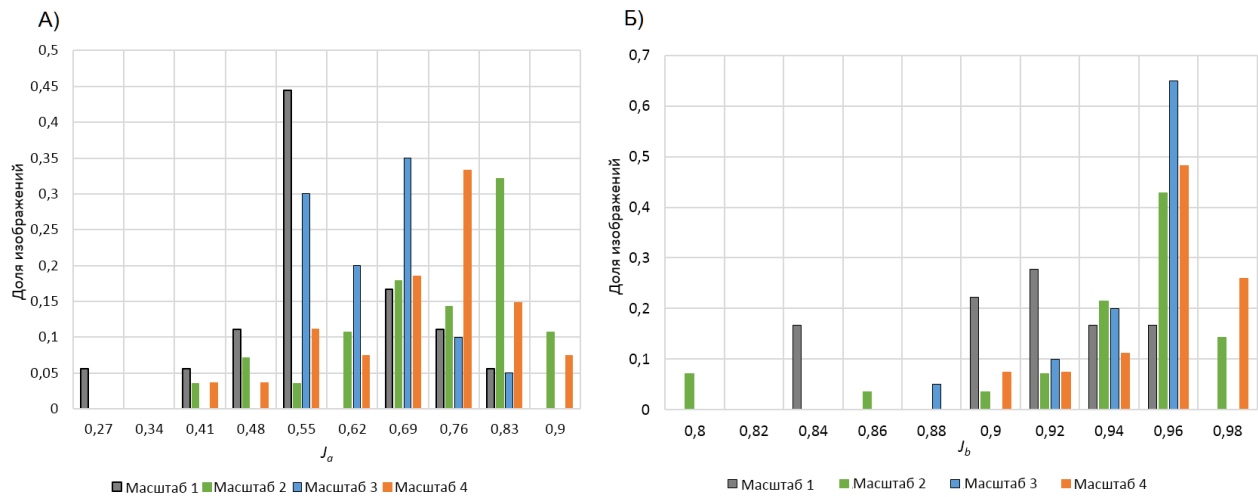


Рисунок 4.6. Распределение индекса J для остей (А) и “тела” (Б) колосьев, сгруппированных по масштабам съемки.

Оценка точности распознавания областей остей J_a находится в интервале от 0.27 до 0.9. Для масштаба 1, распределение точности смещено в меньшую сторону (рис. 4.6) и в среднем составило около 0.549, что закономерно, т.к. это соответствует наибольшему удалению камеры от колоса. Для остальных масштабов: 2, 3 и 4, распределения J_a отличаются в меньшей степени, средние значения: 0.695, 0.607 и 0.676 соответственно. Оценка точности распознавания “тела” колоса J_b распределилась в интервале от 0.8 до 0.98. Средние значения составили: 0.901, 0.928, 0.938, 0.945 для 1, 2, 3, 4 масштабов, соответственно. Можно заметить, что для параметра J_b наблюдается явная тенденция увеличения точности распознавания области “тела” колоса при уменьшении расстояния от колоса до камеры, в отличие от параметра J_a .

Вторым возможным фактором, влияющим на точность работы алгоритма, является тип протокола. Для протокола «на столе», в серии снимков расстояние от камеры до объекта фиксируется, колос расположен горизонтально на поверхности стола и находится в плоскости, перпендикулярной оси объектива. Для протокола «на прищепке» колос может отклоняться от плоскости, перпендикулярной оси объектива в силу его изогнутости или случайных отклонений оси от вертикали при установке колоса на прищепке. Мы проанализировали распределения индекса Жаккара J для изображений, сгруппированных по протоколу (рис. 4.7).

Результаты, представленные на рисунке, показали, что масштаб изображения значимо влияет на точность распознавания как тела колоса, так и остей. Тип протокола оказывают существенное влияние на точность определения тела колоса, но не остей.

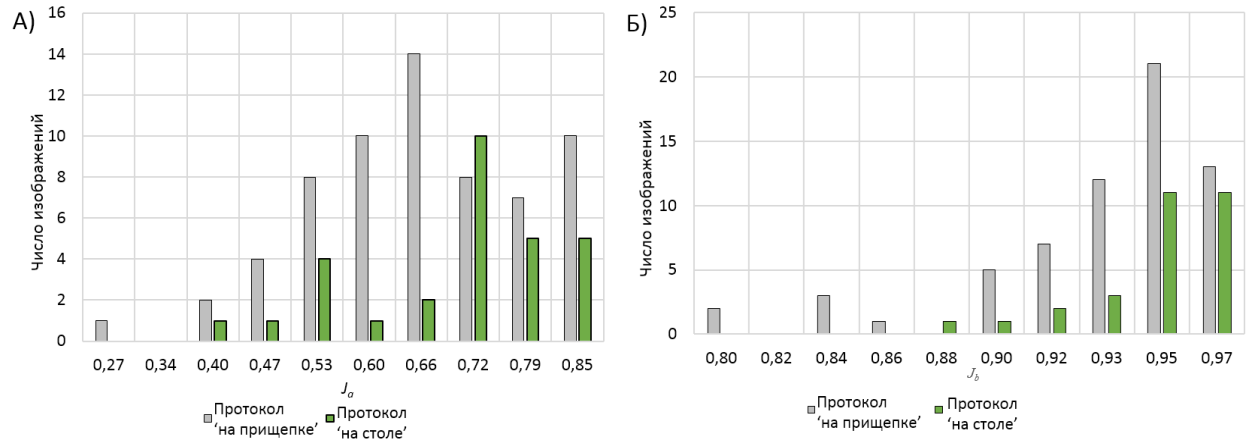


Рисунок 4.7. Распределение значений индексов J остей колосьев (А) и тела колоса (Б), сгруппированные по протоколам «на прищепке» и «на столе».

4.7. Анализ характеристик колосьев и предсказание типа колоса

Различия характеристик разных типов колосьев для модели четырехугольников

Результаты сравнения между длиной колоса, оцененной вручную, и параметром L_e , полученным с использованием 2D-изображений, представлены на рис. 4.8. Коэффициент корреляции Пирсона между двумя значениями составляет 0.808 ($p < 0.01$), MAE=0.75 см, MAPE=0.09. Полученные данные свидетельствуют о хорошем соответствии оценок длины колоса, выполненных вручную линейкой и автоматически, на основе анализа изображений.

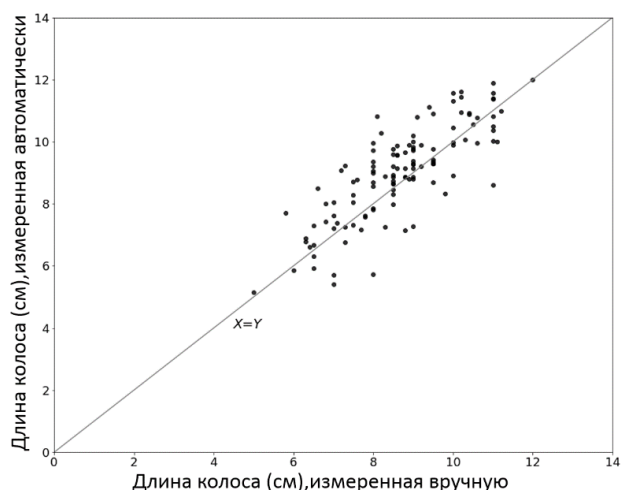


Рисунок 4.8. Диаграмма рассеяния для длины основного колоса для гибридных растений Triple Dirk B × KU 506 F2, оцененных вручную (ось X) и на основе анализа 2D-изображений (ось Y). Сплошная линия $Y=X$.



Рисунок 4.9. Изображения колосьев формы различных типов: компактный (А), нормальный (Б), спельтоидный (В). Каждая панель содержит три фрагмента: на левом изображено исходное изображение колоса, в центре – сегментированное на тело (зеленый цвет), ости (малиновый цвет) и линии сечений (синим). Справа приведено изображение модели четырехугольников для соответствующего колоса с указанием основных параметров модели.

Примеры изображений колосьев формы трех различных типов (компактный, нормальный и спельтоидный) приведены на рис. 4.9. На рисунке также приведены параметры модели для верхнего и нижнего четырехугольника

(отношение к длине колоса параметров x_1 , x_2 и h_1 , h_2). Видно, что для компактного колоса характерна меньшая длина и большее значения отношение параметров h_1 , h_2 к длине колоса (ширина). Для спельтоидной формы колоса наоборот, длина колоса наибольшая, а отношение h_1 , h_2 к длине колоса (ширина) наименьшее, среди колосьев трех типов. Для нормального колоса, эти значения промежуточные. Также интересно, что отношение длины основания колоса к его общей длине (x_1/L_e) у компактного и спельтоидного колоса больше, чем у нормального колоса (~ 0.40 , ~ 0.29 и 0.07 , соответственно). Также следует отметить, что значения геометрических параметров для верхнего и нижнего четырехугольников оказываются близкими. Эти результаты свидетельствуют об адекватности описания формы колоса в рамках предложенной модели четырехугольников и позволяют говорить об адекватном представлении размера и формы колоса в рамках этой модели.

Анализ параметров остистости для выборки колосьев

Мы проанализировали значения площади остей (в мм^2), S_a . Значимые различия распределений S_a на выборке из 46-ти изображений (без повторов в разных проекциях) были выявлены при сравнении изображений колосьев с вариантами остистости “с зачатками остей” и “короткоостые” ($\chi^2 = 22.64$; $p < 0.05$). Также достоверно различались распределения у колосьев с типами остистости “безостые” и “короткоостые” ($\chi^2 = 24.75$; $p < 0.05$); “короткоостые” и “полу-остистые” ($\chi^2 = 18.09$; $p < 0.05$).

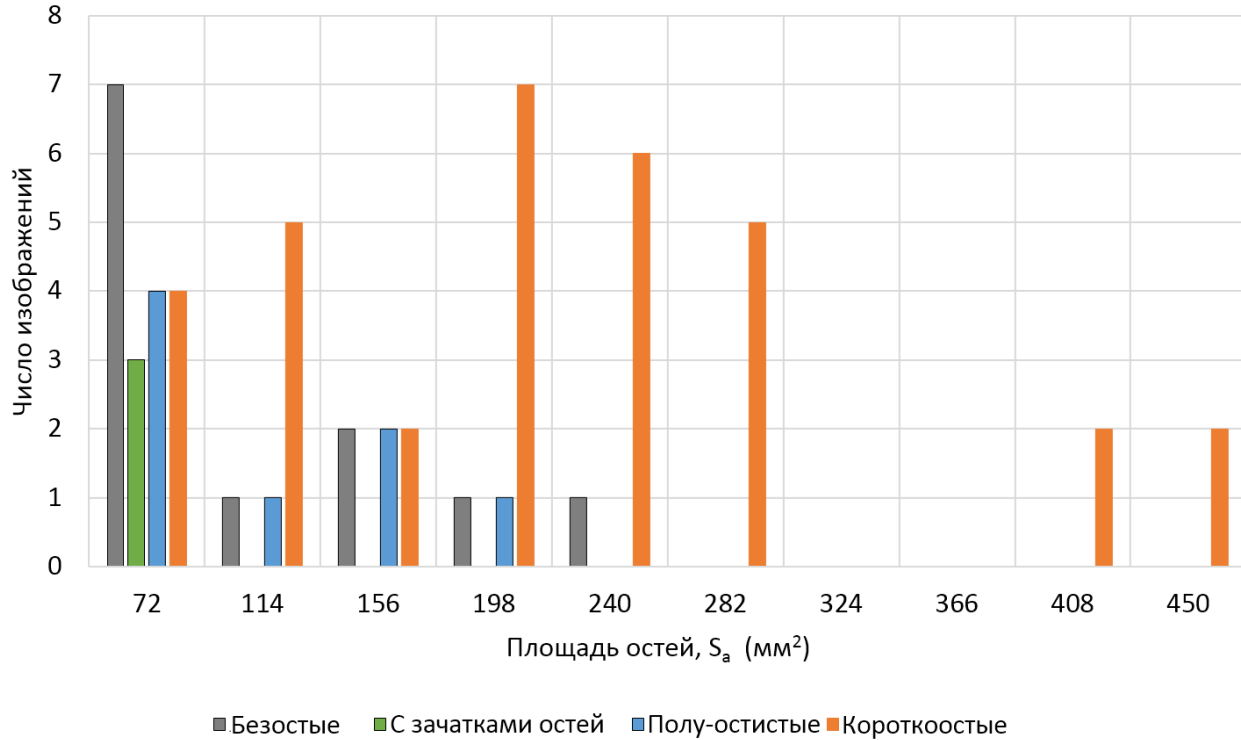


Рисунок 4.10. Гистограмма распределения типов остистости колосьев из выборки по вычисленной площади остей.

Распределение, представленное на рис. 4.10 демонстрирует, что типы остистости колосьев распределены относительно S_a следующим образом: самый большой охват имеют колосья с типом остистости “короткоостые”, которые имеют как малые, так и большие площади остей. Пик количества изображений колосьев с данным типом остистости приходится на средние площади ($180 < S_a < 222$ мм²). Наименьшую площадь остей имеют колосья с типом остистости “с зачатками остей”. Далее по возрастанию площади остей идут “безостые” колосья и “полу-остистые” колосья. Однако, полученные распределения значений S_a для разных типов колосьев достаточно сильно пересекаются: существует лишь один интервал S_a , который содержит один тип остистости колосьев, 264 мм² < S_a , в который попадают изображения только короткоостых колосьев.

Предсказание индекса плотности и типа формы колоса

Распределение характеристик колосьев разного типа в выборке (табл. 4.2) приведено на рис. 4.11. Рис. 4.11 (А) показывает распределение колосьев по индексу плотности, оцененном вручную. Рис. 4.11 (Б) показывает двумерное распределение по длине колоса и количеству колосков.

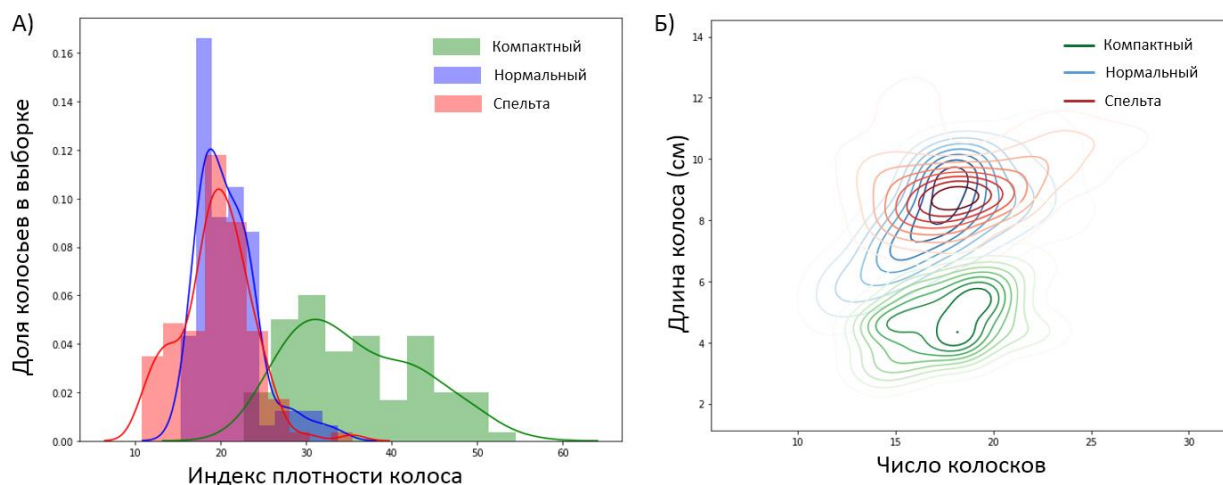


Рисунок 4.11. Распределение характеристик колоса для выборки колосьев разного типа. (А) Гистограммы распределения колосьев по индексу плотности: по оси X отложен индекс плотности колоса, по оси Y – доля колосьев в выборке, распределение спельтоидных колосьев показано зеленым, нормальных – синим, компактных – красным. (Б) Диаграмма рассеяния колосьев в пространстве количества колосков (ось X) и длины колоса (ось Y). Распределение показано линиями рельефа, более насыщенный цвет соответствует более высоким значениям доли колосьев.

Видно, что для компактного колоса характерны высокие значения индекса плотности (от 25 и выше). Значения плотности D для нормальных колосьев находятся в интервале от 15 до 35, для спельтоидных колосьев - от 10 до 30. Видно, что распределения для последних двух типов колосьев существенно перекрываются. Аналогичная картина существенного перекрывания распределений количества колосков и длины колоса наблюдается для этих типов колосьев. Из рис. 4.11 (Б) также видно, что по

длинам колоса и количеству колосков компактные колосья хорошо отделены от спельтоидных и нормальных, в то время как колосья двух последних типов по этим параметрам существенно перекрываются, так что различить их практически невозможно.

Для предсказания значений плотности колоса мы провели отбор переменных, характеризующих морфологию колоса, по нескольким критериям (как это описано в разделе. 4.5). Переменные были ранжированы по итогам усреднения 8 характеристик значимости. После этого, на обучающей и тестовой выборках итеративно, в порядке уменьшения среднего значения значимости по одной добавлялись переменные. Регрессия осуществлялась методом случайного леса. В процессе добавления переменных оценивалось значение MAE для тестовых и обучающих данных.

Обучение моделей осуществилось с использованием 12 наиболее значимых признаков. Количество признаков для обучения было выбрано на основании кривых обучения (рис. 4.12), показывающих зависимость средней абсолютной ошибки предсказания индекса плотности колоса и числом лучших признаков, которые используются для обучения методом случайного леса.

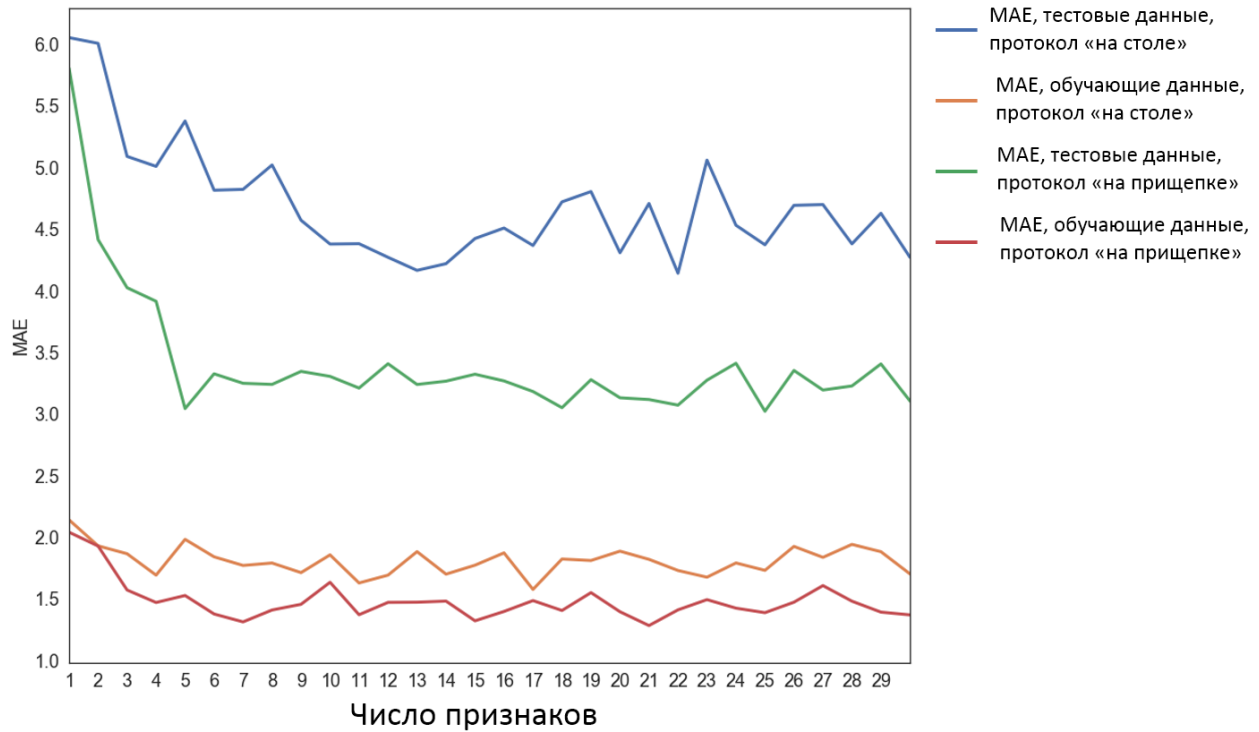


Рисунок 4.12. Зависимость средней абсолютной ошибки предсказания индекса плотности колоса и числом лучших признаков, которые используются для обучения методом случайного леса.

Мы выполнили ранжирование признаков по значимости, при решении задачи предсказания индекса плотности колоса. Результаты отбора лучших 12 признаков для протоколов «на столе» (одна проекция) и «на прищепке» (4 проекции) представлены в табл. 4.3 и 4.4, соответственно.

Таблица 4.3. Оценка мер значимости количественных характеристик колоса для предсказания плотности на основе анализа изображений, полученных по протоколу «на столе». Название характеристик приведено в колонке 1, величины значимости в колонках 2-9, их среднее значение в последней колонке.

Параметр	LinearRegression	Ridge	Lasso	Stability	RFE	RF	Corr	MIC	Среднее
L	0.93	0.07	0.04	0.74	0.81	1	1	1	0.7
R	0	0.75	0	0.68	1	0.06	0.65	0.93	0.51
x_{u2}	1	0.1	0.11	0.61	0.73	0.07	0.5	0.88	0.5
t_{b1}	0	1	1	0.04	0.95	0.01	0	0.34	0.42
P	0	0.07	0.07	1	0.66	0.06	0.49	0.71	0.38
<i>profile_8</i>	0	0.51	0.55	0.86	0.88	0.03	0.07	0.13	0.38
y_{b1}	0	0.7	0.7	0.21	0.92	0.05	0.04	0.37	0.37
y_{u1}	0	0.82	0.78	0.02	0.75	0.03	0.01	0.41	0.35
<i>profile_1</i>	0	0.46	0.45	0.44	0.71	0.18	0.07	0.51	0.35
<i>radial_3</i>	0	0.01	0.01	0.92	0.2	0.21	0.31	0.92	0.32
y_{b2}/L	0	0	0	0.48	1	0.13	0.33	0.6	0.32
C	0	0.2	0	0.34	1	0.05	0.36	0.64	0.32

Таблица 4.4. Оценка мер значимости количественных характеристик колоса для предсказания плотности на основе анализа изображений, полученных по протоколу «на прищепке». В скобках указана проекция колоса, с которой была получена характеристика. Название характеристик приведено в колонке 1, величины значимости в колонках 2-9, их среднее значение в последней колонке.

Параметр	LinearRegression	Ridge	Lasso	Stability	RFE	RF	Corr	MIC	Среднее
x_{b2} (проекция 2)	1	0.15	0.13	0.75	1	0	0.46	0.68	0.52
x_{u2} (проекция 2)	0.98	0.06	0.04	0.87	0.9	0.04	0.63	0.63	0.52
C (проекция 3)	0	0.24	0	0.81	0.53	1	0.62	0.94	0.52
y_{b1}/L (проекция 3)	0	0.01	0	0.9	0.59	0.86	0.86	0.8	0.5
R (проекция 3)	0	0.38	0	0.94	0.64	0.01	0.84	0.96	0.47
L (проекция 4)	0.42	0.06	0.01	0.5	1	0.01	0.79	0.79	0.45
R (проекция 4)	0	0.32	0	0.88	0.52	0.01	0.81	1	0.44
y_{u1}/L (проекция 3)	0	0.01	0	0.94	0.71	0.02	1	0.71	0.42
L (проекция 3)	0.11	0.08	0.02	0.26	1	0.2	0.73	0.76	0.4
y_{u1}/L (проекция 4)	0	0.01	0	1	0.48	0.01	0.84	0.8	0.39
R (проекция 2)	0	0.19	0	0.8	0.49	0.03	0.63	1	0.39
x_{u2} (проекция 1)	0.58	0.05	0	0.52	0.73	0.03	0.43	0.65	0.37

Признаками с высокой значимостью для обоих протоколов оказываются длина и индекс округлости колоса. Интересным является то, что для протокола «на прищепке» присутствуют эти признаки для нескольких различных проекций.

Таблица 4.5. Оценки качества предсказания индекса плотности колоса для четырех (протокол «на прищепке») и одной (протокол «на столе») проекций. Приведены MAE, MAPE и квадрата коэффициента корреляции Пирсона для тестовых и обучающих выборок.

Метрика точности	Протокол «на прищепке»	Протокол «на столе»
MAE (обучающая выборка)	1.48	1.88
MAE (тестовая выборка)	3.33	4.61
MAPE (обучающая выборка)	6.19	7.77
MAPE (тестовая выборка)	13.28	17.80
R^2 (обучающая выборка)	0.95	0.94
R^2 (тестовая выборка)	0.75	0.52

Оценки точности предсказания индекса плотности приведены в табл. 4.5. Средняя абсолютная ошибка предсказания индекса плотности на тестовой выборке (MAE test) для одной проекции равна 4.61, для четырех проекций 3.33. Коэффициент корреляции Пирсона между предсказанными значениями и значениями определенными экспертом для тестовой выборки (R^2 test) равен 0.51 для одной проекции и 0.74 для четырех проекций. Обучение моделей классификации методами логистической регрессии (LinearRegression) и случайного леса (RF).

Таблица 4.6. Оценки точности предсказания формы колоса для одной (протокол «на столе») и четырех (протокол «на прищепке») проекций. « F /лог. Регрессия» и « F / случ. лес» – F мера точности классификации колосьев по форме методами логистической регрессии и случайного леса; « F /лог.регрессия+индекс плотности» и « F / случ.лес+индекс плотности» – F мера точности классификации при добавлении дополнительного признака “предсказанный индекс плотности” методами логистической регрессии и случайного леса.

Метрика точности/метод предсказания	Протокол «на прищепке»	Протокол «на столе»
F /лог. регрессия	0.82	0.78
F /лог.регрессия+индекс плотности	0.83	0.77
F /случ. лес	0.85	0.72
F /случ.лес+индекс плотности	0.84	0.76

Оценки точности предсказания типа колоса приведены в табл. 4.6. Мы пробовали в качестве дополнительного признака добавлять предсказанный индекс плотности (лог.регрессия+индекс плотности, случ.лес+индекс плотности), однако не получили выигрыша в точности. Значения меры качества классификации F формы колоса лучшего метода на тестовой выборке для 4 проекций равна 0.85 (F /случ. лес), для одной проекции 0.78 (F /лог. регрессия). Приведенные оценки точности показывают, что использование информации о четырех проекциях увеличивают точность предсказания формы на 7% и уменьшают среднюю абсолютную ошибку при предсказании индекса плотности колоса на 1.28, что соответствует 4.25%.

В табл. 4.7 представлена оценка значений матрицы ошибок на основании классификации формы колоса на тестовой выборке для четырех проекций,

полученная с использованием метода случайного леса. Видно, что с наименьшим количеством ошибок классификации разделяются колосья типа «компактный» и «спельта». Колосья же типа «нормальный» наиболее часто предсказываются как относящиеся к двум другим типам (самое низкое качество классификации). Интересно отметить, что колосья, относящиеся к типам «нормальный» и «спельта» дают наибольшее количество взаимных ложных предсказаний. Эти результаты согласуются с гистограммами на рис. 4.11, которые построены на основании экспертной аннотации типов колосьев.

Таблица 4.7. Матрица ошибок для классификации формы колоса, полученной на тестовой выборке для изображений 4 проекций (протокол ‘на прищепке’).

	Компактный	Нормальный	Спельта
Предсказанный «Компактный»	158	13	0
Предсказанный «нормальный»	8	79	20
Предсказанный «спельта»	0	20	77

Таким образом, в итоге проведенной работы предложены подходы для автоматической классификации колосьев на три типа по геометрическим параметрам и оценена точность этих методов. Полученные результаты показали, что параметры, характеризующие модель четырехугольников, оказываются наиболее информативными для классификации колосьев, в сравнении с моделями сечений и радиальной моделью.

4.8. Заключение по главе 4

В настоящей главе предложен метод позволяющий:

- распознавать образ колоса на изображениях, полученных по двум стандартным протоколам;
- отделять на изображении ости от тела колоса;
- оценивать количественные характеристики остистости колоса (площадь остей на изображении, количество и среднюю длину остей);
- количественные характеристики формы колоса, как интегральные.

Так и описывающие форму в виде специальных моделей.

Анализ точности выделения области колоса показал, что наиболее сильно влияет на точность сегментации фон/колос и тело колоса/ости масштаб изображения: чем дальше от объектива камеры расположен колос, тем выше ошибка сегментации. Масштаб съемки существенно различался в зависимости от протокола получения изображений. В протоколе «на столе» камера находится значительно ближе к снимаемому объекту. В протоколе «на прищепке» расстояние от объектива до объектов также могло быть различное. При условии, что фокусное расстояние камеры было фиксировано, масштаб зависел только от расстояния до объекта съемки.

Цветокоррекция изображения существенно не влияла на точность сегментации. Вероятнее всего это явилось следствием стандартизованного освещения при съемке колоса. Тем не менее, возможность цветокоррекции, заложенная в наш метод, позволит с большей объективностью оценивать цвета на изображениях и поможет существенно улучшить результаты обработки в условиях слабо контролируемого освещения.

Меньшее среднее значение J_a по сравнению с J_b отчасти связано с малой площадью остей относительно всего колоса и одновременно с большим количеством граничных пикселей для областей остей. Таким образом, цена

ошибки на стадии бинаризации значительно выше для области остей, чем для области “тела” колоса. Кроме того, само ручное маркирование изображений высокого разрешения является трудоемким, и не позволяет до конца избежать флуктуации при выделении границ колоса.

Нами впервые был предложен способ оценки параметров формы колоса. Они условно могут быть разделены на «общие» характеристики, которые отражают форму целиком (например, округлость, площадь колоса, длина). Кроме того, мы дополнительно ввели три специфические модели для описания формы колоса. Две из них (модель сечений и радиальная) были сформулированы из самых общих соображений и их параметры не несут какой-либо специфики типичной формы колосьев пшеницы. Третья модель, описывающая колос в виде двух четырехугольников с общей стороной, может служить более наглядным обобщением формы колоса. То, что эта модель является наиболее адекватной, свидетельствует наличие существенного количества параметров именно этой модели среди параметров, которые значимы для предсказания плотности колоса. Среди значимых параметров преобладают общие характеристики колоса и параметры модели четырехугольников.

Мы оценили точность предсказания плотности колоса и его типа, на основе количественных характеристик, полученных при анализе изображений. Мы выбрали два метода регрессии, логистическую и метод случайного леса, поскольку они не так требовательны к объему обучающей выборки (в нашем случае ~200 колосьев), как нейронные сети или нейронные сети глубокого обучения, когда хороший результат достигается при анализе тысяч и десятков тысяч изображений.

Предсказание плотности колоса с помощью метода логистической регрессии показало, что величина ошибки MAE составляет 4.61 (при

использовании изображения одной проекций колоса). Если использовать для предсказания характеристики, полученные для четырех проекций колоса, то точность предсказания возрастает ($MAE=3.33$), несмотря на одно и то же количество признаков, использованных для предсказания (однако эти 12 признаков включают характеристики, полученные в результате анализа разных проекций колоса).

Точность классификации колосьев на три типа (спельтоидный, нормальный, компактный) с помощью алгоритма случайного дерева (мера F) составила 0.78 для одной проекции колоса и 0.85 для четырех проекций колоса. Интересно, что последнее значение близко к оценке классификации колосьев на 4 сорта, проведенной в работе Би и соавт. (Bi et al. 2010), 0.88 (использовался алгоритм нейронных сетей и 12 входных параметров). Сложностью при решении этой задачи может являться то, что форма колосьев типов спельта и нормальный могут быть похожими, о чем в частности свидетельствует сильное перекрытие распределений плотности колоса (в особенности для колосьев «спельта» и «нормальный», рис. 4.11). Более точному решению проблемы в дальнейшем могут способствовать накопление большего количества аннотированных изображений и использование методов нейронных сетей глубокого обучения.

Полученная в результате анализа информация о морфометрических характеристиках колоса может быть применена в высокопроизводительном, автоматизированном фенотипировании и при проведении селекционных экспериментов.

ГЛАВА 5. БАЗЫ ДАННЫХ ДЛЯ ОПИСАНИЯ ФЕНОТИПА ПШЕНИЦЫ

Методы современной феномики, которые способны быстро, массово и точно оценивать признаки растений позволяют проводить анализ большого количества растений в современных экспериментах по оценке связи между генотипом фенотипом и окружающей средой. В этой связи возникает актуальная задача сохранения полученных данных, поскольку традиционные способы (полевые дневники, лабораторные журналы и даже цифровые таблицы) оказываются неэффективными. Поэтому одной из актуальных задач феномики растений является создание систем хранения фенотипических данных и баз данных (БД), которые будут способствовать ускорению процесса сбора фенотипических данных.

В настоящей главе мы предлагаем результаты работ по созданию баз данных в области феномики пшеницы, направленных на обеспечение проведения высокопроизводительного фенотипирования в ходе селекционно-генетических экспериментов.

Эти работы направлены на сбор данных по параметрам фенотипа растения в процессе эксперимента, для того чтобы сохранить информацию о сроках фаз развития, агротехнических мероприятиях, информации о климатических условиях в месте его произрастания.

5.1. База данных WheatPGE

Модель данных

В процессе селекционно-генетического эксперимента осуществляется сравнении фенотипических характеристик растений, имеющих разные генотипы, условия произрастания, либо воздействия различных факторов среды. В процессе эксперимента (рис. 5.1):

1) осуществляется планирование (выбор генотипов для исследования, места и условий произрастания);

- 2) производится посев семян;
- 3) для контроля метеоусловий производится их мониторинг в месте произрастания растений (влажность воздуха, направление и сила ветра, температура воздуха и почвы, количество осадков, освещенность);
- 4) производятся агротехнические мероприятия (прополка, внесение удобрений и т.п.);
- 5) производится оценка сроков основных фаз роста и развития растений;
- 6) оцениваются фенотипические характеристики растений;
- 7) производится сбор урожая;
- 8) производится обработка результатов наблюдений.

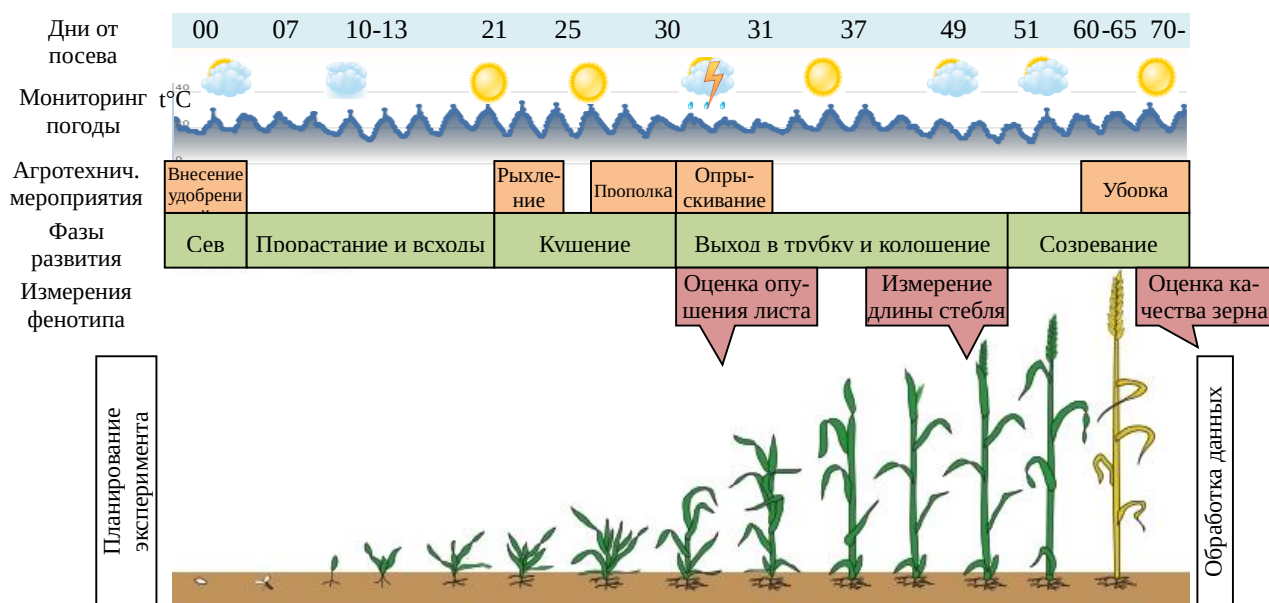


Рисунок 5.1. Общая схема проведения селекционно-генетического эксперимента у пшеницы.

Следует отметить, что при проведении экспериментов на современном

этапе фокус исследований смещается от исследований групп растений (делянок) смещается в направлении исследований отдельных растений. Растение характеризуется уникальным генотипом, фенотипом и местом произрастания. Анализ популяции отдельных растений также позволяет оценить вариации фенотипических признаков, что немаловажно при оценке влияния на эти признаки как генотипа, так и окружающей среды. Поэтому в базе данных целесообразно рассматривать центральным объектом растение. Для него описываются три основных блока информации – (1) генотип, (2) фенотип, (3) место произрастания.

Описание генотипа должно включать такие важные характеристики как индивидуальное название сорта или линии, а также известные для него молекулярные маркеры или гены морфологических признаков. Описание для набора молекулярных маркеров позволит в перспективе также обеспечить анализ экспериментов по картированию количественных признаков (QTL).

Развитие современных технологий фенотипирования позволяет оценивать большое количество разнообразных характеристик растений, с появлением новых технологий и сенсоров. Поэтому в базе данных необходимо учесть возможность расширения описания фенотипа растений за счет новых характеристик. Отметим, что в основе большинства методов фенотипирования растений лежат технологии получения изображений. Поэтому база данных должна иметь возможность сохранять информацию об изображениях в цифровом формате. Измерение параметров фенотипа может производиться в ходе эксперимента как для одного растения, так и для группы растений. Например, для получения экспериментальных образцов при оценке свойств зерна и муки требуется использовать несколько растений, что необходимо учесть при формировании логической структуры базы данных.

В настоящей работе предлагаемый подход был реализован для описания

фенотипа пшеницы в БД WheatPGE (Wheat Phenotype, Genotype, Environment) для пшеницы.

Логическая модель данных включает таблицу растения, связанную с четырьмя блоками информации – генотипом, фенотипом, местом произрастания и экспериментом.

Всего в текущей версии БД содержит 32 таблицы и 55 отношений между ними. Список таблиц и их краткое описание приведены в Приложении, табл. П9. Таблица, описывающая растение содержит список идентификаторов таблиц, описывающих его генетические, фенотипические признаки, место произрастания и условия среды (табл. 5.1). Индексируемым является поле id. Для того, чтобы в базе данных можно было бы работать с признаками, которые идентифицируются для нескольких растений (например, делянка), мы добавили для растения специальное поле: единичное растение («single plant»). В случае анализа единственного растения это поле содержит 1. Если экспериментатор рассматривает при анализе совокупность из нескольких растений и их признаки, то поле содержит 0. При этом для группы растений сохраняются все отношения в базе данных, в том числе и описание характеристик генотипа (считается что это генетически однородная группа растений).

Генотип растения описывается девятью таблицами (рис. 5.2). Они включают описание сорта растения или линии. Генотип связан с рядом таблиц, описывающих молекулярные маркеры. Такая привязка позволяет документировать эксперименты на пшенице, связанные с проведением эксперимента по анализу количественных признаков.

Фенотип растения описывается 15 таблицами (рис. 5.3): таблица, описывающая базовые признаки растения (длина стебля, число колосьев, общую урожайность); таблицы, описывающие структуру урожая (главный и

вторичные колосья); таблицы, описывающие опушение листа; таблица, описывающая длительность стадий развития растения.

Место произрастания описано двумя таблицами, основная информация в которых содержит название, широту, долготу, тип климата, климатические характеристики (среднегодовую температуру, среднегодовую влажность, средние температуры января и июля) (рис. 5.4).

Таблица 5.1. Описание полей таблицы plant (растение) в БД WheatPGE.

Наименование атрибута	Тип данных	Краткое описание
id	int(11)	Идентификатор объекта
sowing_number	varchar(16)	Посевной номер
genotype_id	int(11)	Идентификатор генотипа
phenotype_id	int(11)	Идентификатор фенотипа
environment_id	int(11)	Идентификатор окружения
single_plant	int(11)	Уникальное растение или группа
add_user_id	int(11)	Идентификационный номер пользователя, добавившего запись
edit_user_id	int(11)	Идентификационный номер пользователя, редактировавшего запись
date	date	Дата создания
sowing_date	date	Дата посева

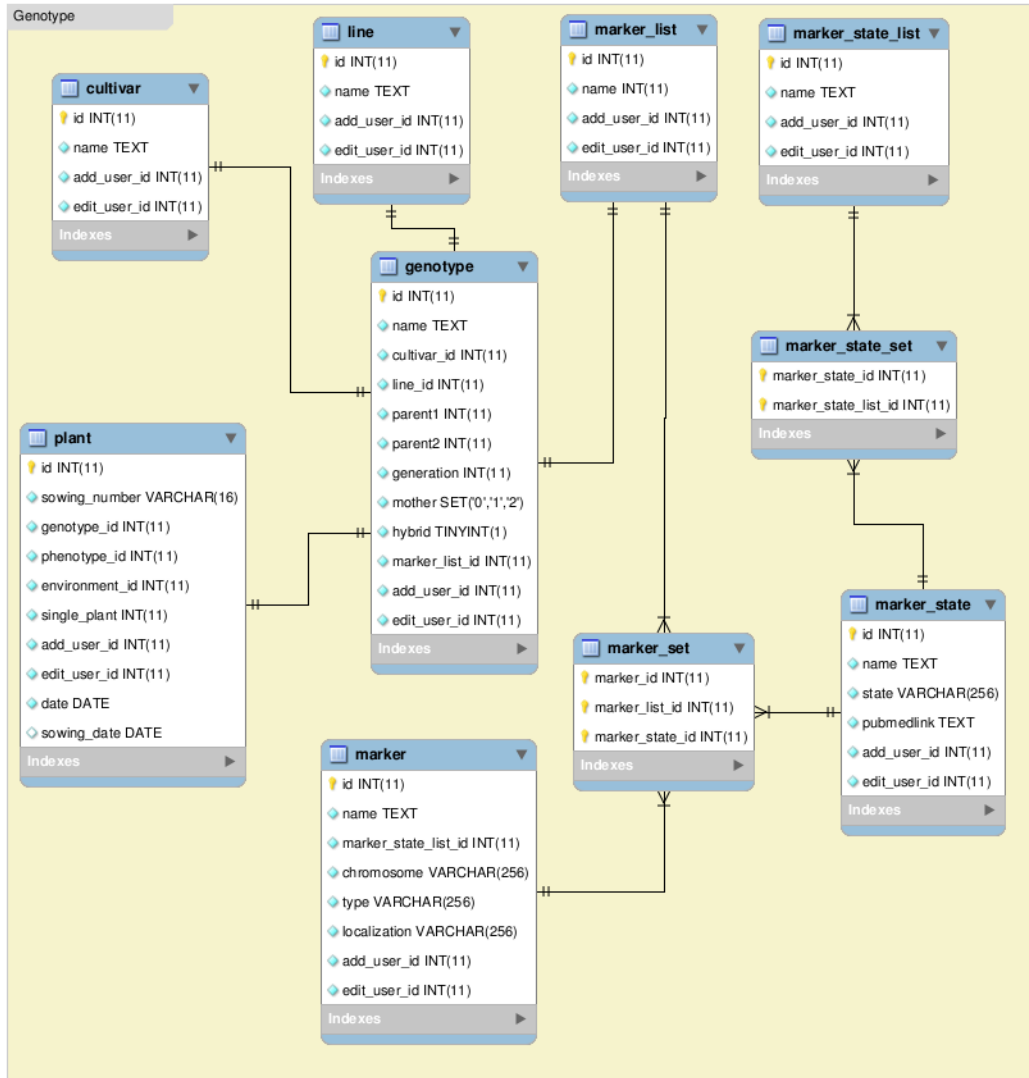


Рисунок 5.2. Таблицы БД WheatPGE и связи между ними, описывающие генотип растения.

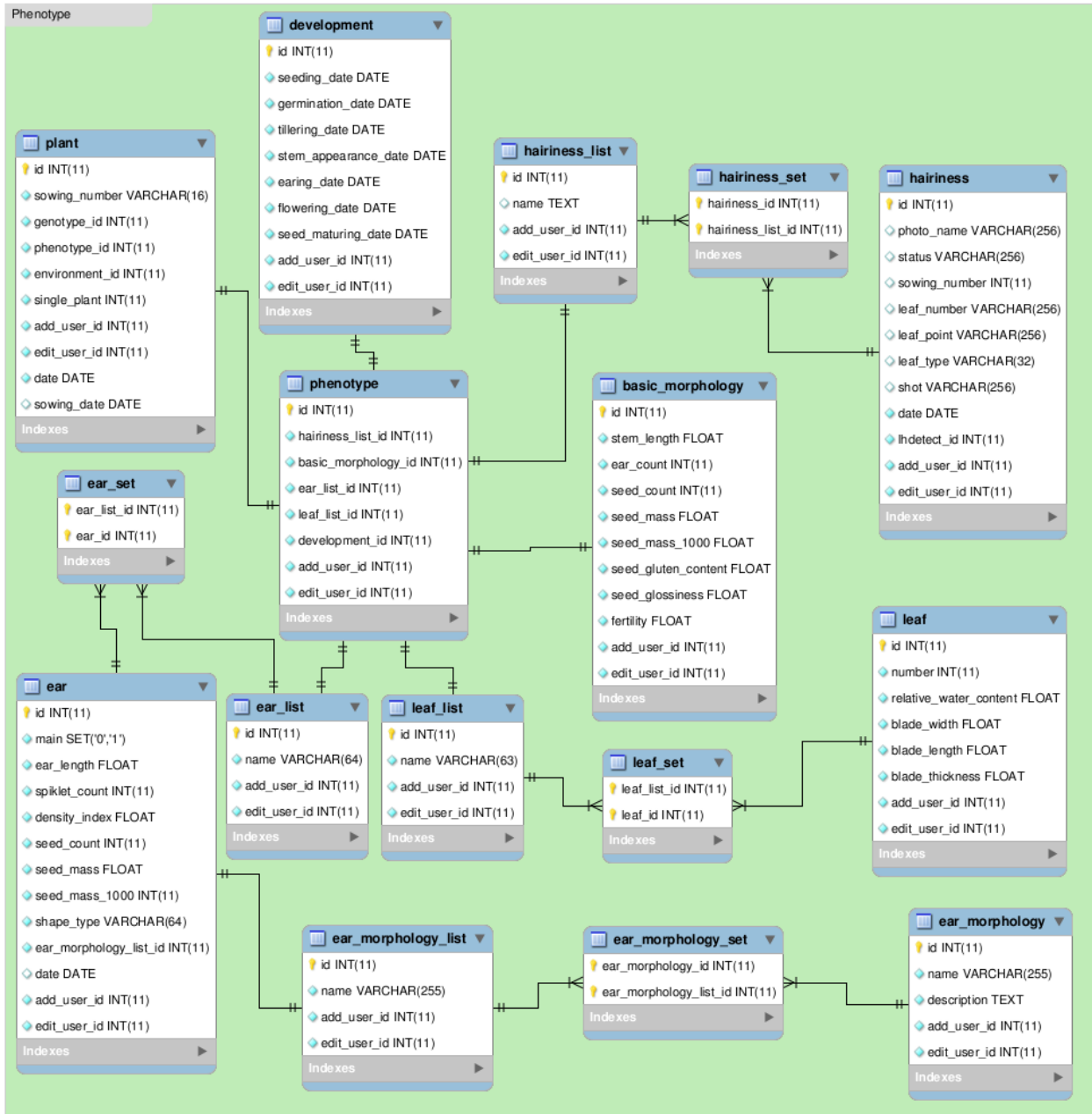


Рисунок 5.3. Таблицы БД WheatPGE и связи между ними, описывающие фенотип растения.

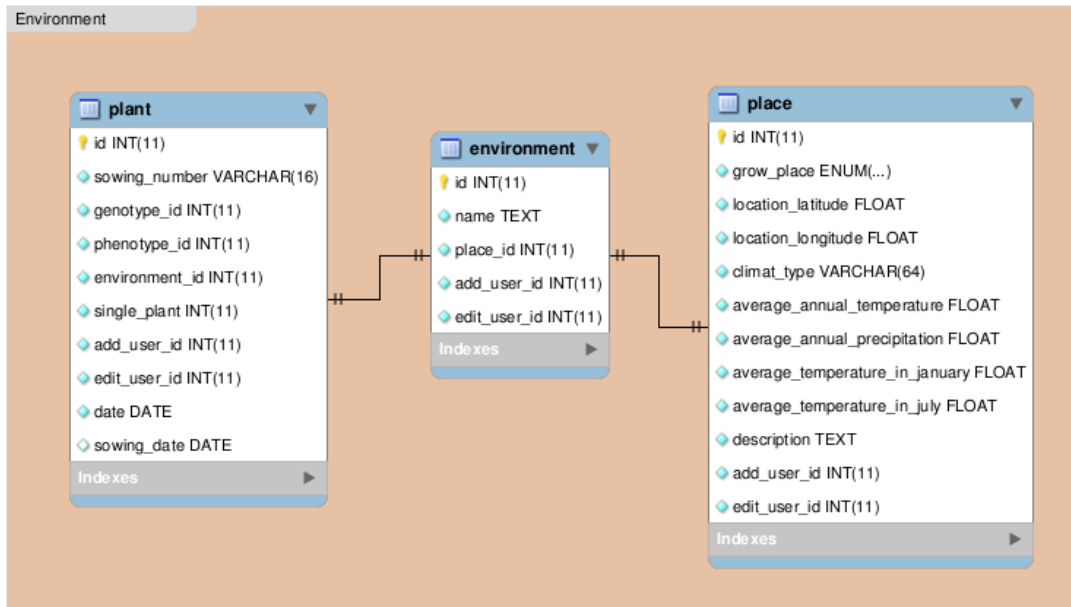


Рисунок 5.4. Таблицы БД WheatPGE и связи между ними, описывающие место произрастания.

Блок информации, связанный с проведением эксперимента, содержит две таблицы (рис. 5.5), описывающих событие и их список для растения. Событие содержит поле названия, типа и значения. Событие описывает проведение измерений над фенотипом, агротехнические мероприятия, или изменение фазы развития растений. Блок информации, описывающий данные пользователя, содержит три таблицы с информацией, позволяющей управлять уровнем доступа пользователя к БД WheatPGE (рис. 5.6).

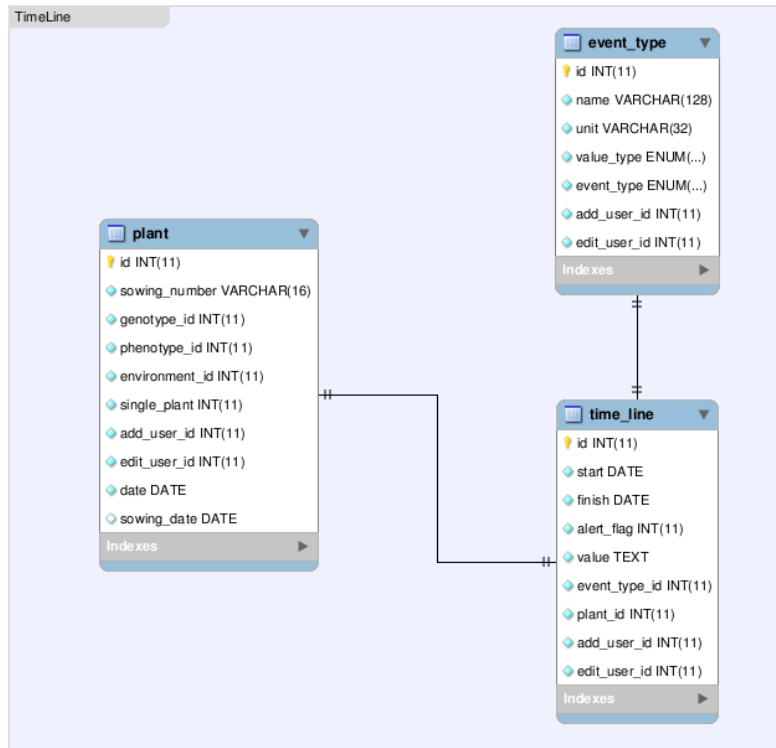


Рисунок 5.5. Таблицы БД WheatPGE и связи между ними, описывающие события в процессе проведения селекционно-генетического эксперимента.

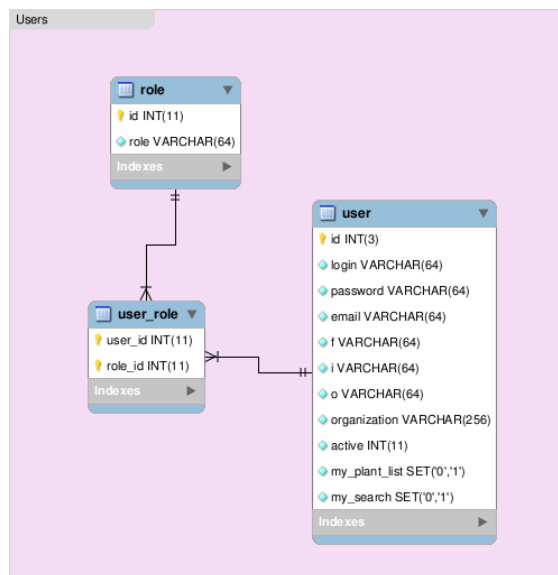


Рисунок 5.6. Таблицы БД WheatPGE и связи между ними, описывающие пользователей.

Технологии реализации БД WheatPGE

Для обеспечения интеграции базы данных и различных методов массового фенотипирования растений, взаимодействие пользователя и БД разработан модуль интерфейса WheatPGE. Интерфейс основан на использовании методологии MVC (Рогачев, 2008) для разработки программного обеспечения. MVC позволяет разделить представление модели данных, интерфейса пользователя и систему управления приложением в виде трех независимых компонентов. Таким образом, изменение каждого из этих компонентов позволяет не затрагивать другие компоненты при изменении кода: трудозатраты на разработку приложения существенно снижаются. Это важно при разработке программ и БД в слабоструктурированных областях знаний, которой является биология.

Технологии MVC позволяют пользователю добавлять в базу произвольные морфологические признаки и параметры окружающей среды без помощи программиста. При этом происходит автоматическое расширение схемы базы данных, создается новая модель, описывающая объекты этого признака. Контроллеры и представления данных, обеспечивающие основные операции при работе с признаком (создание, удаление, редактирование), генерируются автоматически. Мы использовали реализацию MVC на платформе Catalyst (<http://www.catalystframework.org/>).

База WheatPGE содержит описание объектов со сложной структурой. Так, например, описание генотипа составное и включает ряд других объектов, он также является характеристикой объекта «растение». Для представления объектов и отношений между ними в базе данных использована технология реляционной системы управления базой данных (СУБД) наравне с технологиями объектно-реляционного отображения (ORM – object relation mapping). Использование реляционной модели представления данных

обеспечит надежность и производительность, а объектное представление обеспечивает эффективный доступ к данным системы. Для объектно-реляционного представления используется библиотека DBIx (<http://search.cpan.org/dist/DBIx-Class/>), которая обеспечивает работу с базой данных через объектно-ориентированный интерфейс. В качестве СУБД использована MySQL 5, развернутая на сервере баз данных под управлением CentOS Linux.

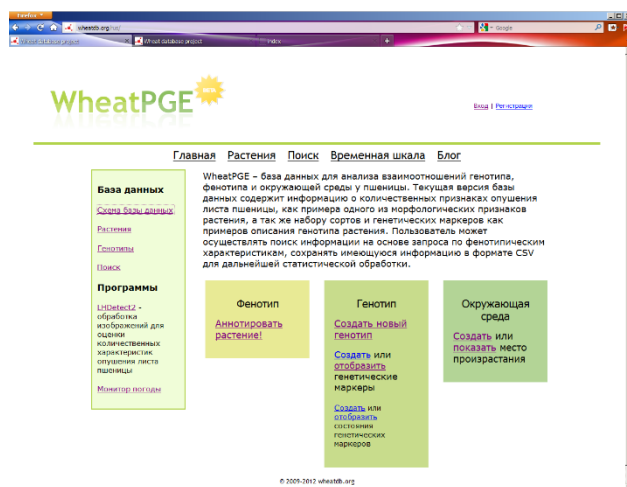
Для визуализации последовательности и времени измерений и агротехнических мероприятий в интерфейсе БД WheatPGE используется библиотека Highstock (<http://www.highcharts.com/products/highstock>). Это библиотека графической визуализации информации, которая зависит от времени (графики, события и т.п.). Она является свободной для некоммерческого использования. Библиотека разработана с использованием языка JavaScript и работает в любом современном интернет браузере, в том числе и на мобильных устройствах.

Интерфейс БД WheatPGE

Главная страница базы данных WheatPGE располагается по адресу <http://wheatdb.org/rus/>. Она содержит краткую информацию о базе данных, ссылки для входа в систему или регистрации, ссылки на основные блоки информации в этой базе данных (рис. 5.7 А).

Для того чтобы просмотреть список растений, информация о которых доступна в базе, необходимо с главной страницы перейти по ссылке «Растения». После этого осуществляется переход на страницу списка растений (рис. 5.7 Б). На этой странице отображается информация о растениях в виде таблицы. Информация различного типа (фенотипические признаки, информация о месте произрастания, генотипе) располагается в разных колонках таблицы и доступна по соответствующим гиперссылкам.

А)



Б)

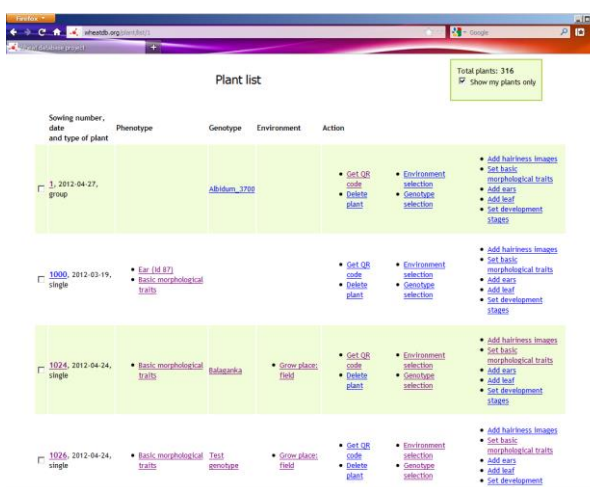


Рисунок 5.7. Интерфейс БД WheatPGE. А) Главная страница доступа к БД; Б) Визуализация списка растений, созданного пользователем.

В крайней правой колонке доступны ссылки для добавления (редактирования) информации о генотипе, фенотипе, месте произрастания. Для каждого растения приводятся ссылки на его QR-код и ссылка для удаления растения из базы данных. Ряд операций доступен для нескольких растений одновременно. Для этого их нужно отметить галочкой слева и внизу страницы выбрать одну из возможных кнопок.

Пользователь может добавить информацию о растении в базу данных. Пример операций по созданию нового растения и занесения информации о его фенотипе – свойствах колоса – приведен на рис. 5.8. Вначале нужно перейти с главной страницы базы по ссылке «Аннотировать растение!» на страницу добавления растения. Появится страница создания новой записи «plant», на которой нужно указать его посевной номер и дату посева.

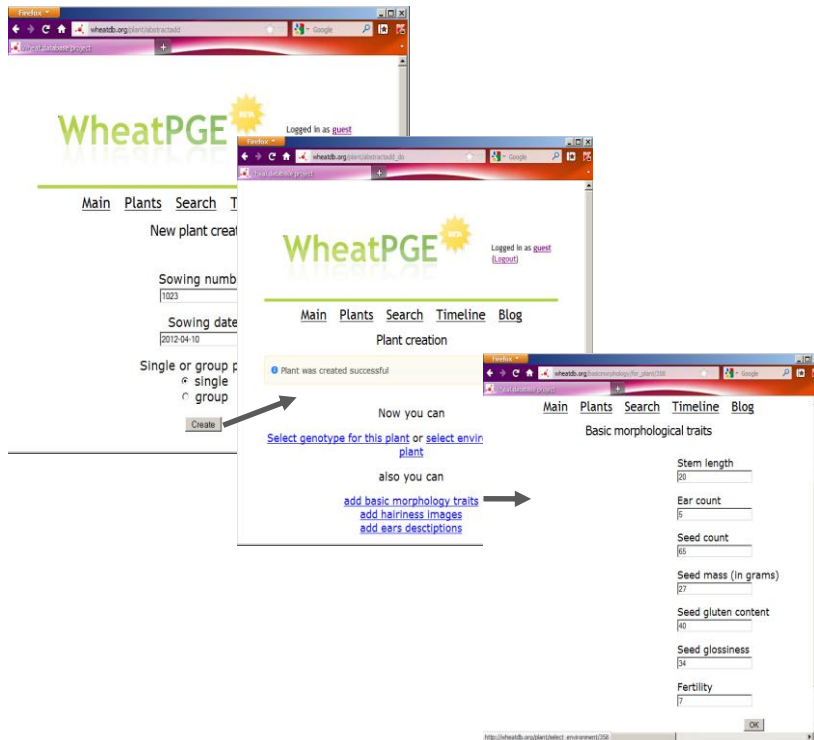


Рисунок 5.8. Добавление информации о растении и его базовых фенотипических признаках.

После нажатия кнопки «Create» осуществляется создание нового растения и переход на страницу его аннотации. На этой странице можно выбрать несколько основных операций: выбрать генотип для растения, выбрать место произрастания, добавить базовые морфологические признаки, признаки урожайности, опушения листа и др. Выбор этих операций осуществляется переходом по соответствующим гиперссылкам, указанным на этой странице. Таким образом, система позволяет вносить информацию о растении и его признаках в базу данных.

Интерфейс базы данных позволяет осуществлять поиск информации о растениях. Для этого необходимо с главной страницы перейти по ссылке «Search» в меню верхнего уровня. При этом откроется страница запроса, показанная на рис. 5.9.

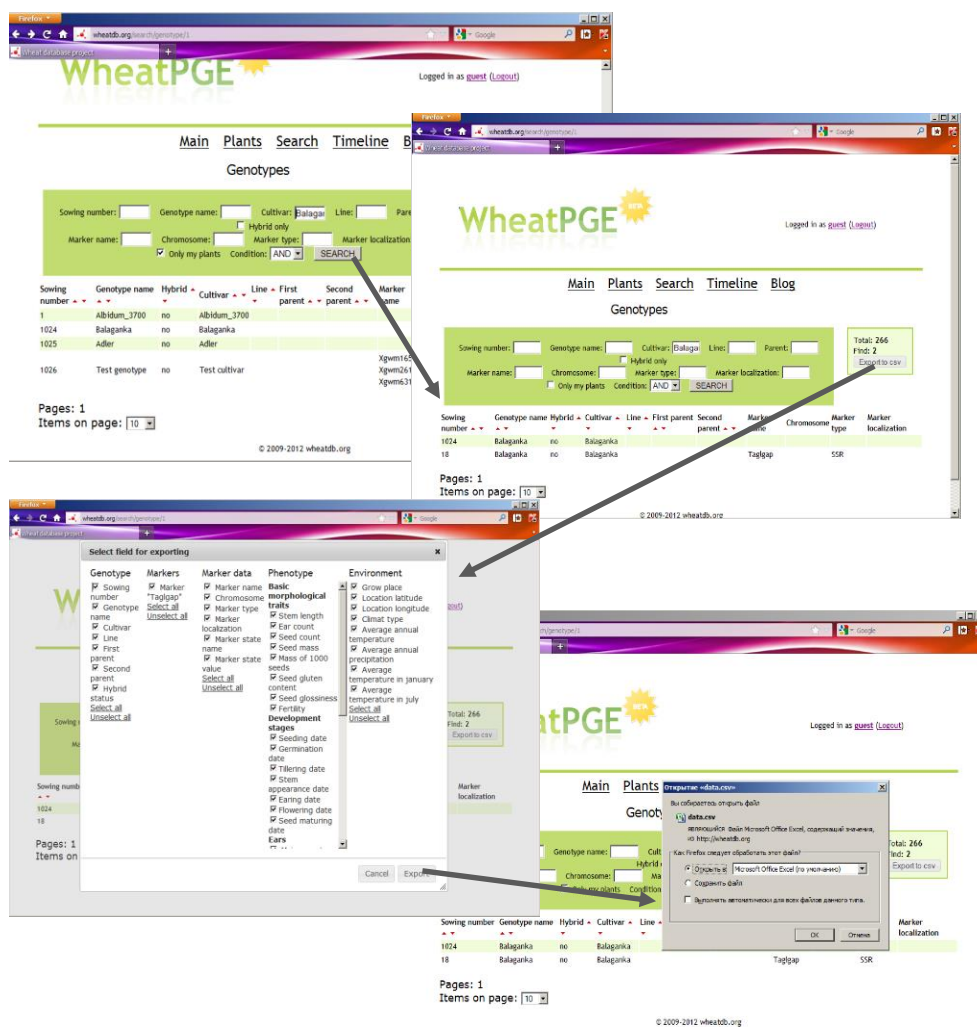


Рисунок 5.9. Поиск и экспорт информации из БД WheatPGE.

На этой странице необходимо ввести в одно из полей или в несколько поля базы данных термины для запроса и нажать кнопку «Search». После этого под панелью запроса появится список растений, удовлетворяющих условию запроса. Информацию о признаках растения можно сохранить в CSV файле, нажав на кнопку «Export to CSV» в верхнем правом углу страницы. При этом появится панель выбора полей базы данных для сохранения. Пользователь может отобразить поля, отметив их галочкой. После нажатия кнопки «Export» на этой панели, система предложит файл на сохранение.

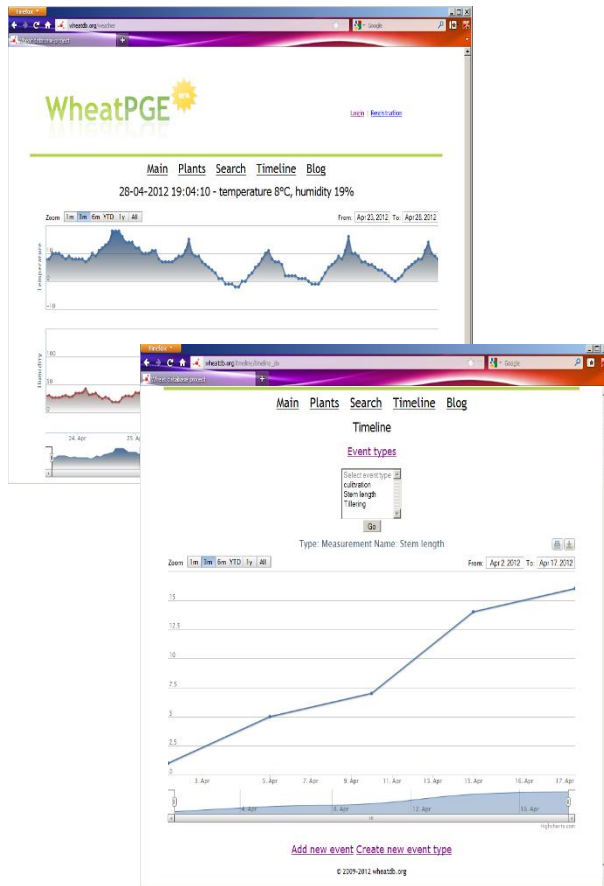


Рисунок 5.10. Примеры отображения событий на шкале времени. В левом верхнем углу - монитор погоды на поле вблизи ИЦиГ СО РАН. В правом нижнем углу – отображение измерения длины стебля растения.

Примеры отображения событий на шкале времени приведены на рис. 5.10. На нем представлены визуализации информации температуры и влажности с датчика, установленного вблизи поля ИЦиГ СО РАН, а также результаты измерения длины стебля для одного из растений в процессе его роста.

Для удобства взаимодействия с базой данных при идентификации растений мы используем систему QR кодов (<http://www.qrcode.com/index-e.html>). QR-код является матричным (двумерным) штрих-кодом, который может быть сканирован камерой мобильного устройства. Информация,

которую он содержит, может содержать до 4296 символов цифр и букв, что достаточно для описания доступа к растению в нашей базе по ссылке <http>. QR-код присваивается в нашей базе каждому растению, может быть распечатан на плотной бумаге и прикреплен к его стеблю. В дальнейшем при измерении параметров растения в процессе эксперимента достаточно считать этот код, открыть в браузере ссылку для этого растения и занести параметры в базу.

Таким образом, интерфейс БД WheatPGE позволяет сохранять информацию об измерениях фенотипов растений в ходе эксперимента непосредственно в базу данных как со стационарного компьютера в лаборатории, так и в поле, минуя записи в журналах.

5.2. База данных SpikeDroid DB

Для сбора, хранения и анализа информации о фенотипических характеристиках колоса пшеницы нами была разработана компьютерная информационная система SpikeDroidDB. Эта БД позволяет структурированно хранить цифровые изображения колоса, производить их аннотацию и предоставляет пользователю гибкую систему для доступа к данным.

Модель данных

Характеристики колоса, которые мы использовали для описания его фенотипа приведены в табл. 5.2. Поскольку фенотип растения формируется под влиянием окружающей среды и генотипа в системе SpikeDroidDB определен ряд понятий, важных при проведении селекционно-генетических исследований, и отношения между ними. Логическая модель данных включает таблицу “растения”, связанную с тремя блоками информации – коллекцией, окружающей средой и фенотипом колоса. Текущая версия базы данных содержит 5 таблиц и 4 отношения между ними (рис. 5.11).

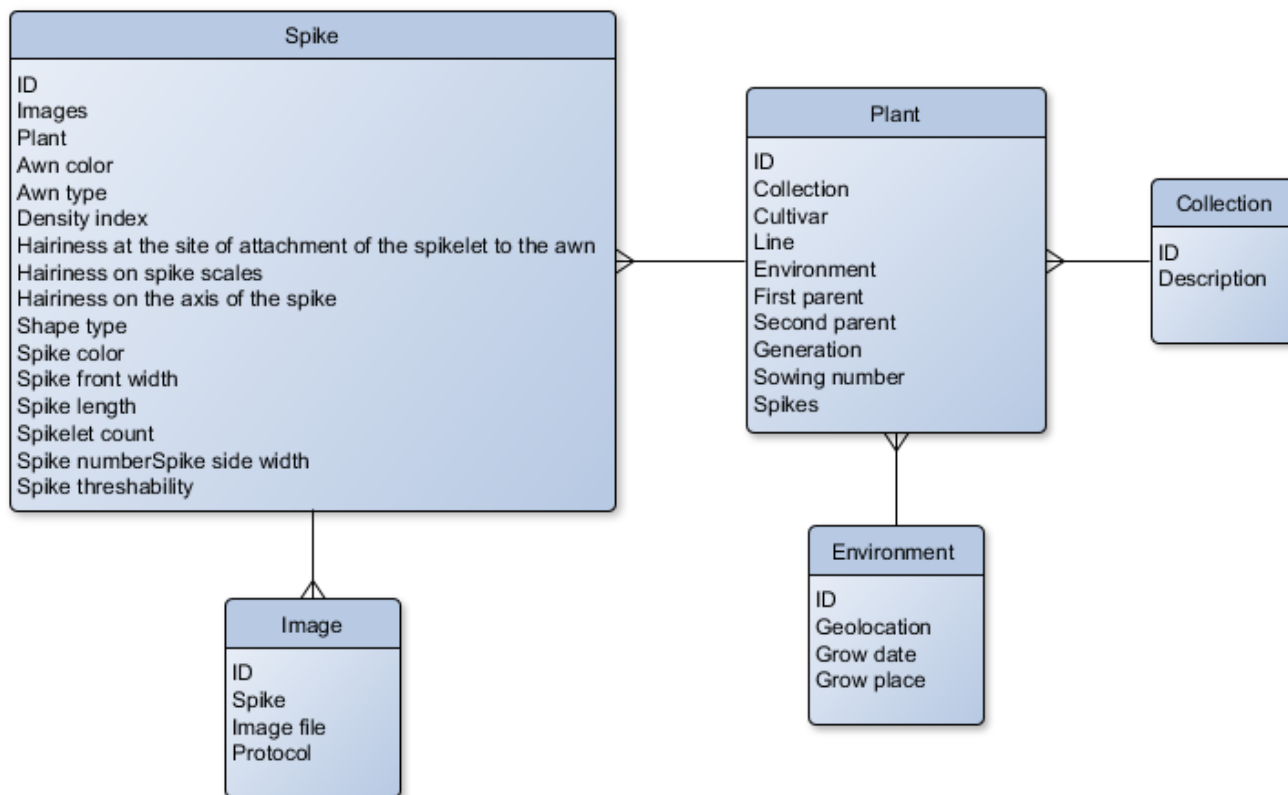


Рисунок 5.11. Модель объектов и отношений базы данных SpikeDroidDB.

Таблица Plant (Растение) включает описание сорта (образца) растения или линии, ссылки на родительские растения, поколение и посевной номер растения. Растение связано с таблицами Collection (Коллекция), Environment (место произрастания) и Spike (описание колоса). Для коллекции указывается её держатель и аффилиация (описание). Место произрастания содержит геоданные, дату вегетации, фенологические данные и данные об условиях выращивания растения. Фенотип колоса описывается набором из 14 признаков, представленных в табл. 5.2. Для признаков колоса пшеницы, содержащийся в нашей базе данных была выполнена привязку к признакам онтологии сельскохозяйственных растений портала CropOntology.org, в тех случаях где это представлялось возможным.

Таблица 5.2. Описание фенотипа колоса.

Наименование атрибута	Тип данных	Краткое описание	Термин онтологии CropOntology.org
Spike number	int	Номер колоса в растении, первый номер присваивается главному колосу	
Spike length	real	Длина колоса, единица измерения: см	CO_321:000056
Spike front width	real	Ширина лицевой стороны колоса, единица измерения: см	
Spike side width	real	Ширина боковой стороны колоса, единица измерения: см	
Spikelet count	int	Количество колосков, единица измерения: штука	CO_321:000058
Density index	real	Плотность колосков. $D = [(A-1) \times 10] / B$ Где: (A-1) – число колосков в колосе без верхушечного колоска; B – длина стержня колоса в см.	CO_321:000055
Spike color	set	Цвет колоса: красный/черный/белый	
Hairiness on spike scales	bool	Опушение на колосковых чешуях: есть/нет (выбор из двух вариантов)	
Hairiness on the axis of the spike	bool	Опушение на оси колоса: есть/нет (выбор из двух вариантов)	
Hairiness at the site of attachment of the spikelet to the axis	bool	Опушение на месте прикрепления колоска к оси: есть/нет (выбор из двух вариантов)	
Awn type	set	Тип остей. Безостый, остистый (длина остей длиннее или равна чем длина	CO_321:000027

		колоса), короткоостистый (длина остей короче, чем длина колоса), полуостистый (верх длина остей длиннее чем вниз длина остей)	
Awn color	set	Цвет остей: красный/белый/черный/фиолетовый/янтарный/смешанный	CO_321:0000960
Spike threshability	bool	Ломкоколосость	CO_321:0000659
Shape type	set	Форма колоса: спельта/норма/компактная	

Технологии реализации системы SpikeDroidDB

Система SpikeDroidDB была разработана на основе системы управления контентом (CMS – content manager system) Drupal 8. В качестве хранилища данных используется реляционная база данных под управлением MySQL СУБД, которая развернута на сервере баз данных под управлением CentOS Linux.

Drupal поддерживает основной функционал, такой как администрирование сайта (управление контентом, регистрация пользователей, управление правами пользователя и доступом к ресурсам), задает MVC модель системы и т.п. Drupal обеспечивает модульный принцип реализации интерфейса и доступа к данным, возможность модифицировать функционирование отдельных его частей без непосредственного вмешательства в код других частей.

Для разработки SpikeDroidDB использовали следующие модули:

- “Conditional fields”. Данный модуль интерфейса позволяет задавать правила визуализации/скрытия полей в зависимости от установленных условий;
- “Display suit”. Данный модуль обеспечивает расширение представлений материалов контента Drupal;

- “Field_group”. Этот модуль интерфейса обеспечивает группировку полей для управления ими как единым целым,
- “Geolocation”. Этот модуль обеспечивает визуализацию и управление полями, описывающими геолокационные данные, что позволяет использовать Google Places API;
- “Charts”. Этот модуль интерфейса используется при создании диаграмм и графиков.

Web интерфейс системы SpikeDroidDB разработан при помощи технологии адаптивной верстки, это позволит работать с системой, используя различные типы устройств: мобильные телефоны, персональные или планшетные компьютеры.

Модуль интерфейса системы SpikeDroidDB

Главная страница системы SpikeDroidDB располагается по адресу <http://spikedroid.biores.cytogen.ru/>. Она содержит краткую информацию о базе данных, ссылки для входа в систему или регистрации и ссылки на основные блоки информации в базе данных. Пользователь может получить доступ к базе данных, зарегистрировавшись на сайте. Зарегистрированный пользователь имеет возможность добавлять и аннотировать собственные растения. Для того чтобы просмотреть список растений, информация о которых доступна в базе, необходимо с главной страницы перейти по ссылке «Plants». После этого осуществляется переход на страницу списка растений (рис. 5.12).

Plants

Sowing number: First parent: Second parent: Generation: F

Apply







Sowing number	First parent	Second parent	Spikes	Generation	Environment	Collection	QR code
8832	Triple Dirk	Triticum yunnanense	5957041040	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	
8833	Triple Dirk	Triticum yunnanense	8784240555	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	
8834	Triple Dirk	Triticum yunnanense	0906429344	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	
8835	Triple Dirk	Triticum yunnanense	8727147590	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	
8837	Triple Dirk	Triticum yunnanense	1494383594	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	
8838	Triple Dirk	Triticum yunnanense	4083614727	F2	Novosibirsk	Wheat ear collection of Goncharov N.P.	

Рисунок 5.12. Визуализация списка растений в БД SpikeDroid.

На этой странице отображается информация о растениях в виде таблицы. Информация различного типа (фенотип колоса, информация о месте произрастания, генотипе и коллекции) располагается в разных колонках таблицы и доступна по соответствующим гиперссылкам. В крайней правой колонке для каждого растения приводятся ссылки на его QR-код. QR-код является матричным (двумерным) штрих-кодом, который может быть сканирован камерой мобильного устройства. Информация, которую он

содержит, может содержать до 4296 символов цифр и букв, что достаточно для описания доступа к растению в нашей базе по протоколу http. QR-код присваивается в нашей базе каждому растению (см. рис. 5.13), может быть распечатан на плотной бумаге и физически прикреплен к растению. После этого для получения или изменения параметров растения достаточно считать этот код мобильным устройством. Таким образом, интерфейс SpikeDroidDB позволяет сохранять информацию об измерениях фенотипов растений в ходе эксперимента непосредственно в базу данных, минуя записи в полевых и лабораторных журналах исследователей.

На странице описания растения (рис. 5.13) отображается информация о генотипе растения, данные о месте произрастания, которые отображаются с помощью метки на карте. Приводятся ссылки на описание фенотипа колоса и коллекции. Приводится QR для доступа к странице растения.


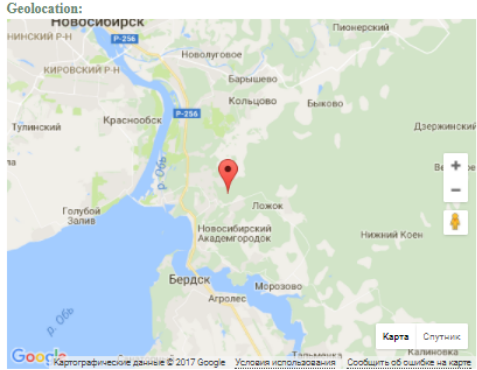
Sowing number: 8832	QR code 
First parent*: Triple Dirk	
Second parent*: Triticum yunnanense	
Generation: F2	
Environment: Grow place: greenhouse	
Grow date: 2017-06-01T19:00:00	
Geolocation: 	
Hybrid: Yes	
Collection: Wheat ear collection of Goncharov N.P.	
Spikes 5957041040	

Рисунок 5.13. Страница описания растения в БД SpikeDroid.

На странице описания фенотипа колоса отображаются характеристики колоса (рис. 5.14). Список всех возможных характеристик представлен в таблице 1. С каждым колосом можно соотнести набор фотографий. Для каждой фотографии указывается протокол, с помощью которого она была получена.






View Edit Delete Manage display Devel	
Plant: 8739	
Spike number: 1	Images  Protocol: pin  Protocol: pin  Protocol: pin  Protocol: pin  Protocol: object table
Spike length: 7.00cm	
Spike front width: 1.10cm	
Spikelet count: 22	
Density index: 0.33	
Spike color: White	
Awn type: awnletted (short)	
Awn color: White	
Shape type: compact	
Hairiness at the site of attachment of the spikelet to the awn: Yes	
Hairiness on spike scales: Yes	
Hairiness on the axis of the spike: No	

Рисунок 5.14. Страница описания фенотипа колоса в БД SpikeDroid.

Информационное содержание SpikeDroidDB

Текущая версия базы данных содержит коллекцию колосьев F2 гибридов от скрещивания австралийского сорта мягкой пшеницы Triple Dirk на образец KU506 китайской пшеницы *Triticum yunnanense*. Коллекция включает в себя 103 растения и 230 фотографий колосьев. Аннотирование колосьев осуществлялось экспертом по списку морфологических характеристик, представленных в таблице 1. В табл. 5.3 представлены результаты распределения колосьев по некоторым признакам.

Таблица 5.3. Разделение колосьев F₂ гибридов от скрещивания австралийского сорта мягкой пшеницы Triple Dirk на образец KU506 китайской пшеницы *T. uinnanense* по форме колоса, остистости, опушению колосовой чешуи, опушению на оси колоса, опушению на месте прикрепления колоса к оси.

Форма колоса	компакт	норма	спельта
Количество, шт	6	35	63
Остистость	безостный	короткоостистый	
Количество, шт	80	24	
Опушение колосовой чешуи	присутствует	отсутствует	
Количество, шт	68	36	
Опушение на оси колоса	присутствует	отсутствует	
Количество, шт	93	11	
Опушение на месте прикрепления колоса к оси	присутствует	отсутствует	
Количество, шт	74	30	

5.3. Заключение по главе 5.

Разработанная БД WheatPGE является не только хранилищем данных о фенотипе, генотипе различных сортов и линий пшеницы, а также условиях их произрастания, но и обеспечивает интерфейс для сбора этих данных в процессе селекционно-генетического эксперимента. Структура базы обеспечивает интеграцию этих данных, поиск информации и ее извлечение в формате, удобном для дальнейшей статистической обработки. Это позволяет экспериментатору затрачивать меньше времени на сбор данных, обеспечивает их сохранность, доступ к данным в любой момент времени.

Для описания характеристик колосьев пшеницы создана система SpikeDroidDB. Эта система позволяет хранить цифровые изображения колоса,

аннотировать их фенотипические характеристики (всего 14 признаков), предоставляет гибкую систему запросов для доступа к данным. В системе SpikeDroidDB для растений взаимосвязанным образом описываются генотип, фенотип, место и условия выращивания. Web интерфейс системы SpikeDroidDB доступен по адресу <http://spikedroid.biores.cytogen.ru/> и позволяет работать с системой, как со стационарных компьютеров, так и с мобильных устройств. С использованием SpikeDroidDB произведена оцифровка и аннотация коллекции колосьев гибридов F2 от скрещивания австралийского сорта мягкой пшеницы Triple Dirk с образцом KU506 китайской пшеницы *T. yunnanense*. Все это позволяет анализировать изменчивость колосьев различных генотипов по форме, длине и ширине.

ГЛАВА 6. ЗАКЛЮЧЕНИЕ

В диссертационной работе создана экспериментально-компьютерная платформа для высокопроизводительного фенотипирования растений – ICGPhenoPlant, которая включает несколько модулей (рис. 6.1).

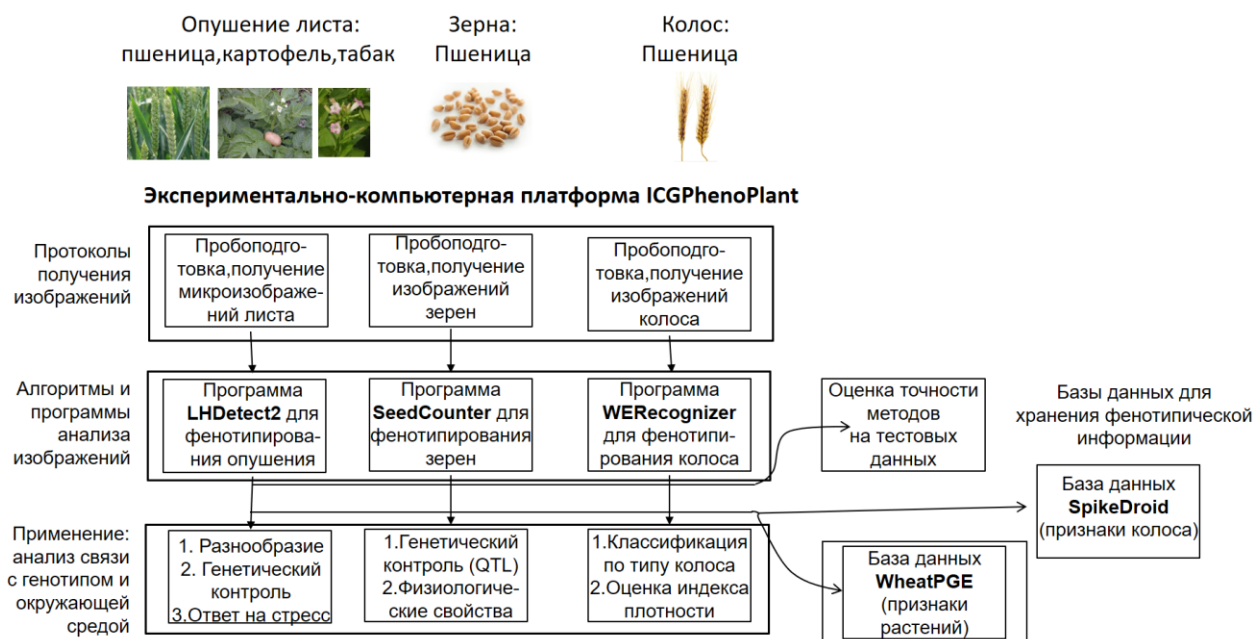


Рисунок 6.1. Структура экспериментально-компьютерной системы ICGPhenoPlant.

Первый модуль: модуль фенотипирования опущения листьев растений на основе микроизображений сгиба листа, который включает протокол пробоподготовки образцов листа для фенотипирования и программу LHDetect2 для оценки количественных характеристик опущения.

С помощью этого метода нами впервые были получены ряд новых важных результатов, включая количественную оценку опущения листа у разных видов пшениц.

Мы впервые оценили фенотипическое проявление для трех генов *H11* и

Hl2^{aesp} и *Hl3*, контролирующих развитие опушения листа у пшеницы, и показали, что ген *Hl2^{aesp}* в большей степени оказывает влияние на длину трихомов. Возможно его функция связана с изменением размера клеток трихомов. Гены *Hl3* и *Hl1* влияют одновременно и на длину трихомов и на их количество. Вероятно, они затрагивают как пути инициации трихомов на листе, так и роста клеток отдельных трихомов.

Наш метод успешно показал себя при анализе изменения количественных характеристик опушения листьев пшеницы при воздействии засухи. Аналогичные выводы позволил сделать анализ опушения листа для растений, выращенных в различных условиях (Дорошков и соавт., 2011): опушение растений сорта Голубка и линии 821 с интрогрессией от *Triticum timopheevii*, выращенных в условиях теплицы оказалось более редким, а трихомы более длинными в сравнении с опушением растений этих же генотипов, но выращенных в полевых условиях.

Было показано, что яровые сорта Голубка и Саратовская 29 со сходным на ощупь опушением листа имеют достоверно различное распределение трихомов по длине и плотности (Дорошков и соавт., 2011). Эти различия были подтверждены генетическим анализом, который показал моногенное и дигенное наследование этого признака у двух сортов, соответственно. Также впервые был проведён массовый анализ длины и плотности опушения листа в коллекции из 47 сортов различного происхождения (Генаев и соавт., 2012). Многомерный статистический анализ показал, что яровые сорта, созданные в Германии, обладают слабым опушением по сравнению с большинством сортов российской селекции.

В целом наши результаты, впервые на количественном уровне полученные для пшеницы, хорошо согласуются с данными для других видов растений: при суровых условиях среды опушение листьев становится более

плотным и коротким, что было показано для большого количества разных видов. Кроме того, в нашей работе впервые у пшеницы была выявлена связь между опушением и физиологическими параметрами растений при воздействии на них условий ограниченного водообеспечения.

Свою производительность и эффективность метод показал при анализе опушения в гибридной популяции F_2 , полученной от скрещивания двух сходно опушенных образцов, сорта Hong-mang-mai и линии с геном опушения, интрогрессированным от вида *Ae. speltoides*. Визуально растения не отличались друг от друга, однако анализ с помощью программы LHDetect2 позволил выявить аллелизм по длине трихом (Genaev et al., 2012). Использование данного метода в моносомном анализе опушения листа у изогенных линий пшеницы позволило идентифицировать новый ген в хромосоме 7D сорта Новосибирская 67 (Дорошков и соавт., 2014).

Таким образом, данные о генетическом контроле и средовой регуляции опушения листа пшеницы были существенно пополнены по сравнению с исследованиями предыдущих десятилетий.

Алгоритм оценки количественных характеристик опушения листьев был применен для анализа опушения картофеля и с изменениями – для анализа опушения листьев табака.

Полученные результаты демонстрируют, что данный метод пригоден для массового анализа опушения растений и позволяет получать уникальные результаты с детализацией количественных характеристик опушения листа, которые ранее не были доступны. В последнее время были разработаны ряд методов для морфометрии опушения. В работе Мирнезами и соавт (Mirnezami et al, 2020) был разработан метод для подсчета трихомов на изображении листа сои. В работе было предложено два метода подсчета трихомов на изображении: алгоритм подсчета вершин трихомов и алгоритм смещения границы листа и

фона. Однако в этой работе авторы не оценивали длины трихомов на изображении. В работе Ролланда и соавт. (Rolland et al., 2022) для классификации листьев табака по интенсивности опушения на 9 классов использовалась нейронная сеть глубокого обучения. Однако результаты этой сети позволяют лишь классифицировать опушение, но не дают возможности оценить его количественные характеристики. Эти работы демонстрируют актуальность задачи изучения опушения листьев растений на основе компьютерного фенотипирования.

Второй модуль (рис. 6.1) включал методы фенотипирования зерен злаков: протоколы для фенотипирования на камеру мобильного устройства или цифрового фотоаппарата и программу SeedCounter для устройств на платформе Android и ПК. Несмотря на то, что методы оценки признаков зерен впервые были разработаны еще в 1980-х годах с появлением первых цифровых фотокамер, они постоянно развиваются как в направлении совершенствования алгоритмов обработки изображений, так и в направлении технологий фенотипирования. Последние, в свою очередь, развиваются в направлении миниатюризации и более высокой мобильности за счет использования сканеров и более компактных фотокамер, создания настольных средств для фенотипирования зерен. Мобильные устройства в этой связи представляют большой интерес, поскольку объединяют в одном корпусе как фотокамеру, так и репрограммируемый процессор. Однако до недавнего времени потенциал мобильных устройств не был использован. В настоящей работе мы впервые предложили практическую технологию фенотипирования зерен на основе использования мобильных устройств, реализованную в виде приложения SeedCounter (Komyshev et al., 2017). Востребованность этого метода оказалась достаточно высокой: в настоящее время это приложение загружено более 1000 раз и использовалось как в обучающих целях, так и научных исследованиях

(Lobo et al. 2019; Kroupin et al., 2020; Kroupin et al., 2019; Науменко и соавт., 2019; Divashuk et al., 2020; Науменко и соавт., 2021; Raviteja и соавт., 2021). На основе предложенных нами идей другими исследователями позднее были разработаны аналогичные приложения для подсчета зерен (Wu et al., 2018; Zhang et al., 2022). Таким образом, выбранное нами направление развития технологий фенотипирования зерен на основе мобильных устройств оказалось перспективным.

Мы продемонстрировали, что на основе высокопроизводительного фенотипирования зерен как в условиях лаборатории, так и вне ее, можно успешно оценивать размеры зерен, их форму и характеристики цвета оболочки. С помощью нашего приложения были выявлены локусы количественных признаков у пшеницы, которые ассоциированы с размером, формой и цветом оболочки зерен. Это позволило локализовать в геноме гены-кандидаты, которые могут быть ответственными за формирование этих признаков у растений.

Исследование признаков размера/формы и цвета оболочки зерен для растений различных годов урожая из генбанка института ИРК, г. Гатерслебен позволило оценить связь между временем хранения семян пшеницы в генбанке и изменчивостью их фенотипических характеристик. Впервые на основе этих уникальных данных было показано, что длительное хранение зерен пшеницы приводит к изменению цвета оболочки так, что она становится более красной и более светлой (что может отражать изменение ее пигментного состава). Дополнительно наш метод продемонстрировал статистически значимую связь между цветом зерен и их всхожестью: более красные зерна имеют худшие показатели всхожести. Несмотря на то, что связь между красной окраской зерен и их всхожестью ранее была известна, в нашей работе мы продемонстрировали количественно на большой выборке зерен разных генотипов и разных годов

урожая. В этом отношении наши результаты являются уникальными.

Все это демонстрирует высокую точность и информативность разработанного нами подхода фенотипирования зерен пшеницы, который позволяет в итоге оценивать связь между фенотипом зерен, генотипом (идентификация QTL) и физиологическим состоянием зерен (процессы деградации пигментов в оболочке в процессе хранения в условиях генбанка).

Интересно, что применение приложения SeedCounter в научных исследованиях выходит за рамки изучения зерен злаков: недавно этот метод был адаптирован для подсчета плодовых мушек дрозофил в задачах оценки их плодовитости (Karova et al., 2020).

Третий модуль направлен на фенотипирование колосьев пшеницы (рис. 6.1). Он основан на двух протоколах съемки колоса и позволяет с использованием программы WERecognizer на основе двумерных изображений определять основные характеристики размера и формы колоса. Дополнительно метод позволяет оценивать характеристики остистости колоса. Мы впервые представили описание формы колоса в модели двух четырехугольников. На ряде примеров показано, что указанные характеристики колоса позволяют адекватно описать как остистость колоса, так и его тип (спельтоидный, нормальный или компактный). Параметры формы и размера колоса позволили также предсказывать плотность колоса. Все это демонстрирует эффективность применения методов компьютерного фенотипирования колосьев и адекватность предлагаемых моделей описания его формы. В рамках продолжения настоящей работы это позволило охарактеризовать сходство и различия по форме колосьев у тетраплоидных и гексаплоидных сортов пшениц (Пронозин и соавт., 2021).

Экспериментально-компьютерная система ICGPhenoPlant (рис. 6.1) помимо методов фенотипирования также включает и ресурсы для хранения и

быстрого доступа к результатам экспериментов по фенотипированию. При этом следует иметь в виду, что сами по себе фенотипические данные не представляют большого интереса для биолога, если они не связаны с данными о генотипе и фенотипе растения. Все это обусловило развитие в рамках диссертации двух баз данных, WheatPGE и SpikedroidDB для хранения данных по фенотипам пшеницы. Созданные в рамках работы прототипы баз данных показали эффективность предложенных методов интеграции разнородных данных о генотипе, фенотипе и окружающей среде.

Полученные в рамках диссертации разработки и результаты свидетельствуют об эффективности предложенных методов фенотипирования растений на основе компьютерного анализа цифровых изображений для успешного решения задач о связи между фенотипом, генотипом и окружающей средой у растений. Они обеспечивают основу для дальнейшего развития методов фенотипирования растений в том числе и с использованием популярных в настоящее время методов глубокого машинного обучения (Genaev et al., 2021; Genaev et al., 2020).

ВЫВОДЫ

1. Разработана экспериментально-компьютерная платформа ICGPhenoPlant для фенотипирования растений, позволяющая с высокой точностью на основе двумерных изображений:

- (а) с помощью приложения LHDetect2 оценивать количественные характеристики опушения листьев пшеницы, картофеля и табака;
- (б) с помощью приложения SeedCounter оценивать 3 признака размера, 4 признака формы и 48 признаков цвета оболочки зерен пшеницы;
- (в) с помощью приложения WERecognizer оценивать характеристики размера и формы колоса пшеницы;
- (г) сохранять информацию о фенотипе, генотипе и месте произрастания растений в базах данных WheatPGE и SpikeDroid.

2. Использование приложения LHDetect2 для анализа опушения пшеницы позволило:

- (а) впервые изучить опушение у 47 образцов пшеницы и ее диких сородичей; выделить 9 характерных типов опушения; установить, что для гексаплоидных пшениц наблюдается положительная корреляция между числом трихомов и их длиной; установить, что характеристики опушения листьев пшеницы значимо ассоциированы с хромосомным набором генома;
- (б) для генов, контролирующих опушение листа пшеницы впервые показано, что ген *HL2*^{aesp} в большей степени влияет на длину трихом, а гены *HL1* и *HL3* влияют одновременно и на длину трихомов и на их количество;
- (в) впервые на основе точной оценки количественных характеристик опушения листа пшеницы программой LHDetect2 показана связь между изменчивостью опушения листа пшеницы и ответом растения на

вододефицит: число трихомов на обеих сторонах листа увеличивается, а их средняя длина уменьшается; изменение длины трихомов связано с нефотохимическим тушением и числом устьиц на нижней стороне листа; изменение числа трихомов на нижней и верхней стороне листа связано с эффективностью использования воды.

3. Использование приложения LHDetect2 для анализа опушения листьев двудольных растений показало, что:

(а) опушение листьев для 35 сортов картофеля демонстрирует широкое разнообразие по количеству трихомов и их длине; наиболее интенсивное опушение демонстрируют сорта Великан, Танай, Кемеровчанин;

(б) линии табака с измененным паттерном экспрессии гена пролиндегидрогеназы PDH и повышенным содержанием пролина по сравнению с контрольным генотипом SR1 имеют достоверно большее количество трихомов на верхней стороне листа, при этом средняя длина трихомов достоверно меньше, чем у растений контрольной линии.

5. Использование приложения SeedCounter для анализа зерен пшеницы растений популяции ITMI позволило:

(а) выявить 212 аддитивных и 34 пары эпистатических QTL для признаков зерен; идентифицировать 29 генов-кандидатов для локусов, ассоциированных с размером семян, и 14 для локусов, ассоциированных с цветом оболочки;

(б) показать, что для зерен мягкой пшеницы, хранившихся в условиях генбанка пигментация оболочки значительно изменяется в зависимости от срока хранения;

(г) выявить значимую отрицательную корреляцию между интенсивностью красной компоненты цвета оболочки и всхожестью зерен.

6. Использование приложения WERecognizer на основе количественных характеристик размера и формы колоса пшеницы позволяет с высокой точностью оценить тип колоса и индекс его плотности.

Список публикаций в рецензируемых журналах и авторские свидетельства по теме диссертации

1. Arif, M.A.R.; Komyshev, E.G.; Genaev, M.A.; Koval, V.S.; Shmakov, N.A.; Börner, A.; **Afonnikov, D.A.** (2022) QTL Analysis for Bread Wheat Seed Size, Shape and Color Characteristics Estimated by Digital Image Processing. *Plants*, 11, 2105.
2. Ibragimova, S., Genaev, M., Kochetov, A., **Afonnikov, D.** (2022). Variability of leaf pubescence characteristics in transgenic tobacco lines with partial proline dehydrogenase gene suppression. *Biologia plantarum*, 66, 24-28.
3. **Afonnikov D.A.**, Komyshev E.G., Efimov, V.M., Genaev M.A., Koval V.S., Gierke P.U., Börner A. (2022) Relationship between the Characteristics of Bread Wheat Grains, Storage Time and Germination. *Plants* 2022, 11, 35.
4. Комышев, Е. Г., Генаев, М. А., **Афонников, Д. А.** (2020). Анализ цветковых и текстурных характеристик зерен злаков на цифровых изображениях. *Вавиловский журнал генетики и селекции*, 24(4), 340-347.
5. Иванова К.А., Комышев Е.Г., Генаев М.А., Егорова А.А., Колошина К.А., Чалая Н.А., **Афонников Д.А.**, Кочетов А.В., Рогозина Е.В., С.В. Герасимова (2019). Оценка количественных характеристик клубнеобразования дикого картофеля на основе анализа изображений клубней с использованием компьютерного приложения SeedCounter. *Вавиловский журнал генетики и селекции*, 19(6), 1-8.
6. Genaev M.A., Komyshev E.G., Smirnov N.V., Kruchinina Y.V., Goncharov N.P. and **Afonnikov D.A.** (2019) Morphometry of the Wheat Spike by Analyzing 2D Images. (2019) *Agronomy*, 9(7), 390.
7. Pshenichnikova, T.A., Doroshkov, A.V., Osipova, S.V., Permyakov, A.V., Permyakova, M.D., Efimov, V.M., **Afonnikov, D.A.** (2019) Quantitative characteristics of pubescence in wheat (*Triticum aestivum* L.) are associated

with photosynthetic parameters under conditions of normal and limited water supply. *Planta*, 249 (3), 839-847.

8. Комышев Е. Г., Генаев М. А., **Афонников Д. А.** Программа для оценки количественных характеристик колоса пшеницы (ОКХК) / The program for the extraction the quantitative characteristics of the wheat spike (WERecognizer) Свидетельство о регистрации программы для ЭВМ N 2019666362 от 09.12.2019, дата приоритета 27.11.2019.
9. Дорошков А.В., **Афонников Д.А.** (2018) Опушение листа у картофеля *Solanum tuberosum*: морфология, функциональная роль и методы исследования. (2018) Вавиловский журнал генетики и селекции. 22(1), 46-53. DOI 10.18699/VJ18.327.
10. Генаев М.А., Комышев Е.Г., Фу Хао, Коваль В.С., Гончаров Н.П., **Афонников Д.А.** (2018) SpikeDroidDB – информационная система для аннотации морфометрических характеристик колоса пшеницы. Вавиловский журнал генетики и селекции. 22(1):132-140. DOI 10.18699/VJ18.340.
11. Komyshev E.G., Genaev M.A., **Afonnikov D.A.** (2017) Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front. Plant Sci.*, 7:1990. doi: 10.3389/fpls.2016.01990
12. Pshenichnikova T.A., Doroshkov A.V., Simonov A.V., **Afonnikov D.A.**, Börner A. (2017) Diversity of leaf pubescence in bread wheat and relative species. *Genet. Res. and Crop Evol.*, 64, 1761-1773. DOI:10.1007/s10722-016-0471-3.
13. Doroshkov, A. V., **Afonnikov, D. A.**, Dobrovolskaya, O. B., Pshenichnikova, T. A. (2016). Interactions between leaf pubescence genes in bread wheat as assessed by high throughput phenotyping. *Euphytica*, 207, 491-500. doi: 10.1007/s10681-015-1520-2
14. Дорошков А.В., Симонов А.В., Сафонова А.Д., **Афонников Д.А.**, Лихенко

- Н.Е., Колчанов Н.А. (2016) Оценка количественных характеристик опушения листьев картофеля с использованием анализа цифровых микроизображений. Достижения науки и техники АПК. Т.30, С. 12-14.
15. **Афонников Д.А.**, Генаев М.А., Дорошков А.В., Комышев Е.Г., Пшеничникова Т.А. (2016) Методы высокопроизводительного фенотипирования растений для массовых селекционно-генетических экспериментов. Генетика. Т. 52. № 7. С. 788-803.
 16. Дорошков А.В., Генаев М.А., **Афонников Д.А.** Протокол анализа количественных характеристик опушения листа картофеля. Вавиловский журнал генетики и селекции. 2016;20(6):863-868. DOI 10.18699/VJ16.218
 17. Дорошков А.В., **Афонников Д.А.**, Пшеничникова Т.А. (2014) Генетический анализ опушения листа у изогенных линий мягкой пшеницы сорта Новосибирская 67. Генетика, Т. 50, N. 2, С. 172–180.
 18. Genaev M.A., Doroshkov A.V., Pshenichnikova T.A., Kolchanov N.A., **Afonnikov D.A.** (2012) Extraction of quantitative characteristics describing wheat leaf pubescence with a novel image-processing technique. Planta, 236, 1943–1954. DOI 10.1007/s00425-012-1751-6
 19. Генаев М.А., Дорошков А.В., Пшеничникова Т.А., Морозова Е.В., Симонов А.В., **Афонников Д.А.** (2012) Информационная поддержка селекционно-генетического эксперимента у пшеницы в системе WheatPGE. Математическая биология и биоинформатика. 2012. Т. 7. № 2. С. 410–424.
 20. Генаев М.А., Дорошков А.В., **Афонников Д.А.** (2012) Программа автоматического определения количественных характеристик опушения листа (ОЛДетект2). // Свидетельство о регистрации программы для ЭВМ N 2012620897 от 21.08.2012 от 5.03.2012, дата приоритета 21.08.2012.
 21. Генаев, М.А., Дорошков А.В., Морозова Е.В., Пшеничникова Т.А.,

- Афонников Д.А.** (2011) Компьютерная система WheatPGE для анализа взаимосвязи фенотип– генотип–окружающая среда у пшеницы. Вавиловский журнал генетики и селекции. 15, 784-793.
22. Генаев М.А., Дорошков А.В., **Афонников Д.А.** (2010) База данных для анализа взаимосвязей генотип-фенотип-окружающая среда у пшеницы (ВитПГЕ). // Свидетельство о регистрации базы данных N 2010620602 от 23.10.2010, дата приоритета 30.06.2010
23. Дорошков А.В., Пшеничникова Т.А., **Афонников Д.А.** (2011) Анализ особенностей морфологии и наследования опушения листа пшеницы *Triticum aestivum* L. с помощью компьютерных методов фенотипирования. Генетика, Т. 47, № 6, с. 839–843.
24. Дорошков А.В., Арсенина С.И., Пшеничникова Т.А., **Афонников Д.А.** (2009) Применение компьютерной обработки изображений микрофотографий листа для анализа опушения пшеницы *Triticum aestivum* L. Информационный вестник ВОГиС. 13, 218-226.

Использованные сокращения

CIE – Международная комиссия по освещенности (International Commission on Illumination).

FN – при бинарной классификации: число отрицательных объектов, распознанных как положительные.

FP – при бинарной классификации: число положительных объектов, распознанных как отрицательные.

HSV – цветовое пространство определяемое компонентами: оттенок (Hue), насыщенность (Saturation), интенсивность (Value).

ISTA – Международная ассоциация по тестированию семян (International Seed Testing Association).

ITMI – Международная инициатива по картированию пшениц (International Triticeae Mapping Initiative).

$L^*a^*b^*$ – цветовое пространство определяемое компонентами: светлота (Lightness), a^* и b^* .

MAE – средняя абсолютная ошибка (Mean absolute error).

MAPE – средняя абсолютная процентная ошибка (Mean absolute percent error).

QTL – локусы количественных признаков (Quantitative trait loci).

RGB – Цветовое пространство определяемое компонентами: красной (Red), зеленой (Green), синей (Blue).

TN – при бинарной классификации: число отрицательных объектов, распознанных как отрицательные.

TP – при бинарной классификации: число положительных объектов, распознанных как положительные.

UPGMA – метод невзвешенной попарной кластеризации (Unweighted pair group method with arithmetic mean).

YCrCb – цветовое пространство определяемое компонентами: яркость

(luminosity), Cr и Сb.

БД – база данных.

РИЛ – рекомбинантная инбредная линия.

СПИСОК ЛИТЕРАТУРЫ

1. Abràmoff M. D., Magalhães P. J., Ram S. J. Image processing with ImageJ //Biophotonics international. – 2004. – V. 11. – № 7. – P. 36-42.
2. Aha D. W., Kibler D., Albert M. K. Instance-based learning algorithms //Machine learning. – 1991. – V. 6. – № 1. – P. 37-66.
3. Ahmad I. S. et al. Color classifier for symptomatic soybean seeds using image processing //Plant disease. – 1999. – V. 83. – № 4. – P. 320-327.
4. Akers C. P., Weybrew J. A., Long R. C. Ultrastructure of glandular trichomes of leaves of *Nicotiana tabacum* L., cv Xanthi //American Journal of Botany. – 1978. – V. 65. – № 3. – P. 282-292.
5. Alemu A. et al. Genome-wide association mapping for grain shape and color traits in Ethiopian durum wheat (*Triticum turgidum* ssp. durum) //The Crop Journal. – 2020. – V. 8. – № 5. – P. 757-768.
6. Ali A. et al. Genetic analysis of wheat grains using digital imaging and their relationship to enhance grain weight //Scientia Agricola. – 2020. – V. 77. – № 6. – P. e20190069.
7. Al-Tam F., Adam H., dos Anjos A., et al. P-TRAP: a panicle trait phenotyping tool // BMC plant biology. – 2013. – V. 13. – P. 122.
8. Amagai Y. et al. Microsatellite mapping of the loci for false glume and semi-compact spike in *Triticum* L //Genetic Resources and Crop Evolution. – 2017. – V. 64. – № 8. – P. 2105-2113.
9. Amer Hamzah M. et al. Nucleotide variations of 9-cis-epoxycarotenoid dioxygenase 2 (NCED2) and pericarp coloration genes (Rc and Rd) from upland rice varieties //3 Biotech. – 2020. – V. 10. – № 3. – P. 1-7.
10. Angeles-Shim R. B. et al. A WUSCHEL-related homeobox 3B gene, depilous

- (dep), confers glabrousness of rice leaves and glumes // *Rice*. – 2012. – V. 5. – № 1. – P. 1-12.
11. Antonious G. F. Production and quantification of methyl ketones in wild tomato accessions // *Journal of Environmental Science and Health, Part B*. – 2001. – V. 36. – № 6. – P. 835-848.
12. Aramaki T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold // *Bioinformatics*. – 2020. – V. 36. – № 7. – P. 2251-2252.
13. Araus J.L., Cairns J.E. Field high-throughput phenotyping: the new crop breeding frontier // *Trends in Plant Science*. – 2014. – V. 19. – № 1. – P. 52-61.
14. Arbuzova V.S., Efremova T.T., Laikova L.I., Maystrenko O.I., Popova O.M., Pshenichnikova T.A. The development of precise genetic stocks in two wheat cultivars and their use in genetic analysis // *Euphytica*. – 1996. – V. 89. – P. 11-15.
15. Arif M. A. R. et al. A SNP-based genetic dissection of versatile traits in bread wheat (*Triticum aestivum* L.) // *The Plant Journal*. – 2021. – V. 108. – № 4. – P. 960-976.
16. Arif M. A. R. et al. QTL Analysis for Bread Wheat Seed Size, Shape and Color Characteristics Estimated by Digital Image Processing // *Plants*. – 2022. – V. 11. – № 16. – P. 2105.
17. Arif R. et al. Genetic studies of seed longevity in hexaploid wheat using segregation and association mapping approaches // *Euphytica*. – 2012. – V. 186. – № 1. – P. 1-13.
18. Arvidsson S., Pérez-Rodríguez P., Mueller-Roeber B. A growth phenotyping

- pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects // *New Phytologist*. – 2011. – V. 191. – № 3. – P. 895-907.
19. Avé D. A., Tingey W. M. Phenolic constituents of glandular trichomes on *Solanum berthaultii* and *S. polyadenium* // *American potato journal*. – 1986. – V. 63. – № 9. – P. 473-480.
20. Baasandorj T. et al. Effect of kernel size and mill type on protein, milling yield, and baking quality of hard red spring wheat // *Cereal Chemistry*. – 2015. – V. 92. – № 1. – P. 81-87.
21. Babben S. et al. Detection of allelic diversity in genes involved in frost tolerance in bread wheat (*Triticum aestivum* L.). // *EUCARPIA Cereal Section – ITMI Joint Conference, Wernigerode, Germany, June 29-July 4, 2014. Book of Abstracts*, p. 223.
22. Backhaus A., Kuwabara A., Bauch M., et al. LEAFPROCESSOR: a new leaf phenotyping tool using contour bending energy and shape cluster analysis // *New phytologist*. 2010. – V. 187. – № 1. – P. 251-261.
23. Bai, J.W.; Xiao, H.W.; Ma, H.L.; Zhou, C. S. Artificial neural network modeling of drying kinetics and color changes of ginkgo biloba seeds during microwave drying process // *Journal of Food Quality*. – 2018. – V. 2018. – P. 3278595.
24. Bai, X. D., Cao, Z.G., Wang, Y., Yu, Z.H., Zhang, X.F., Li, C.N. Crop segmentation from images by morphology modeling in the CIE L* a* b* color space // *Computers and electronics in agriculture*. – 2013. – V. 99. – P. 21-34.
25. Baldocchi D. D. et al. Leaf Pubescence Effects on the Mass and Energy Exchange Between Soybean Canopies and the Atmosphere 1 // *Agronomy Journal*. – 1983. – V. 75. – № 3. – P. 537-543.

26. Barbagallo R.P., Oxborough K., Pallett K.E., Baker N.R. Rapid, noninvasive screening for perturbations of metabolism and plant growth using chlorophyll fluorescence imaging // *Plant Physiology*. – 2003. – V. 132. – № 2. – P. 485-493.
27. Bauriegel E., Giebela A., Geyer M., et al. Early detection of *Fusarium* infection in wheat using hyper-spectral imaging // *Computers and Electronics in Agriculture*. – 2011. – V. 75. – № 2. – P. 304-312.
28. Beale M. H. et al. Aphid alarm pheromone produced by transgenic plants affects aphid and parasitoid behavior // *Proceedings of the National Academy of Sciences*. – 2006. – V. 103. – № 27. – P. 10509-10513.
29. Beavis W.D. QTL analyses: power, precision, and accuracy // *Molecular dissection of complex traits*. – 1998. – V. 1998. – P. 145-162.
30. Behtari B., Luis M. D. E., Nasab A. D. M. Predicting germination of *Medicago sativa* and *Onobrychis viciifolia* seeds by using image analysis // *Turkish Journal of Agriculture and Forestry*. – 2014. – V. 38. – № 5. – P. 615-623.
31. Bensch R., Ronneberger O., Greese B., Fleck C., Wester K., Hulskamp M., Burkhardt H. Image analysis of *Arabidopsis* trichome patterning in 4d confocal datasets / In *Proceedings of the The Sixth IEEE International Symposium on Biomedical Imaging*, 2009. – P. 742-745
32. Bi K. et al. Non-destructive measurement of wheat spike characteristics based on morphological image processing // *Transactions of the Chinese Society of Agricultural Engineering*. – 2010. – V. 26. – № 12. – P. 212-216.
33. Biancucci M. et al. Proline affects the size of the root meristematic zone in *Arabidopsis* // *BMC plant biology*. – 2015. – V. 15. – P. 263.
34. Billiau K., Sprenger H., Schudoma C., et al. Data management pipeline for

- plant phenotyping in a multisite project // *Functional Plant Biology*. – 2012. – V. 39. – № 11. – P. 948-957.
35. Blake V.C., Kling J.G., Hayes P.M., et al. The Hordeum toolbox: the Barley coordinated agricultural project genotype and phenotype resource // *The Plant Genome*. – 2012. – V. 5. – № 2. – P. 81-91.
36. Boden S. A. et al. Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat // *Nature Plants*. – 2015. – V. 1. – P. 14016. DOI 10.1038/nplants.2014.16
37. Börner A. et al. QTL analysis of falling number and seed longevity in wheat (*Triticum aestivum* L.) // *Journal of applied genetics*. – 2018. – V. 59. – № 1. – P. 35-42.
38. Borrill P., Harrington S. A., Uauy C. Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat // *The Plant Journal*. – 2019. – V. 97. – № 1. – P. 56-72.
39. Borrill P., Ramirez-Gonzalez R., Uauy C. expVIP: a customizable RNA-seq data analysis and visualization platform // *Plant physiology*. – 2016. – V. 170. – № 4. – P. 2172-2186.
40. Boughalleb F., Hajlaoui H. Physiological and anatomical changes induced by drought in two olive cultivars (cv Zalmati and Chemlali) // *Acta Physiologiae Plantarum*. – 2011. – V. 33. – № 1. – P. 53-65.
41. Bradski G., Kaehler A. *Learning OpenCV: Computer vision with the OpenCV library*. – O'Reilly Media, Inc., 2008.
42. Bray M.A., Vokes M.S., Carpenter A.E. Using CellProfiler for automatic identification and measurement of biological objects in images // *Current Protocols in Molecular Biology*. – 2014. – V. 109. – P. 14.17.1-14.17.13

43. Breiman L. Random Forests. // *Machine Learning*. – 2001. – V. 45. – № 1. – P. 5–32.
44. Breseghello, F. and M.E. Sorrells, QTL analysis of kernel size and shape in two hexaploid wheat mapping populations // *Field crops research*. – 2007. – V. 101. – № 2. – P. 172-179.
45. Brinton, J. and C. Uauy, A reductionist approach to dissecting grain weight and yield in wheat // *Journal of Integrative Plant Biology*. – 2019. V. 61. – № 3. – P. 337-358.
46. Bucksch A., BurrIDGE J., York L.M., et al. Image-based high-throughput field phenotyping of crop roots // *Plant physiology*. – 2014. – V. 166. – № 2. – P. 470-486.
47. Busemeyer L., Mentrup D., Möller K., et al. Breedvision — A multi-sensor platform for non-destructive field-based phenotyping in plant breeding // *Sensors*. – 2013. – V. 13. – № 3. – P. 2830-2847.
48. Bylesjö M., Segura V., Soolanayakanahally R.Y., et al. LAMINA: a tool for rapid quantification of leaf size and shape parameters // *BMC Plant Biology*. – 2008. – V. 8. – № 1. – P. 82.
49. Camacho, C., et al., BLAST+: architecture and applications // *BMC bioinformatics*. – 2009. – V. 10. – P. 421.
50. Camargo A., Smith J.S. Image pattern classification for the identification of disease causing agents in plants // *Computers and Electronics in Agriculture*. – 2009. – V. 66. – № 2. – P. 121-125.
51. Cardoso M. Z. Herbivore handling of a plant's trichome: the case of *Heliconius charithonia* (L.) (Lepidoptera: Nymphalidae) and *Passiflora lobata* (Killip) Hutch. (Passifloraceae) // *Neotropical Entomology*. – 2008. – V. 37. –

- P. 247-252.
52. Cervantes, E.; Martín Gómez, J.J. Seed shape description and quantification by comparison with geometric models // *Horticulturae*. – 2019. – V. 5. – P. 60.
53. Chaugule A., Mali S.N. Evaluation of texture and shape features for classification of four paddy varieties // *Journal of Engineering*. – 2014. – V. 2014. – P. 617263.
54. Chen D., Neumann K., Friedel S., et al. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis // *The Plant Cell*. – 2014. – V. 26. – № 12. – P. 4636-4655.
55. Chen X. et al. Combining discriminant analysis and neural networks for corn variety identification // *Computers and electronics in agriculture*. – 2010. – V. 71. – P. S48-S53.
56. Chen, K., et al., Genetic and molecular factors determining grain weight in rice // *Frontiers in Plant Science*. – 2021. – V. 12. – P. 605799.
57. Chitwood D.H., Ranjan A., Martinez C.C., et al. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape // *Plant Physiology*. – 2014. – V. 164. – № 1. – P. 259-272.
58. Chung, H.S.; Kim, D.H.; Youn, K.S.; Lee, J.B.; Moon, K. D. Optimization of roasting conditions according to antioxidant activity and sensory quality of coffee brews. *Food Sci. Biotechnol.* 2013, 22, 23-29.
59. Cieplinski, L. MPEG-7 Color Descriptors and Their Applications. In *Computer Analysis of Images and Patterns // CAIP 2001; Lecture Notes in Computer Science*; Skarbek, W., Ed.; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2124, pp. 11–20.

60. Clercx, E.J.M.; El-Lithy, M.E.; Vierling, E.; Ruys, G.J., Blankestijn-De Vries, H.; Groot, S.P.C.; Vreugdenhil, D.; Koornneef, M. Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions *Landsberg erecta* and *Shakdara*, using a new recombinant inbred line population // *Plant Physiol.* – 2004. – V. 135. – P. 432–444.
61. Cobb J.N., DeClerck G., Greenberg A., et al. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement // *Theoretical and Applied Genetics.* – 2013. – V. 126. – № 4. – P. 867-887.
62. Collard B.C.Y., Mackill D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century // *Philosophical Transactions of the Royal Society B: Biological Sciences.* – 2008. – V. 363. – № 1491. – P. 557-572.
63. Comar A., Burger P., de Solan B., et al. A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results // *Functional Plant Biology.* – 2012. – V. 39. – № 11. – P. 914-924.
64. Cornelissen J.H.C., Lavorel S., Garnier E., et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide // *Australian journal of Botany.* – 2003. – V. 51. – № 4. – P. 335-380.
65. Crowell S., Falcão A.X., Shah A., et al. High-resolution inflorescence phenotyping using a novel image-analysis pipeline, PANorama // *Plant physiology.* – 2014. – V. 165. – № 2. – P. 479-495.
66. Cuesta-Marcos A., Szűcs P., Close T.J., et al. Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure // *BMC*

- genomics. – 2010. – V. 11. – № 1. – P. 707.
67. Dar M. I. Naikoo M.I., Rehman F., Naushin F., and Khan F.A. Proline Accumulation in Plants: Roles in Stress Tolerance and Plant Development. // Osmolytes and Plants Acclimation to Changing Environment: Emerging Omics Technologies. / Eds Iqbal N. et al. Springer, New Delhi. 2016. P. 155.
68. Darrigues A., Hall J., van der Knaap E., et al. Tomato analyzer-color test: a new tool for efficient digital phenotyping // Journal of the American Society for Horticultural Science. – 2008. – V. 133. – № 4. – P. 579-586.
69. Dawson-Howe, K. A Practical Introduction to Computer Vision with OpenCV. John Wiley & Sons. 2014.
70. De Vylder J., Vandebussche F., Hu Y., et al. Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects // Plant physiology. – 2012. – V. 160. – № 3. – P. 1149-1159.
71. Debeaujon I., Léon-Kloosterziel K. M., Koornneef M. Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis* // Plant physiology. – 2000. – V. 122. – № 2. – P. 403-414.
72. Delagrangé S., Rochon P. Reconstruction and analysis of a deciduous sapling using digital photographs or terrestrial-LiDAR technology // Annals of botany. – 2011. – V. 108. – № 6. – P. 991-1000.
73. Dell'Aquila A. Development of novel techniques in conditioning, testing and sorting seed physiological quality // Seed Science and Technology. – 2009. – V. 37. – № 3. – P. 608-624.
74. Dell'Aquila A. Red-Green-Blue (RGB) colour density as a non-destructive marker in sorting deteriorated lentil (*Lens culinaris* Medik.) seeds // Seed Science and Technology. – 2006. – V. 34. – № 3. – P. 609-619.

75. Delouche, J.C.; Baskin, C. C. Accelerated aging technique for predicting the seed storage period // *Seed Sci. Technol.* – 1973. – V. 1. – № 2. – P. 427-452.
76. Delwiche S. R., Yang I. C., Graybosch R. A. Multiple view image analysis of freefalling US wheat grains for damage assessment // *Computers and Electronics in Agriculture.* – 2013. – V. 98. – P. 62-73.
77. Divashuk M. G. et al. Effect of gibberellin responsive reduced height allele Rht13 on agronomic traits in spring bread wheat in field experiment in non-black soil zone // *Agronomy.* – 2020. – V. 10. – № 7. – P. 927.
78. Dobrovolskaya O. et al. FRIZZY PANICLE drives supernumerary spikelets in bread wheat // *Plant physiology.* – 2015. – V. 167. – № 1. – P. 189-199.
79. Dobrovolskaya O.B., Pshenichnikova T.A., Arbuzova V.S. et al. Molecular mapping of genes determining hairy leaf character in common wheat with respect to other species of the Triticeae // *Euphytica.* – 2007. – Vol. 155. № 3. – P. 285-293.
80. Doroshkov A. V. et al. Interactions between leaf pubescence genes in bread wheat as assessed by high throughput phenotyping // *Euphytica.* – 2016. – V. 207. – № 3. – P. 491-500.
81. Doroshkov A. V. et al. The evolution of gene regulatory networks controlling *Arabidopsis thaliana* L. trichome development // *BMC plant biology.* – 2019. – V. 19. – № 1. – P. 53.
82. Duan L., Yang W., Huang C., Liu Q. A novel machine-vision-based facility for the automatic evaluation of yield-related traits in rice // *Plant Methods.* – 2011. – V. 7. – P. 44.
83. Duke S. O. Glandular trichomes-a focal point of chemical and structural interactions // *International Journal of Plant Sciences.* – 1994. – V. 155. – № 6.

- P. 617-620.
84. Eberius M., Lima-Guerra J. High-throughput plant phenotyping—data acquisition, transformation, and analysis // *Bioinformatics: Tools and Applications*. P.: Springer New York, 2009. P. 259-278.
85. Edmonds J. M. Epidermal hair morphology in *Solanum* L. section *Solanum* // *Botanical Journal of the Linnean Society*. – 1982. – V. 85. – № 3. – P. 153-167.
86. Ehleringer J. Leaf absorptance and photosynthesis as affected by pubescence in the genus *Encelia* // *Annual Report Director Department Plant Biology, Carnegie Institution*. – 1976. – P. 413-418.
87. Ehleringer J. R., Mooney H. A. Leaf hairs: effects on physiological activity and adaptive value to a desert shrub // *Oecologia*. – 1978. – V. 37. – № 2. – P. 183-200.
88. Ehleringer J. The influence of water stress and temperature on leaf pubescence development in *Encelia farinosa* // *American Journal of Botany*. – 1982. – V. 69. – № 5. – P. 670-675.
89. Ehleringer J., Björkman O., Mooney H. A. Leaf pubescence: effects on absorptance and photosynthesis in a desert shrub // *Science*. – 1976. – V. 192. – № 4237. – P. 376-377.
90. Eliceiri K.W., Berthold M.R., Goldberg I.G., et al. Biological imaging software tools // *Nature methods*. – 2012. – V. 9. – № 7. – P. 697-710.
91. Elshire R.J., Glaubit J.C., Sun Q., et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species // *PloS One*. – 2011. – V. 6. – № 5. – P. e19379.
92. Evers A. D. et al. Predicting milling extraction rate by image analysis of wheat grains // *Aspects of Applied Biology*. – 1990. – № 25. – P. 417-426.

93. Evers, A.D. (2000). Grain size and morphology: Implications for quality. In Wheat Structure, Biochemistry and Functionality, D. Schofield, ed (London: Royal Society of Chemistry), pp. 19–24
94. Exner V., Hirsch-Hoffmann M., Gruissem W., Hennig L. PlantDB—a versatile database for managing plant research // Plant Methods. – 2008. – V. 4. – P. 1.
95. Fabre J. et al. PHENOPSIS DB: an Information System for *Arabidopsis thaliana* phenotypic data in an environmental context // BMC Plant Biology. – 2011. – V. 11. – № 1. – P. 77.
96. Failmezger H. et al. Semi-automated 3D leaf reconstruction and analysis of trichome patterning from light microscopic images // PLoS computational biology. – 2013. – V. 9. – № 4. – P. e1003029.
97. Farris E. et al. Morpho-colorimetric characterization of the Sardinian endemic taxa of the genus *Anchusa* L. by seed image analysis // Plants. – 2020. – V. 9. – № 10. – P. 1321.
98. Feldman M. et al. Genomic asymmetry in allopolyploid plants: wheat as a model // Journal of experimental botany. – 2012. – V. 63. – № 14. – P. 5045-5059.
99. Feldman M., Levy A. A. Genome evolution due to allopolyploidization in wheat // Genetics. – 2012. – V. 192. – № 3. – P. 763-774.
100. Fenner M. Environmental influences on seed size and composition // Horticultural reviews. – 1992. – V. 13. – P. 183-213.
101. Flanders K. L. et al. Insect resistance in potatoes: sources, evolutionary relationships, morphological and chemical defenses, and ecogeographical associations // Euphytica. – 1992. – V. 61. – № 2. – P. 83-111.
102. Franckowiak J. Revised linkage maps for morphological markers in barley, *Hordeum vulgare* // Barley genetic newsletter. – 1997. – V. 26. – P. 9-

- 21.
103. Freimer N., Sabatti C. The human phenome project // *Nature genetics*. – 2003. – V. 34. – № 1. – P. 15-21.
104. Fridman E. et al. Metabolic, genomic, and biochemical analyses of glandular trichomes from the wild tomato species *Lycopersicon hirsutum* identify a key enzyme in the biosynthesis of methylketones // *The Plant Cell*. – 2005. – V. 17. – № 4. – P. 1252-1267.
105. Frisch M., Bohn M., Melchinger A.A. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene // *Crop Science*. – 1999. – V. 39. – № 4. – P. 967-975.
106. Fukunaga K. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 1972.
107. Furbank R.T., Tester M. Phenomics–technologies to relieve the phenotyping bottleneck // *Trends in plant science*. – 2011. – V. 16. – № 12. – P. 635-644.
108. Galmés J., Medrano H., Flexas J. Photosynthesis and photoinhibition in response to drought in a pubescent (var. minor) and a glabrous (var. palau) variety of *Digitalis minor* // *Environmental and Experimental Botany*. – 2007. – V. 60. – № 1. – P. 105-111.
109. Gang D. R. et al. An investigation of the storage and biosynthesis of phenylpropenes in sweet basil // *Plant Physiology*. – 2001. – V. 125. – № 2. – P. 539-555.
110. Gautier M. F. et al. Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid *Triticum* species // *Plant Science*. – 2000. – V. 153. – № 1. – P. 81-91.
111. Gegas V. C. et al. A genetic framework for grain size and shape variation

- in wheat //The Plant Cell. – 2010. – V. 22. – № 4. – P. 1046-1056.
112. Genaev M. A. et al. Extraction of quantitative characteristics describing wheat leaf pubescence with a novel image-processing technique //Planta. – 2012. – V. 236. – № 6. – P. 1943-1954.
113. Genaev M. A. et al. Image-based wheat fungi diseases identification by deep learning //Plants. – 2021. – V. 10. – № 8. – P. 1500.
114. Genaev M., Skolotneva E., Afonnikov D. Application of neural networks to image recognition of wheat rust diseases //2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB). – IEEE, 2020. – P. 40-42.
115. Gershenzon J., Dudareva N. The function of terpene natural products in the natural world //Nature chemical biology. – 2007. – V. 3. – № 7. – P. 408-414.
116. Gianoli E., González-Teuber M. Environmental heterogeneity and population differentiation in plasticity to drought in *Convolvulus chilensis* (Convolvulaceae) //Evolutionary Ecology. – 2005. – V. 19. – № 6. – P. 603-613.
117. Gibson R. W. Glandular hairs providing resistance to aphids in certain wild potato species //Annals of Applied Biology. – 1971. – V. 68. – № 2. – P. 113-119.
118. Gibson R. W., Pickett J. A. Wild potato repels aphids by release of aphid alarm pheromone //Nature. – 1983. – V. 302. – № 5909. – P. 608-609.
119. Gibson R. W., Turner R. H. Insect-trapping hairs on potato plants //PANS. – 1977. – V. 23. – № 3. – P. 272-277.
120. Gill B. S. et al. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium //Genetics. – 2004. – V. 168. – № 2. – P. 1087-1096.

121. Glagoleva A. Y., Shoeva O. Y., Khlestkina E. K. Melanin pigment in plants: Current knowledge and future perspectives //Frontiers in Plant Science. – 2020. – V. 11. – P. 770.
122. Glas J. J. et al. Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores //International journal of molecular sciences. – 2012. – V. 13. – № 12. – P. 17077-17103.
123. Golzarian M.R., Frick R.A., Rajendran K., et al. Accurate inference of shoot biomass from high-throughput images of cereal plants // Plant Methods. – 2011. – V. 7. – № 1. – P. 2.
124. Gomez, L., et al., Seed chitinases // Seed Science Research. – 2002. – V. 2. – № 4. – P. 217-230.
125. Goncharov N. P. Genus *Triticum* L. taxonomy: the present and the future //Plant Systematics and Evolution. – 2011. – V. 295. – № 1. – P. 1-11.
126. Gong Z. et al. Recent developments of seeds quality inspection and grading based on machine vision //2015 ASABE Annual International Meeting. – American Society of Agricultural and Biological Engineers, 2015. – P. 1.
127. Goriewa-Duba K. et al. An evaluation of the variation in the morphometric parameters of grain of six *Triticum* species with the use of digital image analysis //Agronomy. – 2018. – V. 8. – № 12. – P. 296.
128. Granitto P. M., Verdes P. F., Ceccatto H. A. Large-scale investigation of weed seed identification by machine vision //Computers and Electronics in Agriculture. – 2005. – V. 47. – № 1. – P. 15-24.
129. Gregory P, Tingey Ward M, Ave Dirk A, Bouthyette Pierre Y Potato glandular trichomes: a physicochemical defense mechanism against insects.

- In: Maurice B, Green, Paul, Hedin A (eds) Natural resistance of plants to pests, V. 296. American Chemical Society, Washington city. 1986. P. 160–167.
130. Groos C. et al. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white× red grain bread-wheat cross //Theoretical and Applied Genetics. – 2002. – V. 104. – № 1. – P. 39-47.
131. Gu Z. et al. Circlize implements and enhances circular visualization in R //Bioinformatics. – 2014. – V. 30. – № 19. – P. 2811-2812.
132. Guberac V. et al. Hermetically sealed storage of cereal seeds and its influence on vigour and germination //Journal of Agronomy and Crop Science. – 2003. – V. 189. – № 1. – P. 54-56.
133. Guyon I. et al. Gene selection for cancer classification using support vector machines //Machine learning. – 2002. – V. 46. – № 1. – P. 389-422.
134. Guzzon F. et al. Seed longevity of maize conserved under germplasm bank conditions for up to 60 years //Annals of Botany. – 2021. – V. 127. – № 6. – P. 775-785.
135. Hairmansis A., Berger B., Tester M., Roy S.J. Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice //Rice. – 2014. – V. 7. – № 1. – P. 16.
136. Hamaoka N. et al. A hairy-leaf gene, BLANKET LEAF, of wild *Oryza nivara* increases photosynthetic water use efficiency in rice //Rice. – 2017. – V. 10. – № 1. – P. 1-11.
137. Hameed M. et al. Variation in leaf anatomy in wheat germplasm from varying drought-hit habitats //Int. J. Agric. Biol. – 2002. – V. 4. – № 1. – P. 12-16.

138. Hameed M., Ashraf M., Ahmad M. S. A., Naz N. Structural and functional adaptations in plants for salinity tolerance // In Plant adaptation and phytoremediation. / Eds. Ashraf M., Ozturk M., Ahmad M. S. A. Springer Science+Business Media B.V. Dordrecht. 2010.P. 151. doi 10.1007/978-90-481-9370-7_8.
139. Hammer Ø. et al. PAST: Paleontological statistics software package for education and data analysis //Palaeontologia electronica. – 2001. – V. 4. – № 1. – P. 9.
140. Hartmann A., Czauderna T., Hoffmann R., et al. HTPheno: an image analysis pipeline for high-throughput plant phenotyping // BMC bioinformatics. – 2011. – V. 12. – № 1. – P. 148.
141. Hay, F.R. New technologies to improve the ex situ conservation of plant genetic resources. In Plant genetic resources: A re-view of current research and future needs; Ehsan Dulloo M., Ed.; Publisher: Burleigh Dodds Science Publishing, Cambridge, UK, 2021; pp. 185-216.
142. Hayat S. Hayat Q., Alyemeni M.N., Wani A.S., Pichtel J., Ahmad A. Role of proline under changing environments: a review // Plant Signal Behav. – 2012. – V.7. – P. 1456.
143. Herridge R.P., Day R.C., Baldwin S., Macknight R.C. Rapid analysis of seed size in Arabidopsis for mutant and QTL discovery // Plant methods. – 2011. – V. 7. – № 1. – P. 3.
144. Holland S.M. Non-metric multidimensional scaling (MDS). Department of Geology. University of Georgia, Athens, GA. 2008.
145. Hong M. et al. Transcriptomic analysis of seed coats in yellow-seeded Brassica napus reveals novel genes that influence proanthocyanidin biosynthesis //Frontiers in plant science. – 2017. – V. 8. – P. 1674.

146. Honkavaara E., Saari H., Kaivosoja J., et al. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture // *Remote Sensing*. – 2013. – V. 5. – № 10. – P. 5006-5039.
147. Houle D., Govindaraju D.R., Omholt S. Phenomics: the next challenge // *Nature Reviews Genetics*. – 2010. – V. 11. – № 12. – P. 855-866.
148. Huang M. et al. Review of seed quality and safety tests using optical sensing technologies // *Seed Science and Technology*. – 2015. – V. 43. – № 3. – P. 337-366.
149. Huang X. et al. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers // *Theoretical and Applied Genetics*. – 2002. – V. 105. – № 5. – P. 699-707.
150. Huang Y. et al. Characterization of three wheat grain weight QTLs that differentially affect kernel dimensions // *Theoretical and Applied Genetics*. – 2015. – V. 128. – № 12. – P. 2437-2445.
151. Hulskamp M., Schnittger A., Folkers U. Pattern formation and cell differentiation: trichomes in *Arabidopsis* as a genetic model system // *International review of cytology*. – 1998. – V. 186. – P. 147-178.
152. Ilyas M. et al. Drought tolerance strategies in plants: a mechanistic approach // *Journal of Plant Growth Regulation*. – 2021. – V. 40. – № 3. – P. 926-944.
153. Imtiaz M. et al. Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines // *Genetics*. – 2008. – V. 178. – № 3. – P. 1725-1736.
154. Iyer-Pascuzzi A.S., Symonova O., Mileyko Y., et al. Imaging and

- analysis platform for automatic phenotyping and trait ranking of plant root systems // *Plant Physiology*. – 2010. – V. 152. – № 3. – P. 1148-1157.
155. Jahnke S. et al. Pheno seeder-a robot system for automated handling and phenotyping of individual seeds // *Plant physiology*. – 2016. – V. 172. – № 3. – P. 1358-1370.
156. Janampa-Santome M. et al. Influence of seed color on germination performance and storage organ color in the high-Andean crop *Lepidium meyenii* (Maca) // *Archives of Agronomy and Soil Science*. – 2021. – V. 67. – № 8. – P. 1093-1103.
157. Jansen M., Gilmer F., Biskup B., et al. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants // *Functional Plant Biology*. – 2009. – V. 36. – № 11. – P. 902-914.
158. Jantasuriyarat C. et al. Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (*Triticum aestivum* L.) // *Theoretical and Applied Genetics*. – 2004. – V. 108. – № 2. – P. 261-273. DOI 10.1007/s00122-003-1432-8
159. Jofuku K. D. et al. Control of seed mass and seed yield by the floral homeotic gene *APETALA2* // *Proceedings of the National Academy of Sciences*. – 2005. – V. 102. – № 8. – P. 3117-3122.
160. John G. H., Langley P. Estimating continuous distributions in Bayesian classifiers // *arXiv preprint arXiv:1302.4964*. – 2013.
161. Johnson H. B. Plant pubescence: an ecological perspective // *The Botanical Review*. – 1975. – V. 41. – № 3. – P. 233-258.
162. Joosen R. V. L. et al. GERMINATOR: a software package for high-

- throughput scoring and curve fitting of Arabidopsis seed germination //The Plant Journal. – 2010. – V. 62. – № 1. – P. 148-159.
163. Kaehler, A.; Bradski, G. Learning OpenCV 3: computer vision in C++ with the OpenCV library. O'Reilly Media, Inc.: 2016.
164. Kaminuma E., Heida N., Tsumoto Y., Yamamoto N., Goto N., Okamoto N., Konagaya A., Matsui M., Toyoda T. Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis // Plant J. – 2004. – V. 38. – P. 358-365.
165. Kaminuma E., Yoshizumi T., Wada T., et al. Quantitative analysis of heterogeneous spatial distribution of Arabidopsis leaf trichomes using micro X-ray computed tomography // The Plant Journal. – 2008. – V. 56. – № 3. – P. 470-482.
166. Kanehisa M. et al. KEGG: new perspectives on genomes, pathways, diseases and drugs //Nucleic acids research. – 2017. – V. 45. – № D1. – P. D353-D361.
167. Kanehisa M., Sato Y., Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences //Journal of molecular biology. – 2016. – V. 428. – № 4. – P. 726-731.
168. Kang J. H. et al. Distortion of trichome morphology by the hairless mutation of tomato affects leaf surface chemistry //Journal of experimental botany. – 2010. – V. 61. – № 4. – P. 1053-1064.
169. Karpova E. K. et al. Quantifying Drosophila adults with the use of a smartphone // Biology open. – 2020. – V. 9. – № 10. – P. bio054452.
170. Kearsey M.J. The principles of QTL analysis (a minimal mathematics

- approach) // *Journal of Experimental Botany*. – 1998. – V. 49. – № 327. – P. 1619-1623.
171. Kearsey M.J., Farquhar A.G.L. QTL analysis in plants; where are we now? // *Heredity*. – 1998. – V. 80. – № 2. – P. 137-142.
172. Keerthi S. S. et al. Improvements to Platt's SMO algorithm for SVM classifier design // *Neural computation*. – 2001. – V. 13. – № 3. – P. 637-649.
173. Kennedy G. G. Tomato, pests, parasitoids, and predators: tritrophic interactions involving the genus *Lycopersicon* // *Annual Review of Entomology*. – 2003. – V. 48. – № 1. – P. 51-72.
174. Khlestkina E. K. et al. Prospects of molecular genetic approaches in controlling technological properties of wheat grain in the context of the "grain–flour–bread" chain // *Vavilov Journal of Genetics and Breeding*. – 2016. – V. 20. – № 4. – P. 511-527.
175. Khlestkina E. K. Genes determining the coloration of different organs in wheat // *Russian Journal of Genetics: Applied Research*. – 2013. – V. 3. – № 1. – P. 54-65.
176. Khlestkina E. The adaptive role of flavonoids: emphasis on cereals // *Cereal Research Communications*. – 2013. – V. 41. – № 2. – P. 185-198.
177. Kibar H. Determining the functional characteristics of wheat and corn grains depending on storage time and temperature // *Journal of Food Processing and Preservation*. – 2016. – V. 40. – № 4. – P. 749-759.
178. Kibar H., Kılıç İ. Mineral composition and technological properties of einkorn wheat as affected by storage conditions // *Journal of Food Processing and Preservation*. – 2020. – V. 44. – № 12. – P. e14951.
179. Klukas C., Chen D., Pape J.M. Integrated analysis platform: an open-

- source information system for high-throughput plant phenotyping // *Plant physiology*. – 2014. – V. 165. – № 2. – P. 506-518.
180. Köhl K.I., Basler G., Lüdemann A., et al. A plant resource and experiment management system based on the Golm Plant Database as a basic tool for omics research // *Plant Methods*. – 2008. – V. 4. – P. 11.
181. Köhl K.I., Gremmels J. Documentation system for plant transformation service and research // *Plant methods*. – 2010. – V. 6. – P. 4.
182. Kolukisaoglu Ü., Thurow K. Future and frontiers of automated screening in plant sciences // *Plant Science*. – 2010. – V. 178. – № 6. – P. 476-484.
183. Komyshev E. G., Genaev M. A., Afonnikov D. A. Analysis of color and texture characteristics of cereals on digital images // *Vavilov Journal of Genetics and Breeding*. – 2020. – V. 24. – № 4. – P. 340.
184. Konopatskaia I. et al. Spike Morphology Genes in Wheat Species (*Triticum L.*) // *Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences*. – 2016. – V. 70. – № 6. – P. 345-355.
185. Korte A, Vilhjalmsson B.J., Segura V., et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations // *Nature genetics*. – 2012. – V. 44. – № 9. – P. 1066-1071.
186. Kowalski S. P. et al. Purification and characterization of polyphenol oxidase from glandular trichomes of *Solanum berthaultii* // *Plant Physiology*. – 1992. – V. 100. – № 2. – P. 677-684.
187. Kozlov K., Pisarev A., Matveeva A., et al. Image Processing Package ProStack for Quantification of Biological Images // *Proceedings of the 4th International Symposium on Networks in Bioinformatics (ISNB)*. – 2007. – P.

- 204.
188. Krajewski P., Chen D., Cwiek H., et al. Towards recommendations for metadata and data handling in plant phenotyping // *Journal of experimental botany*. – 2015. – V. 66. – № 18. – P. 5417-5427.
189. Kroumova A. B., Wagner G. J. Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants // *Planta*. – 2003. – V. 216. – № 6. – P. 1013-1021.
190. Kroupin P. et al. Effect of dwarfing gene Ddw1 on height and agronomic traits in spring triticale in greenhouse and field experiments in a Non-Black Earth Region of Russia // *Plants*. – 2019. – V. 8. – № 5. – P. 131.
191. Kroupin P. Y. et al. Effects of Rht17 in combination with Vrn-B1 and Ppd-D1 alleles on agronomic traits in wheat in black earth and non-black earth regions // *BMC plant biology*. – 2020. – V. 20. – № 1. – P. 1-17.
192. Kuchlan M. K., Dadlani M., Samuel D. V. K. Seed coat properties and longevity of soybean seeds // *Journal of New seeds*. – 2010. – V. 11. – № 3. – P. 239-249.
193. Kuhl F. P., Giardina C. R. Elliptic Fourier features of a closed contour // *Computer graphics and image processing*. – 1982. – V. 18. – № 3. – P. 236-258
194. Kumar A. et al. Dissection of genetic factors underlying wheat kernel shape and size in an elite× nonadapted cross using a high density SNP linkage map // *The plant genome*. – 2016. – V. 9. – № 1. – P. plantgenome2015.09.0081.
195. Kumar A. et al. QTL analysis for grain colour and pre-harvest sprouting in bread wheat // *Plant science*. – 2009. – V. 177. – № 2. – P. 114-122.

196. Kumar N., Belhumeur P.N., Biswas A., et al. Leafsnap: A computer vision system for automatic plant species identification // Computer Vision—ECCV 2012. Springer Berlin Heidelberg, – 2012. – P. 502-516.
197. Kumari S. et al. QTL mapping for some grain traits in bread wheat (*Triticum aestivum* L.) // Physiology and Molecular Biology of Plants. – 2018. – V. 24. – № 5. – P. 909-920.
198. Kurek K., Plitta-Michalak B., Ratajczak E. Reactive oxygen species as potential drivers of the seed aging process // Plants. – 2019. – V. 8. – № 6. – P. 174.
199. Lacape J-M., Nguyen T.B. Mapping Quantitative Trait Loci Associated with Leaf and Stem Pubescence in Cotton // Journal of Heredity. – 2005. – Vol. 96. – P. 441-444.
200. Lachman J. et al. Genetics and chemistry of pigments in wheat grain—A review // Journal of Cereal Science. – 2017. – V. 74. – P. 145-154.
201. Lang J. et al. Myb10-D confers PHS-3D resistance to pre-harvest sprouting by regulating NCED in ABA biosynthesis pathway of wheat // New Phytologist. – 2021. – V. 230. – № 5. – P. 1940-1952.
202. Lapochkina I. F. et al. Identification of alien genetic material and genes of resistance to leaf rust in wheat (*Triticum aestivum* L.) stocks // Proc. of Tenth Int. Wheat Genetics Symp. – 2003. – V. 3. – P. 1190-1192.
203. Lapochkina I. F. Genetic diversity of “Arsenal” collection and its use in wheat breeding // Genetic Resources of Cultural Plants. – 2001. – P. 13-16.
204. Le Cessie S., Van Houwelingen J. C. Ridge estimators in logistic regression // Journal of the Royal Statistical Society: Series C (Applied Statistics). – 1992. – V. 41. – № 1. – P. 191-201.

205. Lee J.M., Davenport G.F., Marshall D., Ellis T.H., Ambrose M.J., Dicks J., van Hintum T.J., Flavell A.J. GERMINATE. a generic database for integrating genotypic and phenotypic information for plant genetic resource collections // *Plant Physiol.* – 2005. – V.139. – P. 619-631.
206. Lee K. H. et al. Grain width 2 (GW2) and its interacting proteins regulate seed development in rice (*Oryza sativa* L.) // *Botanical studies.* – 2018. – V. 59. – № 1. – P. 1-7.
207. Leifer A. M. et al. Optogenetic manipulation of neural activity in freely moving *Caenorhabditis elegans* // *Nature methods.* – 2011. – V. 8. – № 2. – P. 147-152.
208. Leister D. et al. Large-scale evaluation of plant growth in *Arabidopsis thaliana* by non-invasive image analysis // *Plant Physiology and Biochemistry.* – 1999. – V. 37. – № 9. – P. 671-678.
209. Leonova I. N. et al. Comparative molecular and genetic analysis of *Triticum aestivum* × *Triticum timopheevii* hybrid lines resistant to leaf rust. EWAC newsletter, 2001 // *Proceedings of the 11th EWAC conference, Novosibirsk.* – 2001. – P. 140-143.
210. Li B. et al. Two novel AP2/EREBP transcription factor genes TaPARG have pleiotropic functions on plant architecture and yield-related traits in common wheat // *Frontiers in plant science.* – 2016. – V. 7. – P. 1191.
211. Li J. et al. Glabrous Rice 1, encoding a homeodomain protein, regulates trichome development in rice // *Rice.* – 2012. – V. 5. – № 1. – P. 1-10.
212. Li L., Steffens J. C. Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance // *Planta.* – 2002. – V. 215. – № 2. – P. 239-247.
213. Li N. et al. Control of grain size in rice // *Plant Reproduction.* – 2018. –

- V. 31. – № 3. – P. 237-251.
214. Li N., Li Y. Ubiquitin-mediated control of seed size in plants //Frontiers in plant science. – 2014. – V. 5. – P. 332.
215. Li N., Xu R., Li Y. Molecular networks of seed size control in plants //Annual review of plant biology. – 2019. – V. 70. – P. 435-463.
216. Li S. et al. Dissection of genetic basis underpinning kernel weight-related traits in common wheat //Plants. – 2021. – V. 10. – № 4. – P. 713.
217. Li W. et al. Characterization and fine mapping of the glabrous leaf and hull mutants (gl1) in rice (*Oryza sativa* L.) //Plant cell reports. – 2010. – V. 29. – № 6. – P. 617-627.
218. Li Y. et al. Total phenolic, flavonoid content, and antioxidant activity of flour, noodles, and steamed bread made from different colored wheat grains by three milling methods //The Crop Journal. – 2015. – V. 3. – № 4. – P. 328-334.
219. Li Y.H., Zhao S.C., Ma J.X., et al. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing // BMC genomics. – 2013. – V. 14. – № 1. – P. 579.
220. Liakopolos G., Nikolopoulos D., Klouvatou A., Vekkos K.-A., Manetas Y., Karabouriontis G. The Photoprotective Role of Epidermal Anthocyanins and Surface Pubescence in Young Leaves of Grapevine (*Vitis vinifera*) // Annals of Botany. – 2006. – Vol.98. – P. 257-265.
221. Lijavetzky D., Cabezas J.A., Ibáñez A., et al. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology // BMC genomics. – 2007. – V. 8. – № 1. P. 424.

222. Lin M. et al. Genome-wide association analysis on pre-harvest sprouting resistance and grain color in US winter wheat //BMC genomics. – 2016. – V. 17. – № 1. – P. 1-16.
223. Liu J. et al. The nightshade proteinase inhibitor IIb gene is constitutively expressed in glandular trichomes //Plant and Cell Physiology. – 2006. – V. 47. – № 9. – P. 1274-1284.
224. Lobet G., Draye X., Périlleux C. An online database for plant image analysis software tools //Plant methods. – 2013. – V. 9. – № 1. – P. 38.
225. Lobo A. K. M. et al. Overexpression of calpase decreases Rubisco abundance and grain yield in wheat //Plant physiology. – 2019. – V. 181. – № 2. – P. 471-479
226. Long S. P., Ort D. R. More than taking the heat: crops and global change //Current opinion in plant biology. – 2010. – V. 13. – № 3. – P. 240-247.
227. Loskutov I. G., Khlestkina E. K. Wheat, barley, and oat breeding for health benefit components in grain //Plants. – 2021. – V. 10. – № 1. – P. 86.
228. Lu J., Sang N. Detecting citrus fruits and occlusion recovery under natural illumination conditions // Computers and Electronics in Agriculture. – 2015. – V. 110. – P. 121-130.
229. Luckwill, L. The genus *Lycopersicon*: historical, biological, and taxonomic survey of the wild and cultivated tomatoes; Aberdeen University Press, Aberdeen, Scotland, 1943
230. Luo D., Oppenheimer D. G. Genetic control of trichome branch number in *Arabidopsis*: the roles of the *FURCA* loci //Development. – 1999. – V. 126. – № 24. – P. 5547-5557.
231. Lv Q. et al. Wheat E3 ubiquitin ligase TaGW2-6A degrades TaAGPS to

- affect seed size //Plant Science. – 2022. – V. 320. – P. 111274.
232. Lynch J.P. Roots of the second green revolution // Australian Journal of Botany. – 2007. – V. 55. – № 5. – P. 493-512.
233. Lyshede O. B. The ultrastructure of the glandular trichomes of *Solanum tuberosum* //Annals of Botany. – 1980. – V. 46. – № 5. – P. 519-526.
234. Ma D. et al. Characterization of a cell wall invertase gene TaCwi-A1 on common wheat chromosome 2A and development of functional markers //Molecular Breeding. – 2012. – V. 29. – № 1. – P. 43-52.
235. Ma D. et al. Dynamic Metabolomics and Transcriptomics Analyses for Characterization of Phenolic Compounds and Their Biosynthetic Characteristics in Wheat Grain //Frontiers in Nutrition. – 2022. – V. 9.
236. Ma D. et al. Wheat grain phenolics: a review on composition, bioactivity, and influencing factors //Journal of the Science of Food and Agriculture. – 2021. – V. 101. – № 15. – P. 6167-6185.
237. Ma L. et al. Ta GS 5-3A, a grain size gene selected during wheat improvement for larger kernel and yield //Plant Biotechnology Journal. – 2016. – V. 14. – № 5. – P. 1269-1280.
238. Ma M. et al. Expression of Ta CYP 78A3, a gene encoding cytochrome P450 CYP 78A3 protein in wheat (*Triticum aestivum* L.), affects seed size //The Plant Journal. – 2015. – V. 83. – № 2. – P. 312-325.
239. Maharijaya A., Vosman B. Managing the Colorado potato beetle; the need for resistance breeding //Euphytica. – 2015. – V. 204. – № 3. – P. 487-501.
240. Majumdar S., Jayas D. S. Classification of bulk samples of cereal grains using machine vision //Journal of Agricultural Engineering Research. – 1999.

- V. 73. – № 1. – P. 35-47.
241. Mardis E.R. The impact of next-generation sequencing technology on genetics // Trends in genetics. 2008. – V. 24. – № 3. – P. 133-141.
242. Martín-Gómez J. J. et al. Morphological description and classification of wheat kernels based on geometric models // Agronomy. – 2019. – V. 9. – № 7. – P. 399.
243. Matilla A. J., Carrillo-Barral N., Rodríguez-Gacio M. C. An update on the role of NCED and CYP707A ABA metabolism genes in seed dormancy induction and the response to after-ripening and nitrate // Journal of Plant Growth Regulation. – 2015. – V. 34. – № 2. – P. 274-293.
244. Matsuoka Y. Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification // Plant and cell physiology. – 2011. – V. 52. – № 5. – P. 750-764.
245. Mattioli R., Costantino P., Trovato M. Proline accumulation in plants: not only stress. // Plant Signal Behav. – 2009. – V.4. – P. 1016.
246. McCaig T. N., DePauw R. M., Williams P. C. Assessing seed-coat color in a wheat breeding program with a NIR/VIS instrument // Canadian Journal of Plant Science. – 1993. – V. 73. – № 2. – P. 535-539.
247. McCauley M. M., Evert R. F. Morphology and vasculature of the leaf of potato (*Solanum tuberosum*) // American journal of botany. – 1988. – P. 377-390.
248. McIntosh R. A. et al. Catalogue of gene symbols for wheat // Proceedings of the 12th International Wheat Genetics Symposium. – Yokohama Japan, 2013. – P. 8-13.

249. McLachlan G. J. Discriminant Analysis and Statistical Pattern Recognition. — Wiley Interscience, 2004.
250. Meinshausen N., Bühlmann P. Stability selection // Journal of the Royal Statistical Society: Series B (Statistical Methodology). — 2010. — V. 72. — № 4. — P. 417-473.
251. Meuwissen T.H.E., Hayes B.J., Goddard M.E. Prediction of total genetic value using genome-wide dense marker maps // Genetics. — 2001. — V. 157. — № 4. — P. 1819-1829.
252. Miller M.R., Dunham J.P., Amores A., et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers // Genome research. — 2007. — V. 17. — № 2. — P. 240-248.
253. Miller N. D. et al. A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images // The Plant Journal. — 2017. — V. 89. — № 1. — P. 169-178.
254. Mira S., Estrelles E., González-Benito M. E. Effect of water content and temperature on seed longevity of seven Brassicaceae species after 5 years of storage // Plant Biology. — 2015. — V. 17. — № 1. — P. 153-162.
255. Mirnezami S. V., Young T., Assefa T., Prichard S., Nagasubramanian K., Sandhu K., Sarkar S., Sundararajan S., et al. Automated trichome counting in soybean using advanced image-processing techniques // Applications in Plant Sciences. — 2020. — V. 8. — № 7. — P. e11375.
256. Mohamed-Yasseen Y. et al. The role of seed coats in seed viability // The Botanical Review. — 1994. — V. 60. — № 4. — P. 426-439.
257. Morales F. et al. Trichomes and photosynthetic pigment composition

- changes: responses of *Quercus ilex* subsp. *ballota* (Desf.) Samp. and *Quercus coccifera* L. to Mediterranean stress conditions // *Trees*. – 2002. – V. 16. – № 7. – P. 504-510.
258. Morris C. F. Puroindolines: the molecular genetic basis of wheat grain hardness // *Plant molecular biology*. – 2002. – V. 48. – № 5. – P. 633-647.
259. Nagata T. et al. γ -Radiation induces leaf trichome formation in *Arabidopsis* // *Plant physiology*. – 1999. – V. 120. – № 1. – P. 113-120.
260. Nagel K.A., Putz A., Gilmer F., et al. GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons // *Functional Plant Biology*. – 2012. – V. 39. – № 11. – P. 891-904.
261. Nagel M., Börner A. The longevity of crop seeds stored under ambient conditions // *Seed Science Research*. – 2010. – V. 20. – № 1. – P. 1-12.
262. Nalam V. J. et al. Map-based analysis of genetic loci on chromosome 2D that affect glume tenacity and threshability, components of the free-threshing habit in common wheat (*Triticum aestivum* L.) // *Theoretical and Applied Genetics*. – 2007. – V. 116. – № 1. – P. 135-145.
263. Nasehzadeh M., Ellis R. H. Wheat seed weight and quality differ temporally in sensitivity to warm or cool conditions during seed development and maturation // *Annals of botany*. – 2017. – V. 120. – № 3. – P. 479-493.
264. Neal J. J., Plaisted R. L., Tingey W. M. Feeding behavior and survival of Colorado potato beetle, *Leptinotarsa decemlineata* (Say), larvae on *Solanum berthaultii* Hawkes and an F6 *S. tuberosum* L. X *S. berthaultii* hybrid // *American Potato Journal*. – 1991. – V. 68. – № 10. – P. 649-658.
265. Neal J. J., Tingey W. M., Steffens J. C. Sucrose esters of carboxylic acids in glandular trichomes of *Solanum berthaultii* deter settling and probing by

- green peach aphid //Journal of Chemical Ecology. – 1990. – V. 16. – № 2. – P. 487-497.
266. Next Instruments (2015). Seedcount. Accessed November 2016.
267. Novaro P., Colucci F., Venora G., D'Egidio M.G. Image analysis of whole grains: a noninvasive method to predict semolina yield in durum wheat // Cereal chemistry. – 2001. – V. 78. – № 3. – P. 217-221.
268. Oksanen J. Multivariate Analysis of Ecological Communities in R: vegan tutorial. 2011. available at <http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf>
269. Olgun M. et al. Wheat grain classification by using dense SIFT features with SVM classifier //Computers and Electronics in Agriculture. – 2016. – V. 122. – P. 185-190.
270. Osmont K.S., Sibout R., Hardtke C.S. Hidden branches: developments in root system architecture // Annu. Rev. Plant Biol. – 2007. – V. 58. – P. 93-113.
271. Otsu N. A threshold selection method from gray-level histograms //IEEE transactions on systems, man, and cybernetics. – 1979. – V. 9. – № 1. – P. 62-66.
272. Paproki A., Sirault X., Berry S., et al. A novel mesh processing based technique for 3D plant analysis // BMC Plant Biology. – 2012. – V. 12. – № 1. – P. 63.
273. Parveen A. et al. Genome-wide analysis of RING-type E3 ligase family identifies potential candidates regulating high amylose starch biosynthesis in wheat (*Triticum aestivum* L.) //Scientific reports. – 2021. – V. 11. – № 1. – P. 1-17.

274. Pathare P.B. et al. Colour measurement and analysis in fresh and processed foods: A review. // Food Bioprocess Technol. – 2013 – V. 6 – № 1 – P. 36-60.
275. Paux E., Sourdille P., Mackay I., Feuillet C. Sequence-based marker development in wheat: advances and applications to breeding // Biotechnology advances. – 2012. – V. 30. – № 5. – P. 1071-1088.
276. Pearson T. High-speed sorting of grains by color and surface texture //Applied engineering in agriculture. – 2010. – V. 26. – № 3. – P. 499-505.
277. Pedregosa F. et al. Scikit-learn: Machine learning in Python //the Journal of machine Learning research. – 2011. – V. 12. – P. 2825-2830.
278. Perazza D. et al. Trichome cell growth in *Arabidopsis thaliana* can be derepressed by mutations in at least five genes //Genetics. – 1999. – V. 152. – № 1. – P. 461-476.
279. Pethybridge S. J., Nelson S. C. Leaf Doctor: A new portable application for quantifying plant disease severity //Plant disease. – 2015. – V. 99. – № 10. – P. 1310-1316.
280. Picotte J. J. et al. Plastic responses to temporal variation in moisture availability: consequences for water use efficiency and plant performance //Oecologia. – 2007. – V. 153. – № 4. – P. 821-832.
281. Pirredda M. et al. Genetic and epigenetic stability in rye seeds under different storage conditions: Ageing and oxygen effect //Plants. – 2020. – V. 9. – № 3. – P. 393.
282. Plaisted R. L., Tingey W. M., Steffens J. C. The germplasm release of NYL 235-4, a clone with resistance to the Colorado potato beetle //American Potato Journal. – 1992. – V. 69. – № 12. – P. 843-846.

283. Polunin D., Shtaiger I., Efimov V. JACOBI4 software for multivariate analysis of biological data //bioRxiv. – 2019. – P. 803684.
284. Pomeranz M., Campbell J., Siegal-Gaskins D., et al. High-resolution computational imaging of leaf hair patterning using polarized light microscopy // The Plant Journal. – 2013. – V. 73. – № 4. – P. 701-708.
285. Pott C. et al. Trichomes on the leaves of *Anomozamites villosus* sp. nov.(Bennettitales) from the Daohugou beds (Middle Jurassic), Inner Mongolia, China: mechanical defence against herbivorous arthropods //Review of Palaeobotany and Palynology. – 2012. – V. 169. – P. 48-60.
286. Pourreza A. et al. Identification of nine Iranian wheat seed varieties by textural analysis with image processing //Computers and electronics in agriculture. – 2012. – V. 83. – P. 102-108.
287. Priestley, D.A. Morphological, Structural, and biochemical changes associated with seed aging. In Seed Aging; Priestley D.A., Ed.; Publisher: Cornell University Press, USA, 1986; pp. 125-196.
288. Pshenichnikova T. A. et al. Diversity of leaf pubescence in bread wheat and relative species //Genetic Resources and Crop Evolution. – 2017. – V. 64. – № 7. – P. 1761-1773.
289. Pshenichnikova T.A., Lapochkina I.F., Shchukina L.V. The inheritance of morphological and biochemical traits introgressed into common wheat (*Triticum aestivum* L.) from *Ae. speltoides* Tausch. // Genetic Resources and Crop Evolution. – 2007. – Vol.54. – P. 287-293.
290. Quinlan J. R. Induction of decision trees //Machine learning. – 1986. – V. 1. – № 1. – P. 81-106.
291. Quintana J., Garcia R., Neumann L. A novel method for color correction

- in epiluminescence microscopy //Computerized Medical Imaging and Graphics. – 2011. – V. 35. – № 7-8. – P. 646-652.
292. Rahman M., Blackwell B., Banerjee N., Saraswat D. Smartphone-based hierarchical crowdsourcing for weed identification // Computers and Electronics in Agriculture. – 2015. – V. 113. – P. 14-23.
293. Rajjou L., Debeaujon I. Seed longevity: survival and maintenance of high germination ability of dry seeds //Comptes rendus biologiques. – 2008. – V. 331. – № 10. – P. 796-805.
294. Ram M. S. et al. Development of standard procedures for a simple, rapid test to determine wheat color class //Cereal chemistry. – 2002. – V. 79. – № 2. – P. 230-237.
295. Ranganathan A. The levenberg-marquardt algorithm //Tutorial on LM algorithm. – 2004. – V. 11. – № 1. – P. 101-110.
296. Ren T. et al. Utilization of a Wheat55K SNP array-derived high-density genetic map for high-resolution mapping of quantitative trait loci for important kernel-related traits in common wheat //Theoretical and Applied Genetics. – 2021. – V. 134. – № 3. – P. 807-821.
297. Reshef D. N. et al. Detecting novel associations in large data sets //science. – 2011. – V. 334. – № 6062. – P. 1518-1524.
298. Roberts J.J., Gallun R.L., Patterson F.L., Forster J.E. Effects of wheat leaf pubescence on the hessian fly // J. Econ. Entomol. – 1979. – V. 72. – № 2. – P. 211-214.
299. Roerdink J. B. T. M., Meijster A. The watershed transform: Definitions, algorithms and parallelization strategies //Fundamenta informaticae. – 2000. – V. 41. – № 1-2. – P. 187-228.

300. Rojas Rodas F. et al. Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1→6) rhamnosyltransferase //Plant molecular biology. – 2014. – V. 84. – № 3. – P. 287-300.
301. Rolland V., Farazi M.R., Conaty, W.C. et al. HairNet: a deep learning model to score leaf hairiness, a key phenotype for cotton fibre yield, value and insect resistance // Plant Methods. – 2022. – V. 18. – P. 8.
302. Roussel J. et al. 3D surface reconstruction of plant seeds by volume carving: performance and accuracies //Frontiers in Plant Science. – 2016. – V. 7. – P. 745.
303. R-project: R: A language and environment for statistical computing, available at <http://www.R-project.org>.
304. Sabanci K., Toktas A., Kayabasi A. Grain classifier with computer vision using adaptive neuro-fuzzy inference system //Journal of the Science of Food and Agriculture. – 2017. – V. 97. – № 12. – P. 3994-4000.
305. Sankaran S., Wang M., Vandemark G. J. Image-based rapid phenotyping of chickpeas seed size //Engineering in agriculture, environment and food. – 2016. – V. 9. – № 1. – P. 50-55.
306. Sano N., Marion-Poll A. ABA metabolism and homeostasis in seed dormancy and germination //International journal of molecular sciences. – 2021. – V. 22. – № 10. – P. 5069.
307. Sapirstein H.D., Neuman M., Wright E.H., et al. An instrumental system for cereal grain classification using digital image analysis // Journal of Cereal Science. – 1987. – V. 6. – № 1. – P. 3-14.
308. Sappl P.G., Heisler M.G. Live-imaging of plant development: latest

- approaches // *Current opinion in plant biology*. – 2013. – V. 16. – № 1. – P. 33-40.
309. Sarhan F. Wheat yields: doubling-up on grain numbers // *Nature plants*. – 2015. – V. 1. – № 2. – P. 1-2.
310. Sarker I. H. Machine learning: Algorithms, real-world applications and research directions // *SN Computer Science*. – 2021. – V. 2. – № 3. – P. 160.
311. Schellmann S., Hulskamp M. Epidermal differentiation: trichomes in *Arabidopsis* as a model system // *International Journal of Developmental Biology*. – 2004. – V. 49. – № 5-6. – P. 579-584.
312. Schillinger Jr J. A., Gallun R. L. Leaf pubescence of wheat as a deterrent to the cereal leaf beetle, *Oulema melanopus* // *Annals of the Entomological Society of America*. – 1968. – V. 61. – № 4. – P. 900-903.
313. Schillinger Jr J. A., Gallun R. L. Leaf pubescence of wheat as a deterrent to the cereal leaf beetle, *Oulema melanopus* // *Annals of the Entomological Society of America*. – 1968. – V. 61. – № 4. – P. 900-903.
314. Schillmiller A. L., Last R. L., Pichersky E. Harnessing plant trichome biochemistry for the production of useful compounds // *The Plant Journal*. – 2008. – V. 54. – № 4. – P. 702-711.
315. Schindelin J., Arganda-Carreras I., Frise E., et al. Fiji: an open-source platform for biological-image analysis // *Nature methods*. – 2012. – V. 9. – № 7. – P. 676-682.
316. Schöler F., Steinhage V. Automated 3D reconstruction of grape cluster architecture from sensor data for efficient phenotyping // *Computers and Electronics in Agriculture*. – 2015. – V. 114. – P. 163-177.
317. Seo J. et al. Seed browning in pepper (*Capsicum annuum* L.) fruit during

- cold storage is inhibited by methyl jasmonate or induced by methyl salicylate //Postharvest Biology and Technology. – 2020. – V. 166. – P. 111210.
318. Shepherd R. W. et al. Phylloplanins of tobacco are defensive proteins deployed on aerial surfaces by short glandular trichomes //The Plant Cell. – 2005. – V. 17. – № 6. – P. 1851-1861.
319. Shive J. W. A three-salt nutrient solution for plants //American Journal of Botany. – 1915. – V. 2. – № 4. – P. 157-160.
320. Shoeva O. Y. et al. Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of *Hordeum vulgare* //PloS one. – 2016. – V. 11. – № 10. – P. e0163782.
321. Shrestha R., Matteis L., Skofic M., et al. Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice // Frontiers in physiology. – 2012. – V. 3. – P. 82-91.
322. Shvachko N. A., Khlestkina E. K. Molecular genetic bases of seed resistance to oxidative stress during storage //Vavilov Journal of Genetics and Breeding. – 2020. – V. 24. – № 5. – P. 451.
323. Simmons A. T., Gurr G. M. Trichomes of *Lycopersicon* species and their hybrids: effects on pests and natural enemies //Agricultural and Forest Entomology. – 2005. – V. 7. – № 4. – P. 265-276.
324. Sivanandam, S. N. Genetic algorithm optimization problems / S.N. Sivanandam, S.N. Deepa // Introduction to Genetic Algorithms. – Springer, Berlin, Heidelberg, 2008. – P. 165-209
325. Smolikova G. N., Medvedev S. S. Seed carotenoids: synthesis, diversity, and functions //Russian Journal of Plant Physiology. – 2015. – V. 62. – № 1. – P. 1-13.

326. Smolikova G., Leonova T., Vashurina N., Frolov A., Medvedev S. Desiccation Tolerance as the Basis of Long-Term Seed Viability // Int. J. Mol. Sci. – 2021. – V. 22. – P. 101 .
327. Smykalova I., Grillo O., Bjelkova M., et al. Phenotypic evaluation of flax seeds by image analysis // Industrial Crops and Products. – 2013. – V. 47. – P. 232-238.
328. Sneath PHA, Sokal RR (1973) Numerical Taxonomy. The Principles and Practice of Numerical Classification. W.H. Freeman and Co, San Francisco
329. Sol Genomic Network. available at <http://solgenomics.net/>.
330. Sood S. et al. The major threshability genes soft glume (sog) and tenacious glume (Tg), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci //Theoretical and Applied Genetics. – 2009. – V. 119. – № 2. – P. 341-351
331. Sreenivasulu N., Schnurbusch T. A genetic playground for enhancing grain number in cereals //Trends in plant science. – 2012. – V. 17. – № 2. – P. 91-101. DOI 10.1016/j.tplants.2011.11.003.
332. Strange H. et al. Automatic estimation of wheat grain morphometry from computed tomography data //Functional Plant Biology. – 2015. – V. 42. – № 5. – P. 452-459
333. Su Z. et al. Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.) //Theoretical and Applied Genetics. – 2011. – V. 122. – № 1. – P. 211-223.
334. Swaminathan, M.; Rao, M., Macro-mutations and sub-specific differentiation in *Triticum*. 1961.

335. Szymanski D. B., Lloyd A. M., Marks M. D. Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis* //Trends in plant science. – 2000. – V. 5. – № 5. – P. 214-219.
336. Tahir A. R. et al. Evaluation of the effect of moisture content on cereal grains by digital image analysis //Food Research International. – 2007. – V. 40. – № 9. – P. 1140-1145.
337. Taketa S., Chang C.L., Ishii M., Takeda K. Chromosome arm location of the gene controlling leaf pubescence of a Chinese local wheat cultivar ‘Hong-mang-mai’// Euphytica. – 2002. – Vol. 125. – P. 141–147.
338. Tan, P. N. Introduction to data mining. / P. N. Tan, M. Steinbach, V. Kumar // 1st. – 2005
339. Tanabata T., Shibaya T., Hori K., et al. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis // Plant physiology. – 2012. – V. 160. – № 4. – P. 1871-1880.
340. Tello-Ruiz M. K. et al. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research //Nucleic Acids Research. – 2021. – V. 49. – № D1. – P. D1452-D1463.
341. Thipyapong P., Joel D. M., Steffens J. C. Differential expression and turnover of the tomato polyphenol oxidase gene family during vegetative and reproductive development //Plant Physiology. – 1997. – V. 113. – № 3. – P. 707-718.
342. Tian D. et al. Role of trichomes in defense against herbivores: comparison of herbivore response to woolly and hairless trichome mutants in tomato (*Solanum lycopersicum*) //Planta. – 2012. – V. 236. – № 4. – P. 1053-1066.
343. Tikhenko N. et al. DEFECTIVE ENDOSPERM-D1 (Dee-D1) is crucial

- for endosperm development in hexaploid wheat //Communications biology. – 2020. – V. 3. – № 1. – P. 1-10.
344. Tingey W. M. et al. Green peach aphid resistance by glandular trichomes in *Solanum tuberosum* x *S. berthaultii* hybrids //American Potato Journal. – 1982. – V. 59. – № 6. – P. 241-251.
345. Tingey W. M., Laubengayer J. E. Defense against the green peach aphid and potato leafhopper by glandular trichomes of *Solanum berthaultii* //Journal of Economic Entomology. – 1981. – V. 74. – № 6. – P. 721-725.
346. Tisné S., Serrand Y., Bach L., et al. Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity // The Plant Journal. – 2013. – V. 74. – № 3. – P. 534-544.
347. Tissier A. Trichome specific expression: promoters and their applications. – INTECH Open Access Publisher, 2012.
348. Topp C.N., Iyer-Pascuzzi A.S., Anderson J.T., et al. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture // Proceedings of the National Academy of Sciences. – 2013. – V. 110. – № 18. – P. e1695-e1704.
349. Treutter D. Significance of flavonoids in plant resistance: a review //Environmental Chemistry Letters. – 2006. – V. 4. – № 3. – P. 147-157.
350. Trovato M., Forlani G., Signorelli S., Funck D. Proline metabolism and its functions in development and stress tolerance // Osmoprotectant-mediated abiotic stress tolerance in plants. /Eds. Hossain M., Kumar V., Burritt D., Fujita M., Mäkelä P. New York. Springer. 2019. 10.1007/978-3-030-27423-8_2.
351. Vadez V. et al. LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget //Journal of Experimental Botany. – 2015. – V. 66. – № 18. – P. 5581-

- 5593.
352. Vadez V., Kholova J., Medina S., et al. Transpiration efficiency: new insights into an old story // *Journal of Experimental Botany*. – 2014. – V. 65. – № 21. – P. 6141-6153.
353. van Rijsbergen CJ (1979) *Information retrieval*, 2nd edn. Butterworths, London
354. Vankadavath R.N., Hussain A.J., Bodanapu R., et al. Computer aided data acquisition tool for high-throughput phenotyping of plant populations // *Plant Methods*. – 2009. – V. 5. – № 1. – P. 18.
355. Varshney R.K., Nayak S.N., May G.D., Jackson S.A. Next-generation sequencing technologies and their implications for crop genetics and breeding // *Trends in biotechnology*. – 2009. – V. 27. – № 9. – P. 522-530.
356. Visscher P.M., Thompson R., Haley C.S. Confidence intervals in QTL mapping by bootstrapping // *Genetics*. – 1996. – V. 143. – № 2. – P. 1013-1020.
357. Wagner G. J. Secreting glandular trichomes: more than just hairs // *Plant physiology*. – 1991. – V. 96. – № 3. – P. 675-679.
358. Wagner G. J., Wang E., Shepherd R. W. New approaches for studying and exploiting an old protuberance, the plant trichome // *Annals of botany*. – 2004. – V. 93. – № 1. – P. 3.
359. Wallach R., Da-Costa N., Raviv M., Moshelion M. Development of synchronized, autonomous, and self-regulated oscillations in transpiration rate of a whole tomato plant under water stress // *Journal of experimental botany*. – 2010. – V. 61. – № 12. – P. 3439-3449.
360. Walter A., Studer B., Kölliker R. Advanced phenotyping offers

- opportunities for improved breeding of forage and turf species // *Annals of botany*. – 2012. – V. 110. – № 6. – P. 1271-1279.
361. Wang E. et al. Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance // *Nature biotechnology*. – 2001. – V. 19. – № 4. – P. 371-374.
362. Wang E., Gan S., Wagner G. J. Isolation and characterization of the CYP71D16 trichome-specific promoter from *Nicotiana tabacum* L // *Journal of experimental botany*. – 2002. – V. 53. – № 376. – P. 1891-1897.
363. Wang J., He J., Han Y., et al. An Adaptive Thresholding algorithm of field leaf image // *Computers and electronics in agriculture*. – 2013. – V. 96. – P. 23-39.
364. Wang S. et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array // *Plant biotechnology journal*. – 2014. – V. 12. – № 6. – P. 787-796.
365. Wettlaufer S. H., Leopold A. C. Relevance of Amadori and Maillard products to seed deterioration // *Plant Physiology*. – 1991. – V. 97. – № 1. – P. 165-169.
366. Whan A.P., Smith A.B., Cavanagh C.R., et al. GrainScan: a low cost, fast method for grain size and colour measurements // *Plant methods*. – 2014. – V. 10. – P. 23.
367. White J.W., Andrade-Sanchez P., Gorea M.A., et al. Field-based phenomics for plant genetics research // *Field Crops Research*. – 2012. – V. 133. – P. 101-112.
368. White J.W., Huntb L.A., Bootec K.J., et al. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data

- standards // Computers and electronics in agriculture. – 2013. – V. 96. – P. 1-12.
369. Wiesnerová D., Wiesner, I. Computer image analysis of seed shape and seed color for flax cultivar description // Computers and electronics in agriculture. – 2008. – V. 61. – № 2. – P. 126-135.
370. Wijekoon C.P., Goodwin P.H., Hsiang T. Quantifying fungal infection of plant leaves by digital image analysis using Scion Image software // Journal of microbiological methods. – 2008. – V. 74. – № 2. – P. 94-101.
371. Williams K., Munkvold J., Sorrells M. Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.) // Euphytica. – 2013. – V. 190. – № 1. – P. 99-116.
372. Williams K., Sorrells M. E. Three-dimensional seed size and shape QTL in hexaploid wheat (*Triticum aestivum* L.) populations // Crop Science. – 2014. – V. 54. – № 1. – P. 98-110.
373. Wishart J., George T.S., Brown L.K., et al. Field phenotyping of potato to assess root and shoot characteristics associated with drought tolerance // Plant and soil. – 2014. – V. 378. – № 1-2. – P. 351-363.
374. Wu W. et al. GainTKW: a measurement system of thousand kernel weight based on the android platform // Agronomy. – 2018. – V. 8. – № 9. – P. 178.
375. Würschum T. Mapping QTL for agronomic traits in breeding populations // Theoretical and Applied Genetics. – 2012. – V. 125. – № 2. – P. 201-210.
376. Xu J., Zhao Q., Du P., et al. Developing high throughput genotyped

- chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.) // BMC genomics. – 2010. – V. 11. – № 1. – P. 656.
377. Xu Z., Zhou G. Responses of leaf stomatal density to water status and its relationship with photosynthesis in a grass // Journal of experimental botany. – 2008. – V. 59. – № 12. – P. 3317-3325.
378. Yang C., Ye Z. Trichomes as models for studying plant cell differentiation // Cellular and molecular life sciences. – 2013. – V. 70. – № 11. – P. 1937-1948.
379. Yang J., Zhang J. Grain filling of cereals under soil drying // New phytologist. – 2006. – V. 169. – № 2. – P. 223-236.
380. Yazdanbakhsh N., Fisahn J. High throughput phenotyping of root growth dynamics, lateral root formation, root architecture and root hair development enabled by PlaRoM // Functional Plant Biology. – 2009. – V. 36. – № 11. – P. 938-946.
381. Youssef H. M. et al. Natural diversity of inflorescence architecture traces cryptic domestication genes in barley (*Hordeum vulgare* L.) // Genetic Resources and Crop Evolution. – 2017. – V. 64. – № 5. – P. 843-853.
382. Yu H., Kowalski S. P., Steffens J. C. Comparison of polyphenol oxidase expression in glandular trichomes of *Solanum* and *Lycopersicon* species // Plant physiology. – 1992. – V. 100. – № 4. – P. 1885-1890.
383. Zadebagheri.M., Azarpanah A. and Javanmardi S. Proline Metabolite Transport an Efficient Approach in corn Yield Improvement as Response to Drought Conditions // Intl. J. Farm & Alli Sci. – 2014. – Vol. 3. – P. 453.
384. Zapotoczny P. Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture // Journal of Cereal

- Science. – 2011. – V. 54. – № 1. – P. 60-68.
385. Zapotoczny P., Zielinska M., Nita Z. Application of image analysis for the varietal classification of barley: Morphological features // Journal of Cereal Science. – 2008. – V. 48. – № 1. – P. 104-110.
386. Zarco-Tejada P.J., Berjón A., López-Lozano R., et al. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy // Remote Sensing of Environment. – 2005. – V. 99. – № 3. – P. 271-287.
387. Zdilla, M.J., Hatfield, S.A., McLean, K.A., et al. Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale: a morphometric analysis with neurosurgical considerations // J. Craniofacial Surg. – 2016. – V. 27. – № 1. – P. 222-228.
388. Zhang J, Liu S, Wu W, Zhong X, Liu T. Research on a rapid identification method for counting universal grain crops // PLoS ONE – 2022. – V. 17. – №9. – P. e0273785.
389. Zhang X. et al. Unconditional and conditional QTL analysis of kernel weight related traits in wheat (*Triticum aestivum* L.) in multiple genetic backgrounds // Genetica. – 2014. – V. 142. – № 4. – P. 371-379.
390. Zhang Y. et al. Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus* // Genome biology. – 2022. – V. 23. – № 1. – P. 86.
391. Zhao Y. et al. Effects of ultrasonic pretreatments on thermodynamic properties, water state, color kinetics, and free amino acid composition in microwave vacuum dried lotus seeds // Drying Technology. – 2020. – V. 38. – № 4. – P. 534-544.

392. Zhao Y. et al. Microwave vacuum drying of lotus seeds: Effect of a single-stage tempering treatment on drying characteristics, moisture distribution, and product quality //Drying Technology. – 2017. – V. 35. – № 13. – P. 1561-1570.
393. Zhu T. et al. Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly //The Plant Journal. – 2021. – V. 107. – № 1. – P. 303-314.
394. Беспалова Л. А. Развитие генофонда как главный фактор третьей зелёной революции в селекции пшеницы //Вестник Российской академии наук. – 2015. – Т. 85. – № 1. – С. 9-11.
395. Вавилов Н.И. Теоретические основы селекции. М.: Наука, 1990. 512 с.
396. Генаев М. А. и др. SpikeDroidDB–информационная система для аннотации морфометрических характеристик колоса пшеницы //Вавиловский журнал генетики и селекции. – 2018. – Т. 22. – № 1. – С. 132-140.
397. Генаев М. А. и др. Информационная поддержка селекционно-генетического эксперимента у пшеницы в системе WheatPGE //Математическая биология и биоинформатика. – 2012. – Т. 7. – № 2. – С. 410-424.
398. Гонсалес Р., Вудс Р. Цифровая обработка изображений. Издание 3-е, исправленное и дополненное. Москва: Техносфера, 2012. – 1104 с.
399. Гончаров Н. П. Сравнительная генетика пшениц и их сородичей. – Новосибирск: Академическое изд-во Гео, 2012. – 523 с.
400. Горбань А. Н., Россиев Д. А. Нейронные сети на персональном

- компьютере. – Новосибирский филиал Федерального государственного унитарного предприятия" Академический научно-издательский и книгораспространительский центр" Наука", 1996. – С. 276-276.
401. Давыдов В. А. Количественные характеристики устьичного аппарата растений яровой пшеницы сорта Саратовская 29 при остром дефиците воды // Сельскохозяйственная биология. – 2007. – Т. 42. – № 5. – С. 90-93.
402. Домасев М. В., Гнатюк С. П. Цвет, управление цветом, цветовые расчеты и измерения. СПб., Питер, 2009
403. Дорошков А. В., Афонников Д. А., Пшеничникова Т. А. Генетический анализ опушения листа у изогенных линий мягкой пшеницы сорта Новосибирская 67 // Генетика. – 2014. – Т. 50. – № 2. – С. 172-172.
404. Дорошков А. В., Пшеничникова Т. А., Афонников Д. А. Анализ особенностей морфологии и наследования опушения листа пшеницы *Triticum aestivum* L. с помощью компьютерных методов фенотипирования // Генетика. – 2011. – Т. 47. – № 6. – С. 836-841.
405. Дорошков А.В., Арсенина С.И., Пшеничникова Т.А., Афонников Д.А. Применение компьютерного анализа микроизображений листа для оценки характеристик опушения пшеницы *Triticum aestivum* L. // Информационный вестник ВОГиС. – 2009. – Т. 13 – №1 – С. 218-226.
406. Дорошков А.В., Генаев М.А., Афонников Д.А. Протокол анализа количественных характеристик опушения листа картофеля // Вавиловский журнал генетики и селекции. – 2016. – Т. 20. – № 6. – С. 863-868.
407. Дорошков А.В., Пшеничникова Т.А., Афонников Д.А. Анализ

- особенностей морфологии и наследования опушения листа пшеницы *Triticum aestivum* L. С помощью компьютерных методов фенотипирования. // Генетика. – 2011. – Т. 47. – № 6. – С. 836–841.
408. Ефимов В. М., Ковалева В. Ю. Многомерный анализ биологических данных // Учебное пособие. СПб. – 2008.
409. Журавлёва Н.А. Механизм устьичных движений, продукционный процесс и эволюция. Н.: Наука, 1992. 140 с.
410. Журбицкий З.И. (1968) Теория и практика вегетационного метода. Наука, Москва, 260 с.
411. Ибрагимова С. М. и др. Оценка солеустойчивости трансгенных растений табака, несущих ген P5CS1 *Arabidopsis thaliana* // Генетика. – 2015. – Т. 51. – № 12. – С. 1368-1368.
412. Ибрагимова С.М., Колодяжная Я.С., Герасимова С.В., Кочетов А.В. Частичная супрессия гена пролиндегидрогеназы увеличивает устойчивость растений к различным видам абиотических стрессов // Физиология растений, – 2012. – Т.59. – С. 99.
413. Ильина Л. Г. Селекция яровой мягкой пшеницы на Юго-Востоке. – Изд-во Саратовского ун-та, 1989.
414. Колодяжная Я.С., Титов С.Е., Кочетов А.В., Комарова М.Л., Романова А.В., Коваль В.С., Шумный В.К. Оценка солеустойчивости растений табака *Nicotiana tabacum*, несущих антисмысловой супрессор гена пролиндегидрогеназы. // Генетика. – 2006. – Т. 42. – С. 1.
415. Кочетов А.В., Титов С.Е., Колодяжная Я.С., Комарова М.Л., Коваль В.С. Макарова Н.Н. линский Ю.Ю., Трифонова Е.А. Шумный В.К. Повышение содержания пролина и осмотического давления клеточного сока у трансформантов табака, несущих антисмысловой супрессор гена

- пролиндегидрогеназы // Генетика. – 2004. – Т. 40. – С. 282.
416. Крупнов В.А., Цапайкин А.П. Опушение листьев пшеницы: генетические и экологические аспекты // Сельскохозяйственная биология. Серия «биология растений». – 1990. – № 1. – С. 51-57.
417. Кузнецов В.В., Шевякова Н.И. Пролин при стрессе: биологическая роль, метаболизм, регуляция // Физиология растений. – 1999. – Т. 46. – С. 321.
418. Кумаков В.А. Физиологическое обоснование моделей сортов пшеницы / В.А. Кумаков. – М.: Агропромиздат, 1985. – 270 с.
419. Лихенко И.Е. О взаимосвязи опущения органов растений яровой мягкой пшеницы с хозяйственно и биологически ценными признаками в условиях западной Сибири. Растениеводство и селекция, 2007.
420. Майстренко О.И. Идентификация и локализация генов, контролирующих опущения листа молодых растений мягкой пшеницы // Генетика – 1976. – Т. 12. № 1. – С. 5-15.
421. Мамонтова, В.Н. Селекция и семеноводство яровой пшеницы / В.Н. Мамонтова // Избранные труды.– М.: Колос, 1980. – 287 с.
422. Местецкий Л. М. Математические методы распознавания образов. – 2008
423. Мирославов Е.А. Структура и функция эпидермиса листа покрытосеменных растений. Л.: Наука, 1974. 120 с.
424. Науменко Н. В. и др. Контролируемое проращивание зерновых культур-безопасный способ технологии новых сырьевых ингредиентов //Вестник Южно-Уральского государственного университета. Серия: Пищевые и биотехнологии. – 2021. – Т. 9. – № 3. – С. 53-61.

425. Науменко Н. В., Потороко И. Ю., Калинина И. В. Интенсификация процесса проращивания зерна, используемого для производства хлеба, и его влияние на качество готовых изделий // Food industry. – 2019. – Т. 4. – № 1. – С. 47-54.
426. Нестеров М.А., Афонников Д.А., Сергеева Е.М., и др. Идентификация микросателлитных локусов по данным секвенирования ВАС-клонов и их физическое картирование на хромосому 5В мягкой пшеницы // Вавиловский журнал генетики и селекции. – 2015. – Т. 19. – № 6. – С. 707-714.
427. Пронозин А. Ю. и др. Автоматическое фенотипирование морфологии колоса тетра-и гексаплоидных видов пшеницы методами компьютерного зрения // Вавиловский журнал генетики и селекции. – 2021. – Т. 25. – № 1. – С. 71-81.
428. Рогачев С. Обобщенный Model-View-Controller // RSDN Magazine. – 2008. – № 4. – С. 37-66.
429. Смоликова Г.Н. Применение метода искусственного старения для оценки устойчивости семян к стрессовым воздействиям // Вестник СПбГУ. – 2014. – Сер. 3. – Вып. 2. – С. 82-93.
430. Фасулати С.Р., Иванова О.В., Рогозина Е.В. Комплексная устойчивость картофеля к колорадскому жуку, картофельной коровке и золотистой картофельной нематоды // Защита и карантин растений. – 2011. – № 10. – С. 14-17.
431. Фисенко В.Т., Фисенко Т.Ю., Компьютерная обработка и распознавание изображений: учеб. пособие. - СПб: СПбГУ ИТМО, 2008. – 192 с.
432. Фляксбергер К. А. Пшеницы – род *Triticum* L. рг. р. // Хлебные

- злаки. /Культурная флора СССР.-Т. 1.-М. ; Л.: Сельхозгиз. – 1935. –С. 19-434 с.
433. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. ИД Вильямс. 2004.
434. Хлесткина Е. К., Шумный В. К. Перспективы использования прорывных технологий в селекции: система CRISPR/Cas9 для редактирования генома растений //Генетика. – 2016. – Т. 52. – № 7. – С. 774-787.
435. Хлесткина Е.К. Молекулярные маркёры в генетических исследованиях и в селекции // Вавиловский журнал генетики и селекции. – 2013. – Т. 17. – № 4/2. – С. 53-60.
436. Швачко Н. А., Семилет Т. В., Тихонова Н. Г. Трихомы у высших растений: гомологические ряды в наследственной изменчивости и молекулярно-генетические механизмы //Генетика. – 2020. – Т. 56. – № 11. – С. 1320-1332.
437. Шеффе Х. Дисперсионный анализ. 2-е изд. – 1980.
438. Эсау К. Анатомия растений. М.: Мир, 1969. 564 с.

ПРИЛОЖЕНИЕ

Таблица П1. Склонение слова «трихом» по падежам согласно Всемирному словарю русского языка (<https://ru.worldwidedictionary.org/>)

Падеж	Единственное число	Множественное число
Именительный	трихóм	трихóмы
Родительный	трихóма	трихóмов
Дательный	трихóму	трихóмам
Винительный	трихóм	трихóмы
Творительный	трихóмом	трихóмами
Предложный	трихóме	трихóмах

См. также: Аванесов Р.И., Ожегов С.И. Русский орфографический словарь: около 180 000 слов / Российская академия наук. Институт русского языка им. В. В. Виноградова / О.Е. Иванова, В.В. Лопатин (отв. ред.), И.В. Нечаева, Л.К. Чельцова. — 2-е изд., испр. и доп. — Москва, 2004. — 960 с.

Электронный доступ: <https://wordsonline.ru/dicts/orthography.html>

<https://wordsonline.ru/Т/трихом>

Таблица П2. Список образцов трибы Triticeae использованный для широкомасштабного анализа признаков опушения и полученный из Институт генетики растений и исследований растениеводства Лейбница, Гатерслебен, Германия (**IPK**), Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова, Санкт-Петербург, Россия (**VIR**), Центр Джона Иннеса, Норвич, Великобритания (**ЖС**) и институт Цитологии и генетики СО РАН, Новосибирск, Россия (**ICG**).

№	Геномная конституция	Виды	Номер доступа образца	Генетическая коллекция
1	AA	<i>T. monococcum</i> L. var. <i>nigricultum</i> Flaksb.	TRI 1510	IPK
2	AA	<i>T. monococcum</i> L. var. <i>macedonicum</i> Papag.	TRI 17025	IPK
3	AA	<i>T. monococcum</i> L. var. <i>macedonicum</i> Papag.	TRI 17212	IPK
4	AA	<i>T.monococcum</i> L. var. <i>sofianum</i> Stranski	TRI 28871	IPK
5	AA	<i>T.boeoticum</i> Boiss. subsp. <i>thaoudar</i> (Reut. Ex Hausskn.) Grossh. var. <i>biarpurunivir</i> Gandilyan	TRI 17071	IPK
6	AA	<i>T.boeoticum</i> Boiss. subsp. <i>thaoudar</i> (Reut. ExHausskn.) Grossh. var. <i>biarpurunivir</i> Gandilyan	TRI 17072	IPK
7	AA	<i>T.boeoticum</i> Boiss. subsp. <i>thaoudar</i> (Reut. Ex Hausskn.) Grossh. var. <i>viridalbum</i> A.Filat.	TRI 17098	IPK
8	AA	<i>T.boeoticum</i> Boiss. var. <i>viridialbinigrescens</i> K.Hammer & A.Filat.	TRI 17103	IPK
9	AA	<i>T.boeoticum</i> Boiss. var. <i>viridialbinigrescens</i> K.Hammer & A.Filat.	TRI 17117	IPK
10	AA	<i>T.boeoticum</i> Boiss. subsp. <i>thaoudar</i> (Reut. ExHausskn.) Grossh. var. <i>viridinigrum</i> Flaksb.	TRI 17164	IPK
11	AA	<i>T. boeoticum</i> Boiss. subsp. <i>thaoudar</i> (Reut. ex Hausskn.) Grossh. var. <i>viridifuscum</i> A.Filat.	TRI 17074	IPK
12	AA	<i>T.urartu</i> Tumanian ex Gandilyan var. <i>spontaneoalbum</i> Tumanian	TRI 17118	IPK
13	AA	<i>T.urartu</i> Tumanian ex Gandilyan var. <i>pubescentiurartu</i> K.Hammer & A.Filat.	TRI 17121	IPK
14	AA	<i>T.urartu</i> Tumanian ex Gandilyan var. <i>albonigricans</i> Tumanian	TRI 17136	IPK
15	AA	<i>T.urartu</i> Tumanian ex Gandilyan var. <i>spontaneoalbum</i> Tumanian	TRI 17138	IPK
16	AA	<i>T.urartu</i> Tumanian ex Gandilyan var. <i>nigrum</i> Tumanian	TRI 17141	IPK

17	SS(BB,GG)	<i>Ae. speltoides</i> Tausch.	Ae 102	IPK
18	SS(BB,GG)	<i>Ae. speltoides</i> Tausch.	Ae 1592	IPK
19	SS(BB,GG)	<i>Ae. speltoides</i> Tausch.	Ae 1498	IPK
20	SS(BB,GG)	<i>Ae. speltoides</i> Tausch.	k-389	VIR
21	DD	<i>Ae. Tauschii</i> Coss.	k-865	VIR
22	DD	<i>Ae. tauschii</i> Coss.	k-992	VIR
23	DD	<i>Ae. tauschii</i> Coss.	k-1955	VIR
24	DD	<i>Ae. tauschii</i> Coss.	k-1956	VIR
25	AABB	<i>T. durum</i> Desf.	TRI 28800	IPK
26	AABB	<i>T. durum</i> Desf.	TRI 1703	IPK
27	AABB	<i>T. durum</i> Desf.	TRI 28769	IPK
28	AABB	<i>T. dicoccoides</i>	TRI 18498	IPK
29	AABB	<i>T. dicoccon</i> Schrank ex Schübl.	TRI 2228	IPK
30	AABB	<i>T. dicoccon</i> Schrank ex Schübl.	TRI 15821	IPK
31	AAGG	<i>T. palmovae</i> G. Ivanov	-	ICG
32	AAGG	<i>T. timopheevii</i> Zhuk.	TRI 3433	IPK
33	AAGG	<i>T. timopheevii</i> Zhuk.	TRI 7272	IPK
34	AADD	<i>T. timopheevii</i> Zhuk. ssp. <i>viticulosum</i>	-	ICG
35	AABBDD	<i>T. spelta</i> L.	'Grey'	JIC
36	AABBDD	<i>T. spelta</i> L.	1220029	JIC
37	AABBDD	<i>T. petropavlovskii</i> Udacz. et Migusch.	k-43351	VIR
38	AABBDD	<i>T. petropavlovskii</i> Udacz. et Migusch.	k-3376	VIR
39	AABBDD	<i>T. aestivum</i> L. (cv. Rodina)	-	ICG
40	AABBDD	<i>T. aestivum</i> L. (cv. Chinese Spring)	-	ICG
41	AABBDD	<i>T. aestivum</i> L. (cv. Yanetskis Probat)	-	ICG
42	AABBDD	<i>T. aestivum</i> L. (cv. Diamant 2)	-	ICG
43	AABBDD	<i>T. aestivum</i> L. (cv. Synthetic 6x)	-	ICG
44	AABBDD	<i>T. aestivum</i> L. (cv. Golubka)	-	ICG
45	AABBDD	<i>T. aestivum</i> L. (cv. Novosibirskaya 67)	-	ICG
46	AABBDD	<i>T. aestivum</i> L. (cv. Saratovskaya 29)	-	ICG
47	AABBDD	<i>T. aestivum</i> L. (cv. Hong-Mang-mai)	-	ICG

Таблица ПЗ. Количественные характеристики опушения листьев у родов *Triticum* и *Aegeilops*. n_f – среднее число трихом на изображении; l_m – средняя длина трихомов (мкм), $H_{L/N}$ – индекс опушения.

№ образца	Вид	Кариотип	n_f	l_m , мкм	$H_{L/N}$
1	<i>T. monococcum</i>	AA	8,06±1,14	27,07±3,04	3,33
2	<i>T. monococcum</i>	AA	8,89±1,35	36,04±9,38	4,22
3	<i>T. monococcum</i>	AA	4,22±0,93	29,47±4,82	7,2
4	<i>T. monococcum</i>	AA	17,39±4,3	31,78±5,83	1,78
5	<i>T. boeoticum</i>	AA	15,56±1,93	185,91±50,95	11,77
6	<i>T. boeoticum</i>	AA	32,27±1,81	123,82±21,54	3,79
7	<i>T. boeoticum</i>	AA	13,83±4,17	60,40±27,14	5,03
8	<i>T. boeoticum</i>	AA	28,28±2,56	31,52±2,25	1,12
9	<i>T. boeoticum</i>	AA	42,44±3,24	57,98±6,14	1,36
10	<i>T. boeoticum</i>	AA	45,89±2,83	51,74±3,97	1,13
11	<i>T. boeoticum</i>	AA	32,44±6,27	74,64±8,57	2,31
12	<i>T. urartu</i>	AA	44,5±2,51	64,40±1,43	1,45
13	<i>T. urartu</i>	AA	49,72±2,25	64,74±1,92	1,3
14	<i>T. urartu</i>	AA	53,61±4,18	71,15±5,92	1,34
15	<i>T. urartu</i>	AA	54,33±2,23	61,64±2,09	1,14
16	<i>T. urartu</i>	AA	47,22±5,31	61,89±3,65	1,32
17	<i>Ae. speltoides</i>	SS	6,62±1,15	108,09±10,87	16,24
18	<i>Ae. speltoides</i>	SS	16,67±2,98	154,99±18,39	9,7
19	<i>Ae. speltoides</i>	SS	3,74±0,63	132,87±23,82	34,9
20	<i>Ae. speltoides</i>	SS	6,27±1,26	97,63±17,08	17,62
21	<i>Ae. tauschii</i>	DD	6,23±0,75	231,98±19,21	37,02
22	<i>Ae. tauschii</i>	DD	4,62±0,52	87,23±13,49	18,73
23	<i>Ae. tauschii</i>	DD	6,33±0,86	394,59±21,85	62,14
24	<i>Ae. tauschii</i>	DD	7,20±0,79	231,63±18,36	31,48
25	<i>T. durum</i>	AABB	0,16±0,08	11,20±7,60	--
26	<i>T. durum</i>	AABB	0,08±0,07	7,35±6,86	--
27	<i>T. durum</i>	AABB	0,57±0,23	27,65±11,59	--
28	<i>T. dicoccoides</i>	AABB	12,13±1,30	32,67±2,18	2,75
29	<i>T. dicoccon</i>	AABB	40,1±2,39	90,58±2,27	2,25
30	<i>T. dicoccon</i>	AABB	14,50±0,90	58,77±3,27	4,09
31	<i>T. palmovae</i>	AADD	12,98±1,16	438,63±26,09	33,53
32	<i>T. timopheevii</i>	AAGG	13,43±1,27	710,33±71,57	52,26
33	<i>T. timopheevii</i>	AAGG	19,37±4,25	398,40±52,89	57,66
34	<i>T. timopheevii</i>	AAGG	12,87±1,62	589,15±59,27	21,47
35	<i>T. spelta</i>	AABBDD	0,96±0,31	62,3±29,31	--
36	<i>T. spelta</i>	AABBDD	0,58±0,21	53,72±0,21	--

37	<i>T. petropavlovskii</i>	AABBDD	21,38±3,38	110,78±10,23	5,13
38	<i>T. petropavlovskii</i>	AABBDD	18,48±4,87	97,72±5,28	5,26
39	<i>T. aestivum</i> cv. Rodina	AABBDD	0,54±0,21	15,98±4,92	--
40	<i>T. aestivum</i> cv. Chinese Spring	AABBDD	1,84±1,68	29,20±6,46	--
41	<i>T. aestivum</i> Yanetskis Probat	AABBDD	3,04±0,48	31,41±5,02	10,57
42	<i>T. aestivum</i> cv. Diamant 2	AABBDD	8,43±1,01	33,93±2,42	4,21
43	<i>T. aestivum</i> cv. Synthetic 6x	AABBDD	8,86±1,07	48,72±4,53	5,75
44	<i>T. aestivum</i> cv. Golubka	AABBDD	29,07±1,23	121,97±6,18	4,23
45	<i>T. aestivum</i> cv. Novosibirskaya 67	AABBDD	23,66±1,29	95,36±3,03	4,01
46	<i>T. aestivum</i> Saratovskaya 29	AABBDD	40,46±2,20	147,11±5,90	3,66
47	<i>T. aestivum</i> cv. Hong-mang-mai	AABBDD	28,92±2,16	148,60±9,27	5,14

Таблица П4. Список генов-кандидатов, связанных с QTL признаков размера и формы семян мягкой пшеницы. Колонки содержат название QTL (QTL), положение на хромосоме в сМ (Хром/Поз), идентификатор гена (ID Гена), идентификатор ортогруппы KEGG (ID ОГ), название ортогруппы KEGG (ОГ), номер ЕС, если известен.

QTL	Хром/Поз	ID гена	ID ОГ	ОГ	ЕС
Q.sA-3A	3A/155	TraesCS3A03G0787100	K09286	EREBP; EREBP-like factor	-
		TraesCS3A03G0782400	K09338	HD-ZIP; homeobox-leucine zipper protein	-
		TraesCS3A03G0763900	K00279	CKX; cytokinin dehydrogenase	EC:1.5.99.12
Q.sSo-4A	4A/305	TraesCS4A03G1100100	K19045	BB; E3 ubiquitin-protein ligase BIG BROTHER and related proteins	EC:2.3.2.27
Q.sA-2B.2	2B/208	TraesCS2B03G0313000	K09286	EREBP; EREBP-like factor	-
Q.sCi-2B ^c ; sSo-2B ^c	2B/129	TraesCS2B03G1115600	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-
		TraesCS2B03G1114800	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-
		TraesCS2B03G1105600	K09422	MYBP; transcription factor MYB, plant	-
		TraesCS2B03G1109700	K09286	EREBP; EREBP-like factor	-
		TraesCS2B03G1116700	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-
		TraesCS2B03G1104200	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-
		TraesCS2B03G1109900	K09286	EREBP; EREBP-like factor	-
		TraesCS2B03G1106500	K09338	HD-ZIP; homeobox-leucine zipper protein	-
Q.sW-2D	2D/74	TraesCS2D03G0133000	K09422	MYBP; transcription factor MYB, plant	-
		TraesCS2D03G0143000	K09602	OTUB1; ubiquitin thioesterase protein OTUB1	EC:3.4.19.12
		TraesCS2D03G0143400	K09602	OTUB1; ubiquitin	EC:3.4.19.12

				thioesterase protein OTUB1	
Q.sCi- 2D ^g ; Q.sRo- 2D ^g	2D/58	TraesCS2D03G0107400	K09602	ubiquitin thioesterase protein OTUB1	EC:3.4.19.12
		TraesCS2D03G0107900	K09602	ubiquitin thioesterase protein OTUB1	EC:3.4.19.12
sSo- 7D.2	7D/141	TraesCS7D03G1008800	K09286	EREBP; EREBP-like factor	-
		TraesCS7D03G0987700	K09338	HD-ZIP; homeobox- leucine zipper protein	-
		TraesCS7D03G0983800	K09286	EREBP; EREBP-like factor	-
		TraesCS7D03G0972400	K08869	ADCK, ABC1; aarF domain-containing kinase	-
		TraesCS3A03G0763900	K00279	CKX; cytokinin dehydrogenase	EC:1.5.99.12
sL- 7D ^{bb} ; sRo- 7D ^{bb}	7D/287	TraesCS7D03G1260500	K20547	CHIB; basic endochitinase B	EC:3.2.1.14
		TraesCS7D03G1260300	K20547	CHIB; basic endochitinase B	EC:3.2.1.14
		TraesCS7D03G1260400	K20547	CHIB; basic endochitinase B	EC:3.2.1.14
		TraesCS7D03G1286900	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-
		TraesCS7D03G1260600	K20547	CHIB; basic endochitinase B	EC:3.2.1.14
		TraesCS7D03G1287400	K11593	ELF2C, AGO; eukaryotic translation initiation factor 2C	-

Таблица П5. Список генов-кандидатов, связанных с QTL признаков цвета оболочки семян мягкой пшеницы. Колонки содержат название QTL (QTL), положение на хромосоме в сМ (Хром/Поз), идентификатор гена (ID Гена), идентификатор ортогруппы KEGG (ID ОГ), название ортогруппы KEGG (ОГ), номер ЕС, если известен, идентификатор пути KEGG (ID пути KEGG), название пути KEGG (Название пути KEGG).

QTL	Хром/Поз	ID гена	ID ОГ	ОГ	ЕС	ID пути KEGG	Название пути KEGG
Q.YCrCb_dCCr_1-2A.3	2A/196	TraesCS6A03G0725700	K09840	NCED; 9-cis-epoxycarotenoid dioxygenase	EC:1.13.11.51	map00906	Carotenoid biosynthesis
		TraesCS2A03G0158600	K22772	FG2; flavonol-3-O-glucoside L-rhamnosyltransferase	EC:2.4.1.159	map00944	Flavone and flavonol biosynthesis
		TraesCS6A03G0953500	K13065	HCT; shikimate O-hydroxycinnamoyltransferase	EC:2.3.1.133	map00941	Flavonoid biosynthesis
		TraesCS2A03G0099800	K13066	COMT; caffeic acid 3-O-methyltransferase / acetylserotonin O-methyltransferase	EC:2.1.1.68; 2.1.1.4	map00380	Tryptophan metabolism
Q.Lab_dCb_3-3A ^{i*}	3A/195	TraesCS3A03G0925800	K01904	4CL; 4-coumarate--CoA ligase	EC:6.2.1.12	map00940	Phenylpropanoid biosynthesis
		TraesCS3A03G0925900	K01904	4CL; 4-coumarate--CoA ligase	EC:6.2.1.12	map00940	Phenylpropanoid biosynthesis
Q.RGB_dCB_1-3B.1	3B/269.179	TraesCS3B03G1115600	K12355	REF1; coniferyl-aldehyde dehydrogenase	EC:1.2.1.68	map00940	Phenylpropanoid biosynthesis
HSV_dCH_1-3B ^{m**}	306.179	TraesCS3B03G1278200	K01904	4CL; 4-coumarate--CoA ligase	EC:6.2.1.12	map00940	Phenylpropanoid biosynthesis

HSV_dCH_3-6A ^{t***} ;	6A/246	TraesCS6A03G0725000	K09843	CYP707A; (+)-abscisic acid 8'-hydroxylase	EC:1.14.14.137	map00906	Carotenoid biosynthesis
		TraesCS6A03G0741000	K00021	HMGCR; hydroxymethylglutaryl-CoA reductase (NADPH)	EC:1.1.1.34	map00900	Terpenoid backbone biosynthesis
Q.YCrCb_dCCr_1-6A	6A/340	TraesCS6A03G0953500	K13065	HCT; shikimate O-hydroxycinnamoyltransferase	EC:2.3.1.133	map00940	Phenylpropanoid biosynthesis
		TraesCS2A03G0163500	K00430	peroxidase	EC:1.11.1.7	map00940	Phenylpropanoid biosynthesis
		TraesCS2A03G0164200	K00430	peroxidase	EC:1.11.1.7	map00940	Phenylpropanoid biosynthesis
Lab_dCL_2-6B ^x ; ****	6B/220	TraesCS6B03G0367700	K00430	peroxidase	EC:1.11.1.7	map00940	Phenylpropanoid biosynthesis

*Co-located QTL: Q.YCrCb_dCCb_3-3Aⁱ

** Co-located QTLs: HSV_dCH_3-3B^m; HSV_dCS_1-3B^m; HSV_dCV_1-3B^m; HSV_dCV_3-3B^m; Lab_dCb_1-3B^m; Lab_dCb_2-3B^m; Lab_mb-3B^m; RGB_dCR_2-3B^m; YCrCb_dCCb_1-3B^m; YCrCb_dCCb_2-3B^m; YCrCb_dCCr_1-3B^m; YCrCb_mCb-3B^m

*** Co-located QTL: HSV_dCV_3-6A^t; Lab_ma-6A^t

**** Co-located QTLs: Lab_dCL_3-6B^x; Lab_mb-6B^x; RGB_dCG_2-6B^x; RGB_dCR_2-6B^x; YCrCb_dCCb_2-6B^x; YCrCb_dCY_2-6B^x; YCrCb_dCY_3-6B^x

Таблица Пб. Коэффициенты корреляции Пирсона между значениями признаков зерен пшеницы и годом урожая в различных кодировках (Year01, Year, YearRank). Указаны минимальные и максимальные значения коэффициентов корреляции с кодировкой года YearRank в рандомизационном тесте (Рандом.Тест Мин., Рандом.Тест Макс.), значимость влияния фактора года на значения признаков согласно дисперсионному анализу (Дисп. Анализ Знач., 0- не значимо, 1 - значимо). Жирным шрифтом выделены значения, значимо отклоняющиеся от 0.

Признак	Year01	Year	YearRank	Рандом. Тест Мин.	Рандом.Тест Макс.	Дисп. Анализ Знач.
HSV_mH	0.491	0.514	0.511	-0.051	0.052	1
YCrCb_dCCb_2	0.346	0.289	0.255	-0.055	0.051	1
YCrCb_dCCb_1	0.304	0.295	0.241	-0.069	0.054	1
HSV_dCH_1	0.213	0.216	0.211	-0.061	0.054	1
YCrCb_mCb	0.260	0.223	0.189	-0.055	0.047	1
HSV_dCH_3	0.144	0.175	0.184	-0.048	0.051	1
HSV_dCV_3	0.055	0.113	0.090	-0.069	0.047	0
sRu	0.033	0.048	0.077	-0.054	0.062	0
HSV_dCS_2	0.103	0.103	0.076	-0.057	0.054	1
Lab_dCb_3	0.049	0.090	0.068	-0.050	0.056	1
sW	0.103	0.049	0.062	-0.056	0.063	1
sA	0.070	0.065	0.051	-0.053	0.064	1
sRo	0.095	-0.005	0.040	-0.055	0.060	1
HSV_dCH_2	-0.008	0.016	0.028	-0.062	0.067	0
sL	0.005	0.065	0.025	-0.056	0.060	1
YCrCb_dCCr_3	0.068	0.058	0.018	-0.063	0.054	0
Lab_dCa_3	0.074	0.054	0.005	-0.065	0.060	0
sSo	0.072	0.014	-0.008	-0.059	0.050	1
HSV_dCS_1	-0.029	0.003	-0.019	-0.057	0.058	0
sCi	0.038	-0.030	-0.053	-0.053	0.060	1
YCrCb_dCCb_3	-0.033	-0.072	-0.055	-0.049	0.056	1
HSV_dCS_3	-0.049	-0.044	-0.066	-0.052	0.053	0
RGB_mG	-0.137	-0.141	-0.083	-0.064	0.058	1

RGB_mB	-0.140	-0.172	-0.098	-0.059	0.056	1
RGB_dCB_2	-0.185	-0.208	-0.123	-0.054	0.050	1
RGB_dCG_2	-0.218	-0.205	-0.137	-0.055	0.050	1
RGB_dCG_1	-0.182	-0.206	-0.139	-0.057	0.051	1
YCrCb_dCY_3	-0.234	-0.178	-0.142	-0.049	0.053	0
RGB_dCG_3	-0.239	-0.178	-0.142	-0.060	0.053	0
YCrCb_mY	-0.203	-0.205	-0.145	-0.049	0.070	1
Lab_mL	-0.205	-0.206	-0.146	-0.056	0.071	1
RGB_dCB_1	-0.183	-0.233	-0.153	-0.055	0.062	1
Lab_dCL_3	-0.249	-0.192	-0.156	-0.053	0.064	0
RGB_dCB_3	-0.251	-0.206	-0.169	-0.054	0.054	0
RGB_dCR_3	-0.267	-0.195	-0.174	-0.054	0.050	0
YCrCb_dCY_2	-0.275	-0.262	-0.196	-0.063	0.051	1
Lab_dCL_2	-0.283	-0.270	-0.207	-0.050	0.063	1
YCrCb_dCY_1	-0.256	-0.280	-0.208	-0.047	0.055	1
Lab_dCL_1	-0.261	-0.284	-0.215	-0.057	0.051	1
Lab_mb	-0.292	-0.247	-0.218	-0.062	0.052	1
HSV_dCV_2	-0.284	-0.257	-0.221	-0.061	0.062	1
HSV_mS	-0.236	-0.152	-0.230	-0.054	0.065	1
HSV_mV	-0.309	-0.301	-0.242	-0.060	0.056	1
RGB_mR	-0.308	-0.301	-0.243	-0.056	0.058	1
Lab_dCb_1	-0.334	-0.321	-0.272	-0.056	0.057	1
HSV_dCV_1	-0.310	-0.335	-0.285	-0.054	0.053	1
Lab_dCb_2	-0.377	-0.314	-0.287	-0.053	0.054	1
RGB_dCR_2	-0.397	-0.366	-0.308	-0.060	0.057	1
RGB_dCR_1	-0.372	-0.387	-0.312	-0.060	0.066	1
YCrCb_dCCr_1	-0.566	-0.563	-0.493	-0.054	0.062	1
YCrCb_mCr	-0.583	-0.548	-0.509	-0.052	0.058	1
YCrCb_dCCr_2	-0.592	-0.537	-0.513	-0.063	0.052	1
Lab_dCa_2	-0.652	-0.601	-0.599	-0.063	0.051	1
Lab_dCa_1	-0.677	-0.674	-0.610	-0.062	0.050	1
Lab_ma	-0.691	-0.658	-0.639	-0.067	0.058	1

Таблица П7. Коэффициенты корреляции Пирсона между значениями признаков зерен пшеницы и всхожестью (Всхожесть). Указаны минимальные и максимальные значения коэффициентов корреляции перестановочном тесте (Перест.Мин., Перест.Макс.), бутстреп-тесте (Бутстр.Мин., Бутстр.Макс.). Жирным шрифтом выделены значения, значимо отклоняющиеся от 0.

Признак	Всхожесть	Перест. Мин.	Перест. Макс.	Бутстр. Мин.	Бутстр. Макс.
HSV_mH	0.164	-0.088	0.080	-0.082	0.089
YCrCb_dCCb_2	0.084	-0.102	0.087	-0.112	0.091
YCrCb_dCCb_1	0.081	-0.096	0.099	-0.101	0.079
HSV_dCH_3	0.073	-0.101	0.098	-0.080	0.087
sA	0.067	-0.086	0.101	-0.093	0.089
YCrCb_mCb	0.060	-0.097	0.088	-0.093	0.085
sW	0.057	-0.102	0.095	-0.103	0.101
sL	0.054	-0.095	0.092	-0.087	0.079
HSV_dCH_1	0.036	-0.081	0.092	-0.082	0.102
sSo	0.030	-0.086	0.105	-0.093	0.094
YCrCb_dCCr_3	0.027	-0.092	0.095	-0.119	0.099
HSV_dCH_2	0.026	-0.093	0.112	-0.096	0.085
Lab_dCb_3	0.023	-0.088	0.109	-0.092	0.095
sRo	0.019	-0.086	0.094	-0.106	0.097
sRu	0.018	-0.082	0.095	-0.098	0.100
HSV_dCS_1	0.015	-0.093	0.083	-0.092	0.101
HSV_dCS_2	0.010	-0.093	0.087	-0.087	0.089
sCi	0.002	-0.082	0.077	-0.072	0.099
Lab_dCa_3	-0.001	-0.097	0.101	-0.088	0.092
HSV_dCV_3	-0.007	-0.104	0.083	-0.087	0.092
HSV_dCS_3	-0.026	-0.095	0.101	-0.084	0.083
RGB_mG	-0.027	-0.084	0.081	-0.083	0.099
RGB_dCG_2	-0.041	-0.094	0.090	-0.098	0.099
RGB_mB	-0.041	-0.101	0.091	-0.086	0.096
RGB_dCB_2	-0.042	-0.089	0.110	-0.080	0.078
YCrCb_dCCb_3	-0.042	-0.109	0.105	-0.083	0.094

RGB_dCG_1	-0.045	-0.087	0.091	-0.083	0.088
RGB_dCG_3	-0.046	-0.089	0.097	-0.094	0.078
YCrCb_mY	-0.052	-0.090	0.094	-0.078	0.081
Lab_dCL_3	-0.053	-0.090	0.095	-0.095	0.092
Lab_mL	-0.053	-0.073	0.077	-0.087	0.090
HSV_dCV_2	-0.054	-0.095	0.097	-0.088	0.105
RGB_dCB_1	-0.055	-0.101	0.094	-0.087	0.093
RGB_dCB_3	-0.056	-0.086	0.094	-0.090	0.099
RGB_dCR_3	-0.060	-0.103	0.082	-0.086	0.115
HSV_mS	-0.061	-0.077	0.085	-0.108	0.085
YCrCb_dCY_3	-0.061	-0.103	0.102	-0.094	0.086
Lab_dCL_2	-0.063	-0.090	0.109	-0.078	0.102
Lab_dCL_1	-0.063	-0.087	0.100	-0.096	0.080
YCrCb_dCY_2	-0.067	-0.098	0.095	-0.090	0.094
YCrCb_dCY_1	-0.071	-0.079	0.103	-0.078	0.094
Lab_mb	-0.073	-0.105	0.085	-0.097	0.099
Lab_dCb_1	-0.079	-0.091	0.100	-0.093	0.077
RGB_mR	-0.090	-0.102	0.082	-0.101	0.105
HSV_mV	-0.090	-0.085	0.084	-0.081	0.101
Lab_dCb_2	-0.092	-0.094	0.093	-0.090	0.084
HSV_dCV_1	-0.106	-0.081	0.087	-0.087	0.089
RGB_dCR_2	-0.107	-0.094	0.083	-0.082	0.096
RGB_dCR_1	-0.112	-0.099	0.093	-0.089	0.082
YCrCb_dCCr_1	-0.177	-0.077	0.100	-0.088	0.104
YCrCb_dCCr_2	-0.190	-0.082	0.096	-0.088	0.095
YCrCb_mCr	-0.197	-0.084	0.094	-0.077	0.104
Lab_dCa_1	-0.201	-0.100	0.124	-0.083	0.090
Lab_dCa_2	-0.224	-0.101	0.096	-0.088	0.091
Lab_ma	-0.235	-0.100	0.083	-0.086	0.114

Таблица П8. Параметры, характеризующие форму и размер колоса.

Параметр	Единицы измерения	Описание
L	мм	Длина ломаной линии вдоль стебля (линии стебля)
P	мм	Периметр контура колоса без остей
E _a	мм ²	Площадь, описываемая контуром колоса без остей
A _a	мм ²	Площадь области остей
E _a /L ²	-	Отношение площади колоса к квадрату его длины
C	-	Индекс округлости (circularity, C) равен отношению периметра окружности с площадью, равной площади контура к периметру контура. Индекс отражает насколько форма контура близка к форме окружности. Значение варьируется от 0 до 1.
R	-	Индекс закругленности (roundness, R) равен отношению площади контура к площади окружности с диаметром, равным главной оси вращения контура (Major Axis).
R _g	-	Индекс шероховатости (rugosity, R _g) определяется как отношение периметра контура к выпуклому периметру.
S	-	Индекс компактности/целостности (solidity, S) - это отношение площади контура к площади его выпуклой оболочки.
α _{u1}	градусы	величина наклона ребра AB относительно основания верхнего четырехугольника
α _{u2}	градусы	величина наклона ребра BC относительно основания верхнего четырехугольника
α _{u3}	градусы	величина наклона ребра BD относительно основания верхнего четырехугольника
t _{1u}	-	тангенс угла α _{u1}
t _{2u}	-	тангенс угла α _{u2}
t _{3u}	-	тангенс угла α _{u3}
S _{u1}	мм ²	площадь треугольника АВВ'
S _{u2}	мм ²	площадь трапеции ВВ'С'С
S _{u3}	мм ²	площадь треугольника АВВ'
S _u	мм ²	площадь верхнего четырехугольника
y _{um}	мм	среднее значение высоты верхнего четырехугольника
A _{Ix2}	мм	индекс асимметрии по длинам сегментов
A _{Iy2}	мм	индекс асимметрии по высотам сегментов
A _{Ixy2}	мм	общий индекс асимметрии
Сечения контура колоса (20 + 20 сечений)	мм	Значения длин сечений контура от линии оси стебля колоса

Таблица П9. Список таблиц БД WheatPGE.

Наименование таблицы	Определение
plant	Описание растения
genotype	Описание генотипа
sort	Описание сорта растения
line	Описание генетической линии
marker_list	Список генетических маркеров
marker	Описание генетического маркера
marker_set	Набор маркеров, идентифицированных для растения
marker_state	Признак маркера
marker_state_list	Список признаков маркера
marker_state_set	Набор состояний маркера
phenotype	Описание фенотипа растения
hairiness_list	Список параметров опушения листа
hairiness	Характеристики опушения
hairiness_set	Набор описаний опушения листа
basic_morphology	Основные признаки морфологии растения
leaf_list	Список листьев
leaf_set	Набор описаний листьев
leaf	Описание листа
ear_list	Список колосьев
ear_set	Набор описаний колосьев
ear	Колос
ear_morphology_list	Список колосьев
ear_morphology_set	Набор описаний морфологии колоса
ear_morphology	Морфология колоса
development	Стадии развития растения
environment	Описание параметров окружающей среды
time_line	Описание шкалы времени для регистрации событий
event_type	Описание события
place	Место произрастания
user	Описание атрибутов пользователя
role	Описание входа в базу
user_role	Описание параметров входа пользователя в базу