**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**

**03.01.09 – математическая биология, биоинформатика**

**Диссертация на соискание ученой степени**

**доктора биологических наук**

**Научный консультант**

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

**Новосибирск**

**2022**

**ОГЛАВЛЕНИЕ**

|  |  |
| --- | --- |
| Оглавление  | 1 |
| Введение | 7 |
| Глава 1. Обзор литературы  | 15 |
| 1.1. Задачи феномики растений | 15 |
| 1.2. Получение и анализ изображений | 22 |
| 1.2.1. Получение изображений растений | 23 |
| 1.2.2. Анализ цифровых изображений | 23 |
| 1.2.3. Устранение шумов | 26 |
| 1.2.4. Сегментация изображений | 27 |
| 1.2.5. Представление объектов и морфометрия | 33 |
| 1.2.6. Анализ цвета | 34 |
| 1.2.7. Распознавание образов и методы машинного обучения | 38 |
| 1.3. Фенотипирование органов растений | 43 |
| 1.3.1. Фенотипирование корней | 43 |
| 1.3.2. Фенотипирование листьев | 45 |
| 1.3.3. Фенотипирование соцветий, семян и плодов | 47 |
| 1.3.4. Анализ цвета оболочки злаков | 48 |
| 1.3.5. Фенотипирование побегов | 50 |
| 1.3.6. Интеграция данных в феномике растений | 51 |
| 1.4. Мягкая пшеница – одна из основных культур для селекционно-генетических исследований | 53 |
| 1.5. Заключение по обзору литературы и постановка задачи исследования | 57 |
| Глава 2. Высокопроизводительное фенотипирование опушения листа у растений  | 60 |
| 2.1. Опушение листа у растений  | 60 |
| 2.1.1. Структура и функция опушения листьев растений | 60 |
| 2.1.2. Опушение листа у пшеницы: морфология, функции и генетический контроль | 63 |
| 2.1.3. Морфология опушения листа у картофеля | 66 |
| 2.1.4. Методы фенотипирования опушения листьев у растений | 68 |
| 2.2. Протокол получения микроизображений сгибов листа пшеницы  | 69 |
| 2.2.1. Растительный материал и условия выращивания | 69 |
| 2.2.2. Протокол пробоподготовки и получения изображения сгибов листа | 70 |
| 2.3. Оценка количественных характеристик опушения листа на основе анализа двухмерных микроизображений сгиба  | 78 |
| 2.3.1. Алгоритм автоматической оценки количественных характеристик опушения листа LHDetect | 78 |
| 2.3.2. Алгоритм автоматической оценки количественных характеристик опушения листа LHDetect2 | 81 |
| 2.3.3. Компьютерная программа LHDetect2 для оценки количественных характеристик опушения листа пшеницы | 94 |
| 2.3.4. Статистические методы анализа количественных характеристик опушения листа | 100 |
| 2.4. Разнообразие опушения листьев у мягкой пшеницы и ее сородичей  | 102 |
| 2.4.1. Растительный материал и условия выращивания | 102 |
| 2.4.2. Фенотипирование растений и обработка данных | 104 |
| 2.4.3. Характеристики опушения пшеницы и ее сородичей | 106 |
| 2.4.4. Кластеризация генотипов на основе характеристик опушения | 110 |
| 2.4.5. Связь между характеристиками генома и особенностями опушения листа у родов Triticum и Aegilops | 116 |
| 2.4.6. Разнообразие опушения листа у мягких пшениц и их сородичей и его связь с геномным составом | 117 |
| 2.5. Изучение фенотипического проявления генов опушения листа пшеницы *Triticum aestivum* L. | 121 |
| 2.5.1. Растительный материал и условия выращивания | 122 |
| 2.5.2. Фенотипирование растений и обработка данных | 123 |
| 2.5.3. Анализ опушения листа | 124 |
| 2.5.4. Влияние генов *Hl*1, *Hl*2aesp и *Hl*3 на опушение листа пшеницы | 128 |
| 2.6. Анализ растений замещенных линий пшеницы (*Triticum aestivum* L.) в условиях нормального и ограниченного водообеспечения  | 134 |
| 2.6.1. Растительный материал и условия выращивания | 134 |
| 2.6.2. Фенотипирование физиологических характеристик растений и опушения листьев | 136 |
| 2.6.3. Статистическая обработка физиологических характеристик растений и опушения листьев | 138 |
| 2.6.4. Связь количественных характеристик опушения листа у изогенных и замещенных линий пшеницы с физиологическими показателями растений в условиях нормального и ограниченного водообеспечения | 139 |
| 2.7. Высокопроизводительное фенотипирование опушения листа картофеля | 155 |
| 2.7.1. Условия, материалы и методы для анализа опушения листа у картофеля | 155 |
| 2.7.2. Оценка количественных характеристик опушения листьев картофеля с использованием анализа цифровых микроизображений | 157 |
| 2.8. Высокопроизводительное фенотипирование опушения листа табака | 162 |
| 2.8.1. Генетические линии табака с измененным паттерном экспрессии гена пролиндегидрогеназы  | 164 |
| 2.8.2. Растительный материал и условия выращивания генетически модифицированных линий табака | 165 |
| 2.8.3. Модификация параметров алгоритма для анализа опушения листа табака | 167 |
| 2.8.4. Сравнение характеристик опушения у ГМ-линий и контрольного сорта табака | 172 |
| 2.9. Заключение по главе 2 | 178 |
| Глава 3. Высокопроизводительное фенотипирование зерен пшеницы  | 180 |
| 3.1. Зерна злаков и их характеристики важные для селекционно-генетических исследований | 180 |
| 3.2. Приложение SeedCounter для морфометрии зерен пшеницы | 186 |
| 3.2.1. Протоколы получения изображений зерен пшеницы для фенотипирования с помощью мобильных устройств | 186 |
| 3.2.2. Алгоритм обработки изображений для морфометрии зерен | 187 |
| 3.2.3. Мобильное приложение SeedCounter | 200 |
| 3.3. Поиск локусов количественных признаков для характеристик размера, формы и цвета зерен мягкой пшеницы  | 201 |
| 3.3.1. Растительный материал и условия выращивания | 201 |
| 3.3.2. Протокол фенотипирования зерен | 201 |
| 3.3.3. Оценка характеристик зерен | 203 |
| 3.3.4. Анализ локусов количественных признаков | 207 |
| 3.3.5. Поиск и приоритизация генов в области QTL | 208 |
| 3.3.6. Анализ характеристик зерен в популяции ITMI | 211 |
| 3.3.7. Анализ QTL для характеристик зерен в популяции ITMI | 216 |
| 3.3.8. Приоритизация генов из районов QTL для характеристик зерен | 225 |
| 3.4. Изучение взаимосвязи между характеристиками размера, формы и цвета зерен мягкой пшеницы, временем их хранения и всхожестью | 229 |
| 3.4.1. Биологический материал | 230 |
| 3.4.2. Протокол съемки зерен и обработка изображений | 230 |
| 3.4.3. Статистический анализ | 231 |
| 3.4.4. Кластеризация признаков зерен | 233 |
| 3.4.5. Анализ влияния генотипа и времени хранения на признаки зерен | 236 |
| 3.4.6. Анализ корреляций между годом урожая и признаками зерен | 236 |
| 3.4.7. Анализ связи между всхожестью и признаками зерен | 242 |
| 3.5. Заключение по главе 3 | 248 |
| Глава 4. Высокопроизводительное фенотипирование колоса пшеницы  | 252 |
| 4.1. Колос и его характеристики важные для селекционно-генетических исследований | 269 |
| 4.2. Протокол получения изображений колоса  | 254 |
| 4.3. Идентификация колоса и остей на изображении | 257 |
| 4.3.1. Предварительная обработка изображения | 257 |
| 4.3.2. Бинаризация изображения | 258 |
| 4.3.3. Идентификация остей | 258 |
| 4.3.4. Выбор параметров для выделения областей колоса и остей на изображении | 259 |
| 4.3.5. Количественные характеристики остей | 261 |
| 4.4. Алгоритмы морфометрии колоса | 262 |
| 4.4.1. Идентификация контура колоса и его выпрямление | 262 |
| 4.4.2. Интегральные характеристики формы колоса | 263 |
| 4.4.3. Модель сечений | 264 |
| 4.4.4. Модель четырехугольников | 264 |
| 4.4.5. Радиальная модель | 266 |
| 4.5. Алгоритмы предсказания индекса плотности колоса и типа его формы | 267 |
| 4.5.1. Выборка изображений колосьев | 267 |
| 4.5.2. Методы предсказания характеристик колоса | 268 |
| 4.5.3. Оценка точности предсказания плотности колоса | 270 |
| 4.6. Оценка точности распознавания областей остей и колоса | 270 |
| 4.7. Анализ характеристик колосьев и предсказание типа колоса | 275 |
| 4.7.1. Различия характеристик разных типов колосьев для модели четырехугольников | 275 |
| 4.7.2. Анализ параметров остистости для выборки колосьев | 277 |
| 4.7.3. Предсказание индекса плотности и типа формы колоса | 278 |
| 4.8. Заключение по главе 4 | 285 |
| Глава 5. Базы данных для описания фенотипа пшеницы  | 289 |
| 5.1. База данных WheatPGE | 289 |
| 5.1.1. Концептуальная модель данных | 289 |
| 5.1.2. Логическая модель данных | 292 |
| 5.1.3. Технологии реализации БД WheatPGE | 298 |
| 5.1.4. Интерфейс БД WheatPGE | 299 |
| 5.2. База данных SpikeDroid DB | 305 |
| 5.2.1. Модель данных | 305 |
| 5.2.2. Технологии реализации системы SpikeDroidDB | 308 |
| 5.2.3. Модуль интерфейса системы SpikeDroidDB | 309 |
| 5.2.4. Информационное содержание SpikeDroidDB | 313 |
| 5.3. Заключение по главе 5 | 314 |
| 6. Заключение  | 316 |
| Выводы  | 322 |
| Список публикаций в рецензируемых журналах по теме диссертации | 326 |
| Использованные сокращения | 330 |
| Список литературы  | 332 |
| Приложение | 384 |

**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 маркёр-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. Больший размер выборки позволяет уменьшить ошибку идентификации QTL, а также 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 методов высокопроизводительного компьютерного фенотипирования растений на основе анализа цифровых изображений для оценки характеристик опушения листа, размера, формы и цвета зерен, формы колоса и анализ на этой основе взаимосвязи признаков фенотипа растений с генотипом и ответом на воздействие окружающей среды.

**Задачи**, поставленные для достижения указанной цели, включали:

1. Разработку комплекса методов для компьютерного фенотипирования опушения листьев пшеницы и анализ для сортов и линий мягкой пшеницы и ее сородичей разнообразия опушения, взаимосвязи между характеристиками опушения, аллельным состоянием генов, контролирующих развитие трихом и физиологическим ответом растений на засуху.

2. Модификацию методов фенотипирования опушения листа для картофеля и табака и анализ количественных характеристик опушения для линий табака с модификацией гена пролиндегидрогеназы и для российских сортов картофеля.

3. Разработку комплекса методов для фенотипирования зерен пшеницы, оценки для них характеристик размера, формы и цвета оболочки и поиск связи между характеристиками зерен пшеницы, локусами количественных признаков, длительностью хранения в генбанке и всхожестью.

4. Разработку комплекса методов фенотипирования колосьев пшеницы для оценки количественных характеристик формы, размера, остистости и определение на основе морфологических параметров колоса его типа и плотности.

5. Разработку баз данных для сбора, хранения фенотипических данным пшеницы и доступа к ним.

Научная новизна.

Впервые разработан 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 проведен широкомасштабный анализ характеристик опушения и выделены их основные закономерности и связь с геномной конституцией;

(б) впервые количественно оценено фенотипическое проявление генов, контролирующих опушение листа у пшеницы на генетическом фоне ряда сортов и линий мягких пшениц;

(в) показаны количественные изменения характеристик опушения пшеницы в ответ на засушливые условия произрастания;

Методы фенотипирования опушения листа пшеницы были модифицированы для анализа двудольных растений: картофеля и табака. С их помощью впервые показано:

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

(б) генетических модификации гена пролиндегидрогеназы у табака оказывают влияния на количественные характеристики опушения листьев: длину и плотность трихомов.

Впервые разработан комплекс методов для фенотипирования зерен пшеницы на основе анализа цифровых изображений, полученных с помощью мобильных устройств приложением SeedCounter. С помощью разработанного приложения SeedСounter 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.

Впервые предложен комплекс методов для фенотипирования колосьев пшеницы на основе анализа цифровых изображений, реализованный в виде приложения WERecognizer для настольных персональных компьютеров. Впервые предложена модель двух четырехугольников для описания формы колоса на основе анализа его цифровых изображений и показано, что геометрические параметры колоса позволяют предсказывать его тип и плотность с высокой точностью.

**Положения, выносимые на защиту:**

1. Комплекс методов для фенотипирования опушения листьев пшеницы на основе анализа цифровых микроизображений сгибов листа позволяет с высокой точностью оценивать количественные характеристики опушения: распределение трихомов по длинам, среднее число трихомов, среднюю длину трихомов, соотношение средней длины и количества трихомов.

2. Комплекс методов для фенотипирования зерен пшеницы на основе анализа цифровых изображений приложением SeedCounter позволяет оценивать характеристики формы, размера и цвета оболочки зерен с высокой точностью.

3. Комплекс методов для фенотипирования колосьев пшеницы на основе анализа цифровых изображений позволяет оценивать характеристики остистости колоса а также его формы и размера на основе модели четырехугольников, что позволяет с высокой точностью определять тип колоса и его плотность.

**Теоретическая значимость работы.**

Разработан комплекс методов для фенотипирования важных признаков сельскохозяйственных растений: зерен, колоса (у пшеницы) и опушения листа (у пшеницы, картофеля и табака). 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 методов и баз данных 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 (ИЦиГ СО РАН, г. Новосибирск, 2012, 2014, 2016, 2018, 2020, 2022 гг.); PlantGen (ИЦиГ СО РАН, г. Новосибирск, 2015, 2017, 2019, 2021 гг.); VII Московском международном конгрессе «Биотехнология: состояние и перспективы развития» (г. Москва, 2013 г.); 1-м Международном семинаре “Генетика и геномика растений для продовольственной безопасности” (НГУ, 2016); конференции PhenoDays-2014 (г. Бон, Франция, 2014 г.); XIII Новосибирском инновационно-инвестиционном форуме «Сельхозмашиностроение и переработка агропродукции» (г. Новосибирск, 2018 г.); Школе-конференции по количественной генетике (ВИР, г. Санкт-Петербург, 2018 г.); Конференции «125 лет прикладной ботаники в России» (ВИР, г. Санкт-Петербург, 2019 г.); на VII съезде Вавиловского общества генетиков и селекционеров, посвященному 100-летию кафедры генетики СПбГУ (ВИР, г. Санкт-Петербург, 2019 г.); на II Европейской школе биотехнологов (г. Минск, Белоруссия, 2019 г.); на Международной научно-практической конференции «Развитие сельского хозяйства на основе современных научных достижений и интеллектуальных цифровых технологий «Сибирь – Агробиотехнологии» («САБИТ 2019») (СибНИИРС, п. Краснообск, 2019 г.); форуме Технопром-2019 (г. Новосибирск, 2019 г.); на Третьем Международном форуме «Передовые цифровые и производственные технологии» (СПбПУ, г. Санкт-Петербург, 2021 г.); конференции "В генетику - через разные отрасли науки и практики" (Центр Сириус, г. Сочи, 2022 г.).

**Публикации.**

По теме диссертационной работы опубликовано 24 работы, в том числе из них 12 в зарубежных и российских журналах с импакт-фактором, 8 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 соискатель 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., Пермяковой М.Д., Пермяковым 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., 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, Заключения, Выводов, Списка публикаций по теме диссертации, Списка использованных сокращений, Списка литературы и Приложения. Объем работы составляет 383 листа. В работе 82 рисунка и 29 таблиц. Список литературы включает 428 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 Арифу М.А.Р. за участие в работе на разных ее этапах.

**Глава 6. Заключение**

 Настоящая диссертация посвящена разработке методов высокопроизводительного фенотипирования растений с помощью компьютерных методов, основанных на анализе цифровых изображений. В работе проведена апробация этих методов при решении задач, связанных с поиском связи генотип-фенотип и фенотип-окружающая среда.

 Одними из важных признаков растений, связанных с физиологическим ответом на условия внешней среды является опушение. Мы предложили новый метод фенотипирования опушения листа у пшеницы на основе компьютерного анализа микроизображений сгиба листа. Предложенные нами методы фенотипирования показали свою эффективность и высокую точность при получении количественных характеристик опушения. Метод является уникальным в части полноты описания характеристик опушения. Мы показали, что с его помощью можно успешно решать задачи о взаимосвязи между фенотипом растений (опушением) и генотипом (аллельными комбинациями генов, которые контролируют развитие трихомов). Мы получили ряд результатов, которые без помощи высокопроизводительного фенотипирования было бы невозможно получить (за счет массовости анализа и детализации описания фенотипа).

 Нами впервые был проведен масштабный анализ количественных характеристик опушения листьев у пшеницы и ее сородичей, показана связь свойств опушения и геномной конституции растений (Pshenichnikova et al., 2017).

Мы впервые оценили фенотипическое проявление для трех генов *Hl*1 и *Hl*2aesp и *Hl*3, контролирующих развитие опушения листа у пшеницы, и показали, что ген *Hl*2aesp в большей степени оказывает влияние на длину трихомов. Возможно его функция связана с изменением размера клеток. Гены *Hl*3 и *Hl*1 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. Результаты подтвердили уже известные наблюдения: 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 была выявлена связь между опушением и физиологическими параметрами растений при воздействии на них условий ограниченного водообеспечения.

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*, выращеных 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 трихом 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.

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.

 Предложенный алгоритм оценки количественных характеристик опушения листьев пшеницы с небольшими модификациями был адаптирован нами для анализа опушения картофеля и табака. Мы впервые оценили разнообразие количественных характеристик опушения листьев для российских сортов картофеля.

 Мы показали для генетически-модифицированных линий табака с нарушенной функцией гена пролиндегидрогеназы, что эти линии с повышенным содержанием пролина имеют более интенсивное опушение в сравнении с контрольной линией табака SR1.

 Все это демонстрирует, что методы фенотипирования опушения листа являются эффективным инструментом для выявления связи между характеристиками опушения, генотипом и воздействием окружающей среды.

Важной группой признаков, которые связаны с урожайностью злаков являются признаки зерен. Несмотря на то, что методы оценки этих признаков 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 IPK, 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).

В диссертации представлен новый подход к фенотипированию колосьев пшеницы. Он позволяет на основе двумерных изображений определять основные характеристики размера и формы колоса. Дополнительно метод позволяет оценивать характеристики остистости колоса. В работе впервые было представлено 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.

Одно из важных направлений - это обеспечение хранения и быстрого доступа к результатам экспериментов по фенотипированию. При этом следует иметь ввиду, 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. Разработан комплекс экспериментально-компьютерных методов для фенотипирования опушения листьев пшеницы на основе анализа изображений сгибов листа. Он включает протоколы пробоподготовки образцов листьев, получения их микроизображений, алгоритмы и программы компьютерной обработки полученных изображений. Этот комплекс методов позволяет с высокой точностью по изображениям листьев оценивать количественные характеристики их опушения.

2. Разработанный комплекс методов фенотипирования опушения позволяют решать задачи оценки взаимосвязи «генотип-фенотип» и «генотип-характеристики окружающей среды». С их помощью:

(а) впервые проведен широкомасштабный анализ характеристик опушения для 47 образцов пшеницы и ее диких сородичей, относящихся к разным сортам, линиям и видам: у пшениц выделено 9 характерных типов опушения; показано что для гексаплоидных пшениц (за исключением спельты) наблюдается положительная линейная связь между числом трихом и их длиной; установлено, что характеристики опушения листьев пшеницы значимо ассоциированы с их хромосомным составом;

(б) получены оцененки фенотипического проявления генов *Hl*1, *Hl*2aesp и *Hl*3, контролирующих опушение листа у пшеницы на генетическом фоне ряда сортов и линий мягких пшениц: показано, что ген *Hl*2aesp в большей степени влияет на длину трихом, а гены *Hl*3 и *Hl*1 влияют одновременно и на длину трихомов и на их количество;

(в) Показана взаимосвязь между изменчивостью опушения листа пшеницы и ответом растения на засуху: изменение длины трихом ассоциировано с показателями флуоресценции хлорофилла и числом устьиц на нижней стороне листа; изменение числа трихом на нижней и верхней стороне листа ассоциировано с таким физиологическим параметром, как эффективность использования воды.

3. Методы фенотипирования опушения листьев были использованы для анализа двудольных растений: картофеля и табака. С их помощью показано, что:

(а) опушение листьев для 35 сортов картофеля демонстрирует широкое разнообразие по количеству трихом и их длине: наиболее интенсивное опушение проявляют сорта Великан, Танай, Кемеровчанин, Танго.

(б) линии табака с измененным паттерном экспрессии гена пролиндегидрогеназы PDH и повышенным содержанием пролина по сравнению с контрольным генотипом SR1 имеют большее количество трихом на адаксиальной стороне листа, при этом длина трихом достоверно меньше, чем у растений контрольной линии.

4. Разработан комплекс экспериментально-компьютерных методов для фенотипирования зерен пшеницы на основе анализа их цифровых изображений. Он включает: протоколы получения изображений, в том числе и на камеру мобильного устройства, процедуру цветокоррекции изображений, компьютерный анализ с помощью разработанного нами приложения SeedCounter для оценки 7 количественных характеристик формы и размера зерен и 48 характеристик цвета.

5. С помощью разработанного комплекса методов проведено фенотипирование 114 рекомбинантных инбредных линий мягкой пшеницы популяции ITMI по размеру, форме и цвету оболочки зерен и на основе этого:

(а) выявлено 212 аддитивных и 34 пары эпистатических QTL на всех хромосомах генома пшеницы (кроме хромосом 1A и 5D); продемонстрировано перекрывание расположения QTL, слабое для признаков размера/формы семян и сильное для признаков окраски оболочки; идентифицировано 29 генов-кандидатов для локусов, ассоциированных с размером семян, и 14 для локусов, ассоциированных с цветом оболочки;

(б) показано, что для растений с различными генотипами, выращенных в четыре разных сезона (2003, 2004, 2009 и 2014 гг.) и хранившиеся в условиях генбанка IPK, генотип значимо влияет на изменчивость изученных признаков зерен, а сезон выращивания - на 75% из них (включая все изученные признаки формы и размера);

(в) для большинства признаков цвета зерна обнаружена линейная зависимость между величиной признака и сезоном выращивания: значимые отрицательные зависимости выявлены между сезоном выращивания и краснотой оболочки зерен; значимые положительные зависимости выявлены между сезоном выращивания и светлотой оболочки зёрен (чем больше время хранения зерна, тем более красный и более светлый оттенок имеют зерна);

(в) для генотипов разных сезонов выращивания выявлена значимая отрицательная линейная зависимость между красной компонентой цвета оболочки и всхожестью: чем краснее зерна, тем ниже их всхожесть.

6. Разработан комплекс экспериментально-компьютерных методов для фенотипирования колосьев пшеницы на основе анализа цифровых изображений. Он включает: протоколы получения изображений, процедуру цветокоррекции, а также компьютерный анализ формы и размера колоса на основе геометрической модели двух четырехугольников и общих характеристик размера и формы колоса с помощью разработанного нами приложения WERecognizer.

7. Показано, что предложенные геометрические параметры колоса позволяют с высокой точностью предсказывать его тип («спельта», «нормальный» или «компактный») и плотность.

8. Разработаны Базы данных WheatPGE и SpikeDroidDB, в которых накапливается информация о фенотипических характеристиках пшеницы (опушения листьев, характеристик зерен, колосьев), получаемых с помощью разработанных экспериментально-компьютерных методов; базы данных обеспечивают доступ к информации по фенотипическим признакам пшеницы, характеристикам генотипа и места произрастания.

**Список публикаций в рецензируемых журналах 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 цветовых 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 SeedСounter. 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 characheristics 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):863­868. 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 определеиня 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 A.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.