

На правах рукописи

КОМЫШЕВ ЕВГЕНИЙ ГЕННАДЬЕВИЧ

**РАЗРАБОТКА МЕТОДОВ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ
КОЛИЧЕСТВЕННЫХ ХАРАКТЕРИСТИК, ОПИСЫВАЮЩИХ
ФЕНОТИПИЧЕСКИЕ ПРИЗНАКИ КОЛОСА ПШЕНИЦЫ**

Математическая биология, биоинформатика
03.01.09

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск – 2021

Работа выполнена в лаборатории эволюционной биоинформатики и теоретической генетики Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Научный руководитель:	Афонников Дмитрий Аркадьевич к.б.н., зав. лаб. эволюционной биоинформатики и теоретической генетики ФГБУН «ФИЦ Институт цитологии и генетики СО РАН», г. Новосибирск
Официальные оппоненты:	Ратушняк Александр Савельевич д.б.н., ведущий научный сотрудник, и.о. зав. Лабораторией биомедицинской информатики Федерального исследовательского центра Информационных и вычислительных технологий Сибирского отделения Российской академии наук, г. Новосибирск Козлов Константин Николаевич к.б.н., ведущий научный сотрудник научно-исследовательской лаборатории Института прикладной математики и механики Санкт-Петербургского политехнического университета Петра Великого, г. Санкт-Петербург.
Ведущее учреждение:	ФГАОУ ВО «Севастопольский государственный университет», г. Севастополь.

Защита диссертации состоится « »_____ 2021 г. на утреннем заседании диссертационного совета Д 003.011.01 на базе Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в конференц-зале Института по адресу:
пр. академика Лаврентьева 10, г. Новосибирск, 630090,
тел.: (383) 363-49-06 (1321); факс: (383) 333-12-78;
e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте института www.bionet.nsc.ru.

Автореферат разослан « »_____ 2021 г.

Учёный секретарь диссертационного совета,
доктор биологических наук

Хлебодарова Т. М.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Одной из глобальных современных проблем человечества является рост населения и связанная с этим необходимость в увеличении производства продуктов питания. Согласно прогнозу Продовольственной и сельскохозяйственной организации ООН, производство зерновых культур должно удвоиться до 2050 года, чтобы удовлетворить спрос на продовольствие со стороны растущего населения мира.

В связи с этим возрастает необходимость поиска новых, менее прихотливых к условиям выращивания сортов с более высоким уровнем урожайности и качеством зерна. На решение этих задач направлены современные методы генетики и селекции сельскохозяйственных культур, основой которых является выявление связи между генотипом, окружающей средой и фенотипом (Benfey and Mitchell-Olds, 2008).

Традиционные подходы к фенотипированию являются трудоемкими и дорогостоящими, особенно когда в эксперименте проводится анализ сотен и тысяч растений. Для повышения точности и производительности измерений, а также сокращения трудозатрат и исключения человеческого субъективизма при проведении измерений необходима разработка новых подходов к фенотипированию растений (Hancock, 2014), основанные, прежде всего, на методах анализа цифровых изображений (Li, 2014). В дополнение к разработке методов фенотипирования для хранения полученных данных возникает необходимость в стандартизации описания фенотипа растений в базах данных. Общепринятого стандарта описания результатов экспериментов не существует. Генетики и селекционеры производят измерения характеристик растений вручную и записывают результаты в полевой журнал в свободном формате. При этом не сохраняется каких-либо данных о контексте измерений – с помощью чего производились измерения, с какой точностью и в каких единицах записаны значения. В результате полученные данные невозможно как-либо анализировать, не имея при этом связи с экспериментатором, получившим эти данные.

Одними из важных признаков сельскохозяйственных растений являются показатели урожайности. У злаков они связаны с размером и формой колоса, количеством зерен и их массой (Farooq et al., 2015). Однако высокопроизводительные методики для определения характеристик колосьев и зерен злаков в настоящее время недостаточно развиты. Это приводит к низкой производительности труда селекционера-генетика и затрудняет создание новых высокопродуктивных сортов и линий растений. В этой связи

актуальной задачей для современной генетики растений является создание методов высокопроизводительного фенотипирования, которые позволили бы быстро и точно осуществлять оценку характеристик колосьев и зерен злаков, в частности, пшеницы. Для эффективного хранения результатов фенотипирования необходимо развитие баз данных, в которых бы хранилась разнородная информация: экспертные оценки фенотипа, изображения колосьев, количественные оценки фенотипа, полученные на основе анализа изображений.

Цели и задачи работы

Целью работы является разработка методов автоматического определения количественных морфометрических характеристик колосьев и зерен пшеницы на основе анализа их цифровых изображений.

В соответствии с поставленной целью были сформулированы следующие задачи:

1. Разработка метода морфометрии зерен пшеницы с использованием мобильных устройств.
2. Разработка методов автоматического определения количественных характеристик формы и размера колоса на основе двухмерных изображений и его апробация на примере анализа колосьев пяти видов гексаплоидных пшениц.
3. Разработка базы данных для накопления, хранения и систематизации информации о фенотипических признаках колоса пшеницы.

Научная новизна

Впервые создано мобильное приложение SeedCounter для устройств под управлением ОС Android для подсчета зерен и определения их размеров.

Впервые предложен новый метод автоматического определения количественных морфометрических характеристик колоса пшеницы на основе анализа цифровых двухмерных изображений, который позволяет описывать форму колоса на основе модели двух четырехугольников.

Проведена оценка сходства и различий формы колосьев с использованием количественных характеристик, полученных у 14 образцов растений пяти генотипов мягкой пшеницы на основе параметров модели, оцененных путем анализа двухмерных цифровых изображений.

Разработана компьютерная система SpikeDroid, реализующая технологии хранения разнородной информации, для поддержки работы селекционера по сбору данных по морфометрическим характеристикам колосьев пшеницы и их диких сородичей.

Теоретическая и практическая ценность

Предложенные методы позволяют оценивать такие характеристики колоса пшеницы, как длина, ширина, остистость, количество зерен в колосе, а также характеристики зерен пшеницы (длина, ширина, проецируемая на поверхность площадь, коэффициенты округлости, закругленности, шероховатости и целостности формы контура и др.). Это дает возможность оценивать на основе анализа изображений характеристики урожайности растений с высокой степенью точности. Предложенные методы позволяют исключить субъективизм присущий человеку при проведении измерений, т.к. обработка цифровых изображений производится автоматически. Кроме того, предложенный подход не требует от пользователя специфических знаний в области фенотипирования пшеницы.

Мобильное приложение SeedCounter позволяет выполнять измерения в полевых условиях, сохранять на мобильном устройстве данные, а также отправлять их на сервер посредством сети Интернет, когда он доступен.

Разработанная система SpikeDroid позволяет существенно ускорить процесс массового фенотипирования благодаря автоматизации этапов начиная от получения изображений, заканчивая статистическим анализом занесенных в базу данных параметров.

Собранные воедино в системе SpikeDroid данные с анализируемых изображений, ручного фенотипирования и генотипов позволяют повысить эффективность существующих методов сравнительной генетики пшениц, таких как гибридологический метод анализа, метод возвратных скрещиваний, метод циклических скрещиваний и др.

Преимущества разработанных методов позволяют существенно повысить эффективность селекционно-генетических экспериментов в направлении создания новых высокопродуктивных сортов и линий пшеницы.

Положения, выносимые на защиту

- 1) Методы фенотипирования колосьев и зерен пшеницы на основе анализа цифровых двухмерных изображений, реализованные в виде приложений WERecognizer и SeedCounter, позволяют проводить оценку характеристик продуктивности растений в автоматизированном режиме в массовых селекционно-генетических экспериментах.
- 2) Геометрическая модель колоса, описывающая его форму и размер в виде четырехугольников, предсказывает, что такие характеристики колоса, как длина, размер центральной части, ширина основания, площадь центрального сегмента и основания колоса являются наиболее значимыми для определения вида гексаплоидной пшеницы.

- 3) Компьютерная система SpikeDroid обеспечивает накопление, хранение, систематизацию и поиск информации о фенотипических признаках колоса, полученных из различных источников, а также доступ к ним через Web-интерфейс в сети Интернет.

Апробация работы

Материалы работы вошли в отчеты по грантам РНФ и РФФИ.

Результаты работы были представлены на всероссийских международных научных конференциях и школе молодых ученых в Новосибирске (BGRS/SB 9,10; Международный научный симпозиум «Генетика и геномика растений для продовольственной безопасности» 2016; 3-я Международная конференция «Генетика, геномика, биоинформатика и биотехнология растений», PlantGen, 2015; Актуальные проблемы вычислительной и прикладной математики, АМСА, 2015 и др.) и в Москве (Всероссийская Конференция «50 лет ВОГиС: успехи и перспективы», 2016; международный конгресс «Биотехнология: состояние и перспективы развития» 2019).

Мобильное приложение SeedCounter было использовано в целом ряде научных работ для оценки характеристик урожайности и морфологии зерен пшеницы при решении задач генетики и селекции (Lobo et al., 2019; Kroupin et al., 2019; Kroupin et al., 2020a; Kroupin et al., 2020b; Бабоева и Усманов, 2020; Черноок и соавт., 2019; Divashuk et al., 2019), оценки методики проращивания зерна для пищевой индустрии (Науменко и соавт., 2019), анализа формы и размеров клубней картофеля (Иванова и соавт. 2019). Таким образом, оно получило практическую апробацию в нескольких независимых научных группах как в России, так и за рубежом.

Объем и структура диссертации

Диссертация изложена на 146 страницах. Работа содержит 36 рисунков и 17 таблиц. Список литературы содержит 135 источников.

Публикации

По материалам диссертации опубликовано 17 работ, из них 5 статей в рецензируемых научных журналах, входящих в перечень ВАК, 12 тезисов конференций. Получено два авторских свидетельства.

Личный вклад автора

Основная часть работы выполнена автором самостоятельно. В работах по созданию метода фенотипирования колосьев пшеницы автор принимал участие в разработке алгоритма анализа изображений, ручной разметки изображений, статистическом анализе полученных данных, оценке точности.

В работах по созданию метода фенотипирования зерен пшеницы автор разрабатывал мобильное приложение SeedCounter, проводил оценку точности, проводил статистический анализ полученных данных.

Автор принимал участие в создании модели, описывающих форму колоса пшеницы, проведении вычислительных экспериментов, обсуждении и анализе полученных результатов.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

Обзор литературы состоит из шести разделов и заключения. Первый раздел посвящен объекту исследования - пшенице – одной из наиболее значимых сельскохозяйственных культур, с огромным генетическим потенциалом. В разделе охарактеризован геном пшеницы, описана структура растения, гены, определяющие морфологию колоса и важнейшие характерные черты, тесно связанные с урожайностью. В разделе 1.2 описаны методы обработки биологических цифровых изображений, приведены примеры применения программных средств для анализа данных. В следующем разделе приведены наиболее распространенные библиотеки анализа цифровых изображений в биологии. В разделах 1.4 и 1.5 рассмотрены описанные на данный момент в мировой научной литературе методы фенотипирования колосьев и зерен пшеницы, а также базы данных и онтологии в области феномики, селекции и генетике растений. В разделе 1.6 представлены наиболее распространенные программные пакеты для статистического анализа данных.

Глава 2. Материалы и методы

В качестве растительного материала для морфометрии зерен использовались образцы пяти генотипов пшеницы из коллекции зерновых культур лаборатории хромосомной инженерии злаков сектора генетики качества зерна Института Цитологии и Генетики СО РАН: Аленькая 1102 II-12, 84/98w 99 II-13, Synthetic 6x x-12, Purple Chance 4480 II-03 и Alcedo n-99. Для анализа колосьев использовались растения из коллекции д.б.н. Гончарова Н.П., включающую восемь видов гексаплоидной пшеницы и F2 гибриды от скрещивания австралийского сорта мягкой пшеницы Triple Dirk с образцом KU506 китайской пшеницы *Triticum yunnanense* (син. *Triticum spelta* ssp. *yunnanense* (King ex S.L. Chen) N.P. Gontsch.). В главе приведена и описана общая для морфометрии зерен и колосьев блок-схема алгоритма обработки изображений (рис. 1); описаны методы оценки точности алгоритмов анализа

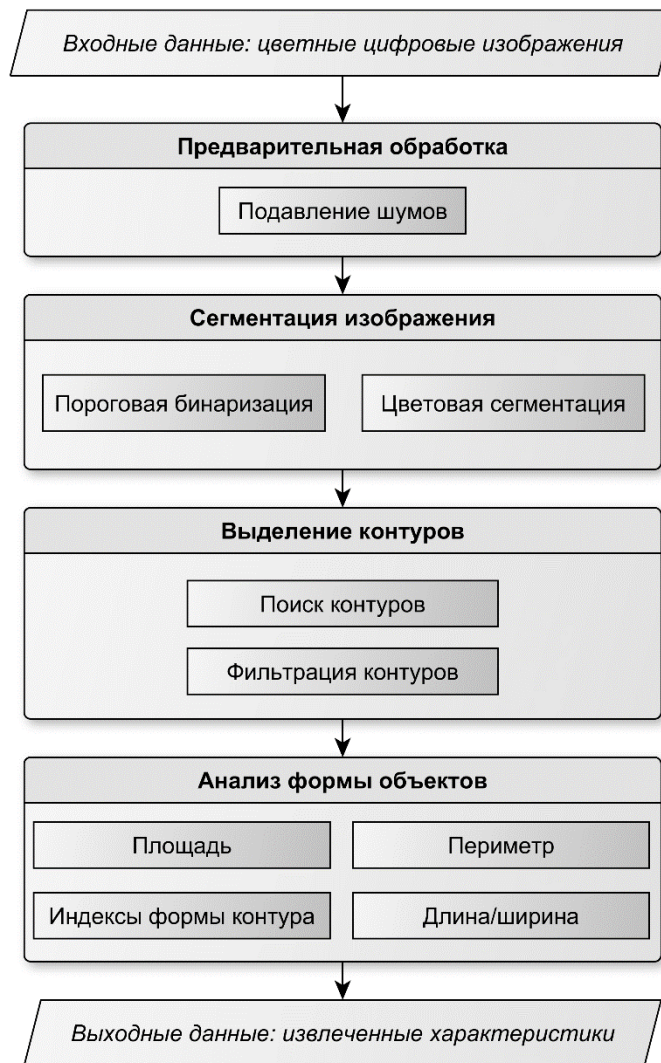


Рисунок 1. Блок-схема алгоритма обработки изображения. Ход алгоритма представлен сверху вниз. Параллелограммами представлены входные и выходные данные. Прямоугольные блоки представляют отдельные шаги алгоритма, объединенные в логические блоки-этапы.

изображений; приведена оценка точности автоматической идентификации контуров - индекс Жаккара J (Jaccard, 1912), а также две оценки точности сравнения наборов скалярных значений: MAE (mean absolute error), средняя абсолютная ошибка и MAPE (mean absolute percentage error), средняя абсолютная относительная ошибка:

$$MAE = \frac{1}{M} \sum_{j=1}^{j=M} |n_j - n'_j|$$

$$MAPE = \frac{100}{M} \sum_{j=1}^{j=M} \left(\frac{|n_j - n'_j|}{n_j} \right)$$

где j - номер измерения, n_j - значение, полученное при фенотипировании вручную, n'_j - значение, полученное на основе анализа изображения, а M -

количество измерений. Чем больше значения MAE и MAPE, тем больше значение ошибки измерений компьютерным методом. Если значения MAE и MAPE близки к 0, то и ошибка низкая.

Приведены использованные в работе методы статистического анализа полученных данных, и соответствующие программные пакеты, включающие: коэффициенты корреляция Пирсона, Кендалла, Спирмена, дисперсионный и кластерный анализ, метод главных компонент. Кратко описаны использованные технологии разработки приложений для платформы Android (раздел 2.5) и методы реализации баз данных (раздел 2.6).

Глава 3. Метод морфометрии зерен пшеницы с помощью мобильных устройств

Метод реализован в виде мобильного приложения SeedCounter для платформы Android. Протокол получения изображений зерен пшеницы определяет условия освещения, подготовку зерен, подготовку фотографируемой поверхности, положение и настройку мобильного устройства. Для получения изображений используется белый лист бумаги стандартного формата (A4 по умолчанию), расположенный на контрастном фоне. На листе бумаги располагаются очищенные зерна пшеницы. Фотографирование производится с захватом всех границ листа (рис. 2).

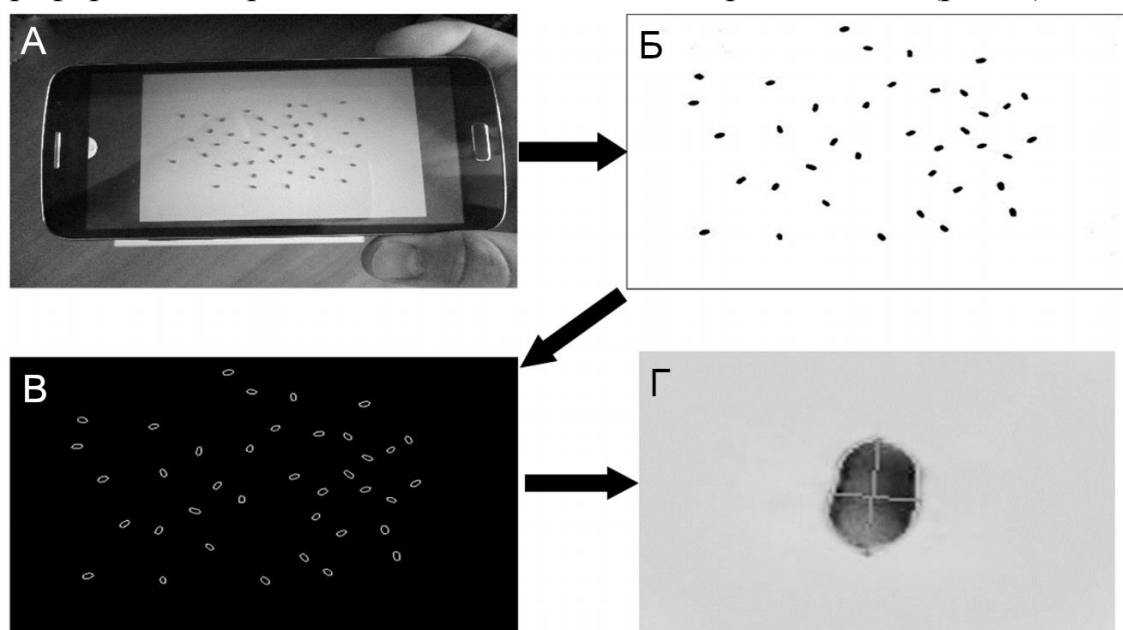


Рисунок 2. Основные этапы распознавания зерна на листе бумаги. Стрелками указан порядок следования этапов. (А) Захват изображения с помощью камеры мобильного устройства и распознавания листа бумаги. (Б) Изображение после аффинного преобразования и бинаризации. (В) Контуры зерна, идентифицированные на изображении. (Г) Изображение зерна с основными осями, показанное скрещенными линиями.

Алгоритм анализа изображений для определения характеристик формы и размера зерен пшеницы состоит из нескольких этапов, включающих распознавание листа бумаги для последующей коррекции перспективы и

определения масштаба, а также идентификацию зерен и их морфометрию (рис.2). Распознавание листа бумаги осуществляется с помощью адаптивной бинаризации изображения в градациях серого, в то время как для идентификации зерен используется сегментация изображения в пространстве HSV для более точного определения границ и снижения искажений, вызванных тенями от зерен. Для определения границ между зернами, которые соприкасаются на изображении используется метод водораздела (Roerdink and Meijster, 2000).

Разработанное мобильное приложение SeedCounter позволяет подсчитывать число зерен, оценивать их длину, ширину, проецируемую на поверхность площадь и прочие характеристики формы контуров зерен.

Помимо непосредственно подсчета и измерения зерен интерфейс программы позволяет настраивать параметры алгоритма обработки изображений и контуров, выбирать различные форматы бумаги и просматривать сохраненные данные (рис. 3).

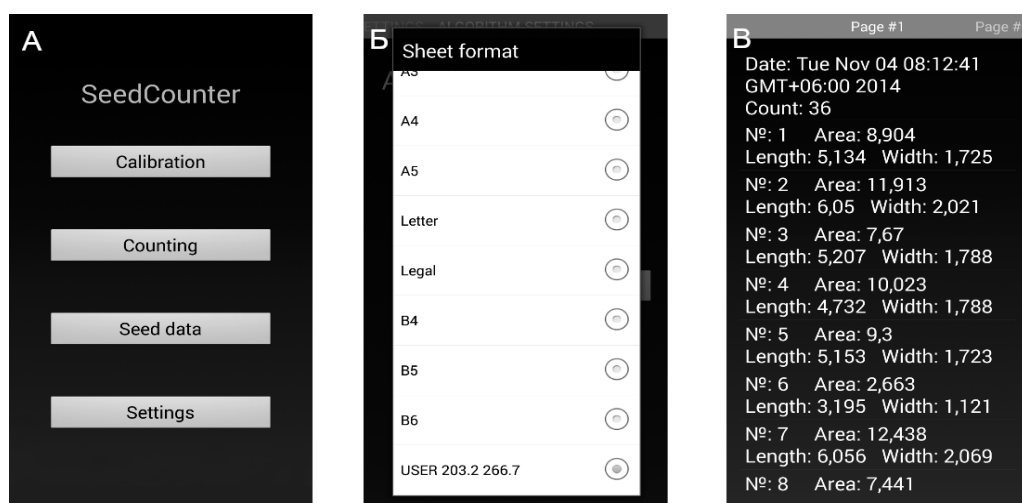


Рисунок 3. Интерфейс приложения SeedCounter. А) Главное меню. Б) Выбор формата листа бумаги. В) Экран вывода результатов измерений (количество зерен и длина/ширина/площадь каждого зерна).

Оценка точности приложения SeedCounter производилась для двух показателей: оценки числа зерен приложением в сравнении с ручным подсчетом и оценки их размеров приложением в сравнении с ручным измерением на основе микроизображений, полученных с микроскопа Carl Zeiss Aioscop 2 plus с цифровой камерой с адаптером AxoCamHRcTV2/3c 0.63. Оценка точности производилась в шести различных условиях освещения (табл. 1) для трех мобильных устройств: двух смартфонов и одного планшета. Значение MAE для количества зерен на листе было около 1, а значение MAPE близко к 2%. Точность оценки размера зерна составляла приблизительно 0,30 мм (средняя для всех серий: 0,31 мм), что составляет приблизительно 8% от

линейных размеров зерна (средний для всех серий: 8,03%) (табл. 2). Точность SeedCounter оказывается незначительно ниже точности широко известной программы SmartGrain (табл. 2), которая, однако, для получения изображений требует использования сканера, обеспечивающего лучшее качество изображений.

Таблица 1. Условия освещения для измерения точности приложения SeedCounter.

Обозначение	Условия освещения	Световой поток (люмен, люкс)	Температура света
L1	Лампа дневного света, 11 ватт	900 лм	4000K
L2	Лампы дневного света: 11 ватт (1 шт.) и 5 ватт (2 шт.)	1700 лм	4000K
L3	Лампы дневного света, 11 ватт и четыре лампы дневного света по 5 ватт	2500 лм	4000K
L4	Лампа дневного света, 11 ватт, четыре лампы дневного света по 5 ватт и галогеновая лампа, 35 ватт	2690 лм	4000K и 2700K
L5	Дневной свет, снаружи, пасмурная погода	(1280 лк)	–
L6	Дневной свет, снаружи, солнечная погода	(656000 лк)	–

Таблица 2. Средние значения точности оценки приложением SeedCounter длины и ширины в различных условиях освещения. MAE – средняя абсолютная ошибка длины и ширины зерна. MAPE – средняя абсолютная относительная ошибка длины и ширины зерна. R, K, S – средние значения коэффициентов корреляции Пирсона, Кенделла и Спирмана для длины и ширины зерна, оцененной SeedCounter и измеренной вручную.

Условия освещения	Среднее MAE	Среднее MAPE	Среднее R	Среднее K	Среднее S
L1	0.297	7.861	0.861	0.610	0.772
L2	0.301	7.795	0.860	0.610	0.775
L3	0.289	7.619	0.866	0.618	0.781
L4	0.322	8.318	0.855	0.601	0.766
L5	0.350	8.535	0.799	0.567	0.735
L6	0.344	8.416	0.825	0.578	0.754
SmartGrain	0.305	6.973	0.917	0.726	0.890

Коэффициенты корреляции Пирсона между длиной зерен, измеренной с помощью микроскопа и длиной измеренной с помощью мобильного приложения SeedCounter во всех экспериментах были не ниже 0,79. Для ширины зерен этот параметр был меньше, но больше 0,67 (рис. 4). Оба коэффициента корреляции были значимыми при $p < 0,01$.

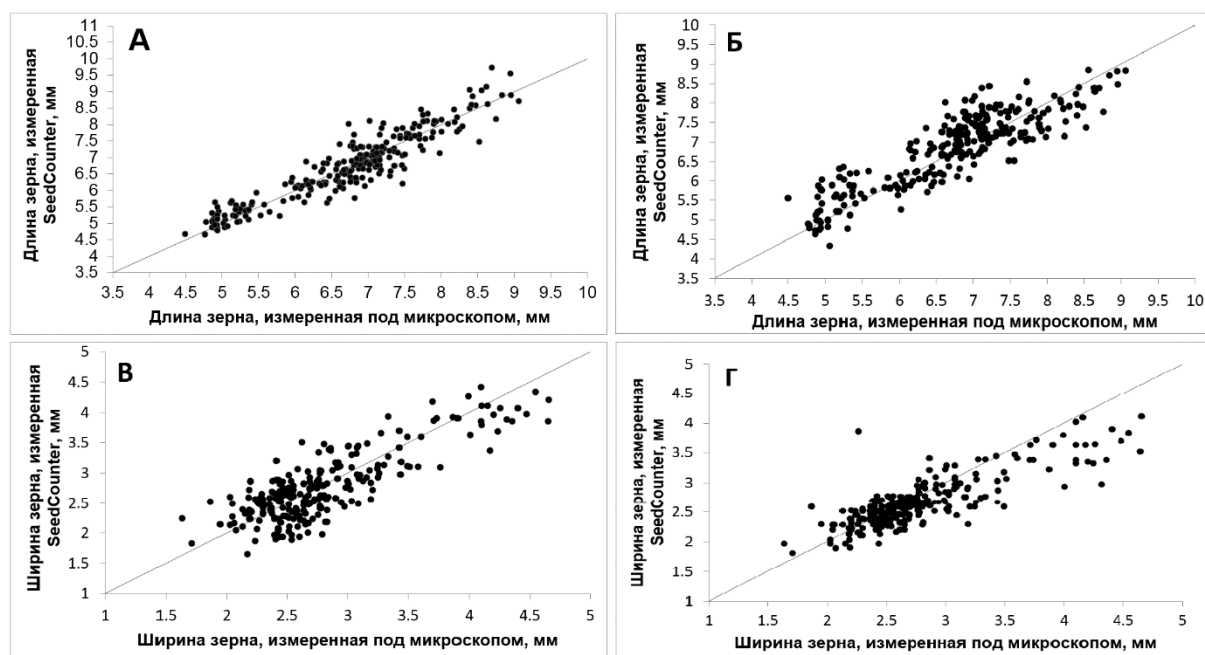


Рисунок 4. Диаграмма рассеивания размеров зерен, измеренный мобильным устройством Samsung (ось Y) относительно размеров, измеренных под микроскопом (ось X). А) длина зерен в искусственных условиях освещения; (Б) длина зерен в естественных условиях освещения; В) Ширина зерен в искусственных условиях освещения; Г) ширина зерен в естественных условиях освещения.

Глава 4 Метод морфометрии колоса пшеницы

Данная глава посвящена разработке метода морфометрии колоса пшеницы на основе изображений, полученных с помощью цифровой фотокамеры. Предложены два протокола получения изображений: «на столе» и «на прищепке» (раздел 4.1). В качестве фона при фотографировании используется синий лист бумаги. В обоих протоколах требуется цветовая шкала ColorChecker Mini Classic target (<https://xritephoto.com/camera>) в области каждого кадра, которая позволяет произвести цветовую коррекцию и определить масштаб полученного изображения. Коррекция цвета позволяет устранить ошибки цветопередачи колоса, возникающих из-за условий освещения. В протоколе «на столе» колос и цветовая шкала располагаются на горизонтальной поверхности (стол) без дополнительной фиксации, что позволяет производить съемку с минимальным оборудованием. В протоколе «на прищепке» колос и цветовая шкала закрепляются на специальном штативе, что позволяет производить съемку в разных проекциях, вращая колос вокруг продольной оси. На рис. 5 приведены примеры изображений, получаемых с помощью двух протоколов.



Рисунок 5. Изображения колосьев, полученные по двум различным протоколам. (А) На столе. (Б) На прищепке.

Алгоритм идентификации колоса на изображении состоит из нескольких этапов: фильтрации шумов, распознавании цветовой шкалы, сегментации в цветовом пространстве HSV. Сегментация выполняется на основе предварительно подобранных параметров – целевых цветов, соответствующих колосу. Подбор данных параметров производился с помощью генетического алгоритма (Whitley, 1994). Далее производится распознавание остей (на основе модифицированного преобразования скелетизации), идентификация контура «тела» колоса, его выпрямление и морфометрия. В работе (раздел 4.3) предложена модель представления формы колоса в виде двух четырехугольников (рис. 6).

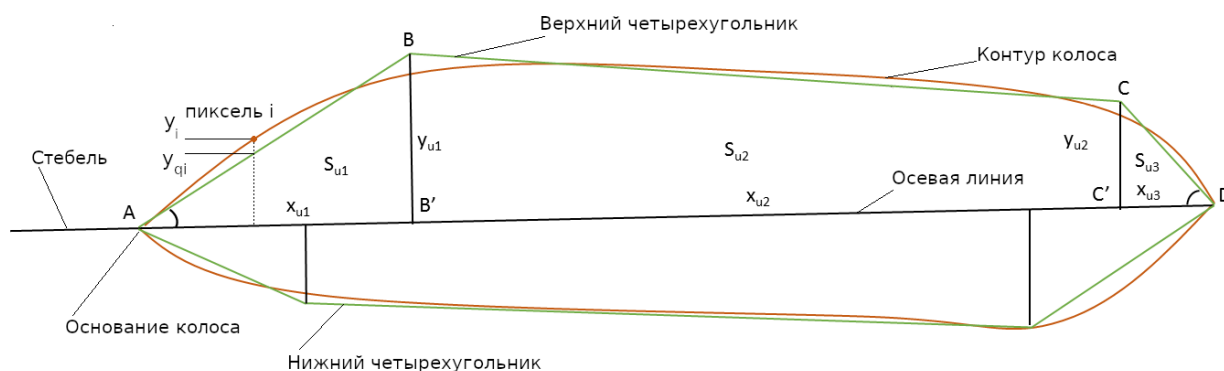


Рисунок 6. Представление формы колоса в виде двух четырехугольников. Черной горизонтальной линией показана осевая линия колоса. Контур колоса показан коричневой линией. Четырехугольники, аппроксимирующие контур колоса, показаны зелеными линиями. Основные параметры, характеризующие геометрию, показаны для верхнего четырехугольника. Пунктирная линия демонстрирует перпендикуляр, который строится от каждого i -го пикселя контура к осевой линии (y_i координата Y для i -го пикселя контура, y_{qi} - координата Y проекции этого пикселя на ребро верхнего четырехугольника).

Данная модель описывается 9 независимыми параметрами: длиной центральной оси колоса, L , параметрами x_{u1} , x_{u2} , y_{u1} , y_{u2} для верхнего и соответствующими параметрами для нижнего четырехугольника (рис. 6).

Дополнительно мы оценивали производные параметры для модели: площадь аппроксимирующих колос четырехугольников (S), площадь трех сегментов четырехугольников (S_1 , S_2 , S_3) и др. Для контура колоса алгоритм позволяет вычислить ряд интегральных характеристик: округлость (Circularity), закругленность (Roundness), цельность (Solidity), площадь контура колоса (Ear area), его периметр (Perimeter). Остистость характеризовали как общую площадь остей на изображении (Awns area). В совокупности мы определяли 25 количественных характеристик колоса.

В разделе 4.4 приведена методика оценки точности распознавания областей остей и «тела» колоса, а в разделе 4.5 описаны результаты анализа характеристик формы колоса для выборки из 160 колосьев 5 видов пшеницы. Мы рассчитали коэффициенты корреляции Пирсона r между основными характеристиками колосьев в этой выборке и с помощью меры близости $d_r = (1 - r)$ провели иерархический кластерный анализ этих характеристик. Результаты приведены в виде дерева на рис. 7. На дереве выделяются четыре кластера признаков. Первый включает характеристики площади и линейных размеров основания колоса; второй включает характеристики формы вершины колоса (длина вершинного сегмента, его площадь, ширину, а также индекс шероховатости контура колоса); третий включает интегральные характеристики формы контура колоса (округлость, закругленность, цельность, площадь контура, периметр и площадь остей); четвертый кластер включает длину колоса и характеристики его центрального сегмента.

Таким образом, в предложенном подходе форма колоса характеризуется тремя основными сегментами: основание, центральная часть и вершина. Характеристики, относящиеся к этим сегментам между собой статистически связаны. Группы этих параметров могут отражать как особенности генетического контроля формы колоса, так и специфику модели описания формы колоса.

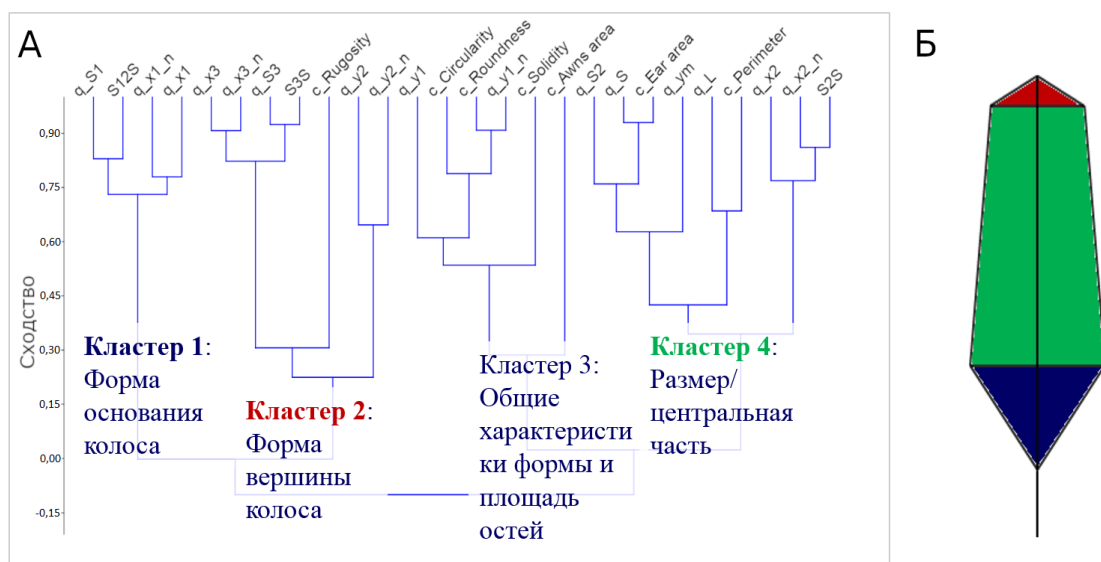


Рисунок 7. А) Результат кластеризации характеристик формы колоса на основе меры близости признаков d_r . Б) Представление формы колоса в виде модели четырехугольников. Синим обозначен сегмент колоса у основания. Зеленым – центральный сегмент колоса. Красным – сегмент колоса у вершины.

С использованием рассчитанных количественных характеристик колоса мы провели анализ формы и размера колосьев для пяти видов пшеницы (см. раздел материалы и методы). Для этого был использован анализ главных компонент (рис. 8), который показал следующее. Разброс количественных характеристик колосьев у каждого вида является существенным, так, что на графике (рис. 8) области, которые занимают эти виды, существенно перекрываются. Тем не менее, у свойств колосьев этих растений наблюдаются и заметные различия. Так большинство растений видов *T. aestivum* на графике представлены в левом верхнем квадранте, а *T. compactum* в правом верхнем квадранте. Вид *T. spelta* представлен только в левой части графика, а виды *T. antiquorum* и *T. sphaerococcum* представлены только в нижней части (рис. 8). Зеленые линии на графике обозначают характеристики и их вклад в первые две компоненты, объясняющие дисперсию. Анализ показал, что изученные генотипы различаются прежде всего по размеру колосьев (соответствующее направление изменчивости на рис. 8 показано голубой стрелкой).

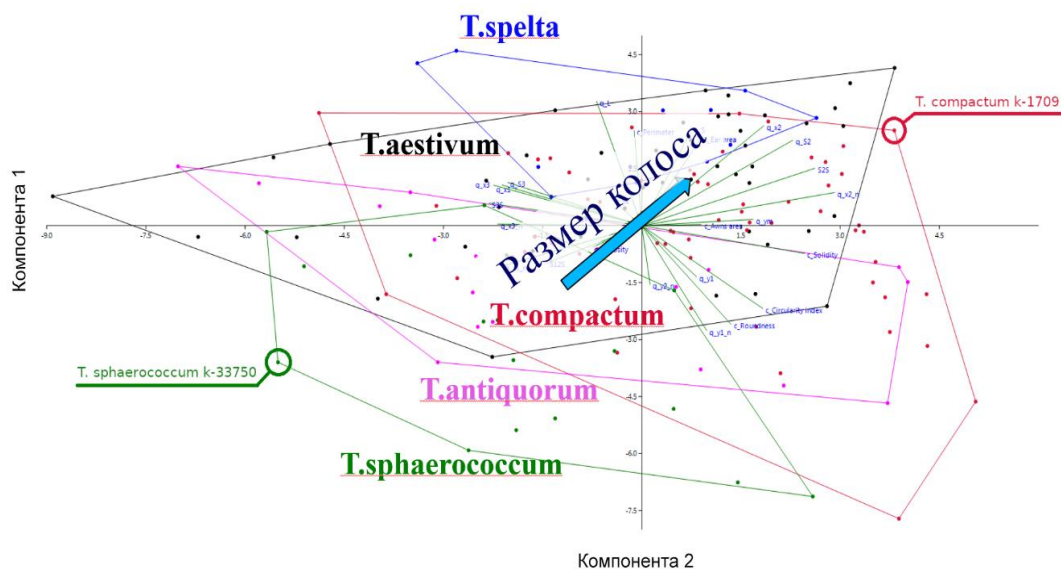


Рисунок 8. Точечный график разложения параметров колосьев на первые две главные компоненты. Черным – образцы вида *T. aestivum*; синим – *T. spelta*; малиновым – *T. compactum*; зеленым – *T. sphaerococcum*; пурпурным - *T. antiquorum*. Зеленой и красной окружностями выделены два контрастных колоса разных видов.

Мы рассмотрели влияние видовой принадлежности на основные вычисляемые характеристики формы колоса. Однофакторный дисперсионный анализ выявил ряд статистически значимо различающихся параметров: периметр контура колоса, площадь контура колоса, площадь остей колоса, индекс округлости, индекс закругленности, индекс целостности формы и др.

Чтобы уточнить влияние видовой принадлежности на некоторые отдельные признаки был проведен анализ попарных сравнений характеристик колосьев для всех возможных генотипов по критерию Манна-Уитни с поправкой Бонферрони на множественную проверку гипотез и на основе однофакторного дисперсионного анализа. Обнаружилось, что большинство генотипов достоверно различаются по таким признакам как длина центрального сегмента, ширина колоса у вершины, длина колоса, площадь центрального сегмента, площадь модели четырехугольников. Однако виды *T. sphaerococcum* и *T. aestivum* выглядят слабо различающимися: они не демонстрируют значимых различий по характеристикам длины центрального сегмента, ширины у вершины колоса и длины колоса (которые различаются для большинства остальных пар). Тем не менее, для таких характеристик как площадь модели четырехугольников и площадь центрального сегмента колосья этих видов значимо различаются.

Глава 5 Информационная система для аннотации морфометрических характеристик колоса пшеницы

Глава 5 посвящена описанию разработанной информационной системы SpikeDroid, позволяющей систематизировать и аннотировать накопленные на основе анализа изображений данные. SpikeDroid позволяет структурированно хранить цифровые изображения колоса, производить их аннотацию, импортировать новые данные и изображения, предоставляет пользователю гибкую систему запросов для доступа к данным и их экспорта. Web-интерфейс системы доступен по адресу <http://spikedroid.biores.cytogen.ru> и позволяет работать с ней, как со стационарных компьютеров, так и с мобильных устройств. Пример страницы интерфейса системы, описывающая характеристики фенотипа колоса приведена на рис. 9.

View Edit Delete Manage display Devel

Plant: [8739](#)

Spike number: 1	Images  Protocol: pin
Spike length: 7.00cm	
Spike front width: 1.10cm	
Spikelet count: 22	
Density index: 0.33	
Spike color: White	 Protocol: pin
AwN type: awnletted (short)	 Protocol: pin
AwN color: White	 Protocol: pin
Shape type: compact	 Protocol: object table
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 9. Страница описания фенотипа колоса системы SpikeDroid. Слева панель фенотипических характеристик колоса. Справа список изображений, полученных по предложенному протоколу съемки.

ВЫВОДЫ

1) Разработано приложение SeedCounter для мобильных устройств и персональных компьютеров, позволяющее оценивать с высокой точностью количество и размеры зерен пшеницы, расположенных на листе белой бумаги стандартного формата.

2) Впервые разработан метод определения количественных характеристик размера, формы и остистости колоса пшеницы на основе анализа двухмерных изображений, реализованный в виде приложения WERecognizer для персональных компьютеров.

3) Впервые предложена геометрическая модель представления колоса в виде двух четырехугольников, позволяющая оценить его форму и размер.

Модель характеризует основные количественные параметры трех сегментов колоса: основания, центральной части и вершины.

4) Анализ 14 генотипов пяти видов гексаплоидных пшениц с помощью предложенной модели позволил выявить признаки колоса, значительно различающиеся для отдельных генотипов: размер центральной части, длину колоса, ширину основания, площадь центрального сегмента и основания колоса. Полученные данные полностью согласуются с существующей классификацией и подтверждают её.

5) Разработана база данных SpikeDroidDB для накопления, хранения, систематизации и поиска информации о фенотипических признаках колоса. Система SpikeDroid позволяет автоматически анализировать изображения колосьев пшеницы, полученных по установленному протоколу, содержит данные экспертной аннотации колоса пшеницы; морфометрические характеристики колоса и зерен; изображения колосьев; данные о образцах растений и их генотипе; описания коллекций образцов, месте их произрастания.

СПИСОК ОСНОВНЫХ РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ:

1. Афонников Д.А., Генаев М.А., Дорошков А.В., **Комышев Е.Г.**, Пшеничникова Т.А. Методы высокопроизводительного фенотипирования растений для массовых селекционно-генетических экспериментов // Генетика. – 2016. – Т. 52. – №. 7. – С. 788-803.
2. **Komyshchev E.G.**, Genaev M.A., Afonnikov D.A. Evaluation of the SeedCounter, a mobile application for grain phenotyping // Frontiers in plant science. – 2017. – V. 7. – P. 1990.
3. Генаев М.А., **Комышев Е.Г.**, Хао Фу, Коваль В.С., Гончаров Н.П., Афонников Д.А. SpikeDroidDB – информационная система для аннотации морфометрических характеристик колоса пшеницы // Вавиловский журнал генетики и селекции. – 2018. – Т. 22. – № 1. – С. 132-140.
4. Genaev M.A., **Komyshchev E.G.**, Smirnov N.V., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A. Morphometry of the Wheat Spike by Analyzing 2D Images // Agronomy. – 2019. – V. 9. – № 7. – P. 390.
5. Пронозин А. Ю. А.А. Паулиш, Е.А. Заварзин, А.Ю. Приходько, Н.М. Прохошин, Ю.В. Кручинина, Н.П. Гончаров, **Е.Г. Комышев**, М.А. Генаев Автоматическое фенотипирование морфологии колоса тетра-и гексаплоидных видов пшеницы методами компьютерного зрения // Вавиловский журнал генетики и селекции. – 2021. – Т. 25. – №. 1. – С. 71-81.