

О Т З Ы В

на диссертацию Афонникова Дмитрия Аркадьевича на тему «Компьютерные методы высокопроизводительного фенотипирования растений», представленную на соискание ученой степени доктора биологических наук по специальности

1.5.8. – математическая биология, биоинформатика.

Диссертация Афонникова Д.А. посвящена разработке компьютерных методов фенотипирования растений на основе анализа их цифровых изображений. Целью диссертационной работы являлось создание экспериментально-компьютерной платформы для высокопроизводительного фенотипирования 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.

Разработка высокопроизводительных методов фенотипирования растений является в настоящее время актуальной для селекционеров и генетиков, поскольку позволяет на основе статистического анализа больших выборок растений выявлять ассоциации между изменениями в геноме и фенотипом растений. В результате могут быть обнаружены новые гены, контролирующие важные признаки, связанные с урожайностью, устойчивостью к биотическим и абиотическим стрессам е растений.

Методы фенотипирования также актуальны для решения задач в области физиологии растений, они позволяют установить особенности ответа растений на стресс, определить физиологические системы растений (органы, ткани), которые определяют сопротивляемость растений абиотическому стрессу.

До недавнего времени, методы фенотипирования основывались на трудоемких протоколах, требующих непосредственного участия квалифицированного эксперта. Эти методы во многих случаях обладали низкой производительностью и давали субъективные оценки. Создание автоматических компьютеризированных подходов на основе анализа цифровых изображений является в силу указанных причин актуальной задачей.

Диссертационная работа Афонникова Д.А. представляет собой комплексное научное исследование, направленное на разработку методов фенотипирования (в рамках создания платформы ICGPhenoPlant), а также на применение этих методов для решения ряда актуальных биологических задач. Работа была проведена на нескольких видах растений: пшенице, картофеле, табаке. Исследовались разнообразные фенотипические характеристики растений: опушение листа, размер, форма и цвет оболочки зерен, размер и форма колоса. При всем разнообразии объектов исследования и фенотипических признаков растений, данная работа характеризуется единством подхода к разработке и применению цифровых методов фенотипирования: вначале разрабатывается протокол фенотипирования, далее – методы и алгоритмы обработки изображений, производится оценка их

точности, затем данные методы используются при проведении массового анализа фенотипов растений для разных сортов и линий растений, в разных физиологических условиях.

Для изучения ряда признаков в процессе работы требовалось создать совершенно новые, уникальные подходы к фенотипированию. Это, в свою очередь, позволило получить ряд новых важных биологических результатов. Так, метод фенотипирования опушения листа пшеницы позволяет с высокой точностью оценивать не только количество трихомов на поверхности листа, но также размер трихомов. Его использование позволило на новом уровне детализации впервые количественно охарактеризовать опушение листьев у представителей пшениц разных видов. Было исследовано проявление трех известных генов, контролирующих опушение у пшеницы при анализе различных сортов и линий пшеницы и показано, что два гена (*HL1*, *HL3*) влияют как на процессы инициации трихомов (их плотность на листе), так и на процессы, контролирующие их размер (среднюю длину). Один из генов (*HL2^{acsp}*) в большей степени влияет на размер трихомов.

В работе также продемонстрировано, что методы фенотипирования опушения листа могут быть эффективно использованы для изучения взаимосвязи изменения опушения и физиологических параметров растений при засухе. Такой анализ с учетом детального опушения характеристик опушения на количественном уровне был проведен впервые.

Оригинальность подхода к анализу признаков опушения листа заключается в универсальности данной разработки: ее оказалось возможным расширить с небольшими доработками для анализа опушения и других видов, картофеля и табака. В результате также были получены новые значимые биологические результаты. Оценено разнообразие характеристик опушения листьев у ряда сортов картофеля российской селекции, показано влияние содержания пролина в растениях на интенсивность и характер опушения ряда генно-модифицированных линий табака.

Другим направлением исследований, представленным в работе, является разработка методов фенотипирования характеристик зерен пшеницы. Несмотря на то, что эта задача в мире давно исследуется и существует целый ряд программ, которые позволяют оценить количество и размер зерен, в диссертации предложен новый подход, который заключается в том, что для фенотипирования зерен предлагается использовать мобильные устройства. В результате становится возможным широкое применение компьютерных методов фенотипирования даже за пределами лабораторий. Предложенное решение оказалось востребованным: приложение SeedCounter для платформы Android, реализующее данный подход, было установлено более 1000 раз и использовалось как в учебных целях, так и для проведения научных исследований.

В дальнейшей работе функциональность метода оценки фенотипа зерен была расширена за счет использования нескольких дескрипторов формы и 48 характеристик цвета (основанных на использовании представления цвета

пикселей в различных цветовых пространствах). Применен алгоритм цветокоррекции изображений на основе использования цветовой шкалы. Все это позволило впервые детально охарактеризовать вариации количественных признаков размера, формы и цвета оболочки зерен у растений популяции ITMI, выявить локусы количественных признаков в геноме пшеницы, контролирующие эти признаки, провести поиск и приоритизацию генов, которые, возможно, участвуют в контроле указанных признаков зерен.

Как и для опушения листьев, признаки зерна были исследованы в их связи с физиологическим состоянием: длительностью хранения в генбанке и всхожестью. Впервые на большом материале (популяция ITMI, 4 года урожая) были получены количественные оценки связи между длительностью хранения зерен и свойствами зерен (размером, формой и цветом), показано, что с увеличением времени хранения краснота и светлота оболочки зерен увеличиваются. Показано, что более красные зерна имеют более низкую всхожесть.

В работе предложен новый подход к фенотипированию колосьев пшеницы на основе анализа изображений. Его новизна заключается в том, что для описания формы колоса в дополнение к широко используемым дескрипторам формы (округлость, шероховатость и др.) используется набор параметров, описывающих форму колоса в виде двух четырехугольников с общей стороной (центральная линия колоса). Показано, что эти параметры могут быть использованы для предсказания типа колоса и его плотности.

В рамках диссертации также были разработаны базы данных, для хранения информации о характеристиках фенотипа, генотипа, окружающей среды, характеристиках колоса. При создании структуры этих баз данных также были использованы новые оригинальные подходы.

Следует отметить, что все разработанные в рамках диссертационной работы методы анализа изображений (программа LHDetect2 для анализа опушения, приложение SeedCounter для анализа зерен, программа WERecognizer) свободно доступны в интернет. Это позволяет использовать их в научных исследованиях и образовании широкому кругу пользователей и свидетельствует о высокой практической ценности полученных в диссертации результатов.

В рамках разработки методов фенотипирования (оценка фенотипических характеристик на основе анализа двумерных изображений) использовался единый методический подход, в котором при разработке метода фенотипирования обязательно уделялось внимание оценке его точности. Такая работа была проведена для всех трех алгоритмов (оценка характеристик опушения, размеров зерен и характеристик колоса). Результаты показали высокую точность предлагаемых подходов на независимых тестовых выборках. Это свидетельствует о высокой достоверности полученных оценок. Косвенным свидетельством этого также могут служить ряд результатов, которые хорошо согласуются с известными данными об изменении фенотипических характеристик под воздействием факторов среды

(повышение плотности опушения листа при воздействии засухи, более низкая всхожесть у зерен с красным оттенком оболочки).

При анализе фенотипических данных для получения достоверных результатов были использованы ряд стандартных подходов (тесты на основе дисперсионного анализа одно- и многофакторного), а также рандомизационные тесты. Использование этих критериев позволяет судить о высокой степени достоверности полученных биологических результатов. Таким образом, все полученные выводы (как методические, так и биологические) являются обоснованными.

Результаты комплексной диссертационной работы с разнообразными биологическими объектами и методами изложены на хорошем уровне, понятным языком с достаточной степенью детализации. Текст 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.

Введение представляет актуальность, цели и задачи работы, положения, выносимые на защиту и прочие формальные характеристики диссертационной работы.

Глава 1 посвящена обзору задач и методов феномики растений с фокусом на методы анализа цифровых изображений. Описаны основные методы и алгоритмы, которые используются при решении этих задач. Даны определения основных метрик для оценки точности алгоритмов. Описано применение этих методов к решению задач массового высокопроизводительного фенотипирования сельскохозяйственных растений.

Глава 2 посвящена решению задач в области фенотипирования опушения листьев растений (пшеницы, картофеля и табака). Вначале дается описание структуры и биологической роли опушения листьев у растений и мягкой пшеницы, в частности. Описан протокол для получения изображений сгибов листа пшеницы для дальнейшего анализа, алгоритмы обработки полученных изображений, реализованные в программе LHDetect2. Описаны результаты оценки точности этих алгоритмов при подсчете трихомов на сгибе листа и определение длины трихомов. Далее описаны результаты анализа характеристик опушения у 47 образцов пшениц от 12 видов родов *Triticum* и *Aegilops*. Затем приводятся результаты анализа фенотипического проявления трех генов (*H11*, *H12^{aesp}*, *H13*), контролирующих опушение листа у пшеницы на основе высокопроизводительного фенотипирования замещенных и почти изогенных линий мягкой пшеницы. Далее приводятся результаты анализа влияния водodefицита на количественные характеристики опушения и количество устьиц на обеих сторонах листьев, физиологические характеристики растений замещенных и почти изогенных линий мягкой пшеницы.

В заключительных разделах второй главы приводятся результаты применения метода фенотипирования опушения листа к анализу картофеля. Описана модификация алгоритма подсчета трихомов на изображении для анализа опушения у табака (трихомы с головками) и его применение к изучению генно-модифицированных линий с повышенным содержанием пролина.

Глава 3 посвящена решению задач в области фенотипирования зерен пшеницы. Описан протокол получения изображения зерен с помощью мобильных устройств, алгоритм анализа этих изображений и оценка его точности. Проведена оценка факторов влияющих на точность определения характеристик зерен (освещение, тип устройства). Представлен протокол для фенотипирования зерен в лабораторных условиях с использованием цветовой шкалы в качестве масштабной линейки и набора стандартных цветов. Далее представлены результаты анализа признаков зерен у растений популяции ITMI (выявление и анализ QTL, приоритизация выявленных генов, оценка связи признаков зерен с длительностью хранения в генбанке и всхожестью).

Глава 4 посвящена решению задач в области фенотипирования колосьев пшеницы. Описан протокол фенотипирования колоса, метод анализа изображений, даны оценки его точности. Описана геометрическая модель, использованная для получения количественных характеристик формы колоса. Приведены результаты разработки метода предсказания плотности колоса и его типа на основе геометрических характеристик модели.

Глава 5 посвящена разработке баз данных. Это база данных для хранения информации о фенотипе, генотипе и месте произрастания растений при проведении селекционно-генетических экспериментов у пшеницы (WheatPGE) и база данных для описания характеристик колоса (SpikeDroid DB).

В заключении систематизированы методические результаты (разработка платформы фенотипирования ICGPhenoPlant) и результаты изучения растений, полученные с ее помощью. Далее представлены Выводы по диссертационной работе.

Текст диссертации позволяет заключить, что эта работа является завершённой, содержит ряд новых научных результатов, как биоинформатических, так и биологических. Научные положения, выносимые на защиту, выводы, сформулированные в диссертации полностью обоснованы, их достоверность и новизна не вызывает сомнений.

Результаты работы опубликованы (24 работы). Это публикации в зарубежных и российских журналах с импакт-фактором, в российских журналах, индексируемых в системе WoS без импакт-фактора, индексируемых в системе SCOPUS и три авторских свидетельства. Автореферат соответствует содержанию диссертационной работы полностью.

К работе есть ряд вопросов и замечаний.

При анализе морфометрических признаков колоса используется очень много параметров. Может, можно было бы уменьшить их число, выбрав другую систему координат, например, эллипсоидную?

При анализе характеристик зерен идентифицировано много QTL и обнаружено существенное перекрытие QTL для разных признаков зерен по положению в геноме. Что можно сказать о скоррелированности этих признаков?

При предсказании индекса плотности колоса наблюдалась значительная разница в величине ошибок на обучающей и валидационных выборках. Не может ли такая большая разница объясняться переобучением моделей?

Также нарушена нумерация выводов (после третьего вывода следует пятый).

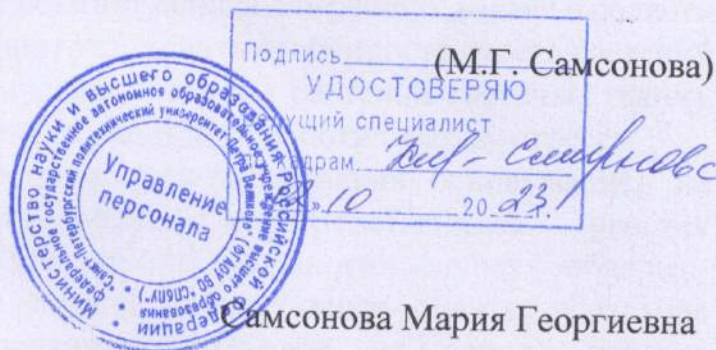
Эти замечания, однако, не влияют на общую высокую оценку работы.

В заключении следует отметить, что диссертационная работа Фонникова Дмитрия Аркадьевича является законченной научно-квалификационной работой, полностью соответствующей требованиям, предъявляемым к диссертациям на соискание ученой степени доктора наук согласно п. 9-14 «Положение о присуждении ученых степеней», утверждённого Постановлением Правительства РФ от 24.09.2013 № 842, а её автор, Фонников Дмитрий Аркадьевич, безусловно достоин присуждения искомой ученой степени доктора биологических наук по специальности 1.5.8. – математическая биология, биоинформатика.

М.С.

Дата: 02 октября 2023 г.

Официальный оппонент



Самсонова Мария Георгиевна

д.б.н., профессор Высшей школы
прикладной математики и
вычислительной физики, заведующий
Научно-исследовательской лабораторией
математической биологии и
биоинформатики ФГАОУ ВО "Санкт-
Петербургский политехнический
университет Петр Великого"

г. Санкт-Петербург

E-mail: m.g.samsonova@gmail.com;

тел. +7 812 290-9645