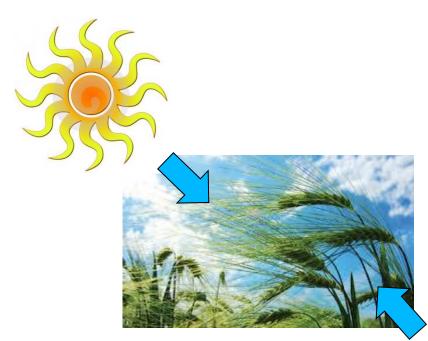
КОМПЬЮТЕРНЫЕ МЕТОДЫ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО ФЕНОТИПИРОВАНИЯ РАСТЕНИЙ

Д. А. Афонников, к.б.н., Институт цитологии и генетики СО РАН г. Новосибирск

> Научный консультант: академик, проф., д.б.н. Колчанов Н.А.

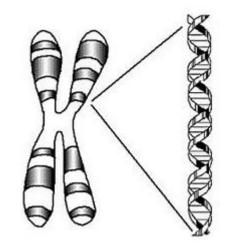


Проблема взаимоотношения генотип-фенотип-окружающая среда в биологии



Необходим учет факторов внешней среды и ответа растений на ее воздействие: известно, что до 50% фенотипических вариаций растений определяются средой.

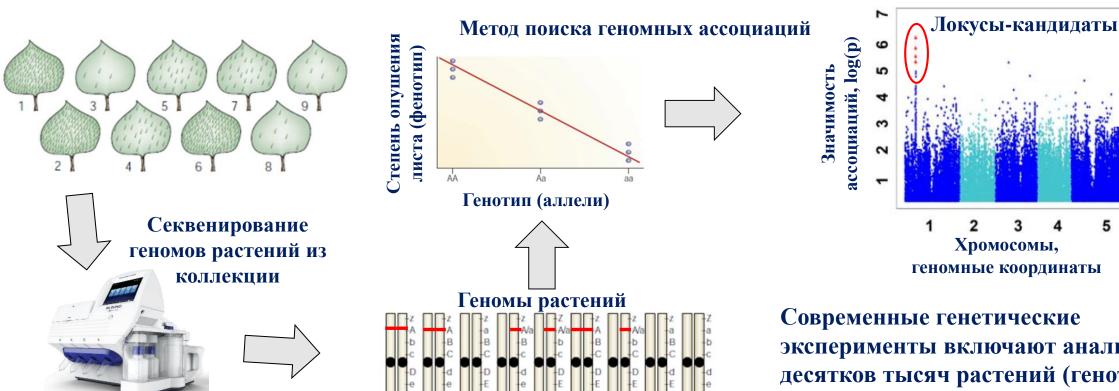
Ключевая проблема биологии — как генотип организма определяет его фенотипические свойства. Для сельскохозяйственных растений важны продуктивность, устойчивость к абиотическим и биотическим факторам среды.





Методы современной генетики: поиск геномных ассоциаций с изменениями фенотипических признаков

Современные методы поиска новых генов основаны на анализе ассоциаций между генетическими вариациями и изменчивостью признаков. Они позволяют определить участки генома, в которых эта связь наиболее значима. Для увеличения точности этих методов необходима обработка большого количества данных.

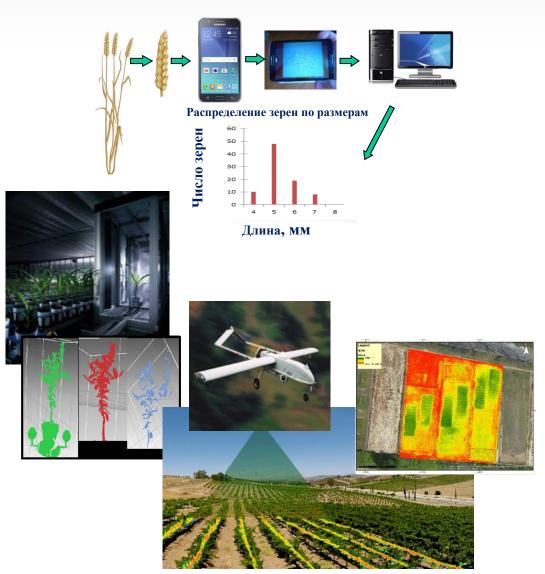


Номер растения

эксперименты включают анализ десятков тысяч растений (генотипов и фенотипов)



Компьютерные методы фенотипирования растений



Современные методы фенотипирования направлены на автоматизированную оценку характеристик фенотипа растений.

- •Основаны на анализе цифровых изображений (2D и 3D)
- •Сохраняют жизнеспособность растений
- •Обеспечивают цифровое (количественное) представление признаков растений
- •Предоставляют для анализа новые признаки
- •Интегрированы с базами данных

При разработке методов фенотипирования необходимо:

- Учесть специфику биологической задачи
- Разработать протокол фенотипирования
- Разработать метод обработки изображений для оценки параметров растения
- Оценить точность метода
- Провести апробацию на реальных задачах анализа больших биологических данных

4





<u>Целью работы</u> 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 исследования



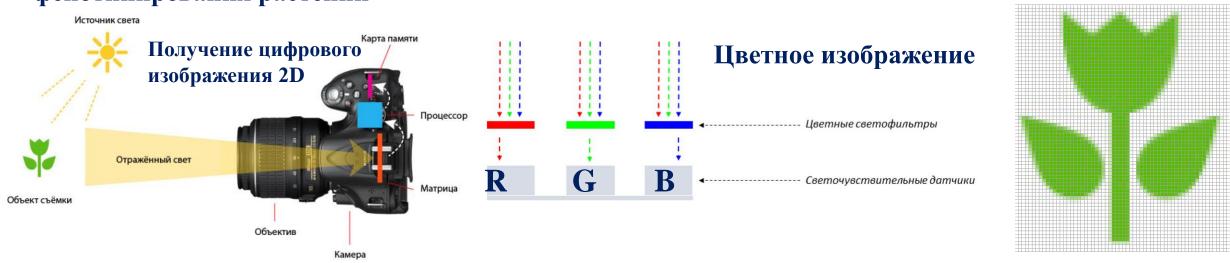
Задачи, поставленные для достижения указанной цели:

- 1. Разработка комплекса экспериментально-компьютерных 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. Применение методов фенотипирования опушения листа для двудольных растений (картофеля и табака) и анализ количественных характеристик опушения у российских сортов картофеля и у линий табака с модификацией гена пролиндегидрогеназы.
- 3. Разработка комплекса экспериментально-компьютерных 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. Разработка комплекса экспериментально-компьютерных методов фенотипирования колосьев пшеницы для оценки количественных характеристик формы, размера, остистости и определение на основе морфологических параметров колоса его типа и плотности.
- 5. Разработка баз данных для сбора, хранения фенотипических данным пшеницы и доступа к ним.



Методы анализа изображений для фенотипирования растений

Двумерные цифровые изображения удовлетворяют большинству условий неинвазивного фенотипирования растений



Основные задачи анализа изображений:

- •Сегментация
- •Выделение контуров (бинаризация)
- •Представление объектов
- •Морфометрия
- •Анализ цвета
- •Классификация объектов
- •Регрессионный анализ

Методы анализа:

- •Методы анализа 2D изображений
- •Методы статистического анализа
- •Методы многомерного статистического анализа
- •Методы машинного обучения
- •Классификация
- •Регрессия



J(A,B)=-

Сегментация изображений

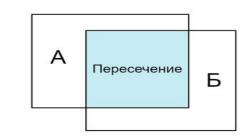
Оценка точности разрабатываемых методов

Подсчет объектов, регрессия

Классификация

Коэффициент Жаккара

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



Средняя абсолютная ошибка

MAE =
$$\frac{1}{N} \sum_{i=1}^{i=N} \frac{|n_i - n'_i|}{n_i}$$

Средняя абсолютная процентная ошибка

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

Коэффициент корреляции Пирсона, *r*

Матрица ошибок

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

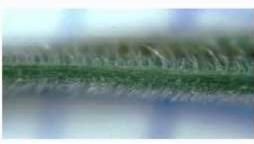
$$pr = \frac{tp}{tp+fp}$$
 $re = \frac{tp}{tp+fn}$

$$F = 2 \frac{pr \cdot re}{pr + re}$$



Опушение листа – важный фенотипический признак растения









Опушение листьев формируется совокупностью эпидермальных образований, которые называются трихомами. На начало работы у мягкой пшеницы были известны только два локуса, связанных с контролем опушения.

Это ген Hl1, локализованный на хромосоме 4В (Майстренко, 1976; Pshenichnikova et al., 2007; Dobrovolskaya et al., 2007) и ген Hl2, локализованный на хромосоме 7В (Taketa et al., 2002; Dobrovolskaya et al., 2007). Предположительно, существует еще один ген на 7D хромосоме (Дорошков и соавт., 2014).

Изучение этого признака было затруднено из-за отсутствия эффективных методов фенотипирования.



Протокол пробоподготовки для анализа опушения листа пшеницы

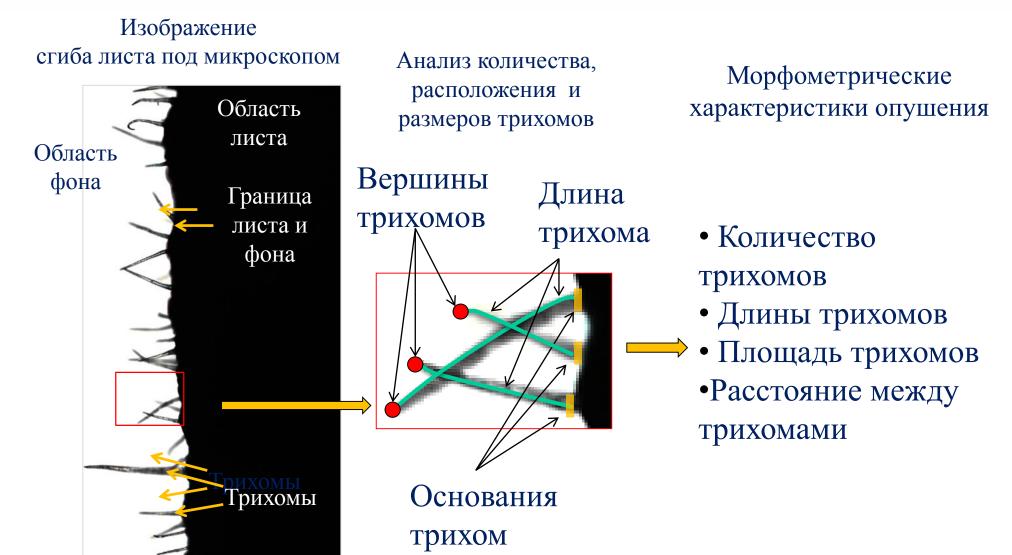


Опушение может быть проанализировано для разных листьев, в разных участках листа, на обеих сторонах. За час удаётся приготовить препараты сгибов листа и получить изображения сгибов для 20 растений (из расчета 1 растение = 3 сгиба = 6 снимков).



Подсчет и морфометрия трихомов на изображении

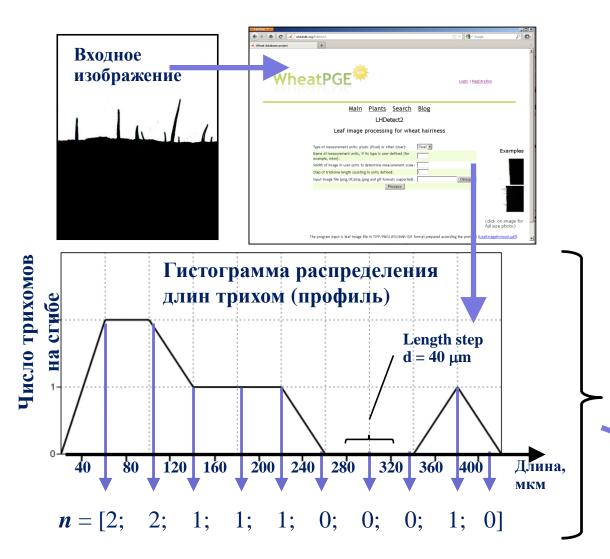
Для идентификации трихомов необходимо определить контур трихома, вершину, основание.





Приложение SeedCounter: подсчет зерен, оценка их размера, формы и цвета

Paзpaбoтaнa программа LHDetect2 для автоматической обработки изображений сгибов листа



Ошибки подсчета трихомов на изображении (идентификация вершин): F=0.85, r=0.986 (автоматическая оценка / ручная оценка).

Точность оценки длин трихомов (автоматическая оценка / ручная оценка): r=0.999

N — число трихом на изображении характеризует плотность опушения.

$$N = \sum_{i=1}^{k} n_i$$

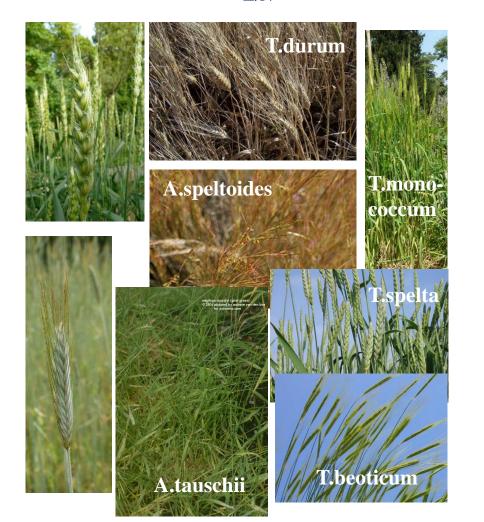
Средняя длина трихом L, характеризует размер трихомов

$$< L> = (\sum_{i=1}^{k} n_i d_i) / k$$



Разнообразие опушения листьев у мягкой пшеницы и ее сородичей

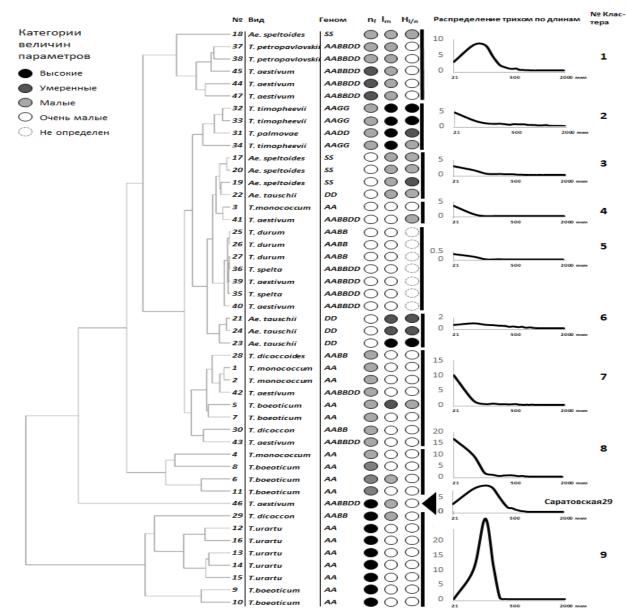
Проанализированы 47 образцов пшеницы разных видов. Количественные параметры: средняя длина трихомов (L, мкм), число трихомов на изображении (N, шт.), отношение длины трихом к их количеству ($H_{L/N}$), анализ гистограмм распределений длин трихом



№	Геномная конституция	Вид	Число образцов
1	AA	Т. топососсит.	4
2	AA	T.boeoticum	7
3	AA	T.urartu	5
4	SS(BB,GG)	Ae. speltoides Tausch.	4
5	DD	Ae. Tauschii Coss.	4
6	AABB	T. durum Desf.	3
7	AABB	T. dicoccoides	1
8	AABB	T. dicoccon	2
9	AAGG	T. palmovae	1
10	AAGG	T. timopheevii	3
11	AABBDD	T. spelta L.	2
12	AABBDD	T. petropavlovskii	2
13	AABBDD	T. aestivum L.	9



Кластеризация пшениц по типам опушения



Выделены 9 кластеров, имеющих специфический тип опушения у мягких пшениц и ее сородичей.

28 из 47 образцов имеют слабое и редкое опушение (кластеры С2-С8).

С2 - тетраплоиды

С5 – без опушения

С3, С4, С6, С7 – редкие трихомы

С1 – интенсивное опушение

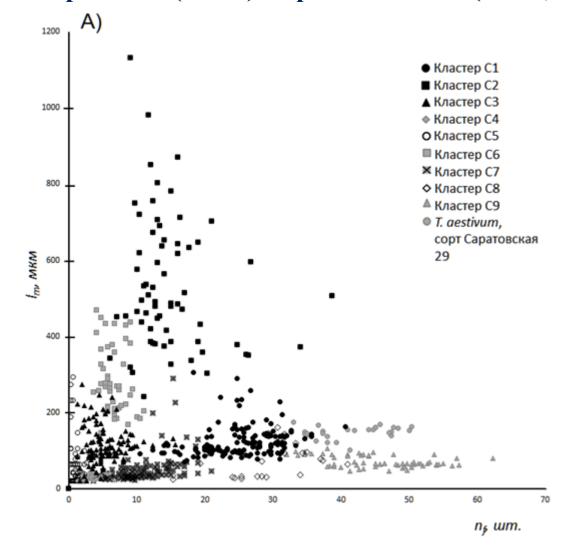
С9 – короткие трихомы (войлочное опушение)

Образец 46- Т. aestivum не относится ни какому из кластеров и формирует специфическое плотное опушение

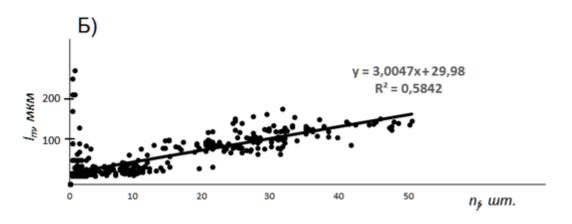


Распределение образцов по числу и количеству трихомов

Выявленные кластеры образцов хорошо различаются на диаграмме для параметров числа трихомов (ось X) и средней длины (ось Y).



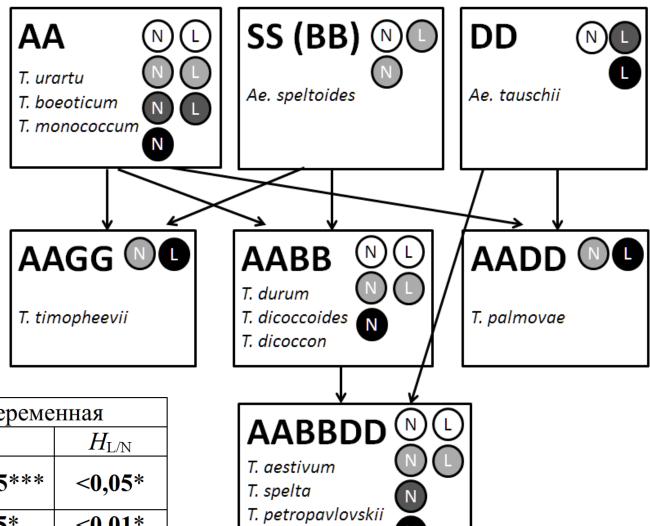
Для гексаплоидных пшениц (T. aestivum, T. petropavlovskii и T. spelta) наблюдается значимая линейная связь между числом трихомов и их длиной r = 0.76 (р << 0.0001). Если исключить из выборки гексаплоидов слабо опушенную T. spelta, коэффициент корреляции увеличивается до r = 0.92; (р << 0.0001).





Связь характеристик опушения и хромосомного состава у пшениц

Хромосомный состав оказывает влияние на все три характеристики опушения. Плоидность значимо связана с длиной трихомов. Плотность опушения значимо зависит от типа генома для диплоидных видов. Длина опушения и индекс опушения $H_{L/N}$ зависит от наличия/отсутствия D-генома.



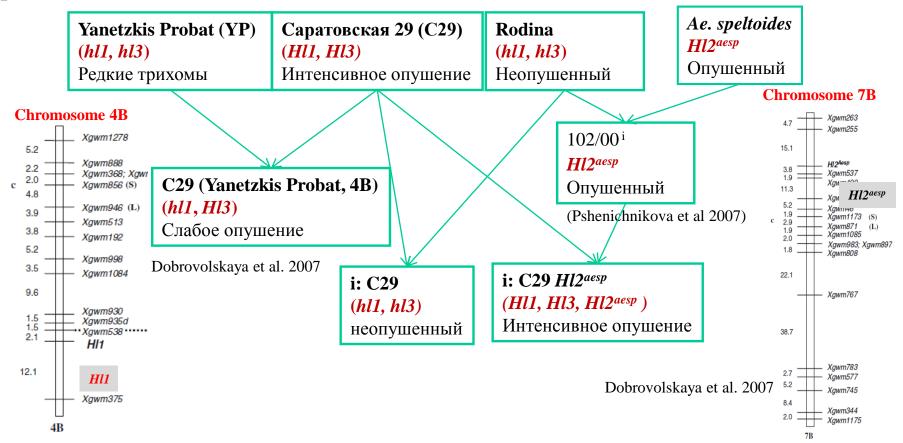
Дисперсионный анализ (р-величины):

Церориение <i>д</i> поромение <i>д</i>	Зависимая переменная			
Независимая переменная	n_{f}	$l_{ m m}$	$H_{ m L/N}$	
Структура генома	<0,005**	<<0,005***	<0,05*	
Уровень плоидности	0,1263	<0,05*	<0,01*	
А геном	<0,05*	0,2982	0,9584	
В (G,S) геном	<0,01*	0,5762	0,05631	
D геном	<0,05*	<0,005**	<0,005**	



Изогенные линии пшеницы для анализа функций генов контроля опушения листа

Генотипы пшеницы-реципиента и донора - носители различных аллелей опушения листьев и их фенотипов. Различаются аллельными состояниями генов *Hl*1, *Hl*2^{aesp}, *Hl*3.

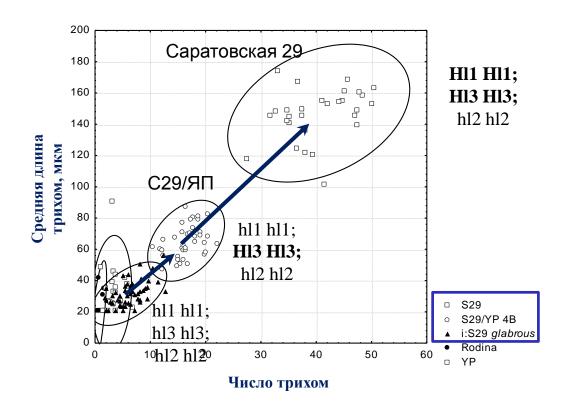


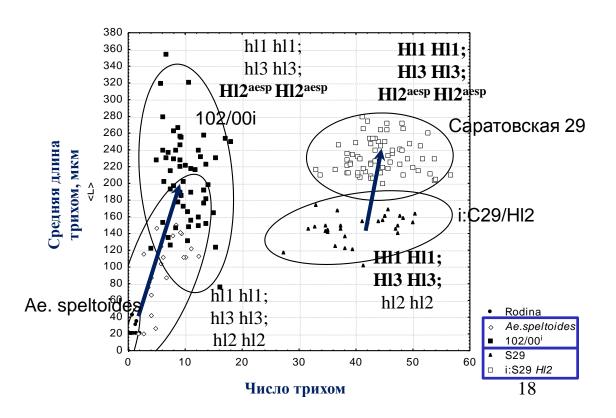
Данный биологический материал уникален для изучения функций генов контроля опушения. Он был предоставлен Пшеничниковой Т.А.



Влияние генов HI1, HI2^{aesp} и HI3 на число и длину трихом

Ген *Hl1* влияет как на увеличение длины трихом, так и на увеличение их количества. (Расположен на 4В хромосоме). Его проявление сходно с проявлением гена *Hl3*. Вероятно, они контролируют пути как инициации трихомов на листе, так и роста клеток отдельных трихомов. Ген Hl2^{aesp} в большей степени оказывает влияние на длину трихом. Возможно его функция связана с контролем роста клеток трихомов.

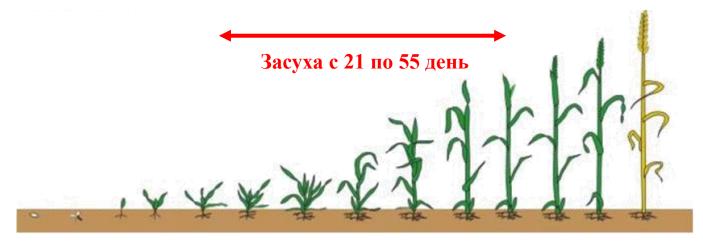






Влияние засухи на растения пшеницы с различными типами опушения

Эксперимент по исследованию влияния засухи на растения пшеницы (совместно с СИФИБР)



1 группа признаков

- 1. Скорость транспирации (Е)
- 2. Устьичная проводимость (G_{H2O})
- 3. нетто-фотосинтез (А)
- **4. WUE**

2 группа признаков

- 1. Потенциальная эффективность ФС II (Fv/Fm)
- 2. Реальная эффективность ФС II (Yield)
- 3. Скорость транспорта электронов (ETR)
- 4. Нефотохимическое тушение флуоресценции (NPQ)

3 группа признаков

Опушение листа

- А) длина трихом TL
- Б) число трихом TN

Длина листа (L)

Ширина листа (W)

Биомасса (Вт)

Количество устьиц на единицу площади SN



Влияние засухи на морфологию опушения

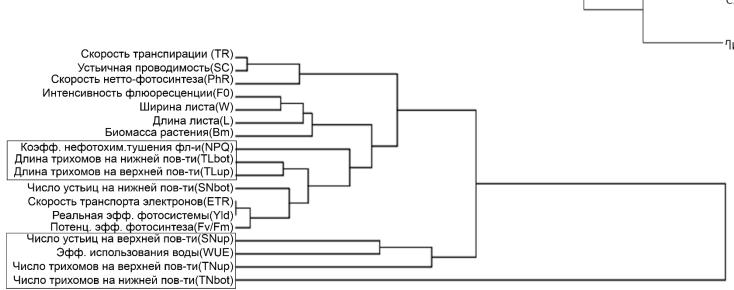
При воздействии засухи плотность опушения верхней стороны листа у опушённых генотипов, за исключением Янецкис Пробат, имеет тенденцию увеличиваться, а длина опушения снижаться. Аналогичная зависимость на нижней стороне зависимость менее выражена.

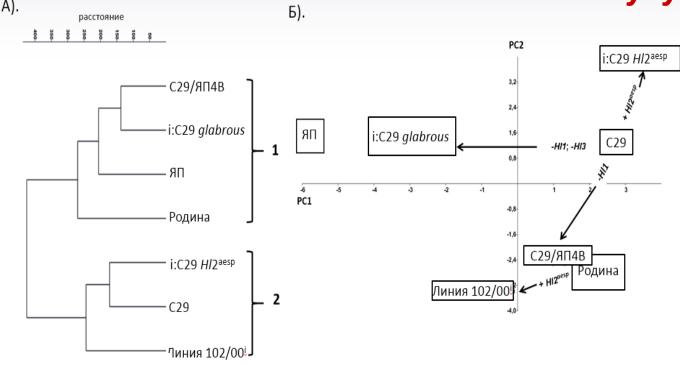
	F	Влияние засухи		
	Генотип	Верхняя сторона	Нижняя сторона	
Родина	hl1, hl2, hl3			
Янецкис Пробат	hl1, hl3	N↑L↑	N ↑	
Саратовская 29	HL1, HL3	N↑L↓	N↑L↓	
Линия 102/00і	HL2 ^{aesp}	N↑L↓	N ↑ L ↓	
С29 ЯП(4В)	hl1, HL3	N ↑ L ↓?	N ↑	
C29 HI2	HL1,HL2,HL3	N ↑ L ↓?	N ↑ L ↓	
С29 Б/о	hl1, hl3	N ↑ -		



Связь признаков опушения с физиологическими параметрами в ответе на засуху

Кластеризация признаков на основе изменения величин под воздействием засухи: длина и число трихомов листа изменяются в ответ на водный стресс согласованно с фотосинтетическими показателями и эффективностью использования воды.





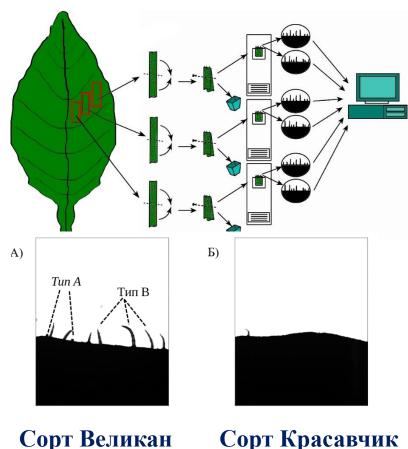
Кластеризация генотипов по их ответу на вододефицит показывает, что опушенные и неопушенные растения ведут себя по-разному. Это демонстрирует важную роль опушения в ответе на засуху для растений пшеницы.



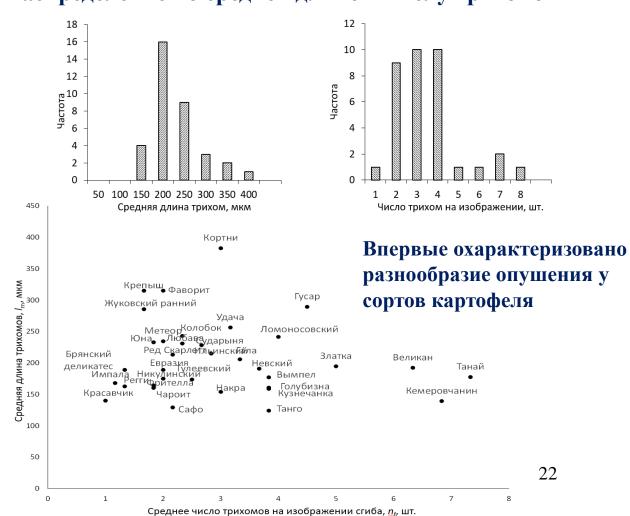
Анализ опушения листьев картофеля

С помощью программы LHDetect2 проведен анализ количественных характеристик опушения у 35 сортов картофеля Российского происхождения. Анализировались трихомы типа В (несекретирующие). Распределение по средней длине и числу трихомов

Протокол пробоподготовки



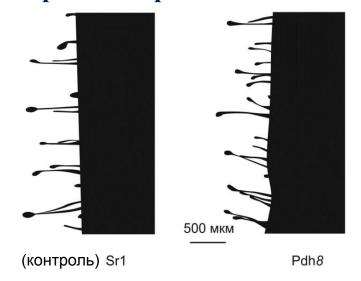
Сорт Красавчик



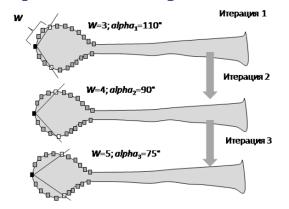


Анализ опушения листьев табака для генотипов с измененной функцией гена пролиндегидрогеназы

Проведен анализ опушения листьев табака для контрольной линии SR1 и ее генетически модифицированных вариантов с нарушенной функции пролиндегидрогеназы (PDH) -> повышенное содержание пролина.



Модификация алгоритма LHDetect2



Точность модифицированного алгоритма (в скобках для прежней версии): MAE=0.78 (4.57); MAPE=7 (52); F=0.90 (0.37).

Результат анализа: Линии с повышенным содержанием пролина демонстрируют повышенное количество трихомов и их меньшую среднюю длину.

23



Фенотипическое разнообразие зерен



[Garg et al., J Cereal Sci (2016), 71, 138-144]

Зерна и их морфологические характеристики

Основные характеристики размера/формы:

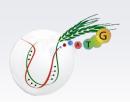
- •количество в колосе
- •объем
- •длина
- •ширина
- •округлость

Связаны с массой зерна, урожайностью, мукомольными свойствами (мелкие округлые зерна дают муку высокого качества).

Основные характеристики цвета оболочки:

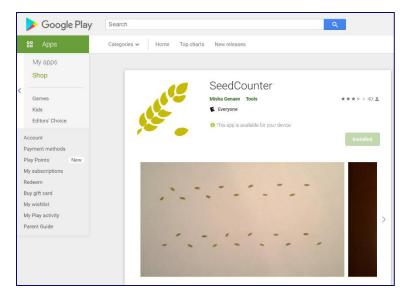
- •краснота
- •светлота

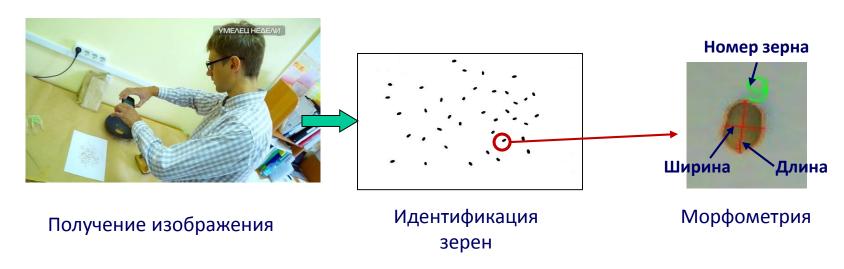
Связаны с пигментным/метаболическим составом (содержанием антоцианов, каротиноидов, флавоноидов). Важны для товарной классификации, характеризуют физиологическое состояние зерна. Связаны с устойчивостью к предуборочному прорастанию/всхожестью (более темные зерна имеют низкую всхожесть).



Приложение SeedCounter: подсчет зерен, оценка их размера, формы и цвета

Предназначено для подсчета и морфометрии зерен злаков. Программа оценивает длину, ширину, площадь, проецируемую на поверхность листа, и другие параметры. Программа получает изображение, идентифицирует белый лист заданного формата (А4 и т.д.), определяет масштаб, определяет контуры зерен на изображении и определяет их размеры в мм, передает результаты на сервер для загрузки в формате ХМL. Есть версия для ПК (реализован анализ цвета).





https://play.google.com/store/apps/details?id=org.wheatdb.seedcounter

Приложение было загружено более 1000 раз пользователями разных стран. Оно было использовано для проведения селекционно-генетических экспериментов в России и за рубежом.



Протокол для съемки цифровой камерой

Съемка зерен на фотокамеру Canon EOS 600D с объективом Canon EF 100mm f/2.8 Macro USM. Зерна располагались на белом фоне. Для контроля цветопередачи и цветокоррекции, определения масштаба использовали цветовую шкалу ColorChecker Classic.

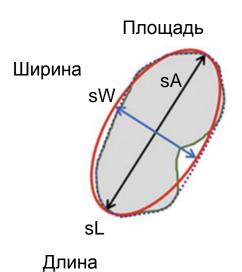






Характеристики размера и формы зерен

Форма и размер проекции зерен на изображении описывались 7 параметрами:



Area: area of ROI (pixels, mm2, um2)

Perimeter (green): length of boundary of ROI

Ellipse (red): fitted ellipse to ROI

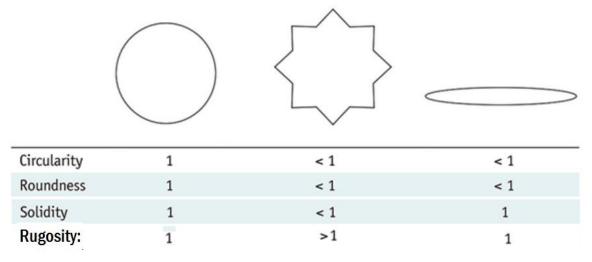
Major axis (black arrow): longest diameter of ellipse

Minor axis (blue arrow): shortest diameter of ellipse

Convex area (dotted boundary)

$$SCi=$$
 Circularity: $\frac{4\pi \times area}{Perimeter^2}$ Округлость

$$SRo=Roundness: \frac{4 \times area}{\pi \times (major axis)^2}$$
 Закругленность



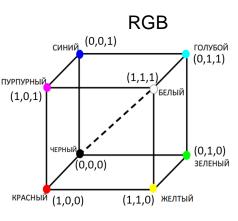
Характеристики цвета оболочки зерен

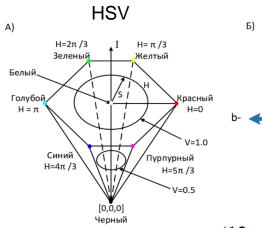
Использовались компоненты цвета в 4 цветовых пространствах. Всего использовались 48 характеристик цвета (12 для средних значений и 36 для доминантных кластеров).

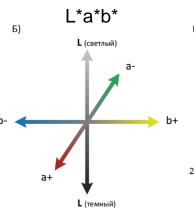
Цветовые пространства:

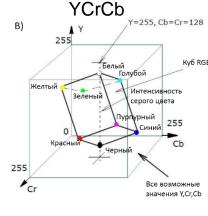
Цвет оболочки зерен неравномерный











Средние значения компонент (12 шт.)

RGB mR - красный

RGB_mG - зеленый

RGB mB - синий

HSV mH - оттенок

HSV_mS - насыщенность

HSV mV - яркость

Lab_ma - краснота

Lab mb – желтизна

Lab_mL - светлота

YCrCb_Cr - краснота YCrCb_Cb - синева YCrCb Y - яркость

Доминантные кластеры (36 шт.)

RGB_dCR_1	HSV_dCH_1
RGB_dCG_1	HSV_dCS_1
RGB_dCB_1	HSV_dCV_1

RGB dCR 2

Lab_dCa_1 Lab dCb 1

Lab_dCL_1

YCrCb_dCY_1

YCrCb_dCCr_1

YCrCb_dCCb_1

28



Биологический материал: популяция ІТМІ

Проводился анализ морфометрических и цветовых характеристик зерен пшеницы для рекомбинантных инбредных линий (РИЛ) картирующей популяции ITMI (International Triticeae Mapping Initiative): Synthetic W7984 \times Opata 85. Зерна хранились при температуре $-18 \pm 2 \circ C$ и влажности $8 \pm 2\%$ в генбанке.

Все растения выращивались в Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany.



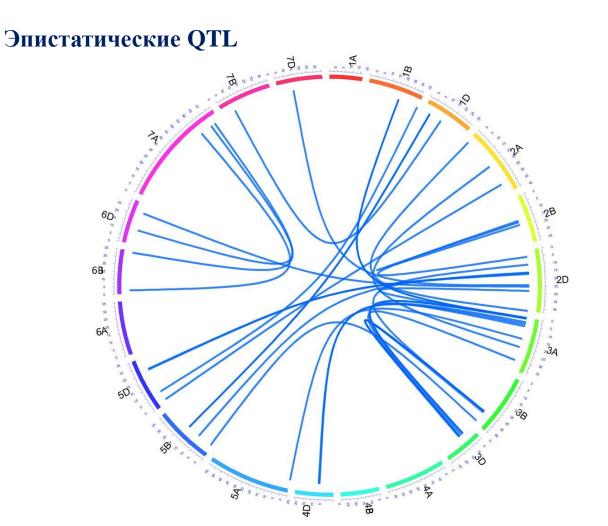
От каждого растения для фенотипирования было взято 20 зерен. Фенотипирование проводилось в марте 2020 г.



QTL для признаков зерен

Обнаружено 212 аддитивных QTL: 20 для признаков размера, 22 для признаков формы и 170 для признаков цвета. Обнаружено 34 пары QTL между которыми выявлены значимые эпистатические взаимодействия.

Аддитивные QTL [Значимость





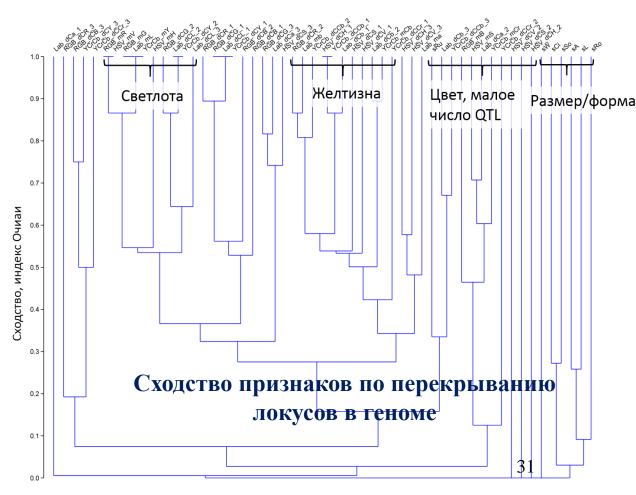
Анализ перекрывания расположения QTL в геноме для разных признаков зерен

Обнаружено существенное перекрывание QTL для разных признаков по положению в геноме (до 12 признаков в одном локусе). Был проведен кластерный анализ признаков по степени перекрывания их QTL в геноме на основе индекса Очиаи. Результаты демонстрируют вырожденность характеристик: многие из них контролируются одними и теми же локусами.

Перекрывание положения QTL в геноме:

TraitName	Chromosom	QTL designation	Position	LeftMarke RightMark Lo	OD
RGB_dCB_3	1B	Q.RGB_dCB_3-1B	126	Kukri_c391JD_c954_2	2.2745
sA	1B	Q.sA-1B	160	AX-945037 BobWhite	13.6347
sA	1D	Q.sA-1D	146	wsnp_Ex_wsnp_Ex_	7.9571
sW	1D	Q.sW-1D	184	IAAV678 RAC875_r	2.0186
RGB_mB	1D	Q.RGB_mB-1Da	287	RAC875_c AX-946066	2.7262
HSV_mH	1D	Q.HSV_mH-1Da	287	RAC875_c AX-946066	2.3565
HSV_mS	1D	Q.HSV_mS-1Da	287	RAC875_c AX-946066	2.2824
Lab_dCa_2	1D	Q.Lab_dCa_2-1Da	287	RAC875_c AX-946066	2.2691
YCrCb_mCr	1D	Q.YCrCb_mCr-1Da	287	RAC875_c AX-946066	2.5588
RGB_dCB_3	1D	Q.RGB_dCB_3-1Db	314	wsnp_Ex_ AX-947721	2.9901
YCrCb_dCY_3	1D	Q.YCrCb_dCY_3-1Db	316	wsnp_Ex_ AX-947721	3.4369
YCrCb_dCCr_3	1D	Q.YCrCb_dCCr_3-1Db	317	wsnp_Ex_ AX-947721	4.4104

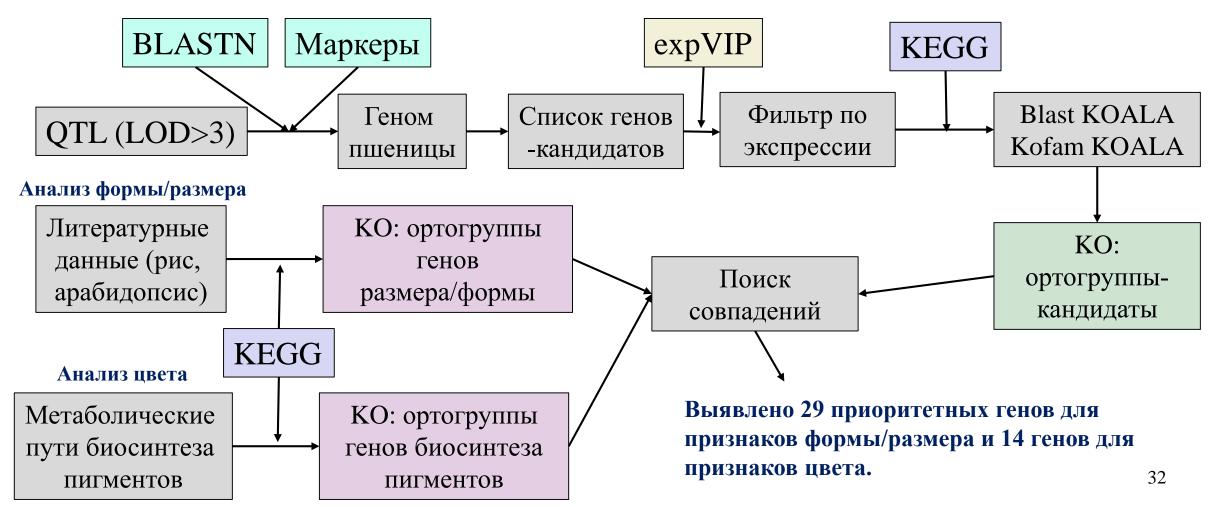
Перекрывание локусов для признаков размера/формы незначительное. Перекрывание для признаков цвета очень существенное.





Приоритизация генов из локусов, ассоциированных с признаками зерен

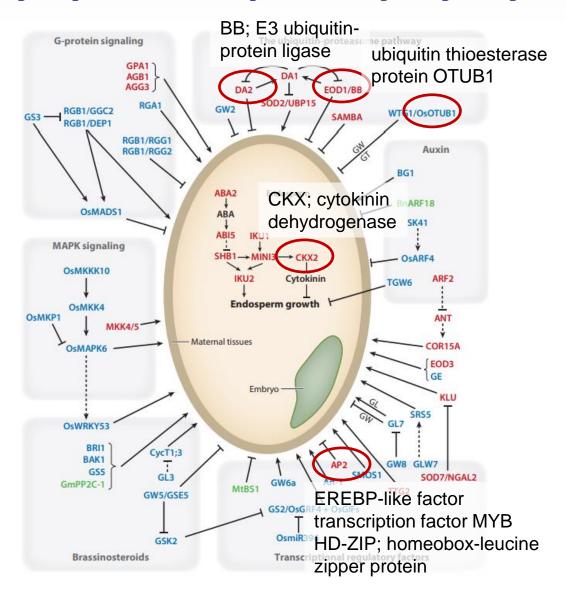
Разработан конвейер для приоритизации генов из локусов, связанных с размером/формой зерен и их цветом на основе использования БД KEGG. Использовали 48 значимых QTL с LOD>3, для которых были идентифицированы позиции маркеров в геномной последовательности пшеницы.



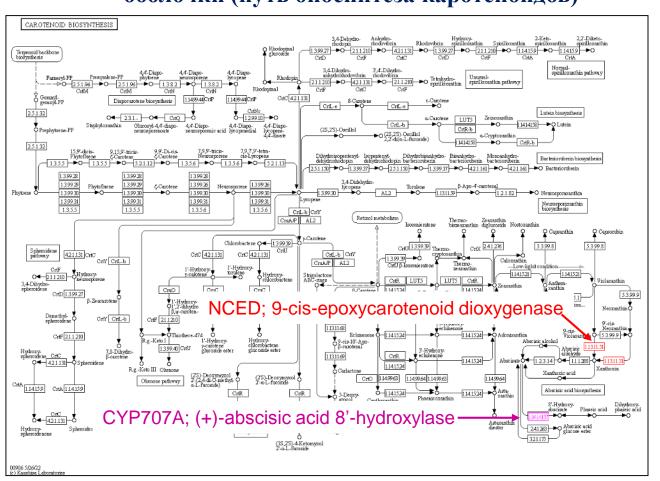


Примеры генов-кандидатов, ассоциированных с признаками зерен

Пример генов, ассоциированных с размером зерна



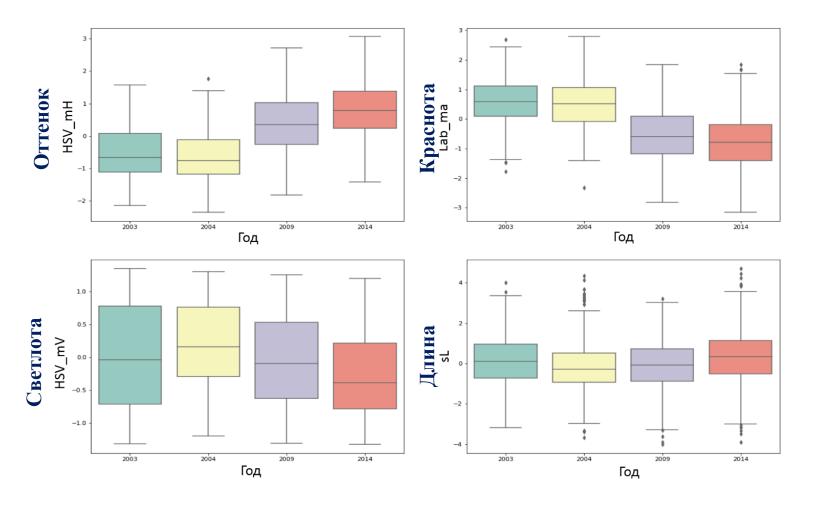
Пример генов, ассоциированных с цветом оболочки (путь биосинтеза каротеноидов)





Связь между характеристиками зерен и сроком хранения

ANOVA: Генотип оказывает значимое влияние на величину 53 признаков из 55. Год урожая оказывает влияние на 34 признака (кроме sW, sA, sRu и компонент для 2 и 3 доминантных цветов). Из значений признака для зерен мы вычитали среднее по генотипу. Искали признаки для которых наблюдается значимая линейная связь между годом урожая и значением признака. Значимость оценивалась рандомизационными тестами (бутстреп, перестановки).



Значимая корреляция выявлена для 44 признаков цвета и 1 признака формы (sRu – не значим согласно ANOVA).

Положительная связь (9 признаков цвета): HSV_mH (оттенок: r=0.51, p<0.05), YCrCb_mCb (синева: r=0.29, p<0.05) и их производные.

Отрицательная связь (35 признаков цвета): Lab (краснота: r=-0.66, p<0.05), HSV_mV (светлота: r=-0.30, p<0.05).

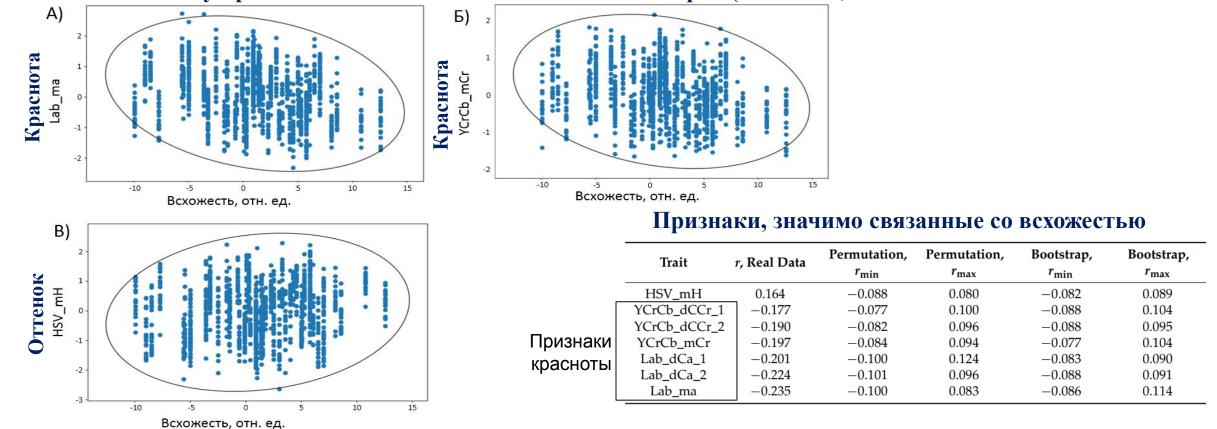
Зерна с более длительным сроком хранения более красные и более светлые (в отличии от эффекта browning при искусственном старении).



Связь между характеристиками зерен и всхожестью

Значимые корреляции со всхожестью были обнаружены только для 7 признаков цвета. Только для одного – положительная (HSV_mH), для остальных - отрицательная. Все признаки с отрицательной зависимостью так или иначе характеризуют красноту.

Зависимость между признаком и всхожестью для отдельных зерен (~1300 шт.):



Более красные зерна имеют более низкую всхожесть.



Колос –орган определяющий продуктивность пшеницы

Характеристики размера и формы колоса – одни из ключевых, связанные с продуктивностью у злаков.



Важные агрономичекие параметры колоса

- Размер
- Форма/тип
- Число зерен
- Плотность колоса
- Остистость

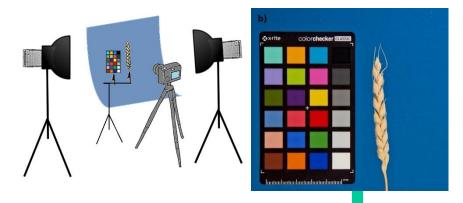
Разнообразие формы колоса у пшеницы





Морфометрия колосьев пшеницы на основе анализа изображений

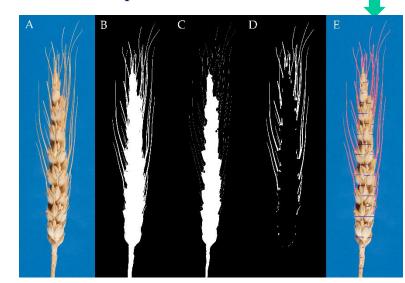
Получение изображений



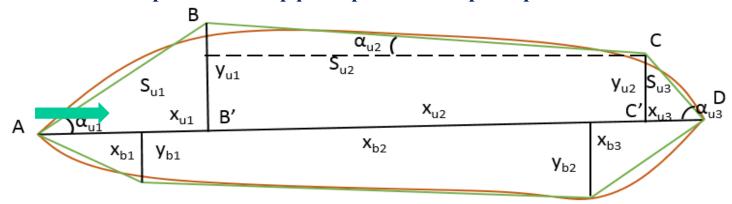
Протокол позволяет получать стандартизированные изображения для автоматического фенотипирования колосьев пшеницы. Точность выделения контура тела колоса *J* 0.925, остей – 0.660.

Форма колоса описана в виде двух четырехугольников. Их параметры позволяют классифицировать колос по типу и оценить его плотность. Кроме того, оценивается остистость колоса.

Сегментация изображения колоса на тело и ости



Определение морфометрических характеристик колоса



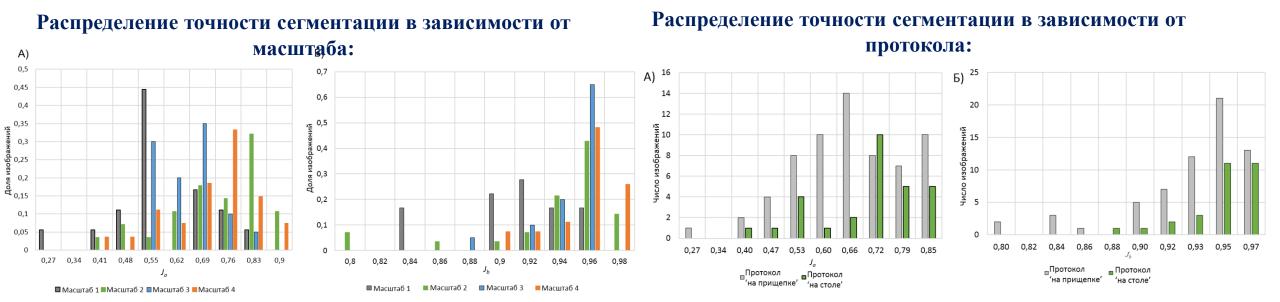
Всего 29 параметров: 9 общих (площадь, длина, остистость), 20 параметров для двуз четырехугольников.

Дополнительно использовалась модель сечений и радиальная. Genaev et al. (2019) Morphometry of the Wheat Spike by Analyzing 2D Images. Agronomy, 9(7), 390.



Оценка точности сегментации колоса и остей

Среднее значение точности распознавания тела колоса на тестовых изображениях $J_{\rm b}=0.925$ и остей $J_{\rm a}=0.660$. Точность распознавания остей оказывается ниже. Кроме того, масштаб изображения влияет на оценку точности: изображения, полученные с большим удалением камеры от колоса (масштаб 1) имеют меньшую точность распознавания колоса.



Сравнение параметров модели для колосьев трех типов

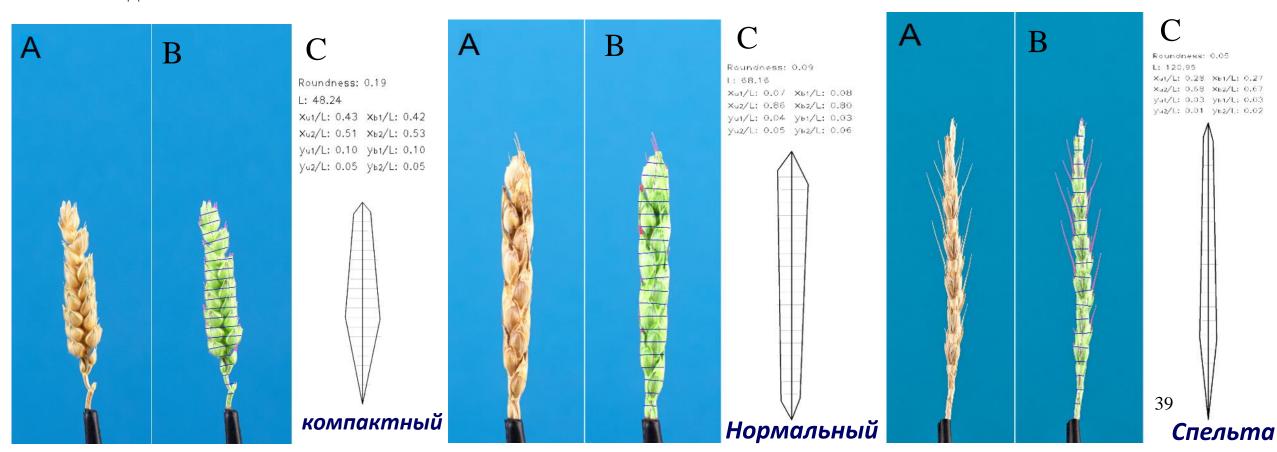
Результат представления колосьев трех различных типов (компактных, нормальных и спельтовых) в виде модели четырехугольников.

А – исходное изображение;

В – выделены тело колоса (зеленый контур) и ости (розовый контур);

С – представление колоса в модели четырехугольников.

Приведены значения округлости, длины колоса (L, мм) и отношения параметров четырехугольника к длине колоса.

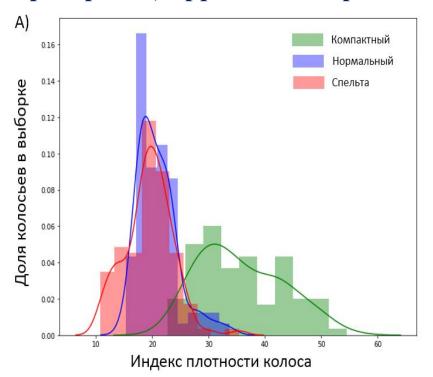


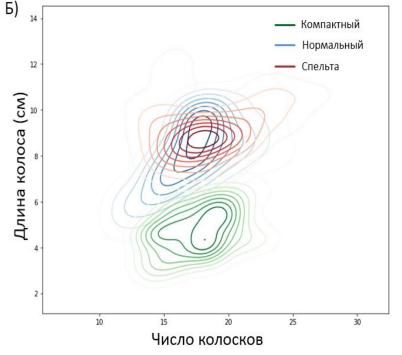


Предсказание типа колоса и его плотности

На основе геометрических характеристик колоса было проведено предсказание его типа (нормальный, спельтоидный и компактный). Дополнительно мы оценивали на их основе индекс плотности колоса: $D = \frac{10(A-1)}{B}$

A-1 — число колосков в колосе без верхушечного колоска, B — длина стержня колоса. При предсказании мы проводили отбор наиболее информативных характеристик на основе 8 критериев (корреляции с предсказываемым параметром, энтропии и пр.).





Распределение колосьев в обучающей выборке по типам колосьев:

- (А) распределение индекса плотности колоса;
- (Б) распределение индекса колосьев в пространстве число колосков/длина колоса



Оценка точности сегментации колоса и остей

Для предсказания индекса плотности колоса и его типа были отобраны 12 наилучших признаков (8 критериев). Использовался метод случайного леса (регрессия и классификация). Использовалась библиотека scikit-learn.

Точность оценки индекса плотности колоса:

Метрика точности	Протокол «на прищепке»	Протокол «на столе»
МАЕ (обучающая выборка)	1.48	1.88
МАЕ (тестовая выборка)	3.33	4.61
МАРЕ (обучающая выборка)	6.19	7.77
МАРЕ (тестовая выборка)	13.28	17.80
R ² (обучающая выборка)	0.95	0.94
R ² (тестовая выборка)	0.75	0.52

Матрица ошибок при предсказании типа колоса:

	Компактный	Нормальный	Спельта
Предсказанный «Компактный»	158	13	0
Предсказанный «нормальный»	8	79	20
Предсказанный «спельта»	0	20	77

Точность оценки типа колоса:

Метрика точности/метод предсказания	Протокол «на прищепке»	Протокол «на столе»
F/лог. регрессия	0.82	0.78
F/лог.регрессия+индекс плотности	0.83	0.77
<i>F</i> /случ. лес	0.85	0.72
F/случ.лес+индекс плотности	0.84	0.76

Для предсказания индекса плотности колоса и его типа были отобраны 12 наилучших признаков (8 критериев).

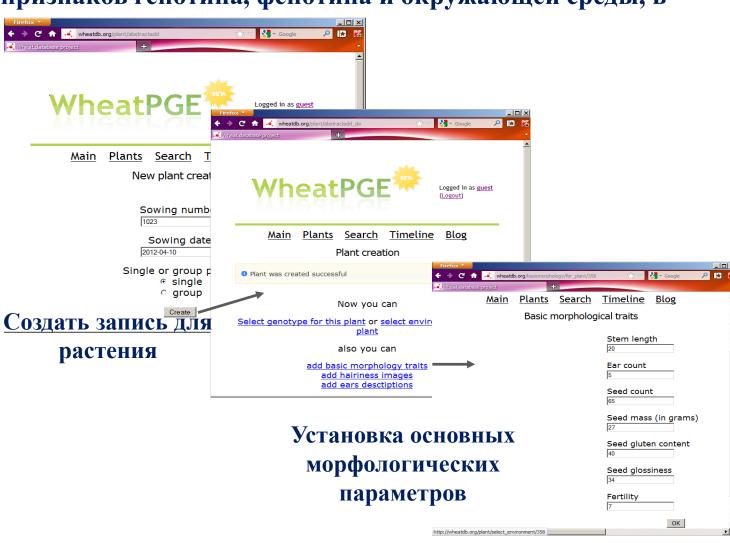


Базы данных в области фенотипирования пшеницы: WheatPGE

Для интеграции получаемых данных с информацией о генотипе и параметрах окружающей среды была разработана система WheatPGE. Центральным объектом базы данных является растение. Растение описывается как совокупность признаков генотипа, фенотипа и окружающей среды, в

которой данное растение произрастает.

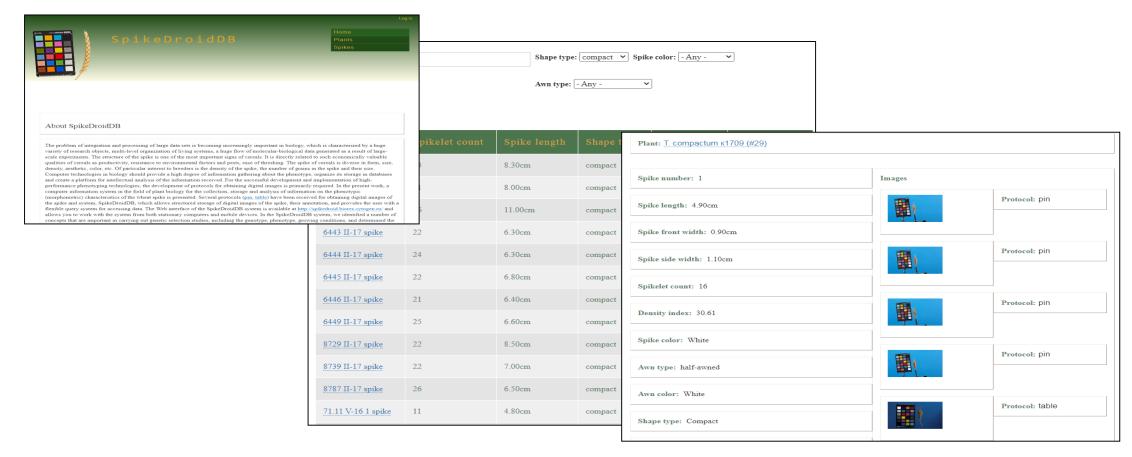






Базы данных в области фенотипирования пшеницы: SpikeDroidDB

Для сбора, хранения и анализа информации о фенотипических (морфометрических) характеристиках колоса пшеницы была разработана компьютерная информационная система SpikeDroid. БД содержит описание характеристик колосьев для отдельных растений (Описание генотипа растений, условия выращивания, место сбора колосьев, экспертное описание, характеристик колоса, фотографии колосьев).



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 Hl1, Hl2aesp и Hl3, контролирующих опушение листа у пшеницы на генетическом фоне ряда сортов и линий мягких пшениц: показано, что ген Hl2aesp в большей степени влияет на длину трихом, а гены Hl3 и Hl1 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 виды.





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

Выводы



- 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);
 - (г) 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 индекс плотности.
- 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.



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.



Публикации по теме работы

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.

- 1. Arif, M.A.R.; Komyshev, E.G.; Genaev, M.A.; Koval, V.S.; Shmakov, N.A.; Börner, A.; <u>Afonnikov, D.A.</u> (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., <u>Afonnikov, D.</u> (2022). Variability of leaf pubescence characteristics in transgenic tobacco lines with partial proline dehydrogenase gene suppression. **Biologia plantarum**, 66, 24-28.
- 3. <u>Afonnikov D.A., Komyshev E.G., Efimov, V.M., Genaev M.A., Koval V.S., Gierke P.U., Börner A.</u> (2022) Relationship between the Characteristics of Bread Wheat Grains, Storage Time and Germination. Plants 2022, 11, 35.
- 4. Комышев, Е. Г., Генаев, М. А., <u>Афонников, Д. А.</u> (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., <u>Афонников Д.А.,</u> Кочетов А.В., Рогозина Е.В., С.В. Герасимова (2019). Оценка количественных характеристик клубнеобразования дикого картофеля на основе анализа изображений клубней с использованием компьютерного приложения SeedCounter. **Вавиловский журнал генетики и селекции**, 19(6), 1-8.
- 6. Genaev M.A., Komyshev E.G., Smirnov N.V., Kruchinina Y.V., Goncharov N.P. and <u>Afonnikov D.A.</u> (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., <u>Afonnikov, D.A.</u> (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., <u>Афонников Д. А.</u> 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., <u>Афонников Д.А.</u> (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.



Публикации по теме работы

- 10. Генаев М.А., Комышев Е.Г., Фу Хао, Коваль В.С., Гончаров Н.П., <u>Афонников Д.А.</u> (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., <u>Afonnikov D.A.</u> (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., <u>Afonnikov D.A.</u>, 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., <u>Afonnikov, D. A.</u>, 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., <u>Афонников Д.А.,</u> 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. <u>Афонников Д.А.,</u> 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., <u>Афонников Д.А</u>. 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., <u>Афонников Д.А.,</u> Пшеничникова Т.А. (2014) Генетический анализ опушения листа у изогенных линий мягкой пшеницы сорта Новосибирская 67. **Генетика**, Т. 50, N. 2, C. 172–180.
- 18. Genaev M.A., Doroshkov A.V., Pshenichnikova T.A., Kolchanov N.A., <u>Afonnikov D.A.</u> (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., <u>Афонников Д.А.</u> (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., <u>Афонников Д.А.</u> (2012) Программа автоматического определения количественных характеристик опушения листа (ОЛДетект2). // Свидетельство о регистрации программы для ЭВМ N 2012620897 от 21.08.2012 от 5.03.2012, дата приоритета 21.08.2012.



Публикации по теме работы

- 21. Генаев, М.А., Дорошков А.В., Морозова Е.В., Пшеничникова Т.А., <u>Афонников Д.А.</u> (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., <u>Афонников Д.А.</u> (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., <u>Афонников Д.А.</u> (2009) Применение компьютерной обработки изображений микрофотографий листа для анализа опушения пшеницы Triticum aestivum L. **Информационный вестник ВОГиС**. 13, 218-226.