

Российская академия наук

УЧРЕЖДЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК  
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РАН

УДК 577.21 004.65 004.932.72

№ госрегистрации

УТВЕРЖДАЮ

Директор ИЦиГ СО РАН,  
академик РАН

Н.А. Колчанов

01.11.2011

ОТЧЕТ

О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

«Разработка информационного ресурса модульного типа для поддержки исследований, проводимых в рамках Технологической платформы "Биоиндустрия и биоресурсы — BioTech2030" в областях агробιοтехнологии и биοинженерии»

Этап работы: «ВЫБОР НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ И ТЕОРЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПОСТАВЛЕННЫХ ПЕРЕД НИР ЗАДАЧ»  
(промежуточный)

ГК № 07.514.11.4052 от 12.10.2011, шифр 011-1.4-514-111-000-018

Научный руководитель  
к.б.н. доцент

Нормоконтролер



подпись, дата

А.В. Кочетов



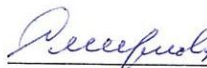







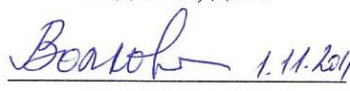
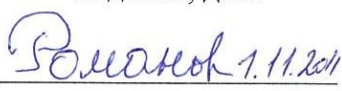

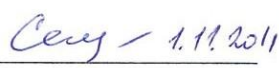




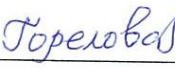









подпись, дата

Н.Н. Козырева

Новосибирск 2011

## СПИСОК ИСПОЛНИТЕЛЕЙ

<p>Научный руководитель, зав. лаб., к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Кочетов А. В. (разделы 1, 3, введение, заключение)</p>
Исполнители		
<p>зав. лаб., к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Афонников Д. А. (разделы 1, 2, 4)</p>
<p>с.н.с. к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Смирнова О. Г. (разделы 2, 3)</p>
<p>с.н.с. к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Ибрагимова С. С. (разделы 2, 3)</p>
<p>с.н.с. к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Добровольская О. Б. (разделы 2, 3)</p>
<p>м.н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Генаев М.А. (разделы 1,3,5)</p>
<p>н.с. к.б.н.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Трифонова Е.А. (раздел 2.1)</p>
<p>н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Федотова В.Д. (раздел 2.1)</p>
<p>н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Обухова Л.В. (раздел 2.1)</p>
<p>м.н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Дорошков А.В. (разделы 1.3, 4.3)</p>
<p>м.н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Волкова О.А. (раздел 3.1)</p>
<p>м.н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Романова А. В. (раздел 3.2)</p>
<p>н.с.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Суслов В.В. (раздел 1.3.1)</p>
<p>ст. лаб.</p>	 <hr style="width: 100%;"/> подпись, дата	<p>1.11.2011 Семенова С. А. (раздел 1)</p>

ст. лаб.		1.11.2011	Стрига В. Н.
	подпись, дата		(раздел 1)
ст. лаб.		1.11.2011	Небылица Н. А.
	подпись, дата		(раздел 1.2)
аспирант		1.11.2011	Горелова В.В.
	подпись, дата		(раздел 2)
н.с. к.б.н.		1.11.2011	Гунбин К.В.
	подпись, дата		(раздел 1.3.5)
программ.		1.11.2011	Рассказов Д.А.
	подпись, дата		(раздел 3)
студент		1.11.2011	Никулин П.С.
	подпись, дата		(раздел 2)
аспирант		1.11.2011	Денисюк В.С.
	подпись, дата		(раздел 1.3.6, 3)
с.н.с., к.б.н.		1.11.2011	Пшеничникова Т.А.
	подпись, дата		(раздел 1.3, 4)
с.н.с., к.б.н.		1.11.2011	Добровольская О.Б.
	подпись, дата		(раздел 1.3, 4)
м.н.с., к.б.н.		1.11.2011	Симонов А.В.
	подпись, дата		(раздел 1.3, 4)
ст. лаб.		1.11.2011	Морозова Е.В.
	подпись, дата		(раздел 1.3, 4)
Нормоконтролер		1.11.2011	Козырева Н.Н.
	подпись, дата		

## Реферат

Отчет 129 с., 10 рис., 15 табл., 203 источника, 1 прил.

Ключевые слова:

БИОТЕХНОЛОГИЯ, БАЗЫ ДАННЫХ, ГЕННАЯ ИНЖЕНЕРИЯ, АГРОБИОЛОГИЯ, СЕЛЕКЦИОННО-ГЕНЕТИЧЕСКИЙ ЭКСПЕРИМЕНТ, ИНФОРМАЦИОННЫЙ РЕСУРС

Объектом исследований является экспериментальный образец информационного портала «Биотехнология растений» (ЭОИП «БР»).

Целью исследований является разработка научно-технического задела по перспективным технологиям в области информационно-телекоммуникационных систем для поддержки НИР в рамках Технологической платформы «Биоиндустрия и биоресурсы — BioTech2030». Целью 1-го этапа НИР является выбор направления исследований и теоретическое исследование поставленных перед НИР задач.

Работа проводится с помощью компьютерных технологий, включающих разработку специализированных баз данных и обеспечение доступа к ним через экспериментальный образец Интернет-портала.

Результаты работы за отчетный этап включают аналитический обзор литературных данных, патентные исследования и обзор доступных Интернет-ресурсов в областях агробиотехнологии и биоинженерии. Результаты анализа позволили обосновать направление и методы исследований, а также выбрать платформы и аппаратные средства разработки программ и способа представления данных в процессе создания ЭОИП «БР». Основными технологическими и технико-эксплуатационными характеристиками являются: доступ к системе через Интернет, включая мобильные устройства, модульная структура, включающая базы данных внешних информационных ресурсов, промоторов и энхансеров растений, системы поддержки проведения селекционно-генетических экспериментов.

Внедрение на данном этапе не планировалось.

Областью применения ЭОИП «БР» являются научные исследования и разработки в области биотехнологии растений.

Экономические преимущества ЭОИП «БР» заключаются в увеличении эффективности проведения исследований в области биотехнологии растений. ЭОИП «БР» представляет собой информационный ресурс модульного типа, что предполагает возможность его дальнейшего развития.

## СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ	7
Сведения о проведении нормоконтроля	7
ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	8
ВВЕДЕНИЕ	9
ОСНОВНАЯ ЧАСТЬ	13
1 Обзор и анализ современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР	13
1.1 Применение биоинформационных технологий для решения биотехнологических задач	13
1.2. Информационные технологии в создании генетически-модифицированных организмов в биотехнологии	14
1.2.1 Актуальность методов трансгенеза в биотехнологии	14
1.2.2 Выбор организма для создания биопродукта	15
1.2.3 Выбор генов мишеней	18
1.2.4 Выбор векторной системы	19
1.2.5 Дизайн генетической конструкции	19
1.2.6 Выбор метода трансгенеза	23
1.2.7 Выбор метода культивирования ГМО	24
1.2.8 Выбор системы очистки продукта	25
1.3 Информационные технологии в агробиотехнологии и селекционно-генетических исследованиях растений	25
1.3.1 Актуальность применения информационных технологий в агробиотехнологии и селекционно-генетических исследованиях растений	25
1.3.2 Пшеница как важная сельскохозяйственная культура	26
1.3.3 Важные селекционно-генетические признаки пшеницы	28
1.3.3 Анализ подходов к высокопроизводительному фенотипированию растений в рамках селекционно-генетических экспериментов	32
1.3.4 Базы данных в области генетики и селекции пшеницы	37
1.3.5 Методы анализа изображений для фенотипирования растений	44
1.4 Выводы и предложения по результатам анализа информационных источников	49

2 Обоснование выбора направления и методики исследований	51
2.1 Выводы и предложения по выбору направления и методики исследований	52
3 Обоснование выбора платформ, аппаратных средств, средств разработки программ, способов представления данных	53
3.1 Выводы и предложения по разделу	55
4 Теоретические исследования	56
4.1 Перечень внешних Интернет-ресурсов (баз данных и программных комплексов), связанных с процессом получения и использования в биотехнологии ГМО, а также с селекционно-генетическими подходами в агrobiотехнологии	56
4.1.1 Внешние Интернет-ресурсы, связанные с процессом получения и использования в биотехнологии ГМО	56
4.1.2. Внешние Интернет-ресурсы, связанные с селекционно-генетическими подходами в агrobiотехнологии	59
4.2 Разработка формата БД промоторов и трансляционных энхансеров для трансгеноза растений	60
4.2.1 Разработка формата БД промоторов для трансгеноза	60
4.2.2 Разработка формата БД трансляционных энхансеров для трансгеноза	67
4.3 Разработка формата описания фенотипических признаков растений в базе данных WheatPGE	70
4.3.1 Форматы описания признаков	70
4.3.2 Алгоритм определения количественных характеристик опушения листа	73
4.4 Разработка формата описания селекционно-генетического эксперимента в базе данных WheatPGE	79
4.5 Выводы теоретических исследований	81
ЗАКЛЮЧЕНИЕ	83
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	85
ПРИЛОЖЕНИЕ (Отчет о патентных исследованиях)	103

## **НОРМАТИВНЫЕ ССЫЛКИ**

В настоящем отчете о НИР использованы ссылки на следующие стандарты:

ГОСТ 2.111-68 Единая система конструкторской документации. Нормоконтроль

ГОСТ 7.32-2001 Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления

ГОСТ 15.011-96 Система разработки и постановки продукции на производство. Патентные исследования. Содержание и порядок проведения

### **сведения о проведении нормоконтроля**

Нормоконтроль был проведен согласно ГОСТ 2.111. в соответствии с требованиями п.п.

3.4 ГОСТ 7.32-2001 (нормоконтролер – Козырева Н.А.)

## ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящем отчете о НИР применяют следующие термины с соответствующими определениями:

Промотор – участок гена, отвечающий за транскрипцию мРНК

Энхансер – участок ДНК или мРНК усиливающий транскрипцию или трансляцию, соответственно

Транскрипция – процесс синтеза пре-мРНК на матрице ДНК

Трансляция – процесс синтеза белка на матрице мРНК

Генотип - генетическая конституция, совокупность генов данного организма, полученная им от родителей.

Геном - характерный для каждого вида организмов гаплоидный (одинарный) набор хромосом; совокупность всех генов (всей ДНК), заключённых в гаплоидном наборе.

Фенотип - особенности строения и жизнедеятельности организма, обусловленные взаимодействием его генотипа с условиями среды.

Пиксель - наименьший логический элемент двумерного цифрового изображения в растровой графике.

Обозначения и сокращения:

$r$  - коэффициент корреляции Пирсона.

$\sigma$  - дисперсия распределения.

ОС – операционная система

СУБД – система управления базой данных

БД – база данных

ГМО - генетически-модифицированные организмы

ДНК – дезоксирибонуклеиновая кислота

мРНК – матричная (информационная) рибонуклеиновая кислота

ПК – программный комплекс

MPI – стандарт передачи сообщений в программировании (Message-Passing Standard)

A, U, G, C – нуклеотиды аденин, уридин, гуанин, цитозин, соответственно

ССТФ - сайты связывания транскрипционных факторов

PCR - от англ. *polymrase chain reaction* полимеразная цепная реакция

CSV - от англ. *comma separated values*, таблица данных, разделенных запятой.



## **ВВЕДЕНИЕ**

Целью научно-исследовательских работ, проводимых в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы» по теме: «Разработка информационного ресурса модульного типа для поддержки исследований, проводимых в рамках Технологической платформы "Биоиндустрия и биоресурсы — БиоТех2030" в областях агробιοтехнологии и биоинженерии» (шифр заявки «2011-1.4-514-111-018») в соответствии с государственным контрактом № 07.514.11.4052 от 12 октября 2011 г. является разработка научно-технического задела по перспективным технологиям в области информационно-телекоммуникационных систем, исследования и разработки по которым осуществляются в соответствии с направлениями технологического развития, поддерживаемыми в рамках Технологической платформы «Биоиндустрия и биоресурсы — БиоТех2030». В рамках проекта планируется разработать экспериментальный образец информационного портала (ЭОИП) «Биотехнология растений», содержащий базы данных для планирования экспериментов в областях генной инженерии и агробιοтехнологии, а также ссылки на внешние информационные ресурсы (базы данных и комплексы программ), в этих областях.

Проведение эффективных теоретических и прикладных исследований в настоящее время невозможно без информационного сопровождения. В специализированной литературе широко обсуждается необходимость использования баз данных и программных комплексов для планирования экспериментов и интерпретации их результатов. В настоящее время в открытом доступе нет информационных Интернет-ресурсов, позволяющих решать задачи в области генной инженерии и агробιοтехнологии (актуальность которых обоснована в аналитическом обзоре), что позволяет определить ЭОИП «Биотехнология растений» востребованным и соответствующим мировому уровню.

Основные выводы, сделанные на основе проведения патентных исследований по тематике проекта, заключаются в следующем. Разрабатываемая тема соответствует мировому уровню техники, поскольку нацелена на создание нового информационного ресурса модульного типа, предназначенного для поддержки научно-исследовательских разработок в областях агробιοтехнологии и создания новых продуктов и биопроцессов с помощью геномных и постгеномных технологий, методов биоинженерии и клеточных технологий. Наибольшую изобретательскую активность в разрабатываемой тематике проявляют фирмы США; эта страна является наиболее перспективной страной для реализации исследуемого объекта техники. В то же время ряд лидирующих фирм в

исследуемой области активно патентуют свои разработки в Российской Федерации (выявлено 19 изобретений).

Конкурентные преимущества нового информационного ресурса позволят рассчитывать на масштабное позиционирование на мировом рынке Российской биоинформационной продукции, поскольку существует устойчивая тенденция увеличения масштабов спроса в области предоставления биоинформационных услуг и рынка принципиально новых биоинформационных продуктов.

Из доступных патентных источников информации не обнаружено действующих патентов на территории России, под действие которых может подпадать исследуемый объект, в связи с чем можно сделать вывод, что объект исследования обладает патентной чистотой в отношении России по состоянию на 27.10.2011.

Целью первого этапа являлись анализ современной научно-технической, нормативной, методической литературы и на этой основе оценка современного состояния решаемой научно-технической проблемы, определение оснований и исходных данных для решаемой научно-технической задачи, а так же проведение подготовительных работ для создания экспериментального образца информационного портала (ЭОИП) «Биотехнология растений». Основными задачами первого этапа являлись:

- 1) проведение патентных исследований;
- 2) аналитический обзор современной научно-технической, нормативной, методической литературы;
- 3) составление перечня внешних Интернет-ресурсов (баз данных и программных комплексов), связанных с процессом получения и использования в биотехнологии ГМО, а также с селекционно-генетическими подходами в агrobiотехнологии;
- 4) разработка формата БД промоторов и трансляционных энхансеров для трансгенеза растений;
- 5) анализ подходов к высокопроизводительному фенотипированию растений в рамках селекционно-генетических экспериментов;
- 6) разработка формата описания фенотипических признаков растений в базе данных WheatPGE;
- 7) разработка формата описания селекционно-генетического эксперимента в базе данных WheatPGE.

На основе результатов, полученных за первый этап выполнения проекта, должны быть сформулированы и обоснованы содержание и основные параметры

экспериментального образца информационного ресурса «Биотехнология растений», разработка которого является задачей НИР в целом.

а) оценка современного состояния решаемой научно-технической проблемы

Анализ решаемой проблемы по данным литературных и патентных источников, а так же доступных Интернет-ресурсов позволил сделать оценку ее современного состояния и показал, что в настоящее время применение информационных технологий при разработке генетически модифицированных организмов все более интенсивно используется как в России, так и за рубежом. Создаются базы данных по генетическим конструкциям, регуляторным районам растений. В этой области особенно важную роль приобретают биоинформационные ресурсы.

Следует отметить, что биоинформатика (комплекс компьютерных методов для обработки биологических данных, поиска закономерностей и разработки технологий предсказания), системная биология (компьютерное моделирование генных и метаболических сетей на основе анализа количественных данных) и синтетическая биология (разработка системы генетических модификаций, позволяющих изменить фенотип организма в заданном направлении – вплоть до создания искусственных живых организмов) считаются основными инструментами качественного развития биотехнологии. Благодаря развитию этих методов совершенствуются методы поиска регуляторных последовательностей в геноме.

В агробиотехнологии информационные технологии используются для широкомасштабных экспериментов по исследованию взаимосвязи фенотип-генотип у растений. Прежде всего, это базы данных, описывающих отношения генотип-фенотип-окружающая среда у растений, а так же системы автоматического контроля за произрастанием растений в теплицах. Актуальность проведения исследований в этом направлении следует из анализа литературных данных, приведенных в аналитическом обзоре.

б) основание и исходные данные для разработки темы

Основанием для разработки темы служит необходимость информационного сопровождения исследований и разработок, проводимых в области генной инженерии и агробиологии в рамках основных направлений Технологической платформы «Биоиндустрия и биоресурсы — BioTech2030». На основе анализа литературных данных были определены основные этапы проведения двух актуальных направлений: генно-инженерных экспериментов и высокопроизводительного фенотипирования в селекционно-генетических опытах в агробиологии и были отобраны этапы,

эффективность планирования которых может быть значительно увеличена при применении специализированных Интернет-ресурсов, что послужило основанием для разработки темы. Был проведен детальный анализ информации, необходимой для обеспечения специализированного информационного сопровождения (например, промоторов и трансляционных энхансеров для генно-инженерных опытов, морфологических параметров растений для фенотипирования), что рассматривается в качестве исходных данных для разработки темы.

в) обоснование необходимости проведения НИР

Был проведен анализ внешних доступных Интернет-ресурсов и была проведена оценка их применимости для решения поставленных в НИР задач. На основе этого анализа было сделано заключение о том, что существующие информационные ресурсы не могут эффективно применяться для решения задач НИР, актуальность которых для проведения исследований в области биотехнологии обоснована аналитическим обзором литературных данных.

д) актуальность и новизна темы, связь данной работы с другими научно-исследовательскими работами.

Актуальность и новизна работы обоснована аналитическим обзором литературных данных, патентных источников и обзором доступных Интернет-ресурсов в области генной инженерии и агробиологии. Согласно полученным данным, поставленные в НИР задачи рассматриваются как актуальные, однако эффективные методы их решения отсутствуют. Связь с другими научно-исследовательскими работами обусловлена самой природой НИР, в рамках которой должен быть создан ЭОИП для информационного сопровождения исследований в области генной инженерии и агробиологии. Собственно применение результатов НИР должно осуществляться в связи с другими НИР, проводимыми в этих областях для решения актуальных задач биотехнологии.

## **ОСНОВНАЯ ЧАСТЬ**

**1 Обзор и анализ современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР**

### **1.1 Применение биоинформационных технологий для решения биотехнологических задач**

Биоинформатика (комплекс компьютерных методов для обработки биологических данных, поиска закономерностей и разработки технологий предсказания), системная биология (компьютерное моделирование генных и метаболических сетей на основе анализа количественных данных) и синтетическая биология (разработка системы генетических модификаций, позволяющих изменить фенотип организма в заданном направлении – вплоть до создания искусственных живых организмов [1]) считаются основными инструментами качественного развития биотехнологии. Для подтверждения этого тезиса можно привести ключевые фразы из нескольких недавних публикаций:

Биоинформатика и методы системной биологии становятся все более востребованными по мере роста объема данных. Методы семантического контекстного поиска нужной информации (text-mining), эффективно организованные базы данных становятся необходимым инструментарием, без которого планирование конкурентоспособных НИР вряд ли возможно. В качестве примера можно привести цитату из недавнего обзора ([2]: “there is so much information available that we simply no longer know what we know, and finding what we want is hard - too hard”. Биоинформатика и системная биология необходимы для реконструкции генных и метаболических сетей [3]. Генные сети и их моделирование лежит в основе интеграции геномных, транскриптомных, протеомных и метаболомных данных [4]. Синтетическая биология использует достижения системной биологии, биоинформатики, молекулярной биологии и белковой инженерии для планирования генетических модификаций, позволяющих изменять фенотип клетки в заданном направлении. Тысячи генетических элементов – белок-кодирующих последовательностей, репрессоров, активаторов, промоторов и т.д. сейчас доступны и они могут применяться для получения новых форм микроорганизмов [5]. Методы системной и синтетической биологии рассматриваются в качестве основных инструментов, позволяющих моделировать метаболические пути, предсказывать направление генетических модификаций [6]. Системная биология (интеграция

количественных экспериментальных данных и компьютерного моделирования) стала важнейшим инструментом развития современной биотехнологии [7].

В качестве примера актуальности применения биоинформатики, системной и синтетической биологии в целях биотехнологии можно привести программу заявок по 7-ой рамочной программе (FP7) Европейского союза в области биотехнологии (Food, Agriculture and Fisheries, and Biotechnology): Area 2.3.6. Emerging trends in biotechnology. КВВЕ.2012.3.6-01: “Systems Biology - ERANET”

В РФ также существуют сильные научные школы, специализирующиеся в биоинформатике и различных разделах системной и синтетической биологии. Близкими к биотехнологии направлениями исследований можно считать сравнительную геномику и реконструкцию генных и метаболических сетей [8-11], моделирование онтогенеза органов и тканей [12,13], систематизацию информации в виде различных баз данных (например, [14]), исследование механизмов эволюционной адаптации микроорганизмов к экстремальным условиям окружающей среды [15] и т.д. Использование этих ресурсов для планирования биотехнологических НИР необходимо для повышения их эффективности.

## **1.2 Информационные технологии в создании генетически-модифицированных организмов в биотехнологии**

### **1.2.1 Актуальность методов трансгенеза в биотехнологии**

Трансгенные организмы различной таксономической принадлежности уже активно используются в различных отраслях современной биотехнологии. Генетические модификации позволяют получать штаммы бактерий и грибов – продуцентов ферментов, аминокислот, биологически активных веществ, также широко используются трансгенные растения. С помощью ГМО развиваются методы получения биотоплива [16-18], разрабатываются технологии оптимизации трансгенеза [19] и наработки биопрепаратов в трансгенных растений (molecular pharming) [20-24], наработки вторичных метаболитов оптимизированными культурами тканей [25].

За последние 20 лет системы наработки биотехнологически значимых продуктов на основе растительных систем составляет все большую конкуренцию на рынке производства рекомбинантных белков фармацевтического назначения [26,27]. Уже первый рекомбинантный белок авидин производимый с помощью растительных систем (кукурузы) появился на рынке. Как системы для наработки фармацевтически значимых белков, растения более безопасны по сравнению с микробами и животными: у них нет

патогенов человека, онкогенных последовательностей ДНК, а также эндотоксинов. К настоящему времени список фармацевтических белков производимых различными растительными системами значительно расширился и включает: антитела, вакцины, ростовые факторы, гормоны.

Не вызывает сомнений тот факт, что направленное внесение серии генетических модификаций, приводящее к получению форм организмов, адаптированных для эффективного решения насущных технологических задач, является магистральным направлением развития современной биотехнологии. Генетические модификации планируются на основе фундаментальной проработки экспериментальной информации из различных областей современной науки. Согласно недавним публикациям, протеомные данные являются важным источником информации для планирования НИР в области биотехнологии [28,29]. Метаболомика является одним из важнейших инструментов, особенно полезным для биотехнологии растений, у которых велико число вторичных метаболитов [30]. Инструментарий метаболомики позволяет решать целый спектр задач, связанных с выявлением генов и QTL у растений, оценке качества биоматериалов (в том числе пищевых продуктов и биобезопасности ГМО): «systems biology driven by metabolome data will aid in deciphering the secrets of plant cell systems and their application to biotechnology» [30]. Метаболическая инженерия уже внесла существенный вклад в развитие производства материалов и реагентов из возобновляемых источников [31-33].

НИР, в которых используют генетически-модифицированные организмы, может включать исследования, в которых применяется очень широкий круг современных технологий – практически все основные генетические, молекулярно-биологические, культуральные, физиологические методы. Согласно современным представлениям, планирование генно-инженерного эксперимента в общем виде требует эффективного планирования и включает ряд последовательных этапов, которые будут описаны ниже.

### **1.2.2 Выбор организма для создания биопродукта**

Выбор организма в ряде случаев задан изначально, если речь идет о специфическом вторичном метаболите, характерном для определенного организма. Однако, в некоторых случаях требуется создание продуцента белка (фермента или фармакологического препарата), для чего могут быть использованы различные подходы. Разные биопродукты характеризуются различными преимуществами и недостатками, среди которых следует выделить степень близости продукта к природному варианту (для

белков важны пост-трансляционные модификации), отсутствие токсичных примесей, сложность выделения и очистки, количественные характеристики синтеза и себестоимость продукции. Например, белки человека можно производить с помощью культур соответствующих клеток и полученный продукт будет практически идентичен натуральному. Однако, стоимость такой продукции будет также наиболее высокой – вследствие требований к стерильности, к отсутствию в культуре вирусов или прионов, а также в результате низкого выхода. С другой стороны, культуры микроорганизмов могут давать высокий уровень биопродукции, однако полученный белок может характеризоваться конформацией, отличной от природного варианта – в частности, это касается пост-трансляционных модификаций. Выбор организма-биопроизводителя в данном случае зависит от экспертной оценки его особенностей, определяющих преимущества и недостатки в рамках решения конкретной технологической задачи.

В настоящее время около сотни видов растений подверглись генетической модификации, выявлены наиболее подходящие виды растений, на основе которых возможно получение фармацевтически значимых белков. Каждый из рассматриваемых объектов обладает как преимуществами, так и недостатками. Первое место среди потенциальных продуцентов рекомбинантных белков занимает табак. Он формирует большую зеленую массу, дает около 1 млн семян с растения. Технология переноса чужеродных генов и их экспрессии у табака хорошо разработана. Табак не является основным пищевым продуктом для человека и кормом для скота, что снижает риск контаминации. Однако у большинства сортов табака повышен уровень токсичных алкалоидов, который разрушается в результате измельчения листьев при экстракции рекомбинантного белка. В то же время суспензионная культура клеток табака ВУ-2 характеризуется низким содержанием алкалоидов.

Среди других видов способных к продукции рекомбинантных белков, относятся такие хозяйственно-важные культуры как люцерна, соя, салат латук. Люцерна и соя способны к фиксации атмосферного азота, что снижает затраты на химические удобрения, люцерна также способна давать возобновляемый урожай зеленой массы до 9 раз в год. Салат латук был использован в качестве продуцента съедобных вакцин и были проведены их клинические испытания. Как отмечается, основной недостаток листовых культур заключается в том, что синтез белка происходит во влажной среде, что делает его нестабильным, что приводит, в конечном счете, к низкому его выходу. Для сохранения наработанных белков требуется быстрое замораживание листьев или их высушивание, а затем экстракция, что повышает стоимость продукта.



Наряду с наработкой рекомбинантных белков в вегетативных органах (листья) растения, потенциальными органами растений накапливающих целевой продукт могут быть семена. Семена способны запасать и сохранять белки при комнатной температуре в течение 3-х лет. Более того в семенах злаков отсутствуют фенольные соединения, присутствующие у табака. Однако общий выход целевого продукта значительно ниже, чем у табака. Среди потенциальных продуцентов рекомбинантных белков в семенах следует отметить рис, пшеницу, кукурузу, а также сою и бобы. Среди овощных культур потенциальными продуцентами рекомбинантных белков являются такие овощные культуры как томат и картофель. Картофель является одним из лучших продуцентов вакцин, антител. Томаты также являются продуцентом вакцин, и они более съедобны. Их можно выращивать в теплице, тем самым увеличивая выход конечного продукта.

Успешно используются также линии суспензионных клеточных культур табака ВУ-2 и риса, где конечный продукт может экскретироваться клетками через поры, если размер секретлируемого продукта не превышает 20-30 килодальтон. Суспензионные культуры клеток обладают быстрой скоростью роста, и следовательно быстрым накоплением целевого продукта, что позволяет быстро оценить эффективность системы экспрессии и качество продукта, оценить затраты.

В целом выбор организма-продуцента должен оцениваться по эффективности наработки целевого продукта, его качества в сочетании с его стоимостью.

Таким образом, выбор организма представляет собой важный шаг, основанный в большей степени на экспертной оценке. Тем важнее в этой области интеграция информации о возможных организмах–продуцентах, которая обеспечила их правильный выбор. К сожалению, в настоящее время ресурсов, которые решали бы такую задачу, нет. Информация может быть получена из целевых обзоров, обычно посвященных конкретным особенностям различных ГМО. Кроме того, некоторые сведения могут быть получены из разрозненных БД трансгенных растений:

- Transgenic Crops (<http://cls.casa.colostate.edu/transgeniccrops/>),
- GMO DB (<http://www.gmo-compass.org/eng/gmo/db/>)
- GM Crop data base [http://cera-gmc.org/index.php?action=gm\\_crop\\_database](http://cera-gmc.org/index.php?action=gm_crop_database)
- Bulgarian Plant Genomics Database <http://bulgenom.abi.bg/Welcome.htm>

В этих базах накоплена информация об известных ГМ культурах, с указанием введенного трансгена, способах трансформации. Однако, для обеспечения большей эффективности решения задачи выбора организма-продуцента необходимо, чтобы эти ресурсы были бы доступны с одного портала.

### 1.2.3 Выбор генов мишеней

При создании ГМО глубина собственно модификаций может варьировать и зависит от поставленной задачи. Существующие технологии позволяют как усиливать экспрессию определенного белка (за счет внесения трансгена в геном), так и снижать или выключать экспрессию (например, с помощью использования нокаутных штаммов, генетического сайленсинга или РНК-интерференции). Процесс выбора прост в тех случаях, когда он задан изначально и планируется получение ГМО, которые несут один трансген – например, при биопродукции фармакологически-значимого (чужеродного) белка или применяемого в биотехнологическом производстве фермента. Однако, может ставиться более сложная задача получения вторичного метаболита, либо ГМО должен характеризоваться дополнительными параметрами (например, присутствием дополнительных специфических белков, участвующих в процессинге основного продукта). В качестве примера можно привести разработки, в которых для повышения выхода определенного метаболита у ГМО выключают метаболические цепи, конкурирующие за субстрат или интермедиаты. В этих случаях необходимо моделирование биохимических контуров и расчет параметров, обеспечивающих оптимальный уровень синтеза.

Информационные ресурсы в этой области: Базы данных, в которых накапливается информация о белковых продуктах, ферментах, регуляторных молекулах, метаболических и генных сетях, содержат информацию, необходимую для выбора направления генетической модификации. На основе реконструкции генных и метаболических сетей возникает возможность идентификации целевых генов, которые могут быть использованы при трансгенезе.

Наиболее известной БД такого типа является KEGG (Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.jp/kegg/>). Этот ресурс был разработан первым (в 1995 году) и является результатом скрупулезной аннотации литературных данных о метаболических путях большого числа различных живых организмов (на настоящий момент – 1370 бактерий, 116 архей и 151 эукариотический организм). Недостатком KEGG является тот факт, что это платный ресурс. База данных MetaCyc <http://metacyc.org/> содержит информацию о более 1790 метаболических путях, более чем для 2000 организмов, аннотированных на основе экспериментальных работ. Примерами более специализированных баз данных метаболических путей также являются AraCyc (<http://www.arabidopsis.org/biocyc/>; БД метаболических путей модельного растения

*Arabidopsis thaliana*), в более широком виде информация о метаболических путях растений представлена в БД PlantCyc (<http://plantcyc.org/>). Однако, для решения задач биотехнологии растений и агробиотехнологии более приспособлена БД MetaCrop (<http://metacrop.ipk-gatersleben.de>), в которой представлена информация об основных метаболических путях хозяйственно-ценных видов растений (*Hordeum vulgare*, *Triticum aestivum*, *Oryza sativa*, *Zea mays*, *Solanum tuberosum*, *Brassica napus*, *Beta vulgaris*). В этой системе рассмотрены комплексные циклы, относящиеся к метаболизму углеводов, кофакторов, нуклеотидов, липидов, аминокислот и т.д.

Существуют еще более специализированные БД сигнальных каскадов и медиаторов клеточного ответа. В качестве примера можно привести базу данных о гормонах арабидопсиса AHD (Arabidopsis Hormone Database). Ресурс содержит диаграммы сигнальных путей и путей биосинтеза гормонов, информацию о генах, участвующих в гормональной регуляции, обеспечивает для каждого гена предсказание возникающих в процессе сплайсинга сайтов связывания с микроРНК, контролирующих стабильность мРНК [34,35].

#### **1.2.4 Выбор векторной системы**

Выбор векторной системы обычно определяется спецификой поставленной технологической задачи и организмом-реципиентом генетической конструкции, с помощью которого будет реализоваться проект. Этот выбор весьма разнообразен (вирусные векторы, плазмиды разных типов, интегрирующиеся в геном конструкты и т.п.).

Информационные ресурсы в этой области: специализированный информационных ресурсов, позволяющих выбирать векторы для решения задач биотехнологии, в свободном доступе нет. База данных VectorDB решала такую задачу на протяжении некоторого времени, однако она уже давно не поддерживается (<http://genome-www.stanford.edu/vectordb/vector.html>). Существующие общественные ресурсы (например, LabLife <http://www.lablife.org/vectordb>) очень ограничены по содержанию.

#### **1.2.5 Дизайн генетической конструкции**

Этот этап включает выбор адекватного промотора, при необходимости – подбор энхансера трансляции и поли(А)-сигнала, оптимизацию кодонного состава. Также необходимо удостовериться в отсутствии ложных сигналов экспрессии: поскольку ДНК

трансгена часто принадлежит организму другой таксономической принадлежности, она характеризуется нуклеотидным составом, отличным от геномной ДНК организма-хозяина (например, ДНК млекопитающих обогащена G+C в сравнении с двудольными растениями). Это, в свою очередь, может привести к присутствию комбинаций нуклеотидов, которые будут распознаваться в клетках организма-хозяина как сигналы экспрессии. В качестве примера можно привести первые линии трансгенных растений, полученные фирмой Monsanto и несущие ген инсектицидного белка *Bacillus thuringiensis*. Эти линии характеризовались практическим отсутствием экспрессии трансгена, поскольку ДНК бактериального происхождения несла комбинации нуклеотидов, распознаваемые в клетках растений как сигналы сплайсинга и полиаденилирования, в результате чего большая часть зрелой мРНК содержала дефектные варианты белок-кодирующей части.

Для обеспечения необходимого паттерна экспрессии перенесенного гена необходимо использовать адекватные регуляторные последовательности и сигналы экспрессии. Промотор является наиболее важным элементом генетической конструкции. В настоящее время генными инженерами активно используется ограниченный набор промоторов, что препятствует развитию экспериментов с трансгенными растениями и недостаточно для решения сложных биотехнологических задач [36-38]. В качестве примера удачных в биотехнологическом плане промоторов можно привести коммерческое применение трансгенных растений со стадиейспецифичным промотором SAG12 арабидопсиса. При использовании в качестве трансгена изопентилтрансферазы *Agrobacterium tumefaciens*, специфическая активность промотора AtSAG12 на поздних стадиях развития у большого числа видов трансгенных растений приводит продлению вегетации, повышению урожая и его сохранности [39-51].

Помимо сигналов транскрипционного контроля, большое значение может иметь оптимизация экспрессии трансгена на посттранскрипционном уровне. В составе эукариотических мРНК часто содержатся сигналы, контролирующие эффективность трансляции или цитоплазматическую стабильность матрицы. Применение таких сигналов в дизайне трансгена может существенно увеличить эффективность трансгенеза. Белок-кодирующая часть чужеродного гена может обладать негативными характеристиками, снижающими эффективность экспрессии. Для достижения эффективного взаимодействия с аппаратом экспрессии организма-реципиента часто бывает необходимо модифицировать структуру трансгена.

Например, в ряде случаев, особенно у злаков, введение интрона в район 5'-конца мРНК целевого гена резко усиливает его транскрипцию. Короткий (173 п.о.) фрагмент промотора глютелина пшеницы 1Вх17 НМW-GS, связанный с интроном гена актина риса, имел такой же уровень экспрессии, как и длинный 1900 bp 1Вх17 НМW-GS промотор без интрона [52]. Промотор гена глютелина пшеницы TaLMW1D1 в комбинации с Ubi-1 интроном кукурузы был использован для наработки легумина гороха в зерновке трансгенной пшеницы [53]. Промотор гена глутатион-трансферазы пшеницы GstA1 в комбинации с интроном гена TaWIR1a проявляет сильную конститутивную активность в эпидермисе и может быть с успехом использован для повышения устойчивости к грибковым заболеваниям у зерновых [54,55].

Приведем описание информационных ресурсов в этой области. Примеры создания стандартной схемы генетической конструкции приведены в базах данных Transgenic Crops (<http://cls.casa.colostate.edu/transgeniccrops/how.html>). В мире разработан ряд баз данных, содержащих информацию о нуклеотидных последовательностях промоторов и сайтов связывания транскрипционных факторов (ССТФ). Большинство из них накапливают информацию о ССТФ. База данных PLACE позволяет находить регуляторные цис-элементы в последовательности ДНК растительных генов [56]. Три взаимосвязанных базы AtTFDB, AtcisDB и AtRegNet информационного сервера AGRIS (Arabidopsis Gene Regulatory Information Server) - содержат информацию о последовательностях промоторов Арабидопсиса, транскрипционных факторах и их целевых генах. Один из трех модулей базы, AtcisDB содержит около 33,000 вышележащих районов аннотированных генов Арабидопсиса с описанием экспериментально проверенных и предсказанных ССТФ [57]. База данных PlantProm database (PPDB) содержит нуклеотидные последовательности промоторов растений с экспериментально проверенным стартом транскрипции [58]. Базы данных Athena и Osiris обеспечивают визуализацию ССТФ и предсказанные структуры промоторов арабидопсиса [60]. Athena содержит нуклеотидные последовательности промоторов размером до 3 kb для предсказанных генов арабидопсиса и нуклеотидные последовательности консенсусов для 105 охарактеризованных ранее ССТФ, импортированных из PLACE и AGRIS. База Osiris содержит нуклеотидные последовательности промоторов, предсказанных ССТФ, аннотацию онтологии генов и данные анализа транскрипции на микрочипах для 24209 генов из генома риса. База данных AthaMap представляет полногеномное картирование потенциальных сайтов связывания ТФ. База содержит сайты для 115 различных ТФ [61]. Навигатор PlantPAN

(Plant Promoter Analysis Navigator) предназначен для узнавания комбинаторных цис-элементов в генах растений [62]. Таким образом, имеющиеся информационные ресурсы преимущественно представляют данные о сайтах связывания транскрипционных факторов и не содержат информацию об экспериментально проверенной промоторной активности фрагментов ДНК растений. В принципе, присутствие ССТФ может быть использовано для выбора промотора при создании генетической конструкции. Однако существует много примеров ошибочности таких предсказаний. Несмотря на то, что промотор GluC не содержит специфичных для экспрессии в эндосперме мотивов (GCN4 AACA, prolamin box), он обеспечивает высокий уровень транскрипции в этой ткани [38]. Ген Pptha1 персика был обнаружен благодаря повышенной экспрессии в условиях холода, но промоторный район этого гена не обеспечивал аналогичный паттерн экспрессии в трансгенных растениях арабидопсиса [63]. Наиболее реальным источником для выбора промотора может быть информация о фрагментах ДНК с определенной транскрипционной активностью, полученной в трансгенных экспериментах [64-67]. Таким образом, мы считаем обоснованным развитие специализированной базы данных промоторов для трансгенеза растений, основанной на аннотировании литературных данных.

Помимо транскрипционного контроля экспрессии трансгенов для эффективного проведения НИР в этой области необходимо планировать и оптимизацию других уровней (пост-транскрипционных). В частности, практически важным является адекватное планирование эффективности трансляции мРНК (обычно генетические конструкции не содержат сайтов сплайсинга и в них используются аутентичные сигналы полиаденилирования, взятые из генов организма-реципиента, поэтому с этой фазой экспрессии трансгена возникает меньше проблем).

Одним из способов управления экспрессией гена являются трансляционные сигналы экспрессии, обычно расположенные в составе 5'- или 3'-нетранслируемых районов мРНК [68]. Некоторые из таких сигналов могут определять общую трансляционную активность мРНК: например, если в составе генетической конструкции используется 5'-НТП мРНК вируса табачной мозаики (размером 68 нуклеотидов), то это в большинстве случаев увеличивает уровень синтеза белкового продукта в несколько раз [69]. Этот энхансер активно используется в биотехнологии растений. Кроме неспецифических усилителей трансляционной активности мРНК, известны специфические сигналы (например, IRE – регулирующий трансляцию мРНК гена ферритина в зависимости от присутствия железа; с этим сигналом связывается

специфический белок IRP, конформация (а активность) которого зависит от присутствия железа, что функционально эквивалентно ситуации транскрипционного контроля и сайтам связывания транскрипционных факторов в промоторах). Однако, в отличие от сайтов связывания транскрипционных факторов, сигналы трансляционного контроля экспрессии малоизучены, что связано со специфическими особенностями одноцепочечных молекул РНК, способных в цитоплазме существовать в виде различных конформеров. Большинство сигналов такого типа представляют собой комбинацию контекстных и структурных элементов, что делает их предсказание очень сложным.

Информационные ресурсы в этой области представлены базами данных UTRsite [70] и TransTerm [71]. Сразу следует отметить, что в этих БД аннотировано всего несколько десятков трансляционных сигналов разной таксономической принадлежности (по некоторым оценкам сигналы этого типа могут присутствовать у 10 – 15% мРНК генов эукариот, то есть аннотировано крайне малое количество). Причина такой недостаточной аннотации заключается в том, что в этих БД в качестве трансляционных сигналов рассматриваются те из них, у которых точно известна тонкая структура (контекстные и конформационные элементы сигнала). При этом для планирования биотехнологических опытов эта информация избыточна, достаточно знать, как изменяется паттерн трансляционной активности мРНК трансгена в присутствии тех или иных участков мРНК с известной трансляционной активностью. Следует отметить, что экспериментальной информации в литературе достаточно много, что говорит о возможности разработки информационного ресурса такого рода.

### **1.2.6 Выбор метода трансгенеза**

В большинстве случаев метод трансгенеза определяется организмом, использованным для получения ГМО и особенностями векторной системы. Однако, в некоторых случаях существует выбор: так, трансгенные растения можно получать с помощью агробактериальной трансформации, бомбардировки частицами с сорбированной на них ДНК, трансфекции протопластов с помощью электропорации или сурфактантов и т.п. Каждый из методов имеет преимущества и недостатки.

По сравнению с доставкой незащищенной ДНК, трансформация с помощью *Agrobacterium tumefaciens* имеет несколько положительных особенностей, таких как внедрение одной или двух копий трансгенной инсерции, возможность доставки фрагментов ДНК с минимальными перестройками [72,73]. Было показано, что у ячменя эффективность агробактериальной трансформации была в два раза выше, чем при

бомбардировке микрочастицами [74]. Все трансгенные линии, полученные агробактериальной трансформацией, имели от одной до трех копий трансгена, в то время как 60% трансгенных линий, полученных бомбардировкой микрочастицами, имели более 8 копий трансгена. В настоящее время считается значительно более правильным получение ГМО с одной копией T-ДНК инсерции, поскольку такие растения характеризуются большей стабильностью экспрессии трансгена в поколениях и у них в геноме реже содержатся гены, «сломанные» трансгенной инсерцией (knockout-гены). Поэтому методы агробактериальной трансформации активно разрабатываются даже для тех видов растений, у которых она затруднена по биологическим причинам (например, злаков) [75-84]. Недостатком метода переноса с помощью агробактерии является возможность использования ограниченного числа генотипов и тканей, обладающих высокой регенерационной способностью [85]. По этой причине трансформация бомбардировкой микрочастицами также часто используется при получении трансгенных злаков в биотехнологии растений [86-89].

В дополнение к методу трансформации с помощью «бомбардировки микрочастицами», разрабатываются новые техники, такие как лазерная микропунктура [90] и микропорация [91]. Электропорация была эффективно использована для доставки ДНК в ткани интактных незрелых эмбрионов пшеницы. Эффективность трансформации при микропорации составила 7.5%, что значительно выше 4.2% при бомбардировке микрочастицами. Однако, у этих методов есть те же самые недостатки, которые есть и у бомбардировки микрочастицами: частая встройка трансгенной инсерции в виде tandemного повтора с большим числом копий, встраивание фрагментов вектора в различные участки генома («замусоривание» генома).

Опишем информационные ресурсы в этой области. Специализированных информационных ресурсов в открытом доступе нет. Все этапы подготовки эксплантов растений к трансформации и способы трансформации представлены на Интернет портале Biotechnology 4u [http://www.biotechnology4u.com/plant\\_biotechnology.html](http://www.biotechnology4u.com/plant_biotechnology.html). в разделе «Plant Biotechnology», однако этот ресурс имеет больше образовательный характер. Существуют обзоры, в которых представлены особенности трансформации для разных культур [72,92-95].

### **1.2.7 Выбор метода культивирования ГМО**

Этот этап связан с особенностями как ГМО, так и технологического процесса. Например, трансгенные растения можно выращивать в теплице, но также можно и



культивировать *in vitro* в виде культуры каллусов или корней (например, при трансгенезе с помощью *Agrobacterium rhizogenes*). ГМ суспензионные культуры могут длительно поддерживаться путем добавления свежих питательных сред в инкубаторах объемом до 1500 литров.

Специализированные информационные ресурсы в открытом доступе отсутствуют. Информация образовательного характера обо всех известных методах культивирования хранится на интернет портале Biotechnology 4u [http://www.biotechnology4u.com/plant\\_biotechnology.html](http://www.biotechnology4u.com/plant_biotechnology.html) в разделах «Plant Cell and Tissue Culture Techniques» и «Regeneration pathways of plants».

### **1.2.8 Выбор системы очистки продукта**

Этот этап практически полностью определяется особенностями собственно продукта. Приведенная последовательность этапов представляет собой общую схему, конкретное наполнение которой в каждом случае может отличаться. Очевидно, что разработать единый информационный ресурс, позволяющий осуществлять эффективное планирование во всех случаях, практически невозможно. Однако, преимуществом предложенной нами модульной схемы является возможность комбинации отдельных ресурсов (баз данных и программных комплексов) для решения конкретной проблемы.

## **1.3 Информационные технологии в агробιοтехнологии и селекционно-генетических исследованиях растений**

### **1.3.1 Актуальность применения информационных технологий в агробιοтехнологии и селекционно-генетических исследованиях растений**

В агробιологии важную проблему представляет увеличение стрессоустойчивости и устойчивости растений к фитопатогенам и, как считают, использование методов системной биологии может привести к быстрому прогрессу [96]. В РФ также ведутся исследования в этом направлении [97-99]. Получение стрессоустойчивых растений за счет комбинации классической селекции и генной инженерии (molecular breeding) позволит расширить площади их выращивания и решить продовольственную проблему [100,101]. Например, одним из направлений развития генной инженерии растений является возможность использования отдельной хромосомы как вектора, несущего чужеродную ДНК [102]. Обсуждаются перспективы использования ГМ растений в сельском хозяйстве – рассматривается возможность 40% - 80% увеличения продуктивности при использовании ГМ [103].

Перспективные направления развития биотехнологии растений рассматриваются в контексте предсказательного моделирования не только клетки или отдельного организма, но и экосистемы в целом (“The analysis of single model systems - plants, animals and bacteria - will finally emerge in the analysis of populations of plants and other organisms and their adaptation to the ecological niche” [104]). Одним из перспективных подходов, лежащих в основе получения новых сортов в генетике и селекции растений, считается картирование локусов, отвечающих за хозяйственно-ценные признаки [105].

### **1.3.2 Пшеница как важная сельскохозяйственная культура**

Мягкая пшеница одновременно является стратегической культурой и уникальным объектом генетических исследований. Знания по частной генетике этой культуры имеют как хозяйственное, так и фундаментальное значение. Поиск новых генов проводится с целью увеличения объёма знаний в области частной генетики пшеницы и расширяет знания о структуре сложного генома данного вида. Одновременно с этим, новые подходы фенотипирования расширяют эффективность селекции, что, в свою очередь, актуально в связи с растущими потребностями производства пищевых ресурсов в мире.

Геном пшеницы отличается большим размером, превосходящим по длине геном человека в несколько раз, насыщен повторами, что затрудняет его сборку при секвенировании. За последние два года появилось множество секвенированных геномных EST - последовательностей для пшеницы, позволяющих проводить поиск по последовательностям. Для пшеницы доступны EST (expressed sequence tags), а также пятикратное покрытие генома мягкой пшеницы фрагментами (<http://www.cerealsdb.uk.net>), что позволяет предсказывать присутствие гомолога исследуемого гена в геноме пшеницы. На данный момент для мягкой пшеницы произведена сборка и привязка последовательностей к цитологической карте отдельных хромосом (<http://www.wheatgenome.org/>).

Высокая специализация клеток, низкая степень регенерации и отсутствие природных способов трансформации также затрудняет работу молекулярных генетиков. Ярким примером этого является трансформация однодольных, которая до последних лет представляла серьёзную проблему. Трансформация при помощи агробактерий для множества двудольных является превосходным методом – она проста и даёт большой выход. Однако, в случае с однодольными метод агробактериальной трансформации играет на равных с прямой и баллистической трансформацией.

Удачная трансформация включает в себя перенос генетической конструкции в клетку, его встройку в геном и успешную экспрессию в ряду клеточных поколений. Первым успешным примером трансформации пшеницы была прямая трансформация протопласта материалом из культивируемых эмбриональных клеток [106]. Этот метод работал с крайне низкой эффективностью, это было связано с трудностями, возникающими при длительном культивировании клеточной культуры и низкой способностью к регенерации, характерных для пшеницы. Однако, в последние годы методы культивирования тканей пшеницы, а также ряд подходов в трансформации стали значительно более эффективны [107,108].

Для локализации генов в геноме пшеницы, отвечающие за формирование важных селекционных признаков в настоящее время используются методы, основанные на анализе генетических маркеров. На данный момент у мягкой пшеницы разработано большое количество разнообразных генетических и молекулярных маркеров. Разнообразие молекулярных маркеров резко возросло за последние десятилетия. Широко применяются маркеры, основанные на гибридизации, такие как полиморфизм длин рестрикционных фрагментов, RFLP- метод (restriction fragment length polymorphism); полимеразной цепной реакции (PCR), такие как метод амплифицированной полиморфной ДНК со случайными праймерами -RAPD (random amplified polymorphic DNA), метод детекции полиморфизма длин амплифицированных фрагментов - AFLP-метод (amplified fragment length polymorphism), детекции длин микросателлитных повторов - SSR (simple sequence repeats), также, в качестве молекулярных маркеров используются ретротранспозоны [109]. В последние годы набирают популярность SNP-маркеры (single nucleotide polymorphism), основанные на анализе однонуклеотидных замен. Достоинства и недостатки тех или иных маркеров детально исследованы ([110-119], <http://wheat.pw.usda.gov/GG2/index.shtml>).

Эти маркеры позволяют проводить картирование генов, отвечающих как за ранговые, так и за количественные признаки. Для этого, на основе генотипов, имеющих полиморфизм по большому количеству маркеров и интересующих исследователя фенотипических признаков, создаются картирующие популяции. Всеобъемлющее резюме важных популяций и их генетических карт, доступных для зерновых, включая пшеницу, недавно было опубликовано [117]. Для отдельных популяций получено 95% покрытие генома молекулярными маркерами [120], однако, полногеномное картирование при помощи маркеров в ряде задач нецелесообразно. Это требует большого времени и значительных ресурсов. Значительно удобнее пользоваться подходом

картирования в две стадии: картирование до группы сцепления при помощи моносомного анализа, а затем уточнение локализации при помощи генетических маркеров.

Важная функция ДНК-маркеров, в частности RFLP это сравнительное картирование геномов родственных видов [121]. Для мягкой пшеницы был показан высокий уровень соответствия локусов молекулярных маркеров и генных локусов хромосом между А, В и D геномами. Дальнейшая оценка степени синтении в *Triticaceae* с использованием ДНК-маркеров показала значительную коллинеарность порядка генов среди субгеномов пшениц и геномов родственных видов [117,122-126]. Сохранение коллинеарности между видами злаков может быть использовано для предсказания локусов генов пшеницы при помощи генетических карт уже доступных для родственных видов [126]. Недавние исследования показали наличие среди EST консервативных SSR-мотивов в ортологичных локусах для пшеницы, ржи и риса [115].

### **1.3.3 Важные селекционно-генетические признаки пшеницы**

К основным признакам пшеницы можно отнести следующие: число дней от всходов до колошения, число стеблей на растении, число продуктивных стеблей на растении, длина главного стебля, длина верхнего междоузлия, длина нижнего (первого надземного) междоузлия, длина колоса, число колосков в колосе, число зерен на растении, масса зерен с растения, плотность колоса, число зерен в колосе, число зерен в колоске, масса зерен с колоса, удельный урожай колоса, масса 1000 зерен, процентное содержание белка в зерне, балл устойчивости к бурой ржавчине.

Одним из важных фенотипических признаков растений является опушение. Под опушением понимают совокупность эпидермальных образований, называемых трихомами. Этот признак разнообразен и широко распространен в растительном мире. Опушение свойственно различным органам растений – листу, стеблю, цветку. Наиболее исследован генетический контроль формирования опушения у *Arabidopsis thaliana*. Однако, различия в структуре трихом, их распределении на поверхности листа, а также структуре самих листьев между двудольными и однодольными, вкпе с отсутствием близких гомологов генов опушения листа *Arabidopsis thaliana* у однодольных позволяет предположить, что генетический контроль опушения листа однодольных имеет серьёзные отличия от такового у двудольных.

Опушение листа у пшеницы *Triticum aestivum* L. состоит из структурно однородных трихом, различающихся по размерам. Оно имеет большое адаптационное

значение. Например, сильное опушение характерно для ряда засухоустойчивых сортов, а для сортов, произрастающих во влажном климате, напротив, характерно очень слабое опушение [127]. Показано влияние опушения листа на влагоудерживающую способность растений яровой мягкой пшеницы [128]. Существуют данные о том, что сильно опушенные сорта пшеницы значительно более устойчивы к поражению жуком-листоедом *Oulema melanopus* L. [129], а также шведской мухой *Mayetiola destructor* [130]. Есть также основания полагать, что опушение листа имеет физиологическое значение. Клетки трихом, основной объем которых вынесен за плоскость эпидермы, могут расширять возможности формирования локальных градиентов тургорных давлений эпидермиса, а также снимают нарастающее в онтогенезе напряжение покровной ткани, увеличивая, таким образом, эффективность работы устьичного аппарата [131]. Поэтому для целенаправленного получения сортов, приспособленных к определенным климатическим зонам, необходимо вести отбор с учетом характеристик опушения листа. С другой стороны, мягкая пшеница является удобным объектом генетических исследований. Что позволяет использовать мягкую пшеницу как модельный объект для исследования генетического контроля опушения двудольных.

На данный момент у мягкой пшеницы известно два гена, ответственных за формирование опушения листа. Ген *HL1* локализован на длинном плече хромосомы 4В сорта Саратовская 29 [132,133]. Ген *HL2* был картирован на коротком плече хромосомы 7В линии сорта Родина с интрогрессией гена контроля опушения от *Aegilops speltoides* Tausch. [133]. Как известно, опушение листа пшеницы обнаруживает большое разнообразие морфотипов [134]. Разнообразные формы опушения листа описаны как для мягкой пшеницы, так и для её диплоидных и тетраплоидных сородичей. Это может свидетельствовать в пользу большего количества локусов, вовлеченных в контроль опушения листа этого злака.

До недавнего времени описание степени опушения листа было основано на визуальной или тактильной оценке. Такая оценка являлась ненадежной и требовала, чтобы вся работа была выполнена одним специалистом. Она практически не позволяет точно описать такие параметры, как количество и соотношение трихом различной длины. Таким образом, существует необходимость более детального исследования, как генетической составляющей, так и влияния условий внешней среды, а также стадий развития на проявление этого адаптивного признака.

Качество зерна, наряду с урожайностью, является наиболее важной характеристикой сортов пшеницы. Это сложный количественный признак, состоящий из множества отдельных показателей, для измерения которых используются различные методики и специальное технологическое оборудование. Большинство этих признаков имеет полигенное наследование, а их выраженность зависит от условий выращивания [135,136].

Разработка цитогенетических методов исследования сложных геномов и создание цитогенетических коллекций мягкой пшеницы позволили оценивать вклад определенных хромосом в отдельные показатели качества зерна и муки. При изучении технологических показателей в наборах межсортовых замещенных линий было обнаружено, что хромосомы различных гомеологических групп вовлечены в генетический контроль данного признака. Результаты также зависели от исходных родительских сортов. Например, в серии межсортовых замещенных линий «Chinese Spring/Cheyenne» было выявлено, что хромосомы 1A, 3A, 3B и 7B, наиболее существенно влияют на определенные параметры качества [137]. При исследовании реципрокных замещенных линий «Cheyenne/Wichita» были выявлены хромосомы 1B и 2D. Использование анеуплоидных серий, таких как дителосомные линии сорта Chinese Spring и моносомные линии сорта Саратовская 29 позволило обнаружить, что отсутствие гомологичной хромосомы или плеча хромосомы влияет на отдельные показатели качества зерна и муки [138].

С помощью генетического материала такого рода была установлена хромосомная локализация генов биосинтеза глиаина и глютеина - запасных белков зерновки пшеницы [139]. Именно они, в основном, формируют клейковину и, обладая различными биохимическими свойствами, во многом определяют ее качество. Эти группы белков контролируются полигенными локусами, расположенными на хромосомах первой (*Gli-1*, *Glu-1* и *Glu-3*) и шестой (*Gli-2*) гомеологических групп. Многими генетическими исследованиями, как с использованием цитогенетических коллекций, так и гибридологического анализа показано, что разнообразие по аллельному составу этих белков связано с различиями по определенным параметрам физических свойств теста [140,141]. К ним относятся сила муки, упругость и растяжимость, определяемые с помощью альвеографа, а также водопоглотительная способность муки, устойчивость теста к замесу и его разжижение, косвенно характеризующие хлебопекарные свойства зерна. На основе этих параметров сорта пшеницы классифицируют как сильные и слабые.

Таблица 1 -Основные признаки, характеризующие качество зерна у пшеницы.

Масса 1000 зерен, г	Определяет размер эндосперма зерновки, из которого будет извлечена мука и процент удаляемых оболочек зерна (отрубей). Таким образом, признак связан с выходом муки при размоле, который в производственных условиях должен быть не менее 70%.
Стекловидность, %	Характеризует консистенцию эндосперма, определяет силу связи между крахмальными гранулами и белками. Признак очень важен для разделения сортов по технологическому назначению на первых этапах уборки зерна. В дальнейшем от него зависит баланс между энергозатратами при размоле (чем тверже, тем больше затраты) и высоким качеством получаемой муки.
Содержание клейковины в зерн, %	Тесно коррелирует с содержанием белка в зерне и, следовательно, определяет питательную ценность зерна и муки. Признак является классифицирующим при отнесение сортов пшеницы к сильным и ценным.
Диаметр частиц муки, микроны	Определяет выход муки из зерна при помоле и пригодность ее либо к хлебопечению, либо для кондитерских изделий. Признак в большой степени определяется генотипом и мало подвержен влиянию среды.
Сила муки, единицы альвеографа	Комплекс свойств, обеспечивающий высокое качество выпекаемого дрожжевого хлеба. Количественное значение признака является классифицирующим при отнесении сорта к типу сильных или ценных. Мука сильного сорта является улучшителем муки слабых сортов и добавляется в них для улучшения хлебопекарных свойств. В значительной мере определяется генотипом
Упругость, мм	Характеристика клейковины и теста, определяющая их способность сохранять физические свойства при расстойке и брожении, высокую газодерживающую способность.
Сбалансированность теста, отношение упругости к растяжимости	Признак определяет эластичность теста, количественный показатель классифицирующий сорта и зерно в целом по торговым классам и пригодности к технологическому назначению.

С помощью межсортовых замещенных линий в хромосоме 5D был обнаружен другой важный ген *Ha* (Hardness of grain), влияющий на мукомольные свойства зерна. Он отвечает за синтез белков пуроиндолинов, которые имеют различный аминокислотный

состав у мягкозерных и твердозерных сортов пшеницы [142]. Этот ген был картирован в коротком плече хромосомы 5D.

Кроме твердозерности мукомольные свойства определяются массой 1000 зерен и их стекловидностью. Первый признак зависит от выполненности и размера зерна пшеницы. Стекловидность вместе с твердозерностью характеризуют силу связи крахмальных зерен с белковыми включениями в эндосперме зерновки и способствуют получению высококачественной муки высших сортов [143]. Кроме этого, важным показателем качества зерна является содержание клейковины, коррелирующее с содержанием белка и определяющее его пищевую ценность [144,145]. Генетический контроль этого признака еще до конца не изучен. Таким образом, качество зерна и муки – комплексный признак, имеющий количественное проявление и определяемый многими генами, многие из которых еще неизвестны.

Создание геномных карт растений, насыщенных молекулярными маркерами, позволяет разделить количественный признак на более простые генетические компоненты – локусы количественных признаков (QTL) и определить их местоположение в геноме. Чем более насыщена карта маркерами, тем эффективнее работа по выявлению QTL. Эта стратегия уже применялась для изучения параметров качества зерна с помощью различных картирующих популяций. Было обнаружено, что некоторые QTL для отдельных технологических параметров совпадают с локусами запасных белков эндосперма и локусом твердозерности *Ha*. Другие были найдены на хромосомах, где ранее не было описано каких-либо структурных генов, ответственных за качество зерна и муки [141,145,146].

Первые карты хромосом для мягкой пшеницы, наиболее плотно насыщенные молекулярными маркерами, были созданы в рамках международного проекта «International Triticeae Mapping Initiative» (ITMI). С ее помощью был изучен целый ряд морфологических, физиологических и стрессозависимых признаков [147-149]. Было подтверждено участие некоторых локусов запасных белков и локуса *Ha* в определении технологических свойств зерна и муки [150].

В таблице 1 приведены основные признаки, характеризующие качество зерна у пшеницы.

### **1.3.3 Анализ подходов к высокопроизводительному фенотипированию растений в рамках селекционно-генетических экспериментов**



Ключевая проблема биологии — изучение взаимосвязи между генотипом и фенотипом у живых организмов. Для этого могут быть эффективно использованы статистические методы выявления взаимосвязи генотип-фенотип. Современные методы секвенирования и генотипирования позволяют быстро и эффективно устанавливать генотип организма и получать информацию о его мутантных аллелях.

Для решения этой задачи применяются компьютерно-экспериментальные подходы, основанные на методе картирования генов, контролирурующих количественные признаки (quantitative trait loci, QTL) [151]. Эти подходы основаны на использовании комбинаций молекулярных (молекулярных ДНК-маркеров и их расположении на хромосоме) и фенотипических данных [152,153].

Другой технологией, позволяющей получать массовым образом данные о наличии маркеров в геноме отдельно взятого растения, является секвенирование нового поколения с использованием оборудования Roche 454, SOLID, Solexa и им подобных [154]. Эти технологии ориентированы на прочитывание коротких последовательностей 30-200 нуклеотидов, но в массовом порядке и с высокой степенью перекрытия, что очень хорошо подходит для определения в последовательности ДНК различных маркеров. Такой эксперимент позволяет картировать до несколько тысяч маркеров в образцах от одного растения за относительно небольшую цену. Таким образом, высокопроизводительное генотипирование (получение большого пула данных по ДНК-маркерам) в анализе взаимосвязи генотип-фенотип сейчас не является проблемой, как это было еще 10-20 лет назад. Узким местом является массовое получение данных по фенотипам растений. Именно поэтому область исследований в биологии, связанная с получением данных о фенотипических признаках растений в массовом порядке (высокопроизводительное фенотипирование) сейчас получает все большее развитие [155].

Современный селекционно-генетический эксперимент использует данные о тысячах и десятках тысяч растений [156,157]. Очевидно, что для выборок такого размера традиционные способы определения большинства фенотипических характеристик (визуальные, тактильные, измерение линейкой и пр.) малоэффективны. Для повышения эффективности решения указанных выше задач в последнее время в мире все более интенсивно используются информационные и телекоммуникационные технологии.

Перед компьютерной феномикой, как и перед любой прикладной областью информатики, возникают три задачи:

- а) получение данных: оцифровка параметров фенотипа растения;

б) хранение данных: разработка методов хранения фенотипических данных и быстрого доступа к ним (так как объем данных может быть чрезвычайно высок);

в) интеграция данных: интеграция полученных данных с информацией о генетических маркерах и геномах растений (необходимая для решения биологических задач по анализу QTL).

д) анализ данных: математическая и статистическая обработка полученных данных и выявления взаимосвязи между генотипом и фенотипом.

Следует отметить, что генотип обуславливает менее 50% вариаций фенотипических данных [153]. Остальная часть фенотипической изменчивости обусловлена изменениями внешней среды, в которой произрастают растения. Именно поэтому полный анализ взаимосвязи генотипа и фенотипа невозможен без учета параметров внешней среды, к которым обычно относят температуру и влажность воздуха, количество осадков, освещенность, характеристики почвы и т.п. Эти данные необходимо так же учитывать при анализе проблем (1)-(4) указанных выше.

Современные методы высокопроизводительного фенотипирования позволяют решать совершенно новые классы задач в биологии растений. В этом разделе мы опишем некоторые успешные применения современных подходов к массовому фенотипированию растений.

Изучая время цветения различных сортов арабидопсиса в различных условиях окружающей среды (полевых и тепличных) и исследуя генетическую природу этого признака с помощью QTL, ученым за два года пришлось фенотипировать более 20000 растений [157]. Фенотипирование проводилось вручную. Несмотря на то, что определение времени цветения и его продолжительность является простым признаком для ручного протоколирования, безусловно, ручное фенотипирование 20000 растений является рутинной и очень временнзатратной задачей. Всё это делает актуальным разработки систем и методов автоматизирующих этот процесс.

Недавно разработанный комплекс PlaRoM (plant root monitoring platform) позволяет осуществлять мониторинг роста корней растений [158]. Система может осуществлять непрерывное наблюдение за развитием корневой системой с высоким пространственным и временным расширением у 50 растений одновременно. PlaRoM позволяет исследовать различные генотипы (сорта) в различных условиях окружающей среды таких как освещенность, температура, питательная среда. Использование технологии позволило зафиксировать различия в динамике роста корней в ночное и дневное время, а так же позволило наблюдать различия в скорости роста при изменении

концентрации различных химических веществ, в частности сахарозы, в питательной среде.

Другой интересный метод позволяет производить оценку биомассы различных злаков [159]. Растение помещается на контрастный одноцветный фон и фотографируется. Компьютерная программа обрабатывает полученную фотографию, отделяет область растения от области фона и подсчитывает её площадь. При этом корреляция между получаемым значением площади и массой растения оказывается достаточно высокой ( $r > 0.96$ ), что доказывает корректность применения данной методики. Преимущества такого подхода заключаются в том, что нет необходимости изымать растение из питательной среды для того, чтобы его взвесить, тем самым мешая его нормальному развитию. Кроме того появляется возможность оценивать биомассу растений в ходе его развития. Подобным способом было фенотипировано тысячи растений.

Для быстрого определения размеров семян у арабидопсиса его семена помещаются под обычный компьютерный сканер, после чего специальное программное обеспечение вычисляет линейные размеры каждого зёрнышка и строит распределение зерен по размеру [160]. Такой метод был использован для получения данных в QTL исследовании. Преимуществами подхода является дешевизна и легкость реализации в любой исследовательской лаборатории.

Получение объективной оценки цвета плодов растений — ещё одна актуальная проблема. Разработанная система TATC (Tomato Analyzer Color Test) производит сканирование разреза плодов фруктов и овощей и даёт точную количественную оценку цвета и монотонности цвета [161]. Для проверки точности был измерен внутренний цвет плодов томата *Solanum Lycopersicum* L. с помощью колориметра и из отсканированных изображений. Оценка показала достоверность получаемых результатов, обусловленную высокой корреляцией ( $r > 0.96$ ). Кроме этого исследователи оценили генотипическую изменчивость связанную со цветом и показали, что доля общей фенотипической изменчивости обусловленной генотипом для цвета и цветовой монотонностью измеренной с помощью TATC была значительно выше чем, когда измерения проводились с помощью колориметра.

Методика массового фенотипирования могут быть с успехом использованы при анализе различных патологий растения, вызванных грибковыми, вирусными или бактериальными заболеваниями. Так используя методы анализа изображений удалось дать количественную оценку заражения листьев сои ржавчиной [162]. Рисунок 1 демонстрирует различную степень зараженности листьев сои, начиная слабым, почти не

видимым для человеческого глаза, заканчивая сильным заражением. Появление подобных методов позволяют давать объективные оценки проявления фенотипических признаков, исключая при этом ошибки вызванные человеческим фактором. Данное исследование было успешно апробировано в лабораторных условиях, но по мнению автором с успехом может проводится и в менее комфортных полевых условиях. Такая возможность в первую очередь обеспечивает колоссальное развитие мобильных устройств, таких как телефоны, смартфоны, карманные компьютеры и планшеты. Исследователь получает возможность, используя цифровую камеру своего мобильного устройства сфотографировать интересующий лист. Затем отправить полученное изображение с помощью технологий мобильного интернета обрабатываться на сервер и получить, в результате, объективную оценку уровня заражения листа ржавчиной. Причем современные технологии позволяют производить подобного рода оценки за считанные минуты или даже секунды. Еще один важный плюс в пользу использования такого подхода — это возможность использования для получения данных менее квалифицированной рабочей силы. Эксперту не обязательно лично идти в поле и собирать данные для своих исследований. Эту работу может выполнить человек не разбирающийся в предметной области, т.к. в роли эксперта теперь выступает программно аппаратный комплекс в виде мобильного устройства и программной реализация оценки признака.



Рисунок 1 - Фотографии, демонстрирующие разную степень зараженности листа сои ржавчиной. Слева на право: сильная, средняя, умеренная, слабая.

В России реализованы подходы в автоматизации исследований в области селекции и генетики растений. Например, в Центре биоинженерии (г. Москва) функционирует современный тепличный комплекс (<http://www.biengi.ac.ru/teplitza.htm>), в котором автоматизирована поддержка условий произрастания растений.

Важной вехой в селекционно-генетических экспериментах на пшенице в СССР и России явился проект ДИАС [163]. Эта была программа скрещиваний 15 сортов яровой

пшеницы (каждого сорта с каждым из других сортов). Она проходила с 1973 по 1983 гг. и включала экологические испытания в течение двух лет родителей и гибридов поколений F1, F2 с измерениями у каждого растения 19-и количественных признаков в 9-и географических точках Сибири. При этом в каждой географической точке проводилась регистрация динамики основных метеохарактеристик среды. В результате исследований была получена информация о более 5 млн. замеров количественных признаков. Данные обрабатывались компьютерными программами на ЭВМ БЭСМ-6 Вычислительного центра СО АН. Программа «ДИАС» дала Западной Сибири и Казахстану 8 новых районированных сортов яровой пшеницы [164]. Выполнение исследований позволило выявить основные закономерности эколого-генетической организации количественных признаков растений.

#### **1.3.4 Базы данных в области генетики и селекции пшеницы**

В последнее время системы информационной поддержки хранения и обработки фенотипических данных и интеграции их с геномами интенсивно создаются для разных организмов. Прежде всего, эти работы ведутся для важных организмов: мыши ([165]; <http://molossinus.lab.nig.ac.jp/phenotype/index.html>), крысы, человека, кукурузы (<http://vphenodbs.rnet.missouri.edu/>). Эти базы данных содержат информацию о фенотипах организмов, в том числе и мутантных, базируются на последовательностях полных геномов, которые секвенированы к моменту создания ресурса, содержат ссылки на аннотации генов, списки маркеров и т.п. геномную информацию.

Подход для поддержки генетических коллекций, основанный на интегрированном описании фенотипа и генотипа растений предложен в базе данных Germinate [166].

Интересным проектом является база данных SGH (<http://solgenomics.net/>), который содержит информацию о фенотипе, генотипе, полногеномных данных и генных и метаболических сетях для пасленовых (к которым относятся томаты, картофель, баклажан, перец и петуния). Эта база данных позволяет производить поиск фенотипа, результатов анализа количественных признаков (QTL), списка маркеров, генов, метаболических сетей. Она тесно интегрирована с геномными данными.

Однако основным недостатком такой системы является то, что она ориентирована на вид, а не на растение. Т.е. для конкретного растения информация в базе данных не представлена. Такой подход характерен и для всех упомянутых выше баз данных. В силу этого, такие системы являются в большей степени информационными (коллекцией результатов, полученных из литературных источников, геномных баз данных и т.п.).

Однако, эти системы не подходят для поддержки генетических экспериментов по селекции важных признаков растения и QTL. Следует отметить, что большинство этих систем используют поддержку хранения информации о фенотипе в виде изображений. Кроме того, методы анализа изображений используются при анализе фенотипа растения все более интенсивно [155,167].

Наиболее перспективный подход – создание систем для поддержки лабораторных экспериментов. Подобным проектом является Chloroplast2010 [156]. Высокопроизводительное фенотипирование, эффективный сбор, хранение большого объема фенотипических данных, их интеграция с геномными данными позволило создать прорывную технологию анализа взаимосвязи между генотипом и фенотипом у *Arabidopsis thaliana* [168]. Однако данная система не позволяет учитывать влияние окружающей среды на развитие фенотипа растения. PlantDB [169] – инструмент на основе Microsoft Access для занесения базовой информации о генотипе и некоторых фенотипических признаках исследуемых растений. Эта база данных, в отличие от предыдущих, ориентирована на описание параметров каждого растения, для которого проводится эксперимент. Однако ее структура не является гибкой и не позволяет расширять описание фенотипа растений. Она так же не позволяет учитывать параметры внешней среды. Другим интересным проектом является система PHENOME для сбора, хранения и анализа данных о фенотипе у томата с использованием карманного компьютера. Исследователь получает возможность быстро описать набор фенотипических признаков растений с помощью специального приложения его карманного компьютера, а затем синхронизировать полученные данные с сервером баз данных [170]. Использование мобильных устройств позволяет существенно повысить эффективность решения задач в области селекционно-генетических экспериментов, особенно для полевых наблюдений. Подход для поддержки генетических коллекций, основанный на информационной поддержке селекционно-генетических экспериментов у растений предложен в базе данных Germinate [166].

Современные подходы направлены на автоматизацию фенотипирования, которая позволяет существенно ускорить процесс сбора данных о важных характеристиках растений, повысить точность их оценок, измерять новые параметры растений, а так же исключить субъективизм человеческой оценки из процесса измерения [171]. В мире так же разрабатываются интегрированные системы автоматизации теплиц. К таким системам относятся комплексы компании Lemnatec ([www.lemnatec.com](http://www.lemnatec.com)) для проведения высокопроизводительного фенотипирования у растений, система Phenopsis для анализа

модельного растения *A. thaliana* ([172]; <http://bioweb.supagro.inra.fr/phenopsis/Accueil.php?lang=En>), система Phenofab ([www.phenofab.com](http://www.phenofab.com)).

Ранее для решения задачи интеграции генотипических и фенотипических данных, а также параметров окружающей среды и для анализа взаимосвязей между генотипом и фенотипом нами была разработана система WheatPGE [173]. Система служит для интеграции разнородных данных о растении, хранении и доступе к информации об отношениях, описывающих различные характеристики растения, его генотипа, фенотипа и факторов внешней среды. Система имеет простой и удобный web-интерфейс и доступна по адресу [www.wheatdb.org](http://www.wheatdb.org). Как и метод автоматического определения опушения система прошла свою апробацию на пшенице, однако является универсальной и с успехом может применяться для других видов растений.

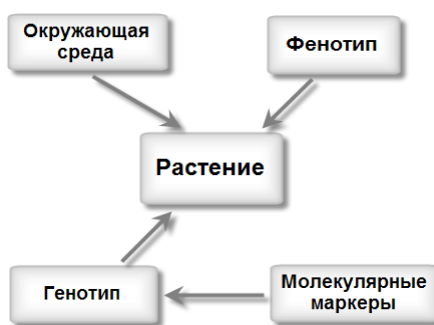


Рисунок 2 - Схема взаимосвязи между основными разделами информации в системе WheatPGE.

Центральным объектом базы данных является растение (рис. 2). Растение описывается как совокупность признаков генотипа, фенотипа и окружающей среды, в которой данное растение произрастает.

Описание генотипа растения содержит следующую информацию: сорт растения, линия (в случае, если растение из чистой линии) или родители (в случае, если растение — гибрид). Для родителей указываются ссылки на генотипы соответствующих растений. Дополнительно для гибридов можно указать поколение и материнское растение. Для генотипа можно определить список молекулярных маркеров (характеристик геномных ДНК, которые определяются экспериментально или могут быть импортированы из

других баз данных). Маркеры объединяются в группы. Для каждого маркера из группы определяется его состояние (например, молекулярная масса или длина). Группа маркеров является характеристикой генотипа растения (рисунок 3). При описании маркера указывается его тип, имя, список состояний и локализация на хромосоме.

Для описания фенотипа растения система WheatPGE позволяет создавать наборы отношений, каждое из которых содержит описание характеристик определенного морфологического признака (опушение листа, длина побега и колоса, количество колосьев, продуктивность и т.п.)

В текущей версии базы данных наиболее полно представлено описание такого морфологического признака как опушение. Для него заданы следующие характеристики: плотность опушения (количество ворсинок (трихом) на единицу площади), вектор распределения трихом по длине. Система позволяет сохранять оцифрованные изображения морфологического признака, если это необходимо. Интерфейс для описания признака позволяет также подключать внешние программы анализа изображения для получения различных его характеристик, например, для получения информации о морфологических характеристиках опушения на основе анализа цифровых фотографий была использована программа LHDetect2. Структура базы данных позволяет легко расширять список анализируемых морфологических признаков растения и модифицировать информацию о них.

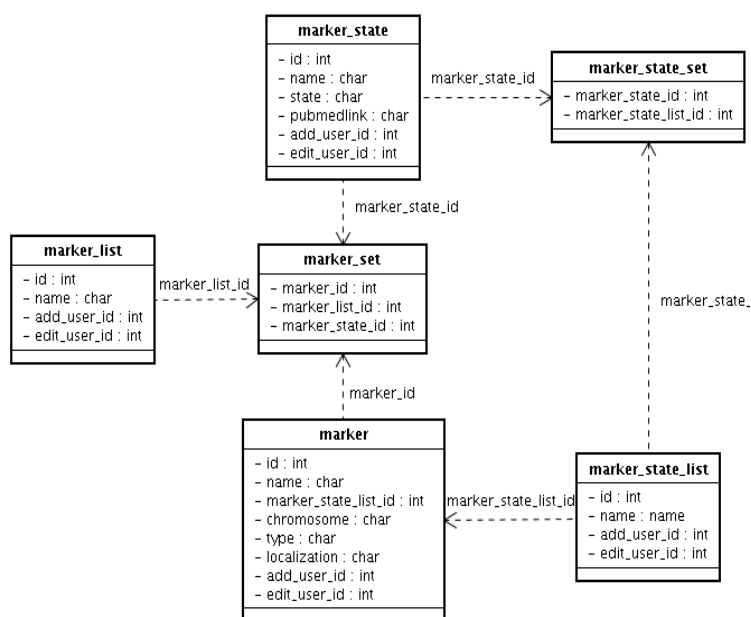


Рисунок 3 - Структурная схема реляционных отношений таблиц, описывающих молекулярные маркеры.



Подобно фенотипу WheatPGE позволяет расширять схему базы данных, добавляя произвольные параметры окружающей среды. Окружающая среда в базе данных может быть представлена набором таких характеристик как место произрастания (теплица или открытый грунт), средняя температура и количество осадков за сезон, дата посева семян и т. п.

Для описания различных характеристик растений пшеницы нами была спроектирована реляционная база данных, которая лежит в основе системы WheatPGE и содержит более 23 таблиц, связанных между собой. В качестве сервера используется MySQL. Для работы с базой данных разработан web-интерфейс, реализованный на основе модуля Catalyst — свободного кроссплатформенного программного каркаса, написанного на языке Perl. В Catalyst заложена методология разработки программного обеспечения MVC, в которой модель данных приложения, пользовательский интерфейс и управляющая логика разделены на три отдельных компонента. В результате модификация одного из компонентов оказывает минимальное воздействие на другие компоненты. Это позволяет добиться эффективной масштабируемости системы. Для связи базы данных с Catalyst используется технология ORM (объектно-реляционная проекция) – технология программирования, которая связывает базы данных с концепциями объектно-ориентированных языков программирования, создавая «виртуальную объектную базу данных». Технология позволяет связывать таблицы базы данных с объектами реального мира, например, объект генотип состоит из 9 связанных таблиц.

Важная особенность нашей системы — возможность для пользователя описывать произвольные морфологические признаки и параметры окружающей среды без помощи программиста. При этом происходит автоматическое расширение схемы базы данных, создается новая модель, описывающая объекты этого признака. Генерируются контроллеры и представления, реализующие базовые возможности работы с признаком (создание, удаление, редактирование). Этот подход имеет существенное ограничение. Семантическое описание нового признака ограничено одним реляционным отношением. Это означает, что описание должно укладываться в одну таблицу базы данных. Тем не менее этого оказывается достаточным, для описания большинства морфологических признаков и параметров окружающей среды, с которыми имеют дело экспериментаторы.

При занесении в базу большого количества гибридных генотипов становится актуальной задача визуализации схем скрещивания растений. Система WheatPGE позволяет автоматически визуализировать схемы скрещивания растений на основе

информации об отношениях родитель-потомок, которые хранятся в базе данных. Схема представляется в виде ориентированного графа. Для размещения графа на плоскости и его рендеринга используется библиотека GraphViz (рис. 4).

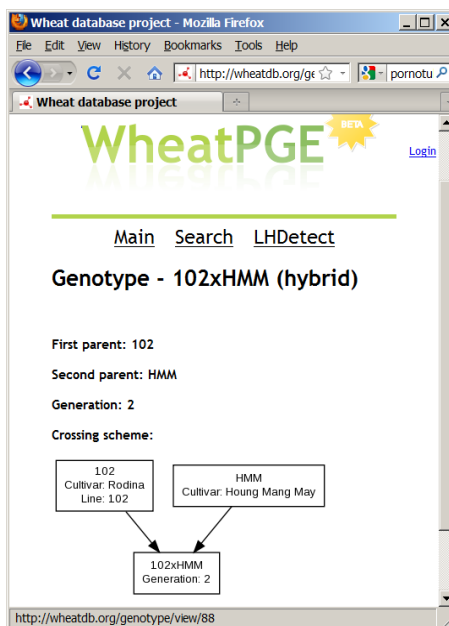


Рисунок 4 - Пример визуализации схемы скрещивания для гибридного генотипа.

Интерфейс системы WheatPGE реализован на основе сервера Apache с модулем mod\_perl под управлением операционной системы CentOS Linux.

Пользователь может получить доступ к базе данных зарегистрировавшись на сайте [www.wheatdb.org](http://www.wheatdb.org). Зарегистрированный пользователь имеет возможность добавлять и аннотировать собственные растения. В каждой таблице базы данных содержатся поля, в которых прописывается идентификационный номер пользователя создавшего запись в таблице и идентификационный номер пользователя отредактировавшего запись в таблице.

Если для работы пользователю требуется возможность аннотировать дополнительные морфологические признаками или параметры окружающей среды ему следует отправить запрос администратору системы с просьбой на расширение модели базы данных.

Пользователю предоставляется возможность просматривать списки генотипов, молекулярных маркеров, параметров окружающей среды и отдельные экземпляры растений, которые содержатся в базе. Кроме этого пользователь имеет возможность осуществлять поиск по растениям, которые содержатся в базе.

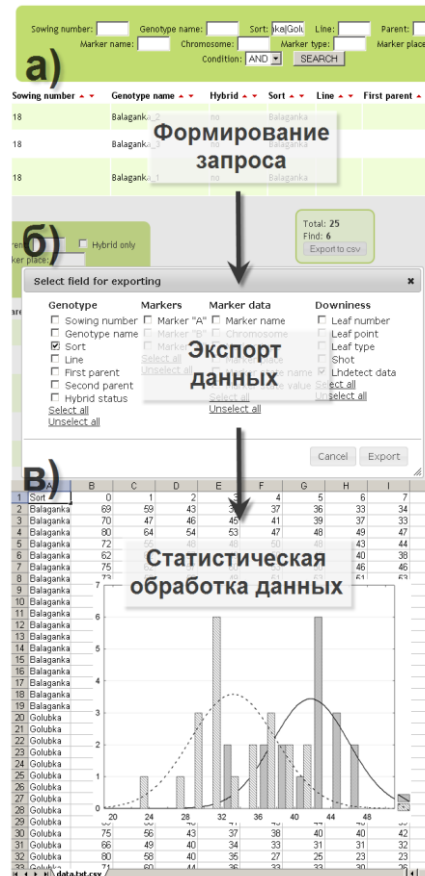


Рисунок 5 - Пример работы с системой WheatPGE. Извлечение данных о взаимосвязи сорта растения и его опушения: (а) Интерфейс формирования запроса отбора растений; (б) Выбор полей базы данных для экспорта в таблицу; (в) вид таблицы в Excel и статистический анализ распределения плотности опушения для сортов Балаганка и Голубка.

Поиск производится по следующим полям:

- посевной номер растения;
- название генотипа растения;
- сорт растения;
- линия;
- является ли растение гибридом или нет;
- название родительского генотипа;
- название молекулярных маркеров, которые присвоены генотипу растения;
- хромосома, на которой локализован молекулярный маркер;
- тип молекулярного маркера;
- положение молекулярного маркера на хромосоме.

При формировании запроса допустимо использование регулярных выражений, например, если необходимо найти в базе все растения двух сортов *Fora* и *Krasa* запросе достаточно написать *Fora/Krasa*.

Результаты любого запроса можно экспортировать в формате CSV с целью их дальнейшего анализа. При экспорте можно указать поля, необходимые для анализа. Экспортировать можно информацию о морфологических признаках растений, молекулярных маркерах и параметрах окружающей среды. Например, для анализа зависимости опущения листа от сорта растения пользователь должен на странице запроса указать список сортов растений, см. рисунок 5(а), которые он хотел бы включить в анализируемую выборку, и указать список характеристик опущения, как показано на рисунке 5(б). В итоге, пользователь получает таблицу данных, в которой строкам соответствуют растения отобранных сортов, представленные в базе, а в колонках приводятся числовые характеристики опущения, как это изображено на рисунке 5(в). Такая таблица может быть далее проанализирована любой программой статистического анализа (Excel, Statistica и другие).

Разработанная система позволяет устанавливать и анализировать взаимосвязь между генетическими и фенотипическими признаками растений и параметрами окружающей среды. Это обеспечивает решение целого ряда важных биологических задач. Например, исследование зависимости морфологических характеристик опущения листа от сорта растения, от места произрастания, поиск генетических маркеров, статистически связанных с тем или иным типом опущения пшеницы и т.п.

В текущем виде, однако, данная система требует расширения полей описания хозяйственно важных признаков растений, включая базовые, касающиеся продуктивности, и качества зерна. Для удобства занесения информации в систему так же представляется необходимым разработать хранение в системе описание селекционно генетических экспериментов.

### **1.3.5 Методы анализа изображений для фенотипирования растений**

Одним из подходов, позволяющих существенно ускорить фенотипирование, является использование анализа цифровых изображений. Они успешно применяются для оценки биомассы растения [159,174], для выявления генов, ответственных за размер зерна у *A. thaliana* [160], для анализа морфологии и развития корня у *A. thaliana* [158,175] и риса [176], в задачах колориметрии у томатов [161] и др. Методы компьютерного анализа были использованы и для анализа морфологии опущения листа [167,177].

В области обработки и анализа медицинских и биологических изображений используются шесть основных классов, методов и алгоритмов: улучшение качества, сегментация, количественный анализ, совмещение, сжатие и обеспечение хранения в базах данных, виртуализация [178].

Улучшение качества. Методы данного класса используются для уменьшения шумов, удаления артефактов, реформатирования и интерполяции, компенсации пространственных искажений и повышения контраста области интереса на изображении. Эффективность указанных методов особенно критична в случаях, когда различие между изображениями анализируемых объектов и тканей в норме и при различных видах патологии являются незначительными.

Сегментация. Целью сегментации изображения является отделение анализируемого объекта, структуры или области интереса от окружающего фона [179]. В отличие от эксперта, у которого визуальная сегментация не представляет особой трудности, автоматическое выделение объектов на биологических изображениях является как правило весьма сложной задачей. Сегментация принадлежит к числу базовых шагов, качество выполнения которых во многом определяет точность, а порой даже саму возможность дальнейшего компьютерного анализа изображения. Методы сегментации базируются на яркостной, градиентной и текстурной информации изображения и используют такие алгоритмы как бинаризация по порогу, морфологическая фильтрация, наращивание областей, активные контуры, деформируемые шаблоны и др. При этом могут привлекаться методы распознавания образов и математической статистики, включая кластерный анализ, многомерное шкалирование, нейронные сети и линейный дискриминантный анализ выборок вокселей.

Сегментация изображений может быть построена с использованием порога. Разделение на классы производится методом сопоставления значения яркости каждого пикселя изображения с заданным значением порога. Приведем краткое описание методов сегментации, которые наиболее часто используются при анализе изображений.

Методы основываются на энтропии, кластеризации, гистограммах, локальных порогах и пиксельной корреляции.

Наиболее распространенным методом локальной пороговой обработки изображения является метод Отса (Otsu). Метод основывается на анализе гистограммы распределения значения яркости пикселей изображения. Алгоритм предполагает наличие двух классов пикселей на изображении и ищет подходящий порог разделения этих классов так, чтобы дисперсия внутри каждого класса была минимальной.

Построение гистограммы производится по значениям отношения количества пикселей с уровнем яркости  $i$  к общему количеству пикселей изображения. Деление диапазона яркостей на два класса производится по пороговому значению яркости  $t$ , где  $t$  - целое значение от 0 до  $L$  (255 после нормализации изображения). Межклассовая дисперсия определяется как взвешенная сумма дисперсий обоих кластеров-классов

$$\sigma_w^2 = \omega_1 \sigma_1^2 + \omega_2 \sigma_2^2 \quad (1)$$

Веса  $\omega_i$  - это вероятности двух классов разделенных порогом, а  $\sigma_{2i}$  - дисперсия этих классов. Оцу показал, что минимизация внутриклассовой дисперсии эквивалентна максимизации межклассовой дисперсии, которая выражается формулой

$$\sigma_b^2 = \omega_1 \omega_2 (a_1^2 - a_2)^2, \quad (2)$$

где  $a_i$  - средние арифметические значения для каждого из классов. Особенность этой формулы заключается в том, что  $\omega_1(t + 1)$ ,  $\omega_2(t + 1)$ ,  $a_1(t + 1)$ ,  $a_2(t + 1)$  легко выражаются через предыдущие значения  $\omega_1(t)$ ,  $\omega_2(t)$ ,  $a_1(t)$ ,  $a_2(t)$  ( $t$  — текущий порог). Эта особенность позволила разработать быстрый алгоритм поиска оптимального порога.

Метод Ниблэк используется на практике для быстрой фильтрации контрастных изображений, на которых отсутствуют сильно зашумленные области с плавными переходами яркости. Идея данного метода состоит в варьировании порога яркости  $T$  бинаризации от точки к точке на основании локального значения стандартного отклонения. В местах плавного перехода яркости метод дает ложные объекты с небольшим шумом. Метод получил свое распространение на практике благодаря его интеграции с этапом постпроцессинга. При этом скорость обработки падает в 3 раза, и количество ошибок сокращается на 20% [180].

Метод Саувола является улучшением метода Ниблэка, предотвращающий наложение шума на объект и дающий более точное отделение объекта от фона. Согласно данному методу изображение обрабатывается с помощью концентрического окна с радиусом  $R$ . Обычно форма окна принимается квадратной. Оно последовательно слева направо сверху вниз накладываются на изображение с шагом равным диаметру. Метод Sauvola может превзойти по быстродействию Niblack применяя только для обработки четких и контрастных изображений, но есть трудности с изображениями, у которых зафиксировано недостаточно освещения, особенно когда значения пикселей объекта находятся близко друг к другу. При обработке тонких пересекающихся линий могут возникать разрывы, поэтому метод хорош для толстых линий и крупных объектов [181].

При определении морфологических характеристик фенотипических признаков по анализу изображений часто приходится иметь дело с задачами анализа формы и размеров

объектов. Кратко опишем основные геометрические характеристики объектов и способы их определения.

Площадь и периметр изображения объекта. Площадь  $S$  изображения объекта вычисляется путём подсчёта числа пикселей изображения, относящихся к объекту. Периметр изображения объекта  $P$  вычисляется после выделения границ объекта путем суммирования множество контурных точек изображения объекта.

Определение радиусов вписанной ( $R_{\min}$ ) и описанной ( $R_{\max}$ ) окружностей складывается из двух этапов. Вначале определяются координаты геометрического центра изображения объекта  $X_{\text{ц}} = xS(x, y)/S(x, y)$ ,  $Y_{\text{ц}} = yS(x, y)/S(x, y)$ , где  $x$  и  $y$  – номера строк и столбцов всех пикселей  $S(x, y)$ , входящих в объект. Затем вычисляются минимальные и максимальные расстояния от центра до границ изображения объекта:  $R_{\max} = r(x, y_{\max})$ ;  $R_{\min} = r(x, y_{\min})$ , где  $x, y$  являются точками периметра. Нормированный признак  $R' = R_{\max}/R_{\min}$  инвариантен к масштабу изображения объекта.

Момент инерции объекта. Термин "моменты инерции изображения объекта" не имеет отношения к механике. Просто для вычисления указанного признака используются математические выражения, аналогичные вычислению моментов инерции материального тела, где вместо значений масс отдельных точек тела подставлены значения освещенностей в соответствующих точках его изображения. Моменты инерции являются информационными признаками для последующего распознавания образов.

Количественный анализ. Методы квантификации обычно применяются к отсегментированным объектам и структурам биологических изображений с целью выделения существенной информации об их размерах, форме, текстуре, морфологии и особенности динамики во времени.

Например, размер и форма мышечных волокон является важным для диагностики нервно-мышечных заболеваний. Было показано, что морфометрические данные образцов биопсии мышц могут на ранней стадии выявить отклонения в характере распределения и размерах волокна. В тоже время ручная сегментация и подсчет волокон требует много времени и является утомительным. Ученым из университета Бонна (Германия) удалось автоматизировать этот процесс, разработав метод для сегментации мышечной ткани и определяя количество и форму содержащихся в ней волокон [182].

Примером изменения особенностей объекта во времени может служить работа, отслеживания моторики поведения червя *Caenorhabditis elegans* [183]. В работе в режиме реального времени осуществлялась обработка цифрового потока, получаемого с камеры, установленной на микроскоп. Программа на основе библиотеки OpenCV быстро, 50

кадров с секунду, анализирует каждый кадр потока. Программа определяет местоположение клеток-мишеней и поручает цифровому микрозеркальному устройству сигнал для освещения клеток-мишеней лазерным светом соответствующей длины волны, чтобы стимулировать или ингибировать активность.

**Совмещение.** Совмещение двух цифровых изображений одного и того же объекта является важным в случае, когда совмещенная карта соответствия может быть использована для последующей обработки или анализа изображения. Совмещаемые изображения могут представлять собой снимки одного и того же объекта, полученные различными путями или снятые в разное время. Например, в процессе диагностики может возникнуть необходимость в совмещении изображений пациента с контрольным изображением здорового человека либо с цифровым анатомическим атласом того или иного органа. Получаемая карта соответствия может использоваться для попиксельного сравнения изображений, оценки изменения формы и мониторинга роста новообразований. Другим примером использования методов совмещения является сегментация изображения. Например, после установления попиксельного соответствия МРТ изображения головного мозга с предварительно размеченным электронным атласом можно автоматически выделить те или иные анатомические структуры анализируемого изображения, используя атлас как шаблон.

**Сжатие, архивирование, хранение и поиск в базах данных.** В последние годы в связи со значительным ростом количества и размеров снимаемых цифровых биологических и медицинских изображений роль этого направления постоянно возрастает. Задачи данного типа носят преимущественно прикладной характер и традиционно решаются в рамках так называемых систем архивирования и передачи медико-биологических изображений PACS (picture archiving and communication system), которые активно используют современные технологии баз данных, компьютерных сетей и грид-систем. Одной из наиболее сложных и наукоемких проблем данного класса является поиск изображений по их содержанию, заданному изображением-образцом. Решение проблемы поиска в базах изображений по образцу призвано обеспечить компьютерную поддержку такому распространенному диагностическому приему как сравнение полученного изображения с предыдущими и/или с похожими снимками и случаями из клинической практики, информация о которых хранится в цифровых архивах.

**Виртуализация.** Используемые здесь методы и алгоритмы развиваются на стыке трехмерной компьютерной графики, систем компьютеризированной диагностики, а



также различного рода тренажеров и образовательных систем типа виртуальной операционной, базирующихся на концепции погружения в виртуальную реальность. Примером одной из достаточно простых, но широко известных систем данного класса является интернет-проект виртуального человека ([http://www.nlm.nih.gov/research/visible/visible\\_human.html#](http://www.nlm.nih.gov/research/visible/visible_human.html#)), поддерживаемый национальной медицинской библиотекой США. Целью проекта является создание полного атласа внутренних органов и всего тела здорового мужчины и женщины в виде трехмерных моделей.

Другим примером визуализации может служить программно аппаратный комплекс служащий для определения количественных характеристик трихом листьев арабидопсиса. Растение помещается в ёмкость цилиндрической формы, где с помощью технологии компьютерной микротомографии (micro x-ray computer tomography) происходит сканирование растения и построение его трёхмерной модели. После построения трёхмерной модели с помощью специализированного программного обеспечения производится её анализ: выделение листьев и определение на них трихом [167].

#### **1.4 Выводы и предложения по результатам анализа информационных источников**

Нами был проведен систематический анализ доступной информации о проведении НИР в областях биотехнологии, требующих использования методов генной инженерии или формализованного описания селекционно-генетических экспериментов (результаты в кратком виде представлены в разделе 1). Для анализа литературы нами была использована система AND-Visio [184].

Анализ показал, что в настоящее время применение информационных технологий при разработке генетически модифицированных организмов все более интенсивно используется как в России, так и за рубежом. Создаются базы данных по генетическим конструкциям, регуляторным районам растений. В этой области особенно важную роль приобретают биоинформатические ресурсы.

В этой задаче интенсивно используются в основном биоинформатические ресурсы и базы данных. Поэтому биоинформатика (комплекс компьютерных методов для обработки биологических данных, поиска закономерностей и разработки технологий предсказания), системная биология (компьютерное моделирование генных и метаболических сетей на основе анализа количественных данных) и синтетическая

биология (разработка системы генетических модификаций, позволяющих изменить фенотип организма в заданном направлении – вплоть до создания искусственных живых организмов) считаются основными инструментами качественного развития биотехнологии. Благодаря развитию этих методов совершенствуются методы поиска регуляторных последовательностей в геноме.

В целом анализ литературных и других источников показал, что в мире существуют десятки информационных ресурсов в области биотехнологий, которые могут служить в качестве основания и исходных данных для разработки нашей темы. Однако основным недостатком на современном этапе является их разрозненность.

В агrobiотехнологиях информационные технологии используются для широкомасштабных экспериментов по исследованию взаимосвязи фенотип-генотип у растений. Прежде всего это базы данных, описывающих отношения генотип-фенотип-окружающая среда у растений, а так же системы автоматического контроля за произрастанием растений в теплицах. При анализе фенотипов упор делается на проведение массовых экспериментов, в обработку и хранение которых все более интенсивно используются методы анализа изображений, мобильные компьютеры и устройства.

Анализ показал, что одной из важнейших сельскохозяйственных культур для России является пшеница. На ее исследованиях сосредоточено большое количество ресурсов, как в России, так и в мире. Основными задачами является выяснение взаимосвязи генотип-фенотип-окружающая среда.

## **2 Обоснование выбора направления и методики исследований**

Круг задач, стоящих перед биотехнологией, очень широк и для проведения НИР в этой области необходимо использовать большое число разнообразных методов в различных областях молекулярной биологии, генетики, физиологии и т.п. Создание универсального информационного ресурса, содержащего специально разработанные оригинальные базы данных и программы, представляется невозможным вследствие их большого числа, технической сложности и разнообразия. Возможным путем для решения этой задачи может быть информационный портал, содержащий специализированные модули, посвященные конкретным направлениям биотехнологических НИР, а также тьюториалы с описанием внешних информационных ресурсов в этих областях. Модульная структура позволяет разрабатывать новые комплексы баз данных и программ по мере необходимости. Описание внешних информационных ресурсов (и соответствующие ссылки) также целесообразно формализовать в виде базы данных, что позволит проводить эффективный поиск подходящих для конкретного случая вариантов.

Для выполнения биотехнологических НИР с помощью методов генной инженерии целесообразно использовать ряд имеющихся в открытом доступе баз данных и программных комплексов. В то же время, существующие информационные ресурсы в большинстве случаев неспециализированы и аннотированная в них информация не оптимизирована для эффективного планирования такого рода. В частности, в литературе содержится большое количество экспериментальных данных о нуклеотидных последовательностях ДНК, обладающих промоторной активностью или функционирующих в качестве трансляционных энхансеров (при расположении в составе некодирующих районов мРНК). Считается, что использование подобных элементов в составе генетических конструкций может существенно расширить арсенал генной инженерии. Эта задача может быть решена с помощью специализированных баз данных промоторов и трансляционных энхансеров. В рамках проекта нами запланирована разработка двух БД, формат которых представлен в разделе 4.2 настоящего отчета.

Анализ доступной информации в области агrobiотехнологии показал, что одной из актуальных задач является анализ взаимосвязи отношений генотип-фенотип-окружающая среда у пшеницы. Учет этих отношений позволит существенно повысить продуктивность селекционно-генетической работы. Наиболее перспективным в этой связи является разработка методов массового фенотипирования, интеграция полученных данных по фенотипам растений с данными по генотипам и

параметрами окружающей среды. Представляется наиболее оптимальным – использовать в качестве основы проекта в этом направлении систему WheatPGE и для решения задач, поставленных в проекте проводить ее расширение.

В области подходов к массовому фенотипированию растений важное значение будут иметь методы анализа изображений. Эти методы позволят существенно автоматизировать процесс фенотипирования и повысить его точность.

С учетом существующей на настоящий момент структуры базы данных WheatPGE представляется актуальным расширить набор признаков растений за счет признаков, характеризующих продуктивность и качество зерна.

С учетом существующей на настоящий момент структуры базы данных WheatPGE представляется актуальным дополнить ее структуру описанием методик проведения селекционно-генетического эксперимента.

С учетом необходимости обеспечения доступа к селекционно-генетическим ресурсам пшеницы в полевых условиях, представляется необходимой реализация возможности взаимодействия с системой через мобильные устройства. Эта опция будет незаменима при работе с базой данных из разных мест культивирования растений.

## **2.1 Выводы и предложения по выбору направления и методики исследований**

На основе аналитического обзора литературных данных, патентных исследований и Интернет-ресурсов сделан вывод об актуальности выбранного направления исследований (информационная поддержка экспериментов по генной инженерии и высокопроизводительному фенотипированию для селекционно-генетических экспериментов), а также предложена методика исследований (использование компьютерных методов для разработки экспериментального образца Интернет-портала, содержащего специализированные базы данных)

### **3 Обоснование выбора платформ, аппаратных средств, средств разработки программ, способов представления данных**

Способ представления данных (информационный Интернет-портал) обусловлен спецификой поставленной задачи, в рамках которой необходимо как разработать специализированные оригинальные базы данных, так и систематизировать ссылки на внешние Интернет-ресурсы. В составе ЭОИП будут содержаться база данных внешних информационных ресурсов (разработка формата которой запланирована на следующем этапе реализации проекта), база данных промоторов для трансгенеза, база данных трансляционных энхансеров для трансгенеза (формат представлен в разделе 4.2 настоящего отчета).

В качестве платформы для выполнения этих трех БД планируется использовать систему Sequence Retrieval System (SRS) [185], специально разработанную для формализованного описания биологических данных по заказу European Bioinformatics Institute (большая часть баз данных EMBO расположена на этой платформе, см. <http://www.ebi.ac.uk/Databases/>). Средства SRS позволяют индексировать большинство полей в карточках блоков и эффективно осуществлять перекрестную связь полей в блоках БД, что необходимо для построения эффективных пользовательских запросов и свободной навигации между полями и карточками в различных блоках. У нас имеется положительный опыт использования SRS для разработки биологических информационных ресурсов. В нашем институте разработан целый комплекс баз данных на этой платформе. К числу этих баз относятся базы данных по регуляторным районам генов [186,187], генным сетям [188], лидерным районам мРНК и т.п. Экспериментальный образец информационного портала «Биотехнология растений» (ЭОИП БР) будет разработан с помощью технологий Web. Выбор обусловлен тем, что такой интерфейс является в высокой степени кросс-платформенным, от персональных компьютеров до мобильных устройств. Возможность использования мобильных устройств является в нашем случае немаловажной, поскольку обеспечит доступ к базе в полевых условиях. Web-интерфейс позволит минимизировать усилия по его разработке для разных платформ и устройств. В нашей работе мы планируем использовать методологию разработки программного обеспечения MVC [189], поддержку которой обеспечивает Catalyst — свободный кроссплатформенный программный каркас для создания web-приложений, написанного на языке Perl (<http://www.catalystframework.org/>). MVC разделяет модель данных, пользовательский интерфейс и управляющую логику приложения на три отдельных компонента. В результате модификация одного из

компонентов оказывает минимальное воздействие на другие компоненты. Такой подход позволяет добиться существенного снижения трудозатрат при работе со слабоструктурированной предметной областью, какой является биология. Важная особенность разрабатываемой системы — возможность для пользователя описывать произвольные морфологические признаки и параметры окружающей среды без помощи программиста. При этом происходит автоматическое расширение схемы базы данных, создается новая модель, описывающая объекты этого признака. Генерируются контроллеры и представления, реализующие базовые возможности работы с признаком (создание, удаление, редактирование).

Web-интерфейс так же будет выполнять функции интеграции базы данных и различных методов массового фенотипирования растений. На практике это будет означать, что исследователь может загрузить в базу, например, фотографии некоторого морфологического признака и получить на выходе числовую оценку характеристик этого признака.

База данных должна будет хранить описание объекта со сложной структурой. Так, например, объект «генотип» состоит из множества подобъектов и в свою очередь может являться свойством объекта «растение». С другой стороны требуется обеспечить должную надежность, производительность и масштабируемость. Для обеспечения легкого расширения схемы базы данных в случае необходимости и представление в базе данных различных комплексных объектов и отношений между ними будет использована технология реляционной СУБД в связке с технологиями объектно-реляционного отображения (ORM – object relation mapping). Использование реляционной модели представления данных обеспечит надежность и производительность, а объектное представление обеспечивает эффективный доступ к данным и масштабируемость системы. Для объектно-реляционного мы используем библиотеку DBIx (<http://search.cpan.org/dist/DBIx-Class/>) которая обеспечивает работу с базой данных через стандартизованный объектно-ориентированный интерфейс. В качестве СУБД может быть использована любая популярная СУБД, поддерживаемая DBIx. Мы остановили свой выбор на MySQL (<http://www.mysql.com/>).

Одно из важных направлений работы – реализация возможности взаимодействия с системой через мобильные устройства. Мы выбрали поддержку работы с устройствами, работающих под управлением ОС Android. Этот выбор обусловлен, прежде всего, относительной дешевизной устройств для этой платформы (производители HTC, Samsung и др.) по сравнению с более дорогими аналогичными устройствами (например,

iPhone), распространенностью их в России. Для того чтобы минимизировать трудозатраты в процессе работы над интерфейсом с мобильными устройствами, мы будем ориентироваться, прежде всего, на работу с Интернет версией портала. Это устранит проблемы кросс-платформенного портирования и позволит быстро реализовать базовые функции работы с системой. Такой подход позволит в принципе сделать систему доступной и с устройств других платформ, а в дальнейшем может быть легко расширен для повышения удобства пользователя за счет функций мобильных устройств.

### **3.1 Выводы и предложения по разделу**

На основе аналитического обзора литературных данных, патентных исследований и Интернет-ресурсов выбраны методы исследований, основанные на технологиях баз данных, Web и анализа изображений. Предложено использовать СУБД SRS, MySQL, для разработки программного обеспечения - Catalyst, Perl, для работы с устройствами мобильного доступа – ОС Android.

## **4 Теоретические исследования**

**4.1 Перечень внешних Интернет-ресурсов (баз данных и программных комплексов), связанных с процессом получения и использования в биотехнологии ГМО, а также с селекционно-генетическими подходами в агробиотехнологии**

**4.1.1 Внешние Интернет-ресурсы, связанные с процессом получения и использования в биотехнологии ГМО**

В рамках проекта был проведен поиск доступных внешних Интернет-ресурсов, которые могут быть использованы для решения задач биотехнологических НИР. Предварительный перечень таких ресурсов приведен ниже в виде формализованного описания по следующей схеме:

1. Название ресурса.
2. Расположение ресурса.
3. Перечень вариантов возможного использования, доступных на момент написания тьюториала.
4. При наличии такой возможности – пример применения данного ресурса для решения какой-либо биотехнологической задачи.

### **1. PlantCyc**

2. <http://plantcyc.org/>
3. База данных метаболических путей 350 видов растений.
4. может использоваться для планирования экспериментов по модификации метаболических путей в заданном направлении

### **1. MetaCrop**

2. <http://metacrop.ipk-gatersleben.de>
3. База данных метаболических путей хозяйственно-ценных видов растений.
4. может использоваться для планирования экспериментов по модификации метаболических путей в заданном направлении

### **1. GenBank**

2. <http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>



3. Данный ресурс может быть использован для поиска последовательностей (генов, промоторов и т.д.) по названию гена и номеру входа в базу.
4. является базовым источником данных о нуклеотидных последовательностях генов.

#### 1. EMBL Nucleotide Sequence Database

2. <http://www.ebi.ac.uk/embl/>

3. Данный ресурс может быть использован для поиска последовательности по названию гена.

4. Релиз базы от 14.09.2011 содержит 220504007 последовательностей. БД расположена на платформе SRS, что позволяет эффективно использовать систему запросов к различным элементам нуклеотидной последовательности на основе информации в Feature Table. Например, используя расширенный поиск в EMBL-Bank с использованием SRS таблицы по адресу <http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+query+-libList+EMBL>, можно найти нуклеотидную последовательность промотора гена, используя название организма и ключевое слово.

#### 1. EBI Dbfetch - database fetch

2. <http://www.ebi.ac.uk/Tools/dbfetch/emblfetch>

3. Dbfetch обеспечивает легкий способ получения записей из различных баз данных в EBI (например, последовательности промотора), используя регистрационный номер (accession number), который приведен в публикации.

#### 1. Webcutter

2. <http://users.unimi.it/camelot/tools/cut2.html>

3. С помощью этого ресурса можно в полной нуклеотидной последовательности найти сайты рестрикции определенных нуклеаз.

#### 1. PlantPromDB – A Database of Plant Promoter Sequences

2.

<http://linux1.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom>

3. Ресурс содержит нуклеотидные последовательности промоторов растений с экспериментально проверенным стартом транскрипции.

### **1. RARGE- RIKEN Arabidopsis + Genome Encyclopedia Promoter Search**

2. <http://rarge.psc.riken.jp/cdna/promoter/index.pl>

3. Ресурс позволяет найти гены и промоторы растений, у которых присутствуют определенные сайты связывания транскрипционных факторов.

### **1. AGRIS - Arabidopsis Gene Regulatory Information Server**

2. <http://arabidopsis.med.ohio-state.edu/>

3. Ресурс содержит информацию о последовательностях промоторов арабидопсиса, транскрипционных факторах и их целевых генах. Один из трех модулей базы, AtcisDB содержит около 330 вышележащих районов аннотированных генов арабидопсиса с описанием экспериментально проверенных и предсказанных цис-регуляторных элементов.

### **1. Osiris - an integrated promoter database for *Oryza sativa* L.**

2. <http://www.bioinformatics2.wsu.edu/cgi-bin/Osiris/cgi/home.pl>

3. Ресурс позволяет получать и визуализировать данные о промоторах риса (*Oryza sativa japonica* strain).

### **1. Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences**

2. <http://www.bioinformatics2.wsu.edu/cgi-bin/Athena/cgi/home.pl>

3. Ресурс позволяет, зная номер гена *Arabidopsis thaliana*, получить и проанализировать последовательность промотора, а также обеспечивает визуализацию сайтов связывания транскрипционных факторов в этом промоторе.

### **1. PLACE – A Database of Plant Cis-acting Regulatory DNA Elements**

2. <http://www.dna.affrc.go.jp/PLACE/signalscan.html>

3. Ресурс позволяет находить регуляторные цис-элементы в последовательности ДНК.

### **1. AHD - Arabidopsis Hormone Database**

2. <http://ahd.cbi.pku.edu.cn>

3. Ресурс содержит диаграммы сигнальных путей и путей биосинтеза гормонов, информацию о генах, участвующих в гормональной регуляции, обеспечивает для

каждого гена предсказание возникающих в процессе сплайсинга сайтов связывания миРНК, контролирующих стабильность мРНК.

### **1. Transgenic Crops**

2. (<http://cls.casa.colostate.edu/transgeniccrops/>),

3. Информация об известных ГМ культурах, с указанием введенного трансгена и способе трансформации.

### **1. GMO DB**

2. <http://www.gmo-compass.org/eng/gmo/db/>

3. Информация об известных ГМ культурах, с указанием введенного трансгена и способе трансформации.

### **1. GM Crop database**

2. [http://cera-gmc.org/index.php?action=gm\\_crop\\_database](http://cera-gmc.org/index.php?action=gm_crop_database)

3. Информация об известных ГМ культурах, с указанием введенного трансгена и способе трансформации.

### **1. MetaCyc**

2. <http://metacyc.org/>

3. База данных содержит информацию о более 1790 метаболических путях, более чем для 2000 организмов, аннотированных на основе экспериментальных работ

### **1. Plasmids & Vector Resources**

2. [http://www.geneinfinity.org/sp/sp\\_plasmidsandvectors.html](http://www.geneinfinity.org/sp/sp_plasmidsandvectors.html),

3. Ресурс содержит набор баз данных по векторам и плазмидам.

### **1. Biotechnology 4u**

2. [http://www.biotechnology4u.com/plant\\_biotechnology.html](http://www.biotechnology4u.com/plant_biotechnology.html)

3. Интернет портале по биотехнологии справочно – образовательного характера

## **4.1.2. Внешние Интернет-ресурсы, связанные с селекционно-генетическими подходами в агrobiотехнологии**

## **1. PHENOPSIS DB**

2. <http://bioweb.supagro.inra.fr/phenopsis/>

3. Ресурс для хранения, поиска и совместного использования данных об арабидопсисе. Позволяет визуализировать и статистически обрабатывать фенотипические данные и анализировать изображения

## **1. GrainGenes 2.0**

2. <http://wheat.pw.usda.gov/GG2/index.shtml>

3. Универсальный информационный ресурс по биологии, генетике, геномике злаковых. Включает ссылки на геномы злаковых, генетические карты, в том числе и пшеницы, фенотипические признаки, сорта, таксономию и публикации в этой области.

## **1. Crop Genebank Knowledge Base**

2. <http://cropgenebank.sgrp.cgiar.org/>

3. Портал по информационной поддержке исследований в области агрономических исследований для таких культур как рожь, кукуруза, рис, пшеница и др. Содержит описание протоколов проведения экспериментов, сбора данных, прочую полезную информацию.

## **1. Gramene**

2. <http://www.gramene.org/>

3. Информационный ресурс сравнительной геномики травянистых растений.

Итого предложено 23 информационных ресурса.

## **4.2 Разработка формата БД промоторов и трансляционных энхансеров для трансгенеза растений**

### **4.2.1 Разработка формата БД промоторов для трансгенеза**

Согласно проведенному нами анализу литературных данных, специализированные БД промоторов для планирования генно-инженерных экспериментов отсутствуют, имеющиеся аналоги приспособлены для решения более широкого круга задач (в основном, фундаментального характера) и не могут быть эффективно использованы для этой цели. С нашей точки зрения, основным недостатком существующих

информационных ресурсов является тот факт, что процедура аннотации промоторов на основе анализа литературных данных в них высокоизбирательна, то есть для отбора промотора для аннотации в БД необходимо выполнение ряда условий (например, картирование сайтов связывания транскрипционных факторов). Однако, для решения биотехнологических задач эта информация не нужна. С нашей точки зрения для выбора промотора в качестве элемента биотехнологической генетической конструкции необходимо и достаточно знать следующее: определенная нуклеотидная последовательность способна направлять транскрипцию репортерного гена в ГМО по определенному паттерну и на определенном уровне. Если такая информация будет оформлена в виде базы данных, то специалист в области генной инженерии может выбирать потенциальные промоторы по следующим полям: (1) организм – донор промотора; (2) организм – реципиент промотора (в котором была оценена его экспрессия); (3) паттерн экспрессии промотора (ткане-, органо-, стадие-специфичность наработки белка-репортера); (4) уровень экспрессии гена репортера, направляемого данным промотором.

Следует отметить, что объем доступной информации этого типа достаточно велик, поскольку характеристика транскрипционной активности промоторов часто используется в процессе изучения функций генов. Контроль экспрессии гена часто изучается с использованием репортерной конструкции, в которой промотор помещается перед репортерным геном (бета-глюкуронидазы *E. coli*, люциферазы, зеленого флюоресцирующего белка). Анализ активности репортерного белка в трансгенных растениях позволяет оценивать функциональные характеристики изучаемого промотора. Описанные в литературе промоторы могут обеспечивать достаточно широкий выбор подходящих паттернов транскрипции. Ниже приведен краткий обзор современной литературы по этой теме. Промотор гена пшеницы TaPT2 обеспечивает специфическую активность репортерного гена GUS в корнях трансгенных арабидопсиса и пшеницы при фосфорном голодании [190]. Промоторы генов глиадинов пшеницы обеспечивают высокую активность GUS в эндосперме трансгенной пшеницы [191,192]. Промотор гена глобулина овса обеспечивает эндосперм-специфичную экспрессию GUS в семенах ячменя (до 10% растворимого белка [193]). Промотор ADP глюкозо-пирофосфорилазы батата обеспечивает высокий уровень экспрессии репортерного гена GUS в клубнях картофеля [194]. Промотор TaPSG719 пшеницы проявляет специфическую активность исключительно в пыльце трансгенного табака [195]. Этот промотор может быть с успехом использован в сельском хозяйстве при создании форм с мужской стерильностью.

В некоторых случаях показано, что такая тканеспецифическая экспрессия является более выгодной, чем конститутивная, которая обеспечивается промоторами 35S вируса мозаики цветной капусты или убиквитина кукурузы [196-198]. Развитие новых методов для повышения устойчивости растений к различным стрессам также требует использование промоторов со специфическим паттерном экспрессии. Например, активность промоторов TaAIDFa и Cor/Lea генов пшеницы, промотора DREB1 карликовой яблони возрастает после воздействия засухи, соли, низких температур и абсцизовой кислоты [199-202]. Эти промоторы могут быть с успехом использованы для повышения устойчивости сельскохозяйственных культур к неблагоприятным воздействиям окружающей среды.

Таким образом, на основе аналитического обзора нами были разработаны основные требования к БД промоторов для трансгенеза. БД должна содержать следующую необходимую информацию:

- а) название организма – донора нуклеотидной последовательности промотора
- б) название гена – донора нуклеотидной последовательности промотора
- в) репортерный ген
- д) паттерн экспрессии промотора (индуцибельность, ткане-, органо-, стадие-специфичность наработки белка-репортера).
- е) уровень экспрессии гена репортера, направляемого данным промотором.

В качестве платформы для БД промоторов для трансгенеза использована SRS (Sequence Retrieval System). Разработанный нами формат включает три блока PROMOTER, SEQUENCE и GENE. Интерфейс SRS позволяет пользователям проводить поиск промотора с определенными характеристиками.

- а) блок PROMOTER (описание собственно промотора) включает 14 полей, описание приведено ниже в таблице 2. Следует отметить, что рабочий язык информационных ресурсов ЭОИП «Биотехнология растений» английский, что необходимо для осуществления взаимодействия с коллегами из других стран. При этом интерфейс ЭОИП и соответствующие тьюториалы будут созданы в том числе и на русском языке.

Таблица 2 - Структура карточки блока PROMOTER {(+) обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации}.

Название поля	Содержание поля
PROMOTER_ID	Идентификатор карточки в блоке PROMOTER (+)
GENE_ID	Ссылка на идентификатор гена в блоке GENE (+)
TARGET SPECIES	Видовое название трансгенного организма (+)
LOCALIZATION	Границы промотора относительно точки отсчета, ссылка на базу GenBank с указанием старта транскрипции или трансляции
DESCRIPTION	Данные о структуре промотора (комментарий) (+)
SEQUENCE_ID	Ссылка на идентификатор последовательности промотора в блоке SEQUENCE (+)
REPORTER	Название гена-репортера
TRANSGENE	Название и видовое происхождение трансгена (+)
STAGE_ORGAN_TISSUE	Название стадий развития, органов и тканей, в которых наблюдается активность промотора (+)
REGULATOR	Индуктор (репрессор), влияющий на активность промотора (+)
COMMENT	Характеристика активности промотора (комментарий) (+)
REFERENCE	Название статьи, из которой была взята информация о промоторе (+)
PUBMED	ссылка на базу данных публикаций Entrez-PUBMED
END	Конец блока описания промотора

б) блок SEQUENCE (описание нуклеотидной последовательности промотора) включает 5 полей, описание приведено ниже в таблице 3.

Таблица 3 - Структура карточки блока SEQUENCE {(+) обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации}.

Название поля	Содержание поля
SEQUENCE_ID	Идентификатор карточки в блоке SEQUENCE (+)
PROMOTER_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке PROMOTER (+)
GENE_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке GENE
SEQUENCE	Нуклеотидная последовательность
END	Конец блока описания последовательности

д) блок GENE (описание гена - донора промотора) включает 14 полей, описание приведено ниже в таблице 4.

Таблица 4 - Структура карточки в блоке GENE {(+) обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации}.

Название поля	Содержание поля
GENE_ID	Идентификатор карточки в блоке GENE (+)
DATE	Дата последнего редактирования
AUTHOR	Имена составителей карточки
GENE	Название и синонимы гена (+)
PRODUCT	Название и синонимы продукта гена (+)
TAXON	Таксономическая принадлежность (+)
SPECIES	Видовая принадлежность (+)
KEYWORDS	Ключевые слова (+)
ACTIVITY	Описание экспрессии гена (комментарий) (+)
PROMOTER_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке PROMOTER (+)
SEQUENCE_ID	Перекрестная ссылка на идентификатор соответствующей карточки в блоке SEQUENCE (+)
REFERENCE	Ссылка на статью, из которой взяты данные (+)
PUBMED	Перекрестная ссылка на базу публикаций Entrez-PUBMED (+)
END	Конец блока описания гена

Ниже приведены примеры аннотации заполнения карточек на основе реальных литературных данных (таблицы 5 - 7)



Таблица 5 - Пример заполнения карточки в блоке PROMOTER.

**PROMOTER\_ID** Ps:TOP2\_P3  
**GENE\_ID** Ps:TOP2  
**TARGET SPECIES** mouse-ear cress (*Arabidopsis thaliana*),  
tobacco (*Nicotiana tabacum*)  
**KEYWORDS** stress response, white light-induced, cotyledon, cold-induced,  
abscisic acid-induced  
**LOCALIZATION** from -362 to +74; GenBank; AF144649; TSS: 609; from 247 to 682  
**DESCRIPTION** The regulatory region upstream of the coding part of the reporter gene  
included a promoter fragment (362 bp upstream of the transcription start site), 60 bp of the  
5'UTR, and 14 bp of the coding sequence of the PsTOP2 gene.  
**SEQUENCE\_ID** Ps:TOP2\_P3S  
**REPORTER** GUS  
**STAGE\_ORGAN\_TISSUE** seedling, cotyledon  
**REGULATOR** white light, cold, abscisic acid  
**COMMENT** The minimal TOP2 promoter that is induced by cold.  
Transgenic *Arabidopsis*  
The -362 TOP2 promoter showed very little stimulation in GUS, if any, in white light, red  
light, far-red light and blue light as compared to dark grown *Arabidopsis* seedlings  
(Hettiarachchi et al., 2003).  
Transgenic tobacco  
The expression of the -362 TOP2 promoter was detectable only in the tobacco cotyledons  
(Hettiarachchi et al., 2003). The -362 TOP2 promoter showed about a 2-fold higher level of  
activity at 12 h, with very little induction, if any, up to 4 h of cold treatment of transgenic  
tobacco seedlings. The -362 TOP2 promoter had no detectable induction by salt or salicylic  
acid treatment. The -362 TOP2 promoter showed very little induction, if any, at 100 microM  
abscisic acid and about 2-fold induction at 150 microM abscisic acid (Hettiarachchi et al.,  
2005).  
**REFERENCE** Hettiarachchi G.H., Yadav V., Reddy M.K., Chattopadhyay S.,  
Sopory S.K. Light-mediated regulation defines a minimal promoter region  
of TOP2. *Nucleic Acids Res.*, 2003, 31(18), 5256-5265.  
**PUBMED** [12954761](#)  
**REFERENCE** Hettiarachchi G.H., Reddy M.K., Sopory S.K., Chattopadhyay S.  
Regulation of TOP2 by various abiotic stresses including cold and  
salinity in pea and transgenic tobacco plants.  
*Plant Cell Physiol.*, 2005, 46, 1154-1160.  
**PUBMED** [15879449](#)  
**END**

Таблица 6 - Пример заполнения карточки в блоке SEQUENCE.

```
SEQUENCE_ID Ps:TOP2_P3S
PROMOTER_ID Ps:TOP2_P3
GENE_ID Ps:TOP2
SEQUENCE
-362 ttaa taaccctagt ttgacactat aaatactaaa gatgctggtg aatgaaagaa
gaaaccaac agatgcctag cgcgtagccc cgaatgccc tctctctca ctctccact
accaacaacc ggataccccc acgtgtagtc caacaaaaac attaaaagac aactgccag
aactgataca acaacacaca ctcacaaaat caccatcctc ctcaccctec atctctctcc
acctgctccc tccactctca atccgccgaa aaagcaccac gccggcaacc acaaacctaa
tgcttctca acctccaatc tccaccctc attctctcc accgttcgtg tttttatcgt
+1
tgttcgtcCt cacctccacc caacgacacc aaaaatcctg cgagcaaac ctaagccgc
cgcaagccca ccatctaccg tc +74
END
```

Таблица 7 - Пример заполнения карточки в блоке GENE.

```
GENE_ID Ps:TOP2
GENE top2
PRODUCT DNA topoisomerase 2, TOP2, TOP2
TAXON Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I;
Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
SPECIES pea (Pisum sativum)
KEYWORDS stress response, salt-induced, cold-induced, abscisic acid-induced, salicylic
acid-induced, white light-induced, red light-induced, far-red light-induced, blue light-induced,
cell proliferation
ACTIVITY The topo II transcript level was maximal in actively growing tissues such as root
tips and young leaves from the apical region as compared to the differentiated tissues such as
the upper part of the root and internodal region of the stem of 7- to 10-day old seedlings. The
topo II transcript was also present in the young and mature flower buds and in immature pods.
The topo II transcript was abundant in proliferative tissues. The level of topo II transcripts
could be stimulated by exogenous application of growth factors that induced proliferation in
vitro cultures. The increase in the topo II transcript was seen within an hour after irradiation of
etiolated seedlings (Reddy et al., 1999).
The expression of TOP2 was more than 10-fold higher in constant white light grown seedlings
as compared to the darkness. The expression of TOP2 was induced up to 8-fold by a broad
spectrum of light (red, far-red and blue). The rate of blue light-mediated induction of TOP2
was slower when compared to the red light- and far-red light-mediated inductions. The
expression of TOP2 was detected to be at the highest level with about 10-fold more as
compared to dark in 7-day-old plants. The transcript level significantly decreased in
14-day-old plants and showed only about 3-fold more expression than the dark in 21-day-old
plants (Hettiarachchi et al., 2003).
In 10-day-old pea seedlings 2.5-fold higher expression of TOP2 mRNA was detected in
shoots after 2 h of cold treatment and this increased level remained the same up to 12 h
exposure to cold.
```

Таблица 7 - Продолжение.

Although the cold-mediated induction of TOP2 in roots was slower, more than 6-fold induction was observed in roots after 12 h exposure to cold. TOP2 mRNA is up-regulated by salinity stress and the response was stronger in roots (7-fold) compared with shoots (3-fold). The expression of TOP2 was up-regulated in response to exogenous abscisic acid (ABA) and the maximum level of induction was observed at 150 microM ABA in both root (4-fold) and shoot (2.5-fold) tissues. Salicylic acid could induce the maximum level of expression of TOP2 at 100 microM in both shoot (2.5-fold) and root (6-fold) tissues. No significant change in the expression of TOP2 in 10-day-old pea seedlings under dehydration stress was detected (Hettiarachchi et al., 2005).

**PROMOTER\_ID** Ps:TOP2\_P1 Ps:TOP2\_P2 Ps:TOP2\_P3 Ps:TOP2\_P4

**SEQUENCE\_ID** Ps:TOP2\_P1S Ps:TOP2\_P2S Ps:TOP2\_P3S Ps:TOP2\_P4S

**REFERENCE** Reddy M.K., Nair S., Tewari K.K., Mudgil Y., Yadav B.S., Sopory S.K. Cloning and characterization of a cDNA encoding topoisomerase II in pea and analysis of its expression in relation to cell proliferation. *Plant Mol. Biol.*, 1999, 41(1), 125-137.

**PUBMED** [10561074](#)

**REFERENCE** Hettiarachchi G.H., Yadav V., Reddy M.K., Chattopadhyay S., Sopory S.K. Light-mediated regulation defines a minimal promoter region of TOP2. *Nucleic Acids Res.*, 2003, 31(18), 5256-5265.

**PUBMED** [12954761](#)

**REFERENCE** Hettiarachchi G.H., Reddy M.K., Sopory S.K., Chattopadhyay S. Regulation of TOP2 by various abiotic stresses including cold and salinity in pea and transgenic tobacco plants. *Plant Cell Physiol.*, 2005, 46, 1154-1160.

**PUBMED** [15879449](#)

**END**

#### 4.2.2 Разработка формата БД трансляционных энхансеров для трансгенеза

Согласно проведенному нами анализу литературных данных, специализированные БД трансляционных энхансеров для планирования генно-инженерных экспериментов отсутствуют, имеющиеся аналоги приспособлены для решения более широкого круга задач (в основном, фундаментального характера) и не могут быть эффективно использованы для этой цели. С нашей точки зрения, основным недостатком существующих информационных ресурсов является тот факт, что процедура аннотации энхансеров трансляции на основе анализа литературных данных в них высокоизбирательна, то есть необходимы детальные знания об их тонкой организации. Однако, для решения биотехнологических задач эта информация не нужна. С нашей точки зрения для выбора трансляционного энхансера в качестве элемента биотехнологической генетической конструкции необходимо и достаточно знать следующее: определенная нуклеотидная последовательность способна направлять трансляцию мРНК репортерного гена в ГМО по определенному паттерну и на определенном уровне. Если такая информация будет оформлена в виде базы данных, то

специалист в области генной инженерии может выбирать потенциальные энхансеры по следующим полям: (1) организм – донор энхансера; (2) организм – реципиент энхансера (в котором была оценена его экспрессия); (3) паттерн трансляционной активности (ткане-, органо-, стадие-специфичность наработки белка-репортера); (4) уровень трансляционной активности мРНК гена репортера, содержащей данный энхансер. Следует отметить, что объем доступной информации этого типа по нашим оценкам достаточно велик и она носит в основном описательный характер. Эта информация полезна для планирования генно-инженерных экспериментов, поскольку трансляционные энхансеры нельзя заменить другими функциональными элементами в структуре генетической конструкции. В предлагаемом формате БД трансляционных энхансеров содержатся два блока (OBJECT и ENHANCER) с оценкой возможности более детального описания – по мере накопления данных.

В качестве платформы для БД трансляционных энхансеров для трансгеноза использована SRS (Sequence Retrieval System). Интерфейс SRS позволяет пользователям проводить поиск энхансера с определенными характеристиками. Ниже представлен формат блока OBJECT. Ниже приведено описание блока OBJECT, включающее 15 полей, представленных в таблице 8.

Таблица 8 - Структура карточки блока OBJECT {(+) обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации}.

ID	Идентификатор карточки блока OBJECT (+)
LOCATION	Расположение энхансера (5'UTR, 3'UTR, CDS) (+)
TYPE	Тип энхансера (stress-specific и др.) (+)
OC	Таксономическая классификация (+)
OS	Название вида (+)
GENE	Название гена (+)
CAP	Наличие кепы на 5'-конце мРНК (+)
POLYA	Наличие поли(А)-участка на 3'-конце мРНК (+)
SQ	Собственно нуклеотидная последовательность энхансера
COMMENTSEQ	Комментарий о происхождении нуклеотидной последовательности и ее расположении в составе генетических конструкций (+)
KEYWORD	Ключевые слова (+)
COMMENT	Развернутый комментарий о специфичности и активности энхансера, эффективности его использования в различных видах организмов реципиентов (+)
LINK_ENH	Ссылка на идентификатор карточки блока ENHANCER
LINK	Ссылка на банк данных нуклеотидных последовательностей (+)
END	Граница карточки

Формат карточки блока ENHANCER включает 9 полей, представленных в таблице 9.

В таблице 10 приведен пример заполнения карточки блока ОБЪЕКТ на основе реальных биологических данных.

Таблица 9 - Структура карточки блока ENHANCER {(+) обозначает индексируемое SRS поле, по которому пользователь может осуществлять поиск информации }.

ID	Идентификатор карточки блока ENHANCER (+)
OBJID	Ссылка на идентификатор карточки блока ОБЪЕКТ
SEQUENCE	Нуклеотидная последовательность функционального района (5'-НТП, 3'-НТП), содержащего энхансер
COMMENTSEQ	Комментарий к структуре экспериментальной конструкции (+)
ORGANISM	Видовое название организма, на котором проводили эксперименты (+)
KEYWORD	Ключевые слова (+)
COMMENT	Развернутый комментарий о специфичности и активности энхансера, эффективности его использования в различных видах организмов реципиентов (+)
REFERENCE	Название статьи и ссылка на БД PubMed (+)
END	Граница карточки

Таблица 10 - Пример заполнения карточки в блоке ОБЪЕКТ.

ID	ADHZM5
LOCATION	5'UTR
TYPE	Stress-specific enhancer
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea
OS	Zea mays
GENE	ADH1, alcoholdehydrogenase I
CAP	Capped
POLYA	Polyadenylated
SQ	ATTTTCTCGCTCCTCACAGGCTCATCTCGTTTGGATCGATTG GTTTCGTAAGTGGTGAAGGACTGAGGGTCTCGGAGTGGATCG ATTTGGGATTCTGTTCGAAGATTTGCGGAGGGGGGCA
COMMENTSEQ	5'UTR of ADH1 gene mRNA
KEYWORD	Enhancer, hypoxia, anoxia, anaerobiosis, stress
COMMENT	It was found that translation of alcoholdehydrogenase mRNA was efficient under oxygen deprivation conditions whereas translation of many other mRNAs was stopped. No changes in mRNA stability were detected so the effect observed could result from the changes in stress-specific translation rate. Deletions of ADH 5'UTR decreased stress-specific translatability: the influence of possible changes in secondary structure was not tested or discussed...
LINK_ENH	Ссылка на идентификатор карточки блока ENHANCER
LINK	EMBL_AC X00580
END	

В таблице 11 приведен пример заполнения карточки блока ENHANCER на основе реальных биологических данных.

Таблица 11 - Пример заполнения карточки в блоке ENHANCER.

ID	E0097
SEQUENCE	ATAGGGAGACCGAATTCGAGTCATTTTCTCGCTCCTCACAGGCTCATCTC GTTTGGATCGATTGGTTTCGTAACCTGGTGAAGGACTGAGGGTCTCGGAGT GGATCGATTTGGGATTCTGTTCGAAGATTTGCGGAGGGGGGCA
COMMENTSEQ	Design of mRNA 5'UTR of GUS reporter gene: first 23 nt were taken from vector sequence followed by 108-nt long 5'UTR of ADH1. In this construct CDS consisted from 18 codons of ADH1 CDS fused to 8 codons derived from vector polylinker sequence and GUS CDS downstream (see LONG). 3'UTR was represented by ADH
ORGANISM	Zea mays
KEYWORD	Hypoxia, stress, 5'UTR, enhancer
COMMENT	Translational efficiencies of reporter mRNAs containing UTR sequences of maize alcoholdehydrogenase gene mRNA were tested in maize protoplasts under normal or oxygen deprivation conditions. No changes in mRNA stability were detected so the effect observed resulted from the changes in stress-specific translation rate. Interestingly, the presence of ADH 5'UTR did not affect translation under aerobic conditions. Generally, deletions of ADH-derived fragments decreased stress-specific translatability: either deletion of first 18 ADH-derived codons or fragments of 5'UTR or 3'UTR. Note, that 5'portion of 5'UTR was presented in all constructions. The influence of possible changes in secondary structure was not tested or discussed. As was found ADH1 3'UTR mRNA increase hypoxia-specific translation 3.5-fold but decrease aerobic translation 3-fold.
REFERENCE	Bailey-Serres J., Dawe R.K. Both 5' and 3' sequences of maize adh1 mRNA are required for enhanced translation under low-oxygen conditions. Plant Physiol. 1996. 112. 685-695 <a href="https://pubmed.ncbi.nlm.nih.gov/8883381/">PMID:8883381</a>
END	Граница карточки

### 4.3 Разработка формата описания фенотипических признаков растений в базе данных WheatPGE

#### 4.3.1 Форматы описания признаков

На основании проведенного обзора литературы мы провели дополнение структуры базы данных в части, касающейся описания фенотипа растения. В качестве наиболее актуальных признаков на данном этапе разработки нами были выбраны признаки, характеризующие продуктивность растения (характеристики колосьев) и его устойчивость к стрессу – опушение.

На рисунке 6 приведена модификация фрагмента логической структуры базы данных WheatPGE, связанного с описанием фенотипических признаков, которая

показывает добавленные нами признаки. Описание характеристик колоса (англ. *ear*) приведено в таблице 12. Всего эта таблица содержит описание 7 признаков колоса. Эта таблица связана с таблицей «Набор колосьев» (*ear\_set*), которая, в свою очередь связана с таблицей «Список колосьев» (*ear\_list*). На таблицу «Характеристика колоса» в свою очередь ссылаются таблица, описывающая список возможных типов морфологии колоса (*ear\_morphology\_list*). Тип морфологии выбирается из «Набора морфологий колоса» (*ear\_morphology\_set*), а описание морфологий хранится в таблице «Морфология колоса» (*ear\_morphology*) (см. рисунок 6).

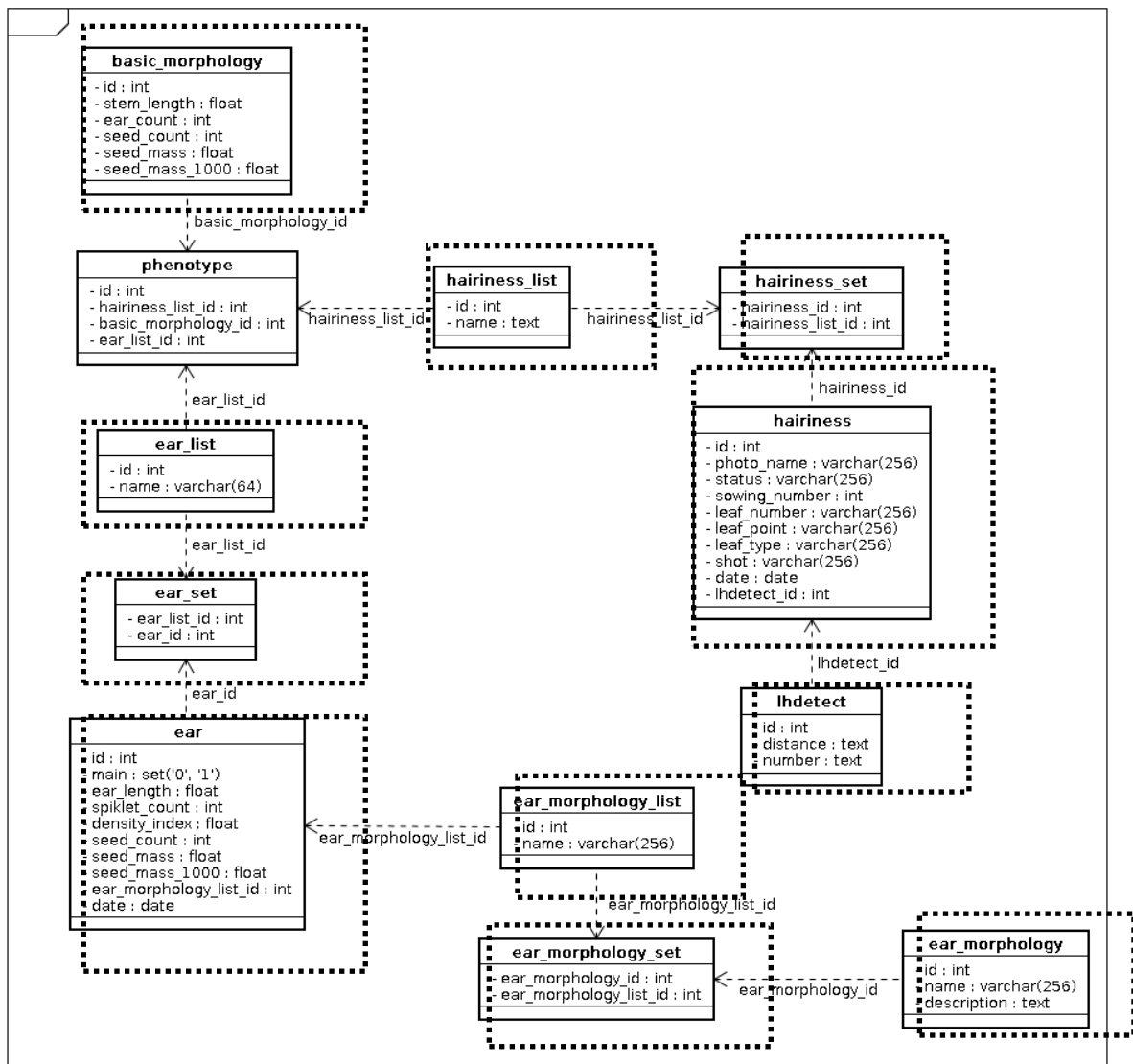


Рисунок 6 - Описание фрагмента логической структуры базы данных WheatPGE, описывающий новые признаки растений, связанные с продуктивностью и устойчивостью к стрессу. Таблицы с добавленными признаками выделены пунктирными линиями.

Таблица 12 - Описание характеристик колоса.

id	Идентификатор
main	Является ли колос главным у растения
ear_length	Длина колоса
spiklet_count	Число колосков
density_index	Индекс плотности
seed_count	Число зерен
seed_mass	Масса зерен
seed_mass_1000	Масса 1000 зерен
ear_morphology_list_id	Ссылка на тип морфологии колоса
date	Дата заполнения
add_user_id	Идентификатор пользователя, создавшего запись
edit_user_id	Идентификатор пользователя, модифицировавшего запись

Описание базовых характеристик растения приведено в таблице 13. Эта таблица содержит 5 признаков связанных с продуктивностью (характеристики колосьев) и формой растения (длина стебля). Таблица описания базовых признаков связана с таблицей фенотипа растения (см. рисунок 6).

Таблица 13 - Описание набора базовых характеристик растения.

id	Идентификатор
stem_length	Длина стебля
ear_count	Число колосьев
seed_count	Число зерен
seed_mass	Масса зерен
seed_mass_1000	Масса 1000 зерен
add_user_id	Идентификатор пользователя, создавшего запись
edit_user_id	Идентификатор пользователя, модифицировавшего запись

Таким образом, модификация структуры базы данных, приведенная на рисунке 6. Позволила описать дополнительно 12 признаков растения, связанных с его продуктивностью и морфологией.

Описание количественных характеристик опушения листа содержит результаты анализа компьютерной обработки изображений, программой LHDetect2, как показано в таблице 14. Метод определения этих характеристик будет описан ниже.



Таблица 14 - Описание набора характеристик опущения листа.

id	Идентификатор
photo_name	Имя файла с изображением
status	Обработан ли файл программой
sowing number	Посевной номер растения
leaf number	Номер листа
leaf point	Место снятия изображения сгиба
leaf type	Тип листа
shot	Номер изображения сгиба
date	Дата заполнения
lhdetect_id	Идентификатор программы обработки данных

#### 4.3.2 Алгоритм определения количественных характеристик опущения листа

Алгоритм получения количественных характеристик опущения листа основан на анализе микроизображений сгиба, примеры которых приведены на рисунке 7. На этих изображениях область листа располагается справа, область фона – слева, граница листа проходит приблизительно посередине снимка вертикально. Трихомы идентифицируются как filamentous выросты, сужающиеся от основания (области прикрепления к поверхности листа) к вершине. Трихомы на изображении могут различаться по длине, содержать воздушные пузырьки, быть значительно изогнутыми. Кроме того, нередко встречаются трихомы, которые на изображении пересекаются. Область фона может содержать пятна от небольших загрязнений. В настоящей работе мы разработали алгоритм, который позволяет по изображениям подобного типа оценивать количественные характеристики опущения.

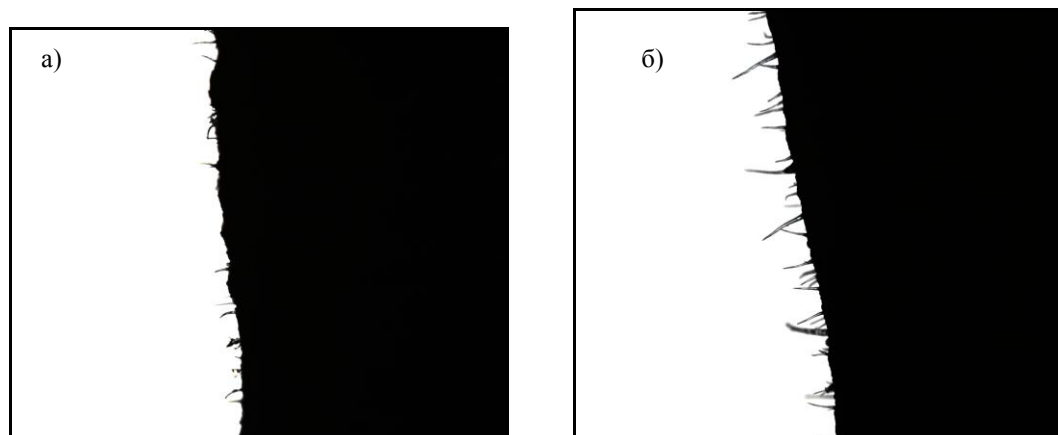


Рисунок 7 - Пример типичных изображений поперечного сгиба листа пшеницы (а) предфлаговый лист растения сорта Фора, являющегося слабоопушенным; (б) предфлаговый лист растения сорта Алтайская 98, демонстрирующий более интенсивное опушение.

При помощи такого метода нами была сформирована база данных изображений сгибов листа пшеницы, в которой содержится 1352 изображения, представляющих 261 растений 88 сортов пшеницы, выращенных как в полевых, так и в тепличных условиях в 2008-2010 годах. База данных изображений доступна по адресу [wheatdb.org](http://wheatdb.org).

При анализе изображений поперечного сгиба листа пшеницы интересны такие их характеристики как:

количество трихом на изображении (характеристика плотности опушения);

длина трихомов и распределение трихом по длинам (характеристики размеров трихом);

среднее значение длины трихом (характеристика среднего размера трихом);

площадь трихом на изображении (характеристика объема трихом).

Алгоритм, разработанный нами для определения этих характеристик, работает в несколько этапов:

а) преобразование изображения в шкалу оттенков серого и удаление шумов;

б) бинаризация изображения на области листа с трихомами и фона;

в) построение границы между листом и фоном;

д) выделение одиночных трихомов и их групп как объектов изображения;

е) анализ размеров и форм объектов.

Опишем эти этапы подробнее.

Удаление шумов. На первом этапе конвертировали изображение из цветного в оттенки серого цвета и подвергали его гауссовой фильтрации с параметром  $\sigma$  для удаления шумов (небольших пятен от загрязнений в области фона и пузырьков в трихомах).

Бинаризация изображения. На следующем этапе мы проводили бинаризацию изображения по цвету для того, чтобы разделить его на область фона (более яркие пиксели) и область листа с трихомами (темные пиксели). Чтобы определить порог для классификации пикселей на фон или лист мы провели анализ интенсивности цвета на 1352 изображениях сгибов листа из нашей базы данных (гистограмма распределения интенсивности цвета приведена в приложении 2). Распределение имело 5 характерных областей по значениям интенсивности: 0-50 пиксели области листа, 50-205 – мелкие пятна в области фона, 205-228 – трихомы и крупные пятна, 228-240 – пиксели листа вблизи границы с фоном, трихомы, не попавшие в фокус, 240-255 – область фона. На первом этапе для бинаризации мы использовали порог по интенсивности равный 240.

Пиксели, имеющие интенсивность цвета выше порога относились к области фона, ниже – к области листа. Пятна в области фона (результат загрязнения стекла) удалялись.

Определение границы листа. Для выделения трихомов на изображении мы вначале определяли границу листа и фона. В качестве аппроксимации границы мы строили линию, которая содержит точки листа, соприкасающиеся с точками фона, и проходит через основания трихомов. Для этого изображение разбивалось на перекрывающиеся горизонтальные полосы размером  $ws=70$  пикселей (147 мкм) с шагом  $s=1$  пиксель по вертикали. В каждой такой полосе выбирался пиксель фона, имеющий максимальное значение координаты  $X$ . Следующий за ним по горизонтали пиксель листа формировал реперную точку границы. После определения реперных точек для каждой полосы на изображении они соединялись ломаной, аппроксимирующей границу. Выступы листа за линию границы на величину  $min\_trichome$  пикселей удалялись (т.е. мелкие бугорки и выступы на поверхности листа в качестве трихомов мы не рассматривали). После определения границы все точки листа, находящиеся справа от нее преобразовывались в точки фона. В результате этого, на изображении оставались лишь объекты, которые соответствовали трихомам.

Выделение объектов и определение вершин и оснований трихомов. После удаления области листа мы вновь проводили бинаризацию изображения по порогу интенсивности цвета 228. Такой порог позволял исключить мелкие пятна вблизи трихомов и трихомы, не попавшие в фокус изображения. Объекты на изображении выделяли при помощи алгоритма поиска контуров, реализованные в библиотеке OpenCV 2.0 (<http://opencv.willowgarage.com/wiki/>). Далее для каждого объекта мы определяли вершины и основания. В случае если объект сформирован несколькими трихомами, он может иметь несколько вершин и оснований, в некоторых случаях числа оснований и вершин могут различаться.

Для поиска вершин трихомов в объекте алгоритм последовательно обходит его границу. На каждом шаге для текущего пикселя границы определяются еще два, удаленных от него на  $N$  соседних пикселей. После этого измеряется угол, образованный тремя этими точками. Если его величина меньше порогового значения  $alpha$ , средняя точка считается потенциальной вершиной трихома.

Из-за того, что ширина трихома вблизи вершины составляет несколько пикселей, определение угла для идентификации вершины по трем точкам может быть неточным. Поэтому для выбора одной вершины из отобранных потенциальных кандидатов мы использовали более робастный, но медленный алгоритм. В среднюю точку мы помещали

окружность радиусом  $R$  и подсчитывали, сколько точек, принадлежащих трихому, попадает внутрь этой окружности. Если эта доля меньше порогового значения *percent*, то потенциальная точка считается вершиной трихома. Вершины трихомов и точки основания трихомов, определенные при помощи такого алгоритма показаны на рисунке 8. Точками, ограничивающие основание трихома, считались точки, граничащие одновременно с областью фона и границей листа и фона.

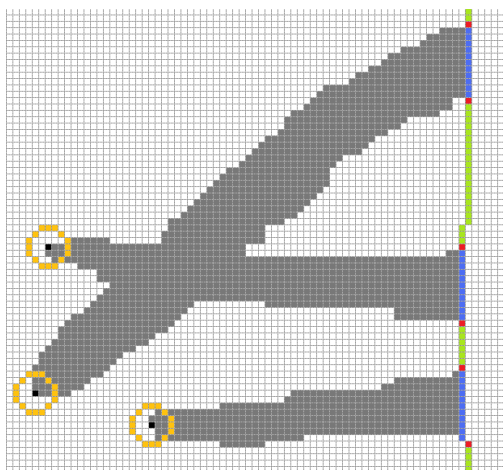


Рисунок 8 - Определение вершин и оснований у трихом. На изображении белым цветом показаны пиксели фона, серым – пиксели трихомов, зеленым цветом показана линия границы листа. Черным цветом показаны пиксели, определенные как вершины, красным – границы оснований трихом, а синим – сами основания. Желтым цветом показаны окружности, позволяющие определить, относится ли ее центральный пиксель к вершине (по доле пикселей внутри, отнесенных к трихому). Показаны два объекта. Верхний является сложным и состоит из двух пересекающихся трихомов, имеет две вершины и два основания. Нижний является простым и имеет одну вершину и одно основание.

Выделение и классификация объектов на простые и сложные, определение длин трихомов. Все выделенные контуры можно разделить простые (состоящие из одного трихома) и сложные (состоящие из нескольких трихомов), рисунке 3. Количество трихомов, содержащееся в объекте, вычисляли как  $\max(H, B)$ , где  $H$  — количество вершин в контуре,  $B$  — число оснований трихом у объекта.

В нашей базе данных изображений было обнаружено всего 20074 объекта, в которых содержится 31037 трихомов. Количество сложных объектов составляет 31.1%, при этом на них приходится 15.4% общей площади трихомов. Число трихомов в сложных

объектах составило 45.2%. Среднее число трихомов в сложном контуре равно 2.8. Таким образом, сложные объекты содержат существенную долю информации о трихомах.

Для определения длины трихома в простых объектах мы используем алгоритм, предложенный в работе Leifer et al [183]. Граница простого объекта разбивается на 2 части: от вершины трихома и до точек, ограничивающих основание трихома. После этого проводится обход точек границы от вершины до основания трихома: для каждой последовательной пары точек из разных частей границы определяется середина отрезка между ними. Серединные точки соединяются ломаной линией, аппроксимирующей центральную линию объекта. Длина центральной линии является оценкой длины трихома (см. рисунок 9).

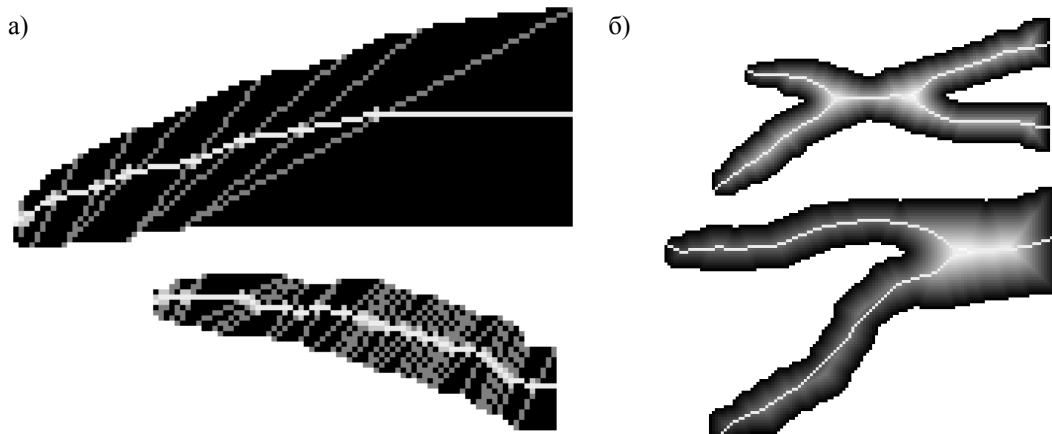


Рисунок 9 - Построение центральной линии для простых и сложных объектов: а) для простых объектов центральная линия аппроксимируется ломаной, проходящая через средние линии отрезков, соединяющих пиксели из верхней и нижней частей границ объекта; б) для сложных контуров среднюю линию формируют пиксели, максимально удаленные от границ объектов.

Для сложных объектов, в которых содержится более одного трихома, используется другой алгоритм. Для каждой точки внутри контура вычисляется расстояние до ближайшей точки границы, значения расстояний нормализуются. Для объекта, определенного контуром, строится граф, в котором каждый пиксель соединён с соседними, а вес связи является средним арифметическим от их расстояний до границы. Центральная линия трихома определялась как путь наименьшего веса в таком графе от узла, являющимся вершиной трихома до узла, являющегося центром основания. Результат работы алгоритма показан на рисунке 9(б).

Для оценки точности определения числа трихомов было отобрано 76 изображений. На каждом изображении области вершин трихомов с диаметром в 15 пикселей были помечены вручную. Мы определяли положение вершин трихомов и если они находились в пределах указанной области, то считалось, что вершина указана правильно. Число вершин соответствовало числу трихомов на изображении.

Подбор оптимальных параметров алгоритма был осуществлён автоматически. Рассматривались следующие параметры:

- Максимальный угол  $\alpha$  (в градусах) кривизны контура, при котором рассматриваемая точка в контуре помечается, как потенциальная вершина трихома. Параметр менялся в диапазоне от 80 до 120 с шагом 1.

- Минимальный размер трихома  $min\_trichome$ . Параметр менял в диапазоне от 4 до 10 с шагом 2.

- Максимальный процент  $percent$  точек листа (в окружности радиуса  $R$  вокруг потенциальной точки) при котором точка рассматривается, как вершина трихома.

- Стандартное отклонение  $\sigma$  при размытии методом Гаусса. Параметр менялся в диапазоне от 0 до 2 с шагом 0.25.

Параметры  $N$  и  $R$  были зафиксированы:  $N=2$ ,  $R=min\_trichome/2$ .

На выборке из 70 изображений для каждого из них оценивали ошибки недопредсказания положения вершины  $fn$  (доля трихомов на изображении, для которых метод не определил вершину в пределах заданной области), перепредсказания  $fp$  (отношение ложно указанных вершин к полному числу трихомов), а так же долю правильно предсказанных трихомов,  $tp$ . На основе этих значений вычисляли оценки  $precision\ pr=tp/(tp+fp)$  и  $recall\ re=tp/(tp+fn)$  и общую меру точности  $Q=2 \cdot (pr \cdot re) / (pr+re)$ . Чем выше  $Q$ , тем меньше ошибок в идентификации трихом на изображении. Максимальное значение  $Q$  равно 1 (в случае если все трихомы идентифицированы правильно и не идентифицировано ни одного лишнего объекта).

Мы так же проводили оценку ошибок в определении числа трихом : рассчитывали среднюю абсолютную ошибку (mean absolute error, MAE) и ее процент к числу трихом (mean absolute percentage error, MAPE) для числа трихом на изображении:

$$MAE = (\sum |n_i - n'_i|) / N, \quad (3)$$

$$MAPE = 100\% \cdot (\sum (|n_i - n'_i| / n_i) / N, \quad (4)$$

где  $N$  – число тестовых изображений,  $n_i$  – число трихом, определенных на изображении  $i$  вручную,  $n'_i$  – число трихом, определенных компьютерным методом. Чем

больше значение параметров MAE и MAPE, тем выше ошибка определения числа трихом на изображении.

При проведении расчетов использовался высокопроизводительный кластер ЦКП «Биоинформатика» (<http://bioinformatics.bionet.nsc.ru/>). При оценке точности подсчета числа трихом были использованы следующие параметры, дающие максимальную точность на выборке из 76 изображений:  $\alpha=105^\circ$ ,  $\text{min\_trichome}=6$  пикселей,  $\text{percent}=30$ ,  $\sigma=1.0$ . Точность  $Q$  при этом составила 0.85, ошибка перепредсказания  $fp=0.07$ , а недопредсказания  $fn=0.19$ . Такая ошибка означает, что при числе 30-40 трихом на лист ошибка в подсчете составит несколько трихом. Ошибки MAE и MAPE составили, соответственно, 3.33 и 11.65%.

#### 4.4 Разработка формата описания селекционно-генетического эксперимента в базе данных WheatPGE

В процессе селекционно-генетического эксперимента исследователь проводит ряд агротехнических мероприятий (посев, прополка, полив, внесение удобрений и т.п.), фиксирует фазы развития растений, условия среды, а так же проводит фенотипирование растений. Эти действия сгруппированы нами по блокам и приведены в таблице 15.

Таблица 15 - Блоки агротехнических мероприятий (формальное описание), определения фаз развития, условий среды и действий по измерению параметров фенотипа.

Информационные блоки мероприятий, проводимых в ходе эксперимента	Содержание действий (последовательно)
"Агротехнические мероприятия"	I. <i>Подготовка поля</i> : осенняя вспашка, весенняя культивация, внесение минеральных удобрений, разбивка опытных делянок
	II. <i>Обработка посевов</i> : прополка, культивация, обработка от ранневесенних вредителей (шведская муха, блошка крестоцветных)
	III. <i>Вторичная обработка посевов</i> : этикетирование, культивация, обработка от летних вредителей (тля, трипсы)
	IV. <i>Борьба с полеганием растений</i>

Таблица 15 – Продолжение.

	V. <i>Уборка урожая</i>
"Фазы развития"	I. <i>Посев</i> . Планирование эксперимента. Подготовка генетического материала
	II. <i>Всходы</i> . Подсчет всхожести, наблюдения за признаками
	III. <i>Кущение</i> . Наблюдения за признаками
	IV. <i>Стеблевание</i> . Наблюдения за признаками
	V. <i>Цветение</i> . Наблюдения за признаками
	VI. <i>Колошение</i> . Наблюдения за признаками
	VII. <i>Налив</i> . Наблюдения за признаками
	VIII. <i>Полное созревание</i> . Анализ структуры урожая, технологических свойств зерна
"Условия среды"	Определение: температуры воздуха, температуры почвы, влажности почвы, количества осадков
"Фенотипирование"	I. <i>Всходы</i> . Пигментация coleoptile и листа
	II. <i>Кущение</i> . Интенсивность и продолжительность кущения, форма листа, опушение листа
	III. <i>Стеблевание</i> . Длительность, сроки, число побегов
	IV. <i>Цветение</i> . Даты цветения, окраска пыльников, опушение листа, учет грибных заболеваний листа и стебля.
	V. <i>Колошение</i> . Морфологические признаки колоса (окраска, опушение, форма), опушение и морфология листа, учет грибных заболеваний листа и стебля.
	VI. <i>Налив</i> . Наблюдения за фазами налива (молочная, восковая, полная спелость), морфологические признаки окраски колоса, стебля листьев.
	VII. <i>Полное созревание</i> . Анализ структуры урожая, отдельного растения – 12 признаков (главный и вторичный колосья, общий урожай). Анализ технологических свойств зерна и муки (масса 1000 зерен, содержание клейковины в зерне, мукомольные показатели, физические свойства муки и теста).



Для того, чтобы внести набор этих действий в базу данных WheatPGE ее логическая модель была изменена в части описания фенотипа и добавления признаков, связанных с проведением эксперимента. Изменения представлены на рисунке 10. Они включают добавление реляционных таблиц описания фаз развития растений, а так же реляционных таблиц описывающих агротехнические мероприятия.

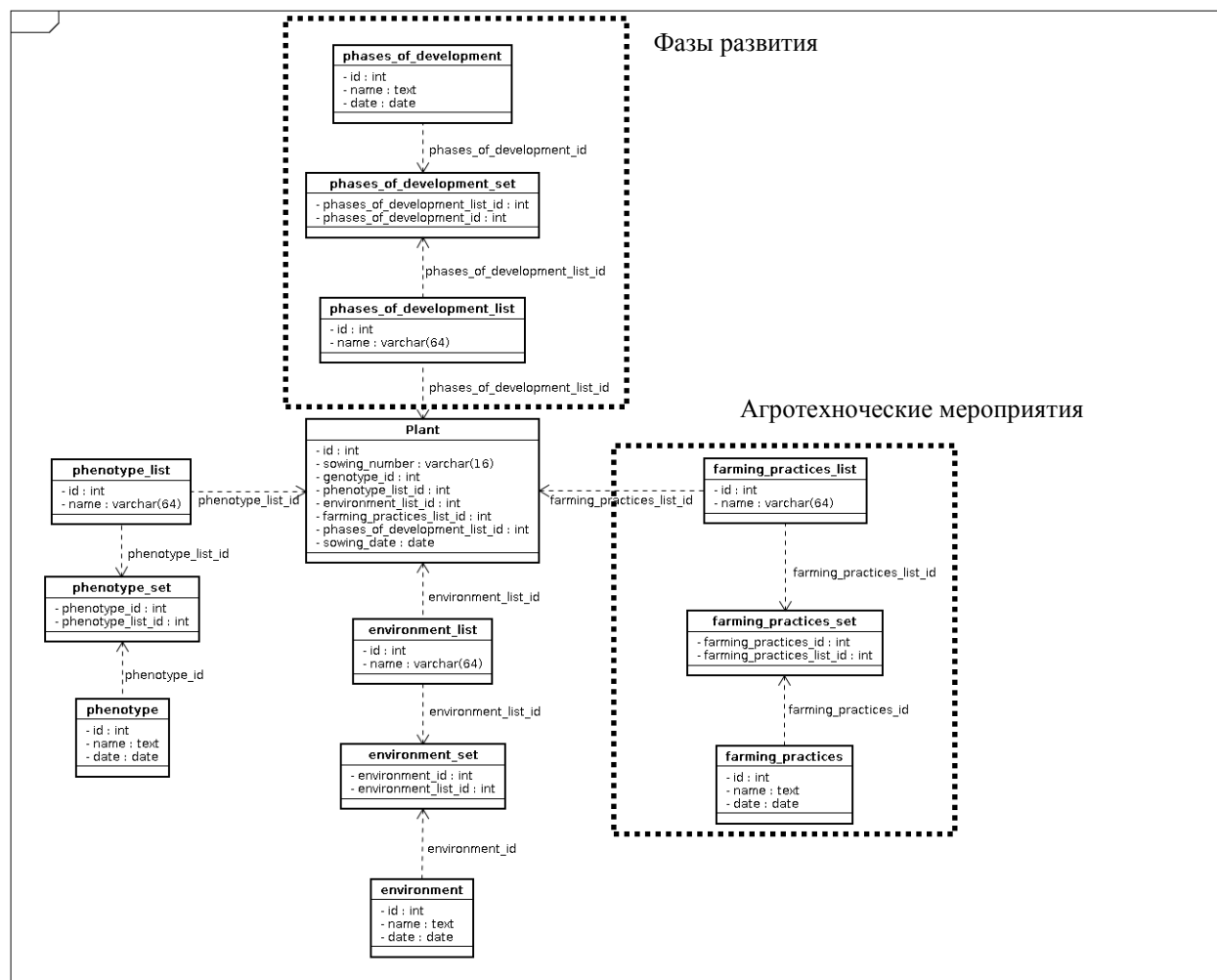


Рисунок 10 - Описание фрагмента логической структуры базы данных WheatPGE, включающий таблицы, описывающие агротехнические мероприятия и фазы развития растений. Эти таблицы с добавленными признаками выделены пунктирными линиями.

#### 4.5 Выводы теоретических исследований

Теоретические позволили сформулировать решение нескольких важных задач первого этапа.

На основе анализа информационных мировых ресурсов в области биотехнологии были отобраны 23 информационных ресурса и разработано представление для этой информации в базе данных внешних Интернет-ресурсов.

Разработаны форматы баз данных промоторов и трансляционных энхансеров для трансгенеза растений.

Разработан формат описания фенотипических растений в базе данных WheatPGE. Это признаки, характеризующие продуктивность растения и его устойчивость к стрессу – опущение. Всего предложено описание 7 признаков колоса, 5 признаков, связанных с продуктивностью растения, и 7 полей, характеризующих описание опущение растения, включая результаты компьютерного изодражения.

Для оценки количественных характеристик опущения предложен алгоритм, основанный на анализе изображения сгиба листа пшеницы.

## **ЗАКЛЮЧЕНИЕ**

### **Краткие выводы по результатам выполнения этапа №1**

В ходе выполнения НИР были решены все запланированные в календарном плане на данный этап задачи и получены следующие основные результаты:

1) Подготовлен аналитический обзор, в котором рассмотрены существующие подходы для планирования генно-инженерных экспериментов и применение высокопроизводительных методов для фенотипирования растений при выполнении селекционно-генетических экспериментов. На основе этого обзора обоснован выбор направления и методики исследований, а также выбор платформ, аппаратных средств и способов представления данных.

2) Проведены патентные исследования по ГОСТ 15.011-96.

3) Составлен перечень внешних Интернет-ресурсов (баз данных и программных комплексов), связанных с процессом получения и использования в биотехнологии ГМО, а также с селекционно-генетическими подходами в биотехнологии.

4) Разработан формат баз данных промоторов и трансляционных энхансеров для трансгенеза.

5) Проведен анализ подходов к высокопроизводительному фенотипированию растений в рамках селекционно-генетических экспериментов.

6) Разработан формат описания фенотипических признаков растений в базе данных WhetPGE.

7) Разработан формат описания селекционно-генетического эксперимента в базе данных WheatPGE.

### **Оценка полноты решения поставленных задач и научно-технического уровня**

Таким образом, все поставленные задачи НИР выполнены в полном объеме, в соответствии с Техническим заданием. Полученные результаты полностью оригинальны. Исследования, в том числе Патентные исследования, проведенные в рамках проекта, показали, что:

Разрабатываемая тема соответствует мировому уровню техники, поскольку нацелена на создание нового информационного ресурса модульного типа, предназначенного для поддержки научно исследовательских разработок в областях агробиотехнологии и создания новых продуктов и биопроцессов с помощью геномных и постгеномных технологий, методов биоинженерии и клеточных технологий.

## **Разработка рекомендаций и исходных данных по конкретному использованию результатов НИР**

Результаты НИР будут представлять собой план проведения НИР в целом (определены направления и методы исследований) и будут использованы в этом качестве.

## **Оценка технико-экономической эффективности внедрения и рекомендации по внедрению**

Конкурентные преимущества нового информационного ресурса позволят рассчитывать на масштабное позиционирование на мировом рынке Российской биоинформационной продукции, поскольку существует устойчивая тенденция увеличения масштабов спроса в области предоставления биоинформационных услуг и рынка принципиально новых биоинформационных продуктов.

В доступных источниках не обнаружено действующих патентов на территории России, под действие которых может подпадать исследуемый объект, в связи с чем можно сделать вывод, что объект исследования обладает патентной чистотой в отношении России по состоянию на 27.10.2011 (последний просмотренный бюллетень «Изобретения. Полезные модели», № 30).

Результаты НИР могут быть рекомендованы к внедрению научными и образовательными организациями России, выполняющими исследования в областях приоритетной деятельности Технологической платформы «Биоиндустрия и биоресурсы — BioTech2030». Высокая актуальность подобных программных средств может служить веским основанием для дальнейшей разработки конкурентно-способного коммерческого программного продукта, как на российском, так и на мировом рынке. Выполнение НИР обеспечит решение задач современной молекулярной биологии, с биотехнологическими проектами в области геной инженерии и агробиологии, что также может рассматриваться как оценка научно-технического уровня.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1) Rad'ko S.P., Il'ina A.P., Bodoev N.V., Archakov A.I. (2007) Synthesis of artificial genome as the basis of synthetic biology. // *Biomed Khim.*, 53, 237-248.
- 2) Attwood T.K., Kell D.B., McDermott P., Marsh J., Pettifer S.R., Thorne D. (2009) Calling International Rescue: knowledge lost in literature and data landslide! // *Biochem J.*, 424: 317-333.
- 3) Werner T. (2008) Bioinformatics applications for pathway analysis of microarray data. // *Curr Opin Biotechnol.*, 19:50-54.
- 4) Lucas M., Laplaze L., Bennett M.J. (2011) Plant systems biology: network matters. // *Plant Cell Environ.*, 34:535-553.
- 5) Picataggio S. (2009) Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. // *Curr Opin Biotechnol.*, 20:325-329.
- 6) Beard D.A. (2011) Simulation of cellular biochemical system kinetics. // *Wiley Interdiscip Rev Syst Biol Med.*, 3:136-146.
- 7) Hubner K., Sahle S., Kummer U. (2011) Applications and trends in systems biology in biochemistry. // *FEBS J.*, 278: 2767-2857.
- 8) Kazakov A.E., Rodionov D.A., Alm E., Arkin A.P., Dubchak I., Gelfand M.S. (2009) Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. // *J Bacteriol.*, 191:52-64.
- 9) Sommer B., Tiys E.S., Kormeier B., Hippe K., Janowski S., Ivanisenko T.V., Bragin A.O., Arrigo P., Demenkov P.S., Kochetov A.V., Ivanisenko V.A., Kolchanov N.A., Hofestädt R. (2010) Visualization and analysis of a cardio vascular disease- and MUPP1-related biological network combining text mining and data warehouse approaches. // *J Integr Bioinform.*, 7, 148.
- 10) Rodionov D.A., Novichkov P.S., Stavrovskaya E.D., Rodionova I.A., Li X., Kazanov M.D., Ravcheev D.A., Gerasimova A.V., Kazakov A.E., Kovaleva G.Y., Permina E.A., Laikova O.N., Overbeek R., Romine M.F., Fredrickson J.K., Arkin A.P., Dubchak I., Osterman A.L., Gelfand M.S. (2011) Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. // *BMC Genomics*, 12 Suppl 1:S3.

- 11) Suvorova I.A., Tutukina M.N., Ravcheev D.A., Rodionov D.A., Ozoline O.N., Gelfand M.S. (2011) Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria. // *J Bacteriol.*, 193:3956-3963.
- 12) Akberdin I.R., Ozonov E.A., Mironova V.V., Omelyanchuk N.A., Likhoshvai V.A., Gorpinchenko D.N., Kolchanov N.A. (2007) A cellular automaton to model the development of primary shoot meristems of *Arabidopsis thaliana*. // *J Bioinform Comput Biol.*, 5:641-650.
- 13) Mironova V.V., Omelyanchuk N.A., Yosiphon G., Fadeev S.I., Kolchanov N.A., Mjolsness E., Likhoshvai V.A. (2010) A plausible mechanism for auxin patterning along the developing root. // *BMC Syst Biol.*, 4:98.
- 14) Kolpakov F., Poroikov V., Sharipov R., Kondrakhin Y., Zakharov A., Lagunin A., Milanesi L., Kel A. (2007) CYCLONET--an integrated database on cell cycle regulation and carcinogenesis. // *Nucleic Acids Res.*, 35:D550-556.
- 15) Gunbin K.V., Afonnikov D.A., Kolchanov N.A. (2009) Molecular evolution of the hyperthermophilic archaea of the *Pyrococcus* genus: analysis of adaptation to different environmental conditions. // *BMC Genomics*, 10:639.
- 16) Feofilova E.P., Sergeeva I.E., Ivashechkin A.A. (2010) Biodiesel-fuel: content, production, producers, contemporary biotechnology. // *Prikl Biokhim Mikrobiol.*, 46:405-415.
- 17) Wiley P.E., Campbell J.E., McKuin B. (2011) Production of biodiesel and biogas from algae: a review of process train options. // *Water Environ Res.*, 83:326-338.
- 18) Ageitos J.M., Vallejo J.A., Veiga-Crespo P., Villa T.G. (2011) Oily yeasts as oleaginous cell factories. // *Appl Microbiol Biotechnol.*, 90:1219-1227.
- 19) Abdeev R.M., Abdeeva I.A., Bruskin S.S., Musiychuk K.A., Goldenkova-Pavlova I.V., Piruzian E.S. (2009) Bacterial thermostable beta-glucanases as a tool for plant functional genomics. // *Gene*, 436: 81-89.
- 20) Wani S.H., Haider N., Kumar H., Singh N.B. (2010) Plant plastid engineering. // *Curr Genomics*, 11:500-512.
- 21) Salyaev R.K., Rigano M.M., Rekoslavskaya N.I. (2010) Development of plant-based mucosal vaccines against widespread infectious diseases. // *Expert Rev Vaccines*, 9:937-946.
- 22) Komarova T.V., Baschieri S., Donini M., Marusic C., Benvenuto E., Dorokhov Y.L. (2010) Transient expression systems for plant-derived biopharmaceuticals. // *Expert Rev Vaccines.*, 9:859-876.

- 23) Peters J., Stoger E. (2011) Transgenic crops for the production of recombinant vaccines and anti-microbial antibodies. // *Hum Vaccin.*, 7:367-374.
- 24) Hassan S.W., Waheed M.T., Lössl A.G. (2011) New areas of plant-made pharmaceuticals. // *Expert Rev Vaccines.*, 10:151-153.
- 25) Bulgakov V.P., Inyushkina Y.V., Fedoreyev S.A. (2011) Rosmarinic acid and its derivatives: biotechnology and applications. // *Crit Rev Biotechnol.* [Epub ahead of print]
- 26) Boehm R. (2007) Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. // *Ann N Y Acad Sci.*, 1102:121-134.
- 27) Xu J., Dolan M.C., Medrano G., Cramer C.L., Weathers P.J. (2011) Towards high-yield production of pharmaceutical proteins with plant cell suspension cultures. // *Biotechnol Adv.* 29(3):278-99.
- 28) Han M.J., Lee S.Y., Koh S.T., Noh S.G., Han W.H. (2010) Biotechnological applications of microbial proteomes. // *J Biotechnol.*, 145:341-349.
- 29) Doyle S. (2011) Fungal proteomics: from identification to function. // *FEMS Microbiol Lett.*, 321:1-9.
- 30) Saito K., Matsuda F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. // *Annu Rev Plant Biol.*, 61:463-489.
- 31) Lee J.W., Kim H.U., Choi S., Yi J., Lee S.Y. (2011a) Microbial production of building block chemicals and polymers. // *Curr Opin Biotechnol.* [Epub ahead of print]
- 32) Lee J.W., Kim T.Y., Jang Y.S., Choi S., Lee S.Y. (2011b) Systems metabolic engineering for chemicals and materials. // *Trends Biotechnol.*, 29, 370-378.
- 33) Misawa N. (2011) Pathway engineering for functional isoprenoids. // *Curr Opin Biotechnol.* [Epub ahead of print]
- 34) Peng Z.Y., Zhou X., Li L., Yu X., Li H., Jiang Z., Cao G., Bai M., Wang X., Jiang C., Lu H., Hou X., Qu L., Wang Z., Zuo J., Fu X., Su Z., Li S., Guo H. (2009) Arabidopsis Hormone Database: a comprehensive genetic and phenotypic information database for plant hormone research in Arabidopsis. // *Nucleic Acids Res.*, 37(Database issue):D975-982.
- 35) Jiang Z., Liu X., Peng Z., Wan Y., Ji Y., He W., Wan W., Luo J., Guo H. (2011) AHD2.0: an update version of Arabidopsis Hormone Database for plant systematic studies. *Nucleic Acids Res.* 39(Database issue):D1123-1129.
- 36) Furtado A., Henry R.J., Pellegrineschi A. (2009) Analysis of promoters in transgenic barley and wheat. // *Plant Biotechnol. J.*, 7:240–253.

- 37) Furtado A., Henry R.J., Takaiwa F. (2008) Comparison of promoters in transgenic rice. // *Plant Biotechnol. J.*, 6:679–693.
- 38) Qu le Q., Xing Y.P., Liu W.X., Xu X.P., Song Y.R. (2008) Expression pattern and activity of six glutelin gene promoters in transgenic rice. // *J. Exp. Bot.*, 59:2417–2424.
- 39) Chang H., Jones M.L., Banowetz G.M., Clark D.G. (2003) Overproduction of cytokinins in petunia flowers transformed with P(SAG12)-IPT delays corolla senescence and decreases sensitivity to ethylene. // *Plant Physiol.*, 132(4): 2174-2183.
- 40) Cowan A.K., Freeman M., Bjorkman P.O., Nicander B., Sitbon F., Tillberg E. (2005) Effects of senescence-induced alteration in cytokinin metabolism on source-sink relationships and ontogenic and stress-induced transitions in tobacco. // *Planta*, 221(6): 801-814.
- 41) Huynh le N., Vantoai T., Streeter J., Banowetz G. (2005) Regulation of flooding tolerance of SAG12:ipt Arabidopsis plants by cytokinin. // *J. Exp. Bot.*, 56(415): 1397-1407.
- 42) McCabe M.S., Garratt L.C., Schepers F., Jordi W.J., Stoopen G.M., Davelaar E., van Rhijn J.H., Power J.B., Davey M.R. (2001) Effects of P(SAG12)-IPT gene expression on development and senescence in transgenic lettuce. // *Plant Physiol.*, 127(2):505-516.
- 43) Merewitz E.B., Gianfagna T., Huang B. (2011) Photosynthesis, water use, and root viability under water stress as affected by expression of SAG12-ipt controlling cytokinin synthesis in *Agrostis stolonifera*. // *J. Exp. Bot.*, 62(1), 383-395.
- 44) Swartzberg D., Dai N., Gan S., Amasino R., Granot D. (2006) Effects of cytokinin production under two SAG promoters on senescence and development of tomato plants. // *Plant Biol. (Stuttg.)*, 8(5):579-586.
- 45) Swartzberg D., Hanael R., Granot D. (2011) Relationship between hexokinase and cytokinin in the regulation of leaf senescence and seed germination. // *Plant Biol (Stuttg.)*, 13(3):439-444.
- 46) Swartzberg D., Kirshner B., Rav-David D., Elad Y., Granot D. (2008) *Botrytis cinerea* induces senescence and is inhibited by autoregulated expression of the IPT gene. // *Eur. J. Plant Pathol.*, 120:289-297.
- 47) Sykorova B., Kuresova G., Daskalova S., Trekova M., Hoyerova K., Raimanova I., Motyka V., Travnickova A., Elliott M.C., Kaminek M. (2008) Senescence-induced ectopic expression of the *A. tumefaciens* ipt gene in wheat delays leaf senescence, increases cytokinin content, nitrate influx, and nitrate reductase activity, but does not affect grain yield. // *J. Exp. Bot.*, 59(2):377-387.



- 48) Wingler A., Brownhill E., Pourtau N. (2005) Mechanisms of the light-dependent induction of cell death in tobacco plants with delayed senescence. // *J. Exp. Bot.*, 56(421):2897-2905.
- 49) Xu Y., Gianfagna T., Huang B. (2010) Proteomic changes associated with expression of a gene (*ipt*) controlling cytokinin synthesis for improving heat tolerance in a perennial grass species. // *J. Exp. Bot.*, 61(12):3273-3289.
- 50) Xu Y., Tian J., Gianfagna T., Huang B. (2009) Effects of SAG12-*ipt* expression on cytokinin production, growth and senescence of creeping bentgrass (*Agrostis stolonifera* L.) under heat stress. // *Plant Growth Regulation*, 57, 281-291.
- 51) Zhang P., Wang W.Q., Zhang G.L., Kaminek M., Dobrev P., Xu J., Gruissem W. (2010) Senescence-inducible expression of isopentenyl transferase extends leaf life, increases drought stress resistance and alters cytokinin metabolism in cassava. // *J. Integr. Plant Biol.*, 52(7):653-669.
- 52) Oszvald M., Gardonyi M., Tamas C., Takacs I., Jenes B., Tamas L. (2008) Development and characterization of a chimaeric tissue-specific promoter in wheat and rice endosperm. *In Vitro Cellular and Developmental Biology*. // *Plant*, 44:1–7.
- 53) Stoger E., Parker M., Christou P., Casey R. (2001) Pea legumin overexpressed in wheat endosperm assembles into an ordered paracrystalline matrix. // *Plant Physiol*, 125(4):1732-42.
- 54) Altpeter F., Varshney A., Abderhalden O., Douchkov D., Sautter C., Kumlehn J., Dudler R., Schweizer P. (2005) Stable expression of a defense-related gene in wheat epidermis under transcriptional control of a novel promoter confers pathogen resistance. // *Plant Mol. Biol.*, 57(2): 271-283.
- 55) Schweizer P. (2008) Tissue-specific expression of a defence-related peroxidase in transgenic wheat potentiates cell death in pathogen-attacked leaf epidermis. // *Mol. Plant Pathology*, 9(1):45–57.
- 56) Higo K., Ugawa Y., Iwamoto M., Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. // *Nucleic Acids Res.*, 27:297–300.
- 57) Yilmaz A., Mejia-Guerra M.K., Kurz K., Liang X., Welch L., Grotewold E. (2011) AGRIS: the arabidopsis gene regulatory information server, an update. // *Nucleic Acids Res.*, 39(Database issue):D1118–D1122.
- 58) Shahmuradov I.A., Gammerman A.J., Hancock J.M., Bramley P.M., Solovyev V.V. (2003) PlantProm: a database of plant promoter sequences. // *Nucleic Acids Res.*, 31:114–117.

- 59) O'Connor T.R., Dyreson C., Wyrick J.J. (2005) Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. *Bioinformatics*, 21:4411–4413.
- 60) Morris R.T., O'Connor T.R., Wyrick J.J. (2008) Osiris: an integrated promoter database for *Oryza sativa* L. // *Bioinformatics*, 24:2915–2917.
- 61) Bülow L., Brill Y., Hehl R. (2010) AthaMap-assisted transcription factor target gene identification in *Arabidopsis thaliana*. *Database (Oxford)* 2010: baq034
- 62) Chang W.C., Lee T.Y., Huang H.D., Huang H.Y., Pan R.L. (2008) PlantPAN: plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. // *BMC Genomics*, 9:561.
- 63) Tittarelli A., Santiago M., Morales A., Meisel L.A., Silva H. (2009) Isolation and functional characterization of coldregulated promoters, by digitally identifying peach fruit cold-induced genes from a large EST dataset. // *BMC Plant Biol.*, 9:121.
- 64) Potenza C., Aleman L., Sengupta-Gopalan C. (2004) Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. // *In Vitro Cell Dev. Biol. Plant*, 40:1–22.
- 65) Jones H.D., Sparks C.A. (2009) Promoter sequences for defining transgene expression. // *Methods Mol. Biol.*, 478:171–184.
- 66) Peremarti A., Twyman R.M., Gymez-Galera S., Naqvi S., Farrq G., Sabalza M., Miralpeix B., Dashevskaya S., Yuan D., Ramessar K., Christou P., Zhu C., Bassie L., Capell T. (2010) Promoter diversity in multigene transformation. // *Plant Mol. Biol.*, 73:363–378.
- 67) Xiao Y.L., Redman J.C., Monaghan E.L., Zhuang J., Underwood B.A., Moskal W.A., Wang W., Wu H.C., Town C.D. (2010) High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in *Arabidopsis* and analysis of their expression patterns. // *Plant Methods*, 6:18.
- 68) Liu Y., Wimmer E., Paul A.V. (2009) Cis-acting RNA elements in human and animal plus-strand RNA viruses. // *Biochim Biophys Acta.* 1789(9-10):495-517.
- 69) Gallie D.R. (2002) The 5'-leader of tobacco mosaic virus promotes translation through enhanced recruitment of eIF4F. *Nucleic Acids Res.* 30(15):3401-3411
- 70) Grillo G., Turi A., Licciulli F., Mignone F., Liuni S., Banfi S., Gennarino V.A., Horner D.S., Pavesi G., Picardi E., Pesole G. (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. // *Nucleic Acids Res.* 38(Database issue):D75-80.

- 71) Jacobs G.H., Chen A., Stevens S.G., Stockwell P.A., Black M.A., Tate W.P., Brown C.M. (2009) Transterm: a database to aid the analysis of regulatory sequences in mRNAs. // *Nucleic Acids Res.*, 37(Database issue):D72-76
- 72) Khurana P., Chauhan H., Desai S.A. (2008) Wheat. In *Compendium of Transgenic Crop Plants, Vol 1. Transgenic Cereals and Forage Grasses* (Kole, C. and Hall, T.C., eds), pp. 83–100, UK: Willey-Blackwell.
- 73) Harwood W.A., Bartlett J.G., Alves S.C., Perry M., Smedley M.A., Leyland N., Snape J.W. (2009) Barley transformation using Agrobacterium-mediated techniques. // *Methods Mol. Biol.*, 478:137-147.
- 74) Travella S., Ross S.M., Harden J., Everett C., Snape J.W., Harwood W.A. (2005) A comparison of transgenic barley lines produced by particle bombardment and Agrobacterium-mediated techniques. // *Plant Cell Rep.*, 12:780-789.
- 75) Patnaik D., Vishnudasana D., Paramjit Khurana P. (2006) Agrobacterium-mediated transformation of mature embryos of *Triticum aestivum* and *Triticum durum*. // *Current Science*, 91(3):307-317.
- 76) Shrawat A.K., Lörz H. (2006) Agrobacterium-mediated transformation of cereals: a promising approach crossing barriers. // *Plant Biotechnol. J.* 4:575–603.
- 77) Supartana P., Shimizu T., Nogawa M., Shioiri H., Nakajima T., Haramoto N., Nozue M., Kojima M. (2006) Development of simple and efficient in planta transformation method for wheat (*Triticum aestivum* L.) using *Agrobacterium tumefaciens*. // *J. Biosci. Bioeng.* 102:162–170.
- 78) Holme I.B., Brinch-Pedersen H., Lange M., Holm P.B. (2008) Transformation of different barley (*Hordeum vulgare* L.) cultivars by *Agrobacterium tumefaciens* infection of in vitro cultured ovules. // *Plant Cell Rep.* 27: 1833–1840.
- 79) Kumlehn J., Hensel G. (2009) Genetic transformation technology in the Triticeae. *Breeding Science* 59: 553–560.
- 80) Shrawat A.K., Becker D., Lörz H. (2007) *Agrobacterium tumefaciens*-mediated genetic transformation of barley (*Hordeum vulgare* L.) // *Plant Sci.* 172:281–290.
- 81) Lange M., Vincze E., Moller M.G., Holm P.B. (2006) Molecular analysis of transgene and vector backbone integration into the barley genome following Agrobacterium-mediated transformation. // *Plant Cell Rep.* 25, 815–820.
- 82) Wu H. X., Sparks C.A., Jones H.D. (2006) Characterisation of T-DNA loci and vector backbone sequences in transgenic wheat produced by Agrobacterium-mediated transformation. // *Mol. Breed.* 18:195–208.

- 83) Hensel G., Kastner C., Oleszczuk S., Riechen J., Kumlehn J. (2009) Agrobacterium-mediated gene transfer to cereal crop plants: Current protocols for barley, wheat, triticale and maize. // *Int. J. of Plant Genomics*, doi:10.1155/2009/835608.
- 84) Chauhan H., Khurana P. (2011) Use of doubled haploid technology for development of stable drought tolerant bread wheat (*Triticum aestivum* L.) transgenics. // *Plant Biotechnol. J.*, 9(3):408-417.
- 85) Shrawat A.K., Good A.G. (2011) Agrobacterium tumefaciens-mediated genetic transformation of cereals using immature embryos. // *Methods Mol Biol.*, 710:355-372.
- 86) Vasil I.K., Vasil V. (2006) Transformation of wheat via particle bombardment.// *Methods Mol. Biol.* 318:273–283.
- 87) Yao Q., Cong L., Chang J.L., Li K.X., Yang G.X., He G.Y. (2006) Low copy number gene transfer and stable expression in a commercial wheat cultivar via particle bombardment. // *J. Exp. Bot.* 57:3737–3746.
- 88) Yao Q., Cong L., He G., Chang J., Li K., Yang G. (2007) Optimization of wheat co-transformation procedure with gene cassettes resulted in an improvement in transformation frequency. // *Mol. Biol. Rep.*34:61–67.
- 89) Shim Y.S., Pauls K.P., Kasha K.J. (2009) Transformation of isolated barley (*Hordeum vulgare* L.) microspores: I. The influence of pretreatments and osmotic treatment on the time of DNA synthesis. // *Genome* 52: 166–174.
- 90) Badr Y.A., Kereim M.A., Yehia M.A., Fouad O.O., Bahieldin A. (2005) Production of fertile transgenic wheat plants by laser micropuncture. // *Photochem. Photobiol. Sci.*, 4: 803–807.
- 91) Liang H., Wu F.S., Wang D.W., Sun D.F., Jia X.. (2005) Wheat transformation by electroporation with ring electrode. // *Yi Chuan Xue Bao.*, 32(1):66-71.
- 92) Dahleen L.S., Manoharan M. (2007) Recent advances in barley transformation. // *In Vitro Cell. Dev. Biol. Plant*, 43:493-506.
- 93) Wang Z.-Y., Ge Y. (2006) Recent advances in genetic transformation of forage and turf grasses. // *In Vitro Cell. Dev. Biol.Plant*, 42:1-18.
- 94) Dunwell J.M. (2009) Transgenic Wheat, Barley and Oats: Future Prospects. // *Methods in Mol. Biol.*, 478: 333-345.
- 95) Kumlehn J., Serazetdinova L., Hensel G., Becker D., Lörz H. (2006) Genetic transformation of barley (*Hordeum vulgare* L.) via infection of androgenetic pollen cultures with Agrobacterium tumefaciens.// *Plant Biotechnol. J.* 4:251–261.

- 96) Pritchard L., Birch P. (2011) A systems biology perspective on plant-microbe interactions: biochemical and structural targets of pathogen effectors. // *Plant Sci.*, 180:584-603.
- 97) Trifonova E.A., Sapotsky M.V., Komarova M.L., Scherban A.B., Shumny V.K., Polyakova A.M., Lapshina L.A., Kochetov A.V., Malinovsky V.I. (2007) Protection of transgenic tobacco plants expressing bovine pancreatic ribonuclease against tobacco mosaic virus. // *Plant Cell Rep.*, 26:121-1126.
- 98) Khadeeva N.V., Kochieva E.Z., Tcherednitchenko M.Y., Yakovleva E.Y., Sydoruk K.V., Bogush V.G., Dunaevsky Y.E., Belozersky M.A. (2009) Use of buckwheat seed protease inhibitor gene for improvement of tobacco and potato plant resistance to biotic stress. // *Biochemistry (Mosc)*, 74:260-267.
- 99) Shukurov R.D., Voblikova V., Nikonorova A.K., Komakhin R.A.V., Komakhina V. A. Egorov T.V., Grishin E.V., Babakov A. (2011) Transformation of tobacco and Arabidopsis plants with *Stellaria media* genes encoding novel hevein-like peptides increases their resistance to fungal pathogens. // *Transgenic Res.* [Epub ahead of print]
- 100) Varshney R.K., Bansal K.C., Aggarwal P.K., Datta S.K., Craufurd P.Q. (2011) Agricultural biotechnology for crop improvement in a variable climate: hope or hype? // *Trends Plant Sci.*, 16:363-371.
- 101) Zhengbin Z, Ping X, Hongbo S, Mengjun L, Zhenyan F, Liye C (2011) Advances and prospects: Biotechnologically improving crop water use efficiency. // *Crit Rev Biotechnol.* [Epub ahead of print].
- 102) Dhar M.K., Kaul S., Kour J. (2011) Towards the development of better crops by genetic transformation using engineered plant chromosomes. // *Plant Cell Rep.*, 30:799-806.
- 103) Skryabin K. (2010) Do Russia and Eastern Europe need GM plants? // *N Biotechnol.*, 27, 593-595.
- 104) Weckwerth W. (2011) Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. // *J Proteomics* [Epub ahead of print]
- 105) Kumar G.R., Sakthivel K., Sundaram R.M., Neeraja C.N., Balachandran S.M., Rani N.S., Viraktamath B.C., Madhav M.S. (2010) Allele mining in crops: prospects and potentials. // *Biotechnol Adv.*, 28, 451-461.
- 106) Becker D., Brettschneider R., Lörz H. (1994) Fertile transgenic wheat from microprojectile bombardment of scutellar tissue. // *Plant J.* 5:299-307.

- 107) Rasco-Gaunt S., Riley A., Cannell M., Barcelo P. and Lazzeri. P. A. (2000) Procedures allowing the transformation of a range of European elite wheat (*Triticum aestivum* L.) varieties via particle bombardment // *Journal of Experimental Botany*, 52(357):865-874.
- 108) Mihály R., Kótai E., Kiss O., Pauk J. (2002) In vitro selection of transformed foreign gene (bar) in wheat anther culture // *Proceedings of the 7th Hungarian Congress on Plant Physiology- 2002*, Vol.46, P:9-10.
- 109) Landjeva S., Korzun V., Börner A. (2007) Molecular markers: actual and potential contributions to wheat genome characterization and breeding // *Euphitica*, 156:271-276.
- 110) Joshi S., Ranjekar P., Gupta V. (1999) Molecular markers in plant genome analysis. // *Curr Sci*, 77: 230–240.
- 111) Gupta P.K., Roy J.K., Prasad M. (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. // *Curr. Sci.*, 80:524–535.
- 112) Langridge P., Lagudah E.S., Holton T.A., Appels R., Sharp P.J., Chalmers K.J. (2001) Trends in genetic and genome analyses in wheat: a review. // *Aust J Agric Res*, 52:1043–1077.
- 113) Korzun V., Ebmeyer E. (2003) Molecular markers and their applications in wheat breeding. In: Pogna NE, Romano M, Pogna EA, Galterio G (eds) *Proceedings of the 10th International Wheat Genetics Symposium*, Istituto Sperimentale per la Cerealicoltura, Rome, Italy, vol 1, P:140–143
- 114) Mohler V., Schwarz G. (2004) Genotyping tools in plant breeding: from restriction fragment length polymorphisms to single nucleotide polymorphisms. In: Lörz H, Wenzel G (eds) *Molecular marker systems in plant breeding and crop improvement. Biotechnology in agriculture and forestry Springer -2004*, Vol 55. P. 23–38.
- 115) Röder M., Huang X., Ganal M. (2004) Wheat microsatellites: potential and implications. In: Lörz H, Wenzel G (eds) *Molecular marker systems in plant breeding and crop improvement. Biotechnology in agriculture and forestry// Springer -2004* vol 55. Springer, P. 255–266.
- 116) Varshney R., Graner A., Sorrells M. (2005) Genic microsatellite markers in plants: features and applications.// *Trends in Biotechnology*, 23:48–55.

- 117) Varshney R., Korzun V., Börner A. (2004) Molecular maps in cereals: methodology and progress. In: Gupta P, Varshney R (eds) *Cereal Genomics*, Kluwer Academic Publishers- 2004, P: 35–82.
- 118) Varshney R., Sigmund R., Börner A., Korzun V., Stein N., Sorrels M.E., Langridge P., Graner A. (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. // *Plant Science*, 168:195–202.
- 119) Хлесткина Е.К. и Салина Е.А. (2006) SNP-маркеры: методы анализа, способы разработки и сравнительная характеристика на примере мягкой пшеницы // *Генетика*, 42(6)Ж725-736.
- 120) Sourdille P., Cadalen T., Guyomarc'h H., Snape J., Perretant M., Charmet G., Boeuf C., Bernard S., Bernard M. (2003) An update of the Courtot x Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. // *Theor Appl Genet*, 106:530–538.
- 121) Devos K.M., Gale M.D. (1997) Comparative genetics in the grasses.// *Plant. Mol. Biol.*, 35:3–15
- 122) Korzun V., Malyshev S., Voylokov A., Börner A. (1997) RFLP-based mapping of three mutant loci in rye (*Secale cereale* L.) and their relation to homoeologous loci within Gramineae. // *Theor Appl Genet* 95:468–473.
- 123) Bailey P.C., McKibbin R.S., Lenton J.R., Holdsworth M.J., Flintham J.E., Gale M.D. (1999) Genetic map locations for orthologous Vp1 genes in wheat and rice. // *Theor Appl Genet*. 98: 281–284.
- 124) Snape J.W., Sarma R., Quarrie S.A., Fish L., Galiba G., Sutka J. (2001) Mapping genes for Xowering time and frost tolerance in cereals using precise genetic stocks. // *Euphytica*, 120:309–315.
- 125) Börner A. (2002) Gene and genome mapping in cereals.// *Cell Mol Biol Lett*, 7:423–429.
- 126) Galiba G., Pecchioni N., Vágújfalvi A., Francia E., Tóth B., Barabaschi D., Barilli S., Crosatti C., Cattivelli L., Stanca M.A. (2005) Localization of QTLs and candidate genes involved in the regulation of frost resistance in cereals. In: Tuberosa R, Phillips RL, Gale MD (eds) *Proceedings of the International Congress: In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution*, Bologna, Italy, 2005 Avenue Media, Bologna, Italy, pp 253–266.

- 127) Крупнов В.А., Цапайкин А.П. (1990) Опушение листьев пшеницы: генетические и экологические аспекты // С.х. биология. Сер. Биол. растений. 1:51–57.
- 128) Лихенко И.Е. (2007) О взаимосвязи опушения органов растений яровой мягкой пшеницы с хозяйственно и биологически ценными признаками в условиях западной Сибири // Растениеводство и селекция. 6:25–31.
- 129) Schillinger J.A., Gallun R.L. (1968) Leaf pubescence of wheat as a deterrent to the cereal leaf beetle, *Oulema melanopus* // *Annals Entomol. Soc. Amer.*, 61(4):900–903.
- 130) Roberts J.J., Gallun R.L., Patterson F.L. et al. (1979) Effects of wheat leaf pubescence on the hessian fly // *J. Econ. Entomol.* 72(2): 211–214.
- 131) Журавлёва Н.А. (1992) Механизм устьичных движений, продукционный процесс и эволюция. Новосибирск: Наука, 1992. 140 с.
- 132) Майстренко О.И. (1976) Идентификация и локализация генов, контролирующих опушения листа молодых растений мягкой пшеницы // *Генетика.* 12(1): 5–15.
- 133) Dobrovolskaya O.B., Pshenichnikova T.A., Arbuzova V.S. et al. (2007) Molecular mapping of genes determining hairy leaf character in common wheat with respect to other species of the *Triticeae* // *Euphytica*, 155 (3): 285–293.
- 134) Вавилов Н.И. Теоретические основы селекции. М.: Наука, 1990. 512 с.
- 135) Hristov N., Novica Mladenov N., Djuric V., Kondic-Spika A., Marjanovic-Jeromela A., Simic D. (2010) Genotype by environment interactions in wheat quality breeding programs in southeast Europe. // *Euphytica*. 174:315–324.
- 136) Gomez-Becerra H.F., Abugalieva A., Morgounov A. Abdullayev K., Bekenova L., Yessimbekova M., Sereda G., Shpigun S., Tsygankov V., Zelenskiy Yu., Pena R.J, Cakmak I. (2010) Phenotypic correlations, G x E interactions and broad sense heritability analysis of grain and flour quality characteristics in high latitude spring bread wheats from Kazakhstan and Siberia. // *Euphytica*, 171:23–38.
- 137) Mansur L.M., Qualset C.O., Kasarda D.D. et al. (1990) Effects of ‘Cheyenne’ chromosomes on milling and baking quality in ‘Chinese Spring’ wheat in relation to glutenin and gliadin storage proteins // *Crop Sci.* 30: 593-602.
- 138) Arbuzova V.S., Ermakova M.F., Popova R.K. Studies of monosomic lines of cv. Saratovskaya 29 on productivity and grain technological properties // *EWAC Newsletter*, Proc.11<sup>th</sup> EWAC Conference, Novosibirsk, Russia. 80-82, 2001.



- 139) Li L., Wang A., Appels R., Ma J., Xia X., Lan P., He Zh., Bekes F., Yan Y., Ma W. A (2009) Maldi-TOF based analysis of high molecular weight glutenin subunits for wheat breeding. // *J. Cereal Sci.* 50:295-301.
- 140) He Zh., Liu L., Xia X., Liu J.J., Pena R.J. (2005) Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread and noodle quality of Chinese bread wheats. // *Cereal Chem.* 82:345-350.
- 141) Ma W., Appels R., Bekes F., Larroque O., Morell M.K., Gale K.R. (2005) Genetic characterization of dough rheological properties in a wheat double haploid population: additive genetic effects and epistatic interactions. // *Theor. Appl. Genet.* 111:410-422.
- 142) Giroux M.J., Morris C.F. (1998) Wheat grain hardness results from highly conserved mutation in the friabilin components puroindoline a and b. // *Proc Natl Acad Sci USA.* 95: 6262-6266.
- 143) Greffeuille V., Abecassis J., Rousset M., Oury F.-X., Faye A., Bar L'Helgouac'h C., Lullien-Pellerin V. (2006) Grain characterization and milling behavior of near-isogenic lines differing by hardness. // *Theor. Appl. Genet.* 114:1–12.
- 144) Ozturk A., Aydin F. (2004) Effect of water stress at various growth stages on some quality characteristics of winter wheat. // *J. Agronomy & Crop Science.* 190:93-99.
- 145) Пшеничникова Т.А., Ермакова М.Ф., Чистякова А.К., Щукина Л.В., Березовская Е.В., Лохвассер У., Рёдер М., Бёрнер А. (2008а) Картирование локусов количественных признаков (QTL), ассоциированных с показателями качества зерна мягкой пшеницы, выращенного в различных условиях среды. // *Генетика.* 44 (1): 74-84.
- 146) Arbelbide M., Bernardo R. (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. // *Theor. Appl. Genet.*, 112: 885–890.
- 147) Lohwasser U., Roeder M.S., Boerner A. (2005) QTL mapping of the domestication traits pre-harvest sprouting and dormancy in wheat (*Triticum aestivum* L.). *Euphytica.* 143: 247–249.
- 148) Kirigwi F.M., Van Ginkel M., Brown-Guedira G., Gill B.S., Paulsen G.M, Fritz A.K. (2007) Markers associated with a QTL for grain yield in wheat under drought// *Mol. Breeding.* 20:401-413.
- 149) Пшеничникова Т.А., Осипова С. В., Пермякова М. Д., Митрофанова Т.Н., Труфанов В.А, Лохвассер У., Рёдер М., Бёрнер А. (2008b) Картирование локусов

- количественных признаков (QTL), ассоциированных с активностью дисульфидредуктазы и липоксигеназы в зерне мягкой пшеницы *Triticum aestivum* L. // Генетика, 44 (5): 654-662.
- 150) Nelson J.C., Andreescu C., Breseghello F. et al. (2006) Quantitative trait locus analysis of wheat quality traits // *Euphytica*. 149: 145-159.
- 151) Mauricio R (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. // *Nat. Rev. Genet.*, 2: 370-381.
- 152) Shavrukov Y., Gupta N.K., Miyazaki J., Baho M.N., Chalmers K.J., Tester M., Langridge P., Collins N.C. (2010) HvNax3--a locus controlling shoot sodium exclusion derived from wild barley (*Hordeum vulgare* ssp. *spontaneum*). // *Funct Integr Genomics*, 10:277-291.
- 153) Kearsey J.M. (1998) The principles of QTL analysis (a minimal mathematic approach). // *J.Exp.Bot.*, 49:1619-1623.
- 154) Mardis E.R. (2008) The impact of next-generation sequencing technology on genetics. // *Trends Genet.* 24:133-141.
- 155) Eberius M., Lima-Guerra J. (2009) High-throughput plant phenotyping – data acquisition, transformation, and analysis. In *Bioinformatics: Tools and Applications*, Ed.: D. Edwards et al., Springer Science+Business Media, LLC 2009, p. 259-278.
- 156) Ajjawi I., Lu Y., Savage L.J., Bell S.M., Last R.L. (2010) Large-scale reverse genetics in *Arabidopsis*: case studies from the Chloroplast 2010 Project. // *Plant Physiol.*, 152:529-540.
- 157) Brachi B., Faure N., Horton M., Flahauw E., Vazquez A., Nordborg M., Bergelson J., Cuguen J., Roux F. (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. // *PLoS Genet.*, 6: e1000940.
- 158) Yazdanbakhsh N., Fisahn J. (2010) Analysis of *Arabidopsis thaliana* root growth kinetics with high temporal and spatial resolution. // *Ann Bot.*, 105:83-791.
- 159) Golzarian M.R., Frick R.A., Rajendran K., Berger B., Roy S., Tester M., Lun D.S. (2011) Accurate inference of shoot biomass from high-throughput images of cereal plants. // *Plant Methods*, 7:2.
- 160) Herridge R.P., Day R.C., Baldwin S., Macknight R.C. (2011) Rapid analysis of seed size in *Arabidopsis* for mutant and QTL discovery. // *Plant Methods*, 7:3.
- 161) Darrigues A., Hall J., van der Knaap E., Francis D.M., Dujmovic N. and Gray S. (2008) Tomato Analyzer-color Test: A New Tool for Efficient Digital Phenotyping. // *J. Amer. Soc. Hort. Sci.*, 133:579–586.

- 162) Di Cui, Zhang Q., Li M., Hartman G.L., Zhao Y. (2010) Image processing methods for quantitatively detecting soybean rust from multispectral images. // *Biosystems engineering*, 107 :186-193.
- 163) Драгавцев В.А., Цильке Р.А., Рейтер Б.Г., и др. (1984) Генетика признаков продуктивности яровых пшениц в Западной Сибири, Новосибирск, СО АН, 1984, 121 с.
- 164) Драгавцев В.А. (2005) Библиография деятелей науки, ГНУ ГНЦ РФ ВИР, СПб, 2005, с. 6.
- 165) Hancock J.M., Adams N.C., Aidinis V., Blake A., Bogue M., Brown S.D., Chesler E.J, Davidson D., Duran C., Eppig J.T., Gailus-Durner V., Gates H., Gkoutos G.V., Greenaway S., Hrabé de Angelis M., Kollias G., Leblanc S., Lee K., Lengger C., Maier H., Mallon A.M., Masuya H., Melvin D.G., Müller W., Parkinson H., Proctor G., Reuveni E., Schofield P., Shukla A., Smith C., Toyoda T., Vasseur L., Wakana S., Walling A., White J., Wood J., Zouberakis M., (2007) Mouse Phenotype Database Integration Consortium. // *Mamm Genome*. 18(3):157-63.
- 166) Lee J.M., Davenport G.F., Marshall D., Ellis T.H., Ambrose M.J., Dicks J., van Hintum T.J., Flavell A.J. (2005) GERMINATE. a generic database for integrating genotypic and phenotypic information for plant genetic resource collections. // *Plant Physiol*. 139:619-631.
- 167) Kaminuma E., Yoshizumi T., Wada T., Matsui M., Toyoda T. (2008) Quantitative analysis of heterogeneous spatial distribution of Arabidopsis leaf trichomes using micro X-ray computed tomography. // *Plant J.*, 56:470-482.
- 168) Lu Y., Savage L.J., Ajjawi I., Imre K.M., Yoder D.W., Benning C., Dellapenna D., Ohlrogge J.B., Osteryoung K.W., Weber A.P., Wilkerson C.G., Last R.L. (2008) New connections across pathways and cellular processes: industrialized mutant screening reveals novel associations between diverse phenotypes in Arabidopsis. // *Plant Physiol.*, 146:1482-1500.
- 169) Exner V., Hirsch-Hoffmann M., Gruissem W. and Hennig L. (2008) PlantDB – a versatile database for managing plant research. // *Plant Methods*, 4:1.
- 170) Vankadavath R.N., Hussain A.J., Bodanapu R. et.al. (2009) Computer aided data acquisition tool for high-throughput phenotyping of plant populations. // *Plant Methods*, 5:18.
- 171) Montes J.M., Melchinger A.E., Reif J.C. (2007) Novel throughput phenotyping platforms in plant genetic studies. // *Trends Plant Sci.*, 12:433-436.

- 172) Fabre J., Dauzat M., Nègre V., Wuyts N., Tireau A., Gennari E., Neveu P., Tisné S., Massonnet C., Hummel I., Granier C. (2011) PHENOPSIS DB: an information system for Arabidopsis thaliana phenotypic data in an environmental context. // BMC Plant Biol., 11:77.
- 173) Генаев МА, Дорошков АВ, Афонников ДА (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.
- 174) Hartmann A., Czauderna T., Hoffmann R., Stein N., Schreiber F. (2011) HTPPheno: an image analysis pipeline for high-throughput plant phenotyping. // BMC Bioinformatics, 12:148.
- 175) Joosen R.V., Kodde J., Willems L.A., Ligterink W., van der Plas L.H., Hilhorst H.W. (2010) GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. // Plant J., 62:148-159.
- 176) Iyer-Pascuzzi A.S., Symonova O., Mileyko Y., Hao Y., Belcher H., Harer J., Weitz J.S., Benfey P.N. (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. // Plant Physiol., 152:1148-1157.
- 177) Bensch R., Ronneberger O., Greese B., Fleck C., Wester K., Hulskamp M., Burkhardt H. (2009) Image analysis of Arabidopsis trichome patterning in 4d confocal datasets. In Proc of the The Sixth IEEE International Symposium on Biomedical Imaging (ISBI'09); 2009:742-745.
- 178) Ковалев В.А. (2008) Анализ текстуры трехмерных медицинских изображений — Минск: Издательство «Белорусская наука», 2008
- 179) Rao A.R., Cecchi G.A. (2009) High-Throughput Image Reconstruction and Analysis (Bioinformatics & Biomedical Imaging), 2009.
- 180) Федоров А. Бинаризация черно-белых изображений: состояния и перспективы развития. 2008, <http://iu5.bmstu.ru/philippovicha/ITS/IST4b/ITS4/Fyodorov.htm>.
- 181) Feng M.-L., Tan Y.-P. (2004) Contrast adaptive binarization of low quality document images. // School of Electrical and Electronic Engineering, Nanyang Technological University. IEICE Electronics Express, Vol. 1, No. 16, pp. 501 – 506.
- 182) Brox T., Kim Y.-J., Weickert J., and Feiden W. (2006) Fully-automated Analysis of Muscle Fiber Images with Combined Region and Edge Based Active Contours, In Proc. Bildverarbeitung fur die Medizin, 2006, 86-90

- 183) Leifer A.M., Fang-Yen C., Gershow M., Alkema M.J., Samuel A.D. (2011) Optogenetic manipulation of neural activity in freely moving *Caenorhabditis elegans*. // *Nat. Methods*, 8:147-152.
- 184) Деменков П.С., Аман Е.Э., Иванисенко В.А. (2008) Associative network discovery (AND) - компьютерная система для автоматической реконструкции сетей ассоциативных знаний о молекулярно-генетических взаимодействиях. // *Вычислительные технологии*, 13, 15-19.
- 185) Zdobnov E.M., Lopez R., Apweiler R. and Etzold T. (2002) The EBI SRS server—recent developments. // *Bioinformatics*, 18:368-373.
- 186) Ananko E.A., Kondrakhin Y.V., Merkulova T.I., Kolchanov N.A. (2007) Recognition of interferon-inducible sites, promoters, and enhancers. // *BMC Bioinformatics.*, 8:56.
- 187) Kolchanov N.A., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Stepanenko I.L., Merkulova T.I., Pozdnyakov M.A., Podkolodny N.L., Naumochkin A.N., Romashchenko A.G. (2002) Transcription Regulatory Regions Database (TRRD): its status in 2002. // *Nucleic Acids Res.*, 30:312-317.
- 188) Ananko E.A., Podkolodny N.L., Stepanenko I.L., Podkolodnaya O.A., Rasskazov D.A., Miginsky D.S., Likhoshvai V.A., Ratushny A.V., Podkolodnaya N.N., Kolchanov N.A (2005) GeneNet in 2005. // *Nucleic Acids Res.*, 33:D425-427.
- 189) Рогачев С. Обобщенный Model-View-Controller [Электронный ресурс]/ Russian Software Developer Network; — Электрон. дан. — М., [199—]. — . — Режим доступа: <http://rsdn.ru/article/patterns/generic-mvc.xml>, свободный. — Загл. с экрана.— Яз. рус..
- 190) Tittarelli A., Milla L., Vargas F., Morales A., Neupert C., Meisel L.A., Salvo-G H., Penaloza E., Munoz G., Corcuera L.J., Silva H. (2007) Isolation and comparative analysis of the wheat TaPT2 promoter: identification in silico of new putative regulatory motifs conserved between monocots and dicots.// *J. Exp. Bot.*, 58(10):2573-2582.
- 191) Van Herpen T.W., Riley M., Sparks C., Jones H.D., Gritsch C., Dekking E.H., Hamer R.J., Bosch D., Salentijn E.M., Smulders M.J., Shewry P.R., Gilissen L.J. (2008) Detailed analysis of the expression of an alpha-gliadin promoter and the deposition of alpha-gliadin protein during wheat grain development. // *Ann. Bot.*, 102(3):331-342.
- 192) Piston F., Marin S., Hernando A., Barro F. (2009) Analysis of the activity of a gamma-gliadin promoter in transgenic wheat and characterization of gliadin synthesis in wheat by MALDI-TOF during grain development. // *Mol. Breed.*, 23(4): 655-667.

- 193) Vickers C.E., Xue G., Gresshoff P.M. (2006) A novel cis-acting element, ESP, contributes to high-level endosperm-specific expression in an oat globulin promoter. // *Plant Mol. Biol.*, 62:195–214.
- 194) Kim T.W., Goo Y.M., Lee C.H., Lee B.H., Bae J.M., Lee S.W. (2009) The sweet potato ADP-glucose pyrophosphorylase gene (ibAGP1) promoter confers high-level expression of the GUS reporter gene in the potato tuber. // *C R Biol*, 332:876–885.
- 195) Chen L., Tu Z., Hussain J., Cong L., Yan Y., Jin L., Yang G., He G. (2010) Isolation and heterologous transformation analysis of a pollen-specific promoter from wheat (*Triticum aestivum* L.). // *Mol. Biol. Rep.*, 37(2): 737-744.
- 196) Qu L.Q., Takaiwa F. (2004) Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. // *Plant Biotechnol. J.*, 2:113–125.
- 197) Tiwari S., Spielman M., Day R.C., Scott R.J. (2006) Proliferative phase endosperm promoters from *Arabidopsis thaliana*. // *Plant Biotechnol. J.*, 4:393–407.
- 198) Eskelin K., Ritala A., Suntio T., Blumer S., Holkeri H., Wahlström E.H., Baez J., Mäkinen K., Maria N.A. (2009) Production of a recombinant full-length collagen type I alpha-1 and of a 45-kDa collagen type I alpha-1 fragment in barley seeds. // *Plant Biotechnol. J.*, 7:657–672.
- 199) Xu Z.S., Ni Z.Y., Liu L., Nie L.N., Li L.C., Chen M., Ma Y.Z. (2008) Characterization of the TaAIDFa gene encoding a CRT/DRE-binding factor responsive to drought, high-salt, and cold stress in wheat. // *Mol. Genet. Genomics*, 280(6):497-508.
- 200) Egawa C., Kobayashi F., Ishibashi M., Nakamura T., Nakamura C., Takumi S. (2006) Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. // *Genes Genet. Syst.*, 81(2): 77-91.
- 201) Kobayashi F., Ishibashi M., Takumi S. (2008) Transcriptional activation of *Cor/Lea* genes and increase in abiotic stress tolerance through expression of a wheat DREB2 homolog in transgenic tobacco. // *Transgenic Res.*, 17(5):755-767.
- 202) Yang W., Liu X.D., Chi X.J., Wu C.A., Li Y.Z., Song L.L., Liu X.M., Wang Y.F., Wang F.W., Zhang C., Liu Y., Zong J.M., Li H.Y. (2011) Dwarf apple MbDREB1 enhances plant tolerance to low temperature, drought, and salt stress via both ABA-dependent and ABA-independent pathways. // *Planta*, 233:219–229.

Российская академия наук

УЧРЕЖДЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК  
ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РАН

УДК 577.21 004.65 004.932.72  
№ госрегистрации

УТВЕРЖДАЮ  
Директор ИЦиГ СО РАН,  
академик РАН

Н.А. Колчанов

01.11.2011

О Т Ч Е Т О ПАТЕНТНЫХ ИССЛЕДОВАНИЯХ

по теме:

«Разработка информационного ресурса модульного типа для поддержки исследований, проводимых в рамках Технологической платформы "Биоиндустрия и биоресурсы — BioTech2030" в областях агробιοтехнологии и биоинженерии»

Этап работы: «Выбор направления исследований и теоретическое исследование поставленных перед НИР задач»  
(промежуточный)

ГК № 07.514.11.4052 от 12.10.2011, шифр 2011-1.4-514-111-000-018

Научный руководитель  
к.б.н. доцент

  
\_\_\_\_\_

А.В. Кочетов

подпись, дата

Нормоконтролер

  
\_\_\_\_\_

Н.Н. Козырева

подпись, дата

Новосибирск 2011

СПИСОК ИСПОЛНИТЕЛЕЙ:

Вед. инженер по патентной и  
изобретательской работе

  
подпись, дата

1.11.2011

Кучумова Л.Я.  
разделы 2, 3  
заключение,  
приложения А-В)


с.н.с., к.б.н.

  
подпись, дата

1.11.2011

Смирнова О.Г.  
(раздел 1)

Нормоконтролер

  
подпись, дата

4.11.2014

Козырева Н.Н.



## СОДЕРЖАНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ, УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ, ЕДИНИЦ, ТЕРМИНОВ	105
ОБЩИЕ ДАННЫЕ ОБ ОБЪЕКТЕ ИССЛЕДОВАНИЯ	107
ОСНОВНАЯ (АНАЛИТИЧЕСКАЯ) ЧАСТЬ	108
1 Технический уровень и тенденции развития объекта исследования	108
1.1 Использование генетически-модифицированных организмов (ГМО) в биотехнологии	111
1.2 Генетическое картирование и фенотипирование в селекции пшеницы	114
1.3 Компании, оперирующие на рынке агrobiотехнологий	122
2 Использование объектов промышленной собственности и их правовая охрана	125
3 Исследование патентной чистоты объекта исследования	127
ЗАКЛЮЧЕНИЕ	128
Приложение А: Задание на проведение патентных исследований	130
Приложение Б: Регламент поиска	131
Приложение В: Отчет о поиске	133

## ПЕРЕЧЕНЬ СОКРАЩЕНИЙ, УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ, ЕДИНИЦ, ТЕРМИНОВ

В настоящем отчете о патентных исследованиях применяются следующие термины, условные обозначения и сокращения с соответствующими определениями:

БД – база данных

ГМО - генетически-модифицированные организмы

ДНК – дезоксирибонуклеиновая кислота

Зарег.- зарегистрирован

КС - компьютерная система

Оп. – дата публикации

ПК – программный комплекс

ФИПС – Федеральный институт промышленной собственности

CDKD – циклин-зависимая киназа D (Cyclin-Dependent Kinase D)

MPI – стандарт передачи сообщений в программировании (Message-Passing Standard)

PDB – банк данных структуры белков (Protein Data Bank)

Двухбуквенные коды для представления наименования стран и международных организаций (стандарт ВОИС):

AU – Австралия

BE – Бельгия

CN – Китай

CH – Швейцария

DE – Германия

IL – Израиль

KR – Корея

EA (EAPATIS) – Евразийское патентное ведомство

EP – Европейская патентная организация

GB – Великобритания

IT – Италия

RU – Россия

SG – Сингапур

US – Соединенные Штаты Америки

WO – Международная патентная организация

## **ОБЩИЕ ДАННЫЕ ОБ ОБЪЕКТЕ ИССЛЕДОВАНИЯ**

**Наименование заказчика:** Министерство образования и науки Российской Федерации

**Исполнитель разработки:** Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, г. Новосибирск

**Начало разработки:** «12» октября 2011 г.

**Окончание разработки:** «01» ноября 2012 г.

### **Назначение, область применения, краткое описание объекта исследования:**

Основным объектом исследования являются информационные ресурсы модульного типа для поддержки исследований, проводимых в областях агробιοтехнологии и биоинженерии.

В рамках проекта планируется создание нового информационного ресурса модульного типа, предназначенного для поддержки научно исследовательских разработок в областях агробιοтехнологии и создания новых продуктов и биопроцессов с помощью геномных и постгеномных технологий, методов биоинженерии и клеточных технологий.

В ходе выполнения проекта будет создан информационный интернет-портал, обеспечивающий для исследователей интеграцию информации по геномным и постгеномным технологиям, методам биоинженерии, клеточным технологиям для создания новых продуктов (биореагентов, биоматериалов, биотоплив) и биопроцессов, а так же агробιοтехнологии.

Будет составлен перечень внешних информационных ресурсов, которые позволяют оптимизировать НИР в вышеперечисленных областях биотехнологии. На основе этой информации будет создана база данных (БД), позволяющая эффективно выбирать инструменты для решения конкретных биотехнологических задач.

Будут разработаны БД для дизайна генетической конструкции в области трансгеноза растений: БД промоторов и БД трансляционных энхансеров.

Будет разработана информационная система для поддержки селекционно-генетических экспериментов у пшеницы. Эта информационная система будет содержать описание важных хозяйственных фенотипических признаков, гибкий интерфейс, позволяющий заносить информацию о параметрах эксперимента и его результатах через Интернет, в том числе и через мобильные устройства на платформе Android.

Будут разработаны методы высокопроизводительного фенотипирования пшеницы, основанные на компьютерном анализе изображений.

## **ОСНОВНАЯ (АНАЛИТИЧЕСКАЯ) ЧАСТЬ**

### **1 Технический уровень и тенденции развития объекта исследования**

Для определения технического уровня и тенденций развития объекта исследования проведены патентные исследования в отношении наиболее технически развитых стран.

Патентные исследования проведены и оформлены в соответствии с ГОСТ Р 15.011-96 Система разработки и постановки продукции на производство. Патентные исследования. Содержание и порядок проведения.

Поиск проведен в соответствии с заданием на проведение патентных исследований от 03.10.2011 г. (приложение А) и регламентом поиска от 04.10.2011 г. (приложение Б).

Результаты поиска представлены в справке о поиске (приложение В).

Для поиска патентов использованы следующие источники информации: FIPS – база российских патентов, USPTO – база американских патентов, Esp@cenet – европейская база патентов, содержащая документы различных стран мира.

Анализ отобранных материалов: в результате поиска обнаружено 197 источников информации, имеющих отношение к разрабатываемому объекту, отобрано для последующего анализа 78 источников (в том числе патентных - 63). Выявлено около 50 организаций, разрабатывающих программное обеспечение для поддержки исследований, проводимых в областях агробιοтехнологии и биοинженерии.

Систематизация и изучение отобранных материалов показало следующее.

Объем данных в области молекулярной биологии, биοинженерии и агробιοлогии в последние годы непрерывно растет. В настоящее время здесь сложилась ситуация, когда экспериментальные технологии в своем развитии значительно опережают биοинформационные средства поддержки, анализа и интерпретации результатов экспериментов.

В частности, активно развиваются экспериментальные методы идентификации молекулярных взаимодействий на самых разных уровнях организации биологических систем: высокопроизводительные биочиповые, протеомные и метаболомные экспериментальные технологии, широко используемые в биомедицине, фармакологии, биотехнологии, агробιοлогии и других областях. Широкомасштабное секвенирование геномов, экспериментальные методы биοинженерии также обеспечивают колоссальный рост молекулярно-биологической информации, которую принципиально невозможно осмыслить и переработать без использования специальных компьютерных средств.

Основная задача интерпретации и анализа результатов первичной обработки молекулярно-генетических данных состоит в выяснении связи генов, белков и метаболитов с функционированием молекулярно-генетических систем, интеграции результатов с молекулярно-биологическими информационными ресурсами.

Средства поиска информации, относящиеся к биотехнологии, биоинженерии и агробиотехнологии, включают поиск данных по литературе, по ключевым словам. Такие средства обеспечивают стандартные возможности поиска в Интернете (поиск, сортировка, кластеризация текстовых документов, ранжирование по количеству и характеру гиперссылок, тематике слов), а также решают задачу извлечения из текстов знаний о предметных областях и установления взаимосвязей между объектами и событиями, описанными в текстах и фактографических базах данных.

Например, в источнике (Патент RU 2335013 С2, оп. 27.08.2008) раскрыт способ и системы для улучшения ранжирования поиска с использованием информации о статье. Техническим результатом является расширение поиска и улучшение его качества.

В источнике (Патент RU 2329533 С2, оп. 20.07.2008) раскрыт способ интерактивного поиска в распределенных вычислительных сетях и информационно-поисковая система для его реализации. Техническим результатом является повышение эффективности интерактивных поисковых процессов, повышение точности, достоверности и объективности поисковых результатов.

В источнике (Патент RU 2420800 С2, оп. 10.06.2011) описан способ поиска на устройствах хранения данных электронных документов, похожих по смысловому содержанию на выбранный документ. Технический результат изобретения заключается в повышении точности и полноты поиска электронных документов. Определяют параметры поиска путем задания правил формирования множества уникальных слов, формируют множество взвешенных уникальных слов и взвешенных связей между ними, на основе которых строят семантическую сеть, производят поиск похожих по смыслу документов путем сравнения семантических сетей двух документов.

В источнике (Патент RU 2419861 С2, оп. 27.05.2011) раскрыт способ ранжирования результатов поиска, раскрыт способ предоставления оценки релевантности документа для документа в сети, а также раскрыт машиночитаемый носитель информации, на котором сохранены исполняемые компьютером инструкции для выполнения способа предоставления оценки релевантности для документа в сети. Помимо этого раскрыты вычислительные способы, содержащие, по меньшей мере, один модуль приложения, который содержит прикладной код для выполнения способов

предоставления оценки релевантности для документа в сети. Техническим результатом является повышение достоверности результатов поиска.

В источнике (Патент RU 2419857 С2, оп. 27.05.2011) раскрыт способ определения подобия объектов и, в частности, определение подобия на основании связей между объектами. Способ включает: для каждого типа, если подобие этого типа основано на внутритиповой связи, предоставление функции внутритипового подобия для каждой такой связи, которая измеряет подобие между объектами этого типа; если подобие этого типа основано на межтиповой связи, предоставление функции межтипового подобия для каждой такой связи, которая измеряет подобие между объектами этого типа; и предоставление функции подобия, которая измеряет подобие между объектами этого типа, основываясь на любых функциях внутритипового подобия и любых функциях межтипового подобия для этого типа. Для каждой связи предоставление данных, которые определяют эту связь между объектами, ассоциированными с этой связью; одновременно решение предоставленных функций подобия на основании связей, определяемых обеспеченными данными; и сохраняют подобия, основанные на одновременном решении предоставленных функций подобия. Техническим результатом является увеличение точности поиска за счет вычисления межтипового подобия.

В источнике (патентная заявка US 2004068381 (A1), оп. 08.04.2004), представлен метод обработки баз данных в биоинформатике. Метод ориентирован на сервер с многопользовательским обращением к данным. Сервер получает последовательность для сравнения и анализа от пользователя и сохраняет ее в очереди задач. Если существуют другие последовательности для сравнения и анализа, сервер считывает последовательность в текущем порядке из базы данных, чтобы сравнить ее со всеми последовательностями, хранящимися в очереди. То есть, сервер обращается к базе данных один раз, чтобы использовать ее для всех запросов пользователя, обрабатываемых в настоящее время. Поскольку сервер обращается к базе данных только один раз для каждого запроса пользователя, средняя стоимость системы и время отклика уменьшаются.

Однако известные информационные сервисы не ориентированы на конкретные предметные области (в частности, на биоинженерию и агробиотехнологию). В силу этого они не способны обеспечить более высокие и критически необходимые уровни технологической поддержки автоматического анализа текстов и баз знаний в перечисленных выше предметных областях.

Для успешной разработки специализированного информационного ресурса модульного типа для поддержки исследований, проводимых в областях агробιοтехнологии и биоинженерии, необходимо разработать информационное обеспечение работ по двум приоритетным направлениям: «Использование генетически-модифицированных организмов (ГМО) в биотехнологии» и «Генетическое картирование и фенотипирование в селекции пшеницы»

### **1.1 Использование генетически-модифицированных организмов (ГМО) в биотехнологии**

ГМО (трансгенные животные, растения, грибы и бактерии) активно используются для решения различных задач, включающих наработку новых биотехнологических и фармакологических препаратов. Развитие существующих представлений о метаболических и генных сетях позволит программировать создание ГМО с заданными свойствами, с помощью которых будут достигнуты качественно новые возможности в создании биопродуцентов. В связи с этим, важной задачей является разработка информационного обеспечения биотехнологических НИР, связанных с использованием ГМО.

Выявлено активное патентование способов получения трансгенных растений, обладающих хозяйственно- ценными признаками с использованием информационных ресурсов.

В источнике (Патент RU 2376377, оп. 20.12.2009) описано устойчивое к стрессу растение пшеницы, где указанное растение пшеницы трансформировано молекулой нуклеиновой кислоты, которая кодирует орнитинаминотрансферазу (OAT). Раскрыт способ защиты растения пшеницы от стресса, включающий стадию введения молекулы нуклеиновой кислоты в растение пшеницы, где молекула нуклеиновой кислоты кодирует орнитинаминотрансферазу (OAT). Представлена конструкция нуклеиновой кислоты, включающая промотор, выделенный из растения, и ген орнитинаминотрансферазы (OAT), где указанная конструкция способна трансформировать растение пшеницы таким образом, что указанное растение пшеницы становится толерантным к стрессу.

В источнике (Патент RU 2382079, оп. 20.02.2010) описана область промотора, под контролем которого можно специфически выявлять трансгены в эпидермисе растений. Раскрыта рекомбинантная молекула нуклеиновой кислоты, которая содержит в себе описанные промоторы. Представлены трансгенные растения и растительные клетки, которые трансформировались с помощью раскрытой молекулы нуклеиновой кислоты.

Предложен способ получения представленных растений и клеток. Изобретение позволяет получать растения, устойчивые к возбудителям болезней.

В источнике (Патент RU 2277586, оп. 10.06.2006) раскрыт способ получения растений пшеницы, устойчивых к вредителям, включающий трансформацию клеток растений твердой или мягкой пшеницы, по крайней мере, двумя генами, кодирующими продукты, блокирующие питание вредителя, где, по крайней мере, один из указанных генов функционально связан с промотором, обеспечивающим конститутивную экспрессию продукта гена в растении, и по крайней мере второй из указанных генов функционально связан с промотором, обеспечивающим тканеспецифичную экспрессию продукта гена в растении.

В источнике (Патент RU 2428480, оп. 10.09.2011) раскрыт способ повышения резистентности трансгенных растений или растительных клеток к воздействиям патогенов, заключающийся в том, что в растение или растительную клетку вводят и обеспечивают там экспрессию последовательности ДНК, кодирующей белок с лейцинобогатым повторяющимся доменом (LRR) и/или киназной активностью.

В источнике (Патент RU 2425151, оп. 27.07.2011) раскрыт способ повышения устойчивости растения к стрессовым условиям и увеличения эффективности использования и/или потребления удобрений растением, включающий экспрессирование в растении экзогенного полинуклеотида, кодирующего определенный полипептид.

В источнике (Патент RU 2384621, оп. 20.03.2010) раскрыт способ производства трансгенного растения с повышенной урожайностью, включающий введение в растение или растительную клетку кодирующей CDKD нуклеиновой кислоты, содержащей NHTALRE звено, где X означает любую аминокислоту, и культивирование растительной клетки в условиях, усиливающих регенерацию и рост зрелого растения. Изобретение позволяет повышать урожайность растений.

Для информационной поддержки исследований в области биоинформатики, биотехнологии и биоинженерии разрабатываются специальные компьютерные программы (системы), ориентированные на предметные области.

Известен ряд программ по биоинформатике, распространяемых международным научным сообществом бесплатно (например <http://bioinformatics.ru/>). В настоящий момент разработано несколько компьютерных систем, защищенных патентами.

Например, в источнике (Патентная заявка US 2003220820 (A1), оп. 27.11.2003) представлена система и метод для анализа и визуализации геномной информатики. Геномные данные организмов из различных источников интегрируют в реляционную



базу данных, которая постоянно расширяется и обновляется из открытых источников. Пользователи могут визуально организовывать эту информацию и выбирать интересующую информацию с помощью графического интерфейса.

В источнике (Патентная заявка US 2002082869 (A1), оп. 27.06. 2002) предоставлен метод обеспечения и обновления информации, основанный на геноме индивида, где геном извлекается из биологических образцов индивидуума. Метод состоит из следующих шагов: секвенирование генома; интерпретация последовательности ДНК для идентификации медицинских показаний или заболеваний, ассоциированных с геномом; запись состояний или болезней на запоминающее устройство; информирование индивидуума с этими состояниями или болезнями, для которых его генетические тесты оказались положительными; отслеживание информации о здоровье и предоставление этой информации индивидуумам.

Активно патентуются программные системы и методы для поиска и анализа трансгенов в трансформированном организме.

В источнике (Патентная заявка US 2011015084 (A1), оп. 20.01.2011) представлена компьютерная система высокой пропускной способности для определения сцепления различных полинуклеотидов, а также метод для анализа трансгенов в трансформированном организме. Представленный метод обеспечивает определение сцепления между различными трансгенными полинуклеотидами в трансформантах и секвенирование районов ДНК, связанных с различными трансгенными полинуклеотидами. Также представлен метод для идентификации трансгенных растений, содержащих вставку трансгена в нежелательном районе генома.

В источнике (Патентная заявка US 2010289800 (A1), оп. 18.11.2010) представлены система и метод для обеспечения доступа к документации, связанной с получением трансгенных растений. Этот метод может включать отображение интерактивной карты для доступа к данным, связанным с процессом трансформации у растений, выбор пользователем в интерактивной карте хотя бы одного из множества процессов трансформации растений и показ документации, описывающей процесс трансформации растений. Документация может включать дизайн лабораторного эксперимента и экспериментальные данные, связанные с процессом трансформации растений.

Выявлено много источников, касающихся разработки БД и компьютерных систем для информационной поддержки экспериментов в биотехнологии и биоинженерии.

Например, известна компьютерная система «PathwayStudio, представляющая собой интегрированную систему, описывающую функции белков и их взаимодействия.

Система «PathwayStudio» позволяет представлять взаимодействия между объектами в виде сетей взаимодействий. Работа с системой может включать такие шаги как реконструкция сетей взаимодействий между генами, белками, метаболитами, заболеваниями и молекулярно-генетическими процессами, поиск сигнальных путей в сетях, анализ экспрессии генов с помощью наложения данных анализа микрочипов, а также математическое моделирование динамики полученных сетей.

Разработана компьютерная система «ЭНДВизио» (свидетельство RU 2008615929, зарег. 11.12.2008, позволяющая реконструировать ассоциативные сети на основе семантического анализа биологических данных.

Для систематизации биотехнологической и биоинженерной информации создаются различные БД.

В источнике (Свидетельство RU 2007620101, зарег. 01.03.2007) описана БД Трансген-Промотор (ТГПром), предназначенная для накопления информации, необходимой для планирования генно-инженерных экспериментов с целью создания растительных организмов с качественно новыми или улучшенными свойствами. База содержит данные об экспрессии биотехнологически-значимых генов растений, активности их промоторов и делеционных мутантов этих промоторов, индукторах, влияющих на активность промоторов и нуклеотидных последовательностях описанных промоторов. База может быть использована для подбора ткане-, органо- и стадиоспецифических растительных промоторов, чувствительных к различным индукторам, с целью обеспечения заданного пользователем паттерна транскрипции трансгена при решении конкретных генно-инженерных задач

В источнике (Свидетельство RU 2010620602, зарег. 13.10.2010) описана БД для анализа взаимосвязей генотип-фенотип-окружающая среда у пшеницы (ВитПГЕ), предназначенная для интеграции разнородных данных о растении. Она хранит в себе отношения, описывающие различные характеристики растения, его генотип, фенотип и некоторые факторы внешней среды. База данных позволяет однозначно устанавливать взаимосвязь между генетическими и фенотипическими признаками растений и параметрами окружающей среды.

В источнике (Lee JW, et. Al. Plant Physiol., 2005, 139, 619-631.) описана БД «Germinate», предназначенная для информационной поддержки селекционно-генетических экспериментов у растений.

Известна база данных «Sol Genomic Network» (<http://solgenomics.net/>), которая содержит информацию о фенотипе, генотипе, полногеномных данных и генных и

метаболических сетях для пасленовых (к которым относятся томаты, картофель, баклажан, перец и петуния). Эта БД позволяет производить поиск фенотипа, результатов анализа количественных признаков (QTL), списка маркеров, генов, метаболических сетей. Она тесно интегрирована с геномными данными.

Актуальной задачей на сегодняшний день является обеспечение информационной поддержки исследований в области агробιοтехнологий, в частности, селекционно-генетических экспериментов у пшеницы, основой которых является генетическое картирование и фенотипирование признаков у пшеницы.

## **1.2 Генетическое картирование и фенотипирование в селекции пшеницы.**

В последние годы экстремальные погодные условия (жаркое засушливое лето, холодная зима) все чаще наблюдаются на территории России, что негативно сказывается на урожайности сельскохозяйственных культур, поэтому получение новых сортов, устойчивых к экстремальным условиям среды актуально, в особенности для таких культур как пшеница.

Активно патентуются новейшие компьютерные технологии для целенаправленного поиска генов, ответственных за те или иные признаки устойчивости растений к экстремальным условиям, для последующего введения этих генов в существующие сорта растений через генные модификации.

Например, в источнике (Патентная заявка WO 0164890 (A2), оп. 07.09.2001) раскрыт метод поиска генов (нуклеиновых кислот), индуцируемых при защите, и кодируемых ими белков. Изобретение раскрывает методы, связанные с изменением уровня экспрессии генов, усиливающих устойчивость растений к патогенам. Изобретение также раскрывает рекомбинантные кассеты, клетки хозяина, трансгенные растения и композицию антители.

В источнике (Патентная заявка US 2006183137 (A1), оп. 17.08.2004) представлены компьютерная программа и кластеры генов растений, которые регулируются стрессом, выделение кодирующей и регуляторных областей этих генов, сравнение регуляторных элементов этих генов с консенсусными последовательностями. Представлены рекомбинантные полинуклеотиды, содержащие стресс-регулируемые гены растений, и трансгенные растения, содержащие эти гены. Представлены методы использования генов растений, регулируемых стрессом, для создания растений, имеющих селективное преимущество в условиях стресса; методы идентификации агентов, влияющих на

активность регуляторных элементов, контролирующих стрессоустойчивость у растений; и методы, определяющие, было ли растение подвергнуто стрессу.

В результате разработки методов полного секвенирования геномов и их биоинформатической аннотации, быстро развиваются технологии генотипирования растений. Становится возможным за относительно короткое время прочитать последовательность генома организма и идентифицировать на ней локализацию всех маркеров и генов. И хотя для многих сельскохозяйственных культур секвенирование генома затруднено в силу повторности хромосом (полиплоидия), ожидается, что в течение нескольких ближайших лет эта задача будет решена, в том числе и для пшеницы.

Например, в источнике (Патентная заявка US 2004203041 (A1), оп. 14.10.2004) представлена технология и компьютерная система для предсказания генов. Эта технология использует анализ вероятностей для выявления особенностей кодирования в последовательностях нуклеиновых кислот, включая определение вероятности состояний для каждого нуклеотида в последовательности нуклеиновых кислот, определение кодирующей цепи, определение открытых рамок считывания, определение вставок и делеций локусов, определение расположения экзонов, а также определение последовательности белка.

В источнике (Патентная заявка US 2011033854 (A1), оп. 10.02.2001) описана технология секвенирования длинных фрагментов последовательностей. Данное изобретение включает в себя методы и их сочетания для подготовки длинных фрагментов геномной ДНК, для обработки геномной ДНК методами анализа длинных секвенированных фрагментов, а также программное обеспечение и алгоритмы обработки и анализа полученных данных о последовательностях.

В источнике (Патентная заявка US 20004002816 (A1), оп. 01.01.2004) раскрыт компьютерный метод обнаружения сходства между ДНК последовательностями. Данный метод хорошо подходит для «масштабных» сравнений, таких как сравнение мутаций у миллионов фрагментов последовательностей множества геномов растений. Позиционное хеширование проводится путем разделения проблемы сравнения последовательностей по их структурам на подзадачи, что делает возможным распараллеливание, при помощи чего большое число узлов в компьютерном кластере или компьютерной системе одновременно задействованы для решения данной проблемы. Хеширование позиций может быть использовано сам по себе или для фильтрации в сочетании с множеством других алгоритмов сравнения последовательностей.

В источнике (Патентная заявка GB 2477703 (A) , оп. 10.08.2011) представлен метод и система для анализа данных о нуклеотидных последовательностях. Система секвенирования и метод генерации индексирующих ключей для одного или более данных о последовательностях основаны на скрытых значениях последовательностей из образцов данных о последовательностях и/или одном или нескольких шаблонах данных.

В источнике (Патентная заявка KR 20070115964 (A) , оп. 06.12.2007) раскрыты система и метод для выполнения непарного сравнения биологических последовательностей, включающий в себя новый метод  $C_{10}$ , являющийся непарным методом подсчета, применяемым в самостоятельном модуле VaSSA-I. Этот метод позволяет получить значительно больше информации о последовательностях и сходстве между ними, чем стандартные биоинформационные техники.

Однако, для того, чтобы решение задач, основанных на массовом генотипировании было эффективным, необходимо развивать подходы для быстрой и точной оценки фенотипа. Современный селекционно-генетический эксперимент использует данные о тысячах и десятках тысяч растений. Очевидно, что для выборок такого размера традиционные способы определения большинства фенотипических характеристик (визуальные, тактильные, измерение линейкой и пр.) малоэффективны.

Именно поэтому область исследований в биологии, связанная с получением данных о фенотипических признаках растений в массовом порядке (высокопроизводительное фенотипирование) сейчас получает все большее развитие

В основе технологий высокопроизводительного фенотипирования растений лежит разработка компьютерных методов оцифровки их параметров. Цель заключается в максимальном исключении человека из процесса измерения. Это позволяет: (а) существенно ускорить процесс снятия данных за счет его автоматизации; (б) увеличить точность оценки фенотипических параметров, устранив субъективизм и неточность измерений, присущие человеку; (в) получать оценки новых характеристик фенотипа, наблюдение или оценка которых ранее были недоступны.

Активно патентуются компьютерные системы высокопроизводительного анализа фенотипа у растений.

Например, в источнике (патентная заявка WO 2010129168 (A2), оп. 11.11.2010) описан метод целевого фенотипирования признаков в экспериментах по селекции растений. Метод включает определение условий среды и вероятности вариантов фенотипических признаков для каждого эксперимента по селекции растений, используя условия среды и предсказанные состояния культуры, выбор множества селекционных

экспериментов для коллекции замеров фенотипических признаков, основанных на вероятности вариации фенотипических признаков, и коллекцию замеров фенотипических признаков из множества селекционных экспериментов.

Одним из подходов, позволяющих существенно ускорить фенотипирование, является анализ цифровых изображений. Он успешно применяется для оценки биомассы растения (Hartmann A, et. al. BMC Bioinformatics, 2011, 12, 148; Golzarian MR, et. al, Plant Methods, 2011, 7, 2), для выявления генов, ответственных за размер зерна у *A. thaliana* (Herridge RP. et. al, Plant Methods, 2011, 7, 3), для анализа морфологии и развития корня у *A.thaliana* (Joosen A. et. al, Plant J., 2010, 62, 148-159; Yazdanbakhsh N. et. al, Ann Bot., 2010, 105, 783-791) и риса (Iyer-Pascuzzi AS. et. al, Plant Physiol., 2010, 152, 1148-1157), в задачах колориметрии у томатов (Darrigues A, et. al, J. Amer. Soc. Hort. Sci., 2008, 133, 579–586) и др.

В источнике (Bensch R. et. al, In Proc of the The Sixth IEEE International Symposium on Biomedical Imaging (ISBI'09); 2009, 742-745; Kaminuma E et. al, Plant J., 2008, 56, 470-482) раскрыты методы компьютерного анализа морфологии опушения листа у пшеницы.

В источнике (патент EP 1322788 A2, оп. 03.01.2002) описан метод анализа изображений для фенотипирования набора мутантных клеток. Метод характеризует клеточную и субклеточную архитектуру мутантных аллелей, выращенных в разнообразных условиях, используя различные морфологические и молекулярные маркеры, комбинированные с автоматическим получением изображений и анализом. Фенотипические особенности могут включать цитоскелет, органеллы, клеточную морфологию, состояние репликации ДНК, взаимосвязь этих признаков друг с другом и т.д. На основе этих признаков может быть создан количественный «отпечаток» для каждого фенотипа. Эта количественная информация о фенотипе делает возможным создание баз данных, которые обеспечивают связывание генотипа с фенотипом. Гены, охарактеризованные таким образом, могут быть сгруппированы в функциональные группы, пути, более высокие уровни сборки белков и т.п.

Высокопроизводительное фенотипирование, эффективный сбор, хранение большого объема фенотипических данных, их интеграция с геномными данными позволяют создавать технологии анализа взаимосвязи между генотипом и фенотипом.

В источнике (Патентная заявка US 201100148 (A1), оп. 13.01.2011) представлен метод для улучшения эффективности программы селекции растений с целью изменения фенотипических признаков, для которых установлена связь с генетическими маркерами.

Генетические значения генома отдельного индивида вычисляются на основе индивидуальных маркеров генотипа и связей, установленных между генетическими маркерами и фенотипическими признаками. Индивиды и схемы селекции в дальнейшем выбираются на основе генетических значений генома и их распределения у потенциальных потомков, полученных на основе схемы селекции. Представленные схемы и компьютерные программы могут быть использованы в селекции растений и создании трансгенных растений.

Известны компьютерно-экспериментальные подходы, основанные на методе картирования генов, контролирующих количественные признаки (quantitative trait loci, QTL). Например, в источнике (Kearsey JM, J.Exp.Bot., 1998, 49, 1619-1623; Mauricio R., Nat. Rev. Genet., 2001, 2, 370-381) используют комбинацию молекулярных и фенотипических данных, однако анализ QTL тем точнее и надежнее, чем больше данных использовано при оценке корреляций и чем выше точность использованных данных.

Лидирующие компании, оперирующие на рынке агробιοтехнологии, разрабатывают интегрированные системы автоматизации фенотипирования, которые позволят существенно ускорить процесс сбора данных о важных характеристиках растений, повысить точность их оценок, измерять новые параметры растений, а также исключить субъективизм человеческой оценки из процесса измерения.

Например, компания Lemnatec создает программные комплексы для проведения высокопроизводительного фенотипирования у растений. Известна также система «Phenopsis» для анализа модельного растения *A.thaliana* (Fabre J, et. al., BMC Plant Biol., 2011, 11, 77), а также система «Phenofab» ([www.phenofab.com](http://www.phenofab.com)).

Использование мобильных устройств позволяет существенно повысить эффективность решения задач в области селекционно-генетических экспериментов, особенно для полевых наблюдений.

Например, в информационной системе «Phenome» (Vankadavath RN, et. al., Plant Methods, 2009, 5, 18) для сбора информации о признаках растений используются карманные компьютеры (PDA), что позволяет собирать большое количество данных в полевых условиях (результаты измерений анатомических признаков растений, их плодов, устойчивости к заболеваниям), с последующим занесением собранной информации в центральную базу данных.

В России также проводятся работы по автоматизации исследований в области селекции и генетики растений. Например, в ООО «Центр биоинженерии» (г. Москва)

функционирует современных тепличный комплекс, в котором автоматизирована поддержка условий произрастания растений.

Зарубежные агробиотехнологические компании переходят на принципиально новый уровень исследований с привлечением достижений хромосомной инженерии и молекулярной генетики, что позволяет создавать генотипы пшеницы с заданными хозяйственно-ценными признаками.

Например, в источнике (Патентная заявка US 2011113506 A1, оп. 12.05.2011) описан метод геномного дизайна для определения ДНК маркеров от M1 до M5 для каждой целевой области (генетического локуса), при этом, ДНК маркер M2 определен на конце последовательности, предшествующей генетическому локусу, или расположенной перед ним, ДНК маркер M1 расположен перед ДНК маркером M2, ДНК маркер M4 расположен на конце последовательности, расположенной в конце генетического локуса или после него, ДНК маркер M5 расположен за ДНК маркером M4, и ДНК маркер M3 расположен в целевой области (генетическом локусе); и проектируется геном так, чтобы в области замены, содержащей генетический локус, в хромосоме оригинального сорта, в котором производят замещение на фрагмент хромосомы чужеродного сорта, границы переносимой области перед генетическим локусом заключались между ДНК маркером M1 и ДНК маркером M2, а после генетического локуса между ДНК маркером M4 и ДНК маркером M5.

Таким образом, в настоящее время существует ряд компьютерных программ и баз данных для поддержки исследований в области агробиотехнологии и биоинженерии, однако, эффективные программно-информационные системы на основе интегрированного анализа данных генотипирования и фенотипирования, **отсутствуют**.

Созданные в результате выполнения проекта компьютерные программы будут пригодны для внедрения и массового применения, как в России, так и за рубежом с целью получения новых знаний, проведения прикладных исследований и опытно-конструкторских разработок в таких областях, как системная биология, агробиотехнология, биоинженерия и др.

В ходе выполнения проекта будет создан информационный интернет-портал, обеспечивающий для исследователей интеграцию информации по геномным и постгеномным технологиям, методам биоинженерии, клеточным технологиям для создания новых продуктов (биореагентов, биоматериалов, биотоплив) и биопроцессов, а так же агробиотехнологии.



Основные компании, работающие в области агробιοтехнологии, биоинженерии и биоинформатики, локализованы, преимущественно в США, Европе, Канаде, Индии, Японии и Китае, а также в стремительно развивающихся странах Азиатско-Тихоокеанского региона.

Среди компаний, успешно развивающих бизнес в области биоинформатики и биотехнологии, лидирующие позиции занимают такие фирмы, как: Genedata AG (US), GeneGo (US), Gene Network Sciences (US, GB), Genomatica (US), Gateway Inc.(US), Real Time Genomics Inc. (US), Complete Genomics Inc. (US), Ip Genesis Inc. (US), Adobe Systems Inc. (US), Applied Biosystems (US), Genedata AG (CN), BGI (CN) и др.

Среди компаний, успешно развивающих бизнес в области агробιοтехнологии лидирующие позиции занимают такие фирмы, как: Advanta Seeds UK Ltd (GB), World Wide Wheat (US), BioHorizon (IN), Mendel Biotechnology (US), Monsanto (US), CIMMIT (MX), NuPlant (AU), RAGT (DE), Vilmorin (FR).

Наибольшую изобретательскую активность проявляют фирмы США (выявлено 27 изобретений).

На основании исследования географии патентования следует отметить, что ряд изобретений ведущих разработчиков в данной области проходят процедуру патентования в Российской Федерации (выявлено 19 изобретений)

В связи с тем, что, как правило, наиболее перспективными для коммерческой реализации являются страны, где наблюдается активная регистрация и выдача патентов иностранным заявителям, справедливо допустить, что наиболее перспективными странами для реализации исследуемого объекта техники являются США.

Анализ выявленных патентов зарубежных стран показал, что наиболее активное патентование наблюдается по следующим направлениям исследований: разработка программных комплексов, обеспечивающих хранение и передачу масштабных потоков биологических текстов, источником которых являются агробιοтехнологии и биоинженерия; системы и методы для поиска и анализа трансгенов в трансформированном организме; компьютерные технологии для целенаправленного поиска генов; технологии анализа взаимосвязи между генотипом и фенотипом у растений.

Анализ источников научно-технической и патентной информации показал, что основными тенденциями развития объекта исследования являются:

1) усовершенствование методов информационной и телекоммуникационной поддержки селекционно-генетических экспериментов у растений;

2) усовершенствование программных комплексов, обеспечивающих хранение и передачу масштабных потоков биологических текстов, касающихся агробιοтехнологии и биоинженерии;

3) разработка методов высокопроизводительного фенотипирования, основанных на компьютерном анализе изображений основных признаков растений;

4) создание программных комплексов для анализа взаимосвязи между генотипом и фенотипом у растений и окружающей средой;

5) разработка компьютерных методов оцифровки параметров фенотипирования растений для более быстрой и точной оценки фенотипа;

6) усовершенствование компьютерных технологий для целенаправленного поиска генов, ответственных за те или иные признаки устойчивости растений к экстремальным условиям,

7) создание баз данных, позволяющих производить поиск фенотипа, результатов анализа количественных признаков (QTL), списка маркеров, генов, метаболических сетей и интегрированных с геномными данными;

8) разработка баз данных промоторов и трансляционных энхансеров для трансгенеза;

9) разработка информационных ресурсов для использования в дизайне генетических конструкций для оптимизации экспрессии трансгенных конструкций;

10) разработка компьютерных систем для автоматизации фенотипирования, которые позволят существенно ускорить процесс сбора данных о важных характеристиках растений, повысить точность их оценок, измерять новые параметры растений;

11) разработка методов проведения информационной поддержки селекционно-генетических экспериментов через мобильные устройства.

Поскольку в последние годы ведется активное патентование технологий генотипирования и фенотипирования пшеницы и ее гибридов, выявлены лидирующие компании, оперирующие на рынке агробιοтехнологий. В России в настоящее время аналогичные компании отсутствуют, в связи с ориентированностью российских потребителей агробιοтехнологической и биоинженерной продукции на западных производителей. Анализ деятельности данных компаний представлен в таблице 1.

### **1.3 Компании, оперирующие на рынке агробιοтехнологий**

**Таблица 1** Компании, оперирующие на рынке агробиотехнологий

Компании, оперирующие на рынке агробиотехнологий	Основные продукты, бренды	На какие отрасли ориентирована продукция	Интернет адрес компании
Syngenta AG (Англия)	Технологии генотипирования и фенотипирования признаков у пшеницы	Сельское хозяйство, наука	<a href="http://www.syngenta.com/">http://www.syngenta.com/</a>
Advanta Seeds UK Ltd (Англия)	Генотипирование и фенотипирование сортов и гибридов пшеницы	Сельское хозяйство, наука	<a href="http://www.advantaseeds.com/">http://www.advantaseeds.com/</a>
World Wide Wheat (США)	Генотипирование и фенотипирование сортов и гибридов пшеницы Создание новых форм растений	Сельское хозяйство, наука	<a href="http://www.worldwheat.com">http://www.worldwheat.com</a>
CIMMIT (Мексика)	Технологии генотипирования и фенотипирования признаков у растений Создание новых форм растений	Сельское хозяйство, наука	<a href="http://www.cimmyt.org/">http://www.cimmyt.org/</a>
Monsanto Technology LLC (США)	Технологии генотипирования и фенотипирования признаков у растений Создание новых форм растений	Сельское хозяйство, наука	<a href="http://www.monsanto.com">http://www.monsanto.com</a>
BioHorizon (Индия)	Технологии генотипирования и фенотипирования сортов и гибридов растений Создание новых форм растений	Сельское хозяйство, медицина, наука	<a href="http://www.biohorizon.com/">http://www.biohorizon.com/</a>
Mendel Biotechnology (США)	Генотипирование признаков у пшеницы. Создание новых форм растений	Сельское хозяйство, наука	<a href="http://www.mendelbio.com/">http://www.mendelbio.com/</a>
NuPlant (Австралия)	Генотипирование пшеницы Хромосомное маркирование	Сельское хозяйство, наука	<a href="http://www.nuplant.com">http://www.nuplant.com</a>
Chany Biotech (Китай)	Генотипирование и фенотипирование сортов и гибридов пшеницы Создание новых форм растений	Сельское хозяйство, наука	<a href="http://www.chanybio.com/">http://www.chanybio.com/</a>
RAGT	Технологии	Сельское	<a href="http://www.ragtse">http://www.ragtse</a>

(Германия)	генотипирования и фенотипирования признаков у растений	хозяйство, наука	<a href="http://mences.com">mences.com</a>
Vilmorin (Франция)	Генотипирование пшеницы Создание новых форм растений	Сельское хозяйство	<a href="http://www.vilmorin.com/">http://www.vilmorin.com/</a>
Hartmann's company (США)	Технологии генотипирования и фенотипирования сортов и гибридов растений Создание новых форм растений	Сельское хозяйство	<a href="http://www.hartmannsplantcompany.com/">http://www.hartmannsplantcompany.com/</a>
China Cabs Dalian Co., Ltd (Китай)	Генотипирование и фенотипирование признаков у с/х растений Создание новых форм растений	Сельское хозяйство	<a href="http://dlcabs.en.alibaba.com/">http://dlcabs.en.alibaba.com/</a>
Boreal Plant Breeding Ltd (Финляндия)	Технологии генотипирования и фенотипирования сортов и гибридов растений	Сельское хозяйство	<a href="http://www.borealfi.fi/index.php?">http://www.borealfi.fi/index.php?</a>
	Компьютерные системы для трансгенеза	Сельское хозяйство, наука	<a href="http://www.pioneer.com/home/site/about/terms-of-use">http://www.pioneer.com/home/site/about/terms-of-use</a>
ООО «Центр биоинженерии» (Россия)	Автоматизация исследований в области селекции и генетики растений	Сельское хозяйство	<a href="http://www.biengiac.ru/teplitza.htm">http://www.biengiac.ru/teplitza.htm</a>

Обобщая проведенные патентные исследования, можно сделать вывод, что они соответствуют заданию на их проведение: получены достоверные данные об уровне техники и тенденциях развития исследуемого объекта, а также выявлены компании, оперирующие на рынке агротехнологий.

Результаты патентных исследований могут быть использованы для обоснования оптимальных путей достижения НИР, в частности: при решении вопросов правовой охраны объектов промышленной (интеллектуальной) собственности, при прогнозировании и планировании НИР согласно задачам первого этапа, выборе направлений исследований для получения запланированных в НИР результатов.

## 2 Использование объектов промышленной (интеллектуальной) собственности и их правовая охрана

В результате анализа применимости в объекте исследования известных объектов интеллектуальной собственности, выявлены патенты, которые могут быть использованы при разработке данной темы в качестве аналогов и прототипов. Предложения по их использованию сформированы в таблице 2.

**Таблица 2** Анализ применимости в объекте исследования известных объектов промышленной (интеллектуальной) собственности

Вид промышленной собственности, наименование объекта пром. собственности. Патентообладатель (страна, фирма)	Номер охранного документа, классификационный индекс, номер и дата подачи заявки (страна, номер заявки и дата конвекционного приоритета)	Наименование составных частей объекта исследования, в которых могут быть использованы объекты пром. собственности	Возможность и целесообразность использования объекта пром. собственности или причины отказа от использования	Ожидаемый эффект
1	2	3	4	5
<b>Изобретение</b> «Программы селекции растений с целью изменения феноти-пических признаков» <b>Заявитель:</b> Syngenta Participations AG (US)	Заявка US 2011010148A1 G 06F 19/18 оп. 13.01.11	Программа селекции растений	Целесообразно использовать	Повышение эффективности программы
<b>Изобретение</b> «Метод целевого фенотипирования признаков в экспериментах по селекции растений» <b>Заявитель:</b> (US)	Заявка WO 2010129168A2 A 01H1/04 оп. 11.11.10	Программа селекции растений	Целесообразно использовать	Повышение эффективности и точности оценки фенотипа.
<b>Изобретение</b> «Система и метод анализа изображений для	Патент EP 1322788 A2 C 12N 15/10 G 06F 19/00	Создание базы данных, связывающей генотип с	Целесообразно использовать	Повышение точности определения генотипа и

фенотипирования мутантных клеток» <b>Заявитель:</b> Cytokinetics Inc. (US)	оп. 03.01.02	фенотипом		фенотипа у растений.
База данных «Трансген-Промотор (ТГПром)» <b>Заявитель:</b> ИЦиГ СО РАН (RU)	Свидетельство № 2007620101 Зарег. 01.03.07	При решении конкретных генно-инженерных задач.	Целесообразно использовать	Повышение эффективности генно-инженерных экспериментов
<b>База данных</b> «База данных для анализа взаимосвязей генотип-фенотип-окружающая среда у пшеницы (ВитПГЕ)» <b>Заявитель:</b> ИЦиГ СО РАН (RU)	Свидетельство № 2010620602 Зарег. 13.10.10	Оптимизация и заполнение базы данных	Целесообразно использовать	Увеличение скорости обработки данных

Обобщая проведенные патентные исследования, можно сделать вывод, что они соответствуют заданию на их проведение: проведен анализ применимости в объекте исследования известных объектов интеллектуальной собственности и сформированы предложения по их использованию при разработке данной темы.

Результаты патентных исследований могут быть использованы для обоснования оптимальных путей достижения НИР, в частности: выявленные патенты могут быть использованы в качестве аналогов и прототипов для выбора оптимального пути при разработке информационного ресурса для использования в областях агробιοтехнологии и биоинженерии.

### 3 Исследование патентной чистоты объекта исследования

При проверке патентной чистоты объекта исследования в отношении России в результате патентного поиска не обнаружено действующих патентов, под действие которых может подпадать проверяемый объект.

Таблица 3. Выводы о патентной чистоте объекта техники

Страна проверки	Результаты проверки			Номер и дата отчета о патентных исследованиях, организация-исполнитель
	Обладает или нет патентной чистотой («Да», «Нет») с указанием даты публикации последних просмотренных патентных материалов			
	Изобретения (полезные модели)	Промышленные образцы	Товарные знаки	
1	2	3	4	5
Россия	«Да». Обладает патентной чистотой в отношении России по состоянию на 27.10.2011 (Офиц бюл. «Изобретения. Полезные модели» № 30, 2011	Проверке не подлежат	Проверке не подлежат	№ 1 от 01.11.2011 ИЦиГ СО РАН

Обобщая проведенные патентные исследования, можно сделать вывод, что они соответствуют заданию на их проведение: получены достоверные данные о патентной чистоте исследуемого объекта.

Результаты патентных исследований могут быть использованы для обоснования оптимальных путей достижения НИР, в частности: разрабатываемый объект беспрепятственно может использоваться на территории Российской Федерации без нарушения прав третьих лиц.

## ЗАКЛЮЧЕНИЕ

1) Проведенные патентные исследования соответствуют заданию на их проведение: получены достоверные данные об уровне техники, тенденциях развития и патентной чистоте исследуемого объекта, выявлены компании, оперирующие на рынке агробιοтехнологий; проведен анализ применимости в объекте исследования известных объектов интеллектуальной собственности и сформированы предложения по их использованию при разработке данной темы.

2) Основные компании, разрабатывающие или использующие программные средства для обеспечения разработок в областях агробιοтехнологии и биоинженерии, локализованы, преимущественно, в США, Европе, Канаде, Индии, Японии и также в стремительно развивающихся странах Азиатско-Тихоокеанского региона.

Данные компании должны быть объектом постоянного внимания разработчиков с целью своевременного выявления созданных ими новых биоинформационных продуктов или тенденций их развития.

3) Ведущими странами, занимающимися исследованиями в исследуемой области являются США, Китай, Великобритания, Германия, Япония, Сингапур, Индия, Франция, Италия и др. страны, где исследованиям в области агробιοтехнологии и биоинженерии придается ранг национальных программ.

4) Специализированный рынок агробιοтехнологий в России в настоящее время находится в начальной стадии формирования в связи с ориентированностью российских потребителей биотехнологической и генно-инженерной продукции на западных производителей.

5) Наиболее активное патентование объекта исследования наблюдается по следующим направлениям исследований: разработка программных комплексов, обеспечивающих хранение и передачу масштабных потоков биологических текстов, источником которых являются агробιοтехнологии и биоинженерия; системы и методы для поиска и анализа трансгенов в трансформированном организме; компьютерные технологии для целенаправленного поиска целевых генов; технологии анализа взаимосвязи между генотипом и фенотипом у растений.

6) Анализ сложившейся патентной ситуации в области разработки программного обеспечения для агробιοтехнологий и биоинженерии, в основе которой лежит патентование изобретений у себя в стране и за рубежом, позволяет сделать следующие выводы:



а) наибольшую изобретательскую активность проявляют фирмы США;

б) наиболее перспективной страной для реализации исследуемого объекта техники являются США;

в) ряд лидирующих фирм в исследуемой области активно патентуют свои разработки в Российской Федерации (выявлено 19 изобретений)

7) Разрабатываемая тема соответствует **мировому уровню техники**, поскольку нацелена на создание нового информационного ресурса модульного типа, предназначенного для поддержки научно исследовательских разработок в областях агробιοтехнологии и создания новых продуктов и биопроцессов с помощью геномных и постгеномных технологий, методов биоинженерии и клеточных технологий.

8) Конкурентные преимущества нового информационного ресурса позволят рассчитывать на масштабное позиционирование на мировом рынке Российской биоинформационной продукции, поскольку существует устойчивая тенденция увеличения масштабов спроса в области предоставления биоинформационных услуг и рынка принципиально новых биоинформационных продуктов.

9) Из доступных патентных источников информации не обнаружено действующих патентов на территории России, под действие которых может подпадать исследуемый объект, в связи с чем можно сделать вывод, что объект исследования обладает патентной чистотой в отношении России по состоянию на 27.10.2011 (последний просмотренный бюллетень «Изобретения. Полезные модели», № 30).

10) На последнем этапе проведения НИР при разработке результатов интеллектуальной деятельности (далее – РИД), способных к правовой охране, будут проведены дополнительные патентные исследования в соответствии с ГОСТ Р 15.011-96.

ПРИЛОЖЕНИЕ А

(обязательное к отчету о патентных исследованиях)

ФОРМА ЗАДАНИЯ НА ПРОВЕДЕНИЕ ПАТЕНТНЫХ ИССЛЕДОВАНИЙ



**ЗАДАНИЕ №1**

на проведение патентных исследований

**Наименование проекта:** Государственный контракт «Разработка информационного ресурса модульного типа для поддержки исследований, проводимых в рамках Технологической платформы "Биоиндустрия и биоресурсы — BioTech2030" в областях агробιοтехнологии и биоинженерии»

**Шифр темы:** 2011-1.4-514-111-018

**Этап работы:** этап № 1 промежуточный, «Выбор направления исследований и теоретическое исследование поставленных перед НИР задач», сроки его выполнения: 12.10.2011 – 01.12.2011

**Задачи патентных исследований:** обеспечить получение достоверных данных об уровне техники, тенденциях развития и патентной чистоте исследуемого объекта, выявить компании, занимающиеся аналогичными разработками и провести анализ применимости в объекте известных объектов интеллектуальной собственности.

**Календарный план**

Виды патентных исследований	Подразделения-исполнители (соисполнители)	Ответственные исполнители	Сроки выполнения патентных исследований.	Отчетные документы
Поиск, отбор и систематизация отобранных материалов для анализа	Отдел защиты интеллектуальной собственности Лаб. генной инженерии	Кучумова Л.Я. Смирнова О.Г.	05.10.2011 – 18.10.2011	Справка о поиске
Исследование уровня техники и тенденций развития объекта исследования.	Отдел защиты интеллектуальной собственности Лаб. генной инженерии	Кучумова Л.Я. Смирнова О.Г.	19.10.2011 – 11.11.2011	Отчет о патентных исследованиях

Начальник отдела защиты интеллектуальной собственности \_\_\_\_\_ Сурнина Н.Ю. 03.10.2011  
личная подпись \_\_\_\_\_ дата

Руководитель подразделения-исполнителя работы \_\_\_\_\_ Кочетов А.В. 03.10.2011  
личная подпись \_\_\_\_\_ дата

## ПРИЛОЖЕНИЕ Б

(обязательное к отчету о патентных исследованиях)

### ФОРМА РЕГЛАМЕНТА ПОИСКА

#### РЕГЛАМЕНТ ПОИСКА № 1

**04.10.2011**

Дата составления регламента

**Наименование проекта:** Государственный контракт «Разработка информационного ресурса модульного типа для поддержки исследований, проводимых в рамках Технологической платформы "Биоиндустрия и биоресурсы — БиоТех2030" в областях агробιοтехнологии и биоинженерии»

**Шифр темы:** 2011-1.4-514-111-018

**Номер и дата утверждения задания:** № 1 от 03.10.9.2011

**Этап работы:** Этап № 1, промежуточный. «Выбор направления исследований и теоретическое исследование поставленных перед НИР задач»

**Цель поиска информации:** обеспечить получение достоверных данных об уровне техники, тенденциях развития и патентной чистоте исследуемого объекта, выявить компании, занимающиеся аналогичными разработками и провести анализ применимости в объекте известных объектов интеллектуальной собственности.

#### **Обоснование регламента поиска:**

Ведущими странами, занимающимися аналогичными разработками, являются США, Великобритания, Франция, Германия, Индия, Китай, Россия. Глубина поиска принята 20 лет. Электронный поиск будет проводиться в патентных базах данных ЕПВ (ESPACE- EP-B, esp@cenet), США (US PTO), Канады (SIPO Canada PO), PCT Online, Российского патентного ведомства (fips/ru). Изобретения отечественных авторов будут изучаться в патентной базе ФИПС "Изобретения, Полезные модели" с 1994 г. Зарегистрированные информационные продукты будут изучаться в Офиц. Бюл. Роспатента «Программы для ЭВМ, базы данных и топологии интегральных микросхем». Основной поиск будет проведен в сети Интернета.

Поиск будет проведен по следующим ключевым словам и их сочетаниям: Компьютерные программы, базы данных, биоматериалы, генетические конструкции, генетически-модифицированные растения, трансгенные растения, промоторы растений, промоторы для трансгенеза растений, целевые гены, метаболические пути, энхансеры, векторы, трансформация растений, биотехнология растений, клеточные технологии, агробιοтехнология, биоинженерия

В зарубежных базах данных поиск будет проведен по следующим ключевым словам и их сочетаниям:

Computer programs, database, biomaterials, genetic construct, genetically modified plants, transgenic plant, plant promoter, promoters for plant transgenesis, target gene, metabolic pathway, enhancer, vector, plant transformation, method of plant transformation, plant biotechnology, cellular technology, plant producer, agrobiotechnology.

**Начало поиска** 05.10.2011 **Окончание поиска** 11.11.2011

Предмет поиска	Страна поиска	Источники информации, по которым будет проводиться поиск									
		патентные		НТИ		конъюнкту рные		другие		Ретро- спекти вность	Наименова ние информаци -онной базы (фонда)
1	2	Наименование	Классифика ционные рубрики: МПК (МКИ)	Наименование	Рубрики УДК	Наимено вание	Код товара	Наименован ие	Классиф икацион ные индексы		
Использование генетически-модифицированных организмов биотехнологии Генетическое картирование и фенотипирование в селекции пшеницы Компьютерные программы для информационной поддержки селекционно-генетических экспериментов у растений	Россия Великобрит ания Германия Франция Япония США Канада Индия Китай	Официальный бюллетень Роспатента «Изобретения. Полезные модели». Реф. сборник ФИПС Роспатента «Изобретения стран мира»	A01H1/00 A01H1/04 C 12N15/09 C 12N15/11 C 12N15/12 C 12N15/29 C 12N15/31 C 12N15/63 C 12N15/82 C12Q 1/68 G 01N 33/48 G 01N33/68 G 06F17/30 G 06F17/40 G 06F19/00- G 06F19/24	J. Biophysical; J. of Nucleic Acids; J. of Mol. Biology; Science; J. American Chemical Society J. Biotechnol Plant Cell Rep. Plant Methods Plant Physiol Trends Plant Sci.	-	-	-	Инф.бюл. Роспатента «Программ ы для ЭВМ, базы данных и топологии интегральн ых микросхем	-	991 – 011	esp@cenet USPTO EAPATIS RUPAT WIPO PAJ PatSearch CIPO (Ca- nada PO) Ресурсы Интернет: Google Yandex

Руководитель подразделения-  
исполнителя работы

  
личная подпись

Кочетов А.В. 03.10.2011  
дата

Начальник отдела защиты  
интеллектуальной собственности

  
личная подпись

Сурнина Н.Ю. 03.10.2011  
дата

## ПРИЛОЖЕНИЕ В

(обязательное к отчету о патентных исследованиях)

### ФОРМА ОТЧЕТА О ПОИСКЕ

**В.1** Поиск проведен в соответствии с Заданием директора ИЦиГ СОРАН Колчанова Н.А. № 1 от 03 октября 2011 г. и Регламентом поиска № 1 от 04.10.2011

**В.2** Этап работы: этап № 1, промежуточный. «Выбор направления исследований и теоретическое исследование поставленных перед НИР задач»

**В.3** Начало поиска: 05.10.2011 г. Окончание поиска 11.11.2011 г

**В.4** Сведения о выполнении регламента поиска: Регламент поиска выполнен в полном объеме.

**В.5** Предложения по дальнейшему проведению поиска и патентных исследований. На последнем этапе НИР при разработке результатов интеллектуальной деятельности (далее – РИД), способных к правовой охране, будут проведены дополнительные патентные исследования в соответствии с ГОСТ Р 15.011-96.

**В.6** Материалы, отобранные для последующего анализа

В процессе поиска обнаружено 197 документов, имеющих отношение к теме. После исключения из списка патентов-аналогов и патентов, касающихся тематики запроса лишь косвенно, из них отобрано 78 документов. Сведения о наиболее релевантных источниках, отобранных для детального анализа, приведены в Таблице В.6.1.

Таблица В.6.1. Патентная документация

Предмет поиска	Страна выдачи, вид и номер охранного документа, МПК, дата публикации	Заявитель (патентообладатель), страна	Название изобретения	Сведения о действии охранного документа
1	2	3	4	5
Компьютерные системы и методы для биоинженерии и агробιοтехнологии	Заявка US 2011015084A1 С 12Q 1/68 G 06F 19/00 оп. 20.01.11	Monsanto Technology LLC (US)	Система и метод для идентификации трансгенных растений	Временная охрана
То же	Заявка US 2005124010A1 С 12N 15/82 G 06F 19/00 оп. 09.06.05	Short Jay M (US)	Метод создания новых трансгенных организмов с желаемыми свойствами	Действует
То же	Заявка US 2010289800A1 G 06F 3/048	(US)	Система для доступа к данным, связанным с	Действует

	оп. 18.11.10		процессом трансформации у растений	
То же	Патент AU 4162201 A C 12N 15/29 G 06F 19/00 оп. 07.09.01	(US)	Новые защитные гены и их применение	Действует
То же	Заявка US 2006183137A1 C 12N 15/29 G 06F 17/00 оп. 17.08.06	Syngenta Participations AG (US)	Кластеры генов растений, регулируемые стрессом	Действует
То же	Заявка US 2011010148A1 G 06F 19/18 оп. 13.01.11	Syngenta Participations AG (US)	Программы селекции растений с целью изменения фенотипических признаков	Временная охрана
То же	Заявка US 2011081708A1 C 07H 21/04 G 06F 19/00 оп. 07.04.11	Genscript Holdings Hong Kong Ltd. US	Система и метод оптимизации последовательностей генов для повышения уровня экспрессии белков	Временная охрана
То же	Заявка WO 2010129168A2 A 01H1/04 оп. 11.11.10	(US)	Система и метод целевого фенотипирования признаков в экспериментах по селекции растений	Действует
То же	Патент EP 1322788 A2 C 12N 15/10 G 06F 19/00 оп. 03.01.02	Cytokinetics Inc. (US)	Система и метод анализа изображений для фенотипирования мутантных клеток	Действует
То же	Заявка US 2011033854A1 C 12Q 1/68 оп. 10.02.11	Complete Genomics Inc. (US)	Методы секвенирования длинных фрагментов последовательностей	Временная охрана
То же	Заявка CN 101751517 A C 12Q 1/68 оп. 23.06.10	Shenzhen BGI Res. Inst (CN)	Система и метод для быстрого картирования короткой последовательности в геноме	Временная охрана
То же	Заявка US 2007161024A1 C12Q 1/68	Agency Science Tech & Research (SG)	Метод анализа идентификационной сигнатуры гена (GIS)	Действует

	оп. 10.02.11			
То же	Патент GB 2477703 A G 06F 19/00 оп. 10.08.11	Real Time Genomics Inc.(US)	Метод и система для анализа данных о последовательнос-т ях	Действует
То же	Патент US 7809510 B2 G 01N 33/48 оп. 01.01.04	Ip Genesis, Inc. (US)	Метод позицион-ного хеширования для проведения поиска сходств в последовательности	Действует
Системы и методы получения трансгенных растений	Патент RU 2376377 C2 C 12N 15/82 оп. 20.12.09	Grain Ayotek Ostreylia Pty Ltd (AU)	Система и метод получения трансгенных растений пшеницы, устойчивых к стрессу.	Действует
То же	Патент RU 2382079 C2 C 12N 15/82 оп. 20.02.10	Univ. Zurich (CH)	Промотор для экспрессии трансгенов в эпидермисе растений	Действует
То же	Патент RU 2277586 C2 C 12N 15/09 оп. 10.06.06	Гапоненко А. К. (RU)	Способ получения трансгенных растений пшеницы, устойчивых к вре- дителям.	Действует
То же	Патент RU 2428480 C2 C 12N 15/82 оп. 10.09.11	BASF LANT ciencia GmbH (DE)	Способ повышения резистентности трансгенных расте-ний к патогенам	Действует
То же	Патент RU 2425151 C2 C 12N 15/82 оп. 27.07.11	Evgen Ltd. (IL)	Способ получения трансгенных растений пшеницы, устойчивых к стрессу.	Действует
То же	Патент RU 2384621 C2 C 12N 15/82 оп. 20.03.10	KROPDIZAYN N.Fi. (BE)	Способ получения трансгенных растений с повышенной урожаемностью	Действует
Компьютер-н ые програм-мы для биоинформа-т ики	Заявка US 2004068381 A1 G 06F 17/30 оп. 08.04.04	Daewood Educational Foundation (US)	Метод обработки баз данных в биоинформатике	Действует
То же	Патент RU 2335013 C2 G 06F 17/30 оп. 27.08.08	Google Inc. (US)	Способ и системы для улучшения ранжирования поиска с использо-ванием информации о статье	Действует
То же	Патент RU 2425566 C2 G06 F 17/30	Microsoft Corporation (US)	Использование обратной связи с пользователем для	Действует

	оп. 20.07.11		улучшения результатов поиска	
То же	Патент RU 2419861 C2 G 06F 17/30 оп. 27.05.11	Microsoft Corporation (US)	Система и способ ранжирования результатов поиска	Действует
То же	Патент RU 2419857 C2 G 06F 17/30 оп. 27.05.11	Microsoft Corporation (US)	Способ определения подобия объектов на основании гетерогенных связей	Действует
То же	Патент RU 2392660 C2 G 06F 17/30 оп. 20.06.11	ГОУ ВПО «Мордовский Гос. универ-ситет» (RU)	Система и способ поиска информации в массиве текстов	Действует
То же	Патент RU 2348072 C1 G 06F 17/30 оп. 27.02.09	Злыгостев А. (RU)	Способ оценки степени раскрытия понятия в тексте для поисковых систем	Действует
То же	Патент RU 2343537 C2 G 06F 17/30 оп. 10.01.09	Microsoft Corporation (US)	Компьютерный поиск с помощью ассоциативных связей	Действует
То же	Патент RU 2305314 C2 G 06F 17/30 оп. 27.08.07	ООО «ЦКМ» (RU)	Способ поиска и выборки информации из различных баз данных	Действует
То же	Патент RU 2282237 C2 G 06F 17/30 оп. 20.08.06	Военная академия связи (RU)	Система управления маршрутизацией текстовых документов в сети обработки данных	Действует
То же	Заявка US 2009012928A1 G 06F 17/30 оп. 08.01.09	Lussier Yves A (US)	Метод создания биоинформацион-ной базы данных	Действует



**Таблица В.6.2 - Научно-техническая информация**

Предмет поиска	Наименование источника	Автор, фирма (держатель) технической документации	Место и орган издания, год, номер, страницы источника (утверждения, депон. источника)
Информацион-ные ресурсы модульного типа для поддержки исследований в областях агро-биотехнологии и биоинженерии.	Towards the development of better crops by genetic transformation using engineered plant chromosomes.	Dhar M.K., Kaul S., Kour J.	Plant Cell Rep., 2011, 30, 799-806.
	PHENOPSIS DB: an information system for Arabidopsis thaliana phenotypic data in an environmental context.	Fabre J., Dauzat M., Nègre V. et all	BMC Plant Biol., 2011, 11, 77.
	Accurate inference of shoot biomass from high-throughput images of cereal plants.	Golzarian M.R., Frick R.A., Rajendran K. et all	Plant Methods, 2011, 7, 2.
То же	Biotechnological applications of microbial proteomes.	Han M.J., Lee S.Y., Koh S.T. et all	J. Biotechnol., 2010, 145, 341-349.
	HTPheno: an image analysis pipeline for high-throughput plant phenotyping.	Hartmann A., Czauderna T., Hoffmann R. et all	BMC Bioinformatics, 2011, 12, 148.
	New areas of plant-made pharmaceuticals.	Hassan S.W., Waheed M.T., Lössl A.G.	Expert Rev. Vaccines, 2011, 10, 151-153.
	Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems.	Iyer-Pascuzzi A.S., Symonova O., Mileyko Y. et all	Plant Physiol., 2010, 152, 1148-1157.
	Quantitative analysis of heterogeneous spatial distribution of Arabidopsis leaf trichomes using micro X-ray computed tomography.	Kaminuma E., Yoshizumi T., Wada T. et all	Plant J., 2008, 56, 470-482.
То же	Use of buckwheat seed protease inhibitor gene for improvement of tobacco and potato plant resistance to biotic stress.	Khadeeva N.V., Kochieva E.Z., Tcherednitchenko M.Y. et all	Biochemistry (Mosc), 2009, 74, 260-267.
То же	Transient expression systems for plant-derived biopharmaceuticals.	Komarova T.V., Baschieri S., Donini M. et all	Expert Rev. Vaccines, 2010, 9, 859-876.
	GERMINATE. A generic database for integrating genotypic and phenotypic information for plant genetic resource collections.	Lee J.M., Davenport G.F., Marshall D. et all	Plant Physiol., 2005, 139, 619-631.

Предмет поиска	Наименование источника	Автор, фирма (держатель) технической документации	Место и орган издания, год, номер, страницы источника (утверждения, депон. источника)
	Novel throughput phenotyping platforms in plant genetic studies.	Montes J.M., Melchinger A.E., Reif J.C.	Trends Plant Sci., 2007, 12, 433-436.
То же	Transgenic crops for the production of recombinant vaccines and anti-microbial antibodies.	Peters J., Stoger E.	Hum. Vaccin., 2011, 7, 367-374.
	Computer aided data acquisition tool for high-throughput phenotyping of plant populations.	Vankadavath R.N., Hussain A.J., Bodanapu R. et all	Plant Methods, 2009, 5, 18.
	Agricultural biotechnology for crop improvement in a variable climate: hope or hype?	Varshney R.K., Bansal K.C., Aggarwal P.K. et all	Trends Plant Sci., 2011, 16, 363-371.