

КОМПЬЮТЕРНАЯ СИСТЕМА WheatPGE ДЛЯ АНАЛИЗА ВЗАИМОСВЯЗИ ФЕНОТИП–ГЕНОТИП–ОКРУЖАЮЩАЯ СРЕДА У ПШЕНИЦЫ

М.А. Генаев¹, А.В. Дорошков¹, Е.В. Морозова¹, Т.А. Пшеничникова¹,
Д.А. Афонников^{1,2}

¹ Учреждение Российской академии наук Институт цитологии и генетики
Сибирского отделения РАН, Новосибирск, Россия, e-mail: mag@bionet.nsc.ru;

² Новосибирский государственный университет, Новосибирск, Россия

В настоящей работе для решения задачи интеграции генотипических и фенотипических данных, а также параметров окружающей среды и для анализа взаимосвязей между генотипом и фенотипом у пшеницы представлена система WheatPGE. Система служит для интеграции разнородных данных о растении, хранения информации об отношениях, описывающих различные характеристики растения, его генотипа, фенотипа и факторов внешней среды и доступа к ней. Система имеет web-интерфейс и доступна по адресу www.wheatdb.org. Возможности использования системы в генетическом эксперименте продемонстрированы на примере анализа количественных характеристик опушения листа пшеницы. Используя WheatPGE, мы провели классификацию опушения листьев 47 сортов яровой пшеницы методом многомерного шкалирования. Результаты анализа позволили выявить различия между опушением листа у растений российских и немецких сортов.

Ключевые слова: пшеница, интеграция данных, фенотип, генотип, окружающая среда, опушение листа, база данных.

Введение

Одна из важнейших проблем биологии – изучение взаимосвязи между генотипом и фенотипом у живых организмов. Для решения этой задачи в последнее время все более широко применяются компьютерно-экспериментальные подходы выявления и картирования генов, контролирующих количественные признаки (quantitative trait loci, QTL). Они основаны на использовании ассоциаций между молекулярными ДНК-маркерами и их расположением на хромосоме и фенотипическими данными организмов (Kearsey, 1998; Shavrukov *et al.*, 2010). Применение этих подходов требует анализа информации как по генотипам, так и по фенотипам организмов. Поэтому в последнее время возрастает интерес к разработке систем информационной поддержки хранения и обработки фенотипических данных и их интеграции с геномной информацией. Эти работы наиболее

систематически ведутся для видов, являющихся базовыми генетическими моделями, например, мышь (Hancock *et al.*, 2007; Grubb *et al.*, 2009; NIG Mouse Phenotype Database), крыса (Twigger *et al.*, 2007), человек (Kahraman *et al.*, 2005), кукуруза (VPhenoDBS: Maize). 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 другую геномную информацию.

Создание систем и подходов, интегрирующих информацию о генотипе, фенотипе и окружающей среде, – актуальная задача не только для модельных организмов, но и для важных сельскохозяйственных культур. Для разных видов растений в последнее время ведется активная разработка таких систем. Например, подход, необходимый для управления генетическими коллекциями, основанный на интегрирован-

ном описании фенотипа и генотипа растений, предложен в базе данных GERMINATE (Lee *et al.*, 2005). Интересным проектом является база данных SGH (Sol Genomic Network), которая содержит информацию о фенотипе, генотипе, полногеномных данных и генных и метаболических сетях для пасленовых (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. Важно отметить, что большинство описанных выше систем используют поддержку хранения информации о фенотипе в виде изображений, поскольку методы анализа изображений используются при анализе фенотипа растения все более интенсивно (Kaminuma *et al.*, 2004; Дорошков и др., 2009; Eberius, Lima-Guegta, 2009).

Однако основным недостатком таких подходов является то, что такие базы данных ориентированы на вид, а не на отдельное растение, которое является минимальной экспериментальной единицей в генетических исследованиях. Информация, характеризующая конкретное растение, в базе данных не представлена. Такой подход свойственен и для всех упомянутых выше баз данных, что делает их в большей степени информационными ресурсами (коллекцией результатов, полученных из литературных источников, геномных баз данных и т. п.). Это не позволяет использовать их для поддержки генетических экспериментов, обеспечивающих базу знаний для селекции по важным агрономическим признакам растений, и молекулярного картирования локусов количественных признаков (QTL).

Следует отметить, что, по некоторым оценкам, генотип обуславливает до 50% фенотипических вариаций (Kearsey, 1998). Остальная часть фенотипической изменчивости обусловлена влиянием факторов внешней среды. Эту изменчивость также необходимо учитывать при анализе данных.

Создание систем для поддержки лабораторных экспериментов является наиболее перспективным подходом в развитии интегрированных междисциплинарных направлений исследований. Подобным проектом является Chloroplast2010 (Ajjawi *et al.*, 2010). Высокопро-

изводительное фенотипирование, эффективный сбор, хранение большого объема фенотипических данных, их интеграция с геномными данными позволили создать прорывную технологию анализа взаимосвязи между генотипом и фенотипом у *Arabidopsis thaliana* (Lu *et al.*, 2008). Однако данная система не позволяет учитывать влияние окружающей среды на развитие фенотипа растения. PlantDB (Exner *et al.*, 2008) – 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 также не позволяет учитывать параметры внешней среды.

Примером системы, учитывающей параметры среды в ходе экспериментов, является PHENOPSIS DB (Fabre *et al.*, 2011), которая создана для документирования и накопления данных, полученных в ходе экспериментов по анализу *Arabidopsis thaliana*. В этой системе используются автоматизированные боксы для выращивания растений, в которых поддерживаются необходимые параметры окружающей среды: температура, влажность воздуха, интенсивность света и др. Эти параметры в режиме реального времени передаются на управляющий сервер системы PHENOPSIS DB, где происходит их хранение и последующая статистическая обработка. В ходе эксперимента производится автоматическое фотографирование различных органов растения (корни, листья, цветки). Цифровые изображения сохраняются на сервере, обрабатываются для получения количественных характеристик фенотипа, которые заносятся в базу данных системы. Генотип каждого растения представлен экотипом, линией, описанием семян, из которых растение было выращено, и другими параметрами. Однако такая база данных не подходит для исследования растений в естественных им полевых условиях, что особенно важно для работы с сельскохозяйственными культурами. Таким образом, назрела необходимость создания базы данных, с одной стороны, ориентированной на отдельное растение как объект генетического эксперимента, а с

другой – учитывающей его генотип и условия окружающей среды.

В СССР и России важной вехой в селекционно-генетических экспериментах на пшенице явился проект ДИАС (Драгавцев и др., 1984). 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 F_1 , F_2 с измерениями у каждого растения 19 количественных признаков в 9 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 (Драгавцев, 2005). Выполнение исследований позволило выявить основные закономерности эколого-генетической организации количественных признаков растений.

В настоящей работе для решения задачи сбора, интеграции, хранения и статистической обработки информации о растениях пшеницы мы предлагаем компьютерную систему **WheatPGE (Wheat Phenotype, Genotype and Environment)**. Система хранит различные отношения, описывающие характеристики отдельного растения, и позволяет однозначно устанавливать взаимосвязь между генотипическими и фенотипическими признаками растений, а также параметрами окружающей среды. Применение системы позволит автоматизировать получение данных о взаимосвязи генотипа, фенотипа и окружающей среды у пшеницы, способствуя тем самым эффективному созданию новых сортов пшеницы с улучшенными свойствами.

Материалы и методы

Для описания различных характеристик растений пшеницы нами была спроектирована реляционная база данных, которая лежит в основе системы **WheatPGE**. Текущая модель данных представлена 27 таблицами, связанными между

собой. В качестве сервера используется MySQL. Для работы с базой данных разработан веб-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 ориен-

тированного графа. Для размещения графа на плоскости и его рендеринга используется библиотека GraphViz (рис. 1).



Рис. 1. Пример визуализации схемы скрещивания для гибридного генотипа.

Серверная часть системы WheatPGE работает под управлением web-сервера Apache с модулем mod_perl и операционной системы CentOS Linux.

Структура базы данных

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

Генотип

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 направление скрещивания. Для генотипа можно определить список молекулярных мар-

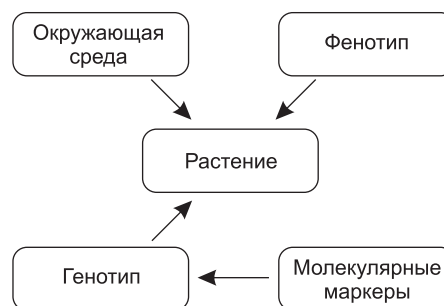


Рис. 2. Схема взаимосвязи между основными разделами информации в системе WheatPGE.

керов (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 (рис. 3). При описании маркера указываются его тип, имя, список состояний и локализация на хромосоме.

Фенотип

Для описания фенотипа растения система WheatPGE позволяет создавать наборы отношений, каждое из которых содержит описание характеристик определенного морфологического признака (опушение листа, окраска органов, форма колоса и др.), а также структуры урожая (длина побега и колоса, количество колосьев, продуктивность и др.).

В текущей версии базы данных наиболее полно представлено описание такого морфологического признака, как опушение листа. Для него заданы следующие характеристики: плотность опушения (количество 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. Для получения информации о количественных характеристиках опушения на основе анализа цифровых фотографий была использована программа LHDetect (Дорошков и др., 2009).

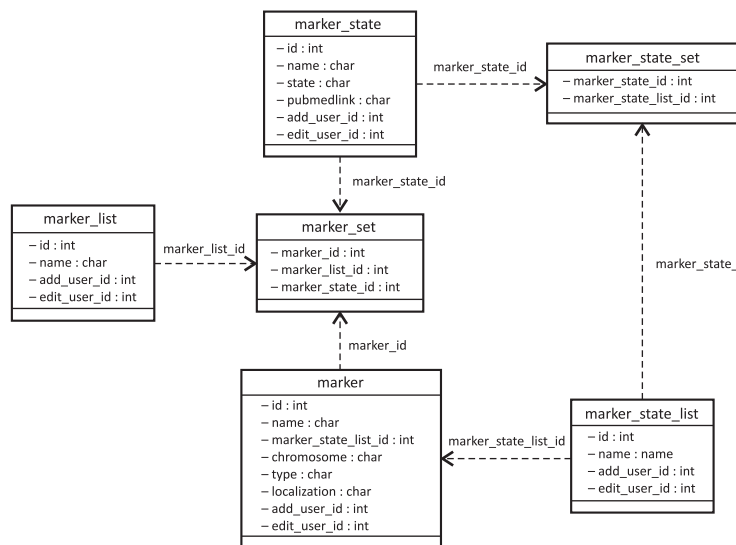


Рис. 3. Структурная схема реляционных отношений таблиц, описывающих молекулярные маркеры.

Структура базы данных позволяет легко расширять список анализируемых морфологических признаков растения и модифицировать информацию о них.

морфологические признаки или параметры окружающей среды, ему следует отправить запрос администратору системы с просьбой на расширение модели базы данных.

Окружающая среда

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.

Авторизация пользователей и разделение прав доступа

Пользователь может получить доступ к базе данных, зарегистрировавшись на сайте 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

Пользовательский интерфейс

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.

Поиск производится по следующим полям:

- полевой номер растения;
- название генотипа растения;
- сорт растения;
- линия;
- является ли растение гибридом или нет;
- название родительских генотипов;
- 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 Фора и Краса, в запросе достаточно написать *Фора|Краса*.

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 опушения листа от сорта растения пользователь должен на странице запроса указать список сортов растений (рис. 4, а), которые он хотел бы включить в анализируемую выборку, и указать список характеристик опушения (рис. 4, б). В итоге пользователь получает таблицу данных, в которой строкам соответствуют растения отобранных сортов, представленные в базе, а в колонках приводятся числовые характеристики опушения (рис. 4, в). Такая таблица может быть далее проанализирована любой программой статистического анализа (MS Excel, Statistica и др.).

Результаты и обсуждение

Текущая версия системы содержит 309 аннотированных растений, 88 генотипов и 1335 микроизображений сгиба листа пшеницы, позволяющих анализировать опушение – один из важных фенотипических признаков растения (Крупнов, Цапайкин, 1990).

Ранее с использованием подходов компьютерной обработки изображений, интегрированных в систему **WheatPGE**, нами была исследована морфология опушения листа ряда сортов пшеницы (Дорошков и др., 2011). В частности, было показано, что сорта Голубка и Саратовская 29, имеющие схожий тип опушения, различаются в соотношении трихом разной длины. Генетический анализ показал, что сорт Саратовская 29 несет 2 гена, а сорт Голубка – 1 ген опушения листа. Первый сорт по сравнению с сортом Голубка обладает большей плотностью трихом, которые имеют большую длину. На примере сорта Голубка исследовано влияние условий внешней среды на формирование опушения. Показано, что в полевых условиях растения этого сорта формируют большее количество трихом, но меньшей длины. Следует отметить, что проведение столь детального анализа опушения

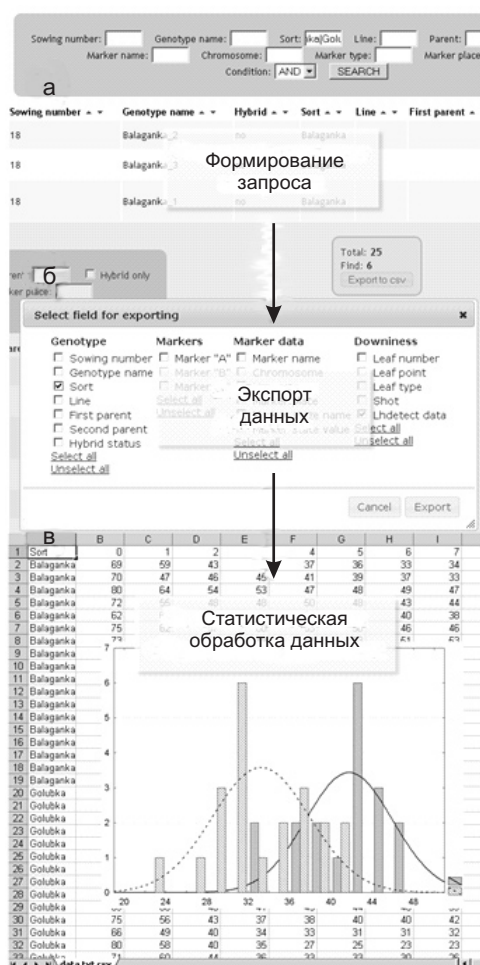


Рис. 4. Пример работы с системой WheatPGE. Извлечение данных о взаимосвязи сорта растения и его опушения.

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.

листа стало возможно исключительно благодаря разработке и внедрению современных методов получения и обработки данных.

В настоящей работе с использованием разработанной компьютерной системы мы выполнили классификацию опушения различных сортов пшеницы методом многомерного шкалирования. Было отобрано 47 различных сортов, содержащихся в базе данных системы WheatPGE.

Эти сорта были выращены в условиях поля в окрестностях Новосибирска в 2008 г. Для определения количественных характеристик опушения растений нами были использованы

их предфлаговые листья. Сформировавшиеся здоровые листья были отобраны, и из них были сделаны препараты сгиба листа, для которых по описанной ранее методике (Дорошков и др., 2009) были получены микроизображения. Всего проанализировано по 6 микроизображений для разных растений каждого сорта.

Определение количественных характеристик опушения на основе цифрового анализа микроизображений производилось с помощью программы LHDetect2 (LHDetect2), которая является улучшенной версией программы LHDetect (Дорошков и др., 2009) и интегрирована в систему WheatPGE. Для каждого микроизображения сгиба листа рассчитывалась гистограмма распределения трихом по длине в пределах 0–2500 мкм с шагом 31,5 мкм. На основе гистограммы для каждого изображения строился вектор n , элемент n_i которого соответствовал числу трихом с длиной в интервале от $31,5 \cdot (i-1)$ мкм до $31,5 \cdot i$ мкм, $i = 1, \dots, 80$. Значения элементов n для растений одного и того же сорта усреднялись. Таким образом, была сформирована таблица данных, содержащая 47 строк, в которой каждая строка соответствовала определенному сорту пшеницы, а каждый столбец соответствовал среднему числу трихом в i -м интервале длины для растений этого сорта.

Данные были проанализированы при помощи метода неметрического многомерного шкалирования (Oksanen, 2011). Этот метод позволяет построить расположение точек, соответствующих сортам растений, на плоскости так, что взаимные расстояния между ними приблизительно пропорциональны евклидовым расстояниям между векторами n для этих сортов в многомерном пространстве, т. е. чем ближе друг к другу на плоскости располагаются точки, тем более сходное распределение трихом по длинам у соответствующих им сортов пшеницы. При анализе данных мы использовали математический аппарат, реализованный в пакете *vegan*, написанный на языке программирования R (**R-project**).

Результат классификации показан на рис. 5. На диаграмме хорошо выделяются два кластера. Первый располагается в верхнем левом углу диаграммы и состоит из сортов яровых пшениц, созданных в разное время на территории Поволжья и Сибири. Они имеют густое опушение листа. Во второй кластер (расположен справа) попали немецкие современные и стародавние сорта из коллекции Генбанка Института генетики растений (г. Гатерслебен Германия), имеющие слабое опушение листа.

Эти данные согласуются с предположением о том, что сильное опушение характерно для

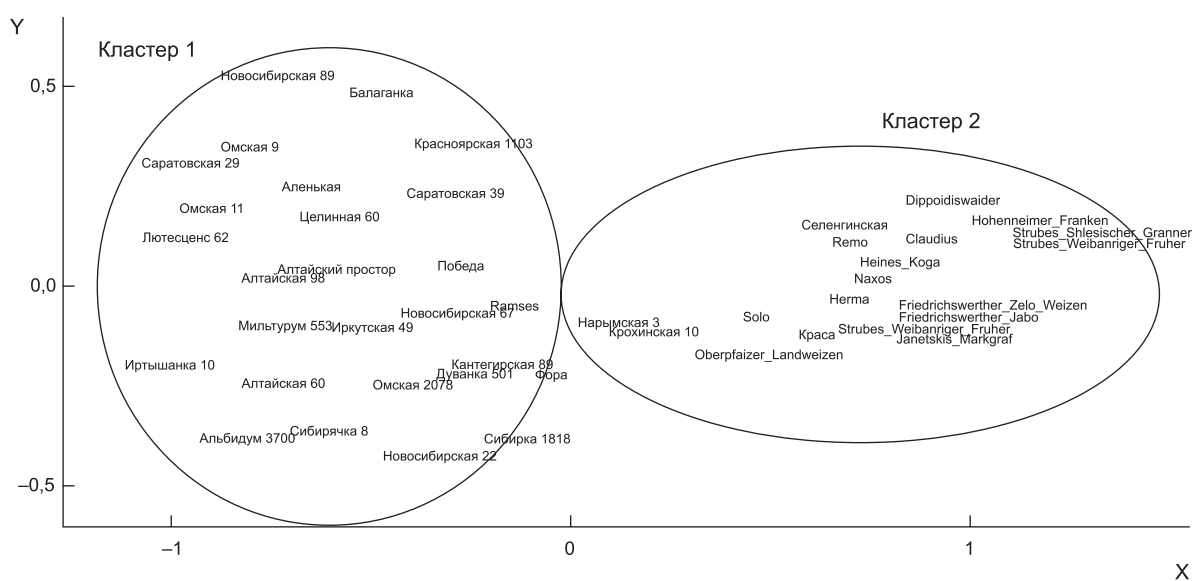


Рис. 5. Классификация 47 сортов пшеницы по опушению листа, полученная с помощью метода неметрического многомерного шкалирования.

Два кластера сортов, имеющих сходный тип опушения, выделены овалами и подписаны.

ряда засухоустойчивых сортов, относящихся к степной экологической группе; для сортов, полученных в условиях большего увлажнения, напротив, характерно более слабое опушение (Крупнов, Цапайкин, 1990).

Более детальный анализ распределения трихом по длинам у растений из разных кластеров показал следующее. Сорт Hohenheimer Franken, попавший во второй кластер (высокое значение координаты X и близкое к нулю значение координаты Y), содержит небольшое количество трихом. Сорта Балаганка и Лютесценс 62 (низкое значение координаты X) хорошо опушены. Таким образом, можно предположить, что ось X на рис. 5 отрицательно коррелирует с интенсивностью опушения: чем выше значение координаты X, тем слабее опушение листа. Мы рассчитали коэффициент корреляции Пирсона между значениями координаты X указанных сортов, средней длиной трихом $\langle L \rangle$ и средним числом трихом $\langle N \rangle$. Оказалось, что эти коэффициенты значимы ($p < 0,05$) и отрицательны ($r(X, \langle L \rangle) = -0,99$; $r(X, \langle N \rangle) = -0,82$). Это подтверждает наше предположение о взаимосвязи между осью X и интенсивностью опушения листа. Отметим, что между значениями $\langle L \rangle$ и $\langle N \rangle$ в нашей выборке также существует значимая положительная корреляция: $r(\langle L \rangle, \langle N \rangle) = 0,78$ ($p < 0,05$). Такая зависимость между средней длиной трихом и их средним числом на листе была обнаружена нами ранее для растений двух генотипов пшеницы (Дорошков и др., 2011) и может являться проявлением специфического механизма генетического контроля формирования опушения у пшеницы.

Отметим, что анализ корреляций значений оси Y и параметров $\langle L \rangle$ и $\langle N \rangle$ не выявил значимой связи между ними ($r(Y, \langle L \rangle) = 0,03$; $r(Y, \langle N \rangle) = 0,03$).

Интересно отметить, что растения кластера 1 на диаграмме (рис. 5) распределены в левой полуплоскости практически равномерно, в то время как растения кластера 2 формируют сильно вытянутое вдоль оси X облако. Оказалось, что коэффициенты корреляции между значениями $\langle L \rangle$ и $\langle N \rangle$ у растений этих двух кластеров сильно отличаются. Для кластера 1 корреляция не значима ($r_1(\langle L \rangle, \langle N \rangle) = 0,06$, $p > 0,05$). Для второго кластера она выше, чем для всей выборки сортов ($r_2(\langle L \rangle, \langle N \rangle) = 0,80$, $p > 0,05$). Это

демонстрирует возможные генетические различия в формировании опушения у российских и немецких сортов. По-видимому, генетический механизм формирования опушения у сортов из Германии является сходным. Число генов, контролирующих опушение, невелико, и они одинаковым образом действуют и на формирование длин трихом, и на их число.

Генетические механизмы формирования опушения у российских сортов более разнообразны. Опушение у них может формироваться большим числом аллелей, действующих независимо друг от друга, при этом число трихом и их длина контролируются независимо и специфично для каждого генотипа, что отражается в отсутствии корреляций между этими признаками. Данное предположение подтверждается тем, что 4 сорта из российской коллекции (Селенгинская, Нарымская 3, Краса, Крохинская 10) оказались в одном кластере вместе с немецкими сортами. Это говорит о том, что в процессе их создания (гибридизация, многократные отборы) из генотипов были утеряны некоторые гены, контролирующие густое опушение листа.

Также интересно отметить, что сорта, имеющие в своей родословной сорт Саратовская 29 (Омская 9, Омская 11, Новосибирская 89) или общее место происхождения (Лютесценс 62), расположены в отдельной общей группе. Это означает, что они имеют сходный характер опушения: самое густое из всех представленных сортов.

Таким образом, проведенная кластеризация демонстрирует свою применимость для анализа экспериментальных результатов и может быть с успехом использована для классификации и всестороннего описания образцов в генетических коллекциях, а также новых линий и сортов пшеницы.

Заключение

Разработанная система WheatPGE позволяет устанавливать и анализировать взаимосвязь между генетическими и фенотипическими признаками растений и параметрами окружающей среды у пшеницы. Это обеспечивает поддержку целого ряда важных биологических задач. Прежде всего, это задачи подбора генетического материала для различных экспериментов, поиск

и картирование генов, отвечающих за определенные признаки, особенности их фенотипического проявления, зависимости от различных факторов, таких, как условия внешней среды, стадии развития и т. п. Задачи, связанные с описанием генетических коллекций и работой с ними, также могут осуществляться с использованием средств системы WheatPGE.

На данный момент возможности анализа изображений системы WheatPGE позволяют быстро и эффективно проводить исследование опушения листа пшеницы, устанавливают зависимость морфологических характеристик опушения от сорта растения, места произрастания. Они облегчают поиск генетических маркеров, статистически связанных с тем или иным типом опушения пшеницы.

Сорта из Генбанка Института генетики растений любезно предоставлены доктором Андреасом Бернером (Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung, г. Гатерслебен, Германия).

Работа поддержана программами РАН Б.25 и А.П.6, интеграционным проектом СО РАН 119, Министерством образования и науки Российской Федерации, ГК 07.514.11.4003, ГК 07.514.11.4023, ГК 07.514.11.4052 и грантом РФФИ 11-04-01748а.

Литература

- Дорошков А.В., Арсенина С.И., Пшеничникова Т.А., Афонников Д.А. Применение компьютерного анализа микроизображений листа для оценки характеристик опушения пшеницы *Triticum aestivum* L. // Информ. вестник ВОГиС. 2009. Т. 13. № 1. С. 218–226.
- Дорошков А.В., Пшеничникова Т.А., Афонников Д.А. Анализ особенностей морфологии и наследования опушения листа пшеницы *Triticum aestivum* L. с помощью компьютерных методов фенотипирования // Генетика. 2011. Т. 47. № 6. С. 836–841.
- Драгавцев В.А., Цильке Р.А., Рейтер Б.Г. и др. Генетика признаков продуктивности яровых пшениц в Западной Сибири. Новосибирск: СО АН, 1984. 121 с.
- Драгавцев В.А. Библиография деятелей науки. СПб: ГНУ ГНЦ РФ ВИР, 2005. С. 6.
- Крупнов В.А., Цапайкин А.П. Опушение листьев пшеницы: генетические и экологические аспекты // С.-х. биология. Сер. «Биология растений». 1990. № 1. С. 51–57.
- Ajjawi I., Lu Y., Savage L.J. et al. Large-scale reverse genetics in Arabidopsis: case studies from the Chloroplast 2010 Project // Plant Physiol. 2010. V. 152. P. 529–540.
- Eberius M., Lima-Guerra J. High-Throughput Plant Phenotyping – Data Acquisition, Transformation, and Analysis // Bioinformatics: Tools and Applications / Ed. D. Edwards et al. Springer Science+Business Media, 2009. P. 259–278.
- Exner V., Hirsch-Hoffmann M., Gruissem W., Hennig L. PlantDB – a versatile database for managing plant research // Plant Methods. 2008. V. 4. P. 1.
- Fabre J., Dauzat M., Nègre V. et al. PHENOPSIS DB: an Information system for *Arabidopsis thaliana* phenotypic data in an environmental context // BMC Plant Biol. 2011. V.11. P. 77.
- Grubb S.C., Maddatu T.P., Bult C.J., Bogue M.A. Mouse phenome database // Nucl. Acids Res. 2009. N 37. P. D720–D730.
- Hancock J.M., Adams N.C., Aidinis V. et al. Mouse phenotype database integration consortium // Mamm. Genome. 2007. V. 18. N 3. P. 157–163.
- Kahraman A., Avramov A., Nashev L.G. et al. PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics // Bioinformatics. 2005. V. 21. P. 418–420.
- Kaminuma E., Heida N., Tsumoto Y. et al. Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis // Plant J. 2004. V. 38. P. 358–365.
- Kearsey J.M. The principles of QTL analysis (a minimal mathematic approach) // J. Exp. Bot. 1998. V. 49. P. 1619–1623.
- Lee J.M., Davenport G.F., Marshall D. et al. GERMINATE. a generic database for integrating genotypic and phenotypic information for plant genetic resource collections // Plant Physiol. 2005. V. 139. P. 619–631.
- LHDetect2: Leaf image processing for wheat hairiness, available at <http://wheatdb.org/lhdetect2>.
- Lu Y., Savage L.J., Ajjawi I. et al. New connections across pathways and cellular processes: industrialized mutant screening reveals novel associations between diverse phenotypes in Arabidopsis // Plant Physiol. 2008. V. 146. P. 1482–1500.
- NIG Mouse Phenotype Database, available at <http://molossinus.lab.nig.ac.jp/phenotype/>.
- Oksanen J. Multivariate Analysis of Ecological Communities in R: vegan tutorial. 2011, available at <http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf>
- R-project: R: A language and environment for statistical computing, available at <http://www.R-project.org>.
- Sol Genomic Network, available at <http://solgenomics.net/>.

Shavrukov Y., Gupta N.K., Miyazaki J. *et al.* HvNax3—a locus controlling shoot sodium exclusion derived from wild barley (*Hordeum vulgare* ssp. spontaneum) // *Funct. Integr. Genomics*. 2010. V. 10. P. 277–291.

Twigger S.N., Shimoyama M., Bromberg S. *et al.*

RGD Team. The rat genome database, update 2007—easing the path from disease to data and back again // *Nucl. Acids Res.* 2007. N 35 (Database issue). P. D658–D662.

VPhenoDBS: Maize. available at <http://vphenodbs.mnet.missouri.edu/>

WheatPGE: A SYSTEM FOR ANALYSIS OF RELATIONSHIPS AMONG PHENOTYPE, GENOTYPE AND ENVIRONMENT IN WHEAT

M.A. Genaev¹, A.V. Doroshkov¹, E.V. Morozova¹, T.A. Pshenichnikova¹, D.A. Afonnikov^{1,2}

¹Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia,
e-mail: mag@bionet.nsc.ru;

²Novosibirsk State University, Novosibirsk, Russia

Summary

The WheatPGE system, presented here, has been designed to solve the problem of integration of genotypic and phenotypic data and parameters of the environment, as well as to analyze the relationships between the genotype and phenotype in wheat. The system is used to consolidate miscellaneous data on a plant for storing and processing various morphological traits and genotypes of wheat plants as well as data on various environmental factors. The system is available at www.wheatdb.org. Its potential in genetic experiments has been demonstrated in quantitative indices of wheat leaf pubescence. We have classified 47 common wheat varieties according to leaf pubescence by multidimensional scaling. The results reveal differences between leaf pubescence in Russian and German wheat varieties.

Key words: wheat, data consolidation, phenotype, genotype, environment, leaf pubescence, database.