

Перевод на английский язык <https://vavilov.elpub.ru/jour>

Приоритизация биологических процессов на основе реконструкции и анализа ассоциативных генных сетей, описывающих ответ растений на неблагоприятные факторы внешней среды

П.С. Деменков^{1, 2}, Е.А. Ощепкова¹, Т.В. Иванисенко¹, В.А. Иванисенко^{2, 3} 

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

³ Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

 salix@bionet.nsc.ru

Аннотация. Методы приоритизации или ранжирования кандидатных генов по их важности в соответствии с заданными критериями, основанными на анализе генных сетей, широко применяются в биомедицине для поиска ассоциаций генов с заболеваниями, предсказания биомаркеров, фармакологических мишеней и т.д. При этом наблюдается тенденция их использования и в других областях знаний, в частности в растениеводстве. В значительной степени это обусловлено развитием технологий для решения задач маркер-ориентированной и геномной селекции, требующих знаний о молекулярно-генетических механизмах, лежащих в основе формирования хозяйственно ценных признаков. Новым направлением для изучения молекулярно-генетических механизмов является приоритизация биологических процессов с применением анализа ассоциативных генных сетей. Ассоциативная генная сеть – это гетерогенная сеть, в качестве вершин которой наряду с молекулярно-генетическими объектами (гены, белки, метаболиты и т.д.) могут быть представлены сущности более высокого уровня (биологические процессы, заболевания, факторы внешней среды и т.д.), связанные между собой регуляторными, физико-химическими или ассоциативными взаимодействиями. С использованием разработанного нами ранее метода осуществлена приоритизация биологических процессов по степени их связи с генными сетями, представленными в базе знаний SOLANUM TUBEROSUM и описывающими ответ растений на повышенное содержание кадмия, солевой стресс и условия засухи. Результаты приоритизации свидетельствуют о том, что фундаментальные процессы, такие как экспрессия генов, посттрансляционная модификация, деградация белков, программируемая клеточная смерть, фотосинтез, передача сигналов, ответ на стресс, играют важную роль в общих молекулярно-генетических механизмах ответа растений на различные неблагоприятные факторы. С другой стороны, среди специфических для устойчивости к засухе была выявлена группа процессов, связанных с развитием семян (seeding development). Процессы, связанные с ионным транспортом (ion transport), вошли в список специфических для ответа на солевой стресс, а связанные с метаболизмом липидов (phospholipid degradation – деградация фосфолипидов) – для ответа на кадмий. Ключевые слова: база знаний SOLANUM TUBEROSUM; Gene Ontology; *Arabidopsis thaliana*; методы text mining; ассоциативные генные сети; центральность вершин; сетевые методы приоритизации.

Для цитирования: Деменков П.С., Ощепкова Е.А., Иванисенко Т.В., Иванисенко В.А. Приоритизация биологических процессов на основе реконструкции и анализа ассоциативных генных сетей, описывающих ответ растений на неблагоприятные факторы внешней среды. *Вавиловский журнал генетики и селекции*. 2021;25(5):580-592. DOI 10.18699/VJ21.065

Prioritization of biological processes based on the reconstruction and analysis of associative gene networks describing the response of plants to adverse environmental factors

P.S. Demenkov^{1, 2}, E.A. Oshchepkova¹, T.V. Ivanisenko¹, V.A. Ivanisenko^{2, 3} 

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

 salix@bionet.nsc.ru

Abstract. Methods for prioritising or ranking candidate genes according to their importance based on specific criteria via the analysis of gene networks are widely used in biomedicine to search for genes associated with diseases and to predict biomarkers, pharmacological targets and other clinically relevant molecules. These methods have also been used in other fields, particularly in crop production. This is largely due to the development of technologies to solve problems in marker-oriented and genomic selection, which requires knowledge of the molecular genetic mechanisms underlying the formation of agriculturally valuable traits. A new direction for the study of molecular genetic mechanisms is the prioritization of biological processes based on the analysis of associative gene networks. Associative gene networks are heterogeneous networks whose vertices can depict both molecular genetic objects (genes, proteins, metabolites, etc.) and the higher-level

factors (biological processes, diseases, external environmental factors, etc.) related to regulatory, physicochemical or associative interactions. Using a previously developed method, biological processes involved in plant responses to increased cadmium content, saline stress and drought conditions were prioritized according to their degree of connection with the gene networks in the SOLANUM TUBEROSUM knowledge base. The prioritization results indicate that fundamental processes, such as gene expression, post-translational modifications, protein degradation, programmed cell death, photosynthesis, signal transmission and stress response play important roles in the common molecular genetic mechanisms for plant response to various adverse factors. On the other hand, a group of processes related to the development of seeds ("seeding development") was revealed to be drought specific, while processes associated with ion transport ("ion transport") were included in the list of responses specific to salt stress and processes associated with the metabolism of lipids were found to be involved specifically in the response to cadmium.

Key words: knowledge base SOLANUM TUBEROSUM; Gene Ontology; *Arabidopsis thaliana*; text mining methods; associative gene networks; centrality of vertices; network-based prioritization methods.

For citation: Demenkov P.S., Oshchepkova E.A., Ivanisenko T.V., Ivanisenko V.A. Prioritization of biological processes based on the reconstruction and analysis of associative gene networks describing the response of plants to adverse environmental factors. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2021;25(5):580-592. DOI 10.18699/VJ21.065

Введение

Быстрое развитие экспериментальных высокопроизводительных методов значительно расширило возможности генерации больших наборов геномных, транскриптомных и протеомных данных в научных исследованиях, что, в свою очередь, обусловило высокую актуальность разработок биоинформатических методов, позволяющих осуществлять интерпретацию омиксных данных на уровне ключевых генов, так и молекулярно-генетических механизмов. Одним из широко используемых подходов к компьютерному анализу выявленных в экспериментах наборов генов относятся методы приоритизации (Raj, Sreeja, 2018). Методы приоритизации решают задачу ранжирования исследуемого множества генов (или других объектов, например заболеваний) на основе критериев, характеризующих их близость к множеству из заданной обучающей выборки. Обучающая выборка, в зависимости от задачи приоритизации, состоит из генов, ассоциированных с заболеваниями или фенотипическими признаками, наборов дифференциально экспрессирующихся генов и т. д. Чем выше показатель близости в относительных единицах, тем больший приоритет имеет анализируемый объект в качестве кандидата, обладающего теми же свойствами, что и объекты из обучающего множества. Такие методы применяются в биомедицине для выявления генов-кандидатов, ассоциированных с заболеваниями (Tranchevent et al., 2016), биомаркеров заболеваний (Jha et al., 2020), потенциальных фармакологических мишеней (Cesur et al., 2020), для перепрофилирования лекарств (Pushrakom et al., 2019). В животноводстве и растениеводстве методы приоритизации нашли применение для анализа геномных данных, связанных с решением задач маркер-ориентированной и геномной селекции (Arruda et al., 2016; Crossa et al., 2017; Kochetov et al., 2017; Kolchanov et al., 2017; Cai et al., 2019; Voss-Fels et al., 2019; Sun et al., 2020), анализа локусов количественных признаков (Bargsten et al., 2014; Schaefer et al., 2018; Lin et al., 2019).

Особое место среди методов приоритизации заняли подходы, основанные на анализе графов генных сетей, включая сети белок-белковых взаимодействий, метаболические сети, сети передачи сигналов, сети ген-заболевание и т. д. В таких сетевых методах близость исследуемых генов к обучающей выборке оценивается с помощью

различных топологических характеристик графа генной сети. Методы анализа структуры графов генных сетей для решения задач приоритизации можно разделить на три большие группы (Shim et al., 2017; Raj, Sreeja, 2018): 1) методы, основанные на выявлении вершин-хабов с помощью оценок показателей центральности вершин в графе (Cho et al., 2016); 2) основанные на сетевой диффузии, включая случайное блуждание (Chen et al., 2011; Shim, Lee, 2015; Le, Pham, 2017; Lysenko et al., 2017); 3) основанные на идентификации функциональных модулей (кластеров или подсетей) (Jia et al., 2011; Leung et al., 2014). Все эти методы направлены на выявление генов (или иных сущностей), которые важны для изучаемого процесса или фенотипа. Важность информации о топологии сетей для оценки функциональной значимости генов была продемонстрирована на примере показателя центральности вершин в сети белок-белковых взаимодействий дрожжей *Saccharomyces cerevisiae* (Jeong et al., 2001). Авторы показали, что делеция вершин, характеризующихся большим числом связей в сети белок-белковых взаимодействий, чаще оказывалась летальной по сравнению с делециями других вершин.

Часто используемой характеристикой значимости вершины в структуре сети являются различные показатели центральности, включая центральность по степени (degree centrality), которая представляет собой число связей данной вершины с другими вершинами сети (Freeman, 1978), центральность по близости к центру графа (closeness centrality), рассчитываемую как обратная величина суммы длин всех наикратчайших путей, проходящих через вершину (Sabidussi, 1966), и центральность по посредничеству (betweenness centrality), представляющую собой количество наикратчайших путей, проходящих через вершину (Freeman, 1977).

Ранее нами было предложено понятие ассоциативной генной сети (Ivanisenko V.A. et al., 2015) как расширенной генной сети, в качестве вершин которой наряду с молекулярно-генетическими объектами (гены, белки, метаболиты и т. д.) могут быть представлены сущности более высокого уровня (биологические процессы, фенотипические признаки, заболевания, факторы внешней среды и др.), связанные между собой регуляторными, физическими, химическими или ассоциативными взаи-

модействиями (Ivanisenko V.A. et al., 2015). С помощью методов автоматического анализа текстов научных публикаций и фактографических баз данных были созданы базы знаний: ANDSystem, содержащая ассоциативные генные сети для животных и человека (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020), а также SOLANUM TUBEROSUM (Сайк и др., 2017; Иванисенко и др., 2018), содержащая ассоциативные генные сети растений. На основе анализа ассоциативных генных сетей из баз знаний ANDSystem и SOLANUM TUBEROSUM разработаны методы приоритизации генов человека, ассоциированных с заболеваниями (Saik et al., 2018, 2019), и генов картофеля, вовлеченных в формирование селекционно-значимых признаков (Деменков и др., 2019), соответственно. Метод приоритизации базируется на оценке показателя центральности по степени вершин, соответствующих генам, в анализируемых ассоциативных генных сетях.

В настоящей работе для приоритизации биологических процессов предложен подход, основанный на расчете показателя центральности вершин в ассоциативной генной сети. На примере модельного растения *Arabidopsis thaliana* (резуховидка Таля) с использованием базы знаний ассоциативных генных сетей SOLANUM TUBEROSUM проведена приоритизация биологических процессов, связанных с ответом растений на неблагоприятные факторы: содержание в почве кадмия, условия засухи, солевой стресс.

Материалы и методы

Программно-информационная система SOLANUM TUBEROSUM (Сайк и др., 2017; Иванисенко и др., 2018) состоит из трех основных блоков:

- 1) блок автоматического анализа текстов научных публикаций и фактографических баз данных, предназначенный для извлечения информации о взаимоотношениях между объектами с помощью семантико-лингвистических шаблонов. Для извлечения знаний используются методы и программные инструменты ANDSystem (Ivanisenko V.A. et al., 2015), настроенные на исследуемую предметную область. Настройка на предметную область, в частности, включала создание новых семантико-лингвистических шаблонов, учитывающих специфику текстов научных публикаций в области биологии растений и растениеводства;
- 2) база знаний, содержащая словари объектов и информацию о взаимосвязях между объектами, экстрагированную с помощью средств блока 1, в виде ассоциативной генной сети (графа, вершины которого соответствуют объектам, а ребра – связям заданного типа). В базе знаний SOLANUM TUBEROSUM представлены словари молекулярно-генетических объектов (гены, белки, метаболиты, микроРНК), биологических процессов Gene Ontology (Gene Ontology Consortium, 2019), фенотипических признаков, болезней и возбудителей заболеваний картофеля и модельных организмов. Для описания взаимосвязей между этими объектами в ассоциативных генных сетях используются более 25 различных типов связей, соответствующих физическим взаимодействиям, каталитическим реакциям,

регуляции, участию, ассоциациям (взаимосвязям, природа которых не установлена) и др.;

- 3) программа ANDVisio (Demenkov et al., 2012), предназначенная для реконструкции и анализа ассоциативных генных сетей, связанных с изучаемой проблемой, на основе обращения к базе знаний SOLANUM TUBEROSUM. Доступ к SOLANUM TUBEROSUM может осуществляться через интернет по адресу <https://www-bionet.sccc.ru/and/plant/>.

Статистика о наполнении базы знаний SOLANUM TUBEROSUM для трех видов организмов представлена в табл. 1.

Принципиальная схема всех этапов работы, выполняемой в ходе приоритизации, показана на рис. 1. Детальная схема формирования базы знаний SOLANUM TUBEROSUM (см. рис. 1, а) описана в работах (Сайк и др., 2017; Иванисенко и др., 2018). Алгоритм приоритизации состоит из нескольких шагов (см. рис. 1, б), которые полностью автоматизированы в программе ANDVisio. Программа ANDVisio предоставляет пользовательский интерфейс доступа к базе знаний SOLANUM TUBEROSUM, обладающий широким спектром различных инструментов реконструкции, графической визуализации и анализа ассоциативных генных сетей.

На первом шаге осуществляется автоматическая реконструкция ассоциативной генной сети, связанной с изучаемой проблемой, которая задается интересующим пользователя биологическим процессом из базы данных Gene Ontology. Для решения этой задачи используется реализованная в ANDVisio программная процедура «реконструкция сети». В окне интерфейса процедуры выбирается имя биологического процесса Gene Ontology из предоставленного списка (содержит полный перечень биологических процессов Gene Ontology, обновляемый ежегодно). В результате исполнения процедуры ассоциативная генная сеть, связанная с заданным биологическим процессом, будет автоматически реконструирована и визуализирована в графическом виде. Алгоритм реконструкции включает обращение с помощью MySQL запросов к базе знаний SOLANUM TUBEROSUM, в результате чего формируется список генов, для которых в базе знаний представлена связь с заданным пользователем биологическим процессом. Пользователь имеет возможность выбрать связи, выявленные путем анализа текстов научных публикаций или экстрагированные из фактографических баз данных. Гены из сформированного списка являются вершинами в искомой ассоциативной генной сети. Формальное описание алгоритма приведено на рис. 1, б. С помощью этого алгоритма проводилась реконструкция сетей для трех биологических процессов: “response to cadmium ion” (ответ на кадмий), “drought tolerance” (устойчивость к засухе) и “response to salt stress” (ответ на солевой стресс).

На втором шаге для каждой реконструированной ассоциативной генной сети осуществлялась приоритизация биологических процессов согласно алгоритму, показанному на рис. 1, б. Приоритизация представляет собой автоматизированный итерационный процесс, выполняемый процедурами в программе ANDVisio. Коротко он может

Таблица 1. Количество основных типов связей, включая молекулярно-генетические взаимодействия и ассоциации, представленных в базе знаний SOLANUM TUBEROSUM для трех организмов

Тип связи	<i>Solanum tuberosum</i> /potato	<i>Arabidopsis thaliana</i> /arabidopsis	<i>Zea mays</i> /maize
Association (ассоциация)	4783	294 520	15 280
Catalyze (катализ)	1140	26 842	1203
Regulation* (регуляция)	346	20 213	1120
Interaction (взаимодействие)	69	16 589	178
Involvement (участие)	3857	81 067	5906

* Показано суммарное число связей для следующих типов: upregulation (усиливающая регуляция), downregulation (подавление), regulation (регуляция), activity upregulation (регуляция, усиливающая активность), activity downregulation (подавление активности), activity regulation (регуляция активности), expression upregulation (регуляция, усиливающая экспрессию), expression downregulation (подавление экспрессии), expression regulation (регуляция экспрессии), transport upregulation (регуляция, усиливающая транспорт), transport downregulation (подавление транспорта), transport regulation (регуляция транспорта).

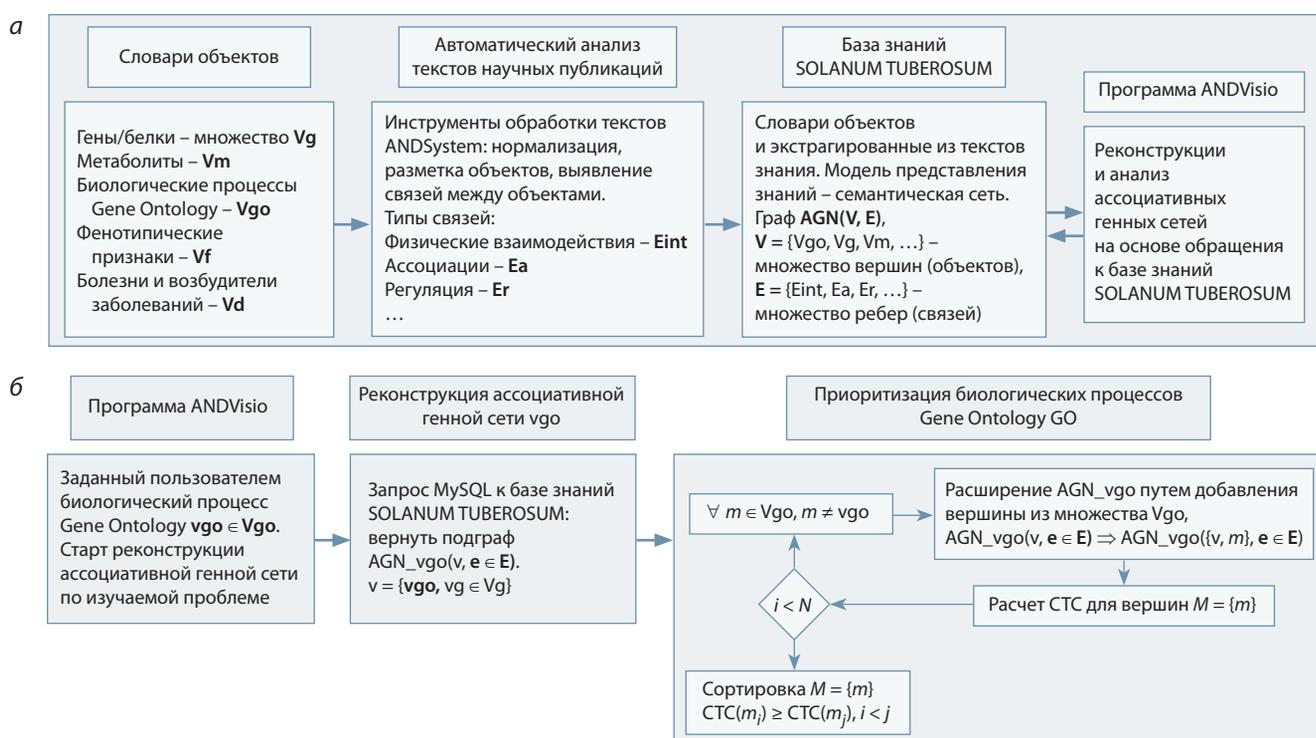


Рис. 1. Принципиальная схема всех этапов работы, выполняемой в ходе приоритизации:

а – создание базы знаний SOLANUM TUBEROSUM; б – приоритизация биологических процессов Gene Ontology (Gene Ontology Consortium, 2019).

быть описан следующим образом: последовательно перебираются все биологические процессы Gene Ontology, представленные в словарях базы знаний SOLANUM TUBEROSUM. Для каждого процесса производится: расширение ассоциативной геной сети, реконструированной на первом шаге алгоритма, путем добавления в сеть новой вершины, соответствующей текущему процессу; расчет показателя CTC (cross-talk centrality) для добавленной вершины; сортировка всех проанализированных процессов по величине значений CTC. Рассчитанные значения CTC ранжировались по степени убывания; таким образом, биологические процессы с наибольшим показателем получили максимальный приоритет.

Показатель CTC для вершины i рассчитывался по формуле

$$CTC_i = N_i/M, \quad M \neq 0,$$

где N_i – число вершин, с которыми i -я вершина связана ребром в анализируемом графе ассоциативной геной сети; M – общее число вершин ассоциативной геной сети; i принимает значения от 1 до M . Область значений показателя CTC лежит в пределах от 0 до 1. Значение 0 возникает в случае, когда рассматриваемая вершина в сети не связана с другими вершинами. Характеристика N_i известна как показатель центральности вершины по степени (Freeman, 1978). Использование этого показателя обусловлено предположением, что чем больше вершин оказываются связаны с данной вершиной, тем большее влияние она может оказывать на ассоциативную геной сеть и описываемый ею молекулярно-генетический механизм в целом. Для оценки значимости центральности вершины (биологического процесса) применяли подход для расчета перепредставленности биологических про-

цессов (Subramanian et al., 2005). Согласно этому подходу, вероятность (p -value) того, что данная вершина будет иметь в ассоциативной геновой сети N_i или более связей по случайным причинам, оценивалась с использованием гипергеометрического распределения с поправкой на множественное сравнение Беньямини-Йекутили (Benjamini, Yekutieli, 2001).

Для характеристики выявленных процессов был проведен кластерный анализ методом MCL (Markov Cluster algorithm) (van Dongen, Abreu-Goodger, 2012), с применением меры семантической близости биологических процессов Ванга (Wang J.Z. et al., 2007).

Результаты и обсуждение

С использованием базы знаний SOLANUM TUBEROSUM была проведена реконструкция ассоциативных геновых сетей *A. thaliana*, описывающих взаимодействие генов с биологическими процессами, связанными с ответом растений на неблагоприятные факторы окружающей среды, включая засуху, солевой стресс и повышенное содержание кадмия. Для исследования потенциальных молекулярно-генетических механизмов, лежащих в основе функционирования реконструированных геновых сетей, выполнена приоритизация биологических процессов по показателю центральности их взаимодействий с генами/белками сети.

Ответ растений на условия засухи

Реконструированная в базе знаний SOLANUM TUBEROSUM ассоциативная геновая сеть биологического процесса “drought tolerance” (устойчивость к засухе) включала 292 вершины (белка), которые были связаны между собой посредством 440 ребер (рис. 2). Для упрощения схемы взаимодействий гены, соответствующие белкам, не отражены. Относительно большое число белков, которые оказались связанными с термином “drought tolerance”, может объясняться тем, что устойчивость растений к за-

сухе обусловлена различными аспектами физиологии растений. Так, поисковый запрос в PubMed по ключевым словам “drought” (засуха) и “plants” (растения) выдает более восемнадцати тысяч публикаций.

Следует заметить, что анализ, основанный на автоматической обработке данных научных публикаций, несомненно, связан с проблемой изученности предметной области. Вследствие этого наши выводы могут констатировать только наличие какой-либо связи согласно данным литературы, но не могут утверждать отсутствие связи на основании того, что связь не обсуждается в текстах.

Среди биологических процессов, связанных с белками в ассоциативной геновой сети “drought tolerance”, оказалось 208 процессов с уровнем значимости Q -value < 0.05 . Их показатели центральности в ассоциативной сети находятся в интервале от 0.0067 до 0.58. Выявленные биологические процессы были разбиты на 22 кластера по семантической близости (рис. 3).

Список из 12 биологических процессов с наибольшим количеством связей с другими объектами геновой сети “drought tolerance” приведен в табл. 2.

В результате проведенной приоритизации можно выделить такие фундаментальные процессы, как “transcription” (транскрипция), “signaling” (передача сигнала), “gene expression” (экспрессия генов), попавшие в кластеры “post-translational modifications”, “cellular processes” и “cellular metabolism” соответственно. Процессы, связанные с транскрипцией, трансляцией, экспрессией генов, широко обсуждаются в контексте транскрипционных факторов и их участия в ответе растений на неблагоприятные условия окружающей среды (Leng, Zhao, 2020).

Важное значение имеют также процессы передачи сигналов, индуцируемые внешними и эндогенными факторами. В частности, с высоким рейтингом оказался процесс “ABA signaling” (сигнальный путь абсцизовой кислоты) (см. табл. 2). Показатель CTC для этого процесса

Таблица 2. Ранжирование биологических процессов по их потенциальной связи с процессом “drought tolerance” согласно анализу ассоциативной сети

№ п/п	Биологический процесс	Название кластера	Число связей*	CTC	p -value	Q -value**
1	Transcription	post-translational modifications	173	0.59	7E-64	3E-61
2	Signaling	cellular processes	121	0.41	1E-81	7E-79
3	Gene expression	cellular metabolism	121	0.41	2E-73	1E-70
4	Signaling pathways	signaling pathways	107	0.37	1E-68	5E-66
5	Transcriptional control	transcriptional control	106	0.36	1E-27	1E-25
6	ABA signaling	signaling pathways	98	0.34	1E-86	1E-83
7	Flowering	regulation of leaf senescence	84	0.29	8E-50	2E-47
8	Phosphorylation	cellular metabolism	74	0.25	4E-20	4E-18
9	Degradation	cellular metabolism	69	0.24	5E-33	7E-31
10	Cellular processes	cellular processes	59	0.20	2E-61	5E-59
11	Translation	post-translational modifications	59	0.20	6E-24	6E-22
12	Response to stress	response to stimulus	58	0.20	5E-33	7E-31

Примечание. Здесь и в табл. 3 и 4: * число связей биологического процесса с белками/генами в ассоциативной сети; ** значимость с поправкой на множественное сравнение Беньямини-Йекутили.

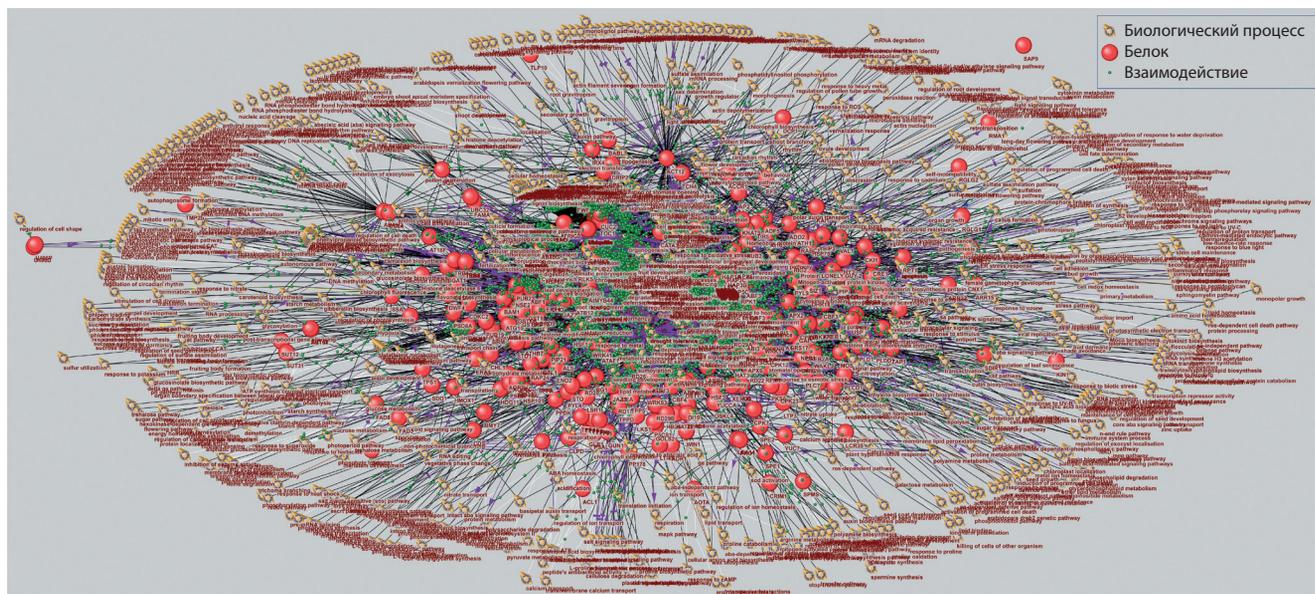


Рис. 2. Ассоциативная генная сеть “drought tolerance” для *Arabidopsis thaliana*, включающая в качестве вершин белки и биологические процессы.

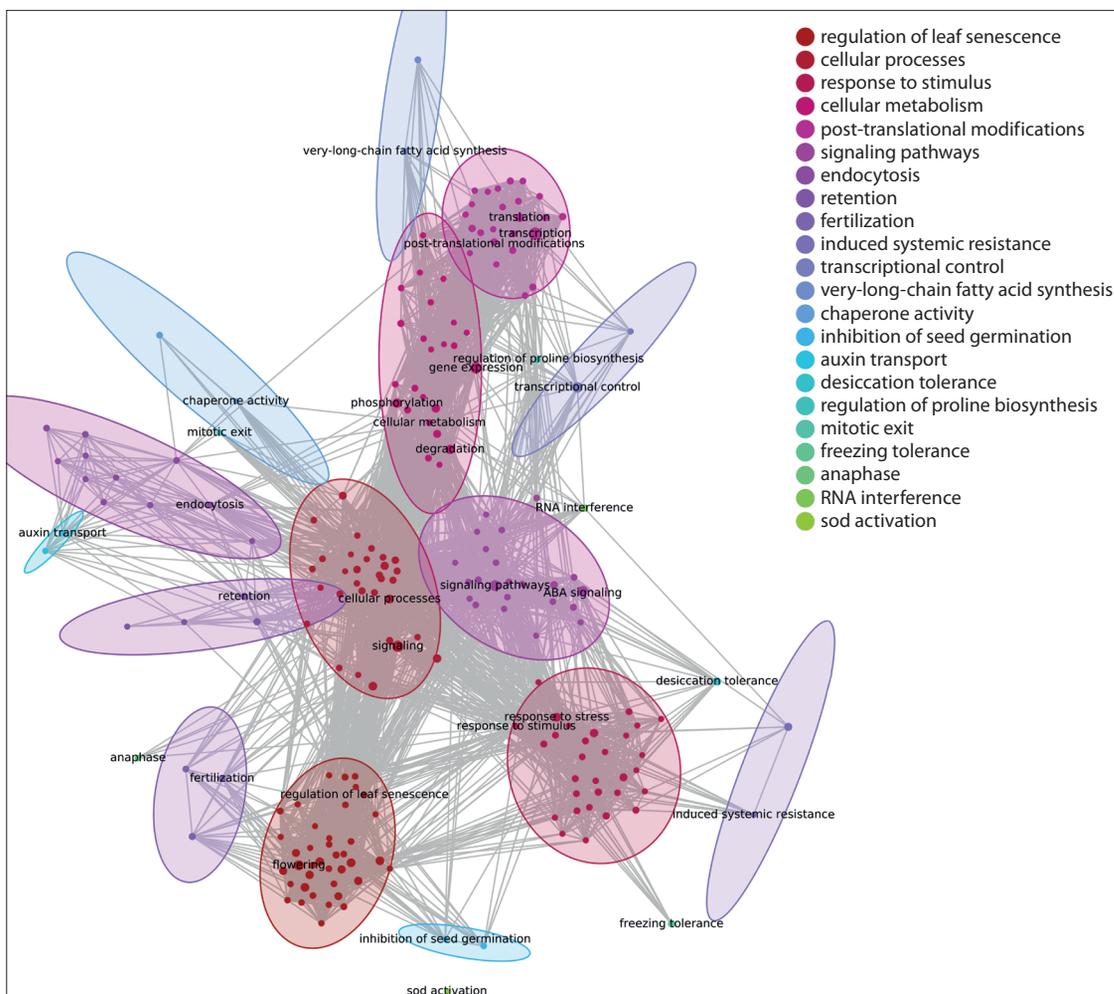


Рис. 3. Кластеризация по семантической близости биологических процессов, значимо перепредставленных в ассоциативной генной сети “drought tolerance”.

Овалами обведены вершины из одного кластера. Цвет вершин и овалов соответствует цвету кластера в легенде.

равен 0.34. Это означает, что при расширении анализируемой ассоциативной геновой сети “drought tolerance” путем добавления вершины, соответствующей биологическому процессу “ABA signaling”, с данной вершиной окажутся связанными в сети 34 % всех вершин. Довольно много исследований показывают высокую значимость АВА в реакциях растений на засуху (около трех тысяч публикаций в PubMed). АВА синтезируется в листьях в ответ на сигнал о дефиците воды, передающийся от корней к побегам, и принимает участие в механизмах засухоустойчивости, таких как закрытие устьиц, наработка белков осмотической защиты и др. (Takahashi et al., 2018). АВА участвует также в регуляции сроков цветения (“flowering” находится в списке приоритетных процессов, см. табл. 2). Например, в условиях засухи ABF3 (ABA-responsive element binding factor 3) и ABF4 с помощью NF- γ C (Nuclear transcription factor γ subunit gamma) усиливают экспрессию *SOCI* (suppressor of overexpression of constans 1), что способствует ускорению цветения растений, позволяя закончить жизненный цикл в более ранние сроки (Hwang et al., 2019).

Еще один процесс, связанный с ответом растений на засуху и расположенный в верхних строках результатов приоритизации (см. табл. 2), – это фосфорилирование, один из основных способов передачи регуляторных сигналов в клетке. Так, например, на *A. thaliana* показано, что для активации транскрипции генов, чувствительных к засухе, необходимо фосфорилирование транскрипционного фактора RD26 киназой BIN2 (brassinosteroid insensitive 2) (Jiang et al., 2019). В исследовании засухоустойчивого сорта рапса (*Brassica napus* L.) выявлена важная роль фосфорилирования β CA1 (beta carbonic anhydrase 1) в регуляции фотосинтеза в условиях засухи (Wang L. et al., 2016).

При анализе ассоциативной геновой сети ответа растений на условия засухи отдельного обсуждения заслуживает попадание в список приоритизируемых биологических процессов таких, как устойчивость растений к холодному и солевому стрессу. Хотя данные процессы и не вошли

в топ-список процессов, имеющих наиболее высокий приоритет по отношению к ассоциативной геновой сети «устойчивость к засухе» (см. табл. 2), они имели достаточно высокий рейтинг (СТС = 0.1). Действительно, устойчивость растений к разным неблагоприятным факторам окружающей среды часто обусловлена одними и теми же молекулярно-генетическими механизмами. Так, на кукурузе (*Zea mays*) было показано, что экспрессия транскрипционного фактора (ТФ) MYB (MYB3R) индуцируется и при засухе, и при солевом стрессе, способствуя устойчивости растений к этим факторам окружающей среды (Wu et al., 2019). На сое (*Glycine max*) было показано, что другой ТФ семейства MYB – MYB118, тоже играет важное значение в устойчивости растений к засухе (Du et al., 2018). Его экспрессия так же, как и MYB3R, индуцируется при засухе и солевом стрессе, а вновь синтезированный ТФ MYB118 усиливает экспрессию стресс-ассоциированных генов для ответа растений на эти стрессы. На хлопке (*Gossypium* spp.) продемонстрирована важная роль белков семейства циклин-зависимых киназ (CDK) в ответе растений на засуху и солевой стресс (Magwanga et al., 2018). Механизмы ответа растений на засуху пересекаются с механизмами ответа на холодный стресс. Например, в обоих случаях в сосудистой ткани листьев индуцируется синтез и мобилизация абсцизовой кислоты, необходимой для регуляции закрытия устьиц (Agurla et al., 2018).

Ответ растений на засоление

Ассоциативная геновая сеть “response to salt stress” (ответ на солевой стресс) содержала 81 вершину и 102 связи (рис. 4). Выявлено 12 наиболее приоритетных процессов, показатель центральности которых был не ниже 0.2 (табл. 3). Как и в случае с анализом ассоциативной геновой сети устойчивости к засухе, все процессы, попавшие в список наиболее приоритетных, оказались статистически значимыми.

Таблица 3. Ранжирование биологических процессов по их потенциальной связи с процессом “response to salt stress” согласно анализу ассоциативной сети

№ п/п	Биологический процесс	Название кластера	Число связей*	СТС	p-value	Q-value**
1	Transcription	RNA synthesis	30	0.34	1.8E-5	8.0E-4
2	Signaling	cellular processes	28	0.32	4.5E-16	1.3E-13
3	Salt tolerance	response to stress	28	0.32	6.5E-32	3.1E-29
4	Gene expression	cellular lipid metabolism	24	0.27	6.4E-11	5.9E-9
5	Phosphorylation	cellular processes	22	0.25	5.9E-7	3.1E-5
6	Signaling pathways	signaling pathways	20	0.23	1.7E-9	1.2E-7
7	Response to salt	response to stress	18	0.20	5.3E-25	1.9E-22
8	Flowering	respiration	17	0.19	3.6E-8	2.2E-6
9	Seed germination	respiration	16	0.18	4.2E-12	5.2E-10
10	Drought tolerance	response to stress	15	0.17	1.2E-9	9.2E-8
11	Response to stress	response to stress	15	0.17	1.6E-8	5.2E-10
12	Degradation	cellular processes	14	0.16	1.4E-5	6.0E-4

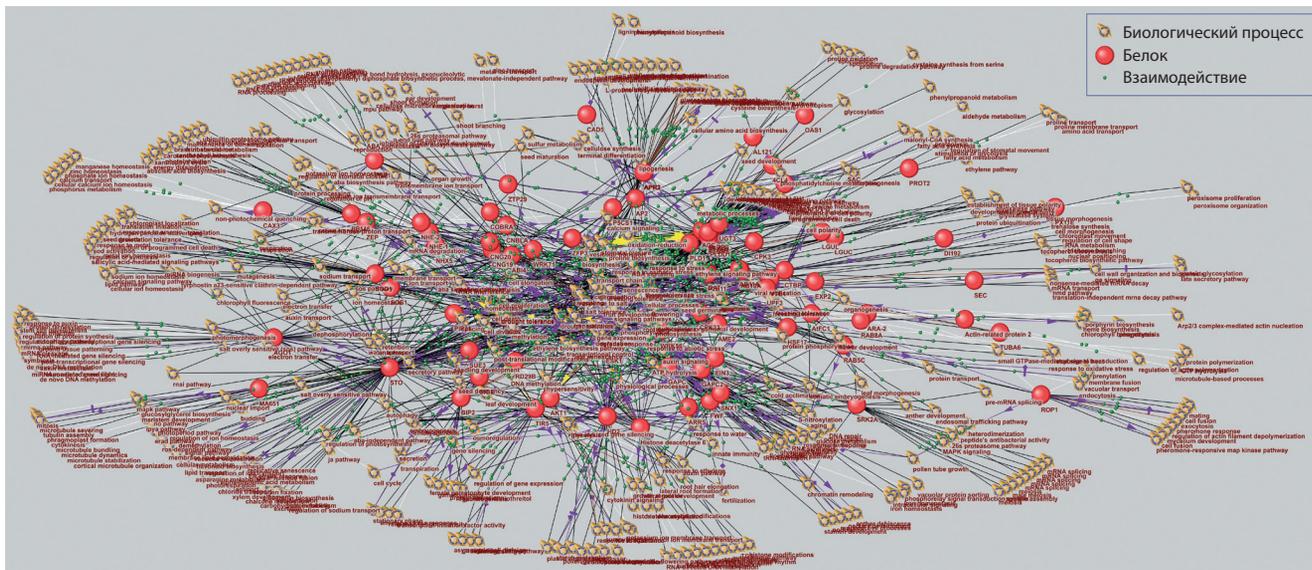


Рис. 4. Ассоциативная генная сеть “response to salt stress” для *Arabidopsis thaliana*, включающая в качестве вершин белки и биологические процессы.

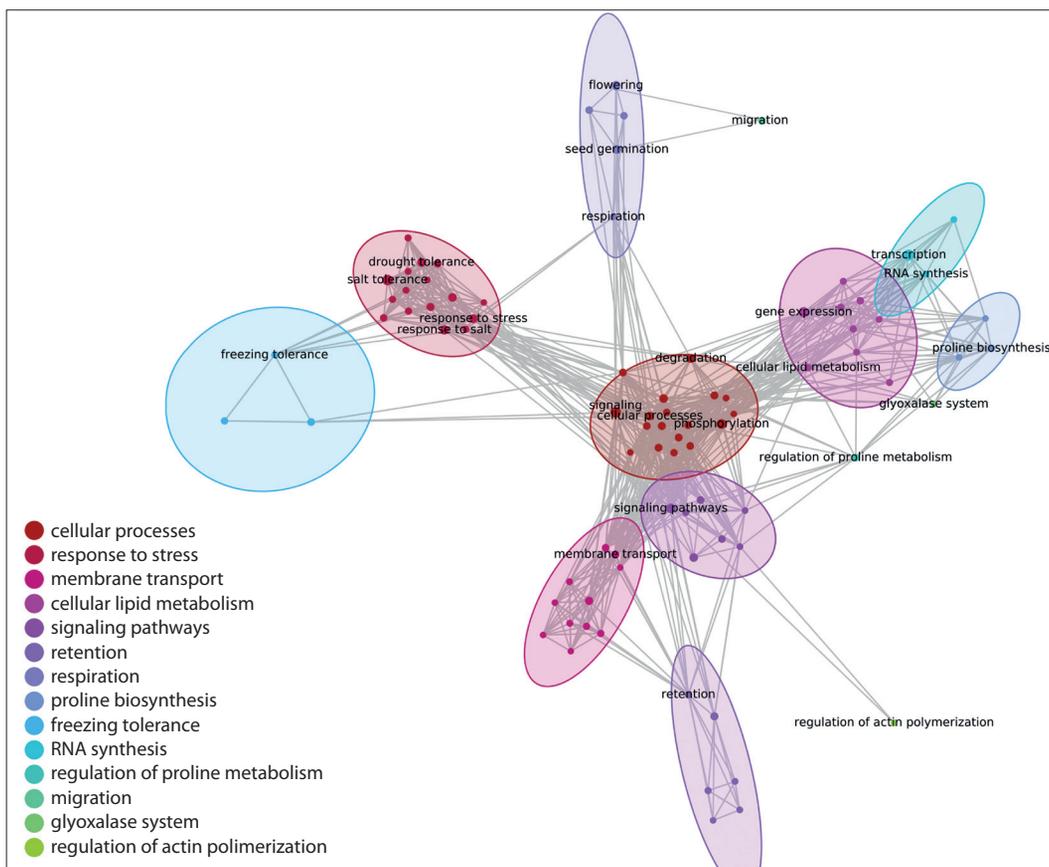


Рис. 5. Кластеризация по семантической близости биологических процессов, значимо перепредставленных в ассоциативной генной сети “response to salt stress”.

Овалами обведены вершины из одного кластера. Цвет вершин и овалов соответствует цвету кластера в легенде.

Среди биологических процессов, связанных с белками в ассоциативной генной сети “response to salt stress”, оказалось 85 с уровнем значимости Q -value < 0.05. Их показатели центральности в ассоциативной сети лежали в интервале от 0.02 до 0.34. Выявленные биологические

процессы были разбиты на 14 кластеров по семантической близости (рис. 5).

Список из 12 биологических процессов с наибольшим количеством связей с другими объектами генной сети “response to salt stress” приведен в табл. 3.

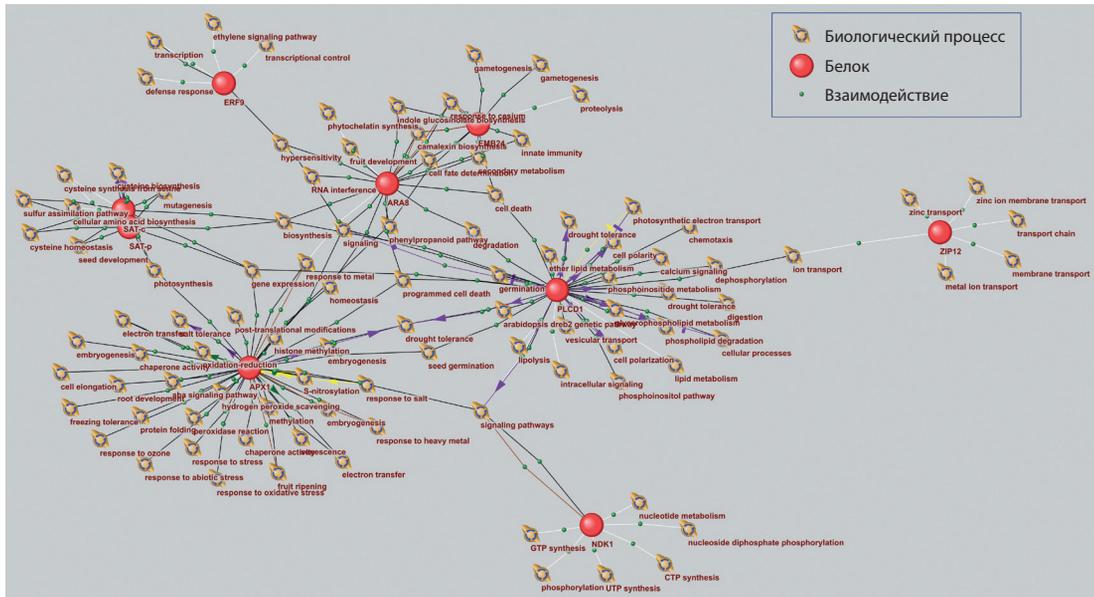


Рис. 6. Ассоциативная генная сеть “response to cadmium ion” для *Arabidopsis thaliana*, включающая в качестве вершин белки и биологические процессы.

Проблематика устойчивости к солевому стрессу очень широко представлена в научной литературе. Поисковый запрос в PubMed по ключевым словам “salt” и “plants” выдает более двадцати четырех тысяч публикаций. В результате проведенного анализа в списке высокорейтинговых процессов среди прочих оказался, например, “seed germination” (прорастание семян), попавший в кластер “respiration” (см. табл. 3). Так, на примере *Stylosanthes humilis* авторы показали, что повышенное содержание солей подавляет прорастание семян, при этом снижается продукция этилена, а наработка абсцизовой кислоты увеличивается (Silva et al., 2018). На арабидопсисе (*Arabidopsis thaliana*) показано, что старение листьев в ходе нормального развития растения и вызванное солевым стрессом имеют общие пути передачи сигнала, опосредованные пероксидом водорода (Allu et al., 2014).

Ответ растений на присутствие кадмия

Кадмий является токсичным веществом для большинства растений и животных (Genchi et al., 2020). Он может содержаться в почве, вызывая стрессовые реакции у растений, подавляя рост корней и побегов, снижая скорость фотосинтеза и потребления питательных веществ (Genchi et al., 2020; Kaуa et al., 2020). Повышенный интерес к проблеме обусловлен тем, что некоторые растения склонны к гипераккумуляции кадмия. Это открывает перспективы их использования как для очистки почв, так и в других промышленных целях (Küpper, Leitenmaier, 2013). Однако механизмы, лежащие в основе ответа растений на кадмий, остаются до сих пор плохо изученными. В базе знаний SOLANUM TUBEROSUM для *A. thaliana* с биологическим процессом “response to cadmium ion” (ответ на кадмий) оказались связаны 9 генов (рис. 6).

Среди биологических процессов, связанных с белками в ассоциативной генной сети “response to cadmium ion”, оказалось 28 с уровнем значимости Q -value < 0.05. Цент-

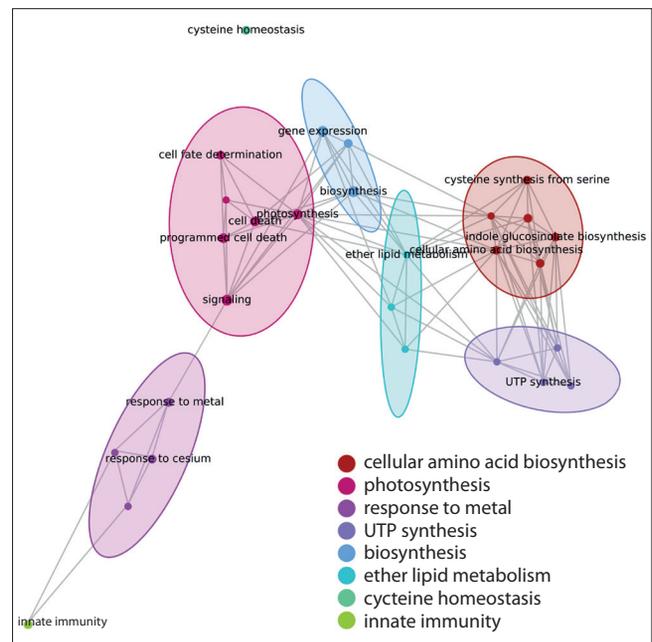


Рис. 7. Кластеризация по семантической близости биологических процессов, значимо перепредставленных в ассоциативной генной сети “response to cadmium ion”.

Овалами обведены вершины из одного кластера. Цвет вершин и овалов соответствует цвету кластера в легенде.

ральность этих биологических процессов в ассоциативной сети лежала в интервале от 0.11 до 0.44. Выявленные биологические процессы были разбиты на 8 кластеров по семантической близости (рис. 7).

Список из 12 биологических процессов с наибольшим количеством связей с другими объектами генной сети “response to cadmium ion” приведен в табл. 4. Среди биологических процессов, имеющих высокий рейтинг в генной

Таблица 4. Ранжирование биологических процессов по их потенциальной связи с процессом “response to cadmium ion” согласно анализу ассоциативной сети

№ п/п	Биологический процесс	Имя кластера	Число связей*	CTC	p-value	Q-value**
1	Signaling	photosynthesis	4	0.44	5.0E-4	1.0E-2
2	Biosynthesis	biosynthesis	4	0.44	3.1E-4	8.4E-3
3	Gene expression	biosynthesis	4	0.44	1.0E-3	1.7E-2
4	Programmed cell death	photosynthesis	3	0.33	7.8E-5	2.6E-3
5	Cell death	photosynthesis	3	0.33	5.1E-4	1.3E-2
6	Photosynthesis	photosynthesis	3	0.33	7.5E-4	1.3E-2
7	Indole glucosinolate biosynthesis	cellular amino acid biosynthesis	2	0.22	1.9E-5	9.7E-4
8	Cysteine synthesis from serine	cellular amino acid biosynthesis	2	0.22	3.8E-5	1.4E-3
9	Cysteine biosynthesis	cellular amino acid biosynthesis	2	0.22	1.6E-4	4.6E-3
10	Camalexin biosynthesis	cellular amino acid biosynthesis	2	0.22	2.9E-3	4.0E-2
11	Cellular amino acid biosynthesis	cellular amino acid biosynthesis	2	0.22	3.7E-3	4.4E-2
12	Cell fate determination	photosynthesis	2	0.22	2.8E-5	1.1E-3

сети ответа на кадмий, оказались два процесса, связанных с клеточной гибелью: “programmed cell death” (программированная клеточная гибель) и “cell death”, попавшие в кластер “photosynthesis”. Важность процессов программируемой клеточной гибели (“cell death”) в ответ на кадмий подтверждается десятками публикаций. Например, у *Genipa americana* L. кадмий вызывает морфофизиологические изменения и запрограммированную гибель клеток (Souza et al., 2011), а на культуре клеток томатов показано, что кадмий вызывает запрограммированную гибель клеток с помощью каспазоподобных протеаз (Iakimova et al., 2008).

В исследованиях молекулярных механизмов ответа растений на неблагоприятные факторы большой интерес может представлять выявление биологических процессов, которые являются как общими, так и специфическими для них. Ниже мы рассмотрим примеры общих и специфических процессов.

Сравнительный анализ реконструированных ассоциативных генных сетей

При сравнении результатов анализа ассоциативных сетей “response to salt stress” и “drought tolerance” было выявлено: 41 общий биологический процесс, 40 специфических для сети “response to salt stress” и 156 – для сети “drought tolerance”. При рассмотрении всех трех ассоциативных генных сетей общими оказались пять биологических процессов: “signaling”, “cell death”, “programmed cell death”, “gene expression” и “photosynthesis”. Интересно отметить, что в таблицы биологических процессов с наибольшим количеством связей с другими объектами (см. табл. 2–4) преимущественно попали общие процессы, имеющие статистически значимую связь с двумя или всеми тремя рассмотренными генными сетями ответа растений на неблагоприятные факторы.

При выявлении биологических процессов, специфичных для ответа растений на засуху по сравнению с генными сетями ответов на засоление и кадмий, можно выделить группу процессов, связанных с развитием семян: “seeding development”, “seed development”, “seed dormancy”, “seed maturation”, “regulation of seed size”, “inhibition of seed germination”. И действительно, опубликовано много статей, демонстрирующих влияние засухи на эти процессы. Так, было показано, что недостаток воды снижает скорость и длительность созревания семян чечевицы, приводя к уменьшению их размера (Sehgal et al., 2019). Аналогичные исследования на линиях сои продемонстрировали, что условия засухи влияют на качество подготовки семян и последующую всхожесть (Wijewardana et al., 2019).

При определении биологических процессов, специфичных для ответа растений на засоление, можно отметить процессы, связанные с ионным транспортом: “ion transport”, “membrane transport”, “sodium transport”, “cation transport”, “transmembrane proton transport”, “sodium ion transmembrane transport”, “potassium ion homeostasis”. Действительно, адаптация растений к условиям избытка соли зависит от способности выводить ионы Na^+ и Cl^- либо от повышения устойчивости к осмотическому стрессу и накоплению ионов в тканях (Munns, Tester, 2008).

Среди обнаруженных биологических процессов, специфичных для ответа растений на присутствие кадмия по сравнению с генными сетями ответов на засуху и засоление, отметим несколько процессов, связанных с метаболизмом липидов: “ether lipid metabolism”, “phospholipid degradation”, “glycerophospholipid metabolism”. Известно, что перекисидация липидов – одно из проявлений токсичности действия кадмия. На томате было показано, что кадмий провоцирует сильные изменения в составе липидов, вызывая преждевременное старение листьев (Djebali et al., 2005).

Заключение

Молекулярно-генетические механизмы устойчивости картофеля к неблагоприятным условиям еще слабо изучены. Использование накопленных знаний о модельном растении *A. thaliana* может пролить свет на картину молекулярных взаимодействий в генных сетях ответа картофеля на стрессовые условия. Такой подход используется и в работах других исследователей. Так, например, Ramšak с коллегами (Ramšak et al., 2018) сначала построили модель иммунных сигнальных путей в *A. thaliana*, а затем наложили на эту модель данные по *S. tuberosum* с целью получения новых знаний об иммунных сигнальных путях картофеля.

База данных Gene Ontology (Gene Ontology Consortium, 2019) содержит информацию о биологических процессах и генах, вовлеченных в их функционирование. С использованием терминов Gene Ontology могут быть описаны молекулярно-генетические механизмы формирования селекционно-важных признаков растений, знания о которых необходимы для развития современных подходов маркер-ориентированной и геномной селекции. Однако большой интерес также представляет анализ генных сетей, построенных на основе анализа литературных источников, содержащих данные о связи генов с изучаемыми процессами.

Анализ результатов приоритизации биологических процессов, которые потенциально могут быть вовлечены в ответ растений на рассматриваемые неблагоприятные факторы (ответ растений на кадмий, засоленность и условия засухи), выполненный с помощью реконструкции ассоциативных генных сетей, показал хорошее соответствие известным литературным данным. В числе биологических процессов, получивших высокий приоритет, оказались фундаментальные процессы, относящиеся к экспрессии генов, посттрансляционной модификации белков и деградации, а также процессы, связанные с клеточной смертью. Важное место в механизмах ответа растений на неблагоприятные факторы занимают различные пути передачи сигналов.

При сравнении результатов приоритизации биологических процессов, помимо поиска общих процессов, ассоциированных с каждым из трех рассмотренных неблагоприятных факторов внешней среды, интерес представляет выявление специфических процессов, статистически значимо связанных только с одной из исследуемых ассоциативных сетей. Среди биологических процессов, специфически вовлеченных в ответ на засуху, помимо прочих, выделяется довольно большая группа, связанная с развитием семян (“seeding development”, “seed dormancy” и др.). В ответе на солевой стресс среди специфических процессов можно выделить группу процессов, ассоциированных с ионным транспортом (например, “ion transport”, “transmembrane proton transport”, “sodium ion transmembrane transport” и др.). А в ответе на кадмий специфичными оказались процессы липидного метаболизма (например, “phospholipid degradation” и др.).

Список литературы / References

Деменков П.С., Сайк О.В., Иванисенко Т.В., Колчанов Н.А., Кочетов А.В., Иванисенко В.А. Приоритизация генов картофеля, вовлеченных в формирование селекционно-значимых признаков, с использованием базы знаний SOLANUM TUBEROSUM.

Вавиловский журнал генетики и селекции. 2019;23(3):312-319. DOI 10.18699/VJ19.501.

[Demenkov P.S., Saik O.V., Ivanisenko T.V., Kolchanov N.A., Kochetov A.V., Ivanisenko V.A. Prioritization of potato genes involved in the formation of agronomically valuable traits using the SOLANUM TUBEROSUM knowledge base. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2019;23(3):312-319. DOI 10.18699/VJ19.501.]

Иванисенко Т.В., Сайк О.В., Деменков П.С., Хлесткин В.К., Хлесткина Е.К., Колчанов Н.А., Иванисенко В.А. База знаний SOLANUM TUBEROSUM: раздел по молекулярно-генетической регуляции метаболических путей. *Вавиловский журнал генетики и селекции*. 2018;22(1):8-17. DOI 10.18699/VJ18.325.

[Ivanisenko T.V., Saik O.V., Demenkov P.S., Khlestkin V.K., Khlestkina E.K., Kolchanov N.A., Ivanisenko V.A. The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2018;22(1):8-17. DOI 10.18699/VJ18.325. (in Russian)]

Сайк О.В., Деменков П.С., Иванисенко Т.В., Колчанов Н.А., Иванисенко В.А. Разработка методов автоматического извлечения знаний из текстов научных публикаций для создания базы знаний SOLANUM TUBEROSUM. *С.-х. биология*. 2017;52(1):63-74. DOI 10.15389/agrobiology.2017.1.63rus.

[Saik O.V., Demenkov P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. Development of methods for automatic extraction of knowledge from texts of scientific publications for the creation of a knowledge base SOLANUM TUBEROSUM. *Selskokhozyaystvennaya Biologiya = Agricultural Biology*. 2017;52(1):63-74. DOI 10.15389/agrobiology.2017.1.63eng.]

Agurla S., Gahir S., Munemasa S., Murata Y., Raghavendra A.S. Mechanism of stomatal closure in plants exposed to drought and cold stress. In: Iwaya-Inoue M., Sakurai M., Uemura M. (Eds.). *Survival Strategies in Extreme Cold and Desiccation*. Advances in Experimental Medicine and Biology. Vol. 1081. Singapore: Springer, 2018;215-232. DOI 10.1007/978-981-13-1244-1_12.

Allu A.D., Soja A.M., Wu A., Szymanski J., Balazadeh S. Salt stress and senescence: identification of cross-talk regulatory components. *J. Exp. Bot.* 2014;65(14):3993-4008. DOI 10.1093/jxb/eru173.

Arruda M.P., Lipka A.E., Brown P.J., Krill A.M., Thurber C., Brown-Guedira G., Dong Y., Foresman B.J., Kolb F.L. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 2016;36(7):84. DOI 10.1007/s11032-016-0508-5.

Bargsten J.W., Nap J.P., Sanchez-Perez G.F., van Dijk A.D. Prioritization of candidate genes in QTL regions based on associations between traits and biological processes. *BMC Plant Biol.* 2014;14:330. DOI 10.1186/s12870-014-0330-3.

Benjamini Y., Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 2001;29(4):1165-1188. DOI 10.1214/aos/1013699998.

Cai Z., Guldbandsen B., Lund M.S., Sahana G. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics*. 2019;20(1):255. DOI 10.1186/s12864-019-5638-9.

Cesur M.F., Siraj B., Uddin R., Durmuş S., Çakır T. Network-based metabolism-centered screening of potential drug targets in *Klebsiella pneumoniae* at genome scale. *Front. Cell. Infect. Microbiol.* 2020;9:447. DOI 10.3389/fcimb.2019.00447.

Chen Y., Jiang T., Jiang R. Uncover disease genes by maximizing information flow in the phenotype-interactome network. *Bioinformatics*. 2011;27(13):i167-i176. DOI 10.1093/bioinformatics/btr213.

Cho A., Shim J.E., Kim E., Supek F., Lehner B., Lee I. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biol.* 2016;17(1):1-16. DOI 10.1186/s13059-016-0989-x.

Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., de Los Campos G., Burgueño J., González-Camacho J.M., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X.,

- Gowda M., Roorkiwal M., Rutkoski J., Varshney R.K. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 2017;22(11):961-975. DOI 10.1016/j.tplants.2017.08.011.
- Demenev P.S., Ivanisenko T.V., Kolchanov N.A., Ivanisenko V.A. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol.* 2012;11(3):149-161. DOI 10.3233/ISB-2012-0449.
- Djebali W., Zarrrouk M., Brouquisse R., El Kahoui S., Limam F., Ghorbel M.H., Chaïbi W. Ultrastructure and lipid alterations induced by cadmium in tomato (*Lycopersicon esculentum*) chloroplast membranes. *Plant Biol. (Stuttg.)*. 2005;7(4):358-368. DOI 10.1055/s-2005-837696.
- Du Y.T., Zhao M.J., Wang C.T., Gao Y., Wang Y.X., Liu Y.W., Chen M., Chen J., Zhou Y.B., Xu Z.S., Ma Y.Z. Identification and characterization of *GmMYB118* responsive to drought and salt stress. *BMC Plant Biol.* 2018;18(1):320. DOI 10.1186/s12870-018-1551-7.
- Freeman L.C. A set of measures of centrality based on betweenness. *Sociometry*. 1977;40:35-41. DOI 10.2307/3033543.
- Freeman L.C. Centrality in social networks conceptual clarification. *Social Networks*. 1978;1(3):215-239. DOI 10.1016/0378-8733(78)90021-7.
- Genchi G., Sinicropi M.S., Lauria G., Carocci A., Catalano A. The effects of cadmium toxicity. *Int. J. Environ. Res. Public Health*. 2020; 17(11):3782. DOI 10.3390/ijerph17113782.
- Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019;47(D1):D330-D338. DOI 10.1093/nar/gky1055.
- Hwang K., Susila H., Nasim Z., Jung J.Y., Ahn J.H. *Arabidopsis* ABF3 and ABF4 transcription factors act with the NF-YC complex to regulate *SOC1* expression and mediate drought-accelerated flowering. *Mol. Plant*. 2019;12(4):489-505. DOI 10.1016/j.molp.2019.01.002.
- Iakimova E.T., Woltering E.J., Kapchina-Toteva V.M., Harren F.J., Cristescu S.M. Cadmium toxicity in cultured tomato cells – role of ethylene, proteases and oxidative stress in cell death signaling. *Cell Biol. Int.* 2008;32(12):1521-1529. DOI 10.1016/j.cellbi.2008.08.021.
- Ivanisenko T.V., Saik O.V., Demenev P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics*. 2020;21(Suppl.11):228. DOI 10.1186/s12859-020-03557-8.
- Ivanisenko V.A., Demenev P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019; 20(Suppl.1):34. DOI 10.1186/s12859-018-2567-6.
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenev P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst. Biol.* 2015;9(Suppl.2):S2. DOI 10.1186/1752-0509-9-S2-S2.
- Jeong H., Mason S.P., Barabási A.L., Oltvai Z.N. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41-42. DOI 10.1038/35075138.
- Jha M., Roy S., Kalita J.K. Prioritizing disease biomarkers using functional module based network analysis: a multilayer consensus driven scheme. *Comput. Biol. Med.* 2020;126:104023. DOI 10.1016/j.compbiomed.2020.104023.
- Jia P.L., Zheng S.Y., Long J.R., Zheng W., Zhao Z.M. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*. 2011;27:95-102. DOI 10.1093/bioinformatics/btq615.
- Jiang H., Tang B., Xie Z., Nolan T., Ye H., Song G.Y., Walley J., Yin Y. GSK3-like kinase BIN2 phosphorylates RD26 to potentiate drought signaling in *Arabidopsis*. *Plant J.* 2019;100(5):923-937. DOI 10.1111/tpj.14484.
- Kaya C., Ashraf M., Alyemeni M.N., Ahmad P. Responses of nitric oxide and hydrogen sulfide in regulating oxidative defence system in wheat plants grown under cadmium stress. *Physiol. Plant*. 2020;168(2):345-360. DOI 10.1111/pp.13012.
- Kochetov A.V., Glagoleva A.Y., Strygina K.V., Khlestkina E.K., Gerasimova S.V., Ibragimova S.M., Shatskaya N.V., Vasilyev G.V., Afonnikov D.A., Shmakov N.A., Antonova O.Y., Gavrilenko T.A., Alpatyeva N.V., Khiutti A., Afanasenko O.S. Differential expression of NBS-LRR-encoding genes in the root transcriptomes of two *Solanum phureja* genotypes with contrasting resistance to *Globodera rostochiensis*. *BMC Plant Biol.* 2017;17(Suppl.2):251. DOI 10.1186/s12870-017-1193-1.
- Kolchanov N.A., Kochetov A.V., Salina E.A., Pershina L.A., Khlestkina E.K., Shumny V.K. Status and prospects of marker-assisted and genomic plant breeding. *Herald of the Russian Academy of Sciences*. 2017;87(2):125-131. DOI 10.1134/S1019331617020113.
- Küpper H., Leitenmaier B. Cadmium-accumulating plants. In: Sigel A., Sigel H., Sigel R. (Eds.). Cadmium: From Toxicity to Essentiality. Metal Ions in Life Sciences. Vol. 11. Dordrecht: Springer, 2013;373-393. DOI 10.1007/978-94-007-5179-8_12.
- Le D.H., Pham V.H. HGPEC: a Cytoscape app for prediction of novel disease-gene and disease-disease associations and evidence collection based on a random walk on heterogeneous network. *BMC Syst. Biol.* 2017;11(1):61. DOI 10.1186/s12918-017-0437-x.
- Leng P., Zhao J. Transcription factors as molecular switches to regulate drought adaptation in maize. *Theor. Appl. Genet.* 2020;133(5):1455-1465. DOI 10.1007/s00122-019-03494-y.
- Leung A., Bader G.D., Reimand J. Hypermodules: identifying clinically and phenotypically significant network modules with disease mutations for biomarker discovery. *Bioinformatics*. 2014;30:2230-2232. DOI 10.1093/bioinformatics/btu172.
- Lin F., Fan J., Rhee S.Y. QTG-Finder: a machine-learning based algorithm to prioritize causal genes of quantitative trait loci in *Arabidopsis* and rice. *G3: Genes, Genomes, Genetics. (Bethesda)*. 2019; 9(10):3129-3138. DOI 10.1534/g3.119.400319.
- Lysenko A., Borojevich K.A., Tsunoda T. Arete – candidate gene prioritization using biological network topology with additional evidence types. *BioData Min.* 2017;10:22. DOI 10.1186/s13040-017-0141-9.
- Magwanga R.O., Lu P., Kirungu J.N., Cai X., Zhou Z., Wang X., Diouf L., Xu Y., Hou Y., Hu Y., Dong Q., Wang K., Liu F. Whole genome analysis of cyclin dependent kinase (*CDKF4*) gene family in cotton and functional evaluation of the role of *CDKF4* gene in drought and salt stress tolerance in plants. *Int. J. Mol. Sci.* 2018;19(9):2625. DOI 10.3390/ijms19092625.
- Munns R., Tester M. Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 2008;59:651-681. DOI 10.1146/annurev.arplant.59.032607.092911.
- Pushpakom S., Iorio F., Eyers P.A., Escott K.J., Hopper S., Wells A., Doig A., Williams T., Latimer J., McNamee C., Norris A., Sanseau P., Cavalla D., Pirmohamed M. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* 2019; 18(1):41-58. DOI 10.1038/nrd.2018.168.
- Raj M.R., Sreeraj A. Analysis of computational gene prioritization approaches. *Procedia Comput. Sci.* 2018;143:395-410. DOI 10.1016/j.procs.2018.10.411.
- Ramšak Ž., Coll A., Stare T., Tzfadia O., Baebler Š., Van de Peer Y., Gruden K. Network modeling unravels mechanisms of crosstalk between ethylene and salicylate signaling in potato. *Plant Physiol.* 2018;178(1):488-499. DOI 10.1104/pp.18.00450.
- Sabidussi G. The centrality index of a graph. *Psychometrika*. 1966;31: 581-603. DOI 10.1007/BF02289527.
- Saik O.V., Demenev P.S., Ivanisenko T.V., Bragina E.Y., Freidin M.B., Goncharova I.A., Dosenko V.E., Zolotareva O.I., Hofstaedt R., Lavrik I.N., Rogaev E.I., Ivanisenko V.A. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med. Genomics*. 2018;11(Suppl.1): 15. DOI 10.1186/s12920-018-0331-4.
- Saik O.V., Nimaev V.V., Usmonov D.B., Demenev P.S., Ivanisenko T.V., Lavrik I.N., Ivanisenko V.A. Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema

- using an analysis of associative gene networks with ANDSysTem. *BMC Med. Genomics*. 2019;12(Suppl.2):47. DOI 10.1186/s12920-019-0492-9.
- Schaefer R.J., Michno J.M., Jeffers J., Hoekenga O., Dilkes B., Baxter I., Myers C.L. Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell*. 2018;30(12):2922-2942. DOI 10.1105/tpc.18.00299.
- Sehgal A., Sita K., Bhandari K., Kumar S., Kumar J., Vara Prasad P.V., Siddique K.H.M., Nayyar H. Influence of drought and heat stress, applied independently or in combination during seed development, on qualitative and quantitative aspects of seeds of lentil (*Lens culinaris* Medikus) genotypes, differing in drought sensitivity. *Plant Cell Environ*. 2019;42(1):198-211. DOI 10.1111/pce.13328.
- Shim J.E., Lee I. Network-assisted approaches for human disease research. *Animal Cells Syst*. 2015;19:231-235.
- Shim J.E., Lee T., Lee I. From sequencing data to gene functions: co-functional network approaches. *Animal Cells Syst*. 2017;21(2):77-83. DOI 10.1080/19768354.2017.1284156.
- Silva N.C.Q., de Souza G.A., Pimenta T.M., Brito F.A.L., Picoli E.A.T., Zsögön A., Ribeiro D.M. Salt stress inhibits germination of *Stylosanthes humilis* seeds through abscisic acid accumulation and associated changes in ethylene production. *Plant Physiol. Biochem*. 2018;130:399-407. DOI 10.1016/j.plaphy.2018.07.025.
- Souza V.L., de Almeida A.A., Lima S.G., de M. Cascardo J.C., da C. Silva D., Mangabeira P.A., Gomes F.P. Morphophysiological responses and programmed cell death induced by cadmium in *Genipa americana* L. (Rubiaceae). *Biometals*. 2011;24(1):59-71. DOI 10.1007/s10534-010-9374-5.
- Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., Paulovich A., Pomeroy S.L., Golub T.R., Lander E.S., Mesirov J.P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*. 2005;102(43):15545-15550. DOI 10.1073/pnas.0506580102.
- Sun C., Dong Z., Zhao L., Ren Y., Zhang N., Chen F. The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnol. J*. 2020;18(6):1354-1360. DOI 10.1111/pbi.13361.
- Takahashi F., Kuromori T., Sato H., Shinozaki K. Regulatory gene networks in drought stress responses and resistance in plants. *Adv. Exp. Med. Biol*. 2018;1081:189-214. DOI 10.1007/978-981-13-1244-1_11.
- Tranchevent L.C., Ardeshirdavani A., ElShal S., Alcaide D., Aerts J., Auboeuf D., Moreau Y. Candidate gene prioritization with Endeavour. *Nucleic Acids Res*. 2016;44(W1):W117-W121. DOI 10.1093/nar/gkw365.
- van Dongen S., Abreu-Goodger C. Using MCL to extract clusters from networks. *Methods Mol. Biol*. 2012;804:281-295. DOI 10.1007/978-1-61779-361-5_15.
- Voss-Fels K.P., Cooper M., Hayes B.J. Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet*. 2019;132(3):669-686. DOI 10.1007/s00122-018-3270-8.
- Wang J.Z., Du Z., Payattakool R., Yu P.S., Chen C.F. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007;23(10):1274-1281. DOI 10.1093/bioinformatics/btm087.
- Wang L., Jin X., Li Q., Wang X., Li Z., Wu X. Comparative proteomics reveals that phosphorylation of β carbonic anhydrase 1 might be important for adaptation to drought stress in *Brassica napus*. *Sci. Rep*. 2016;6:39024. DOI 10.1038/srep39024.
- Wijewardana C., Reddy K.R., Krutz L.J., Gao W., Bellaloui N. Drought stress has transgenerational effects on soybean seed germination and seedling vigor. *PLoS One*. 2019;14(9):e0214977. DOI 10.1371/journal.pone.0214977.
- Wu J., Jiang Y., Liang Y., Chen L., Chen W., Cheng B. Expression of the maize MYB transcription factor ZmMYB3R enhances drought and salt stress tolerance in transgenic plants. *Plant Physiol. Biochem*. 2019;137:179-188. DOI 10.1016/j.plaphy.2019.02.010.

ORCID ID

P.S. Demenkov orcid.org/0000-0001-9433-8341

Благодарности. Реконструкция и анализ ассоциативных генных сетей выполнены в рамках проекта ИЦиГ СО РАН «Создание новых конкурентоспособных сортов картофеля с использованием методов маркер-ориентированной и геномной селекции» (№ АААА-А20-120070990014-7) в рамках КПНИ «Развитие селекции и семеноводства картофеля». Разработка методов приоритизации осуществлялась за счет бюджетных проектов № 0259-2021-0009 и 0259-2021-0012.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 14.12.2020. После доработки 21.06.2021. Принята к публикации 21.06.2021.