

Приоритизация генов картофеля, вовлеченных в формирование селекционно-значимых признаков, с использованием базы знаний SOLANUM TUBEROSUM

П.С. Деменков^{1,2}✉, О.В. Сайк¹, Т.В. Иванисенко¹, Н.А. Колчанов¹, А.В. Кочетов¹, В.А. Иванисенко¹

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

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ e-mail: demps@bionet.nsc.ru

Развитие высокоэффективных технологий в геномике, транскриптомике, протеомике и метаболомике, а также новых технологий в сельском хозяйстве привело к «информационному взрыву» в биологии растений и растениеводстве, включая производство картофеля. Лишь небольшая часть информации попадает в формализованные базы данных (например, Uniprot, NCBI Gene, BioGRID, IntAct и др.). Один из основных источников достоверных биологических данных – научная литература. Объем литературы велик, так, известная база данных PubMed содержит более 18 тыс. тезисов статей, посвященных картофелю. Эффективное получение знаний, представленных в таком количестве неформализованных текстовых документов, требует применения современных интеллектуальных методов анализа. Однако в литературе нет свидетельств широкого применения интеллектуальных методов автоматического извлечения знаний из научных публикаций по таким культурам, как картофель. В рамках настоящей работы использовалась разработанная нами ранее база знаний SOLANUM TUBEROSUM (<http://www-bionet.sysbio.cytogen.ru/and/plant/>). Интегрированная в базу знаний информация о молекулярно-генетических механизмах, лежащих в основе селекционно-значимых признаков, способствует ускорению идентификации генов-кандидатов для селекционно-значимых характеристик картофеля и разработки диагностических маркеров для селекции. В статье выполнен поиск новых потенциальных участников молекулярно-генетических механизмов устойчивости к неблагоприятным факторам у растений. Приоритизация генов-кандидатов показала, что гены *PHYA*, *GF14*, *CNIH1*, *RC11A*, *ABIS*, *CPK1*, *RGS1*, *NHL3*, *GRF8* и *CYP21-4* наиболее перспективны для дальнейшей проверки их связей с устойчивостью к неблагоприятным факторам. В результате проведенного анализа выявлено, что молекулярно-генетические взаимоотношения, ответственные за формирование значимых сельскохозяйственных признаков, являются комплексными и включают множество как прямых, так и опосредованных взаимодействий. Построение ассоциативных генных сетей и их анализ с использованием базы знаний SOLANUM TUBEROSUM – это основа поиска генов-мишеней для направленного мутагенеза и маркер-ориентированной селекции сортов картофеля, обладающих ценными сельскохозяйственными признаками. Ключевые слова: картофель; *Solanum tuberosum*; ANDSystem; анализ текстов; база данных; методы автоматического извлечения знаний из текстов; приоритизация генов; ассоциативные генные сети.

Для цитирования: 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

Prioritization of potato genes involved in the formation of agronomically valuable traits using the SOLANUM TUBEROSUM knowledge base

P.S. Demenkov^{1,2}✉, O.V. Saik¹, T.V. Ivanisenko¹, N.A. Kolchanov¹, A.V. Kochetov¹, V.A. Ivanisenko¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

✉ e-mail: demps@bionet.nsc.ru

The development of highly efficient technologies in genomics, transcriptomics, proteomics and metabolomics, as well as new technologies in agriculture has led to an “information explosion” in plant biology and crop production, including potato production. Only a small part of the information reaches formalized databases (for example, Uniprot, NCBI Gene, BioGRID, IntAct, etc.). One of the main sources of reliable biological data is the scientific literature. The well-known PubMed database contains more than 18 thousand abstracts of articles on potato. The effective use of knowledge presented in such a number of non-formalized documents in natural language requires the use of modern intellectual methods of analysis. However, in the literature, there is no evidence of a widespread use of intelligent methods for automatically extracting knowledge from scientific publications on cultures such as potatoes. Earlier we developed

the SOLANUM TUBEROSUM knowledge base (<http://www-bionet.sysbio.cytogen.ru/and/plant/>). Integrated into the knowledge base information about the molecular genetic mechanisms underlying the selection of significant traits helps to accelerate the identification of candidate genes for the breeding characteristics of potatoes and the development of diagnostic markers for breeding. The article searches for new potential participants of the molecular genetic mechanisms of resistance to adverse factors in plants. Prioritizing candidate genes has shown that the *PHYA*, *GF14*, *CNIH1*, *RC11A*, *ABI5*, *CPK1*, *RGS1*, *NHL3*, *GRF8*, and *CYP21-4* genes are the most promising for further testing of their relationships with resistance to adverse factors. As a result of the analysis, it was shown that the molecular genetic relationships responsible for the formation of significant agricultural traits are complex and include many direct and indirect interactions. The construction of associative gene networks and their analysis using the SOLANUM TUBEROSUM knowledge base is the basis for searching for target genes for targeted mutagenesis and marker-oriented selection of potato varieties with valuable agricultural characteristics.

Key words: potato; *Solanum tuberosum*; ANDSystem; text mining; knowledge base; automatic extraction of knowledge from texts; prioritization of genes; associative gene networks.

For citation: 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 (in Russian)

Введение

Картофель (*Solanum tuberosum* L.) имеет высокую пищевую, техническую, кормовую ценность и является одной из важнейших сельскохозяйственных культур. Такая пищевая ценность картофеля достигается благодаря большому содержанию углеводов, аскорбиновой кислоты, солей калия, кальция, магния и других микроэлементов, а также хорошей усвояемостью белков. Картофельный крахмал служит сырьем для производства спирта, патоки, декстринов, глюкозы, мальтозы и ряда других продуктов для химической промышленности (Khlestkin et al., 2018). Крахмал клубней картофеля также широко используется в бумажной, текстильной и других отраслях промышленности (Kraak, 1992; Ellis et al., 1998; Jobling, 2004).

Развитие высокопроизводительных технологий в области геномики, транскриптомики, протеомики и метаболомики, а также новые технологии в сельском хозяйстве привели к «информационному взрыву» в биологии растений и растениеводстве. Лишь небольшая часть информации попадает в формализованные фактографические базы данных (например, NCBI Gene, Uniprot, IntAct, BioGRID и др.). Один из основных источников надежных биологических знаний – научная литература. Хорошо известная база данных PubMed содержит более 18 тыс. рефератов статей, посвященных картофелю, что делает чрезвычайно трудным для исследователей вручную анализировать такие объемы информации.

Отсутствие единого ресурса, интегрирующего всю доступную информацию, приводит к сильному усложнению работ, связанных с выявлением связей между наборами данных, описывающих важные и практически полезные свойства растений, их строение и процессы на молекулярном уровне (Хлесткин и др., 2017). Таким образом, снижается эффективность использования полученных результатов, не полностью используется генетический потенциал природных ресурсов.

Проблема обработки больших и сверхбольших данных возникает все чаще в различных сферах деятельности человека (Kilicoglu, 2017). В настоящее время большое распространение получили методы автоматического интеллектуального анализа текстов (text mining). Обычно эти методы разделяют на две группы: методы, основанные на правилах, и методы, использующие машинное

обучение. Методы, основанные на правилах, позволяют достигать высокой точности извлечения информации, однако полнота извлечения имеет относительно низкий уровень (Aggarwal, 2012). Другой подход к автоматическому извлечению информации – методы машинного обучения. Эти методы не требуют использования вручную созданных правил и получили широкое распространение в последнее время. К недостаткам таких методов можно отнести требование наличия больших обучающих выборок, содержащих размеченные вручную тексты.

Методы автоматического анализа текстов нашли широкое применение при решении различных задач биомедицины, системной и интегративной биологии (Friedman et al., 1999; Meystre et al., 2008; Cao et al., 2011; Shetty, Dalal, 2011; Rebholz-Schuhmann et al., 2012; Li et al., 2013; Wei et al., 2013; Sarker et al., 2015).

В литературе встречается описание использования методов автоматического интеллектуального анализа текстов только для модельных растений. Например, система PLAN2L (Krallinger et al., 2009) содержит результаты автоматического извлечения информации из полнотекстовых публикаций по *Arabidopsis thaliana* о белок-белковых взаимодействиях и генетической регуляции, а также об ассоциациях генов с некоторыми клеточными процессами и процессами развития (цветка, корня и т. д.). E. da Costa с коллегами (2018) создали интерактивную систему, которая позволяет выявлять вредителей и болезни риса на основе информации, полученной от фермеров в виде коротких текстовых сообщений (SMS).

Ранее нами была разработана компьютерная платформа для комплексного интеллектуального анализа научных публикаций в области картофелеводства – база знаний SOLANUM TUBEROSUM, доступная по адресу <http://www-bionet.sysbio.cytogen.ru/and/plant/> (Сайк и др., 2017; Иванисенко и др., 2018). Программные средства этой платформы обеспечивают решение всех необходимых шагов для автоматического извлечения и формализованного представления в базе знаний информации по генетике, ДНК-маркерам, селекции, семеноводству, диагностике заболеваний, средствам защиты и технологиям хранения картофеля. Для доступа к этим данным, построения запросов и визуализации результатов служит графический пользовательский интерфейс SOLANUM TUBEROSUM.

Автоматический анализ текстов проводился адаптированными методами системы ANDSystem (Demenkov et al., 2012; Ivanisenko et al., 2015; Saik et al., 2016).

Интеграция сведений о молекулярно-генетических механизмах, лежащих в основе селекционно-значимых признаков, способствует ускорению идентификации генов-кандидатов для селекционно-значимых характеристик картофеля и разработки диагностических маркеров для селекции.

В настоящее время в биоинформатике для выявления генов-кандидатов, потенциально вовлеченных в какой-либо признак и/или биологический процесс, широко используются методы приоритизации (Chen et al., 2009). Анализ генных сетей – один из таких подходов. Ранее нами были разработаны критерии приоритизации генов, основанные на анализе структуры ассоциативных генных сетей ANDSystem (Янкина и др., 2018; Saik et al., 2018). В настоящей работе была проведена приоритизация генов, направленная на выявление перспективных кандидатов для исследования их связи с устойчивостью к неблагоприятным факторам.

Материалы и методы

База знаний SOLANUM TUBEROSUM доступна в интернете по адресу <http://www-bionet.sssc.ru/and/plant/>. В ее состав входят три основных модуля.

Модуль автоматического анализа текстов (text mining) используется для извлечения из научных публикаций информации о взаимоотношениях между объектами. Модуль основан на программных инструментах ANDSystem (Ivanisenko et al., 2015). Программные инструменты ANDSystem проводят многоэтапный анализ текстов, состоящий из предобработки текстов, извлечения информации о взаимоотношениях между объектами с помощью семантико-лингвистических шаблонов и сохранения результатов в формализованном виде. Существующая на сегодняшний день версия ANDSystem работает только с текстами на английском языке. Кроме инструментов анализа текстов, ANDSystem содержит средства для сбора и интеграции информации из фактографических баз данных.

Модуль базы данных SOLANUM TUBEROSUM состоит из разделов Dictionary (словари объектов и понятий) и Associative networks (информация о взаимоотношениях между объектами и понятиями).

Раздел Dictionary включает:

- молекулярно-генетические данные по картофелю и модельным растениям (гены, белки, метаболиты, микроРНК, биологические процессы);
- генетические биомаркеры;
- сорта картофеля;
- селекционно-значимые качества, хозяйственно ценные признаки и потребительские свойства картофеля и модельных растений;
- физиологические, фенотипические признаки и болезни картофеля;
- молекулярно-генетические данные по возбудителям заболеваний и вредителям картофеля (гены, белки, метаболиты, биологические процессы);
- генетические маркеры резистентности к средствам защиты растений;

- молекулярные мишени для химических средств защиты растений;
- биотические факторы окружающей среды;
- абиотические факторы окружающей среды (почва, влажность, температура, свет, воздух, климат и микроклимат и т. д.);
- методы и технологии:
 - селекция;
 - диагностика заболеваний;
 - защита от заболеваний;
 - возделывание, переработка и хранение картофеля.

Раздел Associative networks содержит:

- физические взаимодействия (молекулярные комплексы белок/белок, белок/лиганд, белок/ДНК);
- химические взаимодействия (каталитические реакции и процессы) по типу субстрат–фермент–продукт;
- регуляторные взаимодействия и ассоциации (регуляция экспрессии гена, регуляция активности белка, ассоциация ген/признак и т. д.);
- взаимоотношения между понятиями селекции, фенотики и семеноводства, заболеваниями, приемами диагностики и средствами защиты.

Модуль визуализации и биоинформатики используется для интерактивного построения ассоциативных генных сетей и их анализа с применением методов биоинформатики.

Реконструкция ассоциативных генных сетей выполнялась с помощью программы ANDVisio (Demenkov et al., 2012) с использованием информации из базы знаний SOLANUM TUBEROSUM (Сайк и др., 2017; Иванисенко и др., 2018).

Приоритизация генов проводилась на основании показателя центральности CTC (cross-talk centrality), рассчитываемого с использованием функции Intelligent Filtration программы ANDVisio по формуле:

$$CTC_j = N_j/M,$$

где N_j – число связей j -го гена/белка с участниками ассоциативной генной сети; M – число вершин ассоциативной генной сети (Янкина и др., 2018). При ранжировании генов-кандидатов сортировка проводилась по убыванию значения показателя CTC. Таким образом, гены с наибольшим показателем CTC получают максимальный приоритет.

Результаты

На основании информации из базы знаний SOLANUM TUBEROSUM были проведены реконструкция и анализ ассоциативных генных сетей, описывающих биологические процессы, вовлеченные в формирование селекционно-значимых сельскохозяйственных признаков, таких как устойчивость к неблагоприятным факторам окружающей среды, ответ на различные стрессы (избыток солей, холод, засуха, повышенная температура). Ассоциативная генная сеть устойчивости к неблагоприятным факторам показана на рисунке. Сеть включает 542 гена, 544 белка, 34 биологических процесса и 2406 взаимодействий между ними.

В табл. 1 представлен список биологических процессов, которые отвечают за устойчивость к неблагоприятным факторам у картофеля. Из данных, представленных в

табл. 1, видно, что наибольшее число генов и белков ассоциировано с процессом «ответ на окислительный стресс» (Gene Ontology идентификатор – GO:0006979). У растений окислительный стресс наблюдается при действии большинства неблагоприятных факторов среды, включая заморозки, засуху, засоление почв, высокие температуры и патогены (Mittler, 2002; Ramirez et al., 2018).

Для картофеля уже выполнен ряд исследований, в которых ведется обсуждение возможности создания линий растений, устойчивых к различным неблагоприятным условиям среды, полученных путем модификации биологических процессов, представленных в табл. 1 (Jones et al., 2014; Kikuchi et al., 2015). Например, показано, что трансгенный картофель, в котором экспрессировались гены Cu- и Zn-супероксиддисмутаза томата, обладал повышенной устойчивостью к окислительному (Perl et al., 1993), а также холодному и солевому (Shafi et al., 2017) стрессам. В работе Р. Моппеуеу с коллегами (2013) обсуждаются связь 14 генов картофеля с устойчивостью к засухе и возможности их использования для создания трансгенных растений. Была изучена связь между генами картофеля *ACS4*, *ACS5* и ответом на биотический стресс (Schlaghnauffer et al., 1997). Создание трансгенных линий картофеля с различным уровнем экспрессии гена *PHYV* откроет возможность для изучения механизмов устойчивости картофеля к воздействию повышенных температур (Трагеро-Мозос et al., 2018).

Приоритизация, выполненная с использованием показателя центральности СТС (cross-talk centrality), позволила выявить гены-кандидаты, наиболее перспективные для дальнейшего изучения их связи с устойчивостью к неблагоприятным факторам окружающей среды, а также ответом на различные стрессы. В табл. 2 приведены 10 наиболее приоритетных генов-кандидатов. Эти гены отранжированы согласно значениям показателя СТС, который выражает степень связности генов в генной сети, показанной на рисунке.

На первом месте в табл. 2 оказался ген *PHYA*, который кодирует фоторецептор фитохром А, участвующий в различных биологических процессах, в том числе в контроле циркадного ритма, цветения и движений листьев в ответ на воздействие света с различными длинами волн

Таблица 1. Топ 10 биологических процессов, обеспечивающих устойчивость к неблагоприятным факторам окружающей среды, упорядоченных согласно числу ассоциированных с ними генов

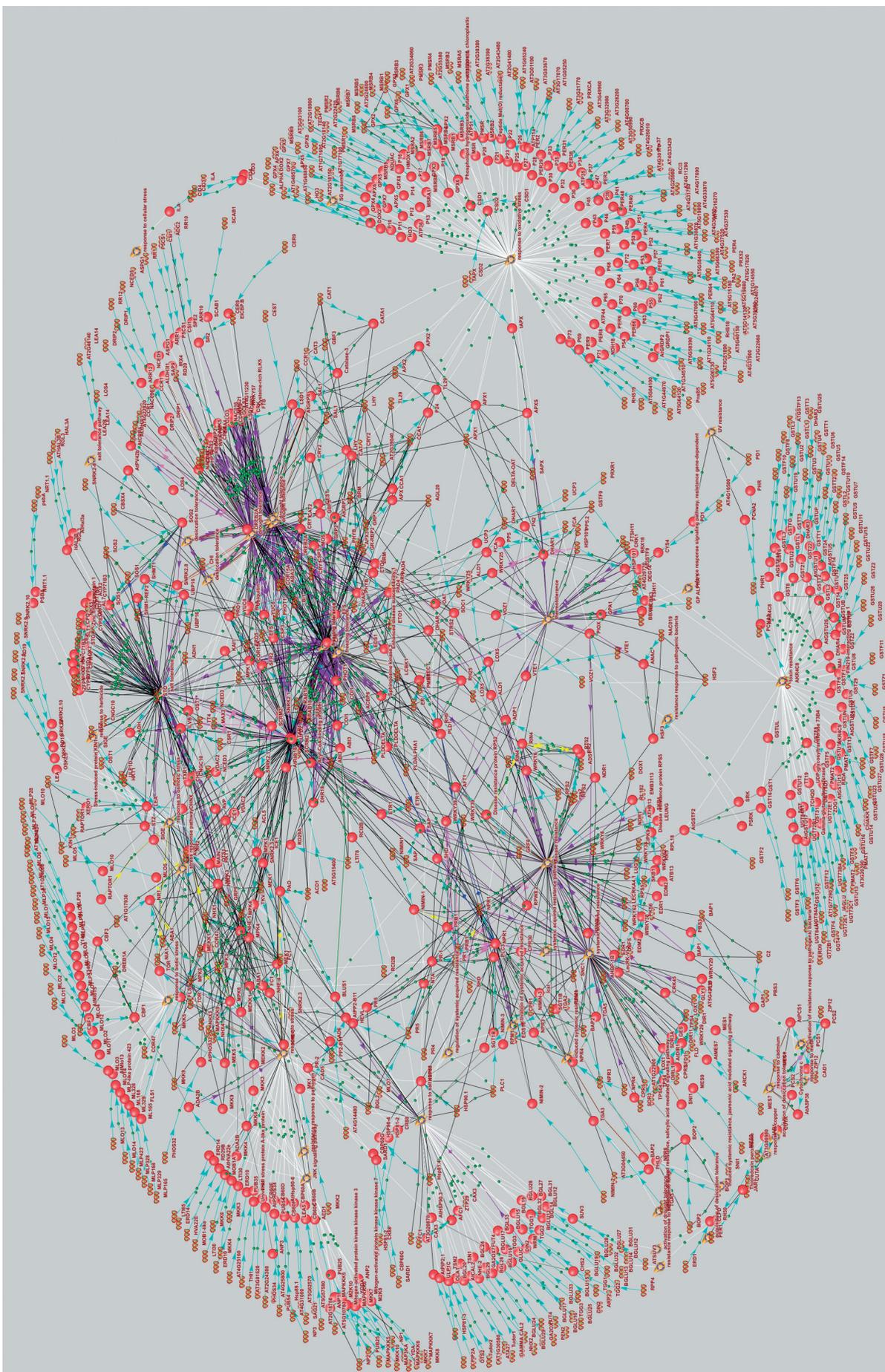
Название биологического процесса	Идентификатор Gene Ontology	Гены/белки
Ответ на окислительный стресс	GO:0006979	118
Ответ на биотический стресс	GO:0009607	71
Устойчивость к токсинам	GO:0009636	68
Устойчивость к засухе	GO:0009633	62
Устойчивость к засолению	GO:0009651	49
Устойчивость к заболеваниям	GO:0009614	39
Устойчивость к холоду	GO:0009409	29
Приобретенная системная устойчивость к стрессам	GO:0009627	26
Ответ на осмотический стресс	GO:0006970	23
Устойчивость к высоким температурам	GO:0010286	14

(Yanovsky et al., 2000). В работе (Sawers et al., 2005) обсуждаются возможности использования фитохромов в программах селекции зерновых культур для разработки сортов, которые обладают устойчивостью к негативным факторам роста в условиях загущенного посева. Другие примеры исследования влияния мутаций фитохромов на фенотипы растений приведены в работах (Zhang et al., 2013; Chen et al., 2014). В исследовании (Zhang et al., 2013) изучалось влияние мутаций *phyB* у *A. thaliana* на ряд фенотипических признаков растений, а в работе (Chen et al., 2014) показано, что потеря функциональной активности РНУС у пшеницы приводила к изменениям циркадного ритма и резкой задержке цветения в условиях длинного светового дня.

Второе, четвертое и девятое места заняли гены из семейства 14-3-3-подобных белков (*GF14*, *RC11A* и *GRF8* соответственно). 14-3-3-подобные белки регулируют клеточный цикл, апоптоз, иммунные процессы, метаболизм азота и углерода, участвуют в регуляции синтеза крах-

Таблица 2. Десять наиболее приоритетных генов-кандидатов, потенциально вовлеченных в устойчивость к неблагоприятным факторам

Ранг	Ген	Название белка	Центральность СТС
1	<i>PHYA</i>	Фитохром А	0.011514615
2	<i>GF14</i>	14-3-3-подобный белок	0.011514615
3	<i>CNIH1</i>	Антипортер CNIH1	0.011514615
4	<i>RC11A</i>	14-3-3-подобный белок GF14 пси	0.010628875
5	<i>ABI5</i>	Гомолог транскрипционного фактора ABI5 bZIP-типа	0.009743136
6	<i>CPK1</i>	Кальций-зависимая протеинкиназа С	0.008857396
7	<i>RGS1</i>	Регулятор сигнального пути G-белка типа 1	0.008857396
8	<i>NHL3</i>	NDR1/HIN1-подобный белок 3	0.008857396
9	<i>GRF8</i>	14-3-3-подобный белок GF14 каппа	0.008857396
10	<i>CYP21-4</i>	Пептидил-пролил-цис-транс изомераза CYP21-4	0.007971656



Ассоциативная генная сеть устойчивости к неблагоприятным факторам у картофеля.
Гены обозначены спиральками; Белки – красными шарами; биологические процессы – коричневыми овалами; взаимодействия между ними – линиями.

мала, производства аденозинтрифосфата, детоксикации пероксидом и в некоторых других биохимических путях. Развитие растений и прорастание семян контролируются факторами, активация которых происходит при взаимодействии с 14-3-3-подобными белками (Fulgosi et al., 2002). Снижение уровня белка GF14 приводило к увеличению содержания кальция, крахмала и отношения растворимых сахаров к крахмалу в клубнях картофеля, а также к значительному росту содержания метионина, пролина и аргинина в клетках картофеля (Świądrych et al., 2002). Было показано, что подавление экспрессии гена *GF14e* методом РНК интерференции приводило к повышению устойчивости риса к вирулентному штамму бактериального фитопатогена *Xanthomonas oryzae* pv. *oryzae* (Хоо) (Manosalva et al., 2011). Известно, что растения с мутациями в генах *RC11A* и *GRF8* проявляли повышенную устойчивость к заморозкам (Catalá et al., 2014; Liu et al., 2017).

Третье место получил ген *CNIH1*, который у растений кодирует белок, взаимодействующий с транспортером натрия НКТ1 и обеспечивающий правильное расположение транспортера на мембране аппарата Гольджи (Rosas-Santiago et al., 2017). В работе (Wudick et al., 2018) изучалось влияние мутаций в гене *CNIH1* на пылецу и гомеостаз кальция у *A. thaliana*. Интересно отметить, что был получен сорт картофеля Юбилей Жукова, который обладал повышенной соле- и засухоустойчивостью, благодаря сверхэкспрессии вакуолярного Na^+/H^+ -антипортера NHX2 (Беляев и др., 2011).

Гомолог транскрипционного фактора ABI5 bZIP-типа (*ABI5*) занял пятое место. ABI5 имеет большое значение для прорастания семян, которое регулируется абсцизовой кислотой (Finkelstein, 1994; Lopez-Molina, et al., 2002). Транскрипционный фактор ABI5 bZIP-типа участвует в активации генов, ответственных за накопление белков при развитии семян. Снижение экспрессии гена *ABI5* активирует рост меристемы (Lopez-Molina et al., 2002). У *A. thaliana* мутации в гене *ABI5* были связаны со сниженной чувствительностью к абсцизовой кислоте, а также к солевому и осмотическому стрессу при прорастании (Finkelstein, Lynch, 2000; Carles et al., 2002; Tezuka et al., 2013).

На шестом месте был ген *CPK1*, который кодирует кальций-зависимую протеинкиназу С, участвующую в иммунном ответе, устойчивости к грибковым заболеваниям и патогенам (Gravino et al., 2015). Известно, что мутации в гене *CPK1* *A. thaliana* приводили к гиперчувствительности к солевому стрессу и засухе, в то время как линии трансгенных растений с повышенной экспрессией *CPK1* обладали значительной устойчивостью к солевому стрессу и засухе (Huang et al., 2018).

Седьмое место занял ген *RGS1*, кодирующий отрицательный регулятор сигнального пути G-белка типа 1. Известно, что экспрессия этого гена снижается в ответ на дефицит воды (Campbell et al., 2012). В работе (Chen et al., 2006) для *A. thaliana* было показано, что трансгенные растения, в которых был сверхэкспрессирован ген *RGS1*, обладали повышенной устойчивостью к засухе.

На восьмом месте оказался ген *NHL3*, кодирующий NDR1/HIN1-подобный белок 3, вовлеченный в ответ на

воздействие патогенов (Chong et al., 2008). Трансгенная линия *A. thaliana*, в которой наблюдалась повышенная экспрессия гена *NHL3*, показала повышенную устойчивость к патогену *Pseudomonas syringae* pv. *tomato* DC3000 (Varet et al., 2003).

Десятое место в табл. 2 занял ген *CYP21-4*, кодирующий циклофилин, локализованный в аппарате Гольджи, который участвует в толерантности к окислительному стрессу (Park et al., 2017). Авторы рассматривают сверхэкспрессию гена *CYP21-4* в сельскохозяйственных культурах как новый перспективный способ увеличения продуктивности и урожайности растений. Для картофеля и риса показано, что трансгенные растения, в которых сверхэкспрессируется ген *CYP21-4*, имели повышенную урожайность, стебли и корни растений были длиннее, а листья толще. У картофеля образовывалось большее число клубней и они были крупнее, а микроклубни формировались быстрее по сравнению с растениями дикого типа (Park et al., 2017).

Заключение

В настоящей работе осуществлен поиск новых потенциальных участников молекулярно-генетических механизмов устойчивости к неблагоприятным факторам у растений. Приоритизация генов-кандидатов показала, что наиболее перспективными для дальнейшей проверки их связей с устойчивостью к неблагоприятным факторам являются гены *PHYA*, *GF14*, *CNIH1*, *RC11A*, *ABI5*, *CPK1*, *RGS1*, *NHL3*, *GRF8* и *CYP21-4*. Проведенный анализ показал, что молекулярно-генетические взаимоотношения, ответственные за формирование значимых сельскохозяйственных признаков, являются комплексными и включают множество как прямых, так и опосредованных взаимодействий. Визуализация этих взаимодействий в виде ассоциативных генных сетей и их анализ с использованием базы знаний SOLANUM TUBEROSUM – основа для поиска генов-мишеней для направленного мутагенеза и маркер-ориентированной селекции сортов картофеля, с устойчивостью к неблагоприятным факторам окружающей среды.

Список литературы / References

- Беляев Д.В., Носова А.Е., Кривошеева А.Б., Терешонок Д.В., Юрьева Н.О., Рудас В.А., Кучук Н.В. Экспрессия вакуолярного антипортера ячменя повышает солеустойчивость и засухоустойчивость трансгенного картофеля. Сборник трудов конференции «Современные аспекты генетической инженерии растений». Киев, 30 мая–1 июня 2011.
- [Belyaev D.V., Nosova A.E., Krivosheeva A.B., Tereshonok D.V., Yur'eva N.O., Rudas V.A., Kuchuk N.V. Expression of the vacuolar antiporter of barley increases the salt tolerance and drought tolerance in transgenic potatoes. Proceedings of the Conference "Modern Aspects of Plant Genetic Engineering". Kiev, May 30–June 1, 2011. (in Russian)]
- 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/VJ18.32.
- [Ivanisenco T.V., Saik O.V., Demenkov P.S., Khlestkin V.K., Khlestkina E.K., Kolchanov N.A., Ivanisenco V.A. The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways. Vavilovskii Zhurnal Genetiki i

- Selektzii = Vavilov Journal of Genetics and Breeding. 2018;22(1): 8-17. (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. (in Russian)]
- Хлесткин В.К., Пельтек С.Е., Колчанов Н.А. Гены-мишени для получения сортов картофеля (*Solanum tuberosum* L.) с заданными свойствами крахмала. С.-х. биология. 2017;52(1):25-36. DOI 10.15389/agrobiology.2017.1.25rus.
- [Khlestkin V.K., Peltek S.E., Kolchanov N.A. Target genes for development of potato (*Solanum tuberosum* L.) cultivars with desired starch properties. Selskokhozyaystvennaya Biologiya = Agricultural Biology. 2017;52(1):25-36. (in Russian)]
- Янкина М.А., Сайк О.В., Иванисенко В.А., Деменков П.С., Хуснутдинова Э.К. Оценка методов приоритизации генов внешнего пути апоптоза в качестве кандидатов, ассоциированных с большим депрессивным расстройством. Генетика. 2018;54(11): 1338-1348.
- [Yankina M.A., Saik O.V., Ivanisenko V.A., Demenkov P.S., Khusnutdinova E.K. Evaluation of prioritization methods of extrinsic apoptotic signaling pathway genes for retrieval of the new candidates associated with major depressive disorder. Genetika = Genetics (Moscow). 2018;54(11):1338-1348. (in Russian)]
- Aggarwal C.C., Zhai C. Mining Text Data. Springer Science & Business Media, 2012.
- Campbell M.A., Gleichsner A., Hilldorfer L., Horvath D., Suttle J. The sprout inhibitor 1, 4-dimethylnaphthalene induces the expression of the cell cycle inhibitors KRP1 and KRP2 in potatoes. *Funct. Integr. Genomics*. 2012;12(3):533-541. DOI 10.1007/s10142-011-0257-9.
- Cao Y., Liu F., Simpson P., Antieau L., Bennett A., Cimino J.J., Ely J., Yu H. AskHERMES: An online question answering system for complex clinical questions. *J. Biomed. Inform.* 2011;44:277-288. DOI 10.1016/j.jbi.2011.01.004.
- Carles C., Bies-Etheve N., Aspart L., Léon-Kloosterziel K.M., Koornneef M., Echeverria M., Delseny M. Regulation of *Arabidopsis thaliana* *Em* genes: role of ABI5. *Plant J.* 2002;30(3):373-383.
- Catalá R., López-Cobollo R., Castellano M.M., Angosto T., Alonso J.M., Ecker J.R., Salinas J. The *Arabidopsis* 14-3-3 protein RARE COLD INDUCIBLE 1A links low-temperature response and ethylene biosynthesis to regulate freezing tolerance and cold acclimation. *Plant Cell*. 2014;26(8):3326-3342. DOI 10.1105/tpc.114.127605.
- Chen A., Li C., Hu W., Lau M.Y., Lin H., Rockwell N.C., Martin S.S., Jernstedt J.A., Lagarias J.C., Dubcovsky J. Phytochrome C plays a major role in the acceleration of wheat flowering under long-day photoperiod. *Proc. Natl. Acad. Sci.* 2014;111(28):10037-10044. DOI 10.1073/pnas.1409795111.
- Chen J., Bardes E.E., Aronow B.J., Jegga A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucl. Acids Res.* 2009;37(Suppl. 2):W305-W311.
- Chen Y., Ji F., Xie H., Liang J. Overexpression of the regulator of G-protein signalling protein enhances ABA-mediated inhibition of root elongation and drought tolerance in *Arabidopsis*. *J. Exp. Bot.* 2006;57(9):2101-2110.
- Chong J., Le Henanff G., Bertsch C., Walter B. Identification, expression analysis and characterization of defense and signaling genes in *Vitis vinifera*. *Plant Physiol. Biochem.* 2008;46(4):469-481. DOI 10.1016/j.plaphy.2007.09.010.
- da Costa E., Tjandrasa H., Djanali S. Text mining for pest and disease identification on rice farming with interactive text messaging. *Int. J. Elec. & Comp. Eng. (IJECE)*. 2018;8(3):1671-1683.
- Demenkov 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 Biology*. 2012;11(3-4):149-161. DOI 10.3233/ISB-2012-0449.
- Ellis R.P., Cochrane M.P., Dale M.F.B., Duffus C.M., Lynn A., Morrison I.M., Prentice R.D.M., Swanston J.S., Tiller S.A. Starch production and industrial use. *J. Sci. Food Agric.* 1998;77(3):289-311. DOI 10.1002/(SICI)1097-0010(199807)77:3<289::AID-JSFA38>3.0.CO;2-D.
- Finkelstein R.R. Mutations at two new *Arabidopsis* ABA response loci are similar to the *abi3* mutations. *Plant J.* 1994;5:765-771.
- Finkelstein R.R., Lynch T.J. The *Arabidopsis* abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell*. 2000;12(4):599-609.
- Friedman C., Hripsak G., Shagina L., Liu H. Representing information in patient reports using natural language processing and the extensible markup language. *J. Am. Med. Inform. Assoc.* 1999;6:76-87. DOI 10.1136/jamia.1999.0060076.
- Fulgosi H., Soll J., de Faria Maraschin S., Korthout H.A., Wang M., Testerink C. 14-3-3 proteins and plant development. *Plant Mol. Biol.* 2002;50(6):1019-1029.
- Gravino M., Savatin D.V., Macone A., De Lorenzo G. Ethylene production in *Botrytis cinerea*- and oligogalacturonide-induced immunity requires calcium-dependent protein kinases. *Plant J.* 2015;84(6): 1073-1086. DOI 10.1111/tpj.13057.
- Huang K., Peng L., Liu Y., Yao R., Liu Z., Li X., Yang Y., Wang J. *Arabidopsis* calcium-dependent protein kinase AtCPK1 plays a positive role in salt/drought-stress response. *Biochem. Biophys. Res. Commun.* 2018;498(1):92-98.
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov 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.
- Jobling S. Improving starch for food and industrial applications. *Curr. Opin. Plant Biol.* 2004;7(2):210-218. DOI 10.1016/j.pbi.2003.12.001.
- Jones J.D., Witek K., Verweij W., Jupe F., Cooke D., Dorling S., Tomlinson L., Smoker M., Perkins S., Foster S. Elevating crop disease resistance with cloned genes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2014;369(1639):20130087. DOI 10.1098/rstb.2013.0087.
- Khlestkin V.K., Peltek S.E., Kolchanov N.A. Review of direct chemical and biochemical transformations of starch. *Carbohydr. Polymers*. 2018;181(1):460-476. DOI 10.1016/j.carbpol.2017.10.035.
- Kikuchi A., Huynh H.D., Endo T., Watanabe K. Review of recent transgenic studies on abiotic stress tolerance and future molecular breeding in potato. *Breed. Sci.* 2015;65(1):85-102.
- Kilicoglu H. Biomedical text mining for research rigor and integrity: tasks, challenges, directions. *Brief. Bioinform.* 2017, Jan 1. DOI 10.1101/108480.
- Kraak A. Industrial applications of potato starch products. *Ind. Crops Prod.* 1992;1(2-4):107-112. DOI 10.1016/0926-6690(92)90007-I.
- Krallinger M., Rodriguez-Penagos C., Tendulkar A., Valencia A. PLAN2L: a web tool for integrated text mining and literature-based bioentity relation extraction. *Nucleic Acids Res.* 2009;11; 37(Suppl. 2):W160-W165. DOI 10.1093/nar/gkp484.
- Li C., Liakata M., Rebholz-Schuhmann D. Biological network extraction from scientific literature: state of the art and challenges. *Brief. Bioinform.* 2013;15(5):856-877. DOI 10.1093/bib/bbt006.
- Liu Z., Jia Y., Ding Y., Shi Y., Li Z., Guo Y., Gong Z., Yang S. Plasma membrane CRPK1-mediated phosphorylation of 14-3-3 proteins induces their nuclear import to fine-tune CBF signaling during cold response. *Mol. Cell*. 2017;66(1):117-128.
- Lopez-Molina L., Mongrand S., McLachlin D.T., Chait B.T., Chua N.H. ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J.* 2002;32:317-328.
- Manosalva P.M., Bruce M., Leach J.E. Rice 14-3-3 protein (GF14e) negatively affects cell death and disease resistance. *Plant J.* 2011; 68(5):777-787.

- Meystre S.M., Savova G.K., Kipper-Schuler K.C., Hurdle J.F. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb. Med. Inform.* 2008;35: 128-144.
- Mittler R. Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 2002;7(9):405-410. DOI 10.1016/S1360-1385(02)02312-9.
- Monneveux P., Ramírez D.A., Pino M.T. Drought tolerance in potato (*S. tuberosum* L.): can we learn from drought tolerance research in cereals? *Plant Sci.* 2013;205:76-86.
- Park H.J., Lee A., Lee S.S., An D.J., Moon K.B., Ahn J.C., Kim H.S., Cho H.S. Overexpression of Golgi protein CYP21-4s improves crop productivity in potato and rice by increasing the abundance of mannosidic glycoproteins. *Front. Plant Sci.* 2017;8:1250.
- Perl A., Perl-Treves R., Galili S., Aviv D., Shalgi E., Malkin S., Galun E. Enhanced oxidative-stress defense in transgenic potato expressing tomato Cu, Zn superoxide dismutases. *Theor. Appl. Genet.* 1993;85(5):568-576.
- Ramirez E., Ducreux L.J., Redfern C., Morris W.L., Wiese C., Morris J.A., Paterson C., Hedley P.E., Hancock R.D., Taylor M. A reversible light- and genotype-dependent acquired thermotolerance response protects the potato plant from damage due to excessive temperature. *Planta.* 2018;247(6):1377-1392.
- Rebholz-Schuhmann D., Oellrich A., Hoehndorf R. Text-mining solutions for biomedical research: enabling integrative biology. *Nat. Rev. Genet.* 2012;13:829-839. DOI 10.1038/nrg3337.
- Rosas-Santiago P., Lagunas-Gomez D., Yáñez-Domínguez C., Vera-Estrella R., Zimmermannová O., Sychrová H., Pantoja O. Plant and yeast cornichon possess a conserved acidic motif required for correct targeting of plasma membrane cargos. *Biochim. Biophys. Acta Mol. Cell Res.* 2017;1864(10):1809-1818. DOI 10.1016/j.bbamcr.2017.07.004.
- Saik O.V., Ivanisenco T.V., Demenkov P.S., Ivanisenco V.A. Interaction of the hepatitis C virus: Literature mining with ANDSystem. *Virus Res.* 2016;218:40-48. DOI 10.1016/j.virusres.2015.12.003.
- Sarker A., Ginn R., Nikfarjam A., O'Connor K., Smith K., Jayaraman S., Upadhaya T., Gonzalez G. Utilizing social media data for pharmacovigilance: A review. *J. Biomed. Inform.* 2015;54:202-212. DOI 10.1016/j.jbi.2015.02.004.
- Sawers R.J., Sheehan M.J., Brutnell T.P. Cereal phytochromes: targets of selection, targets for manipulation? *Trends Plant Sci.* 2005; 10(3):138-143.
- Schlaghhauser C.D., Artega R.N., Pell E.J. Sequential expression of two 1-aminocyclopropane-1-carboxylate synthase genes in response to biotic and abiotic stresses in potato (*Solanum tuberosum* L.) leaves. *Plant Mol. Biol.* 1997;35(6):683-688.
- Shafi A., Pal A.K., Sharma V., Kalia S., Kumar S., Ahuja P.S., Singh A.K. Transgenic potato plants overexpressing SOD and APX exhibit enhanced lignification and starch biosynthesis with improved salt stress tolerance. *Plant Mol. Biol. Rep.* 2017;35(5):504-518.
- Shetty K.D., Dalal S.R. Using information mining of the medical literature to improve drug safety. *J. Am. Med. Inform. Assoc.* 2011;18: 668-674. DOI 10.1136/amiainjnl-2011-000096.
- Świądrych A., Prescha A., Matysiak-Kata I., Biernat J., Szopa J. Repression of the 14-3-3 gene affects the amino acid and mineral composition of potato tubers. *J. Agric. Food Chem.* 2002;50(7):2137-2141. DOI 10.1021/jf0112825.
- Tezuka K., Taji T., Hayashi T., Sakata Y. A novel *abi5* allele reveals the importance of the conserved Ala in the C3 domain for regulation of downstream genes and salt tolerance during germination in *Arabidopsis*. *Plant Signal. Behav.* 2013;8(3):e23455.
- Trapero-Mozos A., Ducreux L.J., Bita C.E., Morris W., Wiese C., Morris J.A., Taylor M. A reversible light- and genotype-dependent acquired thermotolerance response protects the potato plant from damage due to excessive temperature. *Planta.* 2018;247(6):1377-1392. DOI 10.1007/s00425-018-2874-1.
- Varet A., Hause B., Hause G., Scheel D., Lee J. The *Arabidopsis NHL3* gene encodes a plasma membrane protein and its overexpression correlates with increased resistance to *Pseudomonas syringae* pv. *tomato* DC3000. *Plant Physiol.* 2003;132(4):2023-2033.
- Wei C.-H., Kao H.-Y., Lu Z. PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res.* 2013;41:W518-W522. DOI 10.1093/nar/gkt441.
- Wudick M.M., Portes M.T., Michard E., Rosas-Santiago P., Lizzio M.A., Nunes C.O., Campos C., Damineli D.S., Carvalho J.C., Lima P.T., Pantoja O. CORNICHON sorting and regulation of GLR channels underlie pollen tube Ca²⁺ homeostasis. *Science.* 2018;360(6388): 533-536.
- Yanovsky M.J., Izaguirre M., Wagmaister J.A., Gatz C., Jackson S.D., Thomas B., Casal J.J. Phytochrome A resets the circadian clock and delays tuber formation under long days in potato. *Plant J.* 2000; 23(2):223-232. DOI 10.1046/j.1365-3113x.2000.00775.x.
- Zhang J., Stankey R.J., Vierstra R.D. Structure-guided engineering of plant phytochrome B with altered photochemistry and light signaling. *Plant Physiol.* 2013;161(3):1445-1457. DOI 10.1104/pp.112.208892.

ORCID ID

A.V. Kochetov orcid.org/0000-0003-3151-5181

Благодарности. Работа выполнена за счет средств, выделенных на выполнение плана научно-исследовательской работы ИЦиГ СО РАН на 2018 год в рамках комплексного плана научных исследований «Развитие селекции и семеноводства картофеля» (№ 0259-2018-0010-C-01).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 30.11.2018. После доработки 20.12.2018. Принята к публикации 11.01.2019.