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. Saik[1], T.V. Ivanisenko[1], N.A. Kolchanov[1], A.V. Kochetov[1], V.A. Ivanisenko[1]

[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*, *RCI1A*, *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.

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

П.С. Деменков[1, 2] ✉, О.В. Сайк[1], Т.В. Иванисенко[1], Н.А. Колчанов[1], А.В. Кочетов[1], В.А. Иванисенко[1]

[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 публика-

П.С. Деменков, О.В. Сайк, Т.В. Иванисенко
Н.А. Колчанов, А.В. Кочетов, В.А. Иванисенко

Приоритизация генов картофеля
с использованием базы знаний SOLANUM TUBEROSUM

2019
23•3

ций 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 базе 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*, *RCI1A*, *ABI5*, *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.

Introduction

Potatoes (*Solanum tuberosum* L.) have high nutritional, technical, feed value and are one of the most important crops. The high nutritional value of potatoes is achieved due to the high content of carbohydrates, ascorbic acid, salts of potassium, calcium, magnesium and other trace elements, as well as good digestibility of proteins. Potato starch is a raw material used in the production of alcohol, molasses, dextrins, glucose, maltose, as well as many other products for the chemical industry (Khlestkin et al., 2018). Potato tuber starch is widely used in the paper, textile and other industries (Kraak, 1992; Ellis et al., 1998; Jobling, 2004).

The development of high-performance technologies in the field of genomics, transcriptomics, proteomics, and metabolomics, as well as in agriculture led to an "information explosion" in the plant biology. At the same time, only a small amount of information gets into the formalized factographic databases (for example, NCBI Gene, UniProt, IntAct, BioGRID, etc.). One of the main sources of reliable biological data is scientific literature. The well-known PubMed database contains more than 18000 abstracts of articles devoted to potatoes, which makes the manual analysis of such data extremely difficult for researchers.

The lack of unified resources that integrate all available information leads to a strong complication of tasks related to the identification of relationships between the data sets that describing the important and practically useful properties of plants, their structure and the processes on the molecular level (Khlestkin et al., 2017). Thus, the efficiency of the use of obtained results decreases as well.

The problem of the processing of large and extra-large amounts of data is becoming increasingly common in various areas of human activity (Kilicoglu, 2017), making the methods of automated text analysis (text-mining) more and more popular. These methods can be divided into two main groups: methods based on manually created semantic rules and templates, and methods which are using machine learning approaches. Methods based on semantic rules and templates normally have a high level of accuracy, but the completeness of the extracted information often is relatively low (Aggarwal, 2012). Another approach for automatic information retrieval is the use of machine learning techniques. These methods do not require manually created rules and are widely used recently. At the same time, one of the main disadvantages of

such methods is the need for extensive training sets, which are often impossible to obtain without manual analysis.

The text-mining methods found a wide use in the dealing with various problems of biomedicine, systems and integrative biology (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).

At the same time, most of the published scientific literature contains information regarding the application of text-mining approaches only to the model plants. For example, the PLAN2L system (Krallinger et al., 2009) contains results of the automatic extraction of information from protein-protein interactions and genetic regulation from the full-text articles dedicated to *Arabidopsis thaliana*, as well as some data describing associations of genes with some cellular and developmental processes (flower, root, etc.). Da Costa and colleagues (2018) developed an interactive system that allows identification of pests and diseases of rice based on information obtained from farmers in the form of short textual messages (SMS).

Previously, we developed a computer platform for integrated intellectual analysis of scientific publications in the field of potato growing – the SOLANUM TUBEROSUM knowledge base, available at http://www-bionet.sysbio.cytogen.ru/and/plant/ (Saik et al., 2017; Ivanisenko et al., 2018). The software of this platform provides the automatic extraction and formalized representation of information in the base of knowledge, including the genetics data, DNA markers, breeding, seed production, diagnosis of diseases, methods of protection and potato storage technologies. The developed graphical interface to the SOLANUM TUBEROSUM knowledge base provides user access to the data, execution of user-specified queries and visualization of obtained results. Automated analysis of texts was carried out by using the adapted methods of the ANDSystem tool (Demenkov et al., 2012; Ivanisenko et al., 2015; Saik et al., 2016).

The integration of knowledge about the molecular-genetic mechanisms underlying inside the significant for breeding traits can help to accelerate the identification of candidate genes essential for the important breeding characteristics of potatoes, as well as the development of diagnostic markers for breeding.

At present, prioritization methods are widely used in bioinformatics to identify candidate-genes which are potentially involved in the trait and/or biological process

(Chen et al., 2009). Analysis of gene networks is one of such approaches. Previously, we developed criteria for the prioritization of genes, based on the analysis of the structure of the associative gene networks of ANDSystem (Saik et al., 2018; Yankina et al., 2018). In this work, the prioritization of genes was aimed at identifying promising candidates to study their relationship with resistance to adverse factors.

Materials and methods

The SOLANUM TUBEROSUM knowledge base is available at http://www-bionet.sscc.ru/and/plant/. The base consists of three main modules.

The text-mining module is used for the extraction of information about the interactions between objects from the texts of scientific publications. The module is based on the ANDSystem software tool (Ivanisenko et al., 2015). The ANDSystem provides a multi-stage text analysis, consisting of preprocessing of texts, retrieval of information describing the relationships between the objects based on the semantic-linguistic templates, and the presenting of the results in a formalized form. The current version of ANDSystem works only with English texts. In addition to the text analysis tools, ANDSystem also contains tools for the collecting and integrating of information from the external factographic databases.

The module of the SOLANUM TUBEROSUM database consists of the two sections: Dictionary (dictionaries of objects and terms) and Associative networks (information about the relationships between objects and terms).

The Dictionary section includes:
• molecular genetic data for potatoes and model plants (genes, proteins, metabolites, miRNA, biological processes);
• genetic biomarkers;
• potato varieties;
• properties significant for breeding, economically valuable traits and consumer properties of potatoes and model plants;
• physiological, phenotypic traits and diseases of potatoes;
• molecular genetic data on pathogens and potato pests (genes, proteins, metabolites, biological processes);
• genetic markers of resistance to plant protection products;
• molecular targets for plant protection chemicals;
• biotic environmental factors;
• abiotic environmental factors (soil, humidity, temperature, light, air, climate, and microclimate, etc.);
• methods and technologies:
– breeding;
– diagnosis of diseas;
– protection against diseases;
– cultivation, processing, and storage of potatoes.

The Associative networks section contains:
• physical interactions (molecular complexes protein/protein, protein/ligand, protein/DNA);
• chemical interactions (catalytic reactions and processes) such as a substrate-enzyme-product;
• regulatory interactions and associations (regulation of gene expression, regulation of protein activity, gene/traits association, etc.);
• the interactions between the terms of breeding, phenomics and seed production, diseases, diagnostic techniques and methods of protection.

The module of visualization and bioinformatics is used for interactive construction of associative gene networks and their analysis using bioinformatics methods.

Associative gene networks were reconstructed using the ANDVisio program (Demenkov et al., 2012) using information from the SOLANUM TUBEROSUM knowledge base (Saik et al., 2017; Ivanisenko et al., 2018).

Gene prioritization was carried out on the basis of the cross-talk centrality index (CTC), calculated using the ANDVisio program's Intelligent Filtration function using the formula:

$$CTC_j = N_j/M,$$

where N_j is the number of links of the j-th gene/protein with the participants of the associative gene network; M is the number of vertices of the associative gene network (Yankina et al., 2018). When ranking candidate genes, the sorting was performed in descending order of the CTC value. Thus, the genes with the highest CTC score receive the highest priority.

Results

Using the information from the SOLANUM TUBEROSUM knowledge base, we performed the reconstruction and analysis of associative gene networks describing biological processes involved in the formation of selective agricultural traits, such as resistance to adverse environmental factors, response to various stresses (excess salt, cold, drought, high temperature). The reconstructed associative gene network of resistance to adverse factors is provided in Figure. The network includes 542 genes, 544 proteins, 34 biological processes and 2406 interactions between them.

Table 1 contains the list of biological processes that are responsible for resistance to adverse factors in potatoes. From the table, it can be seen that the largest number of genes and proteins is associated with the "response to oxidative stress" Gene Ontology process (Gene Ontology identifier – GO:0006979). In plants, the oxidative stress is observed under the majority of unfavorable environmental factors, including the cold, drought, soil salinization, high temperatures and pathogens (Mittler, 2002; Ramirez et al., 2018).

A number of studies for potatoes discussing the possibilities for creating the plant lines resistant to various adverse environmental conditions, which can be obtained by modifying the biological processes presented in Table 1 have already been conducted (Jones et al., 2014; Kikuchi et al., 2015). For example, it was shown that transgenic potatoes in which the Cu- and Zn-superoxide dismutase genes of tomato were expressed had increased resistance to oxidative (Perl et al., 1993), as well as to cold and salt (Shafi et al., 2017) stresses. P. Monneveux et al. (2013) discussed the relationship of 14 potato genes to drought tolerance and the possibilities of their use for the development of transgenic plants. The relationship between the *ACS4*, *ACS5* potato genes and the response to biotic stress has been studied by C.D. Schlagnhaufer (1997). The creation of transgenic potato lines with different levels of expression of the *PHYB* gene can open up the possibilities to study the mechanisms of resistance of potatoes to the effects of elevated temperatures (Trapero-Mozos et al., 2018).

The prioritization of genes carried out using the CTC (cross-talk centrality) index, allowed to identify candidate genes which are most promising for further study of their relationship

П.С. Деменков, О.В. Сайк, Т.В. Иванисенко
Н.А. Колчанов, А.В. Кочетов, В.А. Иванисенко

Приоритизация генов картофеля
с использованием базы знаний SOLANUM TUBEROSUM

2019
23•3

with resistance to adverse environmental factors, as well as response to various stresses. Table 2 contains the top 10 of such candidate genes, ranked according to the values of the CTC indicator, which reflects the degree of gene connectivity in the gene network presented in Figure.

From Table 2 it can be seen that the first place belongs to the *PHYA* gene that encodes the photoreceptor phytochrome A participating in various biological processes, including the control of the circadian rhythm, flowering and leaf movements in response to exposure to light with different wavelengths (Yanovsky et al., 2000). R.J. Sawers et al. (2005) discussed in their paper the use of phytochromes in crop breeding programs for developing varieties resistant to negative growth factors under thickened sowing conditions. Other examples of studies of the effects of phytochrome mutations on plant phenotypes are the works performed by (Chen et al., 2013; Zhang et al., 2013). Thus, J. Zhang et al. (2013) demonstrated the effect of *phyB* mutations in *A. thaliana* on a number of plant phenotypic traits, while J. Chen et al. (2013) showed that loss of PHYC functional activity in wheat could lead to changes in the circadian rhythm and a sharp delay in flowering during the long daylight hours.

The second, fourth and ninth places belong to the genes from the 14-3-3-like proteins family (*GF14*, *RCI1A*, and *GRF8*, respectively). These proteins regulate the cell cycle, apoptosis, immune processes, nitrogen and carbon metabolism, and are involved in the regulation of starch synthesis, ATP production, detoxification by peroxide and in some other biochemical pathways. Also, the plant development and seed germination are controlled by factors which are activated by interacting with 14-3-3-like proteins (Fulgosi et al., 2002). Świędrych et al. (2002) showed that decrease in the level of GF14 protein leads to an increase in calcium, starch and an increase in the ratio of soluble sugars to starch in potato tubers, as well as to the significant increase of methionine, proline, and arginine in potato cells. It was shown that the suppression of *GF14e* gene expression by the RNA interference method could lead to the increased resistance of rice to the virulent strain of the *Xanthomonas oryzae* pv. *oryzae* (Xoo) bacterial phytopathogen (Manosalva et al., 2011). It is known that plants

Table 1. Top 10 biological processes that provide resistance to adverse environmental factors, ordered according to the number of associated genes

| Name of the Gene Ontology biological process	Gene Ontology ID number	Genes/ proteins
Response to oxidative stress	GO:0006979	118
Response to biotic stimulus	GO:0009607	71
Response to toxic substance	GO:0009636	68
Obsolete drought tolerance	GO:0009633	62
Response to salt stress	GO:0009651	49
Obsolete disease resistance	GO:0009614	39
Response to cold	GO:0009409	29
Systemic acquired resistance	GO:0009627	26
Response to osmotic stress	GO:0006970	23
Heat acclimation	GO:0010286	14

with mutations in the *RCI1A* and *GRF8* genes are having increased resistance to the low temperatures (Catalá et al., 2014; Liu et al., 2017).

The third place was given to the *CNIH1* gene, which encodes the plants protein interacting with the sodium transporter HKT1 and providing the correct location of the transporter on the Golgi apparatus membrane (Rosas-Santiago et al., 2017). In the work of M.M. Wudick et al. (2018), the effect of mutations in the *CNIH1* gene on pollen and calcium homeostasis in *A. thaliana* was studied. It is interesting to note that potato cultivar Yubiley Zhukova, which has enhanced salt and drought tolerance, was obtained due to the overexpression of the vacuolar Na^+/H^+ antiporter NHX2 (Belyaev et al., 2011).

The homolog of the transcription factor ABI5 bZIP-type (*ABI5*) was ranked fifth. ABI5 plays an important role in seed germination, which is regulated by abscisic acid

Table 2. The top 10 high-priority candidate genes potentially involved in resistance to adverse factors

Rank	Gene	Name of protein	Cross-talk centrality
1	*PHYA*	Phytochrome A	0.011514615
2	*GF14*	14-3-3-like protein	0.011514615
3	*CNIH1*	Antiporter CNIH1	0.011514615
4	*RCI1A*	14-3-3-like protein GF14 psi	0.010628875
5	*ABI5*	ABI5-bZIP transcription factor ABI5 homolog	0.009743136
6	*CPK1*	Calcium-dependent protein kinase 1	0.008857396
7	*RGS1*	Regulator of G-protein signaling 1	0.008857396
8	*NHL3*	NDR1/HIN1-like protein 3	0.008857396
9	*GRF8*	14-3-3-like protein GF14 kappa	0.008857396
10	*CYP21-4*	Peptidyl-prolyl cis-trans isomerase CYP21-4	0.007971656
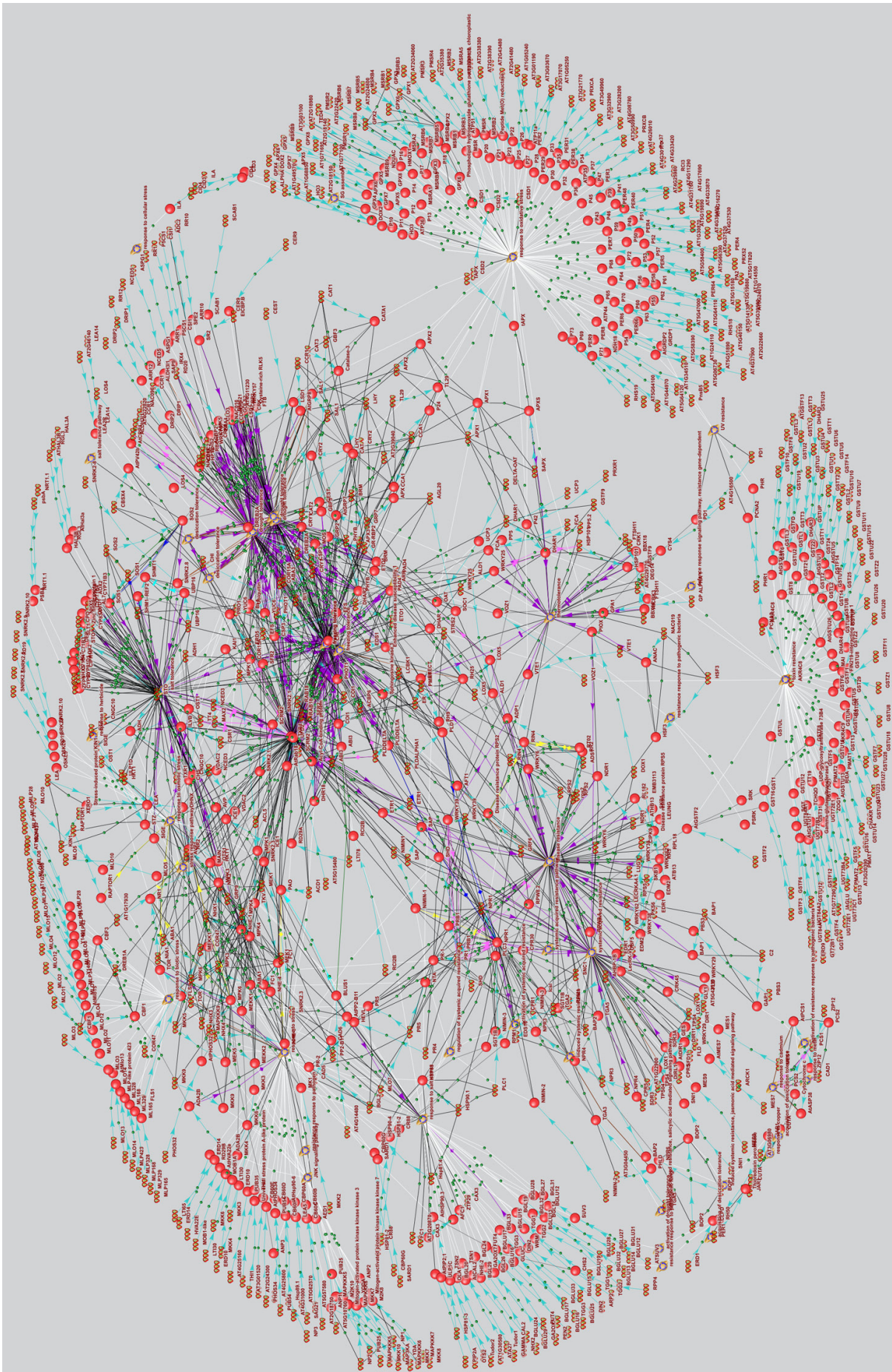
Associative gene network of resistance to adverse factors in potato.

Genes are presented by spirals, proteins by red balls, biological processes by brown ovals, and the interaction by lines.

П.С. Деменков, О.В. Сайк, Т.В. Иванисенко
Н.А. Колчанов, А.В. Кочетов, В.А. Иванисенко

Приоритизация генов картофеля
с использованием базы знаний SOLANUM TUBEROSUM

2019
23 • 3

(Finkelstein, 1994; Lopez-Molina et al., 2002). The ABI5 bZIP-type transcription factor is involved in the activation of genes responsible for the accumulation of proteins during seed development. It is known that the reduction of the expression of the *ABI5* gene activates the meristem growth (Lopez-Molina et al., 2002). Mutations of the *ABI5* gene in *A. thaliana* are associated with reduced sensitivity to abscisic acid, as well as to salt and osmotic stress during the germination (Finkelstein, Lynch, 2000; Carles et al., 2002; Tezuka et al., 2013).

In sixth place was the *CPK1* gene, it encodes a calcium-dependent protein kinase C, which is involved in the immune response, resistance to fungal diseases and pathogens (Gravino et al., 2015). Mutations of *CPK1* gene in the *A. thaliana* are known to cause hypersensitivity to salt stress and drought, while transgenic plant lines with increased expression of *CPK1* showed significant resistance to salt stress and drought (Huang et al., 2018).

The seventh place was taken by the *RGS1* gene, which encodes the negative regulator of the signaling pathway of G-protein type 1. It is known that the expression of this gene decreases in response to water deficiency (Campbell et al., 2012). A. Chen et al. (2006) showed for *A. thaliana* that transgenic plants with over-expressed *RGS1* gene have an increased drought tolerance.

The *NHL3* gene appeared to be in the eighth place; this gene encodes NDR1/HIN1-like protein 3, which is involved in response to pathogens (Chong et al., 2008). The *A. thaliana* transgenic line, in which increased expression of the *NHL3* gene was observed, showed the increased resistance to the pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (Varet et al., 2003).

The tenth place in Table 2 was taken by the *CYP21-4* gene, encoding cyclophilin, localized in the Golgi apparatus, which is involved in tolerance to oxidative stress (Park et al., 2017). The authors consider the over-expression of the *CYP21-4* gene in crops as a new promising way to increase the productivity of plants. For potatoes and rice, it has been shown that transgenic plants in which the *CYP21-4* gene is over-expressed have increased yield, the stems and roots of the plants are longer, and the leaves are thicker. Also, such potatoes produced a bigger number of tubers of a larger size, and the microtubers were formed faster than in wild-type plants (Park et al., 2017).

Conclusion

In the current work, a search for new potential participants of molecular genetic mechanisms of resistance to adverse factors in plants was carried out. Prioritization of candidate genes has shown that the *PHYA*, *GF14*, *CNIH1*, *RCI1A*, *ABI5*, *CPK1*, *RGS1*, *NHL3*, *GRF8* and *CYP21-4* genes are the most promising for the further study of their relationship with resistance to adverse factors. The performed analysis reviled that the molecular-genetic relationships responsible for the formation of significant agricultural traits are complex and include many direct and indirect interactions. The representation of these interactions in the form of associative gene networks and their analysis using the SOLANUM TUBEROSUM knowledge base can be the basis for the search for target genes important for targeted mutagenesis

and marker-oriented selection of potato varieties resistant to adverse environmental factors.

References

Aggarwal C.C., Zhai C. Mining Text Data. Springer Science & Business Media, 2012.

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)

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., Antieaua 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 elongaion 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., Hripcsak 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.

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 Selektsii = Vavilov Journal of Genetics and Breeding. 2018;22(1): 8-17. (in Russian)

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. 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)

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., 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)

Saik O.V., Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Interactome 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.

Schlagnhaufer C.D., Arteca 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/amiajnl-2011-000096.

Świędrych 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.

П.С. Деменков, О.В. Сайк, Т.В. Иванисенко
Н.А. Колчанов, А.В. Кочетов, В.А. Иванисенко

Приоритизация генов картофеля
с использованием базы знаний SOLANUM TUBEROSUM

2019
23•3

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 Ca^{2+} homeostasis. Science. 2018;360(6388): 533-536.

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)

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.1365-313x.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