

Original Russian text [www.bionet.nsc.ru/vogis/](http://www.bionet.nsc.ru/vogis/)

# 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<sup>1,2</sup>, E.A. Oshchepkova<sup>1</sup>, T.V. Ivanisenko<sup>1</sup>, V.A. Ivanisenko<sup>2,3</sup> 

<sup>1</sup> Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

<sup>3</sup> Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia

 [salix@bionet.nsc.ru](mailto:salix@bionet.nsc.ru)

**Abstract.** Methods for prioritizing 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 Seleksii = Vavilov Journal of Genetics and Breeding*. 2021;25(5): 580-592. DOI 10.18699/VJ21.065

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

П.С. Деменков<sup>1,2</sup>, Е.А. Ощепкова<sup>1</sup>, Т.В. Иванисенко<sup>1</sup>, В.А. Иванисенко<sup>2,3</sup> 

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

<sup>2</sup> Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

<sup>3</sup> Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия

 [salix@bionet.nsc.ru](mailto: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.

## Introduction

The rapid development of high-performance experimental methods has significantly expanded the ability to generate large sets of genomic, transcriptomic and proteomic data in scientific research. This, in turn, has increased the relevance of bioinformatics methods that allow researchers to interpret omic data, both at the level of key genes and at the level of molecular genetic mechanisms. Among the widely used approaches in computational analysis of gene sets identified in experiments are prioritization methods (Raj, Sreeja, 2018), which rank the studied genes (or other objects, such as diseases) by characterising their proximity to a set from a given learning sample. Depending on the prioritization problem, the training set may consist of genes associated with diseases or phenotypic traits or sets of differentially expressed genes, for example. The higher the proximity in relative units, the greater the priority of the analysed object as a candidate possessing the same properties as objects in the training set. Such methods are used in biomedicine to detect candidate genes associated with diseases (Tranchevent et al., 2016), disease biomarkers (Jha et al., 2020), potential pharmacological targets (Cesur et al., 2020) and drug republic (Pushpakom et al., 2019). In animal husbandry and crop production, prioritization methods have been applied to analyse genomic data related to marker-oriented and genomic selection (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), as well as Raspanic loci analysis (Bargsten et al., 2014; Schaefer et al., 2018; Lin et al., 2019).

A special place among the prioritization methods is occupied by approaches based on the analysis of genetic network graphs, including network protein interactions, metabolic networks, signal transmission networks, and networks of genes associated with disease. In such network prioritization methods, the proximity of the studied genes to the training sample is estimated using various topological characteristics of the genetic network graph. Methods for analysing the structure of genetic network graphs to solve prioritization tasks can be divided into three large groups (Shim et al., 2017; Raj, Sreeja, 2018): (1) methods based on identifying hubs using centrality indicators of vertices in the column (Cho et

al., 2016); (2) methods based on network diffusion, including random wandering (Chen et al., 2011; Shim, Lee, 2015; Le, Pham, 2017; Lysenko et al., 2017); and (3) methods based on identifying functional modules (clusters or subnets) (Jia et al., 2011; Leung et al., 2014). All of these methods are aimed at identifying genes (or other entities) that are important for the phenotype or process being studied. The importance of information about the topology of networks in assessing the functional significance of genes was demonstrated by the example of the centrality of the vertices in a *Saccharomyces cerevisiae* protein–protein interactions network (Jeong et al., 2001). The authors showed that the deletion of vertices with a large number of connections in the network of protein–protein interactions was fatal more often than the deletion of other vertices.

Among the frequently used characteristics to determine the importance of the vertex in the network structure are centrality indicators, including Degree Centrality, which is the number of links this vertex has with other vertices in the network (Freeman, 1978); centrality proximity to the centre of the graph (Closeness Centrality), which is the reverse amount of the sum of the lengths of all the shortest pathways passing through the top (Sabidussi, 1966); and centrality (Betweenness Centrality), which is the number of the shortest pathways passing through the vertex (Freeman, 1977).

Previously, we proposed the concept of an associative gene network (Ivanisenko V.A. et al., 2015), an extended gene network whose vertices can represent not only molecular genetic objects (genes, proteins, metabolites, etc.), but also higher-level factors (biological processes, phenotypic traits, diseases, factors of the external environment, etc.) related to regulatory, physical, chemical or associative interactions (Ivanisenko V.A. et al., 2015). Automatic analysis of texts of scientific publications and factual databases was used to create knowledge bases: ANDSystem, which contains associative gene networks for animals and humans (Ivanisenko V.A. et al., 2015, 2019; Ivanisenko T.V. et al., 2020), and SOLANUM TUBEROSUM (Saik et al., 2017; Ivanisenko T.V. et al., 2018), which contains associative gene networks of plants. Analysis of associative gene networks from the knowledge bases ANDSystem and SOLANUM TUBEROSUM was used to develop methods for

prioritizing human genes associated with diseases (Saik et al., 2018, 2019) and the potato genes involved in breeding and the development of agriculturally meaningful traits (Demenkov et al., 2019), respectively. The prioritization method was based on the assessment of the centrality indicator according to the degree of vertices corresponding to genes in the analysed associative gene networks.

In this work, an approach is proposed to prioritize biological processes by calculating the vertex centrality indicator in the associative gene network. This work uses the example of the model plant *Arabidopsis thaliana* and the information about the biological processes associated with the response of plants to the unfavourable factors cadmium content, drought conditions and salt stress in the knowledge base SOLANUM TUBEROSUM.

## Materials and methods

SOLANUM TUBEROSUM (Saik et al., 2017; Ivanisenko T.V. et al., 2018) consists of three main blocks.

The first (block 1) is an automatic analysis of texts of scientific publications and factual databases that is designed to extract information about the relationship between objects using semantic–linguistic templates. To extract knowledge, ANDSystem software tools configured to the subject area under study are used (Ivanisenko V.A. et al., 2015). The setting of the subject area includes the creation of new semantic–linguistic templates that take into account the specifics of the texts of scientific publications in the field of biology of plants and crop production.

The second block is the knowledge base containing object dictionaries and information about the relationships between objects extracted in block 1 in the form of an integrated associative gene network (a graph in which the vertices correspond to objects and the ribs indicate the specified types of connections). The SOLANUM TUBEROSUM knowledge base contains dictionaries of molecular genetic objects (genes, proteins, metabolites, microRNAs), Gene Ontology biological processes (Gene Ontology Consortium, 2019), phenotypic traits, diseases and pathogens of potatoes and model organisms, among other information. To describe relationships between these objects in associative gene networks, more than 25 different types of interactions are used. These correspond to, for example, physical interactions, catalytic reactions, regulation, participation and associations.

In block 3, ANDVisio (Demenkov et al., 2012), designed to reconstruct and analyse associative gene networks related to the problem under study, is applied based on the knowledge base of SOLANUM TUBEROSUM. The SOLANUM TUBEROSUM knowledge base can be accessed via the Internet at <https://www-bionet.ssc.ru/and/plant/>.

The number of interactions in the SOLANUM TUBEROSUM knowledge base for three plant species is shown in Table 1.

The schematic diagram of all stages of the prioritization is shown in Fig. 1. A detailed description of the development of the SOLANUM TUBEROSUM knowledge base (see Fig. 1, a) can be found in previous studies (Saik et al., 2017;

Ivanisenko T.V. et al., 2018). The prioritization algorithm consists of several steps (see Fig. 1, b), which are fully automated in the ANDVisio program. The ANDVisio program provides the user interface for access to the SOLANUM TUBEROSUM knowledge base, and has a wide range of tools for reconstruction, graphic visualisation and analysis of associative gene networks.

In the first step, the associative gene network for the problem being studied is automatically reconstructed as specified by the user, incorporating biological processes from the Gene Ontology database. To achieve this, the “Reconstruction of the Network” procedure is implemented in ANDVisio. In the interface window for this procedure, the name of the Gene Ontology biological process is selected from the list provided (which contains all of the Gene Ontology biological processes, updated annually). Upon the execution of this procedure, an associative gene network related to a given biological process will be automatically reconstructed and visualised in graphical form. The reconstruction algorithm includes an appeal using MySQL requests to the SOLANUM TUBEROSUM knowledge base, which generates a list of genes with a connection to the specified biological process. The user can choose whether to use interactions identified by analysing the texts of scientific publications or those extracted from factual databases. Genes from this list become vertices in the desired associative gene network. The formal description of the algorithm is shown in Fig. 1, b. Using this algorithm, networks were reconstructed for three biological processes: “Response to Cadmium Ion”, “Drought Tolerance” and “Response to Salt Stress”.

In the second step, for each reconstructed associative gene network, biological processes were prioritized according to the algorithm shown in Fig. 1, b. Prioritization is an automated iterative process performed by the ANDVisio program. Briefly, all Gene Ontology biological processes in the SOLANUM TUBEROSUM knowledge base dictionaries are sequentially analysed. The analysis included the calculation of the CTC (cross-talk centrality) indicator. The biological processes were ranked according to their CTC values thus, biological processes with the highest indicator received maximum priority.

The CTC indicator for the vertex  $i$  was calculated using the formula:

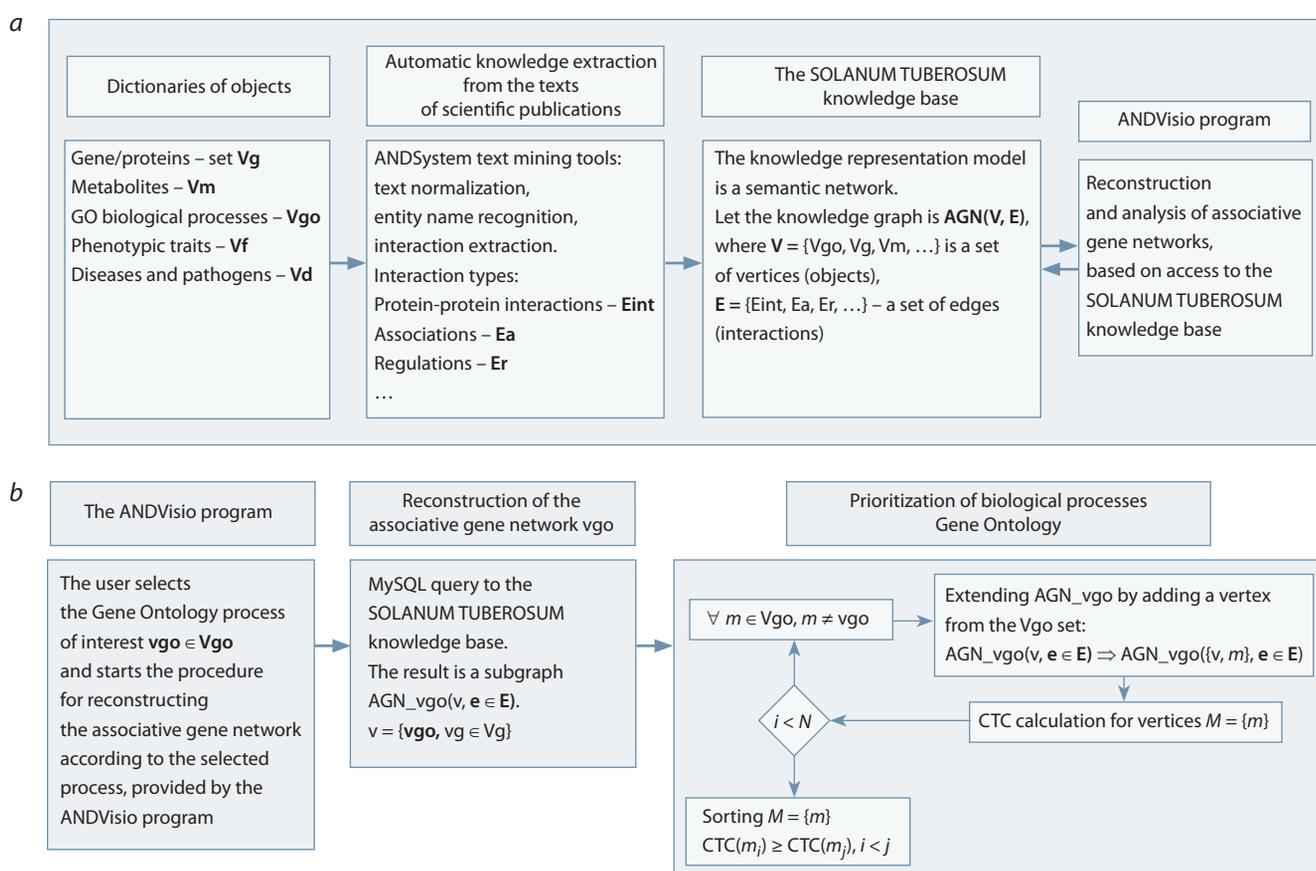
$$CTC_i = N_i/M, \quad M \neq 0,$$

where  $N_i$  is the number of vertices to which the  $i$  vertex is connected by an edge in an analysed column of an associative gene network,  $M$  is the total number of vertices of the associative gene network, and  $i$  takes values from 1 to  $M$ . The values of the CTC indicator range from 0 to 1, with the value of 0 occurring when the peak under consideration on the network is not related to other vertices. The characteristic  $N_i$  is known as the indicator of the Degree centrality of the vertex (Freeman, 1978). This indicator is used because it is assumed that the larger the vertices associated with this vertex, the greater its influence in the associative gene network and the molecular genetic mechanism as a whole. To assess the significance of the centrality of the vertex (biological process), an approach to calculate the overrepresentation of biological processes was used (Subramanian et al., 2005). Using this approach,

**Table 1.** Number of basic types of interactions, including molecular genetic interactions and associations, presented in the knowledge base of SOLANUM TUBEROSUM for three plant species

Interaction type	<i>Solanum tuberosum</i> /potato	<i>Arabidopsis thaliana</i> /arabidopsis	<i>Zea mays</i> /maize
Association	4783	294 520	15 280
Catalyze	1140	26 842	1203
Regulation*	346	20 213	1120
Molecular interaction	69	16 589	178
Involvement	3857	81 067	5906

\*The total number of links is shown for the following types of interactions: upregulation, downregulation, regulation, activity upregulation, activity downregulation, activity regulation, expression upregulation, expression downregulation, expression regulation, transport upregulation, transport downregulation and transport regulation.



**Fig. 1.** Schematic diagram of all steps performed during prioritization: *a*, creating the SOLANUM TUBEROSUM knowledge base; *b*, prioritization of Gene Ontology biological processes (Gene Ontology Consortium, 2019).

the probability ( $p$ -value) that this vertex will have  $N_i$  or more links by random chance in the associative gene network was evaluated using a hypergeometric distribution adjusted to the Benjamini–Yekutieli multiple comparison procedure (Benjamini, Yekutieli, 2001).

For the characterisation of the identified processes, a cluster analysis was carried out by the Markov Cluster algorithm (Van Dongen, Abreu-Goodger, 2012), using the semantic metric for biological processes by Wang and colleagues (Wang J.Z. et al., 2007).

## Results and discussion

Using the SOLANUM TUBEROSUM knowledge base, we reconstructed the associative gene networks of *A. thaliana* that described the interaction of genes with biological processes related to plant responses to adverse environmental factors, including drought, salt stress and increased cadmium content. To study the potential molecular genetic mechanisms underlying the reconstructed gene networks, biological processes were prioritized based on the centrality of their interactions with the network’s genes/proteins.

### Plant responses to drought

The associative gene network for the biological process “drought tolerance”, reconstructed in the SOLANUM TUBEROSUM knowledge base, included 292 vertices (proteins) interconnected by 440 ribs (Fig. 2). To simplify the representation of the interactions, the genes corresponding to the proteins are not illustrated. The relatively large number of proteins associated with the term “drought tolerance” can be explained by the fact that plant resistance to drought is conditioned by numerous aspects of plant physiology – a search query in PubMed for the keywords “drought” and “plants” yields more than 18,000 publications.

It should be noted that the analysis based on automatic processing of scientific literature data depends on how deeply the subject area has been studied. Thus, our conclusions can indicate the presence of an interaction, according to the published data, but cannot assert the absence of an interaction simply because the interaction is not discussed in the literature.

Among the biological processes associated with proteins in the associative gene network “drought tolerance”, there were 208 processes with a *Q*-value below the significance level of 0.05. Their centrality scores ranged from 0.0067 to 0.58. The identified biological processes were divided into 22 clusters according to semantic proximity (Fig. 3).

A list of 12 biological processes with the largest number of links with other objects in the drought tolerance gene network is shown in Table 2.

The prioritization enables the identification of such fundamental processes as “transcription”, “signaling” and “gene expression”, which fell into the clusters “post-translational modifications”, “cellular processes” and “cellular metabolism”, respectively. The processes associated with transcription, translation and gene expression are widely discussed in

the context of transcription factors and their participation in plant responses to adverse environmental conditions (Leng, Zhao, 2020).

Signaling processes induced by external and endogenous factors also play an important role. In particular, the process “ABA signaling” (abscisic acid signaling pathway) was highly rated (see Table 2). The CTC level for this process was found to be 0.34. This means that when the analysed associative gene network “drought tolerance” is expanded by adding the “ABA signaling” vertex, 34 % of all vertices will be associated with it. Many studies have demonstrated the high importance of ABA in plant responses to drought (approximately 3000 publications in PubMed). It has been shown that ABA is synthesised in leaves in response to a water deficiency signal transmitted from roots to shoots and is involved in mechanisms of drought resistance, such as stomata closure and production of osmotic defence proteins (Takahashi et al., 2018). ABA is also involved in the regulation of flowering time (“flowering” is on the list of priority processes, see Table 2). For example, it was shown that under drought conditions, ABA-responsive element binding factor 3 (ABF3) and ABF4, along with nuclear transcription factor Y subunit gamma (NF-YC), enhance the expression of suppressor of overexpression of constans 1 (SOC1), thus promoting acceleration of plant flowering, allowing the plant to complete its life cycle at an earlier date (Hwang et al., 2019).

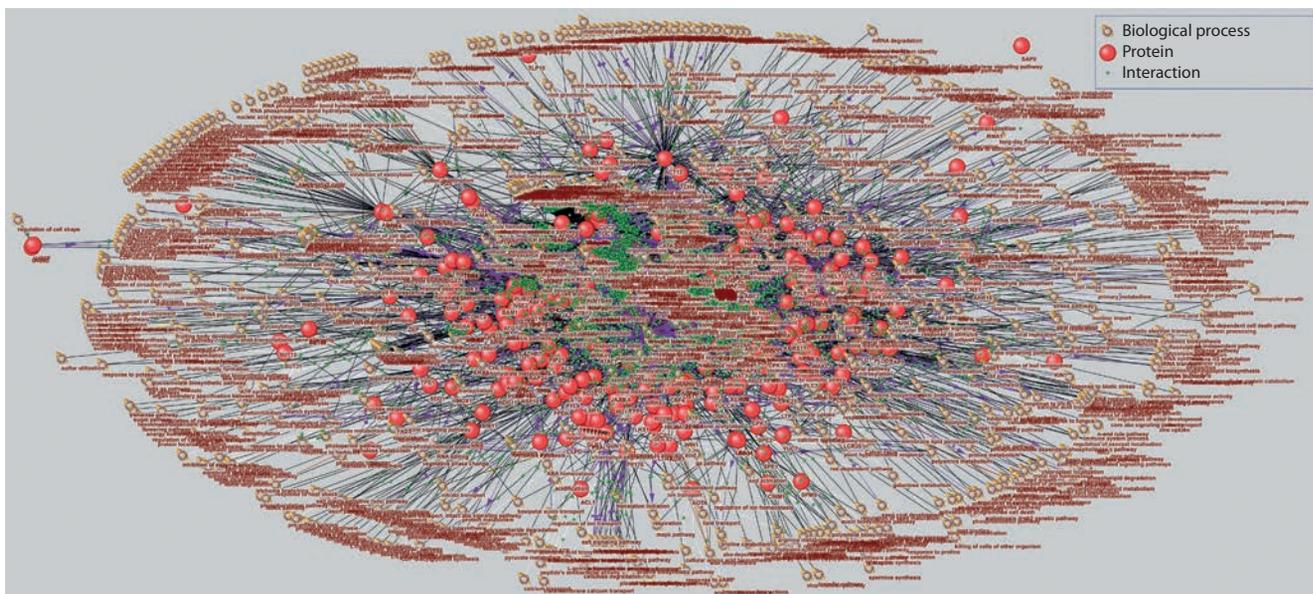
Another process associated with the response of plants to drought, located in the upper lines of the prioritization results (see Table 2), is phosphorylation (one of the main ways to transmit regulatory signals in a cell). For example, it was shown in *A. thaliana* that the transcriptional activation of genes sensitive to drought requires phosphorylation of the transcription factor RD26 by the kinase brassinosteroid in-

**Table 2.** Ranking of biological processes according to their potential relationship with the “drought tolerance” process

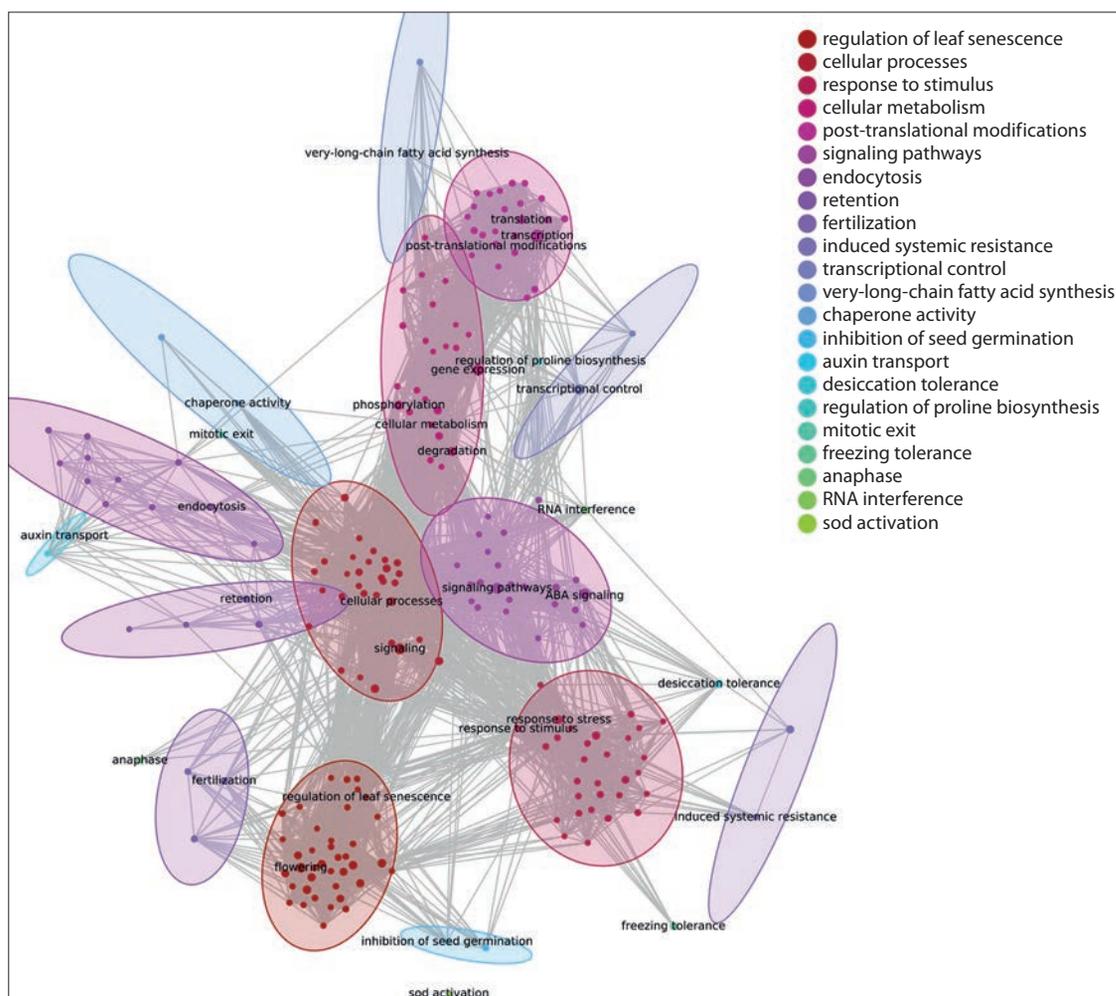
No.	Biological process	Cluster name	Number of links*	CTC	<i>p</i> -value	<i>Q</i> -value**
1	Transcription	post-translational modifications	173	0.59	7E-64	3E-61
2	Signaling	cellular processes	121	0.41	1E-81	7E-79
3	Gene expression	cellular metabolism	121	0.41	2E-73	1E-70
4	Signaling pathways	signaling 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

\* The number of connections between the biological process and proteins/genes in the associative network.

\*\* Significance adjusted for multiple Benjamini–Yekutieli comparisons.



**Fig. 2.** Associative gene network “drought tolerance” for *Arabidopsis thaliana*, including proteins and biological processes as vertices.



**Fig. 3.** Semantic proximity clustering of biological processes significantly overrepresented in the “drought tolerance” associative gene network.

Ovals outline vertices from one cluster. The colour of the vertices and ovals matches the colour of the cluster in the legend.

sensitive 2 (BIN2) (Jiang et al., 2019). In a study of a drought-resistant rapeseed cultivar (*Brassica napus* L.), an important role was shown for beta carbonic anhydrase 1 (βCA1) phosphorylation in photosynthesis regulation under drought conditions (Wang L. et al., 2016).

The position of biological processes such as plant resistance to cold and salt stress in the list of prioritized processes in the analysis of plant response to drought conditions deserves special discussion. Although these processes were not at the top of the list of processes with the highest priority for the associative gene network “drought tolerance” (see Table 2), they had a rather high rating (CTC = 0.1). Indeed, plant resistance to various adverse environmental factors is often mediated by the same molecular genetic mechanisms. For example, it was shown in corn (*Zea mays*) that the expression of the transcription factor MYB (MYB3R) is induced during both drought and salt stress, contributing to plant resistance to these environmental factors (Wu et al., 2019). In soybean (*Glycine max*), it was shown that another transcription factor in the MYB family, MYB118, also plays an important role in plant resistance to drought (Du et al., 2018). Its expression, like that of MYB3R, is induced under drought and salt stress, and the newly synthesised MYB118 enhances the expression of stress-associated genes that mediate the plant response to these stresses. In cotton (*Gossypium* spp.), an important role was shown for proteins in the cyclin-dependent kinase family in plant response to drought and salt stress (Magwanga et al., 2018). The mechanisms of plant response to drought overlap with the mechanisms of response to cold stress: in both cases, the synthesis and mobilisation of abscisic acid is induced in the vascular tissue of the leaves, which is necessary for the regulation of stomatal closure (reviewed in Agurla et al., 2018).

### Plant response to salinity

The “response to salt stress” associative gene network contained 81 vertices and 102 links (Fig. 4). Twelve of the highest-priority processes with a centrality index of at least 0.2 were identified (Table 3). As in the case of the drought resistance associative gene network, all of the processes included in the list of the highest priority processes are statistically significant.

Among the biological processes associated with proteins in the “response to salt stress” associative gene network, 85 had a *Q*-value of <0.05. Their centrality scores ranged from 0.02 to 0.34. The identified biological processes were divided into 14 clusters according to semantic proximity (Fig. 5).

A list of the 12 biological processes with the largest number of connections with other objects in the “response to salt stress” gene network is shown in Table 3.

Salt stress tolerance is widely represented in the scientific literature – a PubMed search for the keywords “salt” and “plants” yields more than 24,000 publications. The analysis yielded a list of top-rated processes that included “seed germination”, which fell into the “respiration” cluster (see Table 3). For example, the authors showed in *Stylosanthes humilis* that increased salt content suppresses seed germination and results in decreased ethylene production and increased abscisic acid production (Silva et al., 2018). Leaf senescence during normal plant development and caused by salt stress occurs via common hydrogen peroxide-mediated signaling pathways, as shown in *A. thaliana* (Allu et al., 2014).

### Plant response to the presence of cadmium

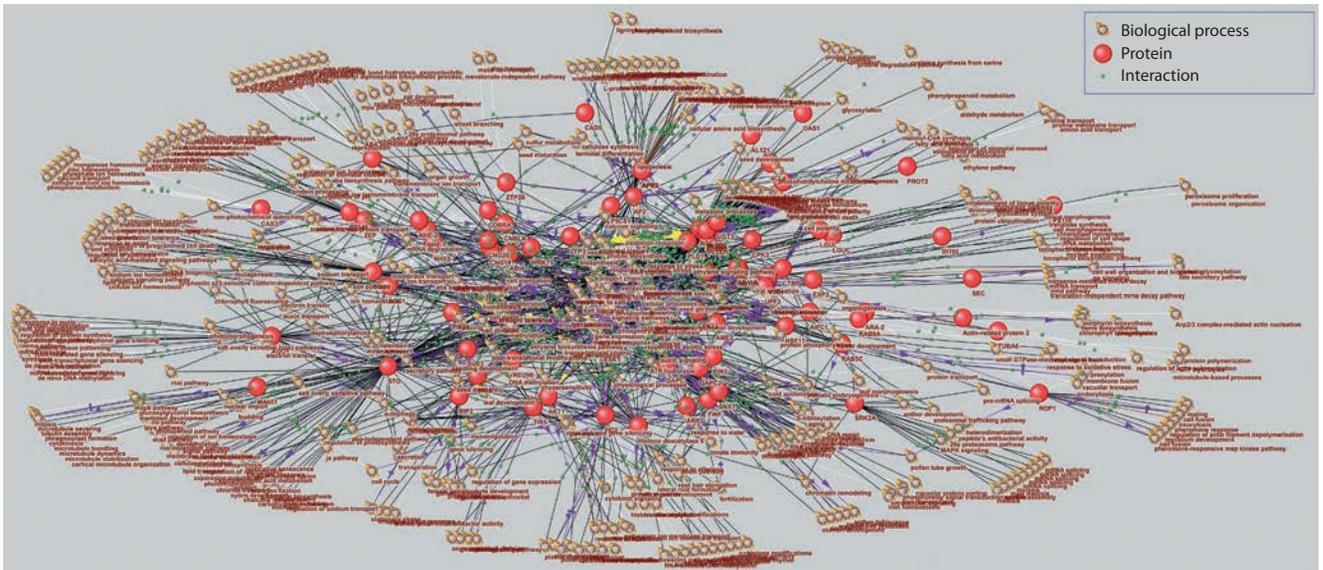
Cadmium is toxic to most plants and animals (Genchi et al., 2020). It can be present in soil, causing stress responses in plants, suppressing the growth of roots and shoots and re-

**Table 3.** Ranking of biological processes according to their potential relationship with the “response to salt stress” process

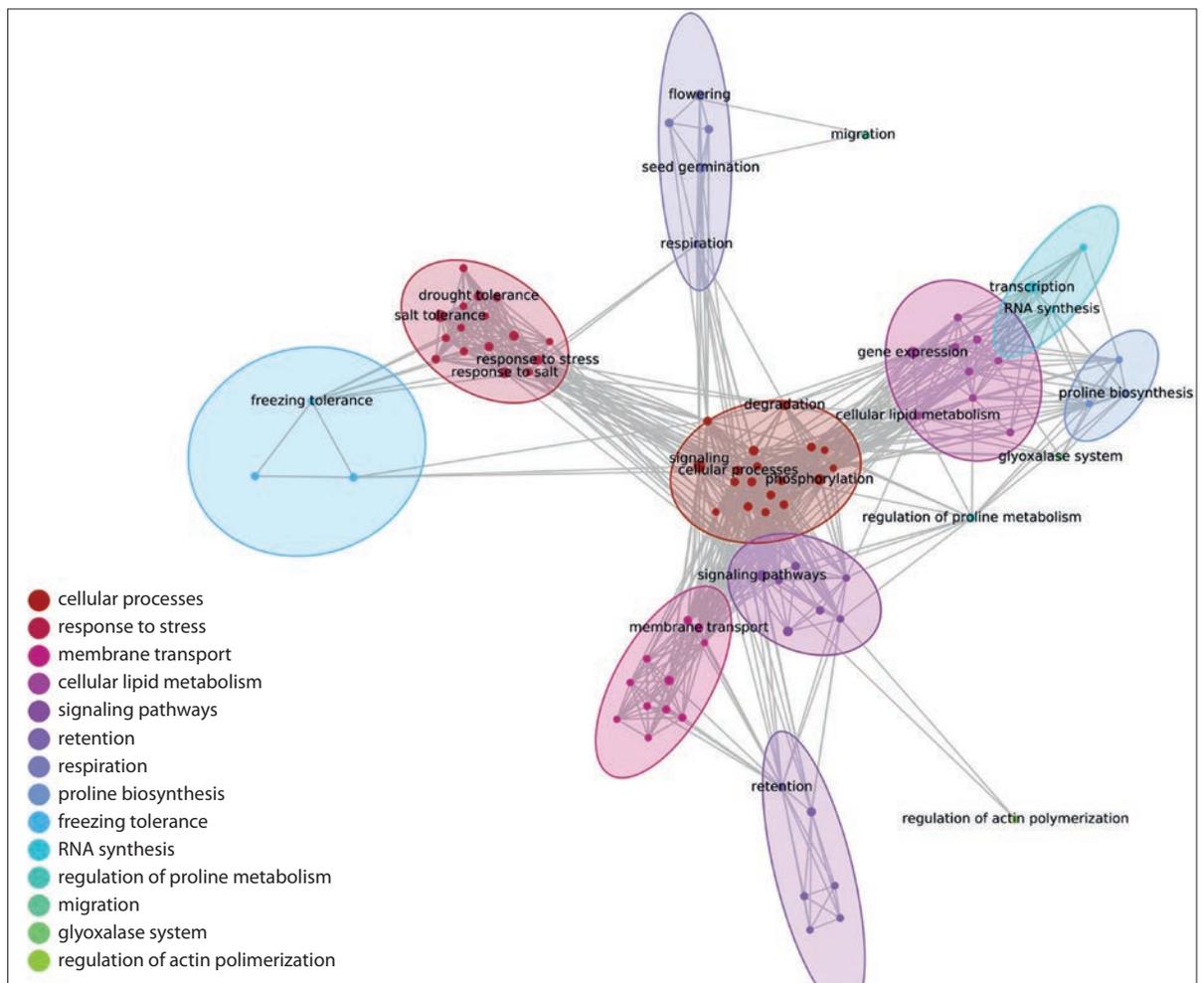
No.	Biological process	Cluster name	Number of links*	CTC	<i>p</i> -value	<i>Q</i> -value**
1	Transcription	RNA synthesis	30	0.34	1.8E–5	8.0E–4
2	Signaling	cellular processes	28	0.32	4.5E–16	1.3E–13
3	Salt tolerance	response to stress	28	0.32	6.5E–32	3.1E–29
4	Gene expression	cellular lipid metabolism	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

\* The number of connections between the biological process and proteins/genes in the associative network.

\*\* Significance adjusted for multiple Benjamini–Yekutieli comparisons.

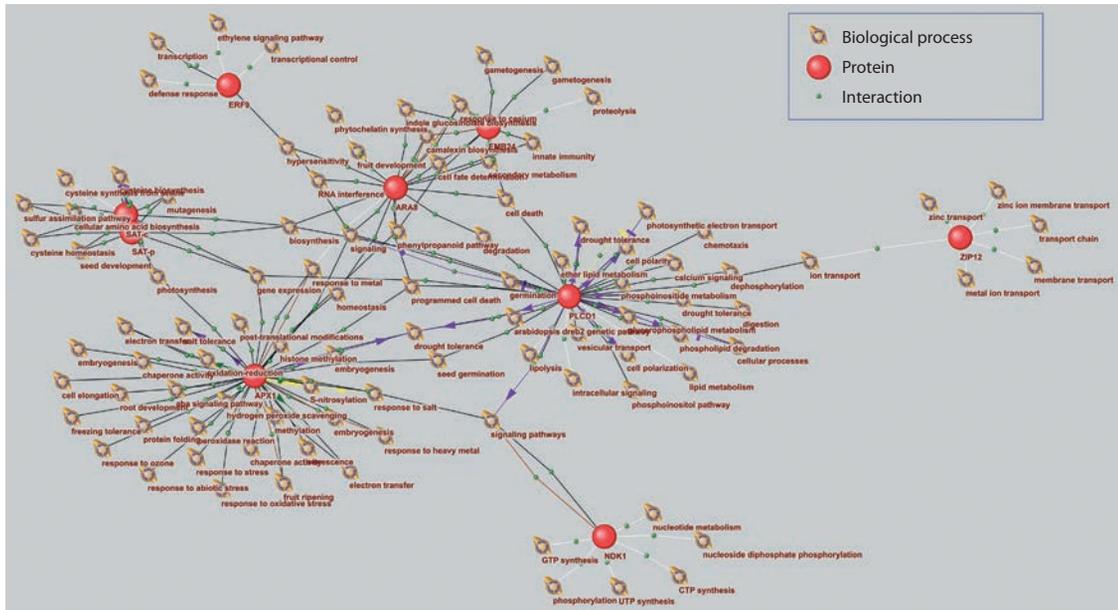


**Fig. 4.** Associative gene network “response to salt stress” for *Arabidopsis thaliana*, including proteins and biological processes as vertices.

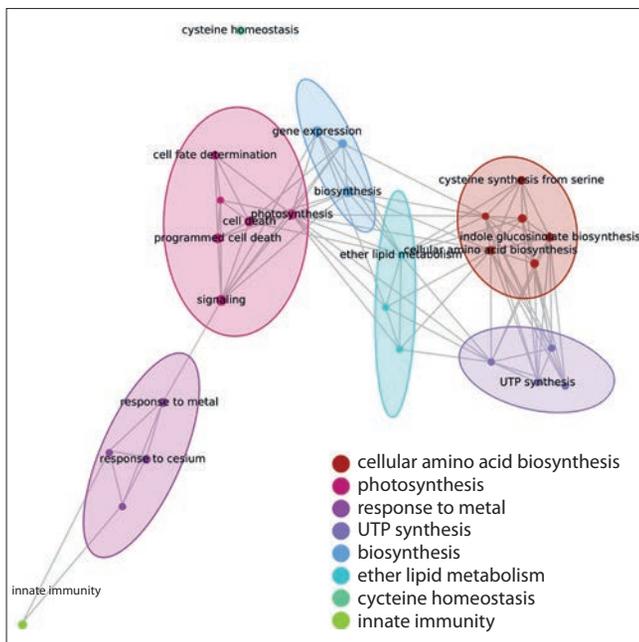


**Fig. 5.** Semantic proximity clustering of biological processes significantly overrepresented in the “response to salt stress” associative gene network.

Ovals outline vertices from one cluster. The colour of the vertices and ovals matches the colour of the cluster in the legend.



**Fig. 6.** Associative gene network “response to cadmium ion” for *Arabidopsis thaliana*, including proteins and biological processes as vertices.



**Fig. 7.** Semantic proximity clustering of biological processes significantly overrepresented in the “response to cadmium ion” associative gene network.

Ovals outline vertices from one cluster. The colour of the vertices and ovals matches the colour of the cluster in the legend.

ducing the rate of photosynthesis and nutrient consumption (Genchi et al., 2020; Kaya et al., 2020). The increased interest in the problem is due to some plants being prone to cadmium hyperaccumulation, which raises prospects for their use for both soil cleaning and other industrial purposes (Küpper, Leitenmaier, 2013). However, the mechanisms underlying

plant response to cadmium remain poorly understood. Nine genes were associated with the biological process “response to cadmium ion” in the SOLANUM TUBEROSUM knowledge base for *A. thaliana* (Fig. 6).

Among the biological processes related to proteins in the “response to cadmium ion” associative gene network, 28 had a *Q*-value below 0.05. Their centrality scores ranged from 0.11 to 0.44. The identified biological processes were divided into 8 clusters according to semantic proximity (Fig. 7).

A list of the 12 biological processes with the largest number of connections with other objects in the “response to cadmium ion” gene network is shown in Table 4. Among the biological processes with a high rating in the cadmium response gene network, there were two processes associated with cell death (“programmed cell death” and “cell death”), which were in the “photosynthesis” cluster. The importance of programmed cell death in response to cadmium has been confirmed in dozens of publications. For example, it was shown that cadmium causes morphophysiological changes and programmed cell death in *Genipa americana* L. (Souza et al., 2011); moreover, in tomato cell culture, it was demonstrated that cadmium induces programmed cell death using caspase-like proteases (Iakimova et al., 2008).

In studies of the molecular mechanisms of plant responses to unfavourable factors, it may be of great interest to identify both general and specific biological processes. Below we consider examples of general and specific processes.

### Comparative analysis of reconstructed associative gene networks

Comparison of the “response to salt stress” and “drought tolerance” associative networks revealed 41 common biological processes, 40 processes specific to the response to salt stress

**Table 4.** Ranking of biological processes according to their potential relationship with the “response to cadmium ion” process

No.	Biological process	Cluster name	Number of links*	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

\* The number of connections between the biological process and proteins/genes in the associative network.

\*\* Significance adjusted for multiple Benjamini–Yekutieli comparisons.

network and 156 processes specific to the drought tolerance network. Examination of all three associative gene networks revealed five common biological processes: signaling, cell death, programmed cell death, gene expression and photosynthesis. Notably, the tables of biological processes with the greatest number of connections with other objects (see Tables 2–4) mainly include general processes that have a statistically significant relationship with two or all three of the gene networks of plant responses to unfavourable factors.

Identification of biological processes specific to the plant response to drought compared with the responses to salinity and cadmium has revealed a group of processes associated with seed development: “seedling development”, “seed development”, “seed dormancy”, “seed maturation”, “regulation of seed size” and “inhibition of seed germination”. Indeed, many articles demonstrating the impact of drought on these processes have been published. It was shown that a lack of water leads to a decrease in the rate and duration of maturation of lentil seeds, leading to a decrease in their size (Sehgal et al., 2019). Similar studies on soybean lines clearly showed that drought conditions affect the quality of seed preparation and subsequent germination (Wijewardana et al., 2019).

When identifying biological processes specific to the response of plants to salinity, a group of processes associated with ion transport was distinguished: “ion transport”, “membrane transport”, “sodium transport”, “cation transport”, “transmembrane proton transport”, “sodium ion transmembrane transport” and “potassium ion homeostasis”. Indeed, the adaptation of plants to conditions of salt excess depends on their ability to remove Na<sup>+</sup> and Cl<sup>-</sup> ions or to increase resistance to osmotic stress and ion accumulation in tissues (Munns, Tester, 2008).

Comparing the biological processes specific to the plant response to the presence of cadmium with those involved in

responses to drought and salinity highlights several processes associated with lipid metabolism: “ether lipid metabolism”, “phospholipid degradation” and “glycerophospholipid metabolism”. It is known that lipid peroxidation is one of the manifestations of the toxicity of cadmium. For example, in tomatoes, it has been shown that cadmium induces substantial changes in lipid composition, causing premature aging of leaves (Djebali et al., 2005).

## Conclusion

The molecular genetic mechanisms of potato resistance to unfavourable conditions remain rather poorly understood. The accumulated knowledge about the *A. thaliana* model plant can shed light on the molecular interactions in the gene networks involved in the response of potatoes to stress conditions. This approach is also used in studies by other researchers. For example, Ž. Ramšak and colleagues (2018) built a model of immune signaling pathways in *A. thaliana*, then superimposed the data obtained onto *S. tuberosum* to gain new knowledge about the immune signaling pathways in potatoes.

The Gene Ontology database (Gene Ontology Consortium, 2019) contains information on biological processes and the genes involved in their functioning. Using Gene Ontology terms, molecular genetic mechanisms involved in the breeding and important traits of plants, knowledge of which is necessary for the development of modern approaches for marker-oriented and genomic selection, can be described. However, the analysis of gene networks built by analysing literature sources containing data on the relationship between genes and the studied processes is also of great interest.

Analysis of the prioritization of biological processes that may be involved in the response of plants to unfavourable factors (cadmium, salinity and drought conditions), performed

using the reconstruction of associative genes, showed good agreement with the well-known literature data. Among the biological processes that received high priority were fundamental processes related to the expression of genes, post-translational modification of proteins and degradation, as well as processes associated with cell death. The various ways to transmit signals occupies an important role in the mechanisms of plants' responses to adverse factors.

When comparing the biological processes identified as playing a role in responses to unfavourable conditions, in addition to finding common processes associated with all three of the unfavourable factors in the external environment, we also identified processes that were statistically significantly connected with only one of the associative networks studied. Among the biological processes specifically involved in response to drought is a rather large group associated with the development of seeds (for example, "seeding development" and "seed dormancy"). Among the processes involved specifically in the response to saline stress are a group of processes associated with ion transport (for example, "ion transport", "transmembrane proton transport" and "sodium ion transmembrane transport"). Finally, processes involved in lipid metabolism (such as "phospholipid degradation") were specific to the plant response to cadmium.

## References

- 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., Gulbrandsen 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 phenome-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.
- 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 Biol.* 2012;11(3):149-161. DOI 10.3233/ISB-2012-0449.
- 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 Selekcii = Vavilov Journal of Genetics and Breeding.* 2019; 23(3):312-319. DOI 10.18699/VJ19.501.
- Djebali W., Zarrouk 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* responses 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 *SOCl* 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., Demenkov 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 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 Selekcii = Vavilov Journal of Genetics and Breeding.* 2018;22(1): 8-17. DOI 10.18699/VJ18.325. (in Russian)
- Ivanisenko V.A., Demenkov 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., 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.

- 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/tj.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/ppl.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., Boroevich 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 (CDK) 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., Guillems T., Latimer J., McNamee C., Norris A., Sanseu 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., Sreeja 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., Demenkov 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., 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.
- Saik O.V., Nimaev V.V., Usmonov D.B., Demenkov 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  $\beta$  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

**Acknowledgements.** Reconstruction and analysis of associative gene networks were carried out within the framework of the ICG SB RAS project "Creation of new competitive potato varieties using marker-based and genomic selection methods" (AAAA-A20-120070990014-7) within the framework of the KPNI "Development of potato breeding and seed production". The development of prioritization methods was carried out at the expense of the budget projects 0259-2021-0009 and 0259-2021-0012.

**Conflict of interest.** The authors declare no conflict of interest.

Received December 14, 2020. Revised June 21, 2021. Accepted June 21, 2021.