Вавиловский журнал генетики и селекции. 2018;22(1):8-17
DOI 10.18699/VJ18.325

База знаний SOLANUM TUBEROSUM: раздел по молекулярно-генетической регуляции метаболических путей

Т.В. Иванисенко[1], О.В. Сайк[1], П.С. Деменков[1], В.К. Хлесткин[1, 2], Е.К. Хлесткина[1, 2], Н.А. Колчанов[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

Быстрое развитие высокопроизводительных геномных, транскриптомных, протеомных и метаболомных технологий обусловило эффект «информационного взрыва» в области биологии растений. На сегодняшний день число научных публикаций только по одной важнейшей сельскохозяйственной культуре *Solanum tuberosum* L. (картофель) превышает 1.5 млн. Эффективный доступ к знаниям, распределенным по такому множеству неформализованных естественно-языковых текстовых источников, требует применения специальных компьютерных интеллектуальных методов анализа текстов (text mining). Однако в литературе нет данных о широком использовании интеллектуальных методов автоматического извлечения знаний из научных публикаций по сельскохозяйственным культурам, таким как картофель. Ранее нами была разработана пилотная версия базы знаний SOLANUM TUBEROSUM, представляющая собой компьютерную платформу для комплексной интеллектуальной обработки больших данных, включая: 1) автоматический анализ текстов научных публикаций и фактографических баз данных, направленный на экстракцию информации по генетике, маркерам, селекции, семеноводству, диагностике возбудителей заболеваний, средствам защиты и технологиям хранения картофеля; 2) формализованное представление извлеченной информации в базе знаний; 3) пользовательский доступ к этим данным; 4) анализ и визуализацию результатов запросов. В онтологии базы знаний SOLANUM TUBEROSUM представлены словари молекулярно-генетических объектов (белков, генов, метаболитов, микроРНК, биомаркеров и др.), сортов картофеля и их фенотипических признаков, болезней и вредителей картофеля, биотических и абиотических факторов окружающей среды, агробиотехнологий возделывания, а также технологий переработки и хранения картофеля. В статье дано описание текущей версии базы знаний SOLANUM TUBEROSUM, полученной в результате расширенного анализа научных публикаций по молекулярно-генетической регуляции метаболических путей у картофеля, а также модельных растительных организмов (кукурузы, риса, арабидопсиса). Всего было проанализировано около 9000 полнотекстовых статей и более 130000 рефератов PubMed. С помощью автоматического анализа текстов научных публикаций выявлено более 59000 фактов о молекулярно-генетических взаимодействиях и генетической регуляции, а анализ фактографических баз данных позволил выявить более 380000 таких взаимодействий у рассмотренных организмов. При этом оказалось, что к *Solanum tuberosum* L. относится около 3 % экстрагированных фактов о молекулярно-генетических взаимодействиях и генетической регуляции. Таким образом, включение сведений о хорошо изученных модельных видах при извлечении информации о молекулярно-генетической регуляции метаболических про-

The SOLANUM TUBEROSUM knowledge base: the section on molecular-genetic regulation of metabolic pathways

T.V. Ivanisenko[1], O.V. Saik[1], P.S. Demenkov[1], V.K. Khlestkin[1, 2], E.K. Khlestkina[1, 2], N.A. Kolchanov[1], V.A. Ivanisenko[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia

Rapid development of high-performance genomic, transcriptomic, proteomic and metabolic technologies led to an information explosion in the field of plant biology and agrobiology. To date, the number of scientific publications on only one of the most important agricultural crops of *Solanum tuberosum* L. (potato) has exceeded 1.5 million. Effective access to knowledge distributed over such a multitude of non-formalized natural language textual sources requires the use of special computer-assisted intelligent methods of data mining (text-mining). However, in the literature, there is no data on the application of intellectual methods of automatic knowledge extraction from publications on agricultural crops, such as potato. Previously we have developed a pilot version of the SOLANUM TUBEROSUM knowledge base. SOLANUM TUBEROSUM is a computer platform for complex intellectual processing of large data bodies, including (1) automatic analysis of scientific publications and databases for extraction of information on genetics, markers, breeding, diagnostics, protection and storage technologies for potato, (2) formalized representation of extracted information in the knowledge base, (3) user access to these data, (4) analysis and visualization of query results. The ontology of the SOLANUM TUBEROSUM knowledge base contains dictionaries of molecular genetic objects (proteins, genes, metabolites, microRNAs, biomarkers); phenotypic characteristics of potato varieties; potato diseases and pests; biotic/abiotic environmental factors; potato agrobiotechnologies. This article describes the current version of the SOLANUM TUBEROSUM knowledge base developed from an extensive analysis of scientific publications on the molecular-genetic regulation of metabolic pathways in potatoes, as well as model plant organisms (maize, rice, *Arabidopsis thaliana*). In total, about 9,000 full-text articles and more than 130,000 abstracts of PubMed were analyzed. With the help of automatic analysis of scientific publications, more than 59,000 facts on molecular genetic interactions and genetic regulation were identified, and the analysis of

Received 20.11.2017
Accepted for publication 11.01.2017
© AUTHORS, 2018

e-mail: salix@bionet.nsc.ru

цессов является важным и позволит предсказывать гены-ортологи у картофеля и проводить их дальнейшую идентификацию и выделение на основе гомологии. Сконструирована ассоциативная сеть генетической регуляции биосинтеза крахмала у картофеля, включающая 33 метаболита, 36 белков, 6 метаболических путей и 132 взаимодействия между ними, 86 из которых описывают каталитические реакции, а остальные – регуляторные события. Сконструированная сеть является основой для поиска генов-мишеней для направленного мутагенеза и маркер-ориентированной селекции сортов картофеля с заданными свойствами крахмала. Тестовая версия базы знаний SOLANUM TUBEROSUM доступна по адресу http://www-bionet.sysbio.cytogen.ru/and/plant/.

Ключевые слова: картофель; *Solanum tuberosum* L.; база знаний; ANDSystem; биосинтез крахмала; ассоциативные генные сети; генетическая регуляция.

factual databases revealed more than 380,000 such interactions in the examined organisms. It turned out that about 3 % of extracted facts about molecular genetic interactions and genetic regulation were related to *Solanum tuberosum* L. Thus, the inclusion of information on well-studied model species during the extraction of information on the molecular-genetic regulation of metabolic processes is important. It allows prediction of orthologous genes in potato and their further identification and analysis based on homology. An associative network of genetic regulation of starch biosynthesis in potatoes, including 33 metabolites, 36 proteins, 6 metabolic pathways and 132 interactions between them, 86 of which describe catalytic reactions, and the rest – regulatory events, was reconstructed. The reconstructed network is the basis for the search for target genes for directed mutagenesis and marker-oriented selection of potato varieties with specified starch properties. The trial version of the SOLANUM TUBEROSUM knowledge base is available at http://www-bionet.sysbio.cytogen.ru/and/plant/.

Key words: potato; *Solanum tuberosum* L.; knowledge base; ANDSystem; starch biosynthesis; associative gene networks; genetic regulation.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

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

HOW TO CITE THIS ARTICLE:

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. DOI 10.18699/VJ18.325 (in Russian)

Картофель (*Solanum tuberosum* L., сем. Пасленовые) представляет собой важнейшую сельскохозяйственную культуру, имеющую высокую пищевую, кормовую и техническую ценность. Пищевая значимость его во многом обусловлена высоким содержанием углеводов (главным образом крахмала), хорошей усвояемостью белков картофеля, значительным содержанием аскорбиновой кислоты, солей калия, кальция, магния, других микроэлементов. Картофельный крахмал служит сырьем для спиртовой и крахмало-паточной промышленности, из него производят декстрины, глюкозу, мальтозу для пищевой индустрии и ряд полупродуктов для химической промышленности (Khlestkin et al., 2018). Кроме того, крахмал клубней картофеля широко используется в бумажной, текстильной промышленности и других отраслях (Kraak, 1992; Ellis et al., 1998; Jobling, 2004).

В настоящее время, в связи со стремительным развитием высокопроизводительных геномных, транскриптомных, протеомных и метаболомных технологий, а также новых технологий в сельском хозяйстве, наблюдается «информационный взрыв» в биологии растений и растениеводстве, в том числе в картофелеводстве. Однако только небольшая часть данных, полученных в результате научных исследований, представлена в формализованном виде в фактографических базах, таких как GeneBank, Uniprot, IntAct, BioGRID и т. д. Основная же часть продуцируемых знаний описана в миллионах научных статей, представляющих собой так называемые неструктуированные текстовые данные на естественном языке. В частности, количество научных публикаций только по *Solanum tuberosum* L. превышает полтора миллиона. Такая разобщенность информации затрудняет установление связей и корреляций между наборами данных, описывающих практически полезные и важные свойства растения, его строение и процессы на молекулярном уровне, например сведения о картофельном крахмале – от генетической регуляции его биосинтеза до применения в различных отраслях промышленности (Хлесткин и др., 2017). Таким образом, снижается эффективность использования полученных данных, растет число упущенных возможностей рационального использования генетического и метаболического потенциала природных ресурсов.

В современном мире проблема обработки больших объемов текстовой информации играет огромную роль в самых разных областях жизни человека. Это способствовало развитию компьютерных интеллектуальных методов автоматического анализа текстов (text mining) (Kilicoglu, 2017). Все методы автоматического анализа текстов можно разделить на две большие группы: методы, основанные на правилах, и методы, использующие машинное обучение. Методы, основанные на правилах, позволяют достичь высокой точности извлечения информации, однако имеют относительно низкие значения полноты извлечения (Aggarwal, Zhai, 2012). Альтернативным подходом к автоматическому извлечению информации, не требующим применения вручную созданных правил, являются методы машинного обучения, которые получили широкое применение в последние годы. Среди таких методов часто используются наивный байесовский классификатор, деревья решений, метод условных случайных полей (conditional random fields – CRF) (Uzuner et al., 2011) и структурированные опорные вектора (structured support vector machines – SSVM) (Tang et al., 2013), а также методы глубокого обучения, основанные на нейронных сетях (Collobert

et al., 2011). К недостаткам этих методов можно отнести требование наличия больших обучающих выборок, содержащих размеченные тексты.

Методы автоматического анализа текстов нашли наиболее широкое применение при решении различных задач биомедицины, системной и интегративной биологии (Rebholz-Schuhmann et al., 2012), включая поддержку клинических решений (Friedman et al., 1999; Cao et al., 2011), курирование биологических/биомедицинских баз данных (Wei et al., 2013), инспекцию фармпрепаратов (Sarker et al., 2015) и др. В качестве источника текстовых данных использовались рефераты, полнотекстовые статьи и патенты (Shetty, Dalal, 2011; Li et al., 2013), а также электронные карточки пациентов (Meystre et al., 2008), текстовые данные в социальных сетях (Sarker et al., 2015) и др.

В области биологии растений интеллектуальные методы автоматического извлечения знаний широко применялись только для анализа статей по модельным организмам. Например, web-доступная система PLAN2L (Krallinger et al., 2009) содержит результаты автоматического извлечения информации из полнотекстовых публикаций по *Arabidopsis thaliana* о белок-белковых взаимодействиях и генетической регуляции, а также ассоциациях генов с некоторыми клеточными процессами и процессами развития (цветка, корня и т. д.).

Ранее нами впервые в мире была разработана компьютерная платформа для комплексного интеллектуального анализа научных публикаций в области картофелеводства – база знаний SOLANUM TUBEROSUM (Сайк и др., 2017). Программные средства этой платформы обеспечивают решение всех необходимых шагов для автоматической экстракции и формализованного представления в базе знаний информации по генетике, ДНК-маркерам, селекции, семеноводству, диагностике возбудителей заболеваний, средствам защиты и технологиям хранения картофеля. Графический пользовательский интерфейс SOLANUM TUBEROSUM позволяет осуществлять доступ к этим данным, проводить анализ и визуализацию результатов запросов.

Для автоматического анализа текстов осуществлена адаптация и интеграция модуля text-mining программно-информационной системы ANDSystem (Demenkov et al., 2012; Ivanisenko et al., 2015; Saik et al., 2016a), предназначенной для экстракции медико-биологических знаний из научных публикаций с помощью синтаксико-семантических правил. ANDSystem использовалась при проведении анализа данных высокопроизводительных протеомных экспериментов в области биомедицины (Momynaliev et al., 2010; Пастушкова и др., 2015а, б; Larina et al., 2015; Pastushkova et al., 2015), а также при анализе тканеспецифического эффекта нокаута генов и поиске потенциальных мишеней для лекарственных препаратов (Petrovskiy et al., 2015). С помощью ANDSystem были выявлены новые регуляторные молекулярно-генетические механизмы коморбидных взаимоотношений между различными патологиями человека (Bragina et al., 2014, 2016; Glotov et al., 2015; Saik et al., 2016b), а также показано, что данная система пригодна для выявления молекулярно-генетических механизмов жизненного цикла патогенов (Popik et al., 2015).

Одной из важнейших задач развития созданной базы знаний SOLANUM TUBEROSUM являются автоматическая экстракция и формализованное представление в базе знаний информации, касающейся молекулярно-генетических механизмов, лежащих в основе хозяйственно ценных признаков, генетической регуляции метаболических путей. Интеграция этих сведений должна способствовать ускорению идентификации генов-кандидатов для селекционно значимых характеристик картофеля и разработки диагностических маркеров для селекции. К числу селекционно значимых характеристик картофеля относятся свойства крахмала, влияющие на пригодность сортов картофеля к переработке. Оценка большинства характеристик крахмала является трудоемкой задачей, поэтому использование на ранних этапах селекционного процесса диагностических ДНК-маркеров, связанных с различными свойствами крахмала, способствовало бы ускорению и удешевлению процесса получения сортов, пригодных к переработке.

Целью настоящей работы было извлечение данных из расширенного анализа научных публикаций по молекулярно-генетической регуляции у картофеля и построение на основе полученных данных ассоциативной семантической сети молекулярно-генетической регуляции метаболизма крахмала. Параллельно с информацией по картофелю извлекалась аналогичная информация по более изученным видам растений (кукуруза, рис, арабидопсис), так как сведения об этих хорошо изученных с молекулярно-генетической точки зрения видах растений будут полезны для дальнейших экспериментальных исследований, направленных на выделение новых генов-кандидатов хозяйственно ценных признаков картофеля. Разрабатываемая база знаний SOLANUM TUBEROSUM может быть полезна широкому кругу специалистов в области биологии растений, в том числе генетикам, селекционерам, фитопатологам, биоинформатикам и др.

Материалы и методы

Подробное описание структуры базы знаний SOLANUM TUBEROSUM (http://www-bionet.sysbio.cytogen.ru/and/plant/) приведено в работе (Сайк и др., 2017). Она включает три основных модуля, представленных на рис. 1.

Модуль автоматического анализа текстов (text mining) научных публикаций и фактографических баз данных предназначен для автоматической экстракции информации о взаимоотношениях между объектами согласно онтологии базы знаний. Этот модуль использует программные средства ANDSystem (Ivanisenko et al., 2015), позволяющие автоматически производить всю необходимую предобработку текстовых данных, включая преобразование исходного текста в понимаемый формат, разделение на предложения, нормализацию, морфологический и синтаксический анализ, разметку поименованных сущностей (имен объектов). Для разметки поименованных сущностей в ANDSystem реализован комплексный алгоритм на основе словарей, также осуществляющий разрешение проблем синонимии, омонимии, анафории и кореферентных ссылок. Необходимость решения перечисленных проблем связана с особенностями естественного языка. Например, авторы часто используют неточное написание

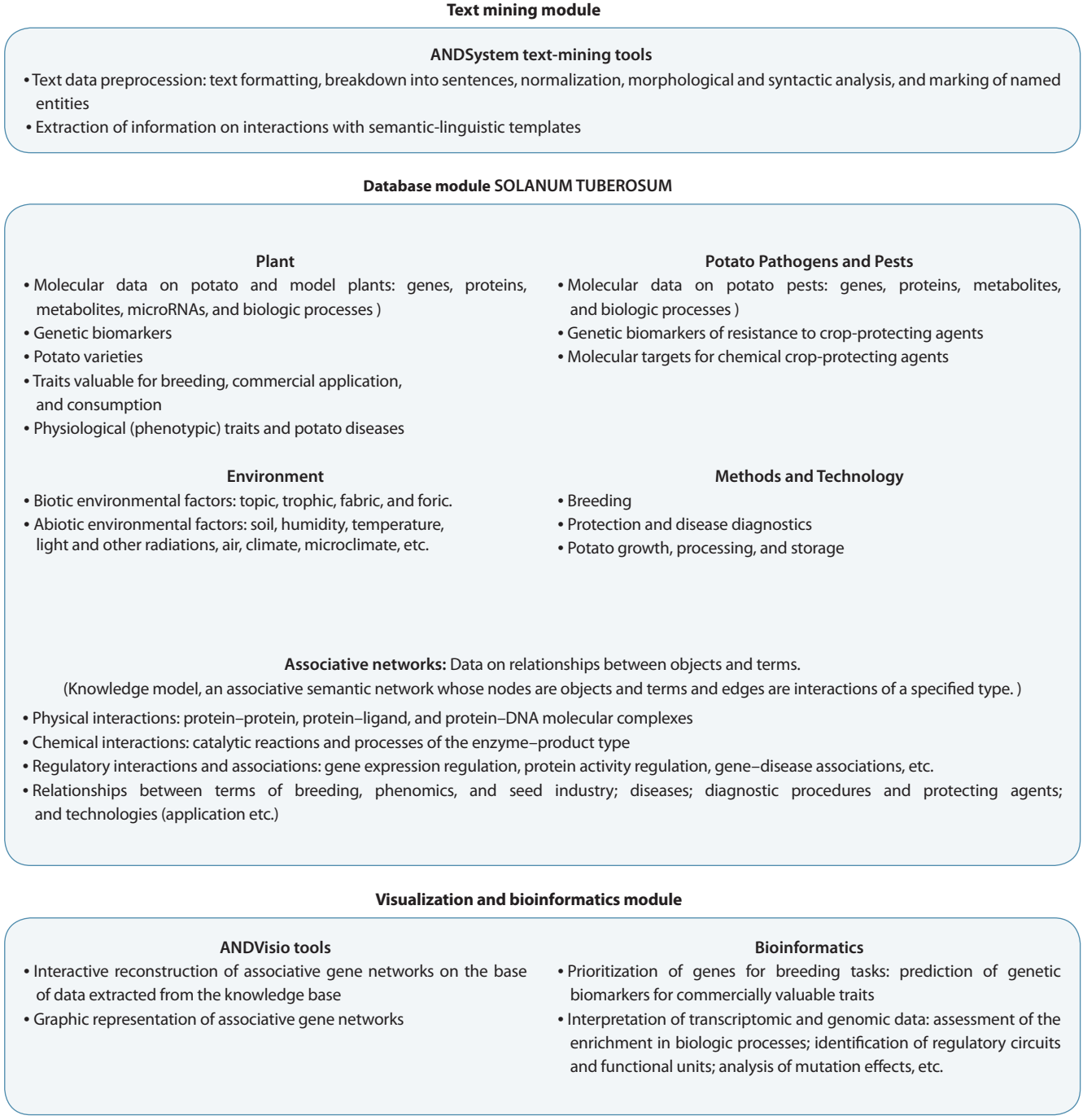
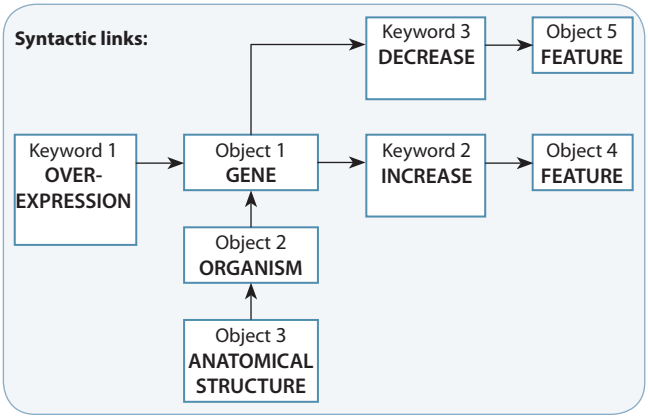


Fig. 1. Modules of the information platform «SOLANUM TUBEROSUM» for integrated intellectual processing of huge bodies of scientific data.

названий генов или других объектов по сравнению с их общепринятыми именами (перестановки, добавления или удаления слов и т. д.). Кроме того, имена разных объектов часто совпадают друг с другом, особенно в случае коротких сокращений. Эти трудности преодолеваются путем расширения числа синонимов и оценки связи имени с контекстом статьи. Стилистические приемы, такие как анафоры, эпифоры, кореферентные ссылки, ведут к неявному упоминанию в предложениях имен объектов, что представляет другую проблему для их распознавания. Решение данной проблемы в ANDSystem осуществляется в пределах только отдельно взятых предложений с помощью специальных лингвистических правил.

Подготовленные таким образом тексты далее используются для извлечения информации о взаимодействиях между объектами с помощью семантико-лингвистических шаблонов. Информация о выявленных взаимодействиях классифицируется по организмам также с применением специальных шаблонов, распознающих имя организма в анализируемом тексте. Следует отметить, что настоящая версия ANDSystem настроена на анализ только текстов на английском языке.



Semantic links:

Object 1 name from the **GENE** glossary
Object 2 name from the **ORGANISM** glossary
Object 3 name from the anatomical structure glossary
ANATOMICAL STRUCTURE
Object 4 name from the **FEATURE** glossary
Object 5 name from the **FEATURE** glossary
Keyword 1 keywords from the **OVER-EXPRESSION** glossary
Keyword 2 keywords from the **OVER-EXPRESSION** glossary
Keyword 3 keywords from the **DECREASE** glossary

Interactions:
Interaction 1: <**Object 1**>-INCREASE-<**Object 4**>
Interaction type 1: **Up-regulation**
Attribute 1 of interaction 1: <Object 2>
Attribute 2 of interaction 1: <Object 3>
Interaction 2: <**Object 1**>-DECREASE-<**Object 5**>
Interaction type 2: **Down-pregulation**
Attribute 1 of interaction 2: <Object 2>
Attribute 2 of interaction 2: <Object 3>

Fig. 2. An exemplary semantic-linguistic template identifying participants and attributes of an event of starch and glucose regulation in potato tubers in a text.

Sentence:
Here, we show that ***over-expression*** of **SnRK1** in **potato tubers** causes a significant ***increase*** in **starch content** and a ***decrease*** in **glucose levels**

Interaction 1:
Interaction type 1: **Up-regulation**
Participant Object 1: **SnRK1 | sucrose non-fermenting-1-related protein kinase-1**
Object type 1: **GENE**
Participant Object 2: **starch content**
Object type 2: **FEATURE**
Attribute 1 of interaction 1: Organism = **potato | S. tuberosum**
Attribute 2 of interaction 1: Anatomical structure = **tuber**

Interaction 2:
Interaction type 1: **Down-regulation**
Participant Object 1: **SnRK1 | sucrose non-fermenting-1-related protein kinase-1**
Object type 1: **GENE**
Participant Object 2: **glucose level**
Object type 2: **FEATURE**
Attribute 1 of interaction 1: Organism = **potato | S. tuberosum**
Attribute 2 of interaction 1: Anatomical structure = **tuber**

Fig. 3. An exemplary computer output describing the result of applying a semantic-linguistic template for identification of participants and attributes of an event of starch and glucose regulation in potato tubers.

Семантико-лингвистические шаблоны ANDSystem имеют сложную структуру, в которой можно выделить две отдельные части: синтаксические связи в предложении между именами объектов и ключевыми словами, а также семантические связи между объектами (рис. 2). Семантические связи содержат всю необходимую информацию о взаимодействиях, в которых участвуют объекты.

На рис. 3 приведен пример результата автоматического анализа предложения «Here, we show that over-expression of SnRK1 in potato tubers causes a significant increase in starch content and a decrease in glucose levels, resulting from a dramatic increase in the level of expression and activity of two key enzymes involved in the starch biosynthetic pathway: sucrose synthase and ADP-glucose pyrophosphorylase», взятого из работы (McKibbin et al., 2006), с помощью семантико-лингвистического шаблона, предназначенного для экстракции информации о событиях генетической регуляции содержания крахмала и глюкозы в клубнях картофеля (см. рис. 2). Выявленные взаимодействия показали, что ген «sucrose non-fermenting-1-related protein kinase-1 (SnRK1)» участвует в двух взаимодействиях: 1) положительная регуляция содержания крахмала «starch content» и 2) отрицательная регуляция содержания глюкозы «glucose level». При этом для каждого из взаимодействий были определены два атрибута, характеризующие организм (Organism = potato | S. tuberosum) и анатомическую структуру (Anatomical structure = tuber).

На следующем шаге анализа информация, экстрагированная с помощью методов автоматического анализа текстов, автоматически помещается в **базу данных**, которая является центральным модулем программно-информационной системы SOLANUM TUBEROSUM. В структуре базы данных можно выделить два больших раздела: 1) словари объектов и понятий (Dictionary); 2) раздел, содержащий информацию о взаимодействиях этих объектов и понятий между собой (Associative networks). Раздел Dictionary, в свою очередь, состоит из четырех больших подразделов: 1) растения (Plant); 2) патогены и заболевания картофеля (Potato Pathogens and Pests); 3) факторы окружающей среды (Environment); 4) методы и технологии (Methods and Technology). Подраздел Plant содержит словари молекулярно-генетических объектов (белки, гены, метаболиты, микроРНК, биологические процессы, биомаркеры и др.), генетических биомаркеров, сортов картофеля, селекционно значимых, хозяйственно ценных и потребительских свойств, а также фенотипических признаков и болезней картофеля. В Potato Pathogens and Pests содержатся словари возбудителей заболеваний и вредителей картофеля, а также молекулярно-генетические данные по этим организмам (гены, белки, метаболиты, биологические процессы). Кроме того, подраздел включает словари по генетическим биомаркерам резистентности у этих организмов к средствам защиты растений, а также

Table 1. Numbers of publications (full-text papers and PubMed abstracts) on various crops used for automated analysis in populating the SOLANUM TUBEROSUM database

Journal*	*Solanum tuberosum*/potato	*Arabidopsis thaliana*/Arabidopsis	*Zea mays*/maize	*Oryza sativa*/rice
Plant Biotechnol. J.	41	174	124	245
Plant J.	110	3387	420	627
Plant Mol. Biol.	171	1300	665	719
Mol. Biol. Rep.	12	102	34	80
BMC Genomics	30	272	154	330
Total number of full-text articles	364	5235	1397	2001
Total number of PubMed abstracts**	7549	44285	38030	44845

* Journals from which full-text papers were analyzed.
** The analysis involved abstracts from all journals annotated by PubMed in which relevant organisms were mentioned.

известные молекулярные мишени для химических средств защиты растений. В Environment представлены словари биотических и абиотических факторов окружающей среды. Подраздел Methods and Technology предназначен для хранения словарей по методам селекции, методам защиты картофеля и диагностики заболеваний, а также технологиям возделывания, переработки и хранения картофеля.

В разделе Associative networks сосредоточены знания, экстрагированные из текстов научных публикаций и фактографических баз данных. В качестве модели знаний используется ассоциативная семантическая сеть, вершинами которой являются объекты и понятия, а ребрами – взаимодействия заданного типа. В онтологии базы знаний представлено более 25 различных типов связей, описывающих молекулярные взаимодействия, а также взаимоотношения между понятиями селекции, феномики и семеноводства, заболеваниями, приемами диагностики и средствами защиты, технологиями. Связи подразделяются на физические межмолекулярные взаимодействия, химические взаимодействия по типу субстрат–фермент–продукт, посттрансляционные модификации белков (фосфорилирование, гликозилирование и т. д.), широкий круг регуляторных взаимодействий, включая регуляцию экспрессии генов транскрипционными факторами, регуляцию активности и стабильности белков и т. д., ассоциативные связи. Взаимоотношения между понятиями селекции, феномики и семеноводства, заболеваниями, диагностиками и средствами защиты, технологиями определяются различного типа регуляторными связями (положительная и отрицательная регуляция), а также такими связями, как применение, назначение и др.

Модуль визуализации и биоинформатики предназначен для работы пользователей с базой знаний. Программа ANDVisio служит интерфейсом для реализации пользовательских запросов к базе данных SOLANUM TUBEROSUM, а также обеспечивает графическую визуализацию ассоциативных генных сетей, построенных на основе результатов этих запросов. ANDVisio была разработана нами ранее как модуль визуализации для системы ANDSystem (Demenkov et al., 2012).

Биоинформатические методы анализа предназначены для решения широкого круга задач по приоритизации генов, предсказанию маркеров, планированию экспериментов и др. на основе информации, представленной в базе данных SOLANUM TUBEROSUM. В частности, проведена интеграция известного пакета программ GUILD (http://sbi.imim.es/web/index.php/research/software/guildsoftware) для решения задач приоритизации на основе анализа структуры графа генных сетей (Guney, Oliva, 2012). Еще один класс биоинформатических методов, реализованный в SOLANUM TUBEROSUM, основан на оценках обогащенности биологических процессов генами, идентифицированными в эксперименте (например, при транскриптомном анализе). Такие методы широко применяются в известных компьютерных системах, предназначенных для интерпретации экспериментальных данных, например DAVID (Huang et al., 2008), PANTHER (Thomas et al., 2006; Mi et al., 2015), GORILLA (Eden et al., 2007, 2009) и др.

Результаты и обсуждение

Статистика базы знаний по теме «генетическая регуляция»

В настоящей версии базы знаний осуществлено обновление информации, относящейся к молекулярно-генетической регуляции метаболических процессов у *S. tuberosum*, а также аналогичной информации у более изученных видов растений (*Arabidopsis thaliana*, *Zea mays* и *Oryza sativa*) – в качестве источника недостающих у картофеля сведений. Внесение в базу знаний SOLANUM TUBEROSUM (http://www-bionet.sysbio.cytogen.ru/and/plant/) дополнительных данных о молекулярно-генетической регуляции универсальных метаболических путей за счет анализа информации по модельным видам позволит предсказать гены-ортологи у картофеля и проводить их дальнейшую идентификацию и выделение на основе гомологии. Всего при решении этой задачи было проанализировано более 130000 рефератов статей и более 8000 полнотекстовых статей для указанных четырех видов растений (табл. 1). Суммарная статистика базы знаний SOLANUM TUBEROSUM по теме «генетическая регуляция» показана в табл. 2.

Автоматический анализ текстов научных публикаций позволил выявить более 59000 фактов о молекулярно-генетических взаимодействиях и генетической регуляции, при этом около 3 % фактов относилось к *S. tuberosum*. Это подтверждает значимость включения сведений о хорошо изученных

Table 2. Statistics of the SOLANUM TUBEROSUM database for the Genetic Regulation topic

Interaction type	*Solanum tuberosum*/potato	*Arabidopsis thaliana*/Arabidopsis	*Zea mays*/maize	*Oryza sativa*/rice
Association	1405	44542	2577	6917
Catalyze	4450	109336	4266	27952
Regulation	120	3125	221	497
Interaction	15	30257	53	160
Involvement	3276	112999	5124	31316
Number of genes	32995	38125	789	32638
Number of proteins	412	15288	53293	3755

модельных видах при извлечении информации о молекулярно-генетической регуляции метаболических процессов.

Дополнительно к данным, извлеченным из литературы, информация о молекулярно-генетических взаимодействиях была дополнена данными из фактографических баз (всего более 380000 взаимодействий). В частности, из базы данных BioGRID (Stark et al., 2006) была экстрагирована информация о белок-белковых взаимодействиях, из базы AmiGO 2 (Carbon et al., 2009) – информация о вовлеченности генов в биологические процессы Gene Ontology, а из базы KEGG (Kanehisa, Goto, 2000) – участие белков в каталитических реакциях. Интересно, что соотношение количества молекулярно-генетических взаимодействий, представленных в этих базах данных по картофелю и трем модельным организмам, оказалось примерно равным аналогичному соотношению, найденному при анализе данных в научной литературе (около 3 %).

Ассоциативная семантическая сеть молекулярно-генетической регуляции метаболизма крахмала

Гранулы крахмала в клубнях картофеля состоят из двух полисахаридов – амилозы и амилопектина. Оба соединения являются полимерами глюкозы, но различаются, кроме всего прочего, по молекулярному весу и топологии полимерных цепей. Амилоза представляет собой линейную малоразветвленную полимерную цепь остатков α-глюкозы, соединенных между собой (1→4) гликозидными связями, и имеет молекулярную массу 10^5–10^6 а.е.м. Амилопектин состоит из разветвленных цепочек остатков α-глюкозы, соединенных как (1→4), так и (1→6) гликозидными связями, и имеет молекулярную массу 10^7–10^9 а.е.м. Биосинтез крахмала в клубнях картофеля включает на первом этапе превращение сахарозы в глюкозо-6-фосфат в цитоплазме через серию химических реакций, катализируемых ферментами сахарозосинтаза (sucrose synthase, EC 2.4.1.13), УДФ-глюкопирофосфорилаза (UDP-glucose pyrophosphorylase, EC 2.7.7.9) и фосфоглюкомутаза (PGM – phosphoglucomutase, EC 5.4.2.2). Далее глюкозо-6-фосфат/фосфатный транслокационный белок (glucose-6-phosphate/phosphate translocator protein) переносит глюкозо-6-фосфат в амилопласт, где он превращается в глюкозо-1-фосфат при участии пластидной фосфоглюкомутазы (phosphoglucomutase, EC 5.4.2.2). Затем аденозиндифосфат-глюкоза-пирофосфорилаза (ADP-glucose pyrophosphorylase, EC 2.7.7.27) катализирует превращение глюкозо-1-фосфата и АТФ в АДФ-глюкозу и неорганический пирофосфат. АДФ-глюкоза выступает как донор активированной глюкозы для различных крахмалсинтаз (starch synthases, EC 2.4.1.21), которые продуцируют амилозу. Крахмалсинтаза, связанная с крахмальными гранулами (granule-bound starch synthase), ответственна за синтез амилозы, тогда как растворимые крахмалсинтазы I–III отвечают за синтез амилопектина. Ветвление амилопектина осуществляет ветвящий фермент SBE (starch-branching enzyme, EC 2.4.1.18), который отщепляет от неразветвленной цепочки фрагмент и переносит его к шестому атому углерода глюкозы (Zhang et al., 2017). Путь биосинтеза крахмала описан во многих источниках, в то время как регуляция этого процесса до сих пор остается малоизученной (Van Harsselaar et al., 2017).

На рис. 4 представлена ассоциативная сеть, описывающая путь биосинтеза крахмала и генетическую регуляцию этого процесса, построенная с помощью базы знаний SOLANUM TUBEROSUM. Ассоциативная сеть включает 33 метаболита, 36 белков, 6 биологических процессов Gene Ontology и 132 взаимодействия между ними, 86 из которых описывают каталитические реакции, а остальные – регуляторные события. Так, например, на сети показана активация аденозиндифосфат-глюкоза-пирофосфорилазы белком NADP-зависимой тиоредоксинредуктазой C (NTR3), которая приводит к накоплению крахмала (Jenner et al., 2001; Michalska et al., 2009; Geigenberger, 2011). Известно, что экспрессия генов *LOB*, *TIFY5a* и *WRKY4* положительно коррелирует с экспрессией генов *SuSy4* и *GPT2.1* (Van Harsselaar et al., 2017). На рис. 4 показана также положительная регуляция экспрессии генов сахарозосинтазы и аденозиндифосфат-глюкоза-пирофосфорилазы белком SnRK1 (Purcell et al., 1998; Slocombe et al., 2002; McKibbin et al., 2006) и сахарозой (Salanoubat, Belliard, 1989; Müller-Röber et al., 1990).

Еще одним примером регуляторных взаимодействий может служить регуляция экспрессии гена сахарозосинтазы *Sus1* маннитолом, который имитирует эффект осмотического стресса у растений (Déjardin et al., 1999), или активация экспрессии гена *PHS1* абсцизовой кислотой (Quettier et al., 2006). Другим метаболитом, вовлеченным в регуляцию метаболизма крахмала, является ауксин. Ауксин способен регулировать целый ряд биологических процессов, включая фотосинтез (Xing, Xue, 2012), органогенез (Furutani et al., 2007), старение (Zhu, Davies, 1997), развитие пыльцы (Ni et al., 2002), морфогенез корня и формирование проростков (Ljung et al., 2005), а

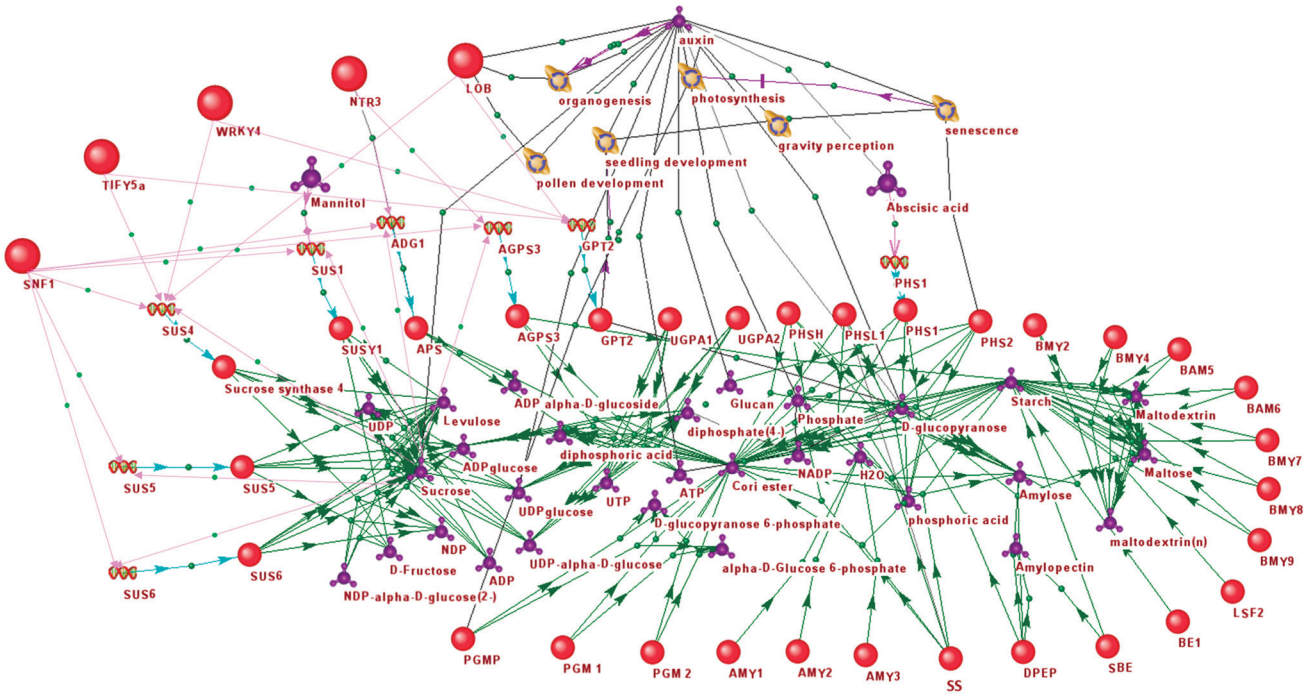


Fig. 4. Associative network of the genetic regulation of starch biosynthesis.

также геотропизм растений (Friml et al., 2002). В работах (Obata-Sasamoto, Suzuki, 1979; Roumeliotis et al., 2012) показано, что уровень ауксина резко возрастает в столоне перед формированием клубней и остается относительно высоким на начальных стадиях роста клубней, а затем снижается, что указывает на стимулирующую роль ауксина в образовании клубней. Исследования (Miyazawa et al., 1999) показывают, что в табаке *Nicotiana tabacum* снижение уровня ауксина приводит к накоплению крахмала. Известно, что экспрессия транскрипционного фактора LOB зависит от паттернов ауксина (Guilfoyle, Hagen, 2007), однако механизм этого взаимодействия остается не до конца изученным (Lee et al., 2015). Показано, что ауксин способен индуцировать биосинтез абсцизовой кислоты (Hansen, Grossmann, 2000). В работе (Mishra et al., 2009) говорится о том, что глюкоза может влиять на гены, вовлеченные в биосинтез ауксина, белки-транспортеры ауксина, рецепторы ауксина и участников сигнального пути ауксина. Обсуждается также способность сахарозы повышать уровень ауксина (Lilley et al., 2012; Boycheva et al., 2015).

Таким образом, регуляция биосинтеза крахмала включает множество взаимодействий, как прямых, так и опосредованных. Визуализация этих взаимодействий в виде ассоциативной сети с использованием базы знаний SOLANUM TUBEROSUM позволяет полнее их учитывать в ходе построения биохимических путей превращений, описания регуляторных функций метаболитов и других исследований, тем самым повышая эффективность использования ранее полученных данных.

Заключение

Текущая версии базы знаний SOLANUM TUBEROSUM включает в себя информацию по молекулярно-генетической регуляции метаболических процессов у картофеля, а также модельных растительных организмов (кукурузы, риса, арабидопсиса), полученную в результате расширенного анализа 9000 полнотекстовых статей и более 130000 рефератов PubMed. С помощью автоматического анализа текстов научных публикаций выявлено более 59000 фактов о молекулярно-генетических взаимодействиях и генетической регуляции, а анализ фактографических баз данных позволил выявить более 380000 таких взаимодействий у рассмотренных организмов. При этом оказалось, что к *Solanum tuberosum* L. относится около 3 % экстрагированных фактов о молекулярно-генетических взаимодействиях и генетической регуляции. Таким образом, включение сведений о хорошо изученных модельных видах при извлечении информации о молекулярно-генетической регуляции метаболических процессов является важным и позволит предсказывать гены-ортологи у картофеля и проводить их дальнейшую идентификацию и выделение на основе гомологии. Сконструированная ассоциативная сеть генетической регуляции биосинтеза крахмала у картофеля служит основой для поиска генов-мишеней для направленного мутагенеза и маркер-ориентированной селекции сортов картофеля с заданными свойствами крахмала.

Acknowledgments

This work was supported by the program of the Federal Agency for Scientific Organizations “Development of potato breeding and seed production”, project 0324-2017-0062.

Conflict of interest

The authors declare no conflict of interest.

References

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

Boycheva S., Dominguez A., Rolcik J., Boller T., Fitzpatrick T.B. Consequences of a deficit in vitamin B6 biosynthesis de novo for hormone homeostasis and root development in Arabidopsis. Plant Physiol. 2015;167(1):102-117. DOI 10.1104/pp.114.247767.

Bragina E.Y., Tiys E.S., Freidin M.B., Koneva L.A., Demenkov P.S., Ivanisenko V.A., Kolchanov N.A., Puzyrev V.P. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. Immunogenetics. 2014;66(7-8):457-465. DOI 10.1007/s00251-014-0786-1.

Bragina E.Y., Tiys E.S., Rudko A.A., Ivanisenko V.A., Freidin M.B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. Infect. Genet. Evol. 2016;46:118-123. DOI 10.1016/j.meegid.2016.10.030.

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.

Carbon S., Ireland A., Mungall C.J., Shu S., Marshall B., Lewis S., AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25(2):288-289. DOI 10.1093/bioinformatics/btn615.

Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K., Kuksa P. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 2011;12:2493-2537.

Déjardin A., Sokolov L.N., Kleczkowski L.A. Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in *Arabidopsis*. Biochem. J. 1999;344(2):503-509. DOI 10.1042/bj3440503.

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.

Eden E., Lipson D., Yogev S., Yakhini Z. Discovering motifs in ranked lists of DNA sequences. PLoS Comput. Biol. 2007;3(3):e39. DOI 10.1371/journal.pcbi.0030039.

Eden E., Navon R., Steinfeld I., Lipson D., Yakhini Z. *GOrilla*: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10:48. DOI 10.1186/1471-2105-10-48.

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.

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.

Friml J., Wiśniewska J., Benková E., Mendgen K., Palme K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in *Arabidopsis*. Nature. 2002;415:806-809. DOI 10.1038/415806a.

Furutani M., Kajiwara T., Kato T., Treml B.S., Stockum C., Torres-Ruiz R.A., Tasaka M. The gene *MACCHI-BOU 4/ENHANCER OF PINOID* encodes a NPH3-like protein and reveals similarities between organogenesis and phototropism at the molecular level. Development. 2007;134(21):3849-3859. DOI 10.1242/dev.009654.

Geigenberger P. Regulation of starch biosynthesis in response to a fluctuating environment. Plant Physiol. 2011;155(4):1566-1577. DOI 10.1104/pp.110.170399.

Glotov A.S., Tiys E.S., Vashukova E.S., Pakin V.S., Demenkov P.S., Saik O.V., Ivanisenko T.V., Arzhanova O.N., Mozgovaya E.V., Zainulina M.S., Kolchanov N.A., Baranov V.S., Ivanisenko V.A. Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome). BMC Syst. Biol. 2015;9(Suppl.2):S4. DOI 10.1186/1752-0509-9-S2-S4.

Guilfoyle T.J., Hagen G. Auxin response factors. Curr. Opin. Plant Biol. 2007;10(5):453-460. DOI 10.1016/j.pbi.2007.08.014.

Guney E., Oliva B. Exploiting protein-protein interaction networks for genome-wide disease-gene prioritization. PLoS ONE. 2012;7(9): e43557. DOI 10.1371/journal.pone.0043557.

Hansen H., Grossmann K. Auxin-induced ethylene triggers abscisic acid biosynthesis and growth inhibition. Plant Physiol. 2000;124(3): 1437-1448. DOI 10.1104/pp.124.3.1437.

Huang D.W., Sherman B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2008;4(1):44-57. DOI 10.1038/nprot.2008.211.

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.

Jenner H.L., Winning B.M., Millar A.H., Tomlinson K.L., Leaver C.J., Hill S.A. NAD malic enzyme and the control of carbohydrate metabolism in potato tubers. Plant Physiol. 2001;126:1139-1149. DOI 10.1104/pp.126.3.1139.

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.

Kanehisa M., Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30. DOI 10.1093/nar/28.1.27.

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. DOI 10.15389/agrobiology.2017.1.25rus. (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.

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; 37(Suppl.2):W160-W165. DOI 10.1093/nar/gkp484.

Larina I.M., Pastushkova L.Kh., Tiys E.S., Kireev K.S., Kononikhin A.S., Starodubtseva N.L., Popov I.A., Custaud M.A., Dobrokhotov I.V., Nikolaev E.N., Kolchanov N.A., Ivanisenko V.A. Permanent proteins in the urine of healthy humans during the Mars-500 experiment. J. Bioinform. Comput. Biol. 2015;13(1):1540001. DOI 10.1142/S0219720015400016.

Lee H.W., Cho C., Kim J. *Lateral Organ Boundaries Domain16* and *18* act downstream of the AUXIN1 and LIKE-AUXIN3 auxin influx carriers to control lateral root development in Arabidopsis. Plant Physiol. 2015;168(4):1792-1806. DOI 10.1104/pp.15.00578.

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.

Lilley J.L., Gee C.W., Sairanen I., Ljung K., Nemhauser J.L. An endogenous carbon-sensing pathway triggers increased auxin flux and hypocotyl elongation. Plant Physiol. 2012;160(4):2261-2270. DOI 10.1104/pp.112.205575.

Ljung K., Hull A.K., Celenza J., Yamada M., Estelle M., Normanly J., Sandberg G. Sites and regulation of auxin biosynthesis in Arabidopsis roots. Plant Cell. 2005;17(4):1090-1104. DOI 10.1105/tpc.104.029272.

McKibbin R.S., Muttucumaru N., Paul M.J., Powers S.J., Burrell M.M., Coates S., Purcell P.C., Tiessen A., Geigenberger P., Halford N.G. Production of high-starch, low-glucose potatoes through over-expression of the metabolic regulator SnRK1. Plant Biotechnol. J. 2006;4(4):409-418. DOI 10.1111/j.1467-7652.2006.00190.x.

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.

Mi H., Poudel S., Muruganujan A., Casagrande J.T., Thomas P.D. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2015;44(D1):D336-D342. DOI 10.1093/nar/gkv1194.

Michalska J., Zauber H., Buchanan B.B., Cejudo F.J., Geigenberger P. NTRC links built-in thioredoxin to light and sucrose in regulating starch synthesis in chloroplasts and amyloplasts. Proc. Natl. Acad. Sci. USA. 2009;106:9908-9913. DOI 10.1073/pnas.0903559106.

Mishra B.S., Singh M., Aggrawal P., Laxmi A. Glucose and auxin signaling interaction in controlling *Arabidopsis thaliana* seedlings root growth and development. PLoS ONE. 2009;4(2):e4502. DOI 10.1371/journal.pone.0004502.

Miyazawa Y., Sakai A., Miyagishima S.Y., Takano H., Kawano S., Kuroiwa T. Auxin and cytokinin have opposite effects on amyloplast development and the expression of starch synthesis genes in cultured bright yellow-2 tobacco cells. Plant Physiol. 1999;121(2):461-470. DOI 10.1104/pp.121.2.461.

Momynaliev K.T., Kashin S.V., Chelysheva V.V., Selezneva O.V., Demina I.A., Serebryakova M.V., Alexeev D., Ivanisenko V.A., Aman E., Govorun V.M. Functional divergence of *Helicobacter pylori* related to early gastric cancer. J. Proteome Res. 2010;9(1):254-267. DOI 10.1021/pr900586w.

Müller-Röber B.T., Kossmann J., Hannah L.C., Willmitzer L., Sonnewald U. One of two different ADP-glucose pyrophosphorylase genes from potato responds strongly to elevated levels of sucrose. Mol. Gen. Genet. 1990;224:136-146.

Ni D.A., Yu X.H., Wang L.J., Xu Z.H. Aberrant development of pollen in transgenic tobacco expressing bacterial iaaM gene driven by pollen- and tapetum-specific promoters. Shi Yan Sheng Wu Xue Bao. 2002;35(1):1-6.

Obata-Sasamoto H., Suzuki H. Activities of enzymes relating to starch synthesis and endogenous levels of growth regulators in potato stolon tips during tuberization. Physiol. Plant. 1979;45(3):320-324. DOI 10.1111/j.1399-3054.1979.tb02591.x.

Pastushkova L.Kh., Kononikhin A.S., Tiys E.S., Dobrokhotov I.V., Ivanisenko V.A., Nikolaev E.N., Larina I.M., Popov I.A. Urine proteome study for the evaluation of age dynamics in healthy men. Uspekhi gerontologii = Advances in Gerontology. 2015б;28(4):294-700. (in Russian)

Pastushkova L.Kh., Kononikhin A.S., Tiys E.S., Nosovsky A.M., Dobrokhotov I.V., Ivanisenko V.A., Nikolaev E.N., Novoselova N.M., Custaud M.A., Larina I.M. Shifts in urine protein profile during dry immersion. Aviakosm. Ekolog. Med. 2015;49(4):15-19.

Pastushkova L.H., Kononikhin A.S., Tiys E., Obraztsova O.A., Dobrokhotov I.V., Ivanisenko V.A., Nikolaev E.N., Larina I.M. Identification of biological processes on the composition of the urine proteome cosmonauts on the first day after long space flights. Rossiyskiy fiziologicheskiy zhurnal im. I.M. Sechenova = I.M. Sechenov Physiological Journal. 2015a;101:222-237. (in Russian)

Petrovskiy E.D., Saik O.V., Tiys E.S., Lavrik I.N., Kolchanov N.A., Ivanisenko V.A. Prediction of tissue-specific effects of gene knockout on apoptosis in different compartments of human brain. BMC Genomics. 2015;16(Suppl.13):S3. DOI 10.1186/1471-2164-16-S13-S3.

Popik O.V., Petrovskiy E.D., Mishchenko E.L., Lavrik I.N., Ivanisenko V.A. Mosaic gene network modelling identified new regulatory mechanisms in HCV infection. Virus Res. 2015;218:71-78. DOI 10.1016/j.virusres.2015.10.004.

Purcell P.C., Smith A.M., Halford N.G. Antisense expression of a sucrose nonfermenting-1-related protein kinase sequence in potato results in decreased expression of sucrose synthase in tubers and loss of sucrose-inducibility of sucrose synthase transcripts in leaves. Plant J. 1998;14:195-202. DOI 10.1046/j.1365-313X.1998.00108.x.

Quettier A.L., Bertrand C., Habricot Y., Miginiac E., Agnes C., Jeannette E., Maldiney R. The phs1-3 mutation in a putative dual-specificity protein tyrosine phosphatase gene provokes hypersensitive responses to abscisic acid in *Arabidopsis thaliana*. Plant J. 2006;47(5):711-719. DOI 10.1111/j.1365-313X.2006.02823.x.

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.

Roumeliotis E., Kloosterman B., Oortwijn M., Kohlen W., Bouwmeester H.J., Visser R.G., Bachem C.W. The effects of auxin and strigolactones on tuber initiation and stolon architecture in potato. J. Exp. Bot. 2012;63(12):4539-4547. DOI 10.1093/jxb/ers132.

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 the knowledge base SOLANUM TUBEROSUM. Selskokhozyaystvennaya biologiya = Agricultural Biology. 2017;52(1):63-74. DOI 10.15389/agrobiology.2017.1.63rus. (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. 2016a;218:40-48. DOI 10.1016/j.virusres.2015.12.003.

Saik O.V., Konovalova N.A., Demenkov P.S., Ivanisenko T.V., Petrovskiy E.D., Ivanisenko N.V., Ivanoshchuk D.E., Ponomareva M.N., Konovalova O.S., Lavrik I.N., Kolchanov N.A. Molecular associations of Primary Open-Angle Glaucoma with potential comorbid diseases (POAG-associome). Biotecnología Aplicada. 2016b;33(3):3201-3206.

Salanoubat M., Belliard G. The steady-state level of potato sucrose synthase mRNA is dependent on wounding, anaerobiosis and sucrose concentration. Gene. 1989;84:181-185. DOI 10.1016/0378-1119(89)90153-4.

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.

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.

Slocombe S.P., Laurie S., Bertini L., Beaudoin F., Dickinson J.R., Halford N.G. Molecular cloning of SnIP1, a novel protein that interacts with SNF1-related protein kinase (SnRK1). Plant Mol. Biol. 2002;49:31-44.

Stark C., Breitkreutz B.J., Reguly T., Boucher L., Breitkreutz A., Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34:D535-D539. DOI 10.1093/nar/gkj109.

Tang B., Wu Y., Jiang M., Denny J.C., Xu H. Recognizing and encoding disorder concepts in clinical text using machine learning and vector space model. Working Notes for CLEF 2013 Conference. 2013;1179.

Thomas P.D., Kejariwal A., Guo N., Mi H., Campbell M.J., Muruganujan A., Lazareva-Ulitsky B. Applications for protein sequence function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. 2006;34(Suppl.2):W645-W650. DOI 10.1093/nar/gkl229.

Uzuner O., South B.R., Shen S., DuVall S.L. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J. Am. Med. Inform. Assoc. 2011;18:552-556. DOI 10.1136/amiajnl-2011-000203.

Van Harsselaar J.K., Lorenz J., Senning M., Sonnewald U., Sonnewald S. Genome-wide analysis of starch metabolism genes in potato (*Solanum tuberosum* L.). BMC Genomics. 2017;18(1):37. DOI 10.1186/s12864-016-3381-z.

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.

Xing M., Xue H. A proteomics study of auxin effects in *Arabidopsis thaliana*. Acta Biochim. Biophys. Sin. (Shanghai). 2012;44(9):783-796. DOI 10.1093/abbs/gms057.

Zhang H., Hou J., Liu J., Zhang J., Song B., Xie C. The roles of starch metabolic pathways in the cold-induced sweetening process in potatoes. Starch-Stärke. 2017;69:1-2. DOI 10.1002/star.201600194.

Zhu Y.X., Davies P.J. The control of apical bud growth and senescence by auxin and gibberellin in genetic lines of peas. Plant Physiol. 1997;113(2):631-637.