














doi 10.18699/vjgb-25-129

SmartCrop: база знаний молекулярно-генетических механизмов адаптации риса и пшеницы к стрессовым факторам

П.С. Деменков , Т.В. Иванисенко , М.А. Клещев , Е.А. Антропова , И.В. Яцык , А.Р. Волянская ,
А.В. Адамовская , А.В. Мальцева , А.С. Вензель , Х. Чао , М. Чен , В.А. Иванисенко  

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

² Факультет биоинформатики, Колледж естественных наук, Чжэцзянский университет, Ханчжоу, Китай

 salix@bionet.nsc.ru














Аннотация. Изучение молекулярно-генетических механизмов реакций растений на специфические условия роста и стрессовые факторы – одно из приоритетных направлений исследований, нацеленных на создание новых сортов сельскохозяйственных культур, в частности риса и пшеницы. К числу таких факторов относятся абиотические стрессы (высокие или низкие температуры, засуха, засоление, загрязнение почвы металлами), биотические стрессы (патогены, вредители), а также реакции растений на регуляторные факторы (удобрения, гормоны, элиситоры и другие соединения). Современные исследования в области генетики растений основаны на понимании того, что формирование любых фенотипических характеристик (молекулярно-генетических, биохимических, физиологических, морфологических и др.) контролируется генными сетями – группами согласованно функционирующих генов, взаимодействующих через свои продукты (РНК, белки и метаболиты). Ранее с целью реконструкции генных сетей, значимых для биологии и биомедицины, нами была разработана интеллектуальная компьютерная система ANDSystem, предназначенная для автоматизированного извлечения знаний из текстов научных публикаций и баз данных. В настоящей работе, используя адаптированную версию ANDSystem для растений, мы создали базу знаний SmartCrop для решения задач, связанных с изучением молекулярно-генетических механизмов взаимодействий «генотип–фенотип–среда» для сельскохозяйственно ценных культур риса и пшеницы. SmartCrop предназначена для помощи исследователям в решении таких задач, как интерпретация результатов омиксных экспериментов на растениях: установление связей между наборами генов и биологическими процессами, фенотипическими признаками и др.; реконструкция генных сетей, описывающих отношения между молекулярно-генетическими объектами и понятиями в селекции, феномике, семеноводстве, фитопатологии; выявление регуляторных и сигнальных путей, ответных реакций растений на специфические условия роста и биотические и абиотические стрессы; прогнозирование генов-кандидатов для генотипирования; поиск маркеров для маркер-опосредованной селекции; выявление потенциальных мишеней (генов и белков) для субстанций, влияющих на растения (контролирующих процессы прорастания семян, вегетативного роста, эффективного поглощения питательных веществ и улучшения устойчивости к стрессовым факторам).

Ключевые слова: база знаний SmartCrop; ANDSystem; извлечение знаний из текстов; искусственный интеллект; молекулярно-генетические механизмы; рис; пшеница; ассоциативные генные сети; абиотические стрессы; биотические стрессы; взаимодействия генотип–фенотип–среда; омиксные технологии; длинные некодирующие РНК; маркер-опосредованная селекция; адаптация растений; стрессоустойчивость

Для цитирования: Деменков П.С., Иванисенко Т.В., Клещев М.А., Антропова Е.А., Яцык И.В., Волянская А.Р., Адамовская А.В., Мальцева А.В., Вензель А.С., Чао Х., Чен М., Иванисенко В.А. SmartCrop: база знаний молекулярно-генетических механизмов адаптации риса и пшеницы к стрессовым факторам. *Вавиловский журнал генетики и селекции*. 2025;29(8):1221-1234. doi 10.18699/vjgb-25-129


Финансирование. Работа ПСД, ТВИ, МАК, ЕАА, ИВЯ, АРВ, АВА, АВМ, АСВ и ВАИ поддержана российско-китайским грантом Российского научного фонда № 23-44-00030. Работа МЧ (MCh) и ХЧ (HCh) поддержана Национальным фондом естественных наук Китая (№ 32261133526).

SmartCrop: knowledge base of molecular genetic mechanisms of rice and wheat adaptation to stress factors

P.S. Demenkov , T.V. Ivanisenko , M.A. Kleshchev , E.A. Antropova , I.V. Yatsyk , A.R. Volyanskaya ,
A.V. Adamovskaya , A.V. Maltseva , A.S. Venzel , H. Chao , M. Chen , V.A. Ivanisenko  

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

² Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou, China

 salix@bionet.nsc.ru

© Деменков П.С., Иванисенко Т.В., Клещев М.А., Антропова Е.А., Яцык И.В., Волянская А.Р., Адамовская А.В., Мальцева А.В., Вензель А.С., Чао Х., Чен М., Иванисенко В.А., 2025

Контент доступен под лицензией Creative Commons Attribution 4.0

Abstract. The study of molecular genetic mechanisms of plant responses to specific growth conditions and stress factors is a central focus of scientific research aimed at developing new valuable crop varieties, particularly rice and wheat. These factors include abiotic stresses (high or low temperatures, drought, salinity, soil metal contamination), biotic stresses (pathogens, pests), as well as plant responses to regulatory factors (fertilizers, hormones, elicitors, and other compounds). Modern research in plant genetics is based on the understanding that the formation of any phenotypic characteristics (molecular genetic, biochemical, physiological, morphological, etc.) is controlled by gene networks – groups of coordinately functioning genes interacting through their products (RNA, proteins, and metabolites). Previously, we developed the ANDSystem intelligent technology designed to extract knowledge from scientific publication texts for the reconstruction of gene networks in biology and biomedicine. In this work, using an adapted version of ANDSystem for plants, we created the SmartCrop knowledge base designed to address challenges related to studying molecular genetic mechanisms of genotype-phenotype-environment interactions for agriculturally valuable rice and wheat crops. SmartCrop is designed to assist researchers in solving tasks such as interpreting omics technology results (establishing connections between gene sets and biological processes, phenotypic traits, etc.); reconstructing gene networks describing relationships between molecular genetic objects and concepts in breeding, phenomics, seed production, phytopathology, diagnostics, protective agents, etc.; identifying regulatory and signaling pathways of plant responses to specific growth conditions and biotic and abiotic stresses; predicting candidate genes for genotyping; searching for markers for marker-assisted selection; and identifying potential targets for substances (including external factors) affecting plants to ensure timely and uniform germination, better vegetative growth, efficient nutrient uptake, and improved stress resistance.

Key words: SmartCrop knowledge base; ANDSystem; text mining; artificial intelligent; molecular genetic mechanisms; rice; wheat; associative gene networks; abiotic stress; biotic stress; genotype-phenotype-environment interactions; omics technologies; long non-coding RNAs; marker-assisted selection; plant adaptation; stress resistance

For citation: Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Antropova E.A., Yatsyk I.V., Volyanskaya A.R., Adamovskaya A.V., Maltseva A.V., Venzel A.S., Chao H., Chen M., Ivanisenko V.A. SmartCrop: knowledge base of molecular genetic mechanisms of rice and wheat adaptation to stress factors. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2025;29(8):1221-1234. doi 10.18699/vjgb-25-129

Введение

Рис (*Oryza sativa* L.) и пшеница (*Triticum aestivum* L.) – важнейшие сельскохозяйственные культуры, обеспечивающие продовольственную безопасность значительной части мирового населения. Обе культуры известны своей высокой питательной, технической и кормовой ценностью (Shewry, Hey, 2015). В современных условиях производство этих культур сталкивается с серьезными вызовами. Экстремальные погодные явления, неблагоприятные климатические изменения, болезни растений и вредители приводят к значительным потерям урожая (Lesk et al., 2016). Преодоление указанных трудностей невозможно без изучения молекулярно-генетических механизмов устойчивости растений к неблагоприятным биотическим и абиотическим факторам, что требует анализа комплексных систем, включающих сложные сигнальные, регуляторные, транспортные и метаболические пути (Mittler, 2006; Nykiel et al., 2023).

Эффективным инструментом для изучения таких механизмов являются генные сети, контролирующие молекулярно-генетические процессы, определяющие формирование фенотипических признаков и функционирование биологических процессов, в том числе ответные реакции растений на стресс.

Современная концепция генных сетей охватывает не только молекулярные компоненты (РНК, гены, белки и метаболиты), но и широкий спектр сущностей различной природы, включая заболевания, биологические процессы и факторы окружающей среды. Такой тип генных сетей получил название ассоциативных генных сетей (Ivanisenko V. et al., 2015). По своей структуре они представляют собой граф знаний, интегрирующий информацию о взаимодействиях между разнородными объектами,

участвующими в функционировании молекулярно-генетических систем или влияющими на них. В агробиологии и растениеводстве анализ генных сетей успешно применяется для исследования таких экономически ценных признаков, как устойчивость к болезням и вредителям, толерантность к абиотическим факторам и урожайность (Virlovet et al., 2018; Chen et al., 2020).

Реконструкция генных сетей растений представляет собой сложную задачу, требующую обработки огромных массивов данных и интеграции разрозненной информации из научных публикаций, включая данные о регуляторных, транспортных и каталитических процессах, а также о связях между генетическими особенностями, фенотипическими проявлениями и факторами окружающей среды. Для извлечения таких знаний используют методы автоматического анализа текстов (text-mining), основанные как на классических компьютерных подходах (на основе словарей, синтактико-лингвистических правил и шаблонов, статистически значимой совстречаемости и т. д.), так и на методах машинного обучения (Ivanisenko T. et al., 2014; Shrestha et al., 2024; Zhang et al., 2024).

В свою очередь, алгоритмы машинного обучения, применяемые для построения и анализа генных сетей, можно разделить на следующие направления: методы обучения с учителем, методы обучения без учителя, обучение с частичным контролем, а также гибридные подходы. Методы обучения с учителем используют предварительно размеченные данные для построения предсказательных моделей, например, для идентификации ключевых регуляторов или предсказания функциональных взаимодействий между генами растений (Ni et al., 2016). Обучение без учителя позволяет выявлять скрытые паттерны в больших объемах информации, что важно, например, при

кластеризации генов по схожести их экспрессии или выявлении модулей генных сетей. Методы обучения с частичным привлечением учителя комбинируют преимущества обоих подходов, используя как размеченные, так и неразмеченные данные, что особенно актуально в условиях ограниченного количества хорошо аннотированных данных (Yan, Wang, 2022). Гибридные подходы объединяют различные методы машинного обучения, а также традиционные подходы биоинформатики, что позволяет эффективно бороться с недостатками этих методов, когда они используются по отдельности (Guindani et al., 2024; Ivanisenko T. et al., 2024). Так, интеграция методов распознавания в текстах поименованных сущностей по словарю с методами машинного обучения способствует их более точной идентификации (Do et al., 2018; Ivanisenko T. et al., 2020).

В последние годы глубокое машинное обучение достигло значительных успехов благодаря внедрению архитектуры трансформера и механизмов внимания, что обеспечило качественный прогресс в обработке естественного языка и анализе биологических последовательностей (Vaswani et al., 2017). Существенное развитие получил и анализ генных сетей с применением графовых нейросетей, среди которых архитектура GraphSAGE (Hamilton et al., 2017) позволяет эффективно обучать модели на больших гетерогенных графах за счет агрегации признаков соседних узлов. Перспективным направлением является также использование больших языковых моделей, таких как Gemma-2-9b-it (Gemma Team, Google DeepMind, 2024), обеспечивающих высокое качество семантического анализа научных текстов и валидации извлеченных взаимодействий.

Для реконструкции и анализа генных сетей растений разработан ряд специализированных ресурсов. К ним относятся: PlantRegMap (Tian et al., 2020), предназначенный для анализа регуляторных взаимодействий транскрипционных факторов; STRING (Szklarczyk et al., 2021), позволяющий исследовать белок-белковые взаимодействия; платформа KEGG PLANT (Kanehisa, 2013), объединяющая информацию о метаболических путях различных видов растений; ресурс Plant Reactome (Naithani et al., 2020), содержащий детальные данные о сигнальных и метаболических путях модельных растительных организмов. Для визуализации и анализа генных сетей широко применяется программа Cytoscape (Otasek et al., 2019), имеющая обширный набор плагинов для работы с биологическими данными. База данных ncPlantDB предоставляет комплексную информацию для анализа регуляторных сетей, включая данные о тканеспецифичной экспрессии некодирующих РНК на уровне отдельных клеток и их взаимодействиях (Cheng et al., 2024; Liu et al., 2025). Интеграция подобных омиксных ресурсов формирует эффективную платформу для реконструкции генных сетей сельскохозяйственных культур (Chao et al., 2023).

Ранее нами разработана когнитивная программно-информационная система ANDSystem (Ivanisenko V. et al., 2015, 2019; Ivanisenko T. et al., 2020, 2022), предназначенная для полного цикла инженерии знаний в медико-биологической области. База знаний системы содержит свыше 50 миллионов взаимодействий для различных организмов.

В области биологии растений с помощью ANDSystem была создана база знаний по генетике картофеля SOLA-NUM TUBEROSUM (Сайк и др., 2017; Иванисенко и др., 2018; Деменков и др., 2019), проведены реконструкция и анализ регуляторной генной сети, контролирующей функции клеточной стенки в листьях *Arabidopsis thaliana* L. при водном дефиците (Volyanskaya et al., 2023), а также разработан метод приоритизации биологических процессов на основе реконструкции и анализа ассоциативных генных сетей (Деменков и др., 2021).

Применение системы автоматизированной реконструкции ассоциативных генных сетей ANDSystem для анализа микроРНК-регуляции адаптации мягкой пшеницы (*Triticum aestivum* L.) к условиям водного дефицита позволило предложить новые кандидатные микроРНК (MIR7757, MIR9653a, MIR9670, MIR9672b), представляющие интерес для дальнейших экспериментальных исследований механизмов адаптации растений к недостаточному увлажнению (Клещев и др., 2024).

В другой работе (Антропова и др., 2024) с использованием ANDSystem была реконструирована молекулярно-генетическая сеть ответа риса (*Oryza sativa*) на поражение *Rhizoctonia solani* в условиях избытка азота, что позволило выявить три потенциальных механизма, объясняющих снижение устойчивости растения к патогену. Идентифицированы ключевые регуляторные пути: OsGSK2-опосредованный каскад, сигнальный путь OsMYB44–OsWRKY6–OsPR1, а также путь, включающий SOG1, Rad51 и гены PR1/PR2. Кроме того, выделены перспективные для селекции маркеры – 7 генов, регулирующих широкий спектр стрессовых ответов, и 11 генов-модуляторов иммунной системы. Дополнительный анализ некодирующих РНК (Антропова и др., 2024) позволил выявить 30 микроРНК, мишенями которых являются гены, входящие в реконструированную генную сеть. Для двух из них (Osa-miR396 и Osa-miR7695) обнаружено около 7400 уникальных длинных некодирующих РНК с различными индексами коэкспрессии, что может указывать на сложную архитектуру посттранскрипционной регуляции в условиях азотного стресса.

Целью настоящей работы была адаптация системы ANDSystem для создания базы знаний SmartCrop, интегрирующей данные о молекулярно-генетических механизмах и ассоциативных генных сетях стрессового ответа риса и пшеницы на основе интеллектуального анализа научных публикаций и фактографических баз данных. Работа включала создание предметной онтологии и оптимизацию интеллектуальных методов извлечения знаний из научных текстов с использованием семантико-лингвистических шаблонов и предобученных больших языковых моделей. Разработанная онтология SmartCrop представлена набором взаимосвязанных словарей, описывающих: молекулярно-генетические объекты (гены, белки, метаболиты, микроРНК), биологические процессы, фенотипические признаки и болезни, патогены, генетические биомаркеры, маркеры устойчивости к средствам защиты растений, молекулярные мишени химических средств защиты, биотические и абиотические факторы, средства защиты растений, а также сорта с их экономическими ценными и потребительскими свойствами.

В результате автоматизированного анализа научных публикаций сформирована база знаний SmartCrop, интегрирующая более 10 миллионов взаимодействий между описанными в онтологии сущностями.

Материалы и методы

Информационные ресурсы. Для создания базы знаний SmartCrop мы использовали программно-информационную платформу ANDSystem (Ivanisenko V. et al., 2015, 2019; Ivanisenko T. et al., 2020, 2022) и ее информационные и биоинформатические технологии.

Настройка методов ANDSystem на предметную область осуществлялась с использованием онтологии, включающей специализированные словари объектов и описание типов их взаимодействий. Основу генетической и геномной информации при составлении словарей составили: база данных NCBI Gene (<https://www.ncbi.nlm.nih.gov/gene>), специализированная база данных по рису Oryzabase (<https://shigen.nig.ac.jp/rice/oryzabase>), база данных микроРНК miRBase (<https://www.mirbase.org>), база данных по коэкспрессии длинных некодирующих РНК ncPlantDB (<https://bis.zju.edu.cn/ncPlantDB/>), база данных однонуклеотидных полиморфизмов dbSNP (<https://www.ncbi.nlm.nih.gov/snp>) и база данных по злаковым культурам GrainGenes (<https://wheat.pw.usda.gov/GG3>).

Для стандартизации терминологии использовались онтологии: Gene Ontology (<http://geneontology.org>), Crop Ontology для пшеницы и риса (<https://cropontology.org>), а также коллекция генетических ресурсов ВИР (<https://www.vir.nw.ru>).

Химические соединения и метаболиты были аннотированы с помощью базы данных ChEBI (<https://www.ebi.ac.uk/chebi>). Информация об устойчивости к гербицидам была получена из International Herbicide-Resistant Weed

Database (<http://www.weedscience.org>), а данные о пестицидах – из EU Pesticide Database (https://food.ec.europa.eu/plants/pesticides/eu-pesticides-database_en).

Для извлечения знаний из текстов применялись семантико-лингвистические шаблоны ANDSystem, а также вновь созданные шаблоны, учитывающие специфику предметной области. Дополнительно для извлечения знаний использовались методы искусственного интеллекта, включая графовые нейросети GraphSAGE (Hamilton et al., 2017) и большую языковую модель Gemma-2-9b-it (Gemma Team, Google DeepMind, 2024).

Оценка точности. Для оценки качества разметки имен объектов в тексте применялась F1-мера, представляющая собой гармоническое среднее между точностью (Precision) и полнотой (Recall):

$$F1 = 2 \cdot (Precision \times Recall) / (Precision + Recall),$$

$$Precision = TP / (TP + FP),$$

$$Recall = TP / (TP + FN),$$

где TP – истинно положительные, FP – ложноположительные, FN – ложноотрицательные результаты.

Результаты

Схематическое представление основных блоков программно-информационной системы SmartCrop показано на рис. 1.

Модуль предметно-ориентированной онтологии SmartCrop

Создание онтологии предметной области было ключевым этапом разработки SmartCrop. Предметно-ориентированная онтология формирует концептуальную модель проблемной области и включает словари объектов и типы их взаимодействий. На основе этих словарей из текстов и

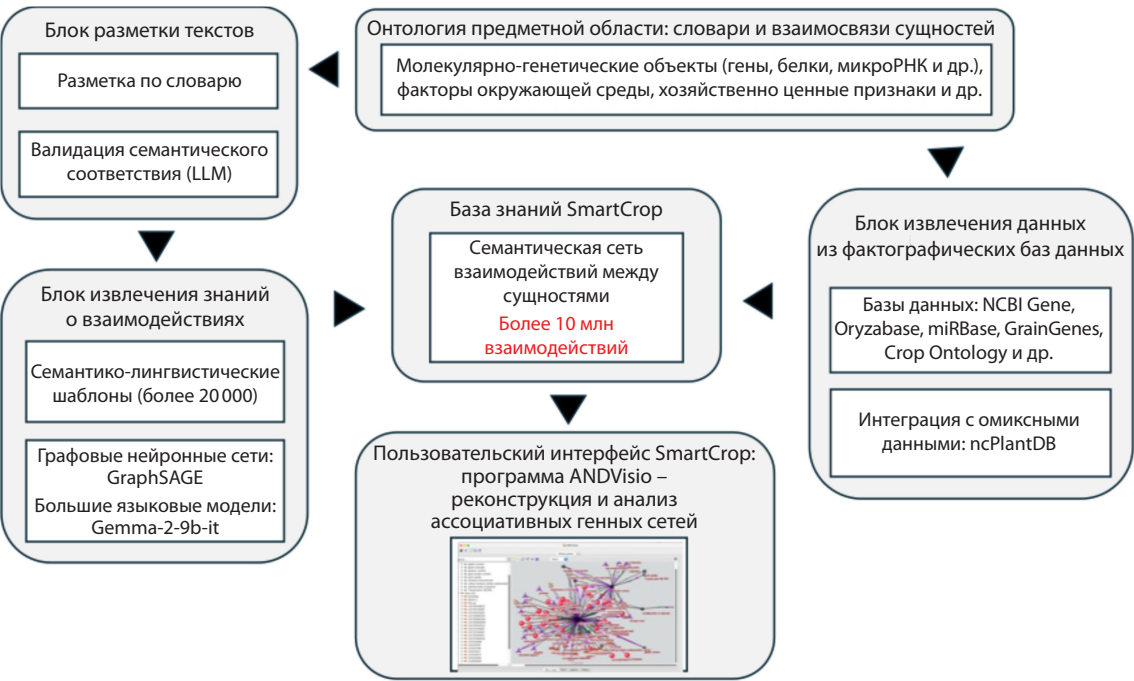


Рис. 1. Схематическое представление архитектуры программно-информационной системы SmartCrop.

фактографических баз данных извлекается информация о взаимодействиях между конкретными объектами. Текущая версия онтологии SmartCrop содержит 15 словарей различных объектов (табл. 1), составленных путем извлечения названий сущностей из специализированных баз данных и существующих онтологий.

Типы взаимодействий. В системе SmartCrop определены 16 типов отношений между объектами онтологии. Все взаимодействия в системе характеризуются определенной направленностью и подразделяются на несколько основных групп. К физическим взаимодействиям относятся процессы образования как кратковременных молекулярных комплексов, так и устойчивых соединений между белками и метаболитами.

Химические взаимодействия включают каталитические реакции типа субстрат–фермент–продукт, процессы протеолитического расщепления белков, а также различные посттрансляционные модификации белков, такие как фосфорилирование и гликозилирование.

Особую группу составляют регуляторные взаимодействия, охватывающие регуляцию экспрессии генов транскрипционными факторами, модуляцию активности и функций белков, контроль транспорта белков и метаболитов, а также регуляцию стабильности и деградации белков. Важной особенностью является то, что регуляторные взаимодействия определяют также взаимосвязи между молекулярно-генетическими объектами, биологическими процессами и фенотипическими признаками. При этом каждое регуляторное событие может характеризоваться эффектом усиления или ослабления соответствующего процесса.

Отдельно выделяются экспрессия и коэкспрессия генов. Продукты экспрессии – белки и некодирующие РНК. Коэкспрессия – одновременная экспрессия генов, обусловленная общими регуляторными механизмами при изменении клеточных условий. Дополнительно система учитывает ассоциативные связи, включающие неклассифицированные взаимодействия между различными объектами онтологии.

Блок разметки текстов на основе объектов онтологии

Распознавание молекулярно-генетических объектов в научных текстах представляет собой сложную задачу, обусловленную спецификой их наименований. Наш опыт работы с ANDSystem показывает, что значительная часть ошибок при автоматической реконструкции ассоциативных генных сетей связана с неточной идентификацией поименованных сущностей (Ivanisenko T. et al., 2022). Причиной таких ошибок являются использование авторами аббревиатур, семантическая многозначность терминов, а также разнообразные лингвистические особенности научных текстов. В публикациях нередко модифицируются стандартные наименования объектов, изменяются пунктуация и порядок слов, варьируются грамматические формы, применяются сокращения либо допускаются технические опечатки (Pearson, 2001; Krallinger et al., 2015; Islamaj et al., 2021).

Для повышения точности распознавания нами разработан процесс, включающий два этапа: 1) первичное сопоставление названий со словарем онтологии; 2) вторичная

Таблица 1. Словари, представленные в онтологии SmartCrop

Словарь	Количество объектов
Гены/белки риса (<i>O. sativa</i>)	45198
Гены/белки пшеницы (<i>T. aestivum</i>)	155761
МикроРНК риса (<i>O. sativa</i>)	604
МикроРНК пшеницы (<i>T. aestivum</i>)	122
Генетические биомаркеры (<i>T. aestivum</i>)	862
Полиморфизмы QTL (<i>O. sativa</i>)	1987
Полиморфизмы QTL (<i>T. aestivum</i>)	1266
Сорта риса (<i>O. sativa</i>)	14377
Сорта пшеницы (<i>T. aestivum</i>)	25501
Метаболиты	74838
Биологические процессы	122477
Хозяйственно ценные признаки	234
Фенотипические признаки	2386
Болезни, патогены и вредители	1065
Маркеры устойчивости к СЗР	861
Биотические факторы	710
Абиотические факторы	496
Средства защиты растений, гербициды (СЗР)	1336
Молекулярные мишени для химических СЗР	14
Длинные некодирующие РНК	6546

верификация соответствия каждого размеченного имени объекта его типу на основе контекстуального анализа документа с помощью нейронных сетей.

Процесс верификации реализован следующим образом: языковая модель преобразует контекст (около 400 слов), который содержит анализируемый объект, замещенный специальным тэгом-маской, в векторное представление. Используя это представление, нейронная сеть выполняет двоичную классификацию, определяя, соответствует ли контекстуальное окружение термина его типичному употреблению.

Для объектов из онтологии ANDSystem (гены, белки, метаболиты и др.) точность классификации приведена в ранее опубликованной статье (Ivanisenko T. et al., 2022). Для новых словарей SmartCrop была проведена ручная экспертная оценка качества разметки (табл. 2) по результатам анализа 1000 случайно отобранных документов из баз PubMed и PubMed Central.

Результаты оценки показали высокую точность разметки для большинства словарей, за исключением наименований сортов риса и пшеницы. Идентификация сортовых названий растений является известной комплексной задачей, обусловленной рядом факторов, включая значительное пересечение терминов с общеупотребительной лексикой и антропонимами, а также отсутствие унифицированного стандарта в номенклатуре новых сортов (Do et al., 2018; D'Souza, 2024).

Таблица 2. Оценка точности разметки имен объектов по новым словарям

Словарь	F1-мера
Сорта (<i>O. sativa</i>)	0.103 (0.88)*
Сорта (<i>T. aestivum</i>)	0.274 (0.919)*
Хозяйственно ценные признаки	0.789
Абиотические факторы	0.748
Биотические факторы	0.927
Болезни, патогены и вредители	0.881

* В скобках указана точность после фильтрации типа распознанных имен с использованием нейронной сети.

Для решения данной проблемы было проведено обучение специализированной языковой модели, ориентированной на задачу контекстной классификации терминов. Обучение выполнялось в соответствии с методологией, ранее описанной в нашей работе по повышению точности идентификации восьми типов молекулярно-генетических сущностей, включая белки, гены, метаболиты и клеточные компоненты (Ivanisenko T. et al., 2022). Интеграция разработанной модели позволила существенно повысить точность распознавания наименований сортов (F1-мера) – до 0.88 у риса и 0.919 у пшеницы.

Блок извлечения знаний

Блок извлечения знаний из текстов научных публикаций реализует три основных этапа: 1) первичное извлечение знаний с использованием семантико-лингвистических шаблонов; 2) реконструкцию первичной семантической сети; 3) ее расширение с применением графовых нейросетей и больших языковых моделей. Дополнительно для расширения семантической сети используется блок извлечения данных из фактографических баз, содержащих структурированную информацию, что позволяет получать дополнительные сведения о взаимодействиях между объектами.

Семантико-лингвистические шаблоны представляют собой структурированные записи, содержащие метаданные о типах объектов и характере их взаимодействий. Они включают два основных компонента: 1) синтаксические связи, описывающие порядок расположения объектов и ключевых слов в предложении через регулярные выражения, и 2) семантические связи, определяющие тип взаимодействия между объектами. Регулярные выражения применяются для поиска соответствий в расположении имен объектов в размеченных предложениях текста. При обнаружении соответствия конкретные имена объектов из текста сопоставляются с идентификаторами шаблона.

Для каждого типа взаимодействий разработаны специализированные группы шаблонов с уникальными синтаксическими правилами. База знаний содержит более 18000 шаблонов ANDSystem для типов взаимодействий, отраженных как в онтологии ANDSystem, так и в SmartCrop, а также более 3000 шаблонов, специально разработанных для онтологий риса и пшеницы. Эффек-

тивность метода извлечения взаимодействий на основе шаблонов была подтверждена в ходе разработки системы ANDSystem (Ivanisenko V. et al., 2015).

Применение графовых нейросетей и больших языковых моделей. На втором этапе на основе знаний, извлеченных с помощью шаблонов, строился первичный граф знаний (семантическая сеть), который использовался для обучения графовой нейросети. После обучения выполнялось предсказание отсутствующих ребер в графе знаний. На третьем этапе большие языковые модели применялись для валидации этих предсказаний на основе анализа текстов научных публикаций, содержащих в разметке объекты с предполагаемым взаимодействием (Ivanisenko T. et al., 2024).

Интеграция с омиксными данными

Некодирующие РНК (нкРНК) – обширный и функционально разнообразный класс РНК-молекул, не транслирующихся в белки, но выполняющих ключевые регуляторные функции в клетке. Особый интерес представляют днРНК, поскольку они участвуют в регуляции экспрессии генов на различных уровнях – от модификации стабильности и трансляции мРНК до включения в сложные сигнальные каскады (Statello et al., 2021; Supriya et al., 2024).

Известным специализированным ресурсом по коэкспрессии нкРНК у растений, в том числе днРНК риса, является база данных ncPlantDB (<https://bis.zju.edu.cn/ncPlantDB/>). Она содержит сведения о тканеспецифичной экспрессии нкРНК на уровне отдельных клеток и об их предполагаемых взаимодействиях, полученные с помощью современных методов одноклеточного транскриптомного анализа (Cheng et al., 2024; Liu et al., 2025). Интеграция SmartCrop с базой ncPlantDB позволила использовать данные о коэкспрессии нкРНК, а также их связи с микроРНК для обогащения реконструируемых генных сетей.

Модуль анализа и визуализации генных сетей

В качестве пользовательского графического интерфейса SmartCrop, предназначенного для реконструкции и анализа генных сетей на основе информации из базы знаний SmartCrop, используется программа ANDVisio (рис. 2).

Программа ANDVisio (Demenkov et al., 2012) изначально была разработана как компонент системы ANDSystem и впоследствии адаптирована для интеграции со SmartCrop. Она предоставляет исследователям широкий набор инструментов для структурного и функционального анализа генных сетей, в том числе:

- многовариантные алгоритмы раскладки графа;
- систему многопараметрической фильтрации;
- механизмы поиска путей и циклов;
- инструменты расчета показателей центральности узлов;
- средства оценки обогащения биологических процессов генами сети;
- дополнительные методы сетевого анализа.

База знаний SmartCrop

База знаний системы реализована в виде семантической сети (графа знаний), интегрирующей данные, извлеченные как из научных публикаций, так и из фактографиче-

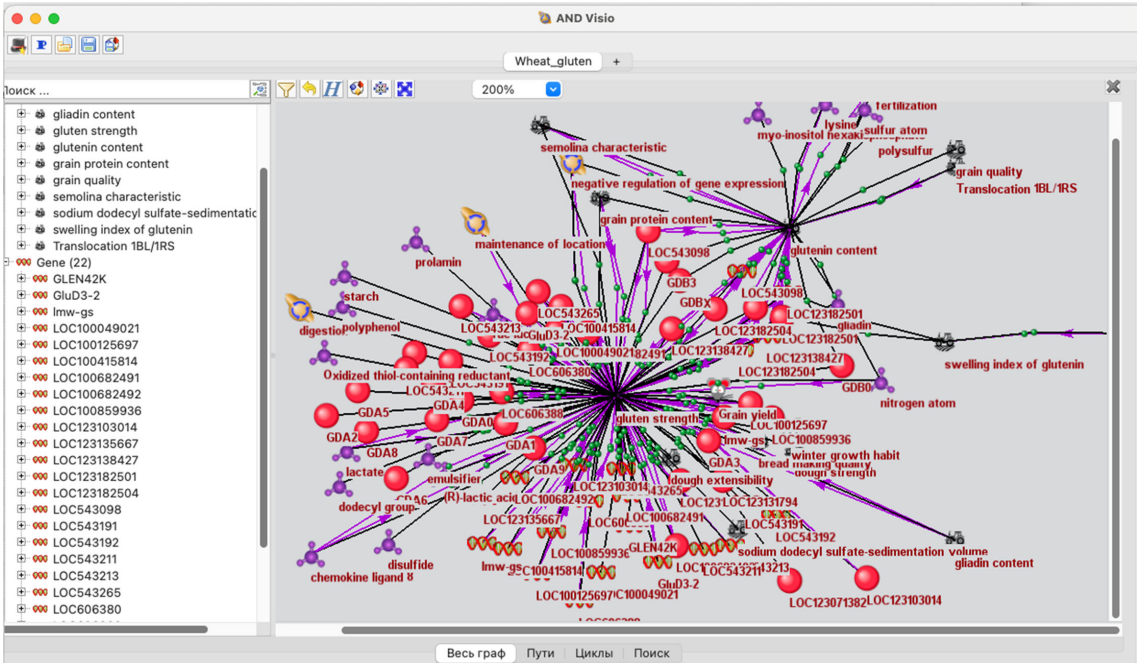


Рис. 2. Образ экрана с изображением интерфейса программы ANDVisio.

ских баз данных. В структуре графа узлы соответствуют объектам предметной онтологии, а ребра отражают различные типы взаимодействий между ними.

Наполнение базы знаний осуществлялось посредством систематического анализа научной литературы, включая рефераты из PubMed и полнотекстовые статьи из открытого ресурса PubMed Central. Проанализированные публикации охватили период с 1970 по 2024 г., при этом основным критерием отбора было наличие упоминаний пшеницы или риса. Детальная статистика по количеству зафиксированных взаимодействий в базе знаний SmartCrop приведена в табл. 3.

Обсуждение

Для демонстрации возможностей SmartCrop рассмотрим два сценария ее применения: анализ экспериментальных омиксных данных и планирование экспериментов.

Анализ экспериментальных омиксных данных

В качестве примера интерпретации омиксных данных проведена функциональная аннотация дифференциально экспрессирующихся генов (ДЭГ) мягкой пшеницы при солевом стрессе. Для анализа был использован набор из 5829 ДЭГ, выделенных из базы NCBI GEO (GSE225565, Alyahya, Taybi, 2023) для тканей корня мягкой пшеницы (*Triticum aestivum* L., сорт Saudi) в ответ на засоление.

Результаты анализа перепредставленности объектов SmartCrop (биологические процессы, фенотипические признаки, хозяйственно ценные признаки, патогены) для этого набора ДЭГ и их белковых продуктов приведены в табл. S1 Приложения¹. Всего была выявлена достоверная ($p\text{-value} < 0.05$, с поправкой Бонферрони) перепредстав-

Таблица 3. Статистика базы знаний SmartCrop по взаимодействиям между объектами для пшеницы и риса

Тип взаимодействия	Количество взаимодействий	
	Рис (<i>Oryza sativa</i>)	Пшеница (<i>Triticum aestivum</i>)
Ассоциация	189650	207415
Регуляция экспрессии	22472	26474
Регуляция активности	3991	23411
Регуляция деградации	1442	4415
Регуляция транспорта	830	1320
нкРНК/миРНК регуляция	2125036	5814
Регуляция процессов	23274	23766
Каталитические реакции	2267	5798
Экспрессия	86353	311964
Физическое взаимодействие	8551	11810
Маркер (marker)	435	226

ленность 217 терминов, описывающих биологические процессы (тип объекта Pathway), 50 фенотипических признаков (Phenotype), 9 хозяйственно ценных признаков (Agrophenotype) и 38 видов патогенных организмов. Примеры объектов, входящих в число статистически наиболее значимых характеристик, приведены в табл. 4.

Анализ перепредставленных биологических процессов показал, что исследуемый набор ДЭГ связан как с ответом на солевой стресс, так и с реакцией на водный дефицит. Это отражает адаптационные механизмы растений к состоянию так называемой физиологической

¹ Табл. S1 Приложения см. по адресу:
<https://vavilovj-icg.ru/download/pict-2025-29/appx46.xls>

Таблица 4. Характеристики мягкой пшеницы, статистически значимо связанные с набором ДЭГ и их белковых продуктов при солевом стрессе, выявленные с помощью системы SmartCrop

Объект (термин)	Число связанных генов и белков	p-value	Объект (термин)	Число связанных генов и белков	p-value
Биологические процессы			Хозяйственно ценные признаки		
Response to water deprivation	37	2.73E–66	Seed longevity	2	3.96E–08
Cell death	31	1.12E–54	Grain thickness	2	1.77E–07
Cold acclimation	23	4.52E–41	Grain length	7	1.62E–20
Hyperosmotic salinity response	23	2.67E–38	Grain protein content	7	1.42E–17
Seed germination	20	7.21E–34	Fiber quality	2	1.23E–06
Фенотипические признаки			Патогены		
Stomatal closure rate	2	8.84E–47	<i>Fusarium</i> sp.	22	2.22E–59
Cell membrane stability	2	1.45E–31	<i>Fusarium culmorum</i>	10	1.23E–27
Grain moisture content	2	5.46E–31	<i>Fusarium oxysporum</i>	10	2.39E–26
Seed length	2	2.32E–30	<i>Botrytis cinerea</i>	10	8.63E–26
Na ⁺ uptake	2	1.98E–29	<i>Fusarium pseudograminearum</i>	10	2.45E–25

сухости, возникающей при невозможности эффективного поглощения воды из-за высокого осмотического давления окружающей среды. К числу таких адаптаций относится закрытие устьиц, обеспечиваемое быстрым повышением уровня абсцизовой кислоты (Verma et al., 2016; Zhao et al., 2021). Соответственно, среди достоверно перепредставленных объектов оказались как фенотипический признак «скорость закрытия устьиц» (stomatal closure rate, см. табл. 4), так и сигнальные пути, связанные с абсцизовой кислотой (см. табл. S1), что подтверждает их важную роль в ответе на засоление.

Следует отметить, что важным преимуществом базы знаний SmartCrop по сравнению с широко используемыми ресурсами для функциональной аннотации генов (DAVID, Gene Ontology, ShinyGO и др.) является возможность анализа связей генов и белков не только с биологическими процессами, молекулярными функциями, клеточными компонентами и путями KEGG, но и с широким спектром абиотических и биотических факторов среды, фенотипических признаков, хозяйственно значимых свойств и патогенов. Такая интеграция позволяет оценивать перепредставленность различных типов объектов для исследуемого набора генов, существенно расширяя возможности функциональной аннотации и обеспечивая выявление генов с плеiotропными эффектами. Последнее особенно важно для маркер-ориентированной селекции, поскольку отбор по одному целевому фенотипическому признаку или генетическому маркеру может затрагивать сразу несколько других, нецелевых признаков.

В частности, результаты функциональной аннотации ДЭГ мягкой пшеницы при солевом стрессе показали их ассоциацию не только с биологическими процессами, такими как ответ на засоление и водный дефицит, и процессами, контролирующими прорастание семян, но и с хозяйственно ценными признаками, отражающими качество зерна (см. табл. 4). Например, аквапорины, кодируемые генами *LOC543267*, *LOC100037645*, *LOC123093445*

и др., обеспечивают селективный транспорт молекул воды, участвуют в поддержании ионного состава клеток и регулировании водно-солевого баланса при повышенном засолении (Ayadi et al., 2019), а также способствуют перемещению воды и растворенных веществ внутри семян, что играет ключевую роль в процессе их прорастания (Noai et al., 2020).

Функциональность SmartCrop не ограничивается анализом перепредставленности. Система также позволяет реконструировать ассоциативные сети белков и генов, значимо связанных с перепредставленными объектами, и проводить поиск их регуляторов, что обеспечивает более глубокое понимание молекулярных механизмов этих связей и выявление их специфики в определенных условиях эксперимента.

В качестве примера была реконструирована генная сеть регуляции устойчивости растений к гиперосмотическому стрессу (GO:0042538 hyperosmotic salinity response) (рис. 3). Согласно данным SmartCrop, в формировании реакции пшеницы на гиперосмотический стресс участвуют 95 генов и 119 белков, включая аквапорины и натрий-водородные транспортеры (sodium/hydrogen exchangers), которые играют ключевую роль в регуляции внутриклеточного pH, водного баланса и гомеостаза ионов натрия (Gupta et al., 2021). Избыточные ионы натрия, поступающие из окружающей среды, удаляются из цитоплазмы в апопласт и вакуоли в обмен на ионы водорода с помощью трансмембранных Na⁺/H⁺ транспортеров (Zhao et al., 2021).

В сеть также входят пероксидазы и каталазы, участвующие в антиоксидантной защите при абиотическом стрессе, транскрипционные факторы семейств MYB и WRKY, дегидрины (LOC123125487, LOC100141381 и др.), белки холодового шока (LOC123080042, LOC543252, LOC542792) и белки DELLA, которые, подавляя сигнальный путь гиббереллина и взаимодействуя с сигнальными путями жасмоновой кислоты, повышают толерантность растений к абиотическому стрессу, включая засоление

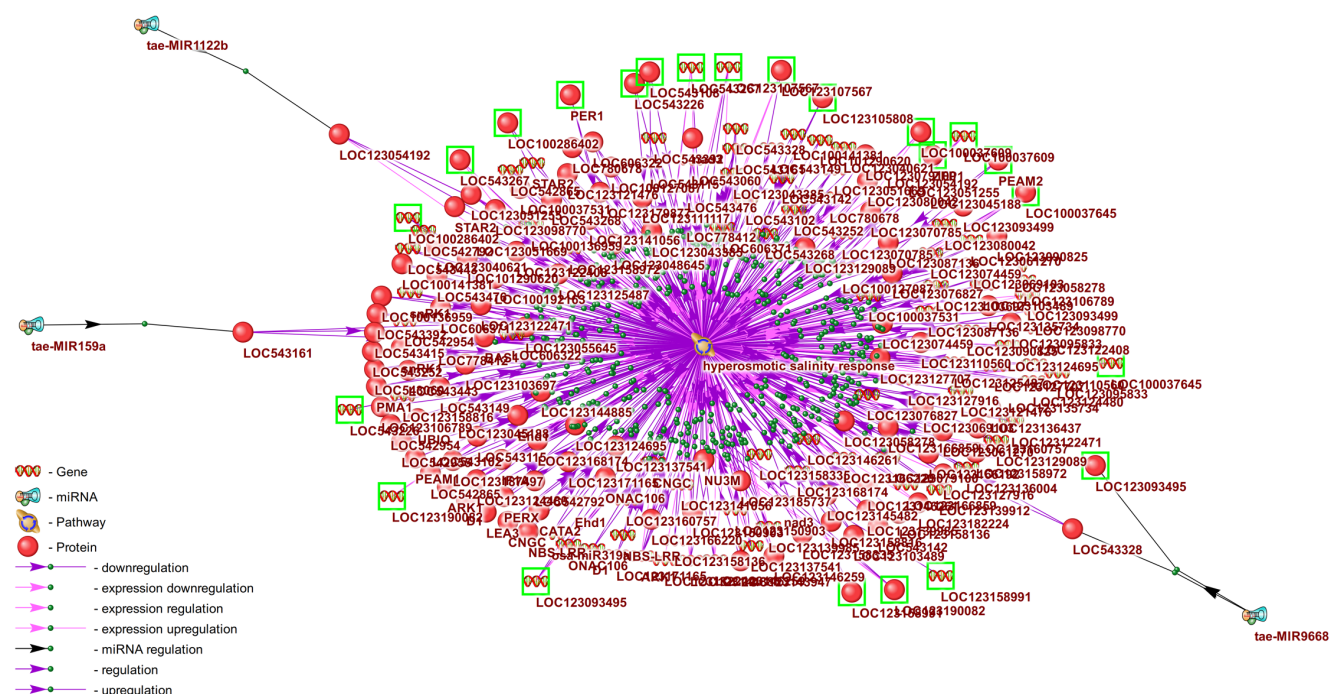


Рис. 3. Гены, белки и микроРНК, участвующие в регуляции ответа мягкой пшеницы на гиперосмотический стресс.

Зеленой рамкой отмечены гены, дифференциально экспрессирующиеся в корнях мягкой пшеницы в ответ на засоление почвы, а также их белковые продукты.

(Colebrook et al., 2014). Дополнительно в сети представлены кальций-зависимые протеинкиназы – ключевые элементы кальциевых сигнальных каскадов, активируемых при абиотическом стрессе.

Кроме того, по данным SmartCrop, в регуляцию ответа на гиперосмотический стресс вовлечены микроРНК *tae-MIR159a*, регулирующая экспрессию транскрипционного фактора TaMyb3 (LOC543161), и *tae-MIR1122b* и *tae-MIR9668*, мишенями которых являются аквапорины LOC123054192 и LOC123093495 соответственно.

Из общего списка генов, участвующих в регуляции реакции на гиперосмотический стресс, лишь девять демонстрировали дифференциальную экспрессию в тканях корня мягкой пшеницы при экспериментальном солевом стрессе в исследовании (Alyahya, Taybi, 2023). В этот перечень входят гены, кодирующие аквапорины, пероксидазы, каталазы, а также серин/треонин-протеинкиназу CTR1 (LOC100286402). Таким образом, в условиях эксперимента, описанного в статье (Alyahya, Taybi, 2023), преимущественно активировались сигнальные пути, связанные с антиоксидантной защитой.

Ассоциативная сеть, реконструированная в SmartCrop, включает эти девять ДЭГ и их белковые продукты, регуляторные белки, а также две микроРНК – *tae-MIR159a*, регулирующую экспрессию транскрипционного фактора TaMyb3 (LOC543161), и *tae-MIR9668*, мишенью которой является аквапорин LOC123093495. Данная сеть представлена на рис. 4.

Интересно отметить, что транскрипционный фактор TaMyb3 (LOC543161), являющийся мишенью микроРНК *tae-MIR159a*, в свою очередь выступает негативным регулятором экспрессии ряда генов, кодирующих перокси-

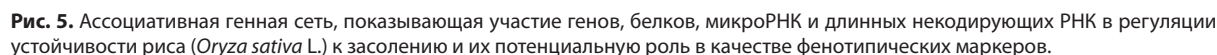
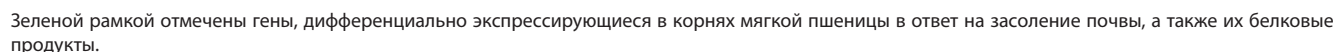
дазы. Подавление экспрессии этих ферментов приводит к повышенному накоплению перекиси водорода в тканях и, как следствие, к снижению устойчивости растений к засолению (Wei et al., 2021). В данном случае реализуется «кассетно-каскадный» принцип регуляции с участием микроРНК, при котором микроРНК контролирует экспрессию своей мишени – транскрипционного фактора, регулирующего целый набор генов, вовлеченных в реакцию на абиотический стресс (Клещев и др., 2024).

Транскрипционные факторы семейства MYB хорошо известны как участники регуляции ответов на различные абиотические стрессы, включая засоление (Kong et al., 2021; Wang S. et al., 2021). В частности, они участвуют в регуляции биосинтеза флавоноидов-метаболитов, необходимых для защиты клеток от оксидативного стресса (Wang X. et al., 2021).

Применение SmartCrop для решения задач планирования экспериментов

В качестве второго примера применения системы SmartCrop был проведен поиск перспективных генов и фенотипических маркеров для последующей маркер-ориентированной селекции и геномного редактирования, направленных на повышение устойчивости риса (*Oryza sativa* L.) к засолению почвы.

Согласно данным базы знаний SmartCrop, маркерами устойчивости к засолению могут служить такие признаки, как содержание хлорофилла, форма семян, содержание метаболитов 3'-метоксиапигенина и 5,7,4'-тригидроксиз-3'-метоксифлавона. Устойчивость риса к засолению регулируется 30 генами и соответствующими 30 белковыми продуктами (рис. 5). Помимо генов, в регуляцию вовле-



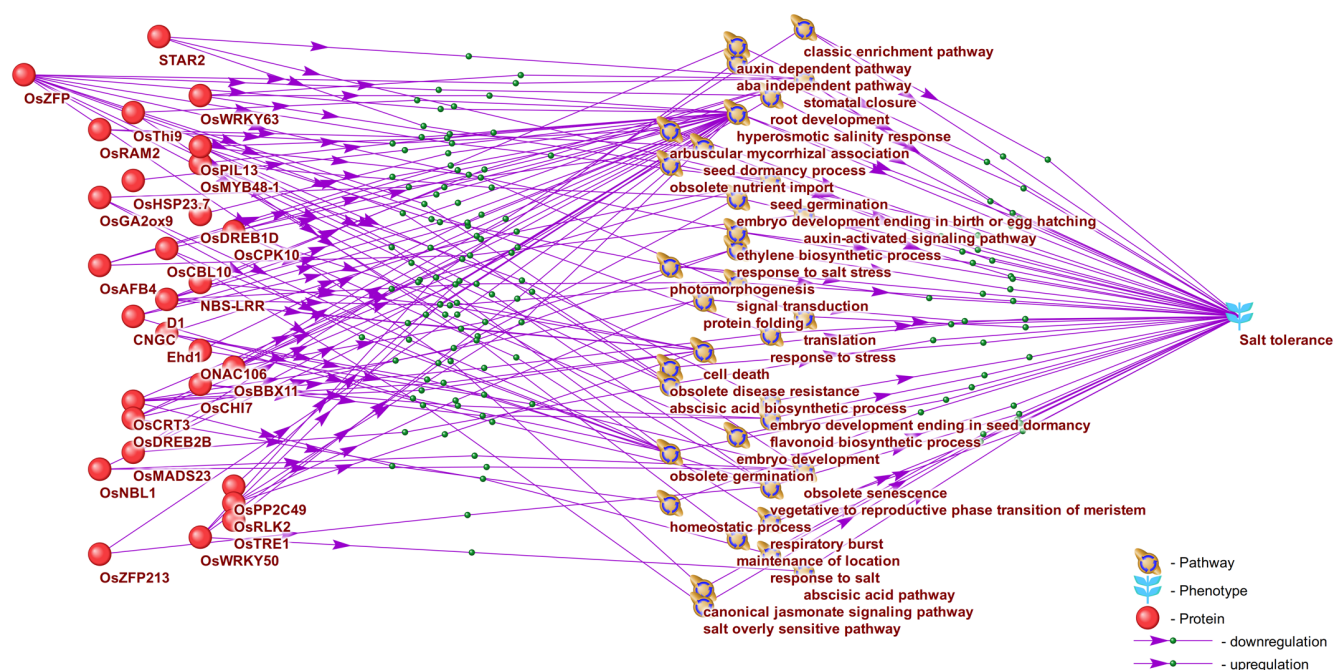


Рис. 6. Регуляторные связи генов, связанных с устойчивостью риса к засолению, с биологическими путями, вовлеченными в регуляцию данного признака.

ченны микроРНК *osa-MIR444f* и *osa-MIR444e*, мишенью которых является транскрипционный фактор OsMADS23, а также *osa-MIR444e* (мишень – ауксиновый рецептор OsABF4). Транскрипционный фактор OsBBX11, известный регулятор устойчивости к засолению (Lei et al., 2023), является мишенью для микроРНК *osa-MIR319a* и *osa-MIR396c*.

Длинные некодирующие РНК (lncRNA, днРНК) – это молекулы длиной более 200 нуклеотидов, которые регулируют экспрессию генов на транскрипционном, посттранскрипционном и эпигенетическом уровнях, тем самым модулируя ответ растений на различные абиотические и биотические факторы, включая засоление (Sun X. et al., 2018). lncRNA могут взаимодействовать с ДНК (хроматином, промоторами и энхансерами), белками, мРНК и микроРНК. Один из важных механизмов функционирования длинных некодирующих РНК – ингибирование микроРНК, что делает lncRNA ключевыми регуляторами их активности (Saha et al., 2025).

По данным SmartCrop, микроРНК *osa-MIR396c* взаимодействует с 508 длинными некодирующими РНК, из которых шесть (*LNC-Os02g06395*, *LNC-Os03g08620*, *LNC-Os03g25810*, *LNC-Os07g13605*, *LNC-Os08g32435*, *LNC-Os09g33385*) коэкспрессируются не только с *osa-MIR396c*, но и с 42 другими микроРНК риса (см. рис. 5). Это указывает на их потенциальную роль как ключевых участников регуляции устойчивости риса к абиотическим стрессам, включая засоление.

Из 30 генов, регулирующих устойчивость к засолению, шесть (*OsPIL13*, *OsNBL1*, *OsABF4*, *OsCPK10*, *OsCRT3*, *OsBBX11*) контролируют содержание хлорофилла. Остальные 24 гена ранее не связывались с известными маркерами устойчивости риса к засолению и, таким образом,

представляют собой перспективные кандидаты для поиска новых генетических маркеров данного признака.

Следует особо отметить, что приоритизация генов для маркер-ориентированной селекции и геномного редактирования требует учета специфичности их регуляторного действия, поскольку отбор по одному целевому признаку может оказывать влияние на другие хозяйственно ценные характеристики. Анализ показал, что гены и белки, регулирующие устойчивость к засолению, ассоциированы с 67 другими фенотипическими признаками, включая биомассу, площадь листьев, морфологию зерна и др., что отражает эффект плейотропии.

Три гена – *OsPIL13*, *Ehd1* и *OsGA2Ox3* – являются позитивными регуляторами устойчивости к засолению и таких хозяйственно значимых признаков, как качество зерна, период покоя семян и длина зерна. Это делает их перспективными кандидатами для селекции и геномного редактирования, поскольку воздействие на них может одновременно повышать солеустойчивость и улучшать качество зерна. В то же время гены *OsWRKY63*, *OsRAM2* и *OsABF4* повышают устойчивость риса к засолению, но при этом связаны с негативной регуляцией периода покоя семян, содержания белка в зерне и устойчивости растений к *Fusarium graminearum* и *F. pseudograminearum*, что необходимо учитывать при селекции.

Согласно данным SmartCrop, 21 ген участвует только в регуляции устойчивости к засолению и не связан с регуляцией хозяйственно ценных признаков или устойчивости к патогенам, что делает их подходящими для целенаправленной селекции на повышение солеустойчивости.

Другое важное обстоятельство, которое надо учитывать при выборе генов для маркер-ориентированной селекции и/или геномного редактирования, – это потенциальная раз-

нонаправленность их регуляторного действия, поскольку продукты генов могут как стимулировать, так и ослаблять биологические процессы, вовлеченные в позитивную или негативную регуляцию целевого признака. Для оценки такой разнонаправленности с помощью модуля «Мастер путей» программы ANDSystem были выявлены регуляторные связи белковых продуктов 30 генов, ассоциированных с устойчивостью риса к засолению, с биологическими процессами, которые, в свою очередь, участвуют в регуляции данного признака (рис. 6).

Среди генов, регулирующих устойчивость к засолению, особый интерес представляют *OsMYB48-1*, *OsCPK10*, *OsCBL10*, *OsDREB2B*, *OsRAM2* и *NBS-LRR*, обладающие однонаправленным действием в виде положительной регуляции ключевых процессов, обеспечивающих толерантность к солевому стрессу (hyperosmotic salinity response, stomatal closure, ABA independent pathway и др.). Высокая степень связанности этих генов с целевым признаком в сочетании с однонаправленным характером регуляции позволяет ожидать, что их использование в маркер-ориентированной селекции или при геномном редактировании окажет более прямое и выраженное влияние на повышение солеустойчивости по сравнению с другими кандидатами.

Заключение

Созданная база знаний SmartCrop является специализированной версией программно-информационной системы ANDSystem, адаптированной для задач генетики и селекции риса и пшеницы. Она интегрирует сведения о широком спектре объектов – генах, белках, метаболитах, некодирующих РНК, биологических процессах, селекционно значимых и фенотипических признаках, патогенах, а также биотических и абиотических факторах и их взаимосвязях. Такая архитектура обеспечивает широкие возможности для исследования молекулярно-генетических механизмов стрессоустойчивости растений, а также для отбора генов, генетических маркеров и фенотипических признаков в рамках маркер-ориентированной селекции сельскохозяйственных культур.

Примеры применения SmartCrop для функциональной аннотации дифференциально экспрессирующихся генов мягкой пшеницы при солевом стрессе и для планирования экспериментов по повышению солеустойчивости риса с использованием маркер-ориентированной селекции продемонстрировали высокую эффективность системы и ее потенциал в решении прикладных задач селекции и геномного редактирования.

Список литературы / References

Антропова Е.А., Волянская А.Р., Адамовская А.В., Деменков П.С., Яцк И.В., Иванисенко Т.В., Орлов Ю.Л., Чо Х., Чэнь М., Иванисенко В.А. Поиск перспективных генетических маркеров, ассоциированных с молекулярными механизмами снижения устойчивости риса к *Rhizoctonia solani* при избытке азотных удобрений, методом реконструкции и анализа генных сетей. *Вавиловский журнал генетики и селекции*. 2024;28(8):960-973. doi 10.18699/vjgb-24-103

[Antropova E.A., Volyanskaya A.R., Adamovskaya A.V., Demenkov P.S., Yatsyk I.V., Ivanisenko T.V., Orlov Y.L., Haoyu Ch., Chen M., Ivanisenko V.A. Computational identification of promis-

ing genetic markers associated with molecular mechanisms of reduced rice resistance to *Rhizoctonia solani* under excess nitrogen fertilization using gene network reconstruction and analysis methods. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):960-973. doi 10.18699/vjgb-24-103]

Деменков П.С., Сайк О.В., Иванисенко Т.В., Колчанов Н.А., Кочетов А.В., Иванисенко В.А. Приоритизация генов картофеля, вовлеченных в формирование селекционно-значимых признаков, с использованием базы знаний SOLANUM TUBEROSUM. *Вавиловский журнал генетики и селекции*. 2019;23(3):312-319. doi 10.18699/VJ19.501

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

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

[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]

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

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

Клещев М.А., Мальцева А.В., Антропова Е.А., Деменков П.С., Иванисенко Т.В., Орлов Ю.Л., Чо Х., Чэнь М., Колчанов Н.А., Иванисенко В.А. Реконструкция и компьютерный анализ генной сети, отражающей роль микроРНК в регуляции ответа пшеницы на засуху. *Вавиловский журнал генетики и селекции*. 2024;28(8):904-917. doi 10.18699/vjgb-24-98

[Kleshchev M.A., Maltseva A.V., Antropova E.A., Demenkov P.S., Ivanisenko T.V., Orlov Y.L., Chao H., Chen M., Kolchanov N.A., Ivanisenko V.A. Reconstruction and computational analysis of the microRNA regulation gene network in wheat drought response mechanisms. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(8):904-917. doi 10.18699/vjgb-24-98]

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

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

Alyahya N., Taybi T. Comparative transcriptomic profiling reveals differentially expressed genes and important related metabolic pathways in shoots and roots of a Saudi wheat cultivar (Najran) under salinity stress. *Front Plant Sci*. 2023;14:1225541. doi 10.3389/fpls.2023.1225541

- Ayadi M., Brini F., Masmoudi K. Overexpression of a wheat aquaporin gene, *TdPIP2;1*, enhances salt and drought tolerance in transgenic durum wheat cv. Maali. *Int J Mol Sci.* 2019;20(10):2389. doi 10.3390/ijms20102389
- Chao H., Zhang S., Hu Y., Ni Q., Xin S., Zhao L., Ivanisenko V.A., Orlov Y.L., Chen M. Integrating omics databases for enhanced crop breeding. *J Integr Bioinform.* 2023;20(4):20230012. doi 10.1515/jib-2023-0012
- Chen T., Nomura K., Wang X., Sohrabi R., Xu J., Yao L., Paasch B.C., Ma L., Kremer J., Cheng Y., Zhang L., Wang N., Wang E., Xin X.F., He S.Y. A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature.* 2020;580(7805):653-657. doi 10.1038/s41586-020-2185-0
- Cheng M., Zhu Y., Yu H., Shao L., Zhang Y., Li L., Tu H., ... Orlov Y.L., Chen D., Wong A., Yang Y.E., Chen M. Non-coding RNA notations, regulations and interactive resources. *Funct Integr Genomics.* 2024;24(6):217. doi 10.1007/s10142-024-01494-w
- Colebrook E.H., Thomas S.G., Phillips A.L., Hedden P. The role of gibberellin signalling in plant responses to abiotic stress. *J Exp Biol.* 2014;217(1):67-75. doi 10.1242/jeb.089938
- 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(3-4):149-161. doi 10.3233/isb-2012-0449
- Do H., Than K., Larmande P. Evaluating named-entity recognition approaches in plant molecular biology. In: Multi-disciplinary Trends in Artificial Intelligence. MIWAI 2018. Lecture Notes in Computer Science. Vol. 11248. Springer, Cham., 2018;219-225. doi 10.1007/978-3-030-03014-8_19
- D'Souza J. Agriculture named entity recognition – towards FAIR, reusable scholarly contributions in agriculture. *Knowledge.* 2024; 4(1):1-26. doi 10.3390/knowledge4010001
- Gemma Team, Google DeepMind. Gemma 2: improving open language models at a practical size. *arXiv.* 2024:2408.00118. doi 10.48550/arXiv.2408.00118
- Guindani L.G., Oliveirai G.A., Ribeiro M.H.D.M., Gonzalez G.V., de Lima J.D. Exploring current trends in agricultural commodities forecasting methods through text mining: developments in statistical and artificial intelligence methods. *Heliyon.* 2024;10(23):e40568. doi 10.1016/j.heliyon.2024.e40568
- Gupta A., Shaw B.P., Sahu B.B. Post-translational regulation of the membrane transporters contributing to salt tolerance in plants. *Funct Plant Biol.* 2021;48(12):1199-1212. doi 10.1071/FP21153
- Hamilton W.L., Ying R., Leskovec J. Inductive representation learning on large graphs. *arXiv.* 2017. doi 10.48550/arxiv.1706.02216
- Hoai P.T.T., Tyerman S.D., Schnell N., Tucker M., McGaughey S.A., Qiu J., Groszmann M., Byrt C.S. Deciphering aquaporin regulation and roles in seed biology. *J Exp Bot.* 2020;71(6):1763-1773. doi 10.1093/jxb/erz555
- Islamaj R., Wei C.H., Cissel D., Miliaras N., Printseva O., Rodionov O., Sekiya K., Ward J., Lu Z. NLM-Gene, a richly annotated gold standard dataset for gene entities that addresses ambiguity and multi-species gene recognition. *J Biomed Inform.* 2021;118:103779. doi 10.1016/j.jbi.2021.103779
- Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. Text mining on PubMed. In: Chen M., Hofestädt R. (Eds) Approaches in Integrative Bioinformatics. Springer, 2014;161-170. doi 10.1007/978-3-642-41281-3_6
- 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(S11):228. doi 10.1186/s12859-020-03557-8
- Ivanisenko T.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. The new version of the ANDDigest tool with improved AI-based short names recognition. *Int J Mol Sci.* 2022;23(23):14934. doi 10.3390/ijms232314934
- Ivanisenko T.V., Demenkov P.S., Ivanisenko V.A. An accurate and efficient approach to knowledge extraction from scientific publications using structured ontology models, graph neural networks, and large language models. *Int J Mol Sci.* 2024;25(21):11811. doi 10.3390/ijms252111811
- Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol.* 2015;9(Suppl.2):S2. doi 10.1186/1752-0509-9-s2-s2
- Ivanisenko 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(S1):34. doi 10.1186/s12859-018-2567-6
- Kanehisa M. Molecular network analysis of diseases and drugs in KEGG. In: Mamitsuka H., DeLisi C., Kanehisa M. (Eds) Data Mining for Systems Biology. Methods in Molecular Biology. Vol. 939. Humana Press, 2013;263-275. doi 10.1007/978-1-62703-107-3_17
- Kong W., Sun T., Zhang C., Deng X., Li Y. Comparative transcriptome analysis reveals the mechanisms underlying differences in salt tolerance between *indica* and *japonica* rice at seedling stage. *Front Plant Sci.* 2021;12:725436. doi 10.3389/fpls.2021.725436
- Krallinger M., Rabal O., Leitner F., Vazquez M., Salgado D., Lu Z., Leaman R., ... Alves R., Segura-Bedmar I., Martínez P., Oyarzabal J., Valencia A. The ChEMDNER corpus of chemicals and drugs and its annotation principles. *J Cheminform.* 2015;7(S1):S2. doi 10.1186/1758-2946-7-s1-s2
- Lei L., Cao L., Ding G., Zhou J., Luo Y., Bai L., Xia T., ... Xie T., Yang G., Wang X., Sun S., Lai Y. *OsBBX11* on *qSTS4* links to salt tolerance at the seedling stage in *Oryza sativa* L. ssp. *Japonica*. *Front Plant Sci.* 2023;14:1139961. doi 10.3389/fpls.2023.1139961
- Lesk C., Rowhani P., Ramankutty N. Influence of extreme weather disasters on global crop production. *Nature.* 2016;529(7584):84-87. doi 10.1038/nature16467
- Liu L., Liu E., Hu Y., Li S., Zhang S., Chao H., Hu Y., Zhu Y., Chen Y., Xie L., Shen Y., Wu L., Chen M. ncPlantDB: a plant ncRNA database with potential ncPEP information and cell type-specific interaction. *Nucleic Acids Res.* 2025;53(D1):D1587-D1594. doi 10.1093/nar/gkae1017
- Mittler R. Abiotic stress, the field environment and stress combination. *Trends Plant Sci.* 2006;11(1):15-19. doi 10.1016/j.tplants.2005.11.002
- Naithani S., Gupta P., Preece J., D'Eustachio P., Elser J.L., Garg P., Dikeman D.A., ... Bolton E., Papatheodorou I., Stein L., Ware D., Jaiswal P. Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res.* 2020;48(D1):D1093-D1103. doi 10.1093/nar/gkz996
- Ni Y., Aghamirzaie D., Elmarakeby H., Collakova E., Li S., Grene R., Heath L.S. A machine learning approach to predict gene regulatory networks in seed development in Arabidopsis. *Front Plant Sci.* 2016;7:1936. doi 10.3389/fpls.2016.01936
- Nykiel M., Gietler M., Fidler J., Prabucka B., Labudda M. Abiotic stress signaling and responses in plants. *Plants.* 2023;12(19):3405. doi 10.3390/plants12193405
- Otasek D., Morris J.H., Bouças J., Pico A.R., Demchak B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185. doi 10.1186/s13059-019-1758-4
- Pearson H. Biology's name game. *Nature.* 2001;411(6838):631-632. doi 10.1038/35079694
- Saha C., Saha S., Bhattacharyya N.P. LncRNAOmics: a comprehensive review of long non-coding RNAs in plants. *Genes.* 2025;16(7):765. doi 10.3390/genes16070765
- Shewry P.R., Hey S.J. The contribution of wheat to human diet and health. *Food Energy Secur.* 2015;4(3):178-202. doi 10.1002/fes3.64

- Shrestha A.M.S., Gonzales M.E.M., Ong P.C.L., Larmande P., Lee H.S., Jeung J.U., Kohli A., Chebotarov D., Mauleon R.P., Lee J.S., McNally K.L. RicePilaf: a post-GWAS/QTL dashboard to integrate pangenomic, coexpression, regulatory, epigenomic, ontology, pathway, and text-mining information to provide functional insights into rice QTLs and GWAS loci. *GigaScience*. 2024;13:giae013. doi 10.1093/gigascience/giae013
- Statello L., Guo C.J., Chen L.L., Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*. 2021;22:96-118. doi 10.1038/s41580-020-00315-9
- Sun X., Zheng H., Sui N. Regulation mechanism of long non-coding RNA in plant response to stress. *Biochem Biophys Res Commun*. 2018;503(2):402-407. doi 10.1016/j.bbrc.2018.07.072
- Supriya P., Srividya G.K., Solanki M., Manvitha D., Prakasam V., Balakrishnan M., Neeraja C.N., Rao C.S., Sundaram R.M., Mangrauthia S.K. Identification and expression analysis of long non-coding RNAs of rice induced during interaction with *Rhizoctonia solani*. *Physiol Mol Plant Pathol*. 2024;134:102389. doi 10.1016/j.pmpp.2024.102389
- Szklarczyk D., Gable A.L., Nastou K.C., Lyon D., Kirsch R., Pyysalo S., Doncheva N.T., Legeay M., Fang T., Bork P., Jensen L.J., von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605-D612. doi 10.1093/nar/gkaa1074
- Tian F., Yang D.C., Meng Y.Q., Jin J., Gao G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res*. 2020;48(D1):D1104-D1113. doi 10.1093/nar/gkz1020
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. *arXiv*. 2017. doi 10.48550/arXiv.1706.03762
- Verma V., Ravindran P., Kumar P.P. Plant hormone-mediated regulation of stress responses. *BMC Plant Biol*. 2016;16(1):86. doi 10.1186/s12870-016-0771-y
- Virlouvet L., Avenson T.J., Du Q., Zhang C., Liu N., Fromm M., Avramova Z., Russo S.E. Dehydration stress memory: gene networks linked to physiological responses during repeated stresses of *Zea mays*. *Front Plant Sci*. 2018;9:1058. doi 10.3389/fpls.2018.01058
- Volyanskaya A.R., Antropova E.A., Zubairova U.S., Demenkov P.S., Venzel A.S., Orlov Y.L., Makarova A.A., Ivanisenko T.V., Gorshkova T.A., Aglyamova A.R., Kolchanov N.A., Chen M., Ivanisenko V.A. Reconstruction and analysis of the gene regulatory network for cell wall function in *Arabidopsis thaliana* L. leaves in response to water deficit. *Vavilov J Genet Breed*. 2023;27(8):1031-1041. doi 10.18699/vjgb-23-118
- Wang S., Shi M., Zhang Y., Xie X., Sun P., Fang C., Zhao J. FvMYB24, a strawberry R2R3-MYB transcription factor, improved salt stress tolerance in transgenic Arabidopsis. *Biochem Biophysical Res Commun*. 2021;569:93-99. doi 10.1016/j.bbrc.2021.06.085
- Wang X., Niu Y., Zheng Y. Multiple functions of MYB transcription factors in abiotic stress responses. *Int J Mol Sci*. 2021;22(11):6125. doi 10.3390/ijms22116125
- Wei T., Guo D., Liu J. PtrMYB3, a R2R3-MYB transcription factor from *Poncirus trifoliata*, negatively regulates salt tolerance and hydrogen peroxide scavenging. *Antioxidants*. 2021;10(9):1388. doi 10.3390/antiox10091388
- Yan J., Wang X. Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *Plant J*. 2022;111(6):1527-1538. doi 10.1111/tpj.15905
- Zhang D., Zhao R., Xian G., Kou Y., Ma W. A new model construction based on the knowledge graph for mining elite polyphenotype genes in crops. *Front Plant Sci*. 2024;15:1361716. doi 10.3389/fpls.2024.1361716
- Zhao S., Zhang Q., Liu M., Zhou H., Ma C., Wang P. Regulation of plant responses to salt stress. *Int J Mol Sci*. 2021;22(9):4609. doi 10.3390/ijms22094609

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 18.08.2025. После доработки 01.10.2025. Принята к публикации 06.10.2025.