

Предсказание методами системной биологии наиболее перспективных генов-мишеней для селекции на устойчивость к окислительному стрессу C_3 и C_4 культурных злаков

А.В. Дорошков^{1, 2}✉, А.В. Бобровских^{1, 2}

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Активные формы кислорода (АФК) – один из ключевых повреждающих факторов для живых организмов. АФК производятся в реакциях нормального метаболизма, в стрессовых условиях их выработка повышается. Улучшение характеристик ферментативной системы антиоксидантной защиты культурных растений позволит повысить их устойчивость к абиотическим стрессам, таким как засоленность, засуха и холод. Однако компоненты системы вырождены – каждая реакция катализируется серией ферментов, кодируемых разными генами. Выбор наиболее важных компонентов позволит ускорить нахождение оптимальной селекционной стратегии для улучшения свойств всей системы у хозяйственно ценных видов растений. В настоящей работе впервые проведен системно-биологический анализ особенностей молекулярной эволюции и характеристик экспрессии генов, принадлежащих к четырем функциональным группам ферментов антиоксидантной защиты (APX, GPX, SOD и CAT), у шести представителей C_3 и трех представителей C_4 злаковых растений. Выделены и проанализированы 25 ортологических групп генов. Выявлены шесть ортологических групп с наиболее высоким уровнем экспрессии и наибольшим давлением стабилизирующего отбора для дальнейшей верификации и использования в селекции. Эти шесть ортологических групп, предположительно, вносят большой вклад в функционирование антиоксидантной системы изученных C_3 и C_4 злаковых растений. Показано, что интеграция эволюционных характеристик и экспрессионных данных представляет перспективный подход к предсказанию важных для селекции генов в геномах хозяйственно ценных растений.

Ключевые слова: активные формы кислорода; антиоксидантная система; отношение K_a/K_s ; уровень экспрессии; гены-кандидаты.

Using the methods of systems biology for predicting perspective target genes to select C_3 and C_4 cereals for oxidative stress resistance

A.V. Doroshkov^{1, 2}✉, A.V. Bobrovskikh^{1, 2}

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

Reactive oxygen species (ROS) are some of the most damaging factors for living systems. Cells produce ROS during normal metabolism reactions, but ROS production increases under stressful conditions. Improving the antioxidant system in cultivated plants will increase their tolerance to abiotic stresses, such as salinity, drought and cold. However, the biochemical components of the system are redundant, for each reaction is catalyzed by a series of enzymes encoded by different genes. Choosing the most perspective components of this system will help speed up evaluating the optimal breeding strategy for improving abiotic stress tolerance in economically valuable plants. In the present research article, we present the results of an integrative analysis of evolution- and expression-related characteristics. The work was carried out on a series of genes that belong to 4 functional groups (APX, GPX, SOD and CAT) of enzymatic components of the antioxidant defense system in six species of C_3 cereal plants and 3 species of C_4 cereal plants. As a result, 25 groups of orthologous genes were evaluated and described. The highest gene expression level and the greatest pressure of purifying selection were found to characterize six groups. These genes were chosen for further verification and use in breeding. Because these genes undergo the most conservative evolution and have the highest level of mRNA expression, we may assume that they contribute a lot to the antioxidant system functioning of the C_3 and C_4 cereal plants studied. We have shown that the integration of evolutionary characteristics and expression data represents a promising approach to predict target genes for plant breeding.

Key words: reactive oxygen species; antioxidant system; K_a/K_s ratio; expression level; candidate genes.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Дорошков А.В., Бобровских А.В. Предсказание методами системной биологии наиболее перспективных генов-мишеней для селекции на устойчивость к окислительному стрессу C_3 и C_4 культурных злаков. Вавиловский журнал генетики и селекции. 2018;22(1):122-131. DOI 10.18699/VJ18.339

HOW TO CITE THIS ARTICLE:

Doroshkov A.V., Bobrovskikh A.V. Using the methods of systems biology for predicting perspective target genes to select C_3 and C_4 cereals for oxidative stress resistance. Vavilovskii Zhurnal Genetiki i Selektcii = Vavilov Journal of Genetics and Breeding. 2018;22(1):122-131. DOI 10.18699/VJ18.339 (in Russian)

Набор современных методов генетики и селекции позволяет эффективно улучшать хозяйственно ценные качества культурных растений, особенно в части признаков, зависящих от одного-двух генов «главного эффекта». Одновременно с этим селекция растений в условиях постоянно растущей потребности в пищевых ресурсах вынуждает использовать все возможности улучшения пищевых культур в комплексе. Расширение спектра используемых генов и целостное понимание функционирования сложных биологических подсистем, ассоциированных с признаками, будут способствовать более эффективной селекции.

Важное направление селекции растений – повышение их устойчивости к стрессовым факторам среды. Существенным поражающим фактором во время стресса являются активные формы кислорода (АФК). В нормальном метаболизме растений существует множество источников АФК – реакции фотосинтеза, дыхания, окисления различных субстратов. Опасность АФК состоит в инициации каскадных реакций, которые усиливают продукцию гидроксидов и липидных пероксидов, возникающих в мембранах клетки (Noctor, Foyer, 1998). Повреждение компонентов клетки под действием АФК называется окислительным стрессом. Установлено, что патогены, ранения или неблагоприятные условия окружающей среды приводят к многократному увеличению производства АФК с участием NAD(P)H-оксидаз (Hammond-Kosack, Jones, 1996; Orozco-Cardenas, Ryan, 1999; Gill, Tuteja, 2010; Das, Roychoudhury, 2014).

К числу наиболее важных поставщиков АФК у растений относятся реакции фотосинтеза. Известно, что некоторые культурные злаки, такие как кукуруза, отличаются особой формой фотосинтеза – C_4 , первичным продуктом которого является не 3-фосфоглицериновая кислота, а щавелевоуксусная (Hatch et al., 1987, 2005). Этот путь более эффективный, однако требует большей инсоляции, что само по себе усиливает выработку АФК. Вместе с тем для C_4 пути практически не характерно фотодыхание (Ehleringer et al., 1991; Sage et al., 2012), а связанное с ним окисление гликолевой кислоты в пероксисомах служит основным источником токсичного пероксида водорода в фотосинтезирующей растительной клетке (Sage et al., 2012).

Для борьбы с окислительным стрессом живые организмы в процессе эволюции развили специальную систему ферментов антиоксидантной защиты. Оценочное время появления первых ферментов антиоксидантной защиты в процессе эволюции соответствует появлению фотосинтеза – примерно 2.4 млрд лет назад (Hohmann-Marriott, Blankenship, 2011). У современных растений присутствует множество ферментов антиоксидантной защиты, находящихся в тесных взаимодействиях (Gill, Tuteja, 2010). Основные пути нейтрализации АФК в растительной клетке представлены на рис. 1.

В процессе метаболизма клетки и воздействия стрессовых факторов накапливаются побочные продукты – частично восстановленные формы атмосферного кислорода: синглетный кислород, супероксидный радикал (см. рис. 1), пероксид водорода, гидроксильный радикал и др. Далее они преобразуются супероксиддисмутазой (SOD) в пероксиды, которые в дальнейшем расщепляются ка-

талазой (CAT), глутатионпероксидазой (GPX) либо вовлекаются аскорбатпероксидазой (APX) в аскорбат-глутатионовый цикл (см. рис. 1). Дополнительные акцепторы электронов для упрощения на схеме не приведены.

Активность антиоксидантной системы увеличивает устойчивость растений к высоким температурам (Wang et al., 2014), засоленности (Gill et al., 2013; Khaliq et al., 2015) и другим стрессам (Osipova et al., 2016). Следует отметить, что каждый фермент системы антиоксидантной защиты представлен в геноме растений в нескольких копиях. Например, в геноме *Arabidopsis thaliana* присутствует не менее 25 генов ферментов только этих четырех классов (SOD, CAT, GPX, APX) (Mittler et al., 2004). В связи с многокопийностью генов возникает дополнительная сложность в использовании маркер-опосредованной селекции растений на устойчивость к окислительному стрессу.

Перспективным решением является использование огромного массива геномных данных, накопленных как для модельных, так и для хозяйственно ценных растений. Прежде всего, это геномные последовательности и различные полногеномные экспрессионные эксперименты. Используя эту информацию, становится возможным проанализировать молекулярную эволюцию, выявить ортологические группы генов для серии видов, сопоставить характеристики накопления замен для каждой отдельной копии гена каждого фермента, а также соотнести их уровни экспрессии. Неоднородность относительных скоростей накопления замен в серии копий свидетельствует в пользу различного давления стабилизирующего отбора и позволяет выявить более эволюционно консервативные копии, которые, как правило, являются наиболее функционально важными (Zhang, 2003). Вместе с тем различный уровень экспрессии однородных генов (например, кодирующих ферменты, катализирующие одну и ту же реакцию) также свидетельствует в пользу неодинаковой функциональной нагрузки: чем выше уровень экспрессии, тем потенциально более высокий вклад вносит соответствующая копия (Li et al., 2005).

Культурные злаки тоже различаются по биохимии фотосинтеза. Такие виды, как пшеница (*Triticum aestivum*), рис (*Oryza sativa*) и ячмень (*Hordeum vulgare*), имеют C_3 фотосинтез, тогда как кукурузу (*Zea mays*) и сорго (*Sorghum bicolor*) отличает C_4 фотосинтез (Карпилов, 1960; Sage et al., 2012). Потенциально C_4 фотосинтез снижает нагрузку на ферменты антиоксидантной системы, утилизирующие пероксид (APX, SOD, CAT), поэтому относительные уровни экспрессии компонентов антиоксидантной системы также могут различаться между видами.

В настоящей работе применен системный подход для поиска и предсказания генов-мишеней, ассоциированных с устойчивостью к окислительному стрессу, основанный на привлечении эволюционных характеристик генов и уровней экспрессии их мРНК у C_3 и C_4 культурных злаков. В связи с различиями в биохимии фотосинтеза нами изучены полногеномные экспрессионные данные для растений обоими типами фотосинтеза. Полученные данные позволяют выявить наиболее консервативные и наиболее высокоэкспрессирующиеся копии генов-ферментов антиоксидантной защиты у девяти видов злаков (*Triticum aestivum*, *T. urartu*, *Aegilops tauschii*, *Hordeum vulgare*, *Zea mays*,

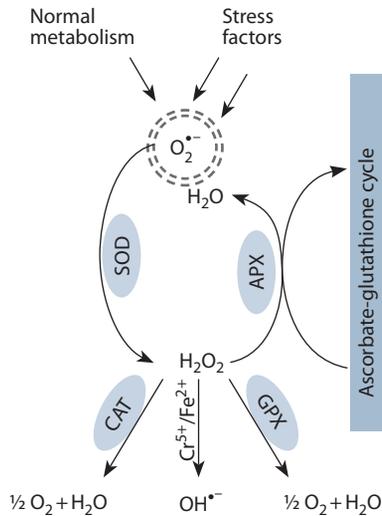


Fig. 1. Main pathways of ROS neutralization
See text for details.

Brachypodium distachyon, *Sorghum bicolor*, *Setaria italica*, *Oryza sativa* ssp. *japonica*) с целью дальнейшего использования их в селекции.

Материал и методы

Выборка видов. Для анализа особенностей эволюции генов антиоксидантной системы у C_3 и C_4 культурных злаков была сформирована следующая выборка видов растений:

- C_4 злаки: *Zea mays*, *Sorghum bicolor*, *Setaria italica*;
- C_3 злаки: *Oryza sativa* ssp. *japonica*, *Brachypodium distachyon*, *Hordeum vulgare*, *Triticum aestivum*, *T. urartu*, *Aegilops tauschii*;
- внешняя группа 1 (двудольные растения): *Arabidopsis thaliana*, *Vitis vinifera*, *Solanum lycopersicum*, *S. tuberosum*;
- внешняя группа 2 (одноклеточная зеленая водоросль): *Chlamydomonas reinhardtii*.

Такая выборка позволила изучить характер событий дупликации у представителей злаковых растений и ряда видов двудольных растений.

Поиск гомологов генов. Для поиска ДНК и белковых последовательностей в геномах и протеомах растительных организмов была использована база данных PLAZA 3.0, а для *T. aestivum*, *T. urartu*, *Ae. tauschii* – дополнительно ENSEMBL PLANTS (<http://plants.ensembl.org>). Поиск по базам данных выполнялся алгорит-

мом Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1990). На первом этапе для каждого исследуемого гена производили реципрокный поиск гомологов в базе данных с пороговым значением $E_{value} = 10^{-7}$. При поиске в PLAZA учитывали также ее встроенную автоматическую привязку к гомологичным группам. Исходя из выравнивания, выбирали такие белки, которые содержали в себе не менее половины последовательности целевого белка. Такие белки считали гомологами исходного белка.

Множественное выравнивание и филогенетический анализ. Анализ молекулярной эволюции проводили при помощи пакета программ SAMMEM v. 0.82 (Gunbin et al., 2012). Исходная выборка гомологичных последовательностей подвергалась множественному выравниванию последовательностей алгоритмом Mafft 6.717 (Kato, Toh, 2008) с использованием матрицы BLOSUM 62 (Henikoff S., Henikoff J., 1992). После этого осуществлялась экспертная оценка выравнивания и выделение консервативных мотивов. Построение модели замен аминокислот на основе множественного выравнивания проводилось алгоритмом ModelEstimator (Arvestad, 2006). Построение филогенетического дерева на основе модели замен осуществлялось алгоритмами: FastTree 2.1.1 (Arvestad, 2006) – для оценки топологии и Phym1 (Guindon, Gascuel, 2003) – для окончательной оптимизации. Для проверки устойчивости узлов ветвления дерева использовали многократную генерацию выборок методом bootstrap. Визуализация деревьев и анализ топологии выполнены в программах FigTree v.1.4.2 (Rambaut, 2008) и Archaeopteryx (Han, Zmasek, 2009; Zmasek, 2015).

Выделенные клады, состоящие из последовательностей однодольных злаковых растений, были затем соотнесены с внешними группами двудольных растений и последовательностями водоросли. Исходя из топологии крупных клад на филогенетических деревьях и их видового состава, судили о ранних событиях дупликации. Эволюционно более поздние дупликации, возникшие после дивергенции цветковых растений на двудольные и однодольные, оценивали путем подсчета количества последовательностей ферментов в соответствующих кладах. Подсчет соотношения замен осуществлялся с использованием *Ka/Ks Calculation tool* (<http://services.cbu.uib.no/tools/kaks>) для клад однодольных растений. На вход сервиса подавалось предварительно оцененное экспертом на предмет протяженных делеций множественное выравнивание каждой клады последовательностей однодольных растений. Реконструкция топологии филогенетических отношений и последовательностей предков проводилась сервисом *Ka/Ks Calculation tool*. Полученная топология сопоставлялась с независимо рассчитанной нами. Отношения частот K_a/K_s для всех ветвей были извлечены и использованы для дальнейшего расчета средних значений и доверительных интервалов и приведены в виде «ящиков с усами».

Анализ полногеномных экспрессионных экспериментов. Поиск экспрессионных данных осуществляли в базе GEO DataSets NCBI (<https://www.ncbi.nlm.nih.gov/gds>). Для этого выбирали эксперименты, в которых присутствовало по три повторности тотальной РНК либо мРНК у растений с нормальным генотипом в стандартных условиях в корне и/или в листе. Были выбраны следующие эксперименты: *H. vulgare* – GSE56437 (лист), GSE6325 (корень); *O. sativa* – GSE14300 (корень и лист); *Z. mays* – GSE50191 (корень и лист). Далее проводился корреляционный анализ уровня экспрессии изучаемых генов между корнем и листом, а также между видами растений. В работе привлечены данные об уровне экспрессии генов в экспериментах с использованием как ДНК-микрочипов (GSE56437, GSE6325, GSE14300), так и высокопроизводительного секвенирования (GSE50191) на платформе Illumina. Данные высокопроизводительного секвенирования предварительно нормировали на общий размер библиотеки (число ридов в эксперименте) и длину целевого гена (UTR + CDS), было рассчитано значение FPKM (fragments per kilobase million). Известно, что микрочиповые данные и данные высокопроизводительного секвенирования имеют разные размерности. Поэтому для сопоставления их между разными экспериментами была использована суммарная доля мРНК каждой копии фермента, отнесенная к общему количеству мРНК функционально однородных ферментов, например APX. Такая величина универсальна,

ее можно рассчитать как из микрочиповых данных, так и из данных высокопроизводительного секвенирования.

Для расчета коэффициентов корреляции Пирсона, применения *t*-критерия Стьюдента, вычисления квантилей и построения графиков использовался пакет PAST STATISTICS (Hammer et al., 2001).

Результаты

Для выявления соответствия генов ферментов антиоксидантной защиты у разных видов растений был проведен поиск гомологов и реконструкция филогенетических отношений. Результаты приведены в таблице, а топологии деревьев – в Приложениях 1–4¹.

В геномах 14 видов растительных организмов было найдено 110 гомологичных последовательностей APX. Филогенетические отношения гомологов приведены в Приложении 1. При анализе топологии дерева было выделено восемь ортологических групп последовательностей злаков – А–Н (см. Приложение 1). Следует отметить, что в основном последовательности в ортологических группах (кладах) представлены единичными копиями, однако в кладах В и С часть организмов не представлена (например, *T. urartu*, *Ae. tauschii*), а в кладе D произошла дупликация гена у *O. sativa*. Клады А и D имеют собственную внешнюю группу из последовательностей двудольных растений, а клады Е и F, а также G и Н являются результатом дупликаций после обособления однодольных растений и имеют по одной общей внешней группе на пару (см. Приложение 1).

Для анализа молекулярной эволюции GPX был сформирован набор гомологичных последовательностей в количестве 110 шт. Филогенетические отношения гомологов приведены в Приложении 2, а данные о соответствии последовательностей – в таблице. Последовательности злаков кластеризуются в семь монофилетических групп. В основном последовательности в кладах представлены единичными копиями, однако *Z. mays* имеет молодые дупликации генов в кладах С и Е. Клады А, В, F, G имеют собственные внешние группы из последовательностей двудольных растений, а клады С, D, Е являются результатом дупликаций после обособления однодольных растений и имеют одну общую внешнюю группу двудольных (см. Приложение 2). Фрагмент дерева, содержащий клады F и G, имеет слабую бутстрэп-поддержку базальной топологии. В связи с этим последовательности двудольных растений отнесены в настоящей работе к кладе GPX G на основе дополнительного анализа эволюционных расстояний между последовательностями двудольных и однодольных растений в рассматриваемом фрагменте дерева.

Далее был сформирован набор из 100 гомологичных последовательностей SOD. Филогенетические отношения гомологов приведены в Приложении 3, а данные о составе ортологических групп и соответствии последовательностей между видами представлены в таблице. Последовательности злаков кластеризуются в семь монофилетических групп. *A. thaliana* и *Z. mays* имеют относительно молодые дупликации генов (после обособления эволюционных линий изучаемых видов друг от друга) в кладах

А, В, F и С, Е соответственно. Клады А, В, F, G имеют собственную внешнюю группу из последовательностей двудольных растений. В кладах В и Е произошли относительно молодые дупликации генов (*Z. mays*, *O. sativa*, а также во внешних группах – *A. thaliana*, *S. lycopersicum*, *S. tuberosum*, *V. vinifera*).

Для филогенетического анализа CAT был сформирован набор из 39 гомологичных последовательностей. Филогенетические отношения гомологов приведены в Приложении 4, а данные о последовательностях – в таблице. Последовательности злаков кластеризуются в три монофилетические группы. При этом лишь клада С имеет соответствующую внешнюю группу из последовательностей двудольных.

Затем для каждой монофилетической группы злаковых растений был сделан расчет отношения несинонимичных замен к синонимичным (K_a/K_s). Результаты представлены на рис. 2, *б–з*.

Для каждой диаграммы приведена схема филогенетических отношений клад. Средние значения K_a/K_s для всех клад не превосходят 0.5, что говорит о стабилизирующем характере отбора каждой из копий в выборке видов. Следует отметить, что давление отбора на разные копии различно, для APX и SOD значения K_a/K_s между копиями достоверно различаются (вероятность принадлежности к одной генеральной совокупности $P < 0.0005$; Приложение 5), а для GPX и CAT значения этой характеристики между копиями не показывают высокой достоверности (см. рис. 2 и Приложение 5). Анализ значений K_a/K_s позволяет выделить копии, находящиеся под более сильным давлением стабилизирующего отбора ($\langle K_a/K_s \rangle < 0.2$). Для APX это представителем клады Е, F, G, для SOD – В и Е, для GPX такую тенденцию проявляют последовательности клады Е. Примечательно, что между копиями CAT не выявляется заметных различий по значениям K_a/K_s , они находятся под относительно высоким давлением стабилизирующего отбора ($\langle K_a/K_s \rangle < 0.2$).

В дальнейшем анализе использовали полногеномные экспрессионные данные, с помощью которых рассчитывалась доля мРНК каждой копии фермента во всем объеме мРНК фермента для каждой из четырех функциональных групп (APX, GPX, SOD, CAT) в тканях корня и листа по отдельности. Данные по экспрессии ортологичных генов в разных видах растений во всех случаях выявили положительную корреляцию ($P \ll 0.0005$). Графики рассеяния усредненных по корню и листу значений экспрессии ортологичных генов и соответствующие коэффициенты корреляции приведены в Приложении 6. Интересно отметить, что, несмотря на разницу в биохимических характеристиках фотосинтеза между C_3 злаками – рисом (*O. sativa*), ячменем (*H. vulgare*) и C_4 – кукурузой (*Z. mays*), наблюдается высокий уровень корреляции между рассматриваемыми значениями экспрессии всех гомологов генов, вовлеченных в данную работу (см. Приложение 6). Таким образом, можно заключить, что относительный уровень экспрессии генов, принадлежащих к одной кладе, не выявляет существенных различий у изученных видов.

Поскольку данные по экспрессии изученных генов хорошо соотносятся друг с другом, мы привели усредненные по всем изученным видам значения (см. рис. 2, *и–м*). Мож-

¹ Приложения 1–6 см. по адресу:
<http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx2.pdf>

Genes for antioxidant protection analyzed in the study

Enzyme	Clade	<i>Zea mays</i>	<i>Sorghum bicolor</i>	<i>Setaria italica</i>	<i>Oryza sativa</i> <i>ssp. japonica</i>	<i>Brachypodium</i> <i>distachyon</i>	<i>Hordeum vulgare</i>
APX	A	–	SB07G027300	SI006G20910	–	BD3G40330	HV39901G00010
	B	ZM05G30510	SB04G022560	SI001G20200	OS02G34810	BD3G45700	HV205777G00010 HV39745G00010
	C	ZM02G14900	SB06G017080	SI007G10710	OS04G35520	BD5G10490	HV1566829G00010
	D	–	SB08G004880	SI007G29170	OS12G07830 OS12G07820	BD4G41180	HV158476G00020
	E	ZM01G11990 ZM09G23350	SB01G038760	SI009G45750	OS03G17690	BD1G65820	HV40855G00010
	F	ZM02G41460	SB02G044060	SI002G45330	OS07G49400*	BD1G16510	HV56215G00010
	G	ZM04G35080	SB07G024880	SI006G24130	OS08G43560	BD3G42340	HV137092G00010
	H	–	SB06G001970	SI007G02570	OS04G14680	BD5G03640	HV1569071G00010
GPX	A	ZM06G15260	SB10G005820	SI004G04770	OS06G08670	BD1G47140	–
	B	–	SB00G00200	SI008G10560	OS11G18170	–	–
	C	ZM01G14850 ZM01G14890	SB01G035940	–	–	–	–
	D	ZM04G20660	SB04G032520	SI001G27470	OS02G44500	BD3G51010	HV63919G00010
	E	ZM02G07380 ZM10G22600	SB06G024920	SI007G20040	OS04G46960	BD5G18000	HV54906G00010
	F	ZM09G20790	–	SI009G41050	OS03G24380	BD1G61930	HV1584327G00010
	G	–	SB06G033440	SI003G01000	–	BD5G26725	–
SOD	A	ZM09G06320	SB10G001280	SI009G12380	OS06G02500	BD1G50550	HV56996G00010
	B	ZM06G08270 ZM09G05010	SB10G002970	SI004G03040	OS06G05110	BD1G51140	HV2547265G00020
	C	ZM06G21620 ZM08G13670	SB09G011450	SI003G29220	OS05G25850	BD2G30580	HV6255G00010
	D	ZM01G08150	SB01G042660	SI009G50380	OS03G11960	BD1G69680	HV2550555G00010
	E	ZM07G29600 ZM09G21130 ZM01G15520	–	SI002G43250 SI009G41550	OS03G22810 OS07G46990	BD1G18340	HV160480G00010
	F	ZM01G36030	SB07G023950	SI006G25290	OS08G44770	BD3G43070	HV1577939G00010
	G	ZM02G06820	–	–	OS04G48410	–	–

<i>Triticum aestivum</i>	<i>Triticum urartu</i>	<i>Aegilops tauschii</i>	<i>Arabidopsis thaliana</i>	<i>Solanum lycopersicum</i>	<i>Solanum tuberosum</i>	<i>Vitis vinifera</i>	<i>Chlamydomonas reinhardtii</i>
-	TRIUR3_01974	F775_01406	AT4G32320	SL08G059760	ST08G011590	VV04G12170	
6AL_470886_AA1497540 6BL_499717_AA1589930 6DL_526775_AA1691860	TRIUR3_33614	-	-	-	-	-	
2AL_095116_AA0307060 2DL_158409_AA0517920 2BL_130399_AA0410390	-	-	-	-	-	-	
5DL_433012_AA1398100 U_642188_AA2112960 5AS_392962_AA1266880	TRIUR3_14147	F775_27748	AT1G77490 AT4G08390	SL06G060260 SL11G018550	ST06G019350 ST11G014780	VV18G01950	
4DL_342793_AA1122380 4AS_307333_AA1019630 4BL_320270_AA1033310	TRIUR3_20326	F775_32980					CR06G10160 CR09G08770
U_641465_AA2095890 2BS_146384_AA0464000 2DS_179678_AA0608540	TRIUR3_16101	F775_32894	AT1G07890 AT3G09639	SL09G007270 SL06G005150 SL06G005159	ST09G006729	VV08G01790 VV06G04679	
7AS_569494_AA1817480 7BS_593978_AA1955430 7DS_624295_AA2060370	TRIUR3_11069	F775_30828					
2AS_112531_AA0340050 2BS_146663_AA0470350 2DS_180302_AA0610730	TRIUR3_26505	F775_05651	AT4G35970 AT4G35000	SL02G083620 SL01G111510	ST02G026930 ST01G048890	VV04G00420 VV03G04820	
-	TRIUR3_07828	-	AT2G25080 AT4G31870	SL08G006720	ST08G007960	VV04G13240	
-	-	-	AT2G31570 AT2G43350	SL06G073460	ST06G027530	VV05G00410	
-	-	-				-	
U_641028_AA2082690 6DL_526629_AA1688510	TRIUR3_16000	F775_26338	AT4G11600	SL08G080940	ST08G029520	-	CR10G08850 CR10G05020 CR03G12200 CR08G00160
2BL_130949_AA0420440 2DL_157956_AA0503330	TRIUR3_07893	F775_05616				-	
4BS_328368_AA1086960 U_641855_AA2105830 4AS_306489_AA1008990	-	-	AT3G63080 AT2G48150	SL09G064850	ST09G015140	VV07G04720	
-	TRIUR3_18588	F775_31639	AT1G63460*	SL12G056240* SL12G056230*	ST12G024020* ST12G024030*	VV02G06550*	
4AL_289135_AA0965570 7AS_571265_AA1846380 7DS_625535_AA2065460	TRIUR3_20218	-	-	-	-	-	
7AS_570531_AA1837180 4AL_291448_AA0994630 7DS_622963_AA2047880	TRIUR3_00248	-	AT5G23310 AT4G25100 AT5G51100	SL02G021140 SL06G048410 SL03G095180	ST02G006550 ST06G015010 ST06G015020 ST03G018550	VV10G06370 VV16G01530	
2DL_158719_AA0524930 2BL_131439_AA0427700 2AL_094166_AA0293620	TRIUR3_09843	F775_26675	AT3G56350 AT3G10920	SL06G049080	ST06G013090	-	CR13G07570 CR16G10140 CR17G01600 CR12G01760
4BL_321588_AA1062590 4AS_306564_AA1010270 4DL_342699_AA1119920	TRIUR3_22623	F775_29877	AT1G08830	SL03G062890	-	VV08G08520	CR10G04050 CR02G06330
2AS_114204_AA0365000 2BS_148548_AA0493640 2DS_179233_AA0605480	-	-	AT2G28190	-	-	VV14G04290	
7AL_557214_AA1778230 7DL_602770_AA1968180 7BL_578150_AA1890160	-	-	AT5G18100*	SL11G066390	ST11G024220	VV06G08800	
-	-	-	AT1G12520	SL08G079830	-	-	

End of table

Enzyme	Clade	<i>Zea mays</i>	<i>Sorghum bicolor</i>	<i>Setaria italica</i>	<i>Oryza sativa</i> <i>ssp. japonica</i>	<i>Brachypodium</i> <i>distachyon</i>	<i>Hordeum vulgare</i>
CAT	A	ZM05G16210	SB10G030840	SI004G29580	OS06G51150	BD1G29800	HV7888G00010
	B	ZM01G02380	–	SI009G57040	OS03G03910	BD1G76330	HV71788G00010
	C	ZM04G41390	–	SI001G12120	OS02G02400	BD3G01470	HV1561906G00010

Rows group orthologous genes. Clades are indicated according to the topology of phylogenetic trees in Supplements 1–4. Asterisks indicate sequences whose assignment to an orthologous group has not been reliably proven. Genes promising for breeding programs are shown in boldface (see Discussion).

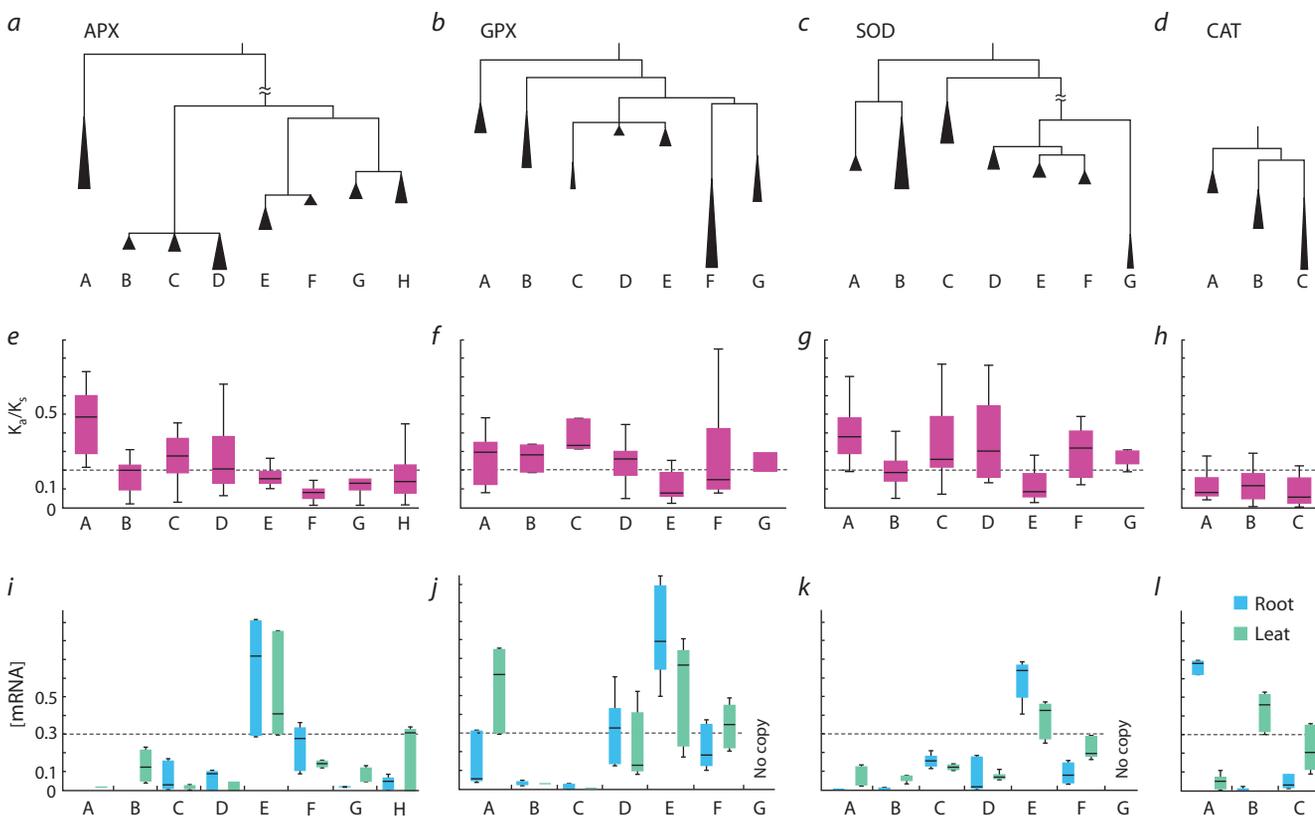


Fig. 2. Phylogenetic relationships among the recognized clades of monocotyledonous species based on (a) APX, (b) GPX, (c) SOD, and (d) CAT; (e–h) K_a/K_s values for each clade; and (i–l) relative levels of mRNAs for corresponding enzyme copies in root and leaf tissues. Topologies of phylogenetic trees are shown in full in Supplements 1–4.

но заметить, что уровень экспрессии для каждой кланды также различен. Отметим, что для разных тканей нередко наблюдается схожий характер экспрессии изучаемых генов. Корреляция значений уровня экспрессии генов в корне и листе достоверна ($P < 0.005$) во всех случаях. Высокий относительный уровень экспрессии ($([mРНК]) > 0.3$) APX и SOD наблюдается для генов кланды E в корне и листе. Гены GPX и CAT выявляют тканеспецифичный характер экспрессии. В корне имеют высокую экспрессию

копии GPX D и E, а в листе – GPX A, E и F. Для корня характерна высокая относительная экспрессия генов CAT A, а для листа – CAT B и C. При этом CAT A отличается высоким относительным уровнем мРНК в тканях корня, а CAT B и CAT C – в тканях листа.

Для более комплексного понимания функциональной важности генов данные характера накопления замен (K_a/K_s) были объединены с усредненными значениями экспрессии в корне и листе. По характеру накопления за-

<i>Triticum aestivum</i>	<i>Triticum urartu</i>	<i>Aegilops tauschii</i>	<i>Arabidopsis thaliana</i>	<i>Solanum lycopersicum</i>	<i>Solanum tuberosum</i>	<i>Vitis vinifera</i>	<i>Chlamydomonas reinhardtii</i>
7AL_556567_AA1765730							
7DL_602975_AA1973160	TRIUR3_21127	-			-	-	
7BL_576925_AA1860110			AT1G20630*	SL04G082460*			
5AL_376544_AA1238210			AT4G35090*	SL02G082760*			
4BL_320451_AA1039710	-	F775_28042					CR09G11950
4DL_343177_AA1131150					ST02G026140	VV18G12260	
U_640838_AA2076560					ST04G035910		
6DS_545065_AA1749890	TRIUR3_31187	F775_32212	AT1G20620	SL12G094620			
6BS_513206_AA1634480							

мен из 25 ортологических групп можно выделить десять, выявивших более сильное давление стабилизирующего отбора: APX E–H; GPX E, F; SOD E; CAT A–C. Одновременно с этим наиболее высокий уровень экспрессии характерен для APX E, SOD E, GPX E и тканеспецифичных форм CAT A (ткани корня), CAT B и C (ткани листа).

Обсуждение

Активные формы кислорода – один из важнейших повреждающих факторов для живых систем. Основой молекулярной механики антиоксидантной защиты растений являются специализированные ферменты, утилизирующие АФК и пероксиды. Улучшение характеристик этой системы позволит повысить устойчивость культурных растений к абиотическим стрессам, таким как засоленность, засуха и холод (Gill et al., 2013; Khaliq et al., 2015; Osipova et al., 2016). В настоящей работе впервые проведен системный анализ особенностей молекулярной эволюции и уровней экспрессии генов четырех функциональных групп ферментов – компонентов системы антиоксидантной защиты (APX, GPX, SOD и CAT).

При реконструкции филогенетических отношений выявлено 25 групп ортологичных генов у злаков, что несколько превосходит количество соответствующих ферментов у наиболее изученного семейства капустоцветных, в то же время в большинстве случаев установлены соответствия (см. таблицу). Таким образом, со времени появления первых ферментов системы антиоксидантной защиты в процессе эволюции – примерно 2.4 млрд лет назад (Hohmann-Marriott, Blankenship, 2011), до появления цветковых растений (более 100 млн лет назад) произошло усложнение состава ферментативной системы, сохранившееся в эволюционных линиях изученных видов. При анализе отношения несинонимичных замен к синонимичным (K_a/K_s) установлено, что характер накопления замен соответствует стабилизирующему отбору. Вместе с тем достоверные различия этого значения между копиями свидетельствуют в пользу неодинакового давления отбора на гены, входящие в разные ортологические группы (клады).

Нами впервые показано, что паттерны экспрессии генов антиоксидантной системы на эволюционных расстояниях менее 50 млн лет (время дивергенции рассчитано по дан-

ным 28 исследований, доступных в системе (<http://www.timetree.org/>) являются достоверно скоррелированными. Это говорит об относительно высокой устойчивости экспрессионных характеристик изучаемой системы на текущем этапе эволюции. Ранее была показана следующая тенденция: паттерны экспрессии в одних и тех же органах между разными видами животных тем лучше скоррелированы, чем меньше время дивергенции между этими видами (Peremyslov et al., 2011).

Значения K_a/K_s для разных копий ферментов антиоксидантной системы были отмечены в пределах от 0.08 до 0.45. Для генов с относительно низким давлением стабилизирующего отбора ($(K_a/K_s) \in (0.2; 0.45)$) наблюдался низкий относительный уровень экспрессии – $\langle [mPНК] \rangle \in (0.3; 0.92)$. Гены с относительно высоким давлением стабилизирующего отбора ($(K_a/K_s) \in (0.08; 0.2)$) имеют как низкий, так и высокий относительный уровень экспрессии мРНК (см. рис. 2). Это согласуется с мировыми исследованиями в данной области. В ряде работ установлена связь между уровнем экспрессии мРНК набора генов и их консервативностью (Wright et al., 2004; Ingvarsson, 2007). На примере голосеменных отмечено, что экспрессия мРНК вносит значимый вклад в частоты используемых кодонов (De La Torre et al., 2015).

В свою очередь, высокому давлению стабилизирующего отбора подвергаются наиболее важные для жизнеспособности гены, что свидетельствует о том, что характер накопления замен является важным индикатором при выборе наиболее перспективных для селекции генов. При интеграции экспрессионных данных и характеристик накопления замен предсказание генов, которые вносят наибольший вклад в работу антиоксидантной системы растений, становится более однозначным.

В результате из 25 ортологических групп было выделено шесть наиболее приоритетных групп генов-кандидатов для дальнейшей верификации и использования в процессе селекции. Это APX E, SOD E, GPX E и тканеспецифичные CAT A (ткани корня), CAT B и C (ткани листа) (см. таблицу). Соответствующие копии для каждого из изученных видов растений и хромосомная локализация генов приведены в таблице. Для всех видов, кроме представителей родов *Triticum* и *Aegilops*, хромосомная локализация

генов дана в цифровой кодировке (первое число после буквенного идентификатора вида) и отделена буквой G от идентификатора локуса (второе число). Можно заметить, что гены антиоксидантной защиты представлены на различных хромосомах C₃ и C₄ злаков (см. таблицу). Для кукурузы *Z. mays* найдено и проанализировано 27 гомологов генов APX, GPX, SOD и CAT. Гены *Z. mays* в ортогологических группах APX E, SOD E и GPX E представлены в двух и трех копиях. Одновременно с этим паралоги с низким уровнем экспрессии и относительно высоким уровнем K_a/K_s, такие как, например, APX A и APX D, отсутствуют.

Номенклатурные названия генов мягкой пшеницы начинаются с кодировки хромосомы. Следует отметить, что изученные гены мягкой пшеницы найдены на хромосомах гомеологических групп 2 и 4–7. Предсказанные гены-кандидаты с наибольшим уровнем экспрессии мРНК GPX и SOD находятся на хромосомах 2-й гомеологической группы, а CAT – на хромосомах 4–7-й гомеологических групп. Это соотносится с классическими работами, в которых анализировались анеуплоидные линии сорта Чай-низ Спринг: основные локусы, ассоциированные с активностью SOD, были локализованы на хромосомах 2-й гомеологической группы, а ассоциированные с CAT – на хромосомах 2-й и 7-й гомеологических групп (Neuman, Hart, 1986; Li et al., 1999; Wu et al., 1999). Это свидетельствует в пользу того, что за полиморфизм в активности антиоксидантной системы могут быть ответственны аллельные варианты именно структурных, а не регуляторных генов. Современные исследования подтверждают эти данные, а также позволяют выявить большее количество ассоциированных локусов, например локус, ассоциированный с активностью APX на хромосоме 7D, и локус, ассоциированный с активностью SOD на хромосоме 4D (Osipova et al., 2016).

Дикие сородичи мягкой пшеницы являются перспективным источником генов устойчивости к стрессовым факторам (Trethowan, Mujeeb-Kazi, 2008; Reynolds et al., 2009). Один из самых актуальных на данный момент источников – *Ae. tauschii*, донор D-генома для мягкой пшеницы, который имеет высокое генетическое разнообразие (Jia et al., 2013). Этот вид вовлечен нами в анализ совместно с *T. urartu*. Информацию о наборе генов антиоксидантной защиты для этих видов можно будет в дальнейшем использовать в селекции для контроля уровня экспрессии ключевых генов у интрогрессивных форм совместно с оценкой их физиологического эффекта.

Таким образом, в настоящей работе проведен широкомасштабный поиск гомологов генов антиоксидантной защиты APX, GPX, SOD и CAT, анализ их эволюционных характеристик и паттернов экспрессии. В результате предсказаны шесть структурных генов-компонентов антиоксидантной системы с наиболее высоким уровнем экспрессии и наибольшим давлением стабилизирующего отбора. Полученные данные требуют дальнейшей экспериментальной верификации. Показано, что интеграция эволюционных характеристик и экспрессионных данных позволяет проводить приоритизацию генов в многокопийных генетических системах и предсказывать перспективные мишени для селекции.

Acknowledgments

The development of the approach and data analysis were supported by the Russian Science Foundation, project 17-74-10198. Use of equipment of the Bioinformatics Shared Access Center for the analysis was supported by State Budgeted Project 0324-2016-0008

Conflict of interest

The authors declare no conflict of interest.

References

- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J. Mol. Biol.* 1990;215(3):403-410. DOI 10.1016/S0022-2836(05)80360-2.
- Arvestad L. Efficient methods for estimating amino acid replacement rates. *J. Mol. Evol.* 2006;62(6):663-673. DOI 10.1007/s00239-004-0113-9.
- Das K., Roychoudhury A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. *Front. Environ. Sci.* 2014;2:53. DOI 10.3389/fenvs.2014.00053.
- De La Torre A.R., Lin Y., Van de Peer Y., Ingvarsson P.K. Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in *Picea* gene families. *Genome Biol. Evol.* 2015;7(4):1002-1015. DOI 10.1093/gbe/evv044.
- Ehleringer J.R., Sage R.F., Flanagan L.B., Pearcy R.W. Climate change and the evolution of C4 photosynthesis. *Trends Ecol. Evol.* 1991;6(3):95-99. DOI 10.1016/0169-5347(91)90183-X.
- Gill S.S., Tajrishi M., Madan M., Tuteja N. A DESD-box helicase functions in salinity stress tolerance by improving photosynthesis and antioxidant machinery in rice (*Oryza sativa* L. cv. PB1). *Plant Mol. Biol.* 2013;82(1-2):1-22. DOI 10.1007/s11103-013-0031-6.
- Gill S.S., Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* 2010;48(12):909-930. DOI 10.1016/j.plaphy.2010.08.016.
- Guindon S., Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 2003; 52(5):696-704. DOI 10.1080/10635150390235520.
- Gunbin K.V., Suslov V.V., Genaev M.A., Afonnikov D.A. Computer system for analysis of molecular evolution modes (SAMEM): analysis of molecular evolution modes at deep inner branches of the phylogenetic tree. In *Silico Biol.* 2012;11(3,4):109-123. DOI 10.3233/ISB-2012-0446.
- Hammer Ø., Harper D.A.T., Ryan P.D. Paleontological statistics software: Package for education and data analysis. *Palaeontol. Electronica.* 2001;4. Available at: http://palaeo-electronica.org/2001_1/past/issue1_01.htm.
- Hammond-Kosack K.E., Jones J.D. Resistance gene-dependent plant defense responses. *Plant Cell.* 1996;8(10):1773-1791.
- Han M.V., Zmasek C.M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics.* 2009;10(1):356. DOI 10.1186/1471-2105-10-356.
- Hatch M.D. C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta – Rev. Bioenergetics.* 1987;895(2):81-106. DOI 10.1016/S0304-4173(87)80009-5.
- Hatch M.D. C₄ photosynthesis: discovery and resolution. In: Govindjee, Beatty J.T., Gest H., Allen J.F. (Eds.). *Discoveries in Photosynthesis.* Springer, 2005:875-880.
- Henikoff S., Henikoff J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA.* 1992;89(22):10915-10919.
- Hohmann-Marriott M.F., Blankenship R.E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* 2011;62(1):515-548. DOI 10.1146/annurev-arplant-042110-103811.
- Ingvarsson P.K. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol. Biol. Evol.* 2007;24(3):836-844. DOI 10.1093/molbev/msl212.
- Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Span-

- naql M., Mayer K., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., International Wheat Genome Sequencing Consortium, Yang H., Liu X., He Z., Mao L., Wang J. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013;496:91-95. DOI 10.5524/100054.
- Karpilov Y.S. The distribution of radioactive carbon 14 amongst the products of photosynthesis of maize. *Trudy Kazanskogo sel'skokhozyaystvennogo instituta = Works of the Kazan Agricultural Institute*. 1960;41(1):15-24. (in Russian)
- Katoh K., Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 2008;9(4):286-298. DOI 10.1093/bib/bbn013.
- Khaliq A., Zia-ul-Haq M., Ali F., Aslam F., Matloob A., Navab A., Husain S. Salinity tolerance in wheat cultivars is related to enhanced activities of enzymatic antioxidants and reduced lipid peroxidation. *CLEAN – Soil, Air, Water*. 2015;43(8):1248-1258. DOI 10.1002/clen.201400854.
- Li W.L., Faris J.D., Chittoor J.M., Leach J.E., Hulbert S.H., Liu D.J., Chen P.D., Gill B.S. Genomic mapping of defense response genes in wheat. *Theor. Appl. Genet.* 1999;98(2):226-233. DOI 10.1007/s001220051062.
- Li W.H., Yang J., Gu X. Expression divergence between duplicate genes. *Trends Genet.* 2005;21(11):602-607. DOI 10.1016/j.tig.2005.08.006.
- Mittler R., Vanderauwera S., Gollery M., Van Breusegem F. Reactive oxygen gene network of plants. *Trends Plant Sci.* 2004;9(10):490-498. DOI 10.1016/j.tplants.2004.08.009.
- Neuman P.R., Hart G.E. Genetic control of the mitochondrial form of superoxide dismutase in hexaploid wheat. *Biochem. Genet.* 1986; 24(5-6):435-446.
- Noctor G., Foyer C.H. Ascorbate and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Biol.* 1998;49(1):249-279. DOI 10.1146/annurev.arplant.49.1.249.
- Orozco-Cardenas M., Ryan C.A. Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. *Proc. Natl. Acad. Sci. USA*. 1999;96(11):6553-6557. DOI 10.1073/pnas.96.11.6553.
- Osipova S., Permyakov A., Permyakova M., Pshenichnikova T., Verkhoturov V., Rudikovskiy A., Rudikovskaya E., Shishparenok A., Doroshkov A., Börner A. Regions of the bread wheat D genome associated with variation in key photosynthesis traits and shoot biomass under both well watered and water deficient conditions. *J. Appl. Genet.* 2016;57(2):151-163. DOI 10.1007/s13353-015-0315-4.
- Peremyslov V.V., Mockler T.C., Filichkin S.A., Fox S.E., Jaiswal P., Makarova K.S., Koonin E.V., Dolja V.V. Expression, splicing, and evolution of the myosin gene family in plants. *Plant Physiol.* 2011; 155(3):1191-1204. DOI 10.1104/pp.110.170720.
- Rambaut A. FigTree, version 1.4: tree figure drawing tool. 2008. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Reynolds M., Foulkes M.J., Slafer G.A., Berry P., Parry M.A., Snape J.W., Angus W.J. Raising yield potential in wheat. *J. Exp. Bot.* 2009;60(7):1899-1918. DOI 10.1093/jxb/erp016.
- Sage R.F., Sage T.L., Kocacinar F. Photorespiration and the evolution of C₄ photosynthesis. *Annu. Rev. Plant Biol.* 2012;63:19-47. DOI 10.1146/annurev-arplant-042811-105511.
- Trethowan R.M., Mujeeb-Kazi A. Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Sci.* 2008;48(4):1255-1265. DOI 10.2135/cropsci2007.08.0477.
- Wang X., Cai J., Liu F., Dai T., Cao W., Wollenweber B., Jiang D. Multiple heat priming enhances thermo-tolerance to a later high temperature stress via improving subcellular antioxidant activities in wheat seedlings. *Plant Physiol. Biochem.* 2014;74:185-192. DOI 10.1016/j.plaphy.2013.11.014.
- Wright S.I., Yau C.K., Looseley M., Meyers B.C. Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Mol. Biol. Evol.* 2004;21(9):1719-1726. DOI 10.1093/molbev/msh191.
- Wu G., Wilen R.W., Robertson A.J., Gusta L.V. Isolation, chromosomal localization, and differential expression of mitochondrial manganese superoxide dismutase and chloroplastic copper/zinc superoxide dismutase genes in wheat. *Plant Physiol.* 1999;120(2):513-520. DOI 10.1104/pp.120.2.513.
- Zhang J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 2003;18(6):292-298. DOI 10.1016/S0169-5347(03)00033-8.
- Zmasek C.M. Archaeopteryx: visualization, analysis, and editing of phylogenetic trees. 2015. Available at: <https://sites.google.com/site/cmzmasek/home/software/archaeopteryx>.