

## **Тема работы.**

Полногеномное изучение экспрессии генов культурных растений  
в ответ на стресс

## **Состав коллектива:**

Бобровских Александр Владимирович; студент новосибирского государственного университета; б/с; б/з avb@bionet.nsc.ru

Дорошков Алексей Владимирович; ассистент кафедры информационной биологии ФЕН НГУ; к.б.н. б/з; ad@bionet.nsc.ru

## **Научное содержание работы:**

### **Постановка задачи.**

**Цель работы:** Изучение особенностей структурной организации и регуляции антиоксидантной системы растений методами системной биологии

### **Задачи:**

- 1) Провести широкомасштабный поиск генов, отвечающих за стрессоустойчивость растений.
- 2) На основе доступных данных полногеномных транскриптомных экспериментов оценить экспрессию найденных генов в различных эволюционных линиях.

### **Современное состояние проблемы (на момент начала работы).**

Низкая устойчивость сельскохозяйственных культур к засухе и холоду является одной из причин плохого урожая. В связи с этим одной из ключевых задач биологии является поиск и описание генов отвечающих за устойчивость растений к этим факторам. Выявление ключевых генов стрессоустойчивости являются важным этапом для дальнейшего выведения новых сортов. Однако большинство генов известных на данный момент выявлены у модельных растений которые не являются сельскохозяйственно значимыми и не могут быть непосредственно использованы селекционерами.

Особую важность имеет изучение известных генетических систем стрессового ответа. Комплексная реакция растений включает в себя систему АФК-сигналинга, тесно связанную с активностью антиоксидантной системы, гормональной регуляцией и вовлекает в работу ключевые семейства транскрипционных факторов (Gill and Tuteja, 2010).

Показана важность антиоксидантной системы в ответ на стрессовые воздействия. В частности, трансгенные линии растений с дополнительными копиями генов антиоксидантной защиты показали большую устойчивость к различным типам стресса, чем исходные линии. Этот факт отражен при работе на рисе (Zhao et al, 2009) (кадмиевый и тепловой стресс), на арабидопсисе (Lu et al, 2007) (солевой стресс). Также показано, что трансгенные линии риса, нокаутные по отдельным генам антиоксидантной защиты, хуже переносят стрессовые условия (Zhang et al, 2013; Wu et al, 2015)

Активность антиоксидантных ферментов разных классов усиливается в ответ на абиотические стрессы. Так, усиление активности аскорбатпероксидазы,

каталазы, супероксиддисмутазы, глутатионредуктазы описано в работе (Shigeoka et al, 2002). Эти факты могут свидетельствовать в пользу ко-регуляции компонентов данной системы в ответ на стресс.

Показана генетическая сложность данной системы. Известно, что система многокопийна: каждый фермент системы антиоксидантной защиты представлен в геноме растений в нескольких копиях и имеет разную клеточную локализацию (цитозоль, митохондрии, пероксисомы, хлоропласты). Например, в геноме *Arabidopsis thaliana* присутствует почти 40 генов, кодирующих антиоксидантные ферменты (Mittler et al, 2004).

Антиоксидантная система растительной клетки является жизненно важной системой, обеспечивающей гомеостаз клетки и стабильный редокс-статус на разных уровнях (от молекулярного до организменного). Система является многокомпонентной в клеточном, биохимическом и генетическом рассмотрении. Требуется комплексный подход к изучению данной системы для выявления закономерностей ее функционирования и динамических характеристик.

Gill S. S., Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants //Plant physiology and biochemistry. – 2010. – Т. 48. – №. 12. – С. 909-930.

F. Y. Zhao, W. Liu, and S. Y. Zhang, “Different responses of plant growth and antioxidant system to the combination of cadmium and heat stress in transgenic and non-transgenic rice,” J. Integr. Plant Biol., vol. 51, no. 10, pp. 942–950, 2009.

Z. Zhang et al., “Gene knockout study reveals that cytosolic ascorbate peroxidase 2 (OsAPX2) plays a critical role in growth and reproduction in rice under drought, salt and cold stresses,” PLoS One, vol. 8, no. 2, 2013.

T. M. Wu, W. R. Lin, C. H. Kao, and C. Y. Hong, “Gene knockout of glutathione reductase 3 results in increased sensitivity to salt stress in rice,” Plant Mol. Biol., vol. 87, no. 6, pp. 555–564, 2015.

S. Shigeoka et al., “Regulation and function of ascorbate peroxidase isoenzymes.,” J. Exp. Bot., vol. 53, no. 372, pp. 1305–1319, 2002

R. Mittler, S. Vanderauwera, M. Gollery, and F. Van Breusegem, “Reactive oxygen gene network of plants,” Trends Plant Sci., vol. 9, no. 10, pp. 490–498, 2004.

### **Подробное описание работы, включая используемые алгоритмы.**

#### Препроцессинг данных RNA-Seq

На первом этапе был проведен контроль качества транскриптомных данных при помощи программы FASTQC v0.11.5 (Andrews, 2010). Затем для удаления адаптеров и фрагментов их последовательностей была применена опция ILLUMINACLIP в программе Trimmomatic-0.36 (Bolger, Lohse, Usadel, 2014), с использованием встроенной PE-библиотеки адаптеров TruSeq3-PE. Для оценки качества данных, прошедших предобработку в программе Trimmomatic-0.36, был повторно проведен контроль качества ридов в FASTQC v0.11.5.

#### Картирование ридов на геном

Предобработанные данные были картированы в программе STAR-2.5.4b (Dobin, Gingeras, 2015) с привлечением Gff-файлов аннотации геномов и референсных сборок геномов из базы данных Plaza3.0

(<https://bioinformatics.psb.ugent.be/plaza/>). Для проверки статистики картирования использовалась программа SAMtools-1.7 (Li и др., 2009), а подсчет числа ридов в генах был осуществлен с применением R-пакета HTSeq-1.7 (Anders, Pyl, Huber, 2015).

Выявление дифференциально экспрессирующихся генов

Анализ дифференциальной экспрессии был проведен в программах DESeq2 v1.18.1 (Love, Huber, Anders, 2014) и edgeR v3.20.9 (Robinson, McCarthy, Smyth, 2010). Для оценки достоверности данных использовалась поправка на множественное сравнение Бенджамини-Хохберга (Ignacio González, 2014). При обработке в DESeq2 v1.18.1 функция автоматической фильтрации ДЭГ с низким уровнем изменения экспрессии была отключена (`independentFiltering = FALSE`). ДЭГ были отфильтрованы в Excel по значению  $q\text{-value} < 0.05$  и разделены на активируемые и подавляемые в ответ на СК по знаку  $\log_2\text{FoldChange}$ .

Уровни экспрессии были нормированы методом Transcripts Per Million (TPM) (Wagner et al., 2012) для сравнения библиотек внутри вида. Для сравнения между видами была выполнена дополнительная нормировка на ген актина.

Andrews S. FastQC: A quality control tool for high throughput sequence data.

Bolger A.M., Lohse M., Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data // *Bioinformatics*. 2014. Т. 30. № 15. С. 2114–2120.

Dobin A., Gingeras T.R. Mapping RNA-seq Reads with STAR // *Curr. Protoc. Bioinforma.* 2015. Т. 51. С. 11.14.1-11.14.19.

Li H. и др. The Sequence Alignment/Map format and SAMtools // *Bioinformatics*. 2009. Т. 25. № 16. С. 2078–207

Anders S., Reyes A., Huber W. Detecting differential usage of exons from RNA-seq data // *Genome Res.* 2012. Т. 22. № 10. С. 2008–2017.

Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. // *Bioinformatics*. 2010. Т. 26. № 1. С. 139–40.

Ignacio González P.B.I.T.P.B.I.M.T.U.T.I.I.I. Tutorial - Statistical analysis of RNA-Seq data. , 2014. 1-75 с.

### **Полученные результаты.**

Составлен список генов, наиболее перспективных для контроля в процессе селекции на устойчивость к окислительному стрессу (таблица 1).

В столбцах приведены латинские названия видов, в строках - аббревиатуры ферментов (APX - аскорбатпероксидаза, CAT - каталаза, GPX - глутатионпероксидаза, SOD - супероксиддисмутаза, MDHAR - монодегидроаскорбатредуктаза, DHAR -дегидроаскорбатредуктаза, GR - глутатионредуктаза), в соответствующих ячейках - ID генов. Данные для пшеницы *Triticum aestivum* получены на основании сопоставления

колокализации ферментов с локусами QTL статьи Osipova et al. 2016 и с высококонсервативными ( $Ka/Ks < 0.20$ ) высокоэкспрессирующимися гомологами других видов, приведенными в таблице 1.

Установлено, что система защиты от АФК в процессе эволюции растений претерпела ряд дупликаций на ранних этапах (до появления многоклеточности). Это привело к ранней специализации копий, включая различную клеточную локализацию.

Анализ паттернов экспрессии мРНК изучаемых ферментов показал, что на межвидовых эволюционных расстояниях менее 100 млн. лет паттерны экспрессии достоверно скоррелированы.

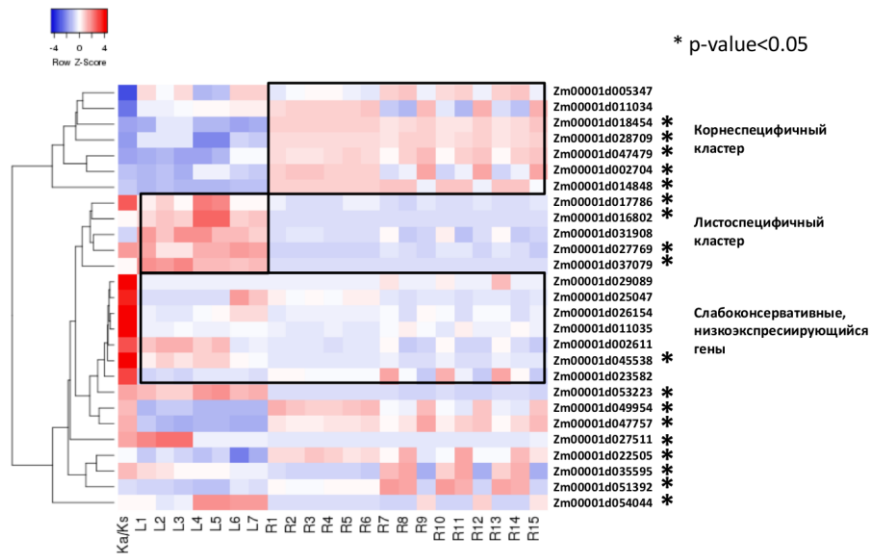
Таблица 1

Организм> //Класс фермента V	<i>Solanum lycopersicum</i>	<i>Populus trichocarpa</i>	<i>Glycine max</i>	<i>Arabidopsis thaliana</i>	<i>Zea mays</i>	<i>Oryza sativa</i>	<i>Hordeum vulgare</i>	<i>Triticum aestivum</i>
APX	AT1G07890	Potri.016G084800	Glyma.11g15680 Glyma.12g07780	AT1G07890	Zm00001d028709	OS07G49400		TRIAE_CS42_2DS_TG ACv1_179678_AA0808 540
CAT	AT4G35090 AT1G20620	Potri.002G009800	Glyma.04g01920 Glyma.06g02040	AT4G35090 AT1G20620	Zm00001d014848 Zm00001d054044	OS06G51150 OS02G02400 OS03G03910	HV7888G00010 HV71788G00010 HV1561906G00010	
GPX					Zm00001d002704 Zm00001d037079	OS04G46960	HV54906G00010	TRIAE_CS42_2DL_TG ACv1_157956_AA0503 330
SOD	AT1G08830 AT4G25100			AT1G08830 AT4G25100			HV1577939G00010	TRIAE_CS42_7DL_TG ACv1_602770_AA1968 180
MDHAR	AT5G03630	Potri.006G114800	Glyma.10g07820	AT5G03630	Zm00001d005347	OS08G05570 OS09G39380	HV39732G00010 HV58410G00010	
DHAR	AT1G19570 AT1G75270	Potri.010G211600	Glyma.10g07820	AT1G19570 AT1G75270	Zm00001d011034	OS05G02530	HV1569254G00010	
GR	AT3G24170 AT3G54660	Potri.001G050000 Potri.003G178200 Potri.015G037800		AT3G24170 AT3G54660	Zm00001d018454 Zm00001d027769	OS02G56850	HV45229G00010 HV232942G00010	

Экспрессия и  $Ka/Ks$  значения для антиоксидантных генов *Solanum lycopersicum* приведена на рисунке 1. В столбце  $Ka/Ks$  отображены значения эволюционной консервативности соответствующих генов, в следующих столбцах экспрессия в экспериментах в листе (L1-L15) и в корне (R1-R6). Названия строк отражают ID генов. Красный цвет соответствует относительно высокому значению характеристики, синий цвет – низкому.

Экспрессия и  $Ka/Ks$  значения для антиоксидантных генов *Zea mays*. В столбце  $Ka/Ks$  отображены значения эволюционной консервативности соответствующих генов, в следующих столбцах экспрессия в экспериментах в листе (L1-L7) и в корне (R1-R15). Названия строк отражают ID генов. Красный цвет соответствует относительно высокому значению характеристики, синий цвет – низкому.

## Иллюстрации, визуализация результатов



**Рисунок 1.** Экспрессия и Ka/Ks значения для антиоксидантных генов *Solanum lycopersicum*. Объяснение в тексте.

### Эффект от использования кластера в достижении целей работы.

Отдельные этапы работы требовали единовременное выделение порядка 100 ГБ оперативной памяти, В связи с чем проведение таких расчетов на личных компьютерах не представлялось возможным. Часть результатов дипломной работы были получены на кластере НГУ.