

ПОИСК НОВЫХ ГЕНОВ, ЭКСПРЕССИРУЮЩИХСЯ В ПРОЦЕССЕ РАЗВИТИЯ ЗЕРНОВКИ, С ПОМОЩЬЮ НАНОПОРОВОГО СЕКВЕНИРОВАНИЯ

**Полховская Екатерина Сергеевна, Болотина Анна Александровна,
Дудников Максим Васильевич, Киров Илья Владимирович**
*Всероссийский научно-исследовательский институт сельскохозяйственной
биотехнологии, Москва, Россия*
eynzeynkreyn@gmail.com

Изучение принципов и закономерностей регуляции биологических процессов, определяющих практически значимые признаки сельскохозяйственных культур, является перспективным для селекции растений. У злаковых культур процесс формирования зерновки относится к наиболее перспективным для исследования, но в то же время характеризуется высокой сложностью и динамичностью. На молекулярном уровне помимо структуры белок-кодирующих генов существенный вклад в рост и развитие организмов, в т.ч. растений, могут вносить механизмы регуляции экспрессии генов с помощью длинных некодирующих РНК (днРНК). На сегодняшний момент транскриптомные изменения во время развития зерна хорошо изучены и многие гены были аннотированы в пшенице, в то время как днРНК, экспрессируемые остаются неисследованными.

Наиболее оптимальным подходом в изучении генов зерновки является использование технологий Oxford Nanopore для анализа РНК, выделенных из семян тритикале на 10 день после цветения. После анализ данных, сгенерированных в процессе секвенирования РНК, обогащённой поли(А) фракцией, получено 1240000 ридов. Используя их, были идентифицированы днРНК, экспрессирующиеся на ранних стадиях развития семян тритикале, большинство из которых ранее не были аннотированы. Были обнаружены экзонные, интронные, антисмысловые и межгенные днРНК. При помощи биоинформатического анализа удалось идентифицировать 796 днРНК тритикале, кодируемые 780 локусами в субгеномах А (167 днРНК), В (212 днРНК) и R (410 днРНК). Паттерн экспрессии идентифицированных днРНК является тканеспецифичным, причем почти половина днРНК демонстрирует максимальный уровень экспрессии во время развития семян. Для дальнейшего анализа были взяты днРНК субгеномов А и В и изучен полиморфизм их генов в коллекции тритикале. Для валидации полученных данных были выбраны 16 днРНК, для которых были разработаны специфичные пары праймеров, после чего проводилась ПЦР на геномной ДНК коллекционных образцов яровой тритикале. В ходе амплификации выявлены структурные вариации в виде инсерций и делеций в генах днРНК в 23 образцах коллекции тритикале. Анализа полиморфизма генов с помощью подобранных праймеров выявил 8 нереперенсных аллелей днРНК среди всей коллекции, из которой одна линия является наиболее вариабельной.

Ключевые слова: тритикале, гены развития зерновки, днРНК, нанопоровое секвенирование.

SEARCH FOR NEW GENES EXPRESSED DURING GRAIN DEVELOPMENT USING NANOPORE SEQUENCING

Ekaterina S. Polkhovskaya, Anna A. Bolotina, Maxim V. Dudnikov, Ilya V. Kirov
All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia
eynzeynkreyn@gmail.com

The study of principles and regularities of regulation of biological processes determining practically significant traits of agricultural crops is promising for plant breeding. In cereal crops the process of seed development is among the most promising for research, but at the same time it is characterized by high complexity and dynamism. At the molecular level, in addition to the structure of protein-coding genes, the mechanisms of gene expression regulation by long non-coding RNAs (lncRNAs) can significantly contribute to the growth and development of organisms, including plants. Nowadays transcriptomic changes during seed development have been well studied and many genes have been annotated in wheat, while lncRNAs expressed remain unexplored.

The best approach in studying grain genes is using of Oxford Nanopore technologies for analyzing RNA isolated from triticale seeds at 10 days after anthesis. After analyzing the data generated by sequencing RNA enriched with poly(A) fraction, 1240000 rds were obtained. Using these, lncRNAs expressed during the early stages of triticale seed development, most of which had not been previously annotated, were identified. Exonic, intronic, antisense, and intergenic lncRNAs were detected. Bioinformatic analysis was used to identify 796 lncRNAs in triticale encoded by 780 loci in subgenomes A (167 lncRNAs), B (212 lncRNAs), and R (410 lncRNAs). The expression pattern of the identified lncRNAs is tissue-specific, with almost half of the lncRNAs showing maximum expression levels during seed development. For further analysis, lncRNAs of subgenomes A and B were taken and polymorphisms of their gene were studied in the triticale collection. For validation obtained data, 16 lncRNAs were selected and analyzed by specific designed primer pairs. Further verification was carried out by PCR on the genomic DNA of spring triticale collection. Amplification revealed structural variations in the form of insertions and deletions in lncRNA genes in 23 samples of the triticale collection. Analysis of gene polymorphism using designed primers revealed 8 non-reference lncRNA alleles among the entire collection, of which one line was the most variable.

Keywords: triticale, seed development genes, lncRNA, nanopore sequencing.