



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
ВСЕРОССИЙСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
СЕЛЬСКОХОЗЯЙСТВЕННОЙ БИОТЕХНОЛОГИИ
(ФГБНУ ВНИИСБ)

На правах рукописи

ПОЛХОВСКАЯ ЕКАТЕРИНА СЕРГЕЕВНА

**СТРУКТУРНО-ТРАНСКРИПТОМНЫЙ АНАЛИЗ ГЕНОВ ПШЕНИЦЫ
И ТРИТИКАЛЕ, ЭКСПРЕССИРУЮЩИХСЯ В ПРОЦЕССЕ РАЗВИТИЯ
ЗЕРНОВКИ, С ПОМОЩЬЮ НАНОПОРОВОГО СЕКВЕНИРОВАНИЯ**

1.5.7 Генетика

Диссертация

на соискание ученой степени
кандидата биологических наук

Научный руководитель:

доктор биологических наук, профессор РАН
Соловьев Александр Александрович

Москва – 2024

ОГЛАВЛЕНИЕ

Список сокращений	4
ВВЕДЕНИЕ	8
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	18
1.1. Характеристика тритикале	18
1.1.1. Происхождение и биологические особенности тритикале	18
1.1.2. Селекционно-генетические аспекты качества зерна тритикале	21
1.2. Транскриптомные особенности развитие зерновки	22
1.2.1 Динамика транскриптома в процессе развития зерновки	25
1.3. Геномные особенности зерновых культур	28
1.4. Запасные белки зерновки пшеницы и тритикале	29
1.4.1. Характеристика глютеинов	29
1.4.2. Низкомолекулярные глютеины (LMW-GS)	31
1.4.3. Высокомолекулярные глютеины (HMW-GS)	32
1.4.4. Полиморфизм генов высокомолекулярных глютеинов	33
1.4.5. Регуляция экспрессии генов глютеинов	36
1.5. Классификация и функции длинных некодирующих РНК	40
1.5.1. Классификация некодирующих РНК	40
1.5.2. Значение днРНК в индивидуальном развитии растений	41
1.6. Секвенирование и аннотация геномов пшеницы, ржи и тритикале	43
1.7. Секвенирование целевых последовательностей	46
РАЗДЕЛ 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ	53
2.1. Материалы для исследования	53
2.2. Методы исследования	53
2.2.1. Выделение ДНК	53
2.2.2. Обогащение геномной ДНК высокомолекулярными фрагментами	55
2.2.3. Дизайн гидовой РНК	55
2.2.4. Синтез и очистка гРНК	56
2.2.5. Подготовка библиотеки nCATS	57
2.2.6. Нанопоровое секвенирование	60
2.2.7. Анализ данных, полученных с помощью секвенирования	60
2.2.8. ПЦР-подтверждение вставки в ген <i>Glu-1Bx</i>	60
2.2.9. ПЦР-амплификация последовательностей генов высокомолекулярных субъединиц глютеинов	60
2.2.10. Валидация делеции в гене <i>Glu-1Bx7</i>	62
2.2.11. Нанопоровое секвенирование ампликонов и анализ полученных данных	62
2.2.12. Выделение тотальной РНК	63
2.2.13. Выделение поли(А)-РНК	63
2.2.14. Подготовка библиотеки для Direct RNA-Seq Nanopore	63
2.2.15. Проведение ПЦР и электрофореза	64
2.2.16. Синтез дц-кДНК	64
2.2.17. Биоинформатический анализ данных секвенирования дц-кДНК	65
ГЛАВА 3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ	67

3.1. Cas9-опосредованное нанопоровое секвенирование (nCATS) генов высокомолекулярных глютеинов	67
3.1.1. Подготовка ДНК библиотеки и секвенирование nCATS	70
3.1.2. Идентификация структурных вариаций гена <i>Glu-1Bу</i>	72
3.1.3. Детекция метилирования	74
3.2. Секвенирование и анализ вариабельности генов высокомолекулярных глютеинов методом Nanopore Amplicon-seq	76
3.2.1. Амплификация полноразмерных генов глютеина	77
3.2.2. Анализ последовательностей аллелей HMW-GS, полученных с помощью ONT Amplicon-Seq	84
3.2.3. Филогенетический анализ аллелей генов глютеинов	88
3.2.4. Преимущества Nanopore Amplicon-Seq для анализа HMW-GS	91
3.3. Анализ транскриптома зерновок тритикале с помощью Нанопорового секвенирования	93
3.3.1. Анализ транскриптов, полученных с помощью Нанопорового секвенирования, по белок-кодирующему потенциалу	94
3.3.2. Характеристика генов, экспрессирующихся на разных стадиях развития зерновки	99
3.3.3. Особенности транскриптома зародыша и эндосперма развивающихся зерновок	104
3.3.4. Полиморфизм генов днРНК в коллекции тритикале	109
ЗАКЛЮЧЕНИЕ	113
ВЫВОДЫ	114
СПИСОК ЛИТЕРАТУРЫ	115

Список сокращений

В/см – вольт на сантиметр

Г – год

Гб – миллиард пар оснований (Gb = gigabase)

гРНК – гидовая РНК (guide RNA).

ДНК – дезоксирибонуклеиновая кислота

днРНК – длинные некодирующие РНК

дц-кДНК – двухцепочечная комплементарная ДНК

Кб – тысяча пар оснований (Kb = kilobase)

кДа – килодальтон

кДНК – комплементарная ДНК

М – моль

Мб – миллион пар оснований (Mb = megabase)

мг – миллиграмм

мМ – миллимоль

мРНК – матричная РНК

мин – минута

миРНК – малая интерферирующая РНК

мкл – микролитры

мкМ – микромоль

мл – миллилитр

нг – нанограмм

нкРНК – некодирующая РНК

п.н. – пары нуклеотидов

пре-рРНК – предшественника рРНК

ПЦР – полимеразная цепная реакция

РНК – рибонуклеиновая кислота

рРНК – рибосомная РНК

РНП-комплекс – рибонуклеопротеидный комплекс
сек – секунда
ТЕ-буфер – Трис-ЭДТА буфер
тракрРНК – трансактивирующая РНК
ТФ – транскрипционный фактор
тыс.п.н. – тысяч пар нуклеотидов
ч – час
АВА – abscisic acid
BLAST - basic local alignment search tool
BS – binding site
C-banding – chromosome banding
CATCH – clone adapted template capture hybridization
CDS – coding sequence
CIP – calf-intestinal alkaline phosphatase
CNVs - copy number variations
CRISPR – clustered regularly interspaced short palindromic repeats
crRNA – CRISPR-RNA
CTAB – cetyltrimethyl ammonium bromide
dATP – Deoxyadenosine triphosphate
DOF – DNA-binding with one finger
DPA – days post anthesis
EB – endosperm box
EDTA – ethylenediaminetetraacetic acid
EE – endosperm + embryo
EN – endosperm
ER – endoplasmic reticulum
FDR – false discovery rate
FLC – flowering locus C
FM – flower meristem
FZP – frizzy panicle

GA – gibberellin
gbM – gene body methylation
GSP – grain softness protein
GCN – N-acyltransferase
Glu – glutenin
GO – gene ontology
GS – glutenin subunit
gtf – gene transfer format
HMW-GS – high molecular weight glutenin subunit
InDel – insertion-deletion
IWGSC – International Wheat Genome Sequencing Consortium
LDMAR – long-day-specific male-fertility-associated RNA
lincRNA – long intergenic non-coding RNAs
LMW-GS – low molecular weight glutenin subunit
lncRNA – long noncoding RNA
LTR – long terminal repeat
LTR-RTS – LTR-retrotransposons
MQ – Mapping Quality
MIPs – microtubule inner proteins
miRNA – microRNA
nCATS – Nanopore Cas9-targeted sequencing
NCBI – The National Center for Biotechnology Information
NGS – next generation sequencing
ONT – Oxford Nanopore Technologies
ONT Amplicon-Seq - Oxford Nanopore Technologies Amplicon Sequencing
PAM – motif - protospacer adjacent motif
P-box – prolamin-box
phasiRNA – phased small interfering RNA
PVP – polyvinylpyrrolidone
Qscore – quality score

RefSeq – reference sequence
RNA-seq – RNA sequencing
SDS-PAGE – sodium dodecyl-sulfate polyacrylamide gel electrophoresis
sgRNA – single guide RNA
siRNA – small interfering RNA
SM – spike meristem
snoRNA – small nucleolar RNA
SNP – single nucleotide polymorphism
SPA – storage protein activator
spp – multiple species
SRE – short read eliminator
TE – transposable element
TFL – terminal flower
TGS – transcriptional gene silencing
TPM – transcripts per kilobase million
Tris – tris(hydroxymethyl)aminomethane
TGSeq – targeted genome sequencing
VRN – vernalization
WGS – whole-genome shotgun
WS – whole seed
WSGAR – wheat seed germination associated RNA

ВВЕДЕНИЕ

Актуальность темы. Зерновка — это важнейший орган злаковых растений, обеспечивающий их высокую ценность для питания людей. Развитие зерновки — один из ключевых и практически значимых биологических процессов, включающий разнообразные физиологические, биохимические и транскриптомные изменения (Rogers et al., 1983; Rangan et al., 2017; Wan et al., 2008). Эти изменения и соответствующие им стадии развития зерновки были хорошо охарактеризованы с помощью морфологических, гистологических и биохимических методов исследований. При этом гены, вовлеченные в развитие и формирование зерновки злаковых, динамика их транскрипции и полиморфизм среди существующих сортов остаются малоизученными. Полученные данные об экспрессирующихся генах недостаточны для отражения полной информации об изменениях транскриптома на разных этапах развития зерновки.

Следовательно, для лучшего понимания транскрипционного ландшафта зерновки требуется учитывать его высокую динамичность, а, значит, проводить исследования в различных временных точках, а также применять технологии секвенирования, основанные на длинных ридах, которые позволяют выявлять наиболее широкий набор экспрессируемых генов, связанных с ключевыми процессами, определяющими развитие зерновки.

Поэтому изучение молекулярно-генетических и транскрипционных особенностей развития зерновки является актуальным и крайне востребованным направлением современных исследований. Быстрая сменяемость транскрипционного ландшафта во время развития зерновки и сложность генома злаковых, многие из которых являются аллополиплоидами (например, пшеница и тритикале) и имеют большой размер генома с многочисленными повторами ДНК, существенно усложняют аннотацию и изучение генов развития зерновки. В тоже время новые технологии

секвенирования длинными ридами (Oxford Nanopore Technology и Pacific BioScience) позволяют улучшить точность аннотации генома и детекции новых транскриптов, включая длинные некодирующие РНК (днРНК). Поэтому использование секвенаторов третьего поколения для характеристики транскриптома и изучение структурных особенностей генов злаковых является важной задачей, решение которой позволит получить новые сведения о генетических основах развития зерновки.

Секвенирование третьего поколения позволило получить информацию о генах, вовлеченных в процесс развития зерновки, в частности, глютеинов. Это открыло широкие возможности для идентификации аллелей глютеинов, которые по-разному влияют на хлебопекарные качества. Однако геномная организация даже хорошо изученных глютеинов остается не полностью охарактеризована ввиду того, что их аллели могут различаться структурными вариациями, расположенными как в кодирующей части гена, так и, что немаловажно, в промоторе.

Помимо сложности и полиморфности самих генов, вовлеченных в процесс развития зерновки, их изучение осложняется геномной организацией, как пшеницы, геном которой состоит из трех субгеномов, и содержит в себе повторяющиеся элементы разной природы, так и тритикале, которая, являясь гибридом, объединяет в себе сложнейшие геномы пшеницы и ржи, что затрудняет аннотацию ее генома.

Именно поэтому появилась необходимость оптимизации новых методов, ранее не использованных на зерновых, таких как секвенирование целых генов. Такие подходы, помимо идентификации структурных вариаций, позволяют проводить профилирование метилирования, что является важным эпигенетическим фактором, влияющим на экспрессию генов.

Степень разработанности темы. Исследования, посвященные изучению биологии развивающейся зерновки, проводились на протяжении 20 века (Villareal et al., 1990, Rogers et al., 1983, Shewry et al., 1992). В частности, были установлены основные запасные белки зерновки пшеницы,

которые представлены проламинами (глиадином) и глютелинами (глютенином). Наибольшее влияние на качественные характеристики муки оказывают глютелины, которые в зависимости от молекулярной массы, делятся на две группы: высокомолекулярные HMW-GS-белки (80-130 кДа) и низкомолекулярные LMW-GS-белки (30–50 кДа) (Gianibelli et al., 2001). HMW-GS представляют собой полиморфное семейство, кодируемое тремя локусами (*Glu-A1*, *Glu-B1* и *Glu-D1*) (Bietz et al., 1975, Payne et al., 1982, Branlard et al., 1985), каждый из которых включает в себя два ортологичных гена, кодирующих субъединицу x -типа с большей молекулярной массой и субъединицу y -типа с меньшей молекулярной массой (Nucia et al., 2019). Разные аллели HMW-GS влияют на хлебопекарные качества пшеницы по-разному (Payne et al., 1987). Аллели глютелинов с большим молекулярным весом, такие как *Glu-A1(1)/(2*)*, *Glu-B1(7+8)/(7+9)/(17+18)*, считаются наиболее ценными. В связи с тем, что повышается интерес исследователей к высокомолекулярным глютелинам из-за их полиморфизма, многие работы направлены на изучение влияния глютелинов на качество хлебопечения с целью исключения возможной отрицательной зависимости между урожайностью и процентным содержанием белка в зерне (Rosa Filho, 1997).

Гены глютелинов, как и многие другие гены в геноме злаковых содержат повторяющиеся элементы разной природы. Поэтому установление полной нуклеотидной последовательности таких генов долгое время оставалось трудоёмкой задачей. Разработка платформ для секвенирования длинными ридами совершила революцию в области сборки геномных последовательностей. Использование данного подхода позволило существенно улучшить сборку геномов злаковых, включая пшеницу, ячмень и рожь (Zimin et al., 2017, IWGSC et al., 2018, Li et al., 2021). Это открыло широкие возможности для поиска новых генов и их аллелей среди разных сортов. С помощью секвенирования длинными ридами были отсекужены разные ткани пшеницы и изучена сравнительная

организация кластеров генов глиадинов, точная сборка которых практически невозможна с помощью коротких ридов (Aury et al., 2022).

Данные нанопорового секвенирования ДНК могут быть использованы для установления паттернов метилирования цитозина в геноме на совершенно новом уровне чувствительности. Данный метод позволяет изучить метилирование цитозина (mC) полноразмерных генов глютеинов с использованием недавно опубликованного алгоритма DeepSignal-plant (https://github.com/PengNi/plant_5mC_analysis), добавляющий тег MM к каждому уникальному выравниванию в bam-файле. Появление методов РНК-секвенирования стало новым этапом в изучении развивающейся зерновки (Rangan et al., 2017; Pellny et al., 2012; Pfeiffer et al., 2014; Yu et al., 2016; Guan et al., 2019). Использование транскриптомных данных от секвенаторов второго поколения (секвенирование короткими ридами) показало, что транскрипционная программа значительно меняется на каждом этапе развития зерновки. Анализ транскриптома 35 генотипов пшеницы на двух этапах развития зерновки (14 dpa и 30 dpa) показал, что на 14 dpa наблюдалась активация генов, связанных с синтезом основных компонентов зерна, включая запасные белки и крахмал. Экспрессия генов снижалась к 30 dpa, но увеличивалось разнообразие генов, связанных с защитой зерновки от стрессовых факторов. Это свидетельствует о том, что гены, влияющие на хлебопекарные качества, активны на ранних стадиях развития зерновки (Rangan et al., 2017).

Секвенирование РНК с помощью технологий третьего поколения (секвенирование длинными ридами) позволяет детектировать транскрипцию в локусах генома, которые трудно поддаются анализу с помощью секвенирования короткими ридами. Применение технологии PacBio SMRT позволило получить информацию о транскриптом развивающейся зерновки пшеницы и использовать полученные данные для улучшения аннотирования генома пшеницы (Dong et al., 2015). Это и другие (Liang et al., 2023) исследования продемонстрировали перспективность использования

секвенирования третьего поколения для поиска новых генов растений, которые не были аннотированы ранее.

Цель исследования – аннотация, характеристика и структурно-транскриптомный анализ генов, экспрессирующихся в процессе развития зерновки тритикале и пшеницы с помощью технологии нанопорового секвенирования (Oxford Nanopore Technologies) РНК/кДНК.

Задачи исследования:

1. Адаптировать и сравнить методы целевого высокопроизводительного нанопорового секвенирования полноразмерных последовательностей генов развития зерновки для анализа вариаций полноразмерных генов HMW-GS и их промоторов в коллекции.

2. Изучить особенности транскриптома на разных стадиях развития и в разных частях зерновки с помощью нанопорового секвенирования кДНК.

3. Идентифицировать ранее неаннотированные белок-кодирующие гены и гены длинных некодирующих РНК, экспрессирующиеся на разных стадиях развития зерновки.

4. Изучить полиморфизм ранее неаннотированных генов в коллекции яровой тритикале.

Научная новизна полученных результатов. Впервые проведен комплексный транскриптомный анализ с использованием секвенирования длинными ридами кДНК и РНК на ранней (10 dpa), средней (15 dpa) и поздней (20 dpa) стадиях развития зерновки тритикале. Впервые идентифицированы 7128 ранее неаннотированных генов, экспрессирующихся на разных стадиях развития зерновки тритикале.

Было продемонстрировано, что Cas9-опосредованное нанопоровое секвенирование (nCATS) можно использовать в качестве потенциального инструмента для таргетного секвенирования целевых генов яровой тритикале. Определены достоинства и недостатки этого подхода. Полученная глубина секвенирования была достаточной для различных задач, включая обнаружение InDels и SNPs, гаплотипирование и профилирование

метиляции. Впервые установлен профиль распределения метилирования цитозина по всей длине генов глютенинов, включая промоторную область, и показано, что данные гены имеют существенное метилирование кодирующей области.

Был впервые оптимизирован метод нанопорового секвенирования ампликонов (ONT Amplicon-Seq) для выявления аллельных вариаций в шести генах HMW-GS. Использование ONT Amplicon-Seq позволило выявить SNPs, InDels, новые аллельные варианты, а также различия как в кодирующей области генов глютенинов, так и в промоторах, которые невозможно выполнить с помощью электрофореза SDS-PAGE. Данный метод, в отличие от nCATS, позволяет проводить анализ варибельности целевых генов у большого числа образцов, но не подходит для установления профилей метилирования целевых генов.

Теоретическое и практическое значение полученных результатов.

По результатам проведённых исследований значительно расширен список известных генов пшеницы и тритикале, экспрессирующихся на разных стадиях развития зерновки. Идентифицированные ранее неаннотированные гены используются для последующего функционального анализа, что позволит глубже понять молекулярные механизмы, вовлеченные в развитие зерновки. Идентифицированные гены могут быть использованы для поиска мишеней для маркер-опосредованной селекции и ускоренного отбора ценных генотипов с улучшенными хлебопекарными качествами. Для решения задач генотипирования в работе были оптимизированы и использованы новые методы секвенирования целых генов, а именно Cas9-опосредованное нанопоровое секвенирование и метод ONT Amplicon-Seq. Результаты, полученные в ходе данного исследования, показывают высокий потенциал применения технологии Oxford Nanopore для анализа генов HMW-GS и оценки их профиля метилирования. Основным преимуществом ONT Amplicon-Seq является его применимость в практической селекции для идентификации известных и выявления новых аллелей целевых генов в

обширном генетическом разнообразии тритикале и пшеницы, а также других видов с большими и сложными геномами.

Методология и методы исследования. Работа выполнена с использованием современного оборудования и применением соответствующих мировому уровню исследований методов молекулярной биологии и биоинформатики. Ключевой инструмент для идентификации ранее неаннотированных генов, новых аллелей, структурных вариаций и эпигенетических характеристик является высокопроизводительная технология Oxford Nanopore.

Основные положения, выносимые на защиту:

1. С помощью Cas9-опосредованного целевого нанопорового секвенирования (nCATS) выявлены структурные вариации (InDels и SNPs) полноразмерных генов высокомолекулярных глютеинов *Glu-1A α* , *Glu-1B α* и *Glu-1B β* гексаплоидной тритикале и показан высокий уровень метилирования кодирующей области гена по сравнению с промоторной.

2. Метод нанопорового секвенирования ампликонов (ONT Amplicon-Seq) позволяет быстро и эффективно проводить генотипирование полноразмерных генов в коллекциях зерновых.

3. Нанопоровое секвенирование транскриптома развивающейся зерновки тритикале идентифицировало 7128 (17,6%) ранее неаннотированных генов, экспрессирующихся на разных стадиях развития зерновки.

4. Значительная часть (более 15%, или 4697) транскриптов, экспрессирующихся в процессе развития зерновки тритикале, относится к длинным некодирующим РНК, гены большинства из них (67,5%, или 3174) не были аннотированы ранее.

5. Анализ полиморфизма генов длинных некодирующих выявил наличие различных аллелей среди 23 сортообразцов яровой тритикале.

Личный вклад соискателя. Соискателем изучены научные труды, посвященные теме исследования, разработаны и проведены экспериментальные работы, освоены методы анализа, проведена интерпретация полученных результатов и подготовка текстов и иллюстраций для публикаций и диссертационной работы.

Степень достоверности. Достоверность научной работы основана на обширном объеме экспериментальных исследований, проведенных на сертифицированном оборудовании с применением методов, соответствующих современным стандартам и общемировым требованиям. Эти исследования отличаются высоким уровнем чувствительности и объективности. Также значительное внимание уделено использованию программного обеспечения, необходимого для проведения биоинформационного и статистического анализа экспериментальных данных. В работе продемонстрирована воспроизводимость результатов в различных условиях экспериментов. Выводы, представленные в диссертации, тщательно обоснованы теоретически и подкреплены экспериментальными данными, что гарантирует их соответствие целям и задачам исследования.

Апробация работы.

Экспериментальные данные, полученные в ходе выполнения диссертационных исследований, были доложены и обсуждены на международных и российских конференциях и конкурсах в том числе: 20-я Всероссийская конференция молодых учёных, посвященная памяти академика РАСХН Г.С. Муромцева «Биотехнология в растениеводстве, животноводстве и сельскохозяйственной микробиологии» (Москва, 2020); Международная научно-практическая конференция. «Научное обеспечение технологического развития и повышения конкурентоспособности в пищевой и перерабатывающей промышленности» (Москва, 2020); конкурс работ молодых ученых ГБС РАН, посвященный году науки и технологий в России (I место, Москва, 2021); VII Всероссийский конкурс научно-исследовательских работ студентов и аспирантов «Наука будущего - наука

молодых» (Новосибирск, 2022); I Международная молодежная конференция. «Генетические и радиационные технологии в сельском хозяйстве» (Обнинск, 2022); Международная конференция «Генетика, биотехнология, селекция, семеноводство, технологии выращивания и переработки тритикале», посвященная 110-летию со дня рождения основателя ВНИИСБ академика Н.В. Турбина» (Москва, 2022); V Вавиловская международная конференция «Тритикале – культура будущего» к 85-летию У.К. Куркиева (Санкт-Петербург, 2022); II Международная молодежная конференция. «Генетические и радиационные технологии в сельском хозяйстве» (Обнинск, 2023); XXIII-я научная конференция молодых ученых "ТРИТИКАЛЕ: генетика, биотехнология, селекция и семеноводство, технологии выращивания и переработки" (Москва, 2023); VII Международная научная конференция Генетика, геномика, биоинформатика и биотехнология растений» (PLANTGEN 2023) (Казань, 2023); VIII Съезд Вавиловского общества генетиков и селекционеров, посвященному 300-летию российской науки и высшей школы (ВОГиС, Саратов, 2024).

Публикации. По материалам диссертации опубликовано 3 научные статьи в журналах первого квартиля и имеется один патент на изобретение.

Диссертационное исследование удовлетворяет требованиям пункта 9 «Положения о порядке присуждения ученых степеней, предъявляемым к диссертациям на соискание ученой степени кандидата биологических наук по специальности 1.5.7. - «Генетика» и соответствует заявленной специальности, о чем свидетельствуют пункты «Паспорта специальности»: 2. Геномы, их структура и функция; 7. Структурная и функциональная геномика. Эволюционная геномика; 8. Эпигенетика: эпигеном/эпипротеом/эпитранскриптом; 10. Молекулярно-генетические механизмы основных биологических процессов; 16. Генетическая/молекулярно-генетическая биоинформатика и методы многомерного анализа; 17. Частная генетика микроорганизмов, растений и животных; 25. Прикладные аспекты генетики.

Объем и структура диссертации. Диссертационная работа изложена на 145 страницах, состоит из введения, 3 разделов, заключения, выводов, списка сокращений, списка использованной литературы и 3 приложений. Текст содержит 11 таблиц, 30 рисунков. Список использованной литературы включает 271 источник, в том числе 254 – на иностранных языках.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Характеристика тритикале

1.1.1. Происхождение и биологические особенности тритикале

Получить первый стабильный амфиплоид между пшеницей и рожью путем спонтанного удвоения числа хромосом удалось немецкому селекционеру Wilhelm Rimpau. В 1888 году путем серии скрещиваний двух родительских форм исследователю удалось получить 12 фертильных зерновок, из которых в дальнейшем были получены однородные по внешнему виду потомки гибридов с сохранением родительских характеристик на протяжении последующих поколений. Позднее Linschau и Oehler в 1935 году и Müntzing в 1936 году доказали, что октоплоидный амфиплоид Rimpau имеет 56 хромосом, полученных из гексаплоидной пшеницы и диплоидной ржи (Mergoum, Macpherson, 2004).

Несмотря на то, что потомки после этих скрещиваний превосходили своих родителей по вегетативной массе, они достаточно сильно отставали от пшеницы по урожайности, главным образом из-за частичной стерильности. Müntzing нашел решение проблемы в 1938 году путем разработки метода удвоения хромосом растений с использованием колхицина. Хотя многие первичные октоплоидные гибриды были получены с помощью данной методики, применение колхицина на самом деле стало наиболее полезным для выведения гексаплоидной тритикале. Müntzing в 1979 году сделал предположение, что без изобретения метода удвоения хромосом весь интерес к тритикале сошел бы на нет после получения октоплоидной формы.

Первичные гексаплоидные формы были получены при скрещивании тетраплоидной пшеницы и ржи, хотя их получение значительно облегчилось после усовершенствования методики с применением колхицина. Термины «первичные» и «вторичные» формы тритикале были впервые предложены Kiss (1974). Такое разделение оказалось общепринятым и позволяет

разделять тритикале, полученные от скрещивания форм одного уровня плоидности (первичные), и вторичные – полученные от скрещивания форм разного уровня плоидности.

Современная классификация в разных странах вариативна, но в основе лежит предложенная Oettler (1998) и основанная на цикле получения: первичная и вторичная; и уровне плоидности: тетра-, гекса- и октоплоидная (Пыльнев, Соловьев, 2008; Пыльнев и др., 2022).

Скрещивание гексаплоидной тритикале с гексаплоидной пшеницей иногда приводит к заменам между хромосомами R-ржи и D-геномом пшеницы. Первоначально информация о подобной замене была отмечена Gustafson и Zillinsky (1973). Они смогли идентифицировать наличие хромосомы 2D при замене хромосомы 2R в португальской линии яровой тритикале Armadillo. Тетраплоидную тритикале (AARR или BBRR), полученную с помощью отбора на гексаплоидном уровне скрещивали с октоплоидной тритикале (AABBDDRR) с целью получения замещений D/A и D/B.

В то же время было выяснено, что единственным замещением, которое присутствует в гексаплоидной тритикале, является 2D/2R, с которой связаны многие желательные характеристики тритикале (Zillinsky, 1980). Однако кариотипированные линии озимой тритикале имеют неизменную структуру хромосом AABBRR (Lukaszewski, Apolinarska, 1987, Seal, Bennett, 1982).

У гибридов с участием линий с 2D/2R-замещением хромосом показан широкий формообразовательный процесс по признакам продуктивности и высоты растений (Соловьев, Вишнякова, 1996; Соловьев, 2000; Дивашук и др., 2010б).

При этом среди коллекционных форм яровой тритикале встречаются генотипы, имеющие разные хромосомные варианты: не только замещения, но и хромосомные перестройки, включая межсубгеномные транслокации (Соловьев, 2000; Дивашук и др., 2010а; Шишкина и др., 2009).

Saranya и Reddy в 2017 году, проводя характеристику 82 гексаплоидных образцов тритикале на основе замены R/D, обнаружили, что замещение хромосом 2R/2D привело к большей озерненности, а замещение 4R/7R оказало промежуточный эффект.

В результате плодотворной работы селекционеров в конце 60-х годов производителям стали доступны первые коммерческие сорта тритикале (Blum, 2014). Растущий спрос на продовольственные ресурсы привел к росту интереса к получению новых сортов тритикале. Несмотря на то, что в процессе выведения пригодных для выпечки сортов все еще встречаются трудности, новые подходы к конечному использованию тритикале в большей степени предполагают её потребление человеком. Использование тритикале в пищу человеком еще не получило широкого распространения. Хотя мука и продукты из тритикале доступны в продаже, их наличие ограничено. В целом, современные сорта как озимой, так и яровой тритикале обладают достаточно хорошим качеством зерна, пригодным к хлебопечению (Грабовец, Кромаль, 2018; Абделькави и др., 2020).

Озимая и яровая тритикале широко возделываются в более чем 40 странах мира. Крупнейшие посевные площади тритикале находятся в Польше (1281600 га), она же и является крупнейшим производителем зерна (5349420 т), Россия в мировом рейтинге по производству тритикале занимает 7 место с урожаем 288000 тонн (<https://www.fao.org/faostat/en/#data/QV>).

В Государственном реестре селекционных достижений РФ допущено к использованию в производстве 102 озимых и 24 яровых сорта российской, белорусской и украинской селекции, из которых 6 и 4 сорта, соответственно, введены в реестр в 2022 году. Стоит отметить, что среди введенных в реестр сортов 2022 года – два, Ботаническая 4 и Тимирязевская 42, – являются результатом селекции ВНИИСБ совместно с Главным ботаническим садом им. Н.В. Цицина (<https://reestr.gossortrf.ru>).

1.1.2. Селекционно-генетические аспекты качества зерна тритикале

Селекция культуры тритикале на качество сопряжена с такими трудностями, как большая фенотипическая изменчивость количественных признаков, характеризующих качество зерна (Villareal et al., 1990).

Одной из основных и трудных задач является селекция на высокий уровень белка с высоким качеством. Для решения такой задачи необходима биохимическая технологическая оценка и генетический анализ селекционного материала (Гуляев, 1998; Казарцева и др., 1990; Сокол и др., 1993).

Поиск генотипов, которые будут отличаться показателями высокого содержания белка и хорошим качеством, является одним из самых перспективных путей создания высокобелковых форм (Dong et al., 1991, Zilic et al., 2011).

Важным вопросом является идентификация генетических закономерностей накопления белка в зерне разных сортов пшеницы и тритикале. Эту проблему возможно решить, используя компоненты высоко- и низкомолекулярных (HMW-GS и LMW-GS, соответственно) субъединиц глютеина для методов генотипирования (Dhaka, Khatkar, 2015). Ряд научных исследований уже используют и приводят в качестве примеров полноценные характеристики отдельных фракций белков (He et al., 2005, Kaya, Akcura, 2014, Ammar et al., 2000).

Было показано, что определенные компоненты запасных белков зерна имеют четкую генетическую детерминацию, а также характеризуются фенотипическими проявлениями, независимыми от внешних условий. (Johansson et al., 2001). Запасные белки зерна используются в качестве маркеров, которые позволяют идентифицировать генотипы с наилучшими характеристиками клейковины (глютена). Пищевая и биологическая ценность тритикале оценивается по многим критериям, поскольку получение высокопродуктивных сортов – трудоемкая задача для любого селекционера. В ряд критериев входят отдельные комплексы запасных белков, содержащиеся в

зерновке, а также различные условия её развития. Получение новых линий с улучшенными характеристиками является энергозатратным и дорогостоящим процессом (Semenov et al., 2018).

Важное место в определении качества зерна занимают два гена локуса *Glu*. В ряде работ показаны локусы *Glu-1* и *Glu-3*, которые кодируют высокомолекулярные (HMW-GS) и низкомолекулярные (LMW-GS) субъединицы глютеинов, соответственно (Krupnov, Krupnova, 2012). Компоненты LMW-GS глютеинов в настоящее время активно исследуются. Сложность заключается в их разнообразии и высоком сходстве с генами глиадинов (Andersen, Lübberstedt, 2003). Изучение высокомолекулярных глютеинов является более трудоемким процессом, чем низкомолекулярных. В основном это связано со сложностью разделения компонентов глютеина и глиадина с помощью белкового электрофореза в полиакриламидном геле (Liu et al., 2005; Park et al., 2011). Однако изучение и совершенствование методов создания новых функциональных маркеров для идентификации специфических генов, кодирующих низкомолекулярные глютеины, считается новым поворотным моментом в науке и одним из новейших направлений исследований. В настоящее время для высокомолекулярных глютеинов существует ряд белковых маркеров, благодаря которым возможно повысить качество зерна (Shewry et al., 1992)

У рода *Triticum* есть дикие родственники, которые отличаются аллельным разнообразием в локусе *Glu-1*. Такая особенность позволяет применять дикие виды для расширения и улучшения генофонда «слабых» по качеству сортов. Стоит отметить, что такой подход применяется для решения проблем качества зерна. (An et al., 2005).

1.2. Транскриптомные особенности развития зерновки

Одним из наиболее важных и практически значимых биологических процессов является развитие зерновки. Он включает в себя различные

физиологические, биохимические и транскриптомные изменения (Rogers et al., 1983; Rangan et al., 2017; Wan et al., 2008).

Выделяют пять этапов развития зерновки, каждый из которых соответствует уникальному набору признаков. На первом этапе (1-7 дней после цветения (dpc)), который называется «ценоцитарный», ткани эмбрионов еще не дифференцированы, и зародыш не может быть отделен от окружающих тканей (рисунок 1). На втором этапе (7-14 dpc) зародыш может быть легко отделен от окружающих тканей, и происходит дифференцировка. В начале третьего этапа (14-21 dpc) появляются зачатки боковых корней, которые хорошо развиваются к концу этапа. На четвертом этапе (21-31 dpc) происходит дальнейший рост и дифференцировка тканей в первичных и боковых корнях. На 5-ой стадии (31-50 dpc) полностью дифференцирован зародыш, сформирована зерновка, наступает полная спелость зерна (Rogers et al., 1983). Считается, что период подразумевает целлюляризацию, налив зерна и созревания.

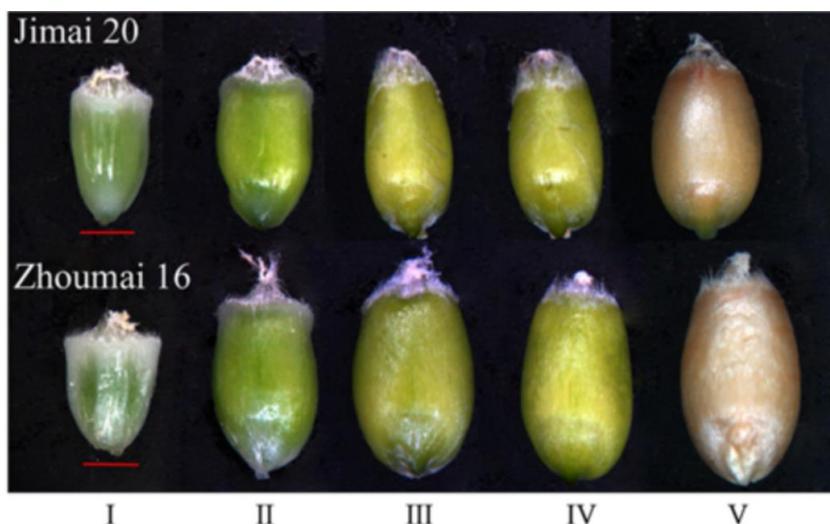


Рисунок 1. Этапы формирования зерновки пшеницы (Guo et al., 2012).

Семена злаков состоят из двух основных тканевых фракций: зародыша и крахмалистого эндосперма/алейрона, на долю которого приходится около 80% зрелой зерновки. Эндосперм является основным местом хранения крахмала и белков. Следовательно, урожайность зерна и качество конечного продукта в значительной степени определяются размером и составом эндосперма.

Как репродуктивный орган, колос пшеницы тесно связан с урожайностью. Например, количество зерен в колосе является одним из важнейших факторов, определяющих урожайность (Sreenivasulu, 2012). Развитие колоса зерновых, таких как пшеница и тритикале, включает в себя последовательное формирование и развитие колосковой меристемы (SM), зачаточной чешуи, цветочной меристемы (FM), тычинок и пестика (Wang, 2011). Недавно были охарактеризованы несколько важных регуляторов, вовлеченных в развитие колосьев пшеницы. Например, *TaTFL1-2D*, ортолог гена *TFL1* арабидопсиса, регулирует сложность колоса пшеницы путем изменения количества колосков, цветков и зерен в колосе, в то время как гены *VRN1*, *FUL2* и *FUL3* выполняют важную функцию в определении идентичности колосковой меристемы, а также влияют на время цветения и высоту растений. У тройного мутанта *vrn1ful2ful3*-null меристема соцветия сформировала нормальную структуру колоса, но затем боковые меристемы образовали вегетативные побеги, дополненные листьями вместо колосков. Эти результаты свидетельствуют о существенной роли этих трех генов в формировании колоса (Li, 2019). Ген *FZP* изменяет структуру колоска, регулируя идентичность SM, поскольку его мутация приводит к эктопическому образованию SM и, следовательно, дополнительных колосков (Dobrovolskaya, 2015). Более того, многие исследования выявляют профилирование экспрессии генов во время формирования колосьев пшеницы (Dong et al., 2015, Rangan et al., 2017, Liang et al., 2023).

Эндосперм образуется в результате второго оплодотворения внутри зародышевого мешка в процессе слияния двух центральных клеточных ядер со спермием, образуя триплоидный эндосперм. Последующее развитие эндосперма злаковых можно разделить на ряд стадий (Evers, 2002). Первой из них является свободное ядерное деление, которое происходит внутри первичной клетки эндосперма с образованием ценоцита, который у пшеницы может содержать более 2000 ядер к 72 часам после оплодотворения (Chojecki, 1986). Затем происходит образование клеток в течение примерно

24 часов путем заложения клеточных стенок, за которым следует период примерно в 10 дней, в течение которого происходит деление, расширение и дифференцировка клеток, дающих характерную структуру эндосперма с общим количеством до 300000 клеток (Evers, 2002; Chojecki, 1986).

1.2.1 Динамика транскриптома в процессе развития зерновки

За последние годы проведено множество исследований по профилированию транскриптома экспрессирующихся генов и клеточных процессов развития семян арабидопсиса *Arabidopsis thaliana* (Le et al., 2010; Бельмонте и др., 2013), риса *Oryza sativa* (Xu et al., 2012; Gao et al., 2013), настурции *Tropaeolum majus* (Jensen et al., 2012) и сои *Glycine max* (Jones, Vodkin, 2013).

В своей работе Laudencia-Chingcuanco D.L. (2007) смог показать распределение генов пшеницы, которые по-разному экспрессируются в процессе развития колоса, что показало значительное перепрограммирование на ранних стадиях развития зерновки, особенно между 7 и 14 dpa, которые совпадает с ранними стадиями налива зерна. В значительной доле (48-55%) выявленных дифференциально экспрессирующихся генов не была установлена биологическая функция. Гены, для которых найдена биологическая аннотация, в большинстве случаев относились к категории клеточного метаболизма макромолекул или образования метаболитов-предшественников, а также энергетического обмена. Около 20% этих генов достигают максимального уровня транскрипции между 14 и 21 dpa, когда происходит налив зерновок. Группа генов (69 транскриптов), которые дифференциально экспрессировались только между 14 и 21 dpa, содержала три взаимодействующих белка 14-3-3b. Белки 14-3-3 – это семейство белков, которые играют регулируемую роль в нескольких процессах развития растений, таких как передача сигнала, точка контроля, апоптоз и пути восприятия питательных веществ (Ferl, 1996; Fulgosi et al., 2002).

Были идентифицированы три гена с неизвестной функцией, которые точно повторяют паттерн экспрессии 14-3-3. Эта коэкспрессия может свидетельствовать об их возможном совместном участии с этими регуляторными белками. Было высказано предположение, что 14-3-3b может ингибировать биосинтез крахмала (Fulgosi et al., 2002; Sehnke et al., 2001). Это согласуется с наблюдением Yu (2016) о том, что накопление транскрипта 14-3-3b обратно пропорционально накоплению крахмала. Среди генов, экспрессия которых значительно повышается только в период 14-21 dpa, есть гены, участвующие в выработке предшественников метаболитов и энергии, а также те, которые, возможно, участвуют в защитных механизмах. У примерно половины генов в этой группе экспрессия значительно снижалась. К ним относятся EF1-а, актин, α -тубулин, гены углеводного обмена и неизвестные гены (Laudencia-Chingcuanco et al., 2007).

Гены, чей пик экспрессии приходится на 3-7 dpa, согласуются с клеточными событиями, происходящими на стадии роста зерновки. Эти гены участвуют в делении клеток и метаболизме нуклеиновых кислот (например, гистонов, тубулинов), фотосинтезе (например, белков, связывающих хлорофилл, ферредоксин) и метаболизме белков (например, белков рибосом, факторов инициации трансляции и удлинения, протеосомных белков) (Воротынцева, 2021). Значительное количество генов в этих кластерах (35-41%) все еще выполняет неизвестные биологические функции в клетке.

Основной метаболической активностью, происходящей на стадии наполнения зерна (14-24 dpa), является синтез и накопление молекул крахмала и белков. Основными запасными белками являются проламины и глютенины, которые являются основными компонентами клейковинного полимера, определяющего экономическую значимость пшеницы. Эти гены демонстрируют резкую скорость увеличения уровня транскрипции между 7 и 14 dpa, пик к 21 dpa, а затем выходят на плато. Также в наливе зерна участвуют: гены запасных белков, гены твердозёрности (Grain Softness Protein (GSP)) и ингибиторы α -амилазы (участвуют в метаболических и, возможно,

защитных функциях). Максимальная скорость накопления белка наблюдается в период от 7 до 14 dpa, что согласуется с результатами, полученными другими учеными (Altenbach, Kothari, 2004; Clarke et al., 2000). Это наблюдение подтверждает гипотезу о том, что транскрипция, процессинг поли(А)-РНК и трансляция генов запасных белков тесно взаимосвязаны в развивающейся зерновке пшеницы, что позволяет предположить, что основной контроль за накоплением запасных белков осуществляется на уровне транскрипции.

Пик уровня транскрипции генов, демонстрирующий паттерн экспрессии развития колоса, приходится на 14 dpa. Они представляют собой в основном метаболические ферменты, участвующие в углеводном и белковом обмене: ортофосфатдикиназу, аспарагиновую протеазу, аспаратаминотрансферазу, аланинаминотрансферазу, сахарозосинтазу 2 и АДФ-глюкозопирофосфорилазу (Tetlow et al., 2004). Колоколообразная экспрессия генов углеводного и белкового метаболизма коррелирует с наблюдаемой скоростью накопления крахмала и белка в зерне, за исключением того, что максимальная скорость накопления этих продуктов наблюдается между 14 dpa и 21 dpa, следуя максимуму экспрессии транскрипта. Следует отметить, что накопление транскриптов, участвующих в биосинтезе крахмала, происходит уже через 3 дня. Это может быть результатом синтеза крахмала в ткани околоплодника, которая на этой стадии все еще фотосинтезирует. Однако присутствие генов биосинтеза крахмала на 3 dpa также может подтвердить морфологическое наблюдение о том, что крахмальные зерна присутствуют уже на ценоцитарной стадии развития эндосперма. В своем обзоре Vilecen et al. (2005) предположили, что белки, синтезируемые в этот период, могут быть вовлечены в реакции на стресс, запрограммированную гибель клеток, регуляцию развития и метаболизм тяжелых металлов.

1.3. Геномные особенности зерновых культур

Сложность сборки геномов пшеницы и тритикале обусловлена не только их большим размером (в 5 раз превышающим размер генома человека), но и высокой долей относительно длинных, почти идентичных повторов, большинство из которых состоят из транспозонов (Li, Zhang, Fellers et al., 2004). Повторяющиеся последовательности и полиплоидия создают серьезные проблемы при создании геномных сборок. Еще одна важная особенность при сборке генома пшеницы и тритикале заключается в том, что они являются гексаплоидными формами и их геномы состоят из трех субгеномов: А, В и D для пшеницы и А, В и R для тритикале, каждый из которых содержит 7 хромосом и имеет много общих областей с высокой степенью сходства.

Таким образом, программы сборки генома сталкиваются с вдвойне сложной проблемой: во-первых, геном с большим количеством повторяющихся элементов разной природы, во-вторых, каждая хромосома существует в 6 копиях с различной степенью внутри- и межхромосомного сходства. Поскольку повторы намного длиннее, чем длина ридов Illumina, появилась необходимость разрабатывать новые способы сборки геномов для идентификации различных повторов.

Наличие повторов делает сборку генома особенно сложной, даже учитывая результаты секвенирования, полученные с помощью длинных ридов и биоинформатических достижений в технологии сборки генома (Alonge et al., 2020).

Несмотря на то, что последние сборки для пшеницы, полученные с использованием различных технологий, являются ценным и проверенным ресурсом, они все еще могут содержать большое количество пробелов (Zimin et al., 2017).

В последнее время увеличивается интерес к технологиям секвенирования Pacific Biosciences (PacBio) и Oxford Nanopore (ONT), т.к.

они позволяют получать длинные риды ДНК и повышать аннотацию сложных геномных сборок. Длинные риды и производительность секвенирования позволяют быстро и экономично собирать многие крупные и сложные геномы (Gnerre et al., 2011; Lam et al., 2011). Сравнения между собранными геномами выявили множество вариаций последовательностей, имеющих важное функциональное значение, которые не были обнаружены путем прямого сопоставления ридов последовательностей с общим эталоном (Gan et al., 2011; Vishara et al., 2015).

Недавние разработки в области подготовки библиотек для секвенирования (Clavijo et al., 2017), алгоритмов сборки и создания полноразмерных скаффолдов значительно улучшили сборку полногеномных фрагментов. К ним относятся подготовка библиотек без использования ПЦР для уменьшения количества ошибок в процессе секвенирования Illumina (Aird et al., 2011), и применение алгоритмов, которые сохраняют аллельное разнообразие во время сборки (Weisenfeld et al. 2014).

1.4. Запасные белки зерновки пшеницы и тритикале

1.4.1. Характеристика глютелинов

Комплекс запасных белков зерна пшеницы и тритикале в большей степени влияет на ее питательную и биологическую ценность. В частности, мукомольные свойства и питательность конечных продуктов во многом определяются составом клейковины – основного запасного компонента в эндосперме зерна. Клейковина представлена белками проламинами (глиадином) и глютелинами (глютенином). Наибольшее влияние на качественные характеристики муки оказывает качественное и количественное содержание нерастворимых в воде белков глютелинов. Глютелины представляют собой высоко агрегированные белки с огромной молекулярной массой, ассоциированные и соединенные между собой межмолекулярными

дисульфидными связями, с помощью которых формируется макромолекулярный каркас клейковины (Одинцова, 2010).

Глютенины в зависимости от молекулярной массы делятся на две группы: высокомолекулярные HMW-GS-белки (80 - 130 кДа) и низкомолекулярные LMW-GS-белки (30 – 50 кДа) (Gianibelli et al., 2001). При замешивании теста, эти глютениновые субъединицы полимеризуются межмолекулярными дисульфидными связями, которые играют главную роль в реологических свойствах теста (упругость, растяжимость, эластичность, энергия теста, время его развития, стабилизации и начала разжижения) (Дивашук и др., 2012).

На текущий момент низкомолекулярные глютенины - LMW-GS мало изучены, по сравнению с высокомолекулярными - HMW-GS. Это связано, прежде всего, с их структурной организацией, которая включает в себя огромное количество полипептидов, что вызывает сложности в изучении генетического контроля качества зерна.

Большинство исследований показывают, что на качественные характеристики макаронных изделий, полученных из сортов твердой пшеницы, влияют низкомолекулярные глютенины, в то время как хлебопекарные свойства мягких сортов пшениц зависят от генов, которые кодируют высокомолекулярные глютенины (Anjum et al., 2007). В тоже время имеются данные о том, что высокомолекулярные глютенины не оказывают такого же сильного влияния на характеристики макаронной продукции из-за их ограниченной генетической изменчивости. (Du Cros, 1987).

Исследования о влиянии низкомолекулярных глутенинов на хлебопекарные качества не дали однозначных результатов (Payne et al., 1987; Liu et al., 2005). В последнее десятилетие низкомолекулярные глютенины стали объектом многих исследований из-за их высокого разнообразия в зерновых культурах (Branlard et al., 2001). В связи с тем, что многие исследования показывают, что высокомолекулярные глютенины по истечении некоторого времени могут потерять свои положительные эффекты для

хлебопечения, многое направлено на изучение влияния низкомолекулярных глютеинов на качество выпечки с целью избегания возможной отрицательной зависимости между урожайностью и содержанием белка в зерне (Rosa-Filho, 1997).

1.4.2. Низкомолекулярные глютеины (LMW-GS)

Преобладающим типом белка в зерновке являются низкомолекулярные субъединицы глютеина LMW-GS. Их доля составляет 20% от общего количества белков глютена (Wieser, Kieffer, 2001).

Локусы низкомолекулярных (LMW-GS) белков (*Glu-A3*, *Glu-B3*) расположены на коротких плечах хромосом 1AS, 1BS рядом с локусами глиадинов *Gli-1* (Payne et al., 1984). Локус *Glu-3* содержит мультигенные семейства, которые включают 30–40 генов (Sabelli, Shewry, 1991).

Предложенная номенклатура Lew et al. (1992) и Cloutier et al. (2001) подразделяет LMW-GS-белки на три группы в зависимости от первого аминокислотного остатка в N-концевых последовательностях. Эти три основные группы первоначально характеризовались их зрелыми пептидными последовательностями, начинающимися изолейцином (LMW-i-тип), серином (LMW-s-тип) или метионином (LMW-m-тип) на их N-конце (Masci et al., 2000). Гены i-типа экспрессируются исключительно в геноме А. Иная классификация, предложенная Carrillo et al. (1990), делит глютеины на 2 типа: LMW-1 и LMW-2, заключается в различиях 13 аминокислот, расположенных в повторяющихся доменах.

Кроме того, низкомолекулярные (LMW-GS) белки дифференцируются по молекулярной массе на тяжелые субъединицы В (40000 – 50000) и более легкие – С (30000 – 40000) (Singh, Shepherd, 1988). Субъединица В является наиболее изученной. Несмотря на то, что С-субъединицы представлены большим разнообразием белковых компонентов, их изучение осложняется трудностью их разделения с глиадинами. Было обнаружено, что аминокислотная последовательность белков типа С почти на 95% идентична

последовательности глиадинов (Masci et al., 2002), за счет этого структура молекул LMW-GS-белков схожа с глиадинами у-типа. N-концевая половина полипептида состоит из повторяющихся последовательностей (богатых пролином, бедных цистеинами), а C-концевая часть – из уникальных последовательностей (бедных пролином, богатых цистеином). Имеется короткий уникальный N-концевой участок (12 - 14 остатков), который может содержать остаток цистеина. Предполагается, что LMW-GS-белки содержат восемь остатков цистеина, участвующих в образовании внутри- и межмолекулярных дисульфидных связей в глютене (Воротынцева, 2021).

1.4.3. Высокомолекулярные глютенины (HMW-GS)

Многие исследования указывают на то, что высокомолекулярные глютенины (HMW-GS) в большей степени влияют на эластичность теста, по сравнению с низкомолекулярными глютепинами (He et al., 2005).

Структура белка HMW-GS кодируется полиморфным геном *Glu*, locus которого (*Glu-A1*, *Glu-B1*) расположен на длинных плечах первой гомеологичной группы хромосом 1AL, 1BL (Payne et al., 1984). Гены высокомолекулярных глютепинов кодируют белки так называемых «парных» субъединиц, различающихся по размеру. Лocus *Glu-B1* кодирует два типа HMW-GS глютепинов: x- и y-тип, молекулярная масса которых равна 83 - 88 и 67 - 74 кДа, соответственно (Caballero et al., 2008a), тогда как locus *Glu-A1*, контролирует синтез высокомолекулярных глютепинов x-типа. Ген *Glu-1Ay* всегда неактивен у культурных гексаплоидных сортов пшеницы и тритикале (Shewry et al., 1992). Однако активные гены *Glu-1Ay* идентифицированы у диких пшеничных видов.

Характерная модель высокомолекулярных глютепинов содержит большой центральный повторяющийся домен (480 – 680 аминокислотных остатков с повторами из 6, 9 и 15 аминокислот), а также терминальные домены: N- (81 – 104 аминокислоты) и C-концами (42 аминокислоты). C-концевая область содержит один остаток цистеина, в отличие от N-концевой

области, содержащей от трех до пяти цистеиновых остатков (Miles et al., 1991). Эти остатки цистеина образуют глютеновый матрикс за счет формирования межмолекулярных связей S-S с другими белковыми компонентами (Li et al., 2004).

Была установлена роль дисульфидных мостиков в структуре высокомолекулярных глютенинов. А именно, высокое количество дисульфидных связей между цистеиновыми остатками делает глютен более крепким и в меньшей степени подверженным растворимости в разных растворителях (Kasarda et al., 1976). Низкомолекулярные глютенины, по сравнению с высокомолекулярными, содержат больше цистеиновых остатков, благодаря чему LMW-GS-белки в большей степени влияют на формирование крупных белковых полимеров, придающих тесту более твердую текстуру. Обогащение LMW-GS глютенинами зерновки приводит к повышению прочности теста (Воротынцева, 2021).

1.4.4. Полиморфизм генов высокомолекулярных глютенинов

Гомеологичные локусы *Glu-A1*, *Glu-B1* очень полиморфны по своей природе, и их аллельные варианты отвечают за различные комбинации HMW-GS-белков в различных сортах пшеницы и тритикале. Полиморфизм высокомолекулярных генов играет важную роль в определении качества муки (He et al., 2002).

В результате древнего события дупликации генов пшеницы в каждом из геномов А, В и D присутствуют два паралога (*x* и *y*), у гексаплоидной тритикале, соответственно, в геномах А и В. Таким образом, локусы *Glu-1* тритикале кодируют субъединицы HMW глютеина *x* и *y*-типа (или *Glu 1-1* и *Glu 1-2* соответственно) и составляют в общей сложности 4 гена *Glu-1* в гексаплоидной тритикале. Гены *Glu-1* не имеют интронов и окружены транспозонами.

Разнообразие изученных образцов яровой тритикале по локусу *Glu-A1* представлено аллельными вариантами, кодирующими исключительно

субъединицы х-типа, *Glu-A1a* и *Glu-A1b*, а также аллелем *Glu-A1c*, при наличии которого в белковом профиле отсутствует HMW-GS, который также не экспрессируется и у тритикале.

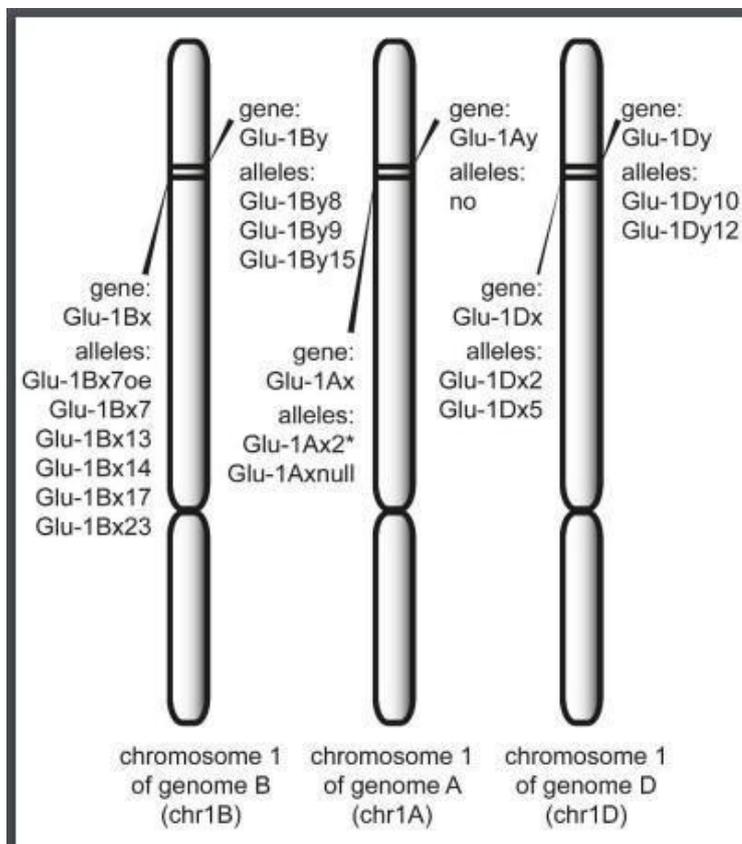


Рисунок 2 – Расположение генов *Glu-1* в трех геномах (A, B и D) гексаплоидной пшеницы и их аллели на хромосоме 1 (Makai et al., 2014)

Локус *Glu-B1* отличается большим полиморфизмом, чем локус *Glu-A1*, и содержит не менее 14 аллелей, различающихся субъединицами различной электрофоретической подвижности. Разнообразие локуса *Glu-B1* представлено следующими аллелями: *Glu-B1a* (7), *Glu-B1b* (7+8), *Glu-B1c* (7+9), *Glu-B1d* (6+8), *Glu-B1e* (20), *Glu-B1f* (13+16), *Glu-B1g* (13+19), *Glu-B1h* (14+15), *Glu-B1i* (17+18) (Wang et al., 2006), *Glu-B1j* (21), *Glu-B1k* (22), *Glu-B1ak* (7*+8*)¹, *Glu-B1c* (7*+9), *Glu-B1u* (7*+8), *Glu-B1al* (7^{OE}+8) (Воротынцева, 2021).

На рисунке 2 представлено расположение генов высокомолекулярных глютенинов гексаплоидной пшеницы, где отображены паралоги генов глютенинов (x- и y-тип), а также указаны их аллели.

Уровни экспрессии генов гомеологов и паралогов *Glu-1* сильно различаются. В целом, гены *Glu-1Bx* имеют самый высокий уровень транскрипции. Гены у-типа имеют наиболее низкую экспрессию. Ген *Glu-1A_у* всегда неактивен как у гексаплоидной пшеницы, так и у ее гибрида тритикале (Shewry et al., 1992). Однако, обнаружены активные гены *Glu-1A_у* у дальних родственников пшеницы (Jiang et al., 2009). Предполагается, что отсутствие активности *Glu-1A_у* является результатом интрогрессии гена *Glu-1A_у* из видов *Triticum urartu*, *Triticum dicoccoides* во время селекции (Margiotta et al., 1996). Инактивация *Glu-1A_у* происходит в результате разных нарушений гена, которые варьируются у сортов. У сорта Chinese Spring была обнаружена вставка ретроэлемента Wis-3 в повторяющуюся область гена *Glu-1A_у* (Gu et al., 2004), а у сорта Cheyenne был идентифицирован стоп-кодон в рамке считывания (Forde et al., 1985).

Различные аллели, кодирующие HMW-GS-белки, по-разному влияют на хлебопекарные качества пшеницы (Payne et al., 1987). Наиболее ценными считаются высокомолекулярные субъединицы, такие как *Glu-A1(1)/(2*)*, *Glu-B1(7+8)/(7+9)/(17+18)*. Полиморфизм локуса *Glu-R1* у гексаплоидной тритикале, а также влияние аллелей этого локуса на хлебопекарные и технологические свойства муки изучены слабо (Груздев, Дудников, Соловьев, 2020).

У тритикале распределение аллелей локуса *Glu-A1* по частоте встречаемости выглядит следующим образом: *Glu-A1b* и *Glu-A1c* примерно в равной степени, аллель *Glu-A1a* встречается реже.

Локус *Glu-A1* является менее полиморфным в сравнении с локусом *Glu-B1*. В локусе *Glu-B1* наиболее редкими являются аллели *Glu-B1e* и *Glu-B1a*, кодирующие субъединицы 20 и 7, соответственно. Наиболее часто встречающимися являются аллели *Glu-B1s* и *Glu-B1r*, обнаруженные только у тритикале (Груздев и др., 2020; Пырсигов и др., 2022).

1.4.5. Регуляция экспрессии генов глютеинов

Запасные белки накапливаются в значительном количестве во время созревания зерновки и служат источником восстановленного азота на ранних стадиях прорастания. Развивающееся зерно пшеницы состоит преимущественно из эндосперма, который является местом синтеза и накопления глютеинов и глиадинов.

Синтез запасных белков зерновки регулируется с учетом особенностей развития, специфики организма и окружающей среды. Установлено, что синтез и накопление запасных белков семян подвергается строгой пространственной и временной регуляции, которая происходит только в эндосперме во время развития семян (Kawakatsu, Takaiwa, 2010).

Координированная экспрессия генов запасных белков контролируется, главным образом, на уровне транскрипции (Soave, Salamini, 1984; Bartels, Thompson, 1986; Heidecker, Messing, 1986; Kreis et al., 1986). Анализ делеции промотора пшеничного гена глютеина HMW *Glu-D1-2*, который кодирует субъединицу 12, установил, что последовательности, необходимые для специфичной для эндосперма транскрипции, находятся в пределах 436 п.н. выше кодона инициации трансляции (Colot et al., 1987; Robert et al., 1989).

Экспрессия всех проламиновых белков следует хорошо охарактеризованному, хотя и меняющемуся, временному паттерну во время развития зерновки. Их транскрипция регулируется транс-действующими факторами, связанными с цис-регуляторными элементами, а также эпигенетическими факторами (Fauteux, Strömviik 2009; Kawakatsu, Takaiwa 2010; Juhász et al., 2011; Wen et al., 2012).

Более ранние исследования показали, что транскрипция генов глиадина и LMW глютеина (*Gli-1,2* и *Glu-3*, соответственно) находится под влиянием метилирования и импринтинга, тогда как экспрессия генов HMW-GS-белков, кодирующих *Glu-1*, менее зависима от этих эпигенетических факторов (Wen et al., 2012). Более тщательный анализ профилей экспрессии

проламиновых генов и их ответов на абиотические стрессы указывает на наличие различных регуляторных механизмов для каждого семейства проламиновых белков (DuPont et al., 2006; Hurkman et al., 2013). Была предложена модель различных механизмов транскрипционной регуляции, основанная на консервативных некодирующих элементах генов LMW-GS-белков, кодирующих *Glu-3* (Juhász et al., 2011).

В транскрипции генов глютеинов участвуют транскрипционные факторы (ТФ), которые связываются специфически с короткими сохраненными последовательностями ДНК (5-15 нуклеотидов), называемыми цис-регуляторными элементами или цис-элементами, которые обычно расположены в проксимальной части промотора генов и характеризуются консенсусным мотивом. У зерновых культур регуляторные механизмы экспрессии генов запасных белков глютеина охарактеризованы, как сеть цис-элементов и взаимодействующих с ними ТФ. Эта сеть состоит из пяти цис-элементов, способных распознавать восемь ТФ, принадлежащих к четырем семействам (*bZIP* семейства Ораque-2 и белки *B3*, *DOF* и *MYB*), которые, как сообщается, являются активаторами генов синтеза запасных белков глютеинов. Кроме того, транскрипционные факторы типа *NAC*, *LEC1* также влияют на экспрессию запасного белка семян (Agarwal et al., 2011; Verdier et al., 2008). В случае пшеницы и тритикале, ни один из этих транскрипционных факторов не экспрессируется исключительно в эндосперме (Dong et al., 2007). Сообщалось, что транскрипционные факторы связывания проламина (WPBF) типа *DOF* конститутивно экспрессируются во всех тканях. Многие сообщения подтверждают, что взаимодействие этих факторов является ключом к тканевой специфичности (Diaz et al., 2005; Qu et al., 2008; Bevan et al., 1993). Сообщалось о взаимодействии *bZIP* и *DOF*, и этот тип взаимодействия, по-видимому, является уникальным для эндосперма однодольных (Agarwal et al., 2011). Также подтверждено взаимодействие транскрипционных факторов *GAMYB* и *DOF*, которые, образуя комплекс, активировали специфическую для эндосперма экспрессию генов (Diaz et al., 2005).

Большинство промоторов генов проламинов, к которым относятся и промоторы генов *Gli-3*, имеют двухсторонний консервативный сайт связывания (BS), называемый эндосперм-боксом (endosperm-box). Он состоит из двух элементов: DOF (DNA-Binding with One Finger) – связывающего мотива, называемого проламиновым боксом (Prolamin-box, P-box), и *GCN4*-подобного мотива, также называемого азотным боксом (Nitrogen-box, N-box). *GLM* мотив распознается *bZIP* ТФ, такими как *BLZ1* и *BLZ2* в ячмене (Vicente-Carbajosa et al., 1998; Oñate et al., 1999) или *SPA* (Storage Protein Activator) в пшенице (Albani et al., 1997). А P-box связывается с ТФ семейства *DOF – PBF* и *SAD* (Vicente-Carbajosa et al., 1997; Mena et al., 1998; Diaz et al., 2005).

Кроме эндосперм-бокса, в промоторах генов проламинов часто присутствуют два дополнительных цис-элемента 5'-ААСА/ТА-3' и 5'-ТАТС/ГАТА-3'. Они способны связываться с ТФ семейства MYB: ТФ *R2R3-MYB* (в частности, *GAMYB*) и *R1MYB* (*MCB1* и *MYBS3*), соответственно, в вегетативных тканях риса, пшеницы и ячменя (Suzuki et al., 1998; Diaz et al., 2002; Chen et al., 2005). И, как было показано, этот комплекс подавляет экспрессию репортерных генов в вегетативных тканях (лист, стебель и корень) трансгенных растений (Yoshihara et al., 1996). Мотив ААСА/ТА наряду с отдельными вариациями P-box, также может модулировать экспрессию генов LMW-GS- белков в эндосперме.

В качестве второй линии регуляции многие ключевые транскрипционные факторы, регулирующие проламины (например, *CBF*, *DOF*, *GAMYB* и *NAC*), поочередно становятся мишенями миРНК, чтобы контролировать переход между фазами закладки и хранения эндосперма (Curaba et al., 2012).

Все гены HMW-GS-белков *x*-типа имеют Cereal-box в положении -450 п.н. (Anderson, Greene, 1989). Промоторы генов *Glu-1Bx* имеют двойной Cereal-box в положении -502 и -448. Количество Cereal-box, по-видимому, пропорционально силе промотора, причем промоторы *x*-типа (содержащие

Cereal-box) сильнее, чем у-типа (без Cereal-box), а *Bx* – самые сильные (содержащие две копии Cereal-box).

Одним из немаловажных факторов регуляции генов, а именно подавления экспрессии, является ДНК-метилирование (Zhang et al., 2018). Множество элементов взаимодействуют в становлении и поддержании метилирования ДНК растительного генома. Различные участки метилирования требуют активности различных цитозин метилтрансфераз, что способствует модификации структуры хроматина при участии ремоделирующих хроматин ферментов и факторов, модифицирующих гистоны. В конечном итоге, формирование гетерохроматина вокруг определенных участков ДНК делает недоступными соответствующие гены для процессов транскрипции. Метилирование относится к эпигенетической регуляции экспрессии генов

Предыдущие исследования показали, что большинство запасных белков, за исключением генов HMW-GS-белков, находятся под строгой эпигенетической регуляцией посредством импринтинга и метилирования ДНК (Gayacharan et al., 2013). В случае генов HMW-GS-белков недавно сообщалось об индуцирующем эффекте гипометилирования в промоторах генов HMW-GS-белков путем облегчения связывания транскрипционных факторов проламинов пшеницы *DOF* (Zhu et al., 2018). Экспрессия генов HMW-GS-белков во время развития эндосперма также активируется ацетилизацией гистонов (Guo et al. 2016). Было показано, что транскрипционный фактор *GAMYB* связывается с промоторами генов HMW-GS-белков, рекрутирует и физически взаимодействует с гистоновой ацетилтрансферазой *GCN5*. Это приводит к повышенному ацетилированию гистонов H3 по остаткам лизина 9 и 14, что индуцирует экспрессию генов HMW-GS-белков (Guo et al. 2016). Позиционирование нуклеосом и локальные структурные изменения хроматина также были предложены в качестве объяснения высокой активности промоторов гена HMW-GS *Glu1Bx7*, содержащих вставку длиной 43 п.н., что привело к повышению экспрессии гена (Geng et al., 2014).

1.5. Классификация и функции длинных некодирующих РНК

1.5.1. Классификация некодирующих РНК

Некодирующие РНК (нкРНК) — это молекулы РНК, которые не кодируют белки. В настоящее время сведения о некодирующих РНК постоянно пополняются, и их однозначная классификация отсутствует (Ma et al., 2013).

Формально среди нкРНК можно выделить длинные некодирующие РНК (днРНК), которые представляют собой разнообразный набор РНК длиной более 200 п.н. (lncRNA, lincRNA и др.), и малые некодирующие РНК (miRNA, siRNA, snoRNA и др.). К малым нкРНК относятся микроРНК, малые интерферирующие РНК, малые ядрышковые РНК и др. (Matera et al., 2007).

За последние несколько лет путем углубленного анализа генома и транскриптома было установлено, что днРНК выполняют важную роль в виде эпигенетических модуляторов и имеют размер от 200 п.н. до нескольких тысяч п.н. У многих различных видов, в том числе растений, животных, дрожжей, прокариот, а также вирусов, были найдены днРНК. Помимо этого, днРНК обладают невысокой экспрессией, что зачастую характеризуется как транскрипционный шум (Chekanova et al., 2015).

Различают несколько классов днРНК, исходя из их положения по отношению к другим генам:

- 1) длинные некодирующие межгенные РНК расположены в рамках геномного интервала между двумя генами и не перекрывают экзоны, кодирующие белок;
- 2) естественные антисмысловые днРНК пересекают любой экзон локуса, кодирующего белок на противоположной цепи, и транскрибируются из антисмысловой цепи;
- 3) интронные днРНК, расположенные в интронах, не имеют перекрытия экзон-экзон;

4) смысловые днРНК транскрибируются из смысловой цепи генов, кодирующих белок и перекрывающихся транскриптов, содержат кодирующий ген в интроне той же цепи (Ma et al., 2013).

1.5.2. Значение днРНК в индивидуальном развитии растений

ДнРНК осуществляют важную роль в большом диапазоне биологических процессов, включая целостность клеточной структуры, клеточный цикл, импринтинг, апоптоз, транскрипцию, трансляцию, сплайсинг и локализацию белка, а также реакции на тепловой шок и перепрограммирование (Ma et al., 2013).

У растений хорошо изучен вклад днРНК в такие важные процессы, как цветение, реакция на стресс, фертильность, фосфатный гомеостаз и фотоморфогенез (Liu et al., 2017; Yuan et al., 2018). Кроме того, было показано, что днРНК:

- 1) участвуют в этапах созревания (процессинге) в более короткие нкРНК;
- 2) являются мишенями для миРНК работают как “губки”;
- 3) снижают активность гистон-модифицирующих ферментов и прямое эпигенетическое молчание через взаимодействие со специфическими хроматиновыми доменами;

4) через модификации белка и белок-белковые взаимодействия регулируют посттрансляционные процессы (Sheldon et al., 2000; Campalans et al., 2004; Dey et al., 2004; Pant et al., 2008; Amor et al., 2009; Neo et al., 2011; Zhu et al., 2012; Marquardt et al., 2014; Liu et al., 2015).

Пример регуляции опосредованной днРНК был найден у *A. thaliana* в процессе перехода от вегетативной к генеративной стадии. Ген *FLC* (англ. Flowering Locus C - Лocus цветения C) выступает в качестве регулятора времени цветения у арабидопсиса, который подавляет индукцию цветения (Csorba et al., 2014). К гену *FLC* была найдена антисмысловая днРНК, называемая COOLAIR, которая транскрибируется в начале фазы яровизации

(Shafiq et al., 2016). Исходя из гомологии был выполнен поиск локуса *FLC* и антисмысловых транскриптов *FLC* у однодольных. Полученные результаты исследования выявили наличие данных транскриптов у шести видов злаков, включая *T. aestivum* (Jiao et al., 2019). Кроме того, обнаружена днРНК, называемая COLDAIR, которая потенциально регулирует экспрессию *FLC*. В дополнение к вышесказанному, она транскрибируется в ответ на холод, но, в отличие от COOLAIR, COLDAIR направлена на смысловую цепь *FLC* (Нео, Sung, 2011). COOLAIR и COLDAIR являются сигналами, определяющими переход растения на новую стадию его развития. На данный момент их точные функции и взаимодействия у других однодольных мало изучены и требуют дальнейших исследований.

Другой пример влияния днРНК в регулировании пути развития представляет собой длинную некодирующую межгенную РНК, которая характерна для длинного дня и связана с мужской фертильностью (LDMAR). Пример такого регулирования продемонстрировали на рисе: если экспрессия LDMAR была меньше конкретного уровня, то детектировались изменения в развитии пыльцы. Мутации, которые вызывают снижение экспрессии LDMAR, приводят к фотопериодочувствительной мужской стерильности у растений, выращенных в условиях длинного дня (Ding et al., 2012; Zhang et al., 2013). Механизм, с помощью которого LDMAR регулирует развитие пыльцы, и факт его проявления у других однодольных растений к настоящему времени неизвестен.

Недавнее исследование определило новую днРНК, обозначаемую как WSGAR (Wheat Seed Germination Associated RNA), которая в свою очередь регулирует прорастание зерновки пшеницы (Guo et al., 2018). Процесс регуляции начинается со специфичной для пшеницы миРНК (miR9678), нацеленной на WSGAR, которая переходит в phasiRNA и препятствует прорастанию зерновки. В другом исследовании обнаружили 177 днРНК, которые отвечали на вещество, блокирующее каналы Ca^{+2} в корнях пшеницы. В ходе работы наблюдалось уменьшение длины корней, а в

дальнейшем и вовсе полное прекращение роста вследствие увеличения количества реагента. В результате было выдвинуто предположение, что все обнаруженные днРНК связаны с ростом корней пшеницы (Ma et al., 2018).

Ученые из США в 2018 году предсказали гены днРНК в *T. turgidum* ssp. *dicoccoides* и провели исследование потенциальных сетей днРНК-миРНК-мРНК. В ходе этого исследования были выявлены 89623 днРНК, из которых 23713 были определены как предполагаемые мишени миРНК (Акпинар et al., 2018). Уже в другой работе у двух сортов дикой пшеницы, Kiziltan и TR39477, и одного сорта твердой пшеницы *T. turgidum* ssp. *durum* (AABB), одомашненного подвида *T. turgidum*, идентифицировали днРНК, 63773, 61823 и 43932 днРНК в Kiziltan, TR39477 и твердой пшенице, соответственно. В результате данного исследования было показано, что 3% идентифицированных днРНК у сорта Kiziltan, 6% идентифицированных днРНК TR39477 и 4% днРНК сорта твердой пшеницы в ответ на засуху различно экспрессировались и проявляли себя как «зависимые от засухи» днРНК, вдобавок ко всему большая часть из них экспрессировалась только в условиях засухи (Cagirici et al., 2017).

Хотя секвенирование следующего поколения дало представление о геномах и транскриптомах многих видов, пока полученные результаты полностью не удалось конкретизировать и выявить клеточные пути, ответственные за устойчивость к стрессу и регуляторные молекулы.

1.6. Секвенирование и аннотация геномов пшеницы, ржи и тритикале

Геном тритикале представляет собой совокупность субгеномов пшеницы и ржи, а, следовательно, вопрос о проявлении генов и локусов количественных признаков, принадлежащих отдельным субгеномам в совокупном геноме, остается не изученным в полном объеме. В случае выявления сходной степени влияния локусов, идентифицированных у субгеномов, открываются широкие перспективы применения уже разработанных молекулярных маркеров непосредственно на тритикале с

минимальными затратами труда и времени на их адаптацию для этой культуры.

Технология PacBio, которая генерирует длинные риды (в среднем 10000 п.н.), доказала свою эффективность в секвенировании таких сложных геномных областей. Геном мягкой пшеницы сорта Chinese Spring был секвенирован и собран с использованием технологий секвенирования Illumina и PacBio. Окончательная сборка содержала последовательности длиной 15 Гб с N50 232 Кб (Huo N. et al., 2018). Тем не менее, эти крупные контиги не были прикреплены к хромосомам, чтобы можно было построить их модели. Данные, полученные на *Ae. fauschita*, использовались для поиска с помощью программы BLAST контигов для выявления их перекрытия.

Отличительной чертой этой сборки является совокупный эффект закрепления 279 неразмещенных скаффолдов (74960 Мб), благодаря чему увеличилось количество неповрежденных LTR-ретротранспозонов на 13%, были выявлены ранее не идентифицированные TE, удалось заполнить пробелы в целевых последовательностях и увеличить длину псевдомолекул приблизительно на 168 Мб (1,19%) по сравнению с предыдущими сборками. Расположение генов в псевдомолекулах IWGSC RefSeq v2.1 более точно отражает дивергенцию субгеномов пшеницы от общего генома предков в том смысле, что в псевдомолекулах IWGSC RefSeq v2.1 было обнаружено большее количество коллинеарных генов и уменьшенная фрагментация синтенных блоков.

Недавно, благодаря совместному использованию разных технологий секвенирования, упорядочивания и алгоритмов сборки генома, были опубликованы последовательности генома гексаплоидной пшеницы, наиболее приближенные к реальному (Zimin et al., 2017, IWGSC et al., 2018). Кроме этого, удалось заново собрать геномы разных сортов и изучить структурные вариации (Walkowiak, 2020).

Несмотря на филогенетическое родство с мягкой пшеницей (*Triticum aestivum*), рожь обладает уникальными характеристиками как по

агрономическим показателям, так и по особенностям организации генома, а также является важным генетическим ресурсом для улучшения пшеницы и тритикале.

Помимо своей важности как культуры, рожь является одним из родителей искусственно полученного вида - тритикале. Стоит отметить, что интрогрессии генома ржи широко присутствуют в современных сортах пшеницы. Так короткое плечо хромосомы 1 ржи (1RS) было интрогрессировано в несколько сотен сортов пшеницы. Фактически, некоторые из наиболее успешных сортов пшеницы несут транслокацию 1BL.1RS, поскольку присутствие 1RS в геноме пшеницы увеличивает как урожайность, так и содержание белка в зерне.

Геном ржи существенно больше, чем у диплоидных видов пшеницы, и, по данным NCBI, составляет 7,7 Гб, из которых мобильные элементы (TEs) занимают примерно 90% генома.

В 2021 году Guangwei Li с соавторами смогли получить высококачественную сборку генома сорта ржи Weining путем объединения ранее полученного набора данных, сгенерированных путем секвенирования с помощью длинных ридов PacBio RS II и коротких ридов Illumina, захвата конформации хроматина (Hi-C), генетического картирования и анализа BioNano. Собранная последовательность генома в общей сложности составила 7,74 Гб, из которых 7,25 Гб было закреплено на семи псевдохромосомах (1R-7R), что составляет 93,67% собранной последовательности генома. Размер собранных хромосом составил от 0,94 Гб и 0,99 Гб (1R и 5R, соответственно) до 1,14 Гб (2R). Анализ сборки выявил около 90% последовательностей, идентифицированных как TEs, из которых LTR-RTS включали элементы Gypsy, Copia и неклассифицированные ретротранспозоны. Кроссгеномные сравнения пшеницы, эгилопса, ржи и ячменя выявили, что LTR-RTS, особенно элементы Gypsy, внесли наибольший вклад в расширение генома ржи сорта Weining.

Применяя генотипирование путем секвенирования, Сао (2022) генотипировал 199 сортов *Triticosecale* spp, где картирование выявило удаление и добавление целых хромосом или хромосомных сегментов. Анализ полиморфизма выявил 434304 надежных SNP-маркера у 153 гексаплоидных сортов тритикале. На основании полученных данных было выявлено 205 генов, связанных с 94 значимыми маркерами, на основе геномов пшеницы и ржи. Некоторые аннотированные функциональные гены были соотнесены с соответствующими признаками.

1.7. Секвенирование целевых последовательностей

В настоящее время эффективные генетические исследования растений уже невозможно представить без этапа расшифровки генома (Varabaschi et al., 2016, Брагина, Афонников, Салина, 2019). Новые технологии селекции растений опираются на информацию о маркерах и ассоциированных с ними количественных признаках, при этом наличие данных о геноме позволяет использовать самый широкий спектр маркеров, включая как повторы (Bai et al., 2017), так и SNPs (Crain et al., 2018; Li et al., 2018, Брагина, Афонников, Салина, 2019).

Не смотря на быстрое развитие технологий секвенирования ДНК "NextGen", для многих экспериментальных протоколов часто возникает необходимость выборочного секвенирования интересующих геномных регионов. Это особенно актуально при попытке параллельно исследовать множество образцов на наличие однонуклеотидных полиморфизмов (SNPs), вариантов числа копий (CNVs) или перестроек ДНК, или исследовать подмножество геномных регионов, а не весь геном, например, экзоны, кодирующие белки.

Высокого уровня покрытия и разрешения полногеномного секвенирования не хватает для обнаружения всех изменений в образцах с большим размером генома. В связи с этим стало развиваться и продолжает набирать обороты целевое (таргетное) секвенирование, которое заключается в

выделении и секвенировании определенной области генома или подмножества генов.

Секвенирование целевого гена (TGSeq) представляет собой набор различных подходов к секвенированию конкретных генов без применения полногеномного секвенирования (WGS), которое является дорогостоящей альтернативой. Стоит отметить, что для видов с большим размером генома, включая пшеницу, тритикеле и лук, TGSeq является оптимальным методом из-за высокой стоимости полногеномного секвенирования и сложности сборки генома.

Для TGSeq разработано несколько подходов, включая секвенирование по Сэнгеру, а также стратегии обогащения генов-мишеней с последующим секвенированием с коротким или длинным ридом.

Секвенирование генов-мишеней по Сэнгеру является методом для секвенирования коротких генов (менее 1 Кб). Прямое секвенирование по Сэнгеру продуктов ПЦР, амплифицированных из гена, может быть затруднено в случае изучения множественных аллелей полиплоидных организмов.

Для растений, которые обладают большим или полиплоидным геномом, существует стратегия уменьшения сложности генома с использованием методики целевого обогащения. Стратегии обогащения мишеней можно разделить на несколько категорий, основанных на различных подходах к захвату интересующего региона ДНК. Некоторые стратегии (называемые "стратегиями на основе гибридизации") используют денатурацию геномной ДНК и отжиг одноцепочечных олигонуклеотидов ДНК или РНК различной длины (называемых "зондами") к интересующему региону. Олигонуклеотиды зонда обычно биотинилированы, что позволяет им связываться с магнитными частицами, покрытыми стрептавидином, и таким образом захватывать сдвинутую комплементарную ДНК в комплексе с частицами. Другие стратегии ("стратегии на основе ПЦР") основаны на амплификации интересующего региона с помощью праймеров и требуют лишь ограниченного количества ДНК.

Разработан ряд коммерческих методик целевого секвенирования с короткими ридами. Они отличаются по способу обогащения библиотеки целевыми последовательностями. Это могут быть стратегии на основе гибридизации, стратегия опосредованной транспозонами фрагментации, стратегия использования молекулярных зондов (MIPs), а также однокомплексное и мультикомплексное обогащение с помощью полимеразной цепной реакции (ПЦР).

После получения коротких ридов их картируют на референсный геном. Поскольку информация о последовательности разбита на миллионы коротких фрагментов, крайне важно правильно их собрать, чтобы не потерять информацию о последовательности. Это особенно сложная задача в регионах с повторяющимися последовательностями или со структурными перестройками (Alkan, Coe, Eichler, 2011).

Чтобы справиться с основными недостатками анализа полученных данных секвенирования короткими ридами, было разработано секвенирование длинными ридами в сочетании с этапами обогащения. В настоящее время существует широкий выбор методов обогащения мишеней для технологий короткого секвенирования (Mamanova et al., 2010), но лишь немногие из них совместимы с платформами секвенирования длинными ридами.

Наиболее распространенным подходом является проведение полимеразной цепной реакции длинных фрагментов (LR-PCR) с последующим секвенированием полученных ампликонов (Ardui et al., 2017; Lode et al., 2018). Однако технически сложно амплифицировать области длиннее 10 тыс.п.н. с помощью LR-PCR, и оба конца целевого фрагмента должны быть известны для разработки праймеров для ПЦР.

Альтернативный подход заключается в проведении гибридизации (Wang et al., 2015) с использованием коротких ДНК- или РНК-зондов, которые используются для извлечения интересующей мишени. Методы, основанные на гибридизации, могут обогащать фрагменты размером до 10 тыс.п.н. на основе

информации о коротком зонде (<200 п.н.), но требуют относительно больших количеств входной ДНК (>500 нг).

И амплификация длинных фрагментов, и методы гибридизации также имеют риск получения химерных ридов во время этапов амплификации при подготовке образца или библиотеки.

Совсем недавно был предложен метод Xdrop для обогащения образцов ДНК целевыми геномными фрагментами с последующим секвенированием длинными или короткими ридами (Madsen et al., 2020). Технология Xdrop основана на выделении длинных фрагментов ДНК в миллионах капель, где капли, содержащие интересующую целевую последовательность, флуоресцентно метятся, амплифицируются и сортируются с использованием проточной цитометрии. Конечным продуктом процедуры Xdrop является обогащенный набор длинных молекул ДНК, который можно исследовать с помощью секвенирования. Xdrop – перспективная технология, но для нее требуется специальное оборудование, и она опирается на этапы амплификации ДНК, которые могут вносить некоторые искажения и артефакты.

Нанопоровое секвенирование ампликонов (ONT Amplicon-Seq) в последнее время используется для изучения микробного разнообразия (Fridman et al., 2023). Например, ONT Amplicon-Seq был использован для секвенирования 4,4 Кб оперон рибосомальной РНК из микробного сообщества. Чтобы снизить частоту ошибок при считывании ONT Amplicon-Seq (обычно она составляет около 5-25%), были использованы уникальные молекулярные идентификаторы (UMI) (Karst, 2021). Применение UMI привело к резкому повышению точности последовательности до ~ 98%.

Различные исследования продемонстрировали влияние грибковой и бактериальной колонизации на гниение древесины с использованием секвенирования ампликонов (Li et al., 2018; Diesh et al., 2023; Luo et al., 2018; Wang et al., 2013), проведенного на платформе Illumina. Напротив, в недавнем исследовании было применено секвенирование ампликонов на платформе MinION, позволяющее таксономически идентифицировать и оценивать виды

грибов и бактерий, связанные с мягкой гнилью и эрозией заболоченной древесины (Yu et al., 2022).

Эксперименты на ячмене (*Hordeum vulgare* ssp. *spontaneum*) демонстрируют применение Нанопорового секвенирования ампликонов для больших и сложных геномов злаков. Например, при исследовании влияния плазмотида на флуоресценцию хлорофилла было проведено секвенирование ампликонов на участке 19409-24572 хлоропластной ДНК ячменя из коллекции В1К (Zheng et al., 2023) с применением ДНК-баркодов. Для получения необходимых данных процедура секвенирования длилась около 24 часов на проточной ячейке Nanopore R9.4.1.

Другой подход к TGSeq, основанный на секвенировании длинными ридами, называемый CATCH (нацеливание на сегменты хромосом с помощью Cas9), был представлен Gabrieli et al. (2018). CATCH реализует Cas9-опосредованное расщепление целевой области генома с последующей очисткой и амплификацией расщепленных фрагментов и секвенированием с помощью нанопора. Технология CATCH (Bennett-Baker, Mueller, 2017) включает стадию импульсного гель-электрофореза, что делает этот метод трудоемким и требующим много времени. Тем не менее, из-за стадии амплификации перечисленные методы TGSeq не позволяют одновременно секвенировать и профилировать метилцитозинные основания, играющие ключевую роль в регуляции транскрипции генов.

Недавно разработаны альтернативные протоколы, использующие систему и не требующие какой-либо стадии амплификации ПЦР (Tsai et al., 2017; Gabrieli et al., 2018; Gießelmann et al., 2019; Gilpatrick et al., 2020; Stangl et al., 2019). Обогащение без ПЦР позволяет исследовать участки генома, которые трудно амплифицировать, например, повторы (Gießelmann et al., 2019; Noijer et al., 2018; Tsai et al., 2017). Также эти участки требуют большого количества ДНК и являются относительно трудоемкими.

Новый метод nCATS (CAS9-опосредованное нанопоровое секвенирование) используется для целевого секвенирования нативных

молекул ДНК. Без этапа амплификации ПЦР nCATS может генерировать данные, подходящие как для секвенирования, так и для профилирования метилирования ДНК целевых генов (Gilpatrick et al., 2020). Метод основан на селективном секвенировании Oxford Nanopore фрагментов ДНК, высвобождаемых после Cas9-опосредованного расщепления тотальной геномной ДНК. В отличие от CATCH, в nCATS не требуется гель-электрофорез для отделения Cas9-расщепленных фрагментов от нецелевых геномных фрагментов. Вместо этого для исключения нецелевой ДНК nCATS используется фермент щелочной фосфатазы кишечника теленка (CIP). CIP выполняет 5'-дефосфорилирование фрагментов геномной ДНК перед стадией расщепления Cas9, что делает концы нецелевых участков ДНК непригодными для лигирования адаптеров для секвенирования. В свою очередь, последующая обработка Cas9/гРНК вводит двухцепочечные разрывы, которые вместе с dA-хвостами делают концы ДНК подходящими для лигирования адаптера.

В природе для функциональной работы Cas9 требуются 2 молекулы РНК. Первая, называемая крРНК, комплементарна участку целевого гена и необходима для узнавания места разрезания последовательности ДНК. Вторая – транскрибирующая РНК (тракрРНК) – является связующим звеном между гРНК и Cas9 белком и образует частичный дуплекс с крРНК. Этот дуплекс является «направляющим» для разрезания ДНК эндонуклеазой Cas9, при этом участок ДНК, где происходит разрезание, должен содержать с 3'-конца PAM-мотив, который выполняет функцию сайта рестрикции. В условиях *in vitro* тракрРНК и крРНК объединяют в одну химерную молекулу гРНК (guide RNA). Когда гРНК находит комплементарный участок в геноме организма, Cas9 разрезает обе цепи ДНК с образованием тупых концов.

Таким образом, адаптеры в основном лигируются с фрагментами ДНК, расщепленными Cas9, что повышает вероятность секвенирования целевых последовательностей с помощью нанопорового секвенирования (Gilpatrick et al., 2020).

Профилирование метилирования целевых генов с помощью nCATS было проведено для млекопитающих (Gilpatrick et al., 2020; Giesselmann et al., 2019). Была предпринята попытка применить nCATS для секвенирования генов растений (López-Girona et al., 2020). Используя 4 гРНК, авторы секвенировали локус *MYB10* яблони (*Malus × domestica*) размером ~7,8 т.п.о. и в результате добились покрытия > 100-кратного целевого охвата и выполнили фазирование гаплотипов. Следует отметить, что геном яблони меньше (730,10 Mb/1C у *Malus × domestica* (Kron, Husband, 2009), чем у многих других агрономически важных видов, включая злаки (Poaceae), размер генома которых колеблется от 276 Mb/1C до 20825 Mb/1C (<https://cvalues.science.kew.org/>, по состоянию на 2 ноября 2021 г.). Например, размеры генома пшеницы (*Triticum aestivum*) и ржи (*Secale cereale*) составляют 16954 и 8624 Mb/1C (Bennett, Smith, 1976), соответственно. Поэтому возникает вопрос, эффективен ли nCATS для секвенирования целевого гена для видов растений с большим размером генома. Кроме того, применим ли nCATS для секвенирования нескольких генов растений (мультиплексирование) за один запуск секвенирования. Наконец, профилирование метилирования генов с использованием данных nCATS до сих пор не применялось у растений.

РАЗДЕЛ 2. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Материалы для исследования

Выделение тотальной РНК с последующим обогащением поли(А) фракции для прямого секвенирования с помощью технологии Oxford Nanopore выполняли из зерновок тритикале линии Л8665 на 10-й день после цветения (dpa) для синтеза дц-кДНК – 15 dpa и 20 dpa.

Для выделения тотальной ДНК для ПЦР с целью изучения днРНК были выбраны четырехдневные проростки 23 сортообразцов коллекции яровой тритикале зарубежной и отечественной селекции. Для секвенирования nCATS генов глютелинов была выделена геномная ДНК из листьев гексаплоидной яровой тритикале линии Л8665.

Для разработки метода одновременного секвенирования полноразмерных ампликонов глютелинов с помощью технологии The Oxford Nanopore было использовано 23 сорта яровой мягкой пшеницы. В том числе, девятнадцать образцов гибридного происхождения были получены из СИММУТ, а также четыре образца различного географического происхождения.

2.2. Методы исследования

2.2.1. Выделение ДНК

Выделение геномной ДНК проводили в вытяжном шкафу ЦТАБ - методом с модификациями (Pucker et al., 2021). Растительный материал гомогенизировали до порошкообразного состояния при помощи ступок с пестиками с добавлением жидкого азота. К полученному порошку немедленно добавляли предварительно нагретый до 75°C лизирующий буфер (2% ЦТАБ; 100 mM Трис-НСl, рН=8.0; 20 mM ЭДТА, рН=8.0; 1,4М NaCl, 0,25% PVP) в пропорции 500 мкл на 100 мкг растительного материала. Полученную смесь переносили в 1,5 мл пробирки с добавлением 30 мкл β-меркаптоэтанола, далее

инкубировали при 75°C в течение 30 мин., периодически переворачивая пробирки. После инкубации к образцам добавляли равный объем смеси (500 мкл) хлороформа: изоамиловый спирт (24:1) для депротенизации. После центрифугирования при комнатной температуре в течение 30 мин. при 13400 g переносили водную фазу в новые 1,5 мл пробирки. Преципитацию комплекса ЦТАБ-ДНК осуществляли добавлением 2 объёмов осаждающего буфера (1% ЦТАБ; 50 mM Трис-НСl, pH=8.0; 10 mM ЭДТА, pH=8.0; 0,125% PVP) с последующим центрифугированием при комнатной температуре в течение 30 мин. при 13400 g. После удаления надосадочной жидкости полученный осадок растворяли в 200 мкл 1M NaCl. Преципитацию ДНК проводили с добавлением 200 мкл изопропанола и последующим центрифугированием при комнатной температуре в течение 30 мин. при 13400 g. Отмывку полученного осадка и стенок пробирок производили добавлением к осадку 500 мкл 75% этилового спирта и центрифугированием при комнатной температуре в течение 10 мин. при 13400 g. После удаления спирта для испарения его остатков пробирки выдерживали открытыми в течение 2 мин. Осадок растворяли в 100 мкл буфера для элюции (10 mM Трис-НСl, pH=8.0; 0,1 mM ЭДТА, pH=8.0) с добавлением 1 мкл РНКазы А (10 ед/мкл) к образцу. Полного растворения осадка и элиминации контаминирующей РНК достигали инкубацией в течение 1 часа при 37°C. Далее проводили очистку от РНКазы добавлением равного объёма (100 мкл) смеси хлороформа: изоамиловый спирт и последующего центрифугирования при комнатной температуре в течение 15 мин. при 13400 g. Водную фазу переносили в новые 1,5 мл пробирки. Преципитацию ДНК проводили добавлением 10 мкл 3M ацетата натрия и 100 мкл изопропанола при последующем центрифугировании при комнатной температуре в течение 15 мин. при 10000 g. Отмывку осадка и стенок пробирок производили добавлением 500 мкл 80% этилового спирта с дальнейшим центрифугированием в течение 5 мин. при 10000 g. ДНК растворяли в 50 мкл буфера для элюции.

Оценку качества выделенной геномной ДНК проводили путём электрофореза в 1% агарозном геле при напряжении 3 В/см.

2.2.2. Обогащение геномной ДНК высокомолекулярными фрагментами

Обогащение геномной ДНК высокомолекулярными фрагментами проводили с помощью набора Short Read Eliminator XL Kit (Circulomics, Baltimore, MD, США). Образец геномной ДНК был доведен до общего объема 60 мкл и концентрации 150 нг/мкл в пробирках 1,5 мкл DNA LoBind Tube. Далее к образцу добавляли 60 мкл буфера SRE XL и центрифугировали при 10000 g в течение 30 мин. при комнатной температуре. После удаления надосадочной жидкости добавляли 200 мкл 70% этилового спирта с дальнейшим центрифугированием при 10000 g в течение 2 мин. при комнатной температуре. К полученному осадку добавляли 50 мкл буфера EB в пробирку и инкубировали при комнатной температуре в течение 20 мин. Для окончательного ресуспендирования осадка полученный образец аккуратно встряхивали вручную.

Концентрацию и качество выделенной ДНК оценивали с помощью спектрофотометра NanoDrop One UV-Vis (Thermo Fisher Scientific, Waltham, MA, USA) и Quantus Fluorometer (Promega, Madison, WA, USA) с использованием системы ДНК QuantiFluor ONE dsDNA (Promega, Madison, Washington, USA).

2.2.3. Дизайн гидовой РНК

гРНК были подобраны на ВАС-последовательности трех генов глютенинов: DQ537335.1 (идентификационный номер в NCBI) для *Glu-1Ax* и DQ537336.1 для *Glu-1Bx* и *Glu-1By*. Для дизайна гРНК использовали программы CRISPRdirect (Naito et al., 2014) и FlashFry (McKenna, Shendure, 2018) (<https://crispr.dbcls.jp/>, accessed on 3 September 2021) с инструментом «discover». Были найдены все позиции гРНК, подобранные на геном. Затем эти

связывающегося с белковой последовательностью Cas9. Все олигонуклеотиды синтезировали в Еврогене (ЗАО Евроген, Москва, Россия).

Для синтеза ДНК-матриц для каждой реакции смешивали следующие реагенты: 2 мкл уникального олигонуклеотида гРНК (1 мкМ); 2 мкл универсального CRISPR_R (1 мкМ); 2 мкл прямого Т7 и 2 мкл обратного Т7 праймеров (по 100 мкМ каждого); 2 мкл 50x dNTP (10 мкМ); 10 мкл 10x буфера Encyclo (ЗАО Евроген, Москва, Россия); 1 мкл ДНК-полимеразы Encyclo (ЗАО Евроген, Москва, Россия); 79 мкл воды, свободной от РНКаз. Олигонуклеотиды амплифицировали в соответствии с программой ПЦР: 98°C - 2 мин, 98°C - 30 сек, 60°C - 30 сек, 72°C - 30 сек, 30 циклов, 72°C - 1 мин.

ПЦР-продукты очищали на колонках набором Cleanup Standard (ЗАО Евроген, Москва, Россия).

Затем проводили *in vitro* транскрипцию при 37°C в течение 2 ч с использованием набора для высокоэффективного синтеза РНК *in vitro* (Биолабмикс, Новосибирск, Россия) согласно протоколу производителя. Продукты *in vitro* транскрипции очищали на колонках набором для выделения РНК с использованием связывающего раствора "ЛИРА", Биолабмикс.

Концентрацию и качество приготовленных гРНК оценивали с помощью Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA), Qubit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) и гель-электрофореза в 2% агарозном геле.

2.2.5. Подготовка библиотеки nCATS

Подготовку библиотеки nCATS для нанопорового секвенирования проводили с использованием протокола SQK-LSK109 (Oxford Nanopore Technologies, Oxford, United Kingdom). Схематично секвенирование nCATS представлено на рисунке 3.

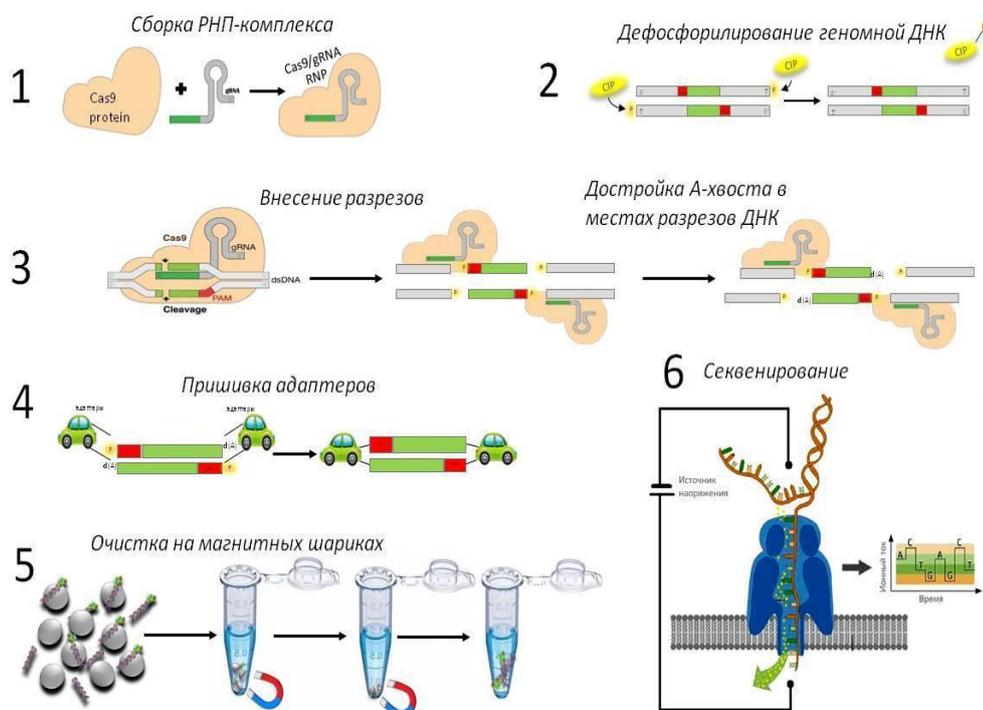


Рисунок 3 – Схема nCATS секвенирования

После синтеза гРНК осуществлялась сборка рибонуклеопротеинового (РНП) комплекса, представляющего собой комплекс РНК-связывающего белка Cas9 с гРНК. Рибонуклеопротеиновый комплекс Cas9/гРНК распознает участки ДНК, идентичные 20 п.н. на 5'-конце гРНК, и нуклеаза Cas9 разрезает обе цепи ДНК с образованием тупых концов.

Перед сборкой РНП комплекса проводили денатурацию гРНК для разрушения вторичных структур. Для этого индивидуальные гРНК (200 нг) объединяли в молярном соотношении 1:1 в объеме 11 мкл воды свободной от РНКаз, далее инкубировали полученные смеси при 95°C в течение 3 мин. и переносили на лёд.

Для сборки РНП-комплекса к 11 мкл смеси гРНК добавляли 1 мкл (20 пмоль) EnGen Spy dCas9 (New England Biolabs, United Kingdom) и 3 мкл (5x) NEB-буфера (New England Biolabs, United Kingdom), полученную смесь инкубировали в течение 30 мин. при комнатной температуре.

В секвенировании nCATS используется щелочная фосфатаза (CIP), которая дефосфорилирует 5'-концы нецелевой ДНК, делая их непригодными для лигирования адаптеров. Для дефосфорилирования к высокомолекулярной

геномной ДНК (3 мкг) добавляли 4 мкл буфера CutSmart (10x), 8 мкл Quick CIP и инкубировали 45 мин. при 37°C. Инактивацию фермента проводили путём нагревания смеси до 80°C в течение 2 мин.

Затем к дефосфорилированной ДНК добавляли 15 мкл предварительно собранного РНП-комплекса, 1,5 мкл dATP (10 мкМ) и 1 мкл Taq-полимеразы. Смесь инкубировали при 37°C в течение 30 мин. для внесения разрезов в ДНК, затем 5 мин при 72°C для построения полиадениновой последовательности в местах разрезов.

Адаптеры для нанопорового секвенирования (AMX) лигировали с расщепленной Cas9 ДНК путем смешивания следующих компонентов: 25 мкл буфера для лигирования LNB (набор LSK109, Oxford Nanopore Technologies, United Kingdom), 5 мкл воды свободной от нуклеаз, 12,5 мкл T4-лигазы ДНК (NEBnext Quick T4 Ligase) и 5 мкл адаптеров (AMX). Смесь добавляли к расщепленной ДНК и инкубировали в течение 20 мин. при комнатной температуре.

Очистку ДНК на магнитных шариках производили путем добавления к расщепленной ДНК TE-буфер (pH 8,0) в соотношении 1:1 и 0,6 объема шариков Ampure XP Beads (Beckman Coulter, № A93881, Brea, California, USA) и инкубировали в течение 10 мин. После удаления супернатанта из смеси ДНК с магнитными шариками смесь дважды промывали буфером для отмывки коротких фрагментов ДНК (Short Fragment Buffer) (набор SQK-LSK109, Oxford Nanopore Technologies, United Kingdom). После промывки образец элюировали в 14 мкл элюирующего буфера (набор SQK-LSK109, Oxford Nanopore Technologies, United Kingdom).

Библиотека для секвенирования была подготовлена путем смешивания 12 мкл элюата, 37,5 мкл буфера для секвенирования и 25,5 мкл загрузочных шариков (набор LSK109, Oxford Nanopore Technologies, United Kingdom).

2.2.6. Нанопоровое секвенирование

Секвенирование проводили на проточной ячейке MinION R9.4.1. Процесс секвенирования осуществляли с помощью программного обеспечения MinKNOW (версия 19.2.2 – Oxford Nanopore Technologies, United Kingdom).

2.2.7. Анализ данных, полученных с помощью секвенирования

Полученные риды картировали на полноразмерные последовательности ВАС с помощью программы minimap2 (Li, 2018). Для обнаружения SNPs использовали программу Nanocaller (Ahsan et al., 2021). Для реконструкции последовательностей аллелей гена *Glu-1B* применяли bcftools (Danecek et al., 2017). Ранжирование ридов по гаплотипам было выполнено с помощью WhatsHap (Patterson et al., 2015).

Идентификация метилирования с использованием необработанных данных ридов с помощью нанопора была выполнена с помощью программного обеспечения DeepSignal-plant (Ni et al., 2021).

2.2.8. ПЦР-подтверждение вставок в гене *Glu-1Bx*

Для проверки 2 InDels (180 п.н. и 12 п.н.), расположенных в гене *Glu-1Bx* и идентифицированных с помощью nCATS, были подобраны праймеры на эти вставки. ПЦР проводили с ДНК-полимеразой Encyclo (ЗАО Евроген, Москва, Россия) согласно инструкции производителя. Программа амплификации: 94°C - 1 мин, 94°C - 1 мин, 60°C - 1 мин, 72°C - 1 мин, 35 циклов, 72°C - 3 мин.

2.2.9. ПЦР-амплификация последовательностей генов высокомолекулярных глютеинов

Подобрали пары праймеров, предназначенные для амплификации промоторной и кодирующей областей. Длины ампликонов составляли от

2600 п.н. (*Glu-1Ax*) до 5000 п.н. (*Glu-1Dy*). Пары праймеров использовали для выявления наиболее длинной консервативной части каждого гена. Все праймеры были разработаны с помощью программного обеспечения Primer 3.0 (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) и проверены PrimerBLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>).

Для оптимизации условий ПЦР использовали смесь ферментов ПЦР, сочетающих корректирующие и высокопроцессивные типы. Для успешной амплификации *Glu-1Bx* и *Glu-1By* была проведена ПЦР с использованием ДНК-полимеразы Encyclo (Evrogen, Москва, Россия) и буфера encyclo, не содержащего магния. Концентрации $MgCl_2$ и ДМСО являются частью оптимизации ПЦР для достижения желаемых результатов. 1 мМ $MgCl_2$ влияет на активность полимеразы, взаимодействия праймер-матрица, общую специфичность и эффективность амплификации ДНК. 3% ДМСО используется в ПЦР для предотвращения образования вторичных структур либо в матрице ДНК, либо в ДНК-праймерах. Мы провели амплификацию четырех длинных фрагментов (*Glu-1Ax*, *Glu-1Ay*, *Glu-1Dx*, *Glu-1Dy*) с использованием полимеразы Biolabmix LR HS и 5% ДМСО в соответствии с инструкциями производителя (Biolabmix, Новосибирск, Россия). Визуализацию полученных результатов ПЦР проводили с помощью гель-электрофореза с использованием 1%-ного агарозного геля с окрашиванием бромистым этидием.

После амплификации полноразмерные продукты *Glu-1Bx* размером 4 Кб были очищены из 1% агарозного геля с помощью CleanUp Standart (Evrogen, Москва, Россия) в соответствии с инструкциями производителя. ПЦР-продукты других субъединиц глютелинов были уравнены по концентрации и объединены в один образец в соответствии с сортами. Объединенные ампликоны очищали с помощью 1,8x магнитных частиц Agencourt AMPure XP (Beckman Coulter, Pasadena, California, USA) в соответствии с инструкциями производителя. Очищенные ампликоны *Glu-1Bx* и объединенные ампликоны выравнивали по концентрации и объединяли

в один конечный образец в соответствии с сортами. Концентрацию и целостность объединенных ампликонов оценивали с помощью Nanodrop (Nanodrop Technologies, Wilmington, California, USA) и Qubit (наборы для анализа dsDNA BR от Qubit, ThermoFisher Scientific, Waltham, Massachusetts, USA) и проверяли с помощью гель-электрофореза.

2.2.10. Валидация делеции в гене *Glu-1Bx7*

Чтобы подтвердить делецию в промоторной области *Glu-1Bx7* (приблизительно 60 п.н.), идентифицированную с помощью ONT Amplicon-seq, были разработаны праймеры, фланкирующие структурную вариацию. ПЦР проводилась с использованием ДНК-полимеразы Encyclo (Evrogen, Москва, Россия) в соответствии с инструкциями производителя. Программа амплификации: 95°C - 5 мин., 95°C - 30 сек., 59°C - 30 сек., 72°C - 35 сек., 35 циклов, 72°C - 3 мин.

2.2.11. Нанопоровое секвенирование ампликонов и анализ полученных данных

Полученные ампликоны были пулированы и баркодированы с последующим секвенированием на проточной ячейке SQK-LSK109. Бейсколлинг был выполнен с помощью Guppy v6.3.8 (Oxford Nanopore Technologies, Oxford, United Kingdom). Полученные риды после ONT Amplicon-seq были выровнены на референсные последовательности генов HMW-GS из NCBI (*Glu-1A*: AF145590.1, KJ531446.1, M22208.2, MF568383.1, EU984510.1; *Glu-1B*: DQ119142.1, FM955452.1, MH108092.1, KC254854.1, JN255519.1, X61026.1, EU137874.1, EF540765.1, KF430649.1; *Glu-1D*: VK006460.1, AV485591.1, X12929.2, JF736016.1) с использованием minimap2 (Li, 2018). Полученные BAM-файлы были визуализированы с помощью JBrowse2.

2.2.12. Выделение тотальной РНК

Для выделения суммарной РНК 100 мг свежесобранного растительного материала измельчали в жидком азоте при помощи пестика и ступки до состояния порошка. Далее добавляли реагент ExtractRNA (ЗАО Евроген, Россия, Москва) и продолжали выделение в соответствии с рекомендациями производителя.

Количество и качество тотальной РНК оценивали с помощью Nanodrop One (Thermo Fisher Scientific, Waltham, Massachusetts, USA) и гель-электрофореза с использованием 1,2% агарозного геля с окрашиванием бромистым этидием. Для устранения примеси ДНК в образцах использовали набор реактивов RNase-Free DNase Set (QIAGEN, Japan).

2.2.13. Выделение поли(А)-РНК

Поли(А) фракция мРНК была очищена из 100 мкг общей РНК с помощью набора Poly(A)Purist Mag Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) в соответствии с инструкциями производителя. Конечную концентрацию поли(А)-РНК измеряли с помощью флуорометра Quantus (Promega Corporation, Madison, WI, USA), Nanodrop One (Thermo Fisher Scientific, Waltham, Massachusetts, USA) и проверяли методом гель-электрофореза (приложение, рисунок 1).

2.2.14. Подготовка библиотеки для Direct RNA-Seq Nanopore

Для секвенирования была подготовлена библиотека из 1 мкг поли(А)-РНК с использованием набора Direct RNA Sequencing kit (SQK-RNA002) (Oxford Nanopore Technologies, United Kingdom). Прямое секвенирование РНК (Direct RNA Sequencing) проводили с помощью MinION и ячейки FLO-MIN106. Бейсколлинг был выполнен с помощью Guppy v4.0.11 (Oxford Nanopore Technologies, Oxford, United Kingdom).

Для сборки транскриптов последовательности геномов А и В *Triticum aestivum* объединялись с последовательностями генома *Secale cereale* в единый fasta-файл. Риды были сопоставлены с полученным файлом fasta с помощью программного обеспечения minimap2 (Li, 2018). Отсортированные файлы bam и gff использовались для визуализации ридов с помощью JBrowse2.

2.2.15. Проведение ПЦР и электрофореза

Благодаря данным биоинформатического анализа с использованием программы Primer3Plus (<https://www.primer3plus.com/>) были подобраны специфичные праймеры для амплификации 7 днРНК. Качество праймеров проверяли с помощью Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). В качестве внутреннего положительного контроля ПЦР использовали пару праймеров на референсный ген актина. В качестве референсной ДНК-матрицы использовали ДНК, выделенную из проростка пшеницы сорта Chinese Spring.

ПЦР проводили в реакционном объеме 25 мкл в ДНК-амплификаторе «T100 Thermal Cycler» (производство комп. Bio-Rad Laboratories Inc., USA). Каждая ПЦР смесь содержала 1 мкл ДНК, 12,5 мкл реакционной смеси, 10,3 мкл H₂O, 0,4 мкл каждого праймера, 0,4 мкл Taq ДНК-полимеразы (ЗАО Евроген, Россия, Москва). Программа амплификации: 95°C - 3 мин, 95°C - 30 сек, 60°C - 30 сек, 72°C - 20 сек, 31 цикл, 72°C - 3 мин.

Разделение молекул нуклеиновых кислот производили в 2% агарозном геле при напряжении 4 В/см.

2.2.16. Синтез дц-кДНК

Для синтеза двуцепочечной кДНК первоначально выделяли тотальную РНК из зерновок с помощью ExtractRNA (ЗАО Евроген, Россия, Москва) в соответствии с инструкцией. После производили синтез дц-кДНК с

применением набора Евроген Mint (ЗАО Евроген, Москва, Россия) в соответствии с инструкцией производителя.

Во время синтеза было подобрано оптимальное количество циклов ПЦР (22 цикла) для достижения экспоненциальной фазы амплификации. Концентрацию и целостность дц-кДНК оценивали с помощью Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA) и Qubit (Qubit dsDNA BR Assay Kits, ThermoFisher Scientific, Waltham, Massachusetts, USA) и проверяли с помощью гель-электрофореза. Синтезированную дц-кДНК очищали с помощью 1,8× шариков Agencourt AMPure XP размером (Beckman Coulter, Pasadena, CA, USA) в соответствии с инструкциями производителя.

2.2.17. Биоинформатический анализ данных секвенирования дц-кДНК

Перед анализом полученных данных объединялись последовательности генома *Triticum aestivum*, которые были загружены с сервера EnsemblPlants (https://plants.ensembl.org/Triticum_aestivum/Info/Index) с последовательностями генома *Secale cereale* в один fasta-файл. Риды, полученные в ходе секвенирования, в соответствии с каждым баркодом были картированы на fasta-файл объединенных геномов пшеницы и ржи с использованием minimap2 с аргументом "-ax splice" (Li, 2018). Мы добавили headers к сопоставленным ридам, и поскольку они были в формате sam, все файлы были преобразованы из sam в bam с помощью SAMtools view, отсортированы с помощью SAMtools sort и отфильтрованы по показателю качества Qscore > 40 с параметром -q 40. Объединение геномных файлов gft было произведено с использованием !cat и с преобразованиями строк в начале этих файлов со следующими аргументами: -n, sed. Сборка и оценка количества расшифровок на основе количества выровненных ридов были выполнены с использованием программного обеспечения StringTie2 (Kovaka et al., 2019), с помощью которого были получены gtf-файлы для каждого баркода и объединены в один. Из отсортированных и отфильтрованных

abundance-файлов были созданы таблицы, которые в последующем были объединены в одну.

После преобразования файлов `gtf` в формат `bed` с использованием `gtf2bed` был применен `bedtool intersect` для идентификации неаннотированных генов с использованием параметра `-v` и аннотированных генов с использованием параметра `-wa`.

Эти данные были отфильтрованы по уровню экспрессии ($TPM > 5$), далее мы смогли получить количество аннотированных и неаннотированных транскриптов.

ГЛАВА 3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

3.1. Cas9-опосредованное нанопоровое секвенирование (nCATS) генов высокомолекулярных глютелинов

Гены глютелинов более чем на 80% состоят из коротких и разнородных повторяющихся мотивов, что делает их секвенирования короткими рядами крайне проблематичным. Поэтому для секвенирования нами был апробирован метод nCATS (Gilpatrick et al., 2020). Для применения данного подхода, были выбраны последовательности генов глютелинов *Glu-1Ax* (DQ537335.1), *Glu-1Bx* (AY367771.1) и *Glu-1By* (DQ537336.1). Для nCATS были подобраны 12 гРНК на промоторную и кодирующую области генов. Сайт посадки гРНК на кодирующую область находился за пределами повторяющегося региона (Рисунок 4). Ожидаемая длина последовательности для целевых областей составляла 3,4 тыс.п.н., 5,1 тыс.п.н. и 3,6 тыс.п.н. для *Glu-1Ax*, *Glu-1Bx* и *Glu-1By*, соответственно.

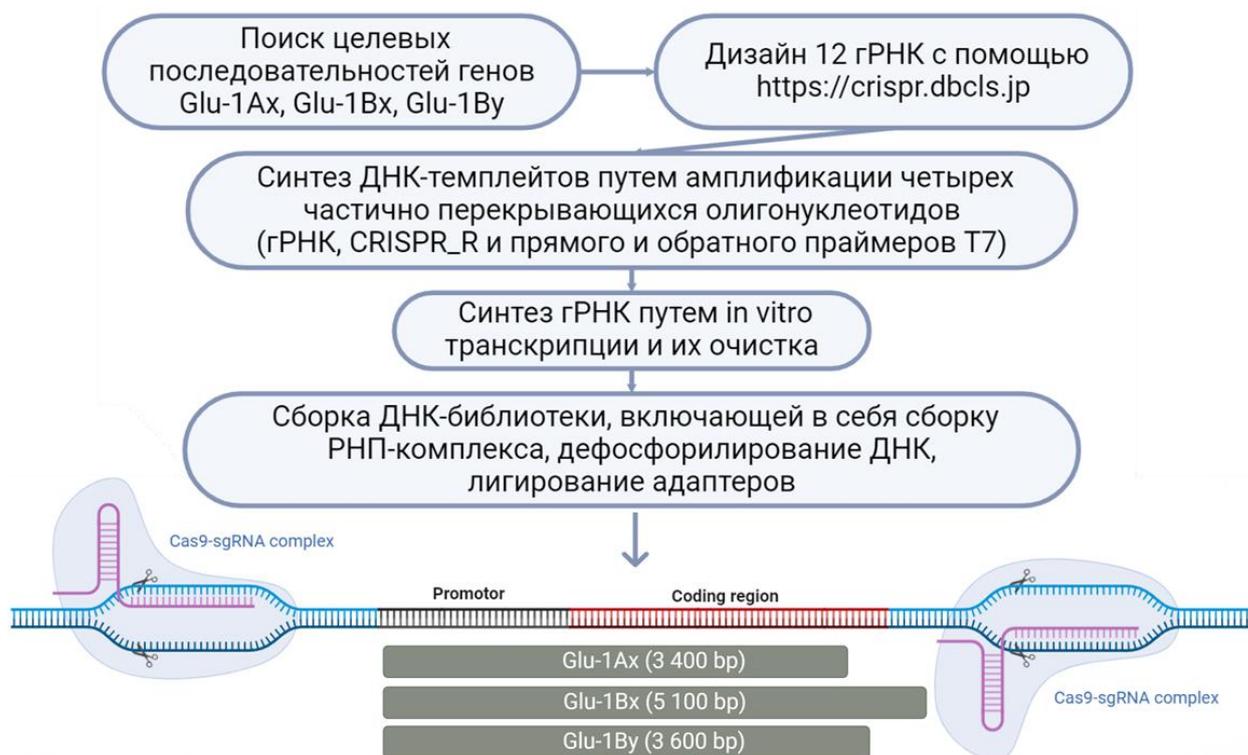


Рисунок 4 – Локализация сайтов узнавания гРНК относительно исследуемых последовательностей генов глютелинов.

Для синтеза гРНК были получены ДНК-матрицы путем амплификации четырех частично перекрывающихся олигонуклеотидов (гРНК, CRISPR_R и прямого и обратного праймеров T7). Результаты электрофореза очищенных ДНК-матриц представлены на рисунке 5.

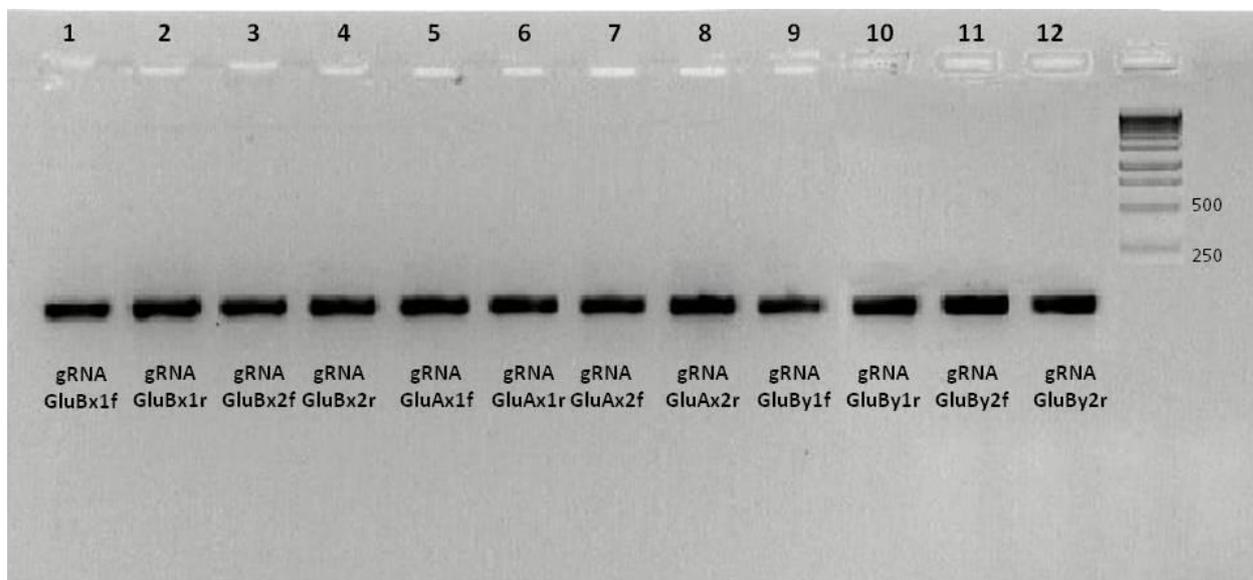


Рисунок 5. – Электрофореграмма 12 синтезированных и очищенных ДНК-матриц.

Затем была проведена *in vitro* транскрипция с использованием набора для высокоэффективного синтеза РНК *in vitro* (Биолоабмикс, Новосибирск, Россия). Полученные после очистки гРНК представлены на рисунке 6. Конформация синтезированных гРНК в 2% агарозном геле представляет собой 2 продукта размерами около 100 п.н. и 200 п.н. Несмотря на то, что на гель наносился одинаковый объем каждой гРНК, можно отметить различную интенсивность полученных фрагментов, что указывает на разную концентрацию полученных гРНК.

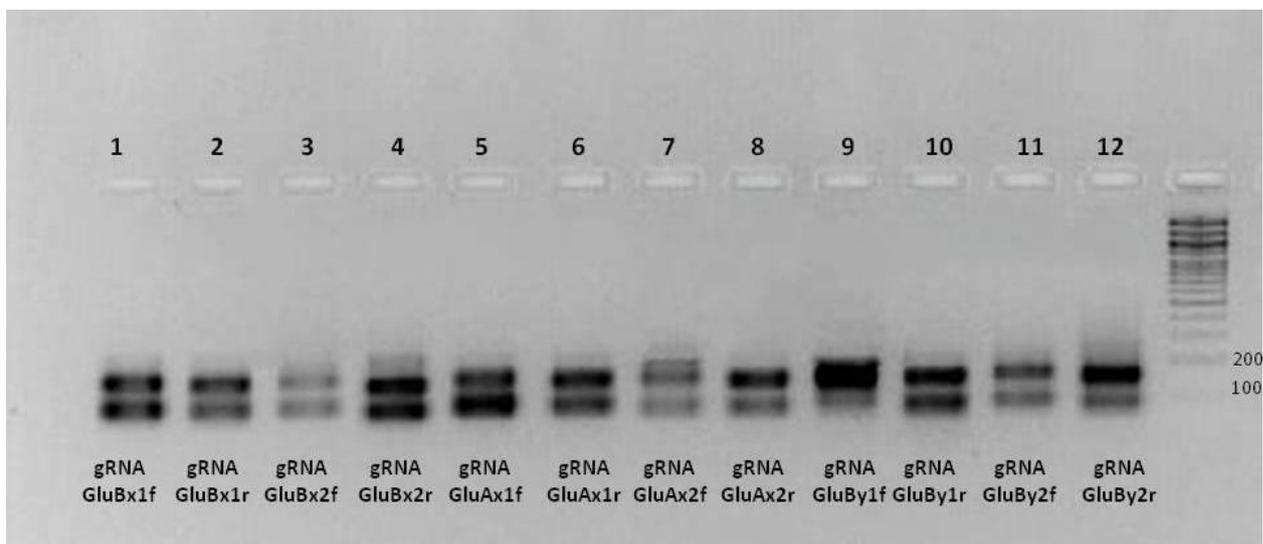


Рисунок 6 – Электрофореграмма 12 синтезированных и очищенных гРНК

Средняя концентрация гРНК составила около 400 нг/мкл (таблица 2), но наибольшие концентрации были получены для гена *Glu-1By*, что в большей степени зависит от первоначально разработанных и синтезированных олигонуклеотидов. Однако для сборки РНП-комплекса и дальнейшего секвенирования концентрации всех гРНК были приведены к единому значению – 11 нг.

Таблица 2. - Концентрация синтезированных гРНК

№ п/п	Название целевого гена	Название гРНК	Конц. РНК нг/мкл	A260/280	A260/230
1	<i>Glu-Bx</i>	gRNAGluBx1f	328,5	2,21	2,29
2		gRNAGluBx1r	386,1	2,18	2,32
3		gRNAGluBx2f	452,6	2,23	2,32
4		gRNAGluBx2r	462,6	2,30	2,57
5	<i>Glu-Ax</i>	gRNAGluAx1f	306,6	2,24	2,46
6		gRNAGluAx1r	467,5	2,22	2,49
7		gRNAGluAx2f	423,5	2,26	2,35
8		gRNAGluAx2r	287,7	2,09	2,33
9	<i>Glu-By</i>	gRNAGluBy1f	446,2	2,23	2,34
10		gRNAGluBy1r	500,1	2,33	2,44
11		gRNAGluBy2f	462,2	2,24	2,20
12		gRNAGluBy2r	334,9	2,24	2,03

3.1.1. Подготовка ДНК библиотеки и секвенирование nCATS

После синтеза всех гРНК была произведена подготовка ДНК-библиотеки, включающая в себя сборку РНП-комплекса, дефосфорилирование ДНК, ее расщепление путем добавления РНП-комплекса, лигирование адаптеров и очистку на магнитных шариках.

В ходе работы было выполнено два запуска секвенирования nCATS на секвенаторе MinION со смесью всех 12 гРНК. С помощью нанопорового секвенирования было получено 120681 высококачественных (Qscore > 8, N50 = 3,1 Кб) ридов. Было подсчитано количество целевых ридов путем поиска сходства прочитанных последовательностей с референсными генами глютеинов и обнаружено 7, 8 и 17 ридов для локусов *Glu-1Ax*, *Glu-1Bx* и *Glu-1By*, соответственно. Следовательно, ~ 0,03% полученных ридов были целевыми.

Несмотря на то, что общее количество целевых ридов было довольно низким, расчетная степень обогащения для трех целевых генов варьировалась от ~200× до ~645× (таблица 3), учитывая размер генома гексаплоидной тритикале (~17 Гб) и общую длину полученных ридов (~ 547 Кб, 0,02× покрытие генома тритикале).

Таблица 3. - Общая информация о результатах секвенирования nCATS

Локус	Количество целевых ридов	Скорость обогащения
<i>Glu-1Bx</i> DQ537336.1:199 854..204 146	8	~200×
<i>Glu-1By</i> DQ537336.1:29 265..35 054	17	~645×
<i>Glu-1Ax</i> DQ537335.1:265 694..270 243	7	~200×

Большая часть полученных ридов охватывала всю целевую область (рисунок 7), представляя полезную информацию для идентификации структурных вариаций генов глютеинов тритикале линии Л8665.

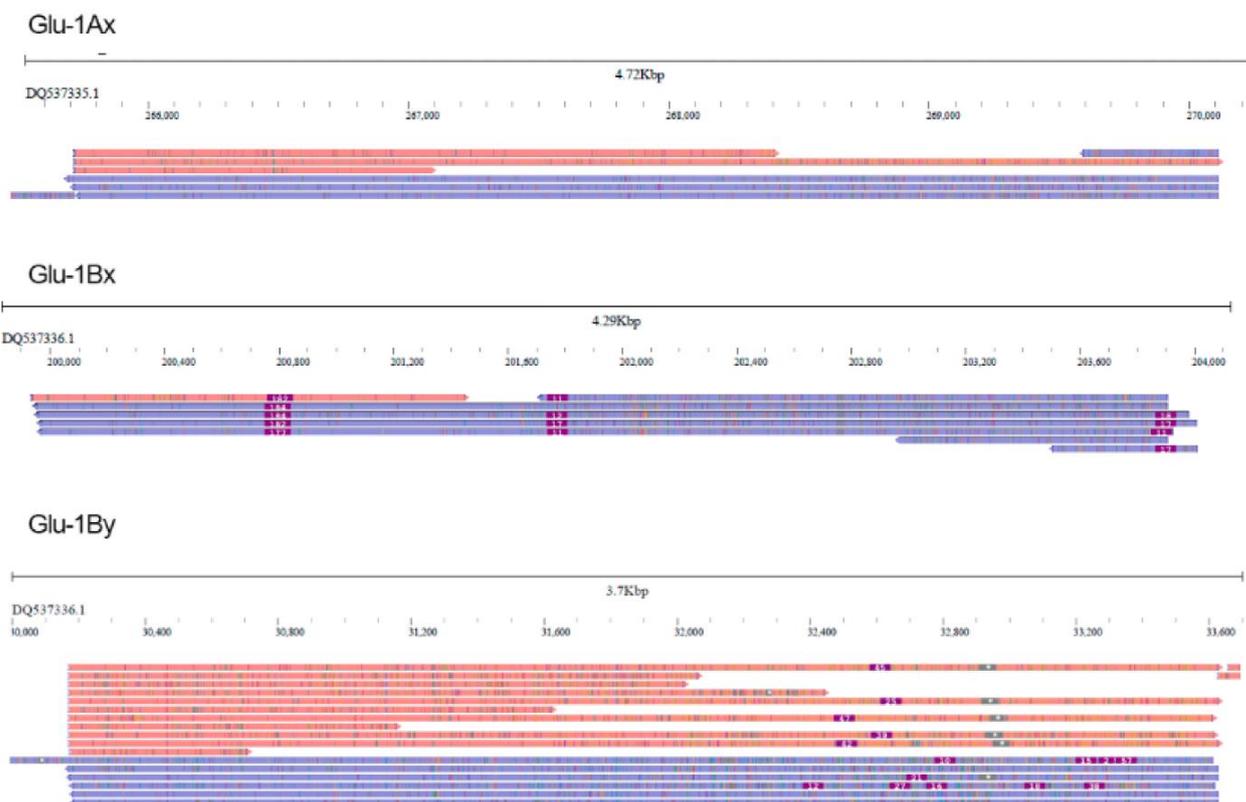


Рисунок 7 – Выравнивание ридов nCATS на референсные последовательности генов.

Удалось обнаружить две инсерции, присутствующие в гене *Glu-1Bx* линии тритикале Л8665. Одна инсерция длиной ~180 п.н. была расположена в промоторной области, а вторая инсерция размером 12 п.н. была расположена в кодирующей области *Glu-1Bx* (рисунок 7).

Таблица 4. - Праймеры, используемые для проверки InDels в последовательности *Glu-1Bx*

Название праймера	Последовательность праймеров	Длина вставки, п.н.
Glu_x_prom F	caaccatgcatagaagaagctc	180
Glu_x_prom R	ccttcttgggggtggcaga	
VxUnique1_350F	ccctgctgcaagaagttac	12
VxUnique1_350R	tggcctggatagtagaccctg	

Для подтверждения наличия инсерций в гене *Glu-1Bx* были разработаны праймеры на обнаруженные участки и проведена ПЦР с геномной ДНК тритикале линии Л8665 и пшеницы сорта Chinese Spring (таблица 4).

Результаты ПЦР и секвенирования по Сэнгеру также подтвердили наличие вставок в вариантах *Glu-1Bx* в линии Л8665 (приложение, рисунок 2).

Сравнение последовательности варианта гена *Glu-1Bx* линии Л8665 с ранее секвенированными генами глютеинов показало, что исследуемая линия тритикале Л8665 несет аллель *Glu-1Bx14*.

3.1.2. Идентификация структурных вариаций гена *Glu-1Bx*

Затем была проведена идентификация SNPs для гена *Glu-1Bx*, так как он имел наибольшую глубину секвенирования (~ 15×) (рисунок 8).

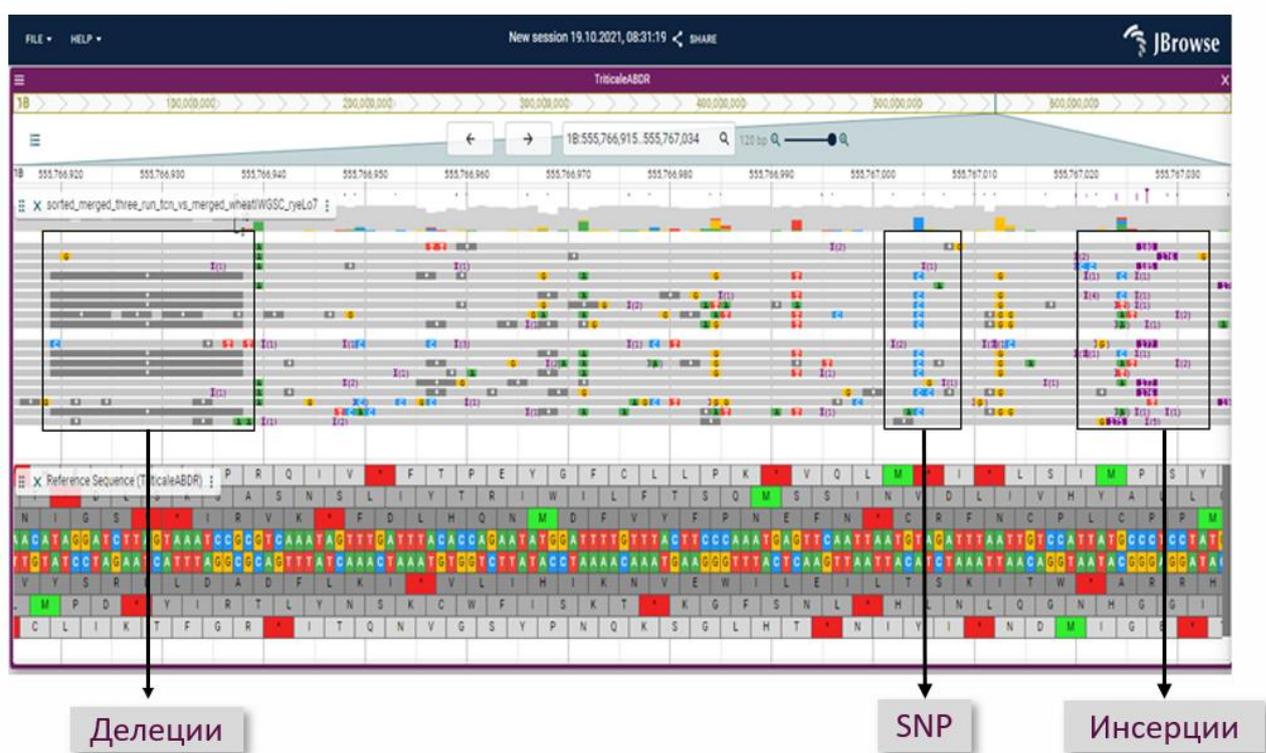


Рисунок 8 – Выравнивание ридов на референсные последовательности гена *Glu-1B* и поиск InDels и SNPs

В целом было обнаружено 222 SNPs, отличающих вариант гена *Glu-1Bx* яровой тритикале линии Л8665 от варианта этого гена у мягкой пшеницы сорта Chinese Spring. Затем был использован WhatsHap для гаплотипирования ридов на основе SNPs, обнаруженных в данных нанопорового секвенирования. Благодаря этому была создана индивидуальная последовательность гаплотипа

*Glu-1B*у и проведен поиск сходства с использованием BLAST с последующим филогенетическим анализом, ограниченным уникальной частью CDS глютеина (~ 300 п.н.). Этот анализ показал, что аллели гаплотипов HP1 и HP2 сгруппированы с известными генами *Glu-1B*у пшеницы, но они не полностью схожи с известными аллелями *Glu-1B*у, что позволяет предположить, что они могут быть новыми вариантами гена *Glu-1B*у (рисунок 9).



Рисунок 9. Гаплотипирование гена *Glu-1B*у на основе SNPs.

3.1.3. Детекция метилирования

Нанопоровое секвенирование методом nCATS позволяет получить уникальную информацию о метилировании ДНК. С помощью этого метода удалось изучить метилирование цитозина (mC) полноразмерных генов глютеинов с использованием алгоритма DeepSignal-plant (https://github.com/PengNi/plant_5mC_analysis).

В настоящее время не существует прямых способов представления графиков метилирования на уровне ридов для данных DeepSignal-plant, поэтому был разработан специальный скрипт DeepS2bam_converter для добавления тега MM к каждому уникальному выравниванию в bam-файле. Это позволило отображать информацию о метилировании по нанопоровому секвенированию в режиме ридов (рисунок 10).

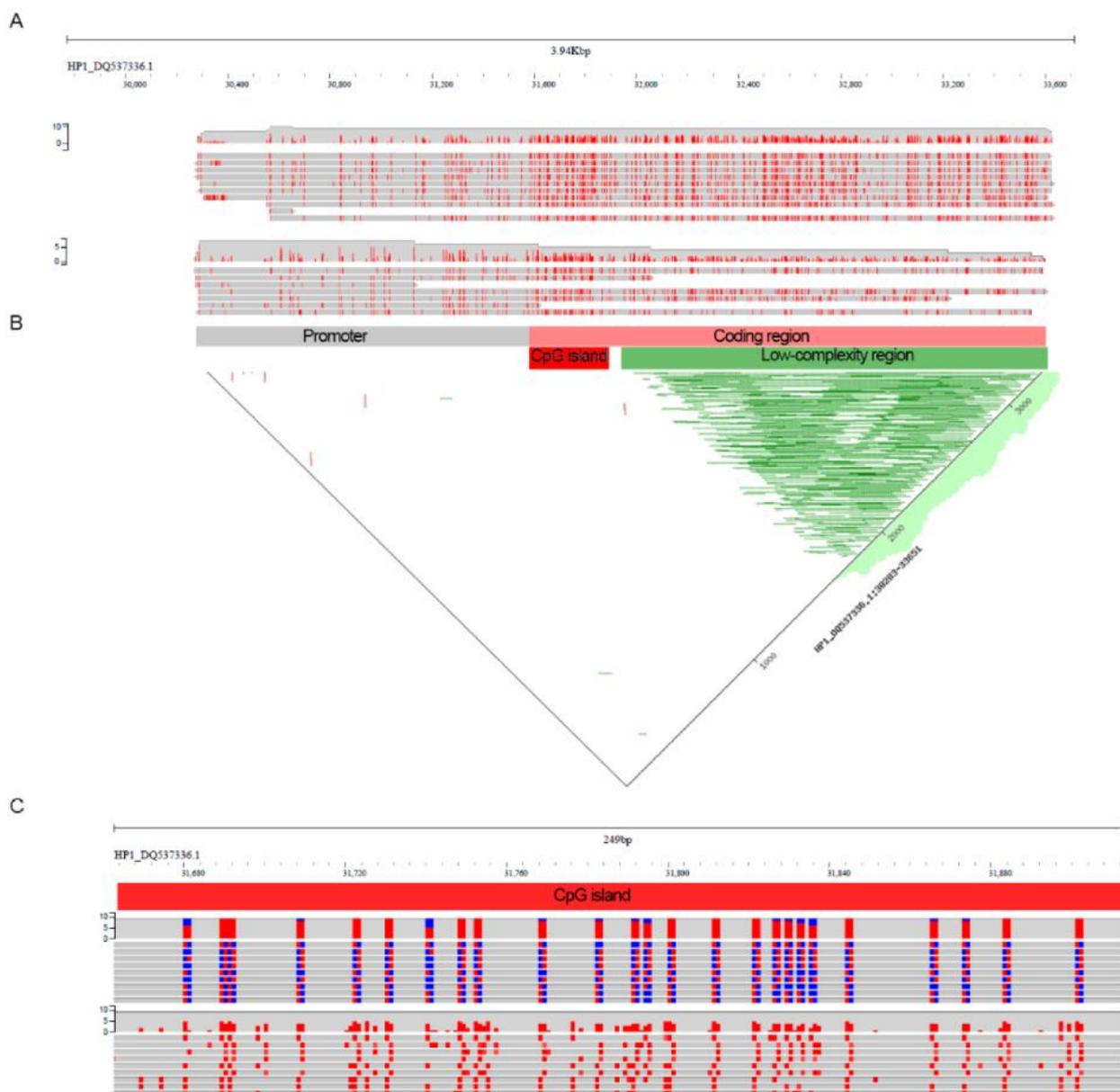


Рисунок 10 – Профиль метилирования гена *Glu-1β*, красным отмечены участки метилирования. (A) Распределение метилирования по гену *Glu-1*, полученное при анализе ридов с помощью инструмента DeepSignal metilation. Визуализация была выполнена в Jbrowse2, установленном на локальном сервере. (B) Точечный график, показывающий повторяющиеся части гена *Glu-1β* (зеленые линии). (C) Увеличенная часть гена *Glu-1β* с островком CpG. Метилированный (красный) и неметилированный (синий) цитозин CpG области показаны на верхней панели. На нижней панели показаны только метилированные цитозины.

В целом, профиль метилирования для всех трех генов был схожим: низко метилированная промоторная область и сильно метилированная кодирующая последовательность. Эти результаты показывают, что гены глютелинов имеют повышенное метилирование кодирующей области гена.

Ранее методом бисульфитного секвенирования было установлено, что метилирование промоторной области генов глютеинов выше во флаговых листьях по сравнению с развивающимися зерновками (Zhou et al., 2021). При этом профиль метилирования кодирующей области генов глютеинов до сих пор не изучен. В нашей работе было показано, что промотор генов глютеинов метилирован значительно ниже, чем кодирующая область. Такое распределение метилирования по транскрибируемой части генов глютеинов напоминает метилирование "gene body" (gbM = gene body methylation). GbM часто характерно для экспрессирующихся консервативных генов (Bewick, Schmitz, 2017). Тем не менее, гены глютеинов не соответствуют этим характеристикам, поскольку они имеют специфические для эндосперма паттерны экспрессии и демонстрируют высокую вариабельность, что подтверждается SNP-анализом аллелей гена *Glu-1By*, который выявил 222 SNPs, которые помогли идентифицировать два аллеля. Однако анализ профиля метилирования ДНК генов с высокой экспрессией в эндосперме риса показал, что эти гены часто имеют gbM (Zemach et al., 2010), что коррелирует с полученными нами результатами для тритикале.

3.2. Секвенирование и анализ вариабельности генов высокомолекулярных глютеинов методом ONT Amplicon-seq

Одновременное секвенирование генов глютеинов и их промоторов интересны тем, что полученные данные могут быть использованы для изучения вариабельности генов и их регуляторных последовательностей. Используя три последовательности генов глютеинов, мы продемонстрировали, что nCATS является подходящим инструментом для анализа только одного сорта, также следует ожидать низкой глубины секвенирования, а следовательно, требуется больше запусков для лучшего покрытия целевых фрагментов. Альтернативным методом секвенирования является подход, основанный на ПЦР-амплификации. Этот метод становится особенно привлекательным при оценке вариаций, как в кодирующей части

генов глютенинов, так и в промоторных областях. В нашем исследовании были разработаны праймеры для амплификации шести субъединиц глютенинов (*Ax/y*, *Bx/y*, *Dx/y*). Длина последовательностей составляла от 2,6 Кб для *Glu-1Ax* до 5,0 Кб для *Glu-1Dx*.

3.2.1. Амплификация последовательностей полноразмерных генов глютенинов

Для секвенирования целевых генов методом ONT Amplicon-Seq была выделена ДНК из половинок зерновок 23 сортов (таблица 5) гексаплоидной пшеницы (AABBDD).

Таблица 5. 23 изучаемых сорта яровой пшеницы

№	Название	Происхождение
1	Mexican collection №23	GLE*2//PBW343*2/TUKURU, CIMMYT
2	Mexican collection №35	LONG CHUN 15*2/4/CROC_1/AE.SQUARROSA (224)//YACO/3/MUNIA, CIMMYT
3	Mexican collection №59	GLE//PRL/2*PASTOR/3/OMSKAYA-33, CIMMYT
4	Mexican collection №65	CROS_1/AE.SQUARROSA (205)//KAUZ/3/SASIA/4/CHUAN MAI 28, CIMMYT
5	Mexican collection №66	LUTESCENS 258.92.3/GLE, CIMMYT
6	Mexican collection №67	PAVLODARSKAYA-93/SUNSTATE, CIMMYT
7	Mexican collection №70	SONATA/VORB, CIMMYT
8	Mexican collection №79	UDACHA/3/PASTOR//HXL7573/2*BAU, CIMMYT
9	Mexican collection №147	STEPNAYA 15/3/QING HAIBEI/WBLL1//BRBT2, CIMMYT
10	Mexican collection №150	STEPNAYA 16/5/TUI//2*SUNCO/SA1166/3/TUI/4/FINSI, CIMMYT
11	Mexican collection №151	STEPNAYA 16/5/TUI//2*SUNCO/SA1166/3/TUI/4/FINSI, CIMMYT
12	Mexican collection №152	STEPNAYA 16/5/TUI//2*SUNCO/SA1166/3/TUI/4/FINSI, CIMMYT
13	Mexican collection №153	LUTESCENS 1085/7/TOB/ERA//TOB/CNO67/3/PLO/4/VEE#5/5/KAUZ/6/FRET2, CIMMYT
14	Mexican collection №178	OMSKAYA 37/5/SERI*3//RL6010/4*YR/3/PASTOR/4/BAV92, CIMMYT

№	Название	Происхождение
15	Mexican collection №187	LUTESCENS 210.99.10/4/MILAN/SHA7/3/CROC_1/AE.SQUARROSA(224)//OPATA, CIMMYT
16	Mexican collection №215	53.94.98.2/3/T.DICOCCON PI94625/AE.SQUARROSA (372)//3*PASTOR/4/GVK 1369.2, CIMMYT
17	Mexican collection №217	FITON 42/3/T.DICOCCON PI94625/AE.SQUARROSA (372)//3*PASTOR/4/GVK 1857.9, CIMMYT
18	Mexican collection №220	LUTESCENS 196.94.6*2/VORB, CIMMYT
19	Mexican collection №223	LUTESCENS 196.94.6*2/VORB, CIMMYT
20	Glenlea	Pembina/Bage//CB10, Plant Science Department, university of Manitoba, Winnipeg, Man., Canada
21	Бомбона	Danko Hodowla Roślin Sp. z o.o., Choryń, Poland
22	Агата	Artyomovka/Milturum 63, Federal Research Center Nemchinovka, Moscow region, Russia
23	Канюк	Belvoir///Mexikaner//Sappo/Sokrates, Secobra Recherches s.a.s (Centre de Bois Henry, Maule, France)

Для амплификации генов-мишеней были разработаны специфичные праймеры на консервативные участки промоторной и белок кодирующей областей и проведена амплификация целевых генов 23 сортов пшеницы (рисунок 11, таблица 6).

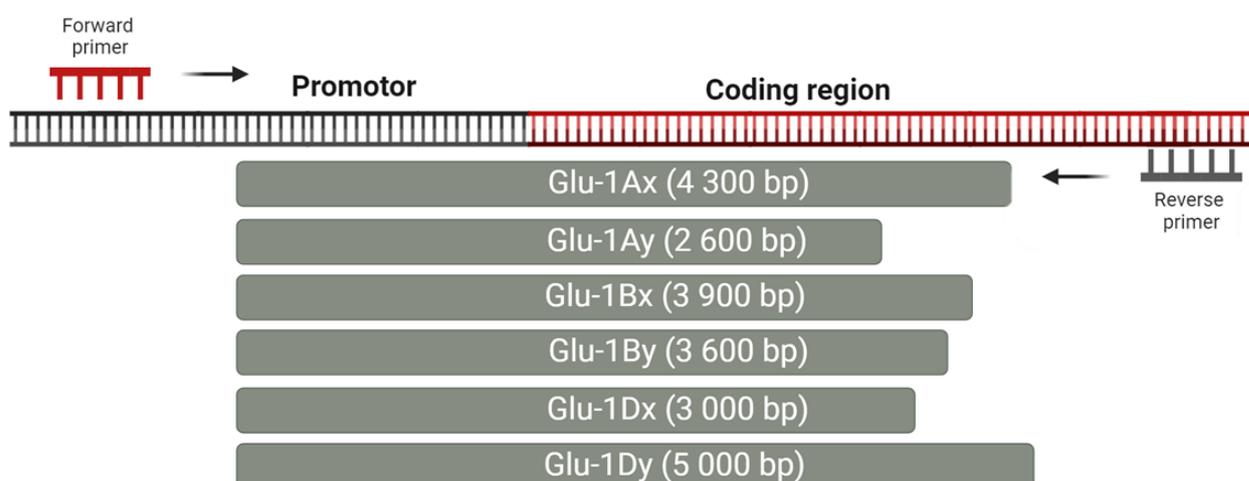


Рисунок 11. Визуальное представление расположения праймеров, разработанных в этом исследовании, нацеленных на полиморфные аллельные варианты генов глютеина. Последовательности были выровнены с помощью BLAST для создания пар праймеров для амплификации как промоторных, так и белок кодирующей областей.

Таблица 6. Последовательности разработанных праймеров для генов *Glu-1*

Название праймера	Последовательность 5'-3'	Ожидаемая длина ПЦР-продукта (п.н.)
<i>Glu-1Ax</i>	F: CTCCATCTGTTTCTTTGACCAAGTTCACAAC	4 300
	R: CTGGTGCTCCGCGCTAACATGG	
<i>Glu-1Ay</i>	F: GGGAAAGTGGA AAAAATCCCCTATTCTGGTG	2 600
	R: GATGGTGCGCCTTTGCCACC	
<i>Glu-1Bx</i>	F: GGGGCAGGGAAGAAACACTTGG	3 900
	R: GAGGCACAGGTGAGCGCACAAT	
<i>Glu-1By</i>	F: TGGACGATGGGAGATGAAAGTGA	3 600
	R: СТАСАСТGGCTAGCCGATAATGCG	
<i>Glu-1Dx</i>	F: GAGTGGCCGTAGATTTGCAAAAGGC	3 000
	R: СТАAAGTGCATGCATGCCTAAGCACCATG	
<i>Glu-1Dy</i>	F: CCATGGCAGCCGTGTTGACGTA	5 000
	R: GCTGGAAAGCTTTTCATTTTGCAATTATTATGGG	

Чтобы добиться успешного синтеза ПЦР-продукта для каждой субъединицы, необходимо было индивидуально оптимизировать условия амплификации. Это связано с тем, что гены глютенинов обладают сложными и полиморфными структурами, а следовательно, программа амплификации подбиралась индивидуально для каждой субъединицы (таблица 7).

Таблица 7. Программа амплификации для каждой пары праймеров для генов *Glu-1*

	Encyclo polymerase		Biolabmix LR HS polymerase			
	<i>Glu-1Bx</i>	<i>Glu-1By</i>	<i>Glu-1Ax</i>	<i>Glu-1Ay</i>	<i>Glu-1Dx</i>	<i>Glu-1Dy</i>
Первичная денатурация	95°C - 4 мин	95°C - 4 мин	94°C - 4 мин	94°C - 4 мин	94°C - 4 мин	94°C - 4 мин
Денатурация	95°C - 20 с	95°C - 20 с	94°C - 20 с	94°C - 20 с	94°C - 20 с	94°C - 20 с
Отжиг праймеров	62°C - 30 с	62°C - 30 с	61°C - 30 с	67°C - 30 с	62°C - 30 с	61°C - 30 с
Элонгация	72°C - 4 мин	72°C - 4 мин	68°C - 3 мин	68°C - 2 мин	68°C - 3 мин	68°C - 3 мин
Циклы	31	35	33	30	30	30
Финальная элонгация	72°C - 4 мин	72°C - 4 мин	68°C - 4 мин	68°C - 4 мин	68°C - 3 мин	68°C - 3 мин

Длины последовательностей для этих целевых областей были следующими: 4,3 тыс.п.н. и 2,6 тыс.п.н. для *Glu-1Ax* и *Glu-1Ay*, 3,9 тыс.п.н. и 3,6 тыс.п.н. для *Glu-1Bx* и *Glu-1By*, а также 3,0 тыс.п.н. и 5,0 тыс.п.н. для *Glu-1Dx* и *Glu-1Dy*, соответственно. Полученные продукты ПЦР затем оценивали с помощью электрофореза с использованием 1% агарозного геля (рисунок 12).

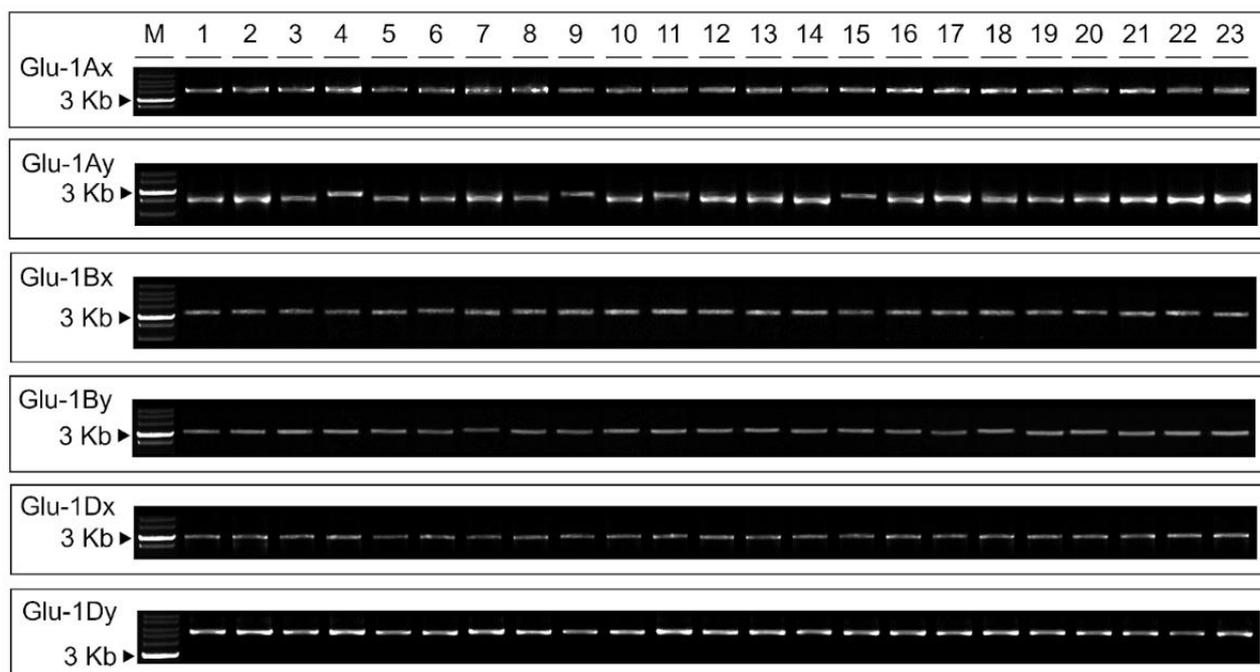


Рисунок 12. Результаты амплификации генов высокомолекулярных глютенинов (*Glu-1Ax*, *Glu-1Ay*, *Glu-1Bx*, *Glu-1By*, *Glu-1Dx*, *Glu-1Dy*) в коллекции мягкой пшеницы. Представленные сорта мягкой пшеницы: МС№23, МС№35, МС№59, МС №65, МС№66, МС№67, МС№70, МС№79, МС№147, МС№150, МС№151, МС№152, МС№153, МС№178, МС№187, МС№215, МС№217, МС№220, МС№223, Glenlea, Bombona, Agatha, Kanyuk. ДНК-маркер Sky-High (Biolabmix, Novosibirsk, Russia) (M).

Из-за наличия неспецифичной примеси в ПЦР-продукте *Glu-1Bx*, требовалась очистка целевого продукта из геля. Контроль качества показал, что полученная концентрация продуктов (*Bx*) после очистки на колонках составляла 8 - 10 нг/мкл во всех образцах. Концентрация остальных ампликонов (*Ax/y*, *By*, *Dx/y*) в ПЦР-смеси до очистки варьировалась в пределах 816 – 840 нг/мкл (таблица 8). Ампликоны были в равной

концентрации объединены в одну пробу в соответствии с сортом и подверглись очистке на магнитных шариках.

Таблица 8. Концентрация очищенных ПЦР-продуктов

№ образца	Концентрация объединенных образцов (Ax/y, By, Dx/y), нг/мкл	Концентрация Bx, нг/мкл
1	30,1	8.8
2	40,2	8.1
3	31,8	8.5
4	34,1	9.6
5	34,4	8.6
6	32,8	8.6
7	32,9	9.4
8	31,3	9.6
9	31,0	8.6
10	34,2	9.4
11	37,4	8.3
12	38,3	7.5
13	38,4	8.7
14	38,7	8.3
15	37,5	8.3
16	35,3	7.4
17	40,7	8.2
18	39,4	10
20	34,6	8.9
30	39,6	8.8
39	35,8	8.4
50	35,8	9.0
61	30,8	10.0

На последующем этапе был проведен повторный цикл разведения, но на этот раз с использованием смеси ранее полученных ПЦР-продуктов. Эта оптимизация была необходима для достижения желаемого количества амплифицированных молекул-мишеней в исследуемом геномном материале. Концентрацию и качество объединенных ампликонов оценивали с помощью Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA) и Qubit (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Для контроля качества объединенная проба, состоящая из ампликонов шести генов глютенинов, загружалась в 1% агарозный гель (рисунок 13). На электрофореграмме можно видеть 2 фрагмента ПЦР-продуктов, которые соответствуют ожидаемым размерам амплифицированных участков от 3000 п.н. до 5000 п.н.

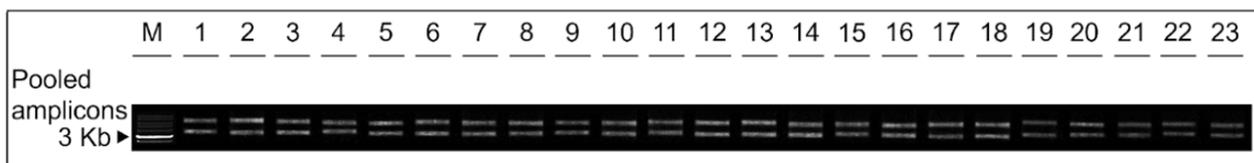


Рисунок 13. Электрофореграмма объединенных ампликонов генов глютенинов (*Glu-1Ax*, *Glu-1Ay*, *Glu-1Bx*, *Glu-1By*, *Glu-1Dx*, *Glu-1Dy*) в коллекции мягкой пшеницы. Представленные сорта мягкой пшеницы: МС№23, МС№35, МС№59, МС №65, МС№66, МС№67, МС№70, МС№79, МС№147, МС№150, МС№151, МС№152, МС№153, МС№178, МС№187, МС№215, МС№217, МС№220, МС№223, Glenlea, Bombona, Agatha, Kanyuk. ДНК-маркер Sky-High (Biolabmix, Novosibirsk, Russia) (М).

Результатом амплификации стали 138 ПЦР-продуктов (6 генов глютенинов для 23 образцов коллекции). Ампликоны каждого образца пшеницы были очищены и объединены для получения окончательных образцов для баркодирования (рисунок 14). Примечательно, что в процессе амплификации использовались нефосфорилированные праймеры. Фосфорилирование ДНК было проведено с использованием метода PНК, который переносит γ -фосфат из АТФ на 5'-конец нуклеиновых кислот в нефосфорилированных продуктах ПЦР.

Затем фосфорилированные ПЦР-ампликоны каждого образца пшеницы были баркодированы, объединены в одну пробирку и секвенированы. Секвенирование проводилось последовательно по 11-12 баркодов (образцов коллекции) за цикл (рисунок 14).

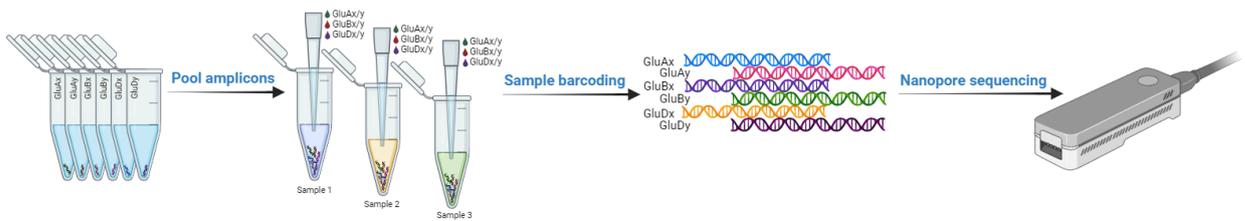


Рисунок 14. Схема, изображающая идентификацию HMW-GS с помощью ONT Amplicon-Seq.

Анализ ридов показал, что глубина покрытия 50×-150× была получена для каждого ампликона в соответствии с образцом пшеницы всего через один час секвенирования. В общей сложности мы получили ~215000 ридов со значением N50 около 3,2 тыс.п.н. (рисунок 15).

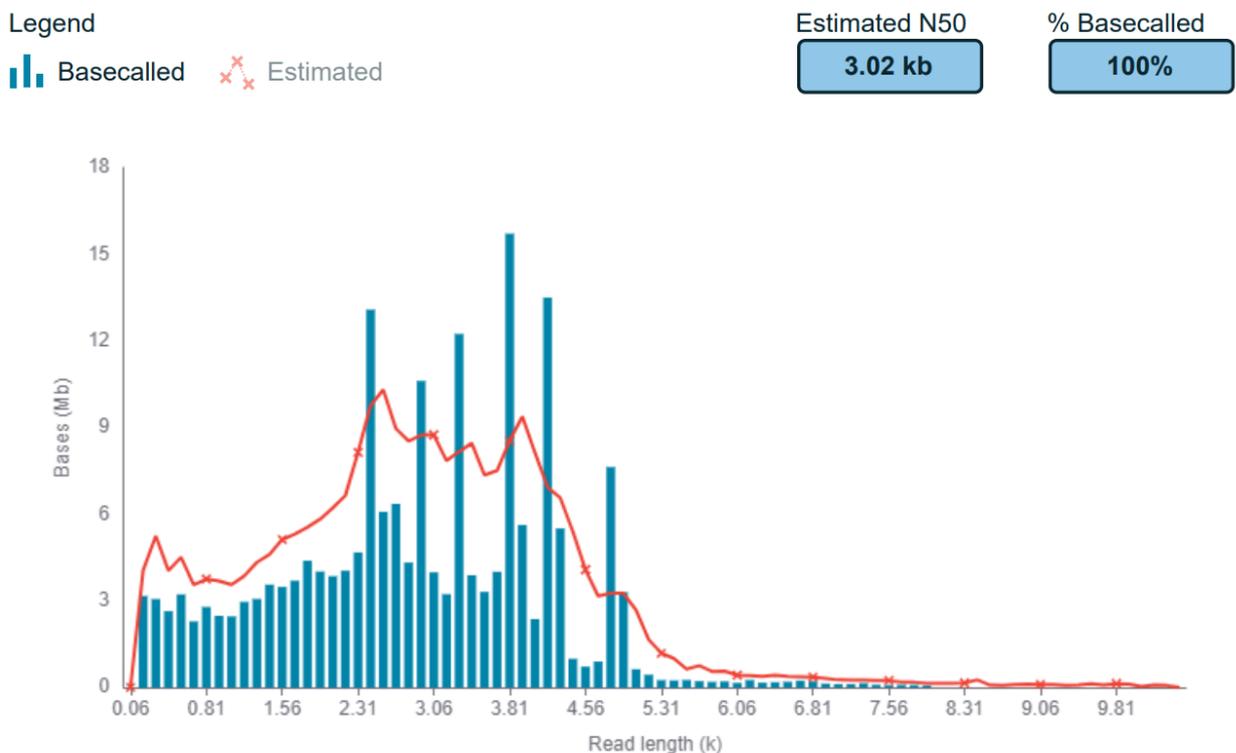


Рисунок 15. Гистограмма, показывающая суммарную длину ридов ONT Amplicon-Seq.

После первого запуска было получено 97333 ридов, а после второго - 117338. Эти риды впоследствии сравнивались с референсными последовательностями различных аллелей глютенинов. Результаты демонстрируют, что ONT Amplicon-Seq может быстро генерировать достаточное количество данных за короткое время.

3.2.2. Анализ последовательностей аллелей HMW-GS, полученных с помощью ONT Amplicon-Seq

Для идентификации аллельных вариантов каждого из шести генов HMW-GS полученные риды были картированы на референсные последовательности генов HMW-GS (*Glu-1A*: AF145590.1, KJ531446.1, M22208.2, MF568383.1, EU984510.1; *Glu-1B*: DQ119142.1, FM955452.1, MH108092.1, KC254854.1, JN255519.1, X61026.1, EU137874.1, EF540765.1, KF430649.1; *Glu-1D*: BK006460.1, AB485591.1, X12929.2, JF736016.1). Затем были идентифицированы SNPs и InDels, а также присвоены соответствующие идентификаторы аллелей HMW-GS, если они были ранее известны и депонированы в GenBank.

Анализ вариации *Glu-1Ax* в 23 образцах пшеницы показал, что 15 образцов имели аллель *Glu-1Ax2* (M22208.2), два образца несут аллель *Ax-null* и пять образцов имели аллель *Glu-1Ax1*. Последовательности генов *Glu-1Ax* двух образцов пшеницы (МС №147, МС №151) имела SNP (3136C -> T) в кодирующей области гена, отличающие их от известных последовательностей *Glu-1Ax1* (рисунок 16). Эта замена не является синонимичной, что приводит к замене аминокислоты пролин (Pro) на серин (Ser). Новый аллель *Glu-1Ax* был назван *Glu-1Ax1-T*.

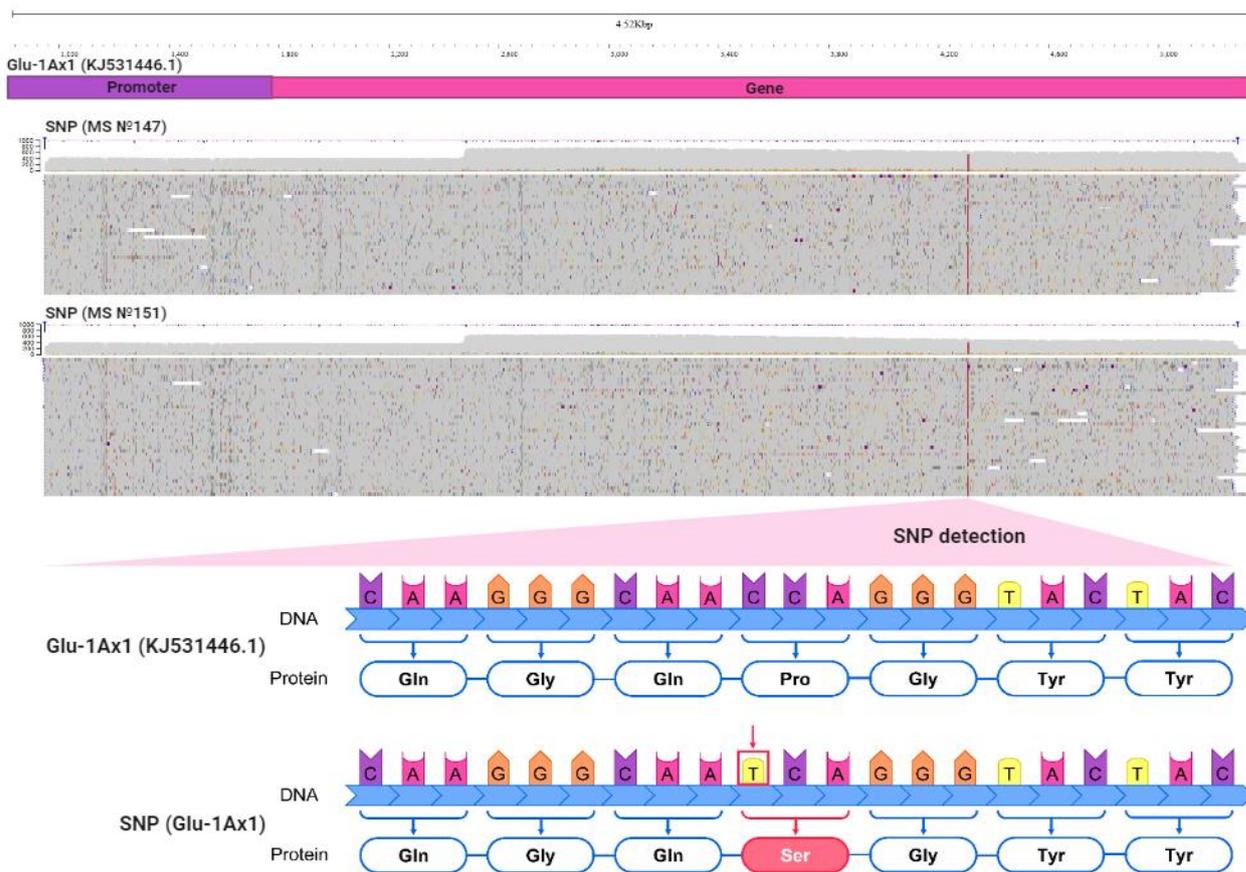


Рисунок. 16. Идентификация SNP в кодирующей части гена *Glu-1Ax1* по сравнению с референсной последовательностью (KJ531446.1). В кодирующей области гена *Glu-1Ax1* была идентифицирована несинонимичная замена, приводящая к изменению аминокислотной последовательности.

Ген *Glu-1A α* обычно неактивен у культивируемых гексаплоидных сортов пшеницы, что делает невозможным его анализ с помощью SDS-PAGE. Однако ONT Amplicon-Seq позволил проанализировать последовательности этой субъединицы. Было обнаружено два аллеля *Glu-1A α* в исследуемой коллекции пшеницы. Аллели были идентичны двум известным аллелям, депонированным в GenBank: аллель *Glu-1A α -d* (MF568383.1 (Luo et al., 2018)) и аллель *A1 γ /Td-s* (EU984510.1, *Glu-1A-2* (Jiang et al., 2009)). Аллель *Glu-1A α -d* был детектирован у шестнадцати образцов в коллекции пшеницы. Остальные семь образцов несли аллель *Glu-1A-2*, который был классифицирован как неактивный (Jiang et al., 2009).

Сравнение полученных последовательностей гена *Glu-1Bx* с ранее опубликованными последовательностями выявило наличие разнообразной панели аллелей *Bx* в нашей коллекции пшеницы: *Bx7* (17 образцов, 73%), *Bx17* (2 образца, 9%), *Bx13* (2 образца, 9%), и *Bx14* (2 образца, 9%). Используя данные ONT, была обнаружена делеция размером ~60 п.н. и SNP (С -> G), расположенные в промоторной области аллеля *1Bx7* (рисунок 17).

Таблица 9. Последовательности разработанных праймеров для идентификации делеции в промоторе гена *Glu-1Bx7*.

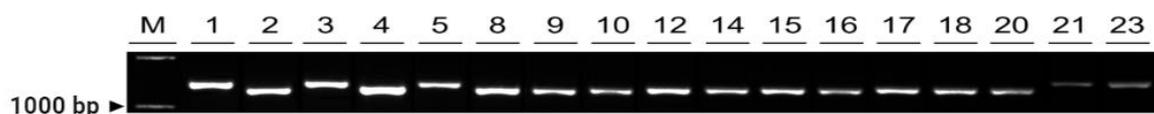
Название праймера	Последовательность 5'-3'	Ожидаемая длина ПЦР-продукта, (п.н.)
<i>Glu-1Bx7</i> prom	F: CGTTTCATGCTTAATCCAATGATATCGACT	830
	R: AATTAACATAGGATCTTAGTAAATCCGCGTC	

Чтобы подтвердить наличие делеции, были разработаны уникальные пары праймеров и проведен ПЦР-анализ (таблица 9). Результаты ПЦР и секвенирования по Сэнгеру подтвердили наличие делеции в вариантах *Glu-1Bx7* в образцах пшеницы (рисунок 17).

A. Flanking target region by design PCR primers



B. Visualization of PCR results with Glu-1Bx7_del primers in agarose gel



C. Sanger sequencing

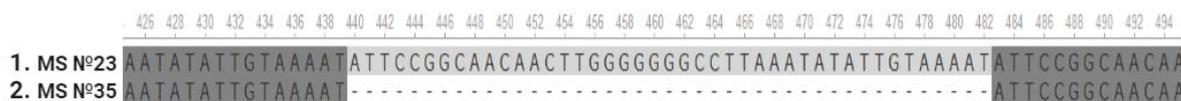


Рисунок. 17. Расположение праймеров, разработанных для анализа структурной вариации (делеции) промотора *Glu-1Bx7* в коллекции мягкой пшеницы. (А) Разработка праймеров для фланкирования делеции в области промотора. (В) Результаты ПЦР со специфическими праймерами для детекции делеции *Glu-1Bx7*. Представленные сорта пшеницы: 2. МС № 35, 4. МС № 65, 8. МС № 79, 9. МС № 147, 10. МС № 150, 12. МС № 152, 14. МС № 178, 15. МС № 187, 16. МС № 215, 17. МС № 217, 18. МС № 220, 20. МС № 223 (with 60 bp deletion); 1. МС № 23, 3. МС № 59, 5. МС № 66, 21. Glenlea, 23. Agatha (without deletion). (С) Идентификация делеции после многократного выравнивания последовательностей, полученных секвенированием по Сэнгеру. МС означает “Mexican Collection”. ДНК-маркер Step50 plus (Biolabmix, Novosibirsk, Russia) (M).

Анализ последовательностей *Glu-1Bx*, полученных с помощью ONT Amplicon-Seq, показал, что образцы пшеницы МС №151 и МС №67 несли аллели, схожие с известными аллелями *By16* и *By19* (рисунок 18) соответственно. Эти два аллеля различаются на 27 п.н, а соответственно, делеция 9 аминокислот слишком мала, чтобы быть обнаруженной с помощью SDS-PAGE, что указывает на большую надежность ONT Amplicon-Seq в

дифференциации аллелей HMW-GS. Анализ остальных 21 образцов пшеницы показал, что 8 (35%) образцов имели аллель *Glu-1By8*, 9 образцов (40%) имели аллели *Glu-1By9*, а *Glu-1By15* (9%) и *Glu-1By8* (9%) по два образца.



Рисунок. 18. Структурные изменения были обнаружены путем сопоставления полученных ридов с референсной последовательностью. Анализ включал сопоставление секвенированных данных с аллелем, идентифицированным с помощью SDS-PAGE (*Bx16*) и подтвержденным с помощью BLAST-анализа (*Bx19**).

Для генов *Glu-1D* были выявлены следующие распределения по аллелям: 8 образцов (35%) имели аллели *Glu-1Dx2* и *Glu-1Dy12*, и 15 образцов (65%) имели аллели *Glu-1Dx5* и *Glu-1Dy10*.

Таким образом, полученные результаты ясно показали, что ONT Amplicon-Seq может идентифицировать аллели генов HMW-GS с высокой точностью. Более того, используя данные ONT, мы обнаружили ранее неизвестный аллель *Glu-1Ax1-T*, который может быть изучен в будущем.

3.2.3. Филогенетический анализ аллелей генов глютенинов

Для изучения эволюционных взаимосвязей между аллелями, кодируемыми *Glu-1A*, *Glu-1B* и *Glu-1D* (рисунок 19) был проведен филогенетический анализ. В наборе из 23 образцов яровой пшеницы был идентифицирован 21 различный аллель HMW-GSs. Примечательно, что два

из этих аллелей отличались от тех, которые ранее были задокументированы в общедоступных базах данных.

Последовательности HMW-GSs позволили получить ценную информацию об эволюционной истории. Результаты выявили четкое разделение между генами *x*-типа и *y*-типа HMW-GSs в филогенетическом дереве, образующими шесть первичных кладов, каждая из которых соответствует определенной субъединице.

Внутри клады *y*-типа гены *Dy* и *Bu* были распределены по отдельным подгруппам. Кроме того, ген *Bu15* сформировал свою собственную отдельную ветвь, хотя он поддерживал более тесную связь с генами *Bu8*, *Bu9*, *Bu16* и *Bu19*. Гены *Glu-1Ay* также были сгруппированы в отдельную подгруппу.

В кладе генов *x*-типа гены *Glu-1Bx* были сгруппированы в определенную подгруппу, в то время как гены *Dx* и *Ax* были отнесены к другой подгруппе. Примечательно, что среди генов *Glu-1Bx* была идентифицирована новая последовательность с делецией в промоторной области *Glu-1Bx7*. Эта уникальная последовательность была сгруппирована отдельно, но в пределах той же подгруппы, что и типичные гены *Glu-1Bx7* и *Glu-1Bx17*.

Последняя клада содержала две подгруппы, *Dx* и *Ax*, из которых субъединица *Ax* содержала аллели генов со структурными вариациями, отличными от канонических. В частности, эти вариации включали *Glu-A1x1-T* с однонуклеотидным полиморфизмом (SNP), расположенным в области гена.

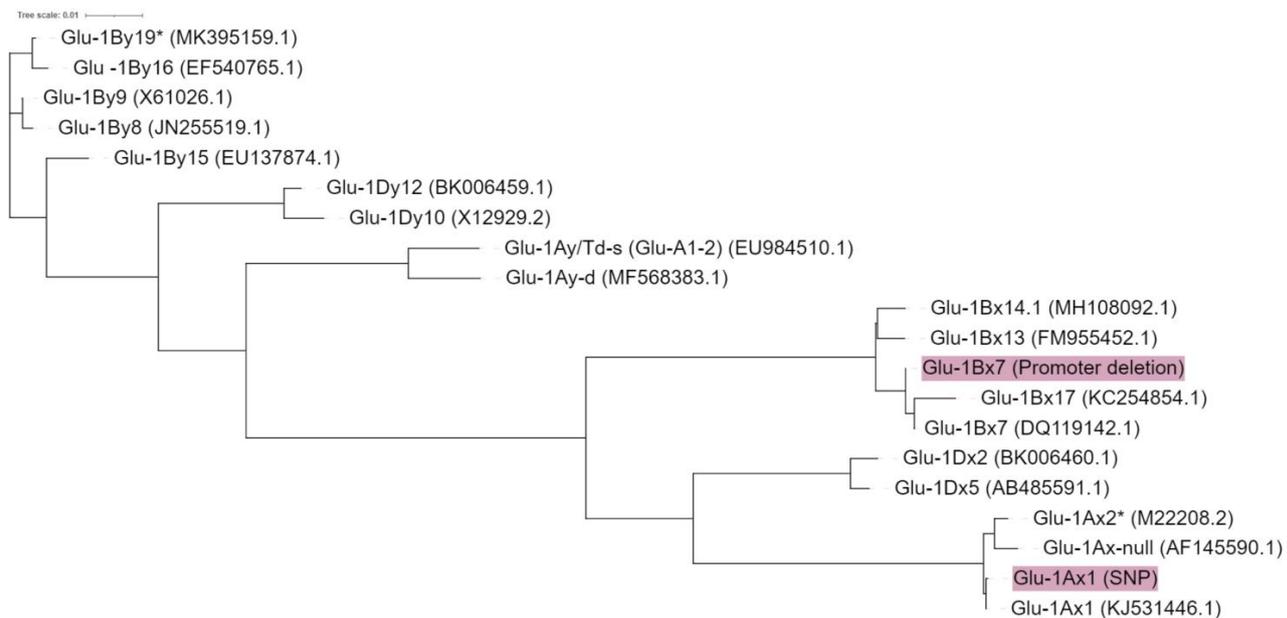


Рисунок 19. Филогенетическое дерево аллелей генов глютелинов (HMW-GS), основанное на ONT Amplicon-Seq.

3.2.4. Преимущества Nanopore Amplicon-Seq для анализа HMW-GS

Использование ONT Amplicon-Seq для секвенирования полноразмерных генов глютенинов в 23 образцах пшеницы позволяет выделить ряд преимуществ данного метода по сравнению с классическим методом белкового электрофореза (SDS-PAGE).

Во-первых, метод позволяет анализировать отдельные растения, что особенно важно при оценке селекционного материала в ранних поколениях гибридов. При этом для выделения ДНК и ПЦР нужно немного растительного материала (100 мг), что даёт возможность проводить анализ на самых ранних этапах развития растений.

Во-вторых, с помощью ONT Amplicon-Seq можно получить полноразмерные риды, что даёт целостную информацию о нуклеотидных последовательностях генов и промоторов. Это позволяет точно идентифицировать известные аллели и выявлять новые, характеризующиеся однонуклеотидными полиморфизмами, делециями и вставками. Такой новый подход потенциально может объяснить случаи, когда высокий уровень глютенинов не приводит к высоким показателям качества хлеба, или наоборот. Например, качество хлеба российских (Rabinovich et al., 1998), турецких (Temizgul et al., 2018) и индийских (Goel et al., 2018) сортов пшеницы связано с аллелем *Glu-D1a* и недавно открытыми *1Dy12.7* (Jiang et al., 2009) и *1Dy12*** (Du et al., 2018). Они обладают той же электрофоретической подвижностью, что и *1Dy12*, но, как и аллель *1Dy10*, являются основным фактором, определяющим технологическое качество. Кроме того, благодаря ONT Amplicon-Seq гены, отвечающие за другие экономически ценные признаки, могут быть проанализированы одновременно с генами HMW-GS (мультиплексирование) в рамках одного запуска секвенирования.

В-третьих, нанопоровое секвенирование ампликонов является более эффективным методом благодаря сокращению времени подготовки

библиотеки. Хотя для этого метода требуются ампликоны интересующего гена, современные инструменты биоинформатики и растущее число секвенированных геномов упростили задачу конструирования праймеров и поиска гомологичных последовательностей у разных видов растений. Это делает ONT Amplicon-Seq мощным инструментом для геномной селекции. Разработка и валидация праймеров занимает всего несколько часов в то время, как продолжительность амплификации с разработанными праймерами зависит от длины целевого участка. В нашем эксперименте время амплификации для каждой субъединицы глютеенинов составляло в среднем около 110 минут. Очистка ПЦР-продуктов и их последующая подготовка к секвенированию требуют точности в связи с тем, что качество секвенирования в значительной степени зависит от этих этапов. Непосредственно процесс секвенирования может занимать менее 1 часа, что достаточно для получения количества ридов, необходимого для анализа аллельного состава глютеенинов (рисунок 20). Секвенирование ампликонов является эффективным подходом, позволяющим упростить и ускорить анализ варибельности целевых генов.



Рисунок 20. Этапы подготовки ампликонов и Нанопоровое секвенирование. Каждый шаг на временной шкале представляет собой один рабочий день.

ONT Amplicon-Seq также имеет некоторые ограничения, которые следует учитывать. Например, высокая частота ошибок, образование химерных ридов и отсутствие возможности профилирования метилирования из-за этапа амплификации. Однако благодаря большому количеству ридов

ампликонов можно минимизировать значительное количество возникающих ошибок.

В целом предложенный нами метод идентификации HMW-GSs является высокоинформативным, экономически эффективным и позволяет получить обширную информацию об аллельном состоянии локусов *Glu-1* и других генов.

3.3. Анализ транскриптома зерновок тритикале с помощью Нанопорового секвенирования

Транскриптом растений динамично изменяется. Многие гены экспрессируются только в определенных условиях. Кроме этого, методы анализа транскриптома имеют свои ограничения для детекции экспрессирующихся генов. Учитывая динамику транскриптома и сложность генома злаковых, детекция генов, экспрессирующихся в процессе развития зерновки и их аннотация, при использовании новых технологий является актуальной задачей. Ряд недавних исследований показывает, что применение разных подходов и технологий секвенирования позволяет значительно расширить границы изучения транскриптома и идентифицировать новые гены. Так, например, глубокий анализ транскриптома пшеницы на разных стадиях позволил обнаружить ген *wbm*, который вовлечён в контроль хлебопекарных качеств пшеницы (Furtado et al., 2015).

Целью этой части диссертационного исследования была идентификация ранее неаннотированных генов А и В геномов тритикале, экспрессирующихся на разных стадиях (10, 15, 20 дней после цветения) развития зерновки и в разных её частях. Для этого мы использовали нанопоровое секвенирование РНК и кДНК.

3.3.1. Анализ транскриптов, полученных с помощью Нанопорового секвенирования, по белок-кодирующему потенциалу.

Отбор зерновок для выделения РНК проводился на 15 и 20 день после цветения (рисунок 21). Кроме этого, каждая зерновка была разделена на две части: эндосперм и зародыш+эндосперм. Из каждой части была выделена РНК и получена кДНК из поли-А+ фракции. В итоге для двух стадий развития, 15 и 20 день после цветения, было получено три образца РНК: (1) РНК, выделенная из целой зерновки (WS15 и WS20); (2) РНК, выделенная из половины зерновки без зародыша, которая включала рассеченный слой алейрона и крахмалистый эндосперм (EN15 и EN20) и (3) РНК из препарированного зародыша с небольшой частью эндосперма (EE15 и EE20) (приложение, рисунок 3).



Рисунок 21 – Морфологическое развитие зерновок яровой тритикале линии Л8665

Секвенирование кДНК проводили в течение 20 часов на проточной ячейке MinION. Чтобы определить транскрипционно-активные области генома, которые ранее не были аннотированы ("неаннотированные гены"), было проведено картирование ридов на геном и проведена сборка транскриптома с помощью программы stringTi2. Секвенирование с помощью технологии Oxford Nanopore (ONT) этих образцов привело к 80000 длинным ридам со средним значением N50 1,3 Кб. В общей сложности для этого

исследования было использовано 6,9 миллиона высококачественных ONT-ридов (Qscore > 8).

Риды гексаплоидной тритикале (ABR) сопоставлялись с ABDR-субгеномами, полученным путем объединения геномов пшеницы IWGSC54 (ABD) (IWGSC, 2018) и ржи Lo7 (R) (Rabanus-Wallace et al., 2021). Поскольку при селекции тритикале возможны интрогрессии генетического материала субгенома D в субгеномы ABR, последовательность D-генома была исключена из биоинформатического анализа. Однако, чтобы уменьшить артефакты, транскрипты, собранные для D-генома, не анализировали в этом исследовании. Для анализа использовались транскрипты с уровнем экспрессии TPM >5 и длиной <5000 п.н. (рисунок 22).

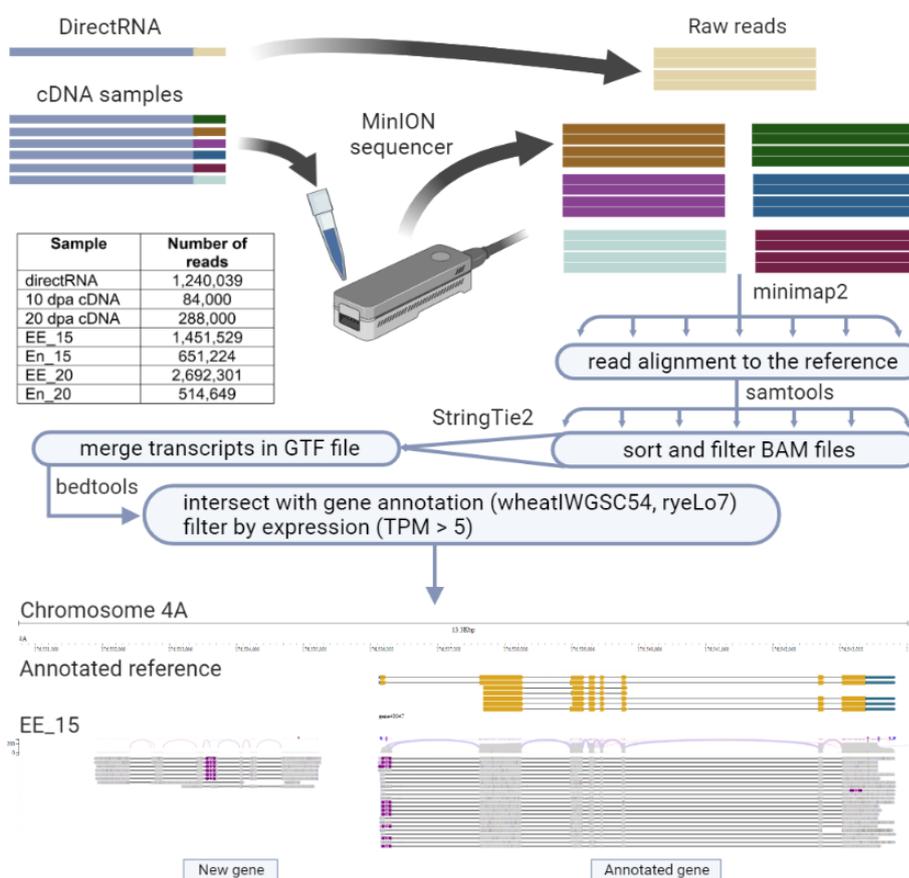


Рисунок 22 – Нанопоровое секвенирование и последующий биоинформатический анализ поиска ранее неаннотированных генов тритикале. Шесть секвенированных образцов кДНК были объединены с данными прямого секвенирования РНК. Визуализация области 4А хромосомы с ранее неаннотированными генами в JBrowser2

В соответствии с последней аннотацией геномов A, B (IWGS54) и R (Lo7) были определены гены, расположенные в межгеномном пространстве и не имеющие совпадений с аннотированными. При использовании данных критериев было идентифицировано 7128 (17,6%) генов как ранее неаннотированные («новые») в геномах A, B и R. Из них 2340, 2947 и 1841 генов были расположены в геномах A, B, и R, соответственно (рисунок 23).

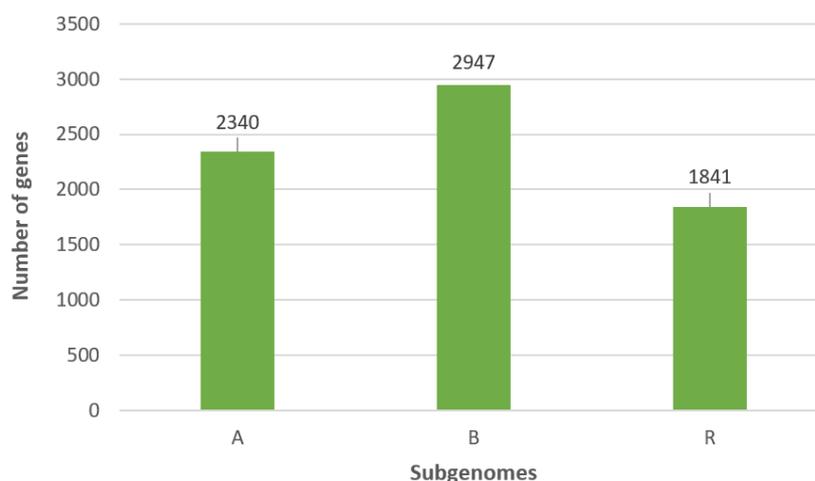


Рисунок 23 – Гистограмма количества ранее неаннотированных генов, расположенных в трех субгеномах яровой тритикале.

Геном R содержит наибольшее количество генов, экспрессирующихся во время развития зерновки тритикале (13458 генов, рисунок 24), и наименьшее количество ранее неаннотированных генов.

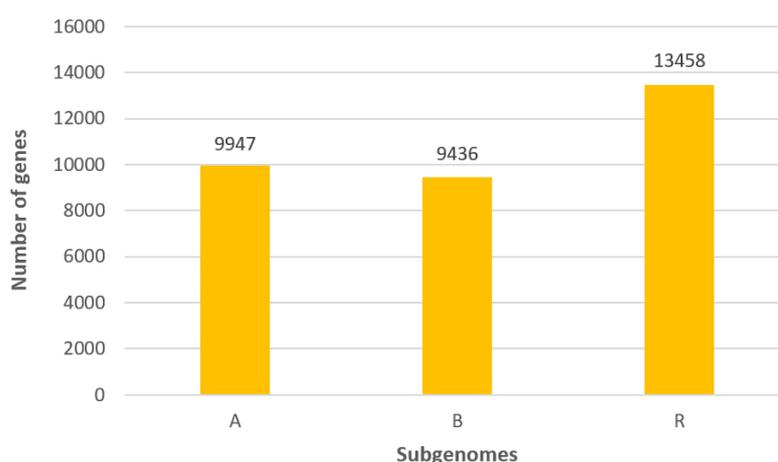


Рисунок 24 – Гистограмма количества аннотированных генов, расположенных в трех субгеномах яровой тритикале.

Осуществлялась оценка потенциала кодирования белка с помощью трех биоинформатических инструментов (PLE, LncFinder, CNCI). Транскрипты, определенные как некодирующие белок одновременно тремя инструментами, классифицировались как днРНК.

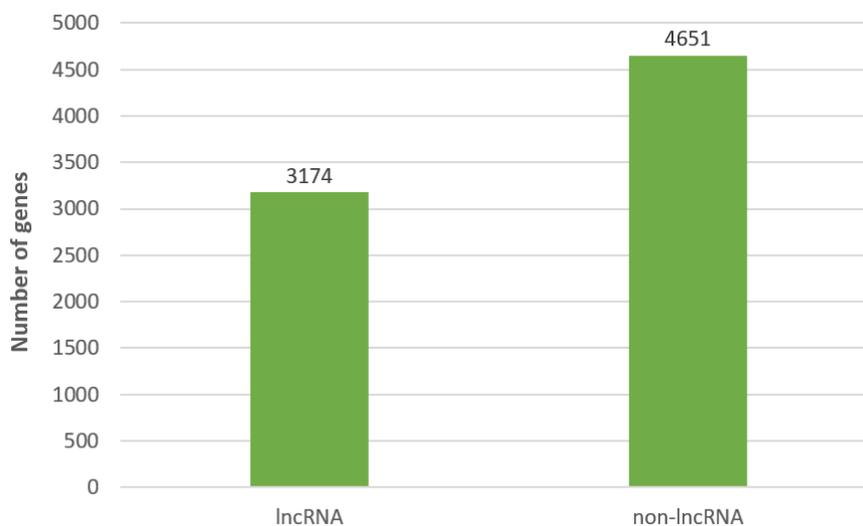


Рисунок 25 – Количество транскриптов, классифицированных как днРНК, экспрессируемых из ранее неаннотированных генов.

В общей сложности было идентифицировано 4697 днРНК (рисунок 25), из них 3174 днРНК были транскрибированы из неаннотированных, а 1523 днРНК - аннотированных генов (рисунок 26). Доля днРНК (40,5%) в новых транскриптах была в 10 раз выше (критерий χ^2 Пирсона с поправкой Йетса, p -value $< 2,2 \times 10^{-16}$), чем в наборе аннотированных (4,1%).

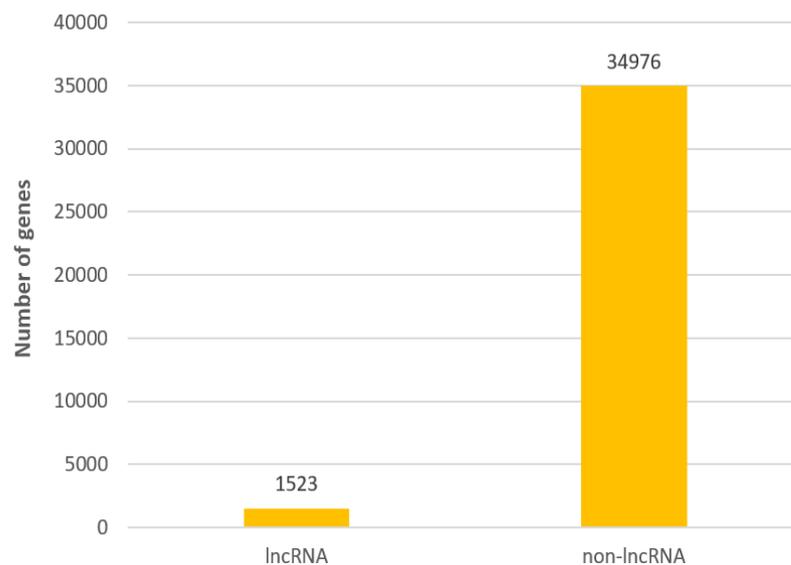


Рисунок 26 –Количество транскриптов, классифицированных как днРНК, экспрессируемых из аннотированных генов.

Анализ данных показывает, что некоторые днРНК недооценены в растениях (Kirou et al., 2020). Таким образом, было продемонстрировано, что днРНК составляют более 10% транскриптома развивающихся зерновок тритикале, что свидетельствует о их различной роли в биологических процессах.

3.3.2. Характеристика генов, экспрессирующихся на разных стадиях развития зерновки

Развитие зерновки – чрезвычайно динамичный процесс, включающий в себя несколько транскриптомных стадий. В данном исследовании мы стремились расшифровать гены, которые уникально экспрессируются на ранней, средней и поздней стадиях развития зерновки. Для этого сравнивались значения экспрессии всех идентифицированных генов, используя данные о транскриптах на 10 dpa (ранняя стадия, 1 запуск ONT секвенирование РНК, 1240039 ридов (Kirov et al., 2020)), 15 dpa (средняя стадия, ONT секвенирование кДНК, 2186753 ридов) и 20 dpa (поздняя стадия, ONT секвенирование кДНК, 3494950 ридов). Учитывая значение экспрессии $TPM > 2$, был получен набор, состоящий из 6057, 1575 и 3274 генов, специфически экспрессирующихся на ранней, средней и поздней стадиях, соответственно (рисунок 27А). Из всех 39914 выявленных генов были идентифицированы те, которые могут экспрессироваться как на ранней, так и на средней стадиях развития зерновки – 2281, на средних и на поздних – 4769, на ранних и поздних – 2735. 19223 генов экспрессировались на всех стадиях развития зерновки. Доля днРНК среди аннотированных транскриптов составляет 4,6% (879) на ранней, 42,7% (1756) на средней и 31,6% (1739) на поздней стадиях развития зерновки.

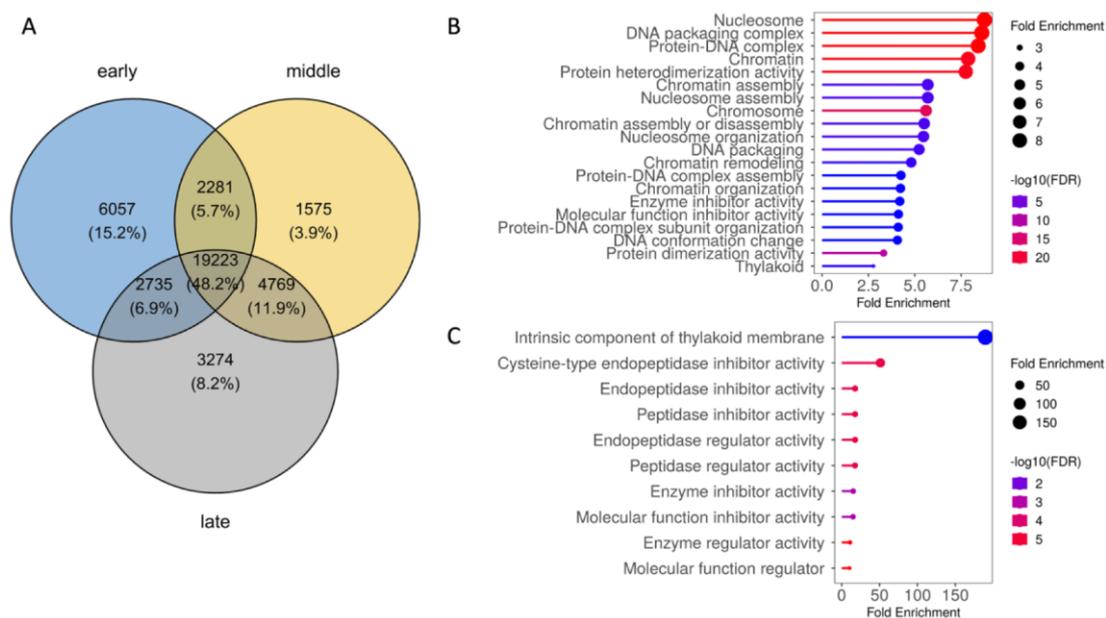


Рисунок 27 – Характеристика транскриптов, экспрессирующихся на ранней, средней и поздней стадиях развития зерновки. (А) Диаграмма Венна генов, экспрессирующихся (TPM > 2) на определенных стадиях развития зерновки тритикале: ранней (10 dpa), средней (15 dpa) и поздней (20 dpa). (В) и (С) Анализ обогащения аннотированных генов, экспрессирующихся на ранней и средней/поздней стадиях развития зерновок, соответственно.

Среди специфически экспрессирующихся на ранней стадии генов в значительной мере ($p\text{-value} < 0.0001$) представлены гены, участвующие в организации, упаковке и ремоделировании хроматина (рисунок 27В). Это отражает интенсивную транскрипцию в процессе дифференцировки клеток на ранней стадии развития зерновок, когда происходит дифференцировка зародыша. Оценка генов, экспрессирующихся на средней и поздней стадиях развития зерновки, показала схожие закономерности: наиболее значительное обогащение было обнаружено для генов, связанных с ингибиторами пептидаз и катаболизмом глюкозы.

Транскриптомное исследование демонстрирует экспрессию генов, участвующих в негативной регуляции активности пептидаз на поздних стадиях развития зерновки тритикале. При изучении роли протеаз в растениях необходимо понимать их высокую активность, которая может негативно сказаться на развитии зерновок и накоплении запасных белков (Rangan et al., 2017). Функционирование этих ферментов регулируется

внутриклеточными ингибиторами, называемыми фитоцистатинами. Для нормального развития зерновых культур необходимо поддерживать баланс между уровнями цистеиновых протеаз и фитоцистатина, который поддерживается антагонистической активностью гиббереллинов (GAs) и абсцизовой кислоты (АВА), регулирующих экспрессию генов, непосредственно участвующих в этих биологических процессах.

Растительные цистеиновые протеазы являются важными элементами в процессе деградации и мобилизации запасных белков злаков (Mikkonen, 1996, Prabucka, 2004, Shi, 2009). Фитоцистатины, в свою очередь, участвуют в регуляции обмена белка во время наполнения зерна или контроле протеолиза во время развития и/или прорастания и наполнения зерновок злаков. Фитоцистатины делятся на 3 основные группы, но наиболее характерной является группа А, ингибиторы которой участвуют в регуляции накопления и мобилизации. Была отмечена специфическая корреляция между паттернами экспрессии генов, кодирующих цистатины ячменя, риса, пшеницы и тритикале в кластере А, и их функциональной активностью в качестве ингибиторов протеаз. Анализ этих цистатинов выявил их наибольшую экспрессию в сухих и прорастающих зернах, и, соответственно, этот факт позволяет предположить, что они могут быть специализированными эндогенными регуляторами ферментов, которые участвуют в накоплении запасяющих белков во время их прорастания.

Для цистатинов пшеницы *WC1-4* (Kuroda et al., 2014) и тритикале *TrcC-1* и *TrcC-4* (Szewińska et al., 2012) наблюдалась тенденция к увеличению экспрессии генов в начале первых 2 недель развития зерна с последующим ее снижением по мере продолжения процесса развития. Различная роль этих ингибиторов подтверждается наличием различных профилей экспрессии во время развития, созревания и прорастания зерновок. Синтез запасных белков в течение формирования зерновок защищен высоким уровнем *TrcC-4* на 10-13-й день после цветения, а его экспрессия на ранней стадии развития ингибирует экспрессию основного фермента *EP8*, участвующего в

прорастании зерновок тритикале (Szewińska et al., 2016), поскольку *TrcC-4* может защитить от неконтролируемого гидролиза запасных белков в прорастающих зерновках тритикале.

Основными представителями фитоцистатинов подгруппы C1 являются фитоцистатины тритикале (*TrcC-7*) (Simińska et al., 2015), ячменя (*HvCPI-6* и *8*), риса (*OC-VI* и *VIII*) (Martínez et al., 2005) и кукурузы (*CC6* и *CC7*) (Massonneau et al., 2005). Основываясь на исследовании Martínez et al. (2009), фитоцистатины ячменя, *HvCPI-6* и *HvCPI-8*, являющиеся продуктивными ингибиторами протеаз C1A, экспрессируются главным образом в зерновках и развивающемся эндосперме. Самые высокие уровни транскриптов *TrcC-7* тритикале и фитоцистатина *CC6* и *CC7* кукурузы, которые регулируют накопление белка и развитие эндосперма, могут быть выявлены на средних стадиях развития зерновки (Simińska, 2015, Massonneau, 2005).

Считается, что фитоцистатины подгруппы C2 экспрессируются на очень низких уровнях, и, соответственно, продукты этих генов также обладают низкой ингибирующей способностью или вообще не способны ингибировать большинство эндогенных активностей цистеиновой протеазы. Однако некоторые исследования показали, что экспрессия гена, кодирующего *WC5*, может происходить на ранних стадиях развития зерновки (Corre-Mengué, 2002), а именно на 8-10-й день после цветения (dra), когда эндосперм дифференцируется на алейроновый слой и крахмалистый эндосперм. Но максимальный уровень экспрессии *WC5* достигается во фракциях, соответствующих зародышу и алейроновому слою. В течение этого периода окружающая клетчатка подвергается дегенеративному процессу, что означает увеличение экспрессии протеолитических ферментов, а также разрушение плазматической мембраны (Domínguez, 1998). Соответственно, *WC5* выполняет функцию защиты алейронового слоя или зародыша от интенсивной протеолитической активности, происходящей в окружающих тканях. Переходная стадия между дифференцировкой эндосперма и созреванием, которая происходит у пшеницы между 12 и 17

дра, означает накопление запасных белков и крахмала (Young, 2000). В это время, приблизительно при 13 дра, белок *WC5* может быть обнаружен вблизи слоя алейрона, следовательно, можно предположить, что этот белок участвует в дифференцировке эндосперма и может контролировать накопление запасных белков. Аналогичную структуру пшеницы *WC5* имеет фитоцистатин *TrcC-5* тритикале, который экспрессируется до 17 дра. В свою очередь, этот цистатин контролирует протеолиз во время развития колоса и накопление запасных белков. Стоит отметить, что мРНК, соответствующая *WC5* или *TrcC-5*, не была обнаружена в побегах, корнях, прорастающих зерновках или молодых растениях, что указывает на то, что фитоцистатины подгруппы C2 также принимают активное участие в регуляции протеолиза во время развития зерновок.

Кроме того, было обнаружено, что гены, кодирующие белки, расположенные в апопласте (поздняя стадия), и белки, являющиеся компонентами тилакоидной мембраны (средняя стадия), чрезмерно представлены на средней/поздней стадиях развития (Рисунок 27С).

Анализ транскриптома развивающихся зерновок яровой тритикале на ранней и средней/поздней стадиях сильно отличаются с точки зрения экспрессируемых генов и их функций.

3.3.3. Особенности транскриптома зародыша и эндосперма развивающейся зерновки

Целью данной части работы было обнаружение различий между наборами генов, экспрессирующихся в разных частях зерновки и в процессе ее развития. Для этого было выполнено ONT-секвенирование кДНК, полученной из РНК, выделенной из двух частей развивающихся зерновок: зародыш + эндосперм (EE) и эндосперм (En). Были проанализированы данные секвенирования для зерновок на 15 (EE_15 и En_15) и 20 день после цветения (EE_20 и En_20) (рисунок 28А).

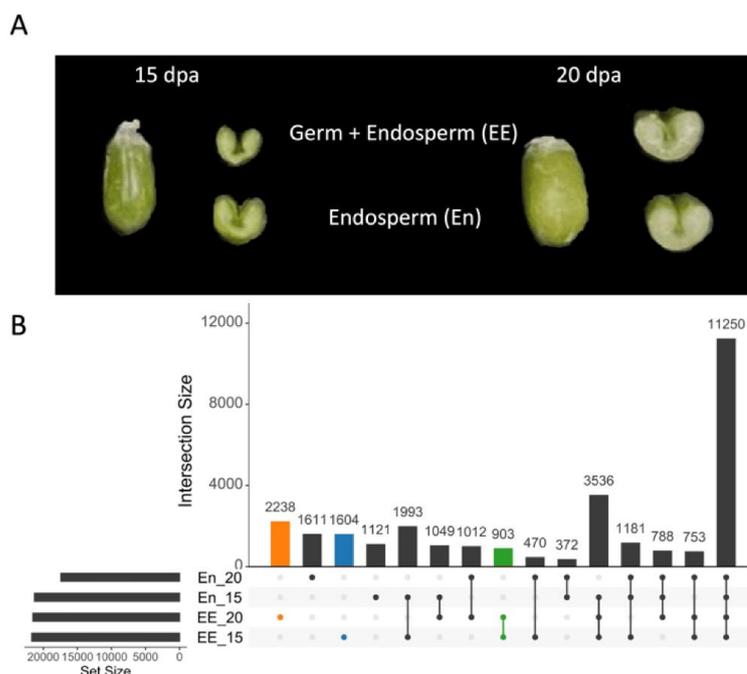


Рисунок 28 – Идентификация генов, экспрессирующихся в зародыше и эндосперме зерновок. (А) Вид целых и препарированных зерновок на 15 и 20 день после цветения, используемых для выделения РНК. (В) График смещения, показывающий пересечение генов, экспрессирующихся в четырех образцах. (EE – зародыш + эндосперм, En - эндосперм)

Используя ограничение значения уровня экспрессии $TPM > 2$, было определено, что 11250 генов экспрессировались во всех образцах (рисунок 28В). Кроме того, было обнаружено 1604, 2238 и 903 гена, которые экспрессировались в EE_15, EE_20 и в обоих образцах, соответственно. Эти гены не имели достаточной экспрессии в образцах эндосперм+зародыш, а

значит они могут отражать транскриптом зародыша на 15 и 20 день после цветения. Анализ обогащения EE_15 генов выявил наборы генов, кодирующих белки комплекса оксидоредуктазы и комплекса мембранных белков, которые вовлечены в процессы фосфорорганического метаболизма. Наиболее широко в наборе генов EE_20 были представлены гены, кодирующие белки рибонуклеопротеинового комплекса, комплекса убиквитинлигазы, и вовлеченные в сайленсинг генов, биогенез рибосом, процессинг РНК и трансляцию (коэффициент обогащения $FDR < 0,003$).

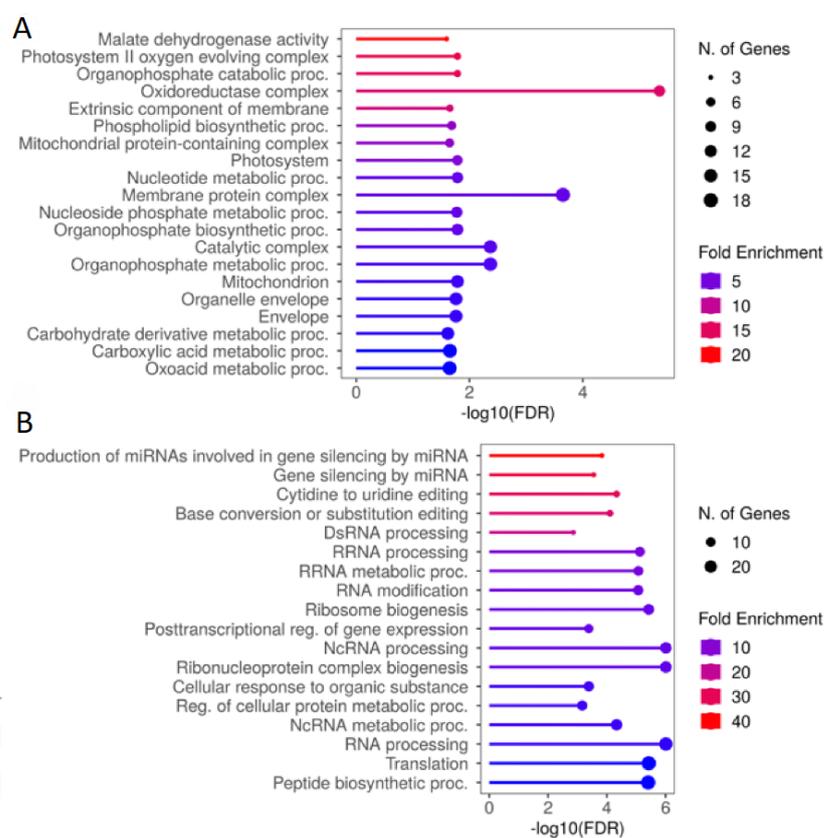


Рисунок 29 – Анализ обогащения аннотированных генов, экспрессирующихся в EE_15 (A) и EE_20 (B).

Среди генов, специфически экспрессирующихся на 15 день после цветения в образцах эндосперм+зародыш (EE), в значительной мере ($p\text{-value} < 0.0001$) представлены гены, участвующие в оксидоредуктазном комплексе (рисунок 29A). Биохимические свойства оксидоредуктазы были впервые исследованы при выделении белка из зрелой зерновки пшеницы (*Triticum aestivum* L.). По субстратной специфичности и ферментативным

характеристикам (оптимум pH и температуры) оксидоредуктаза сходна с аналогичными ферментами тканей животных (Osipova et al., 2005). Предполагаемая физиологическая роль оксидоредуктазы заключается в участии и регуляции SH/SS-статуса белков эндосперма пшеницы и формировании реологических свойств клейковины.

Кроме того, были обнаружены гены, кодирующие мембранный протеиновый комплекс в EE_15. Около трети всех генов, кодирующих белки в растениях и высших организмах, кодируют мембранные белки, и они являются самым большим классом мишеней из-за их фундаментальных функций в клетке в качестве ферментов, транспортеров, ионных каналов и рецепторов (Arginaminpathy et al., 2009). Несмотря на то, что мембранные белки играют решающую роль в различных процессах, их биологическая характеристика для клеток зародыша остается перспективной задачей.

Среди генов специфически экспрессирующихся на 20 день после цветения в тканях зародыш+эндосперм (EE_20) были представлены гены, отвечающие за различные биологические процессы (рисунок 29B). Представлены гены EE_20, вовлеченные в биогенез рибонуклеопротеинового комплекса, процессинг рРНК и биогенез рибосом, который имеет решающее значение для роста растений и акклиматизации. Являясь важным сложным энергоемким процессом (Warner, 1999), биогенез рибосом строго регулируется эндогенными сигналами (Lykke-Andersen et al., 2009; Lafontaine, 2010; Sanchez et al., 2016) и реакцией на изменение окружающей среды (Sinturel et al., 2017), например, температуру (Ohbayashi et al., 2017). В эукариотических клетках аберрантный биогенез рРНК активирует контроль качества РНК в ядре, что запускает более высокое полиаденилирование определенных промежуточных продуктов рРНК и побочных продуктов, катализируемых комплексом полиаденилирования *Trf/Air/Mtr4* (TRAMP; Jia et al., 2011; Lange et al., 2014). Эти промежуточные продукты последовательно разлагаются ядерным экзосомным комплексом (Losh, van Hoof, 2015; Thoms et al., 2015). Дисфункция рибосомального биогенеза

(Ferreira-Cerca et al., 2005, 2007; Tafforeau et al., 2013) приводит к серьезным дефектам развития высших растений (Horiguchi et al., 2011; Weis et al., 2015a, 2015b). Процессинг рибосомных РНК (рРНК) является важным этапом биогенеза рибосом и начинается с транскрипции рДНК. Полученный транскрипт предшественника рРНК (пре-рРНК) подвергается систематической обработке, при которой множественные эндонуклеолитические и экзонуклеолитические расщепления удаляют внешние и внутренние транскрибируемые спейсеры (*ETS* и *ITS*). Сайты процессинга и пути для процессинга пре-рРНК были расшифрованы у *Saccharomyces cerevisiae* и, в некоторой степени, у *Xenopus laevis*, клеток млекопитающих и арабидопсиса (*Arabidopsis thaliana*). В отличие от модельного двудольного вида *Arabidopsis*, созревание рРНК в однодольных культурах остается неизученным. Недавняя работа на рисе (*Oryza sativa*) показала, что РНК-геликаза *TOGR1* (Thermotolerant Growth Required1), гомолог риса *Rrp3* (rRNA-processing protein 3) у *S. cerevisiae* (O'Day et al., 1996) и *DDX47* у *Homo sapiens* (Sekiguchi et al., 2018) необходима для термоустойчивого роста риса и действует в качестве ключевого компонента для гомеостаза рРНК (Wang et al., 2017). Это подчеркивает важность биогенеза рибосом в процессе развития злаковых культур и температурной акклиматизации.

Хорошо известно, что эндосперм представляет собой гипометилированную ткань, и метилирование ДНК важно для накопления запасных белков и биогенеза эндосперма (Zemach et al., 2010). Однако пониженное метилирование ДНК способствует активации ТЕ, создавая опасную среду для зародыша, поскольку активация ТЕs может привести к летальным мутациям и снижению жизнеспособности зерновок. Ранее был идентифицирован LTR ретротранспозон (ТЕ) MIG Ty1/Copia, экспрессирующийся на ранней стадии развития зерновок, хотя, в какой именно ткани экспрессируется MIG, было неизвестно (Kirov et al., 2020). Используя профилирование транскриптома для зародыша и эндосперма,

было показано, что MIG экспрессируется, скорее всего, в эндосперме или одновременно в тканях эндосперма и зародыша. В зародышевой части на поздней стадии развития зерновок выявлена сниженная представленность TE-РНК по сравнению со средней стадией. В то же время в наборе генов, экспрессирующихся на этой стадии, представлены гены транскрипционного сайленсинга. Эти результаты могут свидетельствовать о том, что экспрессия TE на средней стадии развития зерновок запускает активацию посттранскрипционного сайленсинга генов на более поздних стадиях развития зерновок. Последнее может приводить к метилированию TE и защите зародыша от инсерций TE. Таким образом, наши данные согласуются с гипотезой о том, что гипометилирование ДНК эндосперма играет защитную роль для зародыша (Zemach et al., 2010, Hsieh et al., 2009).

Сравнение доли генов, кодирующих днРНК, экспрессирующихся в зародышевой части на 15 и 20 день после цветения продемонстрировало, что гены днРНК чрезмерно представлены среди генов, экспрессирующихся в зародыше на 15 день после цветения по сравнению с 20 днём после цветения (462 из 1604 против 376 из 2238, критерий χ^2 Пирсона с поправкой Йейтса, p -value $< 2,2e^{-16}$).

3.3.4. Полиморфизм генов днРНК в коллекции тритикале

днРНК представляют собой интересную мишень для селекции с помощью маркеров. Поэтому на следующем этапе работы мы выделили некоторые днРНК, для которых предсказали InDels. Чтобы подтвердить, проявляют ли обнаруженные гены полиморфизм в коллекции 23 генотипов тритикале различного происхождения, мы подобрали праймеры на выявленные InDels и показали наличие различных аллелей днРНК в исследуемой коллекции (таблица 10).

Таблица 10 – Коллекционные образцы яровой тритикале

№	Название	Происхождение
1	Dublet	Польша
2	Legalo	Польша
3	Sandro	Швейцария
4	Гребешок	Россия
5	Лана	Белоруссия
6	Памяти Мережко	Россия-Белоруссия
7	Укро	Украина-Россия
8	Ульяна	Белоруссия
9	Хлебодар харьковский	Украина
10	Ярило	Россия
11	131/1656	Россия
12	131/7	Россия
13	6-35-5	Россия
14	C95	Россия
15	C238	Россия
16	C245	Россия
17	C259	Россия
18	V 17/50	Россия
19	Л8665	Россия
20	П 2-16-20	Россия
21	П 13-5-2	Россия
22	П 13-5-13	Россия
23	ПРАГ 551	Россия

Для анализа были выбраны гены днРНК, расположенные в геномах А и В. Был проведен BLAST-поиск по геномам пшеницы (Karusta et al., 2013), отобраны гены днРНК, обладающие полиморфизмом InDel, и были сконструированы праймеры для семи генов днРНК (таблица 14). ПЦР-анализ коллекций тритикале с использованием разработанных праймеров выявил различия в размерах ПЦР-продуктов для семи пар праймеров и геномной ДНК разных генотипов тритикале (рисунок 30).

Таблица 11 – Праймеры, используемые для анализа полиморфизма генов днРНК.

Название праймера	Последовательность 5'- 3'	Длина ожидаемого продукта, п.н.
1A2902	CCATGATTGAAGATGAATTAGATCAG	234
	GATATGCCGGGGTGTACTG	
1A59	TGATTGAAGATGAATTAGATCAGAAGT	208
	TATGAGCGACGACATCTGCC	
3B61	TTTGTTGTTGTCGCACGAGC	302
	ACCTTGATTATGTGGGCCCG	
3B75	CCGTGTTGCTGCACAGAAAT	162
	CCAGAAAAGAAAAGGACAGGCA	
3B82	GACCGACATTGTGACTCCGT	298
	ACACCCAAACAGAGAGGAGA	
6A12	GCACATGTGACGTAGGGACA	168
	TGCAACTATCATAGGGTGTGTGT	
7B70	GTGAAGGGGGTTGGACTCAC	200
	TGTTTTTCGTAGTTTGCACCCA	

Несмотря на то, что при разработке праймеров контролировались специфичность и образование димеров с помощью программного обеспечения, необходимым этапом являлась дополнительная проверка специфичности реакции и последующая оптимизация температуры отжига и концентрации праймеров. В ходе этого была подобрана универсальная температура отжига праймеров, при которой наблюдались целевые продукты

и отсутствие неспецифических продуктов реакции при постановке ПЦР. Анализ результатов проводили с помощью агарозного гель-электрофореза.

В результате проведения ПЦР со специфичными парами праймеров 1A59 и 7B70 амплифицируется продукт ожидаемой длины ~ 200 п.н. Полученные результаты свидетельствуют о том, что у 2 образцов (Памяти Мережко, С95) при использовании праймера 1A59 и 7 образцов с праймером 7B70 (Dublet, Гребешок, Хлебодар харьковский, 131/1656, 131/7, П 13-5-2, ПРАГ 551) наблюдается продукт большего размера относительно амплифицированного на референсной ДНК-матрице (Chinese Spring).

ПЦР со специфичными парами праймеров 6A12 и 3B75 выявляет продукт ожидаемой длины ~ 160 п.н. Однако у линий 131/1656, С95, П-13-5-13 и ПРАГ 551 наблюдается одновременное присутствие двух фрагментов, различных по размеру, что позволяет предполагать о их гетерозиготности по участку, фланкируемому парой праймеров 6A12. Полученные результаты при использовании праймеров 3B75 свидетельствуют о том, что для линий 6-35-5, С95 и С238 наблюдается больший размер продукта относительно амплифицированного на референсной ДНК-матрице (Chinese Spring).

При использовании специфичных пар праймеров 3B61 и 3B82 в результате проведения ПЦР амплифицируется продукт ожидаемой длины ~ 300 п.н. Полученные результаты при амплификации праймеров 3B61 свидетельствуют о том, что для сорта Dublet и линии С245 наблюдается меньший размер продукта относительно амплифицированного на референсной ДНК-матрице (Chinese Spring). У линий 6-35-5 и С259 наблюдается единовременное присутствие двух фрагментов, различных по размеру, что позволяет предполагать о их гетерозиготности по участку, фланкируемому парой праймеров 3B61. По полученным результатам с праймером 3B82 наглядно видно отсутствие фрагментов у сорта Legalo и линий С95, П 2-16-20 и П 13-5-2.

В результате проведения ПЦР со специфичной парой праймеров 1A2902 амплифицируется продукт ожидаемой длины ~ 234 п.н. Полученные

результаты свидетельствуют о том, что для сорта Памяти Мережко и линии С95 наблюдается больший размер продукта относительно амплифицированного на референсной ДНК-матрице (Chinese Spring).

Анализ амплифицированных ПЦР-продуктов показал, что наиболее дивергентным вариантом была линия С95 ввиду того, что при амплификации семи пар праймеров отклонение от референсной ДНК-матрицы (Chinese Spring) наблюдалось в пяти случаях (1А59, 6А12, 3В75, 3В82, 1А2902).

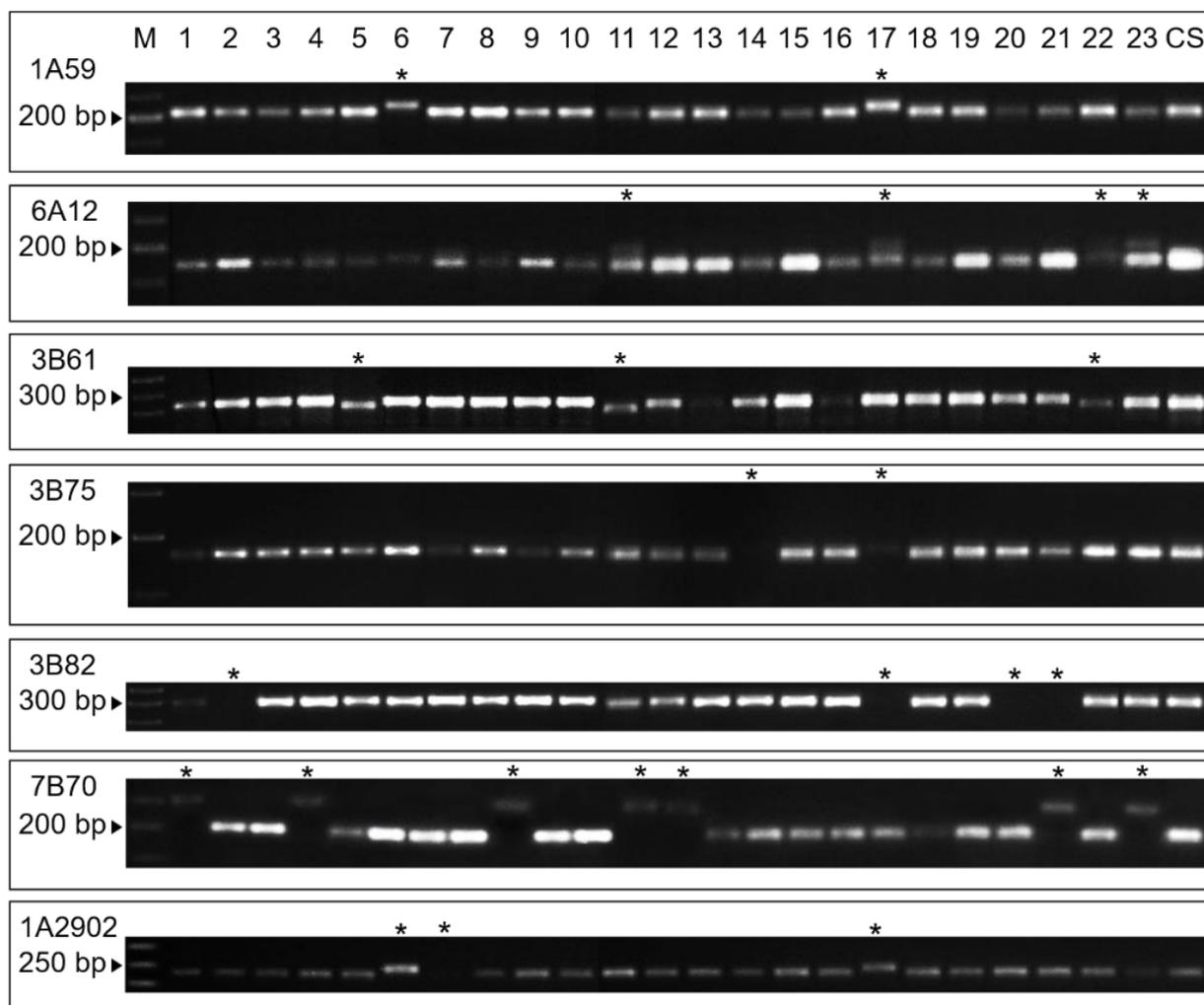


Рисунок 30 – Результаты анализа полиморфизма генов днРНК в коллекции яровой тритикале: Dublet (1); Legalo (2); Sandro (3); Гребешок (4); Лана (5); Памяти Мережко (6); Укро (7); Ульяна (8); Хлебодар харьковский (9); Ярило (10); 131/1656 (11); 131/7 (12); 6-35-5 (13); С238 (14); С245 (15); С259 (16); С95 (17); V17/50 (18); Л8665 (19); P2-16-20 (20); P 13-5-2 (21); P13-5-13 (22); PRAG 551 (23); Chinese Spring (CS). ДНК-маркер Step 50 plus (Biolabmix, Новосибирск, Россия) (M). Нереферентные аллели обозначены звездочкой (*).

ЗАКЛЮЧЕНИЕ

Секвенирование nCATS позволяет проводить эпигенетическое профилирование генов-мишеней у растений, что может помочь в решении важных задач, от разработки маркеров для селекции растений до изучения регуляции экспрессии генов. Это особенно важно для растений с большим размером генома, где секвенирование всего генома является трудоемким и дорогостоящим. Однако невысокая глубина секвенирования означает, что для лучшего охвата целевых фрагментов требуется большее количество запусков, что делает этот метод пригодным только для анализа одного сорта.

Альтернативным методом секвенирования является подход, основанный на ПЦР-амплификации. Этот метод особенно привлекателен при оценке вариаций, как в кодирующей, так и в промоторной областях гена глютеенинов. Однако наличие этапа амплификации делает невозможным оценку профиля метилирования. С помощью ONT Amplicon-Seq за короткое время можно легко обнаружить различные вариации генов у разных растений и в больших коллекциях. Это позволяет точно идентифицировать известные и обнаруживать новые аллели, характеризующиеся мутациями в кодирующих и промоторных областях. Результаты, полученные в нашем исследовании, дают полезную информацию для выбора желательных комбинаций NMW-GS для селекции пшеницы.

Данные секвенирования кДНК и РНК на разных стадиях развития зерновки позволяют выявить существенные различия в динамике транскриптома, что является ценной информацией для улучшения функциональной и структурной аннотации генома. Этот анализ позволил выявить ряд генов, которые не были аннотированы как транскрипционно активные, но могут быть вовлечены в различные биологические процессы. Таким образом, применение целевого секвенирования (ONT Amplicon-Seq) является основным инструментом для быстрого анализа полиморфизма выявленных генов.

ВЫВОДЫ

1. Использование метода nCATS позволяет получить значительное обогащение (до 200х) ридами для целевых генов (*Glu-1Ax*, *Glu-1Bx* и *Glu-1By*) и их промоторов у гексаплоидной тритикале, предоставляя уникальную информацию для детекции структурных вариаций и профиля метилирования, но данный метод не подходит для высокопоточного секвенирования генов у видов с большими геномами.

2. Профилирование метилирования ДНК показало, что гены глютенинов имеют значительно более высокую плотность метилирования цитозинов в кодирующей области по сравнению с промоторным регионом.

3. С помощью ONT Amplicon-Seq охарактеризован полиморфизм и определены аллели шести полноразмерных генов глютенинов (*Glu-1Ax*, *Glu-1Ay*, *Glu-1Bx*, *Glu-1By*, *Glu-1Dx*, *Glu-1Dy*) у 23 сортов пшеницы, идентифицирован новый аллель (*Glu-A1x1-T*) гена *Glu-1Ax* и показано, что ONT Amplicon-Seq является точным и быстрым методом генотипирования полноразмерных генов у растений с большими и сложными геномами.

4. С помощью нанопорового секвенирования кДНК было выявлено 39914 генов, которые экспрессируются на ранней (6057), средней (1575), поздней (3274) стадиях развития зерновки, и 19223 гена, которые экспрессируются на всех стадиях развития зерновки, а также 1604 экспрессировались в образце с зародышевой частью зерновки на 15 дпа, 2238 – в образце с зародышевой частью зерновки на 20 дпа.

5. Нанопоровое секвенирование кДНК позволило идентифицировать 7128 генов как ранее неаннотированные, локализованных в А- (2340), В- (2947) и R- (1847) субгеномах, из которых 3174 (более 40%) кодируют днРНК, что свидетельствует о их значимой роли в процессе развития зерновки.

6. Гены, кодирующие длинные некодирующие РНК, полиморфны в коллекции яровой тритикале и линия С95 является наиболее дивергентной.

СПИСОК ЛИТЕРАТУРЫ

1. Абделькави Р.Н.Ф., Турбаев А.Ж., Соловьев А.А. Технологические свойства зерна яровой тритикале в условиях ЦРНЗ. // Вестник Курской государственной сельскохозяйственной академии. – 2020. – № 5. – С. 87-97.
2. Воротынцева М. В. Степень влияния глютеинов на качество зерна как одного из сложных полигенных признаков рода *Triticum* (обзор) //Труды по прикладной ботанике, генетике и селекции. – 2021. – Т. 182. – №. 1. – С. 168-185.
3. Грабовец А.И., Крохмаль А.В. Тритикале. Ростов-на-Дону, 2018
4. Груздев И.В., Дудников М.В., Соловьев А.А. Идентификация высокомолекулярных субъединиц глютеинов у некоторых образцов яровой тритикале методом SDS-PAGE / В книге: Биотехнология в растениеводстве, животноводстве и сельскохозяйственной микробиологии. Сборник тезисов докладов 20-й Всероссийской конференции молодых учёных, посвященной памяти академика РАСХН Георгия Сергеевича Муромцева. Москва, 2020. С. 93-95.
5. Гуляев Г.В. Селекция озимых тритикале в Польше / Г.В. Гуляев // Селекция и семеноводство. – 1998. - № 2. – С. 55-57.
6. Дивашук М.Г.; Крупин П.Ю.; Белов И.П.; Карлов Г.И. Молекулярно-генетический анализ состава запасных белков в образцах коллекции пшенично-пырейных гибридов // Известия Тимирязевской сельскохозяйственной академии. - 2012. - №5
7. Дивашук М.Г., Крупин П.Ю., Соловьев А.А., Карлов Г.И. Молекулярно-цитогенетическая характеристика линии яровой тритикале 131/7, несущей ржано-пшеничную транслокацию. // Генетика. – 2010а. – Т. 46. – № 2. – С. 211-217.

8. Дивашук М.Г., Соловьев А.А., Карлов Г.И. Влияние отбора по фенотипическим признакам на хромосомную конституцию яровой тритикале. // Генетика. – 2010б. – Т. 46. – № 3. – С. 383-388.
9. Казарцева А.Т. Систематизация признаков качества зерна в селекции озимой мягкой пшеницы / А.Т. Казарцева, Р.А. Воробьева, Н.В. Сокол // Сельскохозяйственная биология. – 1990. - № 5. – С.3.
10. Одинцова Т. И. Структурно-функциональные особенности запасных и защитных белков растений и их использование в генетических исследованиях : дис. – Институт общей генетики им. НИ Вавилова Российской академии наук, 2010.
11. Пыльнев В.В., Коновалов Ю.Б., Березкин А.Н., Малько А.М., Рубец В.С., Буко О.А., Долгодворова Л.И., Хупацария Т.И., Конорев П.М., Баженова С.С., Соловьев А.А. Практикум по селекции и семеноводству полевых культур. // Санкт-Петербург, 2022. Сер. Учебники для вузов. Специальная литература.
12. Пыльнев В.В., Соловьев А.А. Дополнения и изменения в классификации тритикале / В.В. Пыльнев, А.А. Соловьев / Тритикале России: материалы заседания секции тритикале Т67 РАСХН. – Ростов-на-Дону, 2008. – С. 179-181.
13. Пырсигов А.С., Сыксин С.В., Милукова Н.А., Груздев И.В., Полховская Е.С., Гарибян Ц.С., Соловьёв А.А. Применение нового молекулярного маркера и метода SDS-page для установления взаимосвязи аллельного состава генов глютеенинов с качественными показателями зерна тритикале / Кормопроизводство. 2022. № 9. С. 27-33.
14. Сокол Н. В. Эколого-генетический анализ изменчивости признаков качества зерна / Н.В. Сокол, А.Т. Казарцева, В.А. Драгавцев, Ф.А. Колесников // Селекция и семеноводство. – 1993. - № 4. – С.28.
15. Соловьев А.А. Изучение формообразовательного процесса при скрещивании различных форм тритикале. // автореферат дис. ... кандидата биологических наук / Москва, 2000

16. Соловьев А.А., Вишнякова Х.С. Влияние 2R/2D-замещения на проявление некоторых признаков у гибридов F1 тритикале. // В сборнике: Доклады ТСХА. Москва, 1997. С. 3-8.
17. Шишкина А.А., Бадаева Е.Д., Соловьев А.А. Особенности организации кариотипов некоторых форм яровой тритикале. // Известия Тимирязевской сельскохозяйственной академии. – 2009. – № 2. – С. 123-130.А
18. Agarwal P., Kapoor S., Tyagi A. K. Transcription factors regulating the progression of monocot and dicot seed development //Bioessays. – 2011. – Т. 33. – №. 3. – С. 189-202.
19. Ahsan, M.U.; Liu, Q.; Fang, L.; Wang, K. NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks. *Genome Biol.* 2021, 22, 261.
20. Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 12: R18.
21. Akpinar B.A., et al. Chromosome-based survey sequencing reveals the genome organization of wild wheat progenitor *Triticum dicoccoides* //Plant biotechnology journal. – 2018. – Т. 16. – №. 12. – С. 2077-2087.
22. Albani D. et al. The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes //The Plant Cell. - 1997. - Т. 9. - №. 2. - С. 171-184.
23. Alkan C., Coe B. P., Eichler E. E. Genome structural variation discovery and genotyping //Nature reviews genetics. – 2011. – Т. 12. – №. 5. – С. 363-376.
24. Alonge M. et al. Chromosome-scale assembly of the bread wheat genome reveals thousands of additional gene copies //Genetics. – 2020. – Т. 216. – №. 2. – С. 599-608.

25. Altenbach S. B., Kothari K. M. Transcript profiles of genes expressed in endosperm tissue are altered by high temperature during wheat grain development //Journal of Cereal Science. - 2004. - T. 40. - №. 2. - C. 115-126.
26. Altschul S. F., Lipman D. J. Protein database searches for multiple alignments //Proceedings of the National Academy of Sciences. – 1990. – T. 87. – №. 14. – C. 5509-5513.
27. Ammar K., Kronstad W. E., Morris C. F. Breadmaking quality of selected durum wheat genotypes and its relationship with high molecular weight glutenin subunits allelic variation and gluten protein polymeric composition //Cereal Chemistry. - 2000. - T. 77. - №. 2. - C. 230-236.
28. Amor B.B., et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses //Genome research. – 2009. – T. 19. – №. 1. – C. 57-69.
29. An X., Li Q., Yan Y., Xiao Y., Xsam S.L.K., Zeller F.J. Genetic diversity of European spelt wheat (*Triticum aestivum* ssp. *Spelta* L. Em. Thell.) Revealed by glutenin sub-unit variations at the Glu-1 and Gli-3 loci. *Euphytica*. 2005;146(3):193-201. DOI: 10.1007/s10681-005-9002-6
30. Andersen J.R., Lübberstedt T. Functional markers in plants // Trends in Plant Science. – 2003. – T.8. – №11. – C 554-560;
31. Anderson O. D., Greene F. C. The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat //Theoretical and Applied Genetics. – 1989. – T. 77. – C. 689-700.
32. Anjum F.M., Khan M.R., Din A., Saeed M., Pasha I., Arshad M.U. Wheat gluten: high molecular weight glutenin subunits – structure, genetics, and relation to dough elasticity. *Journal of Food Science*. 2007;72(3):56-63. DOI: 10.1111/j.1750-3841.2007.00292.x
33. Ardui S. et al. Detecting AGG interruptions in male and female FMR1 premutation carriers by single-molecule sequencing //Human mutation. – 2017. – T. 38. – №. 3. – C. 324-331.

34. Arinaminpathy Y. et al. Computational analysis of membrane proteins: the largest class of drug targets //Drug discovery today. – 2009. – T. 14. – №. 23-24. – C. 1130-1135.
35. Aury J. M. et al. Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding //Gigascience. – 2022. – T. 11. – C. giac034.
36. Bai B., Wang L., Lee M., Zhang Y., Rahmadsyah, Alfiko Y., Ye B.Q., Wan Z.Y., Lim C.H., Suwanto A., Chua N.-H., Yue G.H. Genome wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol. 2017; 17 (1):93.
37. Barabaschi D. et al. Next generation breeding //Plant Science. – 2016. – T. 242. – C. 3-13.
38. Barnett D. W. et al. BamTools: a C++ API and toolkit for analyzing and managing BAM files //Bioinformatics. – 2011. – T. 27. – №. 12. – C. 1691-1692.
39. Bartels D., Thompson R. D. Synthesis of mRNAs coding for abundant endosperm proteins during wheat grain development //Plant Science. – 1986. – T. 46. – №. 2. – C. 117-125.
40. Belmonte M. F. et al. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed //Proceedings of the National Academy of Sciences. – 2013. – T. 110. – №. 5. – C. E435-E444.
41. Bennett M. D., Smith J. B. Nuclear DNA amounts in angiosperms //Philosophical Transactions of the Royal Society of London. B, Biological Sciences. – 1976. – T. 274. – №. 933. – C. 227-274.
42. Bennett-Baker P. E., Mueller J. L. CRISPR-mediated isolation of specific megabase segments of genomic DNA //Nucleic acids research. – 2017.
43. Bevan M. Et al. Transcriptional control of plant storage protein genes //Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. – 1993. – T. 342. – №. 1301. – C. 209-215.

44. Bewick, A.J.; Schmitz, R.J. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* 2017, 36, 103–110.
45. Bietz, J.A.; Shepherd, K.W.; Wall, J.S. Single-Kernel Analysis of Glutenin: Use in Wheat Genetics and Breeding. *Cereal Chem.* 1975, 52, 513–532.
46. Bilecen K. et al. Triticum durum metallothionein: isolation of the gene and structural characterization of the protein using solution scattering and molecular modeling // *Journal of Biological Chemistry.* - 2005. - T. 280. - №. 14. - C. 13701-13711.
47. Bishara A, Liu Y, Weng Z, Kashef-Haghighi D, Newburger DE, West R, Sidow A, Batzoglou S. 2015. Read clouds uncover variation in complex regions of the human genome. *Genome Res* 25: 1570–1580.
48. Blum, A. (2014). The abiotic stress response and adaptation of triticale—A review. *Cereal Res. Commun.* 42, 359–375. Doi: 10.1556/CRC.42.2014.3.1
49. Boža V., Brejová B., Vinař T. DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads // *PloS one.* – 2017. – T. 12. – №. 6. – C. e0178751.
50. Branlard G., Dardevet R., Saccomano F., Lagoutte F., Gourdon J. Genetics diversity of wheat storage proteins and bread wheat quality. *Euphytica.* 2001;119(1-2):59-67. DOI: 10.1023/A:1017586220359
51. Branlard, G.; Dardevet, M.; Saccomano, R.; Lagoutte, F.; Gourdon, J. Wheat in a Global Environment, Proceedings of the 6th International Wheat Conference, 5–9 June 2000, Budapest, Hungary. *Dev. Plant Breed.* 2001, 9, 157–169. https://doi.org/10.1007/978-94-017-3674-9_18.
52. Buels R. et al. JBrowse: a dynamic web platform for genome visualization and analysis // *Genome biology.* – 2016. – T. 17. – №. 1. – C. 1-12.
53. Caballero L., Martin L.M., Alvarez J.B. Relationships between the HMW- and LMW-glutenin subunits and SDS-sedimentation volume in Spanish hulled wheat lines // *Czech Journal of Genetics and Plant Breeding.* – 2008. – T. 44. – №3. – C. 114-117.

54. Cagirici H. B., Alptekin B., Budak H. RNA sequencing and co-expressed long non-coding RNA in modern and wild wheats //Scientific reports. – 2017. – T. 7. – №. 1. – C. 1-16.
55. Campalans A., Kondorosi A., Crespi M. Enod40, a short open reading frame-containing mrna, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula* //The Plant Cell. – 2004. – T. 16. – №. 4. – C. 1047-1059.
56. Cantacessi C. et al. A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing //Nucleic Acids Research. – 2010. – T. 38. – №. 17. – C. e171-e171.
57. Cao D. et al. Genotyping-by-sequencing and genome-wide association study reveal genetic diversity and loci controlling agronomic traits in triticale //Theoretical and Applied Genetics. – 2022. – T. 135. – №. 5. – C. 1705-1715.
58. Carrillo J.M., Rousset M., Qualset C.O., Kasarda D.D. Use of recombinant inbred lines of wheat for study of association of high-molecular-weight glutenin subunit alleles to quantitative traits: 1. Grain yield and quality prediction tests // Theoretical and Applied Genetics. – 1990. – T.79. – №3. – C. 321-330
59. Chekanova J.A., Long non-coding rnas and their functions in plants //Current opinion in plant biology. – 2015. – T. 27. – C. 207-216.
60. Chen R. et al. Isolation and characterization of genes encoding Myb transcription factor in wheat (*Triticum aestivum* L.) //Plant science. – 2005. – T. 169. – №. 6. – C. 1146-1154.
61. Chen J., Zeng B., Zhang M., Xie S., Wang G., Hauck A., and Lai J. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol.* 166: 252-264.
62. Cho Y. B., Jones S. I., Vodkin L. The transition from primary siRNAs to amplified secondary siRNAs that regulate chalcone synthase during development of *Glycine max* seed coats //PLoS One. – 2013. – T. 8. – №. 10. – C. e76954.

63. Chojecki AJS, Bayliss MW, Gale MD: Cell production and DNA accumulation in the wheat endosperm, and their association with grain weight. *Ann Bot.* 1986, 58 (6): 809-817.
64. Clarke B. C. et al. Genes active in developing wheat endosperm // *Functional & integrative genomics.* - 2000. - T. 1. - C. 44-55.
65. Clavijo B. J. et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations // *Genome research.* – 2017. – T. 27. – №. 5. – C. 885-896.
66. Cloutier S., Rampitsch C., Penner G.A., Lukow O.M. Cloning and expression of a LMW-i glutenin gene. // *Journal of Cereal Science.* – 2001. – T33. - №2. – C. 143-154.
67. Colot V. et al. Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco // *The EMBO journal.* - 1987. - T. 6. - №. 12. - C. 3559-3564.
68. Corre-Menguy F. et al. Characterization of the expression of a wheat cystatin gene during caryopsis development // *Plant Molecular Biology.* – 2002. – T. 50. – C. 687-698.
69. Crain J., Mondal S., Rutkoski J., Singh R.P., Poland J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome.* 2018; 11(1).
70. Csorba T., et al. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization // *Proceedings of the National Academy of Sciences.* – 2014. – T. 111. – №. 45. – C. 16160-16165.
71. Curaba J. et al. miRNA regulation in the early development of barley seed // *BMC plant biology.* – 2012. – T. 12. – C. 1-16.
72. Danecek, P.; McCarthy, S.A. BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics* 2017, 33, 2037–2039.

73. Dey M., et al. Phytohormonal responses in enod40-overexpressing plants of *Medicago truncatula* and rice // *Physiologia plantarum*. – 2004. – T. 120. – №. 1. – C. 132-139.
74. Dhaka V., Khatkar B. S. Effects of gliadin/glutenin and HMW-GS/LMW-GS ratio on dough rheological properties and bread-making potential of wheat varieties // *Journal of Food Quality*. - 2015. - T. 38. - №. 2. - C. 71-82.
75. Diaz I. Et al. The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed
Ding J., et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice // *Proceedings of the National Academy of Sciences*. – 2012. – T. 109. – №. 7. – C. 2654-2659.
76. Diesh C.; Stevens G.J.; Xie P.; Martinez T.D.J.; Hershberg E.A.; Leung A.; Guo E.; Dider S.; Zhang J.; Bridge C.; et al. JBrowse 2: A Modular Genome Browser with Views of Synteny and Structural Variation. *Genome Biol.* 2023, 24, 74. <https://doi.org/10.1186/s13059-023-02914-z>.
77. Dobrovolskaya O, Pont C, Sibout R, Martinek P, Badaeva E, Murat F, Chosson A, Watanabe N, Prat E, Gautier N et al. FRIZZY PANICLE drives supernumerary spikelets in bread wheat. *Plant Physiol.* 2015;167(1):189–199.
78. Domínguez Iribarren F. ETA: Estrategias organizativas y actuaciones (1978-1992). – Servicio Editorial de la Universidad del País Vasco/Euskal Herriko Unibertsitatearen Argitalpen Zerbitzua, 1998.
79. Dong G. Et al. Wheat Dof transcription factor WPBF interacts with taqm and activates transcription of an alpha-gliadin gene during wheat seed development // *Plant molecular biology*. – 2007. – T. 63. – №. 1. – C. 73-84.
80. Dong H. et al. High molecular weight glutenin genes: Effects on quality in wheat // *Crop science*. - 1991. - T. 31. - №. 4. - C. 974-979.
81. Du X.; Hu J.; Ma X.; He J.; Hou W.; Guo J.; Bo C.; Wang H.; Li A.; Kong L. Molecular Characterization and Marker Development for High Molecular Weight Glutenin Subunit 1Dy12** from Yunnan Hulled Wheat. *Mol. Breed.* 2018, 39, 4.

82. Dupont F. M. Et al. Differential accumulation of sulfur-rich and sulfur-poor wheat flour proteins is affected by temperature and mineral nutrition during grain development //Journal of Cereal Science. – 2006. – T. 44. – №. 1. – C. 101-112

83. Evers T, Millar S: Cereal grain structure and development: Some implications for quality. Journal of Cereal Science. 2002, 36 (3): 261-284. 10.1006/jcrs.2002.0435.

84. Food and Agriculture Organization Statistics (<https://www.fao.org/faostat/en/#data/QV>)

85. Fauteux F., Strömviik M. V. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae //BMC plant biology. – 2009. – T. 9. – №. 1. – C. 1-11.

86. Ferl R. J. 14-3-3 proteins and signal transduction //Annual review of plant biology. - 1996. - T. 47. - №. 1. - C. 49-73.

87. Ferreira-Cerca S. et al. Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins //Molecular cell. – 2007. – T. 28. – №. 3. – C. 446-457.

88. Ferreira-Cerca S. et al. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function //Molecular cell. – 2005. – T. 20. – №. 2. – C. 263-275.

89. Flavell R. B., Smith D. B. Variation in nucleolar organiser rRNA gene multiplicity in wheat and rye //Chromosoma. - 1974. - T. 47. - №. 3. - C. 327-334.

90. Forde B.G., Malpica J., Halford N., Shewry P., Anderson O., Greene F. and Miflin B. (1985) The nucleotide sequence of a HMW subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). Nucleic acids research, 13, 6817–6832.

91. Franke R. Crosses Involving W. Rimpau's Triticale. Plant Breeding 107, 303–308 (1991).

92. Fridman E.; Tiwari L.; Prusty M.; Bodenheimer S.; Doron-Faigenboim A.; Yamamoto E.; Kashkush K. Cytonuclear Diversity Underlying

Clock Adaptation to Warming Climate in Wild Barley (*Hordeum Vulgare* ssp. *Spontaneum*). 2023. Available online:

<https://www.authorea.com/users/561760/articles/625014-cytonuclear-diversity-underlying-clock-adaptation-to-warming-climate-in-wild-barley-hordeum-vulgare-ssp-spontaneum> (accessed on 17 February, 2023.).

93. Fulgosi H. et al. 14-3-3 proteins and plant development //Plant molecular biology. - 2002. - T. 50. - C. 1019-1029.

94. Furtado A. et al. A novel highly differentially expressed gene in wheat endosperm associated with bread quality //Scientific reports. – 2015. – T. 5. – №. 1. – C. 10446.

95. Gabrieli T. Et al. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH) //Nucleic acids research. – 2018. – T. 46. – №. 14. – C. E87-e87.

96. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477: 419–423.

97. Gao Y. et al. Transcriptomic analysis of rice (*Oryza sativa*) endosperm using the RNA-Seq technique //Plant molecular biology. – 2013. – T. 81. – C. 363-378.

98. Gayacharan Joel A. J. Epigenetic responses to drought stress in rice (*Oryza sativa* L.) //Physiology and Molecular Biology of Plants. – 2013. – T. 19. – C. 379-387.

99. Geng Y. et al. Expression of wheat high molecular weight glutenin subunit 1Bx is affected by large insertions and deletions located in the upstream flanking sequences //PloS one. - 2014. - T. 9. - №. 8. - C. e105363.

100. Gianibelli M.C., Larroque O., macritchie F., Wrigley C.W. Biochemical, genetic, and molecular characterization of wheat endosperm proteins. *Cereal Chemistry*. 2001;78(6):635-646. DOI: 10.1094/CCHEM.2001.78.6.635

101. Giesselmann P. Et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing //Nature Biotechnology. – 2019. – T. 37. – №. 12. – C. 1478-1481.
102. Gilpatrick T. Et al. Targeted nanopore sequencing with Cas9-guided adapter ligation //Nature biotechnology. – 2020. – T. 38. – №. 4. – C. 433-438.
103. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci 108: 1513–1518.
104. Goel S.; Yadav M.; Singh K.; Jaat R.S.; Singh N.K. Exploring Diverse Wheat Germplasm for Novel Alleles in HMW-GS for Bread Quality Improvement. J. Food Sci. Technol. 2018, 55, 3257–3262.
105. Gu Y.Q., Coleman-derr D., Kong X. And Anderson O.D. (2004) Rapid Genome Evolution Revealed by Comparative Sequence Analysis of Orthologous Regions from Four Triticeae Genomes. Evolution, 135, 459–470.
106. Guan Y. et al. Transcriptome analysis reveals important candidate genes involved in grain-size formation at the stage of grain enlargement in common wheat cultivar “Bainong 4199” //PLoS One. – 2019. – T. 14. – №. 3. – C. e0214149.
107. Guo G., et al. Wheat mir9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling //The Plant Cell. – 2018. – T. 30. – №. 4. – C. 796-814.
108. Guo X. H. et al. Special HMW-GSs and their genes of *Triticum turgidum* subsp. *dicoccoides* accession D141 and the potential utilization in common wheat //Genetic Resources and Crop Evolution. – 2016. – T. 63. – C. 833-844.
109. Gustafson J. P., Zillinsky F. J. Identification of D-genome chromosomes from hexaploid wheat in a 42-chromosome triticales //Proceedings of the fourth international wheat genetics symposium. – Columbia, MO: University of Missouri, 1973.

110. He Z. H. et al. Composition of HMW and LMW glutenin subunits and their effects on dough properties, pan bread, and noodle quality of Chinese bread wheats //Cereal Chemistry. – 2005. – T. 82. – №. 4. – C. 345-350.
111. He Z.H., Lin Z.J., Wang L.J., Xiao Z.M., Wang F.S. & Zhuang Q.S., 2002. Classification on Chinese wheat regions based on quality. Sci Agric Sinica 35: 359–364.
112. Heidecker G., Messing J. Structural analysis of plant genes //Annual Review of Plant Physiology. - 1986. - T. 37. - №. 1. - C. 439-466.
113. Heo J. B., Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA //Science. – 2011. – T. 331. – №. 6013. – C. 76-79.
114. Höijer I. et al. Detailed analysis of HTT repeat elements in human blood using targeted amplification-free long-read sequencing //Human Mutation. – 2018. – T. 39. – №. 9. – C. 1262-1272.
115. Horiguchi G. et al. Differential contributions of ribosomal protein genes to Arabidopsis thaliana leaf development //The Plant Journal. – 2011. – T. 65. – №. 5. – C. 724-736.
116. Hsieh T.-F.; Ibarra C.A.; Silva P.; Zemach A.; Eshed-William L.; Fischer R.L.; Zilberman D. Genome-Wide Demethylation of Arabidopsis Endosperm. Science 2009, 324, 1451–1454
117. Huo N. et al. Dynamic evolution of α -gliadin prolamin gene family in homeologous genomes of hexaploid wheat //Scientific reports. – 2018. – T. 8. – №. 1. – C. 5181.
118. Hu Y. et al. Coordinated regulation of starch synthesis in maize endosperm by microRNAs and DNA methylation //The Plant Journal. – 2021. – T. 105. – №. 1. – C. 108-123.
119. Hung Y.-H.; Slotkin R.K. The initiation of RNA interference (RNAi) in plants. Curr. Opin. Plant Biol. 2021, 61, 102014.
120. Zemach A.; Kim M.Y.; Silva P.; Rodrigues J.A.; Dotson B.; Brooks, M.D.; Zilberman D. Local DNA Hypomethylation Activates Genes in Rice Endosperm. Proc. Natl. Acad. Sci. USA 2010, 107, 18729–18734

121. Hurkman W. J. Et al. Comparative proteomic analysis of the effect of temperature and fertilizer on gliadin and glutenin accumulation in the developing endosperm and flour from *Triticum aestivum* L. Cv. Butte 86 //Proteome science. – 2013. – T. 11. – №. 1. – C. 1-15.
122. Jensen J. K. et al. RNA-Seq analysis of developing nasturtium seeds (*Tropaeolum majus*): identification and characterization of an additional galactosyltransferase involved in xyloglucan biosynthesis //Molecular plant. - 2012. - T. 5. - №. 5. - C. 984-992.
123. Jia H. et al. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex //Cell. – 2011. – T. 145. – №. 6. – C. 890-901.
124. Jiang Q. T. et al. Characterization and comparative analysis of HMW glutenin 1Ay alleles with differential expressions //BMC plant biology. – 2009. – T. 9. – C. 1-13.
125. Jin X.; Fu Z.; Lv P.; Peng Q.; Ding D.; Li W.; Tang J. Identification and Characterization of microRNAs during Maize Grain Filling. PLOS One 2015, 10, e0125800.
126. Jiang Q.-T.; Wei Y.-M.; Wang F.; Wang J.-R.; Yan Z.-H.; Zheng Y.-L. Characterization and Comparative Analysis of HMW Glutenin 1Ay Alleles with Differential Expressions. BMC Plant Biol. 2009, 9, 16, doi:10.1186/1471-2229-9-16.
127. Johansson E., Prieto-Linde M. L., Jonsson J. Effects of wheat cultivar and nitrogen application on storage protein composition and breadmaking quality //Cereal Chemistry. - 2001. - T. 78. - №. 1. - C. 19-25.
128. Juhasz A. Et al. Role of conserved non-coding regulatory elements in LMW glutenin gene expression //plos One. – 2011. – T. 6. – №. 12. – C. E29501.
129. Karst S.M.; Ziels R.M.; Kirkegaard R.H.; Sørensen E.A.; McDonald D.; Zhu Q.; Knight R.; Albertsen M. High-Accuracy Long-Read Amplicon Sequences Using Unique Molecular Identifiers with Nanopore or PacBio Sequencing. Nat. Methods 2021, 18, 165–169, doi:10.1038/s41592-020-01041-y.

130. Kasarda D. D., Bernardin J. E., Qualset C. O. Relationship of gliadin protein components to chromosomes in hexaploid wheats (*Triticum aestivum* L.) //Proceedings of the National Academy of Sciences. – 1976. – T. 73. – №. 10. – C. 3646-3650.
131. Kawakatsu T., Takaiwa F. Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains //Plant biotechnology journal. – 2010. – T. 8. – №. 9. – C. 939-953.
132. Kaya Y., Akcura M. Effects of genotype and environment on grain yield and quality traits in bread wheat (*T. aestivum* L.) //Food Science and Technology. - 2014. - T. 34. - C. 386-393.
133. Kirov I., et al. Nanopore RNA sequencing revealed long non-coding and LTR retrotransposon-related rnas expressed at early stages of triticale SEED development //Plants. – 2020. – T. 9. – №. 12. – C. 1794.
134. Kiss A. Triticale breeding experiments in Eastern Europe//Triticale: proceedings of an international symposium. – IDRC, Ottawa, ON, CA, 1974.
135. Kovaka S.; Zimin A.V.; Perteau G.M.; Razaghi R.; Salzberg S.L.; Perteau M. Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2. Genome Biol. – 2019, – 20, – C.278. <https://doi.org/10.1186/s13059-019-1910-1>.
136. Kreis M. et al. Differential gene expression in the developing barley endosperm //Philosophical Transactions of the Royal Society of London. B, Biological Sciences. - 1986. - T. 314. - №. 1166. - C. 355-365.
137. Kron P., Husband B. C. Hybridization and the reproductive pathways mediating gene flow between native *Malus coronaria* and domestic apple, *M. Domestica* //Botany. – 2009. – T. 87. – №. 9. – C. 864-874.
138. Kubaláková M. et al. Chromosome sorting in tetraploid wheat and its potential for genome analysis //Genetics. – 2005. – T. 170. – №. 2. – C. 823-829.
139. Kuroda M. et al. Molecular cloning, characterization, and expression of wheat cystatins //Bioscience, biotechnology, and biochemistry. – 2001. – T. 65. – №. 1. – C. 22-28.

140. Lafontaine D. L. J. A ‘garbage can’ for ribosomes: how eukaryotes degrade their ribosomes //Trends in biochemical sciences. – 2010. – T. 35. – №. 5. – C. 267-277.

141. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O’Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, et al. 2011. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30: 78–82

142. Lange H. et al. The RNA helicases AtMTR4 and HEN2 target specific subsets of nuclear transcripts for degradation by the nuclear exosome in Arabidopsis thaliana //PLoS genetics. – 2014. – T. 10. – №. 8. – C. e1004564.

143. Laudencia-Chingcuanco D. L. et al. Transcriptional profiling of wheat caryopsis development using cDNA microarrays //Plant Molecular Biology. – 2007. – T. 63. – C. 651-668.

144. Le B. H. et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors //Proceedings of the National Academy of Sciences. - 2010. - T. 107. - №. 18. - C. 8063-8070.

145. Lew E.J.L., Kuzmicky D.D., Kasarda D.D. Characterization of low-molecular-weight glutenin subunits by reversephase high-performance liquid chromatography, sodium dodecyl sulphate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing // Cereal Chemistry. – 1992. –T.69. – №5. – C. 508-515

146. Li C, Lin H, Chen A, Lau M, Jernstedt J, Dubcovsky J. Wheat VRN1, FUL2 and FUL3 play critical and redundant roles in spikelet development and spike determinacy. Development. 2019;146(14):dev175398.

147. Li G. et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes //Nature genetics. – 2021. – T. 53. – №. 4. – C. 574-584.

148. Li H. Minimap2: Pairwise Alignment for Nucleotide Sequences. Bioinformatics 2018, 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.

149. Li W., P. Zhang J.P. Fellers B. Friebe and B.S. Gill. 2004. Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J.* 40:500–511.
150. Li W., Wan Y., Liu Z., Liu K., Liu X., Li B., ... Wang, D. (2004). Molecular characterization of HMW glutenin subunit allele 1Bx14: Further insights into the evolution of Glu-B1-1 alleles in wheat and related species. *Theoretical and Applied Genetics*, 109, 1093–1104.
151. Li Y., Ruperao P., Batley J., Edwards D., Khan T., Colmer T.D., Pang J., Siddique K.H.M., Sutton T. Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Front. Plant Sci.* 2018; 9:190.
152. Liang W. et al. Identification of long-lived and stable mRNAs in the aged seeds of wheat //Seed Biology. – 2023. – T. 2. – №. 1.
153. Liu D., et al., New technologies accelerate the exploration of non-coding RNAs in horticultural plants //Horticulture research. – 2017. – T. 4.
1. Liu J., Wang H., Chua N.H., Long noncoding RNA transcriptome of plants //Plant biotechnology journal. – 2015. – T. 13. – №. 3. – C. 319-328.
154. Liu L., He Z., Yan J., Zhang Y., Xia X., Peña R.J. Allelic variation at the Glu-1 and Glu-3 loci, presence of the 1B.1R translocation, and their effects on mixographic properties in Chinese bread wheats // *Euphytica*. – 2005. – T. 142. – №3. – C.197-204.
155. Lodé L. et al. Single-molecule DNA sequencing of acute myeloid leukemia and myelodysplastic syndromes with multiple TP53 alterations //Haematologica. – 2018. – T. 103. – №. 1. – C. e13.
156. López-Girona E. Et al. CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants //Plant Methods. – 2020. – T. 16. – №. 1. – C. 1-13.
157. Losh J. S., van Hoof A. Gateway arch to the RNA exosome //Cell. – 2015. – T. 162. – №. 5. – C. 940-941.

158. Lukaszewski A. J., Apolinarska B., Gustafson J. P. Introduction of the D-genome chromosomes from bread wheat into hexaploid triticale with a complete rye genome //Genome. – 1987. – T. 29. – №. 3. – C. 425-430.
159. Lukaszewski, A.J.; Curtis, C.A. Transfer of the Glu-D1 Gene from Chromosome 1D of Bread wheat to Chromosome 1R in Hexaploid Triticale. Plant Breed. 1992, 109, 203–210
160. Luo G.; Song S.; Zhao L.; Shen L.; Wang Y.S.; Mechanisms, X. Origin and Heredity of Glu-1Ay Silencing in Wheat Evolution and Domestication. Theor. Appl. Genet. 2018, 131, 1561–1575.
161. Lykke-Andersen S., Brodersen D. E., Jensen T. H. Origins and activities of the eukaryotic exosome //Journal of cell science. – 2009. – T. 122. – №. 10. – C. 1487-1494.
162. Ma K., et al. Genome-wide identification and characterization of long non-coding RNA in wheat roots in response to Ca²⁺ channel blocker //Frontiers in plant science. – 2018. – T. 9. – C. 244.
163. Ma L., Bajic V. B., Zhang Z. On the classification of long non-coding rnas //RNA biology. – 2013. – T. 10. – №. 6. – C. 924-933.
164. Ma X.; Wang Q.; Wang Y.; Ma J.; Wu N.; Ni S.; Luo T.; Zhuang L.; Chu C.; Cho S.-W.; et al. Chromosome Aberrations Induced by Zebularine in Triticale. Genome 2016, 59, 485–492
165. Madhawan A. et al. Identification and characterization of long non-coding RNAs regulating resistant starch biosynthesis in bread wheat (*Triticum aestivum* L.) //Genomics. - 2020. - T. 112. - №. 5. - C. 3065-3074.
166. Madsen E. B. Et al. Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting //Human mutation. – 2020. – T. 41. – №. 9. – C. 1671-1679.
167. Mamanova L. et al. Target-enrichment strategies for next-generation sequencing //Nature methods. – 2010. – T. 7. – №. 2. – C. 111-118.

168. Matera A. G., Terns R. M., Terns M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs //Nature reviews Molecular cell biology. – 2007. – T. 8. – №. 3. – C. 209-220.
169. Margiotta B., Urbano M., Colaprico G., Johansson E., Buonocore F., D'Ovidio R. And Lafiandra D. (1996) Detection of γ -type Subunit at the Glu-A1 Locus in Some Swedish Bread Wheat Lines. Journal of Cereal Science, 23, 203–211.
170. Marquardt S., et al. Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription //Molecular cell. - 2014. - T. 54. - №. 1. - C. 156-165.
171. Martinez M. et al. Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination //Plant physiology. – 2009. – T. 151. – №. 3. – C. 1531-1545.
172. Martínez M. et al. Comparative phylogenetic analysis of cystatin gene families from arabidopsis, rice and barley //Molecular Genetics and Genomics. – 2005. – T. 273. – №. 5. – C. 423-432.
173. Masci S., Rovelli L., Kasarda D.D., Vensel W.H., Lafiandra D. Characterisation and chromosomal localisation of C-type low-molecular-weight glutenin subunits in the bread wheat cultivar Chinese Spring // Theoretical and Applied Genetics. – 2002. – T. 104. – № 2-3. – C. 422-428.
174. Massonneau A. et al. Maize cystatins respond to developmental cues, cold stress and drought //Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression. - 2005. - T. 1729. - №. 3. - C. 186-199.
175. Mena M. et al. An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native Bhordein promoter in barley endosperm //The Plant Journal. - 1998. - T. 16. - №. 1. - C. 53-62.

176. Meng F.; Liu H.; Wang K.; Liu L.; Wang S.; Zhao Y.; Yin J.; Li Y. Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2013, 13, 140.
177. Mergoum M., Macpherson H. G. (ed.). *Triticale improvement and production.* – Food & Agriculture Org., 2004. – T. 179.
178. Metzker M. L. Sequencing technologies—the next generation // *Nature reviews genetics.* – 2010. – T. 11. – №. 1. – C. 31-46.
179. Mikkonen A. et al. A major cysteine proteinase, EPB, in germinating barley seeds: structure of two intronless genes and regulation of expression // *Plant Molecular Biology.* – 1996. – T. 31. – C. 239-254.
180. Miles M.J., Carr H.J., McMaster T., l'Anson K.J., Belton P.S., Morris V.J. et al. Scanning tunneling microscopy of a wheat seed storage protein reveals details of an unusual supersecondary structure // *Proceedings of the National Academy of Sciences of the United States of America.* – 1991. – T. 88. – № 1. – C. 68-71.
181. Ni P.; Huang N.; Nie F.; Zhang J.; Zhang Z.; Wu B.; Bai L.; Liu W.; Xiao C.-L.; Luo F.; et al. Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. *Nat. Commun.* 2021, 12, 5976.
182. Nucia A., Okoń S., Tomczyńska-Mleko M. Characterization of HMW glutenin subunits in European spring common wheat (*Triticum aestivum* L.) // *Genetic Resources and Crop Evolution.* – 2019. – T. 66. – C. 579-588.
183. Ohbayashi I. et al. Evidence for a role of ANAC082 as a ribosomal stress response mediator leading to growth defects and developmental alterations in *Arabidopsis* // *The Plant Cell.* - 2017. - T. 29. - №. 10. - C. 2644-2660.
184. Onate L. et al. Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm // *Journal of Biological Chemistry.* - 1999. - T. 274. - №. 14. - C. 9175-9182.
185. Osipova D. et al. Altered generation of spontaneous oscillations in Alzheimer's disease // *Neuroimage.* - 2005. - T. 27. - №. 4. - C. 835-841.

186. Pant B.D., et al. Microrna399 is a long-distance signal for the regulation of plant phosphate homeostasis //The Plant Journal. – 2008. – T. 53. – №. 5. – C. 731-738.

187. Park C. S. Et al. Influence of allelic variations in glutenin on the quality of pan bread and white salted noodles made from Korean wheat cultivars //Euphytica. – 2011. – T. 180. – №. 2. – C. 235-250.

188. Patterson M., Marschall T., Pisanti N., Van Iersel L., Stougie L., Klau G. W., & Schönhuth A. WhatsHap: weighted haplotype assembly for future-generation sequencing reads //Journal of Computational Biology. – 2015. – T. 22. – №. 6. – C. 498-509.

189. Payne P.I., Holt L.M., Jackson E.A., Law C.N. Wheat storage proteins: their genetics and their potential for manipulation by plant breeding. Philosophical Transactions of the Royal Society. Biological Sciences. 1984;304:359-371.

190. Payne P.I., Seekings J.A., Worland A.J., Jarvis M.G., Holt L.M. Allelic variation of glutenin subunits and gliadins and its effect on breadmaking quality in wheat: Analysis of F5 progeny from Chinese Spring × Chinese Spring (Hope 1A). Journal of Cereal Science. 1987;6(2):103-118.

191. Pellny T. K. et al. Cell walls of developing wheat starchy endosperm: comparison of composition and RNA-Seq transcriptome //Plant physiology. – 2012. – T. 158. – №. 2. – C. 612-627.

192. Pertea G., Pertea M. GFF utilities: GffRead and GffCompare //F1000Research. – 2020. – T. 9.

193. Pfeiffer A. et al. Combinatorial complexity in a transcriptionally centered signaling hub in Arabidopsis //Molecular plant. – 2014. – T. 7. – №. 11. – C. 1598-1618.

194. Poperelya F.O. Three main genetic systems for grain quality in winter bread wheat (Tri osnovni genetichni sistemi yakosti zerna ozimoy myakoy pshenitsi). In: Implementing the potential of cultivars and hybrids from the Breeding and Genetics Institute in the environments of Ukraine. Collection of

scientific papers of the BGI (Realizatsiya potentsiynikh mozhlivostey sortiv ta gibridiv Seleksiyno-genetichnogo institutu v umovakh Ukraini. Zb. Nauk. Prats SGI). Odesa; 1996. P.117-132.

195. Prabucka B., Bielawski W. Purification and partial characteristic of a major gliadin-degrading cysteine endopeptidase from germinating triticale seeds //Acta Physiologiae Plantarum. – 2004. – T. 26. – C. 383-392.

196. Pucker B., Kleinbölting N., Weisshaar B. Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis //BMC genomics. – 2021. – T. 22. – C. 1-21.

197. Qu L. Q. Et al. Expression pattern and activity of six glutelin gene promoters in transgenic rice //Journal of experimental botany. - 2008. - T. 59. - №. 9. - C. 2417-2424.

198. Rabanus-Wallace M.T.; Hackauf B.; Mascher M.; Lux T.; Wicker T.; Gundlach H.; Baez M.; Houben A.; Mayer K.F.X.; Guo L.; et al. Chromosome-Scale Genome Assembly Provides Insights into Rye Biology, Evolution and Agronomic Potential. Nat. Genet. 2021, 53, 564–573. <https://doi.org/10.1038/s41588-021-00807-0>.

199. Rabinovich S.V.; Panchenko I.A.; Parchomenko R.G.; Bondarenko V.N. High-Molecular Weight Glutenin Subunit Composition of Spring Bread Wheats Grown in the Ukraine and the Russian Federation between 1995-97 and Its Connection with Pedigrees. Annu. Wheat Newsl. 1998, 44, 236–251.

200. Rang F. J., Kloosterman W. P., & De Ridder J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biology, 19, 90. <https://doi.org/10.1186/s13059-018-1462-9>

201. Rangan P., Furtado A., Henry R. J. The transcriptome of the developing grain: a resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat //BMC genomics. – 2017. – T. 18. – №. 1. – C. 1-9.

202. Robert M. et al. Effect of the carbon source on biosurfactant production by *Pseudomonas aeruginosa* 44T1 //Biotechnology Letters. - 1989. - T. 11. - C. 871-874.
203. Rogers S. O., Quatrano R. S. Morphological staging of wheat caryopsis development //American Journal of Botany. – 1983. – T. 70. – №. 2. – C. 308-311.
204. Rosa Filho O. Effect of the six glutenin loci on selected bread quality traits in wheat. Corvallis: Oregon State University; 1997.
205. Sabelli P.A., Shewry P.R. Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis // Theoretical and Applied Genetics. – 1991. – T.83. – №2. – C. 209-216.
206. Sanchez C. G. et al. Regulation of ribosome biogenesis and protein synthesis controls germline stem cell differentiation //Cell stem cell. – 2016. – T. 18. – №. 2. – C. 276-290.
207. Seal A. G., Bennett M. D. Preferential C-banding of wheat or rye chromosomes //Theoretical and Applied Genetics. - 1982. - T. 63. - C. 227-233.
208. Sehnke P. C. et al. Regulation of starch accumulation by granule-associated plant 14-3-3 proteins //Proceedings of the National Academy of Sciences. - 2001. - T. 98. - №. 2. - C. 765-770.
209. Sekiguchi K., Ushitani T., Sawa K. Use of redundant sets of landmark information by humans (*Homo sapiens*) in a goal-searching task in an open field and on a computer screen //Journal of Comparative Psychology. – 2018. – T. 132. – №. 2. – C. 178.
210. Semenov O.G., Divashur M.G., Shangeshapvako H.G., Plushikov V.G., Hupacaria T.I., Vvedensky V.V. et al. Specificity of combinations of qualitative and quantitative characteristics of glucovine in genotypes of alloctoplasmatic spruce wheat with allele of wild type *Wx-b1a*. RUDN Journal of Agronomy and Animal Industries. 2018;13(1):14-25.

211. Shafiq S., Li J., Sun Q. Functions of plants long non-coding RNAs // *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. – 2016. – T. 1859. – №. 1. – C. 155-162.
212. Sheldon C.C., Rouse D.T., Finnegan E.J. et al. The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC) // *Proc. Natl. Acad. Sci. USA*. 2000. Vol. 97. P. 3753–3758.
213. Shewry P.R., Halford N.G., Tatham A.S. High molecular weight subunits of wheat glutenin. *Journal of cereal science*. 1992;15(2):105-120. DOI: 10.1016/S0733-5210(09)80062-3
214. Shi C.; Xu L. Characters of Cysteine Endopeptidases in Wheat Endosperm during Seed Germination and Subsequent Seedling Growth. *J. Integr. Plant Biol*. 2009, 51, 52–57.
215. Simińska J., Bielawski W. Identification and expression analysis of a novel phytocystatin in developing and germinating seeds of triticale (\times *Triticosecale* Wittm.) // *Acta Societatis Botanicorum Poloniae*. – 2015. – T. 84. – №. 1. – C. 139-142.
216. Šimková H. et al. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley // *BMC genomics*. – 2008. – T. 9. – №. 1. – C. 1-9.
217. Singh N.K., Shepherd K.W. Linkage mapping of genes controlling endosperm storage proteins in wheat // *Theoretical and Applied Genetics*. – 1988. – T. 75. - № 4. – C. 628-641.
218. Singh N.K.; Shepherd K.W.; Cornish G.B. A Simplified SDS—PAGE Procedure for Separating LMW Subunits of Glutenin. *J. Cereal. Sci*. 1991, 14, 203–208. [https://doi.org/10.1016/s0733-5210\(09\)80039-8](https://doi.org/10.1016/s0733-5210(09)80039-8).
219. Sinturel F. et al. Diurnal oscillations in liver mass and cell size accompany ribosome assembly cycles // *Cell*. – 2017. – T. 169. – №. 4. – C. 651-663. e14.
220. Smith T. F. et al. Identification of common molecular subsequences // *Journal of molecular biology*. – 1981. – T. 147. – №. 1. – C. 195-197.

221. Soave C., Salamini F. Organization and regulation of zein genes in maize endosperm //Philosophical Transactions of the Royal Society of London. B, Biological Sciences. - 1984. - T. 304. - №. 1120. - C. 341-347.
222. Sreenivasulu N., Schnurbusch T. A genetic playground for enhancing grain number in cereals //Trends in plant science. - 2012. - T. 17. - №. 2. - C. 91-101.
223. Stangl A. L. et al. The Health Stigma and Discrimination Framework: a global, crosscutting framework to inform research, intervention development, and policy on health-related stigmas //BMC medicine. - 2019. - T. 17. - C. 1-13.
224. Suzuki A. et al. Rice MYB protein OSMYB5 specifically binds to the AACAA motif conserved among promoters of genes for storage protein glutelin //Plant and cell physiology. - 1998. - T. 39. - №. 5. - C. 555-559.
225. Szewinska J. et al. Molecular cloning and expression analysis of triticale phytocystatins during development and germination of seeds //Plant molecular biology reporter. - 2012. - T. 30. - C. 867-877.
226. Szewinska J.; Siminska J.; Bielawski W. The Roles of Cysteine Proteases and Phytocystatins in Development and Germination of Cereal Seeds. J. Plant Physiol. 2016, 207, 10–21
227. Tafforeau L. et al. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors //Molecular cell. - 2013. - T. 51. - №. 4. - C. 539-551.
228. Temizgul R.; Akbulut M.; Lafandra D. Genetic Diversity of High-Molecular-Weight Glutenin Subunit Compositions in Bread Wheat Landraces Originated from Turkey. Plant Genet. Resour. Charact. Util. 2018, 16, 28–38.
229. Tetlow I. J. et al. Analysis of protein complexes in wheat amyloplasts reveals functional interactions among starch biosynthetic enzymes //Plant physiology. - 2008. - T. 146. - №. 4. - C. 1878-1891.
230. The International Wheat Genome Sequencing Consortium (IWGSC); Appels R.; Eversole K.; Stein N.; Feuillet C.; Keller B.; Rogers J.; Pozniak C.J.; Choulet F.; Distelfeld A.; et al. Shifting the Limits in Wheat Research and

Breeding Using a Fully Annotated Reference Genome. *Science* 2018, 361. <https://doi.org/10.1126/science.aar7191>.

231. Thoms M. et al. The exosome is recruited to RNA substrates through specific adaptor proteins // *Cell*. - 2015. - T. 162. - №. 5. - C. 1029-1038.

232. Tottman D.R. The Decimal Code for the Growth Stages of Cereals, with Illustrations. *Ann. Appl. Biol.* 1987, 110, 441–454.

233. Tsai Y. C. et al. Amplification-free, CRISPR-Cas9 targeted enrichment and SMRT sequencing of repeat-expansion disease causative genomic regions // *BioRxiv*. – 2017. – C. 203919.

234. Tsoi L.C., et al. Analysis of long non-coding rnas highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin // *Genome biology*. – 2015. – T. 16. – №. 1. – C. 1-15.

235. Verdier J., Thompson R. D. Transcriptional regulation of storage protein synthesis during dicotyledon seed filling // *Plant and Cell Physiology*. – 2008. – T. 49. – №. 9. – C. 1263-1271.

236. Vicente Carbajosa J. et al. Barley BLZ1: a bZIP transcriptional activator that interacts with endosperm-specific gene promoters // *Plant Journal*. - 1998. - T. 13. - №. 5. - C. 629-640.

237. Vicente-Carbajosa J. et al. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2 // *Proceedings of the National Academy of Sciences*. - 1997. - T. 94. - №. 14. - C. 7685-7690.

238. Villareal R.L. Advances in spring Triticale breeding// R.L. Villareal, G.Varughece, O.S. Abdalla// *Plant Breed. Rev.* – 1990. – 8. – P.43-90;

239. Walkowiak S. et al. Multiple wheat genomes reveal global variation in modern breeding // *Nature*. – 2020. – T. 588. – №. 7837. – C. 277-283.

240. Wan Y. et al. Transcriptome analysis of grain development in hexaploid wheat // *BMC genomics*. – 2008. – T. 9. – C. 1-16.

241. Wang J. S., Zhang D. Y. Simulation-guided DNA probe design for consistently intraspecific hybridization //Nature chemistry. - 2015. - T. 7. - №. 7. - C. 545-553.
242. Wang K.C., Chang H. Y. Molecular mechanisms of long noncoding RNAs //Molecular cell. – 2011. – T. 43. – №. 6. – C. 904-914.
243. Wang Y, Yu H, Tian C, Sajjad M, Gao C, Tong Y, Wang X, Jiao Y. Transcriptome association identifies regulators of wheat spike architecture. *Plant Physiol.* 2017;175(2):746-757.
244. Wang K.; Zhang X.; Zhao Y.; Chen F.; Xia G. Structure, Variation and Expression Analysis of Glutenin Gene Promoters from Triticum Aestivum Cultivar Chinese Spring Shows the Distal Region of Promoter 1Bx7 Is Key Regulatory Sequence. *Gene* 2013, 527, 484-490. <https://doi.org/10.1016/j.gene.2013.06.068>.
245. Warner J. R. The economics of ribosome biosynthesis in yeast //Trends in biochemical sciences. – 1999. – T. 24. – №. 11. – C. 437-440.
246. Weis B. L. et al. Plant-specific features of ribosome biogenesis //Trends in Plant Science. – 2015. – T. 20. – №. 11. – C. 729-740.
247. Weis B. L. et al. The 60S associated ribosome biogenesis factor LSG 1-2 is required for 40S maturation in Arabidopsis thaliana //The Plant Journal. – 2014. – T. 80. – №. 6. – C. 1043-1056.
248. Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, Sogoloff B, Tabbaa D, Williams L, Russ C, et al. 2014. Comprehensive variation discovery in single human genomes. *Nat Genet* 46: 1350–1355
249. Wen S. Et al. Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health //Proceedings of the National Academy of Sciences. - 2012. - T. 109. - №. 50. - C. 20543-20548.
250. Wick R. R., Judd L. M., Holt K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing //Genome biology. – 2019. – T. 20. – C. 1-10.

251. Wieser H. And Kieffer R. Correlations of the amount of gluten protein types to technological properties of wheat flours determined on a micro-scale // *Journal of Cereal Science*. - 2001. - №34. - C.19-27
252. Xu Y. et al. Os ABCB 14 functions in auxin transport and iron homeostasis in rice (*Oryza sativa* L.) // *The Plant Journal*. – 2014. – T. 79. – №. 1. – C. 106-117.
253. Yang, G.; Deng, P.; Guo, Q.; Shi, T.; Pan, W.; Cui, L.; Liu, X.; Nie, X. Population Transcriptomic Analysis Identifies the Comprehensive LncRNAs Landscape of Spike in Wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2022, 22, 450
254. Yi F. et al. High temporal-resolution transcriptome landscape of early maize seed development // *The Plant Cell*. - 2019. - T. 31. - №. 5. - C. 974-992.
255. Yi R.; Zhu Z.; Hu J.; Qian Q.; Dai J.; Ding Y. Identification and Expression Analysis of microRNAs at the Grain Filling Stage in Rice (*Oryza sativa* L.) via Deep Sequencing. *PLOS One* 2013, 8, e57863.
256. Yoshihara J., Campbell C. T. Methanol synthesis and reverse water–gas shift kinetics over Cu (110) model catalysts: structural sensitivity // *Journal of Catalysis*. – 1996. – T. 161. – №. 2. – C. 776-782.
257. Young T. E., Gallie D. R. Programmed cell death during endosperm development // *Programmed cell death in higher plants*. – 2000. – C. 39-57.
258. Yu Y. et al. Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development // *The Crop Journal*. - 2016. - T. 4. - №. 2. - C. 92-106.
259. Yu Y. et al. ITS3/ITS4 outperforms other ITS region and 18S rRNA gene primer sets for amplicon sequencing of soil fungi // *European Journal of Soil Science*. – 2022. – T. 73. – №. 6. – C. e13329.
260. Yuan J., et al. Stress-responsive regulation of long non-coding RNA polyadenylation in *Oryza sativa* // *The Plant Journal*. - 2018. - T. 93. - №. 5. - C. 814-827.

261. Yuen Z. W. S. et al. Systematic benchmarking of tools for cpg methylation detection from nanopore sequencing //Nature communications. - 2021. - T. 12. - №. 1. - C. 1-12.

262. Zemach, A.; Kim, M.Y.; Silva, P.; Rodrigues, J.A.; Dotson, B.; Brooks, M.D.; Zilberman, D. Local DNA hypomethylation activates genes in rice endosperm. Proc. Natl. Acad. Sci. USA 2010, 107, 18729-18734.

263. Zhang H., Lang Z., Zhu J. K. Dynamics and function of DNA methylation in plants //Nature reviews Molecular cell biology. – 2018. – T. 19. – №. 8. – C. 489-506.

264. Zhang Y.C., Chen Y. Q. Long noncoding rnas: new regulators in plant development //Biochemical and biophysical research communications. - 2013. - T. 436. - №. 2. - C. 111-114.

265. Zheng, P.; Zhou, C.; Ding, Y.; Liu, B.; Lu, L.; Zhu, F.; Duan, S. Nanopore Sequencing Technology and Its Applications. Med. Comm. 2023, 4, e316. <https://doi.org/10.1002/mco2.316>.

266. Zhou Z.; Liu C.; Qin M.; Li W.; Hou J.; Shi X.; Dai Z.; Yao W.; Tian B.; Lei Z.; et al. Promoter DNA hypermethylation of TaGli-2.1 positively regulates gluten strength in bread wheat. J. Adv. Res. 2021.

267. Zhu M. et al. A 3D printed two DoF soft robotic finger with variable stiffness //2018 12th France-Japan and 10th Europe-Asia Congress on Mechatronics. - IEEE, 2018. - C. 387-391.

268. Zhu Q.H., Wang M. B. Molecular functions of long non-coding rnas in plants //Genes. - 2012. - T. 3. - №. 1. - C. 176-190.

269. Zilic S. et al. Characterization of proteins from grain of different bread and durum wheat genotypes //International journal of molecular sciences. - 2011. - T. 12. - №. 9. - C. 5878-5894.

270. Zillinsky F. et al. Triticale: adaptation, production and uses //Span. - 1980. - T. 23. - №. 2. - C. 83-99.

271. Zimin A. V. et al. The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience 6, 1–7 (2017).

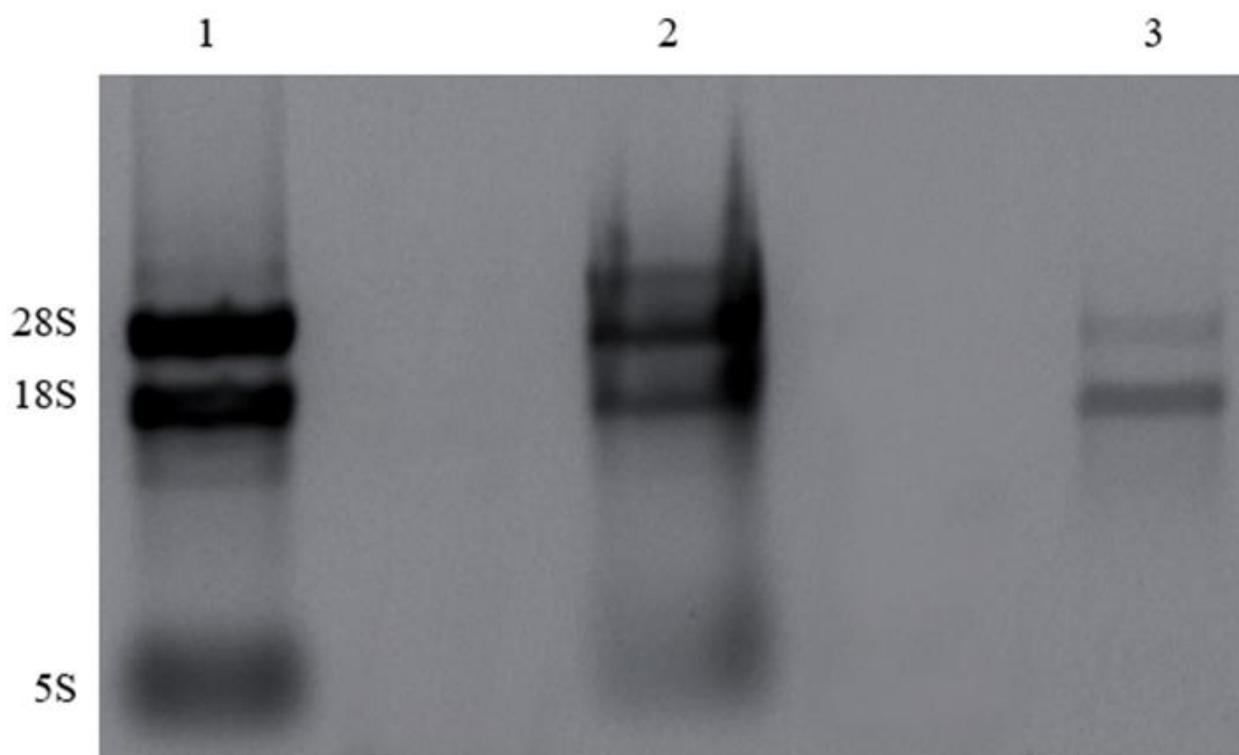


Рисунок 1 – Результаты выделения тотальной РНК с последующим обогащением поли(А) фракцией: 1 – тотальная РНК; 2 – сохраненный супернатант; 3 – обогащенная РНК

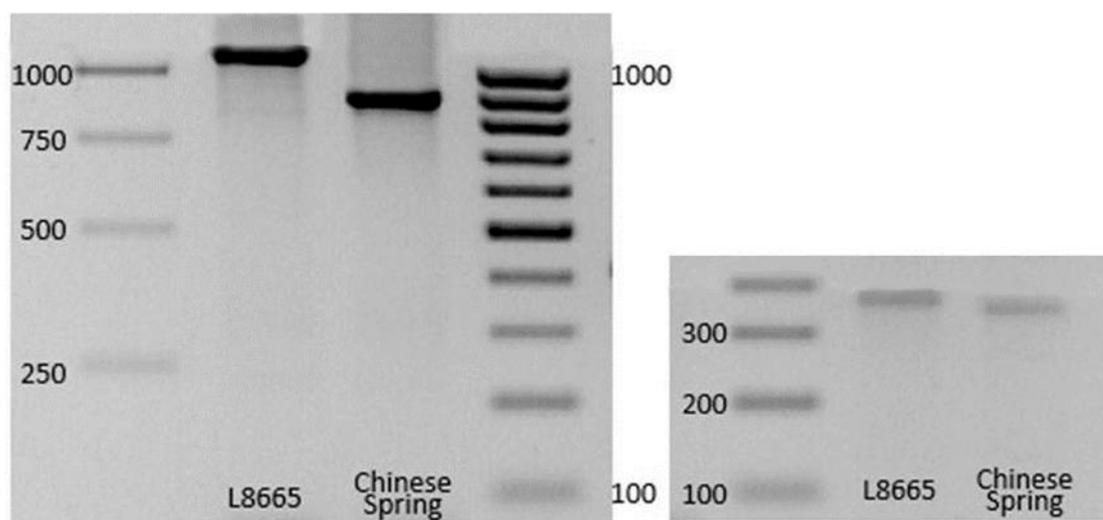


Рисунок 2 – Электрофореграмма идентификации инсерции тритикале линии L8665

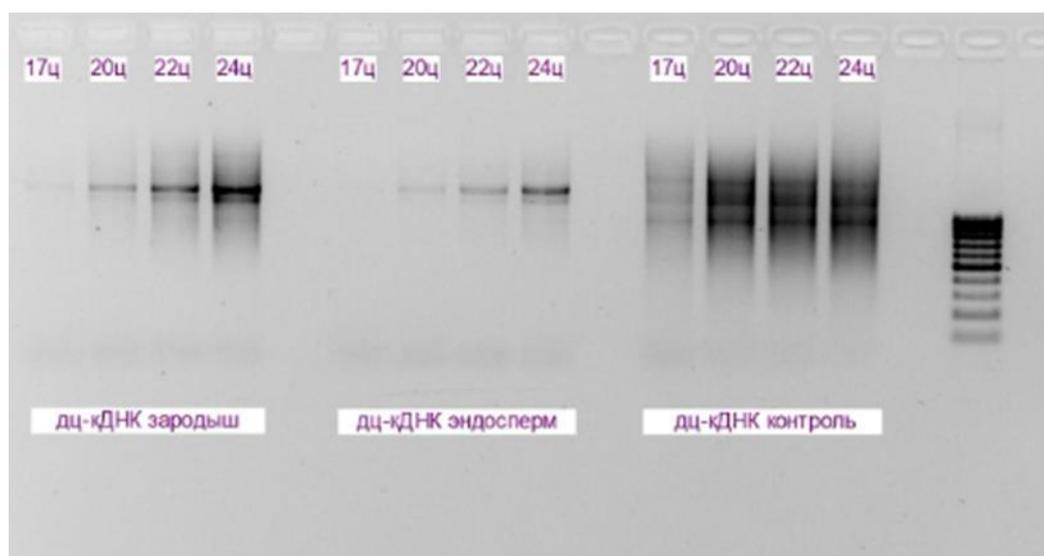


Рисунок 3 – Аналитическая амплификация дц-кДНК на разных циклах.