

Учебно-методические материалы учебного курса дисциплины

**«МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКИЕ ПОДХОДЫ
ДЛЯ СЕЛЕКЦИИ РАСТЕНИЙ»**

Авторы-составители:

Сергеева Е.М., к.б.н., н.с. ИЦиГ СО РАН
Салина Е.А., проф., д.б.н., г.н.с. ИЦиГ СО РАН

Новосибирск

2023

Оглавление

<u>СПИСОК СОКРАЩЕНИЙ</u>	4
<u>ВВЕДЕНИЕ</u>	5
ТЕМА 1.1 Этапы развития молекулярной генетики, интеграция генетики в селекцию.....	6
ТЕМА 1.2 Основные понятия молекулярной генетики.....	7
ТЕМА 1.3 Методы молекулярной генетики растений.....	9
ТЕМА 1.4 Структура генома растений: базовые элементы.....	12
ТЕМА 1.5 Методы цитогенетики для фундаментальных и прикладных исследований.....	15
ТЕМА 1.6 Характеристика и классификация маркеров, Система маркеров RAPD, AFLP, ISSR.....	17
ТЕМА 1.7 Маркеры на основе мобильных элементов.....	22
ТЕМА 1.8 Маркеры на основе однонуклеотидных замен.....	23
ТЕМА 1.9 Маркеры к генам устойчивости и хозяйственно-ценным признакам пшеницы.....	27
ТЕМА 1.10 Генетическое и физическое картирование генов и локусов количественных признаков, ассоциативное картирование.....	29
ТЕМА 1.11 Позиционное клонирование генов. Идентификация аллельных вариантов генов.....	32
ТЕМА 1.12 Методы и подходы к секвенированию генома растений.....	34
ТЕМА 1.13 Проблемы и перспективы использования секвенирования в селекции растений.....	37
ТЕМА 1.14 Маркер-ориентированный отбор для селекции.....	40
ТЕМА 1.15 Геномная селекция.....	41
<u>СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ</u>	44
<u>Приложение. Практические занятия к разделу</u>	45

СПИСОК СОКРАЩЕНИЙ

AFLP – Amplified fragment length polymorphism

BAC – Bacterial Artificial Chromosome

BLAST – Basic Local Alignment Search Tool

CAPS – Cleaved Amplified Polymorphic Sequence

СТАВ – Цетилтриметиламмония бромид

DArT – Diversity Array Technology

FISH – Fluorescence in situ Hybridization

GBS – genotyping-by-sequencing

GISH – Genomic in situ Hybridization

GWAS анализ – genome wide association studies

IRAP – Inter-Retrotransposon Amplified Polymorphism

ISBP – Insertion Site Based Polymorphism

ISSR – Inter-Simple Sequence Repeat

KASP – Kompetitive Allele Specific PCR

LD – linkage disequilibrium (неравновесное сцепление)

LTR – long terminal repeat

MAS (МОП) – Marker assisted selection (маркер-ориентированная селекция)

NGS – next generation sequencing

NIL – Near isogenic line (почти-изогенные линии)

QTL – Quantitative Trait Loci

RAD-seq – restriction-site-associated DNA sequencing

RFLP – Restriction fragment length polymorphism

REMAP – Retrotransposon-Microsatellite Amplified Polymorphism

RAPD – Randomly Amplified Polymorphic DNA

SNP (ОНП) – single nucleotide polymorphism (однонуклеотидный полиморфизм)

SSAP – Sequence Specific Amplification Polymorphism

SSCP – Single Strand Conformation Polymorphism

SSR – simple sequence repeat

ТАЕ – буфер ТрисНСl + ацетат натрия + ЭДТА

ТЕ – буфер ТрисНСl + ЭДТА

ДНК – дезоксирибонуклеиновая кислота

дНТФ – дезоксинуклеозид трифосфат

МЭ – мобильные элементы

ОТ-ПЦР – ПЦР с обратной транскрипцией

П.н. – пар нуклеотидов

ПП – повторяющиеся последовательности

ПЦР – полимеразная цепная реакция

РВ-ПЦР – ПЦР в режиме реального времени

РИЛ – рекомбинантная инбредная линия

РНК – рибонуклеиновая кислота

рРНК – рибосомная РНК

ВВЕДЕНИЕ

Молекулярная генетика – раздел биологии, изучающий молекулярные основы наследственности и изменчивости организмов. Интеграция молекулярной генетики в селекцию растений началось с разработки молекулярных маркеров, которые использовались для генотипирования растений, а также для маркирования генов, определяющих хозяйственно-ценные признаки. Наиболее интенсивно этот процесс стал развиваться в начале 21 века, в первую очередь, в связи разработкой более дешевых и эффективных методов детекции молекулярных маркеров. С другой стороны, возникла необходимость в расширении методов селекции, так как традиционные методы, основанные на гибридизации растений и отбору по фенотипу, уже отставали в скорости получения конкурентоспособных сортов.

В настоящее время эффективное управление селекционным процессом, невозможно представить без таких методов как получение дигаплоидов, молекулярно-цитологической оценки кариотипов, генотипирование растений. Это те инструменты, которыми активно пользуются биотехнологические компании для ускорения селекционного процесса за счет быстрого выравнивания селекционного материал и отбора хозяйственно-ценных линий растений с привлечением молекулярных маркеров.

ТЕМА 1.1. ЭТАПЫ РАЗВИТИЯ МОЛЕКУЛЯРНОЙ ГЕНЕТИКИ, ИНТЕГРАЦИЯ ГЕНЕТИКИ В СЕЛЕКЦИЮ

В первой половине 20 века была идентифицирована молекула-носитель наследственной информации, что положило начало развитию молекулярной генетики. В 1944 г. Эвери, МакЛеод и МакКарти показали, что веществом, вызывающим «трансформацию» невирулентных бактерий-пневмококков в вирулентные, является ДНК (дезоксирибонуклеиновая кислота). В 1952 г. Херши и Чейзом с помощью эксперимента по заражению бактерий радиоактивно меченым фагом T2 показано, что носителем наследственного материала является ДНК. В 1952 г. Розалинд Франклин впервые получена рентгенограмма ДНК, а в 1953 г. Д. Уотсон и Ф. Крик постулировали структуру ДНК как двойную спираль, представленную макромолекулами из сахарофосфатного остова и азотистых оснований. В 1958 г. Ф. Крик сформулировал центральную догму молекулярной биологии, которая заключается в том, что передача генетической информации идет от ДНК к РНК, затем от РНК к белку. Первым методом прямого ферментативного секвенирования ДНК стал метод, предложенный Ф. Сэнгером и Д. Коулсоном в 1975 г. В качестве матрицы в реакции полимеразного копирования использовался одноцепочечный фрагмент ДНК, в качестве праймеров - синтетические олигонуклеотиды. В 1976 г. А. Максамом и У. Гилбертом был разработан метод секвенирования, основанный на специфической химической дегградации фрагмента ДНК, радиоактивно меченного с одного конца. В 1983 г. К. Мюллис разработал метод полимеразной цепной реакции (ПЦР). В 1984 г. А. Джеффрис разработал метод ДНК-фингерпринтинга (ДНК-дактилоскопии), который позволяет различать два индивидуума на основе полиморфизма ДНК. По мере разработки и совершенствования методов анализа ДНК, происходило развитие направления молекулярной генетики растений – разработка систем маркеров, картирование хромосом растений. Так, в 1994 г. была получена генетическая карта с высоким разрешением (726 RFLP маркеров) для риса *O.sativa*. В 2000 г. был секвенирован первый растительный геном – модельного объекта *A.thaliana*. В 2002 г. секвенирован первый геном сельскохозяйственного растений *Oryza sativa*. В 2002 г. были осуществлены первые работы по маркер-ориентированной селекции растений. Разработка в начале 21 века методов современного высокопроизводительного секвенирования позволила осуществить рывок в направлении секвенирования геномов растений – были получены последовательности свыше 300 видов, среди них такие важные возделываемые культуры как соя, пшеница, ячмень, кукуруза, виноград, арахис, яблоня, картофель, кофе, сахарный тростник и др. С применением современных методов секвенирования существенно усовершенствованы

подходы для селекции культурных растений: разработано большое количество молекулярно-генетических маркеров, в частности, высокопроизводительные SNP-чипы; метод полногеномного поиска ассоциаций; маркер-ориентированная и геномная селекция.

Контрольные вопросы по теме 1.1.

1. Кто и в каком году сформулировал центральную догму молекулярной биологии?
2. Кто и в каком году предложил метод ПЦР?
3. В чем различия методов секвенирования по Сэнгеру и по Максому/Гилберту?

ТЕМА 1.2. ОСНОВНЫЕ ПОНЯТИЯ МОЛЕКУЛЯРНОЙ ГЕНЕТИКИ

Растительная клетка имеет размер порядка 100 мкм, размер ядра – примерно 10 мкм. При этом, в ядре каждой клетки растительного организма присутствует набор молекул ДНК, которые содержат всю генетическую информацию, необходимую для дифференцировки в клетки различных органов и тканей. Для развития растительного организма и дифференцировки клеток необходимы воспроизведение (репликация), передача и реализация наследственной информации.

Молекулой – носителем генетической информации у эукариот является ДНК (дезоксирибонуклеиновая кислота). ДНК представляет собой так называемую макромолекулу (биополимер), который состоит из отдельных единиц (нуклеотидов). Нуклеотид состоит из сахара дезоксирибозы, соединенным с молекулой фосфата, и одним из четырех азотистых оснований (аденин, гуанин, цитозин или тимин). Молекула ДНК представляет собой двойную спираль, каждая из нитей которой расположена в антипараллельной ориентации, а нуклеотиды нитей соединены собой водородными связями по принципу комплементарности (А:Т, Г:Ц). Такая протяженная (свыше 20 млн.п.н. у, например, *Arabidopsis*) молекула ДНК формирует хромосому. Совокупность генетического материала, содержащегося в клетке организма, называется геномом, в клетке растений присутствуют ядерный геном (хромосомы ядра), митохондриальный и пластидный геномы в цитоплазме. При этом в ядре клетки молекулы ДНК (хромосомы) находятся в компактизованном состоянии (в виде хроматиновых фибрилл, сформированных непосредственно ДНК и белками укладки – гистонами). Внутри ядра каждая хромосома занимает определенный неслучайный участок – так называемую хромосомную территорию, внутри таких территорий могут присутствовать топологически связанные

домены – участки, внутри которых увеличена частота взаимодействий, что может влиять на регуляцию активности генов.

Реализация генетической информации происходит за счет процесса транскрипции – синтеза молекулы РНК с использованием в качестве матрицы молекул ДНК, с помощью фермента ДНК-зависимой РНК полимеразы. Транскрибируются определенные участки ДНК (гены), которые могут быть как белок-кодирующими (им соответствует так называемая матричная мРНК), так и не кодирующими (например, гены 5S и 45S рРНК, гены тРНК, гены других некодирующих РНК). Затем в цитоплазме на рибосомах происходит процесс трансляции: с мРНК происходит синтез соответствующей последовательности белка из мономеров-аминокислот. Таким образом, происходит передача генетической информации от ДНК к РНК, от РНК – к белку, такое правило реализации наследственной информации называется «центральная догма молекулярной биологии». Совокупность транскрибированных молекул РНК в клетке называется «транскриптом», а совокупность белков «протеом»

Процесс, при котором информация с молекулы ДНК передается на молекулы РНК (и, в случае белок-кодирующих генов, на белок) – называется экспрессией генов. В клетках разных тканей, на разных стадиях развития, при воздействии различных внешних факторов (биотические и абиотические стрессы, освещенность, фитогормоны, редокс-статус и т.д.) – профиль экспрессии генов может меняться, таким образом, транскриптом и протеом являются динамичной фракцией.

Белок-кодирующие гены имеют мозаичную структуру – состоят из экзонов (кодируют белок), интронов, а также 5' и 3'-нетранслируемых областей. Около зоны инициации транскрипции расположена нуклеотидная последовательность, называемая промотором, которая обеспечивает узнавание и связывание РНК-полимеразы. На процесс транскрипции генов оказывают влияние цис-регуляторные элементы (последовательности ДНК), такие как энхансеры, сайленсеры, инсуляторы, расположенные вблизи гена. Также влияние на экспрессию оказывают транс-регуляторные факторы (гены, находящиеся на большом расстоянии, и кодирующие транскрипционные факторы). Факторы транскрипции – белки, контролирующие процесс синтеза РНК путем связывания с цис-регуляторными последовательностями, или непосредственно с РНК-полимеразой.

На физическую доступность РНК-полимеразы к участкам ДНК влияет состояние хроматина, которое определяется степенью метилирования/ацетилирования лизиновых остатков гистонов. Метилированные гистоны формируют участки «закрытого» для транскрипции хроматина, так называемый гетерохроматин. Ацетилированные гистоны формируют «открытый» эухроматин, с ДНК которого может происходить процесс

транскрипции. Также на экспрессию генов оказывают влияние эпигенетические факторы – метилирование ДНК и РНК-интерференция. Метилирование цитозина в области CpG динуклеотидов подавляет транскрипцию гена. Также подавляющее действие оказывает процесс РНК-интерференции: siРНК, короткие двуцепочечные молекулы длиной 20-25 п.н., взаимодействуют с матричной РНК и способствуют ее деградации, предотвращая трансляцию на рибосомах.

Контрольные вопросы по теме 1.2.

1. Как происходит реализация генетической информации?
2. Как называется процесс, при котором информация с молекулы ДНК передается на молекулы РНК?
3. Структура белок-кодирующих генов?
3. Что влияет на физическую доступность РНК-полимеразы к участкам ДНК?

ТЕМА 1.3 МЕТОДЫ МОЛЕКУЛЯРНОЙ ГЕНЕТИКИ РАСТЕНИЙ

Основными методами, применяемыми в условиях лаборатории для изучения генетического материала растений, являются: выделение ДНК, ПЦР (полимеразная цепная реакция), гидролиз эндонуклеазами рестрикции, электрофорез в агарозном или полиакриламидном геле, гибридизация нуклеиновых кислот по Саузерну, клонирование последовательностей ДНК в клетки кишечной палочки *E.coli*.

Для клеток растений, в отличие от клеток животных, характерно наличие клеточной стенки, которую необходимо разрушить для дальнейших манипуляций, необходимых для выделения препарата ДНК. Для разрушения клеточной стенки применяются: растирание в ступке и пестике в присутствии жидкого азота, либо гомогенизация стеклянными или керамическими шариками либо полипропиленовым пестиком. Полученная после гомогенизации суспензия состоит из остатков клеточной стенки, полисахаридов (например, крахмал), белков, ДНК, РНК, и, в ряде случаев, вторичных метаболитов растений (фенольные соединения, алкалоиды, терпеноиды), которые могут существенно затруднять процесс выделения ДНК. Для лизиса мембран, связывания и инактивации белков и других метаболитов проводят инкубацию гомогената с буфером (например СТАВ-буфер; буфер с протеиназой К; гуанидин-тиоцианатный буфер). Следующим шагом является очистка лизата от метаболитов: ее проводят либо с помощью спин-колонки с кремниевой мембраной, либо путем

центрифугирования со смесью фенол-хлороформ. В заключение, препарат ДНК осаждают этиловым либо изопропиловым спиртом, высушивают, и разводят в ТЕ-буфере.

Для аналитической оценки таких параметров ДНК, как примерная длина молекул, качество и количество препарата, используется гель-электрофорез. Электрофорез в агарозном геле позволяет разделять молекулы ДНК длиной от 100 по 25000 п.н. (в зависимости от концентрации агарозы). Электрофорез в акриламидном геле позволяет разделять фрагменты от 5 до 1000 п.н. Принцип гель-электрофореза заключается в том, что фрагменты ДНК являются отрицательно заряженными, и таким образом, в специальной камере с ТАЕ либо ТВЕ-буфером движутся от катода (-) к аноду (+). Гель, сформированный молекулами полимеров агарозы или полиакриламида, представляет собой пористую сетку, в которой более короткие молекулы ДНК движутся быстрее, таким образом, под действием электрического поля происходит разделение молекул ДНК по размеру. Визуализация молекул ДНК в геле проводится с помощью окраски бромистым этидием и экспозицией геля в УФ-излучении. Также электрофорез коротких молекул нуклеиновых кислот может проводиться в капиллярном секвенаторе (автоматический прибор, например, ABI Prism), что, например, используется для секвенирования по Сэнгеру или SSR-генотипирования. В этом случае фрагменты ДНК метят с помощью флюоресцентных красителей, а детекция и анализ результатов проводится автоматизированно.

Одним из классических, но уже мало используемых методов анализа нуклеиновых кислот является Саузерн-блот гибридизация. Этот метод активно использовался в 90-х годах прошлого века для RFLP-картирования. Геномная ДНК различных образцов растений, обработанная эндонуклеазами рестрикции, разделяется в агарозном геле, затем осуществляется денатурация ДНК и одновременно ее перенос из геля на специальную положительно заряженную мембрану. Мембрану с зафиксированной на ней ДНК, гибридизуют с радиоактивно либо флюоресцентно меченным зондом (определенной последовательностью ДНК), молекулы зонда по принципу комплементарности связываются с гомологичной последовательностью геномной ДНК. После детекции радиоактивного или флюоресцентного сигнала можно выявить полиморфизм сигналов гибридизации между различными изучаемыми образцами.

ПЦР (полимеразная цепная реакция), является наиболее часто используемым в условиях лаборатории методом молекулярной генетики. Принцип метода ПЦР заключается в амплификации целевого участка ДНК с помощью фермента Таq ДНК-

полимеразы. Основные компоненты реакционной смеси для ПЦР: (1) матричная ДНК – исследуемый образец; (2) праймеры – олигонуклеотиды длиной 18-22 п.н., (3) буфер, содержащий ионы Mg^{2+} , (4) фермент Taq ДНК-полимераза (либо другая термостабильная полимеразы), (5) дНТФ для построения молекул ДНК. Реакция проводится в приборе, называемом амплификатор, при следующих параметрах – денатурация двуцепочечных молекул ДНК ($94^{\circ}C$), отжиг праймеров на одноцепочечные матричные молекулы ($55-60^{\circ}C$), и элонгация (рост цепи ДНК) при $72^{\circ}C$. Процессы денатурации, отжига и элонгации циклически повторяются (обычно около 35 циклов), таким образом, происходит экспоненциальное увеличение количества целевого фрагмента. Последовательности праймеров, необходимые для конкретной цели, разрабатывают с помощью специальных компьютерных программ (например, Primer3).

Если необходимо исследовать материал в виде не ДНК, а РНК, применяют метод ПЦР с обратной транскрипцией (ОТ-ПЦР). Для этого одноцепочечные молекулы РНК с помощью реакции обратной транскрипции превращают в кДНК (комплементарную ДНК), которая затем используется как матрица для классической реакции ПЦР,

Кроме того, метод ПЦР позволяет не только определять присутствие целевой нуклеотидной последовательности в изучаемом образце, но и измерять количество ее копий. Для этого применяют модификацию метода, называемую кПЦР (количественная), или РВ-ПЦР (в режиме реального времени) – амплификацию проводят на специальном амплификаторе Real Time, а в реакцию вводят флюоресцентную метку (например, SYBR Green), измерение количества которой в каждом цикле позволяет рассчитывать количество копий.

Контрольные вопросы по теме 1.3.

1. В каких случаях используют электрофорез в агарозном геле, а в каких в акриамидном?
2. Каковы основные этапы метода Саузерн-гибридизации?
3. Каковы основные этапы ПЦР (полимеразная цепная реакция)?

ТЕМА 1.4. СТРУКТУРА ГЕНОМА РАСТЕНИЙ: БАЗОВЫЕ ЭЛЕМЕНТЫ

Геном – совокупность наследственного материала (ДНК), присутствующая в клетке организма. Обычно термином «геном» обозначают генетический материал, составляющий гаплоидный набор хромосом. Геномы подразделяют на ядерный (совокупность хромосом ядра) и органелльные (геномы митохондрий и пластид).

Размер гаплоидного генома обозначают термином «значение C» (т.е. содержание ДНК в нереплицированном наборе хромосом в гаметах = 1C), и это значение является постоянным для организма. В исследованиях в области молекулярной генетики размеры молекул ДНК принято измерять в парах нуклеотидов (п.н.). Размеры ядерных геномов растений варьируют в широких пределах – доступные в настоящее время данные показывают различия в размерах геномов у покрытосеменных примерно в 2440 раз. Самый маленький известный геном покрытосеменных растений: плетящееся растение семейства пузырчатковых *Genlisea tuberosa* (61 млн.п.н./1C), самый большой – японский вороний глаз *Paris japonica* (семейство Лилейных) – 149 000 млн.п.н./1C. Средний размер генома покрытосеменных растений составляет 588 млн.п.н. Размеры геномов некоторых возделываемых культур – мягкая пшеница 17 000 млн.п.н., картофель 840 млн.п.н., рис – 400 млн.п.н., соя – 1100 млн.п.н. При этом увеличенный размер генома не означает увеличения морфологической сложности организма – так называемый «парадокс значений C»

Основные факторы, влияющие на увеличение размеров ядерных геномов растений – полиплоидия и накопление повторяющихся последовательностей. Полиплоидия – это кратное увеличение числа наборов хромосом. Автополиплоиды образуются путем кратного увеличения одного и того же генома, а аллополиплоиды – на основе объединения двух или нескольких целых геномов, принадлежащих разным видам и родам (гибридная полиплоидия). По разным оценкам от 30 до 70% современных цветковых являются полиплоидами. Степень полиплоидии обозначается буквой x , число гаплоидных хромосом – n , а разные геномы имеют буквенные обозначения. Например, геномная формула мягкой пшеницы *Triticum aestivum*: $2n = 6x = 42$, AABBDD. Среди культивируемых растений полиплоидами являются мягкая пшеница (аллогексаплоид), картофель (автотетраплоид), банан (автотриплоид), рапс (аллотетраплоид), и др.

Второй источник изменчивости размера генома – амплификация и делеция последовательностей ДНК генома. Геном растений состоит из уникальных и низкокопийных последовательностей ДНК и повторяющейся ДНК. Фракция уникальных и низкокопийных последовательностей ДНК представлена генами, кодирующими белки, необходимые для функционирования организма растений, и связанными с ними

регуляторными элементами. Средняя длина кодирующей части гена у эукариот равна 1 346 п.н. Среднее число белок-кодирующих генов у диплоидных высших растений – около 30 000, суммарно они составляют порядка 1-2% генома. Последовательности ДНК, содержащие гены, неравномерно распределены на хромосоме. Они образуют дискретные кластеры, чередующиеся с участками, состоящими из протяженных блоков повторяющихся последовательностей. При этом плотность генов увеличивается от центромеры к дистальному участку хромосом.

Наиболее представленным компонентом геномов растений являются повторяющиеся последовательности (ПП) ДНК, которые могут составлять до 90-95% генома (как, например, у злаковых). Повторы представляют собой высокогетерогенную группу, представленную тысячами или десятками тысяч семейств, отличающихся по ряду критериев: по длине мотива, уровню копийности и организации в геноме. По уровню копийности повторы делят на высококопийные (свыше 100000 копий/геном) и умеренно повторяющиеся (от 10 до 100000 копий). По уровню организации в геноме повторы разделяются на тандемные (сателлитные) и диспергированные повторы. Сателлитная ДНК состоит из рядов идентичных повторяющихся единиц (мономеров), число которых может варьировать от нескольких сот до миллиона. В зависимости от размера повторяющихся единиц различают микро- и минисателлитную ДНК. Микросателлиты (SSR – simple sequence repeats) имеют длину мономера не более 6 п.н., тогда как у минисателлитов она колеблется в пределах 7–100 п.н. Остальные повторы с длиной мономера выше 100 п.н. относятся к сателлитной ДНК. В геномах злаковых микросателлитная ДНК составляет порядка 2% и в основном представлена короткими высокоповторяющимися последовательностями ди- и тринуклеотидов. Микросателлитные повторы часто расположены в эухроматиновых районах хромосом, и поэтому широко используются в качестве маркеров для построения молекулярно-генетических карт. Для минисателлитной и сателлитной ДНК характерно распределение в виде блоков в гетерохроматине прителомерных или центромерных районов. Некоторые семейства сателлитной ДНК могут составлять порядка 10% генома (например, pSc119.2 ржи).

Диспергированные последовательности ДНК в основном представлены мобильными элементами генома (МЭ). МЭ подразделяют на два класса: элементы класса I (ретротранспозоны), которые перемещаются с помощью механизма «копирования–встраивания» с использованием РНК-посредников; и элементы класса II (ДНК-транспозоны), которые используют механизм «вырезания–встраивания», с образованием либо одно-, либо двухцепочечных разрывов ДНК. Ретротранспозоны, в свою очередь, подразделяются на содержащие длинные концевые повторы (long terminal repeat - LTR)

LTR-ретротранспозоны и non-LTR ретротранспозоны. Выделяют два основных суперсемейства LTR-ретротранспозонов: Gypsy и Copia, различающихся порядком расположения генов обратной транскриптазы и интегразы друг относительно друга. У растений LTR-ретротранспозоны являются преобладающей группой МЭ. Они составляют от 15% генома у *Arabidopsis thaliana* и до 90 % у некоторых представителей Liliaceae. Растения с большими геномами, такие как кукуруза, пшеница, ячмень, могут содержать тысячи семейств LTR-ретротранспозонов. Однако основную массу составляют, как правило, несколько или даже одно семейство ретротранспозонов, например BARE1 у ячменя, Oriе у кукурузы. ДНК-транспозоны представлены в геномах растений в меньшем количестве. Тем не менее некоторые надсемейства, например, САСТА у пшеницы и ячменя, распространились более успешно. В геномах растений ДНК-транспозоны часто расположены в непосредственной близости с генами.

Часть ПП ДНК может иметь важную для организма функцию, например, гены 45S и 5S рибосомальной РНК (рРНК), однако значение в геноме остальных повторяющихся последовательностей неизвестно и является предметом обсуждения. Предполагают, что повторяющаяся ДНК играет важную роль в стабилизации и поддержании структуры хромосом, в «узнавании» и правильном расхождении хромосом во время митоза и мейоза. Лocus некоторых семейств сателлитных ДНК выступают как точки рекомбинации хромосом в мейозе. Также показано, что видообразование у растений часто связано с быстрыми изменениями фракции ПП ДНК.

Контрольные вопросы по теме 1.4

1. Что означает термин «геном»?
2. Из каких элементов состоит геном растений?
3. Какой самый маленький и самый большой геном растений?
4. На какие фракции разделяются повторяющиеся последовательности ДНК?

ТЕМА 1.5. МЕТОДЫ ЦИТОГЕНЕТИКИ ДЛЯ ФУНДАМЕНТАЛЬНЫХ И ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ

Размеры линейных молекул ДНК, составляющих хромосомы растений, могут варьировать. К примеру, размер хромосомы модельного вида *Arabidopsis thaliana* - 30 млн. п.н., при том, что в диплоидном ядре генома содержится 10 хромосом. Геномная ДНК в ядре клетки находится в компактизованном состоянии в виде нуклеопротеиновых

комплексов. Так, двуцепочечная ДНК накручена на нуклеосомы (белковые комплексы из 8 молекул белка гистона), с образованием нуклеосомной нити диаметром 10 нм. Нуклеосомная нить, с свою очередь, формирует фибриллу диаметром 30 нм, которые свернуты в петли, содержащие участки ДНК до 200 т.п.н., закрепленные на ядерном матриксе.

Хромосома компактизована неоднородно – в зависимости от степени метилирования гистонов хроматин может быть «открытым» для транскрипции – эухроматин, так и «закрытым» - гетерохроматин, компактизованные участки, в которых практически отсутствует транскрипция. При окраске методом С-бэндинга участки гетерохроматина интенсивно окрашены.

Структурно хромосому можно разделить на следующие участки: центромеры с прицентромерными районами, теломеры с прителомерными районами, районы вторичной перетяжки (ядрышковый организатор). Эти участки хромосомы являются служебными. Теломеры защищают концы хромосом при репликации ДНК, центромеры участвуют в формировании веретена деления, в компартменте, называемом ядрышковым организатором, происходит синтез рРНК для формирования рибосом. Центромера разделяет хромосому на два плеча – длинное и короткое (L и S).

Хромосомный анализ растений – это комплекс методов, направленных на выявление особенностей хромосомной организации генома вида, идентификацию его хромосом, анализ их функциональной активности. Кариотип (совокупность полного набора хромосом, присущая клеткам данного биологического вида) служит важным таксономическим признаком. У большинства высших растений числа хромосом варьируют в пределах от 10 до 50 ($n = 5-25$). Предполагают, что виды с числом хромосом $n > 10$ являются полиплоидами.

Одним из основных методов для проведения хромосомного анализа (который включает в себя определение хромосомной локализации генных или других ДНК-последовательностей) является метод флюоресцентной *in situ* гибридизации с мечеными ДНК-зондами (маркерными последовательностями). ДНК-зонды метят непосредственно флюорохромами, затем после гибридизации на препарате метафазных хромосом проводят флюоресцентную детекцию сигнала, что дало название методу **FISH - Fluorescence in situ Hybridization**. Метод FISH позволяет локализовать на одном препарате одновременно несколько ДНК зондов.

Одной из модификаций метода *in situ* гибридизации является **GISH (Genomic In Situ Hybridization)**. При проведении GISH меченую геномную ДНК одного из родителей полиплоидного вида или гибрида используют как пробу, а фрагментированную немеченую геномную ДНК другого родителя или самого полиплоида добавляют в гибридизационную

смесь для блокирования кросс-гибридизации. В результате в кариотипе полиплоида проявляется четкая дифференцировка между хромосомами и участками хромосом, унаследованными от разных предков. Показано успешное применение этого метода для *Arabidopsis*, видов рода *Triticum*, *Agropyron* и др.

Метод **Fiber-FISH** (гибридизация на препаратах растянутой ДНК) используется для точной физической локализации генов и разных семейств повторяющихся последовательностей, колокализации зондов относительно друг друга, картирования трансгенов и решения многих других задач.

Генетическая номенклатура для классификации кариотипов базируется на генетическом родстве (гомеологии) хромосом. Анализ мейотической конъюгации хромосом является подходом позволяющим оценить уровень сходства (гомологии) хромосом у гибридных форм растений. Этот метод был впервые предложен О. Розенберг в 1909 году для изучения диплоидных и тетраплоидных видов рода *Drosera* и в дальнейшем получил название метода «геномного анализа». Методы *in situ* гибридизации позволяют получить дополнительную информацию о сходстве геномов различных видов, а также расширить возможности метода геномного анализа. Для идентификации отдельных хромосом внутри кариотипа в качестве маркеров используют видоспецифичные или геном-специфичные последовательности ДНК. Широко используемым типом маркерных последовательностей являются сателлитные (тандемно организованные) последовательности ДНК. Они могут составлять до 10–20% от общего размера генома. Тандемно организованные последовательности ДНК участвуют в формировании таких функционально важных участков хромосом, как теломеры, центромеры и ядрышкообразующие районы. В молекулярной цитогенетике пшеницы чаще всего используют сателлитные геномспецифичные повторы pSc119.2 и pAs1; комбинация этих зондов позволяет распознавать 19 из 21 пары хромосом. Ценными цитогенетическими маркерами являются локусы генов 5S и 45S рРНК, которые позволяют устанавливать гомеологичные отношения хромосом, а также позволяют установить принадлежность некоторых хромосом к одной из трех гомеологических групп (1-я, 5-я и 6-я) у видов трибы *Triticeae* злаковых.

Картирование повторяющихся ДНК не представляет сложности; такие зонды широко применяют для идентификации хромосом самых разных растений. При этом картирование индивидуальных генов представляет значительные трудности из-за избыточного содержания повторов и особенностей организации хромосом растений. Благодаря использованию системы пирамидной амплификации сигнала, а также

совершенствования систем регистрации изображений стало возможным картировать последовательности ДНК длиной менее 3 т.п.н.

Контрольные вопросы по теме 1.5

- 1) Для каких целей используется метод GISH?
- 2) Какую информацию о структуре генома можно получить с использованием FISH?
- 3) В каких случаях используется Fiber-FISH?

ТЕМА 1.6. ХАРАКТЕРИСТИКА И КЛАССИФИКАЦИЯ МАРКЕРОВ, СИСТЕМА МАРКЕРОВ RAPD, AFLP, ISSR, SSR

Молекулярно-генетические маркеры являются методологическим инструментом для изучения генетического разнообразия различных видов растений, используются в фундаментальных (картирование генов, филогенетический анализ) и прикладных (маркер-ориентированная селекция, паспортизация сортов растений) исследованиях. Внедрение современных технологий ДНК-маркеров позволяет обеспечивать такие задачи сельского хозяйства, как повышение урожайности растений, устойчивость к биотическим и абиотическим стрессам.

ДНК-маркеры, или молекулярно-генетические маркеры – полиморфный признак, выявляемый методами молекулярной биологии на уровне нуклеотидной последовательности ДНК для определенного гена или для любого другого участка хромосомы при сравнении генотипов различных особей, пород, сортов, линий. Для того, чтобы последовательность могла быть использована в качестве ДНК-маркера, она должна соответствовать ряду критериев: (1) маркер должен быть сцеплен с изучаемым признаком; (2) последовательность должна быть полиморфной – за счет точечных мутаций, инделей, геномных перестроек, различий в количестве tandemных повторов; (3) маркер должен быть наследуемым (доминантным либо кодоминантным). К молекулярным маркерам применяют термины «локус», «аллель», «доминантный», «кодоминантный». Кодоминантные маркеры позволяют различать одновременно разные аллели, находящиеся в одном геноме, то есть как гомо-, так и гетерозиготы. Если выявляется только один аллель – говорят о доминантном наследовании. Молекулярные маркеры подразделяют на монолокусные и мультилокусные, в зависимости от того, один или много участков ДНК (локусов) используются в качестве маркирующего признака. Монолокусные маркеры наследуются чаще всего по кодоминантному типу, мультилокусные – по доминантному.

Полиморфные последовательности ДНК используются для маркирования различных уровней организации генома: генов, участков хромосом, геномов, пангенома (для совокупности популяций и видов растений). При этом можно выделить следующие генетико-селекционные задачи: 1) генетическое картирование; 2) оценка генетического полиморфизма (гетерозиготность популяции, микроэволюция); 3) филогения и таксономия; 4) генотипирование (фингерпринтинг) особей, линий, семейств, популяций, видов; 5) использование в селекции (MAS, marker assistant selection – маркер-ориентированная селекция); 6) выявление и анализ генов, определяющих важные хозяйственно-значимые признаки, такие как устойчивость к биотическим и абиотическим стрессовым факторам, урожайность, содержание питательных компонентов, продолжительность отдельных фаз развития и др.

В настоящее время насчитывается несколько десятков типов ДНК-маркеров. Их разделяют на три группы, согласно основному методу анализа: маркеры, исследуемые с помощью (1) блот-гибридизации, (2) ПЦР и (3) ДНК-чипов. Данная классификация отражает процесс «эволюции» ДНК-маркеров.

Первое поколение ДНК-маркеров, получивших распространение в 80-е годы прошлого века, основывалось на гибридизации радиоактивно меченых «ДНК-зондов», которые собственно и представляли собой последовательности анализируемых маркеров, на гидролизованную эндонуклеазами рестрикции и разделенную методом электрофореза ДНК образцов. Изобретение метода ПЦР в 1983 г. К. Мюллисом положило задел для разработки использования с 90-х годов и по настоящее время большого количества систем ДНК-маркеров, основанных на применении метода полимеразной цепной реакции. С начала 21го века начал применяться метод ДНК-чипов, основанный на гибридизации ДНК образцов с панелью флюоресцентно меченых зондов, ковалентно пришитых к основанию чипа. За всю историю применения ДНК-маркеры обеспечили бурное развитие генетики и селекции растений: с использованием ДНК-маркеров составлены подробные генетические и физические карты геномов десятков видов растений, проведены исследования в области популяционной генетики, сравнительной генетики, геномики, в филогенетических исследованиях, ДНК-маркеры успешно используются для маркер-ориентированной селекции растений.

RFLP-маркеры (restriction fragment length polymorphism). Представляют собой первое поколение ДНК-маркеров, получивших широкое распространение в 1980-е годы. В основе лежит метод оценки полиморфизма длины рестрикционных фрагментов. Выделенная из растительных тканей геномная ДНК "разрезается" специфичными бактериальными ферментами – эндонуклеазами рестрикции (рестриктазами). Продукты

рестрикции разделяют электрофорезом в агарозном геле, затем с помощью метода Саузерн-блоттинга, и проводят гибридизацию со специфическими фрагментами ДНК (маркерами), меченными радиоактивными зондами. На заключительном этапе с помощью радиоавтографии анализируют положение связанных с зондом фрагментов ДНК. В качестве зондов чаще всего используют клонированные в бактериальных плазмидах отдельные (уникальные или повторяющиеся) последовательности ДНК небольшой длины - от нескольких десятков до нескольких тысяч п.н. RFLP был разработан как первый метод для массового применения, с его помощью составлены генетические карты для многих видов растений.

В 1990-е годы ключевые позиции заняли ПЦР-маркеры. Для постановки реакции необходимы ПЦР необходима пара специально подобранных праймеров, которые являются затравкой для синтеза выбранного маркерного участка ДНК, однако без знания нуклеотидной последовательности исследуемых участков подобрать такую последовательность невозможно. Поэтому до начала широкого распространения геномного секвенирования интенсивно использовались так называемые анонимные ПЦР-маркеры (маркеры, для разработки которых не нужны данные о последовательности ДНК объекта).

К группе анонимных ПЦР-маркеров относятся RAPD, AFLP и ISSR. **RAPD-метод** (*Randomly Amplified Polymorphic DNA*) основан на анализе случайно амплифицированной полиморфной ДНК. Для этого используют праймеры со случайной последовательностью, которые должны отвечать определённым требованиям по соотношению GC-пар (около 60%) и длине (10-12 нуклеотидов). Амплификация фрагментов ДНК осуществляется с использованием единичного короткого праймера, который при низкой температуре отжига связывается с геномной ДНК в двух различных участках инвертированных повторов. Поскольку таких участков в геноме, как правило, большое количество, то в результате реакции образуется множество ПЦР-продуктов размером от 100 до 5000 п.н. Эти продукты при электрофоретическом разделении образуют специфические для отдельных растений или популяций ДНК-паттерны. Большинство RAPD-маркеров являются доминантными (наличие/отсутствие полосы в ДНК-паттерне). Другим недостатком RAPD является слабая воспроизводимость, обусловленная короткой длиной праймеров и низкой температурой их отжига. RAPD-анализ служит своеобразным экспресс-методом для выявления генетического полиморфизма у малоизученных таксономических групп.

Метод **AFLP** (*amplified fragment length polymorphism*) является комбинированным, сочетая в себе гидролиз геномной ДНК рестриктазами (*EcoRI* и *MseI*), лигирование с адаптерами, и двумя последовательными этапами ПЦР. Первый этап ПЦР называется преамплификация, для него используют праймеры, полностью комплементарные к

адаптерам *EcoRI* и *MseI*. При этом образуется большое количество продуктов амплификации, которые невозможно дифференцировать с помощью электрофореза. Во втором этапе ПЦР используют частично комплементарные праймеры, для селективной амплификации. Праймеры содержат радиоактивную или флуоресцентную метку, что позволяет визуализировать фрагменты ДНК после электрофореза в полиакриламидном геле. Получаемый паттерн фрагментов (фингерпринт) для каждого образца ДНК обычно высоко полиморфен и хорошо воспроизводим. AFLP-маркеры также имеют доминантный тип наследования. AFLP-маркеры были успешно использованы для геномного картирования, в популяционных и филогенетических исследованиях. В последние годы эти маркеры используют для исследования мало изученных таксономических групп.

В методе *ISSR (Inter-Simple Sequence Repeat)* используется один или несколько праймеров длиной 15-24 нуклеотида. При этом праймеры состоят из тандемно повторяющихся коротких 2-4 нуклеотидных повторов и несут на 3'-конце последовательность из двух-четырех произвольных нуклеотидов (так называемый "якорь"). Такие праймеры позволяют амплифицировать фрагменты ДНК, которые находятся между двумя достаточно близко расположенными микросателлитными последовательностями (как правило, это уникальная ДНК). В результате амплифицируется большое число фрагментов, представленных на электрофореграмме дискретными полосами (ISSR-фингерпринтинг). Полученные паттерны ПЦР-продуктов видоспецифичны. ISSR-маркеры также относятся к маркерам доминантного типа наследования, полиморфизм которых тестируется по наличию/отсутствию полосы. Аналогично RAPD и AFLP для создания ISSR-маркеров не требуется предварительного знания нуклеотидной последовательности исследуемой ДНК. Метод обладает хорошей воспроизводимостью и наряду с AFLP может быть с успехом использован для выявления межвидовой и внутривидовой генетической изменчивости, идентификации видов, популяций, линий, а в ряде случаев и для индивидуального генотипирования. ISSR-маркеры могут быть использованы также для картирования геномов и маркирования хозяйственно-полезных признаков.

Информация о нуклеотидной последовательности различных групп повторов дает возможность разрабатывать неанонимные ДНК-маркеры, которые локализованы в различных участках хромосом растений. Одним из самых широко применяемых типов маркеров являются микросателлитные, или SSR-маркеры. SSR-маркеры представляют собой пару ПЦР-праймеров к участкам генома, фланкирующих участок микросателлитного повтора (сами праймеры при этом в своей структуре повторов не содержат). **SSR (simple sequence repeats)** – это участки простых повторяющихся последовательностей (микросателлитов), мотивы которых могут состоять из 4, 3, 2 и даже одного нуклеотида;

имеют общую длину, как правило, не более 100 п.н. Они встречаются повсеместно в геномах высших растений: в составе некодирующих последовательностей, регуляторных районов генов, а иногда и внутри генов. Средняя частота встречаемости – каждые 50 т.п.н. Частота встречаемости и распределение микросателлитов по хромосомам различается у растений различных таксономических групп. Источник высокой степени полиморфизма этих последовательностей – в сайт-специфическом варьировании длины кластера повторов, что, в свою очередь, обусловлено различием в числе повторяющихся единиц. Согласно общепринятой точке зрения, данный полиморфизм обусловлен ошибками (эффект "проскальзывания") в процессе репликации или репарации ДНК. Высокий уровень полиморфизма микросателлитов, относительно равномерное их распределение в эухроматиновой части геномов, широкая представленность и кодоминантный характер наследования сделали их чрезвычайно популярными для картирования, в исследованиях генетического полиморфизма. Гетерозиготность их очень высока (часто более 75%). Эксперименты по изучению генетического разнообразия культурного риса обнаружили до 25 аллелей на один микросателлитный локус; при этом общее количество микросателлитных последовательностей составляет порядка 100 000. Около 10 000 SSR-маркеров были локализованы на генетической карте риса.

Несмотря на высокую популярность SSR-маркеров, они имеют и некоторые недостатки. Неравномерность скорости мутирования разных микросателлитов создает определенные сложности для популяционно-генетического анализа. Имеются и технические проблемы, такие как артефакты при проведении ПЦР (за счет эффекта "проскальзывания"), сложности в разработке технологий для автоматического скрининга микросателлитных аллелей. Кроме того, несмотря на высокую плотность микросателлитных локусов в геноме, их бывает недостаточно для тонкого картирования отдельных областей геномов, создания маркеров для локусов количественных признаков (QTL) и решения многих других задач.

Контрольные вопросы по теме 1.6

- 1) Как проводится RFLP анализ?
- 2) Какая информация нужна для разработки RAPD маркеров?
- 3) Какая отличительная особенность AFLP анализа?
- 4) Чем различаются ISSR и SSR методы анализа генома растений?

ТЕМА 1.7. МАРКЕРЫ НА ОСНОВЕ МОБИЛЬНЫХ ЭЛЕМЕНТОВ

Мобильные элементы имеют высокую представленность в геномах растений, составляя от 50 до 90% последовательностей. Подавляющее большинство МЭ в геномах современных растений представлено неактивными дегенерированными копиями, подчиняющимися механизму нейтральной эволюции. Поэтому эти последовательности можно использовать для маркирования геномов и приблизительной оценки филогенетических расстояний между различными таксонами. На основе последовательностей МЭ разрабатываются ПЦР-праймеры для следующих маркерных систем: SSAP (Sequence Specific Amplification Polymorphism), IRAP (Inter-Retrotransposon Amplified Polymorphism), REMAP (Retrotransposon-Microsatellite Amplified Polymorphism), ISBP (Insertion Site Based Polymorphism).

SSAP (Sequence Specific Amplification Polymorphism) является модификацией метода AFLP. ДНК исследуемых образцов расщепляется рестриктазами *Pst*I и *Mse*I, затем лигируется с адаптерами, затем проводится селективная ПЦР, в которой используется меченый праймер, разработанный к последовательности LTR (длинному концевому повтору ретротранспозона), и второй праймер – к одному из адаптеров. В результате амплификации образуются фрагменты ДНК, локализованные между LTR и ближайшим адаптером.

IRAP (Inter-Retrotransposon Amplified Polymorphism) – полимеразная цепная реакция между праймерами, комплементарными последовательностям двух рядом расположенных LTR ретротранспозонов.

REMAP (Retrotransposon-Microsatellite Amplified Polymorphism) – впервые описан как метод ДНК-фингерпринтинга у ячменя. Для ПЦР-анализа используют следующую комбинацию праймеров: к фрагменту LTR-ретротранспозона и рядом расположенному микросателлитному повтору (SSR-праймер). К достоинствам метода относятся относительная простота и потенциально большое количество потенциальных комбинаций праймеров, учитывая большое количество различных ретротранспозонов и вариантов микросателлитов.

ISBP (Insertion Site Based Polymorphism) – данный тип маркеров использует информацию о месте встройки одного мобильного элемента в другой. В геноме растений кластеры МЭ представляют собой множественные инсерции одного МЭ в другой, и в этом процессе могут участвовать различные семейства МЭ, встройки могут происходить в разные сайты – таким образом, большая часть случаев инсерции является уникальной в геноме. Для разработки ISBP-маркеров необходима информация о нуклеотидной

последовательности областей встройки, источником такой информации может быть, как концевое секвенирование ВАС-клонов (участки генома длиной около 100 т.п.н., клонированные в вектор на основе искусственной бактериальной хромосомы), так и случайное (shotgun) секвенирование коротких фрагментов генома. Специфичные праймеры к двух граничащим МЭ подбираются с таким расчетом, чтобы амплифицировать конкретное место инсерции в виде фрагмента известной длины. Этот тип маркеров является относительно новым, появился в 2010 г. благодаря бурному развитию технологий секвенирования, и является дорогостоящим в разработке. ISBP-маркеры были успешно использованы для построения физических карт хромосом при осуществлении проекта по секвенированию генома пшеницы.

Контрольные вопросы по теме 1.7

- 1) Для каких целей используются маркеры на основе мобильных элементов?
- 2) Какие используются маркеры на основе мобильных элементов?
- 3) Какой тип маркеров базируется на информации о месте встройки одного мобильного элемента в другой?

ТЕМА 1.8. МАРКЕРЫ НА ОСНОВЕ ОДНОНУКЛЕОТИДНЫХ ЗАМЕН.

SNP (Single Nucleotide Polymorphism) – однонуклеотидные позиции в геномной ДНК, для которых в популяции существуют различные варианты последовательностей (аллели) с частотой редкого аллеля не менее 1%. В базы данных SNP обычно включают все небольшие изменения геномных последовательностей (небольшие инсерции/делеции, так называемые indels, изменения нескольких нуклеотидов), хотя формально они не соответствуют определению SNP. Обычно SNP представлены двумя аллельными вариантами, хотя встречаются и трехаллельные SNP. SNP чрезвычайно широко распространены в геномах всех организмов. Приблизительная частота их встречаемости один на 1000 нуклеотидов. Никакой другой тип геномных различий не способен обеспечить такую плотность маркеров. Помимо высокой плотности, SNP имеют низкий уровень мутаций на поколение ($\sim 10^{-8}$), что делает их удобными маркерами при оценке крупномасштабных эволюционных событий. Самый точный метод детекции SNP – секвенирование интересующего участка генома. Однако при массовом анализе образцов этот метод дорог и трудозатратен.

Основным достоинством SNPs является возможность использования автоматических методов их детекции, например, использование ДНК-микрочипов

(microarrays). Существующие способы тестирования SNPs можно условно разделить на несколько групп.

CAPS (Cleaved Amplified Polymorphic Sequence). Этот метод относится к ферментативным методам анализа SNPs и аналогичен методу RFLP, но в его основе лежит использование ПЦР. Гидролизу одной или несколькими рестриктазами в этом случае подвергаются продукты амплификации, а не геномная ДНК. Благодаря простоте и надежности метод получил широкое распространение и до сих пор используется для анализа аллельного полиморфизма генов у самых разных объектов. Ограничением метода является то, что с его помощью можно тестировать только известные мутации, затрагивающие сайты рестрикции. CAPS-маркеры считаются вторичными маркерами, так как праймеры для них создаются после выявления SNP на основе анализа нуклеотидной последовательности.

Аллель-специфичная ПЦР. Аллельные варианты различаются за счет того, что 3'-концевой нуклеотид одного из праймеров гибридизуется непосредственно с переменным нуклеотидом (позиция SNP), что обуславливает наличие или отсутствие ПЦР. Специфичность реакции можно повысить, вводя дополнительный, не спаренный нуклеотид во второй или третьей позиции с 3'-конца этого же праймера или используя конкурирующую ПЦР в той же пробирке. Аллель-специфичная ПЦР широко используется для типирования отдельных аллелей генов, или групп аллелей, несущих одинаковые мотивы.

KASP-генотипирование (Kompetitive Allele Specific PCR) – метод, основанный на конкурентной аллель-специфической ПЦР. Позволяет эффективно выявлять биаллельные SNPs а также индели в определенном локусе, подходит для анализа гетерозиготных растений. К образцу ДНК добавляют смесь специфичных к SNP праймеров, и реакционную смесь с флуоресцентной меткой, затем проводят ПЦР с последующим считыванием флуоресценции по конечной точке. Различение аллелей достигается за счет конкурентного связывания аллель-специфичных праймеров.

SSCP, одонитевой конформационный полиморфизм ДНК (SSCP, Single Strand Conformation Polymorphism) эффективен для поиска новых одонуклеотидных замен. Метод SSCP заключается в следующем: после проведения PCR продукт амплификации денатурируют, и полученные одонитевые молекулы разделяют с помощью электрофореза в полиакриламидном геле. Любая нуклеотидная замена в исследуемой последовательности приводит к изменению заряда одонитевого фрагмента, и, следовательно, его электрофоретической подвижности, и может быть тестирована в качестве полиморфного варианта. Этот подход позволяет выявлять нуклеотидные замены, делеции, инсерции на

протяжении всей длины амплифицированного продукта, а не только в местах узнавания рестриктаз, что делает этот подход более информативным. Простота метода, умеренная стоимость и возможность использования флуоресцентной метки и капиллярного электрофореза (ABI Prism 310) позволяет легко автоматизировать процесс и анализировать одновременно большое количество локусов.

Анализ гетеродуплексов. В этом случае денатурированным контрольному и анализируемому ПЦР-продуктам сначала позволяют ренатурировать. Одноцепочечные фрагменты образуют гомодуплексы, если последовательности двух цепей полностью комплементарны, или гетеродуплексы, если они отличаются хотя бы на один нуклеотид. Гомо- и гетеродуплексы имеют разную электрофоретическую подвижность, что используется для их разделения. Для тестирования гетеродуплексов могут быть использованы химические методы расщепления, обеспечивающие практически полное узнавание всех типов гетеродуплексов, что недоступно при использовании современных ферментативных подходов. К недостаткам метода следует отнести сложность процедуры и токсичность используемых реагентов.

Детекция на твёрдых подложках (микрочипах, *microarrays*). Микрочип - это твердый носитель небольшого размера (например, стекло, нейлон или металлическая пластинка) с прикрепленными к нему в определенном порядке короткими олигонуклеотидами или фрагментами ДНК. Прикрепление одного из компонентов реакции к подложке позволяет проводить параллельно множество реакций, пространственно разделив детекцию отдельных SNP. Микрочипы могут применяться как для скрининга мутаций, так и для детекции известных аллелей. О наличии/отсутствии мутации судят по уровню сигнала гибридизации тестируемой ДНК с иммобилизованными на матрице фрагментами. В случае поиска неизвестных SNP на матрице иммобилизуются олигонуклеотиды, содержащие в каждой позиции любой из четырех дезоксирибонуклеозидтрифосфатов (VDA, variant detector array). Олигонуклеотид, полностью комплементарный флуоресцентно-меченому зонду, даст более сильный сигнал, чем отличающиеся от него в одной из позиций. Скрининг уже известных SNP значительно упрощается, так как в каждом случае нужны только те олигонуклеотиды, которые соответствуют известным аллельным вариантам. Дальнейшее усовершенствование метода идет в сторону увеличения количества и плотности расположения олигонуклеотидов на носителе, а также повышения чувствительности и специфичности гибридизации. Существует две системы SNP-чипов: Illumina Infinum и Affymetrix, которые различаются по способу детекции SNP. Чипы SNP разработаны для многих видов сельскохозяйственных растений: картофель (2015), хлопок (2015), арахис (2017) и др. Целый ряд чипов разработан

для анализа геномов пшениц: Illumina Wheat 9K iSelect SNP, Illumina Wheat 90K iSelect SNP, Wheat 50K Triticum TraitBreed, Axiom Wheat 660K SNP и др. Чип Axiom Wheat 660K содержит 660 тысяч SNP с известной физической позицией.

К основным преимуществам ДНК-чипов относятся: возможность использования сравнительно малого количества исходного материала; надежность; удобство маркеров в использовании благодаря высокой плотности и эволюционной стабильности однонуклеотидных полиморфизмов; высокая информативность; возможность полной автоматизации анализа; подходят для геномной селекции; кодоминантный тип наследования. Недостатки ДНК-чипов: наличие технических трудностей, связанных с выявлением SNP; дороговизна создания (что в конечном счете перекрывается экономической выгодой в использовании); необходимость знания последовательностей и фланкирующих областей, а также смещение выборки (чип не позволяет выявлять редкие варианты аллелей).

DArT (Diversity Array Technology) – ДНК-чиповый метод исследования генетического полиморфизма. Данная система применяется для мониторинга состава популяций, изучения генетического разнообразия, идентификации сортов, локусов количественных признаков (QTL) и маркеров для селекции, высокопроизводительного генотипирования, а также составления генетических карт. DArT-технология позволяет осуществлять одновременную детекцию нескольких тысяч ДНК-полиморфизмов, которые возникают за счет наличия однонуклеотидных замен либо коротких инсерций/делеций. Подготовка DArT-чипа происходит следующим образом: референсная ДНК подвергается рестрикции и лигированию, затем полученный материал разделяют на полиморфные и неполиморфные фрагменты; полиморфные фрагменты, отобранные для создания геномной библиотеки, амплифицируют с последующей окраской синим флуорофором и нанесением на стеклянные диски. Полученные чипы используют для гибридизации с целевой ДНК, которая помечена зеленым флуорофором. Далее идет сканирование чипов и измерение интенсивности флуоресцентного свечения для каждого маркера. DArT-технология отличается от SNP тем, что для разработки чипов не требует данных о последовательности генома. По сравнению с AFLP и SSR она является более точной и эффективной, однако применение DArT-технологии требует применения сложного оборудования.

Контрольные вопросы по теме 1.8

1. Какие типы маркеров основаны на детекции SNP?
2. На чем основан метод SSCP анализа?
3. Какие существуют методы детекции однонуклеотидных полиморфизмов?

ТЕМА 1.9. МАРКЕРЫ К ГЕНАМ УСТОЙЧИВОСТИ И ХОЗЯЙСТВЕННО-ЦЕННЫМ ПРИЗНАКАМ ПШЕНИЦЫ

Основные грибные заболевания пшеницы, которые наносят наибольший вред во многих регионах мира, вызываются облигатными биотрофными грибами, такими как бурая (возбудитель *Puccinia triticina* Eriks.), стеблевая (*Puccinia graminis* Pers.: Pers. f. sp. *graminis* Eriks. & E. Hann) и желтая ржавчина (*Puccinia striiformis* Westend.) и мучнистая роса (*Blumeria graminis*). Возбудители грибных болезней приводят к значительным потерям урожая пшеницы и снижению качества зерна, особенно в период эпифитотий. Из-за трудностей, возникающих при фенотипической идентификации генов иммунитета, молекулярные маркеры оказались широко востребованными в генетических исследованиях и селекции пшеницы для решения следующих задач: 1) выявления фрагментов интрогрессий, содержащих локусы устойчивости; 2) идентификации генов устойчивости в образцах с различным генетическим окружением и с различным числом генов; 3) мониторинга переноса генов от донора к реципиенту при создании нового селекционного материала. В настоящее время в геноме пшеницы и ее сородичей идентифицировано более 80 генов устойчивости пшеницы к бурой (листовой) ржавчине (символ Lr), более 60 генов устойчивости к стеблевой ржавчине (символ Sr) и более 50-ти генов устойчивости к мучнистой росе (символ Pm) (<http://www.shigen.nig.ac.jp/wheat/komugi/genes/>). Часть этих генов была картирована с помощью молекулярных маркеров, тесно сцепленных с изучаемыми генами, и диагностическая эффективность маркеров проверена на линиях, контрастных по содержанию генов устойчивости к болезням.

Другой тип маркеров, используемый для изучения и идентификации генов иммунитета к грибным болезням, основан на последовательностях клонированных генов. На данный момент у различных видов растений было клонировано более 50 генов резистентности. Большая часть этих генов относится к классу NBS-LRR (nucleotide binding site-leucine rich repeats, сайты связывания нуклеотидов–лейцин насыщенные повторы) и кодирует белки, действующие как рецепторы на пути передачи сигнала, появляющегося в ответ на воздействие патогена. NBS-LRR домен содержит короткие консервативные последовательности, на основании которых разработаны маркеры RGAs (resistance gene analogs, аналоги генов резистентности). Ряд таких маркеров был использован для картирования генов RGA пшеницы.

Маркеры к другим генам, определяющим хозяйственно значимые признаки, также разрабатываются или на основании их тесного сцепления с изучаемым признаком, или по данным о первичной структуре генов.

Несмотря на то, что многие маркеры к генам, определяющим устойчивость к заболеваниям или хозяйственно-ценные признаки, были проверены на расширенном наборе сортов, селекционных линий и генотипов, только небольшое их число используется на практике в селекционных программах, что часто связано с большой дистанцией от гена до маркера. Поэтому основная задача состоит в идентификации первичной структуры как можно большего числа генов, а если это пока невозможно, то в насыщении фрагмента, содержащего «целевой» ген, как можно большим числом маркеров и уменьшении дистанции между маркерами и геном, что особенно важно в случае близкородственных скрещиваний.

Разработанные ДНК-маркеры, предлагаемых для диагностики генов устойчивости и использования в схемах маркер-ориентированной селекции зерновых культур MAS (marker-assisted selection) (<http://maswheat.ucdavis.edu/>). <http://wheatheadingtime.tilda.ws/>, <https://masrusplants.ru/>

Контрольные вопросы по теме 1.9

1. В каких исследованиях востребованы молекулярные маркеры к генам устойчивости к грибным болезням?
2. Какие символы используются для обозначения генов устойчивости к листовой и стеблевой ржавчине?
3. На основании каких данных разрабатываются маркеры к генам, определяющих устойчивость к заболеваниям и хозяйственно-ценные признаки?

ТЕМА 1.10. ГЕНЕТИЧЕСКОЕ И ФИЗИЧЕСКОЕ КАРТИРОВАНИЕ ГЕНОВ И ЛОКУСОВ КОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ, АССОЦИАТИВНОЕ КАРТИРОВАНИЕ

Генетическая карта – одномерная схема взаиморасположения генетических маркеров в группах сцепления, которые соответствуют индивидуальным хромосомам данного организма с указанием расстояний между ними в сантиморганах, установленных по частоте кроссинговера, однако полученные при этом значения расстояний между ними не соответствуют реальным физическим расстояниям. Поэтому генетические карты можно рассматривать только в качестве первого приближения к реальным физическим картам.

Метод построения генетических карт называется генетическим картированием. При этом устанавливают последовательность расположения различных генетических маркеров (в числе которых могут быть как гены, так и полиморфные последовательности ДНК) по длине всех хромосом. Первая карта сцепления, построенная для пшеницы при использовании RFLP-маркеров содержала в 1.5 раз больше локусов и была в 1.2 раза длиннее прежней классической генетической карты. В зависимости от количества и равномерности локализованных ДНК-маркеров определяется уровень разрешения генетической карты.

Физическое картирование хромосом – установление точной локализации последовательностей ДНК на хромосомах, которое включает в себя точные данные о расстоянии между ними (изменяется в парах нуклеотидов). Физические карты являются важным условием для осуществления проектов полногеномного секвенирования (создание референсной последовательности генома), для позиционного клонирования генов, а также обеспечивают фундамент для улучшения сельскохозяйственных признаков в селекционных программах. В настоящее время термин «физическое картирование» чаще используется для обозначения локализации протяженных ДНК-контигов (групп перекрывающихся последовательностей) в определенном участке хромосомы. Большой размер генома растений требует фрагментации хромосом с образованием набора протяженных (свыше 100 т.п.н.) последовательностей, клонированных в искусственные бактериальные ВАС (Bacterial Artificial Chromosome) вектора – совокупность таких клонов, содержащая целый геном (хромосому), называется геномной (хромосомной) библиотекой. Затем с использованием метода фингерпринтинга клоны геномных библиотек группируют в контиги, которые затем с помощью цитогенетического картирования локализуют («привязывают», anchoring) на хромосоме. Для осуществления цитогенетического картирования используются различные генетические ресурсы растений, включающие в себя наборы анеуплоидных, замещенных, интрогрессивных и делеционных линий.

Для привязки контигов и отдельных клонов к хромосомам используют наборы различных ДНК-маркеров: SSR-маркеры, SNP- и ISBP-маркеры. Маркер, локализованный на хромосоме с помощью генетического и цитогенетического картирования, используется для ПЦР-скрининга ВАС-библиотеки, с целью выделения ВАС-последовательности, которая является физическим носителем данного маркера и представляет собой репрезентацию протяженного участка хромосомы, содержащего этот маркер. Также возможно дополнительная разработка маркеров на основе ВАС-клонов – с целью расширения пула маркеров, а также для разработки маркеров непосредственно к ВАС-клонам для их последующей локализации на хромосоме. Для этого используются

различные методы частичного или полного секвенирования ВАС-клонов – концевое секвенирование по Сэнгеру, а также высокопроизводительное секвенирование. Физическое картирование хромосом используется при построение первой референсной последовательности генома выбранного стандартного образца изучаемого вида (так называемого «золотого стандарта») и чаще используется в случае построение референсного генома для полиплоидных видов растений – например, использовалось Международным Консорциумом является «золотым стандартом» по секвенированию генома пшеницы (IWGSC) для получения референсной последовательности гексаплоида *T. aestivum* ($2n = 6x = 42$).

Локусы количественных признаков (**Quantitative Trait Loci, QTL**) - это районы гена/генов, влияющих на признак, измеряемый по линейной количественной шкале. QTL идентифицируют статистическими методами, интегрируя данные о позиции маркера на карте сцепления с генотипическими и фенотипическими данными. Принцип картирования QTL появилась состоит в том, что если генетические маркеры рассредоточены по всему геному изучаемого организма, то сегрегация этих маркеров может быть использована для оценки эффектов, связанных с QTL. При этом возникает возможность картирования и оценки свойств отдельных QTL. Для картирования QTL необходимы условия сегрегации и рекомбинации. Метод включает в себя поиск ассоциаций между сегрегирующими молекулярными маркерами и изучаемым признаком в экспериментальной популяции, для определения сцепления между маркерами и QTL. Исходными данными для анализа являются расщепляющиеся популяции F₂, поколения возвратных скрещиваний (BC), популяции рекомбинантных инбредных линий (РИЛ), полученные из единственного семени F₂ и близкие к гомозиготности после 5-6 поколений самоопыления, а также серии двойных гаплоидных линий (ДГЛ) или любых расщепляющихся по количественным признакам и по маркерным локусам поколений.

Отличительной чертой проведения генетического картирования и установления ассоциаций маркер/признак, является оперирование большими объемами данных, полученных в результате экспериментальной оценки различных классов генетических маркеров и различного рода картирующих популяций. Для оперативного решения этой проблемы на сегодняшний день существует большое число статистических методологий и подходов, а также информационно-программных средств, подходящих для анализа различных типов популяций и маркеров. Первым и наиболее часто используемым программным обеспечением для построения карт является MAPMAKER, которое было разработано в 1987 году. Это ПО основывается на установлении максимального правдоподобия сцепления между маркером и фенотипом, используя для этого интервальное

картирование, которое имеет дело с несложным QTL и несколькими стандартными популяциями. С тех пор появилось множество пакетов программ, предназначенных для установления ассоциаций между маркерными генотипами и фенотипами признаков: QTL Cartographer, MAPQTL, PLABQTL, QGENE и др. Эти программы сравнивают полученные расщепления и для каждого района хромосомы вычисляют «степень влияния» на количественный признак. В некоторых районах этот параметр может быть выше, достигая максимума в определенных точках. Если значение «степени влияния» в такой точке превышает порог достоверности, говорят о том, что в данной области располагается ген, влияющий на количественный признак. Обычно таких генов находят от 2 до 10, причем степень их воздействия на признак оказывается различной.

Методы картирования QTL были успешно использованы на большинстве сельскохозяйственных культур, включая как томаты, кукуруза, арабидопсис, горох, несмотря на разную степень изученности их геномов и типы молекулярных маркеров, составляющих генетические карты данных объектов.

Одним из необходимых требований, на основе которого строятся современные генетические карты – это наличие ассоциаций маркер-признак и неравновесного сцепления (**LD - linkage disequilibrium**) между ними. LD связано с эволюционно неравным и разнородным уровнем рекомбинации специфичных аллелей в различных локусах, определяющих проявление изучаемой генетической изменчивости в популяции. Неравновесное сцепление может быть установлено статистически и его выявление широко используется для картирования и клонирования генов, определяющих сложные признаки у животных и растений. С изменением направления вектора усилий по идентификации и локализации генов/локусов хромосом от двуродительского скрещивания к естественным популяциям линий или образцов коллекций генетических ресурсов растений, и от традиционного QTL-картирования к LD-ассоциативному анализу, исследователи получили новый мощный инструмент по картированию интересующих их генов. TASSEL – это универсальная программа для анализа признаков посредством установления ассоциаций, эволюционных отношений и сцепления, которая выполняет различные генетические анализы, включая ассоциативное картирование, установление разнообразия и расчет неравновесного сцепления. Анализ LD между генотипами и фенотипами может быть выполнен либо посредством общей линейной модели, либо смешанной линейной модели. Общая линейная модель позволяет пользователям анализировать сложные (комплексные) полевые исследования, взаимодействия с окружающей средой и эпистаз. Смешанная модель специально разработана для обработки полигенных эффектов на множественных уровнях связанностей, включая информацию

родословной. Эти анализы позволяют проводить установление и исследование LD у многих видов растений.

Контрольные вопросы по теме 1.10

1. В чем различие генетического и физического картирования?
2. Какие программы используются при генетическом картировании на двуродительских популяциях?
3. На чем базируется ассоциативного картирование?

ТЕМА 1.11. ПОЗИЦИОННОЕ КЛОНИРОВАНИЕ ГЕНОВ. ИДЕНТИФИКАЦИЯ АЛЛЕЛЬНЫХ ВАРИАНТОВ ГЕНОВ

Позиционное клонирование – подход для целевого выделения последовательности гена, для которого отсутствует идентифицированная мРНК или белок. Для осуществления позиционного клонирования используют тесно сцепленные с геном ДНК-маркеры, которые выявляют с помощью методов картирования с высоким разрешением. Затем происходит привязка этих маркеров к физическим «носителям» генов – ВАС- (или YAC) клонам геномных библиотек, с использованием метода «прогулки по хромосоме». К целевому гену «приближаются» с помощью скринирования (поиска перекрывающихся друг с другом ВАС-клонов), таким образом выявляют контиг предполагаемых носителей целевого гена. Однако у высших растений этот подход может занимать значительное количество времени из-за размеров их геномов. Даже если маркер находится на расстоянии 1сМ от искомого гена, физически это может составлять несколько сотен или тысяч п.н., которые почти невозможно полностью проскринировать из-за большого количества перекрывающихся ВАС-клонов, соответствующих данному участку.

«Прыжки по хромосоме» в определенном смысле являются усовершенствованием этого подхода, но методология поиска маркеров, основанная на использовании почти-изогенных линий (NIL), совместно с точными картирующими методами, представляет собой более сильное орудие для поиска и клонирования генов. Идея заключается в том, что на практически изогенных линиях можно проскринировать несколько тысяч маркеров и найти среди них те, которые находятся на генетической дистанции, соответствующей физическому размеру ВАС-клона. При скринировании библиотек с большими фрагментами ДНК с помощью молекулярных маркеров, как минимум для видов с небольшими геномами, можно совершенно точно отобрать клоны, содержащие целевые гены. Такой подход порой называют «посадка на хромосому» (chromosome landing).

Как правило, отобранный клон содержит множество генов. Одним из способов выявить подходящих кандидатов является биоинформатический анализ нуклеотидной последовательности выявленного клона. Клон также может быть использован для скрининга библиотек кДНК, что приведет к получению дополнительных генов-кандидатов. Для окончательного установления гена, связанного с изучаемым признаком, самым прямым и действенным методом является трансформация. Проведение генетической трансформации позволяет получить практически однозначный ответ о функциональном назначении и роли выявленного гена-кандидата в проявлении изучаемого признака или о сцеплении с ним. И наконец, для некоторых признаков, действенным способом является изучение экспрессии гена-кандидата. С помощью позиционного клонирования были клонированы: ген *ABI3*, ответственный за уменьшение воздействие абсцизовой кислоты у *Arabidopsis*, ген, кодирующий омега-3-десатуразу арабидопсиса, ген *Pto* устойчивости к *Pseudomonas syringae* томата, ген *Axr1* нечувствительности арабидопсиса к ауксину и многие другие.

После постулирования гена-кандидата, для установления связи аллельных вариантов генов хозяйственно ценных признаков с желаемыми качественными (или количественными признаками), необходимо идентифицировать существующие аллельные варианты генов. Так, показана связь полиморфизма нуклеотидной последовательности четвертого экзона гена *VRN-A1* с вариабельностью количественных показателей рядом агрономически ценных признаков озимой пшеницы (таких как чувствительность к продолжительности яровизации, морозостойкость, время колошения). Для идентификации аллельных вариантов требуется секвенирование участка изучаемого гена из разных образцов (линий) – либо секвенирование ПЦР-фрагментов, либо ВАС-клонов, содержащих целевые гены. Применение ДНК-маркеров позволяет не только отслеживать передачу целевых генов, но и проводить выбраковку нежелательных аллелей, чтобы ускорить процесс отбора.

Контрольные вопросы по теме 1.11

1. На каких принципах основан метод позиционного картирования генов?
2. Какими методами можно верифицировать отобранный ген-кандидат?
2. Как определить аллельные варианты гена?

ТЕМА 1.12. МЕТОДЫ И ПОДХОДЫ К СЕКВЕНИРОВАНИЮ ГЕНОМА РАСТЕНИЙ

Последовательность генома является ключом к пониманию биологических процессов, происходящих в организме, а также обеспечивает фундамент для разработки инструментов для генетической селекции и улучшения сельскохозяйственных растений. За последние 20 лет прогресс в секвенировании геномов растений привел к увеличению качества и количества публично доступных геномных ресурсов. Сборка генома модельного объекта *Arabidopsis thaliana* была опубликована в 2000, с тех пор были секвенированы и собраны геномы сотни видов, а данные были помещены в различные репозитории (например, GenBank). Среди сельскохозяйственных растений секвенированы геномы риса, пшеницы, ячменя, винограда, сои, кукурузы, картофеля, и др. Целью проектов по секвенированию геномов является получение референсной последовательности генома, то есть набора последовательностей молекул ДНК для всех его хромосом с суммарной длиной, равной длине гаплоидного генома. Аннотированную сборку последовательности ДНК, соответствующую определенной хромосоме, называют «псевдохромосомой». Поскольку для геномов растений характерно высокое содержание повторяющихся последовательностей, то секвенирование и сборка этих геномов представляет собой нетривиальную задачу, и требует применения других подходов, кроме непосредственно секвенирования (например, физическое и генетическое картирование хромосом).

Методы секвенирования условно делятся на три поколения. Первое поколение, «классические» методы – метод Сэнгера, интенсивно использовавшийся для первых проектов по секвенированию геномов растений. Метод Сэнгера был разработан в 1977 г. и заключается в терминации растущей цепи ДНК включением меченого (флюоресцентной меткой) дидезоксинуклеотида. Набор получившихся фрагментов разной длины затем анализируют с помощью капиллярного секвенатора. Недостаток этого метода в том, что для при проведении одной реакции по методу Сэнгера можно секвенировать одну последовательность ДНК длиной до 1000 п.н. (т.е. низкая производительность), высокие требования к последовательностям ДНК для секвенирования (количество, качество), и высокая стоимость. Тем не менее, данный метод активно использовался для полногеномного секвенирования небольших геномов (бактериофаг φX174 - 5300 п.н., бактерия *Haemophilus influenzae* 1 830 137 п.н.), а также в проектах по секвенированию геномов растений – *Arabidopsis* в 2001 г., ВАС-библиотеки других растений (рис, 2002 г., сорго, 2008 г., соя 2010 г.). В настоящее время используется в условиях лабораторий для секвенирования отдельных коротких последовательностей.

В начале 21 века был разработан ряд методов секвенирования второго поколения – так называемые современные методы параллельного высокопроизводительного

секвенирования (NGS - next generation sequencing). Эта группа методов позволяет одновременно получать последовательности множества (миллионов) фрагментов ДНК одновременно. Существует целый ряд платформ, с разной химией секвенирования – 454-пиросеквенирование, IonTorrent, SOLiD, Illumina. Недостаток этой группы методов – короткие длины секвенированных последовательностей (прочтения), например, у 454-пиросеквенирования максимальная длина прочтения 700 п.н., у Illumina – 300 п.н. В настоящее время среди этой группы методов активно используется только секвенирование на платформе Illumina.

Третья группа методов – «новейшего» поколения – это методы одномолекулярного секвенирования, которые позволяют получать длинные прочтения длиной до 30000 п.н. – PacBio, Oxford Nanopore, MinION. Однако для этой группы методов характерна низкая точность секвенирования, плюс определенные требования к приготовлению образцов ДНК для секвенирования (необходимо выделить длинные молекулы).

Разработка современных методов высокопроизводительного секвенирования привела к прорыву в секвенировании геномов растений. С 2012 г. количество секвенированных геномов лавинно увеличилось, было опубликовано свыше 300 полных геномных последовательностей растений. При этом для секвенирования геномов используется сочетание различных методов и подходов. Так, в 2013 г. была опубликована референсная последовательность генома сахарной свеклы (*Beta vulgaris*), полученная при сочетании платформ секвенирования 454, Illumina и Sanger.

Кроме непосредственно получения последовательностей ДНК, огромное значение для полногеномного секвенирования имеет так называемая сборка (ассемблирование) геномов. Из-за большого количества повторов и полиплоидии сборка генома на основе только коротких последовательностей невозможна. Поэтому, по мере развития методов секвенирования ДНК, менялись и совершенствовались подходы к сборке геномов. Когда основной используемой технологией были короткие прочтения (Сэнгер, Illumina), ключевой стратегией было физическое картирование. Для этого протяженные фрагменты геномов клонировали в ВАС-вектора (ВАСs – Bacterial Artificial Chromosome), которые «привязывали» к генетической карте, а затем получившийся набор клонов секвенировали. Более короткие секвенированные последовательности объединяют в «контиги» (группы перекрывающихся последовательностей), которые затем с использованием данных о локализации маркеров на физической и генетической картах объединяют в «скаффолды». Доступность длинных одномолекулярных прочтений позволила осуществлять сборку длинных контигов, а иногда даже целых хромосом, *de novo*, минуя стадию физического картирования. Так, например, сборка референсной последовательности генома мягкой

пшеницы *T. aestivum* ($2n = 6x = 42$), опубликованная в 2017 г. и полученная с использованием физического картирования с помощью ВАС-клонов, покрывала 78% генома. А скаффолды сборки, полученной в 2018 г. с помощью платформ PacBio и Illumina, покрывали 96% генома.

Нуклеотидные последовательности, полученные в результате проектов по секвенированию, помещают в специализированные базы данных. Самая известная из них это GenBank NCBI, содержащая данные по разным объектам (животные и растения, грибы, бактерии). Специализированные базы для сбора информации по геномным данным растений - PlantGDB, EnsemblPlants. В базах данных псевдохромосомы растений представлены в виде так называемого геномного браузера – пользовательского интерфейса, который позволяет выделять отдельные участки хромосом, и осуществлять поиск по номеру идентификатора гена, либо поиск гомологичных последовательностей с помощью алгоритма BLAST.

Нуклеотидные последовательности при помещении в базу данных, как правило, необходимо аннотировать. Аннотирование – это присвоение данных о возможной биологической информации для последовательности ДНК. Таким образом, используя методы биоинформатики (например, алгоритм BLAST, который ищет совпадение с известными последовательностями, или программы для предсказания генов *de novo*) и существующие базы данных, в последовательности идентифицируют участки, соответствующие повторам и предполагаемым генам, обозначают их координаты.

Контрольные вопросы по теме 1.12

1. Какие методы секвенирования относят к первому поколению и на чем они основаны?
2. Чем отличаются между собой методы секвенирования второго поколения?
3. Какие известны методы одномолекулярного секвенирования?

ТЕМА 1.13. ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ СЕКВЕНИРОВАНИЯ В СЕЛЕКЦИИ РАСТЕНИЙ

В настоящее время для многих сельскохозяйственных культур опубликована референсная последовательность генома. Первой культурой с опубликованной в 2002 г. является рис *O. sativa* ($2n = 2x = 12$). В 2007 г. был секвенирован геном винограда *Vitis vinifera* ($2n = 2x = 38$), в 2008 г. – *Carica papaya* ($2n = 2x = 18$), в 2009 - кукуруза, сорго, огурец и др. – диплоидные геномы с относительно небольшим размером. С 2012 г. началось

резкое увеличение количества секвенированных геномов с использованием технологий NGS. Однако сборки геномов, полученные только с использованием коротких прочтений (NGS, или по методу Сэнгера) так называемое «шотган», или секвенирование методом дробовика, были фрагментированы, и состояли в основном из ген-содержащих участков. При этом значительную часть геномов растений составляют повторяющиеся последовательности, а корректная сборка этих участков на основе коротких прочтений невозможна (гомологичные последовательности повторов объединяются в один кластер, даже в случае, если они расположены на разных участках хромосомы). Так же проблему корректной сборки усложняет и полиплоидия – когда в одном ядре присутствует свыше двух гаплоидных геномов. Для решения проблемы полиплоидии и высокого содержания повторов применяются подходы фрагментирования генома (разделение генома или хромосомы на участки длиной 100 т.п.н., и клонирование их в ВАС-вектора), либо применение новейших технологий секвенирования, которые дают прочтения длиной от 20000 п.н.

Референсная последовательность генома обеспечивает фундамент для дальнейших работ по молекулярной генетике и селекции растений. Получение референсных последовательностей геномов растений и развитие технологий секвенирования открывает новый, так называемый «постгеномный» период исследований, который предоставляет ряд новых возможностей – более эффективный поиск генов-кандидатов хозяйственно ценных признаков, разработка новых маркеров, использование полученных данных для геномной селекции. Секвенирование геномов, транскриптомов и микроРНКомов вывели на новый этап исследования о функционировании наследственного материала.

Например, повторное секвенирование (ресеквенирование) других индивидуумов вида (или других родственных видов), и сравнение полученных данных с референсной последовательностью позволяют определять аллельные вариации, разрабатывать новые маркеры для применения в растениеводстве. Пангеном - совокупность всех генов определенной группы организмов - охватывает все варианты разнообразия генов внутри этой группы и обеспечивает платформу для селекции. Как один из значимых результатов работ по секвенированию и ресеквенированию геномов – выявление SNP. Быстрое определение полиморфизма тысяч генов позволило разработать SNP-чипы, которые могут быть использованы для быстрого и относительно недорогого генотипирования больших выборок генетического и селекционного материала растений.

Другим важным применением NGS-секвенирования является РНК-seq - секвенирование набора молекул кДНК, соответствующих транскрибируемому РНК (может включать мРНК, микроРНК, тРНК, рРНК), одно из наиболее значимых применений –

анализ транскриптома растения на определенной стадии развития. Полученные данные формируют транскриптомную сборку. Секвенирование транскриптомов является важным инструментом для анализа экспрессирующейся части генома, и ключевым моментом для понимания функций генов и установления генетических основ важных сельскохозяйственных признаков. Можно сравнивать транскриптомы разных генотипов растений, из разных тканей, клеток, стадий развития, при разных внешних условиях, таким образом характеризовать гены, контролирующие важные признаки.

Экзомное секвенирование – это метод секвенирования только белок-кодирующих участков генов. Экзом составляет 1–2% генома в зависимости от вида и может быть дополнен функциональными некодирующими элементами (например, микроРНК). Секвенирование экзома включает в себя идентификацию генов, разработку праймеров или чипа для обогащения образцов, и последующее секвенирование полученной библиотеки. Этот метод позволяет секвенировать неэкспрессирующиеся аллели и гены, которые не могут быть найдены с помощью РНК-seq. Однако подходы экзомного секвенирования основаны на существовании высококачественных референсных геномных последовательностей с точной аннотацией. Этот метод позволяет выявлять новые SNP – например, для генома гексаплоидной пшеницы с применением целевого ресеквенирования экзома был разработан 10251 SNP-маркер для восьми сортов.

Целевое NGS-секвенирование областей определенных генов заключается в возможности определения последовательности в конкретной области с гораздо более высоким уровнем покрытия по сравнению с секвенированием по методу Сэнгера. Для осуществления этого метода применяется обогащение образцов (отбор интересующей фракции ДНК), которое заключается в выделении специфических локусов генома (генов, молекулярных маркеров) и может проводиться с помощью ПЦР, метода молекулярной инверсии, гибридизационного обогащения (на микрочипах). В 2013 г. вышла работа по идентификации вариантов геномных последовательностей у 84 сортов картофеля. И для томата, и для картофеля использовалось обогащение и секвенирование генов устойчивости NB-LRR, направленное на обнаружение и аннотирование новых генов.

Другим из применений NGS-секвенирования является метод высокопроизводительного генотипирования, так называемый GBS (genotyping-by-sequencing). Впервые он был успешно применен в 2011 г. для генотипирования 276 рекомбинантных инбредных линий кукурузы и 43 дигиплоидных линий ячменя. Геномная ДНК обрабатывается эндонуклеазами рестрикции, создается библиотека фрагментов, при секвенировании которых получают короткие прочтения (~100 п.н.), объединяемые в контиги, выравнивание которых позволяет обнаружить SNP. Преимущество метода

заключается в том, что его можно применять для анализа образцов не только для видов с уже расшифрованным геномом, но и менее изученных видов, для которых еще не выполнена полногеномная сборка. **Метод RAD-seq** (restriction-site-associated DNA sequencing) схож с методом GBS, но отличается протоколом подготовки библиотек. Метод GBS применяется для выявления взаимосвязи между фенотипом и генотипом на основе анализа двуродительских картирующих популяций (**QTL-анализ**) или выборок сортообразцов (**GWAS анализ** – genome wide association studies). Выявленные в результате этих работ геномные районы, ассоциированные с проявлением того или иного селекционно значимого свойства, сразу маркированы определенным SNP. Если геном изучаемого вида уже секвенирован, то информация об обнаруженном районе генома, ассоциированного с признаком, может быстро привлекаться для поиска гена-кандидата, влияющего на изменчивость по данному признаку.

GWAS-анализ используется для массового поиска ассоциаций между фенотипом и генотипом, этот метод базируется на разработанных SNP-чипах, а также на методах генотипирования, основанных на NGS-секвенировании выборочных фракций генома. Преимущество GWAS-метода заключается в высоком разрешении, вплоть до одного нуклеотида. Ассоциированный с признаком аллель детектируется частотой его наличия в популяции. Исследования установленных геномных районов позволяют идентифицировать, валидировать и маркировать гены-кандидаты. Полученная в результате GWAS-анализа информация используется для программ по маркер-ориентированной и геномной селекции.

Контрольные вопросы по теме 1.13

1. Что дает ресеквенирование геномов?
2. Что такое «пангеном»?
3. Какие преимущество имеет GBS метод относительно методов полногеномного секвенирования?

ТЕМА 1.14. МАРКЕР-ОРИЕНТИРОВАННОЙ ОТБОР ДЛЯ СЕЛЕКЦИИ

Маркер-ориентированная селекция (MAS – marker assisted selection) предполагает использование ДНК-маркеров, тесно сцепленных с целевым геном, совместно с фенотипическим анализом. Маркеры, тесно сцепленные с целевым геном, являются

надежным инструментом для предсказания фенотипа. Отбор нужного аллеля целевого гена осуществляется на основе тесно сцепленного с ним аллеля соседнего маркерного локуса. Большая точность отбора достигается при использовании пары маркеров, расположенных вблизи гена по разные стороны от него (т. е. маркеров, фланкирующих маркеров). Если ген секвенирован и выявлены различия нуклеотидной последовательности разных аллелей данного гена, можно разработать так называемый «внутригенный маркер». Использование такого маркера позволит отбирать нужные генотипы с наиболее высокой точностью. При отсутствии внутригенного или тесно сцепленного с геном ДНК-маркера можно использовать более отдаленные маркеры, однако в таких случаях целесообразно сочетать MAS с последующим фенотипированием. Такой комбинированный подход называется «тандемным» отбором (tandem selection), или маркер-направленным фенотипированием (marker-directed phenotyping). MAS хорошо зарекомендовал себя при беккроссной и линейной селекции, а также при пирамидировании генов. При беккроссной селекции можно вести отбор по внутригенному маркеру (foreground selection), по маркерам, тесно сцепленным с геном (recombinant selection), по генетическому фону (background selection). Два первых метода позволяют вести отбор только по целевому гену, контролируя передачу нужного аллеля от донора реципиенту в череде поколений возвратных скрещиваний. Использование значительно большего числа маркеров, равномерно распределенных по геному, позволяет не только контролировать передачу целевого гена от донора реципиенту, но и ускорять восстановление генома реципиента.

Для использования ДНК-маркеров в селекции по тому или иному признаку требуется информация о нуклеотидных последовательностях генов, контролирующих данный признак, или, по крайней мере, о локализации их в геноме, а также о тесно сцепленных с ними маркерах. Если исходные данные отсутствуют, то необходимые подготовительные исследования могут занять не один год. С целью экономии времени и средств можно проводить молекулярно-генетический анализ и отбор одновременно. Такой подход, сочетает метод беккроссной селекции с QTL-анализом и отбором по генотипу.

Контрольные вопросы по теме 1.14

- 1) Для каких целей используется маркер-ориентированная селекция?
- 2) В каких схемах скрещивания используются маркеры?
- 3) Какое число маркеров следует использовать в зависимости от поставленных задач?

ТЕМА 1.15. ГЕНОМНАЯ СЕЛЕКЦИЯ

Развитие методов современного высокопроизводительного секвенирования и снижение его стоимости открыли возможность массовой реализации программ полногеномного секвенирования. Секвенирование и сравнение геномов разных представителей одного и того же вида позволяют выявлять полиморфные участки генома и разрабатывать маркеры (как правило, SNP), равномерно и плотно покрывающие геном. К настоящему моменту разработаны полногеномные SNP-чипы для автоматического анализа полиморфизма ДНК многих видов растений, имеющих сельскохозяйственное значение. Внедрение методов высокопроизводительного генотипирования сельскохозяйственных объектов открыло путь для применения нового метода селекции, основанного на анализе большого числа ДНК-маркеров, равномерно распределенных по геному, – геномной селекции. Помимо видов, чей геном уже секвенирован, появилась перспектива применения геномной селекции и в отношении тех видов растений и животных, геном которых изучен меньше.

Для геномной селекции, в отличие от MAS, не требуются знания о генах, влияющих на признаки, а значит не нужны многолетние генетические исследования, предшествующие селекционному процессу. Также геномная селекция имеет преимущество при отборе по признакам, имеющим сложный полигенный контроль, тогда как MAS, как правило, эффективна лишь в случае моногенного контроля признаков. Процесс геномной селекции включает три этапа: анализ «тренировочных поколений» (training generations) с использованием методов фенотипирования и генотипирования, выявление корреляций между фенотипом и генотипом, дальнейший отбор по генотипу среди «кандидатов на селекцию» (selection candidates). Установлено, что с помощью ДНК-маркеров можно отбирать устойчивые генные сети, сохраняющиеся в поколениях. Однако необходимыми условиями для успешного осуществления геномной селекции являются адекватное количество тренировочных поколений, используемых маркеров и правильное соотношение числа маркеров и исследуемых генотипов. В настоящее время наиболее активно развивается программа по геномной селекции пшеницы (INRA во Франции, CIMMYT в Мексике).

Методы геномной селекции применяются для различных популяций и сортов пшеницы относительно разнообразных признаков: от содержания элементов в зерне и урожайности до устойчивости к заболеваниям (например, вызываемым различными грибами) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5015948/>). Наиболее активно для обработки данных генотип-фенотип и для отбор генотипов среди «кандидатов на селекцию» используется метод BLUP и его различные модификации — rrBLUP, gBLUP,

egBLUP, wBLUP и другие (<https://link.springer.com/article/10.1007/s00122-022-04227-4>, <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-022-08968-w>, <https://pubmed.ncbi.nlm.nih.gov/24518889/>, <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1439-1>). Данный метод зарекомендовал себя за несколько десятилетий использования в селекции растений и животных. Также используется байесовский метод и его модификации — BRR (Bayesian Ridge Regression, байесовская гребневая регрессия), BL (Bayesian Lasso, байесовское лассо), BA (Bayes A), BB (Bayes B), BC (Bayes C). Относительно недавно стали применяться методы машинного обучения и глубокого обучения в геномной селекции пшеницы (<https://www.mdpi.com/2077-0472/12/9/1406>, <https://www.frontiersin.org/articles/10.3389/fpls.2020.613325/full>, <https://access.onlinelibrary.wiley.com/doi/full/10.1002/tpg2.20119>).

Проведенные сравнения результатов между методами, показали, что в целом они пересекаются и общепринятые методы группы BLUP ничем не уступают байесовским методам и методам машинного обучения (<https://www.frontiersin.org/articles/10.3389/fpls.2021.745379/full>, <https://www.nature.com/articles/s41598-020-60203-2>, <https://link.springer.com/article/10.1007/s00122-022-04227-4>).

Контрольные вопросы по теме 1.15

- 1) Какие этапы работы необходимо выполнить в рамках геномной селекции?
- 2) В чем состоит анализ «тренировочных поколений» (training generations)?
- 3) Какие биоинформатические методы используются в рамках геномной селекции?

СПИСОК РЕКОМЕНДОВАННОЙ ЛИТЕРАТУРЫ:






- 1) Бадаева Е.Д., Салина Е.А. Структура генома и хромосомный анализ растений// Вавиловский журнал генетики и селекции, 2013, Том 17, № 4/2 1017-1043
<https://vavilov.elpub.ru/jour/article/view/219/220>
- 2) Брагина М.К., Афонников Д.А., Салина Е.А. Прогресс в секвенировании геномов растений – направления исследований// Вавиловский журнал генетики и селекции. 2019;23(1):38-48 <https://vavilov.elpub.ru/jour/article/view/1866/1177>
- 3) Быкова И.В., Шмаков Н.А., Афонников Д.А., Кочетов А.В., Хлесткина Е. К. Достижения и перспективы использования методов высокопроизводительного секвенирования в генетике и селекции картофеля//Вавиловский журнал генетики и селекции. 2017; 21(1):96-103 <https://vavilov.elpub.ru/jour/article/view/908/883>
- 4) Леонова И.Н. Молекулярные маркеры: использование в селекции зерновых культур для идентификации, интрогрессии и пирамидирования генов// Вавиловский журнал генетики и селекции, 2013, ТОМ 17, № 2. С. 314-325
<https://vavilov.elpub.ru/jour/article/view/153/155>
- 5) Хлесткина Е.К. Молекулярные маркеры в генетических исследованиях и в селекции. Вавиловский журнал генетики и селекции, 2013, Том 17, № 4/2. 1044-1054.
<http://www.bionet.nsc.ru/vogis/download/17-4/2/21Khlestkina.pdf>
- 6) Чесноков Ю. В., Кочерина Н. В., Косолапов В. М. Молекулярные маркеры в популяционной генетике и селекции культурных растений// Москва : ООО «Угрешская Типография», 2019. — 200 с. <https://www.vniikormov.ru/pdf/molekulyarnye-markery-v-populyacionnoj-genetike2.pdf> , <https://elibrary.ru/item.asp?id=38142726>
- 7) Чесноков Ю.В., Косолапов В.М. Генетические ресурсы растений и ускорение селекционного процесса// Москва : ООО «Угрешская типография», 2016. — 172 с. <https://www.vniikormov.ru/pdf/geneticheskie-resursy-rastenii-i-uskorenie-selekcionnogo-protcessa.pdf> , <https://elibrary.ru/item.asp?id=26550132>
- 8) Чесноков Ю.В. Биохимические маркеры в генетических исследованиях культурных растений: применимость и ограничения (обзор). Сельскохозяйственная биология, 2019, том 54, № 5, с. 863-874. <http://www.agrobiology.ru/5-2019chesnokov-html.html>
- 9) Щербань А.Б. Повторяющиеся последовательности ДНК в геномах растений// Вавиловский журнал генетики и селекции, 2014, ТОМ 18, № 4/1, С. 618-629
<https://vavilov.elpub.ru/jour/article/view/289/291>

Материал к лабораторным и практическим занятиям

Лабораторные занятия 1 и 2: Методы анализа ДНК (Выделение ДНК и ПЦР-анализ, оценка ПЦР фрагментов в агарозном геле)

Для выполнения лабораторных работ, связанных с анализом ДНК, может потребоваться следующее лабораторное оборудование и расходные материалы:

Название прибора/расходного материала	Назначение	Фото прибора
Настольная центрифуга для пробирок объемом 0.5-2 мл, Скорость 14000 об/мин	Для выделения ДНК, РНК, постановки реакций ПЦР	
Центрифуга со сменными роторами: для пробирок 15-50 мл, и пробирок 0,2 мл 5000 об/мин	Для выделения ДНК, работы с клетками <i>E. coli</i> , ПЦР в микропланшетах	
Вортекс/шейкер лабораторный	Для перемешивания содержимого пробирок	

Гомогенизатор-шейкер	Для гомогенизации растительного материала	
Полипропиленовый пестик	Для гомогенизации растительного материала	
Амплификатор	Для Полимеразной Цепной Реакции (ПЦР)	
Водяная баня (либо твердотельный термостат)	Для инкубации пробирок при определенной температуре при выделении ДНК	
Камера и источник питания для гелелектрофореза	Для проведения гелелектрофореза	

Система фотодокументирования гель-электрофорезов	Для визуализации результатов электрофореза	
Автоматические пипетки (дозаторы)	Для приготовления реакции ПЦР, электрофореза: 0,2-10 мкл 20-200 мкл 100-1000 мкл	

Выделение ДНК растений:

Выделение ДНК возможно из различного типа растительного материала: свежих или замороженных в жидком азоте проростков или листьев растений (предпочтительный вариант), из гербарного материала, либо из других частей растений. Выделение ДНК проводится в четыре этапа: (1) гомогенизация растительного материала, (2) инкубация гомогената с буфером, (3) выделение и очистка ДНК; (4) преципитация ДНК.

- 1) Гомогенизация растительного материала. В отличие от клеток животных, клетки растений обладают клеточной стенкой, состоящей из целлюлозы и гемицеллюлозы. Для выделения внутренних компонентов клетки (ДНК, РНК, белки), клеточную стенку необходимо разрушить. Существуют разные способы гомогенизации растительного материала: (а) растирание растительных тканей пестиком в ступке в присутствии жидкого азота; (б) гомогенизация с помощью полипропиленового пестика в буфере для выделения; (в) гомогенизация в автоматической станции и использованием керамических шариков.

Традиционно лучшие результаты дает растирание в ступке в жидком азотом, но данный вариант является самым трудоемким и времязатратным, поэтому при выделении ДНК большого количества образцов предпочтительнее пользоваться шейкером для гомогенизации растительного материала.

- 2) Инкубация гомогената с буфером – на этой стадии происходит лизис клеточной и ядерной мембран, так же частичная очистка от белков, полисахаридов и вторичных

метаболитов. Существуют разные типы буферов для выделения ДНК. Самый широко используемый в лабораторных условиях буфер это СТАВ-буфер (цетилтриметиламмонийбромид).

Примерный состав СТАВ-буфера:

2% СТАВ (*детергент*)

100 мМ Tris HCl pH 8.0

1,4М NaCl

20мМ EDTA (*хелатирует двухвалентные катионы, подавляет активность нуклеаз*)

В случае, если происходит выделение ДНК из объектов, богатых полифенолами либо вторичными метаболитами, буфер может содержать меркаптоэтанол, либо 1% PVP (поливинилпирролидон)

Коммерческие наборы для выделения ДНК (например, DNeasy Plant Mini Kit, Qiagen) обычно содержат буфер с гуанидинтиоцианатом, либо гуанидин гидрохлоридом – хаотропной солью, которая стабилизирует ДНК или РНК, денатурируя нуклеазы.

3) Таким образом, после инкубации с буфером в пробирке оказывается раствор ДНК, РНК, белков, связанных с детергентом, и других метаболитов. Следующий шаг – очистка раствора и выделение ДНК:

(а) Либо добавляют к лизату равный объем смеси фенол:хромоформ (1:3), либо смесью хлороформ:изоамиловый спирт (24:1), затем аккуратно перемешивают и центрифугируют. На этом этапе происходит разделение фаз на органическую (фенол-хлороформ, где остаются большая часть метаболитов), и водную (где присутствует растворенный препарат ДНК). Водная фаза раствора имеет меньшую плотность, и находится в верхней части пробирки, откуда ее аккуратно забирают с помощью пипетки, и переносятся в новую пробирку.

(б) в коммерческих наборах для выделения ДНК используют очистку на спин-колонках с кремниевой мембраной (spin-columns). В растворах с высокой ионной силой, при pH ≤ 7 молекулы ДНК связываются с материалом фильтра и остаются на нем. Остальные компоненты раствора (метаболиты, вода, детергенты) при центрифугировании уходит вниз в виде водной фазы, которую затем выбрасывают. После этого на сухой фильтр добавляют Elution Buffer (с низкой ионной силой и pH ≥ 7), ДНК «отсоединяется» от фильтра и уходит в раствор.

- 4) Преципитация ДНК – осаждение препарата ДНК спиртом в присутствии соли, используется после очистки смесью фенол-хлороформ и хлороформ-изоамиловый спирт, после очистки на спин-колонках используется опционально. В присутствии соли-осадителя (NaCl, NaAc, Kac) и спирта (2,5 объема этанола, либо 0,6 объемов изопропанола) ДНК выпадет в нерастворимый осадок. После центрифугирования спирт удаляют, осадок ДНК высушивают, и растворяют в буфере для хранения (обычно содержит TrisHcl pH 7.5 и ЭДТА).

Подбор маркеров для генотипирования

Информацию о маркерах для изучаемых видов растений, можно взять из соответствующих баз данных. В качестве примера можно привести базы данных для пшеницы: <https://shigen.nig.ac.jp/wheat/komugi/>; <https://wheat.pw.usda.gov/GG3/>; <https://maswheat.ucdavis.edu/>; <http://wheatheadingtime.tilda.ws/>; <https://masrusplants.ru/>.

При выборе маркера необходимо ориентироваться на задачу, которая выбрана для генотипирования, а именно идентификация аллелей конкретных генов или для паспортизации вида. В последнем случае желательно привлекать маркеры, расположенные на разных хромосомах. Важно учитывать тип маркера- доминантный или кодоминантный. Если маркер доминантный, то для генотипирования по одному гену необходимо использовать два маркера, один к доминантной аллели, один к рецессивной.

В таблице 1 приведены варианты маркеров, различающиеся по типу выявления аллелей генов:

- 1) маркер *Intr1*- *Intr1*/B/R3, позволяющий выявлять одновременно две доминантные аллели одного гена *Vrn-B1*, влияющего на сроки колошения пшеницы;
- 2) маркер *L34Plus*, позволяющий выявлять комплекс доминантных генов *Lr34*(*Sr57*/*Yr18*/*Pm38*), определяющих устойчивость к различным листовостебельным заболеваниям;
- 3) маркер *csLV34* (кодоминантный), позволяющий выявлять гомо- и гетерозиготы по генам *Lr34*(*Sr57*/*Yr18*/*Pm38*), определяющих устойчивость к различным листовостебельным заболеваниям;

Таблица 1. Варианты маркеров по типу выявления различных аллелей генов

Ген/аллель	Маркер (тип)	Структура праймеров 5'-3'	*п.н.	Источник
<i>Vrn_B1a</i>	<i>Intr1- Intr1/B/R3</i> (доминантный)	<i>Intr1</i> ATCATCTTCTCCACCAAGGG	1124	<i>Shcherban et al., 2012</i> <i>Fu et al., 2005</i>
<i>Vrn_B1c</i>		<i>Intr1/B/R3</i> CTCATGCCAAAAATTGAAGATGA	737	
<i>Lr34(Sr57/ Yr18/Pm38)</i>	<i>L34Plus</i> (доминантный)	F-TTGATGAAACCAGTTTTTTTTTCTA R-GCCATTTAACATAATCATGATGGA	517	Lagudah et al., 2009
	<i>csLV34</i> (кододоминантный)	F-GTTGGTTAAGACTGGTGATGG R-TGCTTGCTATTGCTGAATAGT	150 (+) 229 (-)	

*диагностический фрагмент ПЦР

Амплификация и визуализация продуктов ПЦР.

Состав реакционной смеси для ПЦР, как правило, содержит следующие компоненты: прямой и обратный праймеры, олигонуклеотиды, Taq-полимеразу, буфер с ионами Mg²⁺ и воду mQ. Разработчики маркеров обычно указывают концентрации праймеров на объем смеси, необходимые для качественной ПЦР-амплификации маркера. Однако для упрощения лабораторной работы по тестированию генетического материала с использованием широкого набора маркеров удобнее использовать универсальную пропись (Таблица 2).

Условия проведения ПЦР для отобранных маркеров представлены в таблице 3.

Таблица 2. Состав ПЦР смеси для генотипирования растительных образцов

Компоненты ПЦР смеси исходные растворы	Рабочий объем (на 25 мкл смеси)
ДНК-матрица 50 нг/мкл	1 мкл
Primer R/F 10 пмол/мкл	0.5 мкл/ 0.5 мкл
dNTP 2мМ	2.5 мкл
10 x буфер для Taq-пол.	2.5 мкл
MgCl ₂ 25мМ	1.2 мкл
ДНК-полимераза 1 ед./мкл	1 мкл
Вода mQ	15 мкл

Таблица 3. Условия проведения ПЦР

Гены	Маркер	Модифицированные условия ПЦР	Рекомендованная система визуализации
<i>Vrn-B1a, Vrn-B1c</i>	Intr1- Intr1/B/R3	94 °С – 2 мин, далее 38 циклов (94 °С – 45 с; 58 °С – 45 с, 72 °С – 70 с); 72 °С – 5 мин	1 % агарозный гель
<i>Lr34(Sr57/Yr18/Pm38)</i>	L34Plus	94 °С – 30 с, далее 35 циклов (94 °С – 1 мин; 55 °С – 45 с, 72 °С – 30 с); 72 °С – 5 мин	1 % агарозный гель
	csLV34	94 °С – 30 с, далее 35 циклов (94 °С – 1 мин; 55 °С – 45 с, 72 °С – 30 с); 72 °С – 5 мин	2 % агарозный гель

Документирование и анализ результатов.

Электрофорез продуктов ПЦР проводится в агарозном геле, концентрация которого подбирается в соответствии с длиной выявляемого фрагмента. В процессе приготовления агарозного геля на 100 мл расплавленной агарозы добавляется 5 мкл 1 % раствора этидия бромида для последующей визуализации фрагментов ДНК в УФ- свете. Для нанесения на гель образцы ДНК смешиваются с буфером (например, 4-кратный буфер для хранения и нанесения образцов ДНК «БиК», БиоЛабМикс).

В качестве примера на рисунке 1 приведены результаты электрофореза 8 образцов ДНК растений после проведения ПЦР согласно таблице 3.

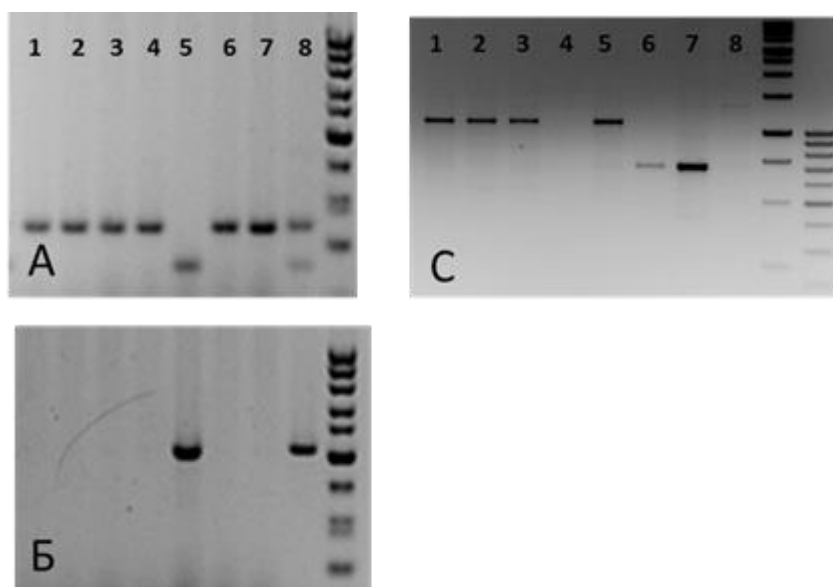


Рис. 1 Электрофореграмма продуктов ПЦР, полученных с использованием маркеров csLV34 (А), L34Plus (Б) и Intr1- Intr1/B/R3 (С). Цифрами сверху обозначены образцы ДНК растений, взятых в анализ. Слева расположен маркер длины.

Интерпретации результатов электрофореза представлена в таблице 4.

Таблица 4. Присутствие генов *Vrn-B1a*, *Vrn-B1c*, *Lr34(Sr57/Yr18/Pm38)* в различных растениях (№1-№8).

ген	маркер	№1	№2	№3	№4	№5	№6	№7	№8
<i>Lr34(Sr57/Yr18/Pm38)</i>	csLV34	-	-	-	-	+	-	-	+/-
	L34Plus	-	-	-	-	+	-	-	+
<i>Vrn-B1a</i>	Intr1- Intr1/B/R3	+	+	+	-	+	-	-	-
<i>Vrn-B1c</i>		-	-	-	-	-	+	+	-

+ наличие доминантного гена, - отсутствие доминантного гена

Согласно полученным результатам с маркерами csLV34 и L34Plus доминантные гены устойчивости к листовостебельным заболеваниям обнаружены у растения №5, растение №8 является гетерозиготой по изучаемым генам.

Согласно полученным результатам с маркером Intr1- Intr1/B/R3 растения №1-№3, №5 несут доминантную аллель гена *Vrn-B1a*, растения №6 и №7 - доминантную аллель гена *Vrn-B1c*, растения №4 и №8 – несут рецессивные аллели этого гена.