

РОЛЬ ДЛИННЫХ НЕКОДИРУЮЩИХ РНК В РАСТЕНИЯХ

© 2025 г. А. Ю. Пронозин^{1, 2, *}, Д. А. Афонников^{1, 2, 3}

¹Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

²Курчатовский геномный центр, Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, 630090 Россия

³Новосибирский национальный исследовательский государственный университет, Новосибирск, 630090 Россия

*e-mail: pronozinartem95@gmail.com

Поступила в редакцию 19.04.2024 г.

После доработки 29.07.2024 г.

Принята к публикации 15.08.2024 г.

Длинные некодирующие РНК (днРНК) представляют собой класс линейных или кольцевых молекул РНК длиной более 200 нуклеотидов без открытых рамок считывания. Экспериментальные исследования показали участие днРНК в регуляции устойчивости к холодовому, солевому, тепловому стрессу, в развитии плодов, корня и листьев. Однако экспериментальные методы являются трудоемкими и дорогостоящими подходами и пока еще не могут быть использованы для массового изучения днРНК в масштабах всего генома. Для этих целей применяют биоинформационные подходы, которые направлены на широкомасштабное распознавание последовательностей днРНК в геномах и транскриптомах. Несмотря на растущее число исследований, посвященных структурно-функциональному анализу днРНК, данный тип молекул по-прежнему остается малоизученным. Это связано со множеством факторов, которые нужно учитывать при идентификации днРНК. Применение пан-геномов и пан-транскриптомов позволит повысить эффективность исследования и общее количество предсказанных днРНК по сравнению с использованием генома одного представителя вида. Настоящий обзор посвящен описанию молекулярных и биологических функций днРНК, экспериментальных и биоинформационных методов идентификации, а также закономерностям эволюции, выявлению и анализу днРНК в масштабах пан-геномов и пан-транскриптомов.

Ключевые слова: длинные некодирующие РНК, микроРНК, регуляция транскрипции, пан-геном.

DOI: 10.31857/S0016675825010016 **EDN:** VFMYSM

Некодирующие РНК (нкРНК) выполняют в геномах растений ряд важнейших функций, связанных с регуляцией экспрессии генов, гомеостазом физиологических параметров растений. Знание функциональных особенностей нкРНК позволяет целенаправленно улучшать различные свойства растений [1–3], что обусловливает интерес к изучению этих молекул. Некодирующие РНК можно разделить на две группы: структурные некодирующие (рибосомные РНК, транспортные РНК, малые ядерные РНК, малые ядрышковые РНК) и регуляторные некодирующие (малые некодирующие РНК и длинные некодирующие РНК) [4]. Наименее изученными из всех перечисленных классов молекул являются длинные нкРНК [5, 6].

Длинные некодирующие РНК (днРНК) являются важным классом некодирующих РНК в живых организмах и представляют собой линейные или кольцевые молекулы РНК, содержащие более

200 нуклеотидов без открытых рамок считывания [4, 6]. Функции и структура днРНК слабо изучены в сравнении с белок-кодирующими генами и генами малых некодирующих РНК. Однако участие днРНК выявлено в регуляции экспрессии генов [7], формировании структуры макромолекулярных комплексов [8], во взаимодействии с белками [9], в патогенезе [10]. Наиболее активно днРНК изучаются экспериментально и биоинформационически у человека и других модельных видов (мыши, крысы, дрозофилы и т. д.). Эти исследования позволяют идентифицировать последовательности днРНК в геномах и транскриптомах, расшифровать важные особенности их структуры, регуляции экспрессии и функционирования, классифицировать их по структуре и функциям. На сегодняшний день эти методы позволили идентифицировать большое количество последовательностей днРНК у животных. Так, согласно базе данных Gencode

[11], в геноме мыши описано около 13000 генов днРНК. В геноме человека насчитывается более 18000 генов днРНК [12]. Изучение днРНК у животных позволило классифицировать их по молекулярному механизму функционирования (промоторные, энхансерные, теломерные, сигнальные и др.) [13–15], положению в геноме (межгенные, инtronные, антисмыловые и др.) [7, 16–18]. Эти исследования приводят к созданию специализированных баз данных, посвященных днРНК и содержащих последовательности десятков тысяч днРНК и их структурно-функциональную аннотацию [19, 20]. Вследствие отсутствия у первичной структуры днРНК четко выраженных функциональных сегментов (как, например, домены у белковых последовательностей) на сегодняшний день предсказание функции днРНК на основе первичной структуры является сложной задачей [15, 21, 22].

Следует отметить, что получение экспериментальной информации о структуре и функции днРНК является трудоемким процессом и требует проведения дорогостоящих экспериментов. Поэтому для изучения этого класса РНК активно используются методы биоинформатики. В частности, они направлены на широкомасштабное распознавание последовательностей днРНК в геномах и транскриптомах [23–25], предсказание их взаимодействий с белками [26], с мРНК и другими некодирующими РНК [27], на регуляцию экспрессии [7], эволюцию [28, 29] и предсказание их структуры и функций [30].

днРНК также известны и у растений. Интересно отметить, что, несмотря на ряд специфических особенностей, функциональные и структурные характеристики днРНК животных и растений имеют много общего [31]. Известно, что у растений днРНК принимают участие в регуляции устойчивости к холодовому, солевому, тепловому стрессу [1–3], оказывают влияние на устойчивость к гипоксии [32], принимают участие в развитии плодов, корня и листьев [17, 33]. Тем не менее до сих пор наши знания о днРНК растений являются менее полными, чем о днРНК животных. Исследования [34, 35] показывают, что последовательности днРНК в процессе эволюции быстро накапливают замены, так что даже с последовательностями из близкородственных видов организмов они имеют слабую гомологию. Применение концепции пан-генома, которая подразумевает охват последовательностей, подверженных структурной вариации и, возможно, отсутствующих в референсной последовательности каждого представителя вида, позволит повысить эффективность исследования и общее количество предсказанных днРНК [36–40].

Настоящий обзор посвящен днРНК растений и включает описание их молекулярных и биологических функций, экспериментальных и биоинформационных методов идентификации,

закономерностям эволюции, выявлению и анализу днРНК в масштабах пан-геномов и пан-транскриптомов.

КЛАССИФИКАЦИЯ днРНК ПО ПОЛОЖЕНИЮ В ГЕНОМЕ

днРНК классифицируют по многим параметрам, однако наиболее часто используют классификацию по положению в геноме относительно известных белок-кодирующих генов. Согласно этой классификации, днРНК делятся на четыре крупных класса [35], рис. 1:

- 1) инtronные — перекрываются с инtronом гена;
- 2) антисмыловые — ориентированы против направления транскрипции гена, кодирующего белок;
- 3) межгенные — расположены между двумя локусами генов;
- 4) смысловые — ориентированы в направлении транскрипции гена, кодирующего белок.

Инициация транскрипции инtronных днРНК происходит в направлении транскрипции гена мишени, терминация происходит в области перекрытия с экзоном гена мишени [41].

Транскрипция антисмыловых днРНК начинается на 3'-конце гена, кодирующего белок. Эти днРНК разделяются на две группы в зависимости от цис- и транс-геометрической изомерии. Транскрипция цис-изомеров антисмыловой днРНК (цис-днРНК) происходит с противоположного конца локуса белок-кодирующего гена, с которым данный транскрипт перекрываеться. За счет этого днРНК имеет высокий уровень комплементарности с данным геном. Транскрипция транс-изомеров антисмыловой днРНК происходит с локуса соседнего белок-кодирующего гена, с которым перекрываеться данный транскрипт. Таким образом, подобные транскрипты могут перекрываться с несколькими генами, но при этом имеют низкий уровень комплементарности с ними [42].

Длина межгенных днРНК превышает 200 нуклеотидов, данные транскрипты не перекрываются с кодирующими белки генами. Особенности транскриптов данного класса: высокая консервативность, выраженная тканеспецифичность, высокая стабильность. В работе [43] биоинформационными методами было показано взаимодействие межгенных днРНК с мобильными элементами. Проведенное исследование предоставляет систематическую оценку вклада мобильных элементов в состав, эволюционное происхождение и регуляцию днРНК.

Смысловые днРНК транскрибируются в направлении транскрипции кодирующих белок генов. Это делает их выделение в отдельный класс днРНК или вообще идентификацию как днРНК

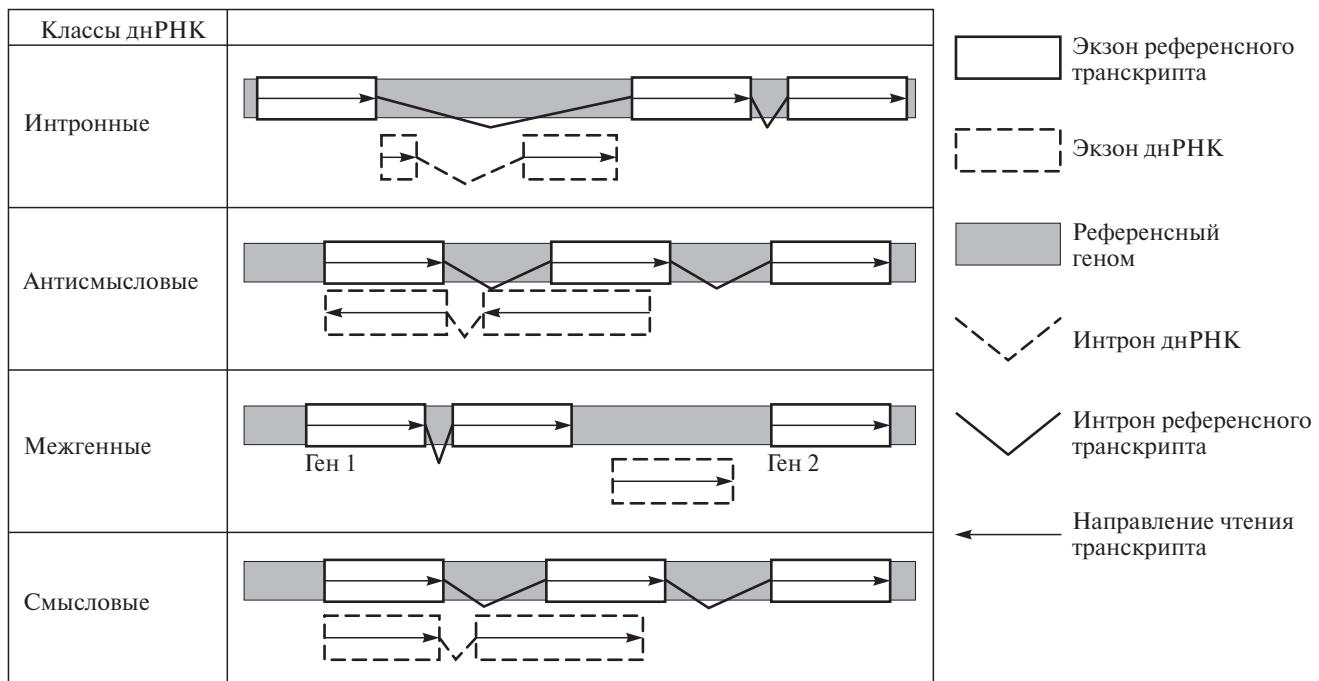


Рис. 1. Классификация днРНК на основе расположения и ориентации относительно ближайших или перекрывающихся генов, кодирующих белки.

достаточно противоречивым. Структурные характеристики таких РНК (длина, количество экзонов и инtronов) могут совпадать со структурой гена, кодирующего белок. Например, в работе [17] предполагается, что транскрипты данного класса днРНК могут полностью перекрываться с генами, кодирующими белок. Однако в работе [41] описываются примеры днРНК, возникших за счет альтернативного сплайсинга или укорачивания первого либо последнего экзона генов. Авторы предполагают, что эти гены могут представлять изоформы белков, кодируемых этими генами, а не днРНК. Например, в работе [34, 41] РНК, перекрывающиеся с экзоном гена, кодирующего белок, не считаются днРНК. Для растений найдено мало данных транскриптов, и ни для одного из них нет функциональной аннотации [41].

МОЛЕКУЛЯРНЫЕ ФУНКЦИИ днРНК

днРНК имеют несколько общих характеристик с генами, кодирующими белки, что затрудняет их идентификацию. Так, большинство днРНК транскрибируется с помощью РНК-полимеразой II. Они имеют сходные эпигенетические профили, наличие сигналов сплайсинга и полигидрирования, а также сходные размеры экзонов и инtronов [44, 45]. Кроме того, некоторые днРНК имеют длину, превышающую 1000 нуклеотидов. Таким образом, они могут содержать открытые рамки считывания (ОРС). Размеры подобных ОРС могут варьироваться, некоторые из них могут

быть достаточно короткими. Например, в работе [46] упоминается, что днРНК в среднем могут содержать несколько транслируемых ОРС, имеющих длину 50 кодонов.

Отличительной особенностью днРНК является способность транскрибироваться РНК-полимеразой IV и V, в особенности у растений [47]. Большинство днРНК локализуются как в клеточном ядре, так и в цитоплазме [48]. Также днРНК демонстрируют низкий уровень экспрессии, при этом экспрессируются более специфично для определенных тканей [49]. Подобное наблюдается и при анализе консервативности структуры днРНК. Сравнение днРНК различных видов растений показало, что доля сходных последовательностей у разных видов мала. Однако днРНК одного вида демонстрируют достаточно высокий уровень консервативности первичной, вторичной и третичной структуры [15].

В работах [14, 15, 50] указывается, что на функционирование днРНК влияет и их внутриклеточная локализация. В работах [4, 15, 50, 51] выделяют пять основных функций днРНК, представленных на рис. 2.

Сигнальные днРНК

днРНК чаще всего экспрессируются в определенных тканях. В некоторых работах отмечается, что днРНК экспрессируется специфически в ответ на различные стимулы [4, 52–54]. В связи с этим

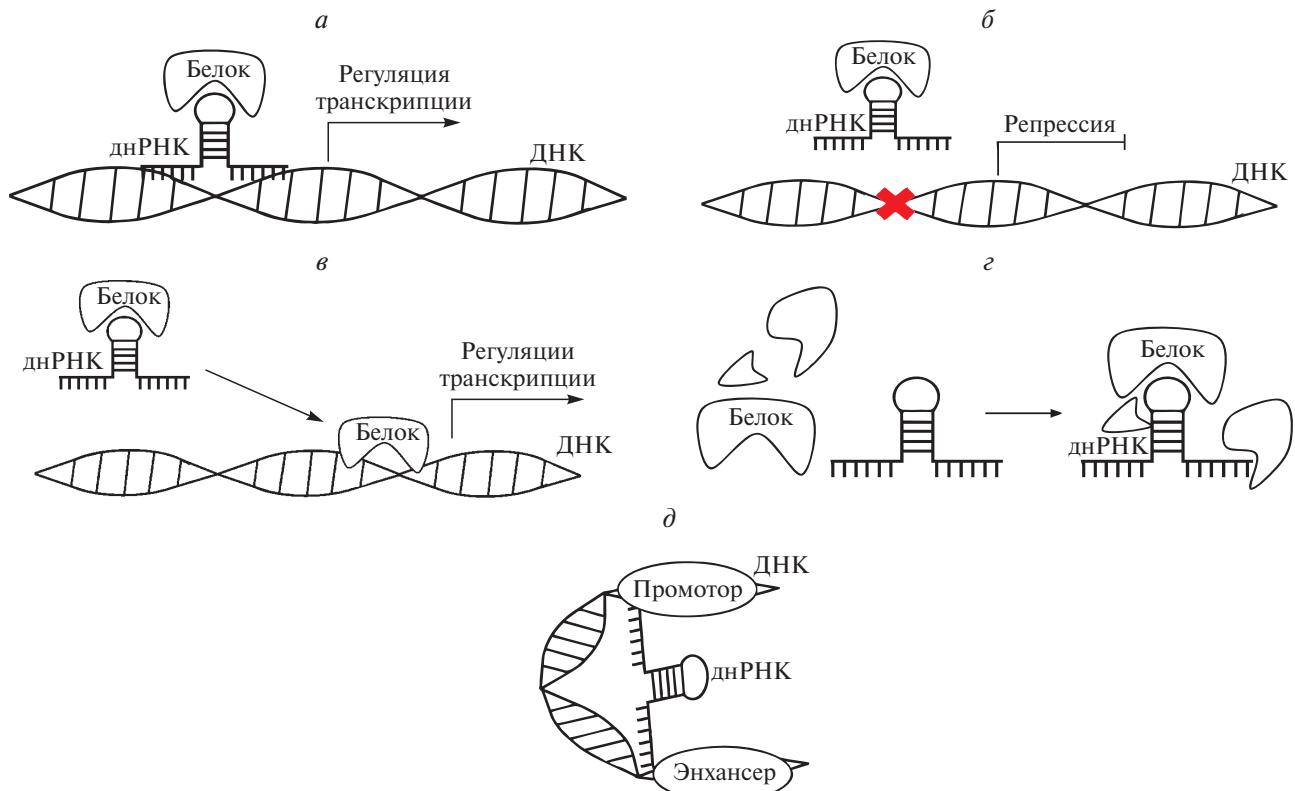


Рис. 2. Молекулярные функции днРНК: *а* – сигнальные, днРНК получают сигнал для взаимодействия с ферментами, модифицирующими хроматин, для регуляции транскрипции; *б* – ловушки, связывают белки и препятствуют их присоединению к генам-мишениям; *в* – направляющие, направляют транскрипционные факторы к местам их функционирования, содействуют необходимому состоянию хроматина; *г* – каркасы, объединяют белки в комплексы и инициируют разные биологические процессы; *д* – энхансеры, могут транскрипционно активировать соседние гены.

можно предположить, что экспрессия некоторых днРНК находится под контролем транскрипционных факторов [55], осуществляется в определенный момент и в определенной ткани и может служить сигналом для запуска регуляторных процессов в клетке [55]. Такая специфическая экспрессия характерна для днРНК, которые называют сигнальными (рис. 2, *а*). Сигнальные днРНК могут выступать в качестве элементов регуляции. Известно, например, что сигнальные днРНК взаимодействуют с модифицирующими хроматин ферментами, такими как гистоновые метилтрансферазы, чтобы подавить экспрессию их генов-мишений посредством образования гетерохроматина [56]. Такие взаимодействия являются частью регуляторных и сигнальных функций данного типа днРНК.

днРНК-приманки или ловушки

днРНК-приманки связывают белки и тем самым препятствуют их взаимодействию с регуляторными районами генов-мишений [55]. Эти днРНК косвенным образом репрессируют транскрипцию, блокируя микроРНК, модифицирующие комплексы

(субъединицы хроматина), каталитические белки и факторы транскрипции [55, 57]. Также днРНК выступают в качестве ловушки для микроРНК, данный механизм взаимодействия подробно описан в подразделе “Механизмы взаимодействия днРНК и микроРНК”.

Направляющие днРНК

днРНК также выполняют функцию группировки/локализации факторов транскрипции в определенных геномных локусах. Такие днРНК называют направляющими, (рис. 2, *в*). днРНК направляют модификаторы хроматина и белковые комплексы или факторы транскрипции к определенным генам-мишениям. Они способствуют локализации регуляторных молекул в определенных локусах для осуществления транскрипционной регуляции [56]. днРНК имеют цис-, транс-геометрическую изомерию. Цис-изомеры направляют изменения хроматина, выступая в качестве комплементарных мишений для малых РНК (таких как микроРНК, малые интерферирующие РНК) [55]. Транс-изомеры

направляют ДНК-мишень посредством РНК–ДНК взаимодействий [58].

днРНК-каркасы

днРНК-каркасы (скаффолды) играют важную роль при сборке макромолекулярных белковых комплексов и инициации различных биологических процессов [59], (рис. 2, *г*). Подобные днРНК способны соединять несколько белков в один рибонуклеиновый комплекс, который позволяет давлять или активировать транскрипцию в зависимости от наличия и природы вовлеченных днРНК и белков [55].

днРНК-энхансеры

днРНК, транскрипция которых осуществляется с энхансерных локусов, называются днРНК-энхансеры (рис. 2, *д*). Как и все днРНК, их последовательности считываются в одном направлении, имеют поли(A)-хвост и сплайсируются [60]. Их следует отличать от энхансерных РНК (эРНК), которые являются продуктом двунаправленной транскрипции, не сплайсированы и не имеют поли(A)-хвоста. Энхансерные днРНК на сегодняшний день довольно плохо изучены. Однако, согласно гипотезе [61], синтез энхансерных днРНК может служить способом поддержания открытого состояния хроматина в области энхансеров для облегчения их взаимодействия с промоторами посредством образования петель. Некоторые исследования показывают вовлеченность энхансерных днРНК в процесс активации экспрессии генов мишени [62, 63]. Также в нескольких работах показано понижение экспрессии генов мишени при нокдауне ближайшей к ним энхансерной днРНК [63, 64].

Механизмы взаимодействия днРНК и микроРНК

Особое внимание стоит уделить взаимодействию днРНК и микроРНК. В работе [65] указывается на участие данных взаимодействий в

регуляции устойчивости к болезням, холодовому, тепловому, биотическому и абиотическому стрессу у растений.

МикроРНК — это малые некодирующие РНК длиной 18–25 нуклеотидов, транскрибируемые из инtronов, некодирующих участков гена или межгенных спейсеров (участков ДНК между генами) [66]. На сегодняшний день выделяют четыре типа взаимодействия днРНК и микроРНК [67], (рис. 3). Более подробно разберем данные механизмы взаимодействия.

Деградация днРНК при встраивании микроРНК

МикроРНК функционируют в комплексе с белками семейства аргонавтов (AGO). Комплекс микроРНК–AGO встраивается в РНК-индуцируемый комплекс подавления экспрессии гена, либо блокируя синтез белка с мРНК, либо приводя к деградации мРНК [68]. Подобный механизм функционирования наблюдается и для днРНК. МикроРНК за счет РНК-связывающих белков встраивается в комплементарный участок днРНК, вызывая деградацию данных молекул [69, 70]. В этом случае днРНК выступает как конкурент мРНК и тем самым адсорбирует часть молекул микроРНК, снижая их воздействие на молекулы потенциальных мишени (мРНК). Важно, что при этом происходит деградация и самой днРНК (рис. 3, *а*).

днРНК-мишени для микроРНК

Впервые днРНК-мишени с эффектом мимикии, или конкурирующие эндогенные днРНК (competitive endogenous RNAs, ceRNAs), были обнаружены в геномах растений [71], (рис. 3, *б*). Механизм мимикии (“эффект губки”) заключается в блокировании взаимодействия микроРНК с мРНК-мишенью за счет связывания данной микроРНК с днРНК-приманкой через частично комплементарные последовательности [72].

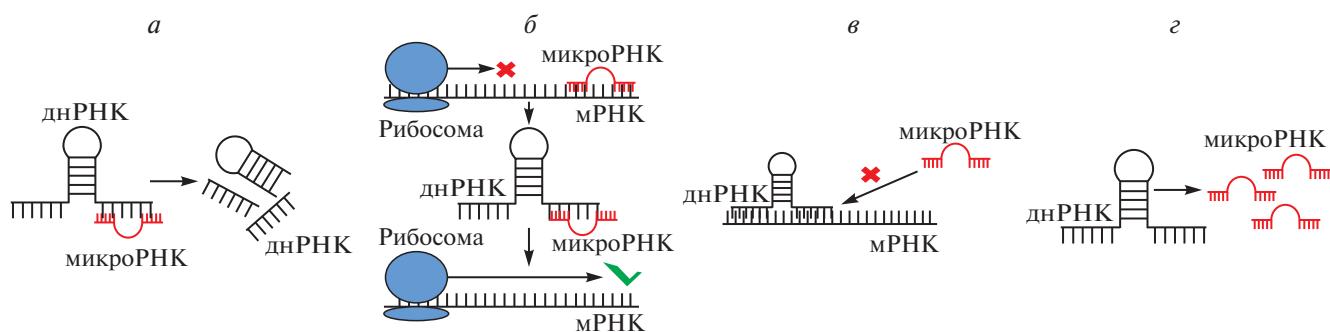


Рис. 3. Механизмы взаимодействия днРНК – микроРНК: *а* – деградация днРНК при встраивании микроРНК; *б* – днРНК выступает в качестве мишени (губки) для микроРНК; *в* – днРНК конкурирует с микроРНК для регуляции уровня экспрессии мРНК; *г* – днРНК выступает предшественником микроРНК (сплайсируется в микроРНК).

На сегодняшний день экспериментально подтверждено влияние эндогенных днРНК на регуляцию синтеза фосфатов в *Arabidopsis thaliana* [72], *Zea mays* [73] и *Medicago truncatula* [74] за счет связывания с микроРНК.

Конкуренция днРНК и микроРНК

Механизмы взаимодействия микроРНК с эндогенными и конкурирующими днРНК схожи. днРНК выполняет роль мишени и снижает концентрацию микроРНК, (рис. 3, в). Однако днРНК может конкурировать и с микроРНК. Этот механизм функционирования днРНК возникает в случае, когда за счет комплементарных оснований они связываются с мРНК, не давая возможности в этих районах мРНК взаимодействовать с микроРНК. Подобный механизм позволяет предотвратить деградацию мРНК, с которой связываются днРНК [75–77].

днРНК-предшественники микроРНК

Некоторые днРНК как первичные транскрипты могут служить предшественниками специфических микроРНК, (рис. 3, г). Малые РНК могут производиться за счет механизма сплайсинга и приводить к усилению посттранскрипционной регуляции соответствующих мРНК-мишеней [35]. Такой механизм функционирования днРНК был экспериментально подтвержден у человека [78]. Показано, что они отвечают за регуляцию развития и прогрессирования опухоли за счет контроля транскрипции микроРНК-супрессоров [79, 80]. Для растений днРНК-предшественники обнаружены биоинформационическими методами. В работах [81, 82] большинство выявленных микроРНК выравниваются на концевые участки последовательности днРНК, указывая на то, что эти днРНК являются их предшественниками.

ЭКСПРЕССИЯ днРНК

На сегодняшний день появляется все больше работ по анализу днРНК у различных организмов, в которых применяются методы исследования экспрессии для данных молекул. Общепринятым считается, что днРНК демонстрируют более низкий уровень экспрессии по сравнению с мРНК, при этом экспрессируются более специфично для определенных тканей [34]. Однако в последнее время начинают появляться работы, демонстрирующие днРНК с уровнями экспрессии, равными таковым для генов, кодирующих белки [16, 83]. Подобные случаи являются единичными, и их результаты требуют экспериментального подтверждения.

Выделяют три направления исследования экспрессии днРНК: анализ экспрессии днРНК в тканях и в ответ на определенный раздражитель;

анализ дифференциальной экспрессии и ко-экспрессии с генами, кодирующими белки; анализ тканеспецифичности днРНК. Данные подходы применяются как для масштабного транскриптомного анализа, так и для анализа небольших выборок последовательностей.

Анализ экспрессии днРНК в тканях является наиболее распространенным направлением исследования. Большинство работ начинают свое исследование с данного направления, постепенно углубляясь в дифференциальную или ко-экспрессию.

Анализ тканеспецифичности днРНК является наименее исследованным, так как нет точного понимания, какой метод определения специфичности экспрессии гена в тканях наиболее точный [84]. Во многих работах специфичным считается ген, имеющий экспрессию только в одной ткани [16, 85, 86]. Одним из первых подходов является подсчет количества тканей, в которых данный ген экспрессируется [87]. Однако данный метод позволяет исследовать только гены с высокой экспрессией, что затрудняет выбор порога фильтрации последовательностей. Также существует проблема близкородственных тканей, таких как части стебля или мозга человека.

С развитием понимания о структуре и функциях днРНК стало появляться все больше подходов к решению данной задачи. Существующие методы делятся на две группы. В первой группе показателем специфичности является число, определяющее, насколько специфично экспрессируется исследуемый ген: 1 – если ген имеет специфическую экспрессию, 0 – если экспрессируется в нескольких тканях – Tau [88], Gini [89], tissue specificity index (TSI) [90]. Во второй группе специфичность экспрессии гена рассчитывается отдельно для каждой ткани – z-score, SPM [91], expression enrichment (EE) [92], при этом уровень специфичности варьируется между разными метриками.

В основном все метрики расчета специфичности экспрессии гена (Tau, EE score, TSI, Gini coefficient) похожи: для расчета используется уровень экспрессии гена в определенной ткани и количество экспериментов, которым данная ткань принадлежит. Например, формула для расчета Tau:

$$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n-1}; \quad (1)$$

$$\hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} (x_i)}, \quad (2)$$

где x_i – экспрессия гена в ткани; n – количество экспериментов с тканью.

Встречаются работы, использующие, помимо перечисленных выше, метрики, основанные на энтропии Шеннона, Jensen-Shannon (JS) [93]:

$$Hg = \sum_{1 \leq t \leq N} -p_{t|g} \log_2 (p_{t|g}); \quad (3)$$

$$p_{t|g} = \frac{\omega_{g,t}}{\sum_{1 \leq t \leq N} \omega_{g,t}}, \quad (4)$$

где $p_{t|g}$ – распределение экспрессии гена в ткани и $\omega_{g,t}$ – экспрессия гена. Таким образом, $p_{t|g}$ показывает отношение экспрессии гена в определенной ткани на сумму экспрессий этого гена во всех тканях.

$$Q_{g|t} = Hg - \log_2 (p_{t|g}), \quad (5)$$

где Hg – энтропия, показывает, насколько ген тканеспецифичный. Если значение энтропии близко или равно 0, то данный ген специфичен для определенной ткани. $Q_{g|t}$ показывает энтропию экспрессии исследуемого гена в одной ткани. Чем выше значение энтропии, тем выше специфичность экспрессии гена в определенной ткани.

БИОЛОГИЧЕСКАЯ РОЛЬ днРНК У РАСТЕНИЙ

Предыдущие разделы показали разнообразие и важность молекулярных функций днРНК у

растений. днРНК могут выступать как генетические и эпигенетические регуляторы экспрессии генов. В зависимости от геометрической изомерии они могут функционировать как цис-регуляторные элементы, взаимодействуя с близко расположеными генами и воздействуя непосредственно на гены в одной цепи, или транс-регуляторные элементы, действуя вдали от места синтеза. днРНК могут препятствовать связыванию транскрипционных факторов с промоторной областью, выступая в качестве репрессоров транскрипции. В результате этих молекулярных процессов контролируются функции растений на более высоком уровне организации: тканей, органов и всего организма. Примеры процессов, которые затрагивает регуляция при помощи днРНК у растений, приведены на рис. 4. К ним относятся: устойчивость растений к различным видам стресса, развитие органов растений, яровизация, развитие плода.

днРНК в процессе яровизации

Цветение является одним из важнейших адаптивных признаков, обеспечивающих переход репродуктивного роста и развития, происходящий в благоприятных условиях в течение жизненного цикла растения. Яровизация является важным механизмом, контролирующим цветение у некоторых видов растений, которые растут в вегетативном состоянии в холодные зимние сезоны и начинают цвети весной [94]. Также яровизация

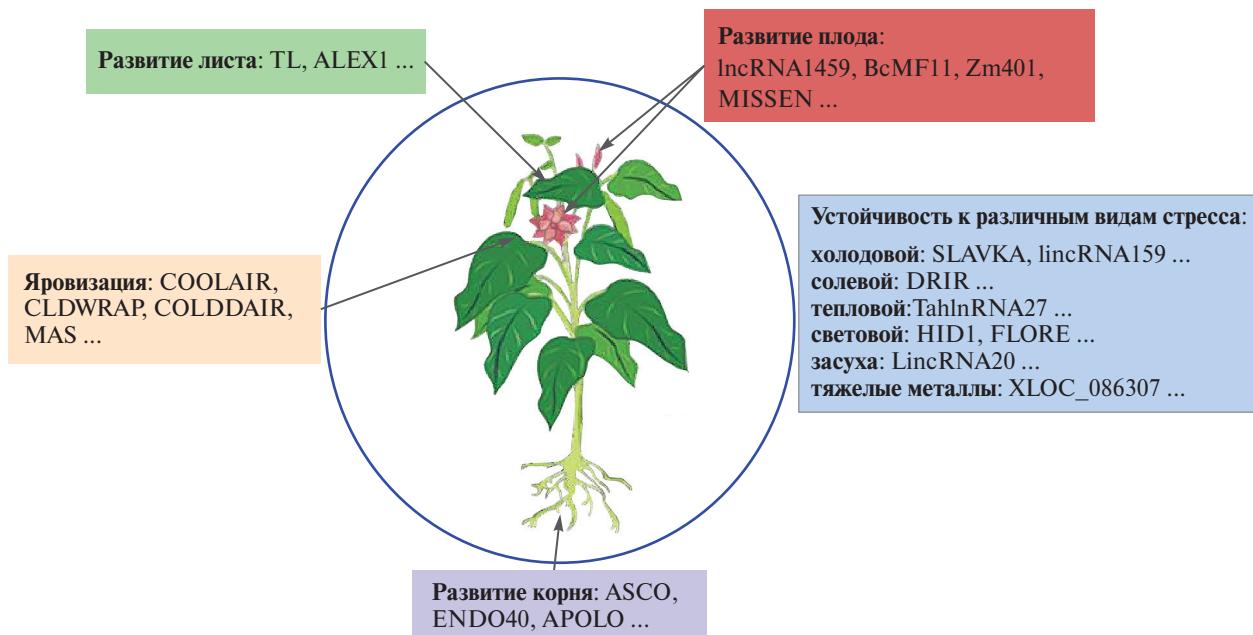


Рис. 4. Биологическая роль днРНК у растений. днРНК принимают участие в регуляции генов, отвечающих за устойчивость растений к различным видам стресса (синий прямоугольник), также влияют на плодородие растений (красный прямоугольник), участвуют в развитии листа, корня (зеленый, фиолетовый прямоугольник) и в процессах яровизации (оранжевый прямоугольник).

является наиболее изученным регуляторным процессом у растений, в котором участвуют днРНК. Они также принимают участие в регуляции гена FLOWERING LOCUS C (*FLC*), отвечающего за устойчивость *A. thaliana* к холодовому стрессу [3]. На сегодняшний день в локусе *FLC* обнаружено два класса днРНК: COOLAIR – днРНК транскрибируется в антисмысловом направлении по отношению к промотору гена *FLC*; COLDAIR – днРНК транскрибируется с первого интрана гена *FLC* в смысловом направлении. днРНК COOLAIR связываются с локусом *FLC* в цис-положении и образуют R-петлю между промотором и первым интраном гена *FLC*, что приводит к подавлению экспрессии данного гена мишени. Данный процесс проходит как на автономном, так и в яровизированном пути [94].

днРНК и влияние на урожайность растений

Развитие нормального пыльцевого зерна имеет решающее значение для оплодотворения и развития семян. Пыльник действует как родительская ткань пыльцевого зерна и регулирует его развитие. Аномальное пыльцевое зерно может привести к развитию растений с мужской стерильностью [95]. Экспериментальные исследования [96, 97] показали вовлеченность днРНК в процесс контроля урожайности. Например, днРНК (LDMAR) риса (*Oryza sativa japonica*) вызывает запрограммированную гибель клеток пыльников в условиях длинного дня (LD, long day) за счет сплайсинга в микроРНК osa-smR5864. Однонуклеотидная мутация микроРНК osa-smR5864 приводит к мужской стерильности растения [96]. Снижение экспрессии днРНК (BcMF11) в *Brassica campestris* препятствует деградации тапетума и приводит к прерыванию образования пыльцевых зерен [97]. Также методами биоинформатики в работе [33] для пяти стадий развития пыльника *Brassica rapa* было выявлено 14 днРНК, ко-экспрессирующихся с 10 известными генами, кодирующими белки, которые играют решающую роль в развитии пыльцы. Как утверждают авторы, выявленные днРНК представляют интерес для дальнейших экспериментальных исследований вовлеченности днРНК в репродуктивные процессы растений.

днРНК в процессе фотоморфогенеза

Молекулярный сигнальный механизм фотоморфогенеза широко изучался у различных видов растений. Во время этого явления ряд белков ассоциируются в качестве первичных и вторичных сигнальных молекул. Среди них выделяют семейство транскрипционных факторов со структурой ДНК-связывающего домена спираль–петля–спираль [helix–loop–helix (bHLH)]. Среди членов семейства выделяют факторы взаимодействия с

фитохромом (PIF), которые выполняют функцию подавления фотоморфогенеза проростков в темное время суток [98]. В работе [77] обнаружена днРНК (HID1), которая функционирует как транс-регуляторный элемент, взаимодействуя с промотором гена *PIF3* (ключевая сигнальная молекула фотоморфогенеза, которая негативно регулирует реакции на красный свет), и ингибитирует его транскрипцию. В результате наблюдается удлинение зародышевого стебля (гипокотиля) у саженцев *A. thaliana*.

днРНК в процессах устойчивости к различным видам стресса

В природе растения подвергаются действию различных стрессовых факторов, таких как соль, засуха, холод, тепло, а также различные вирусные инфекции. Это приводит к проблемам в росте и плодородии у растений. Чтобы адаптироваться и выжить в таких неблагоприятных условиях, растения используют множество механизмов регуляции генов для восстановления клеточного гомеостаза. Современные исследования демонстрируют немаловажную роль днРНК в регуляции экспрессии генов в ответ на стрессовые состояния [99].

Вовлеченность днРНК в процессы устойчивости к абиотическим видам стресса демонстрируется у *A. thaliana* [81], *B. rapa* [100], винограда (Cabernet Sauvignon) [101], *O. sativa* [102] и многих других растений. Например, у днРНК *A. thaliana* выделяют следующие механизмы регуляции устойчивости к холодовому стрессу: модуляция / реконструкция хроматина, например днРНК COOLAIR, описанная выше; альтернативный сплайсинг днРНК TAS1 приводит к образованию варианта транскрипта с нерасщепленным интраном, более устойчивого к холодовому стрессу, чем сплайсированный вариант данной днРНК [103]; транскрипционная регуляция генов, например днРНК SVALKA подавляет транскрипцию гена *CBF1* (ответственного за запуск каскада реакций, направленных на увеличение устойчивости растений к низким температурам) за счет транскрипционной коллизии [104].

днРНК IncRNA973 у *A. thaliana* регулирует устойчивость растений к солевому стрессу путем модуляции экспрессии генов, ответственных за реактивные формы кислорода (SOD, CAT и POD), транскрипционные факторы (MYB5, WRKY46, NAC29 и ERF62) и ионный баланс (NHX7) [105].

Сверхэкспрессия днРНК At5NC056820 у *A. thaliana* повышает устойчивость растения к засухе [106]. днРНК TCONS_00021861 у *O. sativa* за счет эффекта губки взаимодействует с miR528-3p и предотвращает ее взаимодействие с геном-мишенью YUCCA-7 (который участвует в биосинтезе триптофан-зависимых ауксинов и повышает устойчивость к засухе), что активирует синтез

индолил-3-уксусной кислоты (ИУК), повышающей устойчивость к засухе [107].

Подобный механизм наблюдается для днРНК TCONS_00048391, которая является “губкой” для микроРНК *bra-miR164a*, подавляющей экспрессию гена-мишени NAC1 (ответственного за устойчивость к тепловому стрессу у *B. rapa*) [100].

Таким образом, даже внутри одного организма днРНК имеют различные механизмы регуляции устойчивости к абиотическим видам стресса.

ЭВОЛЮЦИЯ днРНК

Эволюционные процессы приводят к изменчивости структуры и функций днРНК. В области изучения эволюции днРНК существует несколько направлений активных исследований. Прежде всего это вопросы, связанные с возникновением новых днРНК в геномах организмов. Одним из таких механизмов является дупликация генов. Появление в геноме дополнительной копии гена открывает свободу для эволюции одной из копий гена. Мутации, возникающие в одной из двух копий и ослабляющие исходную функцию гена, не будут отсеиваться отбором, потому что вторая копия может сохранить прежнюю функциональность. Таким образом, за счет дупликации возникают “псевдогены” (неактивные участки), изоформы и транскрипты днРНК [28].

Существует несколько гипотез возникновения днРНК за счет дупликации. Первая основывается на потере одной из изоформ гена, кодирующего белок (в процессе дупликации), способности кодировать белок. Это объясняется наличием у днРНК регуляторных элементов, сайтов сплайсинга, последовательностей экзонов и последовательностей, имеющих поли(А)-хвост. Наиболее известный пример днРНК, возникшей за счет дупликации гена, кодирующего белок, – человеческая днРНК *Xist* [29]. Вторая гипотеза основана на дупликации некодирующих РНК. Однако подобные случаи встречаются не так часто, как дупликация генов, кодирующих белки, вследствие высокой эволюционной лабильности днРНК [108].

Также одна из гипотез возникновения днРНК связана со встраиванием мобильных элементов (МЭ) в последовательность транскрипта. Мобильные элементы могут способствовать регуляции экспрессии генов при развитии и адаптации, а также служить основными источниками генетических вариаций в эволюции геномов. МЭ присутствуют в экзонах, сайтах начала транскрипции, сайтах полиаденилирования (поли(А)-хвосты) или в некоторых комбинациях этих частей. Исследование, проведенное на восьми растениях (включаящих *A. thaliana*, кукурузу, томат, тополь) [109], показало, что 59.7% днРНК, найденных во всех восьми организмах, имеют по крайней мере один

экзон с частично встроенным МЭ. В исследовании [43] показано участие МЭ не только в появлении днРНК, но и в регуляции и вариабельности днРНК.

Для оценки консервативности днРНК между геномами различных организмов сравнивают их последовательности [110]. Например, при выравнивании 67 тыс. днРНК риса на 39 тыс. днРНК кукурузы доля совпадающих позиций гомологов днРНК равна 20% [111]. При сравнении днРНК для эволюционно более далеких видов, например *A. thaliana* и *O. sativa*, уровень сходства значительно падает вплоть до отсутствия гомологии [109]. Только небольшое количество последовательностей днРНК демонстрирует уровень консервативности, сравнимый с консервативностью генов, кодирующих белки [112]. Одной из таких днРНК является IPS, участвующая в гомеостазе фосфора (Р_и) у *A. thaliana*, *O. sativa*, *S. lycopersicum* и *S. moellendorffii*. IPS содержит консервативный мотив из 24 нуклеотидов в перечисленных видах, который служит мишенью для miR399. IPS первоначально появилась у *S. moellendorffii*, представителя древних споровых высших растений. За время эволюционного развития функция IPS не изменилась, что может указывать на роль данной днРНК в адаптации к сухому процессе миграции из водной среды.

Однако большинство днРНК показывают высокую вариабельность первичной структуры, что подтверждается в работе [113], посвященной анализу консервативности днРНК между 25 цветковыми растениями. Анализ показал, что только 6.79% днРНК консервативны у всех исследуемых растений. В работе [114] был проведен анализ консервативности днРНК (межгенных) риса относительно семи видов растений. В результате была показана низкая консервативность днРНК в сравнении с мРНК, однако более высокая в сравнении с мРНК, содержащими МЭ. Анализ консервативности структурных элементов днРНК показал высокую консервативность экзонов днРНК в сравнении с инtronами. В данных исследованиях показана быстрая эволюция днРНК, которая приводит к низкой консервативности последовательностей. Также отмечается, что процент консервативных днРНК выше среди родственных видов (в пределах одного рода), при этом консервативность резко убывает по мере увеличения эволюционного расстояния [113].

Несмотря на существующие теории возникновения и эволюции днРНК, к ним следует относиться с соответствующим скептицизмом вследствие отсутствия экспериментального подтверждения.

днРНК В ПАН-ГЕНОМАХ И ПАН-ТРАНСКРИПТОМАХ

В природе существует высокое геномное и морфологическое разнообразие у видов, вызванное структурными вариациями. Они включают в себя инсерцию, дупликацию, инверсию, транслокацию, а также вариации копийности генов и их наличия или отсутствия. Данные вариации способны влиять на сегменты ДНК длиной в несколько мегабаз и в некоторых случаях изменять количество хромосом. Вследствие этого один референсный геном не способен охватить все содержимое гена одного вида. Например, в референсных геномах растений отсутствует ряд важных для сельскохозяйственных исследований генов. Для решения данной проблемы используется пан-геном.

Концепция пан-генома подразумевает охват последовательностей, подверженных структурной вариации и, возможно, отсутствующих в референсной последовательности каждого представителя вида. На сегодняшний день существует несколько формулировок термина “пан-геном”, которые представлены в работах [36–40].

В данном обзоре под пан-геномом понимается полный неизбыточный набор генов, представленных в геноме хотя бы одной особи, относящейся к данной таксономической единице [115]. Отметим, что это понятие можно распространить на таксономические единицы любого порядка [116]. Так, была проведена попытка создать общий пан-геном бактерий на основе 573 секвенированных ранее геномов бактерий [117]. Однако в настоящее время понятие пан-геном чаще всего употребляют именно в отношении вида.

Гены “пан-генома” могут быть разделены на две группы.

1. Консервативные гены – присутствуют в геномах всех представителей таксономической единицы.

2. Вариабельные гены – присутствуют в геномах не всех представителей вида. В свою очередь делятся на две категории: гены, представленные в двух и более геномах, и уникальные гены, представленные только в одном индивидуальном геноме.

При увеличении количества вариабельных генов, включенных в какой-либо конкретный пан-геном, общее количество генов в этом пан-геноме будет увеличиваться, в то время как количество консервативных генов будет постепенно уменьшаться. Причем каждая из этих выборок рано или поздно выйдет на плато, после которого она не будет значимо изменяться с добавлением новых геномов. При этом отмечается, что вариабельные гены в среднем эволюционно более молодые, чем консервативные [36]. Этот результат кажется логичным в свете того, что большую часть консервативных составляют гены домашнего

хозяйства. Функции консервативных и вариабельных генов различны. Функции консервативных связаны с важнейшими клеточными процессами, а вариабельных – с окружающей средой и защитными реакциями, рецепторной и антиоксидантной активностью, регуляцией генов и передачей сигнала.

При помощи консервативных и вариабельных генов можно подразделять пан-геномы на две категории – открытые и закрытые. Закрытыми принято считать пан-геномы, которые быстро выходят на плато по количеству генов, т. е. секвенирование геномов новых представителей исследуемого вида не привносит новых генов. Открытые, наоборот, увеличиваются в размере каждый раз при добавлении новых геномов в их состав.

Еще одним из методов анализа генного состава у нескольких представителей какого-либо таксона является анализ его транскриптомов. Нуклеотидные последовательности транскриптов позволяют оценить присутствие генов в геноме только в том случае, если они экспрессируются в какой-либо ткани или органе растения. Таким образом, определение пан-транскриптома можно сформулировать так: пан-транскриптом отражает совокупность всех молекул РНК, присутствующих в данном виде или особи, и состоит из консервативных и вариабельных транскриптов (уникальных для одного вида или особи).

Помимо одноклостидных полиморфизмов (SNP) и множественных структурных вариаций немаловажную роль играют днРНК. Например, полное секвенирование транскриптома показало, что транскрибируется чрезвычайно большая часть генов высших эукариот. Однако более 90% всех транскриптов не транслируются в белки [4, 44, 78, 118] и являются некодирующими последовательностями. Как было описано выше, днРНК выполняют важную роль во многих молекулярных и биологических процессах у растений. Однако, несмотря на растущее число посвященных структурно-функциональному анализу днРНК исследований, данный тип молекул по-прежнему остается малоизученным. Это связано со множеством факторов, которые нужно учитывать при идентификации днРНК. Размер днРНК схож с размером некоторых генов, кодирующих белки, что приводит к ошибкам перепредсказания. Также исследования показывают, что последовательности днРНК перерпевают быструю эволюцию, закономерности которой пока непонятны. Как правило, в процессе эволюции эти последовательности очень быстро накапливают замены, так что даже с последовательностями из близкородственных видов организмов они имеют слабую гомологию [113]. Таким образом, применение пан-геномов и пан-транскриптомов повысит эффективность исследования и общее количество предсказанных днРНК

по сравнению с использованием генома одного представителя вида. Также применение пан-генома позволит повысить точность и полноту исследуемого набора данных, так как пан-геном является совокупностью всех кодирующих и некодирующих последовательностей всех представителей вида. Более того, предсказанные для разных представителей одного вида днРНК могут быть выровнены на один пан-геном данного вида, это позволит произвести дальнейшее исследование вариабельных генов, поможет определить точное количество, тип и функции днРНК, внесенных каждым представителем исследуемого вида. Данный подход также позволит избежать необходимости выравнивать, сопоставлять и объединять результаты для геномов разных представителей одного вида, что существенно сократит время, трудозатраты и вычислительные ресурсы, необходимые для данного исследования. Например, исследование пан-транскриптома люцерны (*Medicago sativa* L.) [119] в тканях листьев, стебля и корней выявило 11677 новых днРНК. Данное исследование позволило выявить 2105 высоко экспрессионных и 1894 низко экспрессионных днРНК, а анализ генов онтологии показал вовлеченность данных днРНК в процессы регуляции устойчивости люцерны к засухе и солевому стрессу. В работе [120] при анализе пан-транскриптома кукурузы (*Zea mays*) было выявлено 644 новых днРНК. Однако на сегодняшний день работы, посвященные исследованию пан-геномов и пан-транскриптомов, в основном направлены на выявление и исследование новых генов, кодирующих белки. Тогда как работ, посвященных исследованию днРНК в масштабах пан-геномов и пан-транскриптомов, не так много, в особенности для растений.

ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ИДЕНТИФИКАЦИИ днРНК

До недавнего времени идентификация днРНК считалась побочным результатом исследований, направленных на выявление генов, кодирующих белки. Однако за последнее десятилетие стало очевидно, как важна роль днРНК в молекулярных и биологических процессах у различных организмов. Это привело к развитию экспериментальных технологий, специально направленных на идентификацию днРНК. На сегодняшний день для выявления и аннотации днРНК широко используют следующие экспериментальные методы [121]:

- 1) анализ РНК-секвенирования – выявление новых транскриптов днРНК;
- 2) микрочипы, разработанные для днРНК;
- 3) SAGE, CAGE – идентификация новых днРНК;
- 4) PARS, SHAPE – предсказание структуры днРНК.

Однако перед началом эксперимента необходимо подготовить днРНК. Для этого применяются различные методы очистки РНК, такие как денатурация путем нагревания, что может негативно повлиять на структуру днРНК [122]. Таким образом, разработаны новые методы, позволяющие избежать денатурации РНК. В большинстве этих подходов используются аффинные метки, которые участвуют в иммобилизации целевых РНК и рибозимов, что обеспечивает более высокую специфичность во время элюции [123]. Метод, предложенный [124], позволяет выявить днРНК с помощью ультрафильтрации и гель-фильтрации. Этот протокол очистки позволяет анализировать днРНК с сохраненными паттернами ко-транскрипционной укладки, а также поддерживает потенциально функциональные структурные элементы [124].

Анализ РНК-секвенирования (RNA-seq) является наиболее распространенным методом идентификации и аннотирования новых днРНК [121]. РНК-секвенирование (1) не требует предварительного знания изучаемых последовательностей, а значит, может быть использовано для изучения транскриптомов немодельных организмов, (2) дает более точные оценки уровня экспрессии и (3) требует меньшего количества РНК для проведения эксперимента [125]. Модификации анализа РНК-секвенирования, такие как цепь-специфичное РНК-секвенирование, позволяют разделять смысловые и антисмысловые классы днРНК [95].

Микрочипы были разработаны для анализа транскриптома, а также количественного и сравнительного анализа экспрессии мРНК между одним или двумя образцами различного происхождения. Стандартные платформы для микрочипов не содержат аннотации днРНК и не подходят для их выявления и анализа. Однако благодаря новым достижениям в технологии микрочипов теперь стало возможным создавать индивидуальные олигонуклеотидные микрочипы [126]. Микрочипы, разработанные для днРНК, используются для исследования профиля экспрессии днРНК в клетках и тканях [121]. Например, в работе [127] применяли микрочип с содержанием 3478 межгенных и инtronных последовательностей в геномах человека, мыши и крысы для поиска днРНК. Исследователи обнаружили 55 новых днРНК с высокой экспрессией, из которых экспрессия восьми была подтверждена в тканях мышей при помощи Нозерн-блот-анализа.

Оба молекулярно-генетических метода – серийный анализ генной экспрессии (SAGE) и кэп-анализ экспрессии генов (CAGE) – основаны на получении и прочтении коротких (обычно длиной 27 нуклеотидов) участков последовательности РНК. Это позволяет одновременно количественно и качественно охарактеризовать экспрессию тысяч различных генов путем анализа их транскриптов

[128]. В отличие от аналогичного метода серийного анализа экспрессии генов (SAGE, superSAGE), в котором фрагменты происходят из других частей транскриптов, SAGE в основном используется для определения точного местоположения сайтов начала транскрипции в геноме. Эти знания в свою очередь позволяют исследователю определять структуру промотора, необходимую для экспрессии гена [128]. Например, метод SAGE был применен в работе [96], посвященной изучению молекулярного патогенеза и биомаркеров предраковых поражений полости рта. В работе были идентифицированы 10 дифференциально экспрессирующихся днРНК, среди которых наибольшую экспрессию имеет днРНК NEAT1, принимающая участие в метастазировании рака.

Методы SHAPE и PARS [129, 130] позволяют получать данные о способе действия и взаимодействиях РНК с другими регуляторными молекулами [121]. В основе метода SHAPE (выборочное ацилирование 2'-гидроксильной группы, анализируемое методом удлинения праймера) лежит использование ангидрида N-метилизатиновой кислоты (NMIA) и его производных для исследования гибких регионов во вторичной структуре РНК. Метод SHAPE был применен для исследования вторичной структуры таких днРНК, как COOLAIR [131] и Braveheart [132]. В методе PARS применяются сразу две нуклеазы: S1 и V1. Первая нуклеаза (S1) расщепляет одноцепочечную ДНК или РНК, вторая (V1) – двухцепочечную. Таким образом, при проведении двух экспериментов параллельно можно получить информацию сразу и о неспаренных, и о спаренных нуклеотидах [130].

Несмотря на перечисленное разнообразие экспериментальных методов идентификации днРНК, данный тип молекул по-прежнему остается малоизученным. В первую очередь это связано с наличием общих характеристик днРНК с белок-кодирующими генами, описанных в разделе “Молекулярные функции днРНК”. Поэтому применение только экспериментальных методов для идентификации днРНК является неэффективным подходом для их исследования. Для массового их изучения в масштабах всего генома лучше применять биоинформационные методы в комбинации с экспериментальными.

БИОИНФОРМАТИЧЕСКИЕ МЕТОДЫ ИДЕНТИФИКАЦИИ днРНК

Биоинформационные подходы основываются на решении следующих задач: предобработка данных, предсказание днРНК и структурно-функциональная аннотация [16, 24, 133]. Однако на сегодняшний день за счет применения технологий секвенирования нового поколения стремительно увеличивается количество экспериментов

по тотальному секвенированию РНК (RNA-seq). Чтобы повысить эффективность анализа транскриптомов для большого количества обрабатываемых данных, активно разрабатываются вычислительные конвейеры, которые позволяют выполнять стандартизованные этапы обработки данных в режиме, не требующем контроля пользователем вплоть до получения конечного результата. Главными достоинствами существующих конвейеров являются возможность сборки транскриптома для исследуемого организма и анализ дифференциальной экспрессии днРНК. Более подробное сравнение функционала биоинформационных конвейеров представлено в таблице П1.

Биоинформационные конвейеры позволяют автоматизировать и ускорить этапы анализа днРНК, перечисленные выше. Таким образом, предобработка данных включает контроль качества прочтений с помощью стандартных программ и выравнивание на геном.

Предсказания днРНК осуществляются программами CPC2 [134], CPAT [135], PLEK [133]. Перечисленные программы основаны на использовании методов машинного обучения (метод опорных векторов) для определения кодирующего потенциала исследуемого транскрипта. Основными критериями, по которым происходит оценка кодирующего потенциала, являются признаки открытых рамок считывания (ОРС). Встречаются случаи сходства признаков ОРС днРНК с ОРС генов, кодирующих белки, однако покрытие или целостность ОРС обычно у днРНК хуже.

Аннотация днРНК включает анализ структурных особенностей (экзон-инtronная структура, положение на хромосоме), экспрессии, ко-экспрессии, гомологичных последовательностей, консервативности. Анализ структурных особенностей днРНК в основном направлен на длину транскрипта, количество экзонов/инtronов и расположение на хромосоме. Анализ консервативности днРНК в работах можно разделить на две группы: консервативность внутри одного организма и консервативность между разными организмами. Консервативными считаются днРНК с уровнем гомологии выше 50% с днРНК других организмов, неконсервативными – с отсутствием гомологии с другими организмами [134].

Однако несмотря на достоинства конвейеров, они имеют свои недостатки. Большинство конвейеров разработаны на основе транскриптомных данных модельных организмов (человека, мыши, мухи). Вследствие этого для исследования другого организма нужно или использовать уже имеющиеся модели, что может внести ошибки в предсказание днРНК, или дополнительно обучать модель предсказания для данного организма, а затем

вносить эти данные в код конвейера, что может привести к неправильной работе программы.

Отметим, что, несмотря на общие черты, между структурой и функциями днРНК растений и животных существуют и отличия [31]. Например, в работе [136], посвященной анализу программ предсказания днРНК, было показано, что программы, разработанные для человека, плохо предсказывают днРНК растений и наоборот, что свидетельствует о различии структурных особенностей днРНК между человеком и растениями. Однако экспериментальных работ, направленных на сравнение структуры днРНК растений и животных, пока нет.

Поэтому для более точного выявления днРНК необходимы дополнительные фильтры последовательностей, не все конвейеры их предоставляют. На сегодняшний день современные биоинформационные конвейеры решают перечисленные выше недостатки. Так, ICAnnoLncRNA [137] и lncRNA-screen [138] позволяют обучать модель предсказания для любого исследуемого вида. Также ICAnnoLncRNA предоставляет несколько фильтров последовательностей (выравнивание на референсный геном, удаление мобильных элементов, фильтрация коротких изоформ и классификации), что позволяет получить наиболее точное количество выявленных днРНК.

При биоинформационном анализе днРНК производится сравнение вновь секвенированных транскриптов с уже известными последовательностями днРНК. Источником информации об этих последовательностях служат базы данных. Для их пополнения используются различные источники: результаты биоинформационической обработки транскриптомных экспериментов и предсказания днРНК, экспериментальные результаты, полученные при анализе литературных источников, совместное использование как биоинформационических методов, так и литературных источников. Базы данных, которые описывают структуру и функции днРНК растений, приведены в таблице П2.

ЗАКЛЮЧЕНИЕ

На сегодняшний день различные исследования демонстрируют важность и незаменимость днРНК во многих клеточных процессах, таких как активация и подавление экспрессии генов, импринтинг и деметилирование ДНК, РНК-интерференция, remodeling хроматина и др. [139]. У растений днРНК участвуют в регуляции устойчивости к различным видам стресса [1–3], оказывают влияние на устойчивость к гипоксии [32], принимают участие в развитии плодов, корня и листьев. Таким образом, все больше повышается интерес к выявлению днРНК в различных видах. Для этого используют методы широкомасштабного распознавания последовательностей в геномах и транскриптомах.

Особый интерес представляет выявление днРНК в пан-транскриптомах и пан-геномах. Во многих работах [119, 120, 140], посвященных исследованию пан-геномов, отмечается более полная и точная информация о выявленных последовательностях генов. Однако, несмотря на растущее количество исследований, посвященных днРНК, их выявление по-прежнему остается проблематичной задачей. Для решения данных задач разрабатывают различные экспериментальные и биоинформационические подходы.

Работа выполнена при поддержке бюджетного проекта FWNR-2022-0020.

Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта людей и животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Deng P., Liu S., Nie X., Wu L.* Conservation analysis of long non-coding RNAs in plants // *Sci. China Life Sci.* 2018. V. 61. P. 190–198.
2. *Wu H.-J., Wang Z.-M., Wang M., Wang X.-J.* Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants // *Plant Physiol.* 2013. V. 161. № 4. P. 1875–1884.
3. *Zhu Q.-H., Wang M.-B.* Molecular functions of long non-coding RNAs in plants // *Genes.* 2012. V. 3. № 1. P. 176–190.
4. *Назипова Н.Н.* Разнообразие некодирующих РНК в геномах эукариот // *Матем. биол. и биоинформат.* 2021. Т. 16. № 2. С. 256–298.
5. *Joshi A., Romanowska J.* Recent advances in computational-based approaches in epigenetics studies // *Epigenetics Methods.* 2020. P. 569–590.
6. *Kim E.-D., Sung S.* Long noncoding RNA: Unveiling hidden layer of gene regulatory networks // *Trends Plant Sci.* 2012. V. 17. № 1. P. 16–21.
7. *Karlik E., Ari S., Gozukirmizi N.* LncRNAs: Genetic and epigenetic effects in plants // *Biotechnol. Biotechnol.* 2019. V. 33. № 1. P. 429–439. <https://doi.org/10.1080/13102818.2019.1581085>
8. *Tsai M.-C., Manor O., Wan Y. et al.* Long noncoding RNA as modular scaffold of histone modification complexes // *Science.* 2010. V. 329. № 5992. P. 689–693. <https://doi.org/10.1126/science.1192002>
9. *Sousa C., Johansson C., Charonet C. et al.* Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA,

- for elicitation of a cell-specific growth response in the alfalfa root cortex // *Mol. Cell. Biol.* 2001. V. 21. № 1. P. 354–366.
<https://doi.org/10.1128/MCB.21.1.354-366.2001>
10. *Medvedeva Y.A., Lennartsson A., Ehsani R. et al.* EpiFactors: A comprehensive database of human epigenetic factors and complexes // *Database*. 2015. V. 2015. P. bav067.
11. *Frankish A., Diekhans M., Jungreis I. et al.* GENCODE 2021 // *Nucl. Acids Res.* 2021. V. 49. № D1. P. D916–D923.
12. *Гордиюк В.В.* Длинные некодирующие РНК – камертон в регуляции клеточных процессов // *Ukr. Biochem. J.* 2014. V. 86. № 2. С. 5–15.
13. *Zhao X., Li J., Lian B. et al.* Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA // *Nat. Commun.* 2018. V. 9. № 1. P. 5056.
14. *Li X., Wu Z., Fu X. et al.* lncRNAs: Insights into their function and mechanics in underlying disorders // *Mutat. Res. Mutat. Res.* 2014. V. 762. P. 1–21.
15. *Ahmad P., Bensaoud C., Mekki I. et al.* Long non-coding RNAs and their potential roles in the vector–host–pathogen triad // *Life*. MDPI. 2021. V. 11. № 1. P. 56.
16. *De Quattro C., Pè M.E., Bertolini E.* Long noncoding RNAs in the model species *Brachypodium distachyon* // *Sci. Rep.* 2017. V. 7. № 1. P. 11252.
17. *Ma L., Bajic V.B., Zhang Z.* On the classification of long non-coding RNAs // *RNA Biol.* 2013. V. 10. № 6. P. 924–933.
<https://doi.org/10.4161/rna.24604>
18. *Chen L., Zhu Q.-H., Kaufmann K.* Long non-coding RNAs in plants: Emerging modulators of gene activity in development and stress responses // *Planta*. 2020. V. 252. № 5. P. 92.
<https://doi.org/10.1007/s00425-020-03480-5>
19. *Griffiths-Jones S.* miRBase: The microRNA sequence database // *MicroRNA Protocols*. 2006. V. 342. P. 129–138.
20. *Amaral P.P., Mattick J.S.* Noncoding RNA in development // *Mamm. Genome*. 2008. V. 19. P. 454–492.
21. *Бейлерли О.А., Гареев И.Ф.* Длинные некодирующие РНК: какие перспективы? // Профилактическая медицина. 2020. Т. 23. № 2. С. 124–128.
22. *Blythe A.J., Fox A.H., Bond C.S.* The ins and outs of lncRNA structure: How, why and what comes next? // *Biochim. Biophys. Acta BBA-Gene Regul. Mech.* 2016. V. 1859. № 1. P. 46–58.
23. *Bryzghalov O., Makałowska I., Szcześniak M.W.* lncEvo: Automated identification and conservation study of long noncoding RNAs // *BMC Bioinformatics*. 2021. V. 22. № 1. P. 59.
<https://doi.org/10.1186/s12859-021-03991-2>
24. *Zhao Q., Sun Y., Wang D. et al.* LncPipe: A nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data // *J. Genet. Genomics*. 2018. V. 45.
<https://doi.org/10.1016/j.jgg.2018.06.005>
25. *Talyan S., Filipów S., Ignarski M. et al.* CALINCA – a novel pipeline for the identification of lncRNAs in podocyte disease // *Cells*. MDPI. 2021. V. 10. № 3. P. 692.
26. *Campalans A., Kondorosi A., Crespi M.* Enod40, a short open reading frame – containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula* // *Plant Cell*. 2004. V. 16. № 4. P. 1047–1059.
27. *Unver T., Tombuloglu H.* Barley long non-coding RNAs (lncRNA) responsive to excess boron // *Genomics*. 2020. V. 112. № 2. P. 1947–1955.
28. *Khorkova O., Hsiao J., Wahlestedt C.* Basic biology and therapeutic implications of lncRNA // *Adv. Drug Deliv. Rev.* 2015. V. 87. P. 15–24.
29. *Duret L., Chureau C., Samain S. et al.* The *Xist* RNA Gene evolved in eutherians by pseudogenization of a protein-coding gene // *Science*. 2006. V. 312. № 5780. P. 1653–1655.
<https://doi.org/10.1126/science.1126316>
30. *Graf J., Kretz M.* From structure to function: Route to understanding lncRNA mechanism // *BioEssays*. 2020. V. 42. № 12.
<https://doi.org/10.1002/bies.202000027>
31. *Golicz A.A., Singh M.B., Bhalla P.L.* The long intergenic noncoding RNA (lncRNA) landscape of the soybean genome // *Plant Physiol.* 2018. V. 176. № 3. P. 2133–2147.
32. *Cheng F., Wu J., Fang L., Wang X.* Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species // *Front. Plant Sci. Frontiers*. 2012. V. 3. P. 30895.
33. *Huang L., Dong H., Zhou D. et al.* Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa* // *Plant J.* 2018. V. 96. № 1. P. 203–222.
<https://doi.org/10.1111/tpj.14016>
34. *Derrien T., Johnson R., Bussotti G. et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression // *Genome Res.* 2012. V. 22. № 9. P. 1775–1789.
35. *Ponting C.P., Oliver P.L., Reik W.* Evolution and functions of long noncoding RNAs // *Cell*. 2009. V. 136. № 4. P. 629–641.
36. *Golicz A.A., Bayer P.E., Barker G.C. et al.* The pangenome of an agriculturally important crop plant *Brassica oleracea* // *Nat. Commun.* 2016. V. 7. № 1. P. 13390.
37. *Meile L., Croll D., Brunner P.C. et al.* A fungal avirulence factor encoded in a highly plastic genomic

- region triggers partial resistance to *Septoria tritici* blotch // *New Phytol.* 2018. V. 219. № 3. P. 1048–1061. <https://doi.org/10.1111/nph.15180>
38. *Alcaraz L.D., Moreno-Hagelsieb G., Eguiarte L.E. et al.* Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics // *BMC Genomics.* 2010. V. 11. № 1. P. 332. <https://doi.org/10.1186/1471-2164-11-332>
39. *Rasko D.A., Rosovitz M.J., Myers G.S.A. et al.* The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates // *J. Bacteriol.* 2008. V. 190. № 20. P. 6881–6893. <https://doi.org/10.1128/JB.00619-08>
40. *Merot-Lanthoene V., Tournebize R., Darracq O. et al.* Development and evaluation of a genome – wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. // *Plant Biotechnol. J.* 2019. V. 17. № 7. P. 1418–1430. <https://doi.org/10.1111/pbi.13066>
41. *Budak H., Kaya S.B., Cagirici H.B.* Long non-coding RNA in plants in the era of reference sequences // *Front. Plant Sci. Frontiers.* 2020. V. 11. P. 441273.
42. *Britto-Kido S. de A., Neto J.R.C.F., Pandolfi V. et al.* Natural antisense transcripts in plants: A review and identification in soybean infected with Phakopsora pachyrhizi SuperSAGE Library // *Sci. World J.* 2013. V. 2013.
43. *Kapusta A., Kronenberg Z., Lynch V.J. et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs // *PLoS Genet.* 2013. V. 9. № 4. P. e1003470.
44. *Лукина С.С., Бурденный А.М., Филиппова Е.А. et al.* Роль длинных некодирующих РНК и ДНК-метилирования в патогенезе рака яичников // *Патол. физиол. и эксперим. терапия.* 2022. Т. 66. № 4. С. 143–156.
45. *Ulitsky I., Bartel D.P.* lincRNAs: Genomics, evolution, and mechanisms // *Cell.* 2013. V. 154. № 1. P. 26–46.
46. *Dinger M.E., Pang K.C., Mercer T.R. et al.* Differentiating protein-coding and noncoding RNA: Challenges and ambiguities // *PLoS Comput. Biol.* 2008. V. 4. № 11. P. e1000176.
47. *Xie C., Yuan J., Li H. et al.* NONCODEv4: Exploring the world of long non-coding RNA genes // *Nucleic Acids Res.* 2014. V. 42. № D1. P. D98–D103.
48. *Pal D., Rao M.R.S.* Long noncoding RNAs in pluripotency of stem cells and cell fate specification // *Long Non Coding RNA Biology* / Ed. Rao M.R.S. Singapore: Springer, 2017. V. 1008. P. 223–252.
49. *Татосян К.А., Зиневич Л.С., Демин Д.Э., Шварц А.М.* Функциональные особенности длинных некодирующих РНК, содержащих последовательности мобильных генетических элементов // *Мол. биол.* 2020. Т. 54. № 5. P. 718–724.
50. *Zhao Z., Zang S., Zou W. et al.* Long non-coding RNAs: New players in plants // *Int. J. Mol. Sci. MDPI.* 2022. V. 23. № 16. P. 9301.
51. *Kopp F., Mendell J.T.* Functional classification and experimental dissection of long noncoding RNAs // *Cell.* Elsevier. 2018. V. 172. № 3. P. 393–407.
52. *Huarte M., Rinn J.L.* Large non-coding RNAs: missing links in cancer? // *Hum. Mol. Genet.* 2010. V. 19. № R2. P. R152–R161.
53. *Hung T., Wang Y., Lin M.F. et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters // *Nat. Genet.* 2011. V. 43. № 7. P. 621–629.
54. *Loewer S., Cabilio M.N., Guttman M. et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells // *Nat. Genet.* 2010. V. 42. № 12. P. 1113–1117.
55. *Wang K.C., Chang H.Y.* Molecular mechanisms of long noncoding RNAs // *Mol. Cell.* 2011. V. 43. № 6. P. 904–914.
56. *Chen J., Wang H., Yao Y.* Experimental study of nonlinear ultrasonic behavior of soil materials during the compaction // *Ultrasonics.* 2016. V. 69. P. 19–24.
57. *Kotake Y., Nakagawa T., Kitagawa K. et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene // *Oncogene.* 2011. V. 30. № 16. P. 1956–1962.
58. *Ye X., Wang S., Zhao X. et al.* Role of lncRNAs in *cis*- and *trans*-regulatory responses to salt in *Populus trichocarpa* // *Plant J.* 2022. V. 110. № 4. P. 978–993. <https://doi.org/10.1111/tpj.15714>
59. *Yang L., Froberg J.E., Lee J.T.* Long noncoding RNAs: Fresh perspectives into the RNA world // *Trends Biochem. Sci.* 2014. V. 39. № 1. P. 35–43.
60. *Lam M.T.Y., Cho H., Lesch H.P. et al.* Rev-Erbα repress macrophage gene expression by inhibiting enhancer-directed transcription // *Nature.* 2013. V. 498. № 7455. P. 511–515.
61. *Kim T.-K., Hemberg M., Gray J.M.* Enhancer RNAs: A class of long noncoding RNAs synthesized at enhancers // *Cold Spring Harb. Perspect. Biol.* 2015. V. 7. № 1. P. a018622.
62. *Kim T.-K., Hemberg M., Gray J.M. et al.* Widespread transcription at neuronal activity-regulated enhancers // *Nature.* 2010. V. 465. № 7295. P. 182–187.
63. *Lai F., Orom U.A., Cesaroni M. et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription // *Nature.* 2013. V. 494. № 7438. P. 497–501.
64. *Melo C.A., Drost J., Wijchers P.J. et al.* eRNAs are required for p53-dependent enhancer activity and

- gene transcription // *Mol. Cell.* 2013. V. 49. № 3. P. 524–535.
65. *Zhang P., Meng J., Luan Y. et al.* Plant miRNA–lncRNA interaction prediction with the ensemble of CNN and IndRNN // *Interdiscip. Sci. Comput. Life Sci.* 2020. V. 12. P. 82–89.
66. *Chen K., Rajewsky N.* The evolution of gene regulation by transcription factors and microRNAs // *Nat. Rev. Genet.* 2007. V. 8. № 2. P. 93–103.
67. *Jin Q., Zhao Z., Zhao Q. et al.* Long noncoding RNAs: Emerging roles in pulmonary hypertension // *Heart Fail. Rev.* 2020. V. 25. P. 795–815.
68. *Huntzinger E., Izaurralde E.* Gene silencing by microRNAs: Contributions of translational repression and mRNA decay // *Nat. Rev. Genet.* 2011. V. 12. № 2. P. 99–110.
69. *Yoon J.-H., Abdelmohsen K., Kim J. et al.* Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination // *Nat. Commun.* 2013. V. 4. № 1. P. 2939.
70. *Mukherjee N., Corcoran D.L., Nusbaum J.D. et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability // *Mol. Cell.* 2011. V. 43. № 3. P. 327–339.
71. *Thomson D.W., Dinger M.E.* Endogenous microRNA sponges: Evidence and controversy // *Nat. Rev. Genet.* 2016. V. 17. № 5. P. 272–283.
72. *Franco-Zorrilla J.M., Valli A., Todesco M. et al.* Target mimicry provides a new mechanism for regulation of microRNA activity // *Nat. Genet.* 2007. V. 39. № 8. P. 1033–1037.
73. *Du Q., Wang K., Zou C. et al.* The PILNCR1-miR399 regulatory module is important for low phosphate tolerance in maize // *Plant Physiol.* 2018. V. 177. № 4. P. 1743–1753.
74. *Wang T., Zhao M., Zhang X. et al.* Novel phosphate deficiency-responsive long non-coding RNAs in the legume model plant *Medicago truncatula* // *J. Exp. Bot.* 2017. V. 68. № 21–22. P. 5937–5948.
75. *Faghihi M.A., Zhang M., Huang J. et al.* Evidence for natural antisense transcript-mediated inhibition of microRNA function // *Genome Biol.* 2010. V. 11. № 5. <https://doi.org/10.1186/gb-2010-11-5-r56>
76. *Kimura T., Jiang S., Nishizawa M. et al.* Stabilization of human interferon- α 1 mRNA by its antisense RNA // *Cell. Mol. Life Sci.* 2013. V. 70. № 8. P. 1451–1467. <https://doi.org/10.1007/s00018-012-1216-x>
77. *Wang Y., Pang W.J., Wei N. et al.* Identification, stability and expression of Sirt1 antisense long non-coding RNA // *Gene.* 2014. V. 539. № 1. P. 117–124.
78. *Cai X., Cullen B.R.* The imprinted H19 noncoding RNA is a primary microRNA precursor // *RNA.* 2007. V. 13. № 3. P. 313–316.
79. *Augoff K., McCue B., Plow E.F. et al.* miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer // *Mol. Cancer.* 2012. V. 11. № 1. <https://doi.org/10.1186/1476-4598-11-5>
80. *Kallen A.N., Zhou X.B., Xu J. et al.* The imprinted H19 lncRNA antagonizes let-7 microRNAs // *Mol. Cell.* 2013. V. 52. № 1. P. 101–112.
81. *Amor B.B., Wirth S., Merchan F. et al.* Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses // *Genome Res.* 2009. V. 19. № 1. P. 57–69.
82. *Hirsch J., Lefort V., Vankerschaver M. et al.* Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts // *Plant Physiol.* 2006. V. 140. № 4. P. 1192–1204.
83. *Lamin-Samu A.T., Zhuo S., Ali M., Lu G.* Long non-coding RNA transcriptome landscape of anthers at different developmental stages in response to drought stress in tomato // *Genomics.* 2022. V. 114. № 4. P. 110383.
84. *Kryuchkova-Mostacci N., Robinson-Rechavi M.* A benchmark of gene expression tissue-specificity metrics // *Brief. Bioinform.* 2017. V. 18. № 2. P. 205–214.
85. *Li L., Eichten S.R., Shimizu R. et al.* Genome-wide discovery and characterization of maize long non-coding RNAs // *Genome Biol.* 2014. V. 15. № 2. P. <https://doi.org/10.1186/gb-2014-15-2-r40>.
86. *Han L., Mu Z., Luo Z. et al.* New lncRNA annotation reveals extensive functional divergence of the transcriptome in maize // *J. Integr. Plant Biol.* 2019. V. 61. № 4. P. 394–405. <https://doi.org/10.1111/jipb.12708>
87. *Subramanian S., Kumar S.* Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome // *Genetics.* 2004. V. 168. № 1. P. 373–381.
88. *Yanai I., Benjamin H., Shmoish M. et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification // *Bioinformatics.* 2005. V. 21. № 5. P. 650–659.
89. *Ceriani L., Verme P.* The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini // *J. Econ. Inequal.* 2012. V. 10. P. 421–443.
90. *Julien P., Brawand D., Soumillon M. et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation // *PLoS Biol.* 2012. V. 10. № 5. P. e1001328.
91. *Xiao S.-J., Zhang C., Zou Q., Ji Z.L.* TiSGeD: A database for tissue-specific genes // *Bioinformatics.* 2010. V. 26. № 9. P. 1273–1275.

92. *Yu X., Lin J., Zack D.J., Qian J.* Computational analysis of tissue-specific combinatorial gene regulation: Predicting interaction between transcription factors in human tissues // *Nucl. Acids Res.* 2006. V. 34. № 17. P. 4925–4936.
93. *Huang X., Li S.Z., Wang Y.* Jensen-Shannon boosting learning for object recognition // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE. 2005. V. 2. P. 144–149.
94. *Marquardt S., Raitskin O., Wu Z. et al.* Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription // *Mol. Cell.* 2014. V. 54. № 1. P. 156–165.
95. *Liu X., Hao L., Li D. et al.* Long non-coding RNAs and their biological roles in plants // *Genomics Proteomics Bioinformatics.* 2015. V. 13. № 3. P. 137–147.
96. *Ding J., Lu Q., Ouyang Y. et al.* A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice // *Proc. Natl Acad. Sci. USA.* 2012. V. 109. № 7. P. 2654–2659. <https://doi.org/10.1073/pnas.1121374109>
97. *Song J.-H., Cao J.-S., Wang C.-G.* BcMF11, a novel non-coding RNA gene from *Brassica campestris*, is required for pollen development and male fertility // *Plant Cell Rep.* 2013. V. 32. P. 21–30.
98. *Kim J., Yi H., Choi G. et al.* Functional characterization of phytochrome interacting factor 3 in phytochrome-mediated light signal transduction // *Plant Cell.* 2003. V. 15. № 10. P. 2399–2407.
99. *Wang J., Meng X., Dobrovolskaya O.B. et al.* Non-coding RNAs and their roles in stress response in plants: 5 // *Genomics Proteomics Bioinformatics.* 2017. V. 15. № 5. P. 301–312. <https://doi.org/10.1016/j.gpb.2017.01.007>
100. *Wang A., Hu J., Gao C. et al.* Genome-wide analysis of long non-coding RNAs unveils the regulatory roles in the heat tolerance of Chinese cabbage (*Brassica rapa* ssp. *chinensis*) // *Sci. Rep.* 2019. V. 9. № 1. P. 5002
101. *Wang P., Dai L., Ai J. et al.* Identification and functional prediction of cold-related long non-coding RNA (lncRNA) in grapevine // *Sci. Rep.* 2019. V. 9. № 1. P. 6638.
102. *Chung P.J., Jung H., Jeong D.H. et al.* Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice // *BMC Genomics.* 2016. V. 17. № 1. P. 563. <https://doi.org/10.1186/s12864-016-2997-3>
103. *Calixto C.P., Tzioutziou N.A., James A.B. et al.* Cold-dependent expression and alternative splicing of *Arabidopsis* long non-coding RNAs // *Front. Plant Sci. Frontiers Media SA.* 2019. V. 10. P. 235.
104. *Jha U.C., Nayyar H., Jha R. et al.* Long non-coding RNAs: Emerging players regulating plant abiotic stress response and adaptation // *BMC Plant Biol.* 2020. V. 20. № 1. P. 466. <https://doi.org/10.1186/s12870-020-02595-x>
105. *Zhang X., Dong J., Deng F. et al.* The long non-coding RNA lncRNA973 is involved in cotton response to salt stress // *BMC Plant Biol.* 2019. V. 19. № 1. P. 459. <https://doi.org/10.1186/s12870-019-2088-0>
106. *Wang X., Fan H., Wang B., Yuan F.* Research progress on the roles of lncRNAs in plant development and stress responses // *Front. Plant Sci. Frontiers Media SA.* 2023. V. 14. P. 1138901.
107. *Chen J., Zhong Y., Qi X.* LncRNA TCONS_00021861 is functionally associated with drought tolerance in rice (*Oryza sativa* L.) via competing endogenous RNA regulation // *BMC Plant Biol.* 2021. V. 21. № 1. P. 410. <https://doi.org/10.1186/s12870-021-03195-z>
108. *Kazemzadeh M., Safaralizadeh R., Orang A.V.* LncRNAs: Emerging players in gene regulation and disease pathogenesis // *J. Genet.* 2015. V. 94. P. 771–784.
109. *Zhu Y., Chen L., Hong X. et al.* Revealing the novel complexity of plant long non-coding RNA by strand-specific and whole transcriptome sequencing for evolutionarily representative plant species // *BMC Genomics.* 2022. V. 23. № S4. P. 381. <https://doi.org/10.1186/s12864-022-08602-9>
110. *Ulitsky I.* Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs // *Nat. Rev. Genet.* 2016. V. 17. № 10. P. 601–614.
111. *Wang H., Niu Q.W., Wu H.W. et al.* Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lnc RNAs associated with agriculture traits // *Plant J.* 2015. V. 84. № 2. P. 404–416. <https://doi.org/10.1111/tpj.13018>
112. *Nitsche A., Stadler P.F.* Evolutionary clues in lncRNAs // *WIREs RNA.* 2017. V. 8. № 1. <https://doi.org/10.1002/wrna.1376>
113. *Sang S., Chen W., Zhang D. et al.* Data integration and evolutionary analysis of long non-coding RNAs in 25 flowering plants: 3 // *BMC Genomics.* 2021. V. 22. № 3. P. 739. <https://doi.org/10.1186/s12864-021-08047-6>
114. *Zhang Y.-C., Liao J.Y., Li Z.Y. et al.* Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice // *Genome Biol.* 2014. V. 15. № 12. P. 512. <https://doi.org/10.1186/s13059-014-0512-1>
115. *Pronzin A.Y., Bragina M.K., Salina E.A.* Crop pangenomes // *Vavilov. J. Genet. Breed.* 2021. V. 25. № 1. P. 57.
116. *Vernikos G., Medini D., Riley D.R., Tettelin H. et al.* Ten years of pan-genome analyses // *Curr. Opin. Microbiol.* 2015. V. 23. P. 148–154.

117. *Lapierre P., Gogarten J.P.* Estimating the size of the bacterial pan-genome // *Trends Genet.* 2009. V. 25. № 3. P. 107–110.
118. *Chekanova J.A., Gregory B.D., Reverdatto S.V. et al.* Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome // *Cell.* 2007. V. 131. № 7. P. 1340–1353.
119. *Medina C.A., Samac D.A., Yu L.-X.* Pan-transcriptome identifying master genes and regulation network in response to drought and salt stresses in Alfalfa (*Medicago sativa* L.) // *Sci. Rep.* 2021. V. 11. № 1. P. 17203.
120. *Jin M., Liu H., He C. et al.* Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation // *Sci. Rep.* 2016. V. 6. № 1. P. 18936.
121. *Chowdhary A., Satagopam V., Schneider R.* Long non-coding RNAs: Mechanisms, experimental, and computational approaches in identification, characterization, and their biomarker potential in cancer // *Front. Genet.* 2021. V. 12. <https://doi.org/10.3389/fgene.2021.649619>
122. *Svergun D.I., Koch M.H.* Small-angle scattering studies of biological macromolecules in solution // *Rep. Prog. Phys.* 2003. V. 66. № 10. P. 1735.
123. *Schön P.* Atomic force microscopy of RNA: State of the art and recent advancements // *Seminars in Cell & Developmental Biology.* 2018. V. 73. P. 209–219.
124. *Chillón I., Marcia M., Legiewicz M. et al.* Native purification and analysis of long RNAs // *Methods in Enzymology.* 2015. V. 558. P. 3–37.
125. *Cheung F., Haas B.J., Goldberg S.M.D. et al.* Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology // *BMC Genomics.* 2006. V. 7. № 1. <https://doi.org/10.1186/1471-2164-7-272>
126. *Au P.C.K., Zhu Q.-H.* Identification of lncRNAs using computational and experimental approaches // *Regulatory RNAs / Eds Mallick B., Ghosh Z.* Berlin; Heidelberg: Springer, 2012. P. 319–340.
127. *Babak T., Blencowe B.J., Hughes T.R.* A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription // *BMC Genomics.* 2005. V. 6. № 1. P. 104. <https://doi.org/10.1186/1471-2164-6-104>
128. *Shiraki T., Kondo S., Katayama S. et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage // *Proc. Natl Acad. Sci. USA.* 2003. V. 100. № 26. P. 15776–15781. <https://doi.org/10.1073/pnas.2136655100>
129. *Merino E.J., Wilkinson K.A., Coughlan J.L., Weeks K.M.* RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE) // *J. Am. Chem. Soc.* 2005. V. 127. № 12. P. 4223–4231. <https://doi.org/10.1021/ja043822v>
130. *Kertesz M., Wan Y., Mazor E. et al.* Genome-wide measurement of RNA secondary structure in yeast // *Nature.* 2010. V. 467. № 7311. P. 103–107.
131. *Hawkes E.J., Hennelly S.P., Novikova I.V. et al.* COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures // *Cell Rep.* 2016. V. 16. № 12. P. 3087–3096.
132. *Kim D.N., Thiel B.C., Mrozowich T. et al.* Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution // *Nat. Commun.* 2020. V. 11. № 1. P. 148.
133. *Li A., Zhang J., Zhou Z.* PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme // *BMC Bioinformatics.* 2014. V. 15. № 1. P. 311. <https://doi.org/10.1186/1471-2105-15-311>
134. *Kang Y.-J., Yang D.C., Kong L. et al.* CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features // *Nucl. Acids Res.* 2017. V. 45. № W1. P. W12–W16.
135. *Wang L., Park H.J., Dasari S. et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model // *Nucl. Acids Res.* 2013. V. 41. № 6. P. e74.
136. *Da Costa Negri T., Paschoal A.R., Alves W.A.L.* Comparison tools for lncRNA identification: Analysis among plants and humans // 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE. 2020. P. 1–8.
137. *Pronozin A.Y., Afonnikov D.A.* ICAnnoLncRNA: A snakemake pipeline for a long non-coding-RNA search and annotation in transcriptomic sequences // *Genes.* MDPI. 2023. V. 14. № 7. P. 1331.
138. *Gong Y., Huang H.T., Liang Y. et al.* IncRNA-screen: An interactive platform for computationally screening long non-coding RNAs in large genomics datasets // *BMC Genomics.* 2017. V. 18. № 1. P. 434. <https://doi.org/10.1186/s12864-017-3817-0>
139. *Ким О.И., Кириченко Е.Ю., Кириченко Ю.Г. и др.* Длинные некодирующие РНК, ассоциированные с онкогенезом: биологическое значение и перспективы применения в диагностике // Клин. лаб. диагностика. 2016. Т. 61. № 1. P. 13–16.
140. *Gao L., Gonda I., Sun H. et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor // *Nat. Genet.* 2019. V. 51. № 6. P. 1044–1051.

Table S1. Comparison of lncRNA bioinformatic analysis pipelines

	LncPipe [1]	LncEvo [2]	CALINCA [3]	lncRNA Detector [4]	lncRNA- screen [5]	ICAnnoLncRNA [6]
Quality control and assembly of the transcriptome	+	+	+	—	+	—
Prediction model training for a non-model organism	—	—	—	—	+	+
Possibility of using data for plants	—	—	—	+	+	+
Prediction of lncRNA using specialized programs	+	+	—	+	+	+
Classification of lncRNAs according to their location relative to protein-coding genes	+	+	+	—	+	+
Removal of transcriptional noise and assembly redundancy	—	—	—	—	—	+
Deletion of sequences that are potential mobile elements	—	—	—	—	—	+
Differential expression analysis	+	—	+	—	+	—
Analysis of lncRNA conservation between different organisms	—	+	+	—	+	+
Analysis of lncRNA distribution in tissues based on its expression	+	—	+	—	+	+

Table S2. Databases for lncRNAs in plants

Title	Description	Data source	Reference
PLncDBv2.0	Plant database with more than 1,246,372 lncRNAs predicted for 80 organisms from chlorophytes to embryophytes. Presents information on expression, tissue specificity, mutations, and developmental stages of lncRNA transcripts in 13834 datasets from different organisms. Allows the user to study the relationship between lncRNAs and epigenetic markers	Methods for lncRNA prediction	[7]
PNRD	Includes from 166 plant species, over 25,739 lncRNAs of 16 types. Offers several analytical tools and includes a proprietary genomic search engine, protein-coding sequence prediction methods and a microRNA prediction method	Literature and databases	[8]
NONCODE v6	The database contains 16 animal species and 549,813 lncRNAs belonging to them, and 23 plant species and 94,697 lncRNA transcripts	Literature and databases	[9]
GREENC v2	Includes more than 495,000 lncRNA transcripts for 94 plant and algal species	Methods for lncRNA prediction	[10]
CANTATAdb v2	The database contains more than 239,631 lncRNAs from 36 plant species and 3 algal species	Methods for lncRNA prediction	[11]
EVLncRNAs v2	Includes 4010 lncRNAs for 124 species of organisms. Provides the user with a network of lncRNA interactions with microRNAs, proteins, genes, and other functional elements. Also contains several lncRNA analysis and prediction programs	Literature	[12]

REFERENCES

1. *Zhao Q. et al.* LncPipe: A Nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data // *J Genet Genomics*. 2018. Vol. 45, № 7. P. 399–401.
2. *Bryzgalov O., Makalowska I., Szcześniak M.W.* lncEvo: Automated identification and conservation study of long noncoding RNAs // *BMC Bioinformatics*. 2021. Vol. 22, № 1. P. 59.
3. *Talyan S. et al.* CALINCA—A Novel Pipeline for the Identification of lncRNAs in Podocyte Disease // *Cells*. MDPI, 2021. Vol. 10, № 3. P. 692.
4. *Shukla B. et al.* lncRNADetector: A bioinformatics pipeline for long non-coding RNA identification and MAPsInc: a repository of medicinal and aromatic plant lncRNAs // *RNA Biol*. 2021. Vol. 18, № 12. P. 2290–2295.
5. *Gong Y. et al.* lncRNA-screen: An interactive platform for computationally screening long non-coding RNAs in large genomics datasets // *BMC Genomics*. 2017. Vol. 18, № 1. P. 434.
6. *Pronozin A.Y., Afonnikov D.A.* ICAnnoLncRNA: A Snakemake Pipeline for a Long Non-Coding-RNA Search and Annotation in Transcriptomic Sequences // *Genes*. MDPI, 2023. Vol. 14, № 7. P. 1331.
7. *Jin J. et al.* PLncDB V2. 0: A comprehensive encyclopedia of plant long noncoding RNAs // *Nucleic Acids Res*. Oxford University Press, 2021. Vol. 49, № D1. P. D1489–D1495.
8. *Yi X. et al.* PNRD: A plant non-coding RNA database // *Nucleic Acids Res*. Oxford University Press, 2015. Vol. 43, № D1. P. D982–D989.
9. *Zhao Y. et al.* NONCODE 2016: An informative and valuable data source of long non-coding RNAs // *Nucleic Acids Res*. Oxford University Press, 2016. Vol. 44, № D1. P. D203–D208.
10. *Gallart A.P. et al.* GREENC: A Wiki-based database of plant lncRNAs // *Nucleic Acids Res*. Oxford University Press, 2016. Vol. 44, № Database issue. P. D1161.
11. *Szcześniak M.W. et al.* CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs // *Plant Long Non-Coding RNAs* / ed. Chekanova J.A., Wang H.-L.V. New York, NY: Springer New York, 2019. Vol. 1933. P. 415–429.
12. *Zhou B. et al.* EVLncRNAs 2.0: An updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments // *Nucleic Acids Res*. Oxford University Press, 2021. Vol. 49, № D1. P. D86–D91.

The Role of Long Non-Coding RNAs in Plants

A. Y. Pronozin^{1, 2, *}, D. A. Afonnikov^{1, 2, 3}

¹*Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090 Russia*

²*Kurchatov Genomic Center, Institute of Cytology and Genetics, Siberian Branch,*

Russian Academy of Sciences, Novosibirsk, 630090 Russia

³*Novosibirsk National Research State University, Novosibirsk, 630090 Russia*

*e-mail: pronozinartem95@gmail.com

Long non-coding RNAs (lncRNAs) are a class of linear or circular RNA molecules longer than 200 nucleotides without open reading frames. Experimental studies have shown the involvement of lncRNAs in the regulation of resistance to cold, salt, and heat stress, and in fruit, root, and leaf development. However, experimental methods are labor-intensive and costly approaches and cannot yet be used for genome-wide mass studies of lncRNAs. For this purpose, bioinformatic approaches that aim at large-scale recognition of lncRNA sequences in genomes and transcriptomes have been applied. However, despite the growing number of studies devoted to the structural and functional analysis of lncRNAs, this type of molecule remains poorly understood. This is due to the many factors that need to be considered when identifying lncRNAs. The use of pan-genomes and pan-transcriptomes will improve the efficiency of the study and the total number of predicted lncRNAs compared to using the genome of a single species representative. This review focuses on describing the molecular and biological functions of lncRNAs, experimental and bioinformatic methods of identification, patterns of evolution, detection and analysis of lncRNAs at the scale of pan-genomes and pan-transcriptomes.

Keywords: long non-coding RNAs, microRNAs, transcription regulation, pan-genome.