

УДК 577.12

СПЛАЙСОСОМНЫЕ ИНТРОНЫ: СВОЙСТВА, ФУНКЦИИ И ЭВОЛЮЦИЯ

Обзор

© 2020 И.В. Поверенная^{1,2*}, М.А. Ройтберг^{2,3,4}

¹ Институт общей генетики им. Н.И. Вавилова РАН, 119991 Москва, Россия;
электронная почта: ipoverennaya@gmail.com

² Институт математических проблем биологии РАН – филиал Института прикладной математики
им. М.В. Келдыша РАН, 142290 Пуцдино, Московская обл., Россия

³ Московский физико-технический институт (Национальный исследовательский университет),
141701 Долгопрудный, Московская обл., Россия

⁴ Национальный исследовательский университет «Высшая школа экономики», 101000 Москва, Россия

Поступила в редакцию 26.03.2020

После доработки 25.05.2020

Принята к публикации 25.05.2020

Для большинства эукариотических генов свойственны сплайсосомные интроны – некодирующие последовательности, которые вырезаются из пре-мРНК генов с помощью специального комплекса сплайсосомы в процессе сплайсинга мРНК. Интроны бывают как в белок-кодирующих, так и в РНК-кодирующих генах, причем могут встречаться как в кодирующей области, так и в нетранслируемых областях генов. В связи с высоким уровнем полиморфизма последовательности интронов крайне изменчивы, поэтому долгое время функции интронов связывали только с альтернативным сплайсингом, а эволюцию интронов обычно рассматривали не как эволюцию отдельной единицы генома, а в рамках эволюции экзон-интронной структуры гена. В представленном обзоре рассмотрены теории происхождения интронов; эволюционные события в экзон-интронной структуре, такие как приобретение, потеря и слайдинг (смещение) интронов; функции интронов, известные к настоящему моменту; а также как изменение таких свойств интрона, как длина и фаза, могут влиять на регуляцию процессов с участием генов.

КЛЮЧЕВЫЕ СЛОВА: сплайсосомные интроны, фаза интрона, длина интрона, эволюция, слайдинг, экзон-интронная структура.

DOI: 10.31857/S0320972520070015

ВВЕДЕНИЕ

Интроны были открыты в 1977 году в двух независимых исследованиях [1, 2], в которых было показано, что гены эукариот, в отличие от генов прокариот, содержат вставочные последовательности, которые удаляются из пре-мРНК вскоре после транскрипции во время созревания мРНК. Такие вставочные последовательности было предложено назвать интронами (от английского «INTRagenic regiON» – «внутригенный регион»), а разделяемые ими фрагменты гена – экзонами («EXpressed regiON» – «экспрессирующийся регион») [3].

Так как интроны относятся к некодирующей ДНК (нк-ДНК), которую раньше называли «му-

Принятые сокращения: ЭИС – экзон-интронная структура; нк – некодирующие последовательности; АС – альтернативный сплайсинг; ППИ – параллельное приобретение интронов.

* Адресат для корреспонденции.

сорной», и считали, что она не несет особой смысловой нагрузки, то долгое время стоял вопрос о функциональной значимости интронов. Сейчас возможные функции интронов принято разделять на три категории: 1) функции, связанные со сплайсингом; 2) общие функции нк-ДНК; 3) хранение регуляторных элементов и белок-кодирующих генов внутри себя [4]. В первом случае, кроме участия интронов в альтернативном сплайсинге (АС) и их вклада в разнообразие и мультидоменность белковых структур, важна также регуляторная роль сплайсинга в других молекулярно-биологических процессах. Например, известно, что мРНК, из которых уже были вырезаны интроны, лучше экспортируются из ядра, чем их копии без вырезанных интронов [5]. Важным механизмом регуляции экспрессии генов, по-видимому, является скорость сплайсинга. Более того, само наличие интронов в гене влияет на регуляцию его экспрессии [6, 7]. Расположение интронов также важно: хотя

большинство интронов локализованы в кодирующей последовательности, точно установлено, что некоторые интроны расположены в 5'-нетранслируемой области, где участвуют в регуляции экспрессии генов у животных и растений [8]. Помимо всего, интроны, видимо, являются энхансерами мейотического кроссинговера между белок-кодирующими последовательностями, поскольку вероятность кроссинговера между экзонами, разделенными длинными интронами, гораздо больше, чем между такими же последовательностями, но без интронов [9]. И наконец, в качестве примера для третьей категории функций интронов можно привести следующее наблюдение: почти во всех эукариотических геномах в 5'-областях генов, кодирующих рибосомальные белки, есть интроны, содержащие регуляторные элементы [10].

В настоящий момент известно три типа интронов: интроны группы I и группы II, а также сплайсосомные интроны. Интроны групп I и II были обнаружены в ДНК некоторых бактерий и клеточных органелл. Интроны группы I были также найдены в рибосомальных РНК простейших и в ядрышках некоторых грибов. Для интронов групп I и II характерно наличие особых РНК структур, благодаря которым они способны к автосплайсингу (то есть, самовырезанию из незрелой мРНК). Основное отличие между интронами обеих групп заключается в разных механизмах автосплайсинга. У таких интронов также есть собственные открытые рамки считывания, содействующие как их удалению из РНК, так и распространению интронов по геному через процесс обратной транскрипции. В настоящий момент идентифицировано ~1500 интронов группы I и 200 представителей группы II [11].

Третий и самый многочисленный тип интронов – сплайсосомные интроны – были найдены в ядерных геномах почти всех известных эукариот (исключение составляет нуклеоморф) [12]. В отличие от интронов группы I и II, сплайсосомные интроны не имеют способностей к автосплайсингу, и для их вырезания из пре-мРНК нужен специальный аппарат – сплайсосома, которая представляет из себя сложный биомолекулярный комплекс, состоящий из 5 цепей малой ядерной РНК и сотни белков.

Несмотря на важные различия между разными типами интронов, сходство механизмов сплайсинга интронов группы II и сплайсосомных интронов указывает на возможность эволюционного родства между ними [13, 14].

Последовательность сплайсосомных интронов обычно квази-случайная и не имеет открытой рамки считывания, хотя внутри очень длин-

ных интронов могут быть вложенные гены, иногда на обратной цепи (подробнее в разделе «Функции интронов»). Длина данных интронов широко варьируется между видами: от нескольких десятков (у простейших) до сотен тысяч нуклеотидов у млекопитающих. Плотность интронов (то есть количество интронов на ген) в эукариотах также сильно варьируется: от нескольких интронов на геном до десятков интронов на ген [4]. Вследствие такого разброса плотности и длин интронов эукариотов условно стали разделять на интрон-бедные (большинство одноклеточных эукариот) и интрон-богатые (животные, растения и т.д.) геномы. Было показано [15], что для организмов с плотностью <3 интронов на 1000 п.н. в среднем характерна небольшая длина интронов и нет значимой корреляции между плотностью и длиной интронов, в то время как для более интрон-богатых геномов можно наблюдать сильную положительную корреляцию. Тем не менее даже среди интрон-богатых организмов выделяются позвоночные, которые имеют значительное количество очень длинных интронов. Несмотря на найденные корреляции, исключения были обнаружены как среди интрон-богатых, так и среди интрон-бедных геномов. Интрон-бедные и интрон-богатые организмы различаются не только разной плотностью интронов, но и распределением интронов в геноме: для интрон-богатых геномов характерно относительно равномерное распределение интронов по геному, в то время как в интрон-бедных геномах интроны перепредставлены в 5'-областях генов [16].

Интересно, что многочисленные интроны были также найдены в длинных нк-РНК. Сравнительный анализ более чем 3000 последовательностей РНК такого типа показал, что, возможно, для длинных нк-РНК свойственно сохранение экзон-интронной структуры (ЭИС) [17].

ИНТРОНЫ И СПЛАЙСИНГ

Сплайсосомные интроны можно разделить на две группы в зависимости от того, какая сплайсосома принимает участие в вырезании интрона из пре-мРНК. Большая часть интронов вырезается с помощью U2 сплайсосомы, оставшаяся часть – с помощью U12 сплайсосомы. Для правильного распознавания и дальнейшего удаления интрона крайне важны консенсусные последовательности на экзон-интронных границах. Наиболее распространенные интроны U2 типа, то есть интроны, вырезаемые U2 сплайсосомой, содержат GT динуклеотид на

5'-донорном сайте сплайсинга (далее ДСС) и AG динуклеотид на 3'-акцепторном сайте сплайсинга (далее АСС).

Поскольку сайты сплайсинга с |GT в ДСС и AG| в АСС встречаются повсеместно, их стало принято называть каноническими. Среди неканонических сайтов самыми частыми являются |GC-AG| и |AT-AC| пары конечных динуклеотидов в интронах U2 типа и U12 типа соответственно. У человека такие сайты сплайсинга встречаются в ~0,9 и ~0,09% интронов [18]. У растений было показано схожее распределение сайтов сплайсинга (в среднем): 98,7% для |GT-AG|, 1,2% для |GC-AG|, 0,06% для |AT-AC| и 0,09% для всех остальных [19]. В геномах млекопитающих доля интронов с другими неканоническими вариантами, которые, как считается, имеют U2-U12-подобные сайты сплайсинга, например |GT-TG| или |AT-AG|, занимает всего порядка 0,02%, хотя, вероятно, их число сильно недооценено из-за несовершенства геномных аннотаций [20].

Катализ вырезания интронов с неканоническими сайтами сплайсинга осуществляет U12 сплайсосома, в то время как интроны с |GT-AG| обычно вырезаются U2 сплайсосомой (хотя известны случаи сплайсинга и U12 сплайсосомой). Вопрос о том, какая сплайсосома была первой, все еще остается актуальным. В данный момент наибольшую поддержку имеет гипотеза, что ранние интроны вырезались с помощью U2 сплайсосомы, тем не менее, согласно данным последних реконструкций, обе сплайсосомы были представлены в последнем общем предке эукариот (last common eukaryotic ancestor — LECA).

В настоящее время известны 2 основных механизма распознавания сигнала сплайсинга: определение по экзону и определение по интрону. Эффективность сплайсинга в первом случае зависит от длины экзонов и не зависит от длины интронов. Обратное верно для второго случая [21]. Можно предположить, что определение по экзону в какой-то момент замещается определением по интрону, так как в исследовании Sverdlov et al. [22] было показано, что в новых интронах сигнал сплайсинга базируется на концах соседних экзонов, но со временем он смещается на концы самого интрона.

В качестве сайта протосплайсинга, т.е. сайта, в который может встроиться интрон, была признана консенсусная последовательность (A/C)AG||G. Выбор сайта протосплайсинга для *de novo* интронов, по-видимому, происходит случайным образом, при этом скорость изменения сайта протосплайсинга не отличается от скорости изменения случайного сайта.

Подверженный ошибкам сплайсинг (вследствие слабого донорного сайта) мог в конечном итоге привести к функциональному альтернативному сплайсингу. Считается, что АС возник еще на ранних этапах эволюции эукариот путем накопления aberrantных сплайс-форм в условиях слабого отрицательного отбора. По-видимому, все альтернативные транскрипты в начале были неработающими, и функциональные АС-формы развивались постепенно и независимо в разных таксонах. При этом считается, что АС в значительной степени способствовал повышению уровня фенотипической сложности у млекопитающих, и что активное накопление интронов дало эволюционное преимущество, так как позволило увеличить потенциально возможное число белковых изоформ. [23].

ПРОИСХОЖДЕНИЕ ИНТРОНОВ

Происхождение сплайсосомных интронов представляет большой интерес как в изучении интронов, как самостоятельных элементов генома, так и в изучении эволюции всего генома в целом. Эволюция экзон-интронной организации эукариотических генов до сих пор является предметом долгого ожесточенного спора между сторонниками двух теорий: раннего и позднего происхождения интронов. Теория раннего происхождения интронов утверждает, что уже на самых ранних этапах эволюции жизни белок-кодирующие последовательности содержали многочисленные интроны, и что эти интроны играли ключевую роль в создании новых белков путем рекомбинации последовательностей, кодирующих отдельные маленькие белковые или пептидные элементы. Теория позднего происхождения интронов гласит, что интроны образовались только в эукариотах и непрерывно накапливались в ходе эволюции эукариот. Недавние открытия показали несостоятельность теории ранних интронов (по крайней мере, в ее начальной формулировке) [24], хотя, по всей видимости, самосплайсирующие интроны существовали уже на ранних стадиях, и, вероятно, именно они стали предками сплайсосомных интронов в эукариотах [4]. Теория поздних интронов стала профилирующей, однако также претерпела некоторые изменения, так как было показано, что общий предок эукариот содержал значительное число интронов, и со временем происходило не только накопление интронов, но и их потеря [4].

Rogozin et al. [4] сформулировали синтетическую гипотезу, в основе которой лежит предположение, что заселение древнего эукариоти-

ческого генома интронами было запущено митохондриальным эндосимбиозом. Согласно этой гипотезе, альфа-протеобактериальный предок митохондрии содержал много интронов группы II (что согласовывается с данными о перепредставленности таких интронов в некоторых альфа-протеобактериях), которые мигрировали в ДНК хозяина, где стали сплайсосомными интронами. Эта гипотеза подтверждается данными реконструкции последнего общего эукариотического предка (LECA), которые показали, что ранние эукариотические геномы на 80% состояли из последовательностей, полученных из интронов группы II [15, 25]. Помимо этого, результаты реконструкции LECA показали, что плотность интронов в LECA сравнима с плотностью интронов в современных умеренно интрон-богатых геномах (предки, как минимум, 3 из 5 таксономических супергрупп — Chromalveolata, Plantae и Unikonta (животные, растения и амебоподобные) — вероятнее всего, имели интрон-богатые геномы, хотя геномных данных об оставшихся супергруппах недостаточно для каких-либо выводов). По всей вероятности, закреплению интронов в геноме должен был способствовать эффект бутылочного горлышка — сокращения генофонда популяции (т.е. ее генетического разнообразия) вследствие прохождения периода, во время которого по различным причинам происходит критическое уменьшение ее численности, в дальнейшем восстановленное. Бутылочное горлышко, в свою очередь, могло быть вызвано симбиозом пре-эукариотической клетки с альфа-протеобактерией [25]. Реконструкция LECA позволила сделать авторам вывод о том, что эволюция эукариотических генов в основном шла по пути потери интронов, значительное приобретение интронов происходило лишь на некоторых этапах разветвления эволюционного дерева. При этом силы отрицательного отбора для потери большинства интронов достаточно только для популяций больших размеров, таких как одноклеточные ($N_e, \sim 10^7-10^8$), а не для высших растений и позвоночных животных ($N_e, \sim 10^5-10^6$) и ($N_e, \sim 10^4-10^5$ соответственно) [26], поэтому первые потеряли почти все интроны, в то время как у вторых большинство интронов все же сохранилось.

ЭВОЛЮЦИЯ ИНТРОНОВ В РАМКАХ ЭКЗОН-ИНТРОННОЙ СТРУКТУРЫ

Позиция интрона — место между конкретными экзонами, в которое встраивается интрон, — также называется сайтом протосплайсинга. Различные эволюционные анализы выя-

вили, что позиция многих интронов является консервативной: в ортологичных генах животных, грибов и растений позиции интронов совпадают в 25–30% случаев [27]. Консервативность позиций интронов говорит о том, что либо сами интроны являются ортологичными, либо интроны любят встраиваться конкретно в этот сайт протосплайсинга. Вторым вариантом называется параллельным приобретением интронов (ППИ) и обычно наблюдается у эволюционно далеких организмов (например, между растениями и животными доля ППИ может достигать до 20% от всех позиций интронов) [28]. Надо отметить, что в 2016 г. вышла работа Roy [29], в которой опровергались ранее показанные случаи ППИ в геномах *Daphnia* и поддерживалась гипотеза, что одинаковая позиция интронов в ортологичных генах означает, что данные интроны произошли от общего предка. Так или иначе, значительное большинство интронов, находящихся в одной и той же позиции ортологичных генов, по-видимому, действительно являются наследственными.

Но несмотря на консервативность позиций, плотность интронов в эукариотах сильно варьируется, и расположение интронов в ортологичных генах не всегда совпадает даже в близких видах [30]. Различные реконструкции ЭИС привели к выводам, что в процессе эволюции генов и интронов происходили как значительные потери, так и значительные приобретения интронов (последние тяжело детектировать), но если процесс потери интронов преобладает на коротких эволюционных расстояниях, то массовые вставки интронов, скорее всего, происходили во время больших эволюционных событий (например, при возникновении царства Животных).

Тем не менее необходимо учитывать, что полученные выводы были сделаны по результатам реконструкции эволюции ЭИС высоко-консервативных генов, так как для других областей генома реконструкцию делать сложнее, и результаты часто получаются не очень точными. А в консервативных белок-кодирующих последовательностях приобретение интронов происходит редко, в то время как для слабо-консервативных белок-кодирующих последовательностей и генов, образованных из мобильных элементов и нетранслируемых регионов, приобретение интронов является довольно частым явлением [31].

Исследования генов-паралогов показали значительное численное превосходство ново-приобретенных интронов над потерянными, что приводит к предположению, что дупликация генов усилила вставку интронов. Сравнение ЭИС у ранних эукариотических генов-паралогов вы-

явило, что позиции интронов в этих генах не консервативны, что, по всей видимости, говорит о том, что эти гены образовались в начале эволюции эукариот, когда происходили интенсивная дупликация генов и активное заселение генов самосплайсирующимися интронами группы II [32].

Молекулярные механизмы приобретения и потери интронов. Как было упомянуто выше, в эволюции эукариотических организмов потери интронов происходили чаще, чем их приобретения [28]. Молекулярные механизмы этих двух основных эволюционных событий, связанных с интронами, остаются не до конца изученными, хотя в настоящий момент уже существуют несколько частично подтвержденных моделей. Основным предложенным механизмом потери интрона считается гомологичная рекомбинация между кДНК, полученными в итоге обратной транскрипции, и геномными копиями соответствующих генов, в результате чего образуется новая аллель без интрона. Этот механизм называют ретротранспозицией или RTMIL (*Reverse Transcriptase-Mediated Intron Loss* – в буквальном переводе потеря интрона через обратную транскрипцию). Экспериментально потеря интрона была показана в геноме дрожжей еще в 1991 году [33].

Среди моделей для приобретения интрона выделяют вставку интрона во время восстановления разрывов двухцепочечной ДНК, вставку интрона группы II, вставку интрон-подобного мобильного элемента, интронизацию (часть экзона становится интроном), перенос интрона из гена-паралога, а также транспозицию интронов – модель с обратным сплайсингом, где интрон, не успевший полностью покинуть сплайсосома, включается в мРНК, не содержащую интронов (при этом данный интрон может быть из другой пре-мРНК). Последние две модели также предполагают гомологическую рекомбинацию, как один из этапов. Lee и Stevens [34] удалось экспериментально подтвердить два случая приобретения интрона через транспозицию интронов в генах дрожжей (до этого события приобретения интрона были только предсказаны с помощью филогенетического анализа). При этом, согласно исследованиям Sverdlov et al. [35] и Tarrío et al. [36], транспозиция интронов не является единственным механизмом вставки интрона для других организмов. Следует отметить, что механизмы массовых вставок интронов (например, у самых ранних эукариот) вполне могут отличаться от механизмов относительно медленного процесса приобретения интронов у ныне существующих эукариот, что усложняет определение точных механизмов вставок.

Интересно, но интронные потери являются наиболее частыми вариациями количества копий генов (CNV), кодирующих белки у человека: 12 986 интронных делеций были обнаружены в 4147 генах (включая 1154 основных генов и 1638 генов, связанных с болезнью) [37]. В итоге можно наблюдать чрезвычайно вариабельность популяции по размеру, где у 40 генов было найдено от 10 до 150 аллелей разной длины. Интронные потери нередки у эволюционно древних генов, которые высоко консервативны на уровне белковой последовательности. Судя по всему, изменчивость экзон-интронной структуры генов, кодирующих белки, вносит важный вклад в изменчивость экспрессии и сплайсинга генов в популяциях человека [37].

Слайдинг интронов. Помимо потери и приобретения существует еще такое событие эволюции интронов, как смещение (слайдинг) – фактически это перемещение интрон-экзонных границ на небольшие расстояния (1–60 п.н.). Существует предположение, что слайдинг возникает посредством альтернативного сплайсинга: в гене появляется новый сайт сплайсинга рядом со старым, что приводит к АС, а затем по каким-то причинам старый сайт сплайсинга исчезает [36].

Вследствие крайней редкости слайдинг интронов долгое время не рассматривался как отдельный объект изучения, а шел в дополнение к другим эволюционным событиям. Например, Yenerall et al. [38] в процессе изучения возможных механизмов приобретения и потери интронов среди 11 видов *Drosophila* обнаружили 189 случаев приобретения интрона, 297 случаев потери и 1 случай слайдинга. Lehmann et al. [39] выдвинули предположение, что некоторые новые позиции интронов в консервативных генах *Drosophila* появились благодаря слайдингу интронов или дупликации экзонов. Возможные случаи слайдинга были также описаны в работах Krauss et al. [40] и Поверенной и соавт. [41]. Поверенная и соавт. [41] также подняли вопрос о связи слайдинга интронов с присутствием неканонических сайтов сплайсинга.

Надо отметить, что само существование слайдинга интрона, как явления, долго было предметом многочисленных споров. Впервые слайдинг постулировался сторонниками гипотезы о раннем происхождении интронов для объяснения удивительного открытия: позиции ортологичных интронов иногда варьируются у ортологичных генов [42]. В то же время сторонники гипотезы о позднем происхождении интронов утверждали, что слайдинг интронов, если и происходит, то мало влияет на разнообразие позиций интронов [43]. Более того, Восо и Csűgös [44] оспаривали ранее показанные случаи слай-

динга, найденные в оомицетах, и заключали, что почти всегда слайдинг интрона является всего лишь артефактом аннотации. Но, согласно результатам статистического анализа позиций интрона в разных семействах генов по методу Монте-Карло, смещение интрона на одну позицию является достоверным эволюционным событием, даже несмотря на свою относительную редкость (встречается менее чем у 5% всех интронов). А вот доказательств о слайдинге интронов на большие расстояния получено не было [45].

Tarrío et al. [36] выдвинули предположение, что позиционное разнообразие интронов обусловлено двумя пересекающимися процессами: фоновым процессом непрерывного перемещения интронов путем слайдинга и всплесками обширного приобретения или потери новых интронных последовательностей. Таким образом, слайдинг является важным источником возникновения новых интронов (по крайней мере в некоторых филогенетических линиях), что подтверждается сравнительным анализом близкорасположенных интронов из 12 геномов рода *Drosophila*, где было показано, что слайдинг интронов является относительно частой причиной новых позиций интронов в данном виде [39].

Возникновение и фиксация случаев слайдинга в процессе эволюции, вероятно, происходит через АС [36]. Согласно данной гипотезе, слайдинг интрона часто встречается в низкоконсервативных регионах белок-кодирующих генов, где повышена частота АС (например, 5'- и 3'-областей многих генов), но редко в консервативных участках белок-кодирующих генов. Для слайдинга интрона на 1 нуклеотид Fekete et al. [46] предложили возможный молекулярный механизм, связанный с участием альтернативно сплайсированного промежуточного соединения: так называемого «stwintron» (в буквальном переводе, сплайсосомные близнецы-интроны) – структуры, в которой один интрон фактически представляет собой два вложенных друг в друга интрона. Пример сдвига интрона на 1 нуклеотид был описан для гена глобина в геномах амфибий и рыб [47].

Фактическое влияние слайдинга на эволюцию эукариотических генов будет точно определено только тогда, когда будут доступны многочисленные наборы близкородственных геномов и разработаны строгие методы статистического анализа.

СВОЙСТВА ИНТРОНОВ

Фаза интрона. Одной из важных характеристик интрона является его фаза – положение

интрона относительно рамки считывания. Как известно, интрон, расположенный в кодирующей последовательности, может прерывать рамку считывания гена в любой позиции: между двумя кодонами, между первым и вторым нуклеотидами одного кодона и между вторым и третьим нуклеотидом кодона. В первом случае, если интрон расположен между двумя кодонами, ему присваивается фаза 0. Интроны, расположенные между первым и вторым или вторым и третьим нуклеотидами, имеют фазу 1 или 2 соответственно.

В ходе целого ряда научных работ, посвященных распределению фаз интронов, выяснилось, что почти во всех изученных организмах наблюдается количественное преобладание интронов фазы 0 над интронами других фаз (исключение – оболочник *Oikopleura* с равномерным распределением интронов разных фаз); более того, если фазу 0 имеют ~50% интронов, то фазу 1 и фазу 2 – 30 и 20% соответственно [42, 48]. Описанная закономерность получила название «правило 50/30/20» или «соотношение 5 : 3 : 2» и объяснялась процессом «перетасовки» экзонов [49], корреляцией интронов фазы 1 и 2 с консервативными участками последовательности [50], и даже особой структурой белка [51]. В рамках теории о раннем происхождении интронов преобладание интронов фазы 0 объяснялось тем, что часть существующих в настоящий момент интронов была представлена еще в последнем общем предке эукариот и прокариот и соединяла близлежащие мини-гены, которые позднее превратились в экзоны. Таким образом, данные интроны должны были иметь нулевую фазу. Однако данная теория никак не объясняла факт большей частоты интронов фазы 1 над интронами фазы 2. А вот, согласно теории поздних интронов, неравномерность распределения фаз интронов отражает неслучайность приобретения интронов в ходе эволюции [52]. Для некоторых видов было показано преобладание сайтов протосплайсинга в фазе 0, тем не менее этим нельзя полностью объяснить преобладание интронов фазы 0, т.к. предполагается, что интроны встраиваются в случайные сайты протосплайсинга.

Еще одно возможное объяснение «правила 50/30/20» было дано Long и Deutsch [53] в 1999 году. Они предложили использовать фазы сплайсосомных интронов для изучения эволюции экзон-интронной организации. Авторы обнаружили сильную корреляцию между распределением фаз интронов и консервативностью последовательности сигналов сплайсинга в близлежащих экзонах: наиболее представленная фаза 0 соответствовала высокому уровню консервативности, в то время как фаза 2, характер-

ная для меньшей части интронов, обычно встречалась в случаях с низким уровнем консервативности. Интроны фазы 1 занимали промежуточное положение. С учетом экспериментально показанного пагубного влияния мутаций в экзонах недалеко от сайтов сплайсинга авторами был предложен возможный генетический механизм эволюции экзон-интронной структуры, предполагающий, что недопредставленность интронов фазы 2 может являться результатом направленной потери интрона, вызванной наличием вредных мутаций [53].

Помимо количественного распределения фаз интронов было обнаружено, что интроны фазы 1 чаще всего встречаются в 5'-концевых участках гена, и их частота существенно снижается к 3'-концу. Строго обратная ситуация была показана для интронов фазы 0 [54].

Астахова и соавт. [55] показали нарушение «золотого правила» распределения фаз в генах с большим содержанием длинных интронов: такие интроны предпочитают иметь фазу 1, а не обычно перепредставленную фазу 0. Надо отметить, что перепредставленность фазы 0 также часто объясняется тем, что интроны фазы 0 создают большую гибкость для процесса кодирования белка, поскольку они не затрагивают концевые нуклеотиды экзонов, где обычно расположены сайты сплайсинга [56].

Длина интрона. Еще одним немаловажным свойством интрона является его длина. Согласно данным Halligan и Keightley [57], последовательности длиной <40 п.н. не могут сформировать полноценный интрон. Средняя длина интрона варьирует в разных организмах, но распределение длин интронов в интрон-богатых организмах имеет правый тяжелый хвост, показывающий наличие очень длинных интронов [55].

Одним из наиболее известных фактов про длины интронов является повышенная длина 1-го интрона [58]. Farlow et al. [59] и Shepard et al. [60] анализировали время протекания сплайсинга в зависимости от длины интрона. Особенности сплайсинга длинных интронов были описаны Sibley et al. [61]. Vinogradov [62] показал, что интроны в конститутивных генах короче, чем в тканеспецифичных. Также было показано, что альтернативные кассетные экзоны фланкируются более длинными, чем конститутивные, интронами; хотя это не наблюдается в случаях альтернативных акцепторного или донорного сайтов сплайсинга [23]. В исследовании Астахова и соавт. [55] была найдена корреляция между длинами соседних интронов в генах животных. Seoighe и Korir [63] показали консервативность суммарной длины интронов в генах эмбрионального развития у млекопитающих.

Длина интрона может оказывать влияние как на регуляцию экспрессии гена, так и на скорость эволюции данного гена. Щепеткова и Гельфанд [64] исследовали корреляцию между сигналами сплайсинга и длинами прилегающих экзонов и интронов. Было показано, что информационное содержание сайтов сплайсинга в генах *Caenorabditis elegans* и *Drosophila melanogaster* зависит от длины прилегающего интрона. Marais et al. [65] и Wu и Hurst [66] исследовали негативную корреляцию между длиной интрона и строгостью эволюционного отбора на уровне аминокислотных последовательностей. Существование негативной корреляции между внутривидовой дивергенцией и длиной интрона у *Drosophila melanogaster* и *Drosophila simulans* было показано Haddrill et al. [67], подтвердившими, что длинные интроны эволюционируют медленнее, чем короткие. Коэволюцию длины интронов можно наблюдать в коэкспрессирующихся генах и в генах, чьи продукты образуют единый белковый комплекс. Наиболее наглядно это можно увидеть в генах, участвующих в клеточном цикле [68].

По утверждению Pozzoli et al. [69] длина интрона и межгенных участков, сама по себе, не играет адаптивной роли в регуляции экспрессии генов, а отражает необходимость сохранения консервативных интронных элементов. Хотя количество некодирующих функциональных элементов может объяснить изменение размера генома, было показано, что на высоко экспрессированные интроны воздействует дополнительное давление, снижающее стоимость их транскрипции. С другой стороны, факт того, что в высоко экспрессируемых генах находятся маленькие интроны, подтверждает гипотезу о транскрипционной эффективности и/или экономии [70, 71]. Косвенным свидетельством может являться исследование [72], в котором было показано, что рептилии и птицы имеют более короткие интроны, чем млекопитающие, и что длина интрона у нелетающих птиц длиннее, чем у летающих, которые тратят большое количество энергии на полет.

Связь между длиной интрона и GC-составом впервые была упомянута Duret et al. [73], проанализировавшими несколько геномов позвоночных: было замечено, что в GC-богатых генах интроны короче. Это наблюдение также подтверждается Versteeg et al. [74]. В то же время в геномах костистых рыб не обнаружили никакой линейной корреляции между длиной интрона и его GC составом [75].

Интересная теория о механизме эволюции длины интрона была предложена Zhang et al. [76]. Авторы показали, что интроны могут ока-

зывать влияние на экспрессию и регуляцию гена посредством взаимодействия с соответствующими последовательностями мРНК. При этом интрон состоит из, как минимум, двух субъединиц (downstream и upstream части), соединенных между собой линкерами. Длина субъединицы довольно консервативна — ~60 п.н., а вот длина линкера может сильно варьировать. Длина всего интрона увеличивается не постепенно, а путем добавления новых структурных субъединиц и линкеров между ними. При этом 5'-область интрона или первая структурная субъединица — это уже зрелая последовательность, в то время как 3'-область — незрелый, т.е. еще эволюционирующий регион. Таким образом, эволюция длины интрона идет от 5' к 3'-концу, создавая новые субъединицы, посредством постепенного умножения фрагментов 3'-области. Интересно, что распределение длины и степень совпадения оптимальных фрагментов согласуются с особенностями последовательностей микроРНК и малой интерферирующей РНК. Эти результаты указывают на то, что взаимодействие между интронами и последовательностями мРНК является разновидностью функционального взаимодействия РНК–РНК [77]. В то же время интроны и мРНК, по-видимому, вместе коэволюционируют для успешного выполнения своих биологических функций.

ФУНКЦИИ ИНТРОНОВ

Наличие интронов в геноме является энергетическим бременем для многих клеток, ведь для транскрипции интронов и их дальнейшего вырезания из пре-мРНК с помощью сложных сплайсосомных механизмов клеткам может потребоваться много энергии. Сохранение интронов в ходе эволюции может быть объяснено только при условии, что преимущества наличия интронов превосходят их негативное влияние на энергетические затраты. Однако, согласно исследованию Lynch [78], интроны — это просто эгоистичные ДНК, заселившие белок-кодирующие гены в эукариотических геномах и сумевшие сохраниться из-за эффекта бутылочного горлышка в популяциях. Тем не менее интроны все же могут давать некоторые преимущества в качестве мутационного буфера в эукариотических геномах, защищая кодирующие последовательности от воздействия случайно возникающих вредных мутаций. Так как интроны в среднем гораздо длиннее экзонов, большинство случайно встречающихся мутаций в генах попадают в интроны и не влияют на последовательности и функции белка. Надо признать, что остает-

ся неясным, насколько буферный эффект последовательностей интрона более выигрышен с точки зрения эволюционного преимущества над энергетическим бременем.

В последнее время во многих исследованиях описываются свидетельства преимуществ, приносимых интронами в эукариотические клетки, для оправдания всех энергетических затрат. В обзоре Jo и Choi [79] функциональные роли интронов разделены на две категории: прямые (такие как регуляция АС, усиление экспрессии генов, контроль транспорта мРНК или сборки хроматина) и непрямые (например, наличие в интронах связанных с признаком однонуклеотидных полиморфизмов или разных генов некодирующих функциональных РНК; изменение длины интрона в вопросе эффективности естественного отбора).

Основную функцию интронов часто связывают с возникновением АС: хотя интроны не вносят непосредственного вклада в протеом, одно их присутствие позволяет увеличить потенциально возможное количество продуктов белок-кодирующих генов благодаря альтернативному сплайсингу.

У многих эукариот, включая млекопитающих, растения, дрожжи и насекомых, интроны могут влиять на экспрессию генов даже при отсутствии сайта связывания для транскрипционных факторов. Этот феномен был назван «интрон-опосредованным усилением экспрессии» [80]. Само присутствие интронов влияет не только на скорость транскрипции, но также на стабильность мРНК и на ее экспорт из ядра. Более того, интроны также могут увеличить эффективность трансляции мРНК [80]. Например, Lim et al. [81] предсказали эффективность трансляции по наличию экзон-интронной структуры в 5'-нетранслируемой области гена.

Вырезанные из пре-мРНК последовательности интронов, по сути, являются нк-РНК, которые быстро разрушаются. Однако, согласно последним исследованиям [82, 83], такие вырезанные интроны успевают быть задействованными в TOR связанной сети сигнала роста в *Saccharomyces cerevisiae*. Систематическое удаление всех известных интронов в генах почкующихся дрожжей указывает на то, что в большинстве случаев клетки с делецией интронов разрушаются при истощении питательных веществ быстрее, чем нормальные клетки. Таким образом, интроны оказываются медиаторами клеточного ответа при голодании.

Длинные интроны, особенно в 5'-области, часто обогащены регуляторными элементами и, таким образом, могут являться предметом отбора [42, 84]. Некоторые интроны содержат не

только различные регуляторные элементы, но и последовательности, участвующие в формировании структуры хроматина, хотя остается неясным, есть ли существенные отличия в таком обогащении регуляторных и структурных элементов между последовательностью интрона и другими нк-ДНК [85]. В состав многих интронов также часто входят разнообразные гены нк-РНК, особенно много генов малой ядрышковой РНК [86] и генов — предшественников микроРНК [87].

В 2013 г. Vakhrusheva et al. [88] на ортологических интронах человека, мыши и рыб, иглобрюха и фугу, рассматривали связь длины интрона, его возможных функций и эволюции. Было показано, что если у интрона есть некий консервативный или регуляторный участок, то его ортолог тоже, скорее всего, будет иметь консервативный или регуляторный участок, даже если последовательности совершенно различаются. Также интроны с сохранившимися консервативными участками имеют большую длину, чем те их ортологи, которые эти участки потеряли.

Известны случаи, когда внутри последовательности интрона находятся отдельные белок-кодирующие гены. Такие вложенные гены были найдены у многих эукариотических организмов и встречаются относительно часто: у человека известно, как минимум, 158 белок-кодирующих генов, расположенных внутри интронов [89], а у плодовой мухи количество вложенных генов составляет почти 10% от общего числа генов [89]. При этом функционально ген-хозяин и вложенный ген могут быть совсем не связаны, их экспрессионные профили также могут мало коррелировать [90]. Сравнительный анализ вложенных генов у нематод, плодовых мух и позвоночных показал, что в ходе эволюции таких генов становится все больше, и их накопление увеличивает сложность организации геномов [91].

Суммируя вышенаписанное, сплайсосомные интроны, исторически относящиеся к «мусорной» ДНК, играют важную роль в жизни многих эукариотических клеток, и их закрепление в генах в ходе эволюции больше не вызывает удивления. Эволюция сплайсосомных интронов неразрывно связана с эволюцией экзон-

интронной структуры эукариотических генов, которая была предметом длительных и интенсивных научных споров. Наиболее актуальная теория происхождения сплайсосомных интронов предполагает, что они образовались из самосплайсирующихся интронов группы II, содержащихся в альфа-протеобактериальном предке митохондрии при заселении древнего предка эукариотических организмов, и с тех пор происходит как постепенная потеря интронов, так и их появление. При этом молекулярные механизмы потери и приобретения интронов могут отличаться у разных организмов. Третье эволюционное событие — слайдинг интронов — случается столь редко, что все еще есть сомнения в его достоверности.

Несмотря на то, что последовательность интрона крайне изменчива в связи со слабым действием отрицательного отбора, такие его характеристики, как фаза и длина, могут обладать неожиданной консервативностью и, в случае длин, могут даже влиять на регуляцию транскрипции.

Хотя основной функцией интронов считается создание потенциала для появления новых вариантов белка, интроны также могут влиять на скорость и эффективность экспрессии гена. Присутствие интронов в генах значительно усложняет не только структуру генома, но и регуляцию связанных с ним процессов, что в итоге создает большую гибкость для адаптации организма и дает достаточно преимуществ над всеми энергическими затратами, потраченными на содержание интронов и их сохранение в ходе эволюции.

Финансирование. Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (грант № 18-34-00932).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических норм. Настоящая статья не содержит описания выполненных авторами исследований с использованием людей и животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Chow, L. T., Gelinis, R. E., Broker, T. R., and Roberts, R. J. (1977) An amazing sequence arrangement at the 5'-ends of adenovirus 2 messenger RNA, *Cell*, **12**, 1-8, doi: 10.1016/0092-8674(77)90180-5.
2. Berget, S. M., Moore, C., and Sharp, P. A. (1977) Spliced segments at the 5'-terminus of adenovirus 2 late mRNA, *Proc. Natl. Acad. Sci. USA*, **74**, 3171-3175.
3. Gilbert, W. (1978) Why genes in pieces, *Nature*, **271**, 501, doi: 10.1038/271501a0.
4. Rogozin, I. B., Carmel, L., Csuros, M., and Koonin, E. V. (2012) Origin and evolution of spliceosomal introns, *Biol. Direct.*, **7**, 11, doi: 10.1186/1745-6150-7-11.
5. Luo, M. J., and Reed, R. (1999) Splicing is required for rapid and efficient mRNA export in metazoans, *Proc. Natl. Acad. Sci. USA*, **96**, 14937-14942.
6. Patel, A. A., and Steitz, J. A. (2003) Splicing double: insights from the second spliceosome, *Nat. Rev. Mol. Cell Biol.*, **4**, 960-970, doi: 10.1038/nrm1259.

7. Patel, A. A., McCarthy, M., and Steitz, J. A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression, *EMBO J.*, **21**, 3804-3815, doi: 10.1093/emboj/cdf297.
8. Chorev, M., and Carmel, L. (2012) The function of introns, *Front. Genet.*, **3**, 1-15, doi: 10.3389/fgene.2012.00055.
9. Fedorova, L., and Fedorov, A. (2003) Introns in gene evolution, *Genetica*, **118**, 123-131.
10. Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.-F., Wellinger, R. J., Chabot, B., and Abou Elela, S. (2011) Introns within ribosomal protein genes regulate the production and function of yeast ribosomes, *Cell*, **147**, 320-331, doi: 10.1016/j.cell.2011.08.044.
11. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinformatics*, **3**, 2.
12. Lane, C. E., van den Heuvel, K., Kozera, C., Curtis, B. A., Parsons, B. J., Bowman, S., and Archibald, J. M. (2007) Nucleomorph genome of *Hemiselms andersenii* reveals complete intron loss and compaction as a driver of protein structure and function, *Proc. Natl. Acad. Sci. USA*, **104**, 19908-19913, doi: 10.1073/pnas.0707419104.
13. Rogers, J. H. (1989) How were introns inserted into nuclear genes? *Trends Genet.*, **5**, 213-216.
14. Cavalier-Smith, T. (1991) Intron phylogeny: a new hypothesis, *Trends Genet.*, **7**, 145-148.
15. Csuros, M., Rogozin, I. B., and Koonin, E. V. (2011) A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes, *PLoS Comput. Biol.*, **7**, 1-9, doi: 10.1371/journal.pcbi.1002150.
16. Sakurai, A., Fujimori, S., Kochiwa, H., Kitamura-Abe, S., Washio, T., Saito, R., Carninci, P., Hayashizaki, Y., and Tomita, M. (2002) On biased distribution of introns in various eukaryotes, *Gene*, **300**, 89-95.
17. Ponjavic, J., Ponting, C. P., and Lunter, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs, *Genome Res.*, **17**, 556-565, doi: 10.1101/gr.6036807.
18. Parada, G. E., Munita, R., Cerda, C. A., and Gysling, K. (2014) A comprehensive survey of non-canonical splice sites in the human transcriptome, *Nucleic Acids Res.*, **42**, 16.
19. Pucker, B., and Brockington, S. F. (2018) Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes, *BMC Genomics*, **19**, 980, doi: 10.1186/s12864-018-5360-z.
20. Alioto, T. S. (2007) U12DB: a database of orthologous U12-type spliceosomal introns, *Nucleic Acids Res.*, **35**, D110-D115, doi: 10.1093/nar/gkl796.
21. Schellenberg, M. J., Ritchie, D. B., and MacMillan, A. M. (2008) Pre-mRNA splicing: a complex picture in higher definition, *Trends Biochem. Sci.*, **33**, 243-246, doi: 10.1016/j.tibs.2008.04.004.
22. Sverdlov, A. V., Rogozin, I. B., Babenko, V. N., and Koonin, E. V. (2003) Evidence of splice signal migration from exon to intron during intron evolution, *Curr. Biol.*, **13**, 2170-2174, doi: 10.1016/j.cub.2003.12.003.
23. Kim, E., Magen, A., and Ast, G. (2007) Different levels of alternative splicing among eukaryotes, *Nucleic Acids Res.*, **35**, 125-131, doi: 10.1093/nar/gkl924.
24. Nguyen, H. D., Yoshihama, M., and Kenmochi, N. (2006) Phase distribution of spliceosomal introns: implications for intron origin, *BMC Evol. Biol.*, **6**, 69, doi: 10.1186/1471-2148-6-69.
25. Koonin, E. V. (2009) Intron-dominated genomes of early ancestors of eukaryotes, *J. Hered.*, **100**, 618-623, doi: 10.1093/jhered/esp056.
26. Charlesworth, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation, *Nat. Rev. Genet.*, **10**, 195-205, doi: 10.1038/nrg2526.
27. Fedorov, A., Merican, A. F., and Gilbert, W. (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes, *Proc. Natl. Acad. Sci. USA*, **99**, 16128-16133, doi: 10.1073/pnas.242624899.
28. Carmel, L., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2007) Patterns of intron gain and conservation in eukaryotic genes, *BMC Evol. Biol.*, **7**, 192, doi: 10.1186/1471-2148-7-192.
29. Roy, S. W. (2016) How common is parallel intron gain? rapid evolution versus independent creation in recently created introns in *Daphnia*, *Mol. Biol. Evol.*, **33**, 1902-1906, doi: 10.1093/molbev/msw091.
30. Mourier, T., and Jeffares, D. C. (2003) Eukaryotic intron loss, *Science*, **300**, 1393, doi: 10.1126/science.1080559.
31. Hong, X., Scofield, D. G., and Lynch, M. (2016) Intron size, abundance, and distribution within untranslated regions of genes, *Mol. Biol. Evol.*, **23**, 2392-2404, doi: 10.1093/molbev/msl111.
32. Sverdlov, A. V., Csuros, M., Rogozin, I. B., and Koonin, E. V. (2007) A glimpse of a putative pre-intron phase of eukaryotic evolution, *Trends Genet.*, **23**, 105-108, doi: 10.1016/j.tig.2007.01.001.
33. Derr, L. K., Strathern, J. N., and Garfinkel, D. J. (1991) RNA-mediated recombination in *S. cerevisiae*, *Cell*, **67**, 355-364, doi: 10.1016/0092-8674(91)90187-4.
34. Lee, S., and Stevens, S. W. (2016) Spliceosomal intronogenesis, *Proc. Natl. Acad. Sci. USA*, **113**, 6514-6519, doi: 10.1073/pnas.1605113113.
35. Sverdlov, A. V., Babenko, V. N., Rogozin, I. B., and Koonin, E. V. (2004) Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion, *Gene*, **338**, 85-91, doi: 10.1016/j.gene.2004.05.027.
36. Tarrío, R., Ayala, F. J., and Rodríguez-Trelles, F. (2008) Alternative splicing: a missing piece in the puzzle of intron gain, *Proc. Natl. Acad. Sci. USA*, **105**, 7223-7228, doi: 10.1073/pnas.0802941105.
37. Rigau, M., Juan, D., Valencia, A., and Rico, D. (2019) Intronic CNVs and gene expression variation in human populations, *PLoS Genet.*, **15**, e1007902, doi: 10.1371/journal.pgen.1007902.
38. Yenerall, P., Krupa, B., and Zhou, L. (2011) Mechanisms of intron gain and loss in *Drosophila*, *BMC Evol. Biol.*, **11**, 364, doi: 10.1186/1471-2148-11-364.
39. Lehmann, J., Eisenhardt, C., Stadler, P. F., and Krauss, V. (2010) Some novel intron positions in conserved *Drosophila* genes are caused by intron sliding or tandem duplication, *BMC Evol. Biol.*, **10**, 156, doi: 10.1186/1471-2148-10-156.
40. Krauss, V., Pecyna, M., Kurz, K., and Sass, H. (2005) Phylogenetic mapping of intron positions: a case study of translation initiation factor eIF2γ, *Mol. Biol. Evol.*, **22**, 74-84, doi: 10.1093/molbev/msh255.
41. Поверенная И. В., Горев Д. Д., Астахова Т. В., Цитович И. И., Яковлев В. В., Ройтберг, М. А. (2017) Слайдинг и вариабельность длины интронов для генов, обогащенных длинными интронами фазы I, *Матем. биология и биоинформ.*, **12**, 302-316, doi: 10.17537/2017.12.302.
42. De Souza, S. J., Long, M., Klein, R. J., Roy, S., Lin, S., and Gilbert, W. (1998) Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins, *Proc. Natl. Acad. Sci. USA*, **95**, 5094-5099.

43. Stoltzfus, A., Logsdon, J. M., Palmer, J. D., and Doolittle, W. F. (1997) Intron “sliding” and the diversity of intron positions, *Proc. Natl. Acad. Sci. USA*, **94**, 10739-10744.
44. Bocco, S. S., and Csűrös, M. (2016) Splice sites seldom slide: intron evolution in oomycetes, *Genome Biol. Evol.*, **8**, 2340-2350, doi: 10.1093/gbe/evw157.
45. Rogozin, I. B., Lyons-Weiler, J., and Koonin, E. V. (2000) Intron sliding in conserved gene families, *Trends Genet.*, **16**, 430-432, doi: 10.1016/s0168-9525(00)02096-5.
46. Fekete, E., Flipphi, M., Ag, N., Kavalecz, N., Cerqueira, G., Scuzzocchio, C., and Karaffa, L. (2017) A mechanism for a single nucleotide intron shift, *Nucleic Acids Res.*, **45**, 9085-9092, doi: 10.1093/nar/gkx520.
47. Roesner, A., Fuchs, C., Hankeln, T., and Burmester, T. (2005) A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals, *Mol. Biol. Evol.*, **22**, 12-20, doi: 10.1093/molbev/msh258.
48. Long, M., de Souza, S. J., Rosenberg, C., and Gilbert, W. (1998) Relationship between “proto-splice sites” and intron phases: evidence from dicodon analysis, *Proc. Natl. Acad. Sci. USA*, **95**, 219-223.
49. Fedorov, A., Suboch, G., Bujakov, M., and Fedorova, L. (1992) Analysis of nonuniformity in intron phase distribution, *Nucleic Acids Res.*, **20**, 2553-2557.
50. Endo, T., Fedorov, A., de Souza, S. J., and Gilbert, W. (2002) Do introns favor or avoid regions of amino acid conservation? *Mol. Biol. Evol.*, **19**, 521-522, doi: 10.1093/oxfordjournals.molbev.a004107.
51. Gilbert, W., de Souza, S. J., and Long, M. (1997) Origin of genes, *Proc. Natl. Acad. Sci. USA*, **94**, 7698-7703.
52. Logsdon, J. M. (1998) The recent origins of spliceosomal introns revisited, *Curr. Opin. Genet. Dev.*, **8**, 637-648.
53. Long, M., and Deutsch, M. (1999) Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns, *Mol. Biol. Evol.*, **16**, 1528-1534.
54. Ruvinsky, A., Eskesen, S. T., Eskesen, F. N., and Hurst, L. D. (2005) Can codon usage bias explain intron phase distributions and exon symmetry? *J. Mol. Evol.*, **60**, 99-104, doi: 10.1007/s00239-004-0032-9.
55. Астахова Т. В., Ройтберг М. А., Цитович И. И., Яковлев В. В. (2014) Закономерности, связанные с распределением длин интронов, *Матем. биология и биоинформ.*, **9**, 482-490.
56. Ruvinsky, A., and Ward, W. (2008) Intron framing exonic nucleotides: a compromise between protein coding and splicing constraints, *Open Evol. J.*, **2**, 7-12, doi: 10.2174/1874404400802010007.
57. Halligan, D. L., and Keightley, P. D. (2006) Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison, *Genome Res.*, **16**, 875-884, doi: 10.1101/gr.5022906.
58. Bradnam, K. R., and Korf, I. (2008) Longer first introns are a general property of eukaryotic gene structure, *PLoS One*, **3**, e3093, doi: 10.1371/journal.pone.0003093.
59. Farlow, A., Dolezal, M., Hua, L., and Schlötterer, C. (2012) The genomic signature of splicing-coupled selection differs between long and short introns, *Mol. Biol. Evol.*, **29**, 21-24, doi: 10.1093/molbev/msr201.
60. Shepard, S., Mc Creary, M., and Fedorov, A. (2009) The peculiarities of large intron splicing in animals, *PLoS One*, **4**, e7853, doi: 10.1371/journal.pone.0007853.
61. Sibley, C. R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M. E., Hardy, J., Modic, M., Curk, T., Wilson, S. W., Plagnol, V., and Ule, J. (2015) Recursive splicing in long vertebrate genes, *Nature*, **521**, 371-375, doi: 10.1038/nature14466.
62. Vinogradov, A. E. (2006) “Genome design” model: evidence from conserved intronic sequence in human-mouse comparison, *Genome Res.*, **16**, 347-354, doi: 10.1101/gr.4318206.
63. Seoighe, C., and Korir, P. K. (2011) Evidence for intron length conservation in a set of mammalian genes associated with embryonic development, *BMC Bioinformatics*, **12**, Suppl. 9, S16, doi: 10.1186/1471-2105-12-S9-S16.
64. Шепеткова И. Л., Гельфанд М. С. (1997) Некоторые статистические особенности сайтов сплайсинга позвоночных и беспозвоночных, *Биофизика*, **42**, 82-89.
65. Marais, G., Nouvellet, P., Keightley, P. D., and Charlesworth, B. (2005) Intron size and exon evolution in *Drosophila*, *Genetics*, **170**, 481-485, doi: 10.1534/genetics.104.037333.
66. Wu, X., and Hurst, L. D. (2015) Why selection might be stronger when populations are small: intron size and density predict within and between-species usage of exonic splice associated cis-motifs, *Mol. Biol. Evol.*, **32**, 1847-1861, doi: 10.1093/molbev/msv069.
67. Haddrill, P. R., Charlesworth, B., Halligan, D. L., and Andolfatto, P. (2005) Patterns of intron sequence evolution in *Drosophila* are dependent upon length and GC content, *Genome Biol.*, **6**, R67, doi: 10.1186/gb-2005-6-8-r67.
68. Keane, P. A., and Seoighe, C. (2016) Intron length coevolution across mammalian genomes, *Mol. Biol. Evol.*, **33**, 2682-2691, doi: 10.1093/molbev/msw151.
69. Pozzoli, U., Menozzi, G., Comi, G. P., Cagliani, R., Bresolin, N., and Sironi, M. (2007) Intron size in mammals: complexity comes to terms with economy, *Trends Genet.*, **23**, 20-24, doi: 10.1016/j.tig.2006.10.003.
70. Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., and Kondrashov, F. A. (2002) Selection for short introns in highly expressed genes, *Nat. Genet.*, **31**, 415-418, doi: 10.1038/ng940.
71. Rao, Y. S., Wang, Z. F., Chai, X. W., Wu, G. Z., Zhou, M., Nie, Q. H., and Zhang, X. Q. (2010) Selection for the compactness of highly expressed genes in *Gallus gallus*, *Biol. Direct*, **5**, 35, doi: 10.1186/1745-6150-5-35.
72. Zhang, Q., and Edwards, S. V. (2012) The evolution of intron size in amniotes: a role for powered flight? *Genome Biol. Evol.*, **4**, 1033-1043, doi: 10.1093/gbe/evs070.
73. Duret, L., Mouchiroud, D., and Gautier, C. (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores, *J. Mol. Evol.*, **40**, 308-317, doi: 10.1007/bf00163235.
74. Versteeg, R., van Schaik, B. D. C., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H. J., and van Kampen, A. H. C. (2003) The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes, *Genome Res.*, **13**, 1998-2004, doi: 10.1101/gr.1649303.
75. Chaurasia, A., Tarallo, A., Bernà, L., Yagi, M., Agnisola, C., and D’Onofrio, G. (2014) Length and GC content variability of introns among teleostean genomes in the light of the metabolic rate hypothesis, *PLoS One*, **9**, e103889, doi: 10.1371/journal.pone.0103889.
76. Zhang, Q., Li, H., Zhao, X.-Q., Xue, H., Zheng, Y., Meng, H., Jia, Y., and Bo, S.-L. (2016) The evolution mechanism of intron length, *Genomics*, **108**, 47-55, doi: 10.1016/j.ygeno.2016.07.004.
77. Sathe, L., Bolinger, C., Mannan, M. A., Dever, T. E., and Dey, M. (2015) Evidence that base-pairing interaction between intron and mRNA leader sequences inhibits initiation of HAC1 mRNA translation in yeast, *J. Biol. Chem.*, **290**, 21821-21832, doi: 10.1074/jbc.M115.649335.
78. Lynch, M. (2002) Intron evolution as a population-genetic process, *Proc. Natl. Acad. Sci.*, **99**, 6118-6123, doi: 10.1073/pnas.092595699.

79. Jo, B.-S., and Choi, S. S. (2015) Introns: the functional benefits of introns in genomes, *Genomics Inform.*, **13**, 112-118, doi: 10.5808/GI.2015.13.4.112.
80. Shaul, O. (2017) How introns enhance gene expression, *Int. J. Biochem. Cell Biol.*, **91**, 145-155, doi: 10.1016/j.biocel.2017.06.016.
81. Lim, C. S., T Wardell, S. J., Kleffmann, T., and Brown, C. M. (2018) The exon-intron gene structure upstream of the initiation codon predicts translation efficiency, *Nucleic Acids Res.*, **46**, 4575-4591, doi: 10.1093/nar/gky282.
82. Parenteau, J., Maignon, L., Berthoumieux, M., Catala, M., Gagnon, V., and Abou Elela, S. (2019) Introns are mediators of cell response to starvation, *Nature*, **565**, 612-617, doi: 10.1038/s41586-018-0859-7.
83. Morgan, J. T., Fink, G. R., and Bartel, D. P. (2019) Excised linear introns regulate growth in yeast, *Nature*, **565**, 606-611, doi: 10.1038/s41586-018-0828-1.
84. Majewski, J., and Ott, J. (2002) Distribution and characterization of regulatory elements in the human genome, *Genome Res.*, **12**, 1827-1836, doi: 10.1101/gr.606402.
85. Glazko, G. V., Koonin, E. V., Rogozin, I. B., and Shabalina, S. A. (2003) A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions, *Trends Genet.*, **19**, 119-124, doi: 10.1016/S0168-9525(03)00016-7.
86. Tycowski, K. T., Shu, M. D., and Steitz, J. A. (1993) A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3, *Genes Dev.*, **7**, 1176-1190.
87. Rearick, D., Prakash, A., McSweeney, A., Shepard, S. S., Fedorova, L., and Fedorov, A. (2011) Critical association of ncRNA with introns, *Nucleic Acids Res.*, **39**, 2357-2366, doi: 10.1093/nar/gkq1080.
88. Vakhrusheva, O. A., Bazykin, G. A., and Kondrashov, A. S. (2013) Genome-level analysis of selective constraint without apparent sequence conservation, *Genome Biol. Evol.*, **5**, 532-541, doi: 10.1093/gbe/evt023.
89. Kumar, A. (2009) An overview of nested genes in eukaryotic genomes, *Eukaryot. Cell*, **8**, 1321-1329, doi: 10.1128/EC.00143-09.
90. Lee, Y. C. G., and Chang, H.-H. (2013) The evolution and functional significance of nested gene structures in *Drosophila melanogaster*, *Genome Biol. Evol.*, **5**, 1978-1985, doi: 10.1093/gbe/evt149.
91. Assis, R., Kondrashov, A. S., Koonin, E. V., and Kondrashov, F. A. (2008) Nested genes and increasing organizational complexity of metazoan genomes, *Trends Genet.*, **24**, 475-478, doi: 10.1016/j.tig.2008.08.003.

SPLICEOSOMAL INTRONS: FEATURES, FUNCTIONS, AND EVOLUTION

Review

I. V. Poverennaya^{1,2*} and M. A. Roytberg^{2,3,4}

¹ Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia; E-mail: ipoverennaya@gmail.com

² Institute of Mathematical Problems of Biology RAS — the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 142290 Puschino, Moscow Region, Russia

³ Moscow Institute of Physics and Technology (National Research University), 141701 Dolgoprudny, Moscow Region, Russia

⁴ National Research University, Higher School of Economics, 101000 Moscow, Russia

Received March 26, 2020

Revised May 25, 2020

Accepted May 25, 2020

Spliceosomal introns contained in most eukaryotic genes are non-coding sequences excised from pre-mRNA by a special complex called spliceosome during mRNA splicing. Introns occur in both protein-coding and RNA-coding genes, and even more, can be found in the coding and untranslated regions of genes. Due to a high rate of polymorphism, the intron sequences may vary greatly and intron function used to be associated only with its presence for alternative splicing. Therefore, intron evolution has been seen not as the evolution of an individual genome element, but rather as the evolution of a gene intron-exon structure. Here we review the intron origin theories: evolutionary events in the exon-intron structure such as intron gain, loss, and sliding; intron functions known to date; and how changes of such intron features as length and phase can affect the regulation of processes involving genes.

Keywords: spliceosomal introns, intron phase, intron length, evolution, sliding, exon-intron-structure