

УДК 577.214.5

«РАЗОРВАННЫЕ» ГЕНЫ И СПЛАЙСИНГ

© 2014 г. Г.М. Дымшиц^{1,2}, О.В. Саблина¹

¹ Специализированный учебно-научный центр НГУ, Новосибирск, Россия,
e-mail: ovsa07@yandex.ru;

² Новосибирский национальный исследовательский государственный университет,
Новосибирск, Россия

Поступила в редакцию 26 мая 2013 г. Принята к публикации 1 февраля 2014 г.

ВВЕДЕНИЕ

После того как была установлена и доказана генетическая роль ДНК, определения гена менялись несколько раз. Первоначальное определение «один ген – один признак» оказалось слишком расплывчатым и в общем случае неверным, так как за развитие одного признака могут отвечать несколько генов, а один ген может оказывать влияние на развитие нескольких разных признаков. Определение «один ген – один фермент» также оказалось неточным, так как очень многие ферменты – субъединичные белки и поэтому являются продуктом нескольких разных генов. А после нахождения нетранслируемых генов, таких как гены транспортных или рибосомных РНК, а также открытия альтернативного сплайсинга матричных РНК определение «один ген – одна полипептидная цепь» также устарело. Для бактерий возможно определение «ген – участок ДНК, кодирующий первичную структуру одной полипептидной цепи или одной rРНК, или одной tРНК, или одной sРНК». Для эукариот ген можно определить как *участок ДНК, по которому образуется функциональная молекула РНК (m, r, t, s)*. Пре-mРНК может подвергаться альтернативному сплайсингу и редактированию, благодаря чему возможно образование нескольких разных полипептидных цепей, закодированных в одном гене.

Практически все бактериальные гены коллинеарны белковому продукту. Подавляющее большинство генов высших эукариот «разорваны». Они состоят из чередующихся участков:

«кодирующих» – экзонов и «некодирующих» – интронов. Поскольку в ряде случаев интроны (или их части) оказывались «кодирующими», а экзоны «некодирующими», то правильнее называть экзонами те участки ДНК, копии которых составляют зрелую РНК, а интронами те внутренние районы гена, копии которых удаляются из первичного транскрипта и отсутствуют в зрелой РНК. Вырезание из пре-РНК копий интронов (в дальнейшем для простоты – *интронов*) и сшивание копий экзонов (в дальнейшем для простоты – *экзонов*) происходит на одном из этапов созревания РНК, названном *сплайсингом* (*splicing – сращивание канатов без узлов*) (рис. 1). Этот процесс был открыт в 1977 г. Ф. Шарпом и Дж. Робертсом (MIT, Массачусетский технологический институт, США) при работе с аденовирусом человека (Нобелевская премия 1993 г.). Несколько разных mРНК содержали три одинаковых экзона и четвертый, пришитый к ним в результате альтернативного сплайсинга; все они считывались с одного гена аденовируса и служили матрицами для синтеза разных белков.

Большая часть эукариотических генов, кодирующих белки, содержат хотя бы один инт-

Экзоны – участки ДНК, копии которых составляют зрелую РНК.

Интроны – участки ДНК, копии которых удаляются из первичного транскрипта и отсутствуют в зрелой РНК.

Сплайсинг – процесс вырезания из пре-РНК копий интронов и сшивание копий экзонов.

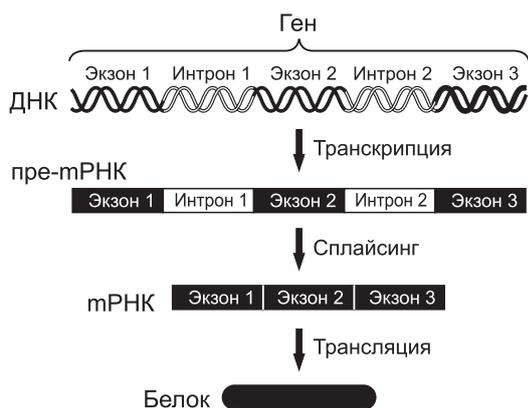


Рис. 1. При созревании мРНК интроны вырезаются, а экзоны сшиваются.

рон. У человека средний ген, имеющий размер 27×10^3 п.н., содержит 8 интронов со средней длиной 1500 п.н. Однако некоторые гены, например гены гистонов, гены отдельных трансмембранных рецепторов, интронов не содержат. У млекопитающих таких безынtronных генов 6 %, у плодовой мушки – 17 %, а у низших эукариот, таких как дрожжи, – 95 %. У низших эукариот всего несколько генов, кодирующих рибосомные и транспортные РНК, содержат по одному интрону.

«Разорванными» могут быть все классы ядерных генов: кодирующие белки, кодирующие гРНК, кодирующие тРНК. Интроны найдены также в митохондриальных генах низших эукариот, в генах хлоропластов и у некоторых бактериофагов. Большая часть генов архей содержит интроны. В последнее время выявлено несколько «разорванных» генов цианобактерий.

ХИМИЯ СПЛАЙСИНГА

У человека и других высших организмов в ядерных генах, кодирующих белки, интроны на своих 5'-концах имеют динуклеотид ГУ, на 3'-концах – АГ. На расстоянии 18–40 нуклеотидов от 3'-конца находится «точка разветвления» – адениловый нуклеотид. Он отделен от АГ полипиримидиновой последовательностью. 2'-ОН группа аденозина атакует 5'-конец интрона с образованием фосфодиэфирной связи 5'–2'. Друг от друга интроны отличаются размером участка от 5'-конца до точки разветвления (от 50 до нескольких тыс. нуклеотидов) (рис. 2).

По существу, **сплайсинг** – это две последовательные реакции трансэтерификации в РНК, проходящие автокаталитически либо при помощи ферментов, или рибонуклеопротеидных комплексов – **сплайсосом**.

Химически сплайсинг представляет собой реакцию трансэтерификации: фосфодиэфирная связь переносится из одного положения в другое. На первом этапе происходит нуклеофильная атака 5'-конца интрона 2'-ОН группой аденилового нуклеотида. На втором этапе освобожденная 3'-ОН группа экзона атакует связь между 3'-концом интрона и следующим экзonom. Две первоначальные фосфодиэфирные 3'–5'-связи на границах экзон–интрон–экзон заменяются на одну 3'–5'-связь между экзонами и одну 5'–2'-связь внутри интрона с образованием формы лассо.

АЛЬТЕРНАТИВНЫЙ СПЛАЙСИНГ

У высших организмов первичный РНК-транскрипт может подвергаться альтернативному сплайсингу. При этом в определенных случаях могут удаляться и некоторые из экзонов вместе с интронами, а некоторые интроны или их части остаются в зрелой мРНК и служат матрицей для синтеза полипептидной цепи (рис. 3). В результате по одному гену может образо-

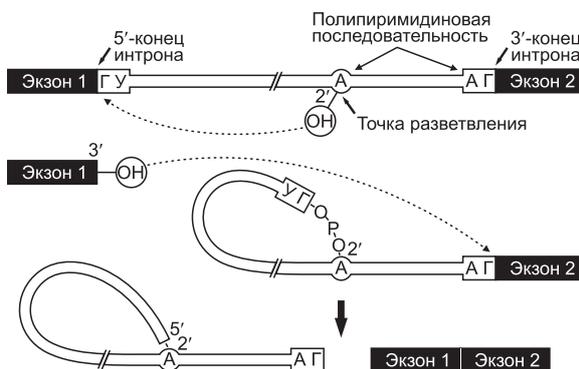


Рис. 2. На концах интронов ядерных генов имеются короткие консенсусные последовательности 5' ГУ-----//-----АГ 3'. Адениловый нуклеотид в точке разветвления своей 2'-ОН группой атакует 5'-конец интрона.

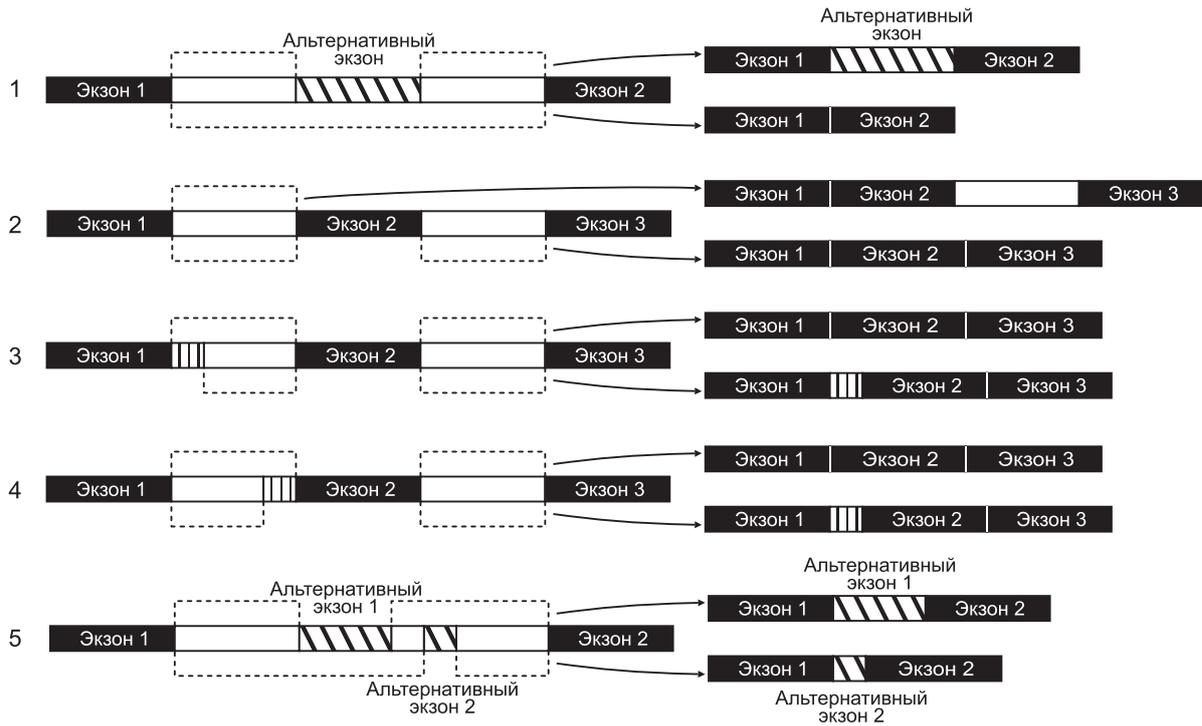


Рис. 3. Альтернативный сплайсинг.

1 – с оставлением или вырезанием экзона; 2 – с оставлением или вырезанием интрона; 3 – с распознаванием разных 5'-концов в интроне; 4 – с распознаванием разных 3'-концов в интроне; 5 – с взаимноисключающим оставлением экзонов.

ваться несколько разных мРНК, по которым синтезируются разные белки. Именно альтернативным сплайсингом объясняется тот факт, что у человека синтезируется не менее 10^5 разных типов белков, кодируемых всего $\sim 20 \times 10^3$ генами. Из всех генов человека, содержащих интроны, более 75 % служат матрицами для синтеза пре-мРНК, подвергаемых альтернативному сплайсингу.

В результате альтернативного сплайсинга по каждому гену высших эукариот может образовываться в среднем три разных мРНК, а по некоторым генам даже несколько тысяч. Так, по гену *Slo*, экспрессирующемуся в волосковых клетках внутреннего уха человека, теоретически возможно образование более 30 000 разных

мРНК (в настоящее время обнаружено около 500 из них). Ген содержит 35 экзонов, 8 из которых альтернативны. Разные формы белка *Slo* определяют различия в частоте звуков, воспринимаемых клетками.

МЕХАНИЗМ СПЛАЙСИНГА ЯДЕРНЫХ мРНК

Известно несколько разных механизмов сплайсинга. Наиболее сложным из них является сплайсинг мРНК ядерных генов высших эукариот. Он осуществляется *сплайсосомай* – рибонуклеопротеидным комплексом, включающим 5 малых ядерных РНК (U1, U2, U4, U5, U6) и более 100 белков. Некоторые белки напрямую участвуют в сплайсинге, большая их часть играет структурную роль и необходима для сборки малых ядерных рибонуклеопротеидных комплексов и их взаимодействий между собой и с пре-мРНК.

Описано несколько десятков белковых комплексов, называемых *SR-белками* (сплайсинг-

Альтернативный сплайсинг – образование нескольких разных зрелых РНК из одинаковых пре-РНК за счет исключения некоторых экзонов и/или оставления частей интронов.

регулирующими), которые различны в разных тканях, а также в одной ткани на разных этапах развития организма. Эти богатые серином и аргинином белки связываются с определенными нуклеотидными последовательностями внутри экзона и сигнализируют о вырезании или сохранении конкретного экзона в мРНК (рис. 4). Если они связываются с экзонным энхансером сплайсинга (ESE), то на 5'-конце соседнего интрона начинается формирование сплайсосомы (рис. 5). Если же другие SR-белки связываются с экзонным супрессором сплайсинга (ESS), тогда блокируется присоединение малых РНК сплайсосомы к фланкирующим интронам, что приводит к вырезанию этого экзона вместе с интронами.

В экзоне после взаимодействия SR-белков с ESE концы фланкирующих его интронов спариваются с малыми РНК. После того как 5'-конец интрона сближается с точкой разветвления, 2'-ОН группа аденилового нуклеотида оказывается в такой конформационной позиции, что с неизбежностью атакует границу экзон-интрон. Приближенный 3'-конец интрона атакуется 3'-ОН группой экзона. Сшивание экзона происходит с точностью до одного нуклеотида, что принципиально важно – при синтезе полипептидов не должно происходить сдвига рамки считывания.

ОПРЕДЕЛЕНИЕ ПОЛА У ДРОЗОФИЛЫ КАСКАДОМ АЛЬТЕРНАТИВНОГО СПЛАЙСИНГА

Первый ген в этом каскаде – *sxl* (sex-lethal). Он одинаков у самцов ($X:A = 0,5$) и самок ($X:A = 1$) и состоит из трех экзона и двух интронов (рис. 6). У самцов в зрелой мРНК содержатся все три экзона, в экзоне 2 (Э2) имеется терминирующий кодон, на котором заканчивается образование белкового продукта, не обладающего функциональной активностью. У самок же Э2 вырезается вместе с фланкирующими его интронами. По мРНК, содержащей Э1 и Э3, образуется белок SXL, который обеспечивает у самок вырезание И1 из транскрипта гена *tra* (transformer). У самцов в мРНК сохраняется часть И1, в которой имеется терминирующий кодон. В результате у самцов белок TRA не образуется, он синтезируется только у самок и вовлекается в альтернативный сплайсинг мРНК

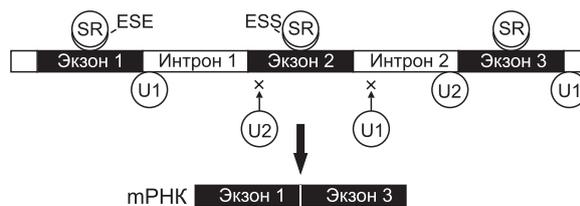


Рис. 4. Маркировка экзона, входящего в зрелую мРНК, и экзона, вырезаемого вместе с фланкирующими интронами.

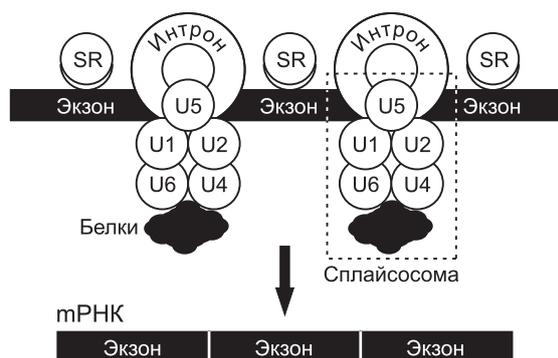


Рис. 5. Образование сплайсосомы.

гена *dsx* (double sex). TRA взаимодействует с белком SR и обеспечивает его посадку на ESE в Э2 транскрипта гена *dsx*, что приводит к оставлению Э2 в зрелой мРНК самок. У самцов Э2 вырезается вместе с фланкирующими его интронами; по Э1 и Э3 образуется «мужской» белок DSX, который репрессирует «женские» гены, в результате чего развитие идет по мужскому типу. «Женский» белок DSX репрессирует «мужские» гены, и развитие идет по женскому типу.

АВТОСПЛАЙСИНГ I ТИПА

Особый механизм сплайсинга, так называемый *автосплайсинг*, был открыт Томасом Чекком (США) в 1982 г. Он работал с инфузорией *Tetrachylena thermophyla*. У нее образуется 35S пре-гРНК длиной 6400 нуклеотидов. Из

Автосплайсинг – процесс вырезания интронов из пре-РНК без участия белков. Каталитической активностью обладают особые третичные структуры самих интронов.

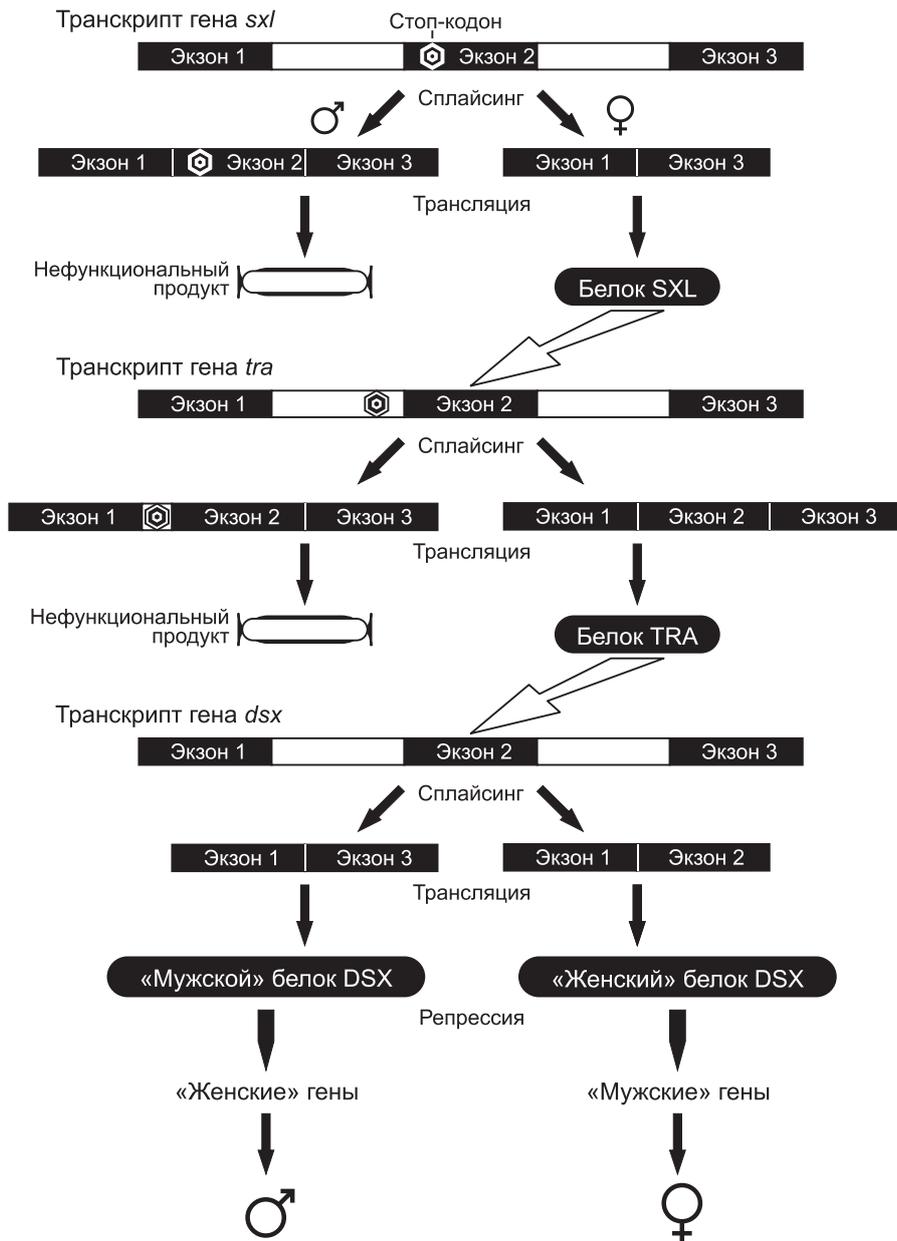


Рис. 6. Развитие самцов или самок дрозофилы определяется каскадом альтернативного сплайсинга.

этой пре-гРНК вырезается внутренний участок длиной 414 нуклеотидов, два экзона сшиваются с образованием 26S гРНК (рис. 7). Оказалось, что это происходит без какого-либо участия белков. Эта пре-гРНК обладает особой третичной структурой, с чем и связана ее каталитическая активность.

Вырезание интрона и сшивание экзонов осуществляются в ходе реакций трансэтерификации. В самом интроне имеется несколько «шпилек», образующих третичную структуру – «карман», в который попадают ГТФ, ГДФ, ГМФ

или гуанозин. 3'-гидроксильная группа гуанозина атакует 5'-конец интрона, что приводит к нуклеофильному расщеплению молекулы РНК. Освобожденная 3'-ОН группа первого экзона реагирует с фосфатной группой на 3'-границе интрона и второго экзона. Экзоны сшиваются, а вырезанный линейный интрон претерпевает две последовательные реакции внутримолекулярной трансэтерификации, которые сопровождаются гидролизом, приводящим к отщеплению сначала 15, а затем 4 нуклеотидов. Образование L-19IVS (linear intervening sequence – линейная

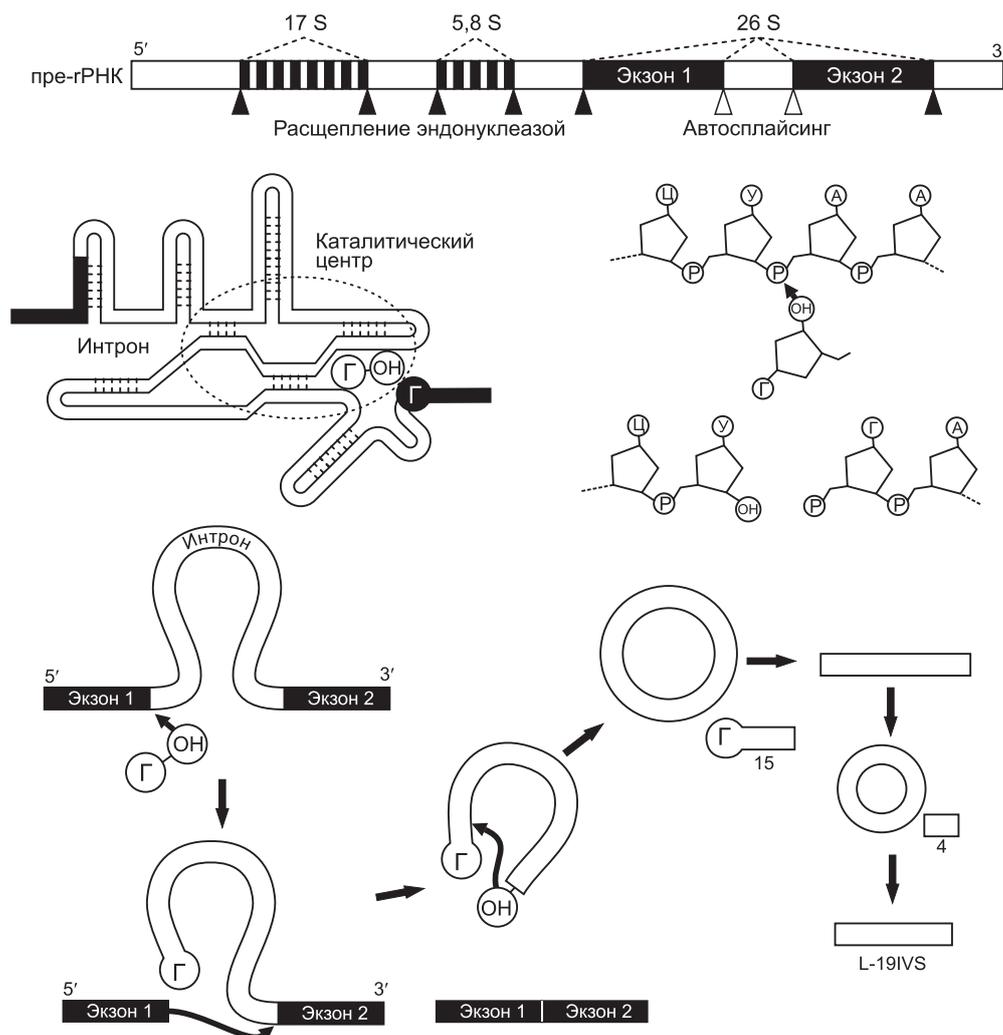


Рис. 7. Образование рибосомных РНК у инфузории. Интрон выбрасывается из 26S rРНК автосплайсингом I типа.

интронная последовательность –19 нуклеотидов) – демонстрация каталитической потенции самого интрона.

Так, впервые было показано, что каталитической активностью обладают не только белки, но и РНК. Молекулы РНК с такой активностью назвали *рибозимами* по аналогии с белковыми ферментами, которые называют также *энзимами*. К настоящему времени описано несколько сотен рибозимов.

АВТОСПЛАЙСИНГ II ТИПА

В 1986 г. в митохондриях дрожжей были открыты интроны, также подвергавшиеся автосплайсингу, но в ходе процесса, отличавшегося

от описанного Т. Чеком. Эти интроны, названные интронами группы II, оказались сходными с ретротранспозонами. В настоящее время показано, что такие интроны довольно широко распространены в митохондриях низших эукариот, в пластидах растений и у зубактерий.

В митохондриях дрожжей вырезание интронов и сшивание экзонов достигается двумя разными путями: автосплайсингом II типа и сплайсингом, катализируемым ферментом *матуразой*. В ходе собственно автосплайсинга II типа атаку 5'-конца интрона осуществляет одна из множества 2'-гидроксильных групп самого интрона. В третичной структуре интрона аденозин, находящийся в одной из 6 шпилек (ближайшей к 3'-концу), сближается с границей

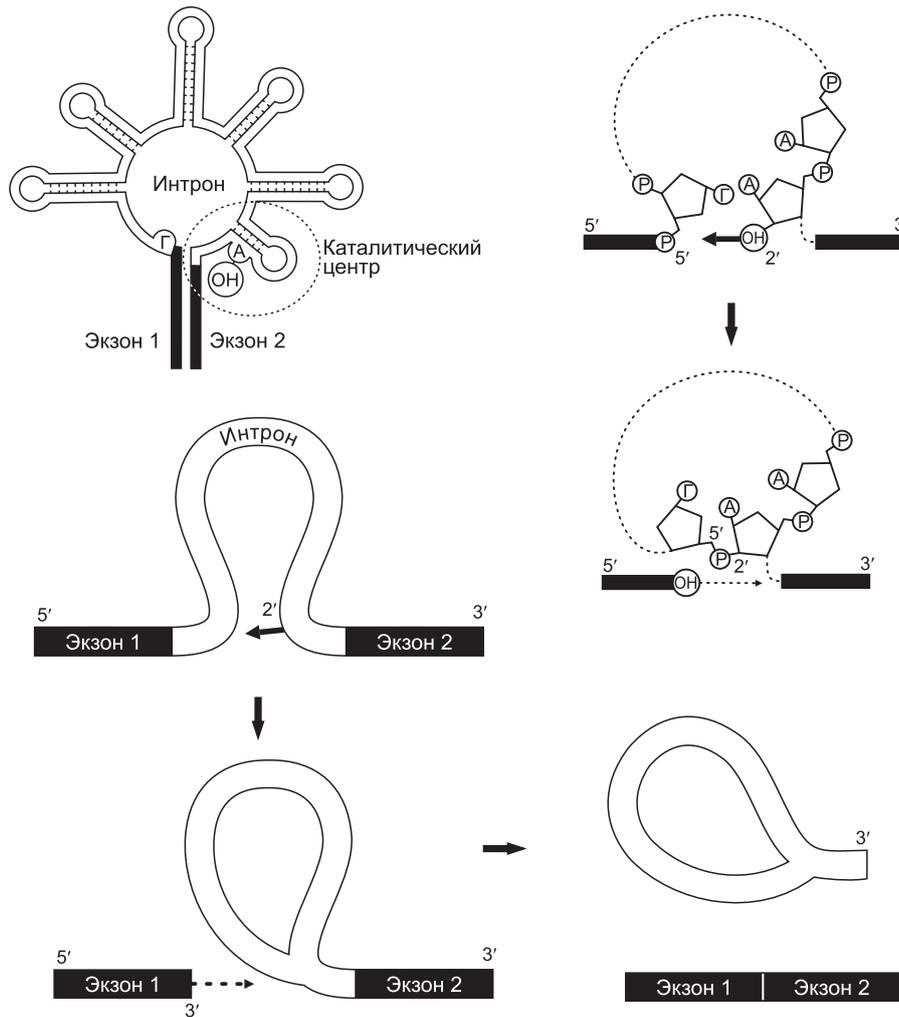


Рис. 8. Интрон, выброшенный автосплайсингом II типа, имеет форму лассо.

эксон–интрон. После образования 5'–2'-фосфодиэфирной связи 3'-ОН группа экзона атакует границу интрона со следующим экзонам. Экзоны сшиваются, выброшенный интрон имеет форму лассо (рис. 8). Такая же картина наблюдается при сплайсинге **mРНК ядерных генов**, но там это осуществляется сплайсосомой, а не автокаталитически.

Другой механизм удаления интронов в митохондриях дрожжей осуществляется специальными ферментами – **матуразами**. Как оказалось, они кодируются в самих интронах (вернее, часть белка кодируется экзонами, а часть – прилежащими участками интронов) генов, **mРНК которых подвергается сплайсингу** (рис. 9). Синтезированный фермент осуществляет сплайсинг, вырезая интрон и тем самым

уничтожая собственную матрицу. Это предотвращает синтез лишних молекул фермента.

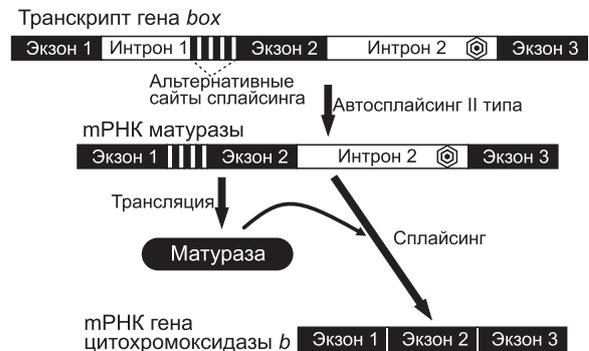


Рис. 9. Матураза – фермент сплайсинга в митохондриях дрожжей – кодируется экзонами и вырезается ею интронами.

РЕТРОИНТРОНЫ

Интроны группы II фактически являются ретротранспозонами, поэтому их называют *ретроинтронами*. Как и другие ретроэлементы, ретроинтроны кодируют белок, обладающий ревертазной (обратно-транскриптазной) активностью. В открытой рамке считывания кодируется белок IEP (**intron-encoded protein**). После трансляции этот белок вступает во взаимодействие с той частью мРНК, которая является рибозимом, и помогает ей принять правильную конформацию. Активированный рибозим осуществляет автосплайсинг и вырезается из РНК. Обладая ревертазной активностью белок IEP обеспечивает синтез ДНК по матрице РНК вырезанного интрона и встраивание этой ДНК в подходящее место. В результате в геноме появляется новая копия интрона группы II (рис. 10). В большинстве случаев бактерии содержат единичные экземпляры ретроинтронов, поэтому их долго не могли обнаружить и считали, что у эубактерий вообще нет сплайсинга. Недавно обнаружено, что термофильная цианобактерия *Thermosynechococcus elongatus* содержит 28 интронов группы II, из которых 17 лишены гена *ier*, но, тем не менее, размножаются при участии «чужого» белка IEP.

Предполагается, что древние эукариоты «заразились» интронами группы II от своих симбионтов-бактерий, которые стали предками митохондрий. В ядерных геномах эукариот интроны группы II размножились и превратились в обычные сплайсосомные интроны. При этом обратно-транскриптазная активность была ими утрачена. Функциональные части интронов группы II, необходимые для сплайсинга, были выведены из их состава и дали начало малым ядерным РНК, которые представляют собой важнейшую составную часть сплайсосом. Другие интроны группы II, сохранившие в своем составе гены обратных транскриптаз, вероятно, дали начало одному из классов ретротранспозонов (не содержащих LTR).

СПЛАЙСИНГ tРНК ДРОЖЖЕЙ

Для транспортных РНК дрожжей характерен особый механизм сплайсинга. В его основе лежит узнавание вторичной структуры tРНК, а

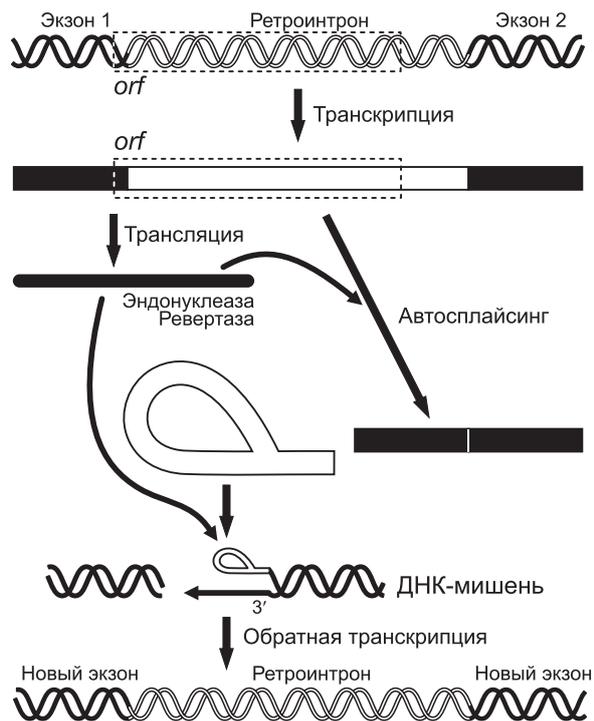


Рис. 10. В интроне группы II закодирован белок, обеспечивающий размножение ретроинтрона.

не первичной структуры интрона (сплайсинг ядерных мРНК) или его особой третичной структуры (автосплайсинг I и II типа). 59 ядерных генов tРНК дрожжей содержат по одному интрону, расположенному на расстоянии одного нуклеотида от 3'-конца антикодона. Длина интрона варьирует от 14 до 60 нуклеотидов. В изоакцепторных tРНК интроны гомологичны, в разных tРНК они не похожи друг на друга. Все интроны содержат последовательность, комплементарную антикодону. Это приводит к формированию необычной конформации антикодоновой петли. Остальная часть молекулы сохраняет свою обычную структуру (рис. 11). Специфическая эндонуклеаза расщепляет пре-tРНК по обоим концам интрона. На 3'-конце первого экзона оказывается 2'-3'-циклический фосфат, на 5'-конце второго экзона – гидроксильная группа. Фосфодиэстераза обнажает 3'-ОН группу, киназа фосфорилирует 5'-конец, лигаза сшивает экзоны с образованием tРНК. 2'-фосфатная группа удаляется фосфатазой. Все ферментативные реакции контролируются последовательностями, не входящими в интрон.

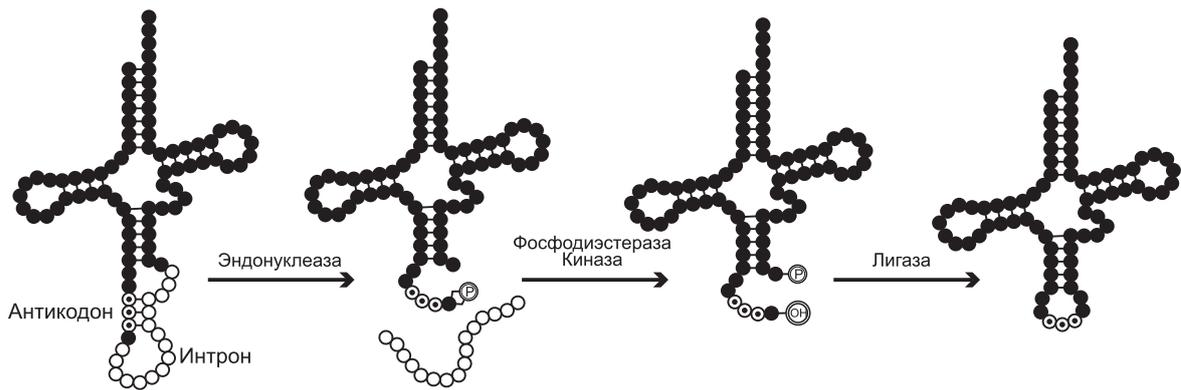


Рис. 11. Сплайсинг tРНК в ядрах дрожжевых клеток.

ТРАНС-СПЛАЙСИНГ

Все рассмотренные выше механизмы касались цис-сплайсинга, т. е. сшивания частей из одной молекулы РНК. *Транс-сплайсинг* – это сшивание экзонов, принадлежащих разным молекулам РНК. Многие отличающиеся друг от друга mРНК трипаносомы имеют одинаковый 35-нуклеотидный 5'-конец. Он является Э1 в SL (spliced leader) РНК. Вырезание (вернее, отрезание) частей интрона осуществляется с помощью сплайсосомы, не содержащей малую ядерную РНК U1. Считается, что присущую U1 функцию связывания 5'-конца интрона принимает на себя собственно SL РНК. Сшитые экзоны образуют зрелую mРНК. Две части интрона не связаны между собой ковалентно, поэтому вместо лассо образуется Y-образная структура (рис. 12).

ИНТРОНЫ И ЭВОЛЮЦИЯ

С увеличением размера генома в ходе эволюции эукариот интроны демонстрируют тенденцию к увеличению при том, что экзоны остаются довольно короткими. У мух и червей средний интрон (приблизительно 150 п.н.) по размеру практически равен экзону. Однако у червей почти нет длинных (более 5 тыс. п.н.) интронов, а у мух они составляют заметную часть (20%). У позвоночных размеры интронов представлены в более широком диапазоне: от сравнимых по размеру с экзонами (менее 200 п.н.) до длин, измеряемых десятками тысяч п.н. и достигающих в отдельных случаях 60 тыс. п.н. В целом у насекомых, птиц и млекопитающих

ген среднего размера примерно в 5 раз длиннее своей mРНК.

Как правило, очень большие гены своими размерами обязаны именно интронам, и их длина никак не связана с величиной кодируемого продукта. У высших эукариот отсутствует корреляция между размером гена, длиной mРНК и количеством экзонов. Размер гена в первую очередь зависит от длины его интронов. Так, ген *C9* (он кодирует один из компонентов системы комплемента) и у человека, и у рыбы фугу содержит 11 гомологичных экзонов. При этом размер гена у фугу 2 200 п.н., а у человека – 90 000 п.н. Объясняется это тем, что средний размер интрона в этом гене у рыбы 105 п.н., а у человека – 3 365 п.н. Предполагают, что длина интронов может играть определенную роль в регуляции экспрессии содержащих их генов, которая значительно сложнее у человека, чем у рыбы. Размер и число интронов в геноме

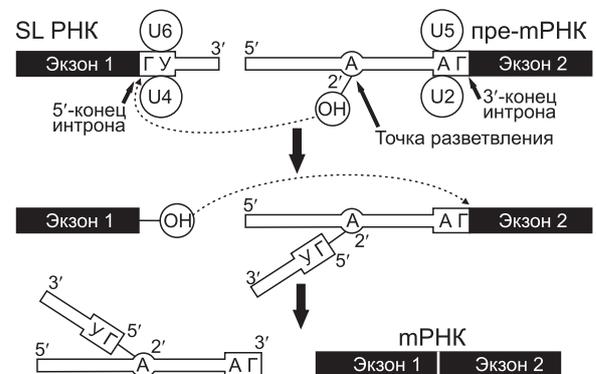


Рис. 12. Транс-сплайсинг у трипаносомы.

человека в расчете на один ген больше, чем у любого другого животного.

Существует две точки зрения на происхождение интронов и, соответственно, процесса сплайсинга – гипотезы так называемых «ранних интронов» и «поздних интронов». Гипотеза «ранних интронов» предполагает, что на самых ранних ступенях эволюции более или менее случайно возникали олигонуклеотиды, часть цепи которых оказывалась способной кодировать олигопептид, а остальные нуклеотиды были «лишними». При объединении таких небольших «прогенов» некодирующие последовательности и дали начало интронам. Впоследствии прокариоты избавились от этих балластных участков, что дало им преимущество – компактный геном, экономию ресурсов и времени при его репликации. Те прокариоты, которые дали начало эукариотам, скорее всего, археи, сохранили интроны, которые впоследствии приобрели многие полезные функции: участие в регуляции экспрессии генов, альтернативный сплайсинг.

Согласно второй гипотезе, интроны были приобретены прокариотами при внедрении мобильных ретроэлементов. Впоследствии эти ретроинтроны оказались в эукариотических клетках в составе геномов митохондрий и пластид. У высших эукариот в результате известного процесса переноса митохондриальных генов в ядро интроны оказались и в ядерном геноме. При этом ими была утрачена способность к кодированию обратной транскриптазы и эндонуклеазы. Функциональные участки, необходимые для сплайсинга, были перемещены в другие места генома и дали начало малым ядерным РНК, которые составляют основу сплайсосом. В пользу этой гипотезы говорит общность механизмов сплайсинга разных типов, консервативность концевых участков интронов.

По-видимому, интроны транспортных РНК дрожжей неродственны остальным интронам. Не исключено, что они имеют другое происхож-

дение, возможно, соответствующее гипотезе «ранних интронов».

ЗАКЛЮЧЕНИЕ

Представления о роли так называемой «некодирующей ДНК» за последние 20 лет претерпели существенные изменения. Интроны, участки, ранее считавшиеся «мусорной ДНК», как оказалось, играют существенную роль в молекулярных механизмах реализации генетической информации. Так, около половины известных микроРНК (miРНК) млекопитающих, участвующих в регуляции экспрессии генов, кодируется в интронах генов белков. Стало ясно, что сплайсинг – явление, свойственное всем группам организмов, включая прокариот, которым до недавнего времени отказывали в наличии этого процесса. Исследования показывают, что интроны, ретротранспозоны и ретровирусы филогенетически связаны друг с другом и являются неотъемлемой частью единого живого мира.

Особенно велико значение интронов и сплайсинга для высших эукариот. Известно, что увеличение количества генов в геноме ограничивается резким усложнением систем регуляции. Возникновение альтернативного сплайсинга позволило обойти это ограничение путем увеличения разнообразия белков без увеличения количества генов. Это значительно расширило эволюционные возможности многоклеточных эукариот, позволив создать новые уровни регуляции экспрессии генов.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

- Льюин Б. Гены. М.: Бином. Лаборатория знаний, 2011. 896 с.
- Breathnach R., Chambon P. Organization and expression of eukaryotic split genes coding for proteins // *Annu. Rev. Biochem.* 1981. V. 50. P. 349–383.
- Guthrie C. Messenger RNA splicing in yeast: clues to why spliceosome is a ribonucleoprotein // *Science*. 1991. V. 253. P. 157–163.
- Sharp P.A. Split genes and RNA splicing // *Cell*. 1994. V. 77. P. 805–815.