Introns are genomic sequences that are removed from the corresponding RNA transcripts of genes. Group I and II introns are both found in some bacterial and organellar genomes, and group I introns are also found in ribosomal RNAs (rRNAs) of protist and fungal nuclei1–3. These two groups have distinct RNA structures that facilitate their self-splicing activity. They also contain internal ORFs, which facilitate both intron removal from RNA transcripts and intron propagation to intronless sites through reverse transcription. In total, around 1,500 group I and200 group II introns have been identified1.

By contrast, a third group of introns — spliceosomal introns — are found in the nuclear genomes of all characterized eukaryotes. They have quasi-random sequences and generally lack ORFs. Their lengths vary widely between species, from just tens of bases in some protists to hundreds of kilobases in mammals. The spliceosome, a complex that comprises five RNAs and hundreds of proteins, removes spliceosomal introns from RNA transcripts, a process that is coupled to several other transcript-processing steps4. Despite important differences between spliceosomal and other introns, similarities between the splicing mechanisms of group II and spliceosomal introns indicate a possible evolutionary relationship between the two5–9.

The timing and causes of spliceosomal intron evolution are matters of great interest in the study of genome evolution as a whole. Spliceosomal introns are absent in prokaryotes and their numbers vary tremendously between eukaryotic species, from fewer than 100 introns per genome in some species to hundreds of thousands per genome in vertebrates and plants (FIG. 1). However, despite the huge numbers of intron gains and/or losses that are implied by these differences, there is less certainty about the mechanisms and forces that underlie intron gain and loss than about any other major class of genetic element. There are millions of known introns in coding regions, but there is only one known intraspecific presence/absence polymorphism10, and there are only three recently inserted introns for which the origins have been confidently traced11–13. Currently, several plausible hypotheses compete to explain the origin of new introns, and no consensus has been reached as to whether introns are positively, negatively or neutrally selected.

The vast interspecific differences in intron number between eukaryotic genomes constitute an important puzzle with which theories about the determinants of genome complexity — for example, selfish genetic elements14–16, organismal complexity15–19 or population size20,21 — must come to terms. Indeed, introns have prominent roles in several ambitious theories of genome evolution. A long-standing theory of intron origin postulates that recombination within introns facilitated the construction of the first full-length genes17,22–38, and a similar theory has been proposed for the origins of many multidomain metazoan genes39,40. Introns have also been suggested to increase fitness by increasing intragenic recombination22,41–43 or to have boosted transcript fidelity in early eukaryotes through nonsense-mediated decay (NMD)44. Differences in intron number have also served as the potential crowning example of the hypothesis that increases in genome complexity are often the results of, or are themselves, deleterious mutations20,21. Therefore, the evolution of spliceosomal introns has broad implications for many fundamental evolutionary questions.

Research into the timing, mechanisms and causes of spliceosomal intron evolution has been extremely active in the past few years, resolving some old controversies and sparking some new ones. Here we discuss recent studies of the rise and fall of intron number through eukaryotic evolution, mechanisms of intron gain and loss, and the evolutionary forces that might be responsible for these changes.

Интроны представляют собой геномные последовательности, удаленные из соответствующих РНК-транскриптов генов. Интроны группы I и II обнаружены в геномах некоторых бактерий и органелл, а интроны группы I также обнаружены в рибосомных РНК (рРНК) ядер протистов и грибов1–3. Эти две группы имеют разные структуры РНК, которые облегчают их активность самосплайсинга. Они также содержат внутренние ORF, которые облегчают как удаление интронов из транскриптов РНК, так и распространение интронов на безинтронные сайты посредством обратной транскрипции. Всего идентифицировано около 1500 интронов группы I и 200 интронов группы II1.

Напротив, третья группа интронов — сплайсосомные интроны — обнаружена в ядерных геномах всех охарактеризованных эукариот. Они имеют квазислучайные последовательности и обычно не имеют ORF. Их длина сильно различается у разных видов: от нескольких десятков оснований у некоторых простейших до сотен тысяч оснований у млекопитающих. Сплайсосома, комплекс, состоящий из пяти РНК и сотен белков, удаляет сплайсосомные интроны из РНК-транскриптов, этот процесс связан с несколькими другими стадиями процессинга транскриптов 4. Несмотря на важные различия между сплайсосомными и др. интронами, сходство между механизмами сплайсинга группы II и сплайсосомных интронов указывает на возможное эволюционное родство между ними 5-9.

Время и причины эволюции сплайсосомных интронов представляют большой интерес для изучения эволюции генома в целом. Сплайсосомные интроны отсутствуют у прокариот, и их число сильно различается у эукариотических видов, от менее чем 100 интронов на геном у некоторых видов до сотен тысяч на геном у позвоночных и растений (FIG. 1). Однако, несмотря на огромное количество прибавлений и/или потерь интронов, которые подразумеваются этими различиями, в отношении механизмов и сил, лежащих в основе прибавления и потери интронов, определенности меньше, чем в отношении любого другого крупного класса генетических элементов. Существуют миллионы известных интронов в кодирующих регионах, но есть только один известный внутривидовой полиморфизм присутствия/отсутствия10, и только три недавно вставленных интрона, происхождение которых достоверно прослежено11-13. В настоящее время несколько правдоподобных гипотез конкурируют за объяснение происхождения новых интронов, и не было достигнуто консенсуса относительно того, выбираются ли интроны положительно, отрицательно или нейтрально.

Огромные межвидовые различия в числе интронов между геномами эукариот представляют собой важную загадку, с которой должны прийти к согласию теории о детерминантах сложности генома, например эгоистичных генетических элементах14–16, сложности организма15–19 или размере популяции20,21. Действительно, интроны играют важную роль в нескольких амбициозных теориях эволюции генома. Давняя теория происхождения интронов постулирует, что рекомбинация внутри интронов способствовала конструированию первых полноразмерных генов 17,22-38, и аналогичная теория была предложена для происхождения многих мультидоменных генов многоклеточных животных 39,40. Также предполагалось, что интроны улучшают приспособленность за счет увеличения внутригенной рекомбинации 22,41-43 или повышают точность транскриптов у ранних эукариот посредством нонсенс-опосредованного распада (NMD)44. Различия в количестве интронов также служили потенциальным завершающим примером гипотезы о том, что увеличение сложности генома часто является результатом или само по себе является вредными мутациями20,21. Следовательно, эволюция сплайсосомных интронов имеет широкое значение для многих фундаментальных эволюционных вопросов.

Исследования времени, механизмов и причин эволюции сплайсосомных интронов были чрезвычайно активны в последние несколько лет, разрешая некоторые старые противоречия и вызывая новые. Здесь мы обсуждаем недавние исследования увеличения и уменьшения числа интронов в ходе эволюции эукариот, механизмы приобретения и потери интронов и эволюционные силы, которые могут быть ответственны за эти изменения.

Cannone, J. J. et al. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. MBC Bioinformatics 3, 2 (2002).

Cannone, J.J., Subramanian, S., Schnare, M.N. et al. 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 (2002). https://doi.org/10.1186/1471-2105-3-2