Journal of Molecular Evolution (2025) 93:474–477
https://doi.org/10.1007/s00239-025-10261-9

MINI-REVIEW

Recursive splicing—a mechanism of intron removal with an unexplored role in the largest genomes

Rachel Lockridge Mueller[1] · **Alexander Nichols Adams**[1]

Received: 26 February 2025 / Accepted: 24 July 2025 / Published online: 1 August 2025

Abstract

Intron splicing is a critical step that pre-mRNA transcripts undergo to become mature mRNAs. Although long thought to occur in a single step, introns are now also known to be removed by a multi-step process called recursive splicing. In recursive splicing, the spliceosome removes the intron one segment at a time with segments defined by discreet sequences called recursive splice sites. As each segment is removed, the remaining downstream intronic sequence is brought into contact with the upstream exon. Recursive splicing can be detected through RNA-seq analysis because it produces a "sawtooth" pattern of read depth across intron length with peaks corresponding to sites in the ephemeral partially spliced introns where the remaining downstream intron segments contact the upstream exon. Recursive splicing can also be detected by RNA lariat sequencing and real-time imaging of single-cell transcriptional and splicing dynamics. These methods have been applied to fruit flies, humans, and mice, revealing that recursive splicing 1) increases in prevalence with intron length, and 2) increases splicing fidelity, particularly in long introns. However, intron lengths in the typically sized genomes of these model organisms fail to represent the diversity that exists across the tree of life. Species with gigantic genomes like salamanders and lungfishes have introns that are ten- to 50-fold longer. Future studies targeting recursive splicing in gigantic genomes will provide a unique perspective on its functional significance and will also reveal whether this splicing mechanism plays a role in overcoming constraints placed on transcriptional capacity and efficiency by enormous introns.

Keywords Intron evolution · Spliceosome · Transcription · Genomic gigantism

Introduction

In a typical eukaryotic cell, DNA is transcribed into a pre-mRNA transcript that must undergo further processing in order to make a mature mRNA transcript. A cap structure is added to the 5′ end, a poly-adenylated tail is added to the 3′ end, and the introns are spliced out (Proudfoot, et al. 2002). In the canonical form of splicing, introns are removed all at once by two reactions catalyzed by the spliceosome. In the first reaction, called branching, the intron is cleaved from the 5′ exon and a lariat-3′ exon intermediate is formed. In the second reaction, called exon ligation, the 5′ and 3′ exons are ligated together and the lariat intron is released (Wilkinson, et al. 2020).

An alternative to this canonical splicing mechanism was later discovered (Gehring and Roignant 2021). This non-canonical form of splicing — called recursive splicing (RS) — works by removing introns in multiple sequential pieces instead of all at once (Hatton, et al. 1998; Burnette, et al. 2005). The mechanism works by splicing out a section of intron flanked by discreet splicing sites, referred to as recursive splice sites (RSS) (Duff, et al. 2015; Kelly, et al. 2015; Sibley, et al. 2015). These subsections are still removed using lariat structures and spliceosomes (Hoppe, et al. 2023). As each is removed sequentially, the remaining downstream intronic sequence is brought into contact with the upstream exon at the RSS (Sibley, et al. 2015; Georgomanolis, et al. 2016).

The ephemeral, partially spliced introns formed as intermediates during recursive splicing can be identified through deep sequencing of total RNA, or rRNA-depleted RNA-seq libraries constructed without using poly-A selection (Sibley,

Handling editor: **David Alvarez-Ponce**.

✉ Rachel Lockridge Mueller
Rachel.Mueller@colostate.edu

[1] Department of Biology, Colorado State University, Fort Collins, CO 80523-1878, USA

et al. 2015). Their existence results in uneven RNA-seq read depths across the entire length of the intron that produce a characteristic sawtooth pattern of depth of coverage with the RSS defining the edges of the sawtooth. Based on the identification of this pattern, recursive splicing has been identified in several model species: In *Drosophila*, several hundred recursive splicing events were identified (Duff, et al. 2015; Joseph, et al. 2018). In humans, 435 putative RSS were identified, 9 of which were verified experimentally (Sibley, et al. 2015). In mice, only 19 recursive splice sites were identified (Moon and Zhao 2022) (Fig. 1).

As analytical methods developed to include intron lariat sequencing and single-molecule imaging, recursive splicing was revealed in ~30% of human genes, in introns of all lengths, and in multiple cell types (Wan, et al. 2021; Hoppe, et al. 2023). Importantly, recursive splicing was found to occur at intronic positions above and beyond the previously identified conserved sites; RSS selection was also shown to be stochastic, removing introns in a variety of different segments defined not at conserved sites, but rather by random selection by the spliceosome of one of many possible RSS (Wan, et al. 2021). This stochastic process produces diverse, transient intermediate RNA molecules that would not appear as a sawtooth pattern in RNA-seq datasets. Instead, they would appear as a pattern of decreasing RNA-seq read depth across the length of the entire intron, the pattern that is also predicted under canonical cotranscriptional splicing. These approaches have not yet been widely applied to species beyond humans.

Within taxa, the presence of recursive splicing has been shown to correlate with intron length; most RSS in *Drosophila* occur in introns longer than 40kb, most RSS in mice occur in introns longer than 51kb, and recursive splicing at both conserved and stochastic sites is more prevalent in longer introns in humans, which also tend to be the first intron (Bradnam and Korf 2008; Sibley, et al. 2015; Pai, et al. 2018; Wan, et al. 2021; Moon and Zhao 2022; Hoppe, et al. 2023). In *Drosophila*, several lines of evidence suggest that recursive splicing is functionally important for intron splicing. RSS are distributed non-randomly across the longest introns, subdividing the longest ones into equal subsections (Burnette, et al. 2005; Joseph, et al. 2018; Pai, et al. 2018). Additionally, splicing becomes increasingly noisy and error-prone as intron lengths increase, and recursive splicing leads to more accurate, albeit slower, splicing of the longest introns (Pai, et al. 2018). In humans, splicing also becomes noisier and more error-prone as intron length increases (Pickrell, et al. 2010). Interestingly, however, initial estimates of recursive splicing in humans and mice — which targeted nervous system tissue, as it is characterized by longer transcript lengths — revealed lower levels than those seen in flies, despite overall longer intron lengths (Sibley, et al. 2015; Joseph, et al. 2018; Moon and Zhao 2022). Thus, the general relationship between intron length and reliance on recursive splicing remains unclear.

To date, all of the work on recursive splicing has examined introns in the typically sized genomes of model organisms. However, natural genome size diversity extends across a much greater range than is seen across humans, mice, and *Drosophila*; salamanders and lungfishes, for example, demonstrate up to a 40-fold increase in genome size relative to humans (Gregory 2025). As genomes evolve towards larger sizes, the amount of protein-coding DNA does not increase in proportion to overall genome size (Aparicio, et al. 2002). Instead, genomic expansion reflects a relative increase in transposable elements and intronic sequence (Smith, et al. 2009; Sun, et al. 2012; Schartl, et al. 2024). On average, intron length in the model salamander *Ambystoma mexicanum* (genome size = 32 Gb) is 13 times longer than in humans, and similar intronic expansion is seen in the Australian lungfish *Neoceratodus forsteri* (genome size = ~50 Gb) (Nowoshilow, et al. 2018; Meyer, et al. 2021). The largest salamander and lungfish genomes are ~120 Gb, suggesting that intron lengths in these taxa may be ~50 times

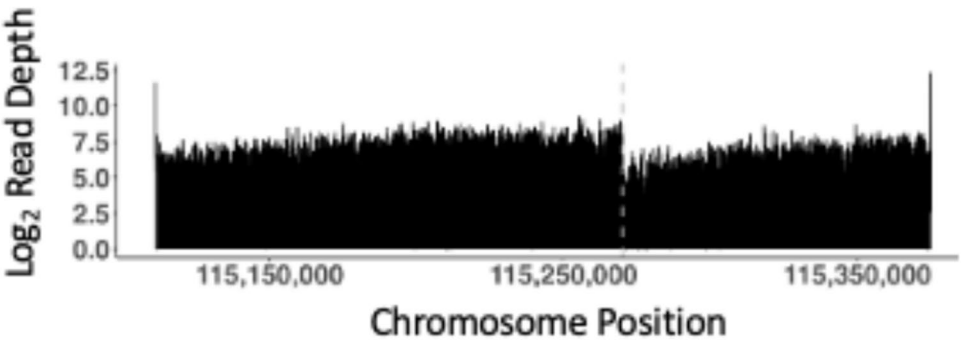

Fig. 1 Example of RSS identification in a human CADM1 intron following the pipeline outlined in (Sibley, et al. 2015). The RNA-seq dataset comprises ~25.9 Gb of Illumina HiSeq 100 bp paired-end sequence data from libraries constructed from rRNA-depleted total RNA extracted from liver tissue (SRX20105218). The edge of the characteristic "sawtooth" is indicated with a dotted line

the lengths of human introns. Considering these extraordinary intron lengths, species with expanded genome sizes provide a powerful system in which to examine the relationship between intron length and recursive splicing. Over the last decade, advances in sequencing and assembly have made gigantic genomes amenable to genomic and transcriptomic analyses (Warren, et al. 2015; Stevens, et al. 2016; Nowoshilow, et al. 2018; Wang, et al. 2021; Schartl, et al. 2024), opening the door for genomic and RNA-seq-based studies of recursive splicing that leverage natural diversity in genome size. More generally, we note that relatively little research has focused on how extreme genomic expansion alters the transcriptional processes of the cell (Sessions and Wake 2021; Taylor, et al. 2024), although increased genome size has been connected to increased nucleus size, cell size, and changes at the tissue and organ levels (Gregory 2001; Marguerat and Bähler 2012; Itgen, et al. 2022).

Based on existing data, one logical hypothesis is that the long introns in gigantic genomes would experience noisy, error-prone splicing if excised as single lariats, and that more widespread reliance on recursive splicing might have evolved to curb this noise. This hypothesis could be tested with RNA-seq read depth-based analysis, using the human genome both as a positive control for validating the RSS identification pipeline, and as a point of comparison representing a typically sized vertebrate genome (Sibley, et al. 2015). Although this approach undercounts total recursive splicing levels, comparing results between the gigantic genome and the much smaller human genome analyzed using the same approach would provide a preliminary test of whether genomic expansion is correlated with increased reliance on recursive splicing (Sibley, et al. 2015). Lariat sequencing as well as dynamic real-time imaging of single-cell transcriptional and splicing dynamics in species with enormous genomes could also be undertaken, enabling more exact comparisons between recursive splicing levels in species with extremely long introns versus more typical (i.e. human) vertebrate intron lengths (Wan, et al. 2021; Hoppe, et al. 2023). Notably, comparative analyses of lungfish and salamander transcriptomes with those of fish with smaller genomes suggest that increases in genome size are associated with increases in pervasive transcription, suggesting a noisier transcriptional environment overall (Fuselli, et al. 2023). Thus, the relationship between intron splicing mechanism, intron length, and tolerable levels of transcriptional error may be complex across evolutionary diversity in genome size and is an important target for future research.

In addition to the intron length variation that exists across species with different genome sizes, there is also intron length variation within genomes. Early zygotic genes in model organisms with smaller genomes (flies, mosquitos, zebrafish, and mice) typically have relatively short intron lengths, reflecting the constraints imposed on total transcription time by rapid cell division during this period of development, as transcription is suppressed during mitosis (Heyn, et al. 2015). In the large genome of the salamander *Ambystoma mexicanum*, developmental genes overall have shorter introns than non-developmental genes, which may also reflect constraints on transcription time throughout development (Nowoshilow, et al. 2018). Somewhat surprisingly, *Drosophila* genes showing evidence for recursive splicing were enriched for developmental processes based on Gene Ontology analysis (Joseph, et al. 2018), suggesting that both speed and accuracy may be especially relevant for genes active during development. Similar analyses have not yet been done in gigantic genomes. The relationship between intron length and recursive splicing may reflect multiple interacting functional demands that vary across large and small genome sizes and is another important target for future research.

Overall, we advocate for the molecular evolutionary biology community to engage in research aimed at the transcriptional and splicing dynamics of genomes from phylogenetically diverse taxa that cover the full range of genome sizes. These data, in turn, will shine light on whether recursive splicing plays a role in driving intron evolution, including accommodating the extraordinary increases in intron size that have occurred in the largest genomes.

Acknowledgments We thank members of A Adams' thesis committee K Hoke, D Sloan, and J Hansen for helpful discussion and feedback throughout the project. Funding for this project was provided by NSF grant 1911585 to R Mueller and Colorado State University.

Funding Directorate for Biological Sciences, 1911585, Rachel Lockridge Mueller.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J-m, Dehal P, Christoffels A, Rash S, Hoon S, Smit A (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297:1301–1310

Bradnam KR, Korf I (2008) Longer first introns are a general property of eukaryotic gene structure. PLoS One 3:e3093

Burnette JM, Miyamoto-Sato E, Schaub MA, Conklin J, Lopez AJ (2005) Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. Genetics 170:661–674

Duff MO, Olson S, Wei X, Garrett SC, Osman A, Bolisetty M, Plocik A, Celniker SE, Graveley BR (2015) Genome-wide identification of zero nucleotide recursive splicing in Drosophila. Nature 521:376–379

Fuselli S, Greco S, Biello R, Palmitessa S, Lago M, Meneghetti C, McDougall C, Trucchi E, Rota Stabelli O, Biscotti AM (2023) Relaxation of natural selection in the evolution of the giant lungfish genomes. Mol Biol Evolut 40:msad193

Gehring NH, Roignant J-Y (2021) Anything but ordinary–emerging splicing mechanisms in eukaryotic gene regulation. Trends Gen 37:355–372
Georgomanolis T, Sofiadis K, Papantonis A (2016) Cutting a long intron short: Recursive splicing and its implications. Front Physiol 7:598
Gregory TR (2001) The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood Cells Mol Dis 27:830–843
Gregory TR (2025) Animal Genome Size Database. http://www.genomesize.com
Hatton AR, Subramaniam V, Lopez AJ (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon–exon junctions. Molecular Cell 2:787–796
Heyn P, Kalinka AT, Tomancak P, Neugebauer KM (2015) Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. Bioessays 37:148–154
Hoppe ER, Udy DB, Bradley RK (2023) Recursive splicing discovery using lariats in total RNA sequencing. Life Sci Alliance. https://doi.org/10.26508/lsa.202201889
Itgen MW, Siegel DS, Sessions SK, Mueller RL (2022) Genome size drives morphological evolution in organ-specific ways. Evolution in press
Joseph B, Kondo S, Lai EC (2018) Short cryptic exons mediate recursive splicing in Drosophila. Nature Struct Mol Biol 25:365–371
Kelly S, Georgomanolis T, Zirkel A, Diermeier S, O'Reilly D, Murphy S, Längst G, Cook PR, Papantonis A (2015) Splicing of many human genes involves sites embedded within introns. Nucleic Acids Res 43:4721–4732
Marguerat S, Bähler J (2012) Coordinating genome expression with cell size. Trends Gen 28:560–565
Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A (2021) Giant lungfish genome elucidates the conquest of land by vertebrates. Nature 590:284–289
Moon S, Zhao Y-T (2022) Recursive splicing is a rare event in the mouse brain. PLoS One 17:e0263082
Nowoshilow S, Schloissnig S, Fei J-F, Dahl A, Pang AWC, Pippel M, Winkler S, Hastie AR, Young G, Roscito JG et al (2018) The axolotl genome and the evolution of key tissue formation regulators. Nature 554:50–55
Pai AA, Paggi JM, Yan P, Adelman K, Burge CB (2018) Numerous recursive sites contribute to accuracy of splicing in long introns in flies. PLoS Genetics 14:e1007588
Pickrell JK, Pai AA, Gilad Y, Pritchard JK (2010) Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genetics 6:e1001236
Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. Cell 108:501–512
Schartl M, Woltering JM, Irisarri I, Du K, Kneitz S, Pippel M, Brown T, Franchini P, Li J, Li M et al (2024) The genomes of all lungfish inform on genome expansion and tetrapod evolution. Nature 634:96–103
Sessions SK, Wake DB (2021) Forever young: Linking regeneration and genome size in salamanders. Dev Dyn 250:768–778
Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, Trabzuni D, Ryten M, Weale ME, Hardy J (2015) Recursive splicing in long vertebrate genes. Nature 521:371–375
Smith J, Putta S, Zhu W, Pao G, Verma I, Hunter T, Bryant S, Gardiner D, Harkins T, Voss SR (2009) Genic regions of a large salamander genome contain long introns and novel genes. BMC Genomics 10:19
Stevens KA, Wegrzyn JL, Zimin A, Puiu D, Crepeau M, Cardeno C, Paul R, Gonzalez-Ibeas D, Koriabine M, Holtz-Morris AE (2016) Sequence of the sugar pine megagenome. Genetics 204:1613–1626
Sun C, Shepard DB, Chong RA, Arriaza JL, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL (2012) LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol 4:168–183
Taylor A, Prasad A, Mueller RL (2024) Amphibian segmentation clock models suggest how large genome and cell sizes slow developmental rate. Integr Organismal Biol 6:obae021
Wan Y, Anastasakis DG, Rodriguez J, Palangat M, Gudla P, Zaki G, Tandon M, Pegoraro G, Chow CC, Hafner M (2021) Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell 184(2878–2895):e2820
Wang J, Itgen MW, Wang H, Gong Y, Jiang J, Li J, Sun C, Sessions SK, Mueller RL (2021) Gigantic genomes provide empirical tests of transposable element dynamics models. Genom Proteom Bioinform 19:123–139
Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJ, Birol I (2015) LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience 4:s13742
Wilkinson ME, Charenton C, Nagai K (2020) RNA splicing by the spliceosome. Annual Rev Biochem 89:359–388