A novel spliceosome-mediated *trans*-splicing can change our view on genome complexity of the divergent eukaryote *Giardia intestinalis*

Ryoma Kamikawa · Yuji Inagaki · Tetsuo Hashimoto

Received: 28 June 2011 / Accepted: 29 September 2011 / Published online: 20 October 2011
© International Union for Pure and Applied Biophysics (IUPAB) and Springer 2011

Abstract Although spliceosomal introns are an abundant landmark in eukaryotic genomes, the nuclear genome of the divergent eukaryote *Giardia intestinalis*, the causative agent of giardiasis, has been considered as “intron-poor” with only five canonical (*cis*-spliced) introns. However, three research groups (including ours) have independently reported a novel class of spliceosomal introns in the *G. intestinalis* genome. Three protein-coding genes are split into pieces in the *G. intestinalis* genome, and each of the partial coding regions was independently transcribed into polyadenylated premature mRNAs (pre-mRNAs). The two pre-mRNAs directly interact with each other by an intermolecular-stem structure formed between their non-coding portions, and are then processed into mature mRNAs by spliceosome-mediated *trans*-splicing. Here, we summarize the recently published works on split introns (“splintrons”) in the *G. intestinalis* genome, and then provide our speculation on the functional property of the *Giardia* spliceosomes based on the putative ratio of splintrons to canonical introns. Finally, we discuss a scenario for the transition from typical GT-AG boundaries to non-typical AT-AC boundaries in a particular splintron of *Giardia*.

Keywords *cis*-splicing · Dynein · *Giardia intestinalis* · Heat shock protein 90 · RNA maturation · Spliceosomal introns · Splintrons · *Trans*-splicing

Abbreviations

HSP90	Heat shock protein 90
OAD- β	Outer arm dynein beta chain
OAD- γ	Outer arm dynein gamma chain
pre-mRNA	Premature mRNA

Introduction

Protein coding sequences in eukaryotic nuclear genomes are often intervened by one or more introns that are excised by the spliceosome, which comprises several small nuclear RNAs and hundreds of proteins (Jurica and Moore 2003). The chemical mechanism for the excision of spliceosomal introns from pre-mRNAs is identical to that for the excision of group II introns, a class of introns in the prokaryotic and organellar genomes (Lambowitz and Zimmerly 2010). Therefore, it is generally believed that the ancestral spliceosomal intron was derived from a group II intron that resided in the ancestral mitochondrial genome (Lambowitz and Zimmerly 2004; Roger and Doolittle 2000). Because mitochondrial endosymbiosis is believed to have taken place at a very early stage of eukaryotic evolution, the “root” of spliceosomal introns can be as deep as that of mitochondria in eukaryotic evolution (Embley and Martin 2006; Martin and Koonin 2006), and these “spliceosomal” introns have undoubtedly contributed to shape the modern eukaryotic genomes. In other words, if a particular lineage diverged from the main trunk of the eukaryotic lineages before the mitochondrial endosymbiosis, then these cells are anticipated to possess a genome with no or few spliceosomal introns.

Giardia intestinalis is an intestinal parasite in humans and various animals. Apart from its significance in public health, this organism has drawn the attention of evolution-

R. Kamikawa (✉) · Y. Inagaki · T. Hashimoto
Center for Computational Sciences and Institute of Biological
Sciences, University of Tsukuba,
1-1-1 Tennoudai,
Tsukuba, Ibaraki 305-8577, Japan
e-mail: ryoma@ccs.tsukuba.ac.jp

ary biologists, as its cellular and genome architectures match those of the hypothesized ancestral eukaryotic cells. For a long time, it was believed that *G. intestinalis* possesses neither a typical mitochondrion with respiratory function nor spliceosomal introns (Adam 2000). We now know that *G. intestinalis* secondarily lost its mitochondria (Roger et al. 1998; Hashimoto et al. 1998; Tovar et al. 2003), and five genes had been found to bear canonical (*cis*-spliced) spliceosomal introns in its genome (Nixon et al. 2002; Morrison et al. 2007; Roy et al. 2011). These findings suggest that this organism does not necessarily represent the ancestral eukaryotes; rather, they lost the typical mitochondria because of their anaerobic lifestyle, and spliceosomal introns in the *G. intestinalis* genome are merely more sparse than those in other eukaryotic genomes. Furthermore, recent studies have identified a novel class of spliceosomal introns in the *G. intestinalis* genome, which challenge our conventional view on *G. intestinalis* and its genome complexity (Kamikawa et al. 2011a).

In the following sections, we briefly introduce spliceosomal introns in a split form (so-called splintrons; Kamikawa et al. 2011a; Kamikawa et al. 2011b), and then, explore the putative properties of *G. intestinalis* spliceosomes. Finally, we discuss the evolution of a particular splintron with non-typical AT-AC boundaries in *G. intestinalis*.

What are splintrons?

We recently identified three splintrons in two functionally indispensable genes in *G. intestinalis* (Kamikawa et al. 2011a, b). For instance, the N-terminal and C-terminal portions of the heat shock protein 90 (HSP90) are encoded in two distinct loci in the same chromosome of *G. intestinalis*, and spliceosome-mediated *trans*-splicing produces the mRNA molecule that encodes the entire HSP90. Mechanisms of *hsp90* gene expression in *G. intestinalis* are shown in Fig. 1. The premature mRNAs (pre-mRNAs) for the N-terminal and C-terminal portions, which bear poly-A tails at their 3' ends, are transcribed independently from the two loci (N-terminal and C-terminal pre-mRNAs). The two pre-mRNAs are likely in close proximity by forming an intermolecular stem structure between the sequence stretch on the 3' non-coding region of N-terminal pre-mRNA and that on the 5' non-coding region of C-terminal pre-mRNA. Furthermore, we noticed that the intron-like motifs conserved among the *Giardia* canonical spliceosomal introns were also present in the non-coding regions of the two pre-mRNAs. These observations strongly indicate that the two pre-mRNAs can behave as a single, continuous pre-mRNA molecule with a spliceosomal intron, and that they are *trans*-spliced into the mature mRNA by spliceo-

somes. Indeed, we successfully identified both the mature mRNAs and the “Y-shaped” RNA molecules excised by spliceosomes. In addition, Nageshan et al. (2011) independently investigated the splintron in the *hsp90* gene and experimentally confirmed that no partial polypeptide corresponding to N-terminal pre-mRNA existed in the *G. intestinalis* cells. Furthermore, we and Roy et al. found two splintrons in the gene encoding the outer arm dynein β subunit (OAD β), which requires the processing of three pre-mRNAs into a single mature mRNA by *trans*-splicing (Kamikawa et al. 2011a, b; Roy et al. 2011). In addition, Roy et al. (2011) found a fourth splintron in the gene encoding the outer arm dynein γ chain (OAD γ) and a fifth *cis*-spliced intron. Splintron *trans*-splicing is apparently indispensable for the regular expression of the four *G. intestinalis* genes, which is distinctive from the major classes of *trans*-splicing known to date (Kamikawa et al. 2011a).

Prior to our work, a single splintron had been identified in the nematode *Caenorhabditis elegans* (Fischer et al. 2008). The mature mRNA encoding ERI-6/7 is generated through spliceosomal *trans*-splicing of two pre-mRNAs, which are separately transcribed from *eri-6* and *eri-7* loci. Importantly, *eri-6* and *eri-7* pre-mRNAs most likely interact with each other via an intramolecular stem structure during the *trans*-splicing reaction, as seen in *G. intestinalis* splintrons. As *G. intestinalis* and *C. elegans* are phylogenetically distantly related to each other, splintrons should have been evolved independently on the branches leading to the two species. Alternatively, as proposed by Blumenthal (2011), there is the possibility of splintrons having emerged anciently in eukaryotic evolution. For better understanding of the evolution of splintrons, we need to examine whether *G. intestinalis* and *C. elegans* are the sole organisms performing splintron splicing in mRNA maturation. Especially, intron data from close relatives of *G. intestinalis*, as well as those from close relatives of *C. elegans*, are indispensable to predict when and how splintrons emerged in eukaryotic phylogeny. Intriguingly, besides *C. elegans*, the genome data are available for several members of the genus *Caenorhabditis*, but no splintron has been identified in their genomes (e.g., *C. briggsae*; Fischer et al. 2008). Thus, splintron *trans*-splicing was most likely established after divergence of the extant *Caenorhabditis* species. On the other hand, the genome data of close relatives of *G. intestinalis* are not sufficient to propose any scenarios for splintron evolution on the branch leading to *G. intestinalis*. Therefore, it is important to conduct a stringent survey of canonical introns and splintrons in the sequence data from the genome project of the diplomonad *Spironucleus vortens*, the close relative of *Giardia* (Joint Genome Institute: <http://genome.jgi-psf.org/>).

Fig. 1 A schematic representation of *trans*-splicing of the splintrons in the *Giardia intestinalis* *hsp90* gene. The N-terminal and C-terminal portions are encoded in two distant loci in the genome. “N-terminal” and “C-terminal” pre-mRNAs with polyA⁺ tails are independently transcribed from the two loci, and physically interact with each other via the intermolecular stem structure between the “left intron piece” and the “right intron piece.” Spliceosome-mediated *trans*-splicing generates mature mRNAs encoding the entire amino acid sequence of an HSP90; the excised intron pieces connect to each other in a “Y-shape.” The “Y-shaped” RNA molecules likely stay in the nucleus and are eventually degraded, while the mature mRNAs are exported from the nucleus to cytoplasm and then subjected to translation

Putative properties of the *G. intestinalis* spliceosomes

In light of the three studies demonstrating splintrons, our views on the evolutionary dynamics of spliceosomal introns and spliceosomes in the *G. intestinalis* genome need to be revised. In the *G. intestinalis* genome, the number of splintrons (four; see Kamikawa et al. 2011a, b; Nageshan et al. 2011; Roy et al. 2011) almost equals that of canonical introns (five; see Nixon et al. 2002; Morrison et al. 2007; Roy et al. 2011). Therefore, we predict that splintrons occupy a large fraction of the entire spliceosomal introns in *G. intestinalis*, albeit the precise numbers of the two types of spliceosomal introns are currently uncertain. We also anticipate that the biochemical properties of spliceosomes in *G. intestinalis* are quite different from those of the well-characterized spliceosomes in yeast and human cells, as the *G. intestinalis* complex needs to carry out both *cis*- and *trans*-splicing with an almost equal efficiency.

The spliceosomes of the nematode *C. elegans* include several components that are exclusively required for *trans*-splicing (Denker et al. 2002). Similarly, spliceosomes of

trypanosomes were proposed to contain unique components that are responsible for “spliced leader” *trans*-splicing—spliceosomal transfer of a short non-coding RNA molecule (spliced leader) to the 5′ end of pre-mRNA molecules (Günzl 2010). Therefore, *Giardia* spliceosomes may have several components that are specific for splintrons *trans*-splicing. Otherwise, some (and, potentially, many) components that are highly conserved among the spliceosomes that predominantly excise canonical introns may be divergent or even replaced by unrelated proteins in the *G. intestinalis* complexes in order to work for both *cis*- and *trans*-splicing. For instance, the *G. intestinalis* spliceosomes may contain unique components dedicated for splintrons *trans*-splicing, such as chaperon-like components that facilitate the physical interaction between two pre-mRNAs transcribed from the two loci in the genome. This conjecture would be, to some extent, consistent with the result presented by Collins and Penny (2005) that many of the spliceosomal components conserved in the yeast and/or human complexes failed to be identified in the survey of the *G. intestinalis* genome.

Evolution of the OAD γ splintron with non-typical boundaries

OAD γ splintron possesses non-typical AT-AC boundaries (Fig. 2a; see also Roy et al. 2011). It is well known that some of the AT-AC introns are spliced by non-canonical U12-type spliceosomes. However, the *Giardia* OAD γ splintron was predicted to be spliced by canonical U2-type spliceosomes (Roy et al. 2011), allowing us to postulate that this splintron once bore the typical GT-AG boundaries. Most simply, two nucleotide substitutions—one at the 5' boundary (GT→AT) and the other at the 3' boundary (AG→AC)—are sufficient to yield the AT-AC boundaries. However, in this case, it is difficult to assume that the two independent substitutions occurred simultaneously. Instead, here, we propose a scenario of how this

“AT-AC” splintron was evolved from an ancestral splintron with GT-AG boundaries through neutral or semi-neutral steps, as schematically presented in Fig. 2b.

The 3' portions of the OAD γ splintron and those of the other introns/splintrons share apparent sequence homology (Fig. 2a). Intriguingly, the former retains a dinucleotide, AG, which is likely to be homologous to the 3' ends of other introns/splintrons (highlighted by closed arrowheads; Fig. 2a). It is also noteworthy that the 3' boundary of the OAD γ splintron is longer by three nucleotides than those of other introns/splintrons (highlighted by dots; Fig. 2a). Considering these aspects, we hypothesize an intermediate state where the splintron has the GT-AG boundaries and an insertion of CAC right after the 3' boundary of the splintron (*center* in Fig. 2b). Luukkonen and Séraphin (1997) experimentally displayed that spliceosomes can excise AT-

Fig. 2 Characterization of the AT-AC splintron in *Giardia intestinalis*. **a** An alignment of the conserved 5' and 3' portions of the five introns and four splintrons in *Giardia intestinalis*. The nucleotides shared among all introns/splintrons are shaded. The branch point adenine is highlighted with an asterisk. Dinucleotides at the 5' and 3' boundaries are shown in bold italics. All introns/splintrons, except that in the gene encoding the outer arm dynein γ subunit (OAD γ), possess the dinucleotide, AG, at the 3' end. The OAD γ splintron also retains the AG dinucleotide (highlighted by closed arrowheads) that is homologous to the dinucleotides at the 3' end of the other introns/splintrons, but it bears an extra trinucleotide CAC (highlighted by dots). Abbreviations: Fd ferredoxin; Rpl7a ribosomal large subunit protein 7a; URF unassigned reading frame; DLC dynein light chain; PNA4 26S proteasome non-ATPase subunit 4; HSP90 heat shock protein 90; OAD β outer arm dynein β subunit; OAD γ outer arm

dynein γ subunit. **b** A scenario for the transition from the splintron with the GT-AG boundaries to that with the AT-AC boundaries found in the OAD γ gene. Exon and splintron sequences are shown in black and gray, respectively. For the exon portion, the putative amino acid sequence is shown above the nucleotide sequence. The dinucleotides at the 5' and 3' ends of the splintron are highlighted in bold italics. We hypothesize that an ancestral form of the OAD γ splintron exists with GT-AG boundaries (*top*). Subsequently, a trinucleotide CAC, highlighted in white, was inserted to the downstream exon (*center*). Corresponding to the CAC insertion in the exon sequence, a single histidine was inserted into the protein in this intermediate state. Finally, a G-to-A substitution (shown with an open arrowhead) at the 5' boundary, coupled with a slide of the 3' boundary from AG to a neighboring AC (an arrow), occurred simultaneously (*bottom*)

AC introns more efficiently than AT-AG introns. Consequently, we propose that, in the last step, (1) a G-to-A substitution at the 5' boundary (open arrowhead) and (2) a slide of the 3' boundary from AG to a neighboring AC (arrow), occurred simultaneously to yield the AT-AC boundaries and to remove the histidine residue inserted in step 2.

G. intestinalis has been considered as a divergent eukaryote with respect to its cellular architecture, molecular machineries, and genome organization. Nevertheless, the finding of splintrons in the *G. intestinalis* genome implies that we have drastically underestimated the complexity of this organism. It is also important to stress that the knowledge from miscellaneous organisms for major research communities in molecular biology and biochemistry, such as *G. intestinalis*, can be significant for understanding the molecular mechanisms ubiquitously found in the Tree of Life.

Acknowledgments R.K. was a research fellow supported by JSPS for Young Scientists (210528). This work was supported by JSPS grants awarded to Y.I. (21370031) and T.H. (20570219).

Conflict of Interest None

References

- Adam RD (2000) Biology of *Giardia lamblia*. Clin Microbiol Rev 14:447–475
- Blumenthal T (2011) Split genes: another surprise from *Giardia*. Curr Biol 21:R162–163
- Collins L, Penny D (2005) Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol 22:1053–1066
- Denker JA, Zuckerman DM, Maroney PA, Nilsen TW (2002) New components of the spliced leader RNP required for nematode *trans*-splicing. Nature 417:667–670
- Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. Nature 440:623–630
- Fischer SEJ, Butler MD, Pan Q, Ruvkun G (2008) *Trans*-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. Nature 455:491–496
- Günzl A (2010) The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 9:1159–1170
- Hashimoto T, Sánchez LB, Shirakura T, Müller M, Hasegawa M (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. Proc Natl Acad Sci USA 95:6860–6865
- Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14
- Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T (2011a) Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated *trans*-splicing. Curr Biol 21:311–315
- Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T (2011b) Splintrons in *Giardia intestinalis*: Spliceosomal introns in a split form. Commun Integr Biol 4:454–456
- Lambowitz AM, Zimmerly S (2004) Mobile group II introns. Annu Rev Genet 38:1–35
- Lambowitz AM, Zimmerly S (2010) Group II introns: Mobile ribozyme that invade DNA. Cold Spring Harb Perspect Biol 3: a003616. doi:10.1101/cshperspect.a003616
- Luukkonen BGM, Séraphin B (1997) The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. EMBO J 16:779–792
- Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. Nature 440:41–45
- Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svård SG, Sogin ML (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 317:1921–1926
- Nageshan RK, Roy N, Hehl AB, Tatu U (2011) Post-transcriptional repair of a split heat shock protein 90 gene by mRNA *trans*-splicing. J Biol Chem 286:7116–7122
- Nixon JE, Wang A, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J (2002) A spliceosomal intron in *Giardia lamblia*. Proc Natl Acad Sci USA 99:3359–3361
- Roger AJ, Doolittle WF (2000) Why introns-in-pieces? Nature 364:289–290
- Roger AJ, Svård SG, Tovar J, Clark CG, Smith MW, Gillin FD, Sogin ML (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA 95:229–234
- Roy SW, Hudson AJ, Joseph J, Yee J, Russell AG (2011) Numerous fragmented spliceosomal introns, AT-AC splicing, and an unusual dynein gene expression pathway in *Giardia lamblia*. Mol Biol Evol. doi:10.1093/molbev/msr063
- Tovar J, León-Avila G, Sánchez LB, Sutak R, Tachezy J, van der Giezen M, Hernández M, Müller M, Lucocq JM (2003) Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. Nature 426:172–176