Molecular Biology and Evolution

Splicing and Transcription Differ between Spore and Intracellular Life Stages in the Parasitic Microsporidia

Erin E. Gill,†‡[1] Renny C.H. Lee,†§[1] Nicolas Corradi,[1] Cameron J. Grisdale,[1] Valerie O. Limpright,[1] Patrick J. Keeling,[1] and Naomi M. Fast*[,1]

[1]Biodiversity Research Centre and Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

†These authors contributed equally to this work.

‡Present address: Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

§Present address: Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

***Corresponding author:** E-mail: nfast@interchange.ubc.ca.

Associate editor: Jennifer Wernegreen

Abstract

Microsporidia are a diverse group of highly derived fungal relatives that are intracellular parasites of many animals. Both transcription and introns have been shown to be unusual in microsporidia: The complete genome of the human parasite *Encephalitozoon cuniculi* has only a few very short introns, and two distantly related microsporidian spores have been shown to harbor transcripts encoding several genes that overlap on different strands. However, microsporidia alternate between two life stages: the intracellular proliferative stage and the extracellular and largely metabolically dormant infectious spore. To date, most studies have focused on the spore. Here, we have compared transcription profiles for a number of genes from both life stages of microsporidia and found major differences in both the prevalence of overlapping transcription and splicing. Specifically, spore transcripts in *E. cuniculi* have longer 5′ untranslated regions, overlap more frequently with upstream genes, and have a significantly higher number of transcription initiation sites compared with intracellular transcripts from the same species. In addition, we demonstrate that splicing occurs exclusively in the intracellular stage and not in spore messenger RNAs (mRNAs) in both *E. cuniculi* and the distantly related *Antonospora locustae*. These differences between the microsporidian life stages raise questions about the functional importance of transcripts in the spore. We hypothesize that at least some transcripts in spores are a product of the cell's transition into a dormant state and that these unusual mRNAs could play a structural role rather than an informational one.

Key words: microsporidia, transcription, intron, genome reduction, *Encephalitozoon cuniculi*.

Research article

Introduction

The typical picture of gene expression in nuclear genomes involves transcription to produce a monocistronic messenger RNA (mRNA), the removal of its spliceosomal introns, and the addition of a 5′ cap and 3′ poly-A tail. These events are highly integrated activities within the eukaryotic cell, coordinated largely by RNA polymerase II (RNAP II) and multiple other proteins (Bentley 2005; Hagiwara and Nojima 2007; Moore and Proudfoot 2009). These events are also highly conserved in evolution: Within most eukaryotes, these core characteristics are generally shared by most genes, and although spliceosomal introns are variable in density between genomes, overall they are very common (although there is a long-standing debate about their origin, potential functions, and evolutionary significance) (Collins and Penny 2005; Rogozin et al. 2005; Koonin 2006).

Microsporidia are unicellular eukaryotes that are obligate intracellular parasites whose hosts include a wide range of animals. Although there is a great deal of variation among species, a generalized microsporidian life cycle proceeds as follows (reviewed in Vavra and Larsson 1999; Keeling et al. 2005; see supplementary fig. S1, Supplementary Material online). The spore is the extracellular infective stage of the parasite. When a spore germinates, a tightly coiled organelle, called the polar tube, everts piercing the spore wall. If the tube contacts the membrane of a host cell, it can be used to inject the spore contents into the host cell's cytoplasm or induce phagocytosis. Within the host, the parasite proliferates before producing more spores that eventually lyse the host cell. The spores may infect adjacent cells or they may be excreted, eventually leading to the infection of a new host.

Microsporidian cells and genomes are both highly derived and reduced, so they have been found to defy many conventions common to other eukaryotes. Transcription has been investigated in two distantly related microsporidia: *Antonospora locustae* and *Encephalitozoon cuniculi* (see supplementary fig. S2, Supplementary Material online). The *E. cuniculi* genome has been completely sequenced and at 2.9 Mbp is very near the lower limit for a eukaryote. The *E. cuniculi* genome encodes a mere ~2,000 protein-coding genes, separated by small intergenic regions, and even the proteins themselves are small in comparison with homologs in other eukaryotes (Katinka et al. 2001). *Antonospora locustae* is a grasshopper parasite with an estimated 5.4-Mb genome that has been partially sequenced (Slamovits et al. 2004; *A. locustae* Genome Project http://gmod.mbl.edu/perl/site/antonospora01?page=intro). Although its genome is roughly

twice the size of that of *E. cuniculi*, those regions that have been examined indicate gene densities similar to that seen in *E. cuniculi* (Slamovits et al. 2004).

Spores of both *E. cuniculi* and *A. locustae* have been shown to possess transcripts containing more than one gene (Williams et al. 2005; Corradi, Burri, and Keeling 2008). These multigene transcripts are distinctly different from prokaryotic operons because many of the genes are not full length or are encoded on opposite strands of the DNA, altogether suggesting that multiple functional proteins could not be produced from such transcripts (Williams et al. 2005; Corradi, Burri, and Keeling 2008). Instead, it is likely that these transcripts are a by-product of genome compaction and result from extremely small intergenic spaces; indeed, it has been shown that transcripts are more likely to contain fragments of adjacent genes when intergenic spaces are small (Corradi, Burri, and Keeling 2008).

Another interesting feature related to compaction is the rarity and nature of spliceosomal introns in these genomes. Only 16 introns in 15 genes were predicted in the *E. cuniculi* genome (Katinka et al. 2001; Vivares et al. 2002), and the genome of *Enterocytozoon bieneusi* has been predicted to encode no introns at all (Akiyoshi et al. 2009). The *E. cuniculi* introns are extremely short (23–52 nt) and are almost exclusively found in ribosomal protein-coding genes (RPGs). Although splicing has never been experimentally demonstrated in microsporidia, *E. cuniculi* and *A. locustae* possess many spliceosomal proteins and at least four small nuclear RNAs (Katinka et al. 2001; Davila Lopez et al. 2008). Spores are also known to contain proteins encoded by intron-containing genes (Brosson et al. 2006). Because these proteins could not be expressed without intron removal, it is reasonable to assume that mRNA splicing has to occur in *E. cuniculi*.

Both transcription and intron splicing are integral to gene expression, but unfortunately, neither has been studied in the actively growing and dividing life stage of microsporidia. In the present study, we have sought to systematically compare both processes between microsporidian spores and intracellular stages (which we collectively refer to as "meronts") and identified stark differences in the regulation of transcription and splicing. Splicing is active in meronts but inactive in spores, and both the abundance and the nature of overlapping transcripts also differ. Overall, spore transcripts are less like typical eukaryotic transcripts compared with those of meronts, leading us to speculate that many spore mRNAs may not function in gene expression.

Materials and Methods

Encephalitozoon cuniculi (genotype II) material was obtained as a generous gift from the Didier lab in the Faculty of Tropical Medicine at Tulane University (New Orleans, LA). Intracellular stages (see supplementary fig. S1, Supplementary Material online), which we cannot separate physically and we collectively refer to as meronts, were harvested from RK-13 (rabbit kidney) cell cultures 48 h after infection and stored in RNAlater (Ambion). *Antonospora locustae* spores were obtained from M&R Durango Inc. (http://www.goodbug.com/). To obtain *A. locustae* meronts, Semaspore Bait was obtained from Seeds of Change (http://www.seedsofchange.com) and fed to locust (*Schistocerca gregaria*) nymphs (kindly provided by Nancy Brard, Faculty of Land and Food Systems, University of British Columbia) during the second instar stage. The locusts were allowed to mature to adulthood and then sacrificed. Their fat bodies were removed and stored in RNAlater (Ambion). All *E. cuniculi* and *A. locustae* spores and meronts were disrupted by grinding in liquid nitrogen. RNA was extracted using Ambion's RNAqueous Kit.

To examine transcription, 31 genes were selected based on their synteny with *A. locustae* (Corradi, Burri, and Keeling 2008), and 5′ nested Rapid Amplification of cDNA Ends (RACE) was carried out using Ambion's FirstChoice RLM-RACE Kit. An annealing temperature of 55 °C was used for all reactions. Minus TAP controls were employed to ensure that there was no contaminating DNA. The 5′ nested primer had a 5-carboxyfluorescein attached to the 5′ end. A list of gene-specific primers used is provided in supplementary table S2, Supplementary Material online. All 5′ nested RACE products were analyzed by capillary electrophoresis (CE) to determine their sizes (Corradi, Gangaeva, and Keeling 2008), and some products were reconfirmed by cloning and sequencing (see below). Results were analyzed using Peak Scanner Software v1.0 (Applied Biosystems). Transcripts were considered to be the same length if the 5′ ends were within 10 bp or less of each other. Transcripts that were of sufficient length to be overlapping with upstream genes (indicated in fig. 1 and supplementary table S1, Supplementary Material online) were cloned, where possible, and sequenced for verification.

To examine splicing, 5′ nested RACE was performed using Ambion's FirstChoice RLM-RACE Kit. Annealing temperatures of 55–65 °C were used for all reactions. A list of primers is provided in supplementary table S3, Supplementary Material online. Products were resolved on agarose gels, resulting bands were excised, and products were gel isolated and cloned using Invitrogen's TOPO TA Kit. At least ten clones of each transcript were sequenced, with the exception of L5, L37, L39, S24, and S29 from *E. cuniculi* spores, where RACE products could not be obtained. Only one sequence was obtained for S30 from *E. cuniculi* spores.

Results

Transcription

Previous studies of *E. cuniculi* spores have demonstrated that a high proportion of loci produce multigene transcripts (Corradi, Burri, and Keeling 2008). To determine whether or not this pattern extends to other life stages, 5′ RACE products were obtained from both spores and meronts for 31 genes and their sizes determined by CE (supplementary table S1, Supplementary Material online).

Transcripts from spores generally had longer 5′ untranslated regions (UTRs) than those from meronts and a greater

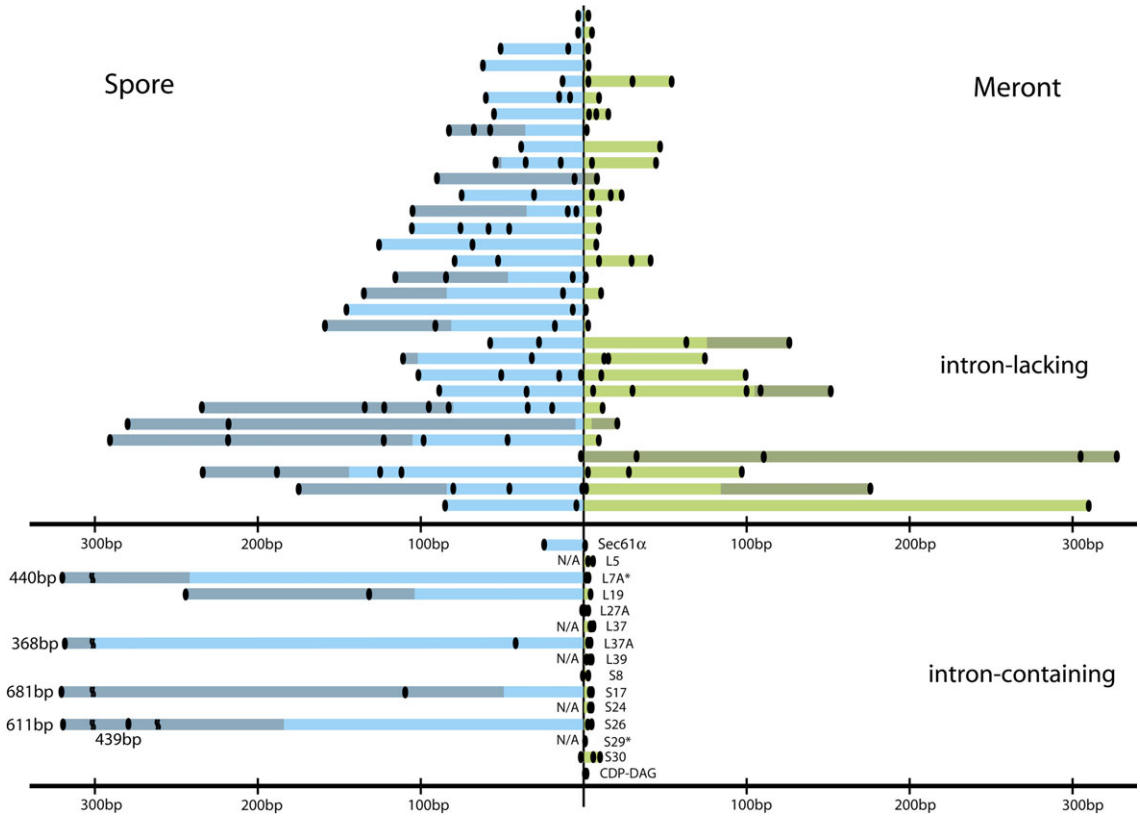

FIG. 1. Transcript 5′ UTR lengths and start sites in *Encephalitozoon cuniculi*. Bar sizes are proportional to transcript UTR lengths derived from intron-lacking and intron-containing genes (upper and lower parts of the figure, respectively). Transcription start sites are indicated by ovals. Portions of bars shaded in gray indicate regions of transcripts that overlap with upstream genes. Asterisks indicate transcripts from genes annotated with predicted introns that are not bona fide (see Results).

tendency to overlap with upstream genes, but the results were highly variable from gene to gene (fig. 1). In 48% of the genes examined, the size of the shortest RACE product was the same in spores and meronts (i.e., within 10 bp on CE; see Materials and Methods); in 42% of the genes, it was smaller in meronts, and in 10% of genes, it was smaller in spores. When comparing the size of the longest RACE product, in 65% of genes, it was longer in spores; in 16% of the genes, it was longer in meronts, and in 19% of the genes, they were identical. The longest 5′ RACE products were also examined to assess overlap with upstream genes. Overlapping transcription occurred in spores exclusively in 39% of the genes, in both spores and meronts in 13% of the genes, and in meronts in only 6% of the genes. Transcripts did not overlap in either stage in 42% of the genes. For more than half the genes examined, one transcription start site in spores was located within 10 bp of a meront start site. Three genes had two shared transcriptional start sites between spores and meronts (see supplementary table S1, Supplementary Material online). The average number of transcription start sites per gene examined was 2.48 in spores and 1.68 in meronts, which is consistent with observations of multiple start sites being common in yeast (Zhang and Dietrich 2005; Miura et al. 2006), mouse, and human (Carninci et al. 2006).

It was also observed that 78% of the meront and 59% of the spore transcripts initiated within 50 bp of an upstream TATA or TTTR sequence. The higher frequency in meronts suggests that this motif might be functionally significant to transcription initiation, although it differs from the predicted promoter elements of polar tube protein-coding genes proposed by Delbac et al. (2001). Consequently, although of interest, the significance of these potential motifs remains unknown, and the overall AT richness of the *E. cuniculi* genome (G + C = 51%; Katinka et al. 2001) is further ground for caution. (For a summary of the results, see supplementary table S4, Supplementary Material online)

Splicing

Transcripts were also examined for those genes annotated as containing introns. As all genes examined possess introns predicted to be located at the extreme 5′ termini of transcripts, 5′ RACE was employed as described above; however, for these genes, the RACE products were cloned and sequenced in order to verify the presence or absence of splicing. Transcripts produced in the spore stage were obtained for 10 of the 15 genes annotated as intron containing (see Materials and Methods). Each product was cloned, and at least ten clones of each product were sequenced. Surprisingly, however, in none of the cDNAs examined were the introns spliced (table 1). Given that spore transcripts are otherwise "mature" (i.e., polyadenylated and capped), the lack of evidence for splicing in this life stage

Table 1. Intron Splicing Patterns and Size Distribution in *Encephalitozoon cuniculi* and *Antonospora locustae*.

Gene	Spore	Meront	Predicted Size (bp)	Actual Size (bp)
*E.*L5	?	✓	38	*26*
*E.*L7A	X	X	29	N/A
*E.*L19	X	✓	31	31
*E.*L27A	X	✓	28	28
*E.*L37	?	✓	31	31
*E.*L37A	X	✓	52	*49*
*E.*L39	?	✓	32	32
*E.*S8	X	✓	31	31
*E.*S17	X	✓	23	23
*E.*S24	?	✓	44	*29*
*E.*S26	X	✓	42	*33*
*E.*S29	?	X	33	N/A
*E.*S30	X	✓	45	45
*E.*Sec61α	X	✓	38	*33*
*E.*CDP-DAG transferase	X/?	✓/✓	43/25	25/25
*A.*L37	X	✓	N/A	23
*A.*L37A	X	✓	N/A	26
*A.*L39	X	✓	N/A	20

NOTE.—The presence of spliced transcripts for each gene is indicated in the spore and the meront. *Encephalitozoon cuniculi* genes are prefixed with an "*E.*," and *A. locustae* genes are prefixed with an "A." The predicted size of each intron is listed with the actual size determined by 5′ RACE. Actual sizes of introns that are smaller than predicted are indicated in italics. Predicted introns that were not spliced in either life stage are shaded in gray. N/A, not applicable.

is unexpected and leads to the provocative possibility that splicing is inhibited in spores (see Discussion). The presence of multiple splicing factors and the protein products of spliced transcripts in spores, however, suggests that splicing must take place in *E. cuniculi*, so we also examined meront mRNAs. Interestingly, both spliced and unspliced transcripts could be identified in 13 of the 15 predicted intron-containing genes (where 1 gene contains two introns; table 1). L7A and S29 were the only genes from which spliced transcripts were not found. 5′ nested RACE products from these transcripts were examined via CE to further assess whether any spliced transcripts were present. Splicing appeared to be absent in both cases (supplementary fig. S3, Supplementary Material online), suggesting these might not contain bona fide introns, a result in line with other studies (Cornman et al. 2009; Lee, Gill, Roy, Fast, unpublished data). Of the 14 spliced introns, 6 were also shorter than previously annotated (table 1).

Meront transcripts for intron-containing genes also stood out because of their consistently short 5′ UTRs compared with those of the intron-lacking gene transcripts examined in the same life stage (fig. 1). Several lines of evidence rule out the possibility that this difference resulted from a technical artifact. First, all transcripts were amplified using the same method (5′ RACE), the difference only lies in the method used to differentiate the products (CE or cloning/sequencing). Second, there is no distinct difference between CE and cloning/sequencing results for transcripts analyzed from the spore stage. Third, RACE fragment lengths identified by CE were corroborated by cloning/sequencing for several genes (those with putative overlapping open reading frames, indicated in fig. 1). Taken together, those meront transcripts arising from intron-containing genes or RPGs appear to have very short 5′ UTRs. We cannot comment on the 5′ UTR length of all RPGs but of the three that lack introns that were examined in this study, L7a and S29 (which lack bona fide introns; see bottom of fig. 1) and L13, all have short 5′ UTRs (fig. 1 and supplementary table S1, Supplementary Material online).

To see if the observed lack of splicing in *E. cuniculi* spores is a general feature of microsporidia, transcripts from the distantly related *A. locustae* were also examined. No spliced transcripts of three intron-containing genes were recovered from *A. locustae* spores, suggesting that the inhibition of splicing in spores could be widespread within this group. A mixture of spliced and unspliced transcripts was recovered from the infected locust tissue, suggesting that splicing also does take place in the proliferative stage of *A. locustae*, as was observed in *E. cuniculi*. Given that the *A. locustae* material harvested from locust tissue was almost certainly a mixture of meronts and spores, in this particular case, it is impossible to tell whether the unspliced transcripts originated from spores, from meronts, or from both.

Discussion

Our results show that an apparent dichotomy exists between mRNAs from microsporidian spores and meronts. Transcripts from spores are unspliced and have longer 5′ UTRs that overlap more frequently with upstream genes. Transcripts from meronts are often spliced and have shorter 5′ UTRs that are less likely to overlap with upstream genes. Spore transcripts were also found to initiate at a higher number of locations compared with meront transcripts and less likely to start in the close proximity of a TATA or TTTR sequence. All these observations raise a number of questions about the function of mRNA in the spore.

Spore transcripts are known to be polyadenylated and possess 5′ caps (Flegel and Pasharawipas 1995; Williams et al. 2005), so the absence of splicing is at odds with the typical situation in eukaryotes, where transcription by RNAP II and RNA processing are tightly interwoven processes that generally occur at the same time (Moore and Proudfoot 2009). The C-terminal domain of RNAP II is able to interact with and coordinate multiple proteins that act to polyadenylate 5′ cap and splice mRNA as it is being transcribed, and multiple lines of evidence have shown that transcription by RNAP II and splicing are very closely linked (see, e.g., Kornblihtt et al. 2004; Bentley 2005; Hagiwara and Nojima 2007). Moreover, components of the U1 small nuclear ribonucleoprotein interact with RNAP II and also help regulate transcription initiation (Kwek et al. 2002; Das et al. 2007), and the *Saccharomyces* cap-binding complex proteins that act in the 5′ capping process also aid in the recruitment of splicing factors to nascent mRNA (Gornemann et al. 2005). The myriad of parallels between animal and yeast transcription and mRNA processing suggest that this system is at least partly conserved among opisthokonts, so there is no reason to assume a priori that microsporidia would differ drastically in these key

processes. Therefore, the degree to which spore transcripts look "unusual" is striking, particularly in comparison with transcripts observed in the meront stage. In contrast, it appears that transcription and splicing in the meront operate in a similar manner to what is observed in other eukaryotes.

It is possible that splicing is suppressed in spores for functional reasons. For example, certain *Saccharomyces cerevisiae* genes are efficiently spliced exclusively during sporulation, and removal of introns from these transcripts leads to protein production that is restricted to the sexual phase of the life cycle (Davis et al. 2000; Juneau et al. 2007). Environmental stresses have also been shown to affect splicing of certain functional classes of genes in yeast (Pleiss et al. 2007). In this latter case, it has been hypothesized that energy is saved during starvation by reducing translation through inhibition of splicing. This could also be the case for microsporidia, but given the general oddities of transcript structure in microsporidian spores, it is unclear whether such broad function can be assigned to splicing regulation in *E. cuniculi*.

Another interesting possibility is that at least some mRNAs in spores remain untranslated and serve a structural/functional role rather than an informational one. Identifying the location of the unspliced and unusual transcripts within the spore would provide important insights into this question as other eukaryotes are known to export unspliced products to the cytosol for degradation (Isken and Maquat 2007). Given that polyribosome structures have been observed in microsporidian spores (see, e.g., Vavra and Larsson 1999), if unspliced microsporidian spore transcripts were also exported to the cytosol, there would be a possibility that they may play a role in polyribosome formation, tethering the ribosomal subunits together. Finally, given the emerging picture of RNA transcription and processing in the spore, there is a distinct possibility that these overlapping and unspliced transcripts might not serve any function at all and that they could represent merely "junk" and unprocessed mRNA molecules. One could imagine that the stepwise shutting down of expression subsystems could result in various by-products: If, for example, the spliceosome and some elements of initiation regulation were shut down early, unspliced transcripts with long and variable 5′ ends could be the last mRNAs to be leftover. Upon germination, spore mRNAs could be degraded, and transcription would replace them with meront-specific transcripts ready to be translated. This scenario raises questions about the potential metabolic differences between the two life stages, as well as between microsporidia and better-characterized systems like yeast.

This study also raises the possibility that the expression of intron-containing genes and ribosomal protein genes is regulated differently in meronts than in other genes based on their very short 5′ UTRs in meronts. This feature is not exclusive to these two types of transcripts, however, so any firm conclusion along these lines is premature. However, splicing is known to regulate gene expression in other systems and could also be playing a role in microsporidia.

Overall, it is clear that although splicing and transcription are key processes that are tightly interconnected in eukaryotes, they operate very differently in the spore and meront life stages of microsporidia. Although overlapping transcription is still present in meronts, it is considerably less prevalent than it is in spore RNA, and spore RNAs are generally more unusual in all other characteristics we observed.

Supplementary Material

Tables S1–S4 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada to N.M.F. (262988) and the Canadian Institutes of Health Research to P.J.K. (MOP 84265). We thank Lisa Bowers, Elizabeth Didier, and Nancy Brard for generous gifts of material; Claudio Slamovits for helpful discussions; and Rachel Elves for technical assistance. P.J.K. is a fellow of the Canadian Institute for Advanced Research and a senior scholar of the Michael Smith Foundation for Health Research. The *A. locustae* Genome Project, Marine Biological Laboratory at Woods Hole, is funded by National Science Foundation award number 0135272.

References

Akiyoshi DE, Morrison HG, Lei S, et al. (11 co-authors). 2009. Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. *Plos Pathog.* 5:e1000261.

Bentley DL. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol.* 17:251–256.

Brosson D, Kuhn L, Delbac F, Garin J, Vivares CP, Texier C. 2006. Proteomic analysis of the eukaryotic parasite *Encephalitozoon cuniculi* (microsporidia): a reference map for proteins expressed in late sporogonial stages. *Proteomics* 6:3625–3635.

Carninci P, Sandelin A, Lenhard B, et al. (41 co-authors). 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet.* 38:626–635.

Collins L, Penny D. 2005. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol.* 22:1053–1066.

Cornman RS, Chen YP, Schatz MC, et al. (11 co-authors). 2009. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *Plos Pathog.* 5:e1000466.

Corradi N, Burri L, Keeling PJ. 2008. mRNA processing in *Antonospora locustae* spores. *Mol Genet Genomics.* 280:565–574.

Corradi N, Gangaeva A, Keeling PJ. 2008. Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics* 91:388–393.

Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. 2007. SR proteins function in coupling RNAPII transcription to pre-mRNA splicing. *Mol Cell.* 26:867–881.

Davila Lopez M, Rosenblad MA, Samuelsson T. 2008. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res.* 36:3001–3010.

Davis CA, Grate L, Spingola M, Ares MJ. 2000. Test of intron prediction reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.* 28:1700–1706.

Delbac F, Peuvel I, Metenier G, Peyretaillade E, Vivares CP. 2001. Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of *ptp1* and *ptp2* genes in three *Encephalitozoon* species. *Infect Immun.* 69: 1016–1024.

Flegel TW, Pasharawipas T. 1995. A proposal for typical meiosis in microsporidians. *Can J Microbiol.* 41:1–11.

Gornemann J, Kotovic KM, Hujer K, Neugebauer K. 2005. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell.* 19: 53–63.

Hagiwara M, Nojima T. 2007. Cross-talks between transcription and post-transcriptional events within a 'mRNA factory'. *J Biochem.* 142:11–15.

Isken O, Maquat LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* 21:1833–1856.

Juneau K, Palm C, Miranda M, Davis RW. 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci U S A.* 104:1522–1527.

Katinka MD, Duprat S, Cornillot E, et al. (17 co-authors). 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414:450–453.

Keeling PJ, Fast NM, Law JS, Williams BAP, Slamovits CH. 2005. Comparative genomics of microsporidia. *Folia Parasitol.* 52: 8–14.

Koonin EV. 2006. The origin of introns and their role in eukaryogenesis: a compromise to the introns-early versus introns-late debate? *Biol Direct.* 1:22.

Kornblihtt AR, De La Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *RNA* 10:1489–1498.

Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A. 2002. U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat Struct Biol.* 9:800–805.

Miura F, Kawaguchi N, Sese J, Toyoda A, Hattori M, Morishita S, Ito T. 2006. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci U S A.* 103:17846–17851.

Moore MJ, Proudfoot NJ. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136:688–700.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. 2007. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol Cell.* 27:928–937.

Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV. 2005. Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief Bioinform.* 6:118–134.

Slamovits CH, Fast NM, Law JS, Keeling PJ. 2004. Genome compaction and stability in microsporidian intracellular parasites. *Curr Biol.* 14:891–896.

Vavra J, Larsson JI. 1999. Structure of the microsporidia. In: Wittner M, Weiss LM, editors. The microsporidia and microsporidiosis. Washington (DC): ASM Press. p. 7–84.

Vivares CP, Gouy M, Thomarat F, Metenier G. 2002. Functional and evolutionary analysis of a eukaryotic parasitic genome. *Curr Opin Microbiol.* 5:499–505.

Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci U S A.* 102:10936–10941.

Zhang Z, Dietrich FS. 2005. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5′ SAGE. *Nucleic Acids Res.* 33:2838–2851.