RESEARCH PAPER

Fly 8:2, 120–125; April/May/June 2014; © 2014 Landes Bioscience

Higher frequency of intron loss from the promoter proximally paused genes of *Drosophila melanogaster*

Evidence consistent with delays in intron splicing as a selective force

Li Jiang, Xue-Nan Li, and Deng-Ke Niu*

MOE Key Laboratory for Biodiversity Science and Ecological Engineering and Beijing Key Laboratory of Gene Resource and Molecular Development; College of Life Sciences; Beijing Normal University; Beijing, PR China

Keywords: intron loss, *Drosophila*, promoter proximally paused genes, splicing, rapid expression.

Abbreviations: PPP, promoter proximally paused gene; UBP, unbound by polymerase II

Although intron losses have been widely reported, it is not clear whether they are neutral and therefore random or driven by positive selection. Intron transcription and splicing are time-consuming and can delay the expression of its host gene. For genes that must be activated quickly to respond to physiological or stress signals, intron delay may be deleterious. Promoter proximally paused (PPP) genes are a group of rapidly expressed genes. To respond quickly to activation signals, they generally initiate transcription competently but stall after synthesizing a short RNA. In this study, performed in *Drosophila melanogaster*, the PPP genes were found to have a significantly higher rate of intron loss than control genes. However, further analysis did not find more significant shrinkage of intron size in PPP genes. Referring to previous studies on the rates of transcription and splicing and to the time saved by deletion of the introns from mouse gene *Hes7*, it is here suggested that transcription delay is comparable to splicing delay only when the intron is 28.5 kb or larger, which is greater in size than 95% of vertebrate introns, 99.5% of *Drosophila* introns, and all the annotated introns of *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. Delays in intron splicing are probably a selective force, promoting intron loss from quickly expressed genes. In other genes, it may have been an exaptation during the emergency of developmental clocks.

©2014 Landes Bioscience. Do not distribute.

Introduction

It is well established that loss of spliceosomal introns is common in eukaryotic evolution.[1-10] However, it is not clear whether these intron losses occur by chance or are driven by selective forces. In principle, intron loss can be positively selected if the existence of that intron is deleterious to the host organism or if its detriments overwhelm its benefits. There are several possible disadvantages of having introns. The first is the accumulation of harmful mutations.[11,12] Most mutations are deleterious. Mutation is inevitable because DNA replication always involves some errors and some DNA damage. For these reasons, the existence of superfluous noncoding sequences is associated with mutational hazard. Consistent with this idea, introns and other noncoding sequences were found to be lost more frequently from genes and genomes with higher mutation rates.[13,14] The second disadvantage of introns is that they impose energy, spatial, and temporal costs on the host organisms. Replication, transcription, and splicing of introns all consume energy. In mammals and *Drosophila*, intron losses have been found to occur preferentially in highly expressed genes.[6,9] Researchers have suggested that this supports a reverse transcriptase model of intron loss.[15] The reverse transcriptase model is an entirely different model that attempt to explain the pattern of intron loss at mutation level. Similar to intron loss, intron size shrinkage could also save energy. In vertebrates, highly expressed genes have been found to have significantly shorter introns than weakly transcribed genes.[16,17] These findings indicate that selection for economy might have driven the shortening of introns. Considering the total expression level of all the cells in an individual animal, a gene expressed in a large organ (like the liver) cost much more energy to transcribe than a gene expressed in a small organ (like the hypothalamus) even if these 2 genes have similar levels of expression at the cell level. If the energy cost operates as an effective force in natural selection, genes

*Correspondence to: Deng-Ke Niu; Email: dkniu@bnu.edu.cn or dengkeniu@hotmail.com
Submitted: 04/28/2014; Revised: 05/21/2014; Accepted: 06/05/2014; Published Online: 07/10/2014
http://dx.doi.org/10.4161/fly.29489

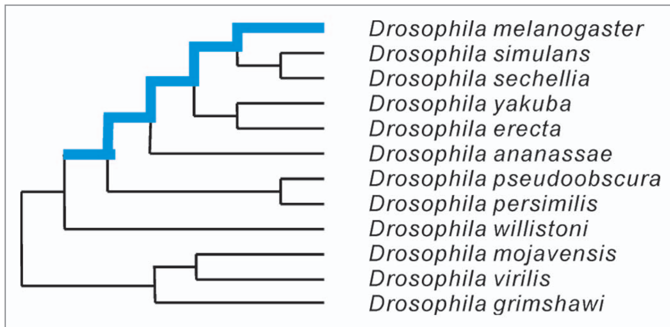

Figure 1. The phylogenetic tree of the 12 *Drosophila* species. The species *Drosophila melanogaster* and its recent ancestors whose intron losses were studied are shown in blue. The branch lengths are not scaled according to substitute rates.

specifically expressed in large organs would be more compact than genes expressed in small organs. However, Huang and Niu did not find any such difference in either humans or mice.[18] The existence of few or small introns in a nucleus might indicate space constraints, especially in organisms with very small nuclei.[19] Positive correlations between genome size and nucleus size have been found among different eukaryotic groups.[20-22] However, the causal relationship is too difficult to interpret. Replication, transcription, and splicing of introns also consume time. The loss of time caused by introns during DNA replication and gene expression has been called intron delay.[23,24] Intron delay is also difficult to examine because it is difficult to find a group of genes that are definitely under time constraints. In a survey including all major groups of eukaryotes, intron density was found to be positively correlated with generation time.[25] This analysis did not consider the phylogenetic dependence of species, which suggests that false-positive results may have been common.[26] Within the human genome, antisense genes that are believed to be rapidly transcribed were found to have significantly shorter introns than other genes.[27] In *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, and *Mus musculus*, genes that rapidly change their expression levels in response to stress were found to have significantly fewer and shorter introns.[28]

Severe stress, like elevated temperature, puts cells at risk. Rapid expression of responding genes, like heat shock proteins (Hsp), is crucial to cell survival.[29,30] To achieve the rapid response, the *Hsp* genes adopted an unusual strategy of induction. Their transcription is initiated without stress stimulation, but it stalls after synthesis of only a short RNA. Upon stress stimulation, the events, like establishing permissive chromatin, recruitment of transcription complex, and initiation of transcription, can be bypassed. These processes are like turning on a racecar's engine before a race. Recent high-throughput analyses have revealed that the promoter-proximal pausing (PPP) of RNA polymerase II is a widespread gene regulation mechanism in metazoans.[31,32] If the time required for the transcription or splicing of introns poses a significant burden, then the speeding up the expression of PPP genes through intron loss should be under positive selection. In *Drosophila melanogaster*, Gilchrist et al. characterized 5529 PPP genes and 6959 genes that were unbound by polymerase II (UBP) in the cell lines that they analyzed.[33] The large sample size of PPP genes in *D. melanogaster* provides an opportunity to determine whether intron delay could act as a selective force promoting intron loss and shrinkage.

Results and Discussion

The present state of introns is determined by both the recent forces that shaped intron evolution and the ancestral state of the introns. A gene that is intron-poor at present might have suffered intron loss driven by some selective forces, like intron delay, or it may have inherited its intron-poor state from an ancestor. For this reason, the present work focused on the evolutionary changes in the introns in *D. melanogaster* rather than on their present sizes and numbers. With 2 reference branches, 186 cases of intron losses were determined in *D. melanogaster* and its recent ancestors using Dollo parsimony (**Fig. 1**).

Frequency of intron loss in promoter proximally paused genes

Among the 5529 PPP genes identified in *D. melanogaster*, 87 genes were found to have lost one or more introns. However, only 44 of the 6959 UBP genes had lost introns. The PPP genes showed an intron-loss frequency 2.5 times that of the UBP genes. Chi-square test showed the difference to be statistically significant ($P = 5 \times 10^{-7}$). In this way, PPP genes were shown to be more likely to lose their introns than UBP genes. This indicates that the intron delay hinders the quick expression of PPP genes, placing the PPP genes under a selective force to lose their introns. However, it is also possible that the PPP genes may have more introns and so they have a higher frequency of intron loss for statistical reasons. To address this issue, the ratio of intron loss, the number of lost introns to the total number of extant introns and lost introns, were compared between PPP genes and UBP genes. As shown in **Table 1**, the ratio of intron loss is significantly higher in PPP genes than in UBP genes (0.67% vs. 0.33%, $P = 5 \times 10^{-6}$). Besides lost introns and conserved introns, there are some ambiguous introns in non-conserved regions. They might be either introns descended from ancient ancestors or introns recently gained. For accuracy, we also tested the difference between PPP genes and UBP genes by confining the extant introns to the introns conserved between *D. melanogaster* and the closest reference species *Drosophila willistoni*. The conclusion that PPP genes have a higher ratio of intron loss still holds (**Table 1**).

In principle, a lower frequency of intron gain could also support the intron delay as a selective force. However, identification of intron gains is risky. Intron gains identified with insufficiently stringent criteria might actually be intron losses.[2,5] Using stringent criteria, only 14 intron gains were identified in PPP and UBP genes. This is a too small of a sample to produce statistically convincing results. Furthermore, clear source sequences could not be identified for any of the 14 intron gains from either the NCBI nucleotide collection or

©2014 Landes Bioscience. Do not distribute.

Table 1. Preferential loss of introns from PPP genes in *Drosophila melanogaster*[a]

	Number of lost introns	Number of all extant introns	Number of conserved introns
PPP[b] genes	119	17 577	3452
UBP[c] genes	58	17 780	2654

[a]Pearson Chi-square test showed that PPP genes have a higher frequency of intron loss than UBP genes either when the all the extant introns were used as the control ($P = 5 \times 10^{-6}$) or when only the conserved introns were used as the control ($P = 0.0059$). [b]PPP, promoter proximally paused. [c]UBP, unbound by polymerase II.

the reference sequences of *Drosophila* repetitive elements from Repbase,[34] even using a very relaxed parameter, E-values < 10^{-5} in the BLAST. According to the criteria suggested by Logsdon et al.,[35] these 14 intron gains could just be regarded as putative cases.

Mutational hazards are unlikely the selective force

According to the mutational hazard hypothesis, genes that have a higher mutation rate are more likely to lose their introns.[14] Here, the synonymous substitution rates (d_s) of the genes that lost introns and those containing only conserved introns were compared. The coding sequence alignments between *D. melanogaster* and *D. willistoni* were first filtered using Gblocks with its default parameters to discard unreliable alignments.[36] The values of d_s were calculated using codeml as included in the PAML package.[37] Unlike previous works that have supported the mutational hazard hypothesis in *Arabidopsis*,[14] the *D. melanogaster* genes that lost introns showed significantly lower d_s values than genes with only conserved introns (Wilcoxon rank sum test, $P = 0.0054$). Five more stringent sets of parameters were tested in Gblocks by changing the maximum number of contiguous non-conserved positions and the minimum length of a block. In some cases, the difference in d_s between the 2 groups of genes became insignificant (Wilcoxon rank sum test, $P > 0.05$). The genes that lost introns did not have significantly higher d_s values than genes with only conserved introns. In this way, the intron losses detected in *D. melanogaster* were unlikely to have been driven by any selective force to reduce mutational hazards.

Intron-splicing delay *vs.* intron transcription delay

In principle, the existence of introns could impose extra time costs on the host organisms during 3 processes: replication, transcription, and splicing. It is currently unknown which if any of these processes is slowed down by introns to any extent significant enough to affect the evolution of those introns. The existence of a large intron would be unlikely to delay chromosomal replication in eukaryotes because eukaryotes generally initiate the replication of a chromosome at more than one origin. There are also many dormant origins that are not normally used but could be activated when required.[38] The preferential loss of introns from PPP genes indicates that the time required for transcription and splicing of introns is main cause of intron delay. If the time required for transcription of introns is a significant burden, then the introns of PPP genes might not only be preferentially lost but would also shrink significantly in size. However, if the time required for splicing of introns is the only significant temporal burden, then PPP genes might decrease intron numbers, but they would not decrease in size. For this reason, changes in intron sizes in *D. melanogaster* were evaluated.

Among the 12 *Drosophila* species, 2727 groups of orthologous introns were detected. Among these, 1390 groups were distributed among 984 PPP genes and 689 groups were distributed among 464 UBP genes. Using the maximum likelihood method, the ancestral size of conserved introns of the 12 *Drosophila* species was estimated. As shown in **Table 2**, the sizes of the introns in both PPP and UBP genes decreased over the course of evolution from the common ancestor of the 12 *Drosophila* species. However, the introns of PPP genes did not shrink more significantly than those of UBP genes ($P = 0.348$, **Table 2**). Meanwhile, no significant differences were detected between the present intron sizes of PPP and UBP genes (Wilcoxon rank sum test, $P = 0.71$). The introns of *D. melanogaster* therefore did not shrink in response to the requirements of rapid gene expression. This indicates that the significant temporal cost of introns must involve splicing rather than transcription.

The present observation is consistent with previous experimental results.[39,40] The inflammatory genes induced by tumor necrosis factor in mouse cells can be defined as early, intermediate, and late genes according to the appearance of their mature mRNAs. Recently, it is revealed that the intermediate and late mRNA productions are mainly due to slowness of splicing, rather than that of transcription initiation or elongation.[40] The gene *Hes7* has a pattern of oscillatory expression during mouse development. Experimental deletion of its introns was found to reduce the delay of its expression by 19 min.[39] The gene *Hes7* (GenBank transcript ID: NM_033041.4) showed 3

Table 2. Changes in the intron size of PPP and UBP genes[a]

	Median	IQR	*P*
PPP genes			
Present size (bp)	62	58–68	8×10^{-13b}
Ancestral size (bp)	64	60–69	
UBP genes			
Present size (bp)	62	58–69	0.0001[b]
Ancestral size (bp)	63	60–69	
Present size/ancestral size			
Changes in PPP genes	0.974	0.904–1.05	0.348[c]
Changes in UBP genes	0.981	0.904–1.05	

[a]Change in the size of an intron was measured by the ratio of its present size to its ancestral size. PPP, promoter proximally paused; UBP, unbound by polymerase II; IQR, interquartile range. Because the data are not normally distributed, their median values and the IQR are presented here. [b]*P* values were calculated using the Wilcoxon rank sum test. [c]*P* values were calculated using the Wilcoxon signed rank test.

©2014 Landes Bioscience. Do not distribute.

Table 3. Intron sizes and duration of transcription[a]

Organisms	No. of introns	No. of introns > 28.5 kb	Intron sizes (bp)		Time required to transcribe introns ≤ Q75 size (min)
			Median	IQR[b]	
S. cerevisiae	313	0	156	91–413	≤0.11
A. thaliana	113 149	0	98	85–156	≤0.04
C. elegans	104 416	1	63	48–321	≤0.08
D. melanogaster	41 140	191	69	60–265	≤0.07
G. gallus	146 269	1817	774	334–1761	≤0.46
M. musculus	174 994	4516	1285	447–3117	≤0.82
H. sapiens	192 482	6207	1406	441–3745	≤0.99

[a]As determined by Singh and Padgett, the time required for transcription is comparable to that required for splicing only when intron size is 28.5 kb or greater.[43] [b]IQR, interquartile range.

introns with a total size of 1,843 bp. The rates of splicing and polymerase II elongation determined by different experimental methods varied within a broad range, 30 s to 10 min, for splicing of an intron.[41-45] They varied within 0.8 to 8 kb/min for transcription of intron sequences.[40,43,46-50] The estimated time required for transcription of the 3 introns of gene *Hes7* ranged from 25 s to 2.3 min, which is much less than the detected 19 min. If it is assumed that the splicing of different introns of the same pre-mRNA molecule may be performed simultaneously, then the 19 min delay could be explained by splicing of the 3 introns at a rate of about 7.5 min per intron, as reported by Singh and Padgett.[43] Singh and Padgett also estimated the rate of polymerase II elongation at about 3.8 kb/min.[43] With these data, it can be deduced that intron transcription delay is comparable to intron splicing delay only for introns 28.5 kb or larger. For a global estimation of the relative significance of intron splicing delay and intron transcription delay, intron sizes were surveyed in 7 eukaryotic model organisms (**Table 3**). In *S. cerevisiae* and *A. thaliana*, no annotated introns are even half 28.5 kb. In *Caenorhabditis elegans*, only one intron is larger than 28.5 kb. Less than 0.5% of the introns in *D. melanogaster* are larger than 28.5 kb. Vertebrates such as *Gallus gallus*, *M. musculus*, and *Homo sapiens* have more large introns, but introns of 28.5 kb are still rare, <5% in all the 3 vertebrates. The time required to transcribe introns of Q75 (the third quartile) size was calculated. For the 75% introns ≤ Q75 in each of the 7 genomes, transcription was found to take less than 1 min (**Table 3**), much less than that the time spent in splicing. In summary, splicing delay was much more significant than transcription delay except in the rare cases of large introns.

Sometimes mutations and even pathogens can have a net beneficial effect on the host genome. For example, the transposed element CORE-SINE not only replicates itself in host genome for its own survival but has also contributed to novel functional elements in its host genomes.[51] The intron delay, if significant, might be utilized in evolution to construct temporal rhythms or timing mechanisms during development.[23,24] The idea has been supported strongly by the experimental deletion of introns from gene *Hes7*, which changed oscillatory expression into steady expression and caused severe segmentation defects.[39] The exaptation of intron delay in developmental clock showed that intron delay plays some role in natural selection, that its effect is not negligible.

Materials and Methods

Orthologs and phylogenetic tree

The genome sequences and annotations of 12 *Drosophila* species, *D. melanogaster* (Release 5.48), *Drosophila ananassae* (Release 1.3), *Drosophila erecta* (Release 1.3), *Drosophila grimshawi* (Release 1.3), *Drosophila mojavensis* (Release 1.3), *Drosophila persimilis* (Release 1.3), *Drosophila pseudoobscura* (Release 2.30), *Drosophila sechellia* (Release 1.3), *Drosophila simulans* (Release 1.4); *Drosophila virilis* (Release 1.2), *D. willistoni* (Release 1.3), and *Drosophila yakuba* (Release 1.3) and the orthologous relationship among the coding genes of these 12 species (fb_2012_06) were obtained from Flybase.[52] The orthologs were filtered further using SynMap with its recommended settings.[53] Some 4273 orthologous groups of coding genes were retained. The protein sequences of each orthologous group were aligned using ClustalW (version 2.1).[54] Using the aligned sequences, the consensus phylogenetic tree was generated using the Protdist flow operated in PHYLIP 3.6.[55] The topology of the phylogenetic tree obtained here (**Fig. 1**) was identical to that constructed in a previous study.[56]

The genome annotations of *S. cerevisiae* (R64-1-1), *C. elegans* (WBcel235), *G. gallus* (Galgal4), *M. musculus* (GRCm38), *H. sapiens* (GRCh37), and *A. thaliana* (TAIR10) were downloaded from Ensembl (ftp://ftp.ensembl.org/pub/release-75/gtf/) and EnsemblPlants (ftp://ftp.ensemblgenomes.org/pub/plants/current/gtf/).

Identification of conserved introns and intron losses

First, the coding sequences were aligned by consulting the alignment of protein sequences, and then the introns were mapped on the alignments. Only intron positions in well-aligned regions were retained for analysis. A well-aligned region across an intron position was defined as a stretch of aligned nucleotides at least 45 bp long at each side of the intron position with an identity not lower than the overall identity of the whole coding sequence alignment. In addition, the intron positions were filtered further by discarding those that have large flanking gaps (≥9 bp gaps

©2014 Landes Bioscience. Do not distribute.

within 45 bp at either side) and introns shorter than 20 bp. Conserved introns were defined as introns present at the same position in all orthologous genes of the 12 *Drosophila* species. The ancestral size of the conserved introns was estimated using the method of maximum likelihood integrated in the R package APE with its default parameters for continuous traits.[57]

At the discordant positions, the Dollo parsimony method integrated in PHYLIP 3.6 were used to identify intron loss and gain.[55] Because of previous observations of the high rate of intron loss relative to intron gain and the high rate of false-positive intron gains,[2,3,6,8] stringent criteria were here adopted for the definition of intron gain events. Intron gain should be supported by clear absence of the intron at the discordant position in at least 4 outgroup branches. The gained introns were also confirmed as actual introns rather than simple insertions using transcriptome data. The splicing of the novel introns in *D. melanogaster* were validated using RNA-seq data (SRR1145619, SRR1197414, SRR1197477), which were retrieved from the Sequence Read Archive of NCBI and mapped to the genome using TopHat v2.0.5 with its default parameters.[58] The intron losses and gains reported by of Farlow et al.[59] were integrated into our data after manual examination of the non-overlapping results. In total, 186 intron losses (177 losses in PPP or UBP genes) and 15 intron gains (14 gains in PPP or UBP genes) were detected in *D. melanogaster* and its recent ancestor.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Financial Disclosure

The authors certify that they have no commercial associations (e.g., consultancies, stock ownership, equity interest, patent-licensing arrangements) that may pose a conflict of interests in connection with the submitted article. All funding sources supporting the work and all institutional or corporate affiliations of the authors have been acknowledged in the article.

Acknowledgments

We thank Karen Adelman for sharing the data on the promoter proximally paused genes of *Drosophila melanogaster*. We would also like to thank Yu-Fei Yang, Tao Zhu, and the anonymous referees for their useful comments. This work was supported by the National Natural Science Foundation of China (grant numbers 31121003 and 31371283). We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

1. Roy SW, Gilbert W. Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci U S A 2005; 102:5773-8; PMID:15827119; http://dx.doi.org/10.1073/pnas.0500383102
2. Roy SW, Penny D. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. Mol Biol Evol 2006; 23:2259-62; PMID:16943250; http://dx.doi.org/10.1093/molbev/msl098
3. Coulombe-Huntington J, Majewski J. Intron loss and gain in *Drosophila*. Mol Biol Evol 2007; 24:2842-50; PMID:17965454; http://dx.doi.org/10.1093/molbev/msm235
4. Roy SW, Penny D. Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. Mol Biol Evol 2007; 24:171-81; PMID:17065597; http://dx.doi.org/10.1093/molbev/msl159
5. Zhang LY, Yang YF, Niu DK. Evaluation of models of the mechanisms underlying intron loss and gain in *Aspergillus* fungi. J Mol Evol 2010; 71:364-73; PMID:20862581; http://dx.doi.org/10.1007/s00239-010-9391-6
6. Yenerall P, Krupa B, Zhou L. Mechanisms of intron gain and loss in *Drosophila*. BMC Evol Biol 2011; 11:364; PMID:22182367; http://dx.doi.org/10.1186/1471-2148-11-364
7. Csuros M, Rogozin IB, Koonin EV. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS Comput Biol 2011; 7:e1002150; PMID:21935348; http://dx.doi.org/10.1371/journal.pcbi.1002150
8. van Schendel R, Tijsterman M. Microhomology-mediated intron loss during metazoan evolution. Genome Biol Evol 2013; 5:1212-9; PMID:23737326; http://dx.doi.org/10.1093/gbe/evt088
9. Coulombe-Huntington J, Majewski J. Characterization of intron loss events in mammals. Genome Res 2007; 17:23-32; PMID:17108319; http://dx.doi.org/10.1101/gr.5703406
10. Zhu T, Niu DK. Mechanisms of intron loss and gain in the fission yeast *Schizosaccharomyces*. PLoS One 2013; 8:e61683; PMID:23613904; http://dx.doi.org/10.1371/journal.pone.0061683
11. Lynch M. The origins of eukaryotic gene structure. Mol Biol Evol 2006; 23:450-68; PMID:16280547; http://dx.doi.org/10.1093/molbev/msj050
12. Lynch M. The Origins of Genome Architecture. Sunderland: Sinauer Associates, Inc., 2007.
13. Lynch M, Koskella B, Schaack S. Mutation pressure and the evolution of organelle genomic architecture. Science 2006; 311:1727-30; PMID:16556832; http://dx.doi.org/10.1126/science.1118884
14. Yang YF, Zhu T, Niu DK. Association of intron loss with high mutation rate in *Arabidopsis*: implications for genome size evolution. Genome Biol Evol 2013; 5:723-33; PMID:23516254; http://dx.doi.org/10.1093/gbe/evt043
15. Fink GR. Pseudogenes in yeast? Cell 1987; 49:5-6; PMID:3549000; http://dx.doi.org/10.1016/0092-8674(87)90746-X
16. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. Selection for short introns in highly expressed genes. Nat Genet 2002; 31:415-8; PMID:12134150
17. Li SW, Feng L, Niu DK. Selection for the miniaturization of highly expressed genes. Biochem Biophys Res Commun 2007; 360:586-92; PMID:17610841; http://dx.doi.org/10.1016/j.bbrc.2007.06.085
18. Huang YF, Niu DK. Evidence against the energetic cost hypothesis for the short introns in highly expressed genes. BMC Evol Biol 2008; 8:154; PMID:18492248; http://dx.doi.org/10.1186/1471-2148-8-154
19. Prachumwat A, DeVincentis L, Palopoli MF. Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. Genetics 2004; 166:1585-90; PMID:15082572; http://dx.doi.org/10.1534/genetics.166.3.1585
20. Andrews CB, Mackenzie SA, Gregory TR. Genome size and wing parameters in passerine birds. Proc Biol Sci 2009; 276:55-61; PMID:18765340; http://dx.doi.org/10.1098/rspb.2008.1012
21. Jovtchev G, Schubert V, Meister A, Barow M, Schubert I. Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. Cytogenet Genome Res 2006; 114:77-82; PMID:16717454; http://dx.doi.org/10.1159/000091932
22. Cavalier-Smith T. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann Bot 2005; 95:147-75; PMID:15596464; http://dx.doi.org/10.1093/aob/mci010
23. Gubb D. Intron-delay and the precision of expression of homoeotic gene products in *Drosophila*. Dev Genet 1986; 7:119-31; http://dx.doi.org/10.1002/dvg.1020070302
24. Swinburne IA, Silver PA. Intron delays and transcriptional timing during development. Dev Cell 2008; 14:324-30; PMID:18331713; http://dx.doi.org/10.1016/j.devcel.2008.02.002
25. Jeffares DC, Mourier T, Penny D. The biology of intron gain and loss. Trends Genet 2006; 22:16-22; PMID:16290250; http://dx.doi.org/10.1016/j.tig.2005.10.006
26. Whitney KD, Garland T Jr. Did genetic drift drive increases in genome complexity? PLoS Genet 2010; 6:e1001080; PMID:20865118; http://dx.doi.org/10.1371/journal.pgen.1001080
27. Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. Human antisense genes have unusually short introns: evidence for selection for rapid transcription. Trends Genet 2005; 21:203-7; PMID:15797613; http://dx.doi.org/10.1016/j.tig.2005.02.003
28. Jeffares DC, Penkett CJ, Bähler J. Rapidly regulated genes are intron poor. Trends Genet 2008; 24:375-8; PMID:18586348; http://dx.doi.org/10.1016/j.tig.2008.05.006
29. de Nadal E, Ammerer G, Posas F. Controlling gene expression in response to stress. Nat Rev Genet 2011; 12:833-45; PMID:22048664
30. Silver JT, Noble EG. Regulation of survival gene hsp70. Cell Stress Chaperones 2012; 17:1-9; PMID:21874533; http://dx.doi.org/10.1007/s12192-011-0290-6
31. Nechaev S, Adelman K. Promoter-proximal Pol II: when stalling speeds things up. Cell Cycle 2008; 7:1539-44; PMID:18469524; http://dx.doi.org/10.4161/cc.7.11.6006
32. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet 2012; 13:720-31; PMID:22986266; http://dx.doi.org/10.1038/nrg3293

©2014 Landes Bioscience. Do not distribute.

33. Gilchrist DA, Fromm G, dos Santos G, Pham LN, McDaniel IE, Burkholder A, Fargo DC, Adelman K. Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. Genes Dev 2012; 26:933-44; PMID:22549956; http://dx.doi.org/10.1101/gad.187781.112
34. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 2005; 110:462-7; PMID:16093699; http://dx.doi.org/10.1159/000084979
35. Logsdon JM Jr., Stoltzfus A, Doolittle WF. Molecular evolution: recent cases of spliceosomal intron gain? Curr Biol 1998; 8:R560-3; PMID:9707398; http://dx.doi.org/10.1016/S0960-9822(07)00361-2
36. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000; 17:540-52; PMID:10742046; http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334
37. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007; 24:1586-91; PMID:17483113; http://dx.doi.org/10.1093/molbev/msm088
38. Blow JJ, Ge XQ, Jackson DA. How dormant origins promote complete genome replication. Trends Biochem Sci 2011; 36:405-14; PMID:21641805; http://dx.doi.org/10.1016/j.tibs.2011.05.002
39. Takashima Y, Ohtsuka T, González A, Miyachi H, Kageyama R. Intronic delay is essential for oscillatory expression in the segmentation clock. Proc Natl Acad Sci U S A 2011; 108:3300-5; PMID:21300886; http://dx.doi.org/10.1073/pnas.1014418108
40. Hao S, Baltimore D. RNA splicing regulates the temporal order of TNF-induced gene expression. Proc Natl Acad Sci U S A 2013; 110:11934-9; PMID:23812748; http://dx.doi.org/10.1073/pnas.1309990110
41. Beyer AL, Osheim YN. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev 1988; 2:754-65; PMID:3138163; http://dx.doi.org/10.1101/gad.2.6.754
42. Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert M-C, Koper M, Karkusiewicz I, Mariconti L, Tollervey D, et al. RiboSys, a high-resolution, quantitative approach to measure the *in vivo* kinetics of pre-mRNA splicing and 3′-end processing in *Saccharomyces cerevisiae*. RNA 2010; 16:2570-80; PMID:20974745; http://dx.doi.org/10.1261/rna.2162610
43. Singh J, Padgett RA. Rates of *in situ* transcription and splicing in large human genes. Nat Struct Mol Biol 2009; 16:1128-33; PMID:19820712; http://dx.doi.org/10.1038/nsmb.1666
44. Huranová M, Ivani I, Benda A, Poser I, Brody Y, Hof M, Shav-Tal Y, Neugebauer KM, Stanek D. The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. J Cell Biol 2010; 191:75-86; PMID:20921136; http://dx.doi.org/10.1083/jcb.201004030
45. Carrillo Oesterreich F, Bieberstein N, Neugebauer KM. Pause locally, splice globally. Trends Cell Biol 2011; 21:328-35; PMID:21530266; http://dx.doi.org/10.1016/j.tcb.2011.03.002
46. Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, et al. A wave of nascent transcription on activated human genes. Proc Natl Acad Sci U S A 2009; 106:18357-61; PMID:19826084; http://dx.doi.org/10.1073/pnas.0902573106
47. Mason PB, Struhl K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol Cell 2005; 17:831-40; PMID:15780939; http://dx.doi.org/10.1016/j.molcel.2005.02.017
48. Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. *In vivo* dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 2007; 14:796-806; PMID:17676063; http://dx.doi.org/10.1038/nsmb1280
49. Zenklusen D, Larson DR, Singer RH. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 2008; 15:1263-71; PMID:19011635; http://dx.doi.org/10.1038/nsmb.1514
50. Boireau S, Maiuri P, Basyuk E, de la Mata M, Knezevich A, Pradet-Balade B, Bäcker V, Kornblihtt A, Marcello A, Bertrand E. The transcriptional cycle of HIV-1 in real-time and live cells. J Cell Biol 2007; 179:291-304; PMID:17954611; http://dx.doi.org/10.1083/jcb.200706018
51. Santangelo AM, de Souza FSJ, Franchini LF, Bumaschny VF, Low MJ, Rubinstein M. Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet 2007; 3:1813-26; PMID:17922573; http://dx.doi.org/10.1371/journal.pgen.0030166
52. McQuilton P, St Pierre SE, Thurmond J, Consortium F; FlyBase Consortium. FlyBase 101--the basics of navigating FlyBase. Nucleic Acids Res 2012; 40:D706-14; PMID:22127867; http://dx.doi.org/10.1093/nar/gkr1030
53. Lyons E, Pedersen B, Kane J, Freeling M. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. Tropical Plant Biol 2008; 1:181-90; http://dx.doi.org/10.1007/s12042-008-9017-y
54. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and clustal X version 2.0. Bioinformatics 2007; 23:2947-8; PMID:17846036; http://dx.doi.org/10.1093/bioinformatics/btm404
55. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 1989; 5:164-6
56. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al.; Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 2007; 450:203-18; PMID:17994087; http://dx.doi.org/10.1038/nature06341
57. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 2004; 20:289-90; PMID:14734327; http://dx.doi.org/10.1093/bioinformatics/btg412
58. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013; 14:R36; PMID:23618408; http://dx.doi.org/10.1186/gb-2013-14-4-r36
59. Farlow A, Meduri E, Dolezal M, Hua L, Schlötterer C. Nonsense-mediated decay enables intron gain in *Drosophila*. PLoS Genet 2010; 6:e1000819; PMID:20107520; http://dx.doi.org/10.1371/journal.pgen.1000819

©2014 Landes Bioscience. Do not distribute.