

Fall 2009

Evolution of Gene Structure in Multicellular Eukaryotes

Maria Hester

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/inquiry>

 Part of the [Cell Biology Commons](#), and the [Molecular Biology Commons](#)

Recommended Citation

Hester, Maria (2009) "Evolution of Gene Structure in Multicellular Eukaryotes," *Inquiry: The University of Arkansas Undergraduate Research Journal*: Vol. 10 , Article 10.

Available at: <http://scholarworks.uark.edu/inquiry/vol10/iss1/10>

This Article is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Inquiry: The University of Arkansas Undergraduate Research Journal by an authorized editor of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

EVOLUTION OF GENE STRUCTURE IN MULTICELLULAR EUKARYOTES

By Maria Hester

Department of Biological Sciences

Faculty Mentor: Douglas Rhoads

Department of Biological Sciences

Abstract

We investigated the patterns of intron conservation in eukaryotes for five different genes. The genes examined were ribosomal proteins L8, S14 and S17, along with elongation factor 2B and triose phosphate isomerase. Intron conservation for S14, S17, and triose phosphate isomerase was determined for 32 species representing the major branches of multicellular eukaryotes. For 25 conserved introns 16 were phase 0, five were phase 1, and four were phase 2. Triose phosphate isomerase had five of nine conserved introns shared between plants and animals, where S14 had one of nine and S17 had one of seven. However, there were two plant S14 introns that could be found in single soil-living organisms from the animal branch, suggestive of horizontal transfer.

Supplementary Figures are available at <http://biscweb.uark.edu/drhoads/pubs/IntronsSupFigs.pdf>

Introduction

Introns are prevalent in all eukaryotes whose genomes have been fully sequenced, though the densities and sizes of the introns vary greatly (Carmel 2007a). Despite the widespread prevalence of introns, little is known about the origins of introns and what role, if any, they played in gene and genome evolution in eukaryotes (Carmel 2007b). Competing theories have been proposed to address these issues, including the "introns early" and "introns late" theories of intron influence on eukaryote evolution (de Souza 1996). These theories attempt to answer the question of whether introns predate eukaryotes or have been acquired more recently during eukaryotic evolution (Logsdon 2004).

Genes in eukaryotes are not just linear sequences that code for proteins. The gene is recognized as the region that is transcribed to make an initial transcript. The initial transcript is processed in the nucleus to add a poly A tail, and specific regions are precisely removed by a protein-RNA complex called the spliceosome. Those portions that are removed are called introns. The remaining portions that are joined together to form the mature mRNA are called exons. As many as 80% of intron positions are conserved across vastly different eukaryote lineages. The other 20% of introns can either be explained by novel insertions or by precise deletions (Coulombe-Huntington 2007).

The conservation and non-conservation of intron position within genes is where the battle between the two competing theories lies. The "introns early" theory is based on the

notion that introns were present before eukaryotes arose from prokaryotes, perhaps even present in the original genome at the origin of life in the protogenote. Since the divergence of prokaryotes and eukaryotes, prokaryotes and some single-celled eukaryotes have streamlined their genome through loss from intron-rich ancestral genes that predate eukaryotic cells (Logsdon 2004). Past studies have shown that early eukaryotic ancestors were relatively rich in introns (Schwartz 2008). According to the exon shuffling theory, introns were essential components of gene evolution as they can increase recombination of exons as gene fragments (Long 1995). Therefore, exons were used as building blocks in the evolution of eukaryotic organisms to create novel genes. Introns persist in eukaryotes as a result of their role in genomic evolution. This theory proposes that introns exist today because they were used historically as a quicker way to form the diversity of genes that are now present (de Souza 1996). The introns early theory holds that introns, through exon shuffling, facilitated the origin of new proteins through recombination. Therefore, introns were an intrinsic element of the first protein encoding genes (Basu 2008).

Exon shuffling would be an extremely effective method to create a large diversity of protein structures (de Souza 1996). The absence of introns in present day prokaryotes is attributed to the complete loss of introns through "genome streamlining" (Basu 2008). In a study using a large database of eukaryotic genes, it was found that at least 19% of the exons present were the result of exon shuffling, and these exons were often found in the conserved regions of ancient genes that are homologous to prokaryotic genes (Long 1995). This data supports the theory that introns were indeed present in prokaryotes at some point in evolutionary history.

In contrast, the introns late theory holds that prokaryotes never possessed introns; introns and the spliceosome emerged during early eukaryotic evolution (Basu 2008). The current distribution of introns can be explained by processes of both gain and loss (Logsdon 2004). Introns are present in certain organisms because molecular processes introduce them faster than the counterselection, or evolutionary drift, mechanisms can remove them. Therefore, they have limited significance in eukaryote evolution and little, if any, function (de Souza 1996). The introns late theory concludes that there have been recent instances of intron insertion into eukaryote genomes. These introns were inserted into preexisting genes at some point in evolution (Long 1995).

A recent study shows that some eukaryotic lineages may still be gaining introns, while others appear only to be losing them. A comparison of the human, dog, rat, and mouse genomes identifies over a hundred instances of intron loss, but no evidence of intron gain, over the last one hundred million years since these organisms diverged (Coulombe-Huntington 2007). However, a genome-wide study of *Drosophila* shows that there has been recent intron gain within the *Drosophila* lineage, with the latest gain occurring around ten million years ago (Coulombe-Huntington 2007). Therefore, the rates for intron gain and loss differ between specific eukaryotic lines.

Intron phase, or the position of the intron relative to codons in the gene, is theorized to be important in showing which of these theories is most valid. Intron phase can be either 0, 1, or 2 (Long, 1995). According to the intron late hypothesis, introns in each phase should have equal distributions, because addition of the intron to a pre-existing gene would have no effect on the coding function of the gene. Conversely, the introns early theory would suppose that a non-random distribution is more likely, favoring phase 0, because exon shuffling would favor introns in phase 0. If most introns were phase 0, then exon shuffling would not alter the protein sequence encoded by the exon (Long, 1995). Phase 0 introns occur between two codons. If the introns early theory is correct, introns should occur in this phase because ancient exons would have been independent units, and phase 0 introns would not have interfered with coding structure of exons after shuffling. A survey of a subset of 296 genes identified 1496 introns where 55% were phase 0 introns, 24% phase 1 introns, and 21% phase 2 introns. This nonrandom distribution of intron phase supports the introns early theory (Long 1995).

A recent study has shown that there may be three distinct modes of evolution of intron/exon structure (Carmel 2007b). The first mode is the primary, balanced mode that operates in all lineages. In this mode, intron gain and loss are strongly and positively correlated. The second mode is one of an elevated rate of intron loss. This mode is prevalent only in certain lineages, such as insects and fungi. The third mode highlights an elevated rate of intron gain, and is seen in the deep, ancient branches of the tree of life. This mode indicates that explosions of intron gain happened at key points in eukaryote evolution, such as the origin of animals. These different modes showcase the fact that it is difficult to determine the main theory that describes how introns arose in evolution, because there are many possible explanations for the current intron distribution.

Genomic data show that there have been approximately twice as many intron losses as intron gains in the past 1.5 billion years of eukaryote evolution (Carmel 2007b). However, because the specific lineages differ widely in the rates of loss and gain, it may be that different genes have significantly different evolution rates for intron gain and loss.

While it may never be known when introns arose, further investigation of intron position and genome evolution will

help to pinpoint better the modes of intron evolution. Many factors may contribute to both the intron's presence and role. Previous work has shown that numerous introns have their genomic position conserved between different taxa, including distantly related taxa such as animals and plants (Carmel 2007a). This would seem to imply that introns occurred very early in evolutionary history. However, there is another possible explanation for this occurrence. These conserved intron positions may occur because of proto-splice sites, which are constrained nucleotide sequences where introns are preferentially inserted (Logsdon 2004). Therefore, these conserved introns may not all have arisen early, but may have been gained later in evolution because of the preference for introns to be accumulated at these sites.

Previous work in this laboratory has catalogued intron position and phase in ribosomal protein S14 (rpS14) for multiple organisms (Nicks 2007). This gene was chosen because it has an important role in ribosome function, and because it is highly conserved in eukaryotes and prokaryotes. Since rpS14 has an essential role in the ribosome for all living organisms, it must be an ancient gene dating back to the earliest origins of life. The organisms are chosen to represent a wide range of the eukaryotic lineages and to represent all of the major branches of the eukaryotic tree of life. The expanding number of sequenced higher eukaryote genomes provides researchers with an opportunity to add additional organisms and to examine additional genes.

To investigate further the introns early theory vs. the introns late theory, several different genes were selected for analysis of intron position and phase from a wide range of eukaryotes. The eukaryotes were chosen to represent all major branches of the tree of life. The first stage of this project extended previous work on rpS17 (Nicks 2007). This protein was chosen because it is a ribosomal protein that is less conserved than rpS14. Whereas rpS14 is a highly conserved and functional component of prokaryote and eukaryote ribosomes, rpS17 has no recognizable homolog in the prokaryote ribosome. Therefore, rpS17 appears to have been added to the ribosome after divergence of eukaryotes and prokaryotes. If rpS17 is less conserved and "newer" than rpS14, it would represent a younger gene and thus might show a different pattern of intron conservation.

Based on initial comparisons of conserved introns in rpS14 and rpS17, we surveyed a few other highly conserved genes for presence of conserved introns in select taxa. Genes examined were ribosomal protein L8 (rpL8), elongation factor 2B (EF2B), and triose phosphate isomerase (TPI). We examined rpL8 because it is conserved in eukaryotes and prokaryotes, but the protein is a component of the large subunit of the protein as opposed to the small subunit. EF2B was included because it is a highly conserved gene that is used during transcription. TPI is a highly conserved and essential component of glycolysis in eukaryotes and prokaryotes. Thus, EF2B and TPI are essential, highly conserved genes that are

not components of the ribosome, and therefore might provide a different perspective on intron evolution.

Analysis of intron position conservation from these diverse genes should allow inferences about the history of introns in these genes and how the evolutionary signal differs among them. Overlaying the results of intron conservation on the eukaryotic tree of life identifies patterns where intron placement corresponds with the phylogenetic relationships. Highly conserved vs. variable intron positions provide information about the role of introns in eukaryote evolution, and contribute significant information to the debate over introns early or introns late.

Materials and Methods

Online genome browsers were used to obtain the protein sequences by using BLAST searches. Where possible, genome browsers were also used to locate intron positions and identify each intron phase. Table 1 provides the specific websites used for each eukaryote. Some of the genome browsers did not provide complete gene structure information or the encoded protein. If this was the case, the sequence was analyzed further using EditSeq and SeqMan software in DNASTar (ver 6.0). EditSeq was used to manipulate and annotate the DNA and protein sequences. SeqMan was used to translate the

Table 1. Eukaryote Organism code and Source of DNA Sequences

Code	Species	Name	Source
Aam	<i>Aureococcus anophagefferens</i>	brown algae	http://genome.jgi-psf.org [No Introns]
Ath	<i>Arabidopsis thaliana</i>	plant	http://www.arabidopsis.org
Bfl	<i>Branchiostoma floridae</i>	lancalet	http://genome.jgi-psf.org
Cel	<i>Caenorhabditis elegans</i>	worm	http://www.sanger.ac.uk
Csp	<i>Capitella species 1</i>	worm	http://genome.jgi-psf.org
Cin	<i>Ciona intestinalis</i>	Sea Squirt	http://genome.jgi-psf.org
Cre	<i>Chlamydomonas reinhardtii</i>	algae	http://genome.jgi-psf.org
Ddi	<i>Dicystidium discoidium</i>	slime mold	http://genome.jgi-psf.org
Dme	<i>Drosophila melanogaster</i>	fruitfly	http://dictybase.org/
Dpu	<i>Daphnia pulex</i>	waterflea	http://genome.ucsc.edu
Fru	<i>Fugu rubripes</i>	puffer fish	http://genome.jgi-psf.org
Hma	<i>Hydra magnipapillata</i>	hydra	http://hydrasome.metazome.net
Hro	<i>Helobdella robusta</i>	leech	http://genome.jgi-psf.org
Lgi	<i>Lotia gigonata</i>	snail	http://genome.jgi-psf.org
Mam	human and Chinese hamster	mammal	http://genome.ucsc.edu
Mbr	<i>Monosiga brevicollis</i>	flagellate	http://genome.jgi-psf.org
Ncr	<i>Neurospora crassa</i>	fungus	(Tyler, 1990 #311)
Ngr	<i>Naegleria gruberi</i>	amoeba[flagellate]	http://genome.jgi-psf.org [No Introns]
Nve	<i>Nematostella vectensis</i>	anemone	http://genome.jgi-psf.org
Olu	<i>Ostreococcus lucimarinus</i>	green algae	http://genome.jgi-psf.org
Pbl	<i>Physcomyces blakesleeanus</i>	zygomycete	http://genome.jgi-psf.org
Pch	<i>Phaeochytrium chrysosporium</i>	basidiomycete	http://genome.jgi-psf.org
Ppa	<i>Physcomitrella patens</i>	moss	http://genome.jgi-psf.org
Ppi	<i>Populus trichocarpa</i>	poplar tree	http://genome.jgi-psf.org
Ptr	<i>Phaeodactylum tricornutum</i>	diatom	http://genome.jgi-psf.org
Sea	<i>Stomoxys calcitrans</i>	stable fly	Genbank Accession AF119387
Smo	<i>Setlaginella moellendorffii</i>	primitive plant	http://genome.jgi-psf.org
Spa	<i>Strongylocentrotus purpuratus</i>	sea urchin	http://www.ncbi.nlm.nih.gov and
Tad	<i>Trichoplax adhaerens</i>	placozoa	http://genome.jgi-psf.org
Tca	<i>Tribolium castaneum</i>	flour beetle	http://www.beetibase.org
Tps	<i>Thalassiosira pseudonana</i>	diatom	http://genome.jgi-psf.org (missing 5' exon)
Tth	<i>Tetrahymena thermophila</i>	ciliate	http://www.ciliate.org
Vca	<i>Volvox carteri</i>	chlorophyte algae	http://genome.jgi-psf.org
Xla	<i>Xenopus laevis</i>	frog	http://genome.ucsc.edu
Xtr	<i>Xenopus tropicalis</i>	frog	http://genome.ucsc.edu

genomic sequence in all three reading frames, view the aligned translations and determine specific intron boundaries.

After all of the intron positions and phases were located for all of the eukaryotes, the program MegAlign was used to align protein sequences utilizing the Clustal W method. Placement and conservation of intron position was correlated with an evolutionary tree based on currently accepted models for the tree for eukaryotes using an accepted evolutionary placement of the eukaryotes (Spiegel and Silberman, personal communication).

Results

For each analyzed gene, a visual alignment of all protein sequences for each gene was assembled. Each protein sequence was annotated for all introns that interrupted coding sequences and noted their phase. From the alignment, conserved introns found in at least two different organisms were identified. Conserved introns were defined as those that were present in the same phase for the homologous amino acid of the protein in more than one organism. These conserved introns were then analyzed further with respect to a currently accepted evolutionary tree for eukaryotes.

Analysis of rpS14 revealed there are nine conserved introns among the eukaryotes investigated. The rpS14 sequences and intron positions are presented in Figure S1. Five of the conserved introns are in phase 0, one is in phase 1, and two are in phase 2. There are several other nonconserved introns from all phases present in this gene. Based on analysis of conserved introns in our eukaryotic tree, intron f was present only in insects. Introns b, d, and h were only found in animal lineages.

In rpS17, seven conserved introns were identified among the thirty eukaryotes studied (Figure S2). For these conserved introns, six are phase 0 and one is phase 2. Based on examination of the conserved intron table, it is clear that introns l and n are only in plant lineages, while introns k, m, and p are found exclusively in animal lineages. Intron o is only found in fungal lineages.

From the analysis of rpS14 and rpS17 it was difficult to discern any consistent pattern. The evolutionary signals from rpS14 and from rpS17 appear to be quite different. Whereas rpS14 has a mixture of phases in its conserved introns, rpS17 has almost exclusively phase 0 conserved introns. It did not appear that many introns were present in all the branches of the eukaryotic tree, which would suggest an introns recent pattern. To further investigate intron patterns, three additional widely conserved eukaryotic genes were surveyed. The genes selected were rpL8, EF2, and TPI. Ribosomal protein rpL8 is conserved between eukaryotes and prokaryotes, but is a component of the large subunit. EF2 is an elongation factor that has been used in other evolutionary studies. TPI has also been used in evolutionary studies and is an essential component of glycolysis in prokaryotes and eukaryotes. Intron patterns were examined for each of these genes from Homo sapiens, Caenorhabditis elegans, and Drosophila melanogaster. These organisms were chosen because they represent critical branches of the eukaryotic tree of life. The pattern of introns in EF2 (Figure S3) only revealed one conserved intron for the three organisms investigated while rpL8 (Figure S4) showed two conserved introns. TPI also had two conserved introns within these three organisms. TPI was chosen for further investigation because, unlike rpL8, it is not a ribosomal protein and therefore could present a different perspective on intron patterns. Intron data for TPI were then assembled from the other organisms from the evolutionary tree.

Investigation of TPI revealed a total of nine conserved introns (Figure S5). Five of the conserved introns were phase 0, three were phase 1, and one was phase 2. Introns q and y are found in plants. The conserved intron found only in animal lineages is u.

Overall, there were 25 conserved introns in rpS14, rpS17, and TPI. Out of these 25 introns, 16 were in phase 0, five were in phase 1, and four were in phase 2. RpS17 favored phase 0 introns heavily, while the other two genes had a mixture of phases in their conserved introns.

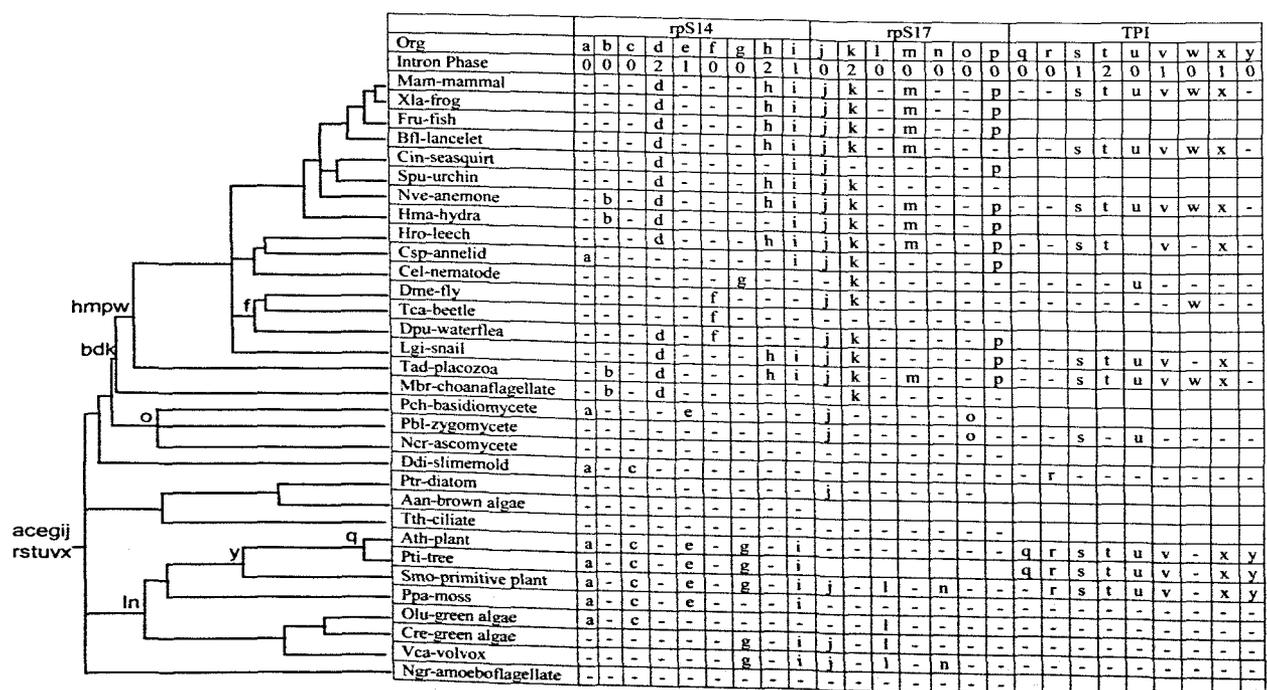
Data were used to construct a table of all the conserved introns arranged by gene, and aligned with the evolutionary tree (Figure 1). This table allows visualization of when the conserved introns arose in evolutionary history, which conserved introns were present exclusively in particular branches, and which were present in different branches. The main deep branch was that between animals and plants. In each gene there were conserved introns shared between plants and animals. In rpS14 introns a, g, and i were shared by both plant and animal lineages. Intron g was only found in nematodes and plants. Intron i was the most widely conserved intron for rpS14 among the organisms used. In rpS17 there was only one conserved intron that was shared by plants and animals – this was intron j. Out of the nine conserved TPI introns, 5 of these were shared in plants and animals. These were introns s, t, u, v, and x; all of these introns were present in approximately the same set of organisms.

Discussion

There are a number of conclusions that can be inferred from these data about trends in intron conservation. The eukaryotic tree of life (Figure 1) was used to identify the most ancient introns to test the relevance of the introns early theory. There were twelve ancient introns in this analysis of rpS14, rpS17, and TPI. Out of these twelve introns, six were in phase 0, five were in phase 1, and one was in phase 2. The introns early theory holds that introns were present before prokaryotes diverged from eukaryotes, and that these ancient conserved introns should be in phase 0. The distribution of these ancient introns does not support the introns early theory, because there were nearly as many phase 1 introns as there were phase 0 introns. While these introns have clearly been present since the beginning of eukaryotes, this intron distribution does not support the strict definition of introns early. Another point to be made about these twelve ancient introns is that only one of these, intron j, comes from rpS17; the youngest of these three genes is present only in eukaryotes. TPI and rpS14 are present in both prokaryotes and eukaryotes, so it is logical that the older genes would contain the most ancient introns.

There were several conserved introns from each gene that were present only in animals. This may be because these introns were simply lost in other branches of the evolutionary tree. There is another possible reason for this; studies have shown that it is possible that widespread intron gain happened only during short periods of eukaryotic evolution that

Figure 1. Phylogenetic analysis of conserved introns for rpS14, rpS17 and TPI. Conserved introns are as indicated in the protein alignments (Supplementary Information). For each conserved intron the intron phase is indicated. Where introns likely first appeared in the tree are indicated on the branches. Three letter organism codes are based on Table 1.



coincided with major evolutionary innovations (Basu 2008). A potential example of this would be an extensive intron gain during the origin of animals (Carmel 2007b); this may be why these introns are conserved only in animal lineages. In rpS14, intron h is only present in animals. However, out of the sixteen animals sequenced, seven of them have lost this intron, which shows the genomes of different animals evolve differently.

The gene for rpS17 has several other conserved introns which appeared only in certain lineages. For example, intron l appears only in plant lineages. This is consistent with the fact that different lineages of eukaryotes lose and gain introns at rates which differ from other lineages (Carmel 2007b). The alternatives are that intron l was present early and lost very early in the branch that gave rise to animals, or that intron l was acquired very early in the branch that gave rise to plants. There are examples of conserved introns appearing only in plant lineages in all three genes investigated in this study. This fact supports the introns late theory, because it is more probable that the plant branch gained these introns than that all other branches lost it.

A peculiarity of introns c, e, g, and r was observed upon further investigation of introns that are almost exclusively in plants. These introns are all present in several plants, but occur in only one member of the animal lineage. For example, in rpS14, intron g is present in plants and in *C. elegans*, a nematode. Nematodes live in soil and therefore are frequently exposed to plants, so it is possible that this plant intron was transferred to this one specific eukaryote, and was incorporated into its rpS14 gene. Introns e and r are both present in plants and in *D. discoideum*, a slimemold. Since slimemolds are soil organisms, a similar scenario could be postulated, with the slimemold naturally acquiring a plant intron through frequent contact in its environment.

Another example of a probable intron gain event is intron f, which appears only in three arthropods: two insects and a crustacean. This intron appears to support the introns late theory, as it was gained only in this lineage within the arthropods during eukaryotic evolution. However, these three arthropods represent very old radiations estimated at 666 ± 58 million years ago (Pisani 2004). Since their divergence, they have maintained these introns with no apparent changes.

Given all the data collected in this study, it is possible to conclude that introns were present in the earliest eukaryotes, and that there have been more recent intron acquisitions. The present study, therefore, suggests the introns late theory is better supported by these data. For example, intron i in rpS14 is ancient and is conserved throughout all the branches of the evolutionary tree. However, it is not present in every single eukaryote we examined, which means that it has been lost in some organisms during evolution. This type of ancient intron supports the fact that introns have been present since the beginning of eukaryotes, and subsequently have been lost by a few eukaryotes along the way. All of the ancient introns that are conserved throughout the tree of life support

this conclusion. A study recently concluded that all events of excessive intron gain were ancient (Carmel 2007b), so it appears that most eukaryotes acquired many of their introns early in their evolution. Then, there are the introns that are conserved only in certain lineages, such as introns q and y in TPI, which only appear in plants. These introns suggest a more recent acquisition of specific introns to specific lineages of eukaryotes. Remarkably, TPI, unlike rpS14 or rpS17, had a preponderance of introns shared in all the major lineages. For rpS14 and rpS17 there are a few introns conserved across eukaryotes, but most introns appear to be acquired in particular lineages. Conversely, TPI has 5 introns shared between plants and animals, two to three introns (q, r and y, although r may be an exception) specific to plants, and only one intron (w) specific to animals. For rpS14 and rpS17, only introns i and j are clearly conserved between plants and animals. Intron a is in a highly polymorphic region of rpS14, while introns c, e, g, are primarily in plants with single occurrences in one animal and may represent horizontal transfer rather than phylogenetic conservation.

Recent studies have expanded on the two extreme themes of introns early vs. late. Current focus is on the full spectrum of ancient, stable introns to recently gained introns (Omilian 2008). However, it is important to continue investigation of introns because their loss and gain is a slow process compared to other genetic characteristics, which allows intron positions to retain a vast amount of information about genome structure and deep evolutionary history (Irmia 2008). Analysis of single cell eukaryotes representing the base of the eukaryote tree would be helpful except that these organisms have apparently streamlined their genomes through removal of most introns. For example, the genome for the budding yeast *Saccharomyces cerevisiae* lacks introns in most genes. A survey of rpS14 genes in yeasts and fungi shows few introns and very little conservation of position (Figure S7). Therefore, we appear restricted to the higher branches of the eukaryotic tree; the multicellular eukaryotes. Further work should be pursued, continuing the analysis of genes for their intron patterns, as more data may clarify the question of the evolutionary history of gene structure.

References

- Basu, M. et al (2008) Evolutionary Dynamics of Introns in Plastid-Derived Genes in Plants: Saturation Nearly Reached but Slow Intron Gain Continues. *Molecular Biology and Evolution* 25(1):111-119.
- Carmel, L. et al (2007) Evolutionarily conserved genes preferentially accumulate introns. *Genome Research* 17(7): 1045-1050.
- Carmel, L. et al (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Research* 17(7) 1034-1044.
- Coulombe-Huntington, J. and Majewski, J. (2007) Intron Loss and Gain in *Drosophila*. *Molecular Biology and Evolution* 24(12): 2842-2850.

- DeSouza, S. et al (1996) Introns and gene evolution. *Genes to Cells* 1: 493-505.
- Irmia, M. and Roy, S. Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucleic Acids Research* 36(5): 1703-1712.
- Logsdon, J. (2004) Worm genomes hold the smoking guns of intron gain. *PNAS* 101(31): 11195-11196.
- Long, M. et al (1995) Intron Phase Correlations and the Evolution of the Intron/Exon Structure of Genes. *PNAS* 92(26): 12495-12499.
- Nicks, Shannon. (2007) Analysis of Gene Structure Evolution for Ribosomal Protein S14 in Eukaryotes. *University of Arkansas Honors Thesis*.
- Omilian, A. et al (2008) Intron Presence-Absence Polymorphisms in *Daphnia*. *Molecular Biology and Evolution* 25(10): 2129-2139.
- Pisani, D. et al (2004) The colonization of land animals: molecular phylogeny and divergence times among arthropods. *BMC Biol.* 2(1).
- Schwartz, S. et al (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Research* 18: 88-103.

Mentor Comments:

According to mentor Douglas D. Rhoads, Maria Hester took on a challenging line of research being pursued in his laboratory and developed her own line of investigation with considerable success. He clarifies as follows:

The work presented in Maria Hester's manuscript was the basis for her honors research in my laboratory. The pursuit of intron evolution patterns in highly conserved genes has always been a great interest to me. However, for many years the numbers of genomes available was so spartan as to not give us anything more than a few examples. With the rapid proliferation of eukaryotic genome projects representing a wide diversity of organisms, we can start to ask questions that were impossible only a few years ago. Maria's work builds upon work that I first began on rpS14. I have continued to mine new S14 genes, as they become available. Another student, Shannon Nicks started working with rpS17, which was the basis of her honors thesis. When Maria chose to pickup this project she greatly expanded the number and diversity of organisms for rpS17 and then did some great investigative work to identify TPI as an alternative. This sort of bioinformatics project has not been attractive to many students because it is so much frustrating computer work. You have to learn different genome browsers, and many of the genome sites don't readily provide the answers we need without detailed further analysis. The project is to try to learn whether the evolution of gene structure with respect to intron placement in conserved genes is a constant. There are competing theories on the timing of the origin of introns and the role of introns in gene evolution. With an ever increasing

diversity of genomes available, we can begin to address some of these puzzles. Maria chose to examine the evolution of gene structure in 23 widely different eukaryotic taxa using 5 different genes of varying levels of conservation. Previous data was for only 2 genes from selected taxa. Maria used genome browsers and other sequence analysis tools to deduce gene structures for ribosomal protein S17 (rpS17), triose phosphate isomerase (TPI), elongation factor 2B (eF2B), ribosomal protein L8 (rpL8), and RNA polymerase subunit 2B (RP2B). Her survey identified rpS17 and TPI as having sufficient numbers of conserved introns for detailed analysis in the entire group of organisms. Over the course of one year, she analyzed these two genes from the different available genomes, mapped introns for placement and codon phase. Her manuscript compares "intron evolution" patterns for rpS17 and TPI to my prior data for rpS14. The three genes tell very different stories, and that is the major conclusion of Maria's honors thesis. It is truly some excellent meticulous work. She has been very dedicated and has tackled some rather daunting tasks in wading through whole genome information, different genome browsers, and difficult user interfaces.