[Frontiers in Bioscience 10, 2412-2419, September 1, 2005]

INTRON POSITION CONSERVATION ACROSS EUKARYOTIC LINEAGES IN TUBULIN GENES

Bagavathi S. Perumal [1], Kishore R. Sakharkar [2], Vincent Tak Kwong Chow [3], Pandjassarame Kangueane [1], and Meena Kishore Sakharkar [1]

[1] NCSV, MAE, Nanyang Technological University, Singapore 639798, [2] BioInformatics Institute, 30 Biopolis Street, #07-01, Matrix, Singapore 138671, [3]Department of Microbiology, National University of Singapore, Singapore 119260

TABLE OF CONTENTS

1. ABSTRACT

A compilation of intron positions obtained from a large number of eukaryotic genomes across orthologous tubulins is explored for molecular evolution. Comparison of intron positions for 41 α, 80 β, and 30 γ tubulin genomic sequences indicates that the putative ancestral tubulin gene contained at least 19, 33, and 52 intron positions distributed at different sites in the coding regions for α, β, and γ tubulins, respectively. Many intron positions are old and are conserved across different eukaryotic lineages and intron distribution patterns are consistent with 'introns-early' hypothesis.

2. INTRODUCTION

The split gene organization of eukaryotic genes has lead to an extensive debate on the origin of introns. Two alternative theories, the 'introns early' or 'the exon theory of genes' and the 'introns late' or the 'insertional theory of introns', argue the presence or absence of introns in primordial genes. The arguments have focused on the positions of introns with respect to protein and gene structures in both views. The 'early' theory proposes that introns are ancient parts of genes that were lost in prokaryotes (1-2), whereas the 'late' model states that introns are derived from transposable elements and were added recently into eukaryotic genes (3-4). A recently proposed view is the synthetic theory of intron evolution that has merged the concepts from both introns early and intron late views (5). Perler *et al* (6) in 1980 characterized the first event of intron loss by examining the rat and chicken insulin genes supporting the intron early theory. Bagavathi *et al* catalogued the intron positions in actin genes from a wide range of eukaryotes and confirmed that introns could be ancestral in origin (7). The intron late view, derives support from the studies by Dibb and Newman (8), that analyzed the tubulin intron positions from eukaryotes at a time only a few sequences were available (28 tubulin genes) and have concluded that introns were gained at proto-splice sites and as a consequence tubulin introns are not ancient. Some

evidences for possible recent intron gains include the evolution of the gene structures of DEAD helicase family genes (9) and for the globin genes of *Chironomos* (10), where a variable intron distribution was observed. Coghlan and Wolfe (11) demonstrated that 122 introns have been gained recently in *Caenorhabditis* genes and that these introns were inserted at proto-splice sites. Thus, several examples exist to support the 'early' and 'late' models.

Recent large-scale studies on intron evolution involved comparative analysis of orthologous genes in different eukaryotic lineages. Comparison of intron positions on a large scale for animal, plant and fungal species revealed, that there exists a remarkable conservation of intron positions even among evolutionarily distant eukaryotic lineages (12). Roy, *et al* performed a similar comparative analysis on intron positions in orthologous genes from vertebrates and showed only a few losses but no gain of introns in mammalian genes (13). The results imply that intron loss dominates at short evolutionary distances. Such comparative genomic analyses help in understanding the molecular and evolutionary mechanisms that possibly resulted in the present day gene architecture in eukaryotic genomes.

Tubulins are a group of ubiquitous proteins and are the principle structural and functional components of eukaryotic microtubules (14). The tubulin superfamily contains six distinct families – α, β, γ, δ, ε, and ζ tubulins of which the three major types, α, β and γ are ubiquitous in eukaryotes and have been characterized in a wide variety of organisms. To evaluate the antiquity of introns in three major types of tubulins - as many as 41 gene sequences belonging to the α, 80 sequences to the β, and 30 sequences for γ are compiled from GenBank and Genome databases. Analysis on the intron position conservation in each type of tubulin groups (α, β, γ) was performed to determine patterns of introns loss or introns gain. Interestingly, all the three types of tubulins share intron positions within and across mammals, invertebrates, plant, and fungal kingdoms hinting on their primordial origin. The possible sites of intron loss for the individual tubulin types have been explained by hypothetical phenograms. The results are generally compatible with the 'introns early' view, with a substantial number of introns positions conserved across diverse phylogenetic lineages. Several obvious events of intron loss and intron sliding are evident.

3. MATERIALS AND METHODS

The gene structures for α, β, and γ tubulins were downloaded from GenBank and Genome data files from NCBI, ftp://ftp.ncbi.nih.gov/. The protein sequences for the tubulins were extracted and clustered into three groups (α, β, γ) based on their annotation. The nucleotide sequences with intron positions were extracted using a Perl script and the 'CDS FEATURE' was identified to locate the intron positions according to a previously described methodology (15). Schematic representation showing the spatial distribution of introns for α, β, and γ tubulins were generated. Hypothetical phenograms were constructed to demonstrate the evolutionary lineage of intron positions in tubulin genes from protists, fungi, alga, plant and invertebrate and vertebrates organisms.

4. RESULTS

Tubulins, as the major structural genes, are essential genes and are highly conserved throughout evolution. This study elucidates the degrees of genetic relatedness and evolutionary relationships between the orthologs of α, β, and γ tubulins in different organisms of varying phyla and hierarchical order, and attempts to correlate these observations with molecular evolution. The spatial distributions of intron positions for 41 α, 80 β and 30 γ tubulin sequences across eukaryotes are compiled and represented schematically (Figure 1, Figure 3, and Figure 5).

Generally, the gene structures of the higher eukaryotes (human, mouse and chicken) show strict conservation of intron positions (within ±10 amino acids positions). The intron distribution patterns observed for the vertebrate genomes are similar in all the three tubulins. For instance, the β tubulins from invertebrates include *D. melanogaster*, *C. elegans*, *C. briggsae*, *D. erecta* and *O. volvulus* (Figure 3). Several of their intron positions are shared with other higher and lower eukaryotes. Notably, the introns at locations 19 and 55 are comparable with the first two introns of higher organisms as well as with a few fungal species. The intron at position 131 of invertebrates is maintained with all the alga and fungi like *S. bovines*, *C. cinerea*, and *P. sajor-caju*. Remarkably, the β tubulin gene structures of the invertebrates *O. volvulus* and *B. pahangi* are similar in having eight introns at identical positions. Furthermore, these introns are not only shared between these two organisms, but they are also conserved with other higher and lower eukaryotes. Likewise, the γ tubulins for invertebrates show many positions conserved with vertebrates, plants and fungi (Figure 5).

The plant α tubulins contained three introns corresponding to positions 31, 109 and 233. Examination of the plant α tubulin genes revealed that there is considerable conservation of intron position with lower eukaryotes for all the three introns. Interestingly, one of the paralogues of *A. thaliana* α tubulin gene structure contained 4 introns at positions 37, 109, 176 and 345. These introns are shared with algae, plants, and higher and lower eukaryotes. For example, the third intron at 176 is an exact match with alga *C. vulgaris* and the marine tunicate *O. diocia*. The fact that the conserved intron positions at 176 and 345 are absent in other eukaryotes, suggests that these introns were lost in orthologous α tubulin genes during evolution.

The plant γ tubulins include three paralogs of *A. thaliana*, *P. patens*, *H. mnioides*, *C. japonicum* and *L. albus*. The common plant intron positions are 43, 110, 147, 281, 299, 334, 359, 374, 417 and 452 and are highly conserved among plants. The first intron at position 43 and the intron at 374 are unique to plants, of which 43 alone is shared with alga *C. reinhardtii*. Several of the intron positions are shared with higher organisms, invertebrates and lower fungi.

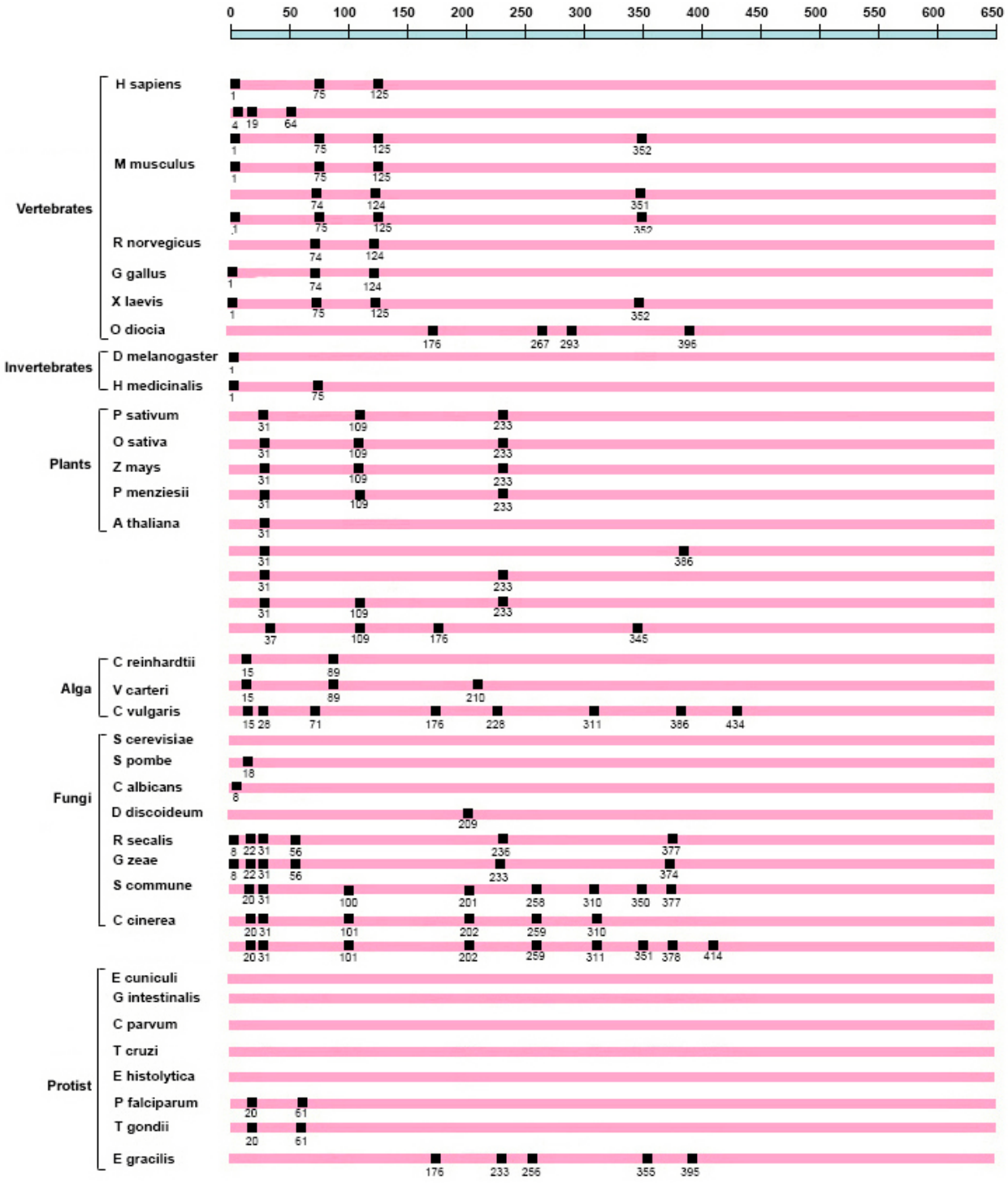

Figure 1. Intron position distributions along the protein length for α tubulins. Introns are shown as black bars and described by codon positions. The scale at the top is in amino acid.

The alga represent the next group of tubulins and in all, it is clear that many of the algal species are generally conserved well with one or the other higher and/or lower eukaryotes. In fig 3, representing the beta tubulins, the two paralogs of *P. agilis* differ by a missing intron position in one of them at codon 6, supporting the idea that duplicate genes (isoforms of tubulins) involves loss of introns.

The α tubulin for fungal species *S. cerevisieae* contain uninterrupted gene structure.

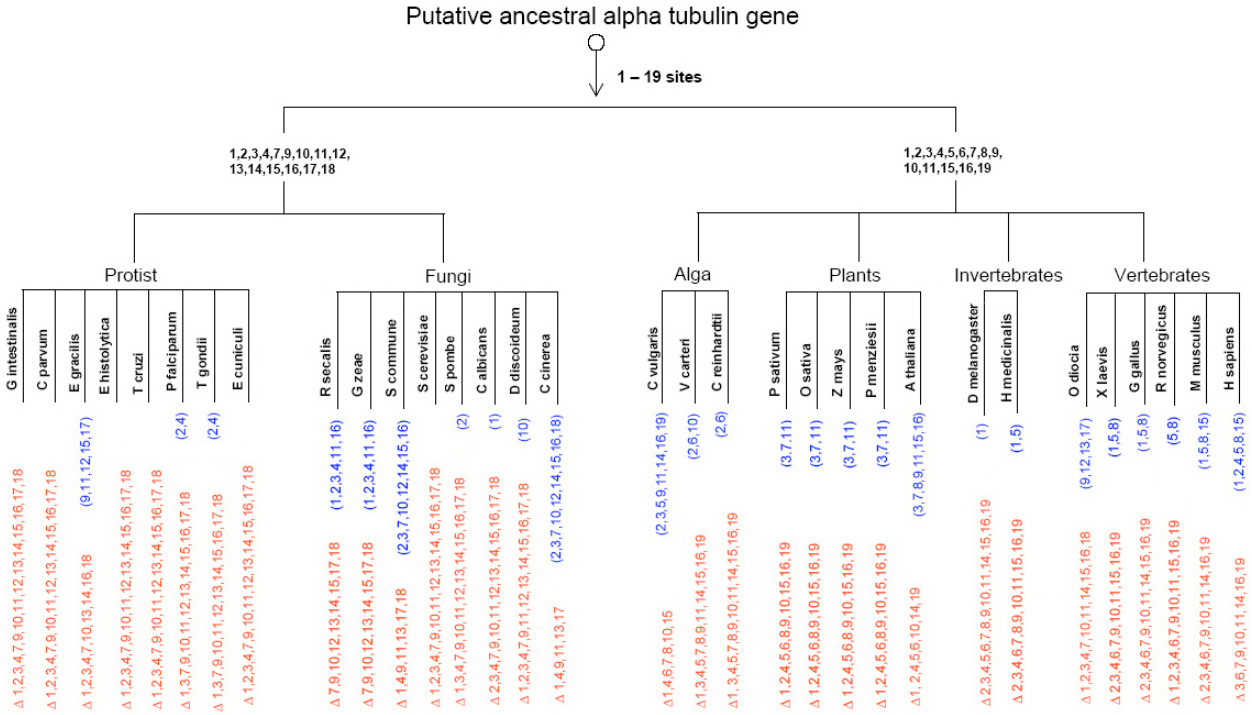

Figure 2. A phenogram for the hypothetical evolutionary lineage of intron sites in α tubulins. Intron positions are given as 19 sites for varying species. Triangles represent deletion of intron sites.

Nevertheless, the genome of the fission yeast *S. pombe* contains a single intron at position 18, that is near identical (± 10 sliding) to 19 of human, 15 of algae and 20 of fungi. The fungal β tubulins comprised about 38 species that exhibited a conserved pattern in their intron locations. Similarly, the γ tubulin fungal species include conserved intron positions at 2, 16, 54, 72,103, 133, 201, 231, 282, 331, 386, 424 and 445. These intron positions are selectively maintained across vertebrates, invertebrates and plants suggesting a highly split, common ancestral gene for γ tubulins. In particular, the introns at 54, 231and 386 for *S. pombe* and *S. japonicus* correspond in position exactly with human introns.

A hypothetical phenogram is constructed to trace the evolutionary lineage of intron sites in α, β, and γ tubulins shown in Figure 2, Figure 4 and Figure 6 respectively. For each lineage we identified the probable intron sites that were lost, based on data derived from the occurrence of intron positions in the present day tubulin gene structures.

5. DISCUSSION

The tubulin family has a number of advantages for studying the evolution of gene structures as: (1) sequence data for both lower and higher eukaryotes are available, and (2) a good conservation of sequences between homologous genes is observed that helps to assign the positions of introns onto the protein. Here, we discuss the implications of gene structure for the possible origin of introns during the evolution of the tubulin family, with particular reference to the conservation of intron positions across phylogenetic distances.

To study gene structure evolution by mapping intron positions onto the protein sequence in each of the tubulin groups, we extracted the intron positions and compared across distantly related eukaryotic lineages ranging from protists to human. Intron position maintenance among diverse taxa may be explained either by conservation of an ancestral intron or by insertion in multiple lineages (16). As more and more different taxa are found with a common intron position, the former possibility becomes more likely (16). Comparative genomic analyses on tubulins also show notable conservation of intron positions confirming their antiquity. Such a conservation of intron positions in orthologous genes from several distinct phylogenetic groups (protists, fungi, alga, plants, nematodes, and vertebrates) shows that, although the numbers of introns vary across different lineages in a particular group of tubulins, the positions of the introns are generally maintained between the orthologs (Figure 1, 3, and 5). This conservation of the intron positions suggests the possibility of an ancestral split gene organization that could have been retained by the different orthologs. Furthermore, comparisons of tubulin genes from distantly related eukaryotes, such as representatives of different animal phyla and kingdoms, suggests substantial loss of introns and thereby intron positions. In some lineages, such as vertebrates or plants, many more introns seem to have been lost. The inferred evolutionary situation is that the common ancestor for eukaryotic tubulins had a large number of introns that were ancestral in nature and

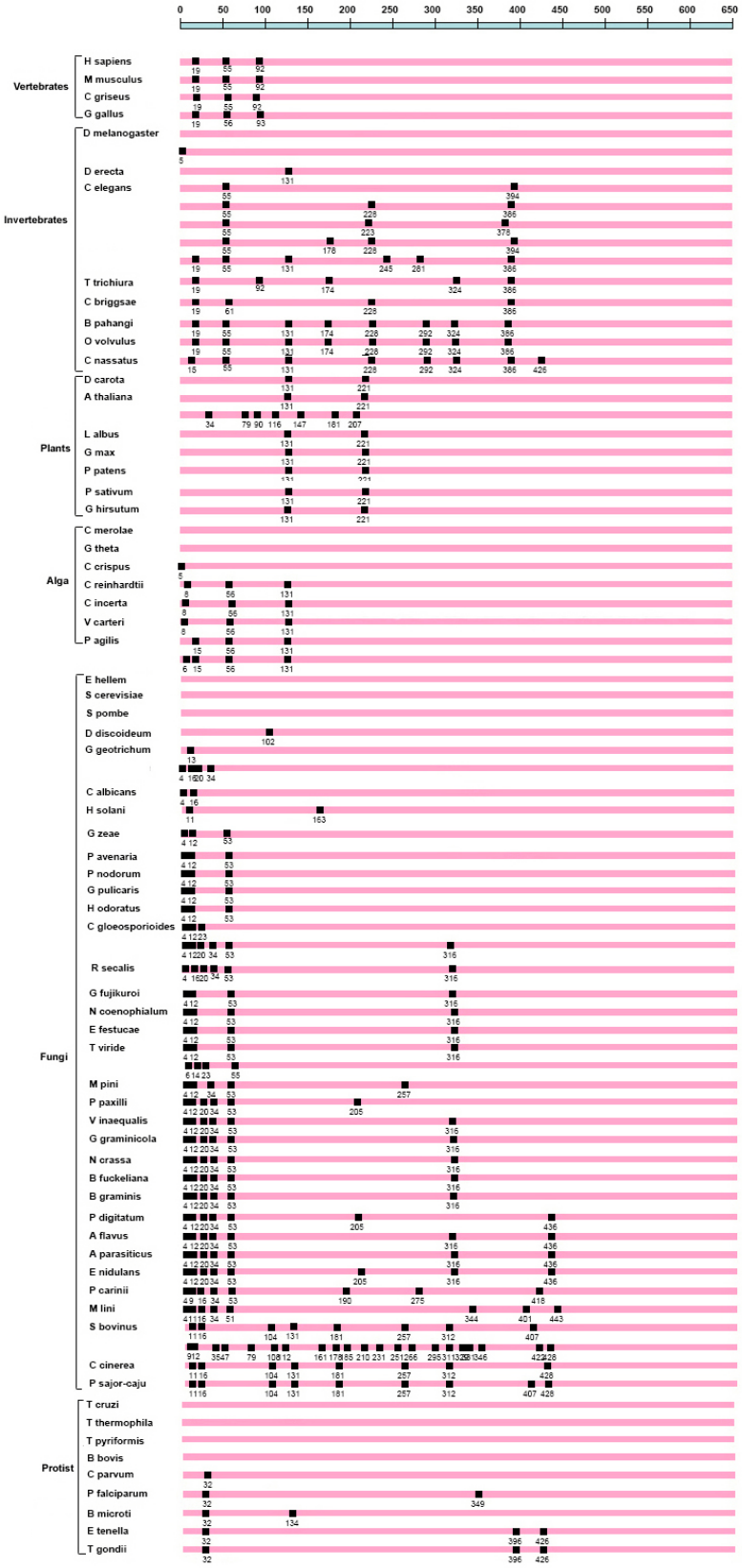

Figure 3. Intron position distributions along the protein length for β tubulins. Introns are shown as black bars and described by codon positions. The scale at the top is in amino acid.

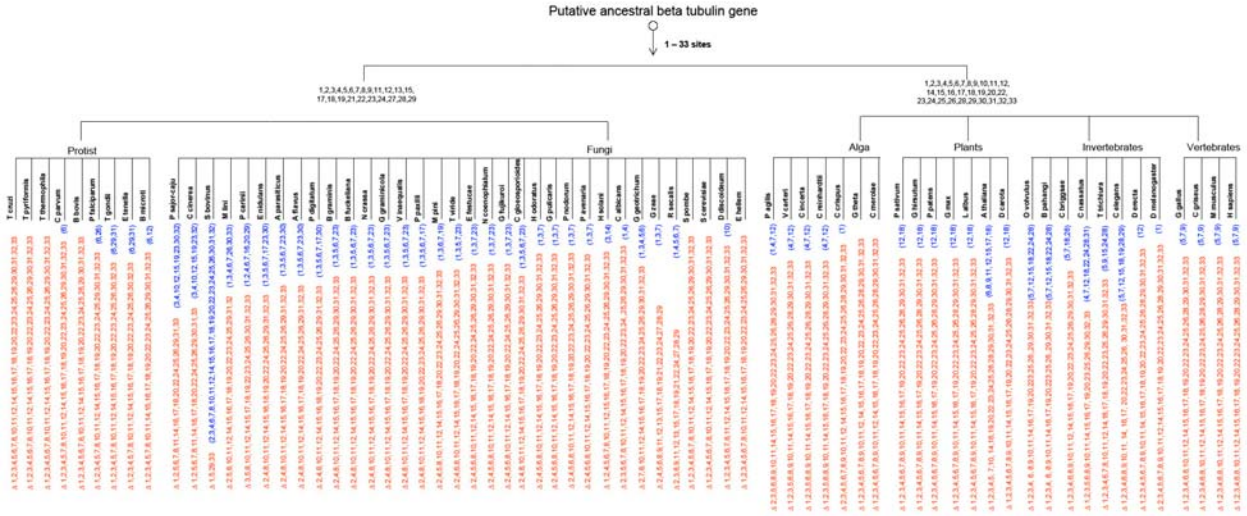

Figure 4. A phenogram for the hypothetical evolutionary lineage of intron sites for beta tubulins. Intron positions are shown as 33 sites for varying species. Triangles represent deletion of intron sites.

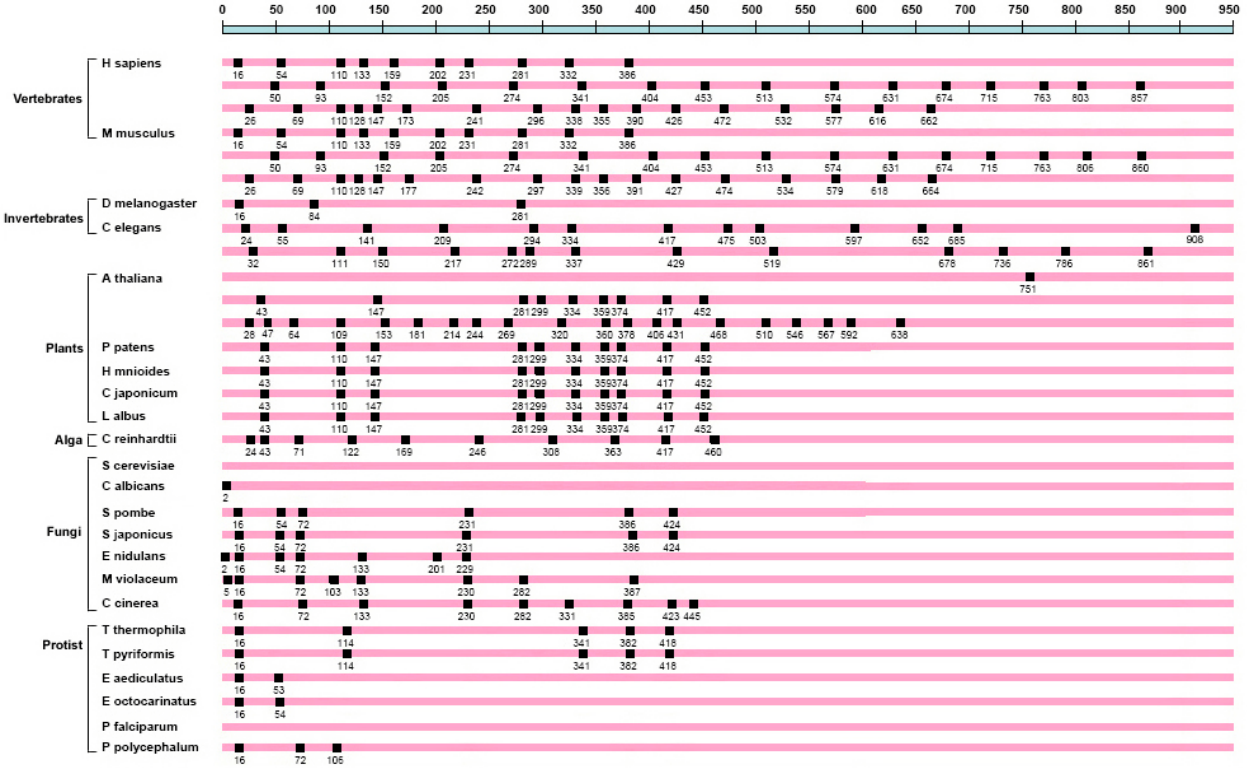

Figure 5. Intron position distributions along the protein length for γ tubulins. Introns are shown as black bars and described by codon positions. The scale at the top is in amino acid.

have been differentially lost in fungi, invertebrate, plants and vertebrates.

Our results also reveal that the distribution of number of introns is generally conserved within specific kingdoms, even though they are highly variant between the kingdoms and for the different tubulin groups studied. Besides, there are also intron positions that are maintained at close positions and these could be treated as instances of intron sliding. The conserved intron positions were also studied for phase distribution (data not shown). Phase conservation is generally observed for conserved intron positions. Phase 0 introns are predominant and most of the introns follow the GT-AG consensus observed at splice

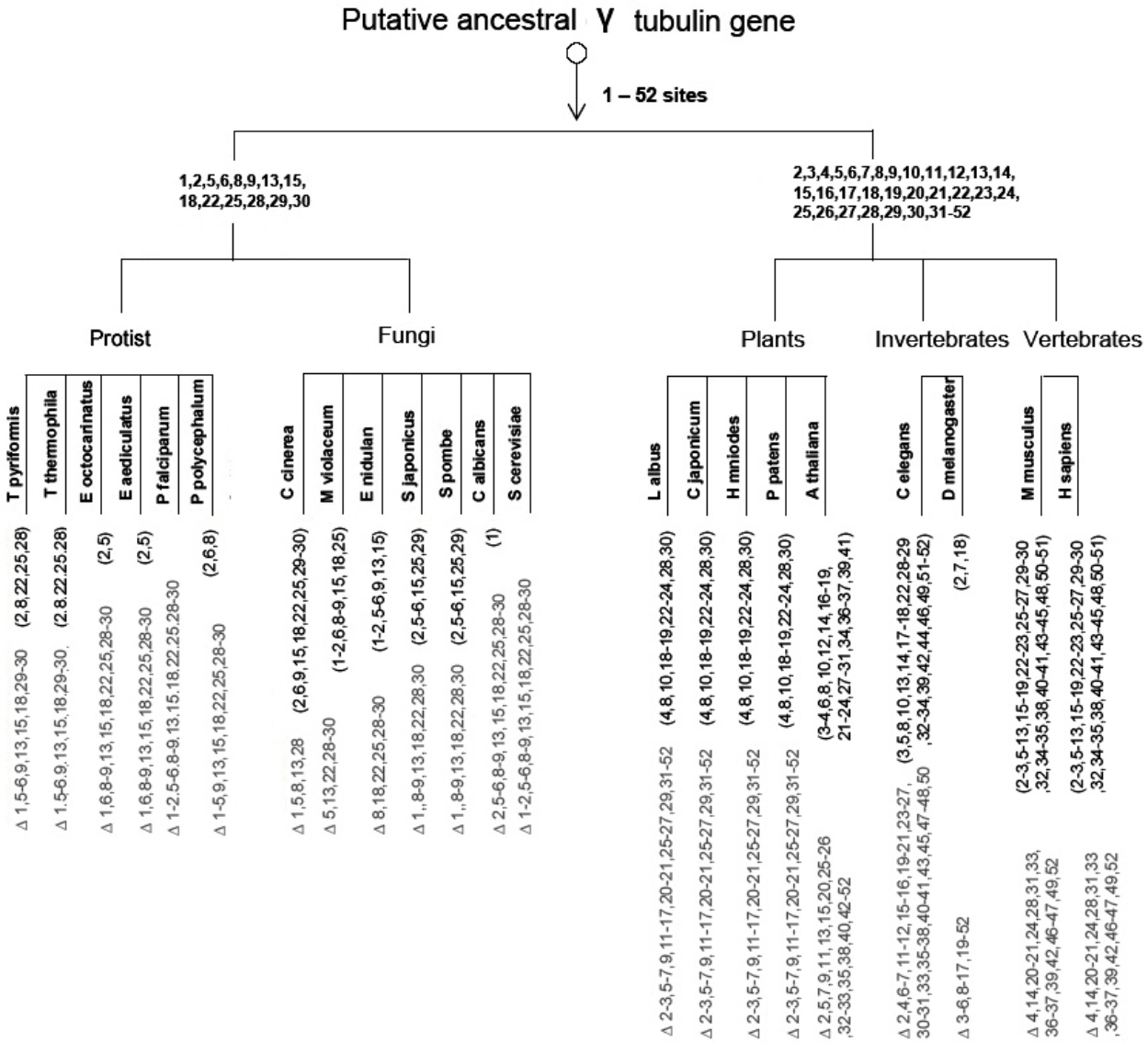

Figure 6. A phenogram for the hypothetical evolutionary lineage of intron sites in γ tubulins. Intron positions are shown as 52 sites for varying species. Triangles represent deletion of intron sites.

sites for spliceosomal introns. It was noted that intron length are highly variant and do not show conservation even between identical intron positions. This is quite acceptable since many reports are available on large standard deviations in mean intron lengths on eukaryotes (17-18). Our results are in line with recent suggestions by Roy *et al* that vertebrate genes show exceedingly few intron differences between the genomes of human, mouse and rat, most of which could be attributed to intron loss (13).

Taking into consideration the issue of intron sliding (±10 positions) and mapping of all the intron positions onto a common denominator shows that generally the intron positions are maintained. Based on the intron positions, it is suggested that a putative ancestral α tubulin gene had at least 19 introns that were common to both lower and higher eukaryotes, which later were lost differentially to produce the observed intron distribution in the modern α tubulins. The putative ancestral β tubulin gene contained at least 33 intron sites that were conserved between the lower and higher eukaryotes, as inferred from their genomic sequences. Similarly, the putative ancestral γ tubulin gene contained at least 52 introns. Many of these introns in all the three tubulins are maintained and the absence of introns at many positions can be attributed to their selective loss in different lineages in the course of evolution.

In order to trace the evolutionary trajectory of the tubulin genes, we generated phenograms for each of the tubulin groups (Figure 2, Figure 4 and Figure 6). The fact that almost all intron positions for tubulins can be clearly attributed to descent from intron loss, explains that introns

are very ancient. These results support the 'intron-early' hypothesis, where introns have always been an integral part of the genome at least for this family of proteins.

6. CONCLUSION

Since the discovery of introns 27 years ago, many instances of intron loss and intron gain have been described. Tubulins are ubiquitous in nature. Here, we analyzed the gene structures for 41 α, 80 β, and 30 γ tubulin genomic sequences to study intron position conservation. Our studies confirm the assumption that introns were present in the common ancestral gene in the case of tubulins and not gained later by insertion into coding regions. This analysis is based on inclusion of a large number of genomic sequences across diverse eukaryotic lineages.

7. ACKNOWLEDGEMENTS

This manuscript is based on the project supported by A*STAR - BMRC grant #03/1/22/19/242.

8. REFERENCES

1. Gilbert W: Why genes in pieces? *Nature* 271, 501-502 (1978)
2. Doolittle W.F: Genes-in-pieces: were they ever together? *Nature* 272, 581- 582 (1978)
3. Orgel L.E. & F.H.Crick: Selfish DNA: the ultimate parasite. *Nature* 284, 604-607 (1980)
4. Cavalier-Smith T: Selfish DNA and the origin of introns. *Nature* 315, 283-284 (1985)
5. De Souza S.J: The emergence of a synthetic theory of intron evolution. *Genetica* 118, 117–121 (2003)
6. Perler F., Efstratiadis A., Lomedico P., Gilbert, W., Kolodner R. & J. Dodgson: The evolution of genes: the chicken preproinsulin gene. *Cell* 20, 555–566 (1980)
7. Bagavathi S & R. Malathi: Introns and protein evolution--an analysis of the exon/intron organisation of actin genes. *FEBS Lett.* 392, 63-65 (1996)
8. Dibb N.J & Newman A.J: Evidence that introns arose at proto-splice sites. *EMBO J*. 8, 2015-2021 (1989)
9. Boudet N, Aubourg S, Toffano-Nioche C, Kreis M & A 6. Lecharny: Evolution of intron/exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis, and Drosophila. *Genome Res*. 11, 2101-2114 (2001)
10. Hankeln T, Friedl H, Ebersberger I, Martin J & ER Schmidt: A variable intron distribution in globin genes of 6. Chironomus: evidence for recent intron gain. *Gene* 205, 151-160 (1997)
11. Coghlan A & Wolfe K.H: Origins of recently gained introns in Caenorhabditis. *Proc Natl Acad Sci U S A*. 101, 11362-11367 2004
12. Fedorov A., Merican A.F. & W.Gilbert: Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc Natl Acad Sci U S A.* 99, 16128-16133 (2002)
13. Roy S.W., Fedorov A & W.Gilbert: Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proc Natl Acad Sci U S A*. 100, 7158-7162 (2003)
14. Oakley B.R: An abundance of tubulins. *Trends Cell Biol*.10, 537-542 (2000)
15. Sakharkar M., Long M., Tan T.W. & S.J. de Souza: ExInt: an Exon/Intron database. *Nucleic Acids Res.* 28, 191-192 (2000)
16. Palmer J.D & J.M Jr Logsdon: The recent origins of introns. *Curr Opin Genet Dev*. 1, 470-477 (1991)
17. Sakharkar M.K, Chow V.T & P. Kangueane: Distributions of exons and introns in the human genome. *In Silico Biol*. 4,387-393 (2004)
18. Deutsch, M. & M. Long: Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res*. 27, 3219-3228 (1999)

Key Words: Tubulin, Intron Loss, Intron Gain, Genome Evolution, Conservation

Send correspondence to: Meena Kishore Sakharkar Ph.D., Nanyang Technological University, Singapore 639798, Tel: +65-6-790-5836, Fax: +65-6-774-4340, E-mail: mmeena@ntu.edu.sg

http://www.bioscience.org/current/vol10.htm