# Sm/lsm genes: a glimpse into early eukaryotic evolution.

Stella Veretnik[1], Christopher Wills[2], Philippe Youkharibache[1] and Philip E. Bourne[1,3]
[1] San Diego Supercomputer Center, University of California San Diego , 9500 Gilman Dr., La Jolla, CA 92093-0537, USA
[2] Section of Ecology and Evolutionary Biology, University of California San Diego , 9500 Gilman Dr., La Jolla, CA 92093, USA
[3] Department of Pharmacology, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA

Sm/lsm genes are key basal components of the spliceosomal machinery. Their association with all five snRNA components (U1, U2, U4, U5 and U6) of the spliceosomal complex is crucial for splicing. Sm/lsm proteins form a donut-shaped ring around the U-rich region of each snRNA that stabilizes the RNA structure and promotes the binding of other U-specific proteins to the snRNP as it forms[1]. Sm proteins interact with U1, U2, U4 and U5, and lsm proteins interact in a similar manner with U6. Sm/lsm proteins are small five-stranded beta-barrels where β-strands 4 and 5 form a wide-open hinge over the rest of the structure and are involved in key interactions that keep barrels together. Thus, the donut-shape ring around the RNA molecule is formed by seven barrels, each barrel interacting with its two neighbors though β-strands 4 and 5 in such a way that β-strands 4 of one barrel interacts with β-strand 5 of the adjacent barrel (see PDBid 1i81).Our analysis of sequence and gene structure of 307 Sm/lsm genes from bacteria, archaea and eukaryota yields two remarkable observations. First, a massive paralogy of Sm/lsm genes occurred in eukaryotes and this process happened very swiftly. Nearly *all* eukaryotes possess 14 distinct copies of the Sm/lsm gene (defying the typical power law distribution based on proteome size), compared to a single copy in bacteria and at most two copies in archaea (in the latter two kingdoms Sm is involved in RNA biogenesis processes other than splicing). The 14 eukaryotic genes (7 coding for Sm proteins and 7 coding for lsm proteins) are significantly distinct from each other, yet each individual Sm/lsm gene is remarkably conserved across all eukaryotic species. Phylogenetic inference based on sequence analysis indicates that this extensive paralogy of the Sm/ lsm genes in eukaryotes happened so rapidly that little trace as to the order of duplication events is left (PAUP, bootstrap analysis; Figure 1). A combination of more sensitive Bayesian inference and functional analysis suggests two closely followed waves of duplication. First, a set of 7 lsm genes is formed, second each lsm gene then undergoes duplication to form its Sm counterpart. Our second observation concerns the introns present in nearly all eukaryotic Sm and lsm genes. Introns positions are remarkably conserved within each Sm/lsm gene and in a few cases also between lsm-Sm counterparts. Since Sm/lsm proteins are a key component of the spliceosomal machinery, we have a unique opportunity to pinpoint the timing of the appearance of functional introns with respect to spliceosome development.

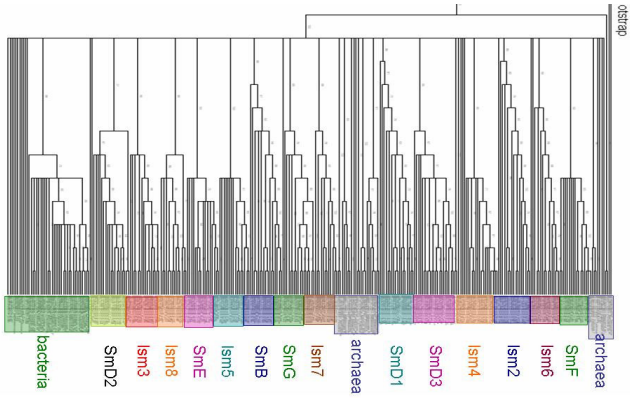

Figure 1. Phylogenetic tree for 307 Sm/lsm genes from Eukaryota, Bacteria and Archaea kingdoms. The analysis is done in PAUP (neighbor-joining, with bootstrapping)

Our analysis brings new insight to early eukaryotic evolution. A rapid and extensive evolution of the basal spliceosomal proteins Sm/lsm is an imprint of some dramatic events that occurred during the development of the early eukaryotic ancestor. The likely event was the invasion of self-splicing type II introns and the subsequent evolution of the machinery to deal with such invasion. Under this scenario, Sm proteins replace the function of the maturase protein carried by the type II introns[2]. By the time of the emergence of the last common eukaryotic ancestor not only is there a full complement of the Sm/lsm genes present, but they carry introns that require the existence of a functional spliceosome. Thus, much of the building and tuning of the spliceosomal machinery happened during the early radiation of the eukaryotic lineage, before the emergence of the last common eukaryotic ancestor.

1. Nagai, K., Muto, Y., Pomeranz Krummel, D. A., Kambach, C., Ignjatovic, T., Walke, S. & Kuglstatter, A. (2001). Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. *Biochem Soc Trans* **29**, 15-26.
2. Martin, W. & Koonin, E. V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41-45.