Insertion of a self-splicing intron into the mtDNA of a triploblastic animal

Y. Vallés, K. Halanych, J.L. Boore

Nephtys longosetosa is a carnivorous polychaete worm that lives in the intertidal and subtidal zones with worldwide distribution (pleijel&rouse2001). Its mitochondrial genome has the characteristics typical of most metazoans: 37 genes; circular molecule; almost no intergenic sequence; and no significant gene rearrangements when compared to other annelid mtDNAs (booremoritz19981995). Ubiquitous features as small intergenic regions and lack of introns suggested that metazoan mtDNAs are under strong selective pressures to reduce their genome size allowing for faster replication requirements (booremoritz19981995Lynch2005). Yet, in 1996 two type I introns were found in the mtDNA of the basal metazoan *Metridium senile* (FigureX). Breaking a long-standing rule (absence of introns in metazoan mtDNA), this finding was later supported by the further presence of group I introns in other cnidarians. Interestingly, only the class Anthozoa within cnidarians seems to harbor such introns. Although several hundreds of triploblastic metazoan mtDNAs have been sequenced, this study is the first evidence of mitochondrial introns in triploblastic metazoans. The *cox1* gene of *N. longosetosa* has an intron of almost 2 kbs in length. This finding represents as well the first instance of a group II intron (anthozoans harbor group I introns) in all metazoan lineages. Opposite trends are observed within plants, fungi and protist mtDNAs, where introns (both group I and II) and other non-coding sequences are widespread. Plant, fungal and protist mtDNA structure and organization differ enormously from that of metazoan mtDNA. Both, plant and fungal mtDNA are dynamic molecules that undergo high rates of recombination, contain long intergenic spacer regions and harbor both group I and group II introns. However, as metazoans they have a conserved gene content. Protists, on the other hand have a striking variation of gene content and introns that account for the genome size variation. In contrast to this mtDNA structure and organization diversity, current genome level studies point to a monophyletic origin of the mitochondria (REFS), raising questions such as: what are the pressures at work shaping the evolution of the mitochondrial genome at "higher" levels? What drives the absence of introns and other non-coding spacers in metazoan mtDNA? What characteristics must have an intron to be maintained in an environment where "extra chromosomes" are usually selected against?

It is believed that the insertion of an intron into a conserved gene plays an important role in its survival within the host genome (REFS). *Nephtys longosetosa*'s intron is no exception and has inserted itself in one of the most conserved genes of the mtDNA, *cox1*. Although the whole mtDNA of *N. longosetosa* was amplified in three overlapping pieces, to ensure that we were in the presence of a mtDNA intron (versus a nuclear pseudogene) we designed primers from the intron outwards and amplified the whole mtDNA in two overlapping pieces. Furthermore, an RNA extraction and later sequencing of *cox1* cDNA showed that the intron is spliced out from the RNA, ensuring a viable and functional *cox1* protein (Fig X). Although the intron is not in frame with the exon, it translates onto a protein 576 aa long. Several features identify it as a group II intron. First, the presence of conserved sequences at the intron boundaries, GUGYG at the 5' end and AY at the 3' end, (versus the consensus sequences among other group II introns GUGYG and AY

respectively). Second the presence of domain V, catalytic core of the ribozymic activity of the intron, identified as well by sequence and structure similarities. Third, the presence of an open reading frame (ORF) encoding for a reverse transcriptase (RT) domain and the partial domain of a maturase (domain X). Finally, even though group II introns have very little sequence similarity, they share a highly conserved secondary structure that consists of six helical domains (I - IV) radiating from a central core. These introns are self-splicing elements, a feature that is essential to have a "minimal" effect on the gene expression of the intronless alleles in which they are inserted (Saldanha 1993). Both RT and X domains function together in the splicing and reverse transcription mechanisms of the intron (Lambowitz 2004). The distribution of these introns is limited to organelles (mitochondria and chloroplast) and bacterial lineages, being absent of the eukaryotic nucleus and of until now of metazoan mitochondrial genomes.

Following this distribution, one possible scenario for the origin of this intron would be the maintenance of an ancient state. Although now group II introns are known to be widespread in the eubacterial "domain", they were first identified from a cyanobacterium and a proteobacterium. This finding suggested at the time that the introns (group I and II) were transferred to eukaryotes by bacterial endosymbiosis (organelles) (REFS). This hypothesis would presume that introns were present in the "ancient" mitochondria and were secondarily lost in metazoan mtDNA. It also assumes that within metazoans it had to occur in multiple instances independently and only be retained in the annelid *N. longosetosa* (in the case of group II introns). This hypothesis is compatible with the fact that no group II introns are found in the nucleus of eukaryotes (except in the nucleus of *Arabidopsis thaliana* that has an integrated remnant of nonfunctional mtDNA containing degenerated group II introns), and with the belief that metazoan mtDNA is under strong selective pressures for smaller genomes.

However the retention of a group II intron in one annelid and the independent loss in the rest of annelids and metazoans seems unlikely. In order to elucidate the nature of the intron we aligned the RT and X domains of the intron's ORF to those of other group II intron ORFs from plants, fungi, virus, retroelements and bacteria. Although in general phylogenetic analyses suggest that most RT domains of group II introns are related to those of LINE1 or non-LTR elements (Lambowitz2004Xiong_&_Eickbush1990EMBOins&outs), we were unable to determine its closest relative due to extreme sequence divergence, obscuring its potential origin. Nevertheless, a more likely scenario would be a recent acquisition of the intron via horizontal transfer (viral or bacterial vectors). Group II introns have an incredibly wide and diverse distribution being most of them at unique positions. Studies in bacteria have shown that these introns have high mobility capabilities and move throughout populations by both horizontal transfer and vertical inheritance. Within plant mtDNA, although vertical inheritance of group II introns seems to be prevalent horizontal transfer has also been documented (palmer). An example is the presence of 23 and 25 group II introns in *Arabidopsis* and *Marchantia* mtDNAs respectively, with only one of the 48 being at a homologous position (ins_outs).

It has recently been suggested that the main differences between the organelle lineages are due to non-adaptive processes such as the mutation rate and random genetic drift (Lynch2006). Because both cnidatians and annelids are the exceptions to the metazoan rule (absence of introns in mtDNA), these taxa are good candidates for elucidating the mechanisms that drive the reduction of genome size in metazoans. The presence of introns in cnidarians and annelids may be considered a genomic version of the "panda's thumb" (Gould, 19XX) that illuminates the evolution of mtDNA in metazoa.
If the are so mobile why are they not in the nucleus?