Concerted evolution of 5S rRNA and trans-spliced leader genes in *Caenorhabditis* species.

Zacharie Cloutier Supervisor: Guy Drouin University of Ottawa

Introduction

Previous studies concerning the evolution of 5S rRNA genes in *Caenorhabditis* nematodes species were based on small chromosomal regions, providing only partial information. Recent development in genomic sequencing have produced complete genome sequences for several *Caenorhabditis* species. The *C. elegans* genome can therefore now be compared to those from the closely related species *C. briggsae*, *C. remanei*, *C. sp11, C. brenneri* and *C. japonica*.

5S ribosomal RNA is part of the large ribosomal subunit in all known organisms except for mitochondrial ribosomes of animals and fungi, and is a well known phylogenetic marker. The gene coding for 5SrRNA (5S) is 119 bp in length and several copies are present in nematode genomes. In addition, it is often linked to the gene coding for *trans* spliced leader 1 (SL1) sequences which are added to the 5' ends of nematode mRNA molecules which are transcribed from genes which are not part of operons[1]. The SL1 gene measures 98 bp, the 22 first of which constitute the functional spliced leader added to the 5' ends of nematode mRNA molecules.

Here, we characterised the organisation of 5S and SL1 genes in six *Caenorhabditis* species. Our aim was to determine whether these two genes are always linked together and whether such linkages are inherited from common ancestors or not.

Objectives

- Characterize the tandem formed by the pairing of gene 5S and SL1 for different *Caenorhabditis* genus.
- Asses the variations of the characteristics of this tandem throughout the specie.
- Use the variation in tandem characteristics to describe the evolution evolution of 5S and SL1 within the phylogenetic group.
- Determine if the 5S-SL1 tandem in different genus is inherited from a common ancestor or the product of independent insertions.

Method

- Available sequences for *C. elegans*, *C. briggsae*, *C. remanei*, *C. sp11*, C. *brenneri* and *C. japonica* were obtained from http://www.wormbase.org/species/all#0--10
- The sequence of the 5S gene[2] was used to determine the intergenic distance (bp) between repeated units of the 5S gene.
- The sequence of the SL1 gene[3] was used to determine intergenic distance (bp) between linked 5S and SL1genes.
- 5S and SL1 genes were identified in all six *Caenorhabditis* genomes using BLAST[4].
- Genomic maps of each contig, chromosome or scaffold on which tandems were found were drawn using PowerPoint.
- The phylogenetic tree of the different *Caenorhabditis* species was obtained from http://www.wormbase.org/.

Results

Species	Number of 5S genes	Number of 5S-SL1 tandem (and %)	Number of complete 5S-SL1 tandem (and %)	5S-SL1 intergenic distance μ ± σ (pb)	5S-5S intergenic distance μ ± σ (pb)
C. japonica	150	150 (100)	150 (100)	161	439
C. elegans	16	13 (81)	12 (92)	172 ± 9	852 ± 19
C. sp11	24	1 (4)	1 (100)	178 ± 0	912 ± 0
C. brenneri	30	16 (53)	9 (56)	350 ± 17	641 ± 29
C. remanei	36	26 (72)	8 (31)	384 ± 24	963 ± 309
C. briggsae	10	6 (60)	5 (83)	183 ± 44	758 ± 120

Figure 1: Phylogenetic tree of *Caenorhabditis* species and the characteristics of repeated 5S-SL1 tandem found in their genomes. The 5S-SL1 tandem percentage is given by the ratio between the number of 5S paired with SL1 and the total number of 5S for each genus. The complete tandem percentage measures the proportion of tandems found with complete 5S and SL1 sequences and with intergenic distance close to the average. μ is the average of the intergenic distance while σ represents the standard deviation. C japonica data was obtained form contig 17147.

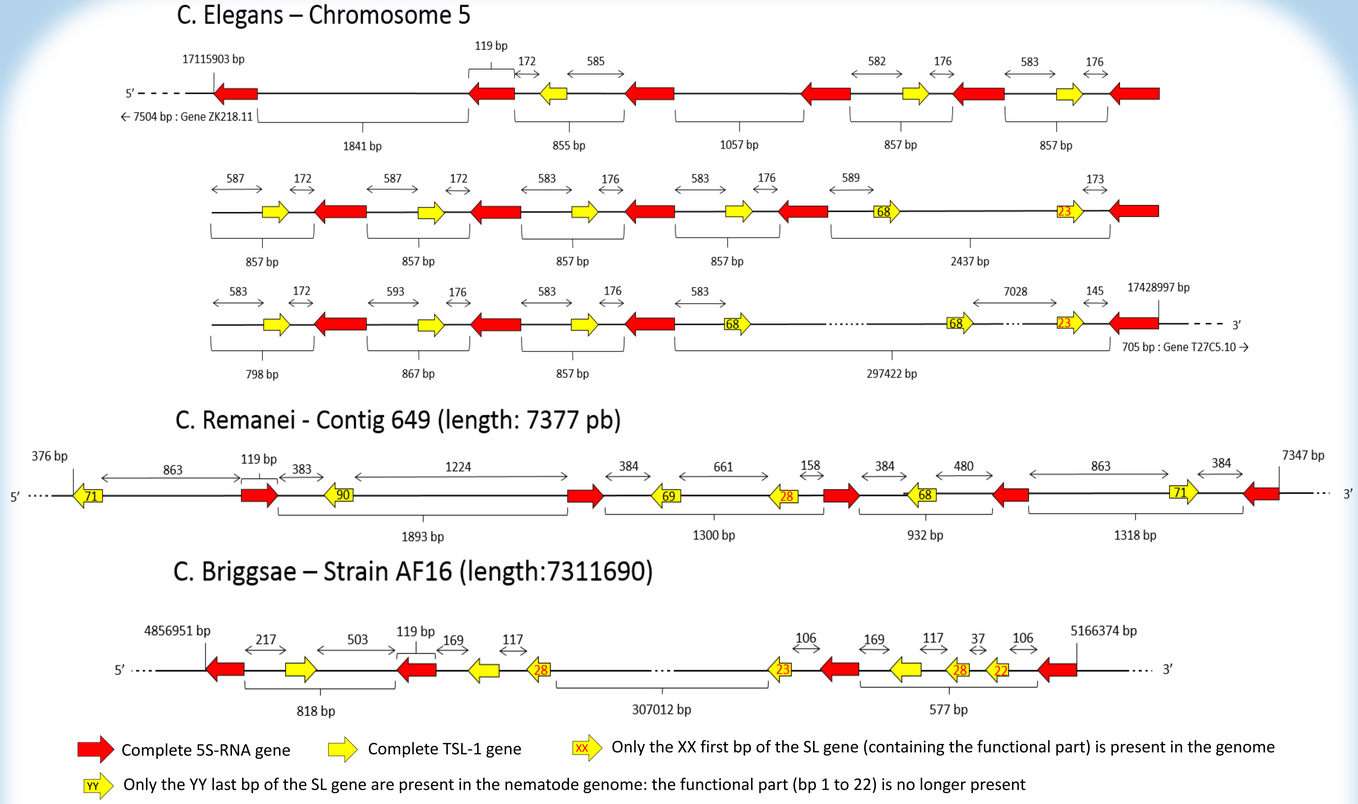

Figure 2: Genomic maps of regions where the 5S and SL1 genes are associated in tandems in the genome of C. elegans, C. remanei and C. briggsae. BLAST was used to determine the location and orientation of 5S (red arrows) and SL1 genes (yellow arrows). Incomplete gene sequences are also shown. Most *C. elegans* genes are organised head-to-head and have a conserved intergenic distance. Contig 649 of C. remanei shows multiple tandems with incomplete SL1 sequences missing the functional part, as well as longer intergenic distances. Strain AF16 of C. briggsae shows an example of repetitive copies of the functional part of SL1 (bp 1 to 22) and the variation of the direction of SL relative to 5S.

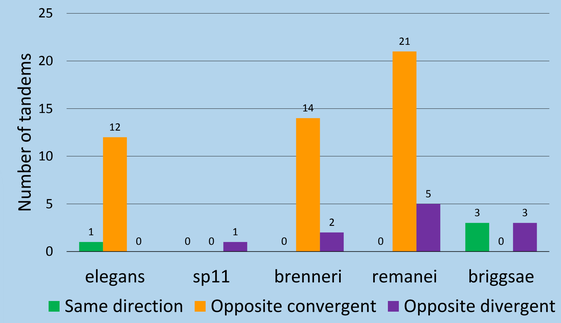

Figure 3: Comparison of the number of tandems with 5S and SL1 in the same direction, in opposite convergent directions and in opposite divergent directions for each genus of *Caenorhabditis*.

Discussion

Our results show that the organisation of 5S and SL1 genes is highly variable between *Caenorhabditis* species. This is so in terms of the number of 5S genes, the number of 5S-SL1 tandems and intergenic distance. Some characteristics seem to be conserved from the original *C. elegans* tandem model shown on Figure 2 such as the intergenic distance for certain genus. The number of 5S repeated units is higher for *C. sp11*, *C. brenneri* and *C. remanei* yet these genus have a decrease in complete tandem percentage, whereas it is not the case for *C. briggsae*. The number of complete tandems seem to decrease further along the phylogenetic tree, but does not respect a similar pattern in terms of direction or intergenic distance. Figure 3 shows that although tandems are mostly found to have 5S and SL1 in opposite convergent direction, some inversions are observed between different genus. Closely related species do not follow a similar trend.

Conclusion

- Segments containing repeated 5S-SL1 tandem genes tend to have a lot of rearrangements regarding the number of tandems, the intergenic and inter-tandem distance and the direction. Isolated tandems in different genus tend to have less modifications, following a steadier evolution.
- 5S and SL1 are not always found in tandem, and SL1 copies are often incomplete, lacking the functional part.
- The tandem variations do not form any particular pattern that can be evidently linked to the phylogenetic relations within the species, indicating independent arbitrary modifications of tandem characteristics for different genus.

Reference

[1] : Evans and Blumental. Trans splicing of polycistronic *Caenorhabditis elegans* pre-mRNAs: analysis of the SL2 RNA. Mol Cell Biol 20: 6659-6667.

[2] : Nelson and Honda. 1985. Gene coding for 5S ribosomal RNA of the nematode *Caenorhabditis elegans*. Gene, 38: 245-251.

[3] : Krause and Hirsh. 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49: 753-761.

[4] : Zhang, Schwartz, Wagner and Miller. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7: 203-14.