

Table S10. Taxa, accession numbers and chromosome location of partial *DIA1L* paralogues.

| Species ^a | Protein accession number | mRNA (or EST) accession number | Genomic DNA accession number | Intron(s) ^b | Comments |
|-----------------------------------|--------------------------|---------------------------------|------------------------------|---|---|
| METAZOA | | | | | |
| Chordata | | | | | |
| Cephalochordata | | | | | |
| <i>Branchiostoma floridae</i> (1) | XP_002593885 | XM_002593839 | NW_003101415 | - | BRAFLDRAFT_214875 Length of predicted protein: 126 amino acids Completeness: incomplete on the carboxy-terminal end. Has an annotated initiation methionine, but this may not be correct. Sequence alignments suggest a close relationship to <i>B. floridae</i> DIA1Lb (data not shown). EST data supports expression (93% identical to EST with GenBank accession number: FE571008). However, this is the same EST supporting the gene below (<i>DIA1L-partial 2</i>). |
| <i>Branchiostoma floridae</i> (2) | XP_002593886 | XM_002593840 | NW_003101415 | - | BRAFLDRAFT_161925 Length of predicted protein: 91 amino acids. Completeness: Incomplete on both ends. Sequence alignments suggest a close relationship to DIA1Lb and 'partial 1.' EST data supports expression (93% identical to EST with GenBank accession number: FE571008). However, this is the same EST supporting the gene above. |
| <i>Branchiostoma floridae</i> (3) | XP_002586953 | XM_002586907 | NW_003101317 | - | BRAFLDRAFT_269049 Length of predicted protein: 230 amino acids. Completeness: annotated as 'complete,' but appears to be missing N-terminal (5' coding) sequence. Sequence alignments suggest a close relationship to DIA1Lb and 'partial-4' below. Expression is supported by a single EST sequence (Genbank expression number: FE569795). |
| <i>Branchiostoma floridae</i> (4) | XP_002613084 | XM_002613038 | NW_003101564 | Yes (but <u>not</u> in DIA1-like region -see adjacent comments) | BRAFLDRAFT_125704 Length of predicted protein: 703 amino acids, annotated as 'complete.' Only the <u>first exon</u> has similarity to <i>DIA1L</i> . In this gene annotation the first exon of DIA1L (encoding 180 amino acids only) has been 'tacked onto' the 5' end of a gene encoding a <i>B. floridae</i> <i>Notch</i> homologue. This most likely represents an incorrect gene annotation. A portion of this sequence is identical to that of <i>B. floridae</i> <i>DIA1L-partial 3</i> (see above), and overlaps the relevant EST sequence (Genbank expression number: FE569795). This suggests these two sequences are incorrectly annotated alleles of the same gene. This is further supported by examination of the surrounding genes, which reveals reverse synteny with the 'partial-3' genomic region (data not shown). Sequence alignments suggest a close relationship of this region to DIA1Lb (data not shown). |
| <i>Branchiostoma floridae</i> (5) | XP_002588671 | XM_002588625 | NW_003101353 | - | BRAFLDRAFT_255455 Length of predicted protein: 115 amino acids. Completeness: Incomplete on both ends. Currently annotated as being expressed, but this is not supported by current EST data. |

^aHomologues of *DIA1* are only present in the Metazoa and *DIA1IL* orthologues in they phylum Echinodermata and subphylum Cephalochordata.^bA dash is used when an absence of introns was found in sequences, due to insufficient data to draw a definitive conclusion of lack in the full-length gene.