

**Table S2.** Taxa, accession numbers and chromosome location of partial *DIA1* orthologues.

Species <sup>a</sup>	Protein accession number or identifier	mRNA (or EST ) accession number or identifier	Genomic DNA accession number or identifier	Chrom <sup>b</sup>	Intron(s) <sup>c</sup>	Comments
<b>METAZOA</b>						
<b><u>Platyhelminthes</u></b>						
<i>Schistosoma mansoni</i>	-	CD093176 & BU770455	-	-	-	Non-overlapping ESTs.
<b><u>Hemichordata</u></b>						
<i>Saccoglossus kowalevskii</i>	-	FF496411	-	-	-	EST sequence.
<b><u>Mollusca</u></b>						
<i>Lottia gigantea</i>	-	FC757266	-	-	-	EST
<i>Mytilus californianus</i>	-	ES736823	-	-	-	EST
<i>Venerupis decussatus</i>	-	AM869986	-	-	-	EST
<b><u>Annelida</u></b>						
<i>Helobdella robusta</i>	HELRO10585	-	scaffold_18		Yes	JGI identifier.
<i>Capitella</i> sp.	-	EY608635	-	-	-	EST
<i>Hirudo medicinalis</i>	-	EY484180	-	-	-	EST
<i>Tubifex tubifex</i>	-	EY450728	-	-	-	EST
<b><u>Arthropoda</u></b>						
<i>Apis mellifera</i>	-	-	NW_001254902	-	-	Genomic sequence found by BLAST, correct gene yet to be annotated.
<i>Bombyx mori</i>	BGIBMGA 004346-PA	BGIBMGA 004346-TA	BGIBMGA 004346	-	Yes	Silkworm genome database identifiers.
<i>Drosophila simulans</i>			Chr2R_random: 2756013,2756201	2R	-	<i>Drosophila</i> species genome BLAST at <a href="http://insects.eugenes.org/species/blast/">http://insects.eugenes.org/species/blast/</a>
<i>Ixodes scapularis</i>	EEC05416	-	DS703671	-	Yes	WGS sequence
<i>Pediculus humanus</i>	EEB19697	-	DS235874	-	Yes	WGS sequence
<i>Triboleum castaneum</i>	XP_001815292	XM_001815240	NW_001092881	-	Yes	Contains gene annotation errors. By similarity to all other <i>DIA1</i> genes, the sequence has regions missing.
<b><u>Chordata</u></b>						
<b><u>Urochordata</u></b>						
<i>Ciona savignyi</i>	-	-	reflig_11 Positions 606,676 to 606,885	-	-	Identifiers from the <i>Ciona savignyi</i> database at the Broad Institute

<i>Halocynthia roretzi</i>	-	DB634200	-	-	-	EST
Vertebrata						
Cephalaspidomorphi*						
<i>Petromyzon marinus</i>	-	-	Contig84862	-	-	ENSEMBL genomic BLAST No further sequence available near start of contig to try to generate a gene model.
Actinopterygii						
<i>Gadus morhua</i>	-	EX722062	-	-	-	EST
<i>Haplochromis matumbi</i>	-	DB869638	-	-	-	EST
<i>Pimephales promelas</i> (a)	-	DT206727, DT232116 & DT131905	-	-	-	Overlapping 5' EST sequences Homologue of <i>D. rerio</i> <i>DIA1a</i> (see data below):  Comparison of 127 amino acids translation product with <i>D. rerio</i> <i>DIA1a</i> gene product- 98% similar 98% identical Protein BLAST E-value 1e-69  Comparison of 127 amino acids with <i>D. rerio</i> <i>DIA1b</i> gene product: 85% similar 97% identical Protein BLAST E-value 1e-55
<i>Pimephales promelas</i> (b)	-	DT148813 & DT1248813	-	-	-	Overlapping 5' ESTs Homologue of <i>D. rerio</i> <i>DIA1b</i> (see data below):  Comparison of 190 amino acids translation product with <i>D. rerio</i> <i>DIA1b</i> gene product- 98% similar 100% identical Protein BLAST E-value 9e-109  Comparison of 190 amino acids with <i>D. rerio</i> <i>DIA1a</i> gene product: 90% similar 99% identical Protein BLAST E-value 2e-100
Tetrapoda						
Aves						
<i>Taeniopygia gutta</i>	-	DV580069	-	-	-	EST
Reptilia						
<i>Anolis carolinensis</i>	-	FG707981	-	-	-	EST

Amphibia						
<i>Xenopus laevis</i>	NP_001089508	NM_001096039	-	-	-	Annotated as a full length protein, but by similarity to other DIA1 genes it is not full length, and lacks sequence encoding the C-terminal ~100 amino acids of DIA1.
Mammalia						
<i>Cavia porcellus</i>	ENSCPOP00000008356	ENSCPOT00000009393	ENSCPOG00000009308	-	Yes	ENSEMBL identifiers
<i>Dasyops novemcinctus</i>	ENSDNOP00000013657	ENSNDOT00000017609	ENSNOG00000017610	-	Yes	ENSEMBL identifiers
<i>Echinops telfairi</i>						
<i>Equus caballus</i>	ENSECAP00000006978	ENSECAT00000009203	ENSECAG00000009060	-	16	ENSEMBL identifiers
<i>Erinaceus europaeus</i>						
<i>Felis catus</i>	ENSFCAP00000005560	ENSFCAT00000005987	ENSFCAG00000005985	-	Yes	ENSEMBL identifiers
<i>Loxodonta africana</i>	ENSLAFP00000003853	ENSLAFT00000004614	ENSLAFG00000004615	-	Yes	ENSEMBL identifiers
<i>Microcebus murinus</i>	ENSMICP00000011901	ENSMICT00000013056	ENSMICG00000013066	-	Yes	ENSEMBL identifiers Is almost full length, except for what appears to be a single nucleotide deletion in exon 1. This is probably a sequencing error.
<i>Mustela putorius</i>	-	GD183634	-	-	-	EST
<i>Myotis lucifugus</i>	ENSMLUP00000002419	ENSMLUT00000002662	ENSMLUG00000002662	-	Yes	ENSEMBL identifiers
<i>Ochotona princeps</i>	ENSOPRP00000005737	ENSOPRT00000006256	ENSOPRG00000006267	-	Yes	ENSEMBL identifiers
<i>Ornithorhynchus anatinus</i>	ENSOANP00000017787	ENSOANT00000017790	ENSOANG00000011229	-	Yes	ENSEMBL identifiers
<i>Oryctolagus cuniculus</i>	ENSOCUP00000010859	ENSOCUT00000012614	ENSOCUG00000012616	-	Yes	ENSEMBL identifiers
<i>Otolemur garnetti</i>	ENSOGAP00000010287	ENSOGAT00000011491	ENSOGAG00000011489	-	Yes	ENSEMBL identifiers
<i>Papio anubis</i>	-	FC112759	-	-	-	EST
<i>Pongo abelii</i>	-	CR856552	-	-	-	EST
<i>Procapra capensis</i>	ENSPCAP00000010296	ENSPCAT00000011030	ENSPCAG00000011077		Yes	ENSEMBL identifiers. Would be full length except for what appears to be a single nucleotide deletion in exon 1. This is probably a sequencing error.
<i>Spermophilus tridecemlineatus</i>	ENSSTOP00000014040	ENSSTOT00000015685	ENSSTOG00000015680	-	Yes	ENSEMBL identifiers
<i>Sus scrofa</i>	-	DN130313	-	-	-	EST sequence.
<i>Tupaia belangeri</i>	ENSTBEP00000013350	ENSTBET00000015387	ENSTBEG00000015399	-	Yes	ENSEMBL identifiers

<sup>a</sup>Orthologues are only present in the Metazoa.

<sup>b</sup>Chromosome location.

<sup>c</sup>Intron(s) disrupting coding sequence (if genomic sequence was available).

\*Classification of extant lamprey as Cephalaspidomorphi (Nelson JS, 1994<sup>¶</sup>) is controversial, and more recently lampreys have been classified in the class Petromyzontida or Hyperoartia [93].

<sup>¶</sup>Nelson JS (1994) *Fishes of the World* (3rd edition). New York: John Wiley and Sons.