

**Table 1. Exemplificative zebrafish genes with extended cDNA 5' region and deduced protein.**

<i>Gene</i> (RefSeq#)	Error type <sup>a</sup>	GenBank EST# Zebrafish <sup>b</sup>	Genomic clone #	Product length new/old (no. of new amino acids)	Kozak sequence old (top)/ new (bottom). Consense <sup>c</sup> : <u>GCCRCCATGG</u>	GenBank EST# Non-zebrafish
<i>selt1a</i> (NM_178290)	ND	CN505709 CK681469 CN018643 <sup>d</sup>	-	196/163 (33, +20%)	ATGA <u>A</u> AGATGC CTGATCATGG	-
<i>unc119.2</i> (NM_205713)	1	CN505408 CK363344 BI710727 <sup>d</sup>	BX465229 BX005137	264/206 (58, +28%)	GAGGCCATGA CGGATAATGA	pp DT261717 <sup>e</sup> pp DT134309 pp DT116366 pp DT263287
<i>nppa</i> (NM_198800)	1, 2	CN176149 CN180261 CO929886 <sup>d</sup>	BX323876	139/106 (33, +31%)	AGCAACATGG TCAGAGATGG	-

<sup>a</sup> (1) extended exon 1; (2) new exon; ND: not determined owing to unavailability of genomic sequence.

<sup>b</sup> GenBank sequences matching extended coding sequence from the new start codon in EST (Expressed Sequence Tag) division.

<sup>c</sup> The two most conserved positions (Kozak, 1999; Kozak, 2002) are underlined; start codon, in bold font.

<sup>d</sup> Only three representative sequences are listed, out of a total of 24 for *selt1a*, 4 for *unc119.2*, and 26 for *nppa*, showing consistent coding sequence extension (see Table online).

<sup>e</sup> pp=*Pimephales promelas* fish.