

zfish	YSRAQFGERLVKELLDMDDCSPTQ-----	518
fugu	YSRAQFGERLVRELELDDCSPTQ-----	498
rat	YSRAQFGERLVRELLGLEDESPA-----	523
mouse	YSRAQFGERLVRELLGLEDSSPAHELLRAMAGPYYQEPPTQGEE--	543
human	YSRAQFGERLVRELLGLEDASPTH-----	524
zfinch	YSRAQFGERLVKELLGLEDISPTH-----	516
chicke	YSRAQFGERLVKELLGLEDTSPSH-----	414
Xenopu	YSRAQFGDRLVRELLGLEDTSPCH-----	592
coelac	YSRAQFGERLVRELLGLEDSSTMH-----	523
lampre	YGRAQFGERLVQQLLRIPANGPATPSPGTGDAPGSTAAPGSTAAP	595
lancel	LSRSDFGAELLRAMAGPYYQEPPTQGEE-----	563

KEY: black carets, vertebrate introns; **blue caret**, lamprey intron apparently orthologous to lancelet although shifted by 3 residues; **red carets**, lancelet introns; **red underline**, lancelet protein sequence derived from genomic tandem repeats (Additional file 1); **red dashes**, 13 residue deletion resulting from exclusion of predicted lancelet exon 5 which is embedded within the 108 bp genomic tandem repeats and would, if included, result in the 57 residue insertion; **magenta underline**, predicted nuclear localization signal not conserved in active *Trichoplusia ni* transposase; yellow highlight, position of four conserved, catalytic aspartates in active piggyBac transposases and homologs including human PGBD1, 2, 3, 4, and 5; **XXX**, regions of known length but undetermined sequence arbitrarily positioned in the clustal alignment; zfish, zfinch, xenopu, coelac, lampre, lancel are zebrafish, zebra finch, *Xenopus tropicalis*, coelacanth, lamprey, and lancelet respectively. Amino acid residues are colored according to the EBI Clustal convention for side chains (red, AVFPMILW; blue, DE; magenta, RK; green, STYHCNGQ; others, grey). To avoid prejudicial judgments regarding the relationship between highly divergent sequences, we refrained from assigning a similarity or homology score to each residue.

Figure 1 (Continued from previous page.)