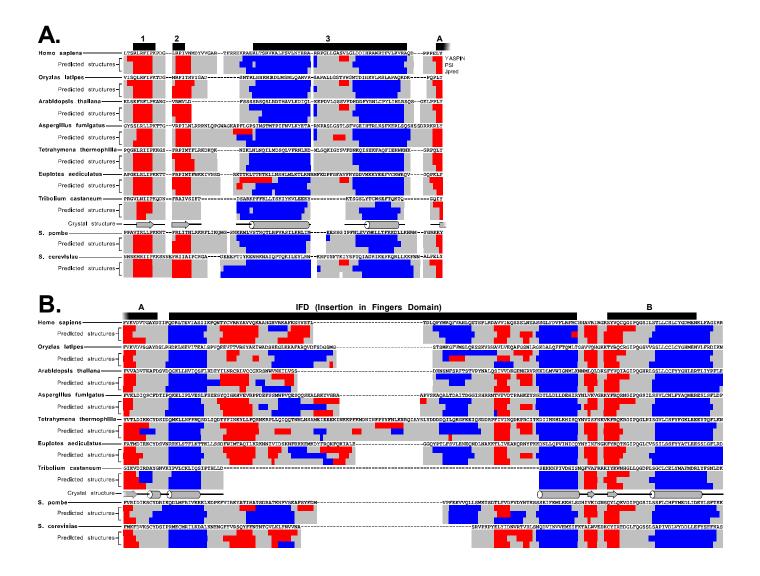
Supplementary Figure Legends

Supplementary Fig. S1. Secondary structure prediction of TERT motifs 3 and IFD. (A) Secondary structure prediction of TERT motif 1, 2, 3, A and B from representative species (*Homo sapiens, Oryzias latipes, Arabidopsis thaliana, Aspergilus fumigatus, Tetrahymena thermaphila, Euplotes aediculatus, Tribolium castaneum, Schizosaccharomyces pombe and Saccharomyces cerevisiae*). Predicted secondary structures (α -helices colored blue, β -sheets colored red, and random coil colored grey) and the crystal structure of *Tribolium TERT* (α -helices denoted by cylinders, β -sheets by arrows, and random coil by a black line) are shown below the amino acid sequence. The primary sequence of the entire RT domain (motif 1-E) was input in three online secondary structure prediction algorithms (YASPIN<http://zeus.cs.vu.nl/programs/yaspinwww>, PSI <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html> and JPred <http://www.compbio.dundee.ac.uk/~www-jpred>), and the output were combined. The predicted structures of motif 1-B are shown. (**B**) Secondary structure prediction of TERT motif A, IFD and B from representative species.

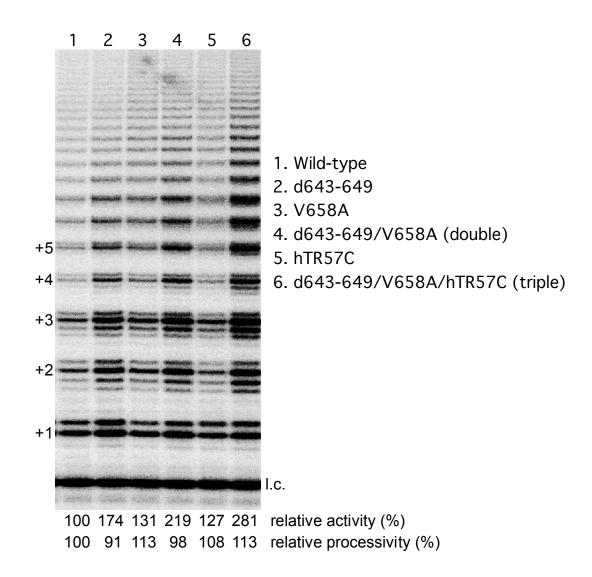
Supplementary Fig. S2. Multiple sequence alignment from motif 2 to motif A of TERT, other RTs and RdRp. The TERT sequences includes additional species along with the representative species from **Fig. 1B**. The RT sequences are grouped as two monophyletic clades based on their evolutionary distance to TERT. The secondary structures (arrows: β -sheets, cylinders: α -helices) shown below the sequences are derived from the respective crystal structures of *Tribolium castaneum* TERT (3DU6), HIV1 RT (1HYS) and FMDV RdRp (1WNE). Motif 2 and A have identity/similarity set at 55% conservation, shaded black/grey with invariant residues are shaded in dark cyan and dark blue respectively. Red/light red shading indicates the motif 3 TERT identity/similarity. An analogous helix-turn-helix sequence within closely realted RTs has been previously identified as motif 2a. The identity/similarity for the residues in motifs 2 and A are derived from all sequences, while the linker/motif 3 identity/similarity is derived from each grouping. Consensus residues with 55% identity are shown below each group.

Supplementary Fig. S3. Additive effects of motif 3 mutations on telomerase activity and processivity. Mutant telomerase were reconstituted in RRL and assayed by using the conventional telomerase activity assay. The telomerase mutants that contain different combinations of the motif 3 mutations (del-643-649, V658A and the double mutation del-643-649/V658A) or the hTR template mutation (57C) were assayed as indicated. The relative activity and processivity of different mutants are shown below the gel. l.c.: loading control. The signal of each repeat added was normalized with dGTP incorporated and the signal of first repeat. Log₁₀[normalized intensity] was then plotted against repeat number. Processivity was derived using equation processivity= $\frac{\ln 2}{(2.303k)}$, where k is the slope of each line.

Supplementary Fig. S4. Multiple sequence alignment of the IFD region of TERT. IFD is divided into IFD-a, IFD-b, and IFD-c based on predicted secondary structure. The predicted secondary structures (as in Fig. S1) of *Homo sapiens* and *Tribolium castaneum* TERT are shown below the amino acid sequence. The secondary structures (arrows: β -sheets, cylinders: α -helices) shown below predicted structures are derived from the crystal structures of *Tribolium castaneum* TERT (3DU6). Shading of residues indicates 55% identity/similarity (black/grey). The four mutations in the yeast, *Saccharomyces cerevisiae* TERT (Est2) shown to decrease processivity are colored red within the yeast sequence. The mutation (790-VVIE-793-4A) in human TERT tested in this study is also colored red.



				motif 2		motif 3		motif	А
					3a	3b	3c	ino (ji)	
Г	Г	Homo sapiens		LREIVNMDYVVGAR	TFRREKRAERLTSRVKALFSVLNYE	RARRFGLLGASVLGI	DDIHRAWRTFVLRVRAQD	PPPELYFVKV <mark>D</mark> VI	TGAYD
		Mus musculus		LR PIVNMSYSMGTR	ALGRRKQAQHFTQRLKTLFSMLNYE	RTKHPHLMGSSVLGM	INDIYRTWRAFVLRVRALD	QTPRMYFVKADVI	TGAYD
		Canis familia Bos taurus	ris	LRPTVNMD1IMGAR	TFHRDKKVQHLTSQLKTLFSVLNYE APPRDKKVQHLSSRVKTLFAVLNYE	RARREGUIGASMLG	DDTHRAWRIEVLRIRAON	PAPQLIFVKVDVI	VGAYD
	inverter	Gallus gallus		LREVVRLSRVVEGQK	LSKESREKKIQRYNTQLKNLFSVLNYE	RTVNTSIIGSSVFGF	DIYRKWKEFVTKVFESG	-GEMPHFYFVKGDVS	SRADD
		Xenopus laevi	S	LREISKISSTLSSO	QSKENQEKKIHHFSSQIRNLFSVLNYE	WNRNCSLIGSSVFGN	DDIYKKWKKFVLDFEKPQ	-VEKLQFYFVKT <mark>D</mark> VK	KGAYD
		Oryzias latip		MREITRVIGAD	SNTRLHHKRIRDLMSMLQAR AKTRLYQSHVRDLLDMLRAC	VRSAPALLGSTVWG	TDIHKVLRSLAPAQKDK	PQPLYFVKVDVS	SGAYD
		Takifugu rubr Danio rerio	ipes	MERITRVIGAD	DTLQYFQSC <mark>VRVL</mark> QNVLSVC	VCSTPSLLGSTVWGM	ITDIHKVLSSIAPAQKEK	PQPLYFVKMDVS	SGAYE
		Ciona intesti	nalis	LRPIMRLRN	ADKMPDMVPFEVIASVLKLV	CSKVEQATGFATOHE	AHIYNGWKLFVLSLAGIP	GRKRLFFVKMDIH	HRCED
-		. Ciona savigny	i	LRPIMKLRN	PKEPSSKMSFNSISSVLKLA	LQRFPDSLGFAIKDF	SDIYCKWRQFLLKLKEIN	-AKRWQLYFVKLDIC	CRCHD
(÷			rotus purpuratus		KKGLSVRLLLQDLFDVLTYH				
motif 3b		Aspergillus n Aspergillus f		VEPTINLERRTLVRSIYA	GKNRYHPAQSVNSAIAPVYSMLNYE	TARENDLLGSSMFS	GOMHSRLKKEKESLMSRGWD	ÖKKETAEAKTOIC	OSCED
ot				VRPILNLRRRMLQPGWA	GKAPFLGP <mark>SINSTMTPIFNVL</mark> KYE GKAPFLGP <mark>SINSTITPIFN</mark> VLNYE	KARNPASLGSTLSFV	GEIHTRLKSFKERLSQSHSS	D-RROPLYFVKLDIC	OSCED
	funç	Neurospora cr	assa	LRPIMNLRRRVLVRGLGS	KGIWGKQKQMLGPSINSVLGPVNSMLKFE	KRERPGRLGGGMFAV	GDIYQRVKEFRARVMSEVSK	KKGKK K FY FVKV <mark>D</mark> VC	OAADD
		L Encephalitozo		ARIISNMS	KARNGRPSINKSIYPEFCVLRHE	THGMLGNSILNE	ISGMYEKLSSYLSRSV	RPLYILKVDLS	SGCED
Ĕ.		Arabidopsis t Iris tectorum		VRMVLDFS	IYFYKSINTSLKELHAVLKTI	KOEHPOLLGSSVEDE	NETHKEWSOPLPKLRGRK	- GELPPLIFVVADVP	AKADD
	ante	Doryanthes ex		VRILANTKVPSKIPLHRN	NNRKRRFVFLKSINSSLKELHAILRRI	KHEHPQALGSSVFGY	DAYRKLYQFLPKVKEGS	- PMMPKVYIVVGDVS	SKAFD
F	- Pro-	L Oryza sativa		MRELVDLR	NNRKRRFVFLK <mark>SINSSLKEL</mark> HAILRRI AK <mark>S</mark> KDANLNKCHLIMKKL	RDEKPEMFGSSVFDY	NNVHQNLSQFISSKRSQL	-MKKLKVYIVVA <mark>D</mark> VS	SKAFD
		Tetrahymena t		FRPIMTFLRKDKQK	niklningi imdsgiyfrni tskisinrkivdikyii <mark>f</mark> rni rkttki <mark>ttntki</mark> inshi <mark>mi</mark> kti	KDMLGQKIGYSVFDN	IKQISEKFAQFIEKWKNK	GRPQLYYVTLDIK	KKCYD
	10	Paramecium ca Euplotes aedi		FREDVTYNRKAK	BKTTKLTINTKLINSHLMIKTL	KNBMEKDPECEAVEN	PEVFSRLEEPKKLWIKL	QQPKTYYITMDIQ COPKIEFATMDIE	QKCYD
	cillia	Oxytricha tri		FRPIMTFNRKIPNQV	GKFQSRMTTNNKLQTAHMMLKNL	KSKMFKHSFGFAVFNY	DDIMKRYENFVQKWKQI	NSPKLYFVAMDIE	EKCYD
		Consensus		RPI	VL	P LGSSV	DI F	LYFVK DV	A D
	Γ	Apis mellifer	a		FRDTDYLFIIFKFIKQI				
		Bombyx mori Tribolium cas	taneum	CRPIVKYRN	DELNKRQKYR <mark>IKEKI</mark> QLI DSARKPFFKLLTSKIYKVLEE	KLLTGNQQ	AKIETAYPKLHSKWLELNK-	PKLYFIKTDLS	SDADG
() A			taneum structure					GQ11G1RVD1	
motif 3b		Caenorhabditi	s elegans	LRELFKRKA	IDKKTETMQWKRLNSMISWC	LERSGVYRHT	RDSCKKVSDFLKKNS	QNSKIIGYTADVS	SKCES
lot	des	Caenorhabditi		FRPIIRRNP	IDKVKDKMHWKKVNSMLTWC	LEKGGETRQ <mark>S</mark> I	NSSCKILLKFLRRND	TFPKLFGYTA <mark>D</mark> VS	SKCES
-	remate	Schizosacchar Saccharomyces			MGSNKKMLVSTNQTLRPVASILKHI DEEEFTIYKENHKNAIQPTQKILEYL				
	()-	Ashbya gossyp			DEEDKREYEDYVRTVIKPVQQILEYL				
	ueast	Kluyveromyces		FRIMAVPCKGV	NIQEFNEFNDYVTNALKPTKIILER	RMRRNTKFTKAF <mark>S</mark> I	LEIPKIVLSYKQRLLLKHG-	KIPNLYLLKF <mark>D</mark> IÇ	QSCYD
	L '	Candida albic	ans		LDTLEKEKREFERYRKEV <mark>LSPV</mark> GQ <mark>IL</mark> RLK	LSKLRDTYESYRASVH <mark>S</mark> S			
_		^{Consensus}		FR I motif 2	L	motif 2a	FL	PKLY K D motif	ACY
						mour za			
		Г	CoprinaCc1	MRPIIPCH	SAIQNPAAKFV <mark>SKNLK</mark> SVSQAPSATYV <mark>SKNLK</mark>	FLIKESPTILHGSKDLAG	KLSNVKLKP	GRRWFFISGDVV	VAYYP
	,		CoprinaPc1	MRPIVPCH	SVSOAPSATYVSKMLK	PLVSROPHVIHCSKHLVF	QLESLQLDK		
	Ĥ					· · ·		HKKAWI V3GDI V	VAFYP
		PLES	AthenaAvM	LRPIVAGM	KSPTIGISKWLDDLLRPLF	NQLALDIAIENGTQLIKS	VERWSDNYL	TSSTTFISMDVT	VAFYP TDLYT
	7	Z officient PLES	AthenaAvM AthenaAvO F msEc67	LRPIVAGM	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI	VERWSDNYI LEKWSKQNI IILNAYKHR	TSSTTFISMDVT NROTOFATMDVS	FDLYT SDLYT
	1	ENdeficient PLES	AthenaAvM AthenaAvO msEc67 msMx65	LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		EN-daficient PLES		LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		<u> </u>	IA-Pa	LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		<u> </u>	IA-Pa b1-Sp	LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		<u> </u>	IA-Pa	LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		group II introns	IA-Pa b1-Sp PtI-An R1Bm R2Dm	LRPIVAGM LRPIVAGL VRTISAPT	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF DRLKDIQRRICDLLSDCRDEIF	NQLALDIAIENGTQLIKS DRLAKETSILNGVELIDI AIRKISNNYSF <mark>GF</mark> ERGKS	IILNAYKHR	TSSTTFISMDVT NRQTQFATMDVS GKQIILNIDLK	FDLYT SDLYT KDF <mark>F</mark> E
		group II introns	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs	L DIVACM L DIVACL	KSPTIGISKWLDDLRPLF 	NQLAIDIAIERGTQLIK DRLAKETSILGVELIDI AIRKISNNYSFGPERGKS VSKLEVVSPQALGEVPGRS ITPEKMSTRSIGPRNMS IYEPIFLDVSHGPRPKRS IYEPIFLDVSHGPRPKRS IYEPIFNTASEGFRRGRS MAEEVADVSYGPRPMR GLTHSISPRQLGFSPGRS NSSINMDPRQLGFLPTDC HIKKLIHHOVGFIFGK	IILNAYKHR IKTGAAPHV CCTATUEVENMF CHTALHQISKW CHSALHQISKW VSWAVCRVINGLNNP TVTALRTILDVSRAS CADNATIVDLVINHSHKHF CADNATIVDLVINHSHKHF		IDLYI SDLYI KDFFF KKCFD KKCFD KGCFD KGCFD SGAFD SKAFD EKAFD
		group II introns	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey	L DIVACM L DIVACL	KSPTIGISKWLDDLLRPLF KSPTIEISRWLDSILRPLF	NQLAIDIAIERGTQLIK DRLAKETSILGVELIDI AIRKISNNYSFGPERGKS VSKLEVVSPQALGEVPGRS ITPEKMSTRSIGPRNMS IYEPIFLDVSHGPRPKRS IYEPIFLDVSHGPRPKRS IYEPIFNTASEGFRRGRS MAEEVADVSYGPRPMR GLTHSISPRQLGFSPGRS NSSINMDPRQLGFLPTDC HIKKLIHHOVGFIFGK	IILNAYKHR IKTGAAPHV CCTATUEVENMF CHTALHQISKW CHSALHQISKW VSWAVCRVINGLNNP TVTALRTILDVSRAS CADNATIVDLVINHSHKHF CADNATIVDLVINHSHKHF		IDLYI SDLYI KDFFF KKCFD KKCFD KGCFD KGCFD SGAFD SKAFD EKAFD
		group I utrons	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV	L PLVAAM L PLVAAL		NOLAIDIAIELGTOIIS DELARETSILGYELIU AIRKISNNYSTÖERGKS VSKLPVSPQAHOPVPGR IVEDISISSI IVEDISISSIGARRKRS MAEEVADVSSYGERRKRS GLTHSISPRQHORSPGRS NSINNDPRGHOFSPGR HIKKLIHHDQYGFIFGKC DVTKAIPKPGPGPELQHG GF	IILDAYKHR IRGAAPHV	TSSTTFISMOUT NRQTGFATMOVS RGRVIKLDLK GRRVVIKLDLK GRSNWFIEVOLK NGCTWMIEGDLK KGCTWMIEGDLK RSCTWINEGDLK RSCTWINEGDLK RSCTVINIIGDLK RSCTVINIIGDLK RSCTVINIIGDLK RSCTVINIIGN RSCTVINIIGN RSCTVINIIGN 	IDLYT SDLYT KDF9E KKCFD KKCFD KGCFD KGCFD SGAFD SKAFD EKAFD 20AFD F
		group I utrons	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV	L DPVACM L DPVACL L DPVACL LLHAP LLHAP LLHAP LLTAP		NOLAIDIAIENGTOIIS DRLAKETSILGVELIDI AIRKISNNYSJÖERGKS VSKLPVSPOAFGPVPGRS IIVEPIFLOVSBÖERKKS IIVEPIFLNSBÖERKKS MAEEVADVSSYÖERPMRN GITHSISPRÖEGSFGRS NSSINNDPRGESLPTOC HIKKLIHHDOVGFIFGK DVTKAIPKFOFGFRLOHG GF	IILDAYKHR IKTGAAPHV	TSSTFFISMOV NRQTGFATMDV3 GRQIILNIDLK GGRNVLKIDLK GGSNMFIEVDLK NGCTWMIEGDLK KGCTWMIEGDLK EQRYVMAIFDUS RSCYIANLDV3 RSCYIANLDV3 KINHMIISIDA D NLSWISIDV3 NLSWISIDV3	IDLYT SDLYT KDF9E KKCFD KKCFD KGCFD SGAFD SKAFD SKAFD EKAFD QQAFD F SAAEY SAAEY
	ETG Antice Manufactories	2000 group introns non-LR territronsposis	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia	L EPVAAM L EPVAAL		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIUD AIRKISNNYSFGPERGKS VSKLPVSPQALGVYGRS IIPPKMNSHSGPRNMS IIPPLFLVSHGPRPKRS IIPPKLSSFGPRFRS SINWDPRQFGFLPHD HIKKLIHNGVGFIFGK GF 	IILNAYKHR IKTGAAPHV CCTAILEVENMF	TSSTTFISMOT NRQTGFATMOV3 GKRVILKIDIK GGRWIFISVDIK GGSNMFIEVDIK NGCTWMIEGDIK KGCTWMIEGDIK RSCYIANLDV3 RSCYIANLDV3 RSCYIANLDV3 NLSWISLDV3 NLSWISLDV3 VILKVMODVK	IDLYT SDLYT KDFFF KKCFP KKCFP KGCFP SGAFD SGAFD SKAFD SKAFD SAAFY SAAFY KTABL
	ETG Antice Manufactories	2000 group introns non-LR territronsposis	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia Tyl	L DPVACM L DPVACL L DPVACL V T_SAPT		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIU AIRKISNNYSFÖERGKS VSKLPVSPQAHGPVPGRS ITVEPIFLOVSFÖRRKRS ITVEPIFLOVSFÖRRKRS MAEEVADVSSYÖERRMRR GLTHSISPRQHGPSPGRS NSSINNUPRQHGFJERGK DVTKAIPKPQFÖERLQHG GF	IILNAYKHR IIKGAAPHV	TSSTTFISMOVI NRQTGFATMOVS ORQVILLNIDLK GGSNWFIEVOLK NGSTWMIEGDIK KGCTWMIEGDIK KGCTWMIEGDIK RCSTVMIEGDIK RCSTVMIEGDIS RCSTVMIEGDIS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS NLSWISIOVS	IDLYT SDLYT KDFFF KKCFP KGCFP KGCFD KGCFD SGAFD SGAFD SGAFD SGAFD SGAFD SAAFD F SAAFY SAAFY SAAFY SAAFY SSAYL SSAYL
	ETG Antice Manufactories	2000 group introns non-LR territronsposis	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia	L EPUVACM		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIUD AIRKISNNYSJÖERGKS VSKLPVSPQAFGPVPGRS ITVEPIFLSVSFGFRKNS ITVEPIFNISSFGFRKNS ITVEPIFNISSFGFFFSGS MAEEVADVSSYGFRFNR GITHSISFGGFFSGSS NSSINNDFRGFFFLQHG GF GF 	IILNAYKHR IKTGAAPHV	TSSTFFISMOV NRQTGFATMDV3 GKQIILNIDLK GGRWVLKLDLK GGRWVLKLDLK NGCTWMLEGDLK KGCTWMLEGDLK KGYVWAIEDIK RCYVWAIEDIK RCYVWAIEDIK RCYVWAIEDIK RCYVWAIEDIK RCYVWAIEDIK RCYVWAIEDIK NLWLSLDV3 NLWHESLDV3 NLWHESLDV3 NLWHESLDV3 NLWHESLDV3 NLWHESLDV3 NLWHESLDV3 NLWYHTCLDIS	IDLYT SDLYT KDF9E KKCFD KGFFN KACFD KGCFD SGAFD SKAFD EKAFD 20AFD F SAAEY SAAEY SAAEY SSAYL SSAYL
	ETG Antice Manufactories	BOD	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV Copia Tyl Gypsy Ty3 Copia	L EPUVACM		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIU AIRKISNNYSTGERGKS VSKLPVSPQAHGPVPGRS ITVEPIFLOVSIGGRRKNS ITVEPIFLOVSIGGRRKNS MAEEVADVSSYGERMRN GLTHSISPROFGESFRGS NSINNDPROFGESFRGS SSINNDPROFGESFLQHG GF	IILNAYKHR IIKGAAPHV		IDLYT SDLYT KDFBP KKCFD KGCFD KGCFD SKAFD
	ETG Antice Manufactories	BOD	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV Copia Tyl Gypsy Ty3 Copia	L EPUVACM		NQLAIDIAIENGTQLIKS DRIAKETSILGVELIU AIRKISNNYSJÖEREGKS VSKLPVSPQALGPVPGRS ITVEPIFLOVSIGFRANKS ITVEPIFLOVSIGFRAKS GRANG GITHSISPRQEGESPGRS NSSINNDPRGEGFLQEG GF VTKAIPKFQEGERLQEG GF SSINNDPRGEGFLQEG SSINNDPRGEGFLQEG SF SSINNDPRGEGFLQEG SF SSINNDPRGEGFLQEG SF SSINNDPRGEGFLQEG SF SSINNDPRGESSINNDPRGESSINNDPSG SSINNDPRGESSINNDPSG SSINNDPRGESSINNDPSG SSINNDPRGESSINNDPSG SSINNDPRGESSINNDPSG SSINNDPSG	IILNAYKHR IKTGAAPHV		IDLYT SDLYT KDF5P KKCFD KGCFD SKAFD SKAFD SKAFD SKAFD SKAFD SAAFY KTAFL SSAFY KTAFL SSAFY KTAFL SSAFY KSGYH HSGYH KSGEW
	ETG Antice Manufactories	COSCI Development Property and the second representation of the	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copla Ty1 Gypsy Ty3 CaMV CoYWV MWTV	L EPUVACM		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIDI AIRKISNNYSPÖERGKS VSKLPVSPQAHGDVPGRB ITDKKMSTHSIGBRCKS IYEDITLDVSHGTRRKRS IYEDITLSSGTRRKRS MAEEVADVSSYGTRRMKN GLTHSISPRQHGDSPGRS NSSINNDPRGFGEPGRQ DVTKAIPKFQFGERLQHG GF	IILINAYKHR- IIKGAAPHV	- TSSTFFISMOT - NRQTGFATMDVS 	IDLYT SDLYT SDLYT KKCFD KKCFD KKCFD SGAFD SGAFD SGAFD SGAFD SGAFD SCAFD SCAFD SCAFD SCAFD SCAFD SCAFD SCAFD SCAFD KCGFW KSCFW KSCFW KSCFW
	ETG Antice Manufactories	COSCI Development Property and the second representation of the	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copla Ty1 Gypsy Ty3 CaMV CoYWV MWTV	L EPVAGM		NQLAIDIAIELGTQLIKS DRLAKETSILGYELIUL AIRKISNNYSTGEBERGKS VSKLPVSPQAHGEVPGRS ITVEPIFLOVSTGERKKS MAEEVADVSSTGERKKS MAEEVADVSSTGERKKS MAEEVADVSSTGERFMRN GITHSISPRQHGESGENC HIKKLIHHDQYGERLQHG GF	IILNAYKHR IKTGAAPHV		IDLYT SDLYT KDFTE KDFTE KGFTP KKCFD KGCFD SGAFD SGAFD SSAFD QQAFD F SAAFY SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY SSAF
	ETG Antice Manufactories	COSCI Development Property and the second representation of the	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copla Ty1 Gypsy Ty3 CaMV CoYWV MWTV	L DPVACM L DPVACL L DPVACL LTAP		NQLAIDIAIELGTQLIKS DRLAKETSILGYELIUL AIRKISNNYSTGEBERGKS VSKLPVSPQAHGEVPGRS ITVEPIFLOVSTGERKKS MAEEVADVSSTGERKKS MAEEVADVSSTGERKKS MAEEVADVSSTGERFMRN GITHSISPRQHGESGENC HIKKLIHHDQYGERLQHG GF	IILINAYKHR IIILNAYKHR CQTAIWEVRNMF CQTAIWEVRNMF CQTAIWEVRNMF CHIALQISKW CHIALQISKW USANGRUINGLINF CHIALQISKW CADNATIVDLVLRHSKHF GMFNITKSINVIQHINRAND TFPEQLHRVNFALEAMEN- TPEQLHRVNFALEAMEN- PFKFAVGNLQSITNLISS- PFKFAVGNLQSITNLISS- PFWFAVGNLQSITNLISS- PFWFAVGNLQSITNLISS- PFWFAVGNLQSITNLISS- PFWFAVGNLGSITNLISS- -FFUGRIDNLSRIGN- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTISSLALD- -YNLGRINTSSL		IDLYT SDLYT KDFTE KDFTE KGFTP KKCFD KGCFD SGAFD SGAFD SSAFD QQAFD F SAAFY SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY KTAGL SSAFY SSAF
	ETG Antice Manufactories	BOD	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CaMW CovMV MMTV HIV1 structure Consensus	L DIVACM L DIVACL VI TLSAPT VI TLSAPT VI TLSAPT VI TLSAPT VI TLSAPT VI TLSAPT VI TLSAPT VI TLSAPT VI TLSA VI	KSPTIGISKWLDDLRPLF 	NQLAIDIAIENGTQLIKS DRLAKETSILGVELIUD AIRKISNNYSJÖERGKS VSKLPVSPQAHGPVPGRS ITVEJIFLOVSHÖRRKKS ITVEJIFLOVSHÖRRKKS MAEEVADVSSYGBRPKKS MAEEVADVSSYGBRPKKS MAEEVADVSSYGBRPKKS MAEEVADVSSYGBRPKLOHG MIKKLIHHDQVGFIFGKG OVTKAIPKFQFGBRLOHG GF	IILINAYKHR IIKIGAAPHV	TSSTTFISMOT NRQTGFATMDV3 GRQIILNIDLK GRRVVLKIDLK GGSNMFIEVDLK NGCTWMIEGDLK KOCTWMIEGDLK RCSTWMIEGDLK RCSTWMIEGDLK RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIEDL RCSTWMIESD NLSWIESLOVS NLSWIESLOVS NLSWIESLOVS NLSWIESLOVS NLSWIESLOVS SKIYSKFDLK SKIYSKFDLK SKIYSKFDLK SKIYSKFDLK SKIYSKFDLK KKSVTVLDVG LD	DLYT SDLYT SDLYT KKCFD KKCFD KKCFD SGAFD SGAFD CKCFD SGAFD CCC SGAFD CCC SGAFD SGAFD SGAFD SGAFD SSAFL
	ETG Antice Manufactories	COSCI Development Property and the second representation of the	IA-Pa b1-Sp PtI-An R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CaVMV MULV MIVI HIV1 structure Consensus	L DPVACM L DPVACL L DPVACL LTAP		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIUD AIRKISNNYSFOBERGK VSKLPVSPQARGPVGRS IVEPIFLOVSFORPKRS IVEPIFLOVSFORPKRS IVEPIFLOVSFORPKRS MAEEVADVSSVGBRPKN GITHSISPRQFGSFGS NSSINNDPRGFGFGRLQRG GF 	IILINAYKHR IIILNAYKHR CQTATHEVRNMF CQTATHEVRNMF CQTATHEVRNMF CGHALSIFTNF CHALRQISKM CHALSIFTNF CHALSIFTNF CHALSIFTNF CHALSIFTNF CHALSIFTNF CADNATIVDLVLRHSHKHF- GWFNICKITNVFALEAMEN PFKAVENLOSETNLLSS PFKAVENLOSETNLLSS PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIQ PFVRSTFILSIVIC PFVRSTFI	TSSTTFISMOT NRQTGFATMDV3 GRVUILIDIK GRRVUIKIDIK GRRVUIKIDIK GGSNMFISVDIK KGTTWMLEGDIK KGTTWMLEGDIK RSCYIANLDV3 RSCYIANLDV3 RSCYIANLDV3 NLGWLSLDV3 NLGWLSLDV3 NLGWLSLDV3 NLGWLSLDV3 NLGWLSLDV3 NLGWLSLDV3 SIIGVFTIDIH SKIYSKPDK SKIYSKPT 	DLYT SDIYT S
	ETG Antice Manufactories	BL DEPUT	IA-Pa bl-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CoMW CoYMV CoYMV MMTV HIV1 HIV1 structure Consensus MS2 TEV	L EPVAGM		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIDI AIRKISNNYSTGERGKS VSKLPVSPQAHGPVPGRS ITVEPIFLOVSIGERKKS MAEEVADVSSTGERKKS MAEEVADVS	IILINAYKHR IIKTGAAPHV	TSSTTFISMOT NRQTGFATMDVS NRQTGFATMDVS GRSNVIKIDIK GRSNVIKIDIK GRSNVFIEVDIK LANYQYVETDIK LANYQYVETDIK RSCITANLDVS RSCITANLDVS KEITANLDVS NLSWISIDVS VILKUHQMDVK NLSWISIDVS VILKUHQMDVK NLSWISIDVS SILYSKPDK SKIYSKPDK SKIYSKPDK GSLATTDLS SGSLATTDLS	DLYT SDIYT S
	ETG Antice Manufactories	BL DEPUT	IA-Pa b1-Sp PtI-An R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CaMW Corsensus HIV HIV1 HIV1 HV1 structure Consensus MS2 TEV FMDV FMDV structure	L DPVACM L DPVACL L DPVACL L DPVACL L DPVACL L LHAP- LLHAP- LLHAP- LLHAP- LLHAP- LL LHAP- LL LHAP- LL LHAP- LL LHAP- LL LHAP- LL LL LL LL LL LL		NQLADDIATERGTQLIKS DRLAKETSILGVELIDI AIRKISNNYSFÖERGKS VSKLPVSPQAFGPVPGRS ITVEPIFLOVSFÖRRKRS ITVEPIFLOVSFÖRRKRS MAEEVADVSSYÖFRAMKR GLTHSISPROFGFSFRAMK GLTHSISPROFGFSFRAMK SINNDPROFGFSFLQHG GF 	IILINAYKHR IIKIGAAPHV	- TSSTTFISMOT NRQTGFATMDVS NRQTGFATMDVS GRQVIKIDLK GRSNVFIEVDLK GRSNVFIEVDLK NGTTWMIEGDIK LANYQYVEIDLK RCTTWMIEGDIK RCTTWMIEGDIK RCTTMICO RCTTMICO RCTTALLOVS RCTTAL	DDLYT KDFFB KDFFB KGCFD KGCFD KGCFD KGCFD EKAFD EKAFD EKAFD EKAFD F F SAAFY KTAPL SSAFY KTAPL SSAFY KTAPL SSAFY F F DCSF F F DCSF F F DCSF SSASD SSASD SSASD SSASD
	ETG Antice Manufactories	COSCI Development Property and the second representation of the	IA-Pa b1-Sp PtI-An R1Bm R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CaMV CorMV MMTV HIV1 structure Consensus MS2 TEV FMDV structure Polio	L EPVAGM L EPVAGL V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TLSAPT V TL V TL	KSPTIGISKWLDDLRPLF 	NQLAIDIAIENGTQLIKS DRLAKETSILGVELIDI AIRKISNNYSFÖERGKS VSKLPVSPQAHGPVPGRS ITVEPIFLOVSFÖERGKS MAEEVADVSSYGERPKRS ITVEPIFLOVSFÖERKS MAEEVADVSSYGERPMRN GITHSISPRQFÖEFGKG MAEEVADVSSYGERPMRN GITHSISPRQFÖEFGKG DVTKAIPKFOFGFRLQHG GF 	IILNAYKHR IIKGAAPHV	TSSTTFISMOT NEQTGFATMOVS OKQIILNIDLK OGRAVIKIDLK OGRAVIKIDLK OGSTMMLEGDLK LANYQYVETDLK EQRYVMAIFDDLQ RSCITANLDYS KINHMISIDAS KINHMISIDAS VILKULSDYS VILKULSDYS VILKULSDYS VILKULSDYS SKIYSKPDLK SKIYSKPDLK SKIYSKPDLK SGLATTDDS QYKNVMDVD-Y VELFAPD-Y	DDLYT KDFGE SDLYT KDFGE KGFD SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAES SGAE SGAE SGAE SGAE SGAE SGAE SGA
	ETG Antice Manufactories	BL DEPUT	IA-Pa b1-Sp PtI-An R2Dm L1Hs Jockey Consensus HBV WHV Copia Ty1 Gypsy Ty3 CaMW Corsensus HIV HIV1 HIV1 HV1 structure Consensus MS2 TEV FMDV FMDV structure	L DPVACM L DPVACL L DPVACL L DPVACL L DPVACL L LHAP- LLHAP- LLHAP- LLHAP- LLHAP- LL LHAP- LL LHAP- LL LHAP- LL LHAP- LL LHAP- LL LL LL LL LL LL		NQLAIDIAIENGTQLIKS DRLAKETSILGVELIUD AIRKISNNYSJÖERGKS VSKLPVSPQALGPVPGRS IVSPLFVSPQALGPVPGRS IVSPLFNDSSGÖRRPKRS IVSPLFNTASIGGRPGRS MAEEVADVSSVGPRMKA GITHSISPRQLGFGSPGRS NSSINNDPRGGFJEFGRG DVTKAIPKFQCGFLQHG GF 	IILINAYKHR IIKIGAAPHV	TSSTTFISMOT NEQTGFATMDV3 OKQIILNIDUK OKQIILNIDUK OGSNWFIEVDLK NCGTWMIEGDIK KGTTWMIEGDIK KCTWMIEGDIK KEYAVGAFLDIC D RSCYIANLDV3 KHIMISIDV3 KHIMISIDV3 NLSWISIDV3 NLSWISIDV3 NLSWISIDV3 NLSWISIDV3 NLSWISIDV3 NLSWISIDV3 NLSWISIDV3 AVFTIDLK AVFTIDLK AVFTIDLK SKIYSKPDLK SKIYSKPDLK GSLATDL3 CSLATDL3 QYKNVNVD-Y CKIFAFD-Y	DDLYT KDFGE SDLYT KDFGE KGFD SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAED SGAES SGAE SGAE SGAE SGAE SGAE SGAE SGA



	motif A	Insertion in Fingers Domain (IFD)	motif B
	IFD-a	IFD-b	IFD-c
Homo sapiens —	<u>IYFVK</u> VDVTGAYDTIEQDRITEVIASIIKPQNTYCVRRYAVVQKA	AHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQ	SSSLNEASSGUFDVFLRFMCHHAVRIRGKSYVOCO <mark>GIPOGS</mark> ILSTLLCSLCYGDMENK
Predicted structures	PN PS		
Mus musculus		DSQGQVHKSFRRQVTTLSDLQPYMGQFLKHLQDSDASALRNSVVIEQS	SISMNESSSSLFDFFLHFLRHSVVKIGDRCYTOCOGIPQGSSLSTLLCSLCFGDMENK
Gallus gallus —		ITPTGKARKLYKRHVSTFEDFIPDMKQFVSKLQERTSLRNAIVVEQ	
Xenopus laevis — — — — — — — — — — — — — — — — — — —		VDPTGRIIKSFKRHVSELADVLPNMKQFVSNQQEKNLLRNTILVEQ	NLLLNESSVKILAVFQQIIRSHILRIKDRYYMQCCCGIPQGSMLSTILCSLCYGDMENA
Oryzias latipes	LYFVKVDVSGAYDSLEHDKUKEVITEALSPVQEEVFTVRHYAKIW	ADSHEGLKKAFARQVDFSDGSMGSTSMKGFVMSLQKSSKVHHAVLVEQ	AFGSNLRGKDALQFFTQMLTGSVVQHGKKTYRQCRGIPQGSVVSSLLCCLCYGHMFNV
Takifugu rubripes		ADSHEGLKKAFIRQADFLEANMGSINMKQFLTSLQKKGKLHHSVLVEQ	
Danio rerio		SDPTRGLRKRFCTKAEMSEPLNMKGFVVDEQVSGRLHDAILVE	RHSSEVRGGDVFQFFQKMLCSYVIHYDQQMFRQVCGIPQGSSVSSLLCNLCYGHM ^D KA
Strongylocentrotus purpuratu	us- LYFVKIDIEKCYDSIKHSKILQIISMLLQGHDKPDEYQLQRYVTV	TRAASASTGLQKRSHRHVTTELSSFHRQLIQMAHHGKIKNAIIINQ	JHTVKVTPKDULQRLKQHVMADVVKSGRKYMWRQDGISQCSILSSLICSFFYAHLFRC
Ciona intestinalis —		LTSATKGKL	ICKVKHCTHCILTTLKYFLNKVTLSVLSKKFVKLH <mark>GIPQGWELS</mark> NLICDIFYGHMPQQ
Apis mellifera		YYTMKSEISKKDNTYHKYFSDLQLPFPSDQFYANLVDN	2WYQLINKSWILESISKYIFYQRIKINKNIHIIGK <mark>EIVQES</mark> MISPINSDIYYNYILNK
Bombyx mori			
Tribolium castaneum			LLDSEKKNFIVDHISNQFVAFRRKI <u>MKWNHGLLOC</u> DP <u>US</u> GC <u>HC</u> ELYMAFMDRL
Predicted structures	PSI		
Crystal struct			
Caenorhabditis elegans	IIGYTADVSKOFSTVNHDVIISIIDRLFSOEHD	IYTVCGKGRNHGGFHKLIFCSAGTELNAHEALRRKMELKGVFNFEV	CYREMSSSTT YSVIRTTLSTYYYKRGPTSWRITK VPOCHPISSN AHMYLNNF POK
Schizosaccharomyces pombe —		YATIHATSDRATKNFVSEAFSYFDMVPFEKVVQLLSMKTSDTLFVDFV	VDYWTKSSSEIFKMLKEHLSGHIVKIGNSOYLOKVGIPOGSILSSFLCHFYMEDLIDE
Saccharomyces cerevisiae	UYEMKEDVKSCYDSIERMECMRILKDALKNENGFFVRSQYFFNT	NTGVLKLFNVVNASRVPKPYELYIDN	VRTVHLSN <mark>ODVI</mark> NVVEMEIFKTALWVEDKCYIREDCLFOGSSLSAPIVDLVYDDLLEF
Saccharomyces castellii		KTKTFKVKNIVNGSRYPKSDEIYIDN	VRTAHLSTEDVTNVLEMELFRTALWVGNQCYLRKDCLFQCSSLSAPIVDLVYDDLLEY
Ashbya gossypii		KHHIIKRKSAVNGNLQASDGSIIIDNS	
Kluyveromyces lactis ———		ASTRLKRLHAVNGDIQFQENDLVIDSV	VKTTVLSKEDVLTVVKMELFKTSIFYRGKCYLRKDELFOCTPLSSVIVDILYDDLLEF
Candida albicans		HKLKKVKTTIDTQYHNLNILSSSRHLSNCKSLVDK	IKTIALQKGNILEVCRSQIYDVVGSV6LHLYKRKREVFQGFSLSSIFCDILYSAMVHD
Yarrowia lipolytica		SKTVLRWFTTSRACNGRSEGALQAFRRLGNKGIIVDEN	
Aspergillus nidulans		NMWPLRKPQQRRTWSKYLQRVGPVGRPENLADAIANGSVVGRRNTVLVDT	
Aspergillus fumigatus — — —		SMWPVQESQQSKALRKYVGRAAPVSKAQALTDAITDGGISHRRNTVFVDTI	
Neosartorya fischeri		SMWPVQESQQSKALRKYVGRAAPVSRAQDLTDAITDGGISHRRNTVFVDTI	PAHKEYSGDYLLDLLDEHIRYNLVKVGRKYFR <mark>ORNGIPQGS</mark> ILSSVLCNLFYAQMFRE
Neurospora crassa —		-NNDDTNIDNPKYKLKPTKRWHSTTTSFPSITTTTTTTSTTSASISILPQQQDSVATTKKNTLFIPSH	3STKLHTSSALLSLAKEHITQNLVKIGKKYYRQKT <mark>GIPQGSVLSS</mark> TLCNYFYADL <mark>F</mark> RS
Encephalitozoon cuniculi		GGELRSRYVCKVTENAETINELMMEPGAFVNKVVKEN	
Arabidopsis thaliana		GKRSNWVNKILVSSDKNSNFSRFTSTVPYNALQSIVVDK	
Doryanthes excelsa		TKGIKTFYEQVSSDHSSSSNDIINFEALTRLRPSYSVYIDQ	
Iris tectorum		CALANCE CONTRACTOR CONTRACTION	
Oryza sativa		ASIGNGNSIYDLSIQLSSGGGIFVDQ	
Tetrahymena thermophila		NNLNSAMEIEEEKINKKPFKMDNINFPYYFNLKERQIAYSLYDDDDQILQKGFKEIQSDDRPFIVINQI	
Paramecium caudatum	TYYITMDIQKCYDTHILQKYLQFIEESKQFSSIYAINKY	HIVSRNNRMLKPSFSMKDLFNILDRTCAIPFNKPQLLQKGYIEHIQKNKQTIIINQ	JFQNSVTFSEFLNSIKNICQNNIVQFENRYFROTLGIPQGLNISGILCSFYLANLFQK
Euplotes aediculatus		KNNIVIDSKNFRKKEMKDYFRQKFQKIALEGGQYPTLFSVLENEQNDLNAKKTLIVEA KNNIVERSNFRKLPIKQYFRYKFQKIGIDGSSYPTLFEILEDEFNDLNMKRTIVEQ	KQRNYFKKDNLLQPVINICQYNYINFNGKFYKQTKGIPQGLCVSSILSSFYYATLJES
Oxytricha trifallax		KNNIIVERSNFRKLPIKQYFRYKFQKIGIDGSSYPTLFEILEDEFNDLNMKRTIIVEQ	<u>JQRKKFPKNDLLQPVLKICQNNYVTFNKKQYKOMKGIPQGLCVSYITSSFYYANIJEN</u>
Giardia intestinalis		KMPKIRTVALPCQHNGCVSMDSISRYLLCKETPKKQPNSMLCP	MHARIYKREDIIKMLELHLLNPLVIHEGGVURLTS <mark>GIPQGS</mark> VVSTVUFNCYSSLLHVQ