

Additional file 10

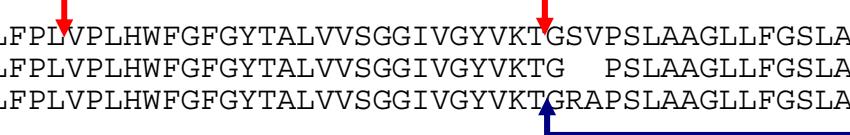
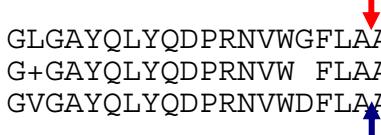
Protein-level alignments of intron gain retrogenes (“Sbjct”) and their parents (“Query”) by GeneWise [S1].

Each alignment shows the protein sequence of the parental gene on the first line, followed by a line indicating the similarity level of the match followed by a line representing the translation of the DNA sequence of the retrogene. If the intervening sequence of the retrogene is long, it is indicated in the square brackets. If the intervening sequence of the retrogene is short and not divisible by 3, it is indicated by '!'. (for specifics, see the description of the output format of GeneWise (<http://www.ebi.ac.uk/Tools/Wise2/>)). Red downward arrows indicate the exon-exon borders of parental genes and the blue upward arrows and connecting lines mark the new intron in the retrogene. In XXyac-R12DG2.2, the retrogene is also annotated to be embedded near the 3' end of an ncRNA gene candidate (LOC100190939, Additional file 12). In CSMD3, the green upward arrows and connecting lines mark the intragenic region in the retroposed sequence. The retrogene is an alternative translation start point (Additional file 12). In WBP2NL, the retrogene is located in its first intron, but could be transcribed at sometimes (Additional file 1 and 2). Besides, the intronic region at XXyac-R12DG2.2 and AC019016.1 could be spliced in two patterns.

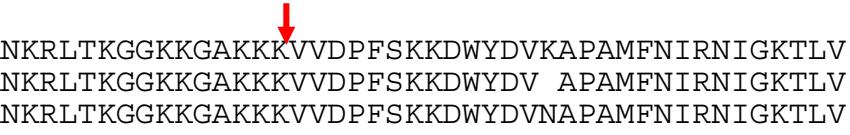
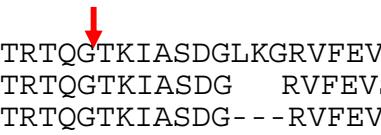
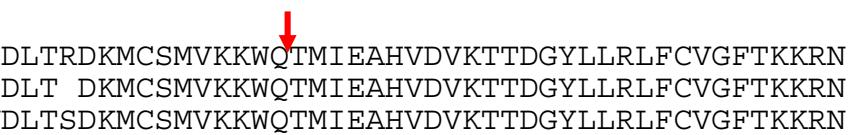
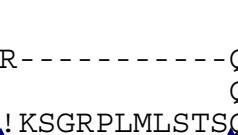
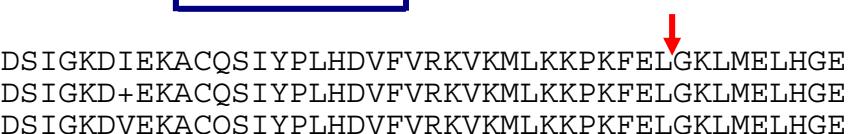
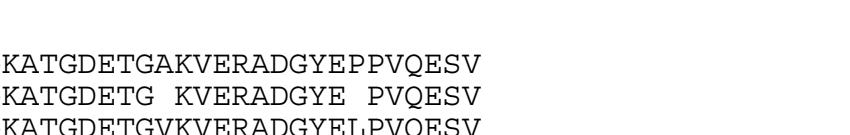
References

- S1. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome Res* 2004,
14(5):988-995.

TMEM14D

Query	1	MEKPLFPLVPLHWFGFGYTALVVSGGIVGYVKTG	SVPSLAAGLLFGSLA
Sbjct	2	MEKPLFPLVPLHWFGFGYTALVVSGGIVGYVKTG	PSLAAGLLFGSLA
			
Query	50	GLGAYQLYQDPRNVWGFLAATSVTFVGVMGRSYYYGKFMPVGLIAGAS	G+GAYQLYQDPRNVW FLAATSVTFVG+MGMRSSYYYGKFMPVGLIAGAS
Sbjct	149	GVGAYQLYQDPRNVWDFLAAATSVTFVGIMGRSYYYGKFMPVGLIAGAS	
			
Query	99	LLMAAKVGVRMLMTSD	
Sbjct	296	LLMAAKVGVRMLMTSD	

RPS3AP5

Query	1	MAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMFNIRNIGKTLV	M V KNKRLTKGGKKGAKKKVVDPFSKKDWYDV APAMFNIRNIGKTLV
Sbjct	2	MVVVDKNKRLTKGGKKGAKKKVVDPFSKKDWYDVNAAPAMFNIRNIGKTLV	
			
Query	50	TRTQGTKIASDGLKGRVFEVSLADLQNDEVAFRKFKLITEDVQGKNCLT	
Sbjct	149	TRTQGTKIASDG RVFEVSLADLQNDEVAFRKFKLITEDVQGKNCLT	
			
Query	99	NFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLRLFCVGFTKKRN	
Sbjct	287	NFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLRLFCVGFTKKRN	
			
Query	148	NQIRKTTSYAQHQQR-----QIRKKMMEIMTREVQTNDLKEVV	
Sbjct	434	NQIRKTTSYAQHQQR QIRKKMMEIMTREVQTNDLKEVV	
			
Query	186	NKLIIPDSIGKDIEKACQSIYPLHDVFVRKVVKMLKKPKFELGKLMELHGE	
Sbjct	583	NKLIIPDSIGKD+EKACQSIYPLHDVFVRKVVKMLKKPKFELGKLMELHGE	
			
Query	235	GSSSGKATGDETGAVERADGYEPPVQESV	
Sbjct	730	G SSGKATGDET GVERADGYE PVQESV	
			

XXyac-R12DG2.2

Query	1	MARGGRGRRRLGLALGLLLLALVLAPRVLRAKPTVRKERVVRPDSELGERP	
Sbjct	2	MARGGR LGLALGLLLLALVLAPRVLRAKPTVRKERVVRPDSELGERP	

Query	50	PEDNQSFQYDHEAFLGKEDSKTFDQLTPDESKERLGKIVDRIDNDGDGF	↓
Sbjct	140	PEDNQSFQYDHEAFLGKEDSK FD+LTPDESKERLGKIVDRIDNDG GF	
Query	99	VTTEELKTWIKRVQKRYIFDNVAKVWKDYDRDKDDKISWEEYKQATYGY	
Sbjct	287	VTTEELKTWIKRVQKRYIFDNVAKVWKDYDRDKDD ISWEEYKQATYGY	
Query	148	YLGNPAEFHDSSDHHTFKMLPRDERRFKAADLNGDLTATREEFTAF LH	↓
Sbjct	434	Y!GNPAEFHDSSDHHTFKMLPRDERRFKAADLNGDLTATPEEFTAF LH	
Query	197	PEEFEHMKEIVVLETLEDIDKNGDGFDQDEYIADMFSHEENGPEPDW V	↓
Sbjct	583	PEEFEHMKEIV LETLEDIDKNGDGFDQDEYIADMFSHEENGPEPDW V	
Query	246	LSEREQFNEFRDLNKDGKLDKDEIRHWILPQDYDHAQAEARHLVYESDK	
Sbjct	730	LSEREQFNEFRDLNKDGKLDKDEIRHWILPQDYDHAQAEARHLVYESDK	
Query	295	NKDEKLTKEEILENWNMFVGSQLTYGEDLTKNHDEL	↓
Sbjct	877	NKDEKLTKEEILENWNMFVGSQLTYGEDLTKNHDEL	

HSP90B2P

Query	8	GLCCVLLT ^{FGS} VRADDEV ^D V ^D GTV ^E EDLGKSREGSRTDDEV ^V QRE ^{EE} AI	↓
Sbjct	24	GLCCVLLT ^{FGS} VRA DEV ^D V+GT ^E EDLGKS ^R +GSR DDEV ^V QRE ^{EE} AI	
Query	57	QLDGLNASQIRELREKSEKFAFQAEVNRMMKLIINSLYKNKE ^I FLRELI	↓
Sbjct	171	QLD ^K LNASQIRELREKLEKFAFQAEVNRMMKLIINSLYENKE ^I FLRELI	
Query	106	SNASDALDKIRLISLTDENALSGNEELTVKIKCDKEKNLLHVTDTGVGM	↓
Sbjct	318	DAL KIRLISLT+ENALSGNEELTVKIKCDK KNLLHVTDTGVGM	
Query	155	TREELVKNLGTIAKSGTSEFLNKMTAQEDGQSTSELIGQFGVGFYSAF	
Sbjct	455	TREELVKNLGTIAKS T+EFLNKMTAQEDGQSTSELIGQFGV FYS F	
Query	204	LVADKVIVTSKHNNDTQHIWESDSNEFSVIADPRGNTLGRGTTITLVLK	↓
Sbjct	602	LVADKVIVTSKHNNDTQYIWESESNEYSVIADPRGNTLGRGTTITLVLK	

Query 253 EEASDYLELDTIKNLVKKYSQFINFPIYVWSSKTE_TVEEPMEEEAAKE
Sbjct 749 EKA_FDYLALDTIKNL!KYSQFINFPIYVWSSKTE_TTVKEP!KEEAAKE

Query 302 EKEESDDEAAVEEEEEKKPKTKV_EKTVWDWELMNDIKPIWQRPSKEV
Sbjct 894 EKEES DEAAVEEEEE++KP_KTKV_EKTV +WELMNDIKPIWQRPS EV

Query 351 EEDEYKAFYKSFSKESDDPMAYIHFTAE_EGEVTFKSILFVPT_SAPRGLFD
Sbjct 1040 E-DEYKAFCKSF_SKESDDPVACIHFTAE_EGEVTFKSILFVPTFVPRGLFD

Query 400 EYGSKKSDYIKLYVRRVFITDDFHDMMPKYLN_FVKGVVDSDDLPLNVS_R
Sbjct 1184 EYGSKKSDYIKLYVR_CVFITDDFRDTMPKNLFVKGVVDSGGLSLNVS_C

Query 449 ETLQQHKLLKVIRKKLVRKTLD_MIKKIADD_KIAD_KYNDTFWKEFGTN_NIKLGVI
Sbjct 1331 ETLQQHKLLKVIRKKLV KTLD_MIKKIAD+_KYNDTFWKEFGTN_NIKLGVI

Query 498 EDHSNRTRLAKLLRFQSSH_HPTDITS_LDQYVERMKEKQDKIYFMAGSSR
Sbjct 1478 EDHSNRT LAKLLRFQSSH_HP DITS_L Q VERMKEKQDKI +MAGSSR

Query 547 KEAESSPFVERLLKKGYEV_IYLT_EP_DEYCIQALPEFDGKR_FQNV_AKEG
Sbjct 1625 KEAESSPFV RLLK+GYEV_IYLT_EP_DEYCIQALPEFDGKR_FQNV_AKEG

Query 596 VKFDESEKTKESREAVEKEFEPLL_NWMKD_KALKDKIEKAVVSQRLTESP
Sbjct 1772 VKFDDSEKTKESHEAVEKEFEPL_NWVKDAIKDKIEKAMVSQCLTESL

Query 645 CALVASQYGWSGNMERIMKAQAYQTGKD_ISTNY_YASQKKT_EINPRHPL
Sbjct 1919 CALVASQYGWSGNMERIMKAQAYQTGK I_STNY+AS+KKT_EINPRHPL

Query 694 IRDMLRRIKEDED_DKTVLDLAVVL_FETAT--LRSGYLLPDTKAYGDRIE
Sbjct 2066 IRDMLRRIKEDED_DKTVLDLAVVL_FETA +RSG LL_PDTKAYGDRIE

Query 741 RMLRLSLNIDPDAKVEEEPEEEPEETAEDTTEDTEQDEDEEMDVGTDEE
Sbjct 2215 RM LSLN PDAKVEE P+EEPEETAED EQD+D+EMDVGTDEE
RMPCLSLNT!PDAKVEE-PDEEPEETAEDK---EQDKDKEMDVGTDEE

Query 790 -EETAKESTAEKDEL
Sbjct 2345 ETAKESTAEKDEL
KQETAKESTAEKDEL

HSP90AA4P

Query	1	MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIINTFSNKEIFLRELI
Sbjct	6	MPEETQTQDQPTEEXEVETFTQAEIAQLMSLFINTFYLNKEIFLRELI
Query	50	SNSSDALDKIRYESLTDPSKLDGKELHINLIPNKQDRRTLIVDTGIGM
Sbjct	153	SNLSDALDKILXESLTDPSKLDGKEPHISLIPNKQDRRTLIVDTGIGM
Query	99	TKADLINNLGTIAKSGTKAFMEALQAGADISMIGQFGVGFYSAYLVAEK
Sbjct	300	TKADLINNLGTI KS TK FME LQAGADISMIGQF VGFYSAY VAEK
Query	148	VTVITKHNDDEQYAWESSAGGSFTVRTDTGEPMGRGTVKVLHLKEDQTE
Sbjct	447	VTVITKHNDDEQYAWESSLRGSFTVRTDTGEPIGHGTVKVLPLKEDQTE
Query	197	YLEERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDEAEEKEDKEEEK
Sbjct	594	YLEER+I EIVKKHS FIGYP+TLFVEK+ DKEVS+DEAEEKEDKEEEK
Query	246	EKEEKESEDKPEIEDVGSDEEEE-KKDGD---KKKKKKIKEKYIDQEE
Sbjct	739	EKEEKESEDKPEIEDVGSDEEEEXKKDGDKKKKKKKIKEKYIDQGE
Query	290	LNKTKPIWTRNPDDITNEEYGEFYKSLTNWDHLAVKHFSVEGQLEFR
Sbjct	886	LN TKPI TR PDD+TN EY EFYKSLT +WED+LAVKHFSVEGQLEFR
Query	339	ALLFVPRRAPFDLFENRKKNNIKLYVRRVFIMDNCEELIPEYLNFIRG
Sbjct	1032	A LFVPR APF+L+E RKKKN IKL RR +IMDNCEELIPEYLNFIRG
Query	388	VVDSEDLPLNI SREMLQQSKILKVKIRKNLVKKCLELFTELAEDKENYKK
Sbjct	1179	VVDSEDLPLNI REMLQ SKILKVKIR NLVKKCLELFTEL E KE Y K
Query	437	FYEQFSKNIKLGIHEDSQNRKKSELLRYYTSASGDEMVS LKDYCTRMK
Sbjct	1326	FYEFSKNIKLGIH DSQN+KKLSELLRYYT ASGDEM V L+DYCTRMK
Query	486	ENQKHIYYITGETKDQVANS AFVERLRKGLEV IYMI EPIDEYCVQQLK
Sbjct	1473	ENQKHIYYITGETKDQVANS +V+RL KGLEV IYMI EPIDEYCVQQLK
Query	535	EFEKGTLVSVTKEGLELPEDEEEKKQEE-----KKTFENLCKIMK
Sbjct	1620	EFEKGTLVSVTKE LELPEDEEEKKQEEGKQKTQKK !KFENLCKIIVK

Query	577	DILEKKVEKVVVSNRLVTSPCCIVTSTYGTANMERIMKAQALRDNSTM DILEK VEKVVVSN+LVTSPCCIV+STYGTANMERIMKAQALRDNST
Sbjct	1769	<u>DILEK!VEKVVVSNQLVTSPCCIVSSTYGTANMERIMKAQALRDNSTT</u>
Query	626	GYMAAKKHLEINPDHS I IETLRQKAEADKNDKSVKDLVILLYETALLSS
Sbjct	1914	GYMAAKKHLEINPDHS + I + TLRQKAE DKNDKSVKDLVILLYETALLSS
Query	675	GFSLEDPQTHANRIYRM IKLGLGIDEDDP TADDT SAAVTEEMP PLEGDD F LE+PQTHANRIYRM KLGLG DEDDP TADDT SAAVTEEMP PLEGDD
Sbjct	2061	DFGLEG P QTHANRIYRM NKLGLG TD EDDP TADDT SAAVTEEMP PLEGDD
Query	724	DTSRMEE
Sbjct	2208	DTSRMEE+
Query	724	DTSRMEE
Sbjct	2208	DTSRMEEK

HSP90AA5P

Query	192	EDQTEYLEERRIKEIVKKHSQFIGYPITLFVEKERDKEVSDDEAEEKED E + EYLEERRIKEIVKKHSQFIGYPITLFVEK+R+K+VSD EAE+KED
Sbjct	22	ERRPEYLEERRIKEIVKKHSQFIGYPITLFVEKRNKQVSDAEAEKKED
Query	241	KEEEKEKEKESEDKPEIEDVGSDEEEEKDGD K K K K K I KEKYIDQE K K++ KES DKPEIEDVGSDEEEEK D K K K K KEKYIDQE
Sbjct	169	-----KRKKKESNDKPEIEDVGSDEEEEKDADKKKKS - KEKYIDQE -
Query	290	LNKT KPIWTRNPDDITNEEYGEFYKSLTNDWEDHLAVKHFSVEGQLE LNKT KPIWTRNP D ITNEEYGEF++SLTN+WEDHLAVKHFSVEGQLE
Sbjct	298	LNKT KPIWTRNPDAITNEEYGEFHQS LTNNWEDHLAVKHFSVEGQLE
Query	337	RALLFVPRRAPFDL FENRK KNN
Sbjct	439	<2----- [441 : 465] -2> RA LFV R APFDL FENRK KNN RA-LFV SRCAPFDL FENRK KNN
Query	361	I KLYVRRVFIMDNCEELIPEYLN FIRGVVDSEDLPLNISREMLQQSKIL I+LYV RVFIM NCEELIPEYLN FIRGV+DSEDLPLNI EMLQQSKIL
Sbjct	533	I QLYVHRVFIMGNC EELIPEYLN FIRGVMDSEDLPLNIFSEMLQQSKIL
Query	410	KVIRKNLVKKCLELFTELAEDKENYKKFYEQFSKNIKLGIHEDSQNRKK KVIRKNLVK L ELF ELAEDK+NYKKF+E++S+NI L + DSQN+KK
Sbjct	680	KVIRKNLVKTYLELFAELAEDKKNYKKFHEKLS ENINL! EYIDSQNQKK
Query	459	LSELLRYYTSASGDEM VSLKD YCTRMKENQK H IYYITGETKDQVANS AF LSELL+ TSAS DEMV L + RMK NQK H IYYITGETKDQVANS AF
Sbjct	825	LSELLKXCTSASRDEM VYLV F-SRMKANQK H IYYITGETKDQVANS AF

Query	508	VERLRKHGLEVIYMIEPIDEYCVQQLKEFEGKTLVSVTKEGLELPEDEE
Sbjct	969	VECLQKHGLEVIYMIELIDKYCVQQLKELESKTVVSVAKEGLELPEDEE
Query	557	EKKKQEEKKTKFENLCKIMKDILEKKVEKVVSNRLVTSPCCIVTSTYG
Sbjct	1116	EKKKQEEKKTKFENLCKIMKD+LEKKV+KVVSN VTSPCCIVTSTYG
Query	606	WTANMERIMKAQALRDNSTMGYMAAKHLEINPDHSIIETLRQKAEADK
Sbjct	1259	WTANMERIMKAQA DNSTMGY+AAKHLEINPDHS IETLRQKAEADK
<hr/>		
Query	655	NDKSVKDLVILLYETALLSSGFSLEDPQTHANRIYRMIKLGLGIDEEDDP
Sbjct	1405	N+KSVKDLVILLY TAL S GFS+EDPQ H N+IYRMIKLGLG+DE DP
<hr/>		
Query	704	TADDTSAAVTEEMPPLEGDDDSRMEEV
Sbjct	1552	TA+D +AA+T+EMPL G DDTSRMEEV

CSMD3

Query	1	MGVDIRHNKDRKVRRKEPKSQDIYLRLLV-KLYRFLARRTNSTFNQVVL
Sbjct	1	MGVD RH KDRKVRRKEPKSQDIYLRL V + ++ TNSTFN+VVL
<hr/>		
Query	49	KRLFMSRTNRPPLSLSRMIRKMKLPGRENKTAVVVGITDDDRVQEVPK
Sbjct	145	KRLFMSRNNRPPLPLSRMIWKMKLPGRENKTAVVGAITDDDRVQEVPK
<hr/>		
Query	98	LKVCALRVTSRARSRLR-----AGGKILTDFDQLALDSPKG
Sbjct	292	LKVCALR+TSRARS I A GK LTF QLALDSPKG
<hr/>		
Query	134	CGTVLLSGPRKGREVYRHFGKAPGTPHSHTKPYVRSKGRKFERARGRA
Sbjct	439	CGVVRLSGPPKGGRGVY-----APET!CKPYIIRSNGRKFERARDRA
<hr/>		
Query	183	SRGYKN
Sbjct	560	<u>SRGYKN</u>

WBP2NL

Query	1	MTDAAVSFAKDFLAGGAAAISKAVAPIERVKLLLQVQHASKQITADK +T+AAVSFAKD LAGGAAAISK AVAPIERVKLLLQVQHASKQITADK
Sbjct	1	ITEAAVSFAKDSLAGGAAAISKMAVAPIERVKLLLQVQHASKQITADK
Query	50	QYKGI IDCVRIPKEQGVLSWRGNLANVIRYFPTQALNFAFKDKYKQI QYKGI DC+V IPKE GVLS W GNLA+VIRYFP QALNF FKDKYKQI
Sbjct	148	QYKGIRDCMVCIPKEHGVL-S-WCGNLASVIRYFP!QALNFTFKDKYKQI
Query	99	FLGGVDKRTQFWLYFAGNLASGGAAGATSLCFVYPLDFARTRLAADVGK FL GVDKR+QFW YFAGNLASGGA GATSLCFVYPLDFA TR+AADV GK
Sbjct	293	FLDGVDKRSQFWRYFAGNLASGGATGATSLCFVYPLDFAHTRVAADV GK
Query	48	AGAEREFRGLGDCLVKIYKSDGIKGLYQGFNVSVQGIIYRAA---YFG AGAERE+RG GDCLVKIYKSDGIK LYQGFNVSVQGII YRAA YFG
Sbjct	440	AGAERELRGFGDCLVKIYKSDGIKSLYQGFNVSVQGII-YRAACC!YFG
Query	194	IYDTAKGMLPDPKNTHIVISWMIAQTVTAVAGLTSYPFDTVRRMMQS YDTAKGMLPDPKNTHIVISWM AQTVTAVAGLTSY F+T+ MMMQS
Sbjct	582	NYDTAKGMLPDPKNTHIVISWM!AQTVTAVAGLTSYSFNTICCYMMQS
Query	243	GRKGTDIMYTGTLDWRKIARDEGGKAFFKGAWSNVLRGMGAFVLVLY G KGTDIMYTGT DCWRKI DEG KA FKGAW +VLRGMGGAFVLVLY
Sbjct	728	GCKGTDIMYTGTDCWRKITHDEGSKAVFKGAWSVLRGMGGAFVLVLY
Query	292	DEIKKY DE KKY
Sbjct	875	DET KKY

AC019016.1

Query	9	AEEFIVGGKYKLVRKIGSGSFQDIYLAINITNGEEVAVKLESQKARHPQL AEFIVG KYKL+ KI SFGDI LAI+ITNGEEV VK ESQ ARHPQL
Sbjct	393	AEFIVGRKYKLLWIKIWCSSFGDICLAI SITNGEEVVVKPESQ!ARHPQL
Query	58	LYESKLYKILQGGVGIPHRYWYQEQKDYNVLVMDLLGPSLEDLFNFC SR LY+S+LYK+LQGGVGIPHRYW GQE+DYNVLVMD G L NFC SR
Sbjct	538	LYKSRLYKMLOGGVGIPHRYWCGQEEDYNVLVMDFSG!-LASKTNFC SR
Query	107	RFTMKTVLMLADQMISRIEYVHTKNFIHRDIKP DNFLMGIGRH C NKL FL RFTMKTVL+LADQMISRIE+VHTKNFIHR IKP NFLMGIGRH C+KL F
Sbjct	681	RFTMKTVL L IADQMISRIEHVHTKNFIHRAIKPGNFLMGIGRH CSK L FR
Query	156	IDFGLAKKYRDNRTRQHI PYREDKNLTGTARYASINAHLGIEQSRRDD M IDFGLAK Y DNRTRQHI PYREDKNLTGT A YAS AHL IEQSRRDD M
Sbjct	828	IDFGLAKTYGDNRTRQHI PYREDKNLTGT AL YAS -HAHLSIEQSRRDD M

Query 205 ESLGYVLMYFNRTS---LPWQGLKAATKKQKYEKISEKKMSTPVEVLCK
Sbjct 972 ESL YVLMYFNRTS LPWQGLKAA +KQKYEKISEKKMSTPV CK

Query 251 GFPAEFAMYLNCRGLRFEEAPDYMMLRQLFRILFRTLNHQYDYTFDWT
Sbjct 1112 GF AEFA+YLN CRGL +E+APDYMML QLF ILFRTL HQYDYTFDWT

Query 300 MLKQKAAQQAASSSGQQAQPTGKQTDKTKSNMKGF
Sbjct 1259 LKQKAAQQAASSSG GQQAQPTGKQ DTK NMKGF
KLKQKAAQQAASSSG!GQQAQPTGKQPDKTGNMKGF

