A Specialized Version of the HD Hydrolase Domain Implicated in Signal Transduction

Michael Y. Galperin*, Darren A. Natale, L. Aravind, and Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Recently, a superfamily of proteins containing a previously undetected domain with predicted metal-dependent phosphohydrolase activity has been described and designated the HD superfamily, after the principal conserved residues implicated in catalysis (Aravind and Koonin, 1998). In the course of our analysis of ancient conserved regions in microbial genomes (Koonin et al., 1998), we found a distinct version of this domain which is encoded in one to three copies in the genomes of Aquifex aeolicus, Borrelia burgdorferi, Synechocystis sp. and Treponema pallidum, but is dramatically expanded in the genomes of Thermotoga maritima and Clostridium acetobutylicum (Figure 1). Compared with the consensus HD domain (Aravind and Koonin, 1998), this version contains a number of additional highly conserved residues; hereinafter we refer to it as the HD-GYP domain, after the characteristic sequence signatures (Figure 1). This domain was also detected in previously uncharacterized proteins from Wolinella succinogenes (Kreis-Kleinschmidt et al., 1995), Bacillus halodurans (Takami et al., 1999), Pseudomonas aeruginosa and Bordetella pertussis (Figure 1). The HD-GYP domain is missing in E. coli and B. subtilis. Remarkably, however, in other γ -proteobacteria, such as Shewanella putrefaciens and Vibrio cholerae, it is present in up to 8 copies (data not shown).

While none of the proteins that contain the HD-GYP domain has ever been characterized experimentally, the spectrum of the domains that are associated with HD-GYP in multidomain proteins (Figure 2) suggests that it is probably involved in signal transduction. In *Synechocystis* sp., both copies of the HD-GYP domain are found in proteins that also contain CheY-like receiver domains of the twocomponent signal transduction system (Pao and Saier, 1995; Volz and Matsumura, 1991). A similar CheY – HD-GYP domain organization is found in two T. maritima proteins, TM0186 and TM1147 (Figure 2). Two other proteins from T. maritima, TM1170 and TM1682, combine the HD-GYP domain with extracytoplasmic ligand-binding domains, which are closely related, respectively, to periplasmic solute-binding protein components of the ATP-dependent transport systems (Tam and Saier, 1993) and the extracytoplasmic part of methyl-accepting chemotaxis proteins of Bacillus subtilis, such as McpA and McpB (Hanlon and Ordal, 1994). Such a combination of an extracytoplasmic ligand-binding domain and a cytoplasmic HD-GYP domain, connected by a transmembrane segment, has the same topology as methyl-accepting proteins and many sensor kinases, which further supports the participation of the HD-GYP domain in signal transduction. Finally, in the Aq_2027 protein from A. aeolicus, the HD-GYP domain is found together with the GGDEF domain (Figure 2). The latter domain has been recently identified in diguanylate cyclases and phosphodiesterases involved in the regulation of cellulose synthesis in Acetobacter xylinum (Tal et al., 1998) and in a variety of bacterial signalling proteins in combination with CheY, PAS, and HAMP domains (Hecht and Newton, 1995; Aravind and Ponting, 1999). The GGDEF domain is often associated with another uncharacterized domain, EAL, in particular, in diguanylate cyclases and phosphodiesterases (Tal et al., 1998; Aravind and Ponting, 1999). Remarkably, however, the combination of the HD-GYP and EAL domains is not seen in any of the currently available microbial genomes. Moreover, the number of copies of the HD-GYP domain in complete genomes generally correlates with prevalence of the GGDEF domain over the EAL domain (Table 1). Furthermore, the HD-GYP family of HD proteins so far is lacking in archaea and eukaryotes and so are the GGDEF and EAL domains. This suggests that HD-GYP domain might be also involved in cyclic diguanylate-mediated signaling. The HD superfamily is related to the cAMP/ cGMP phosphodiesterases that are involved in eukaryotic signalling (Aravind and Koonin, 1998). Therefore is seems plausible that the HD-GYP family proteins are likely to possess a diguanylate phosphodiesterase activity and complement the function that, in the characterized diguanylate phosphodiesterases, is performed by the EAL domain.

Table 1. Distribution of Three Domains Implicated in Signal Transduction in Complete Microbial Genomes

Species ^a	Domains			
	GGDEF	EAL	HD-GYP	
Escherichia coli	19	18	-	
Rickettsia prowazekii	1	1	-	
Bacillus subtilis	4	2	-	
Mycobacterium tuberculosis	1	2	-	
Synechocystis sp.	23	13	2	
Borellia burgdorferi	1	1	1	
Treponema pallidum	1	-	3	
Aquifex aeolicus	11	6	1	
Thermotoga maritima	9	-	9	
Clostridium acetobutylicumb	10	4	8	

^a Genomes of bacteria Haemophilus influenzae, Helicobacter pylori, Chlamydia trachomatis, C. pneumoniae, Mycoplasma genitalium, M. pneumoniae, and archaea Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Archaeoglobus fulgidus, Pyrococcus horikoshii, and Aeropyrum pernix do not contain any of these domains.

^bPreliminary data based on unfinished genome sequence (http:// www.cric.com/sequence_center/bacterial_genomes)

Received October 4, 1999; accepted October 4, 1999.

^{*}For correspondence. Email galperin@ncbi.nlm.nih.gov; Tel. 301-435-5910; Fax. 301-480-9241.

Aq_2027_Aq	16	DAYTRGHTI	ERGVFYAQLIGKELGLTSQEIELLKLGGYTIDVGKVAIPDVILLKPAKLTPEEYEIMKLIVELGYEMVKDLEIP-1	-	
BB0374_Bb	133	ANYHIIHSY	VNTAILTVALGNEMGLNNYKTVELCSIALL <mark>HK</mark> IGFLFIPSKISEKKEALTEEELEIIKKYPIISYKIASTSNLS-1	-	
s]]1624 Sv	155	RYAVGEAS	TTCSOLIEGEGEFIGIEGGOTODIVEAARINDEGLIOVPDETMLKOGPINSTELOVPDUVHVAATIEEPLAGP-1	_	
$e_1r_{2100}e_V$	174	DUETCNET		_	
D112100_Dy	1/4 001	GEETGINHTI			
TM0037_Tma	221	SEFTISET	WRVAVFSEKMAEKAGRNSEDLFIAGLYHDIGKISVSYKILEKPAKLSDEEYAIMKKHVYFSYLILLPYKDE	-	
TM0186_Tma	16/	DMETHRET	ERVGWLSGRIAEEMGMSEVFVTEIQFAAPLHDIGKIGIPDRILLKPGILTPEEFEIMKQHTTIGFRILSRSNSP-3	-	
TM0739_Tma	107	DENGWSHS	QRVARLAERVGKKLGLSEEKLSLLREYAML <mark>HD</mark> VGKIGIEQLMLYTPTRIRIFE-QYPQD <mark>H</mark> TIMGSVFLASLEVL-1	-	
TM1145_Tma	168	SRYTLNHI	YNVAFWAEEIGKIVGLDEKRIAKIYIAGLI <mark>HD</mark> IGKVYIDERILNKEGRITLQEYQEMKK <mark>H</mark> VDYSYNMVKDLMIW-4	-	
TM1147_Tma	181	DYETHEHTI	LRVGRISALVAEAMGCSSEFVELIEKAAPFHDIGKVFIPESILLKRGKLTSDEVEIMKLHTIFGYELLKTSKNP-3	-	
TM1170_Tma	442	DYYTKGHSI	EEAAYYAVEIGRMFDLGDEKLEKLYWAGLLHDIGKIYYPOYVLNKTSKLDEREFELIKIHPVRGYELVKEIEGF-1	-	
TM1467 Tma	251	DEYTRGHSI	EKVAELAKKIAKRIGESPKMVERIYNAAFI DTGKIGIPDHILKKPTILSKEEMMIVKNUPIMGEDILREVDIF-2	-	
ТМ1678 Тта	22		TGKTAVPEHTINKPCKITDEEFETKKUSTVCADLLEEVPEL-1	-	
TM1682 Tma	384	DEPTGN	RRYOFT SKEFA FEMCIDEDLYHKTYIYADI IIIIIIIII CKIKYAKKITA FEWFIMKKHTTWCCFTISCHKELT	-	
TM1600 Tma	201	DDVTCC			
THIUJJJ_IMA	152	UVALDOWA	LRVAD SOCIARRRG PDOVVERTRIAAVIAD GRIGVRGATENKTSKITKEETEVRRPEIGERIISKIEDF-1	-	
TP0/64_Tpa	100	HYALETBA	LRSTLYAVVIGLQLKMQPFKIIELATSCLIHDIGMARVIPKAYTTEGELDPKTQKAIFAPPIISYHILRDHSLP-1	-	
TP08//_Tpa	189	HCERVNEY	CREIAEYLYDINLYP-4-DFVENIAFLAAMHDVGKIGIPDYVLKKRGGLNELEWELMKEHTINGALTLSSYPDP	-	
TP0912_Tpa	201	GFRFAKPAV	VDTAIFSIIVAQHLGFSEKDIFDLVVASLLHDIGMQHVPADILKKSGRLHFQEQQAVDAHTIYAHSYIVNTLKY-2·	-	
Orfl4_Bhal	139	EDYLYHHAV	VSVGLLSAFLGAKLGYEKGEWLKIALAGALADCGMAKLPTSLLTKNGLYTLKEKHEMRD <mark>H</mark> PKYSYQMLKKRPEI-2·	-	
Orf2_Wsuc	54	SKETGFHVI	KRVAEYSKILALG MGLS ELEAELVK VAS PM <mark>HD</mark> IGKVAIPDAILKKPGKLTEEEFEVMKS <mark>H</mark> SRLGYETLKHSNRR-3	-	
Z7f _Vcho		SPYTAGHSI	ERVAVYTDLIARQLAISDADRIWLRRAALIHDIGKIGVSNAILDKPGKLDEVEWRAVQAHAAYTEQILYKLSPF-1	-	
Cl.acetobut.		SHETGHHVI	KRVAEYSNLLALRYGLTOREAEIIKLASPMHDVGKLAIPDSILNKPASITNEEFEVMKT <mark>H</mark> SKIGYDMLKSSNRK-3	-	
Cl.acetobut.		DEYTFYHS	VNVSFYGMLIAKWMNLSEKOIREVISAGLIHDIGKTKIKNSTINKPGRITDEEFEEMKKHPLYGYETIKDDPTT-2	-	
Cl.acetobut.		DNYTAVHS	KSVAFVARFIAEKIKYSEEDCEKVETGALLEDICKIGISDTLINKDLKITEDEVKTIOEUPOTCESILELNEDE-3	-	
Cl acetobut		DEEMLIND		_	
Cl. acetobut.		DEFENDER	Invan Skelaikuusseleikuu 1056 tulu gali Seatinkapiieke viiksisi SSNIIKNITEL-1.	-	
Cl.acetobut.		DDITIONS	VNVAV SLITGIKTKYHKTELFDLCIGALLIDVGKMFIPSETLNKKGELTEEEFKIMKENTVRGYDYEKGLYEI-2	-	
cl.acetobut.		DDYTYIHCI	LDTGIMAAFLGLSMGLNTARIKDLSIAAMDHDIGKTKIPSEIINKKEKLTDQEYATIKNHPAYGRDILNSIGML-2	-	
Cl.acetobut.		SSFTAK	RSISNLAFQVSKYVGYSEEKCLKMKIAGLLHDIGKLAIPTSILDKNGSLTDDEFSIIKSHVYYTKIILDSIEDI-1	-	
Cl.acetobut.		KNNIYRHSV	VNVAAISFILGKWLGMNEEEINLLTYASLL <mark>HD</mark> FGKTQLDNSILKKESSLTPEEYAVYKT <mark>H</mark> PVTAYHLIKEIPDI-2·	-	
Cl.acetobut.		SEYTQRHSS	SDLSEKVEIMADYYKMSHEEKLKLIITAADI <mark>HD</mark> IGKLAVPNSILDSPKKLTKEEFKEIKK <mark>H</mark> PYTTKLALHDITGF-1	-	
P.aeruginosa		DNPRSRHLI	PRIERYVRLLAEHLA-12-AVDLLSKSALLHDTGKVAVPDRVLLNPGOLDAADTALLOGHTRAGRDALASAERR-9	-	
P.aeruginosa		DEYTYMES	VAVCALMIALAROLELPDPLVREAGLAGLIHDIGKMAVPDPILNKPGKLTDPEFGLVRRHPONGARMULDCROV-2	-	
B.pertussis		DPSTATHS	GRVAATAGELGRAYGLDGMATRELELGASTHDTGKIGVADRVLHHPGRILGEDWTHMORESVIGERTTAATGLD-3	_	
- · L		_ ~			
Structure		HHI	ннининининин		
		LENT L VILOT	WYDO TO YN DOWN O DDWN A DOWN A DOWN AND DDWN A WOLWEN W WD DDWN A DDWN AND AL OD	100	2004261
Aq_2027_Aq	ES.	LEVLLY <mark>HQE</mark>	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR	182	2984261
Aq_2027_Aq BB0374_Bb	ES. SI	LEVLLYHQE CLTLLTHKE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NLDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS	182 299	2984261 2688290
Aq_2027_Aq BB0374_Bb s111624_Sy	ES SI AV	LEVLLY <mark>HQE</mark> CLTLLTHKE GEIMRY <mark>HHE</mark>	KYDGT <mark>GYP</mark> FGKKGEETPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA - GKQFDPDIVPYAVRALSR NLDGT <mark>GYP</mark> KGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA - DKKFDKRVLKLIINAISS RWDGS <mark>GYP</mark> DALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK	182 299 322	2984261 2688290 1652854
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy	ES SI AV LA	LEVLLY <mark>HQE</mark> CLTLLTHKE GEIMRYHHE KEIAYS <mark>HQ</mark> E	KYDGTGYPFGKKGEETPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA - GKQFDPDIVPYAVRALSR NLDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA - DKKFDKRVLKLIINAISS RWDGS <mark>GYP</mark> DALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KW <mark>DGTGYP</mark> QGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY - GQHFDPDVVDAFMAIHPT	182 299 322 356	2984261 2688290 1652854 1653077
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma	ES SI AV LA NW	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE	XYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED	182 299 322 356 384	2984261 2688290 1652854 1653077 4980522
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma	ES SI AV LA NW LG	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AEIALTHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NILGTGYPKGLTSENISIESNIIGAASÄYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIIQVADYFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSIED RWDGSGYPRGLKERELPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKIEKE	182 299 322 356 384 335	2984261 2688290 1652854 1653077 4980522 4980682
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0739 Tma	ES SI AV LA NW LG	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AEIALTHHE VPIVRHHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NLDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRMPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD	182 299 322 356 384 335 275	2984261 2688290 1652854 1653077 4980522 4980682 4981264
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0739_Tma TM145_Tma	ES SI AV LA NW LG DV	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AEIALTHHE VPIVRHHE ALWVVOHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NLDGTGYPKGLTSENISIESNIICAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFOIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDDDVVDAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKEEEIPLEARIISVCDYYDVIT-6-WEGRKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KEDGSGYPFGLKEEEIPLEARIISVCDYYDVIT-6-WEGRKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD	182 299 322 356 384 335 275 337	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma	ES SI AV LA NW LG DV DV	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AEIALTHHE VPIVRHHE ALWVQHHE AEVALEUHE	XYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIOVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDFKVVEAFLEIFSD KFDGSGYPFGLKEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKUVEAFLEIFSD	182 299 322 356 384 335 275 337 349	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma	ES SI AV LA NW LG DV DV LA	LEVLLYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AETALTHHE VPIVRHHE ALWVVQHHE AEVALFHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKIEKE NWDGTGYPGLKGEETPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLIRE KYNGKGYPHGLKGEETPLSAQIAAVADSFDAMVSRRPYKIRKSFDEAIREIKLS-GRFSPFVKAFVQTIDE	182 299 322 356 384 335 275 337 349	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM1145_Tma TM1145_Tma TM1147_Tma TM1170_Tma	ES SI AV LA NW LG DV DV LG DV	LEVLLYHOE CLTLLTHKE GEIMRYHHE KEIAYSHOE FLPAVRHHE AEIALTHHE VPIVRHHHE ALWVVOHHE AEVALFHHE AIWIRHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPGLKEEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRYSPFVVKAFVQTIDE RWDGKGYPHGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKIKSFDEAIREIKDLS-GRYSPFVVKAFVQTIDE	182 299 322 356 384 335 275 337 349 608	2984261 2688290 1652854 1653077 4980522 4980682 4981692 4981692 4981695 4981719
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1147_Tma	ES SI AV LA NW LG DV DV LA	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FLPAVRHHE AEIALTHHE ALWVVCHHE ALWVVCHHE AEVALFHHE AIWIRHHE SKWVRHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NLDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRDEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPPRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEEKE NWGGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPFLVAEAFLKIEFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPFEARVLCVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPFEARVLCVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPFEARVLCVADSYQAMRSDRPYKRGKSVEESIQELRRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPDGLKGEEIPFEARVLCVADSYQAMRSDRPYKRALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK	182 299 322 356 384 335 275 337 349 608 419	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0739_Tma TM1145_Tma TM1147_Tma TM1170_Tma TM1467_Tma TM1678_Tma	ES SI AV LA NW LG DV DV LA DV RE TA	LEVILYHQE CLTLLTHKE GETMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVCHHE AEVALFHE AIWIRHHE SKWVRHHE	KYDGTGYPFGKKGEEIPLLARIITTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQASIIELIKDA-DKKFDRRVLKLIINAISS RWDGGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDFKVVEAFLEIFSD KKPGSGYPHGLKEEITPISGIIAAVADSFDAMVSRRPYKNFFDLDVIKEIERLR-GKKFDFKVVEAFLEIFSD KKPGSGYPHGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDFKVVEAFLEIFSD KKPGSGYPHGLKEEITPISGIIAVADSFDAMVSRRPYKNFFDLDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPFERVLCVADSYQAMSDRPYKIKSFDEAIREIKDLS-GRFYSPFVVEAFLEIFSD KFDGSGYDGLKGEEIPFERVLCVADSYQAMSDRPYKICKSVEESIQELRRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPDGLKGEEIPFERVLCVADSYQAMSDRPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIISVADVEALTSDRPYRKAKFEDAALMKKMP-LDQEIVGILEKHLSEFVS	182 299 322 356 384 335 275 337 349 608 419 148	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM145_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1682_Tma	ES SI AV LA NW LG DV DV LA DV RE FA	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE AIWIRHHE AIWIRHHE SKWVRHHE RNIALYHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLDGYGGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKKFDPLLADLTVDLIRE KYNGKGYPHGLKGEEIPFEARVLCVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPFEARVLCVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPFEARVLCVADSYQAMRSDRPYKRALSHEEAKEIMKNMKGTVLDPRLVDLLFEIEK RVDGSGYPGLKGEEIPLEARISSYQADYFAITSDRPYRKALSHEEAKEIMKNMKGTVLDPLLVDLLFEIEK RVDGSGYPGGLKDGEIPLLARIISVCDYQALTSDRPYRKALSHEEAKEIMKNMKGTVLDPLVGLIEFEIEK SYDGTGYPFGLKDEIPLLARISST	182 299 322 356 384 335 275 337 349 608 419 148 550	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM145_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1682_Tma TM1699_Tma	ES SI AV LA NW LG DV DV LA CV FA VA DI	LEVILYHQE CLTLLTHKE GEIMRYHHD KEIAYSHQE FIPAVRHE AEIALTHHF VPIVRHHE ALWVVQHHE AEVALFHHE SKWVRHHE SKWVRHHE RNIALYHE AKIVRHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVAFMAIHFT RUDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVFAFLKLEKE NWDGTGYPDGLKGEEIPLEARIISVCDYJDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLEARIISVCDYYEALTSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYDSPEVVKAFVQTIDE RWDGKGYPDGLKGEEIPLEARIISVADVYEALTSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLSRQIARVUYEDAITSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK SYDGFGYFGLKDGEIPLSRIISVADVYDALTSDRPYRKALSHEEAKEIMKNKFTVLDERLVGLLEKHLSEFVS SYDGTGYPFGLKDDEIPLESRIIAVADAFDAMTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSGQSYPGLRGEEIPLESRIIAVADAFDAMTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE	182 299 322 356 384 335 275 337 349 608 419 148 550 557	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM11467_Tma TM1467_Tma TM1678_Tma TM1682_Tma TM1682_Tma TM1682_Tma	ES SI AV LA NW LG DV LA DV RE FA VA DI DI	LEVILYHQE CLITLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE ALEVALFHHE ALWVVQHHE ALWVVQHHE AKWRHHE SKWVRHHE RNIALYHE AKIYREHHE CVGALEHRE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOA-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLEARIISVCDYZDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLLRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLSAQIAAVADSFDAMVSRRPYKLIKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLSAQIAAVADSFDAMVSRRPYKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEIECG RWDGSGYPFGLKDEIPLSAQIAAVADSFDAMVSRRPYKALSHEEAKEIMKNMFGTVLDPRLVDLLFEIEEG WWGGGYPDGLRGEIPLEARIIAVADYFDALTSDRPYRKALSHEEAKEIMKNMFGTVLDPRLVDLLFEIEFLEG RVDGSGYPFGLKDEIPLARILAVADYFDALTSDRPYRKAMSHRVTLEGD-GRTSPSNFHPKLIEIFRE WYSGGYPGGLKDGEIPLESRIIAVADAFDAMTSDRPYRKAMSHRVARAEGIVDLVRNA-NTQYDAVISRALLFALSF	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323066
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1145_Tma TM1170_Tma TM1170_Tma TM1678_Tma TM1678_Tma TM1699_Tma TM1699_Tma TP0764_Tpa TP0877_Tpa	ES. SI AV LA NW LG DV LA DV RE FA VA DI DI MA	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVCHHE AEVALFHE SKWVRHHE SKWVRHHE RNIALYHHE AKIVRHHE CVGALEHRE KEIALSHE	XYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIELKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIOVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKIEFSD KFDGSGYPFGLKEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPFERVLCVADSYDAMVSRRPYKIRKSFDEAIREIKDLS-GRFYSPFVVEAFLKIEFSD KFDGSGYPGLKGEEIPFERVLCVADSYQAMRSDRPYKRCKSVEESIQELRRNA-GRQFDPVIVEKFIEFIEG RWDGKGYPDGLKGEEIPFERVLCVADSYQAMRSDRPYKRALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIISVADVFALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIIAVADSFDAMTSDRPYKKALSHEEAVRVILEGO-GRTSPSNFHPKLIEIFEG RWDGKGYPDGLKGEEIPLESAIIAVADAFDAMTSDRPYRKAMSHEAALHIKMN-LDQEIVGILEKHLSEFVS SYDGTGYPFGLKDDEIPLESAIIAVADAFDAMTSDRPYRKALSHEEAVRVILEGO-GRTSPSNFHPKLIEIFRE RWDGKGYPDGLKGEIPLESAIIAVADSFDAMTSDRPYRKANSHEAVRVILER-GPQWDPEILKALDYFSG RWDGTGYPFGLKDDEIPLSAIIAVADSFDAMTSDRPYRKANSHEAVRVILEGO-GRTSPSNFHPKLIEFFEG RWDGKGYPDGLKGEIPLESAIIAAVADSFDAMTSDRPYRKANSHEAVRVILEGO-GRTSPSNFHPKLIEFFEG RWDGKYFGLKDDEIPLSAIIAAVADSFDAMTSDRPYRKANSHEAVRVILEGO-GRTSPSNFHPKLIEFFEG RENGLGYPRRLVGEKISLYGKIIAVACSYEAATAPRSYKEMSNAAEGIVDLVRNA-NTQYDAVISRALLFALSF	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355	2984261 2688290 1652854 1653077 4980522 4980682 4981692 4981695 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323169
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1170_Tma TM170_Tma TM1678_Tma TM1678_Tma TM1699_Tma TP064_Tpa TP0912_Tpa	ES. SI AV LA NW LG DV DV LA CV RE FA VA DV SV SV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE AIWIRHHE SKWVRHHE RNIALYHHE AKIVRHHE CVGALERRE GLSVLOHHE	XYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEIPFEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKKFDPKVVEAFLEIFSD KFDGSGYPGGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPGGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKCLS-GKFYSPFVVEAFLEIFSD KFDGSGYPGGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKLLS-GKFDPKVVEAFLEIFSD KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVEAFLEALEFSD KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVEAFLEFLEG RWDGKGYPDGLKGEEIPLEARIISVCDYYAALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPLLVDLLFEILEK RVDGSGYPFGLKDEIPLLARILAVADVFDALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPLLVDLLFEILEK RVDGSGYPFGLKDEIPLEARVICVADSYZAATAPRSYKKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSGQYPDGLRGEEIPLESRIIAVADAFDAMTSDRPYRKAMDRKTALEILRNR-GPQWDPEILKIALDYFSG RENGLGYPRRLVGEKISLYGKIIAVACSYEAATAPRSYKEMKNAAEGIVDLVRNA-NTQYDAVISRALLFALSF RWDGTGYPFKLEGEMIPLSARITSIADVYDALRSERPYKKALSHEEAVRVILEGD-GQSFDPILARVFQKIHTK HWNGKGYPOSLSGNKITIGARILAVACSYEAATAPRSYKEMFSGYQAMKSLLADN-ARREDPVXKAMIGSVGI	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323066 3323189 3323224
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM145_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM167_Tma TM167_Tma TM1678_Tma TM1682_Tma TM1699_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal	ES. SI AV LA NW LG DV LA DV RE FA VA DV RE SI SL	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHHF VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE SKWVRHHE SKWVRHHE RNIALYHHE AKIVRHHE CVGALEHRE KEIALSHHE GLSVLQHHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVAFMAIHPT RIDGYGGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLIRE KYNGKGYPHGLKGEEIPLSAQIAAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPGLKGEEIPLSAQIAAVADSFDAMVSHRPYKRKSVEESIQELRRNA-GRQFDPVIVEKFIEFLG RWDGKGYPGLKGEEIPLSAUICVADSYQAMRSDRPYKRAKSVEESIQELRRNA-GRQFDPVIVEKFIEFLG SYDGTGYPFGLKDGEIPLSAUISVADVFDAITSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK WYSGQYPFGLKDGEIPLESRIIAVADVFDAITSDRPYRKAMKPEDAAALMKKMP-LDQEIVCILEKHLSEFVS SYDGTGYPFGLKDGEIPLESRIIAVADAFDAMTSDRPYRKAMSAALEILERRNE-GPQWDPEILKIALDYFSG RENGLGYPRRLVGEKISLYGKIIAVACSYEAATAPRSYKEMKNAAEGIVDLVRNA-NTQVDAVISRALLFAISF RWDGTGYPFQSLSGNKITIGARILAVTDNFAAMLAPKSYRKFSGYQAMKSLADN-ARRFDPDVIXKAMIQSVGI	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982258 4982276 3323066 3323189 3323224 5672654
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1467_Tma TM1682_Tma TM1682_Tma TM1682_Tma TM1682_Tma TP0764_Tpa TP0877_Tpa TP0912_Tpa Orf14_Bhal	ES SI AV LA NW DV DV LA VA CV FA VA DL SV SV SL	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE AIWIRHHE SKWVRHHE CVGALBHE KEIALSHE GLSVLQHHE LSAVLQHHE LSAVLQHHE	NYDGTGYP FGKKGEEIPILARIYTIADSFEAMTTRIYKRAKSWNEALKELEELA-GKQFDPDIVPAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOA-DKKFDKRVLKLINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLDGYGYP FRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLVLAFHLELFFSD KFDGSGYPFGLKEEEIPFERILKIADTLDAMLSPRSYRNPFPLDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKES-GRFSPVKAFVQTIDE RWDGKGYPHGLKGEEIPLEARIISVCDYDAMSSTAMVSHRPYKRKSFDEAIREIKDLS-GRFSSFVVKAFVQTIDE RWDGKGYPHGLKGEEIPLEARIISVCDYDAAMSSTANVSHRPYKRKSFDEAIREIKDLS-GRFSPVKAFVQTIDE RWDGKGYPDGLKGEEIPLEARIISVADVFAALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLLARILAVADVFDAITSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLLARILAVADVFDAITSDRPYRKAMKPEDAAALMKKMP-LDQEIVGLLEKHLSEFVS SYDGTGYPFGLKDGEIPLLARILAVADAFDAMTSDRPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVSGCGYPGGLKDGEIPLLARILAVADAFDAMTSDRPYRKAMSFEDAAILKKMP-LDQEIVGLLEKHLSEFVS RWSGCGYPGGLKDGEIPLESRIIAVADAFDAMTSDRPYRKAMSFEDAAILMKKMP-NTQYDAVISRALLFALSF RWSGCGYPGSLKGEIFLESRIIAVADAFDAMTSDRPYKKALSHEEAKEIMMCNAKETUDPEIVGLUKALLFALSF RWDGTGYPFKLEGEMTPLSARITSIADVYDALRMERSYKKGFSHEQTTHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPQSLSGNKTTIGARILAVTDNFAAMLAPKSYKKGFSHEQTTHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPQSLSGNKTTIGARILAVTDNFAAMLAPKSYKKSPRYNVLEDLAAKSLEAPPOVIKAMIQSVGI REDGTGYPLGVRQKTHFYQIIAVADVFDAICSORCYKKKKKYRVKLEFIVALLEFERSGRDHTTHVLVLVSOUTQ	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982256 3323066 3323189 3323224 5672654
Aq_2027_Aq BB0374_Bb s1l1624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1147_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1699_Tma TP0764_Tpa TP0877_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc 27f Vcbo	ES. SI AV LA NW LA DV LA TA TA SV SL AA.	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVCHHE ALWVCHHE ALWICHHE SKWVRHHE SKWVRHHE SKWVRHHE CVGALBHE KEIALSHID GLSVLQHHE LSÄVLQHHE AA IAHEHE	NYDGTGYP FGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIELKDA-DKKFDDRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQASIELGGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKLEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKLVEAFLEIFSD KFDGSGYPFGLKGEEIPFERILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLLRE KYNGKGYPHGLKGEEIPFERVLCVADSYQAMRSDRPYKIRKSFDEAIREIKDLS-GRYSPFVVEAFLKIEFEG RWDGKGYPDGLKGEEIPFERVLCVADSYQAMRSDRPYKRCKSVEESIQELRRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPGLKDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYSPFVVEAFLVL FERUGKGYPGLKDEIPLSAQIAAVADSFDAMVSHRPYKRALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIIAVADAFDANTSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDDEIPLSAQIXXIVVDYDALTSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RWDGKGYPGSLKDGEIPLSARIIAVADAFDAMTSDRPYRKAMSDRKALEILRRNE-GPQWDPEILKIALDYFSG RENGLGYPFRLVGEKISLYGKIIAVACSYEAATAPRSYKEMKNAAEGIVDLVRNA-NTQYDAVISRALLFALSF RWDGTGYPFLFKLGEMIPLSARITAYADYDALRMERSYKKGFSHEQTHMILEQS-GSFDPILARVFQKIHTK HWNGKGYPQSLSGNKITIGARILAVTDNFAAMLAPKSYRKFFSGYQAMKSLLADN-ARRFDDDVIKAMTQSVGI REDGTGYPLGVRGQKHFYAQILAVADVYHAMISERKYRKKSPHVVLEDLAKER-FGKFNPRIVHVLVSQITQ KWNGTGYPQCKKGEDIHLYGTIAVADAFDANTSDRYKKARKKKSPYMVLEDLAKER-FGKFNPRIVHVLVSQITQ	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222	2984261 2688290 1652854 4980682 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982258 3323189 3323224 5672654 2126438
Aq_2027_Aq BB0374_Bb sll1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1170_Tma TM1170_Tma TM11678_Tma TM1678_Tma TM1682_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho	ES. SI AV LQ DV DV LQ CV LA DV CV EA SV SL AA.	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE AIWIRHHE VPVVLYHHE RNIALYHHE AKIVRHHE CVGALEHRE GLSVLQHHE LSAVLQHHE AAIAHEHHE AAIAHEHE	KYDGTGYP FGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOD-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEEIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKEEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTDAMLSPRSYRNPFPLDDVIKEIERLR-GKKFDPVVEAFLEIFSD KFDGSGYPGGLKGEEIPIEARVLCVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPFEARVLCVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLEARIISVCDYVALTSDRPYRRALSHEEAKEIMMKNMKGTVLDPRLVDLFEIEK RVDGSGYPGGLKGEEIPLESNIISVADVYEALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLFEIEK RVDGSGYPGGLKGEEIPLESNIISVADVYEALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLFEIEK RVDGSGYPFGLKDEIPLESNIISVADVYEALTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFEE WYSGQGYPGCLRGEIPLESNIIAVADAFDAMTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFEE WYSGQGYPGCLKGKISLGKKIIAVACSYEAATAPRSYKEMSNAAEGIVDLVRNA-NTQYDAVISRALLFALSF RWDGTGYPFKLEGEMIPLSARITAVADYDALRMERSYKKGFSHEQTHMILEQS-GQSFDFILKTALDYFSG RWDGTGYPFKLEGEMIPLSARITSIDVADAFDAMTSDRPYRKAMDRKTALEILRRNE-GPQWDEIIKTALDYFSG RWDGTGYPFKLEGEMIPLSARITSIDVADAFDAMTSDRPYRKAMDRKTALEILRRNE-GPVIVAVISRALLFALSF RWDGTGYPFKLEGEMIPLSARITSIDVAATDNFAAMLAPKSYRKFSHEQTHMILEQS-GQSFDFILARYFQKIHTK RWDGTGYPCSLSGNKITIGARILAVADVYHAMISERKYRKKSPHVLEDLAKER-FGKFNPRIVHVLVSQITQ KWNGTGYPQCKKGEDHLYGRIVALADVFAAISERYRAMPHENDHVLEDLAKER-FGKFNPRIVHVLVSQITQ KWNGTGYPQCKKGEDHLYGRIVALADVFAAISERYRKKKSPMVLEDLAKER-FGKFNPRIVHVLVSQITQ KWNGTGYPQCKKGEDHLYGRIVALADVFAAISERKYKKKSPMVLEDLAKER-FGKFNPRIVHVLVSQITQ	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222 136	2984261 2688290 1652854 4980522 4980682 4981264 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM175_Tma TM1145_Tma TM1145_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut.	ES. SI ALA NW. LG. DV DV LA DV RE FA VA DI SL SL AA.	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHHT VPIVRHHE ALWVVQHHE AEVALFHHE ALWVVQHHE ALWVVQHHE AIWIRHHE SKWVRHHE CVGALEHRE KEIALSHHE GLSVLQHHE AAIAHPHE AAIAHPHE AAIALFHE	KYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOD-DKKFDKRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVAFMAIHPT RIDGYGGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEIPEGRILKIADTLDAMLSPRSYRNFFPLDDVIKEIERLR-GKKFDPLLADLTVDLIRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVAFVQATILE RWDGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVAFVQATILE RWDGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVAFVQATILE RWDGKGYPHGLKGEEIPLEARIISVCDYVEAITSDRPYRKAKSPESIQELRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPGLKGEEIPLEARIISVADYDALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLSAQIAAVADSFDAMTSDRPYRKAKSVEESIQELRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPGLKDGEIPLSAQIXKIVDYDALRSERPYKKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSGQYPGLKDGEIPLEARVICVADSYQAMRSDRPYRKAKSVESIQELRNA-GRQFDPVIVEKFIEFLEG RWDGKGYPGLKDGEIPLEARVICVADSYQAMTSDRPYRKAKSPUTIEGD-GRTSPSNFHPKLIEIFRE WYSGQYPGLKDGEISLGKIIAVACSYEAATAPRSYKEMKAAEGIVDLVNA-NTQVDAVISRALLFALSF RWDGTGYPFKLEGEMIPLSARITSIADVYDALRESKYKKGFSHGUMKSLADN-ARRFDDVIKAMIQSVGI REDGTGYPGSLSGNKITIGARILAVDVFAALSMENSYKKKSPYMVLEDLAKER-FGKFNPRIVHVLVSOITQ KWNGTGYPGKKGEDIHLYGRIVALADVYHAMLSERKYRKKSPYMVLEDLAKER-FGKFNPRIVHVLVSOITQ KWNGTGYPGKKGEDIHLYGRIVALADVFAALGSDRCYKKAWELERIVALLEEEK-GRHFDPLIEVFFENLEE KIDGTGYPGVGKSGLSGNLHIYGRITSIADVFAALGVFAALGSDRCYKKAWELERIVALLEERC-GRYDDVIKAMIQSVGI REPGTGYPGVGLSGNDHIYGRITSIADVFAALGVFAALGTKRVKNAAELIVKLLHFREK-GROPDVIKATINTITTADIFAALGSDRCYKKAWELERIVALLEEK-GRHFDPLIEVFFENLEE KIDGTGYPGVGLSGNDUHIYGRITSIADVFAALGVFAALGTKRVKNAAELIVALLEEK-GRHFDRLIEVFFENLEE	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222 2136 	2984261 2688290 1652854 4980682 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982258 4982276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1699_Tma TM1699_Tma TM0699_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetobut.	ESS SIL AVV LA NWW LGQ DV DV LA A VA DV V LA VA DIL VA SV SIL SA A A A A IL	LEVILYHQE CLITLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHHE VPIVRHHE ALWVVQHE AEVALFHHE ALWVVQHHE AEVALFHHE SKWVRHHE SKWVRHHE CVGALBHRE KEIALSHHE GLSVLQHE LSAVLQHE AATAHEHHE ARMAGAHHE AIIALEHHE KNVLQHE	NYDGTGYP FGKKGEEIPILARIYTIADSFEAMTTRIYKRAKSWNEALKELEELA-GKQFDPDIVPAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKDA-DKKFDKRVLKLIINAISS RWDGSGYPALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSIED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLVAAFMAIHPT RLDGYGYPGLKGEEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLIRE KYNGKGYPHGLKGDEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLIRE KYNGKGYPHGLKGDEIPLEARIISVCDYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPLLADLTVDLIRE KYNGKGYPHGLKGDEIPLEARIISVCDYDVDAFAALSPRSYRNPFPLDDVIKEIERLR-GKDFDPLLADLTVDLIRE KYNGKGYPHGLKGDEIPLARIISVCDYDANDSFDAMVSHRPYKIKSFDEAIREIKDLS-GRYSPFVKAFVQTIDE RWDGKGYPHGLKGDEIPLARIISVCDYDALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLLARIISVCDYDALTSDRPYRKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLLARIISVCDYDALRSERPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIISVCDYDALRSERPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIISVCDYDALRSERPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLARIISVCDYDALRSERPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGGYPFGLKDGEIPLANIVVDALRSERPYKKALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGGYPFGLKDGEIPLANIVVDALRSERPYKKALSHEEAKEIMMKNAKGTVLDPRLVDLLFEILEK RVDGGYPFGLKDGEIPLANIVVDALRSENSKGFSHEQTTHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPGSLSGNKITIGARILAVTDNFAAMIAPKSYKKGFSHEQTTHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPGSLSGNKITIGARILAVTDNFAAMISERKYKKSPYMVLEDLAKER-FGKFNPRIVHVLVSOUTQ KWNGGYPGSLSGNKITIGARILAVTDNFAAMISERKYKKKSPYMVLEDLAKER-FGKFNPRIVHVLVSOUTQ KWNGGYPGSLSGNKITIGARILAVTDNFAAMISERKYKKSPYRAAMPIDKALAIMEENL-HTAIDEFGAALKKAINL RFDGTGYP-GVNGDEISIMTRIITTADIFDALSAERPYRAAMPIDKALAIMEENL-HTAIDEFGAALKKAINL RFDGTGYPGLSGNDHIYGRITSIADVFDALGTVANDANSDRVYKKKNEPFRAFEMFKTEG-ISIFDTDLIKVFLNNINV	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 2222 136 	2984261 2688290 1652854 1653077 4980522 4981682 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982253 4982256 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1147_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1678_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetobut.	ESS SILA AVV LA NWW LG DVV DVV LA DVV LA DVV DVV C VV LA V V V V SV SV SV SV SV SV SV SV SV SV S	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVCHHE ALWVCHHE ALWVCHHE SKWVRHHE SKWVRHHE SKWVRHHE KVILYHHE AKIVRHHE GLSVLQHHE LSÄVLQHHE AAIALBHE AIIALEHHE KNVILQHHE LDMILYHHE	KYDGTGYP FGKKGEEIPLLARIIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOD-DKKFDDRVLKLIINAISS RWDG SGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQASIIELGGEARRGWRDEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDG SGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKLEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKLVEAFLKLEKE NWDGTGYPDGLKGEEIPFERRILVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPFERRILVAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRYSPFVVEAFLKLEKE RWDGKGYPDGLKGEEIPFERRILVAVADSFDAMVSRRPYKIRKSFDEAIREIKDLS-GRYSPFVVEAFLKIEFEG RWDGKGYPGLKGEEIPFERRVLCVADSYQAMRSDRPYKRCKSVEESIQELRRNA-GRQFDPVIVEKFIEFEG RWDGKGYPGLKDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRYSPFVVEAFLVL LEK RVDG SGYPFGLKDGEIPLARIIAVADAFDANTSDRPYRKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RVDG SGYPFGLKDDEIPLESAQIAVADSFDAMTSDRPYRKALSHEEAKEIMKNNKGTVLDPRLVDLLFEILEK RVDG SGYPFGLKDDEIPLESAQIAVADSFDAMTSDRPYRKALSHEEAKEIMKNNMKGTVLDPRLVDLLFEILEK RVDG SGYPFGLKDDEIPLESAQIAVADSFDANTSDRPYRKALSHEEAKEIMKNNKGTVLDPRLVDLLFEILEK RVDG SGYPFGLKDDEIPLESAQIAVADSFDANTSDRPYRKALSHEEAKEIMKNNKGTVLDPRLVVLIEFIEG RWDGKGYPGCKREIVGEKISIYGKIIAVACSYEAATAPRSYKEMKNAEGIVDLVRNA-NTQYDAVISRALLFALSF RWGGTGYPFKLEGEMIPLSARITSIADVYDALRMERSYKKGFSHEQTTHMILEQS-GQSFDPILKIALDYFSG RENGLGYPRVLGEKISIYGKIIAVACSYEAATAPRSYKKKSPHWIEDLAKER-FGKFMPRIVHVLVSOLTQ KWNGTGYPQSLSGNKITIGARILAVTDNFAAMLAPKSYRKFFSGYQAMKSLLADN-ARRFDPDVIKAMIQSVGI REDGTGYPGVGKGEDIHLYRQILAVADVYDALSDRVYKKKSPYNVLEDLAKER-FGKFMPRIVHVLVSOLTQ KWNGTGYPGKKGEDIHLYRGITALAVTDNFAAMLAPKSYRKFSGYAAMSLLAIMEENL-HTAIDPECFAALKKALNL RFDGTGYPGYGKKGEDIHLYGRITALAVFDALGSDRCYKKAWELEIVALLEFREK-GRHFDPRLIEVFFENLEK RINGTGYPGYFGUKADEMGLYSRITALAVADAFDAMTSDRVYKKKSPRAFEMFKTEG-ISIFDTDILRVFLNNINV RYDGMGYPS	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222 136 	2984261 2688290 1652854 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982258 4982276 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1145_Tma TM1170_Tma TM11678_Tma TM1678_Tma TM1699_Tma TM0699_Tma TP0764_Tpa TP0877_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f _Vcho C1.acetobut. C1.acetobut. C1.acetobut.	ESS STILA AVV LA NWW LG DVV LA DV DV LA DV DV RE FA VA VA SV SV SV SV SI AAA SV SI AAA SV KI L L L DI L	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVQHHE ALWVQHHE AIWIRHHE XEVALFHHE CVQALFHHE CVQALFHHE CVGALFHE GLSVLQHHE LSAVLQHHE AAIAEHHE AIIALFHE AIIALFHE AIIALFHE AIIALFHE AIIALFHE AXIVCHE	KYDGTGYP FGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOD-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRDEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIICVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKIEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLKIEKE NWDGTGYPDGLKGEEIPIEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKKFDPKVVEAFLEIFSD KFDGSGYPDGLKGEEIPIEARVICVADSYDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPIEARVICVADSYDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPIEARVICVADSYDAMSDRPYKRALSHEEAKEIMMKNMKGTVLDPLLVDLLFEILEK RVDGSGYPFGLKDEIPIEARVICVADSYDAMSDRPYKRALSHEEAKEIMMKNMKGTVLDPLLVDLLFEILEK RVDGSGYPFGLKDEIPIEARVICVADSYDAMTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEFIEG RWGGKGYPDGLRGEEIPLESRIIAVADSFDAMTSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSGQYPGLKDEIFIEARVICVADSYDAMTSDRPYRKAKSFSHEQTHMILEQS-GQSFDPILKTALDYFSG RENGLGYPFRLVGEKISLYGKIIAVACSYEAATAPRSYKKASHEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSGQYPGGLKDEIPIEARIIAVACSYEAATAPRSYKKASFSHEQTHMILEQS-GQSFDPILKTALDYFSG IRWGGTGYPFKLGEMIPISARITSIADVYDAIRMERSYKKSFSHEQTHMILEQS-GQSFDPILKKALFALFALSF RWGGTGYPFKLGEMIPISARITSIADVYDAIRMERSYKKKSPHVLEDLAKER-FGKFNPRIVHVVSQLTQ KWNGTGYPGLGVRGEIHLYGRIVALADVFDAIGSDRCYKKAWELERIVALLEEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGLSGNITTIGARITSIADVFDALGSDRCYKKAWELERIVALLEEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGSLSGNITTIGARITSIADVFDALGSDRCYKKAWELERIVALLEEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGLSGNITHIYRIISIADVFDALGSDRCYKKAWELERIVALLEEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGSLSGNITTIGARITSIADVFDALGSDRCYKKAWELERIVALLEEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGSLSGNITHIYRIISIADVFDALSDRVYKKKSPFRAFEMFKTEG-ISIFDDTULRVFINNINV RYDGMGYPSGLKGEIPLISRIIKADASDAALSPRSYKKSDINYVISELRKYK-GIIYDPCVAAFIEYLSS SYDGTGYPGLKADENIYRTIATADSVDAMLSPRSYKKSKDINYVISEL	182 299 322 356 384 335 275 337 349 608 419 148 550 557 319 355 368 306 222 136 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM175_Tma TM1145_Tma TM1145_Tma TM1467_Tma TM1682_Tma TM1682_Tma TM1682_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut.	ESS STIA AVV LA NWW LG DV DV DV DV DV DV DV DV DV DV DV DV DV	LEVILYHQE CLILTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE AIWIRHHE SKWVRHHE RNIALYHHE AKIVRHHE CVGALEHRE KEIALSHU GISVLQHHE AAIAHEHHE ANIACHHE ANIACHHE ANIACHHE RNILQHHE AKIVRHHE RIISLQHHE RIISLQHHE	KYDGTGYP FGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELKOD-DKKFDRRVLKLIINAISS RWDGSGYPDALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVAFMAIHPT RIDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR-GKKFDPKVVEAFLEIFSD KFDGSGYPGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKLS-GKFDPLLADLTVDLIAE KYNGKGYPHGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKLS-GKFDPKVVEAFLEIFSD KFDGSGYPGGLKGEEIPLEARIISVCDYYDVLT-6-WEGRTKTHEEALEIKCS-GKFDPLLADLTVDLIAE KYNGKGYPHGLKGEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS-GRFYSPFVVKAFVQTIDE RWDGKGYPDGLKGEEIPLARILSVADVYEAITSDRPYRRALSHEEAKEIMMKNMKGTVLDPLVDLLFEILEK RVDGSGYPGLKGEEIPLEARVLCVADSYQAMRSDRPYKRALSHEEAKEIMMKNMKGTVLDPLVDLLFEILEK RVDGSGYPGLKGEEIPLARIISVADVYEAITSDRPYRKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSCQGYPGLKDEIPLARITSIADVYDALRSERPYKKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE WYSCQYPDGLRGEEIPLESRIIAVADAFDAMTSDRPYRKAMDRKTALEILRNRE-GPQWDPEILKIALDFSG RWDGTGYPFKLEGEMIPLSARITSIADVYDALRMERSYKKGFSHEQITHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPQSLSGNKITIGARILAVADVDALRMERSYKKGFSHEQITHMILEQS-GQSFDPILARVFQKIHTK HWNGKGYPGLSGNKITIGARILAVADVPAALSDRYKKKSPHVLEDLAKER-FGKFNPRIVHVLVSOLTQ KWNGTGYPGLSGNKITIGARILAVADVDALADVFDALGSDRCYKKAWELERIVALLEEK-GRHFDPRUIKAMINSVGI REDGTGYPGLSGNDIHLYGRIVALADVFDALGSDRCYKKAWELERIVALLEEK-GRHFDPRLIEVFFENLEE KIDGTGYPGLSGNDIHLYGRITAIDIFDALSAERPYRAMPIDKALAIMEENL-HTAIDPECFAALKKAINL RFDGTGYPGSLKGEIPLESITATIITTADIFDALSAERPYRAMPIDKALAIMEENL-HTAIDPECFAALKKAINL RFDGTGYPGSGLKGEIPLESIZATIADVFDALGTWYDAMTSDRVYKKKSPFRAFEMFKTEG-ISIFDTDILIRVFINNINV RYDGMGYPSGLKGEIPLISRIIAADVADAFDAMSNDRYKKKSKDRSFAKSCISSAKETSAVKEIKNS-GTQFDPVVAAFISSLDIITG RAGGYPGGYPGINLKGEIPLASSIKADSVAMLSP	182 299 322 356 384 335 337 349 608 419 148 550 319 355 368 306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982256 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1699_Tma TM0699_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut.	ESS SILA AVV LA NWW LG DV DV LA DV DV DV DV DV DV DV DV DV DV DV DV DV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHHE VPIVRHHE ALWVVQHE ALWVVQHE ALWVVQHE ALWVVQHE ALWVVQHE ALWVQHE ALWITHHE CVGALEHRE KEIALSHHE GLSVLQHE ALTALEHE ALTALEHE ALTALEHE ALTALEHE ALTALEHE KNVILQHE ALTALEHE AKIVRHHE IDMILYHE AKIVRHE	XYDGTGYPFGLKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA-GKQFDPDIVPYAVRLSR NIDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASIIELIKDA-DKKFDKRVLKLIINAISS RWDGSGYPALRGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS-GKLYSPEVVEAFLKEEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKFDPKVVEAFLEIFSD KFDGSGYPFGLKEEEIPEGRILKIADTLAMLSPRSYRNPFPLDDVIKEIERLS-GKLYSPEVVEAFLEIFSD KFDGSGYPFGLKEEEIPEGRILKIADTLAMLSPRSYRNPFPLDDVIKEIERLS-GKFDPLVVAFLEIFSD KFDGSGYPFGLKEEEIPEARVICVADSFDAMVSRRPYKKSFDEAIREIKNLS-GKFDPLVVAFVEAFLEI RWDGKGYPDGLKGEEIPLEARVICVADSYQAMSDRPYKRKSFDEAIREIKDLS-GRFYSPFVVAFVQTIDE RWDGKGYPDGLKGEEIPLEARVICVADSYQAMSDRPYKRKSFDEAIREIKNLS-GRFYSPFVVAFVVAFVET RWDGKGYPDGLKGEEIPLSRIISVADVYEALTSDRPYKRALSHEEAKEIMMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLLARILAVADVFDALTSDRPYKKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLSRIISVADVYEALTSDRPYKKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE RWDGGYPFGLKDGEIPLLARILAVADVFDALTSDRPYKKALSHEEAVRVILEGD-GRTSPSNFHPKLIEIFRE RWDGGYPFGLKDGEIPLLARILAVADVFDALTSDRPYKKALSHEEAVRULEGD-GRTSPSNFHPKLIEIFRE RWDGTGYPFGLKDGEIPLLARILAVADVFDALTSDRPYKKALSHEEAVRULEDA-GRGSVPFGLKALLFALSF RWDGTGYPFGLKDGEIPLLARILAVADVFDALSSYRKFSYSYSAYGFSGYQAMKSLLADN-ARRFDPDVIKAMLSKGI RENGLGYPRLVGKIISLYGKIIAVACSYEAATAPRSYKEMKNAAEGIVDLVNA-NTQYDAVISRALLFALSF RWDGTGYPFGLKGEDIHLYGRIVALAVDVYAMISERKYKKSPYMVLEDLAKER - GRYFDPDVIKAMIQSVGI REDGTGYPLGVRGKIHFAQILAVADVYAMISERKYKKKSPYNLEDLAKER - FGKFNPRIVHUVSOLTQ KWNGTGYPLGVGGSDDIHLYGRIVALAVDYDALGSDRCYKKAWELERIYALLEEKE - GRYFDPDVIKAMLSVGI RFDGTGYPLGLSGNDIHIYGRISITAADVFDALGTKRVKKSPFRAFEMFKTEG - ISIFDTDILKVFLNNINV RYDGMGYPSGLKGEDIHLYGRIVALADVFDALGTKRVKKSPFRAFEMFKTEG - ISIFDTDILKVFLNNINV RYDGMGYPSGLKGEDIHLYGRIVALADVDAMTSDRVYKKKSPFRAFEMFKTEG - ISIFDTDILKVFLNNINV RYDGMGYPSGLKGEDIFLISRIIVAVADVADAMTSDRVYKKSKSDINVISELRKYK GISFDYAVEEISKVOAFSKSVIP RYDGSGYPGKKGEISLGFGRISICOVFAXSANRYRLRFKPN	182 299 322 356 384 335 275 337 349 608 557 319 355 368 306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982253 4982256 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s111624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1147_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1678_Tma TM1679_Tma TP0764_Tpa TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut.	ESS SILA AVV LA NWW LG DVV DVV LA DVV DVV LA DVV DVV DVV DVV DVV DVV DVV DVV DVV DV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AETALTHHE VPIVRHHE AETALTHHE VPIVRHHE ALWVCHHE ALWVCHHE SKWVRHHE SKWVRHHE SKWVRHHE KRIALSHID GLSVLQHHE LSÄVLQHHE LSÄVLQHHE LSÄVLQHHE LDMILYHHE ARITALEHHE KNVILQHHE LDMILYHHE AKIVRHHE RIISLQHHE SYWASSHE	XYDGTGYPFGKKGEEIPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELEELA - GKQFDPDIVPYAVRALSR NIDGTGYPKGLTSEN IS IESNIIGAASAYSA ILLDKAYKKSFNSGASIIELIKDA - DKKFDKRVLKLIINAISS RWDGSGYPALIGKEIPLLAQVFQIIDIFTALTSNRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK RWDGTGYPQGLSGDDIPIAARLMAIADVYDALISSRVYKEAVNHEEAVAIEASGY - GUHPDPVVDAFMAIHPT RLDGYGYPFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGYPRGLKEEEIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS - GKLYSPEVVEAFLKLEKE NWDGTGYPDGLKGEEIPLEARIISVCDYYDVIT - 6 - WEGRTKTHEEALEIIKKEA - GKKFDPKVVEAFLEIF5D KFDGSGYPFGLKEEEIPPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLF.GKDFDPLLADLTVDLRE KYNGKGYPHGLKGEEIPLEARIISVCDYYDVIT - 6 - WEGRTKTHEEALEIIKKEA - GKFFDPVVAFVQTIDE RWDGKGYPGLKGEEIPEGRILKIADTLDAMLSPRSYRNPFPLDVIKEIERLF.GKDFDPLADLTVDLRE KYNGKGYPHGLKGEEIPEGRILKIADTLDAMLSPRSYRNPFPLDVIKEIERLF.GKDFDPLADLTVDLRE RWDGKGYPGLKGEEIPEGRILKIADTLDAMLSPRSYRNPFPLDVIKEIERLF.GKDFDPLADLTVDLRE RWDGKGYPGLKGEEIPEGRILKIADTYDAMSDRPYKRKSVEESIQELRNA - GROFDPVIVEKFIEFLEG RWDGKGYPGLKGEEIPLSRIISVADVYEALTSDRPYKKALSHEEAKEIMKNMKGTVLDPRLVDLLFEILEK RVDGSGYPFGLKDGEIPLEARIIAVADAFDAMTSDRPYKKALSHEEAVRVILECO - GRTSPSNFHPKLIEIFRE WYSGCGYPDGLKGEEIPLESRIIAVADAFDAMTSDRPYKKALSHEEAVRVILECO - GRTSPSNFHPKLIEIFRE WYSGCGYPGLKDGEIPLESRIIAVADAFDAMTSDRPYKKALSHEEAVRVILECO - GRTSPSNFHPKLIEIFRE WYSGCGYPGGLKGEIPLESRIIAVADAFDAMTSDRYKKANAEGIVDIVRNA - NTQYDAVISALLFALSF RWDGTGYPFKLECEMIPLSARITSIADVYDALFMERSYKKGFSHEQTTHMILEQS - GQSFDPILARVFQKIHTK HWNGKGYPQSLSGNKITIGARILAVTDNFAAMLAPKSYRKKSFSHQUIAMSILADN - ARRFDPDVIKAMIQSVGI REDGTGYPFGLKGENIPLSARITSIADVYDALGSDRCYKKAWELERIVALLEEEK - GRHFDPRLIEVFENLEE KLDGTGYPLGSVGKGEIPLISRIIATADVFDALGSDRCYKKAWELERIVALEEEK - GRHFDPRLIEVFENLEF KNGGTGYPQGKKGEDIHJGRIVAADVDAMTSDRVYKKKSPRAFEMFKTEG - ISIFDTDILRVFLNNINV RYDGMGYPSGLKGNEIPQCASIVAVADAFDAMTSNRVYRKGKDRSFAVKEIIKNS - GQFDPVLIEVLSNLDK RNGTGYPRIKLEGENISILSRIVAVADYDAMTSDRVYRKSKDINVISELEYKSG GSMFDPDVAAFIEYLSE SYDGTGYPDNLKGEDIPLISRIIKAVADVDAMTSDRVYRKSKDINVVISELEYKGSGSGRISJKSKSVIP RYDGSGYPRGLKGENISJCGRIIS	182 299 322 356 384 335 275 337 349 608 419 4550 557 319 355 368 306 222 2126 	2984261 2688290 1652854 4980682 4980682 4981264 4981692 4981692 4982032 4982253 4982253 4982258 4982276 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1145_Tma TM1147_Tma TM11678_Tma TM1678_Tma TM1678_Tma TM1699_Tma TP0764_Tpa TP0877_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f _Vcho Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut.	ESS STITA AVV LA NWW LG DV DV LA DV DV LA DV DV RE FA VA VA SV SL SV SL SV SL SV SL SV SL SV SL SV SL SV SV SV SL SV SV SV SV SV SV SV SV SV SV SV SV SV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVQHHE ALWVQHHE ALWVQHHE AIWIRHHE XFVVLYHHE RNIALYHHE CVGALEHRE CVGALEHRE CVGALEHRE CVGALHRE ALIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE RIISCHE IIGSVNHE IIGSVHHE IIGSVHHE	YDG TGY PFGKKGEE I PLAR I YT I ADSFEAMTTRE I YKRAKSWNEALKELEELA - GKQFDPD T VPXAVRALSE NIDGTGY PKGLTSEN I SIESNI I GAASAYSAI I LDKAYKKSFNSGASI IELIKDA - DKKFDKRVIKLJI NA ISS RWDGSGY PGLKGEKI PLAQVPQI I DIFTALTSNRRYKEAVNTAQALEI LQGEARRGWRNPEMVEKFVNF INK KWDGTGY PQGLSGDDI PLARIMA I ADVYDALISSRVYKEAMPHEEAVAI I EAGY - GQHFDPDVVDAFMAI HPT RLDGYGY PFRLKGEEMTLEDKI I QVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGY PRGLKEREI PI SGLIVAVADSFDAMYSRPYKNPKPLEEAFREI ESLS - GKLYSPEVVEAFLEKEKE NWDGTGY PDGLKGEE I PLEARI I SVCDY YDVLT - 6 - WEGRTKTHEEALEI I KKEA - GKKFDPKVVEAFLEI FSD KFDGSGY PFGLKGDEI PLEARI I SVCDY YDVLT - 6 - WEGRTKTHEEALEI I KKEA - GKKFDPKVVEAFLEI FSD KFDGSGY PFGLKGDEI PLSAOI I AVADSFDAMVSRPYKNPKPLEEAFREI ESLS - GKLYSPEVVEAFLEKE RWDGKGY PHGLKGDEI PLSAOI I AVADSFDAMVSHRPYNI RKSFDEAI REIKDLS - GRYFSPFVKAFVQT I DE RWDGKGY PDGLKGEEI PLSSAI I SVADVYEALTSDRPYKRGKSVEESI QELRRNA - GRQFDPVI VEKFI EFLEG RWDGKGY PDGLKGEEI PLSRI I SVADVYEALTSDRPYKRALSHEEAKEIMKNMKGTVLDPRLVDLLFEI LEK RVDGSGY PFGLKDDEI PLSRI I SVADVYEALTSDRPYKKALSHEEAKEIMKNMKGTVLDPRLVDLLFEI LEK RVDGSGY PFGLKDDEI PLESTI I AVADAFDAMTSDRPYKKALSHEEAVRVILEDG - GRTSPSNFHPKLI EFFES SYGGTGY PFGLKDDEI PLESTI I AVADAFDAMTSDRPYKKALSHEEAVRVILEDG - GRTSPSNFHPKLI EFFS SYGGTGY PFGLKDDEI PLESTI I AVADAFDAMTSDRPYKKALSHEEAVRVILEDG - GRTSPSNFHPKLI EFFES RWDGGY PFGLKDE FI EAQI VKI VDVYDAL RMERSYKKGFSHEOTTHMI LEQS - GQSFDFI LARVFQKI HTK HWNGKGY PQSLSGNKIT I GARI LAVTDNFAAMLAPKSYKKFSPSGYQAMKSLLADN - ARRFDPDVI KAMI QSVGI REDGTGY PFGLKDGEN I PLSARI TSI ADVYDAL RMERSYKKKFSPYMVLEDLAKER - FGKFNPRI VHVLVSQITT KWNGTGY PGSLKGDEI SLMTRI I TTADI FDALSAERPYRAAMPI DKALAIMEENL - HTAI DPECFAALKKALNL RFDGTGY PLGVRGQKTHFYAOI I AVADAYDAMTSDRVYKKKKSPFAFFEMFKTEG - I SI FDTDI LRVFLNNINV RYDGMGY PSGLKGNEI FYCASI VAVADAFDAMTSDRVYKKKSDRSAVKEI LHFRREK - GRYPPKLI EI FFENLEE SYDGTGY PYGI KADEMGI YSRI I AI ADVYDAMTSDRVYKKKSKDINYVI SELRKYK - GI I YDPKVTOSMLDI LTG RAGGQY PEGLKGNEI SJCGRI I SI COVFTAVSANRSYRLKFKDRAFAVKEI LLSGS - GSKFDPVVDAFASKSVLP RYDGSGY PGLKGNEI SJCGRI I SI LSVADAVDAYDAMTSDRVYKK	182 299 322 356 337 349 6419 148 550 557 319 355 368 306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM175_Tma TM1145_Tma TM1145_Tma TM1467_Tma TM1467_Tma TM1678_Tma TM1699_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut.	ESS SILA AVV AVV LA DV DV DV DV DV DV DV DV DV DV DV DV DV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHHE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE AIWIRHHE RNIALYHHE RNIALYHHE RNIALYHHE AKIVRHHE CVGALEHRE KEIALSHU CGISVLQHE AAIAHE HAIALEHE KNVILQHE AAIAHEHE RNIALCHHE AIIALEHE KNVILQHHE AIIALEHE KNVILQHE GISVLHHE RIISLQHHE GIGVIMHE GLGVIMHE	KYDGTGY PFGKKGEET PLLARTIYT I ADSFEAMTTRR I YKRAKSWNEALKELEELA - GKQFDPDT VPYAVRALSR NIDGTGY PKGLTSEN TSTESNTIGAASAYSATILDKA YKKSFNSGASIIELIKDA - DKKFDKRVIKLIINA KWDGTGY PQGLSGDDIPIAARIMAIADVYDALISSRVYKEANTAQALEILQGEARRGWRPEMVEKFVNFINK KWDGTGY PQGLSGDDIPIAARIMAIADVYDALISSRVYKEANPHEEAVATIEAGY - GQHPPDVVDAFMATHPT RIDGYGY PFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGY PFRLKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSLED RWDGSGY PFRLKGEETPLEARIISVCDYYDVLT - 6 - WEGRTKTHEEALEIIKKEA - GKLYSPEVVEAFLEIFSD KFDGSGY PFGLKEEETPEGRILKIADTLDAMLSPRSYRNPFLDDVIKEIERLA - GKKFDPKVVEAFLEIFSD KYDGSGY PFGLKGEETPLEARIISVCDYYDVLT - 6 - WEGRTKTHEEALEIIKKEA - GKKFDPKVVEAFLEIFSD KYDGSGY PFGLKGEETPLARIISVCDYYDVLT - 6 - WEGRTKTHEEALEIIKKEA - GKFDPPLIADITVDLIKE KYNGKGY PHGLKGDEIPLSAQIAAVADSFDAMVSHRPKIRKSFDEAIREIKDLS - GRFYSPFVKAFVQTIDE KYNGKGY PHGLKGGETPLARTILVADVYADAISDRPYKRGKSVEESIQUELRNA - GRQFDPVIVEKFIEFEEG RWDGKGY PHGLKGGETPLARITAVADVYDALTSDRPYKRAKPEDAAALMKKMP - LDQEIVGILEKHLSEFVS SYDGTGY PFGLKDGETPLARITAVADVFDALTSDRPYKKAKSDEAIRUMNNKGTVLDPRLVDLIFEILEK RVDGSGY PFGLKDGETPLARITAVADVFDALTSDRPYKKAMDRTALEILRRNE - GPQWDPEILKIALDYFSG SYDGTGY PFGLKDGETPLSRIISAADAFDAMTSDRPYKKAMDRTALEILRRNE - GPQWDPEILKIALDFSG RENGLGY PFKLVGEKISLYGKIIAVACSYEAATAPRSYKKKSPKNEDAAALMKKMP - LDQEIVGILEKHLSEFVS SYDGTGY PFGLKDGETPLSRIISAADAFDAMTSDRYKKAMDRTALEILRRNE - GPQWDPEILKIALDFSG RENGGTGY PFGLKDGETPLSRIISAADAFDAMTSDRYKKKSPKNEDQITHMILEGS - GOSFDPILARVFQKIHTK HWNGKGY PQSLSGNKITIGARILAVTDNFAAMLAPKSYKKFSRUQTHLELAKER - FGKFNPRIVHVLVSQLTQ KWNGTGY PQGKKGEDIHLYGRIISAADYDALANDYDALTRERSYKKGSPKOTHENLEEK - GRHPDPRLIEFYFENLEE KIDGTGY PGCSGNDHIYGRITSIADVYDALTRERYKKKSPYRVLEDLAKER - FGKFNPRIVHVLVSQLTQ KWNGTGY PQGKKGEDIHLYGRITSIADVYDALSDRYKKKSPFRAFEMFKTEG - ISIFDTDILKVFLNINV RYDGGY PGGLSGNDHIYGRITSIADVFDALGTKRVYKDAWSINKILLEFFREK - GRYFDPCVLSAALFFENLEE KIDGTGY PGCKSGNDHIYGRITSIADVYDAMTSDRVYKKKSPFRAFEMFKTEG - ISIFDTDILKVLNSLT RYDGSGY PGGLKGNEIPCCASIVAVADAFDAMTSDRVYKKSKDINYVISEIKKSFINGA - GXFPDPVVLYSKIVYDAMTSINVVYDG RYDGSGY PGGLKGNEIPCGASIVAV	182 299 322 356 384 335 275 337 349 608 557 319 355 8306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982253 4982256 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1699_Tma TM0692_Tma TM0699_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut.	ESS SILA AVV LA NWW LG DV DV DV LA DV DV TA DV SV SI SV SV SV SI SA DI L DI KIL DI C SV SI SV SI SV SI SI SI SI SI SI SI SI SI SI SI SI SI	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE FIPAVRHE AEIALTHIF VPIVRHHE AEIALTHIF VPIVRHHE ALWVVQHHE AEVALFHHE ALWVVQHHE ALWVQHHE AIWIRHHE SKWVRHHE GLSVLQHHE AAIALSHHE GLSVLQHHE AAIALSHHE AIIALEHHE AIIALEHHE LDMILYHE LDMILYHE IDGVFNHHE SYWASSHHE GLGVLMHIE TKWASNHE GLGVLMHIE TKWASNHE ROIAYSHE	KYDG TSYP FCKKGEETPLLARIYT TADSFFAMTTRIYYRAKSWNEALKELELLA - GKOFDPDTVPYAVRALSR NLDG TGYPKGLTSENISIESNIT GAASAYSAILLDKAYKKSFNSGASILELIKDA - DKKPDKRVLKLIINAISS RWDG SCYPDALRGKEIPLLAQVFOIIDIFTALTSNRRYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDG TGYPQCLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDG TGYPQCLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTEAVALEAVAITEAGY - GOHFDPUVDAFMAIHPT RLDG YGYPFLKGEEMTLEDKIIQVADYFSALLEERPYRPANTVEKALSMVREEVKKKKLSQEAFELLESSED RWDG SCYPRCLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS - GKLYSPEVVEAFLKEEK NWDG TGYPDGLKGEEIPLEARTISVCDYYDVLT - 6 - WEGRTKTHEEALEIIKKEA - GKKFDPKVVEAFLEIFSD KFDG SGYPFCLKEEEITPECRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR - GKKFDPKVVEAFLEIFSD KYDG SYPFCLKEEEITPECRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR - GKFDPLLADLTVDLLRE KYNGKGYPHGLKGDEIPLSAQIAAVADSFDAMVSHRPYKIRKSFDEAIREIKDLS - GRYSPFVVKAFVQTIDE RWDG KGYPCLKGEEIPFEARVLCVADSYQAMRSDRPYKRGSVEESIQELRNA - GROFDPVIVEKFIEFLEG RWDG KGYPCLKDEIPILARILAVADVFDALTSDRPYRKAKSFDEAIREIKDLS - GRYSPFPVKAFVQTIDE RWDG KGYPCLKDEIPILARILAVADVFDALTSDRPYRKAKSFDEAIREKMKINKGTVLDPRLVDLFIELEK RVDG SGYPFGLKDEIPILARILAVADVFDALTSDRPYRKAMSPEDAAALMKKMP - LDQEIVGILEKHLSEFVS SYDG TGYPFGLKDDEIPIEAQIVKIVDVIDALRSERPYKKALSHEEANEUMKNMKGTVLDPRLVDLFIELFER WYSG GGYPDGLRGEEIPLESRIIAVADAFDAMTSDRPYRKAMDRKTALEIIRRNE - GPQWDPEILKIALDYFSG RENG TGYPFCLKDEIPIEARIIAVDNFAAMLAPKSYKKFSHEQTTHMILEGS - GQSTDPILAVFOKIHTK HWNG KGYPQCKKGEDIHLSARITSIADVYDALRSERYKKKSPHWLEDLAKER - FGKFNPRIVHVLVSQLTQ KWNG TGYPQCKKGEDIHLSARITSIADVFDALGSDRCYKKAWELERIVALLEEEK - GRHPDPLIEVFENLEE KLDG TGYPQCKGEDIHIYGRITALADVFDALGSDRCYKKAWELERIYALLEEEK - GRHPDPLIEVFENLEE KLDG TGYPQCKKGEDIHIYGRITALADVFDALGSDRCYKKAWELERIYALLEEEK - GRHPDPLIEVFENLEE KLDG TGYPQCKKGEDIHIYGRITALADVFDALGSDRCYKKANELERIYALLEEFTALKKALNI RYDG GYPGCKGEDIFIISRIIKAADVDAMSDRVYRKAKSDFRAFFMFKTEG - ISIFDTDILRVFLNNINV RYDG GYPGCKGEDIFIISRIIKAADVDAMTSDRVYRKAKSDFRAFFMFKTEG - ISIFDTDILRVFLNNINV RYDG GYPGCKKGEDIFIISRIIKAADVDAMTSDRVYRKAKSDFRAFFMFKTEG - GSKFDPKVQAFSKVIP RYDG TGYPPCLKGENESSCERISILANDYDAATSDRVYRKAKSDINVYSG SCHPFTNFFFI KING TGYPGC	182 299 322 356 384 335 275 337 49 608 419 148 550 355 368 306 222 212 6 	2984261 2688290 1652854 4980522 4980682 4981264 4981692 4981695 498276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1678_Tma TM1678_Tma TM1679_Tma TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho C1.acetobut. C	ESS SILA AVV LA NWW LG DVV DVV LA DVV DVV LA DVV DVV RE FA VA VA SV SSL SV SSL AAAS DIL SI V SL SA SV SL SSL SSL SSL SSL SSL SSL SSL SSL S	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AETALTHHE VPIVRHHE AETALTHHE VPIVRHHE ALWVCHHE ALWVCHHE ALWVCHHE SKWVRHHE VPVVLYHHE RNIALYHHE ALWVCHHE LSÄVLQHHE LSÄVLQHHE LSÄVLQHHE LSÄVLQHHE LSÄVLQHHE LDMILYHHE AR I IALEHHE KNVILQHHE LDMILYHHE AK I VRHHE RI ISLOHHE SYWASSHHE GLGVLMHE TKWASNHE	KYDG TSY PFCKKGEET PLLARIYT TADSFFAMTTRRIYKRAKSVNEALKELELLA - GKOFDPDTVPYAVRALSR NLDG TGYPKGLTSENISIESNI TGAASAYSAIILDKAYKKSFNSGASIIELIKDA - DKKFDKRVLKLIINAISS RWDG SCYPDALRGKE IPLLAQVFOIIDITTALTSNRYKEAVNTAQALEILQGFARRGWRNPEMVEKKIINA KWDC TGYPQLSGLSGDDIPIAARLMAIADVYDALISSRVYKEANPHEEAVAIIEAGY - GOHPDPDVVDAFMAIHPT RLDGYGYPFELKGEEMTLEDKIIQVADVFSALLEERPYRPANTVEKALSMVREEVKKKLSQEAFELLESSLED RWDG SCYPRGLKEREIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS - GKLYSPEVVEAPLKEEKE NWDG TGYPDCLKGEEIPIEARIISVCDYVDIT - 6 - WEGRTKTHEEALEIIKKEA - GKKFDPKVVEAPLKEIFSD KFDGSGYPFGLKEEEITPEGRILKIADTIDAMLSPRSYRNPFPLDDVIKEIERLR - GKDFDPKVEAPLKEIFSD KFDGSGYPGLKEEEITPEGRILKIADTIDAMLSPRSYRNPFPLDDVIKEIERLR - GKDFDPKVEAPLKEIFSD KFDGSGYPGLKGEEIPERANUCVADSFDAMVSRRPYKIRKSFDEAIKEIKDLS - GREYSPFVKAFVQTIDE RWDG KGYPGLKGEEIPERANICVADSTQAMRSDRPYKRGSVEESIQELRNNA - GROFDPVIVEKFIEFLEG RWDG KGYPGLKGEEIPLESRITSVADVYDALTSDRPYRKALSHEEAKEIMMKMMKGTVLDPRLVDLLFEILEK RVDGSGYPGLKGEIPLESRITSVADVYDALRSERPYKRALSHEEAKEIMMKMMKGTVLDPRLVDLLFEILEK RVDGSGYPGLKDEIPLESRITAVADAFDAMTSDRPYKKALSHEEAKEITMKINMSGTVLOPRLVDLLFEILEK RVDG GYPGCKGGEIPLESRITSVADVYDALRSERPYKKALSHEEAKEITMILEGS - GGSFDPILKIALDYFSG SYDG TGYPFGLKDDEIPIESRITSVADVYDALRSERPYKKALSHEEAKEITMILEGS - GGSFDPILKIALDYFSG RWDG GYPQSLSGNKITIGARITAVACYKAATAPRSYKKKKSPHWIELALARE - FGKFNPRIVHUVSGLI RENGTGYPFSLSGEIJLARITSIADVFDALGSCYKKAWELERIYALLEEK - GROFDPVIKAMIQSVGI RENGTGYPGSLKGEDIHISARITSIADVFDALGSCYKKAWELERIYALLEEK - GROFDPVIKAMIQSVGI RENGTGYPGSLKGEDIHISARITSIADVFDALGSCYKKAWELERIYALLEEK - GROFDPVIKAMIQSVGI RENGTGYPGSLKGEDIHISGRITAIADVFDALGSCKYKKAWELERIYALLEFK - GROFDPVIKAMIQSVGI RYDGSGYPGLKGEDISINTRITTADUFDALSERPYKAAMIDKALLHFREK - GROFDPVIKAMIQSVGI RINGTGYPGSLKGEDIFISSITIAIADVFDALGSVARSYLKKKSPTAFEMFXTG - ISIFDTDILKVFLNNINV RYDGMGYPGSLKGEDIFISSITIAIADVFDALGTKRVYKGKSNESFAVKETIKNS - GSKFDYKVQAFSKSVIP RYDGSGYPGLKGEDISISTIAIADVDAMTSDRVYKKKSPFRAEKTSINSGGICKANFOGYPOVAAFIEVES SYDGTGYPOLKGEDISTYTANSANFYKRKSKSPFNEAFENTRYSGS - GSMFDPFVLERATFFIN RYDGSYPGLKADEISJESECIMGVOJYQALTEDRYRKGLSHEAFKILNHTSDGYICEKAVFQLMASSVIP RYDGSGYPGLKADLFNARIAA	182 299 322 356 337 349 275 337 349 148 550 557 319 355 368 306 222 136 222 136 	2984261 2688290 1652854 4980522 4980682 4981264 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982258 4982276 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0186_Tma TM1147_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM1678_Tma TM1678_Tma TM1678_Tma TM1679_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Bhal Orf14_Chacetobut. Cl.acetobut	ESS SILA AVV LA NWW LG DV DV LA DV DV LA DV DV DV DV DV DV DV DV DV DV DV DV DV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVQHHE ALWVQHHE ALWVQHHE ALWVQHHE SKWVRHHE SKWVRHHE CVGALEHRE CVGALEHRE CVGALEHRE CVGALEHRE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE AIIALHHE CVGALHHE IDGVFNHE GLGVLHHE TKWASNHE CQIAYSHE	KYDG TSY FCKKKGEET PLLARIYTTADSFEAMTTRI YYKAKSYNEALKELELLA - GKOFDPDTVYAVRALSR NLDG TGYPKGLTSENISIESNITGAASÄYSAIILDKAYKKSYNSGASIIELIKDA - DKKFDKRVLKLIINAISS RWDGS GYPDALRGKEIPLLAQVFOIIDITALTSNRYKEAVNTAQALEILQGFARRGWRNPEMVEKVNNFINK KWDG TGYPQLSGDIPIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY - GOHFDPDVVDAFMAIHPT RLDG YGYPFLKGEEMTIEDKIIOVADVFSALLEERPYKPANTVEKALSGWREEVKKKLSOEAFELLESSIED RWDG GYPFGLKGEEIPISGLIVAVADSFDAMVSRRPYKNPKPLEEAFREIESLS - GKLYSPEVVEAFLEIFSD RWDG GYPGLKGEEIPISARIISVCDYVDVIT - 6 - WEGRTKTHEEALEIIKKEA - GKKFDPKVVEAFLEIFSD KFDG SGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR - GKKFDPKVVEAFLEIFSD KFDG SGYPFGLKEEEITPEGRILKIADTLDAMLSPRSYRNPFPLDDVIKEIERLR - GKKFDPKVVEAFLEIFSD KFDG SGYPFGLKGEEIPISAQIAAVADSFDAMVSHRPYKIRKSFDEAIKEIKDLS - GRFYSPFVVKAFVQTDDE RWDGKGYPGLKGEEIPISAQIAAVADSFDAMVSHRPYKIRKSFDEAIKEIKDLS - GRFYSPFVVKAFVQTDE RWDGKGYPGLKGEEIPISAQIAAVADSFDAMVSHRPYKIRKSFDEAIKEIKDLS - GRFSPFVVKAFVQTDE RWDGKGYPGLKGEEIPISAQIAAVADSFDAMVSHRPYKIRKSFDEAIKEIKDLS - GRFSPFVKAFVQTDE RWDGKGYPGLKGEEIPISSIISVADVFALTSDRPYKKAKSHEEAKEIMMKNMKGTVLDPRIVDLLFEILEK RVDG SYPFGLKDGEIPILARILAVADVFDALTSDRPYKKALSHEEAKEIMMKNMKGTVLDPRIVDLLYEI LEK RVDG SYPFGLKDGEIPILSSIIIAVAAFDAMTSDRPYKKALSHEEAKEIMMKNMKGTVLDPRIVDLIVDLLFEI LEK RVDG GYPFGLKDGLIFIEAQIVKIVDVDALRSERPYKKALSHEEAVRVILEGD - GRSPDPILKIALDYFSG RENGLGYPRGLKDEIPISSIIIAVAAFDAMTSDRPYKKALSHEEAVRVILEGD - GOSDPDILKIALDYFSG RENGLGYPRLVGEKISLYGKIIAVACSYEAATAPRSYKKKASHEVTMILLEGS - GOSDPDILKIALDYFSG RENGLGYPRLVGEKISLYGKIIAVACSYEAATAPRSYKKFSFSGYQAMKSLLADN - ARRFDPDVIKAMIOSUGI RKNG GYPFGLSGNDIHIYGRIVALADVFDALGSDRCYKKAWELERIYALLEEEK - GRHFDPRLIEVFFNLEE KIDG GYPFGUSGSGNIHIYAQIAAVADYDALGNFAWKNASHKKSSPRAFEMFKTEG - ISIFDDILAVATVHVIVSQUTQ KNNG GYPGGLSGNDIHIYGRITSIADVFDALGSDRCYKKAWELERIYALLEEK - GRHFDPLLEIFFENLEE KIDG GYPGCSGNGLISISTIAIADYDALSDRVYKKASSPRAFEMFKTEG - ISIFDDILAVATVHVIVSQUTQ KNNG GYPGGLKGREIYAAADYDAABDAMTSDRVYKRASHFREELEYAKAS - GYGPDPLYAAFSSVIP RYDGSGYPGGLKGREISISSIIIAAVADSVDAMSDRVYKRASHFREELEYAKSGSIGVKYVQAFSSVIP RYDG GYPGCKAGELISLSENISLANG AVADAYDAATSDRVYKRASKSUNAELILHANS - GYGPDPLVAA	182 299 322 356 337 349 6419 148 550 557 319 355 368 306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982258 4982276 3323066 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy slr2100_Sy TM0037_Tma TM0186_Tma TM175_Tma TM1145_Tma TM1145_Tma TM1467_Tma TM1682_Tma TM1682_Tma TM1682_Tma TM1699_Tma TP0764_Tpa TP0764_Tpa TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Spacetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Cl.acetobut. Spacetobut. Spacetobut. Cl.acetobut. Spacet	ESS SILA AVV LA NWW LG DVV LA DV DV LA DV DV RE FA SV SV SV SV SV SV SV SV SV SV SV SV SV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE ALWVQHHE ALWVQHHE AIWIRHHE VPVVLYHHE RNIALYHHE AKIVRHHE GLSVLQHHE LSAVLQHHE LSAVLQHHE AIIALEHHE RNIALEHHE RIISLQHHE	KYDGTGYP GKGEETPILARIYTIADSFEAMTTRRIYKRAKSWNEALKELELA-GKQEDPDIYVPAVKALSR NLDGTGYPKGLTSENISIESNIIGAASAYSAIILDKAYKKSFNSGASILELELA-GKQEDPDIYVPAVKALSR WWDGSGYPQLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILGGEARRGWKNPEMVEKFVNFINK KWDGTGYPQLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILGGEARRGWKNPEMVEKFVNFINK KWDGTGYPQLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILGGEARRGWKNPEMVEKFVNFINK KWDGTGYPGLSGDDIPIAARLMAIADVYDALISSRVYKEAVNTAQALEILGGEARRGWKNPEMVEKFVNFINK KWDGTGYPGLSGEETPLSKIIVAVADSFDAMVSRPYKNPKPLEAAFREIESLS-GKLYSPEVVEAFLEISSLED WWDGTGYPGLKGEETPLEARIISVCDYYDALTSCHYPKPANTVEKALSMCHKKKSCQEAFELLESSLED KWDGKGYPGLKGEETPLSAXIIVAVADSFDAMVSRPYKNPKPLEAAFREIESLS-GKLYSPEVVEAFLEIFSD KFDGSGYPFGLKGEETPLSAXIISVCDYYDALTSDRPYKRAKSFDEAIREIKNLS-GRFSPFVVKAFVOTIDE RWDGKGYPDGLKGEETPLSAXIISVCDYADSFDAMVSRPYKNIRKSFDEAIREIKNLS-GRFSPFVVKAFVOTIDE RWDGKGYPGLKGEETPLSAXIISVADYFALTSDRPYKRALSHEEAKEIMMKNMGTVLDPFLVDLIFEILEK RVDGSGYPFGLKDGEIFLARIIAVADYFDALTSDRPYKRALSHEEAKEIMMKNMGTVLDPFLVDLIFEILEK RVDGSGYPFGLKDGEIFLSRIIAVADYFDALTSDRPYKRALSHEEAKEIMMKNMGTVLDPFLVDLIFEILEK RVDGSGYPFGLKDGEIFLSRIIAVADYFDALTSDRPYKKALSHEEAVRVLEGD-GRTSPSNFHPKLIEIFRE RVDGSGYPFGLKDGEIFLSRIIAVADYFDALRSERPYKKALSHEEAVRVLEGD-GRTSPSNFHPKLIEIFRE RVSGGYPFGLKDGEIFLSRIIAVADYFDALRSERPYKKALSHEEAVRVLEDG-GRTSPSNFHPKLIEIFRE RVGGYPGLKDGEKISLYGKIIAVACSYEAATAPRSYKEMKMABEGIVDUVNAN-NTQIDAVISRALLFALSF RWDGTGYPFKLEGEMIFLSARITSIADYDALRSERPYKKALSHEEAVRVLEDLAREF-GQFDPDIKALLFFK REMGLGYPRCLKDEKISIIGUNIAANDYDALSERPYKKALSHEEAVRVLEDLAREF-GRENPRIVHVLVSQLTQ KWNGTGYPGKLGDEHLARIIAAVADVHAMISERYKKKSFMVLEDLAREF-GRENPRIVHVLVSQLTQ KWNGTGYPGLKDEWGLYSRIIAIAADVDAATSDRVYKKKSPFNALEDIAAREF.GROPDPVLAAFIEYFENLEE KLDGTGYPGCKADEWGLYSRIXAADAFDAMTSDRVYKKKSPRAKETIKNS-GTOFDPDVLAAFIEYFENLEE KLDGTGYPGCLKGEDIFLSRITIIAAVADVAAMSDVDAMTSDRVYKKKSPRAKETIKNS-GTOFDPDVLAAFIEYFENLEE KLDGTGYPGCLAGENISLSRIVAVADAYDAMTSDRVYKKKSPRAKETIKNS-GTOFDPDVLAAFIEYFENLEE KLDGTGYPGLKAGELSSILSRIVAVADAYDAMTSDRVYKKKSPRAKETIKNS-GTOFDPDVLAAFIEYFENLEK SYDGTGYPGLKAGELSSILSRIVAVADAYDAMTSDRVYKRALAPNEALEYMANS-GSKFDVKVOAFSKSVIP KNGGGFPECLAGERISLGOFGIISICDVFTAVSANRSDRVFRKALSHVANNILNNK	182 299 322 356 337 349 6419 148 550 557 319 355 355 355 355 355 368 306 2222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981692 4981692 4981695 4981719 4982032 4982253 4982253 4982276 3323026 3323189 3323224 5672654 2126438 3341548
Aq_2027_Aq BB0374_Bb s1l1624_Sy s1r2100_Sy TM0037_Tma TM0186_Tma TM0739_Tma TM1145_Tma TM1147_Tma TM1147_Tma TM1467_Tma TM169_Tma TM1699_Tma TM1699_Tma TP0764_Tpa TP0912_Tpa Orf14_Bhal Orf2_Wsuc Z7f_Vcho Cl.acetobut. Cl.acetob	ESS SITA AVV LA NWW LG DV DV DV DV DV DV DV DV DV DV DV DV DV	LEVILYHQE CLTLLTHKE GEIMRYHHE KEIAYSHQE KEIAYSHQE AEIALTHHE VPIVRHHE AEIALTHHE VPIVRHHE ALWVVQHHE ALWVVQHHE ALWVVQHHE ALWVQHHE AIWIRHHE SKWVREHHE CVGALEHRE KEIALSHE CVGALEHRE KEIALSHE LSAVLQHE AAIAHEHHE AAIAHEHHE AXIALEHHE KNVILQHHE ILGVFNHHE RIISLQHHE ILGVFNHHE RIISLQHHE ILGVFNHHE SYWASSHHE GLGVLMHE RQIAYSHE VDVCLHHE GRIVRHHE	NYDGTGY PERKKGEETPLLARIYTIADSFEAMTTRRIYKRAKSWNEALKELELA-GKQEDPDIYVYAVRALSR NIDGTGYPKGITSENISIESNIIGAASAYSAIILDKAYKKSFNSGASILEIKDA-DKKPDKRVLKLIINAISS RWDGSGYDALRCKEIPLLAQVFQIIDIFTALTSNRYKKEAVNTAQALEILQGEARRGWRNPEMVEKFVNFINK KWDGTGYPQLSGDDIFIAARLMAIADVYDALISSRVYKEAMPHEEAVAIIEAGY-GQHFDPDVVDAFMAIHPT RLOQ'GYPFRLKGEEMTLEDKIIQVADYFSALLEERPYRPANTVEKALSMVEEVKKKKSQEAFELLESSEED RWDGSGYPRGLKEEEIPLEGRILKIADVFSALLEERPYRPANTVEKALSMVEEVKKKKSQEAFELLESSEE KWDGTGYPDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKEDPKVVEAFIKLEKE NWOGTGYPDGLKGEEIPLEARIISVCDYYDVIT-6-WEGRTKTHEEALEIIKKEA-GKKEDPLIADLTVDLLRE KYNGKGYPHGLKGEEIPPERAVLCVADSYGAMRSDRYKRKGSVEESIQEIRRNA-GRYSPFVVEAFIKLEKE SYNGKGYPHGLKGEEIPPEARVLCVADSYGAMRSDRYKRGSVEESIQEIRRNA-GRQFDPVIVEFIEFLEG RWDGKGYPDGLKGEEIPLEARIISVADVYEALTSDRPYRKAKSPEDAAIRKKMF-LDQEIVGILEKHSEFYS SYDGTGYPFGLKDDEIPLARIIAVADYFDALTSDRPYRKAKSPEDAAAIKKMP-LDQEIVGILEKHSEFYS SYDGTGYPFGLKDDEIPLARIIAVADYFDALTSDRPYRKAMSPEDAAAIKKMP-LDQEIVGILEKHSEFYS RWDGKGYPGLKGEEIPLESRIIAVADAFDAMTSDRPYRKAMSPEDAAAIKKMP-LDQEIVGILEKHSEFYS SYDGTGYPFGLKDDEIPLARIIAVADYFDALTSDRPYRKAMSPEDAAAIKKMP-LDQEIVGILEKHSEFYS RWDGKGYPGLKGEEIPLESRIIAVADAFDAMTSDRPYRKAMSPEDAAAIKKMP-LDQEIVGILEKHSEFYS RWDGGYPGLKGEEIPLESRIIAVADAFDAMTSDRYYRKAMSPEDAAIKKMP-LDQEIVGILEKHSEFYS RWDGGYPGLKGEEIPLESRIIAVADAFDAMTSDRYYRKAMSPEDAAIKKMP-LDQS'GOSDPILARVFGKIHK NYSGCGYPGLKGEIPLESRIIAVADAFDAMTSDRYKKKSDINVVILEDIAKER-FGKFNPRIVHVLVSQLTQ KWNGTGYPCSLSGNKITGARILAVGNIKAMISERKYRKKSPYMVLEDLAKER-FGKFNPRIVHVLVSQLTQ KWNGTGYPCSLSGNKITGARILAVADVFDALGTKRVYKDAWSLKILLHFREEK-GRHDPRLIEVFFENLEE KLDGTGYP-GVNGDEISIMTRIITADIFDALGSAERYRAAMPIDKALAIMEENL-HTAIDPECFAALKKALNE RINGTGYPCGLKGEDIFLISSIISICDYFTAVSAMSSKININIKISSIGIITAS-GCGPDPVLAFISLIEVFFENLEE RINGTGYPCGKKGEDIFLISSIISICDYFTAVSAMSDRVYKKKSSPFRAFEMFKTEG-ISIFDTDILRVFINNINV RYDGSGYPRGLKGNEIFLISKIVANDYDAMTSDRVYKKKSSPFRAFEMFKTEG-ISIFDTDILRVFTNNINV RYDGSGYPGCLKGEDIFLISSIISICDYFTAVSAMSYRLFFRYNEALEYIKANSOGIYOLYDAAFSKSVIP RYDGKGYPGLAGENFLISSIISICDYFTAVSAMSSKILFFYNAAYLADSGSGMFDPDFVDEFRNTFFI KINGTGYPFGLAGENFLGERINGVCDIYDAITSDRYKKKSSDINVVISELEENK-GRYPDPFVDERNTFFI	182 299 322 356 384 335 275 337 349 608 557 319 355 306 222 136 	2984261 2688290 1652854 1653077 4980522 4980682 4981264 4981695 4981719 4982032 4982253 4982253 4982253 4982253 4982276 3323066 3323189 3323224 5672654 2126438 3341548

Figure 1. Multiple alignment of HD–GYP domains. The proteins are listed under their names in complete genomes (left column) and their unique gene identification (gi) numbers in the GenBank protein database (right column); the numbers indicate positions of the first and the last residues in each protein, where available, and the distances between the aligned segments. Species name abbreviations are as follows: Aq, *Aquifex aeolicus*; Bb, *Borrelia burgdorferi*; Bhal, *Bacillus halodurans*; Cl.acetobut., *Clostridium acetobutylicum*; Sy, *Synechocystis* sp., Tma, *Thermotoga maritima*; Tpa, *Trepomera pallidum*; Vcho, *Vibrio cholerae*; Wsuc, *Wolinella succinogenes*. Reverse shading indicates most conserved amino acid residues that are probably involved in metal and/or substrate binding. Grey shading indicates conserved uncharged amino acid residues, other conserved residues are in bold. The secondary structure of the HD domain is as predicted by PHDsec program (Rost and Sander, 1993); H indicates predicted α -helical segments, dash indicates a loop or the absence of confident prediction.



Figure 2. Association of HD-GYP domain with other signaling domains. CheY domain (Pao and Saier, 1995; Volz and Matsumura, 1991) and the periplasmic ligand-binding domain of TM1170 (Vyas *et al.*, 1988) are well characterized; GGDEF domain (Hecht and Newton, 1995; Tal *et al.*, 1998) and the MCP-like extracellular ligand-binding domain of TM1682 (Hanlon and Ordal, 1994) are less studied. The transmembrane portion of TM1467 does not show significant similarity to any characterized membrane protein.

Such a function is compatible with the high sequence conservation of this domain as well as its unusual expansion in certain genomes. Indeed, cyclic diguanylate stimulates cellulose synthesis in Acetobacter xylinum in response to the lack of oxygen (Ross et al., 1991). Similarly, multiple HD-GYP domains in T. maritima, C. acetobutylicum, S. putrefaciens and V. cholerae might be involved in signalling the availability of various electron acceptors, including iron and sulfur (Nealson and Saffarini, 1994; Vargas et al., 1998). Remarkably, sugar metabolism in Thermotoga neapolitana has been found to be subject to catabolite repression (Galperin et al., 1997; Vargas and Noll, 1996), although this organism is devoid of the PTS system (Galperin et al., 1996; Nelson et al., 1999) and contains negligible amounts of cAMP (Vargas and Noll, 1996). Thus preferential utilization of certain sugars (e.g., glucose) in Thermotoga should be regulated by an elaborate regulatory system different from those found in, for example, E. coli or B. subtilis. Whatever its exact function, the HD-GYP domain is likely to play a crucial role in this novel regulatory mechanism.

Acknowledgements

Analysis of unfinished genome sequences was made possible by generous submission to the public databases of preliminary sequence data by Genome Therapeutics (*C. acetobutylicum*), the Sanger Centre (*B. pertussis*), The Institute for Genome Research (*S. putrefaciens*, *V. cholerae*) and the *Pseudomonas* sequencing project (*P. aeruginosa*).

References

- Aravind, L., and E. V. Koonin. 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem. Sci. 23: 469-472.
- Aravind, L., and Ponting, C.P. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol. Lett. 176: 111-116.
- Galperin, M. Y., K. M. Noll, and A. H. Romano. 1996. The glucose transport system of the hyperthermophilic anaerobic bacterium *Thermotoga neapolitana*. Appl. Environ. Microbiol. 62: 2915-2918.
- Galperin, M. Y., K. M. Noll, and A. H. Romano. 1997. Coregulation of betagalactoside uptake and hydrolysis by the hyperthermophilic bacterium *Thermotoga neapolitana*. Appl. Environ. Microbiol. 63: 969-972.
- Hanlon, D. W., and G. W. Ordal. 1994. Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*.
 J. Biol. Chem. 269: 14038-14046.
- Hecht, G. B., and A. Newton. 1995. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter*

crescentus. J. Bacteriol. 177: 6223-6229.

- Koonin, E. V., R. L. Tatusov, and M. Y. Galperin. 1998. Beyond the complete genomes: from sequences to structure and function. Curr. Opin. Struct. Biol. 8: 355-363.
- Kreis-Kleinschmidt, V., F. Fahrenholz, E. Kojro, and A. Kroger. 1995. Periplasmic sulphide dehydrogenase (Sud) from *Wolinella succinogenes*: isolation, nucleotide sequence of the sud gene and its expression in *Escherichia coli*. Eur. J. Biochem. 227: 137-142.
- Nealson, K. H., and D. Saffarini. 1994. Iron and manganese in anaerobic respiration: environmental significance, physiology, and regulation. Annu. Rev. Microbiol. 48: 311-343.
- Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, and C. M. Fraser. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature. 399: 323-329.
- Pao, G. M., and M. H. Saier, Jr. 1995. Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. J. Mol. Evol. 40: 136-154.
- Ponting, C. P., J. Schultz, F. Milpetz, and P. Bork. 1999. SMART: identification and annotation of domains from signalling and extracellular protein sequences. Nucleic Acids Res. 27: 229-232.
- Ross, P., Mayer, R., and Benziman, M. 1991. Cellulose biosynthesis and function in bacteria. Microbiol. Rev. 55: 35-58.
- Rost, B., and C. Sander. 1993. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 232: 584-599.
- Takami, H., N. Masui, K. Nakasone, and K. Horikoshi. 1999. Replication origin region of the chromosome of alkaliphilic *Bacillus halodurans* C-125. Biosci. Biotechnol. Biochem. 63: 1134-1137.
- Tal, R., H. C. Wong, R. Calhoon, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana, and M. Benziman. 1998. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. J. Bacteriol. 180: 4416-4425.
- Tam, R., and M. H. Saier, Jr. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol. Rev. 57: 320-346.
- Vargas, M., K. Kashefi, E. L. Blunt-Harris, and D. R. Lovley. 1998. Microbiological evidence for Fe(III) reduction on early Earth. Nature. 395: 65-67.
- Vargas, M., and K. M. Noll. 1996. Catabolite repression in the hyperthermophilic bacterium *Thermotoga neapolitana* is independent of cAMP. Microbiology. 142: 139-144.
- Volz, K., and P. Matsumura. 1991. Crystal structure of Escherichia coli CheY refined at 1.7-A resolution. J. Biol. Chem. 266: 15511-15519.
- Vyas, N. K., M. N. Vyas, and F. A. Quiocho. 1988. Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. Science. 242: 1290-1295.