Supplementary Material Brauchle et al.

Overview and legends

Supplementary Table S1 ……Pages 1- 5
Overview of family membership of all described homeodomain proteins. Also compare to supplementary fig. S4 for the actual alignments.

Supplementary Table S2 ……Page 6
Overview of class membership of all identified homeodomain proteins in 8 more Xenacoelomorpha spp. based on the Cannon et al., 2016 transcriptomes.

Supplementary Figure S3 ……Page 7
1 - ORF sequences of all identified homeodomain proteins of *I. pulchra*……7-14
and identified PFAM domains therein……14-16
2 - ORF sequences of all identified homeodomain proteins of *S. roscoffensis*……16-26
and identified PFAM domains therein……26-28
3 - ORF sequences of all identified homeodomain proteins of *H. miamia*……28-35
and identified PFAM domains therein ……35-37
4 - ORF sequences of all identified homeodomain proteins of *N. westbladi*……37-45
and identified PFAM domains therein……45-47
5 - ORF sequences of all identified homeodomain proteins of *X. bocki*……47-54
and identified PFAM domains therein……54-55
6 - ORF sequences of all identified homeodomain proteins of *Ascoparia sp.*……55-57
7 - ORF sequences of all identified homeodomain proteins of *C. submaculatum*……57-62
8 - ORF sequences of all identified homeodomain proteins of *C. macropyga*……62-72
9 - ORF sequences of all identified homeodomain proteins of *D. gymnopharyngeus*……72-78
10 - ORF sequences of all identified homeodomain proteins of *D. longitubus*……78-84
11 - ORF sequences of all identified homeodomain proteins of *E. macrobursalium*……84-85
12 - ORF sequences of all identified homeodomain proteins of *M. stichopi*……85-92
13 - ORF sequences of all identified homeodomain proteins of *Sterreria sp*……92

Supplementary Figure S4……Page 93
Sequence alignment of all homeodomains relative to described family members of other species.

Supplementary Figure S5 ……Page 122
A phylogenetic tree indicates the family relationships of acoel ANTP genes (colors as in Fig. 1) relative to previously classified human ANTP sequences. The 3 *B. floridae* genes for *Bari*, *Rough* and *Abox* are included since these families are lacking in humans. This tree is rooted and Tale genes are used as an outgroup. All Hs genes are in black and Bf genes in dark red. Orange nodes represent a bootstrap value > 70%.

Supplementary Figure S6 ……Page 123
A phylogenetic tree indicates the family relationships of acoel non-ANTP genes (colors as in Fig. 1) relative to previously classified human non-ANTP sequences. *B. floridae Repo* is included so as to include a member of this family. For the zinc-finger proteins all homeodomains are listed, as well as for DUX proteins. Homeodomain sequences of proteins that are exactly the same are represented only by a single sequence. This tree is rooted and HoxA genes are used as an outgroup. Orange nodes represent a bootstrap value > 70%.

Supplementary Figure S7 ……Page 124
A phylogenetic tree indicating the family relationships of 5 different Xenacoelomorpha ANTP gene complements relative to previously classified ANTP homeobox sequences of 5 other species including *B. floridae, C. gigas, C. elegans, D. melanogaster* and *N. vectensis* (colors as in Fig.3). This tree is rooted and Tale genes are used as an outgroup. Orange nodes represent a bootstrap value > 70%.

Supplementary Figure S8 ……Page 125
A phylogenetic tree indicating the family relationships of 5 different Xenacoelomorpha non-ANTP gene complements relative to previously classified ANTP homeobox sequences of 5 other species including *B. floridae, C. gigas, C. elegans, D. melanogaster* and *N. vectensis* (colors as in Fig. 3). For the zinc-finger proteins all homeodomains are listed, as well as for DUX proteins. Homeodomain sequences of proteins that are exactly the same are represented only by a single sequence. This tree is rooted and HoxA genes are used as an outgroup. Orange nodes represent a bootstrap value > 70%.

Supplementary Figure S9 ……Page 126
Sequence alignment of human HNF1a and the identified Aq homolog. Domains are indicated in boxes.

Supplementary Figure S10 ……Page 127
Sequence alignment of human CERS and Aq Ip and Nv homologs. Domains are indicated in boxes.

Supplementary Table S11 ……Page 128-163
List of homeodomain sequences for the generation of Supplementary Figure S7 and S8

Supplementary Table S1:
Number of Xenacoelomorpha homeodomain genes by class and family in comparison to other selected species

	Hs	Bf	Dm	Ce	Cg	Ip	Sr	Hm	Nw	Xb	Nv
ANTP class											
HoxL subclass											
Cdx	3	1	1	1	1	1	1	1	1	1	1
Evx	2	2	1	1	2	1	1	1	2	1	1
Gbx	2	1	1	0	1	0	0	0	1	1	1
Gsx	2	1	1	0	1	1	1	1	0	1	1
Hox1 / antHox	3	1	1	1	1	1	1[a]	1	1	1	2
Hox2	2	1	1	0	1	0	0	0	0	0	3
Hox3	3	1	3	0	1	0	0	0	0	0	0
Hox4 / CentHox	4	1	1	0	1						0
Hox5 / CentHox	3	1	1	1	1	1[b]	1[b]	1[b]	0	3[b]	0
Hox6-8	8	3	4	1	3						0
Hox9-13/PostHox	16	7	1	3	2	1	1	1	2	1	0
Meox	2	1	1	0	1	1	0	0	1	1	4
Mnx/Exex	1	2	1	1	2	1	1	1	1	0	1
Pdx/Ipf	1	1	0	0	1	0	0	0	0	0	0
Rough	0	1	1	1	1	1	1	2	1	0	1
Total	52	25	19	10	20	9	8	9	10	10	15
NKL subclass											
Abox	0	1	1	1	0	1	1	1	0	0	0
Ankx	0	1	0	0	0	0	0	0	0	0	0
Barhl	2	1	2	2	3	1	1	1	0	1	0
Bari	0	1	1	0	0	1	1[c]	1	0	0	0
Barx	2	1	0	0	1	2	1	1	1	1	1
Bsx	1	1	1	1	1	1[d]	1	1	0	0	0
Dbx	2	1	1	0	1	0	0	0	2	0	0
Dlx	6	1	1	1	1	1	1	2	3	1	1
Emx	2	3	2	1	3	1	1	0	2	1	2
En	2	1	2	1	2	0	0	0	0	0	0
Hhex	1	1	1	1	1	0	0	0	0	1	1
Hlx	1	1	1	0	1	0	0	0	0	0	7
Hx	0	1	0	0	0	0	0	0	0	0	0
Lbx	2	1	2	0	1	0	0	0	0	0	1
Lcx	0	1	0	0	0	0	0	0	0	0	0
Msx	2	1	1	1	1	0	0	0	0	1	1
Msxlx	0	1	1	0	1	1	1	1	1	0	2
Nanog	1	0	0	0	0	0	0	0	0	0	0

Nedx CG13424	0	2	1	0	0	0	0	0	0	0	2
Nk1	2	2	1	1	1	1	1	0	0	1	1
Nk2.1	2	1	1	2	1	3	2	3	1	1	5
Nk2.2	2	1	1	1	1	1	1	1	1	1	
Nk3	2	1	1	0	1	1	1	1	0	0	1
Nk4	3	1	1	1	1	0	0	1	1	1	1
Nk5/Hmx	3	1	1	1	2	1	2	2	1	1	1
Nk6	3	1	1	1	1	1	1	1	1	1	1
Nk7	0	1	1	1	1	0	0	0	0	1	1
Noto/Emxlx	1	1	1	0	1	0	0	0	0	0	2
Tlx	3	1	1	0	2	1	1	1	1	1	2
Vax	2	1	0	1	1	2	2	1	2	0	2
Ventx	1	2	0	0	0	0	0	0	0	0	0
Unknown family				4	1	4	2	2	4	3	24
Total	48	35	28	22	31	23	20	21	21	17	59
PRD											
Alx	3	1	0	0	0	1	1	1	2	1	1
AprdA-E	0	5	0	0	0	0	0	0	0	0	0
Argfx	1	0	0	0	0	0	0	0	0	0	0
Arx/All	1	1	2	1	1	0	0	0	0	0	1
CG11294	0	0	1	0	0	0	0	0	0	0	0
Dmbx	1	1	0	1	1	0	0	0	0	0	6
Dprx	1	0	0	0	0	0	0	0	0	0	0
Drgx	1	1	1	0	1	0	0	0	0	1	0
Dux	24	0	0	0	0	0	0	0	0	0	3
Esx	1	0	0	0	0	0	0	0	0	0	0
Gsc	2	1	1	1	3	1	1	1	1	1	1
Hbn	0	0	1	0	1	0	0	0	0	1	1
Hesx/Anf	1	0	0	0	0	0	0	0	0	0	0
Hopx	1	1	0	0	1	0	0	0	0	0	0
Isx	1	1	0	0	0	0	0	0	0	0	0
Leutx	1	0	0	0	0	0	0	0	0	0	0
Mix	1	0	0	0	0	0	0	0	0	0	0
Nobox	1	0	0	0	0	0	0	0	0	0	0
Otp	1	1	1	0	1	1	1	1	1	1	1
Otx	3	1	1	3	1	1	1	2	1	1	3
Pax3/7	2	1	3	1	1	0	0	0	0	0	2
Pax4/6	2	1	4	2	2	2	2	2	2	2	2
Phox/Arix	2	1	1	1	0	0	0	0	0	0	0
Pitx/Ptx	3	1	1	1	1	2	2	2	1	1	1
Prop	1	1	1	1	1	1	1	1	0	1	0
Prrx	2	1	1	0	0	0	0	0	0	0	0

Repo	0	1	1	0	1	1	1	1	1	1	1
Rax/Rx	2	1	1	1	1	1	1	1	1	1	1
Rhox	3	0	0	0	0	0	0	0	0	0	0
Sebox	1	0	0	0	0	0	0	0	0	0	0
Shox	2	1	1	0	1	0	1	1	0	1	0
Tprx	1	0	0	0	0	0	0	0	0	0	0
Uncx	1	3	2	1	1	1	1	1	0	1	1
Vsx/Ceh10	2	1	2	2	2	2	2	2	1	1	1
Unknown family				1	9	2	1	3	2	2	7
Total	69	27	26	17	30	16	16	19	13	17	33
LIM class											
Ap/Lhx2/9	2	2	1	1	1	2	2	2	2	1	0
Islet	2	1	1	1	1	1	1	2	2	1	1
Lhx1/5	2	1	1	2	1	1	1	1	1	1	1
Lhx3/4	2	1	1	1	1	1	1	1	1	1	1
Lhx6/8	2	1	1	1	2	1	1	1	1	1	1
Lmx	2	1	2	1	2	1	1	1	6	1	0
Unknown family						1	1	1			
Total	12	7	7	7	8	8	8	9	13	6	4
POU class											
Hdx	1	1	0	0	0	0	0	0	0	0	0
Pou1	1	1	0	0	0	0	0	0	0	0	1
Pou2	3	1	2	1	1	2	1	1	0	1	0
Pou3	4	2	1	1	1	1	1	1	1	1	2
Pou4	3	1	1	1	1	1	1	1	1	0	1
Pou5	2	0	0	0	0	0	0	0	0	0	0
Pou6	2	1	1	0	1	0	0	1	1	1	1
Unknown family									1		
Total	16	7	5	3	4	4	3	4	4	3	5
HNF											
AhnF	0	1	0	0	0	0	0	0	0	0	0
Hmbox	1	2	0	1	0	1	1	1	1	1	0
Hnf	2	1	0	0	0	0	0	0	0	0	1
Total	3	4	0	1	0	1	1	1	1	1	1
SINE class											
Six1/2	2	1	1	2	1	1	1	1	0	1	1
Six3/6	2	1	1	1	1	1	1	1	2	2	1
Six4/5	2	1	1	1	1	1	1	1	0	1	2
Unknown family											1

Total	6	3	3	4	3	3	3	3	2	4	5
TALE											
Atale	0	1	0	0	0	0	0	0	0	0	0
Irx	6	3	3	1	4	4	4	3	5	2	1
Meis	3	1	1	1	1	1	1	2	0	2	1
Mkx	1	1	1	0	1	0	0	0	0	0	0
Pbx	4	1	1	3	1	1	1	1	3	1	1
Pknox	2	1	0	0	1	1	1	1	1	1	2
Tgif	4	1	2	0	1	1	1	1	0	1	1
Unknown family					14	1	1				
Total	20	9	8	5	23	9	9	8	9	7	6
CUT class											
Acut	0	1	0	0	0	0	0	0	0	0	0
Compass	0	1	1	1	2	0	0	0	0	0	0
Cux/Cutl	2	1	1	1	1	0	0	0	0	0	0
Onecut	3	1	1	6	1	2	2	1	1	1	1
Satb	2	0	0	0	0	0	0	0	0	0	0
Unknown family					2				2	2	
Total	7	4	3	8	6	2	2	1	3	3	1
ZF class											
Adnp	2	0	0	0	0	0	0	0	0	0	0
Azfh	0	1	0	0	0	0	0	0	0	0	0
Tshx	3	1	0	0	0	0	0	0	0	0	0
Zeb	2	1	1	1	1	0	0	0	0	0	0
Zfhx[d]	3	1	1	1	1	1[e]	1[e]	1[e]	1[e]	1[e]	0
Zhx/Homez	4	1	0	0	1	0	0	0	0	0	0
Unknown family					1	2	2	2			
Total	14	5	2	2	4	3	3	3	1	1	0
CERS class											
Lass	5	1	1	0	1	1	1	1	4	2	(1)
Total	5	1	1	0	1	1	1	1	4	2	1
PROS class											
Pros	2	1	1	1	1	1	1	1	2	1	0
Total	2	1	1	1	1	1	1	1	2	1	0
unknown class	6	3	0	23	3	3	3	1	3	2	4
	260	131	103	103	134	83	78	81	86	74	134

[a] This gene has not been identified in our dataset but was previously identified.

[b] Central Hox genes have uncertain family membership.

[c] This gene has a very different N-terminus. Possibly an assembling issue.

[d] Current sequence runs into a putative intron.

[e] Each Xenacoelomorpha protein has 4 HD's; the number of HD's can vary in other species.

Hs: *Homo sapiens*
Bf: *Branchiostoma floridae*
Dm: *Drosophila melanogaster*
Ce: *Caenorhabditis elegans*
Cg: *Crassostrea gigas*
Ip: *Isodiametra pulchra*
Sr: *Symsagittifera roscoffensis*
Hm: *Hofstenia miamia*
Nw: *Nemertoderma westbladi*
Xb: *Xenoturbella bocki*
Nv: *Nematostella vectensis*

also see supplementary fig. S4 for the actual alignments

Supplementary Table S2

Overview of class membership of all identified homeodomain proteins in 8 more Xenacoelomorpha spp. based on the Cannon et al., 2016 transcriptomes.

Species	#	ANTP	CUT	POU	HNF	LIM	PRD	ZF	CERS	TALE	PROS	SINE	Amb
Ascoparia sp	14	6					2			3			3
Childia submaculatum	52	16	2	3	1	5	10	1		6	1	4	3
Convolutriloba macropyga	76	31	3	4	1	8	16	1		6	1	3	2
Diopisthoporus gymnopharyngeus	49	17	2	1		6	9	2		6	1	2	3
Diopisthoporus longitubus	58	20	2	3		10	9	2	1	6		2	3
Eumecynostomum macrobursalium	11	2		1		2	2			2		1	1
Meara stichopi	67	18	1	4	1	8	15	1	3	8	1	3	4
Sterreria sp	1	1											

#: number of genes based on unique sequences
Amb: number of genes classified in more than 1 class (ambiguous)

Supplementary Figure S3

1 - *Isodiametra pulchra*

>IP_DN10641_ANTP_AMB
EEFKKDKYLSVSKRLELSVALSLTETQIKTWFQNRRTKWKKQVASRMKIAYRSFWTVPNSAGGHQPQFPLPNPFYALGNPLNPL
NPLYLQSVSGGGGFLGAAHSPLDDDDRELDQSDCFSPDDSDDPGNESPLRS
>IP_DN12279_ANTP_Evx
MQSSAFSISSIIDLPRPTQPGTGHVVENETPGESRSPSPDFPAGSQTAHQQQSTGSPPPRCLSPMSAAGMRRYRTAFSKHQQD
SLEMEFRKENYVSRPRRCELAASLNLPETTIKVWFQNRRMKEKRQRQAFTWPPYFDPYYYGYLISRGLIPAPVAQPSIPQMP
>IP_DN12334_PRD_Pax4/6
AVAAAAHCAAAVTDEQQQKKLEEAAGGGGCTGASEEARMQLKRKLQRNRTSFTPEQIDALEKQFEVTHYPDVFARERLAGKI
ALPEARIQVWFSNRRAKWRREEKLRTQKQQQQQQQQQQQQQQ
>IP_DN15610_ANTP_Nk5/Hmx
MQKPKFVSPFSISHILGHSPAAASEPAPAKQVEEQEATKHVIQSKEQLDLLLGIATNPVDSVEASTNRPNEAASPQAQQPKAKQ
NTEVSPSENISPCSVECDTSSDDEDGGDDGDKERRKKTRTVFSRNQVCQLESMFDMKRYLSSSERTGIARDLGLSEQQIKIWFQ
NRRNKLKRQIKTGLEMAGNISAGTVPPPNVNAALLGSFSSQQPSHHSQMLPFPSPTSMYHQLPGSTPNLSLKVDSNPASNFPH
LNFPAPNRTMAPSASQAATVSHTQGTPFSARPSQQNQPADLFSFHNFSMMRSLSGLV
>IP_DN18113_POU_AMB
FSSRRRHTRYIGDWSSDVCSSDLVILLLNRCLMASAGESLDVQKVVLELISEEDGVFRTIEQIKQKLAASDYEARIMIIYAQQKIILD
LKSKIADLVRMKDQTTQSLLDFHDLNRKLEIPVRELVTLEANLQNAALSSPQAINVNPLVVLSNQNQSPNSTGYPAIEPSTMQQ
LLVLAPSTTTPVAKGPIRISNPQVTVACPVGFSPESPDSEKTRSGGVDSTEVNVVTKKFAPQPPAVDDSSILEELEAFAKDFRTRRI
KLGFTQADVGLSLGKMHGSDFSQTTISRFEALNLSFKNMINLKPFMEKWIDDAERHISMVKNDLTQLSDEDQRKLGLFINAGD
PLTNRKRKRRCVIDAHTKKTLEQSFRDKPHPSPDETNALAEALGLDRETVRIWFTNRRYRERRSETPNSTS
>IP_DN19265_PRD_Rax
METERVSSEKSSSQVMEESSSPSSPRQQIYSINSILGIQSLLPSQINTSGCDTTSSTGGQSQGAEQSGVGGADFSDEDKEDEECK
KHRRNRTTFTTFQLHELERAFEKSHYPDVYSREELAARVNLPEVRVQVWFQNRRAKWRRQEKLEQSQLKLNTDFPLATLRASA
AAAKFSALDPWGLGAAASAAFPFIPPTAFFHPYSGGQMTAAAAALSNPFFFGNPSQQSV
>IP_DN26203_ANTP_AMB
MTMQNPLSGYFPHHPSYQQWYQEMSAAAEQQHQMAALQAMYFPQFTQSAPISNDDIVGPLQRSGFIGSCNPWNQSRLKF
VQAPSPPVHRLAPTPGDGEDSDKENIDPRGGSSPPIIRPFSSSPGHLASDAEQTGRSKRQGRTVFTKQQLEILNGHFNHRQYLSL
AERAQLSEQLCLSQKQVKIWFQNMRSKAKRVMNASSRTQPQDLAQKQSQEAFECAGNQVTSALDENHNVSRNESTGESSYF
YQPPTINSFLCPNPDSPECNGANGCHRTTDSDSPSNLVIDC
>IP_DN27657_LIM_Lhx6/8
QKCLRCAACSIPLGNHLTCYVAKDNLIYCKNDYIRYFGIKCAHCHRSVSASDWVRRAKQLVYHLACFACNVCHRQLSTGEEFAL
HESNVLCKQHYLDLLEVPGKATTGSSSSSDKMADLVSNSGGDSSDGNSELDLSMGSPHKSKRSRTTFSEDQLKILQANFNLDSN
PDGQDLERIAQQTGLSKRVIQVWFQNARSRQKKHMTGFSSSNAGKLSMISNGGNSFPGCEDMSPGSASLL
>IP_DN29418_PRD+POU+ANTP_AMB
MQKRREILPVQKKEHLIKVFNVNPYPYPPQRVELARELKLRQETVKVWFRNRRYRLKHTKPKVSKEPKNKKGKQEEEDVNEEEV
LRDQSSPIRREKGEILRDSGRDSRQFSLSPIQSVSFLERERVCRLTPPPTPVTS
>IP_DN31212_CUT_Onecut
APSAFDFNNYNAWSTAATLDSMYNGAAAPMIPPYQNPAVNRSLATAGLAACSAGTRFNPTYPTLNGFYYPTAYTNPYARYPR
GFLGREGEELDTREIAQRVTTELKRYSIPQAVFAEKVLCRSQGTLSDLLRNPKPWSKLKSGRETFRRMGKWLEENEFQRMATLRI
AACKKKEEEQMVVKPKVMKKSRLVFTDLQRRTLRAIFQESQRPSKESQAQIAKQLGLEVTTVSNFFMNARRRCGDRWKPEDK
L
>IP_DN31321_PRD_Vsx
GGGGGGGHSLSHGFSSCLPTAEGFPSHAQASLSTKKKKKRRRHRTIFTSYQIEELEKAFKEAHYPDVYARELLSIKTDLPEDRIQV
WFQNRRAKWRKKEKTWGRAS
>IP_DN36044_ANTP_Barhl
MGDHSSSSCGYSSASEVDSETEQHGKQPSNSFFIDNILSSPPSRDDGPRKKKMRKARTAFTDEQLVKLEQNFERHKYLSVQDRL
DLAAALDLTDTQVKTWYQNRRTKWKRQAAVSLELIA
>IP_DN36933_CUT_Onecut
METFPESIHPAFVSNSANFQLSPRDMFNIAPPAAAFTNPAHRPSDPFFDFASSTLSHDYRPADLYSNPVTQPPPTSSDLASSAAA
RSPGSNYSSSFSPIKNLPPVATFDQPGGGDSFGALFPPANYFYPNQYSRLDFKSDLKSPPNVYPSPYSMDKLPYYHTQSCFYPQS
YDSFESAFAYAAAGTATRQPAATAAVPNIYSSPKLKEECIFTSLATSTQPPSKPRNRNSSRQSGEKEMTMKEVGESGGSELDTRD
VAHRITAELKRYSIPQAVFAEKVLGRSQGTLSDLLRNPKPWGKLKSGRETFRKMEKWLEEPEFQRMSSLRVAACKKKELESLNP

GGLPAASAAAPVQNPGKPHQKKNRLVFTDIQRRTLFAIFRETKRPSKEMQLQISRQLGLEPTTVSNFFMNARRRCGDRWKDPS
DIHIQAKSMRKMTGGGTSKTVNGLQPLI
>IP_DN38710_PRD_Alx
MSTDVVFPSVSTSGSNMVVNSCLYKGPSVPNLLYPPSVNPTYPVGVHSAGGGGAAPPAPTPVTPYWSPALMAAASAASGCTL
NSATVPEKLSPAAAAAADHKSPVPSNNSLSCCADVPEKSPLACSDQYSPAAAADSGEEDGISIDGGSSGGRKKRRNRTTFTQAQ
LEEMERIFQKTHYPDVYVREQLAVKCDLTEARVQVWFQNRRAKWRKKEKHQQLRSGFQSANSFAAESQALRAAAAAAAAAT
DPISQLSGSWSYAGQNQGSNPLAPAAASQNGSNNSGGMPGSYQLPAAASTAPSLFPAANYIMAAQQAAAGYSCLPAGPHN
SSPTSTNTYTFSPQQSHFTGVPAAFLNSSPLNHPASLMAAGSCFFSQSGDFFSPAAAAGSAAAAAAAAGFRLQRSKDMPSGQ
AGGATAHSAPFPTHFIAAAGGGGGMSSYK
>IP_DN40909_ANTP_Nk1
MDDSSSGGRRRSRTAFTYEQLVALENKFRASRYLSVCERLNIALALNLTETQVKIWFQNRRTKFKKQNPSASLDVGKSKASSHH
SAHHANNCHHLKQQQMNVAAATGGPCSLTSSLFPANHHQPHTFARTSGSQFLSSTISPFVPFGTFNGDVSSQ
>IP_DN43104_ANTP_Emx
MLRPPVNAFQTPNIPTMLLESHLQQLPLTLAIPGFHPLLYDPFRHHLSAISSPNQPLNPVASILTERIERANQERETAAAALAAAS
FFLQSHHDTSINGGGGGVPSTCKRVRTAFTATQLAMLERHFLQSPYVVGSQRRKLSSDLGLSETQVKVWFQNRRTKAKRCIVP
STTDAITSQELAIRKTPNNLAGEPDEGFEESAN
>IP_DN44102_ANTP_Cdx
MHLDSNSDAIDSMNRASAPNGWIHHLTQNFAYNSTAPTGIHNPLKFESSPPSSIGAGASLKFEHSPPLADHHTLHHHQQQSAL
HNGSLQNLPQQQQQHFTELRTAAENCSPWLSQTQHYLSAAAVGAAPPSAPLSPSSLHQRSSGNAFAQAGHNGWSSRASTA
PGYPSPHFGMSPHSAEQFAAVAAAGAYSRSTGCNLMSTGANGGSSDGSAAAAAAAMHMPAAYFMSAAAAAHHHASTSPG
GCMPGGGGPGNPFSAVAAAVGYHPAEMYACAAAAAHSAPDHTSSTFRYNPNTGKTRTKDKYRVVYTDRQRAELENEFRSA
QYITIRRKSELAMQVGLSERQVKIWFQNRRAKERKVSRKPGSGG
>IP_DN44510_TALE_Irx
MSFANSVSLVGGTEESVGATALQSAAEYGYATQHDLRQPFFDSSPLTHHEYQLPYFPHNAAAGYFPHDFPAGHGPPAHVMM
SHAPPPPPPTFHNPDTQQLQPTFPQPVGSAGRSESFSGLEFKTEFTHHHMPSNAGGHPAMEELLPRSFPYPPPPPAHYPFPAG
AAPPAQYPFDRLHPTGVPTSTSATAGSNSSSIRRKNASRETTAALKAWLYEHRNNPYPTKGEKIMLAIITKMTLTQISTWFANAR
RRLKKENKMTWTPKNKTSSNRTSQLQPSDSAAAAAPDTERPSAEEKSSPISDKHHQIND
>IP_DN45560_SINE_Six1/2
MSDLGSFPFSADQAACLCHLMEVSSEIKKLEKFLWTLPNYENYQQHENVLRSRAFLAFNEQQYKEVYRIIQSKPYSMNNHFALQ
QLWLNSHYAESENSRKKPLGAVGKYRIRRKFPLPRTIWDGEETSYCFKEKSRNMLRDYYGHNPYPSPREKRDLAEATDLSITQVS
NWFKNRRQRDRALEARP
>IP_DN49678_LIM_Lhx3/4
RRRHTRYIGDWSSDVCSSDLSGRFLRSITTAVASVRNGSQSRILNQKTPKIMDLGHLYHLAPHGLFIYGLNVSPGFRQYGAKCSG
CEKGIAPTEVIRRATDHVYHLECFKCLICNKLLETGEEFFLMENQQLVCKNDYEMAKSKEFEELNGSNKRPRTTITAKQLEILKNA
YKSSSKPARHIREQLSKDTGLDMRVVQVWFQNRRAKEKRLKKDAGRRMAAASSQYFSQQPSHLNSPTTPKGKRSKKNPGSPG
FDISRHAPLLSGFGDLDQGQNQALLNAYQNQLQPPMFFRDALGQNLKPYMPISSGLLSPIVDPNMPDCRYGMPPPPPPPQPP
QNFPMDAPTPPANFDSRLNQVQLTPPNNGSGYQPSLPISSQNSFNPAMMFPDQLSSLHSALTGGRPQNYPPIPTPCKFEPKFE
PGSAGNFMPNNSQQQLENLLMW
>IP_DN51842_PRD_Pitx
MDSFVGNLDFSDLASANINRALIPPALPHSTACSPPFSSGGHVAEPPLHLAPPPTLSYPPASGSAGTSPGSNLHLLKPVSNFSPAT
HELQIKTEHSPQTLESGPQPPPQQPSTGGEDSGAEEHPSPLAPTGLDEDEDKKNKRARRQRTHFTSQQLQELETLFSRNRYPD
MATREEISAWTNLPEAKVRIWFKNRRAKWRKKERNQLQEYKNAAAFGMNFMPAPYDPHDFYSSQLTAGYSAYNNAWTKM
AASPLNPAKTPFPWTFNPAPGNPQNPAAAAAAMVPNVSIPASLNPSPPCPSWGANPATQQYMYGRDCTGLSQFRLKHSPQ
SNPIGGYQPFRPPTSIQPCQYSMDRPL
>IP_DN52339_PROS_Prox
GELLIEYFITSAQDLGRVMQDISMMKEEADASPLDQFEAASTSPSPVPSSPEKASSPAVASAESPVKAPATTSAAASSGATSLPST
NPINSFPYMPTSFPFPPNAAAAFPLAASATAAAAAAFYHENSEHFFHDRKMPFPGPPTSSAINYSDLLPPLPKLRKPNEPVRDR
MNRTPPTISDIPNPFHNLPKPPMMGDKVLPWRFPHPLVGGADPLAVLRHQAMMSGAGMSSAGNFPASHGAQENSWNNE
RIRLQHEVMALKNKMLHMSHNLGVILADQVLTSEAAARINTILSSSDLSFLADDIIVELEKKVKNVIYKVIHGFLTANRPSLGEDG
GEEREAAQRPNLKRKMSPMPYNLNSEDVKFGSANSSPVGLSNGFPSSFCRDRVCGPRTKVTEEGSCTGLSLASAVGSTDNLTA
NHLRKAKLMFFFQRYPTSSTLRQYFTDVKFNRHNNAQLIKWFSNFREYYYIQIEKYAREVVSKVSAERQSTSNPVEVSASSQALL
QLDESSELFKNLNLHYNKSTTFKPPESFLTCVRMTVRQFAATILVGGDSDPAWKKHIYKIISKLDTTIPEAFKTHHLVDNFQCN
>IP_DN52339_PROS_Prox
RRTFVSSLERRIVYSVALEVVDEVVGFEGLWDGGVEFGDDFVDVFLPGGVRVAADEDGGGELPHRHPDAREEALGGFEGRAL
VVVEVEVLEQLTALIQLQQRLTGGRHFHGVASRLPLRAHLAHHFPRVLLDLNVVVLSEVAEPFDELGVVVAVELDVGEVLAEGG
GGGVALEEEHQLRLPQMVRRQVIGRPDRRRQREARAAPLLRNLSSRSADTIAAEAAGESVAEADRRRVRAAELHVLRVEVVRH
GAHLPLQVGSLGRLPLLAPVLPETRPVGGEEAVNHLVDHVLHLLLQLDDDVVGEEGEVGGGEDGVDACGRL

>IP_DN52339_PROS_Prox
MFGVLVIEGGSGSGGGTGGQREGSGRVGREGEAGRHVGEGVDGVGARERRRSRGSSGRGCGRGFDWGLGTGHCWRRCLL
RGGRDRTGRGGGSLKLIQRAGVRFLLHHGDILHHSTQVLSTCNKVFN
>IP_DN54186_PRD_Pax4/6
MPQKGKGHSGVNQLGGVYVNGRPLPDATRQQIIELAQKGARPCDISRMLQVSNGCVSKILGRFYETGSIHPKAIGGSKPRVAT
SDVVRKIAALKRESPSVFAWEIRDRLVNDGSCTADNVPSVSSINRVLRSLQADKMKSSPGPQDMSGCAAAATMSSMYDKLRL
FNQHWPSPAWYNGLHPAQANPAPGHFPSAAAAAVLSHLPASSHQLSPSDVIKPEITSPTEPHLHLNPIDPNRGTTIPPVSNGYP
LASPIEDSEKIEENGSPCSNLSPDSADDDNLDDSGKKKNQRNRTAYSQNQLELLEKEFEKSHYPDVYARERLSEATDIPESKIQIW
FSNRRAKHRREEKIKSNNNNEMPLHPSPLELTNQRIDRMENGGQSDPQTPQSLYQSPYFPYTNSGATSLSITNILQAAPPPAVN
QSTPYSHQLTSPNFGARVAAHQVINPMGALPYSSMSQHTNNPYYNPQASYFHQTQMSFPATSPMDGPTPMSNYWPTRFQ
>IP_DN55409_POU_Pou2
MNLSTCADPHQSRETSFTQQPQPIRSLQSTATQCEIITSSSPVTHHPSSECIHQLLKLLDNELIAGSKKAGIDITSVLLHAQQKVIA
QLLTQPLGDGITHCCPKASDSCTPNSDTKLSKDFAGNRESESDGLIFDLSKGVKREPDVQLARANPLELFNLDRQVPLQRDQQIP
FANGRAGIFKQNLKNESWNLNQDESSRASFLSSASPPEQEGWSRAHQEQLEEMHQFAREFKEKRIRLGYTQSDVGHSLGSVF
GSDFSQTTISRFEALNLSFRNMIKLKPTLQLWIENVHRENRRIQPSSDPEQKPQDANPFMVPRKRKKRSCIDSNCRETLEQLFDQ
NNHPTPDEIVKYSEALNMEPETIRIWFSNRRQKQKRLASPQIDENGHISQMASLASLGGQSINLNYRRMPSPRQDSLLDIRSVQ
PVDIEADTARQDSFLVHTSRDG
>IP_DN55547_LIM+ZF_AMB
MESVINLLSRNHHFVDDSQGDGSSPAQGDSDTQIHLNPLETGTPQDDLEDIQKDLSIVEDVQNEIKNLQVGLDGSETKLDLADS
QKSVNDASDDNSSLDLSSFPSVSELPMELNINHPMTSDKLCLHCHQEFPDPVSLYQHKRYHCLHNPEVLFHHSNSESDNPNDP
KKLPERKKRTVYTDQQLDTLKEYYIRQPNPTSDDIERIASVIGLTKRNVQVWFQNKRARDKKNPETEEQKYFSPIQDLSVKIPKSP
RKSQDLESKLMKSLQEVPNVAVKKELTPTDKPESSPIENVVAPNALYSYQWLSLLGFYQAVAAQSILVSQNHLPINLTPDSSTQS
LETCDESRTTCNEPRDQITCNEESREMNESQNESSEQMEPLDLSMNSNPSSSFGRSFCSPSSSISSLTLNTPNSIKEEHMDTDFLR
LHHKSEATSPEGSGPYFCSQCSKVFQKQSSLIRHIYEHSGKRPYQCTECDKAFKHKHHLTEHSRLHSGERPFQCPNCSKRFSHSG
SYSQHLSQKRCL
>IP_DN56009_ANTP_AMB
MAVAQARPAAGPSSSFMMDSILNNATAAAHSPPAPDERRPRRRRERSERSGGGGGDESEWEESESEQDPAAPRPAGGKKK
NRTAFTPGQIFELEKRFRVQKYLSASERGEFASHLNLTDTQIKTWFQNRRMKHKRQQVDAENLQLAALYGPLFQQSAAALLRA
AAVTGPKPPVTPTDTQLMDPRHRLYLPAAAAAHYPLQSFASWHHLLHLGDKMSQQQHLLYHQPPSPTSKRLPTDTPRVEMQ
LLPTRITRLPLSGQ
>IP_DN56056_LIM_Isl
MLLHHLRYSEMDFDSYIPNLGPELPPPPRHFSGQQSSANNLNIKTESRNESPGGLTDPTLPCEPIKGLDSPPMKLPHCSGCAQPI
LDQFILRVAPDLEWHSTCLKCHQCHVQLDERQTCYLKDGRPYCKLDYLSLFSVRCGKCDQPFEKTDRVMRASTQVFHVDCFTC
SACARRLLPGDQYAMRDDTILCKDDMPDLALEDKEQGTNNNNGTSNAALTLNTSSDLDNQDKLLYSPEEKSPGSKDSPRSDSK
RSSKDGKPTRIRTVLNEKQLHTLRTCYNANPRPDALMREQLVEMTGLSPRVIRVWFQNKRCKDKKRSIAIKQMHQMQAEKNI
NGLNGVPMVAGSPMRHEINDSQFHYSYGLPPHSELDTSYMVGISITQQ
>IP_DN57662_PRD_AMB
MTDNMSSETESELCVAQQHVTRNPPPAANPPQQFFDPVTVDVVEKPPTAAKPPPNPPTDPKPTPGSGDEKQVKAKRHRTRF
TPAQLNELERCFNKTHYPDIFMREELAMRIALTESRVQVWFQNRRAKWKKRKKHSPNMFRSAAGMLLGNPGAPPPFAQSGF
PAGPACQDPFYPHFPNWGNPGSFYPAFAPPTSRADFTRNPFQFPTAGGAADSRTLQFGTSDGVMMNHQSPNGGNVTPNP
FALNPSVTQSFAMMCQSPPSEQIPANEPSDPWRNSSLAALRRKALEHTANSMNTFTSLSSCSFDLR
>IP_DN57889_ANTP_Bsx
GSAGTRTAYLSLGEVEGGGDVDALGGGEIALGLEPPLQAVELPIGEDSAGFSPSVHAHAWVVGHHATQERIVPCETTTSPSEYL
SCDCRTDRRTCTRWVRSGGR
>IP_DN58162_PRD_AMB
MHQQHQFVPSSTGQNSLSLASYLALLPQLSAAAAAAAAQQYPYHLLLQHNSSEEQSAAASKLSPSSPKKHTSCSLASPPSPSAE
SGAGEEEGSGRSEEGGGGAGDLKRRRTRTNFTNWQLEQLELAFHRTHYPDVFVRETLALQLNLMESRVQVWFQNRRAKWR
>IP_DN60252_PRD_AMB
MEDFGSHIYGNLQAQSVGNFQSTATSNEITSPLSDSAVGFKQQQQPSPSKPIKSEAGSNLSPNSGSCAPTPARRRHRTTFTQEQ
LNELEAAFAKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQEKQLTKAFVSPPVMPGMSQPACNQMMRNIYQQAA
NVRPGFSPNGMPFQYQPAMLTPNSANNMPRYSMHNSPTGGSMIP
>IP_DN60553_ANTP_amb
MPGAFSIDDILAKPSDSSPQKLTSPQTPLPTLNISTTASSFFEETMRKFQHCAERTKSLGSFNSPSLSRAEAPEAAREESAGGKFQ
TSYEASELSDDFGFSDSEEENDSALNLKTDANAPEPRRMLIKDTNGQVRELLFPRALDLDRPKRERTSFTPEQLFRLEKEFRANQY
MVGRDRGDLAASLGLSETQVKVWFQNRRTKYKREKQRDIEAKKSLDHSMTHCAAGMFKILDQTSPTLHTYYPGRSFPTAPGF
GQSAAGLQLPGSRTAGGLILPQSAHLPMAPPIMAPFYWMNS
>IP_DN60676_ANTP_AMB

MSAKLGYESESDSEVSSVGEFSDADERQPSAVPVKKSGFFISDILQSDGKASARMDDFPTHGNASLSPDWLLALNELNKPGGV
NNNDSCNDFGSMVGKKTRRRRTAFTQAQLNFLERKFRCQKYLSVADRSDVAEMLRLSETQVKTWYQNRRTKWKRQTSSEER
MAELEGHACSEPDCPGNARGNSTRQSELRSIAGQSELRPELR
>IP_DN60751_ANTP_Gsx
MAKATITEPPAKNTFSIDSLLHDPRSERKEKMEEQTEIQGPELENTEDEVITQSVIDCNSVAARSEAGSTGRWDIPPANETAETA
AYHVRSRISSSWDLLGLLNEARSSFTPITHAHLHTIEALNKLGFMLEHSSQQTSSSRRTRTAFSSSQLMCLESEFESSNYLTRIRRIEI
ANKLHLTEKQVITTPSRNSVPHCINPVTC
>IP_DN60800_ANTP_AMB
MKRSFLMRDLLEEKAAAAAVVPETPLSIHIHSHPDEEDRSSDSPDSGVECSSFFEEIREQDEAELAAAAAAKEDPAAAGEEEEDG
RRASKKQRTAFTPDQVYELERRFSQQKYLSASERAEFASKLQLSDTQIKTWFQNRRMKRKRQQQDMENAQLQAFYSTTGMLI
PPPAHPQGFLPAIDCLPQPTILSHPVLMGQPSAAAAAAAAPSAASAFLQYPYSAAAFQPAAAVSSQSLQCFPPIGTAAALSQH
MPPAFSPFLRNMPHAQTHLNHLC
>IP_DN61228_ANTP_Dlx
MQMNPLSTSPTPRDPQRPEHQPHPQDLLGKAYFPELPHTLHYAASIPGFAQAAGHHPSAPGGFPHYMHPGAAAHAQHPFH
NPHHPAMAGAHYAMAAANQESFLNNPYRNSAAAAFPGYLAGYQPPGSPPNSMYSHMQFQAQQQGRPMPIPYGAPPPVS
PNQINDASRVSQGEKELDDESPKNSSGKGSGKKLRKPRTIYSTVQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQN
KRSKMKKIIKHAHNGASPGSLAYSALSPEEQEALKEEAEDTGGLDNQMGHQIHQQQQQQQPHEPWACQQQNPHDLSQQS
AYDPSNGVAMKREMNTPGVPSPASALPQPPSVSNAHCEQTSGGVANSWYVGDHLPPHAHMIPGQNLPH
>IP_DN61481_ZF_Zfhx
MEDQPRKEETPPRAKSPERLIQALYQQIYLQNCIIMQQSLALHQRQTTSGALALENLVKSQATTKYSPTDEPPPESFVLPIETAFK
ASFPSKDNALSKLFVGINRSCRRDPRDPRKCRVHNTFHQKPSRVQIPPTEDSTKEESPLDYSVDIQNRSSLLTFSQIDESNGSRSEI
PTLDIKEDPAVTPSSESANRSFDSASVTSYLDESPDLRRSFDESSSVSRLADGFISIGMYCRPDVNDSARCSVHRIRHRTRTKLPAA
ELALLKKIYETNPSPSKKQFAELARRTSLPIKVLKVWFQNTRRLTRNTCSGCRKTFVTNHLFAEHFDTCPQKSGSGSHDESFET
>IP_DN62049_SINE_Six3/6
MMHADPKLGPLFLQQMGPAAAAMLNPLLASNPHFLHQLNAAHIQAQQQQQQQQQPPPPPAMTFNPEQVAQVCDTLEES
GDFDRLSRFLWSLPPHLLDSCLKNESLLKARATVYFHNGQFRDLYMLLENNRFKKEYHPKLQAMWLEAHYQEAERLRGRPLGP
VDKYRVRKKYPLPRTIWDGEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGNWFKNRRQRDRAAAAKNR
NCSINGFGSEEGSDKMKALEVDIHDDDDDDDDPIGSDFEDDKDLMSPLSGAGSDPESLNSSKADAQLATSKSEDLSPSSTTSSVT
SPAAESIITPPSKMSPREAPAHAHPLMQSHHLHHLHQQMQQNHHRLSIPTSVFLLQAQAQLAAGRFNPQAVPPHSAPGIPPF
FNPVPGMFRSNPLLMASLAPTLAVSSAPSAASPVSSAAAAATAVSTNSVSSTISPVKPSVFSPVALCNIKDSS
>IP_DN62850_ANTP_AMB
MAQFSIERILQPSPAIPQPENHPVESGLKLLATDPWSRLLSSNLSSLPFTNPDLNTAQLIASLQTRQQQGKKVGSPGHHLASRRQ
RTSFTPDQTLQLELQYERSEYITRSKRFELAESLGLSEGQIKIWYQNRRAKDKRIEKAHLETQLRTTRINASIQGPIQGHIQGLERE
MPMAELQRKLLLGAGLVPFLDPSALMALTKEQKLSPHSSHNISSSDAMFVSRSSAFVSSANY
>IP_DN63567_CERS_Cers
MTHSFWNERFWLPNNITWADLRDTPLAKYPQIEDLLIVIPIALVLFLSRLLFERLLAQPVGVCIGISPRGPRKPVHNAILERVFHTI
TQKPDEKRLKGLEKQLDWSPYQIRQWFARRRAQIRPTKMAKFKESLWKLVYYTGILGYGLFVLSEEDFFYELSKCWTNYPFQPV
STNLRNYYVTELSFYVSLVLCEIVSVKRKGYRQMLTHHLITILLMIFSYVDNFIRIGCVILITHDIADVWLETVKLLNYTKCSRLSNFV
FAIFALHFIATRLVYFPFWIIYSTLVQTLPSFGFFPAYVVFNALLVGLLLLHIYWAALILHVGYTLLTTPASEQSSCTQIYSDSALSEDE
QEYSFHSTPTYLFRISSTLHNGSASGGNSQIFQNGNTSGGNGLIGKNGSLAKSATLVANGNSCFMSGKH
>IP_DN64224_ANTP_Nk3
MFGSGGSNPTTNITSPVSSSSSDFKTPAQMQLDVMRASAAAGVHPYLSLLYYSVATGMLPAHCLSLINQVSPAAAVQRSGGG
GVGSPISSLIGLQNLTALQLPKMPSAAAAKQDSHVPLLKVDTSRPAESSESSAV
>IP_DN64224_ANTP_Nk3
SGRVRPSPTPRCSSWSAASPSRSTSPAPSAPTSPTPSSSPRLRLRSGSKTGGTRRKGSSYRRTRTRHLPTQPAATTHLPTTANSHY
TCSAAAAATPPPTSPLRSRHRPPTSRLRPRCNWM
>IP_DN64595_PRD_AMB
MDSCESSPPKKCKFAKIEDLLGIGQRQDATDSQRLKRDGIFDSDFNEAAVREVGRPVWLQPPALTSRGEAATSPSPTSTSSFSAL
MSLSQAQRKKSRYRTTFSAEQLEQMERAFQLAPYPDVFTREELAARVGLTEARVQVWFQNRRARWRKERQARR
>IP_DN64840_ANTP_Nk2.1
MGVFATPFIFPSQHQQQQHETYGACLSKSPMMDNCNPPNWNFAAAQMGAAAGFMPPAASTGNTFGYCSPYTANFTPNN
MVGQGMMNQQAAAAAMGMYGHAAAATGNPQTAAATAFPGNIGTGGGGSGQDQVSAAAAGISSGATGQSAWDAFTR
QQNWYPTTPGGHNDPRLALWMRSSAAAAAMNPMGFPLGFPGSQDPTKFFATRRKRRVLFTQAQIYELERRFKQQKYLTAA
ERENLAHVINLSPTQVKIWFQNHRYKTKKSKSPSAGGATERGSEGSKSTAAAVSSTSSPNKCSSSEECSSISSDQNQNTLMQSER
SKPSNSENALLNLSNAKSPEDPLHLSTNSLVQDCSLQQWSLH
>IP_DN65023_HNF_Hmbox

MNGNPPEEPKYTIEQVELLRRLRNSGLTKSQIIQALERMDRVDTSGFSIPADEPKINNNNIANFNLMNSSPASEYLRNIALQQAI
AVATLANGSLIGQNKASLQSGLANQRSLAEDLLTPQLKPNGIGGKMSGSSSYLPTTCLQQQVSSVPSYVKREGTGPNVPLGDPS
AIKDDDKDLMALQAMSVVDSKEEIKTFMMCHGIKQHQVAEKTGVSQSYISQFLHQGVWMKDSTRTNIYKWYLDERRKREVS
GEMVVSSRGSSPSLRTSPNLFTSPAVQATSVSPNRTSPNVVSPPSGQPKKRDRFVWKENCVALLEKYYEANPYPDDTRREEIAA
ACNQLNGNSGSGQEERVVTAVKVYNWFANRRKEVKRKNASQEGMKSEPLWSSDQPTLANQMEENSTNEGSALMAALSGN
IPRNSPKHFTSNGFDEMGPCTKVSRASEESTEFLESQLAEVDHTISALVAPDSPSSEIDQFDDNQPMEKAAAAVDFTAA
>IP_DN65413_ANTP_AMB
MQPSDLSPNSLKFSIDSLLCSRTHEESELPEKKSFEDTCRFSWDDFNKSSLTSELDTEGIKGLGLAESGREALSPPKKGHSPPEKKP
ASRKLGRNPRVPFSSTQVAVLEQKYQHKQYLSSIDVAELSSLLNLNDSRVKIWFQNRRARARKEQEMSKIPRTIHSTTRSQQVSS
LGNLNAHLFFPALCASLETAACRAAAAAVTPSSAAAPPLTPVTSGGGGALRLQQTSTNHYFPSTVSLANPPPKPPMGGLFFSS
>IP_DN65824_PRD+ANTP_AMB
MPRFKFQTPQGAIIMEFDESATIYQVRMLLSQSLDRKVKSLLIPAAGTDSVNKLSDSQRMKEVDRRSIILTECEKREEIEEEMRPQ
SSSNPQLSEEGPSTCEATSTQDEERTESAPREESSSPPSSCKKCCGSSNNSSSFLLPFLSSFGTMAPFGGVADSGCPPVLPFNPAM
LHPAVLKYCIHFEIEKFKINIEIFAHAATGTDAAALEKTLTLAAGETLDSNGRFLPVIMHLLKKDDAKVLLYDKTKILLMQMKNAAS
ATVQHPGAQRSVVMDTVNGSEGMKSDSPADMANASLDKSRINLTEMAKLSHHLKQHFINLAQLNAAAQIPIVASSVAHKPSF
LNPADAPHPPAAAAAAATQPTSPGGALSHNNNNHCTPSSELRRARRRTTISKYSKKVMVEWYLNNGRYLTTEGRQLLSNHLG
LTEENVRVFFKNLRRIEKKGEAEGLSLFDMVSFADSAAGGNGYNPYADDTLFDDDPEQDSPPNVKIEGSSA
>IP_DN66046_PRD_Vsx
EVAPKGGGGGGKCSSRRRHRTIFTNSQVEELEKAFEEAHYPDVYQREILSLKTDLPEDRIQVWFQNRRAKWRKKEKRWGKASI
MAEYGLYGAMVRHSLPLPDAILNSEEGDDHDKQPVAPWLLGMHNKSLSARKELASSEAANSNSDSSTIEPDLTDRKSVV
>IP_DN66164_ANTP_Hox9-13(15)
MPPMSVNWPANPPPPIMAQYAEPTATAAALQHVPYDCKPGDNLPRTHMPEMISPGYNQTFFSPEQKFPPTFDHLGAYQW
NPAYTYYPANDPYKSGYPTGYDTSMKRSYDLSPQSRHLNSLNQMQHFPFNQYDMTTLENGVKSESYLQSPVLLADQYSPNKPI
NDHFARINTGGGGGGQLDTQSQLNTLNNNNNSWSVGAAGQQAALRLPTSSQAATPSGQNSPSSSIGDHSLGSPNGSSPSSA
LCNPSSLQNTPTGGLTNLQSRPSNGLAPPSGGVSGTAGSGGGSGAAAAAAAAPNMGWMARNVSRKKRRPYTKTQTLELEKE
FLYNTYITRERRLEIARSLSLTDRQVKIWFQNRRMKNKKQMNGGTPQTMHMVHPGVVNVPMDHCRYDVC
>IP_DN66164_ANTP_Hox9-13(15)
MSLSPKAAPGFSVTDILNPIDELSYKSGLTGGAGLSAMDSRGALNFPAINYSTNQQMGAAAARGINSSYMGNFAGNFPNFTG
NSYCANTADLSFGPACFGYDHFGRQNAVAAAVAAAGNMGAYGAYQPGTQQSPTTDPRLSFSRLMNNNSISSSYSFFNDPHK
SLLASQRRKRRVLFTQAQVYELERRFKAQKYLSAPERETLAHSIGLTPTQVKIWFQN
>IP_DN66164_ANTP_Hox9-13(15)
MTLSPKTGGFSVTDILNPIEESYKRSGGMNSSALNFASINYGQSSSMSSSPYMHMANMHMMNPTYCNGEFNPYFNPGAFSS
AGQGYGGFGAGNMGWGYGGGGAGGGGDPKMSFSRFMNSSSGMTSSFNPMTSSFSLTDPCKMLAASQRRKRRVLFTQAQ
VYELERRFKTQKYLSAPERETLAHAINLTPTQVKIWFQNHRYKCKRASKERSSNQTDSAKGSAEKDPDTDSLVDPNESDIDSNEE
STNPTLHMPIDYTSNAGQIYGTTLAAVAAVGNNSLQPGNNSLQCNNLGYRSSMEDSVPQLPS
>IP_DN66473_CUT+POU_AMB
MYHWLLLDDQSKQSSLQLQSHFSIPTDLSLHTQMSSNQDERIKPSPPEKSEAELGKSESSKRKRARKSFTLQQKSVLSTEFSANQ
RPSRHRISQLASELNLTSASVANFFANSRQRNKHISL
>IP_DN66612_TALE_Irx
MMNRLMSAGFYPPPSSTACLPPVSIEQASALLESRSPATPSAGPGGKPFPPPPQGSILDLAARHAAAAAAAGQHPGGQLPPAF
FSSTAYHPALFAAAAAAAAGQFNPYSGAAEGYPGGMDFNMITRRKNATRETTSALKAWLNDHKKNPYPTKGEKIMLAIITKM
TLTQVSTWFANARRRLKKENKMTWATNGGGGGGVKDDDLSDDEDDDEDHKILEDEEDTSARESGYEDNDEDSEERRRRSSS
VECPSERLHQGSPGSKSESDLLESTCWRGNSSSKLHKPLNVHTKQPLTCTSPTVLSSSAAAPAAPPSTQSPKDSPCLPPQQPKKH
KIWSLADTAMSPPTEKQPQELKIDHHHLRCIPDTLEAGVNLALLHHHHQQQQQQRSYLPPQQQHFSAHAFQNALLSFNSGG
NLPKSY
>IP_DN66907_ZF+ANTP_AMB
QLTNQIQEEEIPISSAIVQISVLRIFFTTSPIDKIQEDFEMGKGGFFMEDILGYQQKDSEQKMKELGQPQPLRTEEIIEYEEDQDIDS
TIDNSECADYDMNAEEDNILDFSAEGTRNLMSSPELSPASKKKRKRRVLFTKNQTYELERRFRQQRYLSAPEREALAAMIDLTPG
QVKIWFQNHRYKCKKSNKEKMGGALDSGAPFFNNPGYPSLGLAYMMQANNATKSHTIDPTRVSIHNPLTAGIIYPQAQRVNP
WNNPVLHQQSYPNSLMAALATKHGIFQALNNATRLSHGFNSTEQMKQTLLSNLIRSQNITNSLPQTSSGPQSEKETNITSSPQ
PKNTA
>IP_DN67235_PRD_AMB
MSSSFGIDALLALSPHHRQSGPQCGAPSPQCGAPSPQCGAPSPQCGAPSPPRSAIPTLQLSVSPQSPGFSPRSAKSASPPQLVP
VSPPPVSAAEYMYPAMLGGGISSDALGLWNLQYGGASPAFGPHLTPHDAPEPEMTGYSQAAAAAVLFHSALIRASLAKRKRR
HRTIFTEEQLAELEAVFVRSHYPDVLAREKVATSLGLKEERVEVWFKNRRAKWRKQKRDAEEKKRLLKLKSK
>IP_DN67846_LIM_Lhx2/9

PACSPFPPMPYPQGLPPPPPADQFMFGMNGPSPQDMMLPHNVPQKAQKGRPRKKKNVHSDDPFHSLYGSDQNQSPHDM
LGMYMNGSVEQQNRAKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRMMMKNQ
ANGGGGGSGKLMEDEFSEDGGAAKECESGDEMMEEMKNPDSPSAESISDFGEMMSPESPDSNPPIDVKESSDSSAAATAVK
PPVLTQSF
>IP_DN67846_LIM_Lhx2/9
MAQPCIAPKLPVESKPSPLVKCEPTNEPTPPAAAAPAGACRECGKPIRDKEYVSTGNDSWHSGCLCCSECGVNVSGEDTCYER
GDRIYCKKDYHKIFNQRSCHRCGQIIPSTEMVQRARSLVFHLTCFTCSICQKTLETGDRIGLQNNEIFCISHSSCENGGGYPMTSC
ANGNIPMSTCINQPIPNEYDSLPPPPPLRQPMMMPPPPLQVSQCGTMYLPHMYQPTQVEFLKSLAASSTAQAPKGRPKKKKA
LEEQMMSYSNQAAAAAAMCDPDGSMQALFTLNDPHRNPHLMGDMRGAGGHGGGGRSKRMRTSFKHHQLQAMKVYFN
ANHNPDAKDLKQLATKTGLTKRVLQVWFQNARAKFRRNLMREQLQASRKEKSSSECAANGGGGGEPNGAEEKAETNEDELL
DESSEDESMMSYSPASSTESHELDTSLEEN
>IP_DN67947_ANTP_Barx
MISQLQFTGFPALWNQSPPFDKPRNPWTSPENQKEQSSKPKKSFLISNILDGDSQDKRLQPEDKTLTSPSDPPVSNPGQLRLFP
GQFEPVNPGPAMQASTGFLNPLFYHPAFLTSHNDLYSSQGLSLDGFALMTPAAAYQYPPGQLSPTSGGGAGLRRCRRSRTVFT
EAQLANLEKQFGNTKYLSTAQRSRIAKSSGLTQMQVKTWYQNRRMKWKKQL
>IP_DN68178_LIM_Lhx1/5
MILNGVGDHVQCENSAAAAITPEGCMNQQQQQPPPVVVMCGGCSLPIMDRYLMNVLDKPWHSDCVRCCDCGAALTDKC
YSRDGRLFCKDDFYRRFSTRCFGCGEAILPKELLRKVKKQLFHLSCFVCSVCGKQLATGDQLYVTPDNKLICRHDYFTIREESKEPE
CGESCEESGHSSCSMSPPQFAHGAPIADQAECIVQPDSSTAAGDPVSGQMLDTEVKQEAEFGELTPCGDLEKHIPELEQSAELP
GADRFLKPENFHMKQEPQPAATATAGSPALECHNDSSDKQEGYTSDDAGGENSDDEKYDDEAGQSDDNKSEFDENELLKSP
KAGCDEDVMEKTKTQVGSKRRGPRTTIKAKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLN
ALGARRAAFFQRRMRGPFDTLPHAADFLGGHGGGGPGLNFFPHQDLPGSEFFAMNQPPFSYDFFPGVHGSAPPPPPPGGH
MSMDNNGDFNPGGNGSAADALRLSSELVAAANFMSHCPPAPPPTEDTLSLSGLGNPHLNIATSDIVLPPSDRFLTSFAKSRGFL
CEGEPTNNVW
>IP_DN68209_ANTP_Tlx
MDLKFSITSILSPVSCHNIPPSDSSPSPATSPSENPRNSGPDPERQGLAAADVAASVLAQIYPGNPPHNPPISSAASFAANQASLL
DFAQHKFLRERFLLAAASLSSGTSPASHPPSFNPAAAASFFRAGHHHHPLGTPLFSLGCLVSPGAGHPGVGGGPPKRKKPRTSF
TRLQICELEKRFHQQKYLASGERATLASRLNMTDSQVKTWFQNRRTKWRRQSSEEREADTKVATQMMLGGLPPSPLDCSGQ
PPATDGSHLSPDSPSDCSDDSDTCC
>IP_DN68501_LIM_amb
MKGNSVAAATSTAGSNSVLKCFACSETIREHTYFKLLDRTWHSSCVKCTTCGVQLPSKCYTRDGGLYLFCQHHFNSDAFSRCSG
CSQRIAPEELVQRIHDYIFHLSCLTCCSCNQHLQTGDNIYIITGAPCRVFCTTHYRSQVAANGGVMKLMPSAGGEWPAPTPVHY
NPNAHFSDPTASQVHAYSYPQQQHGVVYPGHQQGMPQQHHPGSHPQQQQLLCAPGSAVPPPVQHSGNPLSSISQNVVSQ
PSPATVSNHDSNSSDFLFSADSSRPTGNSLPPHATFSQAAGAQLQLEDEKMSDNLSVCSTENDNISSPKPMGSNNNTRKPLKR
EIAEHLRAYFDKNSKPTKPEREELANQLGITERVIQIWFQNRRSKDRRVAGIASGSPRPLSPASTDGNSNSMPHSIAQNENYFM
PHSNSMLMEQDQTAIITPSHLT
>IP_DN68679_TALE_AMB
MKLAVLLGGSASRPPPPAAAAAPRRSPRRPYSRRRSKDQIEEDTVPLRHWLLLHSGHPFPTRVEKKRLARECGLQLKQVEDWFL
NTRRRILKHHQIAHSGRQKKEWTALIRACCRQLQLRPCTPPCTICHDFELSDSDENNNLTHKEEVKADSTEFKHRYLTRRAESM
DAAAALLQLSKN
>IP_DN68679_TALE_AMB
MLSALLVRYRCLNSVLSALTSSLCVRLLFSSESDSSKSWQMVQGGVQGRSCSWRQQARISAVHSFFWRPLWAIWWCLRMRR
RVLRNQSSTCFNCSPHSRASRFFSTLVGNGCPECRRSQWRSGTVSSSIWSLLRLLEYGLLGERLGAAAAAGGGGREALPPSSTAS
FIADAVEA
>IP_DN68900_PRD_Pitx
MLGDLEASDLGITGCCPSDPDSGLSASLAVHSSGHSAAGLDSSGHCAAGLDSSGHCAATLESSGQSAAGLASSPGSIRDSPNKK
HERSSQAESEMESTIKKEELLPVAGLLAPPADGPPQQSAPSPPSAALSRERRQRTQFTSHQLRELESMFARNRYPDMATRERLS
AWVQLSETKVRIWFKNRRAKWRKKERHNTYQTHNMIDSWKPGGAPLLPPTPTAHFGYPSYDPLRSPLVSAAPTFGSAAEAHK
AGSSFMSAQSAFSAPFRNPYGQTSWNFLSSPLQPQFSGSHGIQPFPSHYPDHMFETHLSSFPPSGATPANQSDSLLTQGSAAG
SQPANYFNLENAASCFQVPTSR
>IP_DN69333_TALE_Meis
STTGESLSSPVPVGGAEEAAGPGYRSHHPPGQQLVHQREAAHRPAHDRPIKSRRHALGVRGGGRGPDGLHARHAARRSALPP
PSQSVYYIHSHCNHQKLVKAVL
>IP_DN69333_TALE_Meis
GQQLVSHSLHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQSNRAGMPSAYAAADAAQMGFMLDTQQGGA
PYLRHPSQYTTFTHTATIKNW
>IP_DN69528_TALE_Pknox

MMHTISHQPYNSGGLLTTGSSQAGPTTGGGGGHPHLSTPPQQHMMNQQQMTPPSPQQMHNGTSTPGSTGGAQVMGSS
TPNGAGSQQGNMQSQAGPGTNTPLTSGGSSLGVTDNHQFEQDKKVIFNHPLFPLLQLLFEKCETATNSADCPSSASFDLDIEN
FIHLHERDRKPFYTDNADANDVMIRAIQVLRIHLLELEKVNELCKDFCQRYIACLKNKMQSDSLLGCGGGGGQYPVSTSPNSQL
QQQLAANCGAQSQVNHAALQQPPPSANQLLYPQLSMFPASMAAAAFNQSSPYQMGHMNGLGGHSPGGNGGMLSGSTPI
SQVGLQHQGGSPLHHHLLASHGQEMNPDGSYKKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNN
WFINARRRILQPMLEASNPDSTPKSKKSKNSRPTAANRFWPDNLANGQIQGGSCKTECESPSPTDGGGESTLNGHTPSNNNN
MSLLSPSGLSENSEHDSSNGSGNETLSPSSQRTPGAGGGHGSPSNFLPNGGGGITPFSSSGGAAHITSSAQSSAGGGGALYME
QSS
>IP_DN69723_ANTP_Nk6
MLTHQAALVRHWQQQQIQRQQATLGASSLDINPNQRNDTDSPNLALNANNPMEERNETGIMNEQSNTADSVKKHTRPTFS
GHQIFALEKTFEQTKYLAGPERARLAFALGMSESQVKVWFQNRRTKWRKKHAAEVASTRFKLEGNLRRSTSSDHEVYSDTD
>IP_DN70276_TALE_Pbx
MVVGGVYDYGNHAGIGMDEQQHRFAMAAALMHQQQQMAAVAQQQQDASAQMHAGASVPGGWGHMGMGQPVVQ
DQGVPSSVLPSQSSGMMTSEETRKQQLQEILQQIMTITDQSLDEAQARKQTLNIHRMRPALFSVLCEIKEKTGTFLNMRNQAD
DDAPDPQIVRLDNMLAAEGVSGDGKSPTGMASSSASGAGQPDNTIEHSDYKAKLQEIRRVYHSEHGKYQLACNEFTTHVMN
LLREQSRTRPISPKEIDRMVAIIQKKFNAIQMQLKQGTCETVMILRSRFLDARRKRRNFSKQASEILNEYFYSHLNNPYPSEEVKEE
LAKKCCLSVSQISNWFGNKRIRYKKNIDNPGSYNPNDILLPPGGAATPGVYPATTPSSYSDTDQFSSGQLTPVGSDIDSGGGAPY
NSQQGGAASASSSASAGHDGGASASGHMFQYPPMGVSTSDASGMGAGSAVGAGMSALFDDSLNNWQQAANMARSSAN
QSF
>IP_DN70650_TALE_Meis
MNNLMYEEPPISGGGGGSLPPQYPLTIDSGSYSAPYRPPPPTGSQFAQFPRGQQQQQAADTPGGDPQAASEGQDKEAVYSH
PLYPLLELLFEKCELATCTPRGGGSSDPSGGDPVCSSSSFDEDLRLVSKNILSSAVDKNFYSDPETDNLMLLAIQVLRLHLLELEKVH
ELCDNFCDRYISCLKSKMPIDLVVDDKDPGSSPAHAGRGLDKGRRLSCDSPDDFQPPEEPESVQSPPYQLPPSSDTDTAAHSPP
PVSPKREPAAAADEPASTQSVHQSKSPEMDGTTATATVGSENHSPNTNNNHHHHSSSGSQSELVTSPSSDETSGGGDPKKPQ
KKRGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQV
>IP_DN70650_TALE_Meis
TWRMVRPVSWASCFFCSSDGYGCVRCWKSQARMMLVAVLGKMPRFFCGFLGSPPPLVSSEEGEVTSSDWLPLLLWWWWL
LLVLGEWFSDPTVAVAVVPSISGDLDWWTDWVLAGSSAAAAGSRLGDTGGGLWAAVSVSLEGGSWYGGDCTLSGSSGGW
KSSGESQESRLPLSKPRPAWAGLEPGSLSSTTRSMGILLLRQEM
>IP_DN71368_ANTP_AMB
MNAEEDNILDFSAEGTRNLMSSPELSPASKKKRKRRVLFTKNQTYELERRFRQQRYLSAPEREALAAMIDLTPGQVNILLSYSTK
KILRRVVSSQMLKCLIPVVQVKIWFQNHRYKCKKSNKEKMGGALDSGAPFFNNPGYPSLGLAYMMQANNATKSHTIDPTRVSI
HNPLTAGIIYPQAQRVNPWNNPVLHQQSYPNSLMAALATKHGIFQALNNATRLSHGFNSTEQMKQTLLSNLIRSQNITNSLPQ
TSSGPQSEKETNITSSPQPKNTA
>IP_DN71480_TALE_Tgif
MDSVSQIGNPPPSKSVSPLSGQNSPLSLVTRIESSSTTLRNESSSTTLRNDLGLSTPNLSTLGLSPYQTAASPILYPSLCQATLREWL
QQNPYRALLTSRGQDLTFREQDDVLLASRPGHELTSRGHELTSRGEDESREQPDCSISRQASCVSHQSEENSDDDGSSSLEISLT
TSGESGFSQGSGCHKPPQQQPKEKKKRGSLPKEAICIMRSWLEQHRYHAYPTEQEKLSLAQQTNLTPLQVCNWFINARRRVLP
EIIKNEGNDPAMFTISRKPRTTSLASSDEGVNSPTVKRKTQAPWQPWIDEEKSEKQKKMKRSNSISDLEAALPKLRLPGYNEAHL
DLWYRSQLSPSEKLSPLLNRIPATAGSPISPTGLYYPVASSMYLPIVPARNIQSFPPPTCWKL
>IP_DN71570_TALE_Irx
LLSNPYPTKGEKVMLAIITKMSLTQVSTWFANARRRLKKEGKMTWTCRNTYDDGEDSSSSSNSTDHHDGSPDHFKAAVAATI
ALAEEDEGYNEPSGEPHLHTPIGENAAVQPRRRRLSPEAHLTRRRPLRLTLPTPEHRQHHQAAAADSLMPKPKFWSLVDTALV
CTSDKRRPPTAPSLQAASTQQACT
>IP_DN71570_TALE_Irx
MGVCRCGSPDGSLYPSSSSASAIVAATAALKWSGLPSWWSVLLLLLLLSSPSSYVFRHVHVILPSFLSLLLALANQVLTWVRLILV
MMASMTFSPLVGYGLLRR
>IP_DN71583_PRD_AMB
MKSECSPSRPRRSRTTFTAFQLHQLESIFLSNPYPDVARRDQLAAMLGLTESRIQVWFQNRRAKRRKAEWKTGTCQDHLAPLP
DQMRPPLQWSSGPQFAMTSQMQLPPHLAHTFRTMPFNQNSALPPLNSWANPLLGGPFIPMMRPGLPTNRSLANDFGLQ
>IP_DN72227_LIM_Lmx
MTVSNQQSCSSPPATPSSGETLQLTPSFCNGCHKLIADKFFLQTTDGNCWHENCLTCRACGCYLDQSCFMRNAQPYCKADYD
NLFSGKCSGCLERIPSTELVMRVGGGATYHLACFACSVCSAVLHTGDQFVLRGTHIYCRTHFEKHFPQLCNFPPSQEETSPEKKP
EEPEVKSEPDASPNPEEKLKLDCEENEEQGQNSDNELDSPKDEENSDDENDDSAKDKNSSQQKPCRKRVILTTQQRRAFKAAF
DLSSKPCRKVREQLARETGLSVRVVQVWFQNERAKVKKLAKRQQQQHQNHLEQGLSPPHHMLSDDGNYNMFISTEGGYMH
PAMDGRFPAPPNLIGSDGSMFPPPPHVTNDMFMCPPPPHEGAVEFSEYHSSGLEIKDDLYSSLKMGSGLLNPIDKLYSMQSSY
FTPS

>IP_DN73075_PRD_Otx
MMNVNLYPDMSGYLKGPGGAAAAAASHPYMNPSAMLLTHHHHSAAAAAAHMDSLHSYPAMPYAVPQHHMANPGGPG
GAPGRKQRRERTTFTRTQLEILEQLFSKTRYPDIFMREEVANRINLPESRVQVWFKNRRAKCRQQQQQTDKCEKASPNSGNTT
SPAVANAPTSDKSAAEQQETKPSPVTVKTEPLKLAEPLDILKKAAVVGAGVGYDSLLTKAPHLSPANDDHDNSNNAGLGKEGL
DQDNSISPVPAGYLPSVSSSFVNIWSPAAVGGGNNPSPYSSDPNGGLPRSLAPYTQPPVSQSYFPSAYNPYFDPSYSPYFANSY
QNVAGAGRYGGYTASSAAAAATAADNAALFNSVVDMKNDARNTPGVPSLNEITSWKNY
>IP_DN73593_PRD_AMB
MQLGSPYVQHTNMDSGSFAGSKDGKVRAWRGKRANSIIRSKSREVLTENSIIRSKSREGLSEYSITRSDSRDELTRYPIVRSNSRD
GLTGYLSLPRSNRESDTYQPWQNRPLRPADKENNKSLQDSLDSITIHSSTNKSTYQSTTYLYRKSDSKSLGDLSILSTNSPEISIHFD
DTDSQCSPSPTGTSFRADSKLSAYTEHSGDSTPRFASATPPIIRKTRNPVKLQPSLNTPALSSKTVMLKQRRLAPVLQRKPVIYKSA
IEINCPQILPLHMLTDAEKTLERSFMRDHNPADYMYEIMAEEIGWNLNQVKQWYDRKKATWRSNEGLPDRVGSVKD
>IP_DN73792_TALE_Irx
MVERRPSLRRKAAGLLSGDDRHFPLTADDEFFLKYSSIAGGGGGGGSSSSPASPSAVKESTYPLRVWMEENRANPYPTKNQKLI
LALVTKLTLTQVSTWFANARRRLKKERRGDVDYPEYLPLDLLGPEILTRPVVPYALCLPPAYHTCHECIYPPFQSHLFTPPALTGM
SHATSPSDTSPTLKKGRSILVTSPPLKSPTMRKTKEESAKKTLKSKSIWSLAETALST
>IP_DN75362_ANTP_Hox1
MDYPSTMDYHHHVINDYPIKSPPEITTRANSATATAAYINDYYTNYHQSPNISDYYYPPSNATDNPSPVNSNPGSQFSPQKCYE
YPPNSSEIPAGLYQPPPANPFLPSGYLLSNPQTGYYPGTYHPHTGYPNSVFSRHPAAVNINVSCNIANGNSEEISTADTNSSPSST
LSNTSQNYDNFIHHGEKFGASWGEEGNFGGQTAGEFNPVGVGHMKREVIPAPSGVMVKDELRGLRIEEGGELGQPVGSVAA
GPGTGNMPYSWMKIKRNQPHLQWSKSATAGLPGSAHAYPPGSQTRGGRTNFTNKQLTELEKEFHFNRYLTRARRIEIASSLNL
NETQVKIWFQNRRMKQKKLVKEGKLA
>IP_DN75504_POU_Pou4
MFFAAMSQASSADSKPPQMEQLSMGQSTVSTAGSANPGSTIKYSPPVHFHNTNRPEYKPRPFPPPPTSNFFGSAFEDPFFRAE
SLHMSMGGGGGGGGHPKPEYKPHEYFSSVMADMQQASRCSQLDLLDQINPNFNPAGMTHDQFPHMPGSHLGFHPGSFA
TQQHHGGSALSASHMIPPAMTSHQDTVDADPRELESFAEKFKQRRIKLGVTQADVGKSLSHLKIPGVGSLSQSTICRFESLTLSH
NNMVALKPVLHAWLEEAEAAAKGRLPDKDTCMSGIDKKRKRTSIAAPEKRSLEAYFTVQPRPSGEKIAAIAEKLDLKKNVVRV
WFCNQRQKQKRLKFSAMR
>IP_DN76364_ANTP_AMB
MQGVGYAPPMDLSASDLPAETGCSSRRYRRNRSVFSEKQLADLEKWFVAEKYLTASLRVKIASQCGLTEMQVKTWYQNRRM
KWKKQMQQQGVVDIPTKAKGRPCKRQRPSAESDNEEH
>IP_DN82139_LIM_Lhx1/5
GENSDDEKYDDEAGQSDDNKSEFDENELLKSPKAGCDEDVMEKTKTQVGSKRRGPRTTIKAKQLETLKSAFAATPKPTRHIREQ
LAQETGLNMRVIQVWFQNR
>IP_DN89336_ANTP_Barx
LSYHRLTVCCRLVARPEPGWVRADDAGRRLPVPARPAQPHQRRRRRPETLQTEPHRIHRGSARQPVDGTAEPDREEQRADPD
AGEDLVPEPPHEVEEAGV
>IP_DN89336_ANTP_Barx
THLLLPLHAAVLVPGLHLHLGQPAALRDPAPLCRRQVGELSLGEYGAAPSAASQAGAAAAGGAELAGRVLVGGGRRHQREPI
QAQALRRVCNTPLGGGS
>IP_DN89774_ANTP_Vax
AYYCELKLIISTQLVMSHQLDPAAAAAIKAPEHAIRRISVKDATTGCVRELVFPASLDIDRPKRERTSFTAQQLFLLEEEFLKAQYLV
GRDRTELAKKLNLSETQIKVWFQNRRTKHKREKCKNQCEGAEGVMGHHPGGRSVNQSAGVLQLIERTNPLYAITKSLIDNRDS
KPNQQFPWPEKQQNTQQILRSSGNTAQQPQQQQQYSMSEFWYGMGHVDYPSPSL
>IP_DN100684_PRD_Pax4/6
KKKNQRNRTAYSQNQLELLEKEFEKSHYPDVYARERLSEATDIPESKIQIWFSNRRAKHRREEKIKSNNNNEMPLHPSPLELTNQ
RIDRMENGGQSDPQTPQSLYQSPYFPYTNSGATSLSITNILQAAPPPAVNQSTP

Protein domain identification

trinityID class	(homeoDB blast)	protein domains (Hmmer search)
DN68178_c0_g1_i3|m.106476	LIM	LIM
DN56056_c0_g1_i1|m.42738	LIM	LIM
DN82139_c0_g1_i1|m.231741	LIM	Orbi_VP3 YL1
DN68501_c2_g1_i1|m.110831	LIM	LIM
DN27657_c1_g1_i1|m.11322	LIM	C6_DPF Cytochrome_C554 HypA LIM zf-BED zf-C2H2
DN67846_c0_g4_i1|m.102863	LIM	Homez LIM zinc_ribbon_2

DN49678_c0_g2_i1|m.29979	LIM	LIM
DN67846_c0_g1_i1|m.102860	LIM	Homez
DN72227_c0_g9_i2|m.162480	LIM	C1_1 CDC45 DUF1510 LIM
DN71480_c1_g4_i4	TALE	
DN70650_c1_g5_i7	TALE	
DN68679_c0_g2_i1	TALE	PMAIP1
DN73792_c1_g1_i1	TALE	Homez
DN70276_c2_g3_i3|m.132878	TALE	HTH_19 HTH_3 Isy1 PBC
DN44510_c0_g1_i1|m.23508	TALE	HA
DN71570_c5_g7_i5	TALE	DUF4260 Homez
DN69528_c1_g2_i3|m.122462	TALE	
DN69333_c0_g4_i5	TALE	Pinin_SDK_N
DN66612_c1_g4_i2|m.91737	TALE	CDC45 DUF2457 DUF566 Homez Nop14
DN18113_c0_g1_i1|m.7284	POU	HTH_19 HTH_3 HTH_31 Pou
DN75504_c0_g2_i7|m.225572	POU	HTH_19 HTH_31 Pou
DN55409_c0_g2_i1|m.41119	POU	HTH_19 HTH_3 HTH_31 Pou
DN63567_c1_g1_i1|m.70349	CERS	rve_3 TRAM_LAG1_CLN8
DN26501_c1_g1_i1	SINE	
DN62049_c0_g2_i3|m.62452	SINE	HTH_3
DN45560_c0_g1_i1|m.24666	SINE	Homez HTH_3
DN31212_c0_g1_i1|m.13546	CUT	A1_Propeptide CUT HTH_24 HTH_38 HTH_Tnp_ISL3 MarR_2 Sigma70_r4_2
DN36933_c0_g1_i1|m.17321	CUT	CUT
DN52339_c0_g1_i2	PROS	HPD
DN57662_c0_g1_i1|m.46698	PRD	OAR
DN66046_c0_g1_i4|m.86993	PRD	
DN68900_c3_g1_i2|m.114481	PRD	
DN38710_c0_g1_i2|m.18965	PRD	MRP-L20 Rotamase_2
DN51842_c2_g1_i1|m.33653	PRD	HA
DN71583_c0_g1_i1|m.152092	PRD	
DN100684_c0_g1_i1|m.236544	PRD	HA Statherin
DN31321_c0_g1_i1|m.13615	PRD	
DN73075_c6_g1_i5|m.177200	PRD	DUF3534
DN67235_c0_g3_i1|m.97300_HD1	PRD	
DN67235_c0_g3_i1|m.97300_HD2	PRD	
DN73593_c0_g2_i3|m.184706	PRD	
DN64595_c0_g1_i1|m.76632	PRD	
DN54186_c0_g1_i2|m.38019	PRD	HTH_23 HTH_29 HTH_Tnp_IS1 PAX Statherin
DN58162_c0_g1_i2|m.48062	PRD	
DN60252_c0_g2_i3|m.55039	PRD	HA
DN67235_c0_g4_i4|m.97302	PRD	HA
DN19265_c0_g1_i1|m.7660	PRD	
DN12334_c0_g1_i1|m.4957	PRD	
DN65023_c0_g1_i1|m.79542	HNF	Cro HNF-1_N HTH_19 HTH_23 HTH_3 HTH_31 HTH_37 MerR_1
DN75125_c1_g6_i2|m.217565.1	ZF	
DN61481_c0_g1_i1|m.60009	ZF	zf-C2H2
DN64840_c0_g1_i1|m.78270	ANTP	

DN83515_c0_g1_i1	ANTP	
DN60751_c0_g1_i1|m.56640	ANTP	
DN15610_c0_g1_i1|m.6431	ANTP	
DN60800_c0_g1_i2|m.56990	ANTP	Homez
DN54162_c1_g1_i1	ANTP	
DN12279_c0_g2_i1|m.4929	ANTP	
DN69723_c1_g1_i4|m.125267	ANTP	BsuBI_PstI_RE Homez
DN57889_c1_g2_i1	ANTP	
DN76364_c0_g1_i1|m.230241	ANTP	Stc1
DN17434_c0_g1_i1	ANTP	
DN66164_c5_g2_i1|m.87495	ANTP	
DN60676_c0_g1_i2|m.56432	ANTP	HA
DN64224_c0_g1_i1	ANTP	Mating_C
DN65413_c0_g2_i2|m.82199	ANTP	Homez
DN66164_c5_g1_i1|m.87491	ANTP	Fork_head_N
DN67947_c0_g2_i1|m.104238	ANTP	Homez
DN61228_c2_g1_i1|m.58904	ANTP	
DN62850_c0_g1_i3|m.66453	ANTP	
DN44102_c0_g2_i1|m.23116	ANTP	Homez
DN60553_c0_g2_i1|m.55905	ANTP	Homez
DN26203_c0_g1_i1|m.10708	ANTP	Homez Tetraspannin
DN86741_c0_g1_i1	ANTP	
DN56009_c0_g1_i5|m.42659	ANTP	
DN10641_c0_g1_i1|m.4062	ANTP	
DN75362_c0_g1_i3|m.221830	ANTP	
DN68209_c2_g2_i2|m.107169	ANTP	Homez
DN40909_c0_g1_i1|m.20738	ANTP	
DN89336_c0_g1_i1	ANTP	
DN66164_c4_g4_i4|m.87486	ANTP	FTCD_C
DN36044_c0_g2_i1|m.16521	ANTP	BORA_N Homez
DN43104_c0_g1_i1|m.22332	ANTP	Homez
DN71368_c0_g1_i1|m.147900	ANTP	
DN89774_c0_g1_i1|m.233752	ANTP	CENP-B_N Death Homez
DN55547_c0_g2_i6|m.41438	LIM+ZF	C1_3 C1_4 DUF3153 zf-C2H2 zf-C2H2_4 zf-C2H2_6 zf-C2H2_jaz zf- H2C2_2 Zn_Tnp_IS1595
DN29418_c1_g1_i1|m.12421	PRD+POU+ANTP	
DN66473_c2_g1_i2|m.89822	CUT+POU	DUF2316 DUF3469 FAM163 PRCC
DN66473_c2_g1_i2|m.89822	CUT+POU	DUF2316 DUF3469 FAM163 PRCC
DN65824_c0_g1_i1|m.85443	PRD+ANTP	
DN66907_c1_g1_i2|m.94022	ZF+ANTP	

2 - *Symsagittifera roscoffensis*

>SR_DN25767_LIM+ZF+PRD_AMB
IRSMASLAVNSKSDGPFRIPRSPLKYPQDLPHLTLAEAIQSWKNLIKKLRKSSQGFTASPEPKKLRKVSEMEELKVSPEKQTTKRN
LRRRKKPAIKDASSVTSNMESDESFSGSGKKRKRKNSDQIKVLLRHFKKNPKWDRKTIENAANESGLSISQAYKWGWDRKNKS
SSLEGIDKPSKDEF
>SR_DN42484_ANTP_Tlx
RHPNSHIISPQQHSVDYTGIPGMVGVIGPGVVPGFGSGVGGVSSAGGLMKRKKPRTSFTRLQICELEKRFHQQKYLASGERATL
ASKLNMTDSQVKTWFQNRRTKWR
>SR_DN52363_PRD_AMB
MTSSNILTDNVFFEDDLFKSHTMASLPEMDQKSPSKKKRVIKKITFEQTNILENIFMEDPSWSTKTIKKACKLLGLPRRKVYKWG
YDRKAKKVSFTKVNSISDRDVKLPSNILS
>SR_DN56103_ANTP_Nk1

RSRTAFTYEQLVALENKFRCSRYLSVCERLNIALALNLTETQVKIWFQNRRTKFKKQNPNAPLIDQELLGSSGGSGKRGGCHRST
THHQNHHINNCHHIQQHLHMTSAVVLAN
>SR_DN57798_ZF+POU_AMB
MTTQFTYFNNFRNENGLNWFSNEYEFDNNQFTNMNQNEMLFYFDSKPAMQNEIKPVADSPLQDYAFESSLKNVEKWDNSS
MKCENEEPVLIKTHYDSHSQEGTNSTLEVDTTSEKSLKRGYEDLFLMVEDKNLYKIGNEDDSLTTTNNKDLVNLDKIISDKETDLN
NFIEDMLKNCPCDNLIKKQRIYKAKNIKRKRKTKTQIRILEKELQNNPNWMKEDFKRLSEELSLNRDQVYKWYWDQKKKSDF
>SR_DN61173_PRD_Pax4/6
GEEGEAVGEESKKYSKSQRNRTAYTSEQVDALEKEFEKSHYPDVYARERLSTLTGIPETKIQIWFSNRRAKHRREEKIKTNNNNN
NIDSISSGQLSMKTEAGNHSDNSPSVPGIYGTTNTNVPMPPYFTHYNMHGALTNILQNTQQNSAVHSNNQTHPSLSSAISYNA
APNMPSPNSFTRP
>SR_DN65396_TALE+ZF+CUT+CERS_AMB
CSSDLAAVTAVMGLNTLNGDSNNSELQISAFANCAKSLVNTGNNTKISHDTLKSKIVSAQKSRKNLRAKKRKKTQEEINILEAYFT
QDPAWTRKTVKSLKGELPNLSVDQIYKWGYDKKLLQKK
>SR_DN70982_HNF_Hmbox
GYSAVGQPIQQVKKRDRFVWKESCVCLLEKFYHVNPYPDDAKREEIAQACNQLIDSTGQGHDEKLVTAVKVYNWFANRRKEIK
RKNHFAEQQQAAAATNHM
>SR_DN74225_ANTP_Tlx
GIPGMVGVIGPGVVPGFGSGVGGVSSAGGLMKRKKPRTSFTRLQICELEKRFHQQKYLASGERATLASKLNMTDSQVKTWFQ
NRRTKWRRQSTEEREAERQVATQLMIGMAATAAVQQQQQNQSNSHNLFDDKSLLDSHGEDS
>SR_DN75610_PRD_Vsx
VNSNGGGCYPSASSLLAGSMMGGQGMNGGGPLSLLGKKKKKRRRHRTIFTSYQIEELEKAFGDAHYPDVYAREMLSLKTDLPE
DRIQVWFQNRRAKWRKKEKTWG
>SR_DN76474_PRD_Alx
EFEGEEDFEDSDSIGGDGASTRSRKKRRNRTTFTQGQLEEMERIFQKTHYPDVYVREQLALKCDLTEARVQVWFQNRRAKWR
KKEKHQQLRSGFHSSGGGGGTPFNPGGGNSVAEASLAAHRNDVITQQLSSTWPYSGASGAAPGHLAGAAGSTLLGGNSANP
>SR_DN80548_ANTP_Barx
LQSAINNRRCRRSRTVFSDQQLKGLEKNFCSTKYLTTMQRNRLAKESGLTQMQVKTWYQNRRMKWKKQLLLDGATSIPTKPK
GRPKKILEEQELEEDEEH
>SR_DN87552_SINE_Six1/2
MADVPPFSFNPEQTACICHVMEISNEMTKLEKFMWNIPALDMFQQNESVIRAKAHLAYNDQQYKEVYRLIQSSNFTIQHHNS
LQQLWLNAHYAEAENSRNKSLGAVGKYRIRRKFPLPRTIWDGEETSYCFKEKSRNMLREYYQHNPYPSPREKRDLAENTDLSVT
QVSNWFKNRRQRDRALENRT
>SR_DN89648_ANTP_Nk2.2
MPGFSMDEILGLGSQPQPNNSRSSGDFQTSPLIASSRESGLRLASNDASSDAGSDEVLDYYEEDEANLGDDNMDDDEEDEDE
GIGSGSSHSFGRFPQDGASNSTFGGLGEVDKSSLLTNLMGDKKKRKRRVLFSKSQTFQLESRFRQQRYLSAPEREQLANAINLTP
GQVKIWFQNHRYKCKKSSKDKCSP
>SR_DN93804_ANTP_Bsx
CRMWRRTVQSVRIIGHSNHLGLNPFLGCMFPMAQSCSAMSGVPGLGFNMGVHPGGDPRMRIYRRRKARTVFSDGQLHGL
ERKFQQQRYLSTPERIEIATQYNLTETQASSMSCNYLQIDQT
>SR_DN95229_PRD+ANTP_AMB
MENLFYQEELAGSTESCLNNLSHQLNLEIKQDLSPVGTASVDNFSNQESSSVYANHEEDFEDKMLTLDSSERSQDGDSIENILKK
EIEQCSLDFLVAECCKSKEIKDPVKNSEKMRFKTTKTPLQIQRLQESLIQFPTRFPKKERKKLALEIGLEEVQVYKWYYDNKPKKAN
KKSKKSV
>SR_DN98097_PRD_Vsx
PINLFAPPTLPPLPPHNNNPTSLNSTTAPLNNSSTPSSSNNGNVNINNIQTSNAQNQLTSLAQVNKDALQHVISSSNAAAAAGG
MTFESHNNNPTSLNSATAPLNNSSTPSSSNNGNVNINNIQTSNAQNQLTSLAQVNKDALQHVISSSNAAAAAGGMTFESLV
MGGGVNSNGGGCYPSASSLLAGSMMGGQGMNGGGPLSLLGKKKKKRRRHRTIFTSYQIEELEKAFGDAHYPDVYAREMLSL
KTDLPEDRIQVWFQNRRAKWRKKEKTWGRASIMAEYGLYGAMVRHSLPLPEAILDSAGKLDQEDVKNSVAPWLLGMHRKSL
EAQEALKTEDDNEDVDDEDSENESTKSSPDEEIEKLHDKLKLANQNQARAMTAQI
>SR_DN98018_PRD+CERS_AMB
YNQMNCFGFNQTNDFNESNFFNPYKVATANYMFLFNEEDNLKNEINSFSDFKSTEEMSSKGGSNLSVIAEDNAPLCEEITQITS
YEDDLNSLALDIQSKINNGSIENIVDEALDNVFDGTIEKFRPIQHKKKKTSAQNQYLEQALEANHEWDKAFMQDMAGKLGLKY
RQVYKWYWDKIKKRQAKQEKKASKVKHNKFEDIESAFSSVFN
>SR_DN100159_ANTP_AMB
HCGVPLASSSVSGSISVPPSSSIPSPLALQNFLNKKNRRRSDVAELLCLSETQVKTWYQNRRTKWKRQTSAEERLEELESHRCHE
PDCPANTIATHPTTATATLRTTHNNPNNTQVASGQDNMNNNNSAVSLLNMSHHGPSTLQMDIQHQGTMAPHGIPISLSSAL

ANMRHQQQPLTPQQGGIVGPNGLMFPPPQTQQRNCTSQRKEDLLKVTSGHGASGLVSCAAELSPAMCKVSPASLFMSHQH
TTLLNSFISTGQPGGVAPGSLSPHAELASGKHNPLSPFPLSLPLTLPHPHSGQLAAAALAMKLQLLARHNLSA
>SR_DN100159_ANTP_AMB
MPRARLSRQHHRHPPHNRHSHTPHHTQQPKQHASGQWTGQHEQQQLSSLLTQHVPSWPLYTADGHTTPGHHGSPRHPH
QSIQCTGQHATPAAATHPTAGGHCGPKWTHVPPSPNTTEELHLTAQGRSTQGYFWTWSIGTG
>SR_DN100159_ANTP_AMB
MLSKETAELLLFMLSCPLATCVLFGLLCVVRSVAVAVVGWVAMVLAGQSGSWHLWLSSSSSLSSALVCRFHFVLLFWYQVFTC
VSLRQSSSATSLLRLFFLFRKFCSAKGEGMEEEGGTEMEPETEEDASGTPQW
>SR_DN100159_ANTP_AMB
MAPLHCRWTYNTRAPWLPTASPSVYPVHWPTCDTSSSHSPHSRGALWAQMDSCSPLPKHNRGTAPHSARKIYSRLLLDMEH
RDWLVVRLSCPPPCAKCPLPHSS
>SR_DN100159_ANTP_AMB
MKSEAGDTLHMAGDNSAAQLTSPDAPCPEVTLSRSSLRCEVQFLCCVWGGGNMSPFGPTMPPCCGVSGCCWCRMLASAL
DRLMGMPWGAMVPWCCMSICSVEGP
>SR_DN102254_LIM+ZF+ANTP+PRD_AMB
MNTEFSFDNFRNEDDLPFFSSLPDEDNSDRKLDLSSENTFCLSSDKSYNSQDYSPCKKNINEYVLSESFTEDVDQVPFSGKIPDLN
ALAEIVREQISERGVEHVIEKCLDIELEDEPQPLMRKMVRKSPRQIRELRRAFRITPEWNRKFEEKLAKIVGLTRAQVHKWHFDE
RKRTSQK
>SR_DN106437_ANTP_Bsx
MISRPIHLNSKFSSKMQTKGLLLAAAAGKSPVKTKTDFSITNLLTAEITKTQNKDIPTRKTENPTDWEDDSKHKQVTNINDNSSPI
VNVSSDTEEIPSRTFQIPSSRTPQDIKVKDLQSTTNASSCNTATISNSSEHQDPQRLIDKQLQLLNSSNANKSSHSSNRLTGNNSSI
SDSIYKSLLVQTQQNHYNQQQIHSNGLGGGGHFMDHQGPVTPAHFLQSATIAPPVSGLSTSPMFNSLASTFYKSIGLNPFLGC
MFPMAQSCSAMSGVPGLGFNMGVHPGGDPRMRIYRRRKARTVFSDGQLHGLERKFQQQRYLSTPERIEIATQYNLTETQVK
TWFQNRRMKQKKIFKVGGCSETEEVEEDGETE
>SR_DN106703_TALE_AMB
MRQDAKKTQKPRSKSEARSASFNDNALEVLKEWLEGHLDHPYIERSDILKIQTRTSLSRKQIMGWCTNIRRRRLTIIKDEQGDRK
LVPHDNPESKREQNK
>SR_DN107350_ANTP_AMB
YELYLNNAQDTPKDEDFTLFEGKLVSEYDFESGNLTPTKEQSSPSDDEAFSSHSTQQTHLQVDFEGENPNMDDFLYKIRGEISLK
GVNEVVCDALDIKNGDEFSLEPVELIKVKKRKTKDQIRQLEEEFSKNDDWSKEFMNRLASKLNLEPSQVYKWHWDQISEIGRAS
CRERV
>SR_DN108290_LIM+ZF_AMB
MDSATTNHDQMLPRTDNNETDINVSSFHQQHTTDLLRCVPISLGPENMPTNMRTLIQKLCEQIYMQNVIIAQQSVLLRHNIRF
PRPPNSDRNPDTEAASELPLQSSLGSFTQLHDKTSQNSDACGHPTFSPADESRAINNPQCVFNYSAISPFKDDEGLLPEDNGELS
SQNEINSIPDECYIGAFIAISPTCLPDKDSPTKCEVHKIRHRTRTKLTQSEVVLLKSIYENTPSPSKRQFEELANSTGLPTKVLKVWF
QNTRRLRRNTCPRCKRVFVTNHLFVNHSKQCHNNNNCKNPKINSSPELEN
>SR_DN108331_PRD_AMB
MFTIDSILSSVLPCQPSEFNQSPLSRGLSGSETITQKDGGNAEGLLKQWSSPLSLDKHFSTVTASANRNTVNSDEGQKGPTATSI
PIKFNSRYGFNYSGSSSLPCHSGSARKKSRYRTTFTSEQLDQMERAFECAPYPDVFTREELANRIKLTEARVQVWFQNRRARWR
KERNLRNSSAVVETGGGNRSVGDSIARGNNAEDHSECETPINDCHFGQTKSGSGAEVTKNMGQPTNLKKVNSAHRQTPTQL
QMAANVPLAPFFRNVNNFVNTAPVSVLQENNGAVHTAPYHNLYSLSYLCMYQNLVRHQRSTEMISRQQESMVTKIEQE
>SR_DN109741_ANTP_Vax
MDLKKKCPESVRRIKTKDHNGNERELVVPLGLDIDRPKRERTCFSAEQLFLLEEEFTRGQYMVGRDRTLLAGRLGLTETQIKVWF
QNRRTKQKRESIKPDLARKPFNSISKHDNSDSSLTQDNLTSLSSFFDESKPTQFGRHQVSLYD
>SR_DN110321_ANTP_Nk5/Hmx
ALPIWEGNGRKKKTRTVFSRTQVCQLENMFDLKRYLSSSERATLARDLTLSEQQIKIWFQNRRNKLKRQIKTGFEMAAGNISAG
TIPPPNVNAALLNSLGTTSIPNSMHPHAFNQSFTSYLTTNHHSQKEPHSVLASNLFNSGGTPFAHPGNPLVLEGHGHIGRNPSD
QQAHHMFSKLTPYSSLNQHLSRQTAQNVSSHRHPMTPNTSAQTLNSHGVLSSASVSTSIFPTPSSSSTSAVTITPYPYSYTYPNL
QLFKSLSGLV
>SR_DN111202_PRD_Vsx
ALLSSLPQQVDKSTGQISHKKCRRRHRTIFSVSQVEELERAFEDAHYPDVVQRETLSNKTELPEDRIQVWFQNRRAKWRKKEKK
WGKASIMAEYGLYGAMVRHSLPLPDAILRSREEEEEEDDEED
>SR_DN112643_ANTP_AMB
MKKTKFSIADLITPQSTEETLQFLAQRANLLNSLHIHPNIHTQNFANFNQNLSKNLSMKVSQHESATKNFPNLTDQKFLLQVASK
NLKPKREQTSSTVPNFASSGQNFSSISRSFSPSGAELNSNISFTAGHAASSFDNPTTELLRNRAKSQFIASKATIDHLGNRSSKKTL
DNEKEEPSSSLNEATINLNASVSHSTDGFHNLTPACVQLSPPANSPVGSESSDHIQFSSPSSPAGDGSASSANKRTRKGGVDRKP
RQAYSAKQLEKLEEEFKNDKYLSVSKRLELSVALSLSETQIKTWFQNRRTKWKKQLASRVKLAYRQFWQLPHNPAAAAAAGAM

TFPHALYNFPGNPAQQATIATPGGGLPFASLLPQTTGGLMPFGLFAPPNKTIQSEPTGTSLSLQQLAQLVGSSNPTTLSSLITSGA
NNFGLNNAQIGPLMTPLGLSRLLASSSSQFSSVNDQLVAQFGVKNDINSKNENLIVNDLERELREEVNEFYDDLEDNEEDIQLM
GNEEKNLNDDD
>SR_DN112801_PRD_AMB
MKTKRSNFSIESILAGDENQHTPDAETFGQEFFRSREKSPVVSQEIVQNKLTPGCVAKGRSRESDHLSEHEETFSLLETSGDQTKL
HLTTTETSTKPTNTSDYVTQTADRTVDLVTSSSNSNSCRRSRTTYTPWQLHQLESSFEDNPYPDVGQRDYLAGVMGVTESRIQ
VWFQNRRAKRRKLARVFHGDINEHQLNPNNFHIHHSQAAVSTSIFLPPSHLAEQPMHSINRGASNHTLAYSNMPSCKDRVSSI
TSLPTIDDRLSQESIRGRILHHPGFKNGDWQSDLLASQIPRMNGTFTNNPFINFPTQGRPIMYPQLLESSVFPSINE
>SR_DN113036_PRD_AMB
MTLPNSGQLPFLPGMQHLGALFATNSQAATSNIPSGSQFKPSDSPSVSGPIFTLPLNPNQSSRGVLENAAAAISLNAYNLMFQ
HHMKLLQQQQGGQNFAAERKILEQREATNRMGKQIQSSQVKPPKKGISLTQKDKPTKTESKLSQSAKMSPANVNSCEPLDYTI
GRKSNEVELDSNLDEGDMEVDMSPDDLCDPNDSKRRRTRTNFTNWQLEQLEMAFHRTHYPDVFVRETLAMQLNLMESRV
QVWFQNRRAKWRKMENTRKGPGRPPHNAHPLTCSGEPIVSTNSGSNQGAKHQNQHLSPTETQKQNKALVMSKSILTSKESK
IFQKEESKPMPGVCVANPKFQIFGDKFSFVGPGESNSTDVKKGDRLTLFESTVDMITKRCAQQSFDNNLRFVSPTFLKSSTSPN
MEKMNCHECDDQKKSFRIDSLLDL
>SR_DN113132_PROS_Prox
MEFKTNSPVNPTDISIPIIDASSELKEFIAKSPAEIDASILTNDDQESEELSVVVDDMETEPVIFPKISDQDIKMDPEQPSRNRFTQ
NEFFSDKSKSFSEANSHKIANLTPSCGFVNQSVYNNQNILEQFIGNWAVFHQMLHNKQSAPQVPPNNIRSAEVLASLIQATNLE
MERSRQVAKLQQMSPRLSGALLQASALGGINPGMMQLQPQNMPFPQKTMSFPTFNQQGFQFNSAFMPMQTPGLPMMH
KLATHGKLLPHGYLNPVINAAPFQTTQLFSTTSGQQISPNKHLNQVRGLNSKIIKSTAKEQLFPQSLTKSTISKLNSRTANASLDQK
QNGRLAASCNNFKTPLRTSKGNCGATRLLKEMDNSMMSGVSVVQMTENLTANHLRKAKLMFFYQRYPNSAILRQFFVDVKF
NRSNNAQLIKWFSNFREYYYIQIEKFVRDTLQKHKNKFLSNENRTHSSYSSASDLTNQNKEFSGFSPKRANQNRPFDIESLIGDH
EVTNSEEQFEELRQALTIDKKSDLFKNLNTHYNKTSNFEVPDEFMNTVNVTIQHFAVAVSSGKDRNSGWKKQIYKVISKLDTNIP
HSFRAALGSQLPAQK
>SR_DN113631_PRD+POU_AMB
MSNESLRPISSFLKPLRTSELYPQAETWSVILSQDNDLQLQQSLETTNLVDLPGNRELIEHDYVDPPRPFDIRDEHPYSIDLNDGE
AKFKQVSNILRRYRNIQPGIPFSRLLKMPPNDKSKKKKNRKEKLCDMGPLVPGSLHAQAVMDAEVQIEWIMKARVHLKLSQAE
LCTYIRRNYNIVFDQIQISRLEQKSLTLGKAQEYLKLLYSFLVDVKRTVSPHVKTKIAAQSSRHRCLFNSHQINLLKSLFEVNAFPKK
AELSILANHLRKDRESLRKWFDKRRLNFRRQGRNDLSNGTFDQWKNLLECDFIKAESKELEEQTMSCVIAEGTV
>SR_DN116171_PRD_AMB
MNSIGICSDVPNLHHPISISTSASVEQHILNSVAKAGHLPPNHPLGNNNNPKQLLPLSGTTPQQHNSLSAVSAASPASSEGSVPT
PSCSPVKNSPQGQSPVKLSSNNSGGGGGGGTVVPKPKRHRTRFTPAQLNELERCFNKTHYPDIFMREELAMRIALTESRVQV
WFQNRRAKWKKRKKCSPNVFRGGALLTPGSHPMGFGQFAAFGGDPFVCSYNPFAPDPRQWGSAGGAATTPNNSMTSFTF
PPTFGNPRHSGIGTAGNSINQAAFGFGMETQPGNLGLQSGHSLAPNSPTSVTFPVLNQQAQSFANAMMCANNANVTANG
NIIDGPGSGIPQLGSSIPSPIPVGASNDPSGLMALRRKALEQTASMNSFSALSNCGFDLR
>SR_DN116214_ANTP_amb
RSYKSKSQANSQIQASQLRTTNGSSVYDPWVFVLDTNGQIKELIFPRGLDLDRPKRERTSFTPEQLFRLEKEFSLNQYMVGRDR
GQLATCLGLTETQVKVWFQNRRTKYKREKQREQDTKKSDMDTLAACSMFKILE
>SR_DN116853_PRD_Pax4/6
MNMPHKGHSGVNQLGGVYVNGRPLPESTRSKIVELAHSGARPCDISRLLQVSNGCVSKILGRYYETGTIKPKAIGGSKPRVATQ
DVVNKIADYKQECASIFAWEIRDRLLQDQVCKEDTVPSVSSINRVLRNLSNSTPADHNHNLKHSPGSDSVANVNSGTLLHTPGS
NGAQNSDDATDAATAVANGTYKIMTHYRPNQSPGSNNGWPCSNGGGGKVTPGASPSPFYSYNISPGSQYIAIAAAAAQAAV
QGGFGSCDKKTMGMIGSDMDMEGGVGLSVAASEEARMQLKRKLQRNRTSFTPEQIEALEKQFEVTHYPDVFARERLAGKIN
LPEARIQVWFSNRRAKWRREEKLRSQKQHHHLHQTPTGNTPNTMDK
>SR_DN116892_POU_AMB
MKTFDLEEEIFKNVDADPKDIVPRLLATIKEKVSSSDEEARNMLIFSQQAVIRSLHGKIGALIELKHEEKEAFIDVDQMQQKIEIPV
KSLVELETTLQDVAENTQQPLNVNPIVVLQSGDFDPSMGYPAIEPSTMQQLLILAPTRVTPVCKGPLKLKNPQISVACPVNSDID
TQNPNAETRSAGLNEKDIDMIARKYMIPIPASDESSKLEELESFAKDFRTRRIRLGYTQNDVGLSLGKIYSSDFSQTTVSRFEALNL
SFKNMVNLKPFMEKWINDAERHKTLVEQDTDGLSLEDQKTLEIFQSSKDNLGHRRRKRRTCIDVKTKKALEQFYNDKPHPTHE
ECCAIAEALNQDRNTIRIWFGNRRQRDKRLMQLQNSVFAPRASFASRPSMDVFDNREVVSLDDSMMPTSFDSSELQTMPNF
SQSEAPIMPQINNSEVSTVGSHNNVLQPSESSAFSTPQQNNSAFRPVSQSTQN
>SR_DN117560_CUT_Onecut
MDLISEINFASSASLASENVNNGVNNDASSPTGKNNHSFLNIDHLTNCSSTGDIQQLVGGISIPPTSKQDITGIQQNPSAVCDSA
SFTNFLMCPDIGSQFCSSATQNSFNAWISNPPDKTNSGLNFGTSPFQTNTSANNLYNNVSCIPTSAKFEFNTPSFDFNAYNPAS
AWAPLDMYSNSNMAASYQCQRNISASALSACSPAKFAATYPSINVNGFYYPTTYATPYTRYSRNLLTKDGEELDTRDIAQRITTE
LKRYSIPQAVFAEKVLCRSQGTLSDLLRNPKPWSKLKSGRETFRRMGKWLEEPEFQRMAALRIAACKKKEEEQQVISKPKLSKKS

RLVFTDIQRRTLRAIFQETQRPSKESQVQIGKQLGLEVSTVSNFFMNARRRCGDRWKMDDNCGSKGGPGVRQSLTSGVVQVP
PSLQHHTHTPTAINITTANNLNVITGLNCPTIPTL
>SR_DN118344_ANTP_Hox5
PYYSPASAVAAELGQHYPHSLLPAAAATLSSYNPYYPSMAATQMYQSSGALQSNPALQQITNSATAGSNPMNWHFGFLPTSL
DARSMGANISHHASSGSLKHSQINEAPHHFGHQFKFQGSNPGLDSSNGLMMIASGGEGNGSELDGGDHMNTCHGQISPSH
NQNGSKVFAWMSKRPPTDCKRTRTAYTRFQTLELEKEFHFNRYLTRRRRIEIANLLALTERQIKIWFQNRRMKWKKDNNLKSM
SQIDSITTA
>SR_DN118733_LIM_Lhx1/5
NNNNNNNNNNNNNNNNNNRSAVATEAQIKSPTQFEAHLGTFRTNRANAGRKDTFEDTLTQSEIPPIGDLRKQNPNDANVS
QDIKSEPIAAYQESQNAFVLTSRADFGENNIFDQNSLLFCRSVIDHPENEQNTDEILNKSSSDHELNESMTKNDDILKTSTDEMIK
AEMNKKSSCEMPQRRSSLGNEPNSVLDGYGSDNCGYNSDDEEKLFDEQDEGAEISDDNKSDYDDGDFKSCASPNKEGSDEVI
GGGPKTKSQVGSKRRGPRTTIKAKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGARRA
AFFQRRIRGPFDPNPHNQEFLTPQIPFFAHQDLNNPDFFPMGQPFSYDFFPGVHHSSIGQPPSSIGVDNSEFPMPPTLPTPSSLI
DPLRLHSGRSEERRVGKECRSRWSPY
>SR_DN118860_SINE_Six4/5
MCDFPVSQSNFYFNREQIECIAEVMFQSGDVDRLTRFLWTLPSTELIHGSEVVMKSRAVVAFHTRNFRELYTLLETYSFSLKNHD
LLQDVWYKAHYYEAEKIRGRALGAVDKYRIRRKFPLPKTIWDGEETIYCFKEKSRNKLKESYKKNKYPTPEEKKALAKNSGLTLTQ
VSNWFKNRRQRDRTPHKSGDGNSDSEDEITIDHKGKFISLKHHANNPQTTTNQQHLNSSISFSSTNINNKSSSLNINVSENGSN
TKRNLSSSNSSHTFGKNKRHNDVNAENIPKKMSGTHVKNETSSFFARAHAISPSRNSETIFLNTANTGVNYPAALGLFASNGNV
TNSGKQLSEFVPSGANFGSVLAGSGMSPYHNAATAAYLQSFAKN
>SR_DN118878_PRD_AMB
MGSGGGGGPTPARRRHRTTFTQDQLNELEAAFSKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQEKQLTKAFVSPP
SASVLAPPGGVACNQMMRNIYHQATNGARPTGFPGNPMSGSFHYQPPPMLTPISSQNGPGSNNLSTSPCSMSSRYPMSGA
GGSPGSA
>SR_DN118899_ANTP_Emx
MSTLSSMALLDYFLTSQNSMQQNSLLAANMLLNRQQQSQSRFMTSSPPFTTPQLLTSSYGGCGDPLLSMMGGTHAGGSLNS
ALASMGSQCNSVPPTHMTNASHLHPSIPPQLLMNPSHLPPHFLSNPAAAMCSALLYPHRSKRIRTAFTPAQLAKLETQFARSPY
VVGHQRRQISEELGLTETQVKVWFQNRRTKEKRVLHPEDESILEQESDDECPLTDDDEKSSNTQKPCETNKPISISNKKQETEKK
NMNENEQKKFSALEELLNNSANKELDVEGKLSVNDLEKIAESKFLGCSSEKCSV
>SR_DN119820_ANTP_AMB
MEQNDRNPKKSFLSIDDILSTKEKMNRSFENKIISLQSERNQHSEISPIYNQADILGVQHNADNQNISPSFDMAALAGSNISALID
AVSAVAPFGSLNLDNRELIGMDSVQSRSSAASGVDNESAKYAWSQTNTNQADGAEFSRERHSNFLLSLQQRKHSGPGGGGG
AGSKRGDQLNGRRQRTSFSAEQTIKLELEFQMNEYITRSRRFELADTLGLSENQIKIWYQNRRAKDKRIEKAHVESNIRRQLNNL
AFVGAANPFSLIPSGPNPMNSYIEELSALERLVSSDALSAERLLLANFQIPTKTAIQNIPSSMANKTIIENALSFL
>SR_DN120054_PRD_Gsc
MSSFKIDDILSPPNGFAQHLPFPIKNEHDIPTANSLLGISGLCENNNIVFSKLPLHMQQFLLESVLKNQQFRQNELLQMQQLQHS
LPVSILNSVLQPTVLLADNNISNNKSNNSSLAALVNTVDSIAAHPQLSGGISISPSAHHNHQQQQNNLTAQLNSAFVNQQQQQ
QQQPPFNFSKISPNSGLDVHTSESQLRSEDNHVAYLLKHLLQTHALQNSPRTSTPQQQQQMISLPLQQQDSPQSTLPQFISRPE
LFRNAAAAAASAVATGGTVMAQNSNTAADDLHQQMMTSPGGAGSAGTLGVGVGARTLGVAAAHMQNVALWQGALMR
ASLARRKRRHRTIFSEEQLKELEEAFLKCHYPDVVTREGVAQKLSLKEERVEVWFKNRRAKWRKQKKDAAEIRKHNKTAMLKM
KGKETQQHQPTTIVNTDQPNNMFEKTSPNSNP
>SR_DN120882_PRD_AMB
MRNYKRTRTQFSEDQICALEKLFEESHYPDSSARNELVAATGLTDSKIQVWFQNRRAKARKEKYVAPLTSCYCKKYHNAIEWEA
IPQQSFSSGVAADVQISRDHRANLRSCRLFDQDNIRELTIVSKATSTLDPFSMPPVEQLLPSNVKRSRWSK
>SR_DN121592_HNF_Hmbox
MFSSTNAPAVTNSPTLIQPPTPINCPNANSAPNSPPDLKYTIEQIELLRRLKNSGLTKNQIVQGLSSMEKIDSTGFSTGVEANSPP
GKEIKQELSPNSKGNLLPGTIGQLGSLNSATTGAEYLRSLALQHAFFNAMTNGNFVNGNAHMQQNGIQSSLAQRNLAAALMC
SQGMTIPGIGGLHNQAASPASAGLAGSLVGFHLGSTNAAENSILNQMSHINGGGGANSPVNLKRESMGGVNASLGDPTTISD
DDEQLMQLQSMSVVDTKEEIKNFMMCHGIKQHQVAEKTGVSQSYISQFLHQGVWMKDSTRTNIYKWYLDEKRKREASGD
MVSPVHSNSSRGNSPINEGSNISQTPPTSNTITGQQINNNNNNNNSPVGTNSSKNGASDGGSNGYSAVGQPIQQVKKRDRF
VWKESCVCLLEKFYHVNPYPDDAKREEIAQACNQLIDSTGQGHDEKLVTAVKVYNWFANRRKEIKRKNHFAEQQQAAAATN
HNNLSLLATGHPAGVLPFLLEGSNNKQQQPHPHSLPGLNKNNPGGHLNPQVGTNPTAVVASAASFLDELIEHNRSALQHNQ
QQMQIKNENKADEYGGEDKAFNMCQNGNDVDEIDEPVAKKPKSSLAEQLAEVNSVISELVPGNLENSTNDENDGLRNSYAM
TETHSTHGETETGQQATFDVLNLTSENSSKI
>SR_DN121921_ANTP_Nk3
MSHQQGHHHHHLQHRPKFMQLEDHKGLLGSRSHQLEEKMNLTKNEGVESDDESNEGSSCNEDDEDLLPSPSSTSLISKYQSS
LTPTSADLFLKQSHDCHQETKSSGISGTIEQYEQCLSELQQGGNGGNRRRKKRTRAAFSHAQVFELERRFAQQKYLSGPERSEL

ASALKLTETQIKIWFQNRRYKTKRKQIHGDILPSHCGPLTSSVGENNPHLNHLRKLSAAQLQMALTSQIANSAMRGTQITNSVP
VTSQSINQPALAPPVGFDMGSNPAVSAASQIFANETSAQKTLAQMQLELLRASTVGVNPYLSLLCYSAAAAASGAIFPQTVQP
KAQTIGIGNAQQLANSVSLAQQLANLTGQQNNTISLANVLTGITQPPSTLTTSNALFSSSLNLTCDNILTQSTVKHGHDTKLTMD
SARPDLLLAPVVSTNNI
>SR_DN122029_PRD_Alx
MVINSSSFYKPVLSLTGHPILDLQQHSNISSNSFVAPGASLSTNPFCLPSSGAAGGGGGTAVAPYWSPALIASATNHSLKSDVEM
SKIGVEISDEKLQDKQLSPLLKSEGGATLDRSRDHQTGEGDEFEGEEDFEDSDSIGGDGASTRSRKKRRNRTTFTQGQLEEMERI
FQKTHYPDVYVREQLALKCDLTEARVQVWFQNRRAKWRKKEKHQQLRSGFHSSGGGGGTPFNPGGGNSVAEASLAAHRND
VITQQLSSTWPYSGASGAAPGHLAGAAGSTLLGGNSANPYSSLSASAASNPHQQAQAAAASFFPNYIMAAQQVAVAGYGCL
PQSSVAQQMGSNSMVPGSQFTPSNQTPQFPGMSFLGSGAFAHPSLMSAAAASSFFSSPAGAASAQNTANAHSATPSDLFTT
GGSAATSVPMNNLDPRNSFANFRFNQHKSLDGRW
>SR_DN122139_ANTP_Nk2.1
MESGGNNWNFNSMPNFMGAGGGNSSAGAAMQYPSNPSSFNYPTPSYMQQVAAGMIPSAMNAMYSSSPTGCGGAGMS
GGASVHSGVNHGSFPGANASGAIMDQSFGSGASAIHAFDAFARHQPSAASNWYNAANNPSLAFSNLMRSSAAAAVAYNP
MTGLSLTGFPPGANADPAKFFGLANRRKRRVLFTQAQIHELERRFKQQKYLTAAERESLAHLINLSPTQVKIWFQNHRYKTKKS
KHSSPSSGTASSQNGNSNGNQNDQQGDSESRGTSHGLNNQQQYMQSKPPVKRQDTEVIIKNPQVDDKSSITPILNMSMNP
KGPENYLNDLDSQGDLPADAMTRQDSSSQIDVSANNIVTAAQLRSVAAANHNSLLSFNSENISAPQFGSHGMY
>SR_DN122619_PRD_Pitx
MASISASLGNDSNTIYSSNNNEISSGSQNQTENNPDDNCTATETEQSSTKISANHSPADDVTAHDDAIDEDDEDKKNKRTRRQ
RTHFTSQQLQELETLFARNRYPDMATREEISAWTNLPEAKVRIWFKNRRAKWRKKERNQLQEYKNAAAFGMNFMPAPYDPH
DFYSTQLTAGYYNNWTKMTAAAASPLAPGKAPFPWSFNPNPAAVASGTPAAAMMNAGMSSLNPSPPCPYG
>SR_DN123001_LIM_Lhx6/8
MTKGNISSIFPSEPADTFNSFHLKISKFDDFANPPVPKKPLLTIEEEDPILDPSQFLGTTGLESQNQSSKQIQIPSQQCPPEGVGVE
TCQSCRKAIRDQYLLHVNGKAWHTGCLRCSACMTHLGNQQSCFVGKDAQLYCKSDYSRYFGVRCAKCQRVVSSTDWVRRAK
QLVYHLACFACNICHRQLSTGEEFALHEHHVLCRQHFMELLDIPPAGGGASGANGAGTTKGGSSSGVLVTSSSSSSGAGGCSG
NGSDSNDRADCNSMNGDGPDEPNGCNDLDCGGSLGSPHSHKSKRSRTTFSEDQLKVLQANFNMDSNPDGQDLERIASQTG
LSKRVIQVWFQNARSRQKKHLSTGSSGGNHMNGQQQGTNGGGPPGGMMQGGGAAGANQAAAMFAANASAAKLHHHA
LTFVGPSGFDDVMSPNPVGLM
>SR_DN122932_PRD+ANTP_AMB
MPNRFKFQGPEGSIIMEFDDNCTVFQVRTLLSQTIDRKVLGLVAAADKLSGGSAGTNQSQHSVIRLLDSQRMKEVDRKVVMM
LECEKLAMMTSEGVNKKEPSEGQEAKSGEMETSSTSNDHDSKETSSTGNEPQNTSPKTNSHQSKSPEEKREPQATLTQQPFPS
PFLEAASAGMAPLPFNFNPFLLVPAVLKYCIHFEIDKFNINAEIFSHATMGCNAATMEKVLLLPTGEYMDGSGKFLPIVRHFMQ
CEEGRAKIYSRVKTELKRVAHNAAAATASSGNNGGQGRGLQGHTLPHHSHTSQETPSNQLQNGATTSSEIQSFPGANKQINN
NNMMQHNLNSLNDLFPNSRNTQNVFGALNSLHLLGSMQRDSIDTNPTQNLTPTPASVLSVTSLANATTAGVASNQPQISPN
SSSLSGKASSYFSTSSEVAKAMQQQSSAIDSTSSGQNMKRMRRRTTISRSAKQVMFEWYLSNGRYLTADGRSLLSSHLGLTEEN
VRVFFKNLRRLEKKCEDTGVSLYDALNRSDSEHSSFSDDLNLLTQQSFTEEDEDSSPTMIKLETIE
>SR_DN123029_PRD_Otx
MNLYPSAAEMSYLKQVQQAAAVASHHPAYMQPGAAAHSMLSHHHHPQMDLLHHPMSMNYPVPGAHHQMQGANAIPA
GRKQRRERTTFTRTQLEILEALFSKTRYPDIFMREEVANRINLPESRVQVWFKNRRAKCRQQQQQTQNDKTDKATPSPDNSD
MKKDPTDSQLTPGSQMNSQLQQDTGSPTMKVENVIKNELCDNLLKKTTHVAPDFNDLGNVLKQQQELSSPQDHDIVRNGVL
GAIGSKDMNDISNIARSESSPFLPSVSSSFINIWNPAGAMHNHSSVSSYNPDMNSAAGLGRPSGIPNYNQAAQTSFYPSGYYSF
DTSYAPYFNNSYQTSAAAGAALSAAAASRYPYSVGTDDNTNALFNPITSSAGGLTSEIKQDVIRNPLNDLPPVWKGY
>SR_DN123138_POU_Pou3
MVITSSIHDSNIVASGLVQPIQQCSPNCVPSTNTAASGIGVNAFGESIAAATFGIPTSAPEYSSYLSANMSPIMQHSSAMQCSAV
SATTPNVASSIHPLHHGSALNSSNTSPYGAFKQESLTMTPGVPYSTCLTPPQAASHQAAGWLISPESNLSWNGTTFPYQSSPQC
LIQGGSFFPNGNSAASGIYGAQSPAAINHALSNSNTWSHLAAAGMHPFNGYMTPSRLESPPSNVAVPPYPLSTTPNIHSYYAA
AAAVHANQSLLYGRTLEESALLDCMERECEETPSNDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLS
FKNMCKLKPLLTKWLDEAENNNGYPSALDKIASQGRKRKKRTSIEVSIKNALEHNFHIQPKPSAHEISALANSLQLEKEVVRVWF
CNRRQKEKRMTGICLEDPTGQKHEDMEHGMICEGESSSDEDDFTRREDSDLDERSSGGNSNLISNQVDPKMYHVTSLTGAGY
EYPNEAVTSETLSIMHAKGDISNMPGFSANQSQDSLVHQNGTSNSSNPTGIHVNQLNIICPGMEYVKDDNRSSNHNAFQHSE
KERILLQQLSEQQSNFNCAQYGYQFAN
>SR_DN123163_LIM_amb
MKFISTTAGSGNNSCSDVPIPAGATDKRDKFTPICARCSHDITDESYTKTLDRCWHNSCFVCCACNNKLGPKCYSRDGFVFCQ
MDFRTQAWTLCSGCNKSITPDELVQRIHDYIFHVSCLTCSECKAHLQTGDKIFIVHGYPCKVYCHNDYKQHVIARNQNESAKV
MPAGGQWVNTAQQQAYNQSFSHRPPNPDYISSPSNQESRGVFRPTPLPATSKSHPQMTVGPPASQAIQPSNLAGTMPSNS
YPCQPIAQHPPNYKQEATAKQLASQQDPQQFKQNGHVPNPNHRFKDEQMHSPSSQTAIQANFVTGYTDNTAPPHQQPNTT
QQQLPMPVHGHNGVGHNTTYFQGQSMQQSPAQSSDSNSIDMMDYRQSNPANQPPPPPNRGPPGVLNSSSAFSPCPPASI

VIQRQGGGFAPPQTQHYPNGLAAHPKMVAHIALDDEKSSDNVSVSSENDLQCSPKSGASMGSSTSGGTSSNPNSVSSAGTA
NTPPAAGSFSASGTGTPVSNKRGPGNHLKGPQADYLRSYYLTNQKPSKEERDAIAAEIGVEAKTIATWFQNRRSKEKIRREKAA
AAAASAGVELASENSQSSSIAHSPSASSDTNQPASSMAPSAPVLTQESLSQSMAISGHNLQAIQEHDPTSPAASNLSESSGYAS
YVIKDTNQPITSDYNMNAPNPPQDSDAFHQSNFGFVKMENNQQYTINQQANFDPNGRQVNGQAAGFPATEYYVTDDVTN
VAQQYNAQIVMQSAGERSNTNEAYPSVQSEQFKNFQPQNHQIGINLPGSAQSPGQFAPGAPVQPQIQTSAQQSINLNMGQ
ASTNFVSAQ
>SR_DN123214_TALE_Irx
MQVHRNFDSSRFNLANAQQQSHILQCQLDKVMSEAVGGGAASSAYNPHPSPQCHNISSSSSQSKVCRCAACMNTNSNSGA
KVLTREEEFILKYGLLANNNGSSQPPSSSPNSATAVKENTYPLRVWLEENRDNPYPTKSEKLMLSIVTKMTLTQVSTWFANARR
RLKKERRPEHEFSLHDFYPFPLYQPIIAPHSFYPSGPAAPHIPKLVPQLVNHALLAPPPQPQPQAFQPAIRVLPANQNQSSLISNSI
IQEYLMSHREALNALRSKSSEDNRFIRSAGIETKPADGQERLLELILKSNNKPAEPQPRLSTISPNVTTAGARKTSEKEPVPPKKSN
EMPRNIWSIVETLKKQDNS
>SR_DN123370_ANTP_Hox9-13(15)
NNNNNNNNNNNNIVQHPHNNHTTYDHTNQTAFEFGTSFNSQIQNNAIIHTNNVNFVSPSLLGATTNTSISQSNQSTLVTAVP
PVNQSQAGTPITPTNRNASWSTTNHNTAAADQRNSLAPNVKTCIKSEASLSQGSGSPPRDVSGGSLGVGGCFNSAESSPADSS
TESVGNIMNSPEHAGSISPSSLYAQMTSDPSGVLQGNQQVMSSSGHSQGSHNHHLLMTPGNTSVPPQTPSNGGTNPSSGG
GMAWMARNVSRKKRRPYTKNQTLELEKEFLFNTYITRERRLEIARSLNLTDRQVKIWFQNRRMKNKKQHSGGPGMGPPIPG
HPAMHLMVPQQHPPHHML
>SR_DN123538_TALE_Pknox
GGVGMGVGGHQEMNPDGSFKKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQ
PMLEASNPESVNKTKKNKNNNRPTAVNRFWPDNLAQGQITNSFKGSGGGGGDDTSPHSDDNEGESPPPPNHSHGHRGAEE
GGTTGSGNNNNNHHHLSPNNRTTLGHEGIVMNGTRDNNSMNNLQLLNSPGALSDDCSDDQSSNEVTITPPDRKSTRLNSSH
ITISYAVF
>SR_DN123563_ANTP_Dlx
MQNIINQNHAQLSPSSRESHLRHDSSNELLVGKPPPGFPYNHPDFPQHGSAGFPYNLGGLSQHQTSTIQHHPSPHQSSGGFN
PYMQMAAASGHPGNPHNPQNPFHHAHPNPMMNYAMAVANQDPSFLQHHYRSHAQAAAAAGFPGYLASQMYPQQSGS
PQGPNPYSGAAGAFANQNGLGRGATLTYGSPPPPNHMLDHSPRGDQVLPGEKDLELGDSPKSQQQHQAAHVLGGLGVGM
TGGVGGGGVEAALTKGGKKMRKPRTIYSTIQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKMKKLMKGN
PNARIDMNFDALSPDAKDEEEGDDLNETDEHPIDHNTTANDVIDATIAHSGANQIAPNVTMPNLVLTGQPMSHQTSVAPPG
QPQSPWNMTMPHMGPNHGSHQNPNQHHQATGHHVSPHSMGYLDGANIVEMIANSRATMEAGGAPHIKVEMPPGLDM
TQQGVVAMQHHLGIRPNIMHANPNNQHPDWYHAKGGLPAHLLPNGHNLAQ
>SR_DN123682_ANTP_Nk5/Hmx
MSETLSENKTDLESPTLGKSGFSIADILNSKTLNLQETLKNLSKDLTIATSGPELSWGKNFPISPISSIPVSGTPTSASKNQFPGVPGL
HNNPALSGLHPMLALPQAEVLLQQSGLLKGLSGHAGSPGVGLPPVSGPASVSGLTPTALGGQFVQPADIQLNESAAAALLRW
YSPWALLSGQNGMMPQNSSPIFQQHFANNLQQLTNNFAQKHANDGVGKLPLMPGFVELGAGLASKSSGVIPPKTSQYTSSI
HTPTTPLAALSNPKSLDITSVVSTLAASQIRGGDHSSTGNRANNQSNYKQNNREDSNNSHVRDFVGHVTNHVTNQSIRGSKSA
TTVGAHRGDGTASHQDDNDDDEMMNQDACSDLIEDSDDVIMGGMGGVNDESGGDGEDELDDEEGEKRKKKTRTVFSRSQ
VFQLESTFDVKRYLSSSERANLANSLHLTETQVKIWFQNRRNKWKRQIAAEIEAANASAVQVGSNPGSATAQRLVQVPLQMY
GHHASAGGAPNPHTVLPNPVSSVSSADLQGLKNSLLTSSNGGTNQAALIGQLHAAAAANYQAGIAAGNPYYPSSIGAVMAA
MARQHQHQQSQQQQQNLIPVTTGVNTTSSSNILSSVVTSAS
>SR_DN123752_ANTP_Mnx
MQSETTRVAQDFSGILNSQVSQVSNISKFLIESILNVSAKEAASSQNLQILNAIQGNLADSKSADKNNTEQTADEIEEDTTDDQS
KNKTTIESFQRATSTPKIEELPEKVIRKELNSSFSPNAENDNLSPVRSSENKVSKQKEELQNFQRNLLYSNFGNGGLNGIHQYSGV
PGIAISQNLQNLYKNASSSLAQFTSTTSPSSLVGTQPTNQIAADSLSLFNYLQQAAVLSSSCNPATLLVQSGHVTGAGSVCSPNQ
QMENYSRLLALGMGLDQNNGFVPSFHHPGPLSSPLDGNSSGLGGLQGGNSKCRRPRTAFTSQQLLELENQFRKNKYLSRPKR
FEVATKLMLTETQVKIWFQNRRMKWKRAKKVLSSSSTNNNNLVNNN
>SR_DN124757_LIM_Lhx3/4
MENQQLVCKNDYEMAKSKAEFEDLNGSNKRPRTTITAKQLEVLKSAYKNSSKPARHIREQLSKDTGLDMRVVQVWFQNRRA
KEKRLKKDAGRRLAVASNSQFYASGSGNLTPTSGKKRAKKTTTNILGQPIEIPPGSGFTDADLMAAHSNAAAAMQQHGLFHSS
YGPQISQPVLFRDNFSIPGMQGLPHPPPNHLKPYMPISTGLLSPIGDIGSVAASVSASLANVLPNSHHVDSRFSQLMGAHANP
GYPHIPDNVPTPPNVVNINSFASPNSATNGNSAAEVLASRLANAQVTPPQSAASMYQSAPNASPHLASPIFAHSLNAPPPSGL
VFPHQEPISSLHSALTAGNRNYPPHPWLQNPCKIEPPSANSNIGGATSSSHLVVGHNSSNQLPFLQHNNPIQPNSHQLEGILM
W
>SR_DN124792_ANTP_AMB
MDSNNISEIDTATNSTPAKSMKFSINYLLQKTESTVQQPEPNSEQVEFPIYNPTSGAHHLMSQFAGAHATAGFYRQISQFAPLN
SSLQSLLASQERAKEEEKSVSPREWGDKMAYMSPMKSPSKGCKRSSSADMCLNVQTSQQSPKEDLSPNDIFGKSPSPSDLKHS
PPRSKHSPGLSGTHMHPTSRKLGRNPRVPFSSAQVAILEQKYQVKQYLSSIDVAELSTMLSLNDSRVKIWFQNRRARARKEQQ

MGKSLSASANPNVTITKNMSAQLQNLQTQPNRPAGGNDASAPGSQARWSPSMQSNFNPQFSSANVHNNNNNNNNHNN
HNNPMRPQLNLQPFSPFLSFCNPNANPMTSQLLRPPIGQLLASHNSSHIPPNAFASSAMNLAAQLIQTSRSAQNSNLVDSN
>SR_DN124898_CERS_Cers
MAGVGFWNERFWLPNNITWADLRDTPTDKYPQVEDLALVFPFAFVLGIVRLLFGRLVAQPIGACIGICPRGARKPYPNAILERV
FLTITQRPDEKHLKGLEKQLDCWNQYKIKQWFALRRAQLRPSKMSKFKESLWKLMYYTGIVGYGLYVFSEEDFFYDLAKCWKN
YPFQPLNPSIQNYYSIQLAFYVSLLVGQVVMVRRKAFRHLVMHHVIAIVLLVLSYVDNFIRIGCIILIAHDIADVWLETVKLLNYTK
CGRLTHAVFAIFAIYFVATRLVYFPLWIINSTLVQSLWLPNVGFFPAYVIFNALLLSLLALHVYWALIFFRIGLGMFYSSNSCSATQL
KLKSQTPPMHQQQPPLCNTTQPTATSSITHGFFGGVFGLGGFGEAGWDKSSSALYSDSEISETELNMSVGEFSPTYWFRVMP
AAGGSPAGGVNGTATSGSGATLNGIFPSNIGLNMTTHQNGGASQHHHLNNGAMNSASATANNTDKGTNSNSGNMTTVH
HRRNNSGRNKSSNGGGGCSSSNSSSNSSSTPITRNANALQVQNGHVGGAATTANGHMCH
>SR_DN124898_CERS_Cers
MWSPHPRGLCNIRNLLRCNKVGLLPPLDHQLHPGPVSLASQRWILPCLRHLQRSSSLPPRPSRILGTHILPNRPRHVLFFKLLLSD
SVEAEITNTAHASTAAPIVQHHSAHCDK
>SR_DN124972_ANTP_Evx
MDSFSISNLIKREPVKKMTTESANQFALLNEKRNNSFDQLHIQHTFNRSADVLSEKDGDGDGEGVGTMRKRERMESTESHSSS
DLLLQSDRDVSSSPLPAHDHLSSANASVLGGGGGGGECSIVNGIRRYRTAFSKDQQDILEAEFRKENYVSRPRRCEMAAALNLP
ETTIKVWFQNRRMKEKRQKQTLAWPPFFDPYYYGYLVHRGVIPSQPTPPNPFELQGGNHPSQHSFMPPTAPSSLLAPNNPTG
FKFGHPIFGQNMPLNPNAGALNPTSLPNIPSLSHLL
>SR_DN125302_LIM_Isl
MVISHFTPHLNFGHFSMDFDSFIPPIPPNFTPGSPAIIPSSPQIMFPPVPGSNQNPNLFTRTQAVPQYSHKKETSPSLLAQNDSSL
TPSSLSPPFDGTVGAKKESGAMLSDLSKQHPISNKSLSPVFDSENVLSTNENRESNQSSAGCDNSSPSPLSSATEKIPKCIGCCEPI
LDQFILRVAPDLEWHAHCLKCHQCSTVLDERQTCYLKDGKPYCKLDYIRLFSIRCGKCDLPFEKTDRVMRTMNKVYHVECFSCN
VCFKQLLPGDEYAIRDDLIFCRDDMVELMTSSAAANANQSPNQNSLPTCENPSSAGDPRIAGIPSEQNSPKYGGMQNNSAGN
NLSLSLSSSNDLDRDDKIIFSPDDKSCKDSAVSLSVRGEKRHHSSKDGSSKPTRIRTVLNEKQLHTLRTCYNANPRPDALMREQLV
EMTGLSPRVIRVWFQNKRCKDKKRSIAIKQMHQLQAEKNISGLNGVPMVAGSPMRHDPVLDGQFPPYPHPPYGLPHHPHH
ASEMDPAYMMAFDNGGGGPGSVISDGCLGSPAGSDCSLSLSSD
>SR_DN125354_POU_Pou4
MFFPTNSISLVGASQQINHVSKNSMQSHTSEHLQSLVDCKPIMHNSATADQLQAAAAAGLHAQHGSVTGSAGGSHGHNEFK
YSASPFQIRQRTAFAPPPTSNFFGHPFDDSFFRADSLTSNHMNHMAPKPMDYKPHEYFSTMMADMQNNAARCSQLDLLDQI
APSVQFGGGADQQVNSFHPATHIGFHPSFGHGNMAAAAGLAGPSSFPQITHVPQAEVEADPRELEAFAEKFKQRRIKLGVTQ
ADVGKSLSHLKIPGVGSLSQSTICRFESLTLSHNNMVALKPVLQAWLDEAEAAAKGRLSEKQSSCLSVVDKKRKRTSIAAPEKRSL
EAYFTVQPRPSGEKISAIAEKLDLKKNVVRVWFCNQRQKQKRLKFAAGR
>SR_DN125451_LIM_Lhx2/9
MTPALQNPTSIPPPNLGTPTCPISGLNLSNGPPSLNELASNPMFPGFVHASRLGASNIPIPPLESPLPPCFVPSSMSPPVTNGTNP
GNLGSRPAQSTCHGCMAPIVDNVYLSVADFTTWHVPCLRCSDCNVGLDSEKSCFIKIGKYLCKKCYQRRFKSPSTPSPAVCARC
NNKIHAKEMVMHAKEFVYHVSCFNCTHCSRILTTGDHFGMKDSLIFCRQHFEMLSSPPLNLPGGGSHISHSTGNNPLHPGLQP
PLPQDFNHLPPTSGNPLSAIQSGFNNQPPCSPFNQGHLSQLPLGFPGPDGMPPNGYPLPQGDPNAMMNHPNPPPPKSQKG
RPKKKKQPQEDAFQNLYANDFNSGMHGAPHDAVLGLYMAANNPNGASLEQQNRTKRVRTSFKNHQLRVLKQYFQLNHNP
DAKELKQLSQKTGLSKRVLQVWFQNARAKYRRLMMKNQQNGTNTKFSLDDEMPEDQNMKISESDEELDDHESNKESNTNN
PNSPSAESMSDFGEMLSPDLSPEANPSPTGSFGMSHPDKSLIDESLSGNSNFSLQDAVLNNSKSEPMAAF
>SR_DN125595_ANTP_Cdx
MHLDGGNSEVIDSMSRESIQTGWIHHLAHNFSNVASSALVNNPAGNYPGSGSTIPGLSGNTHHSLKFAADHHSPTGLKLHSD
HSPVDPNNPHSQFTELRTAAQNAASVDAAGAWPPTMPNSYLAVAAASHNQATGMMHPASRNSFAASAVHPNASWAHAR
NAYGTQFTSAEQVAAAYRGCNLNIMNAVVAGGDAAAAQHHHNSTAAAAAAAMHAMPFYMTHPAAGNPMAPGGHPGV
HPGQNPFVSPHHPYAASVAAAEMYGDHSSSTFRYNPTTGKTRTKDKYRVVYTDRQRAELENEFRSAQYITIRRKSELAMQVGL
SERQVKIWFQNRRAKERKVSRKVPGGGNHSSSDIEDSDNEIDDEDEENQLRSPAKHQSLSPGSQNTCHNGVQLPLIKSENNSL
LRPPSNELYGAKMS
>SR_DN125608_CUT_Onecut
MMNPSAFMAASHHGLQQTAPSFIPSNAPGSGSSVINGVDTPGSMNELVSALNNCNNHSSGNLFTPSHQTSSFLQCSPSGHM
LPGSRDVFSGLHNPMAHSPVGIGQSFQSSRQPNGVIGGTDHLFDFTIMPSIDYHRNPIGDLYSGYSISGSNRNTNSLTDISSNLG
GAGLSPTASSRPLMNVGTPTSGYPQFSPIKSQTLSSMVSGNPLDSNASPADRFGLFDPSGYFYPQYNLKPSDLKAPNMYSPYSA
VMSDKFSYYQAPSFYSQGYDSFESAAAFAYASATKPNPLAYSPTKPKEEVGLDSHTISPFLHTNHQQNKPSTKNRNGKSSNSNK
NQNENNSANLTGDSTNGSHVITNKELTEGEVDTRDLAHRITAELKRYSIPQAVFAEKVLGRSQGTLSDLLRNPKPWTKLKSGRE
TFRKMEKWLDEPEIHRMSTLRVAACKKKELEATHGIPNQHQNVMNKPPCHQNMTTPPHTTQQQQQHHSTKKNRLVFTDLQ
RRTLFAIFRETKRPSKEMQITISRQLGLEQSTVSNFFMNARRRCSDRWKEQEHHGMIGQNSKPIRNKHSINNSVNEKNKSLMSI
TPLHSNIA
>SR_DN126115_PRD_Pitx

LDCDAEGEGDGEDGGKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELASRVNLPEVRVQVWFQNRRAKWRRQEKLEQ
NHLKLNADFPLASLRAASAAAGNPSLHGMNKLLCSPLDSWLTATSPAALASLQQSLFPNPSLLNLQPNSVSSSISTPKHNNATS
DPSNLGPNLTSPNFDLTACTNLFTSASATQ
>SR_DN126115_PRD_Pitx
MVSTATPIVKSVSGASDIYPLQPNLLGPNGTGSMLDSHPDQMSPSTGPNRERRQRTQFTSHQLRELESMFARNRYPDMATRE
RLSAWISLSETKVRIWFKNRRAKWRKKERHSATSTPQNPAPSLDAAAAAAAAWKPSSFLHTTPSVPFNYPHYDPLRQSLVSAA
APNSFNVDSKTTSPFIAPANAFGSCSATPFRPYTQPSWNLFSNSLQTQFPGVQAFGAHSYPDPIMDHHLNPFGPSAAAPLGSA
NDTLLGGAPSHPNPYSVKQEFWHELDARANNTANQNSCFSVATSRDRKSV
>SR_DN126162_SINE_Six3/6
MHTDPKLGFLLSQAHANPGAFPPGAAALFAANPMIANHPQLLHQLSAQLAAQMNSAAGGPGGVPGGSMHPGGAPIPHPA
LMGFNPEQVAQVCDTLEESGDFDRLSRFLWSLPPHLLESTMKNESILKARATVYFHNGQFRDLYVLLENNRFKKDYHPKLQAM
WLEAHYQEAERLRGRPLGPVDKYRVRKKYPLPRTIWDGEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGN
WFKNRRQRDRAAAAKNRGTALNNFNSSDNEGSYDKIRSLEIDINDDDDDDDIDLDVSVSDSDDEDRENSAETKKSTSMS
>SR_DN126207_TALE_Irx
MTKHSPLPLSPIAPTITPLPFPPGSHNPAHINHVANMAAAAAARSALGQLLKSHQLMGGAGFNGGSHIGLGGPRPGSGGDG
MGSIPSSGVSPGSGCQLQLRGNGPCFCSSCSLQKVSMSSLSLPHSAGSLPGAPPPTFSTPTTFPSLGQMVAPGMLPLPLLPNPG
MLPLAGGGAGLLPPMMSAGLYGPAGPHSPFFDGGLHPSAADWVALGRRKNATRETTAALKSWLQDHLKNPYPTKGEKVML
AIITKMSLTQVSTWFANARRRLKKEGKMTWTPRNSYDDDEDDEMDDDDDVTALDDDSRSVISKASRPDESSTTIPKPHEVLRV
HVNEDDEGYNDDLFTPKDSSTCNATPKLSLSTTLHGNSSPGHPYPGTLVKCKLPLSPFTNPIQTLDNINFSSHSSQSAFATHANQI
SGQNNVTTIFSNHSQLSTPVEKPRRKTTSNGSKFWSIVDTVSKNEPETQPAEKSTNSPLHLSPTCDDPSPEPLTISETNGAEPIAV
SE
>SR_DN126498_TALE_Irx
MTSSYSTANKQSLANISSGERLVDPRHPENGSDALDRSKYLQNPHANSSNPYPLHGSSKHTYHQLTGGHTQPPQAPGLTPYHY
IHNPQLELDLQSAAGFDYGAAFGNPFESQSAMHYLPDSQFVGYRPADAYALGFESQLHNSFGPNGNSKNAATPQYSYGSHAS
RYGNQSYPPSQQYPDTQLLQPTFAHPSENMNPSRTVADSKVGSRNPQDCNSNIVSTSPSRGHVMTHQEYQQPGAGDPGFN
MQLCAPQNPFMFAPFHPTSYPSQNERLTVGPTEMSSNSQAPNATGSLGSAVATSGSTSSMRRENASRETTAALKTWLFEHRK
NPYPTKGEKIMLAIITKMSLTQISTWFANARRRLKKENKMTWTPKNKTNNNSQSHKENKLNHTAATDNIEATKLSTGEILENEA
DVTFDSLDSELKLEN
>SR_DN126831_ANTP_Nk2.1
MTLSPKNSAAGFSVTDILNPLEEYNKRSSLMDSTALNFASMGYRAQHSSPHNPQVGGNNAAQNAMQNVNASAITSPYMAA
QMAAAGNFGAHSMMAAAANSSAAAAYCNPNAISDINQSFAGLAQYDHFGRHSSMAGANVSGWGYAPSADPRLSFSRLM
HSSNGMGMGSMGSMGMTMSSFGALGDPSKMLSSQRRKRRVLFTQAQVYELERRFKQQKYLSAPEREALAQGINLTPTQVK
IWFQNHRYKCKRASKERQAANSNESGGDKSNGDRNVKLENKDPDADSLVDPNDSDGHVGSPLNDEDFTQDQVKNQLLANY
ANAAHVQQQQQAATQLQTQGYPGQQGLQGNPVTGTLNDMSRCPVIQPALGSVANPVNYSSDANNLALFQNSLGTRW
>SR_DN126921_ANTP_AMB
MDRSRPMILNESQETLRRASNELSPPMTSPISCSNNNTAVRMTPISTPTSGTPPASLSTEKYKSISPTPFNSKKQISYSPTSSDVID
DVTHDIIDDDNDTEDEDDSSDLSRKKNRTAFTPHQIYELEKRFRSQKYLSASERGEFALSLKLTDTQIKTWFQNRRMKHKRQQID
AENLQLAALYGPLIQQQAAAASFLRNSGALFPGPGNGLEHHPFMSLPGSPAASLFYNNPGQFGLSPNENHPSVGLPLSTVFAN
GSPAHVPLPTPVSRNFPGLPGTTPNMTHPLIPNINSIPHEFLKLTSSLVGNNLNFHNQFATKPPHQMPVFSSAAQIMSPIRSCSD
STPILNLPKMEPNFASSDSQKLIVTNATMENQQTVKSHPILTEKLQKLIQCVVHNE
>SR_DN127496_PRD_Pax4/6
MPHKGHSGVNQLGGVYVNGRPLPDATRQQIIELAQKGARPCDISRVLQVSNGCVSKILGRFYETGSIHPKAIGGSKPRVATDG
VVRRIAALKRESPSVFAWEIRDRLIQDGCCTQENVPSVSSINRVLRTLQSEKLKSPTASAQDMNHCSVVSMYDKLRLITPHWAQ
HPASVSAHNPWYNGFPSHPGATATPGGPPGMHHNGFTTAAAAAVLSASMGAAHVPHGHNHDPSFMTHLAAAHQNHNV
NLNNQNPLSLEAAAPGMIELTERTMRLSSISGNSTEGSKGQLQIPTPAGHMPTSQQINQSQLTSPSDSRLQHVHGNNAFSLAN
SMRLVNGHHPMGKLEIDDRHSPASPDNSVTSSPEDDEGEEEGEAVGEESKKYSKSQRNRTAYTSEQVDALEKEFEKSHYPDVY
ARERLSTLTGIPETKIQIWFSNRRAKHRREEKIKTNNNNNNIDSISSGQLSMKTEAGNHSDNSPSVPGIYGTTNTNVPMPPYFTH
YNMHGALTNILQNTQQNSAVHSNNQTHPSLSSAISYNAAPNMPSPNSFTRPTAVPGSNSALTPAGLSAMGTQLVSSTGGNS
GGVPGTTRSLSPMTAAALHYNQSAAALQHNSGACYYNSQQSYFHAAAQAVGGFPAGGSLESVQVNPSIGNHPQYWGGGGI
TSSAARF
>SR_DN127821_TALE_Pbx
MILGNAYDNPQNMDDQHRFALAFLQQQQQQQMNAAAMHSGVDPSGVNMHGGVSVSGASGGWGHLAHQGVPTSVGL
GDGSGQDLGLGSLGNNSGMLTEETRKRQLQEILQQIMTITEQSLDAAQARKQTLNIHRMRPALFSVLCEIKEKTGTFLNMRNQ
NDDDAPDPQIVRLDNMLAAEGVSGDGKSPTGSSTTGGAGQPDNTIEHSDYKAKLGQIREIYHTEMRKYHIACNEFTSHVMNL
LREQSRTRPISPKEIDRMVAIIQKKFNAIQMQLKQSTCEAVMILRSRFLDARRKRRNFSKQASEILNEYFYSHLNNPYPSEEVKEEL
AKKCGLSVSQISNWFGNKRIRYKKNIDNPGSYNPQDILVPPGSTPGAMPSSGSSASINGPPSGGASSSAGGPTGGASGYSDAD

QFSSGQLTPVGSEENAAAAAAMAQQQQQSDGHPLFQYPHSSESSMGVSSSSTGGGASISSLFDDSALASWGASAHAHHSAA
AAAASAANAAAVQDAIARSSVNNF
>SR_DN127948_LIM_Lmx
LELAETLKPKPNIKSEPTDIELTGEVVTSQIKKEKKSENSSPNTSFPGKMTVIMSEDISIWSDKHQQNHSIIGSQSLSELSSDFVPPK
NNLAMNKIPTASNSANSFQPTNPCYPLSRSFPHIGFPSNPLGSLGPIKSLDSSASSSMKENQSFHQNMPRNHAIINRQNQSISA
KNNNNNNNSINQNELPCHLEPQGSSGVAFCEGCGLAIEDRFMLQTTDRQCWHEKCLNCAACGCKLDHTCFIRNSQPYCRKDY
ESLVSGNCTGCELKIASTEMVMRVADNIYHLNCFSCSVCRIVLRTGDQFVQIEEQLYCKTDFEKQFPQFAMPEKQSSSENDNSR
KLFDSIAENEGNSCSNQNSDASDPNLMDHLFKSEHSNHEHESDQIKTETNSDNELDSEPKDDSDDEADLENSKDKNSSIQQKP
CRKRVILTTQQRRAFKAAFDLSSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKIGQETTTATTHSLRIRIGFSTLNG
>SR_DN127996_TALE_Irx
MNRLLVPPTSFYNPNSTAVTLSQMGASNVPNSLQPSPLHQAAIMMHHHQQQQHQAQQLAVSQAAGISNPNEKQSQSM
GGLKIPGSGDNSSTATTHHFPRQLNPVSSSGFDLAAAVSAARSSQVAAALHATSVAHHAQTPHPFVTASQQTPPGMPLGLSH
PPPHSPAMFGQGPHPGAAAAAALFSGHPAAAGASGIHPLLAAAASSGHFPGYHTPDGFGPMDFNMITRRKNATRETTSALK
AWLSEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKMTWAHGSGNEDDDDEDADIDDDGDDDVMNDEK
VNDLNELMSREAFNSGASQLHTAQKPSQLGHHFGPGADRSEERRVGKECRSRWSPYH
>SR_DN128068_LIM_Lhx2/9
MHSHLAPFGHITDNSHRSKRMRTSFKHHQLQAMKQYFNANHNPDAKDLKSLAQKTGLTKRVLQVWFQNARAKFRRNLMR
EQNQDMGSGMSEKSLGDDILDDDLNRRELCSMNGSEGGNDSTFGELDESFSDGVEFDDESMMSYSPDSSTESLGGTNSQET
VCDDEDEISQIGRAS
>SR_DN128068_LIM_Lhx2/9
MHSHLAPFGHITDNSHRSKRMRTSFKHHQLQAMKQYFNANHNPDAKDLKSLILLLELLLFQVWFQNARAKFRRNLMREQNQ
DMGSGMSEKSLGDDILDDDLNRRELCSMNGSEGGNDSTFGELDESFSDGVEFDDESMMSYSPDSSTESLGGTNSQETVCDD
EDEISHD
>SR_DN128406_ANTP_Dlx
PTRRSSDLGGVEAALTKGGKKMRKPRTIYSTIQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKMKKLMKG
NPNARIDMNFDALSPDAKDEEEGDDLNETDEHPIDHNTTANDVIDATIAHSGANQIAPNVTMPNLVLTGQPMSHQTSVAPP
GQPQSPWN
>SR_DN128964_TALE_Pknox
MTSSGGLTHLNNNTQYAPASFDYNDSKIMHSLSHSLLGGNSANNHLIGPQHHLTSPLNGTSNMGGYVQQQPQQNTSPLNM
PLPGTTPQNYNSVMTSGGSLNNSANMKQELSPLSNSTANSLLTAAQPQQLSGGTEGSPSSGGAGSIPLNSAAHLSSGAGTGG
GRASNHNTPGGGGGDTTVVAGSNTPLTTGGSSLGVTDHQQFENDKKVIFNHPLFPLLQLLFDKCETATNSAECASSASFDLDIE
NFIHLHERDKKPFYSDNADVNDVMIRAIQVLRIHLLELEKVNELCKDFCQRYIACLKNKMQSDTLLGCGAAGAPGGALSPTSLH
HQYPLSDSPNSQLQLSNCGLGVNVSHHNALSNTPSGGPHLSSAQLQQQQGGGGQQQGHLSTASSAASAAAMHAALYPQLS
MFPGMAAAAAAASAANFPGYHSPQGHAPPPGTPHSHPHPGSNLHNGGPGGMLSGSTPISQVGLQHPHHHGVMNGGAPP
GVSPLHPHHPHHHLLGGMGGVGMGVGGHQEMNPDGSFKKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTN
LTLLQVNNWFINARRRILQPMLEASNPESVNKTKKNKNNNRPTAVNRFWPDNLAQGQITNSFKGSGGGGGDDTSPHSDDNE
GESPPPPNHSHGHRGAEEGGTTGSGNNNNNHHHLSPNNRTTLGHEGIVMNGTRDNNSMNNLQLLNSPGALSDDCSDDQS
SNEVTITPPPPPGGGLGNSLLSNGGVRGGGGHPHSTSTTTGASLSAGGSFLTNGYDAYSAAASVANSQSQAASLFAAQTSAI
>SR_DN129314_TALE_Meis
MNGLRFEDQPQVPCQYPLTIDSGGGAGGSSTPFESNPGVQPSSLLNANGGYHPTPGSATGSTTQQQSSSAGGGAAQQAYTP
HPGSVSVPPHGTSTGGVLQSGVVPGGHHGSMGGVTPHGQDGCVQDKDSVYNHPLFPLLDLVFEKCELATCTPREPGMSGA
DICSSHSFTEDIHHFSKQLFEGMSNEKSNVYSDQDTDSLMLLAIQVLRLHLLELEKVHELCDNFCERYIECLKTKMPMDLIVEDKE
GTTSSATVSTNSTCLTNSSSSNGALGSMSIKPNRVPSTGNDDTNSTNDFEEFAPPSCNDMNDTDVLSPSLTNFSSSSGGIPRDH
HSNHVTKSTTSNPNNLPDFLPNSFGLSASNIDQSQQSINCNQLPSVHHLSTNQSATSKINQSNSALNISNNVSSTHGNSTQLSL
NGNIGTNNSKHDSNKYNSAIYGKDSSKSTKHGLDKNGNTKKSQSQNSSPEVSDRSDSFSPVNTNNTTNNDSSGGGATNNHN
NNQGVTSPTGSDESMDPKKPQKKRGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQP
MIDQSNRAGMPAAYAADASQMGAFMLDGQQAPYHLRHHPMMHAAATLAHGGNPGADMAAATQFAAGYPGAAAAAAS
GFHHHPSQAAMLIPGHPHAAMMMPPHAAAALHHQHYQQQALMDPMTTQHVMDSIHS
>SR_DN130571_LIM+ZF_AMB
METQVEQPFVCQFLEKPTSQKNPEDISEPLTPADNRNDEENLKIPEESSSGDETELEQDDGASMQMTMTQDPQSNQGAEHLL
YRKSTSPSSPSSASKLSSAVPGSRSKSTSAEQKSQLTSENNHQRSTSNRNSPSIYESSGLPVIQSSIPREISHQIGSKYNQLLNLQSA
KNRFLDSEMLNSVPSNFSISSLASQSVPNSSTVSELKLKLDELMDCSSSEPMSSNSSLPEDFSVEKCHFCEQTFADRIGLYQHKRY
NCSKNPEVLLHQAVAYTKSVQDVAGTQITEKDLSSIIKSEMNSSSFSPSTPMSTCKFGTLSLANQNAQNLERKKRTVYTDKQLET
LREYYVKQPNPNSDDIEKIANNIGLTKRNVQVWFQNKRARDKKNPGIFDYLDSSISATSPTHSVTSEISVKVQSKSPAVEGIDQSP
VSSLVNTNLSVNKVKNEGVLASSPMKPEALISNSNTNKLPSGESNVIPNLYSYQWLSFLSIYQAMAAQSMLNGGLNPASLLQN
QLANLAILNQLKTETATSNLESSEKSPSTKPSPVQSELGGLVTPNKSDASLETSSKASEDEPLDLSLVKTEHEHDMSKSSSTSPGSD
TGSVVNASNGISNDGSISGVSNSSDSLIMDLSSSILPPTLMSPGSQDLNSSASSSPLSLSNSSFLGGLMGHQHNQKTDCGSAAGS

STSSPVESCRSSSPNSLLAAACAANSSHGGPFACTQCKKVFQKQSSLIRHIYEHSGKRPYQCTQCDKAFKHKHHLTEHSRLHSGE
RPFQCPNCAKRFSHSGSYSQHLSQKRCL
>SR_DN130725_ANTP_Nk6
MTALCPSPQNNNNKQSSPGIVTAKLEPSEDRQHMSFQLTPTGENQDSPLLINQPTNPAQIKSPNHMNYTRYPTLNPSNCFSLI
DHQNQFSLLNSASFPTTLQNFQFSHSNSKPTSPKTATQTSPLGHSVSTSVENSQKPVPNNFSRELLEMFYNYNACRYGPQLTSG
FHNMANSEVHTSGHNTMPFSNYPMGGPILPQIQQWASNSPVMNAVSAASNNQNLDSRSDNNSLLDQTESLISGHKSEYVN
NRYTIDKILDSTTSSKDKSKRIQLDTKSSSSAALKGKKERDLSSSTFDSKDFQVHPHNNSLFHHQQAAAASLMLTQHAQAALMR
QWQQGQILNFGQAKDSDQFTDGASRPLPLLNKDNQIQMSIPTSNHGFSLDPFSGANFRSNPGQPLDLTPNDETNPLSLLSSG
GDNDPNFGEGSGVNKKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALGMTESQVKVWFQNRRTKWRKKHAAEVASN
RFKLQQSISSSNNPHSVGHMVGGASDLDHDCPPNTLNLTLGTEDDLDDNCLYSDSD
>SR_DN130717_ANTP_AMB
MKRSFMMVDILSSDSKKESKNEPESSANKVSMLDQNLLEQTVLSVTDSLTIQIEDEATKRQPENDDEEDRTSDSPDSGVECPSF
FEPNVAEGNDTETKVSERPSQPNNPPIRRPSSAHSKGDDNEKFENLDAYDELDAMSSGSRDDTDSVYSGGFSTGSGCKKQRTA
FTPDQIHELERRFAQQKYLSASERADFANKLKLSDTQIKTWFQNRRMKRKRQQQDIESAQLQALYSSALMIPPAPYHLMVAQ
NPANHPGLNMNHLQNPFHSQASPVNFPGHLPGGQNVGVPHLTLGLSEHLPTSQGLMVPPVSREFPQQGGPKSLFLPALGLS
ATFGLTNSSNHSGQNGGVPSMTSSSRALELTSQMNPNIPHMGLGGMFGPPKMSL
>SR_DN132449_PRD+ZF+CUT+CERS+LIM_AMB
AMNQINNYKRQRIAEDLNQIDFTLSSTECGSPNTSNWVSPFAGKENIETGKSKLASKSSLVFKVEKGCSAKRRKHILKFEQVSILE
NYFHIDSSWKMATVEKAAAELGLPVKKVYKWGYDRKHVSRNLGDVAKVNPNDYNELVEEILMLHQRKSVSKRTYKKRSKRAT
KQSNSSSNTNIGDSGLTGLEQRLASIEELAGSDLNKKAQKLSYTQKETLP
>SR_DN142180_TALE_Tgif
MFSLINQQFYEYYEDKTESTVGNELKYSSGITDHGSHLPTYEDPENFSENLHSSNSHLIYLKKQRGTLPKEAVDILRSWLELHRYH
AYPTEQEKISLSEATNLTPLQVCNWFINARRRILPALIKKEGRNPSHYTISRKPRRPNKFSTNRKVNTNFQENPQPIKFWRPWLQ
EEKMQNIQSRKMHCFKMSHIEQILRFRNSQTIKRRWHSLRFSSNLSETAPKREPNSFKSHHSFQI
>SR_DN152752_ANTP_Barhl
MSSAAPIQETALAVQRGEESRSCIDGHQLNESSNSVMSSSADKRSPRFTIDFLLSSHYKENSDVSKTEKTNANQRNMTLDSDN
DSDDSLSQSEHGADFNSTNHSEGAGRGGQSPCFGTGKKHRKARTAFTDEQLAQLEQNFERHKYLSVQDRLDLAASLQLSDTQ
VKTWYQNRRTKWKRQTAVSLEMLAAAERQHNLCTLQTAAL

Protein domain identification

trinityID class	(homeoDB blast)	protein domains (Hmmer search)
DN126162_c1_g1_i3|m.89990	SINE	DUF1421 Homez HTH_3 NOA36
DN87552_c0_g1_i1|m.17761	SINE	Homez RloB
DN118860_c2_g1_i2|m.56407	SINE	Homez
DN113036_c0_g1_i1|m.41100	PRD	
DN126115_c0_g4_i7|m.89715	PRD	
DN120882_c0_g2_i7|m.63873	PRD	
DN123029_c0_g3_i3|m.73095	PRD	Neur_chan_memb SARS_X4
DN112801_c0_g1_i2|m.40593	PRD	
DN52363_c0_g1_i1|m.7385	PRD	HTH_17 HTH_23 HTH_28 HTH_8
DN108331_c0_g1_i2|m.33300	PRD	
DN75610_c0_g1_i1|m.13237	PRD	
DN116171_c0_g1_i2|m.48160	PRD	
DN126115_c0_g3_i1|m.89710	PRD	
DN120054_c0_g1_i1|m.60873	PRD	Membralin
DN122029_c0_g1_i1|m.68599	PRD	
DN118878_c0_g1_i4|m.56464	PRD	HA NAD_binding_1
DN127496_c0_g2_i1|m.98660	PRD	DDE_Tnp_1_assoc HTH_23 HTH_29 HTH_Tnp_IS1 PAX
DN122619_c1_g1_i1|m.71282	PRD	CAF-1_p150 CDC45 DUF3653 SOBP
DN116853_c0_g1_i1|m.50040	PRD	DUF900 HTH_23 HTH_29 HTH_Tnp_IS1 PAX
DN98097_c0_g1_i1|m.23524	PRD	DUF3982 ETRAMP
DN61173_c0_g1_i1|m.9426	PRD	
DN111202_c0_g2_i1|m.37759	PRD	

DN76474_c0_g1_i1|m.13560	PRD	bZIP_2 GP38 MRP-L20
DN128964_c2_g7_i4|m.109271	TALE	TSA
DN142180_c0_g1_i1|m.133360	TALE	Myb_DNA-bind_6 Phage_int_SAM_2
DN129314_c1_g2_i3|m.111748	TALE	
DN127821_c2_g3_i2|m.101192	TALE	Homez HTH_19 HTH_26 HTH_3 IMD PBC
DN123538_c0_g2_i12|m.75677	TALE	
DN126498_c0_g1_i4|m.91973	TALE	DUF702
DN127996_c1_g1_i3|m.101906	TALE	DUF2457 FAM176 HA Homez
DN126207_c1_g2_i3|m.90240	TALE	DUF4611 Homez
DN123214_c0_g1_i1	TALE	Homeobox_KN Homez
DN106703_c0_g1_i2	TALE	
DN123138_c1_g2_i1|m.73697	POU	HTH_31 Pou
DN125354_c0_g1_i1|m.85350	POU	HTH_19 HTH_31 Pou
DN116892_c2_g1_i1|m.50278	POU	Pou
DN124898_c2_g4_i2	CERS	Hap4_Hap_bind SR-25 TRAM_LAG1_CLN8
DN70982_c0_g1_i1|m.11720	HNF	Exonuc_V_gamma FAM183
DN121592_c0_g1_i4|m.66335	HNF	HNF-1_N HTH_19 HTH_3 HTH_31 MerR_1
DN117560_c0_g1_i1|m.52153	CUT	CUT
DN125608_c1_g1_i1|m.86741	CUT	CUT TOM20_plant
DN127982_c1_g1_i4|m.101787	ZF	
DN113132_c0_g1_i1	PROS	ATP-synt_E_2 Collar HPD
DN123163_c0_g1_i1|m.73968	LIM	LIM
DN128068_c3_g1_i2|m.102333	LIM	
DN124757_c1_g2_i2|m.82144	LIM	
DN128068_c1_g1_i1|m.102330	LIM	Homez
DN125302_c3_g2_i1|m.85233	LIM	LIM
DN127948_c0_g1_i1|m.101565	LIM	DUF1510 LIM
DN125451_c0_g1_i1|m.85840	LIM	Homez LIM MLANA Mut7-C
DN118733_c0_g2_i1|m.55910	LIM	
DN123001_c0_g1_i1|m.72616	LIM	LIM zf-C2H2 zf-RING_5
DN106437_c0_g1_i1|m.31009	ANTP	BrkDBD Homez
DN109741_c1_g1_i1|m.35411	ANTP	
DN116214_c0_g1_i1|m.48271	ANTP	Homez MADF_DNA_bdg Peptidase_S49_N
DN112643_c0_g1_i1|m.40300	ANTP	YtxH
DN130717_c3_g1_i1|m.123804	ANTP	Auxin_canalis Homez
DN107350_c0_g1_i2|m.32011	ANTP	
DN110321_c0_g1_i1|m.36339	ANTP	
DN123682_c0_g2_i1|m.76247	ANTP	
DN80548_c0_g1_i1|m.15026	ANTP	Homez
DN123370_c1_g1_i1|m.74729	ANTP	
DN126921_c0_g1_i2|m.94661	ANTP	
DN74225_c0_g2_i1|m.12757	ANTP	Endosulfine Homez IPP-2
DN124972_c0_g3_i4|m.83074	ANTP	
DN42484_c0_g1_i1|m.5667	ANTP	Endosulfine Homez
DN56103_c0_g1_i1|m.8196	ANTP	DUF1577
DN126831_c2_g4_i1|m.94319	ANTP	
DN121921_c0_g2_i1|m.68018	ANTP	ORC5_C
DN93804_c0_g1_i2|m.20732	ANTP	
DN130725_c3_g1_i1|m.123269	ANTP	

DN100159_c0_g1_i1	ANTP	HA HTH_28
DN152752_c0_g1_i1|m.134562	ANTP	Homez
DN123752_c0_g1_i1|m.76614	ANTP	Ndc1_Nup QLQ
DN118344_c1_g1_i1|m.54580	ANTP	
DN87690_c0_g1_i1	ANTP	
DN118899_c0_g1_i1|m.56533	ANTP	Phage_CI_repr
DN122139_c1_g1_i1|m.69070	ANTP	
DN89648_c0_g1_i2|m.18777	ANTP	TRAP_alpha
DN123563_c1_g2_i1|m.75836	ANTP	
DN124792_c0_g1_i1|m.82241	ANTP	
DN125595_c2_g1_i1|m.86363	ANTP	Homez TFIIA
DN128406_c0_g1_i1|m.105028	ANTP	Homez
DN160215_c0_g1_i1	ANTP	

DN113631_c0_g1_i3|m.42241	PRD+POU	HTH_19 Pou
DN108290_c0_g1_i1|m.33236	LIM+ZF	Transp_Tc5_C zf-C2H2
DN122932_c0_g2_i2|m.72688	PRD+ANTP	Tim54 Trypan_PARP
DN65396_c0_g1_i1|m.10252	TALE+ZF+CUT+CERS	Codanin-1_C HTH_32 Tau95
DN95229_c0_g2_i1|m.21775	PRD+ANTP	
DN98018_c0_g1_i1|m.23556	PRD+CERS	TIP120
DN25767_c0_g2_i1|m.3518	LIM+ZF+PRD	HTH_32 SR-25
DN102254_c0_g1_i2|m.26921	LIM+ZF+ANTP+PRD	KorB
DN130571_c0_g1_i2|m.121942	LIM+ZF Homez	Mt_ATP-synt_B zf-C2H2 zf-C2H2_4 zf-C2H2_6 zf-C2H2_jaz zf-H2C2_2 zinc_ribbon_5 Zn-ribbon_8
DN132449_c0_g1_i1|m.132246	PRD+ZF+CUT+CERS+LIM	HTH_17 HTH_30 RIC3 Sugarporin_N
DN57798_c0_g1_i2|m.8573	ZF+POU	

3 - *Hofstenia miamia*

>HM_DN1193_LIM_Lhx6/8
MMTCEALSSFTPISLQQRNGHHFVNESTDFFAPFRQSSVVKNLISSSFNSVGKEILIKNPPVLSSEESDLQKNLLKCSACHNGILD
RYLLRVNGLRYHISCLKCSACQVVLSGQTSCFTRDNSIYCKTDYMRFFGIKCAKCFRSVSSLDWVRRAKQHVYHLACFACDVCH
RQLSTGEEFALHEGRVLCKQHYGELIGTSAEEDKESKDSSEECTEADTTLSPPKTKRSRTTFSEEQLKVLQANFSLDSNPDGQDLE
RIAQMTGLSKRVIQVWFQNARSRQKKHLSIPANGNKLSHSMLSPYECIQEDRITSPLSLNVMYEDSTIV
>HM_DN1275_ANTP_Bsx
MMNNVSFSINNILSNNSNVKATPNKISSHTCQYPVSHNKTIPYSQLPLPWHRRSPIFDYPPAKQTNYDYNMLPSIYKPLALSQL
MRLGHSEMALRLYRRRKARTVFSDSQLQGLEERFKDQKYLSTPERIEMATLLCLTETQVKTWFQNRRMKDKKCCKKAVLSDDD
>HM_DN3061_ANTP_Hox9-13(15)
MPPVVGSEVFMSPAGPWVQNPTYYEQAFINHPGTHTPATHQNMFNAFQHPDKYLNPEAVAHPTYIQDPKISELVTSGFENP
WNMYPPSSHTGYISATSPDYIQPITNDKHYILPPATQPGIHQHSEHAFMYSDIKPAKEEIISHDLPYCDKRTSPAQLPIRDNRFD
MGLSKNSCYSINTSSCNVTPKSSMKESPLESDNSSIDGSPEMPSNPGNMPSLALNQQNPSWLNRNTTRKKRRPYTKTQTLELE
KEFLYNTYITRERRLQIAKSLNLTDRQVKIWFQNRRMKMKKINASQQHNMPSAHLVV
>HM_DN3930_ANTP_AMB
AIGSRINYFNWPVNVIVVICDNLAFHYTGENQESVDMTIYPSVLNHGRSQSKTSFLMDDILSSKDQTQLKAAQVVTHSPEITSFT
QRFSHSTDNDVSPQMDFSGNCSDDSSASKKKNRTAFTPGQICELENRFRLQQYLSTTERGEFASRLKLTDTQVKTWFQNRRM
KLKRQQTDAENMQMQMAFLYDVCTVHPTSLVHQLTSSPLMYQHTHQPLPLRSQSDRLYPPHPFSKLALFNPAHTSSTPS
>HM_DN4748_ANTP_Mnx
IKRVYKSHIQAATSTVRMNPADLKQEPTFSIDSILNCEQPAAKQNICDLLQKNIIPAQHLIFPQLIRNPFVKQHNPLQALAHFPIVN
HQVPFGALIPRHEYNRMLDMQNTSAVQNSLNKSRRPRTAFTSQQLLELENQFKRNKYLSRPKRFEVATKLMLTETQVKIWFQ
NRRMKWKRARKGSIACTKQSTPPITCRPQLPI
>HM_DN7261_POU_Pou4
MMMDQKSHISMHPPISAGVSAGCNPMKYSPFHTAAIRSQPTDNFRRHFQTAPSSNFFTGNPFDDSFFRAEALAAMEACAAK
GANQTCKESSLSAPHQMTYTPHAHMAPPPRCSQSYGEVFDSFDGFGASFQSNMAALTHPITHHHHSGGVSDIEADPRELEAF

AEKFKQRRIKLGETQCDVGKSLSHLKIPGVGSLSQSTICRFESLTLSHNNMVALRPVLQAWLEEAEAAARNRMDKNALLGGIDK
KRKRTSIAAPEKRSLEAYFAVQPRPSGEKIAAIAEKLDLKKNVVRVWFCNQRQKQKRLKFAAFR
>HM_DN7688_PRD_Vsx
MSGPTMNFQQLISSQQPQRSPFGINELLGLDASSDSSRQFNHNLAFFNSMNASTMCPPNYTPASTLFDTTQGYLARGTAPFLP
SIYSSCNSAVKNNLPASFPFTNSSVSSSPATQNALTFSDILTNGSSTTENRNISNLFINGLSKKKKKRRRHRTIFTSLQIDELEKAFSE
AHYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKKEKT
>HM_DN8120_LIM+ZF+ANTP_AMB
TFQQKSGQHTCTACNLRINNVDLLHMKGKYWHTGCLTCSMCATSLTDDESCYFQEGNAYCHLDYAYVFAPACTKCALIIHPSE
WLTSEAGKHFHSECFQRFHGSFIFNQPLSHGEWLTINNNVEPLHYPIGDGKQNYNIERSMHNPPRESKQALDNAQPEHQPHL
ENIQNTQPLHVPPRKSIRKARKRTVFTQQQLAVLQSRFAIKQILSRQEYANLADEIGVKIPIVRVWFQNYRQRSKTNS
>HM_DN8401_ANTP_Nk5/Hmx
MPLSDKKCSSFSIRSLLELPQDLEASLTEEPDLFNSKMADDQSPESSDEDDSITVSSDGDDRKKKSRTVFSRSQVCQLEGTFDQK
RYLSSAERSTLARSLKLTETQVKIWFQNRRNKLKRQIKTGIEAGAINSTGNTVSGVQSGPPNVNAAFLTFNQSPPLFKLPQPTPA
DINPFLSTSAASFPHLSPPLFLTSRSSAFPGTFTDSSFPSQLHPANKFSNFTLVRSLTGLV
>HM_DN9855_LIM_Lhx2/9
MEPLKVQDRKEIKCDGCLKKITERFLYNTIDHYWHQECLVCEDCKVDLSEARSCFTKFGRIYCKADYFRRFGSKCCGKCQQTLEP
HDMIMRAKHLVFHTQCFTCSICQAILRTGDEIAIFNDRILCNKHSDFSSLEKPLLSGENPLQVFTPNEIFRNLSLSQTLQPSNKGRP
RKRKPTQSDDANQFRQHREDGSVYVEGHISHSRSKRMRTSFKNNQLVQMKQYFTMNHNPDAKDLKQLSHKTGLTKRVLQV
WFQNSRAKYRRNMLRDQLQSTRKSDSLSKSGDEEDDNISYDESIDLQTLDSDSLSDGSALSYSPSSSPDPNTLTEESNENIVPTN
P
>HM_DN10228_ANTP_Nk6
GWFIEKILYVKMVQNEWSSNYSIQNLLNKPTDKQCLQPQQQQSMNVAMTNYLLTQQTAIIKQWQIHQQISACSKDIRNPASS
LPIYLPMHPNAEESGGSSKKKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAYALGMSESQVKVWFQNRRTKWRKKHAAD
MAVAKFTLKDYQSDDGSDDSSDETFSTNHIFDKIMICQ
>HM_DN10420_PRD_AMB
MENLISPNQASSNKIYSSLNISPPERKNVLTYAENKTPDVFSDSQSSEDDKSNILKSSTLNSSLTNSNDDKGSKPKRHRTRFSPNQ
LNELERCFGKTHYPDIFMREEMALRIGLTESRVQVWFQNRRAKWKKRKKTTSMFRPHPASFLPSHSGFGFTQLNPSAFADPFC
SLSSDPRSWTQMQTPASAAMTGFVQGSLARSVQSPSQCLLPPSTTPPQTGSTFHHQPFPSHGMVDAEFFDPSSDPNNSWCA
PSIAALRRRAFEHSSVVGGTAATSGGLNSFHSLNQLATNFELR
>HM_DN12035_ANTP_Dlx
LNTSLNHYEMNRSPIESEYHRPMLPYHHTMTFSQPQPPFYPHSYYSQAQQPELFYRPTKPYYPQGQPGLIHHAPNMSFGASTL
YNDSLPLPPPPHYPTPGSTCSNREDASPSMPDLSVGEPVLSPEEKLSSEKSDKHNKKIRKPRTIYSSTQLQQLNRRYNTTQYLALP
ERADLAATLGLTQTQVKIWFQNKRSKMKKIMKQAQANSTNPALSNTNAPASVTTPVLPSQSWDIPSTIKQQPLSPQDSPCAG
LSWYRNSTPHPSIC
>HM_DN12216_ANTP_Dlx
HQFAHNSTHMVDSALSDCDKHSEYEHRIAKGKKIRKPRTIYSTMQLQQLNRRFQRTQYLALPERAELAAALGLTQTQVKIWFQ
NKRSKVKKILKHVHANSNNPSIPARSSPATSPDAIELSNS
>HM_DN12757_TALE_Tgif
MSQDYCTSDDGEQKDIIDSLKMKKKRGNLPKESVEVMKKWLNEHRYNAYPNETEKNQLAEKANLTVLQVCNWFINARRRILP
EMIIRDGGDPNDFTLSRKSNRALKHHTPYSRPAYHHLPTCHYRRCTQPPTSNQKHNIWQPWVKQSGLEKPPRSSFPCSNCSLQ
TVQMPIINQKLLYNQYCYLASGIGLFNEYRRQMKTL
>HM_DN12978_ANTP_AMB
MQLEDCRGGCFNCIFSQMCQDDYNKYIELLANQMNNANMLYSFSTHPFTLYLKRQNALPSINPIPNLITSSLMYDNRLLRKRRK
EGERRQRTTFTTKQANILESEYQRREYLSKQQRQALATSLQLSETQIKIWFQNRRAKDKRTGKSPNDQIRKKSWCKALKIADERV
AETDIVVE
>HM_DN14479_LIM_Isl
MQCAISAASHPLAITFLQCVLKMVNISVDHRRESFQANISTPDAINIFSAYKTSGNERFSPCISLCVGCGEKIVDRYLLRVAPDME
WHTSCLKCDHCGVELGEKTSCYIKHGKAYCALDYIRLFAIKCTRCTGNFQRSDSVMRARNQIYHVECFACAICGKQLLPGDEFAL
REEGLCCLEDNENIERDLETDGTVEEIQSPDDGTSVCSPNPEGKRKSSKSGRVRTVLNEKQLSTLRAYYHANPRPDAPMREQLV
EMTSLSHRVIRVWFQNKRCKDKKRQKALKDNSDVECELSSNANDDDKADNEFSGENDQSVFNIIANDQSYKADCSPPKA
>HM_DN14515_ANTP_Cdx
MHLEALDSHHTLDVMSNRTPPHQAAPGSNNNWFLPASNFHTAYSQQQAASNAAAAAAIKFEHSKFDSPLDSPHFSELSRTG
DNCCSANWINSSYLRTGSPVQRTYHPAWPARTPFGTGLPDYRCNAVATAAATGHHDALGSMGPLYMPSNAGFVGAPTYAD
MYPITTESTTYRINPQTGKTRTKDKYRVVYTDRQRAELENEFRSNQYITIRRKTELASLVGLSERQVKIWFQNRRAKERKVARKG
GVRLSSSADSDISLKSPCPTSP
>HM_DN15356_PRD_Pax4/6

MPHKGHSGINQLGGVYVNGRPLPEATRQRIVNLAHGGARPCDISRILQVSNGCVSKILGRYYETGTIRPKAIGGSKPRVATNDV
VAKVASYKQECPSIFAWEIREKLIQDHVCTEENAPSVSSINRVLRNLANEKTGSTNGEQNLDKLRLLANNQWANPNSIYAAYVN
AGTYGSGALSMAAALPAPELEKKTNCSNANGMEEASEEARMQLKRKLQRNRTSFTQEQIDNLELEFDRTHYPDIFAREKLSEKI
KLAEARIQVWFSNRRAKWRREEKLKNQKQQIANNNNLTNGAGMAVVSPGNSHLSNAIANNLYFQYNSAAITYGYSVDNYNA
TAGTFGPLGAFSSTGQPYPCVPRTSYDMSYLRPKESSQVLAPISSEGFNYSNIPYSEQTGLVAPPLTKMSGSQSTDMSHYWSRI
>HM_DN15813_ANTP_Nk2.1
MSLSPKNTSFSVTDILSQVEESYQQQVVSKQRPMDSALNFAASQMVNPYRQNAQVASASVPTMNGSASTAQYMHGMSGLS
AHSMMGSSYCNGTMADLGQGFTPGLSHYNDHFSRHNPASGWYGTTDPRFSSFSRLMHPSSGMNMASMSLGFPSDPAKFL
SSAQQRRKRRVLFTQAQVYELERRFKQQRYLSAPERETLAHLINLTPTQVKIWFQNHRYKCKRASKDKAATPNSVEDKQNTGIK
SEIKDANSPQPHHVVATSPSLCADLGLDEGVKSSNDYHSASLPTLEPAPNSNLDYATAVSQGCALTMSAGYPSANDAYSGMLQ
TRIW
>HM_DN18872_ANTP_AMB
YKEVITQGGWPSVIMVHNFSIQALIEDIPKSSESSKSEQSVFCMDNIVPYPPNLYVPPYFNVTHTLASLHRRQRRNRKNRRERTT
FTSDQSVKLEIEYQRSEYITRSRRFTLAEMLKMSENQIKIWFQNRRAKDKRIEKAQIDQQLRLMTTFGNVNPIVQYDPCMILWK
QAVDKSLQQTI
>HM_DN20170_PRD_Rax
MSTLNYKKEARKQPILHSITDILGLEESHTDKWSQHLEDESDKSECSEDNDRKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSR
EELATRVQLPEVRVQVWFQNRRAKWRRQEKLEQSQMKINTDFPLTSLKSFIKQPMLPGAGSFHLLSDTWNMSNPSLGNVAL
HQFLPFSGHNTLAASFGSHISAINNIYDNTNKDSLKHVSTSHNSCSDSSKIISSTNLFVG
>HM_DN20211_ANTP_AMB
MPKSFLMRDLLEMKRELSTDISENPNEKPLYISTMETCSSISPETDVESDVEVNEITSLEDDNCSDKLRKKKQRTAFTPEQVFELEK
RFQIQKYLTASERSDFSDSLKLSDTQVKTWFQNRRMKHKRQQQDIQNAHLTAIYSVYGNATLHHVLSANSDLLPSILGDGLPSV
IAPARSQLYNFQNPYTATSHLYSPRKHFSLPLNICT
>HM_DN20967_POU_Pou3
MVVSSEIALSNNCPTSNTSYQNQNTFLGLYHSESSSQHISSDVSHEYDPYSIYPAGVSSPSTFTQSQTCKQEYSSWQQSQLTSSE
SAQLFSGMRWQYPNSDSVSTTGAYSCLDWKPSSYPSYPSTCYQWNPAQTYLAYNGNCASPYAAQSGTAFLFSKEHATTGAEA
AVALPPDWAEEALSGEETPTSDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLAK
WLEEAENNNGTPSSLDKIAAQGRKRKKTYKYRSWY
>HM_DN21081_TALE_Pbx
QPPIACFIKSLIREFYSVRVSLMVVWDKSIMDEQHRLAAVALSRSIPSETLHPYMSHLAVANLSGAPQPPTSVIDHGGLLTAASD
ERKQQIQDILQQIMTITDQSLDEAQARKHTLNCHSMKPALFSVLCEIKEKTGTFLNMRNTAEDDPPDPQIVRLDNMLIAEGVA
GPEKGTNLGQAAGAGQPDNAIEHSDYKAKLQQIRQIYHTELDKYEQACSEFTSHVMNLLKEQSRTRPISPKEIERMVGIIHRKFN
AIQMQLKQSTCEAVMILRSRFLDARRKRRNFSKQASEILNEYFYSHLCNPYPSEEVKEDLAKKCCLSVSQISNWFGNKRIRYKKNI
DNPPNYNGDVSEMGRSGQSSSSMYNMSSNGQYGCDSEYGSGQITPVSGQVSEMENDTSGVITSNEGMATSAVSAGMNAIF
DDNINNWQGNRSTPSAGSNY
>HM_DN21447_TALE_Irx
MSTNVPFVQCTNITTIQNLEHYDKRFAFPENEIGGSTSPYYYPFFDCTQNPAASYPGVCSPQQPCIQASFVPPSFPTSPSSTKLSV
TPTPFPQNFSCENYLPPTSKQQYNPMCNDPNFMPYSTNQLQAQYPFNERLDSQCNIRRKNTSRETTAALKSWLYEHRNNPYP
TKGEKVMLAIITKMTLTQVSTWFANARRRLKKENKMTWTPRNKSSVSTISQIKSTETLKKD
>HM_DN23062_PRD_AMB
MVEDFASLYAMDNSGINVPRSELVDNLQESVQSYKKVKGEPSNPAACPTPARRRHRTTFNQDQLNELEAAFSKSHYPDIYVRE
ELARITKLNEARIQVWFQNRRAKYRKHEKQLSKTLVPPVGSASIISNQCNNIMMRNIYPPNPAVFRSAGSPHSTHYSYGQPMS
HTQVSRYPTQMPSASVSVAYPSAMAAAMPQFGAASMPSHAEMFASHEEWYNKSLSAFRATGAAVHNPLNPMTSAGLISYG
N
>HM_DN24302_ANTP_Gsx
MQHPRNFTIQALLEDQSSQENPKSFHRSPIFASPNTCLTYGMPWFQYIYHNNLYLTNENISIEDGVKNTPILCNYQLTSSQEGAR
RSRTAFTSNQLVMLEKEFNNSKYLTRLRRIDIATSLKLTEKQIKIWFQNRRVKDKKTKPNEEG
>HM_DN24356_PROS_Prox
MSATKSEISSPLDFSPSSSTQSITAIRSVNRTTHLEAFPPTSTAANPAFTASLPSLYPFPYINPFGNNNGISNSHLPYFPFPSSVLPFL
SSLPLPYINERKRTKQNAESLPTMVPYLIKPNDSTHRCFLDNTPLNFKSDLVKTLTDCSSPKRAKLSDQLYSNKCQNETCCSEKREL
KAKLDELSHRLQLSGKCELSNDIVLKEFAKILQKNSREQQENDVNKNTTQLQENCEQNPLLFPQYINNQLLTNRFKHKLCEPRTK
VPDERMLHSLAAETLTANHLKKAKLIFFYQRYPNPSTLRMYFPDVRFNRQNNGQLIKWFSNFREYFYNQIEKYVRELNGNCYN
HTSDVILSEDISGSVLFKSLILHYKKCNTFEIPSSFHEVTRAAIQEFNTSISSHLDKNPSWKKQIYNVIKKYDKPIPDSFKEINFCE
>HM_DN24659_PRD_Otx
MNLYGTTEVAYFKPSHHPAYMNSMLAHSLPQQVDHLVPHAMNHMAYPAASLSHLQTTQQPARKQRRERTTFTRAQLEILES
LFSKTRYPDIFMREEVAMKINLPESRVQVWFKNRRAKCRQQQQQGGNNTTTKDSQESPSSNLLIKTEPGILSSLIPDKVSKESPP

IKAEKDTSSSVSPLPAASPKSDSLGFTPAYNPYPNSHSLCGYSGAEHFPYLSRGAATNPYSPMSATPSTGGYYPNSYSSYITDSGC
YAPYFGNSGRVYPGTTANMPGFGDTFTASTSPVSDSSKLDHITATWKY
>HM_DN24763_TALE_Irx
YWFKWLFCMPTYEGALLICNGQSLTARLTGAVLISLLIVLSRRLDVVKSTDKNAHNNQFSFASILMDVTNSSQMLLMASMQVL
KNQLETKCVESSNILEPSKLSPTLHKAHLMGGSCNCSSCLPPSSVEHAQYLPRVLQMAQMRAAQQPNAPLSQQVLLHPFGFPA
AAFHPAFHIPNATPQFPDGLVAGDWGPLARRKNATRETTAALKTWLNDHLKNPYPTKGEKIMLAIITKMTLTQVSTWFANAR
RRLKKEGKMTWTPRNAYDDDELSDDEICSDEVLGDSAVTTTSGESGQTDKFVDVTGCCTEEVDEGYDESYSKVILSSQQPHVFS
RISDVSPLSDQGSENSEDSTSKTDSAVPPPRKRTKFWCLADIATSNKSSLAG
>HM_DN26989_PRD_Pitx
MDPYPANIDFRDLIPSVVPLVPTHHSLQSEHQLRLHHQGACNTSEPLFNKESEIPPLSLNSHHFSINTLSSSNSKEQTIKEENAGSS
SSTKRETLNPIRLTSASSDADEDDEDEDKNGKKQRRQRTHFTSQQLTELETLFSRNRYPDMATREDISAYILLPESKVRIWFKNRR
AKWRKKERNQLQEYKQAAALGMNFMGYPYEHTELYPTQFSYPSYNNWASKMPSPLSAASKSFPWSFNTGSAPVPAPGASM
MPNMSSSLMTAAPSPPCVGYGNPVATSQYFYRDCAPTAALTASLRLKPSQQNSGSFSYNSPNRNSNSIPPTLTNCQYAVERPT
>HM_DN27245_PRD_Shox
GKELRCPIIICCEMRQELDEKAHCKRNRTQFSTFQIKRLEKLFEETHYPDSYMREELGKRLNLSESRIQVWFQNRRAKCRKYDCV
KEPIPSNPLELFSSVQLYQSPVGKMTSLTPNRELPPVEQPLKATMQKISPSSNVVVDPLGGNSFLLMRHSLINKTNDLEFDSKVH
SMNFVKDLGYQMLYNNMN
>HM_DN27253_PRD_Gsc
MTKKSFTIDSILGRESKTIINSPLKYRPQSAAHNGLSLASYSLLAMGMQSPPSVFHPQSFISSHNFIQHPSLSYISSLEQSLTPPDLW
HSALFRATLQKRTRRHRTIFTEEQLEELEKTFSLTHYPDVQTREQLAAKADLKEERVEVWFKNRRAKWRKQKREVEENLKSGNT
HSTTVKSFERSHELDSSLRALTSPLISDCGSTSGEDVMLNEDLR
>HM_DN27605_ANTP_Hox5
MSSPFFMKASNSAVNANYYQSGFYVPQVDQTGQYQTFSAPQTSPTYSPSTSPLEQLTYSTPYITHHTTYPSDYPTVPQHYNSN
QNTQSMMFENYYPYHAVTPYVMGYELNNSGYCPPPVTSELVDRFHQEDTKPHINSLMPCTRSLDEKEIDKDNNQSPSVNHSL
VGPPTQTAAIFAWMKRNHPPTESKRTRTAYTRHQTLELEKEFHFNRYLTRRRRIEIATLLSLSERQIKIWFQNRRMKWKKDNNL
KTMSQVDAVAM
>HM_DN27964_LIM_Isl
GISENHFRIKEEFSSPNSISEMQSIVNKHDYQQFPGQYPPIKTEQQQPTMHQNSSRYSLCVGCGGLIHDQFILRVAPDLEWHAK
CLRCAACKGTLDDSTSCYVKDGKTYCRKDYFELFSVKCAECKLAFEKTDFIMKAANKIFHVECFRCGICNRQLLPGDEFALRKDGI
YCREDSENLDKNGNSFSEENHEQAELSPDSCLGDDRCGCKQEEKVHTKEKKGKDGKPTRVRTVLNEKQLHTLRTCYNANPRPD
ALMKEQLTEMTALSPRVIRVWFQNKRCKDKKRSLAMKQIHQMQVEKNINGLSGVPMVAASPVRQDLESRSFSTGYHGYGM
HGDMPAGPYMLNYDNGPNSVMSDGIGSPGSDCSLSIPSD
>HM_DN28007_SINE_Six3/6
MTSVNLAALQLYAQHQPLLLTNNNPLLAASAGSAPYLHMPLINFTAEQLATVCDTLEDSGDIDRLARFLWSLPVNPVTCEALNK
HESILRSRATVSFHQGNFRELYHLLENHRFSKEAHPKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKT
HCFKERTRGLLREWYLQDPYPNPAKKRELAHATGLTPTQVGNWFKNRRQRDRAAAAKNRISQNGQFVPNKHLDDEESDFER
DDSSPISIGSNTPPLSPYEDDDIQKPSQNPVMSSSLSFPHKPFMFPHKSPFLHFTASPQHLNPFLLFQQAAAFSKPSPTNSVTSSS
SSTSQQSSPTKPTVFSPVALCNIKDES
>HM_DN28959_LIM_Lhx3/4
MDEISTIKCNTNTSLIAGNGPISSCGLCGDEIFEQIVFSIGDLQWHAKCLICCQCGSLLRDKCFFKANSLYCKEDFFRVFGNKCASC
KKGIAPTEIVRRANDYVYHLGCFECLMCHTQLETGDEFFLLDNQKLVCKTDYHAAKQKEMEEQSHSNKRPRTTINNKQLDVLRE
AYSKSAKPARHIREDLSKKTDLDMRVVQVWFQNRRAKEKRLKKDAGRRSIGASQSYYQSVHSNSNVQGTKRSQRSKQPIKAEK
LNNRKSTFNHNSNFPYLTDDDFEEKMCARSNPYLTPDGGFIKPYLPLSSGLLSPSHHEQFAIAAARLPPPCGDGTGAGNTTQSSS
LPGHPFLQINSPSGQEGQTLLPYSYSTTGFPTQPASELLASFEHQHYPGAHVPPFSEGSLTLHSVLTGARMQSGLNHSGLAGSTP
DSMPYDYTYPSASLLPPASVNSNIPHMLWG
>HM_DN29024_ANTP_AMB
MSVKSRSFSIDNILEQTTKIKSESSCSDANSFVHPNRMTTLSMPSVNMPYLMGPILDIATISNKFQTASSVFTSKQSNQIVRRQR
KPGVERKPRQAYSTKQLEKLEEEFKKDKYLSVSKRLELSVALELTETQIKTWFQNRRTKWKKQMAVRMKLAYHHLWGGLNYLG
PRSSQSATDLSSSMMVLPSNHQSSFNTDIANKDI
>HM_DN29252_PRD_Pitx
MQSSQFVLPNIECGGQKSQTSFGPALYHPLHTPCYPPLTLACTQSQTHSQPALIEEKRVNKEMQELVDSSALIKREYMPPVVNG
KTIVSEEKLAVRGSSPEADSQSSAINRERRQRTQFTSHQLRELETMFARNRYPDMASRERLSAWISLSESKVRIWFKNRRAKWR
KKERNHLHQQMSADTWKTSPPLISTHPFPYTYDPVRQFSQPPNPICQSNPQPVVETSSKGVGMGVVTTTNSNNFHSFRTYNS
AASWNFLASPFQSNQHFAAPFDPLSANTMHSTVPTSYSPGDFFHQPLYYGQSSGLHGDNLLHTSSSDSNCFPLEHR
>HM_DN30383_LIM_Lhx1/5
MQLKSHFQHCQFAPVNEESPSMTVECAGCHKPIYDRYLLNVMDRHWHNQCIQCVDCGVGLTDRCFSRDGLLLCKNDFYRRY
SSRCEACSQAILPNQLLRRVKDHMFHLNCFSCTICTKQLTTGDHIYVTQDNKFICKEDYVMVKNQVVDNDVFTEPERKFNPSSP

FPMRQLSTQLPVNQDVKPSIVKEENNIEPCCVEATPEPPTQTNSDEEIEDEELLGEELEDEIIESDNESVENKSTTASFDDEQDEK
KCEEKTVGSKSKNGNNKRRGPRTTIKAKQLETLKTAFMNTPKPTRHIREQLAMETGLNMRVIQVWFQNRRSKERRMKQLSAI
GARRAAFYHRRLRSYDTRSDLLHQLPYFPTDGSSPDFFHPTHPFTGFDLYSGAPPGTSFLNNSLENASDSLTTPSDIDDPRHAFLP
SEDHRLQLTSEAMLLPAPPTDNIISRYTYSRLHEMRPDPPTIW
>HM_DN30308_CERS_Cers
MSTWFWNERFWLPTNITWADLRSPGGVSNLQPGRLDDLAPALPIAACLLLVRVVLQRLLHFVLPSTQQPTPQPNAILDKVYTSI
ARDPDAVRLAGLAKQLDWSIRQVQVWFIARRALDRPSHAARATDALLRTALHAAIVLQAYPHFGNCTPWELSSCWASYPGPI
PSPGLRRAYTLTLSTLLTFLALAAAQKPPRPWRQLFHHAIAGACLAFSYADNFVTVGALVACILDAPVCFLAAGRLLPLQLLWQLL
YTASFIATRIVWIVGYLLPATLLPTQLGLFPAHFVFNGLLLLMLIVDVTWVWTLLHGLFTSRRDSDLYVSETGSDDEAVSSKETLN
GYAVKSNGHTKPCIAETITSNGDIKPHRRKMASARSM
>HM_DN31526_PRD_Uncx
VLTGDKPQDMPPYSTLGVQQSMMSFSALPYLHASGMLLPNTFHHSPQFGIKSPPSIAAAIKLHSHLQHLHATATRTTKVKTERS
KEQAFSEEAEEDGKRRRTRTNFTSLQLEELERAFAASHYPDVFMRESLGQRLSLLESRVQVWFQNRRAKWRKIEHTRKGPGRP
PHNAHPLSCSGEPIDPTKERQKLLEKQEKKR
>HM_DN32866_SINE_Six1/2
MALNFIYSTEQVCCVTQLLEQSGDIDKLEKFLSLLPGYESLQNNEYVLRAKAVLAYHRASYRELYRLVESHNFSAEHHPVLQTLWL
NAHYKEAEKARNKPLGAVGKYRIRRKFPLPLTIWDGEETSYCFKEKSRNVLRDYYLHNPYPSPREKRDLAEATDLTIIQVSNWFK
NRRQRDRALESRSPPSVDSSGGCVALKTHHPVVVSSLSHDTSEQNSASGEGTLESIFRHDAMVRSTDPHLCPQSPSSNLDSLDG
DDSDSKDLINDCLYNGSYNDSDKSIASSHPFSASEMLQQAAIHSTLASMLMV
>HM_DN34181_ANTP_Barx
WINYKFNMFTPLSVTTDVSNGQLQMQNMLMDFVLKESMESHDKSRGNPVSKSSFLIKDIISPEPNSEKSIENGPYPCGSPNPA
MFHPRLHTVPPIYPVNPVDIITRHDINRKLLYDNNYGLPLHNPIDLSMIGTPPFMKSPAPHSLLYYYKSLSRESERVENIKRCRRSR
TVFTDSQLIGLDKSFKHSKYLNTPERNRLAKHLGLTQTQVKTWYQNRRMKWKKQLIQDGATDIPTKPKGRPRKTC
>HM_DN34160_LIM+ZF_AMB
MSFMGLQQNWYCCELPFFPSRFQPSAWKSNIYLTSEMETDMKNEILEIEKSSTKLENCLPSIDQLSGQTVFPKNDSVNKDLRKF
SCNQCHKAFKYKHHLKEHLRIHSGEKPYRCKNCGKRFSHSGSYSAHTTGKKCIQSPNSYKTNSEHLEESPLKKPAMNGFESGFFP
IVGDQVSAVRYLQLYADSLAALKQQKLEICSDIDQNENHELKEKRLSFHIGNLIGDKCQEKLTAIEDDDSETQRKKRTVFTDFQQK
MLRELYKINSSPTSEEISKLADNIGLSKRSIQVWFQNKRARDKKSRKISNDSFDSDDNCNTDDMKNTRAKDKLFPMFDPLLAEVI
PSRKLLYMNSFQSHNDPVKGMKNLCTRQSFPSKIEAVIYEEPLDLSRKLPLFPENNKSKVSDEEHVLEETFRCSICEKSFQKFSSLQ
RHIYEHSGRRPHVCAICSKAFKHKHHLTEHSRLHSGEKPYECFNCYKRFSHSGSYSQHMSQKRCL
>HM_DN34684_TALE_Meis
MNNQPYDQSSDTPPYSEDVYDPSLRMLPPMGQMRPVASANYAPTQIYPMTQAQNSSHALAHENCKRDKESIYSHPLYPLLEL
IFEKCELATCTPREPGMASADVCSSDSFNEDITMFSKQFFESDKNMFPVDRESDNLMLLAVQVLRLHLLELEKVHELCDNFCQR
YIGCLKSKMPMDLVVEDSDSKTRTFSSSSLEDPPNSVSDEQLSPNSPSSAFYSTPAVSSAPSTYSQPLHLSLPESNFPQPPLCDTSA
HTSPDLIKMSASKDMFPSDRYSSVLPTANKQSRTSSPESTVVKPPTEVCSSPNNGDLKSPTDSNPDSDDPKKPQKKRGIFPKAAT
NIMRTWLFTNLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQSNRAVMGAAASGYAADGQMGAFMIDSQ
QGYLRHSGLPHAMGHEVPSPGAFAPNAPYPTFHPHASPQTAMLIPGHPHHAMMMAHHQSMQHSYHPAQGAASLMDPM
LMDIHS
>HM_DN34694_PRD+LIM_AMB
KDQFPSKESKKTLSKKLNLDYSSVDVWFQNKRARTNSKIFSTSDKVTTAPTPNSLSPSIQSDSIKTSISTGRKARTSITPFQQDQLE
AAFRLNPFPLWSDKDYIASELDLDYSVVQVWFQNKRARSKRKLKKSALDEGFEHKLTIITPSKQRQDENENVSSDTKIFSQELQT
KRINHNDTSTEISHQDSTYFPD
>HM_DN35675_PRD_AMB
MLHDRFLHHSGKASPINLCIESEKMRFSNIQYCPPAISSCQLEKAPRHPSFSSRSANQETKSDNSKENSVYVERNGRKPRVRTFFT
SDQLRNLEKAFSENAYPDVKMRAEICSRLKLRESQVLIWFQNKRARNKKLKQKVDLVASINDKENINIFSKNVSLKTEPLEKNVS
VPYQESINVSDCNLLPKAHQGYASNSPPEEYSLDLHQEFPSSVSISPENFRLKTEIPSNFSVPYQDSASSISTSPDNLALKTETSDDL
SNLYQQSASPQQFISNPSTAHITPFHSLQSYNIVDAYSPLYFAPLNCFGPLTPYQYIPPFYGFPHTSENSSSYKFTNL
>HM_DN36226_LIM_Lhx2/9
MAAVGAAADHQQHQRNKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRNVLKNQ
NSTGIDGKILSPIDIENNENAISAKDEESDEDDLDKLDMSPEDESLVSDFGEFSPDISDSSASPIACTDLSLRIKHEEVAY
>HM_DN36375_TALE_Irx
MVSRIPPGFYSGSNNQMEHPASHFLVQNSSENGPLYPNSLMFAAGNLEFPGRAFSNLVSPSLFSSVPFLTAGSPFPYPDAYGG
VDLNGIARRKNATRETTSALKAWLSEHKKNPYPTKGEKNYAGHYHKDDSNSSEHVVCEC
>HM_DN36375_TALE_Irx
MLAIITKMTLTQVSTWFANARRRLKKENKMTWATRARCETDDEDDDENDTKNSTFDVDIKVSCSDDEFLHSPSTYSNISVSPS
DQKSLTSPKPKIWSLADTAMRPTQKIDDINKNSIFTSHHMAINCQKQLSKSC
>HM_DN36484_ANTP_Nk2.1

HSKLMNPRKRRILFTQAQVYELERRFKMQKYLNAPERESLAQLINLTPTQVKIWFQNHRYKCKRAAKDNSSRGRANSVIEISQS
FVENNDLSINKQVAVKHEKMENNMRCNMPSSGTKMQCTNGNRTQPGQVIQNYPDSEISPYPQTVLPPHIW
>HM_DN36591_ANTP_AMB
MTRYKFQTPSGSIIMEFEEGLQVTQVKTLLSQTLNRQVRSIHTAGLPSNNHSVLADHTKLRDVDCRDTIFVECEPVSDPAVAKPE
DPLITEEAAATPQPCNPSIATETEPQRKSSVDTFLSEQILIPFFSIFGNMAPFCGLGQGLLPFNPVLLQPAVLKYCIHFEIEKHKINIQ
IFAKAAIKGDPAALEKLLELKDQLLDSTGQFLPIITFLLKSDELRIKAYDKVKEELRQIDKEKIQQFLYTAGLQNGLPTTLQSSSNSVIP
IVKPEVPENISVQSVTSSVSPPSDQAKSPVVTSTSSSHVPTTKQTNNFTQASLVNASAVTSPAVAAFTALHLQQAQLLTQQLQQ
QRSALNLSSQMSSNYKPITSVAPTPHHQLHANSTSSHRIRRRTTIPKFAKKVMIDWYLGNGRYLTTEGRQLLASQLQLTEENVR
VFF
>HM_DN36971_PRD_AMB
MGRSNQYKWSQSTTQLTLAVMHCCETNQYATFGYVPMVQPTFYIPDYTADQSRGLTISTEKLLKRKQRRNRTTFTNYQLNTLE
CAFRKTHYPDVFTREELAKKVELTEARVQVWFQNRRAKWRKGEPNDKAMKEFSSDLNCEYFDTKKKAKKSFFINDVLLPTQSS
SYVDEESLGLLSANHFLNTSSQERNIYRMPAWMSPAYSMQQHANL
>HM_DN37189_TALE_Meis
MFALPTHSMKISCTSQKKIFEKEHTETADRDTDNLMLLAIQVLRLHLLELEKVHELCDNFCQRYIGCLKSKMPTDLVVEEKDSKFR
SLSTDSIEDPISCEDELSPQLSSGLASPPSNPVYAALSNPSVVISSATAAPTVSSPNIAISNSNSISPPRTEVSRSSTNSPRSSPEAAN
NVSALDSIADDPKKPQKKRGIFPKAATNIMRSWLFTNLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQTNR
SGKSSRNSALLDAPAEMAAVAYNAEVQSNPLMLNPQHQSYLRHSVMGHALGGHDPNTFSNCPPPSAYHSFSNAPPQTAMLI
PGHPHAAMMMAHHSSLPNPYSHPQFI
>HM_DN37226_HNF_Hmbox
MANQPHLLNGDLKSMNPEEEPKYTIEQIELLRRLKTSGLTKYQVLHALESMDRIEKSAFATTPMETSTSHSLPTISEAPKINSIAPSI
NSIQSVQNDYLRSLALQQTLAAVAMANNHSLLNPLSQLHHNSGSNSSLSGNFANPQSQPPTSSPPTSFTAYSGSIGTIKRTDSYY
NPTLGDPSETNDDDKELIMLQAMPVFDTKEEIKNFMLTHGIKQHQVAEKTGVSQSYISQFLHQGVWMKDSTRTNIYKWYLAE
KRKREGVNGTSSPPFSTSTTAPALSSAKMYMNGNNGTPRISQFKKRDRFVWKDSCVSILEQFYQENSYPDDLKREEIAQACNN
ALELQGQFLNDEKLVTSVKVYNWFANRRKEIKRKQQLELEQNVKNEPSPNSSDCGDSNYESGVSDGATNCQDPISAELDAVN
HSIMALVNPSPVQNCAPVQP
>HM_DN37527_PRD_Pax4/6
MPQKGHSGVNQLGGMYVNGRPLPDSTRQRIIDLAHSGARPCDISRVLQVSNGCVSKILGRYYETGSIKPKAIGGSKPRVATSQV
VSKVALYKRESPSIFAWEIRDKLINDAVCTIDTVPSVSSINRVLRTLQNEKLKMAIPKDSGMYDKLRLFNTHGWVAPGSIPGPPPS
SWYQTPSIYPSGAPALSQFPSPTDIKPIIDERSITVSSPESISSDDLKKKFQRNRTSFTQSQIDALEKEFERTHYPDVYARERLSESTS
LAENKIQVWFSNRRAKWRR
>HM_DN37578_TALE_Pknox
MMHVQTAPFMLDPGAMVTSTGPPPPPSILVPSVGNTTTQNSLRTSTPLGGNATPVPNTAVSGTSLGITDSQSFETDKNYIYGH
PLFPLMALLFDKCEQATNSVECPSSASFEIDIETFIHQQERERKPFFTENAEVDSLMIKAIQVLRIHLLELEKVNELCKDFCQRYIAC
LKGKMHSENILRTDIQVYPQSSSPNSMQVSPMQSASTQPAQHQMAAAALSTQQLLYPPLTMLTPAASHPGGPGSLINSAGAF
MHHAPVVTSHSGYPLPAHVSSLAPMIHGGASQLVPHSAAAGGHLMHPQASLISNSLHDLNPDGSMKKTKRGVLPKQATQIM
RSWLFQHIVHPYPTEDEKRQLAQQTSLTLLQVNNWFINARRRILQPMLEATNPDTSLQKAKKSKTVARPNAANRFWPDNLAQ
GQVAKIEPSSPNDDLSVGSSYGPDVGNSISPDGLSSSSVSPPTVGTPSPSQQSSSMNGQLHISAQSPLQSHEQSGAIFHDTSPSL
>HM_DN37913_POU_Pou6
MEEPSSTSNFDTPALVDVEALIAHASQPSSISIHSSSAAPPQHINISGVEATISPITLSSQIVNDSQIIVIYGINGTQQIFTIPVSSLKSG
TSFNSQLITLPATQSSTLDNISKEADDSKTHHIQNDNSVPSPVKEFDENLKVSADNQVPTSDSTNDQCVLLEAYTTEELINMAIDK
AIDTIGSSTSPYLPQKYSPIAPDTPEARAHFESRIVLKPIEPTKKITIAENPELEYLDDVNIEEIRQYANFFKRKRVCMGLTQTQVGE
MLSVDNKTVYSQSAISRFEKLDVTPKSGQKLKSILEKWLSQCNYKSVNSEEAVDSAMDQRRSRKKRTQFSAETTIFLNDYFEKIT
HPRGFQINEIANTLQLDREVVRVWFCNKRQLMKSKDNSDCPSAESMEKFFRNVSICGLQPISPDPFDEFSDSKNEEEESNCKPV
TPVKSSTPFNSSSENKSDLIHNEDVGFSSDMDLSAFKVYAMENILPIEFEEV
>HM_DN38072_PRD_Alx
RNPEASICKLSSQSRLSEEDSRTESSENKDESFDKDDSSEILANDSDDFADKIDLNCSSRKKRRNRTTFTNLQLEEMEKIFQKTHYP
DVYVREQLANRCELTEARVQVWFQNRRAKWRKRERSQQLQIRSLNHNSICEQTPLIPTATPRSDPVTQHLSSSWPYGSSTSCS
TARTGVAYQALNSATQNFVSSNPFLSSYMSAAMASQPGHFAASPVAFTSQQPLSGMNLINSSFSPSLMGAAASQFLPSTPMD
FFSNSASQFDGRSTSLAALRLKAKEHEASLGGIVSFAKQAIQ
>HM_DN40002_CUT_Onecut
MDSNNSSSNTASGICEQLQPSAAFFAAAVSAPNATTGSFYPRYGDMGSTQRHSTHHDLFFDHLVGDYRNEFYSRGSTPSGYSC
FSSLKPPLPPVSSLTEASPRDFCFTPADANSFYHHHQQYSNAAALAATSAAAMHCSAPDLKPYGTTPPTQQQYFYSPCDHANY
EPYCYNPRSAQLTCLDTTNSSCITRLPCRETNNSKLFEMGLLTDNTCPSPDLKPSRKNSYKSSGNDKASRISASPGKDLEDMDTK
EIAQRITAELKRYSIPQAVFAEKVLGRSQGTLSDLLRNPKPWTKLKSGRETFRKMGKWLKEPEFQRMSSLRLAACKKKEIDQVQP
KVHHPKKIRLVFTDIQRRTLYAIFRETKRPSKEMQQTIGQQLGLDPTTVSNFFMNARRRCMGDRWKDIDDNSGKSHQHRSHG
SKSDHIRKHPIHNGAILPILTPSIMMSH

>HM_DN40476_ANTP_Tlx
MKTNFSISSILTGQSDDEECCGSKSPFSRFEDDDRKTSIHSQLLINANESKDSPFTYSYTFPSTKTSSLTNFTPDIIRSPTLPTNYSQS
AVYPLPNPVMWLEMNQNILRDRLLYNMARRIGHSFPTRSVPKRKKPRTSFTRLQICELEKRFHLQKYLASTERASLAKSLKMTD
AQVKTWFQNRRTKWRRQTSEEREAERQAASR
>HM_DN41482_PRD_AMB
MTCNDFPVTKMLITPPSYNRYNYSNILKPAADEKSFSVDTVSNYRKKKSRYRTTFTVFQLESLEKAFERAAYPDVFAREELARRIN
LSEARIQVWFQNRRAKWRKDKNLNSISPYQNGQKANRNLKSVFQLFPAPKLPAIPSNSSLPPSVSYLACLLGLSEVAYPSINPYL
QQLNAAMILHQQQNSNLLQFKAL
>HM_DN41603_LIM_Lmx
MTIGAFDQSPCDIKPICCEGCRGLIDDRFFLQVQNVSWHERCLKCAICQVGLDSTCFFKNDIPYCKTDYERLFSGKCGSCFQTIPA
NKLVMRAGDIVYHLDCFSCTVCHQQLRKGDEFVLKNRLVYCKIDYEKDAASLITDFKSENLNSPVNTYPSKATPKSECSDFLSDE
DKSSKEDELKENHDSYEEDDELCEKDDRSSGHQKPCRKRVILTTQQRRAFKAAFDVSSKPCRKVREQLARDTGLSVRVVQVWF
QNERAKVKKLAKRQQQQHHSQLDQDSPHRGLEDVPFGMYIPDQFSSHMEIRFNNDIMSPGDPSIPFPPNMVPAINSEFFNSS
FCSALPPISEDVAVSAADIAVPSTCMSEQPLKTVSIDSNMAPPTIMSKNILNPIDKLYSMQSSYFTQLS
>HM_DN42050_ZF_Zfhx
MLKQEVLQFNNEENNQSPNNSSALTTTLCTLCNSNLPDVKSLQQHMIVYHSAFDIKNFAQSLLQTESAKQDSNVSKEKGLTGS
VHQVLERSEISKMERLLSQNPALMTLISNSPERPDKEKQLQELVQTLSQQIQLQQNIILHQMMALQQLSQKDNQTVRHWFDP
KQLRPDSEVLNRFQMQFPTRKNIPTHGIAMDNSVNSAHKKIFKGINRSCRRDHRDPRRCRVHKTYHSHEALLSPNYLALQTNC
SVPVRVAETSPSANKTEPQEEPIDLSANKKEEPQNLRILSHSEQEVKFTCNRPVARGRSQIISGNGFVSIGLLCKPDYKEPHKCMI
HRIRHRVRTRLTAQDLLYLHNMYRRTANPSRRLLAEISNHIGLPLRVIQVWFQNTRRLMKNTCACGKVYVTNKLFDEHLKECTS
RAS
>HM_DN43016_ANTP_Nk5/Hmx
MTMAWVRKGRMVNEKKKTRTVFSRNQVFQLETTFDLKRYLSSSERASLAATLSLTETQVKIWFQNRRNKWKRQLAAELEAA
NIANASSNQNQRLVRMPLLFAESKSSGDGVLTNSAVNSASTNPSLLHINYPLTAYYPGLAGNPSLPMVRQNALNSFV
>HM_DN43708_POU_Pou2
MVCNCAREKIKLAAYKLNFSGNNGVNRSLEILIQTDSDPSEVDRARQQATEMLLAAQQDIIQGLQDELTKARSSDEHSDTFDKC
DESERPAIIAAHGTSGDMNIPVKEIVDLEDKLKNSNLPLSVHPFVVLHPSATQAKPQLVQPSTMQQLIILQPSQATSLLQAPLSSF
AMPYNSVVTPTPVSSSQFTSQSDSTSTTSESFTTNASSPFSSGEPAQLEELEKFAKDFKTRRIKLGFTQGDVGLSLGKLYDNDFSQ
TTISRFEALNLSFKNMCKLKPVLQQWIVDAEKHMQMIRASATSNSTSEEGKSLSALHDPSYKRRKKRASIETKSKLALEKSFSLNS
KPTAEQINLLADQLIMERETVRVWFCNRRQKEKRVPSSEMSTSAEAESAAFNSDVDVANSPGFVIDATMSGSSST
>HM_DN44693_ANTP_Nk2.1
MSLSPKQSSFSVTDILSQVDPLQHSRMDSSFNTFAPYPRQNIQSNSPYMSQLGQSMMGSYCTQSDLTPGFSTGFPSATAGLTY
PEHFARHTAASNWYNGASDRLGFPRLVHPSSVTGMGIGFGPDSTKFMGATQQRRKRRVLFTQAQVYELERRFKQQRYLSAPE
RENLAQLINLTPTQVKIWFQNHRYKYKRASKEKSGESQPLDSNIILPSSPAKDLCIKNEKSISSGSNSDVQSPESVQSLDEIKESEYG
QAVLPPTHTNSIDYNPLPGYPISHDSLQPYAPANLMAAATRIW
>HM_DN54918_ANTP_Hox1
MDAIGNVQSTTHHNHYRSNETISSEHGKWDFSDGYIPSEYYAGHQYVISPSAYHQNCAYTDFNYASQVNYYDYHSTYCTPPSP
LNSEEFRQVYYPTENAYRQPDASLEENRSPSFDPKEQICSNRVINTFTSNPHFAKYSTTTPPCFYSPISLQDSSPNDQLCNTGKIPD
ASPSSNSTQNSSTRSPASSDVISPCDTSVHSAHNTYSWMKIKRNQPNALKSAQQAQFPNQTRGGRTNFTNKQLTELEKEFHFN
RYLTRARRIEIAASLNLNETQVKIWFQNRRMKQKKLLKEGKLGN
>HM_DN56217_PRD_Otx
MTYCSPQCDNFEAVDEYAFIYGCQTQLSNSGKNSSFKGEEHSFVKSTEKPRRERTTFSKNQVLYLEELFTKTRYPDVFMRQEIAD
KLHLTESRVQVWFKNRRAKSRQQTMHSNFKIENLSNLDKNECISRKLEFLNSNSLQCREINQDFPEMNSTLEELSSSIQGHFKYT
>HM_DN58284_ANTP_Evx
MTKSSSFNIASIMEKDPQPKCSAFSKSVDPGPMRTQTALYYQLSDSEQLRRYRTAFTKEQQDKLEAEFKKENYISRPKRCEIASEL
NLPETTIKVWFQNRRMKEKRQRQMYTFSPFVEQTYYTYLLAKGLLPNPMMAQMMPPTPSISNGPLTNIITRYQFTPFQNFEH
DKQFSNVLQNCIAE
>HM_DN58320_ANTP_AMB
NCVSVEFDNLHSVGKYNQLPFHVRMTSFKIDQLLLPSMPTSKSTDTQLASVSPGYSPWKYCDGIGDKIKMDEERQSPKSDDNQ
DCSSSDTNEPRRMLIKDTNGVYKEFLFPRGLDLDRPKRERTSFTPQQLYQLEKEFHKNQYMIGRDRAELAVLLKLTETQVKVWF
QNRRTKYKREKHRETEIRNSKANNFAACNILKLLEQKQISSNYPSSGTFGGLDVPNAYPSAGNTISPDFMLPNSQRSLLFSPFFW
AVNGTV
>HM_DN66848_SINE_Six4/5
AHYSEAEKIRGRPLGAVDKYRIRRKYLLPKTIWDGEETVYCFKEKARRKLKESYKKNKYPTPEEKKSLAKASGLTLTQVSNWFKNR
RQRDRAPTTKSACNSDNEGELCDTNKCEIKCEQSTMNSHQQQSNDVARELAMSAFLSGTFPYQTTPPS
>HM_DN68850_ANTP_Barhl

MSLKRQSATWKRSFLINDLLDEENSSDTESVHADEYAHLGKKKTRKARTAFTEEQLSKLEASFERQKYLSVQDRLDLASSLDLSD
TQVKTWYQNRRTKWKRQAAVNLELFSTKNIPFSYQNSLMRFNGPYQMKQPMYLHSNMLVLKLFADMNERQTAVTQAL
>HM_DN79735_ANTP_Nk3
MTSSLPVSMAKEVDGELIPKFSIDHLLQKNVNLEETRESEKCSDSMYLPHLKQSFIESFSNSNPVNSHFPDMFESNCSKKRKKRT
RAAFSHAQVFELERRFTQQRYLSGPERSEVASNLKLTETQVKIWFQNRRYKTKRKQMQVDAASTLLSPACDTQYCKGPLNMLV
SKHGNQSTQSVTQSDMSHCLQYPTPMASYLARLLPAQYLLGSLSRSATTPFINQLDYGTFFSNLIKATKPKSIETMSSPDLPENS
NSAI
>HM_DN80881_PRD_Vsx
MNKLESTQSTGFGIKEILGLKEPSNFHSMPYLRSPTCLWPLPLLTGGNMWVEPILNSELSQSLLLQSHINQQTLIRLQRNGQTPS
DLSFESNFTKFSQHLSDDICMVTTSRKKSRRRHRTIFTNEQVDELEKIFQTAHYPDVNAREMLSLKTELPEDRIQVWFQNRRAK
WR
>HM_DN83548_ANTP_AMB
MSGKKTSSFLITDILKREYDESSEGEDNEAESTYDESEDDEFRKLEVYTNHLNDAFSSEIEHIKSGVVSKKPRRRRTAFTQAQLNFL
ERKFRCQKYLSVADRNEVATQLALSETQVKTWYQNRRTKWKRQTTPDDRMQELENHSCNEPDCPGNLRTRQTDLLRNPIASL
AEMGIAGSALKGYDENLLNHSTNSILSLPHLWSTDWYKRYFPNPSKINDPSTARLLMRTDLIKNLQNYIYSPMVRNQSV
>HM_DN90735_ANTP_AMB
MFDMAAFISEMKSSNKSPNQKTRKMGFSVKDILGSENEDDSNIEIDDFVDMSSDDESNSTEYRTIEGSDEIHNSSRTGILPTSM
NDLISIRKPKKRKRRVLFSKAQTFELERRFRQQRYLSAPEREHLANLIKLTPTQVKIWFQNHRYKCKKNHHEQRLNYENFLSSSYGI
NASKRYNVQLINKDSAKSFPALSFQCDEPSFSPNTASNFHTAKSILPSHLQVPFLPTQTNMSFLANPSPANFSHLIPQMRAQQLS
ELVKLLHYQNNIISQQQHVKLLKESKFTASSKTTVTQDKD
>HM_DN91365_ANTP_AMB
MTSNSELFLISNLLKEDDHYKFIRKKQRILFHHRQIEQLEDFFAAQPYISAIERQCLSQKTGLTNNQVKIWFQNRRYKEKRLKALQ
TNMQQPKSSQSNGITLHHVCQLFKEAELYTNKEKQERAPENYFQRLLEFNNNCHNNNKPRL
>HM_DN104433_ANTP_AMB
MVISTVQCSLQAVLKKQVQTMSFTIEALLAPESPPNKTSSDEIITGFKSYAWIDCSRYRPPKIDRYKRKDSEYKKRTSRNPRIPFTS
NQVALLEQTFQKKQYLSSMDVSDLSNSLRLSDTRIKIWFQNRRARARRDGFNGAFTVQLMEPHDKLKSLESSTSPSSTSLMKQL
AVTSASSTSSQSSYSAAMAAA

Protein domain identification

trinityID class	(homeoDB blast)	protein domains (Hmmer search)
DN24356_c0_g1_i1	PROS	HPD Mit_KHE1
DN40002_c0_g1_i1|m.24477	CUT	CUT DUF620 HTH_38 Sigma70_r4_2
DN10420_c0_g1_i1|m.2841	PRD	OAR
DN35675_c0_g1_i1|m.16721	PRD	
DN37527_c0_g2_i1|m.19291	PRD	HTH_23 HTH_29 HTH_Tnp_IS1 PAX
DN41482_c1_g1_i2|m.28797	PRD	
DN31526_c0_g1_i1|m.12245	PRD	XLF
DN27253_c0_g1_i2|m.9227	PRD	DUF3437 HA
DN23062_c0_g1_i1|m.7235	PRD	HA
DN7688_c0_g1_i1|m.2075	PRD	
DN24659_c0_g2_i1|m.8154	PRD	
DN36971_c0_g1_i2|m.18384	PRD	DUF1577
DN26989_c0_g2_i1|m.9094	PRD	CDC45 Pal1
DN38072_c0_g2_i1|m.20141	PRD	OAR
DN15356_c0_g1_i1|m.4176	PRD	HTH_23 HTH_29 HTH_Tnp_IS1 PAX
DN29252_c0_g1_i2|m.10492	PRD	
DN27245_c0_g1_i1|m.9220	PRD	
DN56217_c0_g1_i1|m.41307	PRD	DUF3408 DUF413 HTH_23
DN80881_c0_g1_i1|m.43410	PRD	ATG2_CAD XRCC4
DN20170_c0_g1_i1|m.5558	PRD	Upf2
DN37226_c0_g1_i1|m.18864	HNF	HNF-1_N HTH_19 HTH_26 HTH_3 HTH_31 HTH_37
DN28959_c0_g1_i1|m.10298	LIM	LIM Vps39_2

DN36226_c0_g1_i1|m.17408	LIM	Homez Retrotrans_gag
DN27964_c0_g1_i1|m.9540	LIM	LIM
DN30383_c0_g1_i1|m.11355	LIM	DUF3408 DUF581 LIM zf-C2H2 zf-dskA_traR zf-HYPF
DN41603_c0_g1_i1|m.29404	LIM	DUF1510 LIM TBX TNFR_c6
DN14479_c0_g1_i1|m.3858	LIM	LIM
DN1193_c0_g1_i1|m.357	LIM	Cytochrome_C554 DUF1364 LIM MerR_1 XPA_N zf-CHY
DN9855_c0_g2_i1|m.2527	LIM	Homez LIM TNFR_c6
DN34684_c0_g1_i2|m.15334	TALE	
DN21081_c0_g1_i1|m.5973	TALE	DUF1319 HTH_19 HTH_3 IMD NUC130_3NT PBC Sec15 Syntaxin
DN36375_c0_g1_i1	TALE	Homez PGA2 Phage_P2_GpE
DN37578_c0_g1_i2|m.19358	TALE	
DN21447_c0_g3_i1|m.6190	TALE	HA Homez
DN24763_c0_g1_i1|m.8213	TALE	Homez
DN37189_c0_g1_i1|m.18754	TALE	
DN12757_c0_g1_i1|m.3416	TALE	
DN33495_c0_g1_i4|m.14054	ZF	
DN42050_c0_g2_i1|m.31057	ZF	zf-C2H2 zf-C2H2_4 zf-C2H2_jaz
DN30308_c0_g1_i1|m.11388	CERS	PRA1 TRAM_LAG1_CLN8
DN28007_c0_g1_i1|m.9588	SINE	Homez HTH_3
DN66848_c0_g1_i1|m.42105	SINE	CENP-B_N Homez HTH_3 Perm-CXXC
DN32866_c0_g1_i1|m.13539	SINE	Homez
DN7261_c0_g1_i1|m.1968	POU	HTH_19 Pou
DN43708_c0_g1_i1|m.40060	POU	Pou
DN37913_c0_g2_i2|m.19890	POU	HTH_19 HTH_3 Pou Terminase_5
DN20967_c0_g1_i1	POU	HTH_31 Pou
DN79735_c0_g1_i1|m.43259	ANTP	
DN27605_c0_g1_i1|m.9399	ANTP	Nuc-transf
DN10228_c0_g1_i1|m.2704	ANTP	Homez
DN91365_c0_g1_i1|m.44211	ANTP	DUF4332 Homez Phytochelatin USP8_dimer
DN8401_c0_g1_i1|m.2232	ANTP	Homez
DN83548_c0_g1_i1|m.43646	ANTP	HA
DN1275_c0_g1_i1|m.391	ANTP	
DN36591_c0_g1_i1|m.17917	ANTP	
DN68850_c0_g1_i1|m.42447	ANTP	Bul1_C DUF3206 Homez
DN54918_c0_g1_i1|m.41056	ANTP	
DN40476_c0_g2_i2|m.25716	ANTP	Homez
DN14515_c0_g1_i1|m.3892	ANTP	DUF1743
DN12035_c0_g1_i1|m.3241	ANTP	CITED Homez
DN44693_c0_g1_i1|m.40284	ANTP	Homez
DN90735_c0_g1_i1|m.44111	ANTP	
DN104433_c0_g1_i1|m.45397	ANTP	DUF4476 Homez
DN58284_c0_g1_i1|m.41535	ANTP	Terminase_5
DN18872_c0_g1_i1|m.5077	ANTP	
DN3930_c0_g1_i1|m.1095	ANTP	
DN58320_c0_g1_i1|m.41539	ANTP	Homez
DN4748_c0_g1_i1|m.1255	ANTP	QLQ
DN12978_c0_g1_i1|m.3456	ANTP	Homez Oxidored-like
DN43016_c0_g2_i1|m.36485	ANTP	
DN36484_c0_g2_i1|m.17692	ANTP	
DN15813_c0_g1_i1|m.4392	ANTP	

DN34181_c0_g1_i1|m.14867	ANTP	HTH_28
DN20211_c0_g1_i1|m.5588	ANTP	Auxin_canalis Homez
DN29024_c0_g1_i1|m.10351	ANTP	
DN3061_c0_g2_i1|m.947	NTP	
DN24302_c0_g1_i1|m.7955	ANTP	
DN12216_c0_g1_i1|m.3286	ANTP	Homez
DN34160_c0_g1_i1|m.14868	LIM+ZF	DUF2180 DUF2321 SIR2 zf-C2H2 zf-C2H2_2 zf-C2H2_4 zf-C2H2_6 zf-C2H2_jaz zf-H2C2_2
DN8120_c0_g2_i1|m.2184	LIM+ZF+ANTP	DUF1319 DUF164 LIM
DN34694_c0_g1_i1|m.15382	PRD+LIM	SOBP

4 – *Nemertoderma westbladi*

>Locus_49193.0_Transcript_1/0_Confidence_19_Length_3044
MLISERSIKGEELSSQLYQFSGSTGNLDQHLHSVTNCQNCPPPPAMATGAHQKSWSGYPSSYQTFACGKSDIQI
PCPIGDPVMEPEYSYYGYFPSAPAPFPPSNAPFPPANTNFTHTNAPFPPPNTSFPSSNHTPTFSNPPSTSPTLTD
KLPGKSSHWSDYVTSTGAQYCFEPSATSSGGYNPSWINPSDRFSGGFPLYLGGSQSGHFLSGYSGSGSQRYPL
LDTRESSTTYRVNPNNGKTRTKDKYRVVYSDRQRAELESEFRCNQYITIKRKTELAGMVGLSERQVKIWFQNRR
AKERKVARKYTIGGK*
>Locus_48679.0_Transcript_3/0_Confidence_10_Length_1102
MATTNITAESSESRRLGRHYVSVSETAEDDSLDDSQRRYRTAFSRQQLQRLEREFVKENYVSRPKRCELASSLG
LPETTIKVWFQNRRMKDKRQRQALLWPYLDPYYYAYLVSRGIIPVTSPPPGGLPGLPEVTPTIAEYRNKLPTQRLA
PQFHQSGELQAVFEFPALSRESLFRTGDYTFPAGYPSCHPGVPLPPVLNPQP
>Locus_36518.0_Transcript_1/0_Confidence_4_Length_1293
MKYYQTCSMESTILPDDQVGCGRLNRQLDPQTNEKREDDSASMDDNQRRYRTAFSRQQLQRLEREFMKENYV
SRPKRCELAATLGLAETTIKVWFQNRRMKDKRQRQALLWPYLDPYYYAYLVSRGVIPMGTTPPVTSALMATTEQ
PAGGTLDLRQKTIHQSMTAPRIMLPSEPQMGELQAVFEFPSLHRDPISMMPAATAVPSSRDYLFTGFPGQSAPG
MYLDPRLLNSTTNYVAVDSSNYPTVSHSAVTPQKRNTDFTMDAIIGNKNTKSIHQHRSTLTTQ*
>Locus_35873.0_Transcript_1/0_Confidence_5_Length_1977
MNDNSVEYSSFPLSHPSDGSAMNEEEFASANADYRYQKHMAQFGWLPHSYHDYNASTWYQGHHHQALSTHY
PNGYYPATPMDTYGVVDNDRPSEPAEYHVVPNARLFNADSLPVTNNNPTQLYHGNAPYESGSPDVKPFTDSYIN
GTTEPPHQVYSWMRLKRSQPNANKSHEFGYAGQTRGGRTNFTTKQLTELEKEFLYNKYLTRARRVEIAASLNLN
ENQVKIWFQNRRMKQKKLVKEGK*
>Locus_54694.0_Transcript_1/0_Confidence_0_Length_1248
MYTSTTVSEGMINSWPGIAPYYDTEPLGSVRYTNLDNRDSWLINSRSVAAVDPSSCNDEKSGTGLPRGGTAAATI
HYTTGYQNNYSQYGNNLYGNNEDAYNYSTSLFPQHHFTYAADTIDEAVKHSRTLPSYHGNSIDCTTGRDSVHQP
LSHASKESLFALERTRGSDNYVAMKYCPVECSTQQTIAPADPASQLHYQSPHEYSSPDQPEDNSNGNWMTRNP
SRKKRRPYTKYQTLELEKEFLFNTYITRERRSEISRALNLSDRQVKIWFQNRRMKMKKMTQQHNSSSSIGIAGYQ
TLCL*
>Locus_10725.0_Transcript_1/0_Confidence_0_Length_314
SMKYYPVDCSVQQTIAPAEQPSYLSSVQSNVEFSSPELSHTDDNAGSNWMTRNPSRKKRRPYTKYQTLELEKE
FLFNVYITRERRSEISRSLNLTDRQVKIWFQ
>Locus_58015.0_Transcript_1/0_Confidence_0_Length_736
AEYMGDYCADGIPNSANFPFFSCENFSEINNNNIFNGSITAPSNDQFIALDTNCDSVWMEERNPPDQQPIPNCGI
QMVNVVNGVATATQFIYPVAITNRNLHVSDNLEYEDADDNENDAFNKRKDRTAFSKFQIEQLESEFETYNYLTRL
RRFELSKKLSLSERQVKVWFQNRRMKWKRVKGGVAREEQDKKNFLERKMKRISDMEYGLS*
>Locus_58481.0_Transcript_1/0_Confidence_0_Length_480
MSIADLDNWANMLQKCDFASVSGQFSGAQLPLIANLLLTRGYQIGGILGKSRRPRTSYSSKQIKILEEHFVLNRYL
SRPKRFEVATSLNLTETQVKIWFQNRRMKLKRSERILPHTNGSSK*
>Locus_6133.0_Transcript_1/0_Confidence_0_Length_867
TQSTWESAYLLEHIAFYLNTCCNLYNKDLKTMIILIKVLFLKIFLLAIFLAREYIRLSMDLIRNQRPSKCLKSELLSPDC
NQIETITPTVVQSIFTSVEKMCESAELSASRISNVDLLSHPTMQDIVLSNHSSFMRNHLSYIPTLHWRALVRPDNKN
RRQRTTFSSEQTLSLEVEFQRSEYISRSKRYEIAEHLHLTENQVKIWFQNRRAKDKRMEKIQIDQQTRYIASAFAG
LCQSRLPLPY*
>Locus_44999.0_Transcript_2/0_Confidence_12_Length_1575
MSSLVQPTFPLQFSMNNPPTLGLMTPQLPSLCCQGPSRASPFPMHPCDRLTPPLPYPIYPYRYNQHQLVAMDTN
CDYQSAYRNKKCRRTRTVFTELQMLGLEKRFDMQKYLSTPDRLGLSDSLGLTQLQVKTWYQNRRMKWKKEILIC
GGIHVPTKPKGRPKKCQPLEPPGINDIKLADLQSRFRNSRKNNTSEM*
>Locus_50937.0_Transcript_3/0_Confidence_47_Length_2589

MVAYSYSINGMGSPPSNNIQHRVNPTMIPYDTDMGRNAAFMEMYHQPGGYGGNALVSPSPWQHINPAAAHMS
HLLTAGGPGRMPVASAISSNPSHAGNMLNPHDQTVFHNSMGVMNPNGFPYNPYGSNPSQFTINHHPPSLPRSH
PHQPSPVGLGYGYHSQAPPNYPHSTLQPPHHPSSAYMATTVRDSLSISPDSKEDKLLHDPELRLSSKGKKLRKP
RTIYSSYQLSMLNKRFQGTQYLALPERAELAATLGLTQTQVKIWFQNRRSKFKKLLKADPDAMPSEAMQTLLDED
MAGVDDADENAESKPVKQEYLGGQQPQGWDQVLADSRTLETEKHYHGNPTQFPRGSPSLSGNHDNGGGPPS
GLINASYGLTIGNSPLTPLGGYGQFPNPWYSPSQDNVPHLPLLT*
>Locus_54562.0_Transcript_1/0_Confidence_0_Length_921
MVPVHVCRRSSATGRHIRKNDGVIFSNKSTKLCSAKHFIWLQKYISWAACNNINSTMASLSEPPNHAMMEDRKFS
SFLDYSANQSSSYTPSSPGGQVCAYHPPNVPSMYGTMQPYQYHPHGNYRDAPHHPHYPSSAAPYQGSRHAIG
YNGIHYNPLLGNVPSSPITPNVNITNITHSYHYMSNSLQPPTTSQAQSRDDLMFGNCSSTQANTMLRCSLPGADQ
TYESKLMEPQLESDNLHIISHKGKKMRKPRTIYSSLQLQILNKRFQQTQYLALPERAELAASLGLTQTQVKI
>Locus_35817.0_Transcript_2/0_Confidence_6_Length_1968
MVVFNPITTNNYTMNTHHHLGINSLVNHPPTMLPYEAEIGRNATLMDMYQQQPASYGGQSLVNSGSPWQHINSA
AAAHHMSQLFNSGQAGRQPVHSSSAGSAPPGLGYFNGAGEQYPNTMAMMNPYSPYSGSPTANQFTINHHPSQ
QRSGVLPSHHHPSQSSNLPPYHHGGPYPTPPSHGYHHQSLQGVHPASSNYLSPRDQMMDMKDEKLIDGELRQ
TGKGKKIRKPRTIYSSYQLSMLNKRFQGTQYLALPERAELAASLGLTQTQVKIWFQNRRSKFKKLMKADPNSLTS
EVMAKLMEEDRRTSQRDLTSVDSDNETSEESVKSVNIKQEPYAWNNSVLEHNNLDNKLTYQDNNLLPPIGELAN
NNASNSINPPIGTREDLRANINMSVTASTAVHSDNNSMLNYSHHAGISLVNSQQPGLSLGNSPMGGALSGFGQF
SAPPWYNLNQENLPQLSMLT*
>Locus_17210.0_Transcript_1/0_Confidence_1_Length_1444
MVTCEQSVSNTSFRSAMFTIDQLVGTSTSDTSRSSVSMTTGEPAHSLHNGQSVIISPVSPSSTSSYEKPVSRGLV
NSPSRYFLYPNLPPIPNYHSMLTTPMNPVDHARYWGCYINSNSYNSGTLPPVSMATLPTPTQQHPPVVHPPAVY
PTAGHHPVVHHPAVHYPAVSPSLPLPNFPMHQMSQLAANHHLQSSQLSHPNYHLLQSSLIGSPVYRHPILPSNN
MQCYSSAVNGSLGDLLHRKPKRIRTAFTSEQLQRLENEFTSNPYIVGEERRQFAASLRLSETQVKVWYQNRRTK
DKREIAKDEKGCEDEEDCEETGGEPRGENCGDIRGESCRDIGRENDRDGRETRELTGRDIQVQTSRDFRGENC
RDIREQNSRDESRDIDNAIDVVNGKYENNKELNVKSRKRLNSFTEDNNMKLYKISSYNELLYR*
>Locus_44961.0_Transcript_2/0_Confidence_9_Length_1683
MVLSVSPSTMFTIDSLVGPSQVALSQGVRTIPVTPSVPVLATTPRMPHNASGGTVTTPPTIPTTLQCAEYSRQTYL
AHQAVLESYLGYPTLPSLPFTTARSPHPTPIELLQTTGYPHCHPHCVLPTCPHFLPGGALACHHGQIDVSQSLSST
NLHTSYYQSQLAMSPLQLARYQSQLAAPPLQLANYQSMMTPHHPNATYIADLLHKPKRIRTAFTTEQLQRLEDEFI
ANPYIVGEERRRFAANLNLTETQVKVWYQNRRTKDKRETSREEKRENSGVENLGNIGEEKRGSWDGKRETVKIE
SRHELIESEEIVNTQLGSNQSGESVNSQLGGNQSDISNLSRLYSDSNIESTKLRHGSFTNGTHGNSFQCSRKRKV
IESADASGTKNSRIVNTEGKLDIDVSSTH*
>Locus_45472.0_Transcript_2/0_Confidence_5_Length_3205
MSFSPKQTSAFSVSDILTPIDDPFNYKRGMDGLTGPSLGSYRTGPPSQNPMHMGAAAAMNMAAANSAYHAAVH
GMSQLSHPGMGSGCNGTTPSMNELGHYNPYTDPRSSYTGWYGASAAADPRISSEYWASMAGMNMGFSADYN
KILPSQQRRKRRVLFTQAQVYELERRFKQQKYLSAPEREQLAAMIHLTPTQVKIWFQNHRYKMKRQAKEKTPHD
GNQGQQQQNSPRRVAVPVLVKDGKPCTGGANSNTTTGNNNISTTSPTSSNRHQSITNIANNNTAIIPNYSSNFAV
HKQHLDVDSSGGQHRSSCSLNNIAAAGSIAAAGYDHNALASSYLLHGRAW*
>Locus_20731.0_Transcript_1/0_Confidence_1_Length_661
MAAYKRGFSVAELLDLNENADGVPESSRIALTIEEEQRNKSMLLRRGILAAFIREEFANLHSSKPEGSSRGQLWTP
NQEWERRCDQQPQGDTGLGGCGDGVDVSDDTDEDIEVGQEMQTSEECFDGVMEDVDEDGDGTPTGNKKRK
RRVLFSKQQTYELERRFRQQRYLSAPEREHLASIINLTPTQVKIWFQN
>Locus_36577.0_Transcript_2/0_Confidence_2_Length_1846
YRKVNFSLPISVTMILNMENQCPIKEMQSETPENPKKLKLSSPEEEPRRKTIRTSFGIDDILGLSKPGPPSKPPSNH
SAFTPPVSTLPVHPLATHIPSGMASGLLGYQLGPLCPSWLACRGPPTVSLPTLLPPPQFVLPSSGNFGRSGEIVD
TSSGKLFPSSIPFAPNNESEIEGPNDELENKAESDDDELESIEEEVDISDKEDKLDDSKDSDQENKRSYRKKKTRT
VFSRHQVYQLETMFDCKKYLSSSERAGLAATLHLTETQVKIWFQNRRNKLKRQFAADQEARQQALNRMVGSVH
VLYNEGRNVHSIGTPQPGGQGTSRPPHLPIIPPQIGATLPYFPLVFPPKDIPDFQPKLFDHPPTYQD*
>Locus_41714.0_Transcript_1/0_Confidence_6_Length_1153
MEFSLLDDSQMMKTMMTTIGKVPPTHSIEDILGLNKGCHKELHFATETDRFRYCCPEQPRHIPLYHCQRQPEAMA
DNKHRPQQPFPIPTFPLVSKNRLAGSYEYFSQLGVTVPSLNQSHINNIEGKKHTRPTFSGHQIFALEKTFEQTKYL
AGPERGRLAYTLGMSESQVKVWFQNRRTKWRKRAVQESKFKHRVGCGLTDDFAEETVIDDIHPITNNSIDIIEPP
KGSTSTKDICREFIASGIACQDSKTYNENRDPLRS*
>Locus_33755.0_Transcript_1/0_Confidence_1_Length_2054
MLPIEQNNNNDRSGKRFKFSIDSILESPKCATLSDDEQLHHSNNQSPTQHATDIPSQLRLQDIMQSHVVFPGGITP
SPFFIRPAAESLDPTVMATQLGFARVETYPGSGVGPPGFPLWLDTRPGYQGMPQSISLMDRYPLSCRRPGHPYT
NRAPAKRKKPRTSFTRVQVGELEKRFIVQKYLASAERNSLAKGLGMTDAQVKTWFQNRRTKWRRQTQEEREIE
RQATSRYLMGIEGTNNTSPGHTVVHSKSAHPNGTDSDDVSVGD*
>Locus_49354.0_Transcript_1/0_Confidence_14_Length_2077
MVVTQTLMSDANNRPSPNETFLLTARSGFPRHSEFAPVSNFSITDLLQCQQQLQHSPANTMPPAYNNRPYQSFN
HSPNALLHQYRQGLPLLPPQNLPPIHGNTFLPKVAQNRATYQPHQLQAFNARMAFPASTIDSVSMEPPNKYLRH
DPSFYQGDNTLLYQTPRLPNNTITSPTTKHYETPQPKETERFIDNLKPNKPVSCDGYQTSGGVTTYNPSLLLGSV
NMPVTTDSVQDYSSQLNNNVEQNKLDKDANIKTTKVRVNFPSETASTNSTDVTVDCDGSDVSDLVADDQISPTD

SAPGEIKEPRRIKITDSNGVIKEYVFPRGLDLDRPKRQRTSFTAEQLYILEKEFRRNQYMVGKDRGLLAESLDLTET
QVKVWFQNRRTKYKREKTRHEEYHQVNAESLAACNILKMLELPPPPVALIPPTGAPPAAASGISPLGGAISCSPP
RAADHYSKPIAAGAAANMFPR*
>Locus_4231.0_Transcript_1/0_Confidence_0_Length_1276
RRVHGIKLSDCLIFRLLLRFLLTFKSSNEEHKMHPIFPQQQGYLIPPWFNLDTFRTNTTAVPPYYNSFSYISLNSGS
LTNGYVDMMAQNSSNNGHASLTSSHNSYKKPALKFGIDRILAIEDSKDEVCSSRNSRFYNDSHEESTIDRSEPIKN
LELFLIRNPPECDYRMGPFEPAISPLLRDQHPESYLSSTSNNSQFNSRNDSDLLGLVSRQQQHPRDGVKHQSLL
GVDHTEKFQWLQCTRYSPPKIQRTKKYSERKRKLGRNPRVPFTQTQVDHLELKFREAKYLSSLDVADISTHLNLT
DTRVKIWFQNRRARERREKAAESAPPDNSEEIIDIVNIDDETSNHADIKPIDKHNSAKNIISDNFKLTKTVTNQKSTI
DLYKTGNSI*
>Locus_15302.1_Transcript_1/0_Confidence_1_Length_337
RKPRILFNHSQIFQLELRFKHQRYLTANDRRELSDQLNLTQNQIKIWFQNRRYKSKRQEINQSLASYYQHNKQLPT
SNISYWNNDQWFYLTPTPNLYPLMMITQTET*
>Locus_20062.0_Transcript_4/0_Confidence_2_Length_1268
MQHSLSQSTKPISRGLLADSTHPIVTASSFITTYPPLCCLLPPSMPPVCLNVCQCHGKPTHNPLNWTYGHSAQLQ
TFCSGTPPQPVPVASLTHGIYYNTVGDEYRITDNITSSSNQVSPRTQLSEDVSSQTPSNSHTTTTTPCRTIPEGCE
RTGQANVSKFHTYNNQQPLNLIPISIREQICLNKRNKLITNETTISLQPMNNCHGNVPPRVGNGSLSLPESSDSSA
DTPIIGYRRKRTSFSSSQLRELEKHFNRKKYLSVPDRTLLANILHLTQPQIKVWFQNRRAKWKRIKAGSYQAATSG
AMTVSTGHKLVVPIPIHVGPSRTANLDQQRPPLA*
>Locus_49786.0_Transcript_4/0_Confidence_14_Length_1543
NHNNNKQRKMFGNQWDQLPKHAKPIDVDFTMNSRTLNLARQGSTEVKVDDICDHFSPTSGCHNACNDVYSSPP
LLYSPYGDLVEVSSYKGTSLGSMIVRNGPSPTHSTQSCAANSYELSNDNFTHDHNSRNNVSLFQSPMYCRTQPS
HPRSSYSKGFAHTANLHGNTNCTQSISNLKFSIRSILGPNVDDNHETSTLVVGNSSLNDLSEATVSHIPTVTAKNV
ASLADEDLSDLHFPLTHGVNPQPGLGMWNSRALPCTYDFLSTGNIIPSPYFPPTGGAYYWPFASRGKPRRGMLR
RAVFSDAQRRGLERKFRELKYISKPDRKELANKLDLRDSQVKIWFQNRRMKWRNSKERELNSQAANRFNDAGIQ
DHSSNSQGTNDCIDYELLGNAMANGNNLPINYPTGPQLNY*
>Locus_29958.0_Transcript_1/0_Confidence_2_Length_908
MSHGKFTIDTILSDSKSKLTALNKVPPVPVTGLAVSPGTLKITCTPEVPWRRDELHLRMFDRQIFQSQLNFTDQFS
MVKPPPFVFPFNRLSCSPFPRHTWVPPGLCFPGDAPCLQSDICMCYFHSRRNRGWNRVRTVFTNAQRLGLEEQ
FREQKYVSKPDRKELAGKLGLSDSQVKVWFQNRRMKWRNSLKLKAAARNGRRAITVEKDSNNVGIEEESESDD
NI*
>Locus_31941.1_Transcript_1/0_Confidence_3_Length_2634
MESDDDEGGGSTNTPPLPIIFKDEPQSPPELTIDDTPLAPIINNNTYEFTDTVNQYQPPIDDIDNDHTTETPMESTV
TYTSQGQVRRSTRITARREIQAMTQLRARRDRLAKDEQLFARFKLQRLRSRRGRGGVQPAVLVPLPKSLSSKYNI
RDAETPLINTTSAISQSTPSNQITLPAAESPIQLMKKSSPTPPTSLSPIKPSNNQLFAITRNMDGQIVLLPLMPSNSG
NYTVQTSTSAPVGTSVNALTTAASVAVSSVYDGSITSTTSEGSMLGSMRPPPPLVLHRNHSASSSNSAMGKVTP
TKSRLLLPRPSAGATPNALSLIRANNYSRVKQRVMLRFQPTRNRQFLEMPQSSSWSYRNLTGLSTVTVGGTVNG
NDNSRDFISKRQDTMMVRVQQELADMKIPTGVFTKAALNRTRGILQHILKPEGTCQQYLSTLESFLEQPEEQRRN
LYNQSQNSIVEVRLAKRQALFAQGRSKPTRLRIRTPITGRHRASMFQHFLTKSQYLTPVDCSNMSCNLGIPEPAIR
VFFQNLRARVRARTKQQVNDIPTRQTENIILSDGTLT*
>Locus_36321.0_Transcript_1/0_Confidence_7_Length_1157
SDNVENRQQRIGGFITPPSSPDCETTPSKLEQFVNEISEHDDDQSYKRKLEDEIDEYDVDSDEPKRNGKKSRTSF
TPVQVLKLEHKFQTQKYLAAVERGEFAEKLGLTDTQVKTWFQNRRMKHKRVQQDIDNAAMANNYYNRLMISCA
APPSGLFHSGYNDPAPVMLPAAPLMTAYPPITSQQPMLPAMLPPTMDPYSINRIQQHLPGVPSMTSPQFSNSYA
PVQPSIAIGQGFLHPSQTQSPLANTEFWLNSNRIALDVGLPNRMPPLGVYH*
>Locus_45900.0_Transcript_1/0_Confidence_5_Length_1608
MDAYNISLSQWFWDPVFWLPPSPAIQMAWHYMATHSCIRTYYPIPIDLLKRGLPFAIVLLVIRLLLSRFVFRAYAKY
LRVSGSTSHQDHVENQGYQPILESKFAINNYISGSEVEELSNRLGITDIQIRDWFANRRRRGQFSQLEKFNNNLW
LFLCRSVLSFYGILALAGREFSKTSRLKTVWIGYPRHPISEAIYEFYMLEIAFYASQIITSLRHANKRGVKLEILYDSIIL
SMSIAAWVCNYHRMMCLILLAIDVGQPGRHLAKVFHYAHFDVPATFTYIVQLTIYAIFTVYAFSFHIFVTATISFTEIF
HVINCPLSYYGYRASERGGEVAVPSNHEYNSTDLVALPSNHAYYYALVSLIFLVCYHVVTLATNTWMLLRILIHAKI
HGLQAEEKKVN*
>Locus_50700.0_Transcript_1/0_Confidence_27_Length_1758
MVFNFSVEFLAESCARPASGTTQGRVYGLITPPISPDSGNGSPERKLNSQSTDNFNFCFHENMEFNASKLEDIGD
YDDESDVTNKKGCNKKSRTSFTACQVVKLEQKFQSQKYLAAVERGEFAEKLGLSDTQVKTWFQNRRMKHKRV
QQDIDNAAIATNYYNRLVSGCPPPPVYHPVAYPEVIPAHMVPPQSLHYGGFPGLAPSSEAFPSLGAPLTPHLLPP
ATQLPPMGPTQLPPMGPPVQSPTRLGTELWLNANRLPWDMPYPALTAATGVYM*
>Locus_20893.0_Transcript_2/0_Confidence_2_Length_1499
MICYAFMSKRDTFYNKSGSTSRGNGTPHRVQSPNRNADSHTRQIKYINKRSTQFQDQHNSSSYNQKNQHLYVQ
PALFQALKTTSHSTQSKQPQKGSQERHTSKNPLHCNIYSQVQQPKDQQSGTHTLQSHYPYNRAQQSNSQQTTS
HSYQRQPLCERPTRSPDALQLATDSCGHLKSDPICCKESSHNFREYPPTFHHQSSTDFYPRSPLLPGLPGRIVLR
SPRPYIMGHCSAQLSHPSFPNIPSYGFPSGISHPSYSSPPPRVVYSSPPHSGSSYGFDPASSADGSDLQPRSAA
ASLVSQTSSIVDQLDKISDSRPGNQPETVLVKDVSDSRKRSLSVSETDRKEPKRILVTDSSGAVKEYVFPAALDLD
RPKRQRTCFSAEQLYVLEQEFIKNQYMVGKDRTLLAQSLTLSETQV*

>Locus_55290.0_Transcript_1/0_Confidence_1_Length_1438
MRIDDIMNSTTSINNNQQFLNPNLYYSNLSSSTSSQDAVITTRASTFNYYPRFAGSLPNSGYPTDIANMSTVSSHP
TAPPSFVHSYLPHAAETFVPSSYLSASSNHLVGISPFPQVASFTNIHSTSGCGTNLQNSSPADVCDGYSPRSTSD
NIDIKQSPVSNGSPTINKNAPNPHTTDRDHHSISHTDRDHHARESMITQLNNNINDYNKKELEDRKIDSDDDLTDD
VITKVAKKRRNRTTFTSYQLDELERIFQKTHYPDVYCREQLAARCNLTEARVQVWFQNRRAKWRKKERFQQINH
ASRLSPYDIPWMGGHTATAAGSQWSYQLPPTAAGLHHTIAQYYPSYLTGTGAAQALHTGAPTLHSGTAALHHSL
STGAAVSDAYYPTDFRSSSIAALRLKAREHQVAMDLINSTATPTANSHNPQY*
>Locus_40829.0_Transcript_1/0_Confidence_6_Length_1163
RVHGDVITKVAKKRRNRTTFTSYQLDELERIFQKTHYPDVYCREQLAVRCTLTEARVQVWFQNRRAKWRKKERF
HQITQNRLSPFDGTWGANHNAAVASQWGYHTTPLPPTALPGISQYYAPYIGGGGGAALHPLGVPAGTPCAPGTT
DSPCYPTDFRTSSIAALRLKAREHQVAMGLFNPHGNAMAPTHTSHYAIS*
>Locus_40214.0_Transcript_2/0_Confidence_4_Length_1637
MSRSNMGVSGFSIDSLLGSSSSSPVSTSLSSSTPTYPSLLTIPPSLDPQIGIGFNAAHTQSSILDLHNATSPLLPRS
NQFLLPPHPSTIPPSQSLFPCYYSAHPSLYAFPSSYTEFLMHNALLRNPKRKRRHRTIFSEEQLDELETTFAKTHY
PDVLLREQLALRVELKEERVEVWFKNRRAKWRKYKRDTETKKESSSGETKNELKADISTSSKLRECSTTANSTEV
VGELVSTQTGPGSRYEVPVGEPASRYGIPVAEGDNSDSYLSSDNGGEHHLDETEMVDVEI*
>Locus_53544.0_Transcript_15/0_Confidence_41_Length_3588
MEPANITMEDKSSHYKIEDEGGVSDSDDKANKPKRHRTRFTPAQLNELERCFTKTHYPDIFMREELAVRIGLTES
RVQVWFQNRRAKWKKRKKTTNVFRTPGALLPGHGLSGFSGLSSDSLYPFTPSFDTRWHPSFSSPFPKHTDTFS
RFSPGTYQPPAHMMHPSHPATPSSLGFSHHPVSDACVFPPPPPLPIQSLAVDHDQPPTLTGMYSSASNLQPQTS
SAFRNSPPLSAMGCLEASSAYSTCVNPLLTRGDNDDNWPGSSIASLRRRALEHAATFSGNCAHPLVPTANGITLT
DIRDESWLSRRHNEYQR*
>Locus_48073.0_Transcript_2/0_Confidence_17_Length_2165
MAWNYDSVAYLKGAYGLSGSMDLLHHPAMAGYASQYPQLSSGFTSGPPPTVNGQPPRKQRRERTTFTRAQLE
VLESLFTKTRYPDIFMREEVAMKINLPESRVQVWFKNRRAKCRQQQQTGSQSSGGTNNTTKKSPSSGATTVA
KKRPTTPQTSGLKLLQTPPGTHHSGMNNLSEGNLLLCKTEPEMEIPPAPCSPLYKSGSHGVLQSCHSNYSAGAP
SSSLWSPASFPTMAEYFAGSGCMQRGSSGNGPVYPTQQTTPGAGQFYPQTSGYCGSSGTFEYLPGNMAPSG
ASHMSQSHVAYVGYPGRYADSSQCSIEYGPKEQSSPWKFQTL*
>Locus_49939.0_Transcript_2/0_Confidence_19_Length_2983
MYDKLRLFNSQAWTHPGAWYPGPGSCMAPAAPHYPSMLQHPSLVQMPRPDLLKEENDNGASSSGEEGGDSE
EARMRLKRKLQRNRTSFTPEQIDALEKEFERTHYPDVFAREKLAEKIGLPEARIQVWFSNRRAKWRREEKLRSQ
RRSQNDSGVTMNSAGLAAPSHTAAFNPGNLYSTAAPSGFQGYGDYSGFSHLPAYSMGMAQNSAANAYSCMGS
SASRYAPDPLSSYRPPAAPAQFQSHGSHLHPGSANGSSTGLLSPGMSVPLQVPGGGMDMTHHSMATQYWAS
RIQ*
>Locus_24274.0_Transcript_1/0_Confidence_4_Length_1313
MMPLKKSLKSRLQRDRGNNGVNQLGGVYVNGRPLHESTRRRIVELAHNGARPCDISRLLQVSNGCVSKILGRYY
ETGDIKPKIIGGSKPRVATNKVVQKINQYKQDSPSIFAWEIREKLIEDGLCNADTVPSVSSINRVLRSISIDDMAECS
RNSTLYHHGNTSDNLFYDKLQFLQPSSWLLPGVPTLEPNNKSPVDNLINSFIPESLSFSSAFTTDKGKCEFSPSDK
ETAATTDLNDSDYSRNSDDNDEAGTDRLKLKKKLQRNRTSFSVEQIDALESTFNESHYPDSNERERLAEKVGIPE
AKIQVWFSNRRARWRREDRTSLSLSLNKYQHDILANTIQQQHNTSQLRGGTSKNELQQ
>Locus_27707.0_Transcript_1/0_Confidence_3_Length_1324
MASSGLQLEFSPECNALPLNNVDHLGLHNSNGFSFRSVGNQNCCVSSGTRNNSTTGATGSKHSDGGAASKTPD
CFRQNPNRPSTDVKKESGPTAATAAGDDKQQSETVDKKKRPRRQRTHFTSQQLQELESMFTRNRYPDMATRE
EIAAWTGLTESRVRVWFKNRRAKWRKKERNQLCTDYKGPFGPSQFASYFNPYDDSLYAGAYTTYNSWAAKVAS
SPLAGKSTFPWTPSPQQFYLSSPNSTLGGAASCPYGSGSSQYSVYRDCSPATGSLRFKPSPPSGPVPQMPTKS
GGYFGYNSLRQSPPSVGPNATAFSACQYTSTMQGGGLSSMKIDGF*
>Locus_29086.0_Transcript_1/0_Confidence_4_Length_1565
MLANAYCSSSSNSLSPKKGGHSINDILGIGSRFHDDGSSEETQSLTKHGDLLTSPNGDDIVHDDSDYGLDDEDEM
MKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYAREEIALKINLPEVRVQVWFQNRRAKWRRQEKMETSIYKLP
DFSHFRGHGGHLDGFFPSPKCSHCSTSATRLRCHGSRCCSQSLWNIWPLLYGSFFSRSQWSI*
>Locus_21928.0_Transcript_1/0_Confidence_0_Length_1379
HYISNLCCSISVIYKLLEMNTQDMLLVGHPSAHHVRSVTTMHPSPSSKSTRFAIQELLGLNSRSSSPISRIQPTNIP
MITHHPFNYFGGSLPQSSPTAHTLAFPCNLPSPTYFRPAFPPLMCSSASSLFAAAATKQQLLDSSNVHLNIDNSCT
GSSTGSYGLSPDIYSGSNKRKKKRRHRTIFTALQLEELEAAFGDAHYPDVYARELLSIKTDLPEDRIQVWFQNRRA
KWRKKEKTWGKSTVMAEYGLYGAMVRHSLPLPDAILQSACETDVKESCAPWLLGMHKKSIEAQHKLDDVCIESE
LSNDDVIEDPKIRYLNNIGVSGNIVGDNNDRQPATITYIHRHSTETEIPNISKNNNEFKLIEIDDQWGGSSTIADKPTT
DSSAVI*
>Locus_54213.0_Transcript_49/0_Confidence_50_Length_4112
MYLSPINQYQSLSPGFVTPIEMSDGHISQTPTGRFNETMATAALSTPTNQHRATPTRQIQPDFPPQNDMFNSLFT
MNNVIETPVSKHYSPQFAYNRRTHNISYGMARRSLCDENYMPPSKIQRIVNFTGDLSNVQFSNSDGVSSEPIVTA
VEDNLQKCPLLIYQQHGAPAKNVRFMGSNHAIVNHSAGDCGNNISGWNLSRKRKQSICLGPEQMGATELTNLEK
LPQRDIKRKLFPKNVPILAHGETVPASIECAVEATRISYDFIDIPVQKKREKKVFDPMQIALLEHNYMEIRYPDLQRR
KELAAVVGIDEKRIQVWFQNRRARTRRKHPDVLEMDIVSHRVRNVEHFKPLLTKHSAPDGARGWSKKKVKGKVN
ATRESQRSDERCKLLEAPKTSNVGTNTTCYSHSTATSATVVVSPDAEKSSISIVLSEDGASVIGSQYGRSGEFAA

SDDSVDCHSRTVSFGDTQIAVAQSPAPTNIARHETVGSYDFMFTPGPCEGNTNLSTSYQANTSTKNNFASNNSS
ANIDIVNPNIGNSNLNIAAFIPSSANNTIEYAHIDGNTVECAPYFVPLTINIPTPDSVPTEQPHPVAIPPGALQTPLNS
WDSFDSVIVPVAVGSDPGVVEVPDSEASLMSDVNDVLENKSSSPYVYENPSCNPVVFSDTYVTSQYTYDTNNQL
LLNGVQNFSTMSDPVNQGFQLQQGCQETEISADSNGIKYCYYNNGYPGNGAYSSYSGMNYNNGYMSDTLQES
NMNTFVANAPPSYTACTMLAFLARIASQHQAMETLTNNYVYNVT*
>Locus_18856.0_Transcript_1/0_Confidence_1_Length_716
MALNYSIDYLLGNQEDVQFPCYKTSSNNNNDSSCDDGDCIEICHIRNKKPRHRTTFTSWQLQEMERAFERAPYP
DSPMREQLAINIGLTDARVQVWFQNRRAKWRKQTPLVQRSRDELNKLYNEALIKYCKLTPVVIGEQTAVAASLAN
DGKEPVLTLEHDSSDYGITIESKNDE*
>Locus_47651.0_Transcript_2/0_Confidence_7_Length_269
INDTTIEHDSNCISRGVATAFTCDESYSSPKKIMRIDNNQQQAPLREAEERSKSNKPLLPKGIPPPLHCNAISKSISK
DNNTRERKSTTNLNDDDVNGGMTSRSSDVAGLSTAISPTTSESGGLGALDTSAASEVLQSRHYTFNSMNKVNYV
KKRVKKVFEPGQIALLEHNYILNQYPSPQGRKALSNEVGIDEKRIQVWFQNRRARTRKKHPNNSELAVVSPRVRH
AEHLRRYLANHDKAIVRTTRTRCTRPIKRMKSIQKRSESDTRNISQQHEFYTEPPQDLSCGMDNTYPTMSPLPPN
QEYMRPTDVPFNQPQCLPSYGYYHPPPFNPYPMYNSQFPDGQMHNGAVNESEFSHVFAQRDATNATA*
>Locus_54455.0_Transcript_1/0_Confidence_0_Length_772
NAEYMGGPPFPNPFQMPTPNPDMHISSPTKNQKGRPRKKKTNRMLDPNIHSLYSDPSAYMTHDGMMDCGGGL
YLPGGTGSRSSCQDVHSRTKRVRTSFKNPQLRVLKQYFQLNHNPDAKELKQLSQKTSLSKRVLQVWFQNARAK
YRRQVVKGSKGSLDGEGELKTLEDCHPDDLSDGDMEFQEPDKEPGSTGAENPDMDNSGQGSLSPSMTLSPSS
ESVAEFADMVSPYMLSPNGPTLEQTTDGIPTQNSLY*
>Locus_51594.0_Transcript_4/0_Confidence_29_Length_2823
MTALDQNHNPEVMDPNNPYHQGNLSPGGHGNSRTKRVRTSFKHHQLRTMKSYFNLNHNPDAKDLKQLSQKTC
LSKRVLQVWFQNARAKYRRLMLKQQNGHSGDGTEDGGETQSDADNVGTSVDSPNGVDDDMDEDDDIMSTLSS
CQSMNSSGGGDNPSPGTSDMSNSQLSEMGGPHMINSAPTLAELHAMQQHQPAISPTLILPNLTVTIPHSMIPPTI
PGLPIHSSPIGTYHHSVASLISMEHNMRRASTVQGFPY*
>Locus_51594.0_Transcript_2/0_Confidence_12_Length_1330
MNDIMSSGIPIPFNEHWRASPITFANGRKIKQEIKKECKSLLNCNDSSYLSPPAQLILRSSNPNIKVNPDDKPSQCA
GCLGQILDRYYLHAVGKQWHLHCLKCSQCKIPLDSELTCYTKDDNIYCKEDYFKRFSAKRCARCNTGIGGRELIM
RAKEYVYHLHCFSCATCHKALHTGDEFGMRDAVIFCSEHINEGVVYCTDQSHTQWSGDLKQEVCAQVNCSRTN
VRTENYIAADGVKHSAIILSGCNRTRHSLSRKKDKQSNKMIGQYLNKRRHRTMFHRYQLEQLKESFKTNSNPSSI
DVEYIAKSVNSTKRIIQVWFQNARSKQQRRLKGRTVANGYNSPTMIQFACEDNSMNRENGNNSPTMIQFACEEN
SMNRENCNN
>Locus_45303.0_Transcript_4/0_Confidence_14_Length_1909
MKTRGGSKIYHVDCFRCSTCSVQLITGDKFVVRDDGIYCEEDNQVIERAELVEFTNNNNNNNTKTSAMQSDQSV
TLQSRSHGSTKSSPSSDHKPTRVRTVLNEKQLSTLRTCYAANARPDALMREQLSDMTGLSGRVIRVWFQNKRC
KDKKRSMHIKSQMELQGKQHSMTGVAMVASSPVRQDANLHGNQVEVQSFQPLWSHGGPVDDNSYMSISNSM
SKGDLHSDYMVYDTQLSSSPLSDISSNGDSIASPLLLYGH*
>Locus_22501.1_Transcript_1/0_Confidence_1_Length_2011
MMIRDDNVIMPIYNNNNELMVNKTTINKLRTPLCVGCGSLILDPFVLRVAPDLEWHSSCLKCAECHCFLDERATCF
VKDGKIYCKRDYCRLFGSKCAKCGMGFDKTADLVMRTSNRIYHVDCFRCTSCNVPLVTGDKFVIRDDGIYCEED
NLSLQERCDDFTNNNNDNNNINNNKNNKESVQSSSNNKESSGTTSRSSSKDHKPTRVRTVLNEKQLQTLRQCY
AANSRPDALMREQLSDMTSLSGRVIRVWFQNKRCKDKKRSIHLKNQIEMQNKQHSMNGVAMVASSPVRQDSSL
HGNQVDVQSFQPMWAHHGAVDETYLSAISSNPLKSDLHSDYMVTTYDDNQLSTSPMSDISSNGESMPSPLQNIL
YGH*
>Locus_61503.0_Transcript_1/0_Confidence_0_Length_420
INNNFDTKNQNECSTEEIDDFDDDIKDDDDDDESIIDDKKSTKDGDSKANGSKRRGPRTTIKAKQLDTLKAAFDAT
PKPTRHIREQLAQDTGLNMRVIQVWFQNRRSKERRMTHMRELGVHRSAFLRHNRRLIALGAPS
>Locus_38224.0_Transcript_1/0_Confidence_1_Length_1306
FVRSGKLFCRDDFFGRFGTRCAGCDKGIPSTEIVRRAQDKTYHLDCFVCVLCRRELATGDEFYLMEDNKLMCKA
DYELARQRAADGSDASDLSNSQKRPRTTITAKQLETLKTAYSNSPKPARHIREQLSHDTGLDMRVVQVWFQNRR
AKEKRLKKDAGRGRWGANSFTHHGISRPSKRSRIMNRSDILDNMDSPPSSPLSCFSGEFDSSMENNNNGNKLQ
HNMSPSECSTPPSADLSVSTIGHHNVSPSTTNAVPVDAIFPPPISTSATEWLAPPPIVSPWDFPLYFLHNY*
>Locus_50222.0_Transcript_6/0_Confidence_13_Length_1916
MNRCDALQYERETLSIQSEISMKREPPSHAQCHGCKTDILDRYLLSVSGGGESFWHEGCLMCCECGIPLNNVCY
VKQREYYCKHDYLRLFGTKCHVCCQLVGPRELVMRVKELVFHLSCFNCAWCCKHLEKGDEFIINEGQLMCKTDL
DKMSFNVDGLSEDDSGCSDELFENSSPSGSDKSTAFRSSNKLKKHYLVNRSKRVILTTTQRRIFRTAFDVSQKPS
RKVKERLADETGLPVRIVQVWFQNERAKGKKLAKRQLHKMAIKSSRERSEDVSSGSPVYQLDAPMVGPNRPFPL
VNIGYPNFESIHPHFVFGSDPSRRRFPNQINPEITRNATGALGDLVHSYFRGCF*
>Locus_14051.0_Transcript_1/0_Confidence_0_Length_1064
QSTWGIEFIVKRYHADINAMTESSGVSNRVCRGCGYQIIDKFIINIGDSVWHEGCLTCCVCYEELSGSCYTRGDSL
YCKSDYRSVFGARCRYCSLLLRPDDLVLRADGNIFHEDCFRCAICDTVLRKGEVVAIGAQGRLYCKHDADNSIESI
ESRSNYSEVTSDEEETMSRRKTTKTKGVGRKRVILTHNQRRVLRAAFIVSQKPSRKVKEDLGGATQLPVRIVQV
WFQNERAKVKKLALRERSLNGSDNLCCYGNAVNNDTNIWQTSYWEGSSDGSDSGYNNSLHSYTDEHLCLVHV
NN*

>Locus_49053.0_Transcript_1/0_Confidence_12_Length_2438
MTVTGMDLIPGQTDTANNNCSQLPHGGVLGGGGAGDFKREVCYGCQTYIDDKYILKVNDKSWHEQCMICAECR
TPLSDSCYVKDRSVYCKDDYERLYGARCHVCLDAIPPKELVMKATGYVYHLKCFACVLCGERLKQGDEFILKDNR
LLCKTDFEREEAEEKLRETHQQVLMAGAPSSMSVGEGLSMAPAVASAPPPHSLPLNPSLMVRPAASLGSSLIML
PDPSGGAQGIPNLPISHLRPTLPPSQLPQQPVGLPPIDPLSSMSMKPSSLGGIMPGGAPPGGLPPGTLIPGSLPP
GSLPTSGGIGINNNNKKAMSLDKDGSRKRVVLSPEQRKIFQLAFDASAKPSRKVREDLANKCSVSMRFIQVWFQ
NQRAKMKKLAAKGTDINFPLSAFDDMVDGKLNDSGEHDSLMDTGGLNDDDDDLAMDVESLEDGGGGGDHHKG
FHLRDSPLPTPTGGVPPTASDMVLYPPSNPLLSSHLSPSNGPEPLTTSIPSIHMNNQSSICSPMLNPSLNV*
>Locus_7973.0_Transcript_1/0_Confidence_0_Length_1041
MMDDQNEIPELTVCLGCNQPIHDQYILQIDSTHWHEECLTCRICLTPMTGDRCYQRNGHIYCPTDYHTLFTAHCR
RCEAALKPKDLVMKCGNGDIYHVGCFRCSVCNILFIKGDRVIIIHDGSLYCEEDATFQSLIDGADGVSLDGSSKIHH
SLLPKKPRRSRRIKRTGRKRVLLTNPQRRIFRSAFTISDKPCRRVKEGLAMMTQLPIRIVQVWFQNERAKMRKLS
MKEECPSSMILNSIDDDEDYDLEAEQEDTSDSMINTDDFEYEEYFKNEKLTTLDYNCNVETLYYL*
>Locus_48370.0_Transcript_1/0_Confidence_2_Length_1736
MNETSVISEVKSEGSHMSQNVANGAATAMCHGCQLDIGDSHLLHVSDRRNADNCGTLWHEQCLRCSECDMSL
DEVCYIKNSDFYCRADYQRLFGAICHGCRQPMATDEYVMRVKIWVYHLQCFNCLWCGRRLQKGDKFVQQENAII
CKSHLGKMSSIQIIEGLSEDDSSASDLYEISSPSCETSTKVARGDGMVPTSKLRRALQSAHNNKRVNLNNMQRHI
LLAAYETSHKPTRSMKETLAEQTGLPMRIVQVWFQNERAKCKKMAKRQQQINKSQRCSMKSDRYSSSYNVPNH
SSNMFLYFLPKCDETMNGNRSDNPYSIDDSYTRCNNINEESQIATMKNSYPAASNIDLNQNHGYFQ*
>Locus_52392.0_Transcript_2/0_Confidence_13_Length_3474
MKATGYVYHLKCFACVMCGDRLKQGDEFILKDNRLLCKTDFEREELEEKLRESQQRGLLSGIPPSISAAAISAGFQ
LQTTMPISQHSMNPSSLMVRSAGLNSLPPGAPGSLVMLPDGTTAQIGCPPQGVACGPSAPLLGAGSLMPPGTPT
MVGASQLGLAGGGMMLQQQTDSLNSSLQMGMKVGPPSAGGQTPNNIVIANIKKNSNLDKDGSRKRVVLTPEQR
KIFQQAFDMSPKPSRKVREQLASECTVSMRFIQVWFQNQRAKMKKLASKGGDLSAYGGLYDDMGMDGKLNDS
SEHDLLMENAGLMDDDDDLNMDAESLEDGGNGSNELSGYGLSDSPLPSFVPQQFPPSQLQTTTGTSPTTAGG
MPSLFPVNSTSHHLIEPSSTLPSLHNINNQTQQQPINSPLINSSVN*
>Locus_27996.0_Transcript_1/0_Confidence_2_Length_2302
MVSLINTPAFSDTGPVPSALQYASLYASPAVDYPATSSLSGAYGNQQSYPPPTSTQTLLGQSAYKLDYLGTASQL
SSYYPGNQSASWLTSPSTGDYWHHDIKSYGELKHSANSEPIKQLVQELKSPSIPDLSKSYSTTHDLKAFNMPPPV
RDEVPYKPYYLSSPYSMWNSGSSSALNSSTHFMPYSSSNPNHLSSYSSSAYGSYSTYPMNSAYNMSTLVSPGL
GYGPRGDPDYPSHLSDEEAPSSDDLEAFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFESLQLSFKN
MCKLKPLLAKWLEEAEVNNTSPTAIDKLASQGRKRKKRTSIEMSVKGALENHFMKQAKPCAPEIQALADSLQLDK
EVVRVWFCNRRQKEKRMTPLYGCEDDGSLDGEGEEKVSPRTPVSVKDEFVTSVSNYVSTFASNVNSSGDFPV
GDRMYSNSESSSDVMKGDSHNAMYDSYHQNGLQSSLSSLYSGRPDIDPKSYDIDHQPLSHISTAVITPNPNLLYP
GLNYPIISYPSLPSNQYISSPTMSAKQPQNSAPMSSANSDINALTYNSSPM*
>Locus_9311.1_Transcript_1/0_Confidence_0_Length_709
PASLHHTVVGSLPSIQHCPDLDTDPRELEKFSEHFKQRRIKLGVTQSDVGRALANLKIPGVGSLSQSTICRFESLT
LSHNNMVALKPILHTWLEEAERVSMETSLITAERKRKRTSIASPEKRSLEAYFAVQPRPSSEKIGQIADKLDLKKNV
VRVWFCNQRQKQKRLKFSAMR*
>Locus_52441.0_Transcript_6/0_Confidence_5_Length_1244
IISRIYSDNLPVLSTDSLLDIDDISGNPNTIQNTNSMATVDGINLDEIKNFARLFKLKRLQLGLSQAQIGLALTKQEGP
LYSQSAICRFEKLDLTPRSAVKFKPVLERWLLEYDKAIRAIPQHEHVDNICDTSSLELQRKRKRRTSFPPAAIDILNS
YFLKNSHPSGHEFTQLSRLLEHDREVIRVWFCNKRQAMRHYALNRSTNNNSYNNNNNFL*
>Locus_53834.0_Transcript_19/0_Confidence_35_Length_2674
MYVPLSLPKPSASIPNNTNKLITFVDNINPIQISNLGNDEPLRGVDVGVDDIVFTPEGNDTQQPSHTQQNESSRKC
DVVSPKRQPRPLCLTINSITNHPKSSKTIESYSPSINPQSPRTITHPPTDTQQSQQHVSFCSTNSPQLLQDSNILTS
EFNLPQFAPTHIPQPRARYLLSPDVLQSCSKPPDIRTLPSTRRPNDTLGTDTALQLEANRLEHLERFSKEFKETRIR
LGFTQSDVGSAVGRLWGNDFSQTTISRFEALNLSFNNMLKLKPLLEKWLSEEAGCKMEESCEEGVRIEFETPAFL
SESQETLRRRKRRTSLNPDVREKLEAHFEVNSRPGANQIFVISEHLSLDREVVRVWFCNRRQKIKRVNSSTAATR
TATITTPTTSGTASTTLTYTPIVLHSILPPTQPLFH*
>Locus_48953.0_Transcript_27/0_Confidence_27_Length_1864
MQASKYTIEQIELVRRLRYTGINKSQVLHIWDSLELYENEPNSLLLGHPSPRSSGPVVPTMGVFSTPPSVDIQSKL
EPQSSKFIADSTHMQGPPGQPSMSPESIVSDTHWSDESATSGVTNNENRSHNPPGGYTSNYNSRTAGDSMLLC
QPRVHDAQWFNKEDIGDDDIELRELLCKPVLEVIEEMKTFMINNKITQNSVADQASVSQGYISQFFCRGIALRATT
QMLLYKWYISEKRNVVNHGVSRFSADTARSKPRCAPRRRERFVWREECMPVLDKFFNENQYPDDRKREDIASA
CNGVLEAPGVHIPGERLVTPAKVYNWFVNRRNSYKIKMRQMACGSNWDESLECQSPLNERKFTPDEMMDLVR
QQQSIRNMDLDYPKEMDNESDNQPSSYLGNASGYQRNEHTPTTAEYREVLPVLDYDVSQTNGTQCAVDG*
>Locus_36924.0_Transcript_1/0_Confidence_0_Length_774
PTLPSLNFTPSQVASVCQTLEDSGEIDRLARFLWSLPINPSALDSFNQHESVLRCRAIVAFHQGNFRELYHLLENY
RFTKDSHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKERTRNLLREWYLQD
PYPNPQKKRELAAATGLTPTQVGNWFKNRRQRDRAAAAKNSVNGSISTTMDDDDDDDELMLD*
>Locus_42137.0_Transcript_1/0_Confidence_3_Length_1922
MLLFPMFPLPGLGGAPGTPFRLPPPGMGGPPPNMLPPPPPAGLPLPPPPPLPCLNFSPDQVASVCQTLEDSGEI
ERLARFLWSLPINPGAHESFNQHESILRSRAIVAFHQGSFRELYHLLENYRFTKDSHGKLQAMWLEAHYQEAEKL

RGRPLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKERTRNLLREWYLQDPYPNPQKKRELAGATGLTPTQVGNW
FKNRRQRDRAAAAKNSVNGSISAALDDDDDDLMLE*
>Locus_28871.0_Transcript_1/0_Confidence_6_Length_1464
MFDQLSTTDQKVRDHSMVFPQKMITPGYPLINPISQRGMIMGLNNNIIGGDPLWSSMQSRYDLHKPPPRGSSGV
RLPFNFTAMDRLACLYRNGLHPQLHGNHHPMDYSCTIDASLGLRRKNATRETTSPLKAWLYEHRKNPYPTKAEKI
MLAIITKMTLTQVSTWFANARRRLKKENKMTWIPRNRYGDDIMRDSHTTTNNNLDSSNYQRSLNNVRNDSDDFD
NIDADCAESVMSDSDDEEILRRYPQISTNINNMVGSRYGERECVEDSSESKETTSAEWSSTNNNNDSSTSAEKP
KIWSIVELATSTKSNNDHVLKTDNLEQQYRHWLNEVYQNSLMSPSGTGIDLSNSKHSQQISSQEMNYNFNRNGA
LPLITPLNINTIGRIPNI*
>Locus_10005.0_Transcript_1/0_Confidence_0_Length_1138
GCFDTLLPSLPNPPISLPMTLPPPPPLFGIQQSRLLHPDYRLYGDSRTPTSQLLYPGRFGMDDWISGISVGCRRK
NATRETTSALKTWLADHRKNPYPTKAEKIMLAVITKMTLTQVSTWFANARRRLKKDNKVEWGMDCSPSDLQMQ
GSEDELVHISSSFSGNGRTEMETEPVSSCNSAAVSCCNNAAITCCNKSGDILLDKTTESSTDENTANSTANPNKQ
TAKSTKIWSIVDLAVSDKQMDADRSDEEASVIKLELPQTNPAVNKLDSTIANRDIRNCLF*
>Locus_26582.1_Transcript_1/0_Confidence_0_Length_1190
NGTNASILNGASKSSNTPALTVNCQQTLEAMGLPLQFCQLLCMSDSNNDPRYSINIQDLVENWRQFQQAAQANN
LDPYYLAFYHSALMDSGMGSGMRRKNSSRETTQYLKSWLKHHQKNPYPTKGEKIMLAILTKMTLTQVSTWFANS
RRRIKKDCRVSIMNNCGAIGGGGSCSTGDQSALSNNNRSPSEEETEGCEGDANEYFQSPSTLSPPISGQTKVRP
PSTTVDTFWLAGMNASSANLLVNSHSDRIAPSCDMPGATNNCSKNVEYFKMKSHEFDEVDINCFRNNRSVSSAA
KTISSSIKQLQPSHICNSSAVNNSIVPESSHIPAIAVVPTSSTMLVKEKFINSASSFEQQFLSSCCSDTTVSGHLPQS
PPVVRNFRSIDDLIH*
>Locus_36182.0_Transcript_1/0_Confidence_4_Length_1378
MMDFQQPHSSSYTTAALLAPSTTTNFSESDHRLNERTPPNCNLNSPPISRDGDCVNTAISNPLIPNRFPVINLHSR
NSPFNGIEFHHPYLPFGDAGSGFTPLGMNGIPNFPPFVDPLKSSMIGPDLWHKLPHHPLYHPYNGPPIGPYPYPY
GALDLANGIRRKNATRETTSTLKAWLYEHRKNPYPTKGEKIMLAVITKMTLTQVSTWFANARRRLKKENKMTWTP
RNRSGIIDCNDDNDNLNVMEVIDNMDDTDIDINIDDCDQSDDEYDHTDNDHNGSPRMNRDTNHTTDKPKIWSLA
DTAAVNDTHKDGLINNEKEAIITNINELTNHRRLINEPLNSHINLSLWHRMVTMQAQHHQRGVPLSIQPTQSGINH
MIGFNQHTPTFHNTLPYSTSTSKSTTLNSVSESTITMATDI*
>Locus_48316.0_Transcript_1/0_Confidence_8_Length_1528
MTWNGTPTNQFTYTTSDPLTNPIVSFNPPMTRGLCCEHGRLFTNDMNGAPSSGSCQCLSQSDKSRPRLMMSP
MLLPTTHKPLPGYETDQTHVDHDMSITNNGITEHSAFYTPLVQSSYEALKDPVPSSETWRNHQQMGPSQMYNY
ENTLQGCMPSSMSGGQYPYTSPAQYHYNSSPPYSYTSPMEFTPGIRRKNATRETTSTLKAWLYEHRKNPYPTK
GEKIMLAIITKMTLTQVSTWFANARRRLKKDNKMTWTPRNRCGQIQETLSEPDDVVDKYGDINDRTMRRTVTPAA
NTKSSNDTSEHNEVSSSCRNTDNVNVTDEDCDSDDSMKDAMEDKPKPKIWSLAQTAADRACNSTGAGNPSYTT
LMNKLPFSNSNRLPEFQAANATPSNGSNIISQMEIRSGLSASGYYQPLSYHHPLYNEQFHCEVPPSTTVCHTMSF
TY*
>Locus_52332.1_Transcript_9/0_Confidence_46_Length_2474
MLDSGIMFDNSVPDVAINGGQFSKADMREALQQIMSVTNVSLDEAQHWKAALNAHRMRPALYDVLCEIKQKHGT
FLHMRASPEDNSSDPQLDRLNNMLRAEGVDGPEINPGTGGPGQPDNAIEHTDYRDKLQQIRMIYRQELDKYETA
CNEFTTHVMNLLREQSRTRPISSKEIDRMVEIIHKKFNTIQIQLKQSTCEAVMILRSRFLDARRKRRNFNKTATEVL
NEYFYSHLSNPYPSEEVKEDLAKKCNLSISQVSNWFGNKRIRYKKNIGKAQEEANLYAAKAAAQGIHEDDSPSRS
DSPISNHGSLRESPSDAFIAAHSQPQFSTVVPNGYSENQADNLRQMIQTPNYHDDSQSFDLQYAPAGVATGVSQ
TPSSMATGSTVPTDVHQMAHQQAAAPTSIFESGRRMELWSSQCPTTINSQSFNAYP*
>Locus_52332.1_Transcript_11/0_Confidence_46_Length_2489
MLDSGIMFDNSVPDVAINGGQFSKADMREALQQIMSVTNVSLDEAQHWKAALNAHRMRPALYDVLCEIKQKHGT
FLHMRASPEDNSSDPQLDRLNNMLRAEGVDGPEINPGTGGPGQPDNAIEHTDYRDKLQQIRMIYRQELDKYETA
CNEFTTHVMNLLREQSRTRPISSKEIDRMVEIIHKKFNTIQIQLKQSTCEAVMILRSRFLDARRKRRNFNKTATEVL
NEYFYSHLSNPYPSEEVKEDLAKKCNLSISQVSNWFGNKRIRYKKNIGKAQEEANLYAAKAAAQGIHEDDSPSRS
DSPISNHGSLRGYFVDNPYFQESPSDAFIAAHSQPQFSTVVPNGYSENQADNLRQMIQTPNYHDDSQSFDLQYA
PAGVATGVSQTPSSMATGSTVPTDVHQMAHQQAAAPTSIFESGRRMELWSSQCPTTINSQSFNAYP*
>Locus_52537.0_Transcript_2/0_Confidence_21_Length_3983
MIPSGYMNQGDIQGGGAATVTAPYDTPPSGEAMLAPRKRELSDILHQIMSITDQSLDEAQAKKAALSSHRMRPAL
YDVLCEIKQKHGTFLNMRTTQEEDPPDPQLVRLDNMLVAEGVSGPEKGPGTGGPGQPDNAIEHSDYRDKLQQIR
MIYRQELEKYETACNEFTTHVMNLLREQSRTRPISGREIDRMVSIIHKKFNAIQIQLKQSTCEAVMILRSRFLDARR
KRRNFNKNATEFLNEYFYSHLNNPYPSEETKEELAKKCGLSISQVSNWFGNKRIRYKKNIGKAHEEASLYAAKAA
SHPLPGGTDSPTHSDSPHSTHGSLVGYFSDTRYLTDSPTDPFAPPPSTEGQMTIQSELYNIGAASNGYHENQAE
NLRQMIQTPNFHEGQPFDLHQSYGAAAQTGGQTHGPGPHTGTAAQPSDMNIRRIDPNVTRFEPVDFTKERSSS
SRDIAS*
>Locus_45317.0_Transcript_1/0_Confidence_7_Length_2004
MHSENLLRNDAFNNYGNPVQHAISPQEMQCHASLVGSSLPSPGNNQLIYQPVAMVSPPGSFQLHQSHGHAIHG
STPLSQIGVTQQQQQQQQPLQPSHLDPSMSINHQAMLMDHMLGGVGVDNCSMDPKKVSKRGILPKHATSVMR
SWLFQHIVHPYPTEDEKRQLAGQTGLTLLQVNNWFINARRRILQPMLEASNPESLPKPKKAKTAGRPAQQRFWP
DSLAALGNQETGGVTHQYPVNGDMDCGQTGSSSSPGSLHPTVAPLPPHTGPSPHTPGTGPLQSGR*
>Locus_46342.0_Transcript_1/0_Confidence_9_Length_2142

RVHGDFNSTQIINSSNFKTCLFEKKKRTNKINYNLVFTGQCLQRKRIYNCNMYAMEGFGDLYSGHLHFNPDPLHH
AASRDTMKYEFGGPPDFTKAPNLASSFFDQTDYRMDPYYNHSSAAAYGIQHHHSGVPTNNYTTLTPLQPLPPIST
VTEKLLQPPHPYDMMSHAHHHHQQSPAVSHHHYSTSMMAAAAAHHQRGPASFAAPTMQSHHGASAFETAQAA
AMARLPQSAGHMMGFNGFHGQVRADSNRPLHATAEISYKEIDEVKTRDVAQQVMTELKRYSIPQAVFAEKVLCR
SQGTLSDLLRNPKPWSKLKSGRETFRRMWKWLQEPEFQRMSSLRLAGTVSCKKKEEEQTRPKCLGPRKQRLV
FTDLQRRTLHAIFKETQRPSKEMQLQIASQLGLQLTTVSNFFMNARRRSVDKWKDECNANVLHK*
>Locus_35933.0_Transcript_2/0_Confidence_4_Length_3307
LLDQLPQGLAEVDEVKRTFESLQNEIVSLQSRIYRSESMFLDIQRRIHPLPDPVPSLELSSHNGHSSTQLVENRRL
EIDNGVYRERLINTVSELNQCKSELNKYKEADGNLLYSSKRDTETNNDSETSINNDRNETVDVQGSAEETIAADG
ADGGRIWNKGRQLLPGCGRKLSRRKTQPVKIKVEPQFLNMWEESGNNADPEANEPTAHAYAGGDSQESNHISP
GGSELNIFPDDYNSDVGMANSLSPLPPIDDDYRVTIKPTCQSALPPSVITVEGTVDTCQLVDYIKELLHKNNIGQKL
FGLKVLGLSQGSVSEILSKPKPHRKLTAKGREPYLRMLQWAMNPQGIDELARLASVSSSMLVNNKIFPHATLPDV
SEFTHRVSGFASGESPVAPETSNEFNEIVTTSTGNFEPQTPTEMAIRKILDKAKTEMDYERAATQKLVMHIRNGHP
IGDPISPTESELMCEAACFPSPEDCGEGSGENWANPQQEQTPSIIYNNNPSASSSVLGVPSMNNSVLGVPSNFH
GGGLVNLPNIGGGFNDSGFVCESSMDHAGVSTYNDVCVAKSESCASGGSPSCPPTSTTPVLYPPLNLANVKVE
VSEPEDTVGPFERDYHFGDPSLLPSCSVISVSPAVIRKAALLPLNHEDIVEFSDLDTVELTQQVKERLVRHAVSQR
MFGEVLGLSQGSISEILAKPKPFNKLTSKGQEPFIRMALWMRDVDGINKLKERHIQKTVQSHQMQVTLQQQHDK
QSEIRRIQNIRTKSKHVATVPSRPLFNRFQPIKPKPVGSSCNTNLLAYSPTSNSVASVSTQYSLTPSQDINLPGKDI
VDTVYPRSISHSSNSSGAESVDDRTIVEMSNQLGADRSAIDINPNMQLYMMQLDTKALCDEVRCVLNRHSIAQRV
FGEVILGVTQGSVSDLLNRPKSWDKLSLKGKEPYLRMRDWVRSASRVAQLRELQECRRMRKVDRFLLRGESGP
YKDDTWAPKSKRLRYSLDDDSKRFLLSVYQHVTHPSQDTISRLATHLNMQPATVYNWFQNRRQRFKVKDYYNSI
VDGGLSHSDIMKQLSGHLNVLSNLYSSDNVNSSIAGSRINEYNIIGANSAGIG
>Locus_ZFHX
MIIPTLSSPMLQLPENILQTGSSPLLQIGTKVVGMKPHCDGYQEIPLKRPRTRITDDQLKVLRTHFDINNSPNDDQIA
NISQETVLPVKVIKHWFRNTLFKERQRNKDSPYNFNNPPSTPLAGSATESNRNIPQVNNLISHEIKEEVLENIATIPT
HLKSEVFVKKENSLENIVDNVRAKVIKEEPLDSKGYLEHQAQSTNLVQSSSLPMYSNYSLSGTTNSLISPYNFPSY
TAGFGTSTTYPHLYSNSSSTSSETSYKDTTSHMFHTADEQSRYDLSKRAGRTRFTDWQIQKLQEFFEINAYPKDD
DLDTLSKDLSLNPRVIVVWFQNQRQKARKSFENLPMENDDDSSGRYFKTPNLNYQCKECSHVFQRFIDLIKHQK
EVCFKDSFDHHSPSNQTDKIEENSKDGGGKTLSNMEYPCNKCTMVFNSLPDWEDHQQLHLMPAYGPTYYPSLM
NASYPMMYMSGMENYLPNMGALTSSSIASPFSMNSQLLSNSDCKTTNSSSNIGFQIGQEKYCTTDDGGLNSSKD
RRLRTSITPKQLNTLYKWFKENSNPSRKMLEIISVDVSLKKRVVQVWFQNTRARQRKASVRQTESQMGSCRCPF
CFSLFNSEVVLDIHMRNKHSDAMTDIYNQGLPIPKGIHVTAGPFVGDNSANMNNSDVVDNAKNYSDDVMDYEES
ELVIDDYGYNLCGKGDGDDDDFEDNDDPNEMEQGTSYAGGAALDLTKAIGNNDTEAPFNGSYNNIYKSFEANLS
SPARQGTLVTDRNRRVSTPISVSVQSSTANAMNDDSESSLTHVDMEYSDGDPTSPLKLRDDKRYRTQLSHLQIR
VLKACYVDYKTPTMNECEELGKTIGLAKRVVQVWFQNARAKDKKSKQCGIFHTGGQMDTKPPSQCLLCNVKYS
VHLTGRDHVFGSKHLETLVKRSADVSCLAIKGSPPTAVAEPVLPQPVANCPTTYPEPTLPKAVGVAPVTPVTSDV
ASSTNPSAAYSSMMTAAACYPFIMGMNPYVSGNPFVNPSMMFASQDGAQKQHTNDMLRLMQQQYMSAAMAT
YPNLFASGSYSNNPPATATTGSNSQYMCNRCRLTFDTKELALAHQRSICFYGLSVNEADTITHLTPEVLTALMGT
GNALTSVTQTSSNQITKTESS*
>Locus_51994.0_Transcript_4/0_Confidence_17_Length_2917
MAVISQWFWNERFWLPDNVTWADLKSDEFHHYPEMSDLYIVLPIAIFIYTLRIALEKMVFSRIGVMFGIRSSSNQLR
PIDNLVLERVYLSVSRTPDDSRLKGLSKQLDWSEKQVEAWFHLRRAQDRASRLQKFNETLWRFIFYLCIFVYGCR
VQLHHKFYEDLRSCWLGYPLLAIDDYLYRYYIVTLAFYCSLMISQFTDIRRKDFWIMLTHHIVTLFLVTFSWMVNFV
RIGSVIIAIHDASDVFLELAKLLRYIKYLMLCNISFAMFTIVFFVTRLYIYPVYVIPSIYNSPDIVGFHNSWYIFFALTCSL
VLMHAFWAYLILKVVYSIAFRGKTDDSRSDSESDMCLEEDGNRRGGGGGDDKSDGSNGFAHQTYN*
>Locus_44536.0_Transcript_4/0_Confidence_10_Length_1995
RVHGDTTAVLSLLQQYNYSRPTTGRLDRLLQLPDYNLTMSNLSQSLYDTIWNEKRWLAPNHTWDELVNDPGVTL
RRPEARDIKDAFICGLCLLLIRYFIETCIATPVATKLAGIKAKKPSNGGLINPVLEEAYGKCTRPNREAIKGMAKKAD
MEADRVERWFRRRRQQDQPSTLDKFISYFWKTLYHITMLVWGFYLIQQEGIFPHTRKAWLDYPATISPCLHWFY
MVQLGYYLKEIPMHMKESRRKDNLIMFVHHVLTINLIAFSWLFDYVRIGLLIVVVHDVMDPVAYTAKVLHYAKLDYI
ARPLVYVLVAVFTVSRLYIFPYFVLYSTIFESLDTFDCYIPAVPESFQTPRSFRLPLISWWSLFNSQLLLLLVFSAVW
GVGFAKIAYRIVRGLKAQDNRSETELSSDENGGTPTRYQNGPVLQGDTSIKG*
>Locus_40648.1_Transcript_1/0_Confidence_2_Length_1479
YICIKEMAAISQWFWSERFWLPDNVTWADLKNDGTTTYPEASALLAALPLALLIYLIRVLSERLVFSNIGIMFGIRSS
NHLRAVDNLVLERVYASVSRNPDGNRIKGLSKQLDWNEKQVQAWFHLRRAQDRSSRLQKFNETIWRFLFYFFIF
MFGCQIQSNHGYIYNLKVCWQDYPFLNVPYDVHMYYIVSLAFYCSLLISQFTDIRRKDFWVMFTHHIVTVFLISFS
WAANFIRIGCCIIVIHDSSDVLLELAKLLRYIKYMSFCNVAFGLFTATFFVTRLYFYPVYLVHSIHTSTAVVGFYQSW
YVFMALSVSLVFMHFYWSILILKVIYNIVIRSKTEDSRSDSESDLREAEEDDDREGSGNSNGSVPSGYN*
>Locus_62117.0_Transcript_1/0_Confidence_0_Length_466
IILPSEDTMATHLQYDNGFWDPRVWLPRNHTWSDLKNIPGLTSYAEVGDLIYCLVTGFILLTIRHIVQHYLANTIATK
YFQIKPKQRNSKSIHPLLEQVYKDKSSPPDKVIEELAKKLDSEPPKIRRWFSRRHRHDQPGVLDKFNTYFWKTGY
HT
>Locus_44308.0_Transcript_1/0_Confidence_11_Length_2316

MDPDSIDTVIGKQDVYHYPQFINPHNSFLYASATTVPSCSDCLPCGMIDSSTKMAGQPIRRKKRRNKRPRVDVGL
PRNSYTLLKTPACGSNHAVDMKSFVAHRTTGSGYRSNQLNFNCYPFSRLSNFYAGNTLSQISRLKEVVHADEEK
VQDLSIKPKHFDKIGDVKKPDPEFNNNCRSGHFMSSSSGLDSRPLLSSYKQVNALNYDVTPIGTINDMTQWSGSP
KSTGTDGTGRSMDENLNKRLYDNVTGDELSEDDDVNIDDVDSDEDNCKLTDDSTPMLVDDNDVRDTNGMSMN
YPEDLMFSAAAVGSLRLHSNNNNNVETGVERNNFLCRIEKEYFSAQIAHMKQQLDALEKTFSKFYGEDILQIGNP
GNGPCQLIKHGNLPYQLRCHGNHPSSHLDGYCNELSQPNPHGNQLLQQSSHSNQLAQLSRQVARREHQRTSR
CSELGDIRVKVADSSPMQQVSTNQGWATNLTPNHLKKAKLMFFYTRYPSSAVLRLHFPDVRFSRHNNAQLIKWF
SNFREFFYLQIEKFSRQAISDGVTMATNIVIIEDSDLFRSLNLHYNKSNQFHVPKSFFTAVQATAQEFFTAVTLNKD
NDASWKKQIYKIINKLDSIIPLCLRSVNFIEDLERGCLGNSLAE*
>Locus_51275.0_Transcript_2/0_Confidence_33_Length_3000
MNSNNSPIFSTEAVTDYTHPELVDLNTTVINQPIRNLKRRNKRLRVDFGLSRNGYSKRRGLDNLLVEQPPPENGC
YDYQINQWSQQLTENYVSCRPADDVFKSDMEFDQDKVYDLSMKPTHKTVEDHNVKIINVDSVCNYERFIPIEGPR
TSNNHSSLIHTDVTTVSLDTDVKHGHGLQSHSGCDVIERYEDESIICDVTKPPYDKLCVDDVANNFNIELTTDSYK
MANEEDLGKSFPLANTHIATTSNASSIPQYLDEKEKYIAEILRMKEQLLALEETFSRLYGEDPPHLGCHGNRPIRLG
NYGTQSTQVGSQGNQTTQLGRHGNQTNQMGSHGNEPPQIGCSDNHLTQIGSRGNQSTKVGSHGSQLAQLGS
HDNKASKLCSQGNQLSQLNNHENSGVGEDIVEVLSCALTKAVDGAFKKYIPIIYNNNHMGGTLPVVSPPAQRHIA
PTSRDEVQIILKPVGGAVDTSAFQCRPIPLIEYRHQPVECSSRKPLEAPEPIPVTDQPVRVDSCPPYNNKVMDTVS
APIRRCLQGGSARDKKAQRLLVQRRCPDGIGKDNVKVTDNTSIQTASFWTSNLTPEHLKKAKLMFFYTRYPTSAM
LRINFPDVRFTRHINAQLIKWFSNFREYFYIQIEKFARQSVVDGAAIATDIVVNEDSALHRALNLHYNKANLLQIPKS
FVTAVTTTTQEFFFAVFHNKDRDPSWKKRIYKIITKLDCSIPLSFRSPNFMEELERSYQHEYTQ*
>Locus_53646.0_Transcript_1/0_Confidence_15_Length_4742
MHARFRFGLSLTQDNAASSLLKMLSEKVPSKLSVTDAGTTEQASSSLLYADTVNLGFASTKFDALKTTSKVRGPT
PNNPRASQTRKDLLRQKYSNRGPHAFNIDYKQRVIDRLMSLGANQYTGLVPKGAYSKLSAEYRVGHTTVQRFW
VQHVRDCSPNGSLPTLLEGQPPSGVDKVLNLLNQPRPSPNAKRGFYHHAINLLDAAQENPGQKVKLIPVTSDKP
RSTPSTTTVRSSTTNPSVVLHPLPTLSTSVLPPQTPPTLSGPFVPSTVSLAQPARCSWTDSNQRDSDDDADDSA
SDSVDKDPAYNKSITIKKEHIDPPQQCELPKLQCNTRATVDVKPLPVHLASNTARTTIEPPIDTSSAPLRRSARNSG
RRLRYTEDGFVVCSTQDDSTSRALVGSGSINSVNQNVLQFVRNAADIPTENHVDIARGSLPSSSNSNSSLGNTNI
CYSNSASGVTSLANDKNQTTIFEFATTPEGQLILTPVVRGASDPITPSSTASLQPQTSTELEIPTKVADDSKGKSLV
SNFKIPRPNYLSARNNPGSYFGKYRMASTNGTTGNKSSQVNWLSYGVNSVKLPADSIADDDLQPGEDNGEQAS
LETEKAPQPPTQVESQSVTVGPEQSPPPNRMEDHLNRYLASVSADKRAIFAERRPAVAAAQRAPQNRGLRSATV
AVRRPVRILPNVSGTSRNSVHHYYRALTTKYSKDALTARSVLGTTSATETDNPPATTAKLIPESQLSISQSLPKYG
DFAMSGESSTHAIKRNQIANCVQKELDDMKIPSGVFTKAVLNKTRGILQHVLKPEGTCQHYLPVISDFLSKPLEER
RELYNASQNNIVEGRIAKRKAMFEQGRTKPARLRARTVISQKDRAIMFQHFLTRNQYLTANELLHLSRSLDIPQPTI
RIFFQNLRARIRARTRTR*
>Locus_39214.0_Transcript_1/0_Confidence_6_Length_1426
MSVNLLQPCVHSSLDREHYSGYRSYYSNATTHLASIKNPTQEADHRPTVRSLKQTRYRILPGKPTYPRLSDTKRD
SSANSDVEFTKSGVGIAGSTSKSISCPRLRTHSVESIGSSRSFLSAQSNVSRITGKPTYFSVFNRKESPSFGNVDE
LPARNPYPKYQLPSDDEDDFSTPLSHSPEFHSSQGSLTNLNDTTSAGMKTTDLNISIDSADSRSCSRTSGYKRRP
YPTYYRGVGTVYPSPSDSPGCLPSSATIITTSSTDHYPPWNTAALYRPRTAVSSVEAKPTTVNLNLDSEQLSVLER
AYKWNKCPSAQEMVDISNNTDLSTSYVKAWFENRLSEWRRREGLPARSSSVFDRFRK*
>Locus_32393.0_Transcript_2/0_Confidence_5_Length_1036
QSTWGMVLPNYMGNTYRSYNKHFDQSTIKGGHLLTPPTVQTSNSLSYVCSGDNQTLSSETICSSGYDVMGTVIH
IKDEPADDVVDVTPEGNCDADDYNCKDVLDRIRLEIRSMNVRESVFITACLGEDVAKSIEQLTPVGVPLDIYGYLR
RNWSLNTDWNARLCYFFDKQPYQRAAKYHQVARQRTHKPMIPIQPEQLHTEAIYNSVEIRPSLSVYNRDQPLVN
QESSSHDTRMDYTNSSNQSRPYYRQKERRHRTTFSEHAKSVLTQYFMTKTHYLSPTERMSLANQVNLPVETVRI
YFKNLRSKNKNFLKPSKSFGGGKS*
>Locus_44118.0_Transcript_2/0_Confidence_9_Length_1342
MAVNYLQPRVYGPTRHTLHGSIAAPTVSKSQTNRSQTVSVPLTDTTKSVSVRPSVRTLKQTRGRIQPGKPTYPRI
TESTGFDTNRNRLSSKSLSCPRLRTQSVESIGSHRSFASTASNPQRIQTTYFSVFNRKDSPSFGNVDQLPSKNPY
PRYQLASEDDDDFLTPPSRSPEFHSSQGSLSNLAGLESPPAARLRVPDLLDRSADSCDRLRSRSVTPNSKRYPR
FYQGTSQHRERTPESLSYRPTNLESTPDARSRTHANIARSTPSSRTFTLETPPSSRTYDVAVSPTPTTDTTTHSA
AGGVSMRYPPWNSEAFYRNRIHRIVAKPSTAEMELSSHQLSVLESAYKLNKCPDTDELSILSEKANMSTSLVKTW
YENKLSEWRRREGLPARGSSVFDRFR*

	predicted class	PFAM domains
Locus_10725.0	ANTP	
Locus_15302.1	ANTP	DUF1374, Homez, RcbX,TUDOR
Locus_17210.0	ANTP	
Locus_20062.0	ANTP	
Locus_20731.0	ANTP	
Locus_20893.0	ANTP	

Locus_29958.0	ANTP	Homez
Locus_31941.1	ANTP	CUT
Locus_33755.0	ANTP	
Locus_35817.0	ANTP	BROMI
Locus_35873.0	ANTP	HTH_23
Locus_36321.0	ANTP	Homez
Locus_36518.0	ANTP	
Locus_36577.0	ANTP	Ax_dynein_light, CDC45, DUF2115
Locus_41714.0	ANTP	Luteo_P1-P2
Locus_4231.0	ANTP	Homez
Locus_44961.0	ANTP	Transpeptidase
Locus_44999.0	ANTP	DUF3078
Locus_45472.0	ANTP	
Locus_45900.0	ANTP	Homez
Locus_48679.0	ANTP	Death
Locus_49193.0	ANTP	Med3
Locus_49354.0	ANTP	
Locus_49786.0	ANTP	ARF7EP_C
Locus_50700.0	ANTP	Homez
Locus_50937.0	ANTP	
Locus_54562.0	ANTP	
Locus_54694.0	ANTP	HTH_23
Locus_58015.0	ANTP	C6; Homez
Locus_58481.0	ANTP	
Locus_6133.0	ANTP	DUF382, DUF3870, DUF4445, Imm7
Locus_18856.0	PRD	HA
Locus_21928.0	PRD	
Locus_24274.0	PRD	HTH_23, HTH_29, HTH_Tnp_IS1,PAX
Locus_27707.0	PRD	DUF1631, HTH_32, QLQ
Locus_29086.0	PRD	
Locus_40214.0	PRD	Adeno_PV
Locus_40829.0	PRD	OAR
Locus_47651.0	PRD	
Locus_48073.0	PRD	
Locus_49939.0	PRD	MRP-L20
Locus_53544.0	PRD	ActA, OAR
Locus_54213.0	PRD	
Locus_55290.0	PRD	OAR
Locus_10005.0	TALE	Homez
Locus_26582.1	TALE	Homez
Locus_28871.0	TALE	Homez
Locus_36182.0	TALE	Homez
Locus_45317.0	TALE	
Locus_48316.0	TALE	Homez, MRP-S32
Locus_52332.1	TALE	DUF1319, Homez, HTH_19, PBC
Locus_52537.0	TALE	DUF1319, Homez, HTH_19, PBC, HTH_23, HTH_26, HTH_3
Locus_14051.0	LIM	LIM
Locus_22501.1	LIM	LIM, RFXA_RFXANK_bdg
Locus_38224.0	LIM	DUF581, LIM
Locus_45303.0	LIM	LIM, RFXA_RFXANK_bdg, Treslin_N
Locus_48370.0	LIM	LIM, ORC5_C
Locus_50222.0	LIM	LIM
Locus_52392.0	LIM	LIM
Locus_54455.0	LIM	Homez

Locus_61503.0	LIM	CDC45, MmgE_PrpD, Orbi_VP3, RRN3, YL1
Locus_7973.0	LIM	LIM
Locus_49053.0	LIM	DUF164, LIM
Locus_27996.0	POU	HTH_31, Pou
Locus_35933.0	CUT	CUT, HTH_23, HTH_3, Med9, Terminase_5
Locus_46342.0	CUT	CDK2AP, CUT
Locus_36924.0	SINE	Homez, HTH_23, HTH_3
Locus_42137.0	SINE	
Locus_40648.1	CERS	TRAM_LAG1_CLN8
Locus_44536.0	CERS	TRAM_LAG1_CLN8, DBC1
Locus_51994.0	CERS	TRAM_LAG1_CLN8
Locus_44308.0	PROS	HPD
Locus_51275.0	PROS	HPD, RTX
Locus_48953.0	HNF	HNF-1_N, HTH_19
Locus_50181.0	ZF	Glycos_transf_N, HA, zf-C2H2, zf-C2H2_4, zf-C2H2_6
Locus_9311.1	PRD+ZF+ANTP	HTH_19, HTH_31, POU
Locus_62117.0	PRD+CERS	FKS1_dom1
Locus_53646.0	PRD+ANTP	CUT
Locus_53834.0	PRD+ANTP+POU	HTH_19, HTH_31, POU
Locus_52441.0	CUT+ANTP	HTH_31, Pou
Locus_51594.0	LIM+ZF	LIM, TNFR_c6, Homez
Locus_39214.0	PRD+ANTP	Homez
Locus_44118.0	LIM+ANTP+PRD	Homez
Locus_32393.0	PRD+ANTP	

5 – *Xenoturbella bocki*

>DN63819_c0_g1_i1lm.182148
SLENKFKLTRYLSVCERLNLALSLSLTENQVKIWFQNRRTKWKKQNPGRDITAPELNQPSATSLIHSRESFLSPGL
YFGGN
>DN14752_c0_g2_i1lm.34886
KEYTRENYVSRPRRCELAAQLNLPETTIKVWFQNRRMKDKRQRQAMSWPHLDPNYYAYLVSRGVLPAVFPYPT
NNPGLVGPSPGLSGGTPLPYYPGFHHHNHPTAVSSAPPSFPPLNPAAAA
>DN115451_c0_g1_i1lm.1141018
DPYLHTAAYNNLLYATANGYRGPHTPVLFQFGITSHDSLQSAKRTRTAFSSTQLLNLEKEFSRNSYLTRIRRIELA
QSLDLSEKQVKIWFQNRRVKQKKLLMSPSRADEAMDRTEPSANIDNP
>DN2325_c0_g1_i1lm.5825
VPDYPTSSSTGPTAAAYQNCYFQATPDPHQPTFFTNRPESNTFDYTSNTMRVEYASAQSGPHVTRHRPDGDKL
SPGIISDRPTSDGSASSNGGGTLSPHSCSGKDTGGIASAASSCYNASPDFPYSWMKIKRSQPRAVSKNGFMPGY
PSRSRGGRTNFTNKQQTELEKEFHFNRYLTRARRIEIAASLSLNETQVKIWFQNRRMKQKKLLKEGKLAALC*
>DN124767_c0_g1_i1lm.1167403
GGGGDPAGQTCFYTPPQDALEMTAGDLAGVTALVGGGDKDDSNCSEYPHHYATAPLTPQERELPGSPRSTALS
TPNSNVVIYPWMRKGHSNTARLVGNGESKRTRTAYTRQQVLELEKEFHYNKYLTRRRRIEIANALLLTERQIKIWF
QNRRMKYKKENKLTGNADLNNEGIDLTTGGRQDSFMRQAAEALNASGTAFNIENSKQHSACQIANHHGGVMLN
PLGESSADDMSGGDGGANH
>DN84275_c0_g2_i1lm.368677
QQQQQQQQQQQQQQQQHSFHTTNGGGHVVANALEQYNSPDSWKGRFDADPEQHLLQYQHAQQQQHNHHHI
QQQQSQLQENNTSVTLFPWMRSASGVNRKRGRQTYTRYQTLELEKEFHYNQYLTRRRRIEIATGLGLTERQIKI
WFQNRRMKFKKENKIGNKRSVDYAESGSTSPPSMTGDAMVND*
>DN3733_c0_g1_i1lm.8775

MSVAATGTSMANWAPINSYYMDPFISSRDASAAATCNTAARNADLSRAVFSSGPAASAGIHPDKACGLDFAPAS
AGGIGSSPYGATASGLVSQYQATQCMGADPTGAYRGYQTTMPGRDGHGAGVVDHLDAHGSFIGTPNAVGTSK
HQDVYYGQRQLHWAYGTTPGQIFNHDTTPNFKTYLYNDSGRIFSQTGLPNHVGHAGNFAAATGAGHVASPASG
TSPHFGGDAKRSQVCAAFCCTESPTGLPIHTSPNTIMDTTAFGTKPLADLTDPYGADSILVKPDTPPQSVDHVHD
SSNSTDIGKNSSNNANWLTRNPTRKKRRPYTKFQTLELEKEFLFNTYITRERRLEISRALNLTDRQVKIWFQNRRM
KMKKMNSQQQCHSS*
>DN76496_c0_g1_i2lm.266175
HEDMTVAWSDDSNDRVSVMGPCTFAPRTLPGANTSTEIIMDQVLPDRSNPRSCYVDESHQGRWSRENQVTRR
GERTDRATMYTPHIPPTIYPWMTKSYSTKSVYCRRKRSSYDRLQCLELEKEFHFNRYLSKRRRSQISKHLGLTER
QVKVWFQNRRMKHKKDRQRQLQTR*
>DN44680_c0_g3_i1lm.102754
LSDINDDGKDVGDAQSNPCSPLNGSMVGSENNSSGLKKKKPRKARTAFTDDQLARLETSFERQKYLSVQDRMD
LAANLNLTDTQVKTWYQNRRCDSLKTI*
>DN81549_c2_g2_i6lm.325476
HYTDHPKHVSSYPARSHGPIYRPPSGSVPLCMRRVGYLGVDDDDLCGGSSFYDDRQQKCRRSRTVFTESQLVG
LERRFADQKYLSTPDRITLADQLRLSQIQVKTWYQNRRMKWKKQVPI*
>DN90457_c0_g2_i1lm.535164
IALAQQQQQQQQQQQQQQAQQAQQQQQQQDQQQNQNQTVTNGTSAQSPLTISTQHLTAQQALSIQNAVKG
VPGARIENSPTPGDEPSQLEELEAFAKMFKQRRIKLGFTQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCK
LKPLLQKWLEDADASLSNPQSILGTPPTPDPLGRRRKKRTSIETNVRVALEKAFLSNPKPTSEEIQMLAESLCMEK
EVVRVWFCNRRQKEKRINPPTSGGLLMHVDSLKPPTMVDSPMKSVDGSMKAQVTYQ*
>DN105712_c0_g1_i1lm.1111542
MIDQSGRPHSKLPKLSTENEHNSRPRSGTTTERMHTANPMFTVDALISRPTNGHVSPSPALSLPYTSISSSPPPF
DSFGLHKSFYEHKYLSNSSSLGFGPWSLGAQPGQPQAPIYYVPRAMLCSPAGRSVSPVSSRDGYSSFRPWMSP
IQNRLALHPSALPGFYHRAGDGSPFLPPVTRKPKRVRTAFSPSQLLRLEHEFEKNHYVVGEERKQLANVLNLSET
QVKVWFQNRRTKYKRQKLEEKQDGDGGGNSPDHHMAECNTDADDSRHEDIDQQPQSPASLNEDEDEPDNKT
SDVEQNSA
>DN63160_c0_g2_i2lm.177407
RFIILLNILLFPGSTSETSAAQKCSLRKHKPNRKPRTPFTTSQLLALERKFRHKQYLSIAERAEFSSSLSLTETQVKI
WFQNRRAKAKRLQEAEIERLKLAARPYLHHTLGFCLTGASNDMMYGYHSLGHPGPLFQLPPLSSHLHGLSAPHG
FSRNDITSPGHYVFPVEARAVDTSAASDADRTSSCNSVY*
>DN131037_c0_g1_i1lm.1186210
SLENKFKLTRYLSVCERLNLALSLSLTENQVKIWFQNRRTKWKKQNPGRDITAPELNQPSATSLIHSRESFLSPGL
YFGGN
>DN93010_c0_g1_i1lm.661361
MSISPKQTTPFSVTDILSPIEETYKKVASMDAFNLGSPAMPSAYRPAQGTQSLQHSMNGMNPAVNSQYHVHVPQ
LTHAMGTAYCNGSVSDFTAHYNDSVGVRQHAASGWYGAAAPDPRFSISRLMHQPPSMNMNMSALSFHDGTKM
LPTQQKRKRRVLFTQAQVYELERRFKQQKYLSAPEREHLASLINLTPTQVKIWFQNHRYKCKRAAKDKAAQEQQ
QQHSPRRVAVPVLVKDGKPCSSTSDCSQHDGNGTASVVTSSQSCMSSGPDSPLPGVATEAVAHTPSALNTALN
CAVNNTGQVNSGLDSCTMNTVTNNTSSCSVTTMDFGSQNASLTSPYLLPGRAW*
>DN101095_c0_g1_i1lm.1098549
QTYELERRFRQQRYLSAPEREHLASIIRLSPTQVEVWFQNHRYKLKRSRQEKGLDMNPLPSPRRVAVPVLVRDG
KPCQPGMNQMVKPQDPMQLQHNLSSQVNMNNALSPVNSM
>DN50727_c0_g1_i1lm.123834
ASCWLPQTWFRLVFIACWGNAQFVTSDMTGKFSISALLGLDDNYSEAKARVLVSSKCYGGSSPTKSAPTPPTFT
SSSEDRYCRNSTENKLTIRPEKPGRSGELSDEDNQAEDISLDIDGTAKGTVAFNLTPSPPPAPVVIPSDGAGPVDS
MGCSNSNGMKRQRRKPRVLFTQAQVYELERRFKQQRYLSAPEREHLAHLLKLSSTQVKIWFQNRRYKCKRQRT
EQMSRGHVVGHEFLNTMTGSGSSLAAALSGYPHFIQQPPRHVSVPVIIK
>DN92869_c0_g1_i2lm.652722
MSQEDSDSKTASGSKNADMKKDDVVAKKKPRHSFMISSILGDVDSSDDRRNNNKSPAKRNHASADAMISSSLEP
EYPESQRLHVTPTLVPAFATHPAEATSWYSWFHSTHSNHPHVEPSLSRALTLHKHSHSPVSAHTHYLSDKEDAA
GELEQKQCPTNCSLGTSERSSSPKSTNDDNEGQNAVSVDNDKDDDEHSDTSGHGGQKKKNRRKKKTRTVFSR
SQVFQLESTFDMKRYLSSSERAGLAASLHLTETQVKIWFQNRRNKWKRQLAAELEAANMAQAAQRMVRVPILYH
ENSSQGLHASDTLNHAVTPHLPFYPVSYQTGSAHFPPLSVVRPPLSGLV*
>DN75963_c0_g1_i1lm.260766
MEESCANSNNIASYRSSVRNEYLNSQKGCSENENYPISQYRNPFPTNNCSSPDNDDGIVLHQPRMTSKINNNSA
TTPHGINDILSRPDRFNCQPPTQTHRLPPFCLEGAICLGPSGPIGFHPDHVRPVFSSPAAVYWGSPGVLQDTSLR
DNIRASTGSYNTAVNRSGTQLEPSGKDSSKARKHTRPTFSGHQIFALEKTFEQTKYLAGPERRRLAFSLGMTESQ
VKVWFQNRRTKWRKKHAAEMAIAKKNHDSVAEEHMMSVSSSDNEEYLDDD*
>DN135951_c0_g1_i1lm.1198937
HPYQSRTPPKRKKPRTSFTRLQIMELEKRFHKQKYLASAERSSLAKDLKMTDAQVKTWFQNRRTKWRR
>DN90720_c0_g1_i1lm.547254
NYVISGPNYVISGPNYVISGPNNVISHPNQVIPGPNHHLISGPPHGSLVPDMTSSDVGKRMKFSIDSLIASSGTGYD
TRGCHDYTPAVLDTPYQHYPVGINPPGPFPWLPNVTSLNFPHFLGGLPYPGVVGGFPLSPYHLSNNKHETTTSV

LNGSIPAHTAHTVGNHSNSPTRGGPHQHIVNDPPRCGGNGGVDATLMGRPSPPSTNSQYILAQSYCPNATGGSI
YNYPPHTSRDTNPPTTPSNTATSLGGVEQGSGVYPSISGPSLLDLDTARPLPPPVTADDTTTEDRLPSHSPTAVD
PCRDLSTSSPQGPIPAFQPVGKNTVRHGPLVAVPGSPMITHAPQFPATFAMDYSNTDETAVRNSSVSPGSGSMY
SDDDVSDSGDMDSMLYMGSNCKDSDSENNRVSSTSLGKSRRRRTAYTSEQLLQLEREFHGKKYLSLTERSHIA
KNLKLTEGQVKIWFQNRRAKWKRVKAGMANRNRQVTGGSKIVVPIPVHVNRVAIRNQQQTQQPRHPVANQTAA
*
>DN86948_c1_g1_i1lm.424178
PAIKTGKGSKPRKERTAFTKHQIRELEREFTKHNYLTRLRRYEIAVSLELSERQVKVWFQNRRMKWKRVKGAAR
SREETAKFLTDSERRLRVLESPSSSTGNPITPPKLPTPDIPVSPYHHTPDVMQSAPMTSHPVEQQQAFPCYLENA
VKVFDALPVSTSGEVLGPVSTAVGVYPQLHHTGQNSDNSSTCSDSP
>DN69228_c0_g1_i2lm.213920
MRLKSFYINDILASDRPRPTSSGIPPDRLLNQAEIDSTFFTNEESRTGGDLRVCRRSRNSSRCYNEQNPIQELRKL
TGRVLYKLEERRSDETNGDRNSRTILSLCLFGDRERKLRSKRRRTSFSQEQIRLLEKKFEEKQYLTAKERDDIACD
LDLKASQVITWFQNKRSKIRRETEEAK
>DN94680_c0_g1_i1lm.774060
MSNFHPIINNGGHQQSRIFCPPSPHMFLHPSMAIGPVSSDVQMVIPSVSVPMGYFYPIQTVPRTWHSHVYVNMR
DTPTPHLIKHILELTATDDESEHVEVDAVRSVRTMPPDVANEFIPRPDKPSDVLPPGAGFTQQQHRLPKLQPIHVV
TTDVSLQQQHPDGDMAPRGSAKLSRNFSQPYRVKVEEMPLSPEQDLPGHCRSLPSELTSDTPPQIIKVDSVVSD
LSSKESPALLHKKKRQMTRKERKKARTVFTGRQIFELERQFEAKKYLSAHERAEMAALLNVTDQQVKIWFQNRR
TKWKRLDKLKNHSDLREDELSSPKTKARKLIISGKYGTLRNSPRHGGSEPLISSPSSSMRRRLNLDGDTVSGNSR
GQYAVCSKGGSAPANNNNDESYVQASINMDLGLFHHQQHYGVSQNGVSSARFLECSEDGEIDDVDEDVAEDD
DEDEEGGGELTIQGRVDATTDGNPSTTTCDTKP*
>DN41925_c0_g1_i1lm.92873
ETFLSSSTSDHQNCAMRVSPISDGDSESSSTFEHASLSKLVVGASNLKNSSSDGKRMRTVFTYKQLQKLEKVFQ
DQQYMVGEDRCRLADSLGLSETQVRVWFQNRRIKWRKETLKDQKAKLALLEIRSRNEADTLEEAEKEAA
>DN122354_c0_g1_i1lm.1160697
AWRISSTANDQPRCESGVGVNTSVGIKKQSTSSVGIISHQLFYYNHLCYPVIKTTNDLLTQVTMVVAATSTPTPTD
MQSSAFSVENLAKSSSTDRVILSPPLSGPSPKASSPTSFFVKDIVGLSSSDHEDDTATALDSSYESYSDKCDDVA
SECERGDTEFNSKRKKKTRTAFSPSQIYELEKRFRLQKYLAASERGEFADKLSLTDTQVKTWFQNRRMKYKRQH
QDAENAQLAALYNAYGQPVPMVPTMRSNVSPHAAPISMNQQTAFTPPTYPIRPTALSVPVSHALQAHIPMPALNS
ITPTEVHLSSAAQNIGITYPPPTHMLQHPAVTNAFHCGPMMHGLPSPYHNFHAHPSMLPTHLSQHPTMINAFTTPI
LGQ*
>DN94983_c0_g2_i4lm.798955
MSACHSPTALPLYTPGPVFGHTYATPRQVAHSMAPKPTTSFFIDDILGLRNGDNFGECVPRDFPTGTSAAAVPPT
NIYRMDYTLPDTRHQLGGGFPGTTPPGTDAPFPLAAAGRPPTDDFALPLFSTYPAMLPGTLRSPYGQLVYTTAA
GAGAGVSHGHMPPSVGSGGSCNAPAHPGFPKEEIYQTAGLQWNIYPTVRMAQKRKGGQVRFSNEQTLKLEKT
FHKQKYLSPSERKKIAKSLQLTERQVKTWFQNRRAKWRRLEQESQKSTNDTNNGFNVDCRDRENNAVEFDDID
VETV*
>DN84186_c0_g2_i1lm.367331
MEKYPSMSGSAGHQIAGMVANHQHPSPPNSYRHRANSISAYPAAFNPVPAATGLVVVAGISPPAVSGQGSSPS
PGPGPSAAELNKNPQPGGGTAVTATGASHSIDDILGKSPAVSGKQDRLDGENHRSSREKPVKRERSESPHRAD
KEFVLLRKDDSPRLNLTSSLGCTSPDCSNSDGSPKSFPNDACTDSSPSELKLTDDVTIKRKKRRNRTTFTSFQLD
ELERVFQKTHYPDVYLREQLALRADLTEARVQVWFQNRRAKWRKKERLQQMQMRQQLAHQGYDIPLTAPRPE
GYAQIQANAQWQASCGQICQPVTPGAAMVGTAACMSPQSNVPTFMGISNANQLQQLHAAAGIQAGSLATAVSA
VQQYPQASMTMVGSPMGNIVNPHCQPPGGVINSPSHLDYTRDFCQQSADRRSSSIAVLRMKAKEHSAALGMMT
GYPQLYAAANGTIQSVDTSTVKF*
>DN48278_c0_g1_i1lm.115742
MSSSHQHSVFSIDNILAPSRSPTHNSQPHIPAALPPTSAYPPTVNLYQHHQHRHYPSSPHSVSAMGYSLIPPPANI
PNLAFATMLPPHLHPSSPFHFDYWQHHLLRSQRRKRRHRTIFTEEQLEELERTFEKTHYPDVLLREQLAVKVDLK
EERVEVWFKNRRAKWRKQKRENDERLRKDTPDTQNGSPTKCTKVAMDDRTSPSPQIKDGDQARPGPSSNTPE
AEVAPADVQDPYSKSDSSFHLTDSDNDLSLSPCSSAGEHSRSPVVVCRDKLLANGSLD*
>DN86933_c0_g1_i1lm.424448
TKCSGPMMDTNYPIHLKTESGSNHLLLSSSSSSKQSLPHLVDSLHSSSGDGLSQHPEGVHGKFNDKHDDHGEN
GNKTDDDKPSKPKRHRTRFTPAQLNELERSFNKTHYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTT
NVFRTPGSLLPSHGLTLPGQFSSSLGDPLGCTFHTPESRWAAMSGMTSMGGGPLSLGAASFQRQSLGNSLSQP
GMQMNSSGQCSVSLAGLHVANSTPPMYQASAFPGPAVTSDGLSPTSPSPLSVTSHLTCGGSPEESDMWRGSSI
ASLRRKALEHAASFGSMNSMNALGGCNLDFR*
>DN55770_c0_g1_i2lm.144367
MFGDMAYLKGHPYSVNVNGLSIGPSPNVDLLHPAMSYPAWCSVSAATQPQNQGRKQRRERTTFTRAQLEVLEA
LFSKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCRQQAQQQQNGGQPKQRPKKKTTPPPNESTPNLTTTS
TIQTQYKPSTVSSSITTANSIWSPANLPPAVSSDPLGSSSSCMGGRSPYSMGGGAQAPTGPQCYAQNYPATPYF
SNIDCSSYLSPMQLGNSSFNQMGSSSMSSMSSMAASMSPHMGQNSGYSQLTPSQGISRAGAGSCNPAGGMT
GLTADCTNISDYKDPAWKFQML*
>DN77439_c0_g1_i2lm.275737

MPHKDLCVGHSGVNQLGGVFVNGRPLPDGTRQKIIELAHSGARPCDISRILQVSNGCVSKILARYYETGSIKPKAI
GGSKPRVATPQVVTKISQFKRECPSIFAWEIRDRLLSETICNAENIPSVSSINRVLRNIANEKPILASSCESLQDKLR
LFNTGSWRAPWFPTAPTPYTASLDMQCLPPKTETDPDELIRSASSHSSDTQTSSATTADEEARMRLKRKLQRNR
TSFTPEQIDSLEKEFERTHYPDVFARERLAKDIDLPEARIQVWFSNRRAKWRREEKLRNQSRQEGCNGAAHPLN
ATLTGTLYHPSLHPHHQPGANITGPGVPGGRADYGALSAHHVASFPPSGAGLSPVSLQASTPSVSYSCMMPPN
PAVSGGRTYDHLACHRAGGSNSYTNHHHGQAMLPSHMRAAHAQGGSSQTGLISPGVSVPVQVPGSPCSEYTQ
YHWPRIQ*
>DN86691_c1_g2_i5lm.418460
MRLKRKLQRNRTSFTPEQLESLEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRLQ
QESVKHNNSIPNHMTINSSFSHNNIYHHQPPLQHHSTVMAPPPTCSSSSIADSYSAALPPVPNFSMSNSIPNIQPQ
NSSSYSCMLPNAPRTYDPMGYSRASHQLGHQNPHQGAFQSPVTSSAMTAHSTGLISPGVSVPVQVPGGSIDMA
NMASHHPQYWPRVQ*
>DN49217_c0_g1_i1lm.118984
MEPMQVSLSLADLVQVTPAVASVTGATQSVAVGSVQSRGIKPPSAVSGRVTGMASCGTDSSGRAAGSTNATYT
TPECKGSKSGSDKQSRDSSHSKGDNTSTVAISDDDPKKKKRQRRQRTHFTSQQLQELEATFARNRYPDMATRE
EISVWTNLTEARVRVWFKNRRAKWRKKERNQVQEYKTSLGAQFNTLMPTGYDDTFYSGYHAGYNYNNWASKV
SATSGGLSAKGLPWPLNSTLQPAAQQMCFSATPSTIGNSMAVPNVNVAGANLNTLGGNLNSLASPSSTCPYASP
SPHQYMYGECSSSIASLRLKAKQHATLGYPSPMRQPGLSACQYAVADRPV*
>DN88387_c0_g1_i3lm.464016
VKVSTNLIPTIDKVTNMTEEQRVPDERGPATADKRKCETDFSGLYIDPNPRAPTGLHEHHHHDTYFTVADHDGAF
KKLKPSDAGSGLNNATQCPTPARRRHRTTFTQEQLNELEAAFAKSHYPDIYVREELARITKLNEARIQVWFQNRR
AKYRKQEKQLSKALSPVIPTCNGMMRNIYPPPVPHTQRGYPHPHQYPTNAMSRYPPQMTTSTYSPMAQFTMPS
AAPMSMGRNPVHVDTSGVVGAEDDWYNKSLTALRMNHHSHTNLSAPVLQYQT*
>DN70414_c0_g3_i2lm.220060
MASTFNEDVLKIPVSCVSSVDGNKGRHTSHSIREILGLEELPGQNMAPEPTSAAGAHQLNSTKSSKRRHLFHFAS
DSRKDVNVSLALNDNTVDSSSGKSEFSNKQRDRPSPGDRIQFLDKPHSKEPKKKTRYRTTFTAWQLEEMERAFE
RAPYPDVFQREELAMKISLSEARVQVWFQNRRAKWRKLKPAKSSTEIAKLYAPYLNETSPHRRHDTATDPTDHP
NEPRDHSPLPTSPVNGMTSYVSEPWSAMQAATSTTDRIPSTTFSNFANFVNQNGSRIPIQSFFAPTLPYLYASGP
GQPPMYRGYSCINPMGLNSHVLGSDVIINDMYSKGGAFANKPDFLMRGERRSAGSSANTQMDGVCRMDSD*
>DN67189_c0_g1_i1lm.197813
MVHLSHGQQATQPMLQRNMAVSIVENSPRSLNATPSPSKETASETPRQDEYDYEECESIAGNAPGQNLTHTIDAI
LGLKTATPKKNGINNNDGESALFGSKEDENSRSPDASPPQTLMSPNASDSHESVRYRLVANNGTQSNRHDDDR
DSRPVSHSHPSISQENSSEYKSDSGEDDGLSENQPKKKHRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAM
KVALPEVRVQVWFQNR
>DN14020_c0_g1_i1lm.33082
ETVKTSRFSHSFMEGLAAFVSKSFHAVYNDSDGPADQVAARSKRPTDSFMDASGNHVQNIDELVTEDPGQKAR
AHNGPQTPNPGDDQGGIDHYRKMNKHRESQSNGRATNHTSHAEDSDPGVLSPILKNPAENKRTPNNTSSHNNN
NNNNNSKSIKQRRTRTNFTLEQLEELEKLFDETHYPDAFMREELSSRLGLSEARVQVWFQNRRAKC
>DN12917_c0_g1_i1lm.30558
RNSKRRRTRTNFNSWQLEELEKAFLVSHYPDIFMREALAMRLDLIESRVQVWFQNRRAKWRKKEHTRKGPGRP
PHNAHPLTCSGTPIEPEELERKEKDRQEKKKKRSLEKEMMREKRKIEKQLHQTKKETSDNDEGVTTTGDATEGS
SQTSKQGQKLEKVVKTLKKGKCRKKSRVSKKNATMIERKSALFTIDNILSSETVKPTSIVTKRTNPIPGLCGGLTGG
ERRVIGGENCE
>DN86983_c0_g1_i1lm.424471
MTREVEIMSHITNSSTTRVAAPSSLAPASSQAHQRSPFAIHELLGLSQRHTVRSSEKNSMHNAHHLHHQLQVSSC
LGKPVCFPDPAMTMAYFSKPHFIPGLGASLQSPLLMSQKHDLSQNLSMGGNISPEDISEEKSGLGGGQGKKKKK
RRRHRTIFTSYQLEQLEKAFTEAHYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKKEKCWGRSTVMAEYG
LYGAMVRHSLPLPESIINSAKNGIEDSCAPWLLGMHKKSMEAQGQLTEDSDPSASPPIPLDHPQMASLAAAGENA
NNDSDEIRSNSIAQLRAKAKEHSEKVLNHAIQ
>DN122491_c0_g1_i1lm.1161007
SINDHSQPQNVLRGIHPIISQSTLRYFIILLISSINYAMQQQGFNNRATGKVTTAEITKDFRDWVNSFQLTQSTQENP
AANPTLTDDDEINEIHSALPCTETTNGNRRPAKRRNRMSFQPEQRAILEEAYRVNKYPDILHRAWISQMTNIPDLRI
QIWFQNRRAKSRKREFKYLPPKGSLALNQGSPSYNRPIGLYLRSQSRRLPNEVPIFPNIIQPNNNSPIEPSRTVIVP
ADDSAEQLSQLQSSMNNRAKNICDEEWEEQLGHFLGCCPPNQVNHPIHPAHQYDQNH
>DN102158_c0_g1_i1lm.1101472
LGLRPICHTPNQQSAFQPVSFSSHEFDDGMIRRKQRRNRTTFTLHQLEQLETAFAQTHYPDVFTREELALKINLTE
ARVQVWFQNRRAKWRKHERQKEVSIQHDKSATSSNTDPISRTLPGSSGGNSNHVIDS
>DN122218_c0_g1_i1lm.1160059
FLDNFRVFVIFEPAMQHTAVSSYNVESLLTYTPPSPAVTSHISSPPTTFQGAGPEISSAGLTSCYVPTISEEILMDE
MMYNRRRQRRNRTTFTPQQLEELEQLFSKTHYPDIFVREDLAMRINLSEARVQVWFQ
>DN83632_c0_g1_i2lm.357974
MIRVREPHDSTRRARVASSASIAPADVERKSPCVCSEEKKTTMSPFSVEHLLSKEIRPGASESNNNLSCRKCQTA
SSRRDRAEQSGFSYIPVQPALNSSPANKTDSTCLIHVNGTCCSNADSDAARSHDVRAARSGSEIANSDDDVDVIS
ESDEESAGFDDIALKDHHHSRKVRRSRTTFTTFQLHHLERAFEKSHYPDVFTREELAASLHLSEARVQVWFQNR

RAKWRKTERANKHESGTGVPIYGFLDGDVMPSRVVCQYPNSIWPGEVPPTYGLPFLALPHMWPRIGHAWGRTG
VPFYPSVYQSSAQYSAEVAPVSVPPHPALREAKNMYPDSERSYSSQQHIHKPRNSSRSPEISQPVAANIASSRVL
DIVTPGSPTEGNVDMRI*
>DN63074_c0_g1_i2lm.176628
RRFYLKSSNVLGLHQNENENMDAGFNSNQPRQKRMRTSFKHHQLRAMKSYFSINHNPDAKDLKQLAQKTGLTK
RVLQVWFQNARAKFRRNMLKEQSQQSSGQSQGSDTGLGLAPNSSSSETHDLDLDHVSTTSLTELSTNPSPTLP
DLTSPSLPDYHTNIQDENTSQTSLSDMFS*
>DN91806_c1_g2_i4lm.597628
MNGNATDVYLTDMDDVSPKKKRRISMCVGCGSQIHDQYILKVAPDLEWHAACLKCAECQCFLDETCTCFVKDGK
TYCKRDYIRLFGTKCTKCSLGFSKTDFVMKARTKIFHLECFRCVACSRQLVPGDEFALREDGLFCKADHEVVERS
SSASENGSNGLQLAAQPTSARESNSIRPHVHKQSGQKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEM
TALSPRVIRVWFQNKRCKDKKKSILMKQIQQQENGKDPSRAALQGINNMNGVPMVAASPVRHDGGLQANPVDV
QSYQPWQALSEYALQKEMEQAPFQQLGVSLQILVRR*
>DN95395_c0_g2_i1lm.832504
MVVNCAGCERPILDRFLLNVLDRAWHVKCVQCVDCNCFLQEKCFSREGKLYCKTDFVRRFGTKCAGCLSAISPH
ELVRRARNLVYHVTCFQCMICHKQLTTGEELYVVDDCKFICKDDYLNHRNNNAPPEKSPDSTIDHEIKSDGELDIK
ETESELGTVDAKENSHPTSESEPESKDGSGLTSAGNSEGSEQSTPGAKRRGPRTTIKAKQLETLKAAFAATPKPT
RHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGARRQFFNRTNPRRMRALRMGLRSPSELENNPDLL
NAGQFGYYPDAGVPGEYYSPQTFYDFFPGQQPPTSEHRNDFLPPMSQAGLDSGLVNPALVPLPPDNGQYLTMS
DIGPQTGITSCTPPSPAMDIHSNMDNFSRSPSRGNIPTHSFVGLGQPQDLSSADQMVW*
>DN11429_c0_g2_i1lm.25212
MNSFDFISSRCPITRKGSSGMNLALPLPTTPTYKSSPLQPIDANGNTQYGPTNTLPTCAGCDEPILDRFILKVLDRC
WHGKCLSCSDCHMQLSDKCFARDGKVFCREDFFNRFGTKCAGCEKGIPPTQIVRRAQDNVYHLECFACIICKRQ
LATGDEFYLMENNKLVCKTDYEMAKTKAGVDPDSNNPSKRPRTTITAKQLETLKNAYSNSPKPARHVREQLSNE
TGLDMRVVQVWFQNRRAKEKRLKKDANRQRWGQFFSGSGSVNKRSRGRSRHDRKEKIVNEEPKDFQDIPSDD
DDDDDTVDMDIDIELEMDGNDSFEIDSRLGKDSPCRTDVQTPVTLTPPGSIPPPPCSIPPPSHIQPPPLTSNGPTY
HENQPPFSKLPPQCGGGLAYTLSPPPQGMPGGHQHLPPHLAYGEGIPTPMAIGHGHPAFSNPMRNIIGGGEMP
PGSQNMAFHHEFGTPTASWLDEVPHAPY*
>DN81776_c0_g1_i4lm.328858
MTLERQSSCYTKDNSIYCKLDYIRQFGTKCSKCTRNIQPNDWVRRAKNHVYHLACFACYSCHRQLSTGEEFALQ
DSRVLCKTHYIEMLEKANTTASTKGENGQLSDDSKASRAKRVRTTFTEEQLQILQANFNLDSNPDGQDLERIAQL
TGLSKRVTQVWFQNSRARQKKNNNNMNNLKDSPSKILGRDKRRTTTT*
>DN87622_c0_g4_i2lm.442537
MKFNNNNNSSCLVEMKMTMDTGLSTTVNNNNNNNQITGCGLIKIEKEVCFGCSRSIEDRYLLKVMGNSWHEHCL
QCDVCKIPLSRSCYAKDAKLYCKSDYEKLFGTKCFVCQMSIPSNELVMRALGNVYHLHCFTCIMCCQRLQKGDE
FVVKDGHLYCKMDFEKELIGVMQNGDDDDQIPLSDDKEQDCPQSLVSLDGKDSPSSSSIQESGQIQSSGRSGKT
NADGKPCRKRIILTTQQRRIFKAAFEVSSKPCRKVRESLAAETGLSVRVVQVWFQNQRAKVKKLARRQQQEQEA
AAQAATIPGTGRGRNGRGQRSNDRGMKDRDSDCIDMMDFEQMKHPDALHPPLLSEMMRADGHYPPQIPMEG
QFPPPLLPPPMQESMLYPNSIAMSQSDFNEMFSRNGSPMVDNMSHDGLTFTMQGNHSPPISSCPPIGPPNSLG
QLNNPIDKLYDMQSSYFNAE*
>DN87006_c0_g1_i2lm.427067
IALAQQQQQQQQQQQQQQAQQAQQQQQQQDQQQNQNQTVTNGTSAQSPLTISTQHLTAQQALSIQNAVKG
VPGARIENSPTPGDEPSQLEELEAFAKMFKQRRIKLGFTQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCK
LKPLLQKWLEDADASLSNPQSILGTPPTPDPLGRRRKKRTSIETNVRVALEKAFLSNPKPTSEEIQMLAESLCMEK
EVVRVWFCNRRQKEKRINPPTSGGLLMHVDSLKPPTMVDSPMKSVDGSMKAQVTYQ*
>DN83679_c0_g1_i1lm.358416
MATTSAGYPIHDLSQIPSSVPNSNNSINMSMSLLSEGTPPHQISPTSYQVHHSQKMADFLSPHHQQQQALQNGH
PGLHSGLHHQWVGLSPHNDTSGWGTGMSPTTATLLNQDIKPSPFQQHHTREDIHPHLHSSQNSLYNRVASHQS
FHPSSAPWNTVSCSSSYMPSSNGSPLSYAHPGYTMNGMLANHGQAMHPHSPPLLQDIGHRGDMDHHPDDEEE
APTSDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLAKWLEEADKN
NGCPSTIDKIAAQGRKRKKRTSIEVTVKGVLESHFQKQPKPSAQDITNLAESLQLEKEVVRVWFCNRRQKEKRMT
PNILNGNGNDENSMSNLNNNLTPSPPYSSPIQATHPALQPPVAAPQHIQ*
>DN93613_c0_g1_i1lm.701698
ATVKLPNATSVVAAAQTHINNGGAGPAIGVPSANISTLGLSASNSNQLNSTLTQLNSSPHSLSSNLEQLTGLAGVT
VSMDGSVAHLSNGITLANNCIATVNSSHITSTNQQANNAAHTPTGANTGTISSTDANTVDGINLEEIKEFAKAFKM
RRLSLGLTQTQVGQALSQSDGPSYSQSAICRSASDYRFEKLDITPKSAQRIKPVLERWMTEAEARHKNGNNINIP
NLAGLEPIKKRKRRTSFTPAALEILNVHFEKNTHPSGSEMTEIASKLNYDREVVRVWFCNKRQALKNTIKKLKQSD
GCLTESSLADGGIS*
>DN90583_c0_g1_i1lm.541096
MPSEGCLFTIEQIELLRRLRNTGLTKEQIVSALDTMTRVERELNSLSESIEVHEITRAPVVSSAAVTATTSVSTNHH
QPEIPTQSLVSQFTHPQRSSSQSPASTSPSPISPLSGGETGVRVASSCSVTTNRGSNSSMMERLDTATCFRDIAD
TDPELKFLASLSLQTTKGLLKDFLQSHNLMQKDLAGDIGITQSLASVFLNGPNALRPSTMLTILKWYVWKKRLYEQ
EAATQSPERRTTNPVPQTTENVSLPRLQAPSSSTTVAQNSTHKRMLQFLKLKKFRSRINWDPSAVNILEHHFNTK

NRFPNAFERETIAANINGACHSKGRKVTAHHIQTWFNNRRNKRTGVSSLSQKTNRGYQHTQTHPQIPSTSTNQL
PSPGHVTSPSLHLQQQQQQQHQQQSLIQQASGF
>DN63639_c0_g1_i1lm.180328
MLTDADRYSDSLPTFGFTQEQVACVCEVLQNSGDIERLARFLWSLPACEHLHKNESVLKAKSVVAFHRGNFKEL
YKVLESNTFHPLNHPKLQELWLKAHYIEAEKIRGRPLGAVGKYRVRRKFPLPRTIWDGEETSYCFKEKSRTVLRE
WYSHNPYPSPREKRELSDGTGLTVTQVSNWFKNRRQRDRAAEAKERESSDAEGKGKGYTKNRKDHDDDNDV
DSVGNDKLDDMSDGDDDIDEDDCGLGVGQTSCPIMNNNNNNNNESGHTTVIPHTS
>DN23152_c0_g1_i1lm.52491
MFQPQLPQPAPVHPAPLSPAFGLPMINFTPSQIANVCETLEESGDIDRLARFLWSLPINPATCEALNKHESILRARA
TVAFHQGQFRELYSLLEHHRFTKESHAKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQK
THCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQVGNWFKNRRQRDRAAAAKNRLQQQQLLQPPDDL
EKQSCPGSCSSSSDRDNGHHHHHHHHRVSSPITCSTSLSGLSDSDISPSAIPRDLADASP*
>DN97213_c0_g1_i8lm.1016479
MGHPDVRMSGKEMLLLTPNILASICEILDQTEDTARMRNLFNALPDYPYVKYVLDQHEEILSARATVLFHDAKFKE
LYRILEMHHFSAKYHPKLQNLWLDAHYKEAEVVKGQPLCAVSKYRVRKKYPLPYTIWDGEAQTTRLFQKNTLSVL
KDFYQSDPYPNLNKIKDLARATRLSPIQIRTWFKNRRQRDRIAAAAAVSSSSPSSSTSSLHNHLNDEHNSSSPNS
EH*
>DN120045_c0_g1_i1lm.1152712
EETVYCFKEKSRQALKECYRRNRYPTPEEKRSLAKMTGLTLTQVSNWFKNRRQRDRAPKGDEPGPTSIRSNLPS
HGGSDDGFCRDNRLFDECRTPGPSATILGHGHTGHYAHTPLFPHYV
>DN63799_c0_g1_i1lm.181371
MAYPHAELGSRPIARLPVDIGGGHLAGAPMGSTLQWSPDFLARGTPLPRMAPFPTTLNQEQHISYGSPHKPPTIG
SSPYLQVMWPEGVSYQDLQALQYHPYAFIDPSQDPRALQMAGFSEQEIMGMMRKKSATRETTNALKAWLREHR
RNPYPTKSEKIMLAIITKMTLTQVSTWFANARRRLKKEQKELEITSDTRDLNDQYITSSTESNRAVSVSPNLKIMTH
QIYHETTDTGSRDMAPHHSLNVDSDDDNVSSPILSRSSSVCSSDATTVAVCSDQKVSKPKIWSVDDIINNKDSRG
DRLHRSNHLIYRMPKTETELRSGSLSPPEHIALAANHHNVGL*
>DN84998_c0_g1_i1lm.382419
MIMMHTQFGSPYAAQMLVSSTATTPCCDGTRTITGPDGTPRPTVCSCQYDGRGMLLRPPLAGFPIYSTAYPDHG
YPVSLTGADTAALYAPGTSPYDLKDGNPSLAQPTACYPYGAAVLAGPPPQYPFHDGYPGMDFPGVARRKNATR
ETTSALKAWLHEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKMTWSPRNRTDDENKDSPGEN
ADQRNDGDKDRMEITTDTELDDIKVSDDERDLEIDCDSRINVTSPDSLADDREFPKPSLKSPLRLTTHDNEHESRP
STPNKPGTTECRAAKSLKLDDAPSSKNGPLLIAAPSVARHSADMQSSTSRNQETQPAKPRIWSLADTATSTDEPK
PVRSVAYSLWTPQNADYGFHHSSFNGLSRWPTNGLLQPSPVALSSSLSAHTSPSSAAVAAAPGQPHLANTVPIIY
QRSPFLSVSLKDNIHSALDKGTNVVSP*
>DN64910_c0_g1_i1lm.187887
NGASMNEASGNGTMTSTPASFNCYETSADTDSKDKQPEVYRRPTSKTKLKNESDRDDDSKRHQKKRGIFPKSA
TNVMRAWLFQHLSHPYPSEEQKKQLAQETGLTILQVNNWFINARRRIVQPMIDQSNRAAAMAVRANTSSYGGES
QHIG
>DN79707_c0_g1_i1lm.301745
MFRGMCTGPGGPSATNHVSYSLQQYDDSLSAAAAAAAAHVGYGASTMNPSTYDPSVHVASSRSSLQSATSGL
SQYLEHSHNISRPQAHSSHHSNYLNSPTNYTDVHKRDKELIYSHPLFPLLALVFEKCELATCTPREPGSAGDVCS
SESFNEDIAVFSKQIQVEKPYFTANPDVDSLMVQAIQVLRFHLLELEKVHELCDNFCQRYINCLKGKMPLDLVVED
KDTPVGNTRSRSASCNSLDDAFGSSSAWEHQQDESPPLYHNLTAPPPQQSPMTPEDNVSTRSSSADMVPMVS
NNNNNNNTTTSNQNKQMPPTASTQSCSIETSNTNNCSETDTSTEGSLTSPKDKSIDDTEPEEDPKKYQKKRGIFP
KSATNILRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQSNRANMAGSNAYVPENQP
LATGYLMDQQAQMHVRHSGIHPMSGHEMNSTMGSSTASYHHAQLPQLHHPVHSQAMLIPGHPHAMMMAHHP
SVHQHHTMGLQATQSTAGFIHHDAMAHQVMDIHAS*
>DN68543_c1_g2_i5lm.209442
MMSAGGGGPPPQQGLLSGHNGQLGGGQRVSEGGAACDDHAPAANRKQEIGEILQQIMNITDQSLDEAQARYG
ERKHTLNCHRMKPALFSVLCEIKEKTGTFLNMRNSQEEEPPDPQLIRLDNMLIAEGVAGPEKAPGGGGPGQPDN
AIEHSDYRAKLQQIRQIYHTELEKYEQACNEFTTHVMNLLREQSRTRPISPKEIERMVSIIHKKFNAIQVQLKQSTCE
AVMILRSRFLDARRKRRNFNTNATQILNEYFYSHLSNPYPSEEVKEELARKCGLTVSQISNWFGNKRIRYKKNIDS
GSPTPTGYAQQHTPQPSGNTSDTQVSASNSAIVSVSSQSGQPDMFIPMSASGQVTNGYDDSAEVSSLRHVMNQ
SGSFSSNDSLASHSQGSLYGASSHYSQFSNPTVAGSAEQMASGLPTPSSVEQMYDRISSWRSSPAVTNPPANS
GMPTPPNY*
>DN93733_c0_g1_i1lm.707538
VEIDCADKKPFKCTICNVSYSQGSTLDIHMRSVLHQSRAAKADINLNAIQNNKPESAARAISPAPNQQDECCHVSK
SDSIEYSCTPTPTETSTPVPKHIHSPQISSEAPKSSHSVLSMVESSVNSSSLASSMNSSNDIQKIHLQLLQQQQQL
LFSQQQHNLFPDFAGFRSYLDPLALFQAQSMGTPTVYSCNHCSSIFTNQESLMQHQQLHCFLSDINVNNNSVSL
PITSISANTSCNAIQAHSSSKTTQKTKCVQRQILRDIGLDIVMHYAENSIGGLSETLKEAASKSVDPQLGIFKCNHCV
KSFSSIWILKTHEEEVHSQLISTDDVEKMANIIREEIDKSLSDDSQDTSLNNTTSSTGAGSATPLIDHMTTPLIPMTT
PLIPPPPPLAPIPQPSVMFAPPFPPPPSIRQPYETIYPNVLEHFEAQYRLQQQSQQQSHTKRPRTRITDEQLKILRA
YFDINNSPSEEQIIEMSEKSGLPQKVIKHWFRNTLFKERQRNKDSPYNFNNPPATVLTTTDHGYTATILTTTSDGS
KRTLTTASTASTPTSKIQSVKTESQSPKPSSHGSKLKAYSEKYHEHNAEIKPLSLESTLKVESGTMRDMNEDKLG

HVEKTMNFHKTSLLERDQIGGGGRHAVPMITSDNMIPSVSSSVFSQSPLFLPSSLVATSSIVAYPHFSLSTGAMGV
TSTPTLPTAASMSGFVYPTTSTSPLQTLASMSDSKPMESRRAGRTRFTDFQIKRLQEFFDTNAYPKDDDLDRLSK
ELNLNPRVIVVWFQNARQKARKSFENLPPLEGDSGGEGRYIRTPGLNYQCKKCMLVFQRFIDLIKHQRQQCFRE
DDTKCSAATAEPSTPTTGHTRRSSSSYMSGGESNSSNSSSKPPDTSPTESSADVMTYPCNKCSMVFPKLDLWQ
EHQRLHLLNSFNPFMPSIMDNPYGLVYMPYGDLNMNSMATASASGLPPAMTALSNAASLVATSQASSSQSCNP
TTPTSNSPSPHHSACTPQPASNAKPSKDAVVISQMASASEPEPPRDRRLRTTITPEQLEILYEKYQENSNPSRRM
LDIISREVGLKKRVVQVWFQNTRARQRKGQIRAVGPIYCHRRCPYCRAMFKAKTALDAHIGARHWDKVAAADRS
GCLAGKDDQAPNEPVNGDYDDMNVDDDMAHDDMAHDDSLEMERIESENISFHSDSHEENGSANLMDKGHSSS
ADHHPDKASASFISDRSDEKCCQDEEVVDDVDMFDDCDDKLVIDHSDNDGRNKNCSESRDDLDCSSNKSYGSE
HGHAPSDTRTSTPVSRGLSSPVSSGDDMYSNITNFDNSRNSSQMVSPSGTSSPRCSTSPSNLKGKHNDGKRFR
TQMSNTQLKVLKHCFTEYRTPTMQECELLGSCIGLQKRVVQVWFQNARAKEKKSKLSYAKTFGVLNPPFETPKQ
ECKLCGVKYSIQMTVRDHIFTEMHINKLKQSVGGQTPSEKSYIDSLTGMSQVLSGSQDAHQVPFQSPSTSSSTSS
FFTPTQISPISNMQTSHLSPLTPYLPPPPLRSPQSSHVSPPTLPKIKQEVLDEPKISEHSKLDSSGSSNASQSSSNK
SIPYDKSSPSAYSSGAPSIANFVPSASGFPFLTSAMPSYYMQYPFLQPVINPVDFQLPFAHASMYGRLMGESAAP
HTSTSPPSDTSSVSDEKASGETAKCEKHFTNTASQRYMCRKCLMVFEEKSVAVAHQQLVCYCGQIVIPDQTLLTL
SSSESSRMNGEYSFDKWVHHNGEMGILSGSKVHERESNSDCGKHKDLKSNEGHESNNIASSYITSSNEQRAAK
NGTDY*
>DN87293_c2_g2_i1lm.434168
RAHTSVTETDIIMAVYGSWSFWNEKFWLPHNVTWADLRNDETHTYPQADDLILPIGIAFILLLVRFCFERVIAIPFAN
LLGLKIPARHTPEDNTILEKVYRSVSKQPDAKRIEGLAKQLDWSERQVQRWFRRRRNSDRPSVLKKFKESCWRF
AFYLSVFIYGVWTLHDAEWASDTTMCWRGYPMLNISDKLYWYYMIELAFYWSLTISQFFDVKRRDFWQMFVHHI
ATIMLLSYSWVTNFVRIGSLTLIVHDCADFFLEAAKMFNYAKWENMCNVIFLFFAIAFFATRTVIFPFWIIYSAMFESI
SLLAFFPAYYLFSVLLLTLQALHLFWGYLIVAMVLNMIRSKEINVQDTRSDSEEETTNGSEQDITNADGDHAVTNG
RM*
>DN112933_c0_g1_i1lm.1132014
SSGPGPANAAAGMLQTLYDYFWWERLWLPVNLTWADLEDRDGRVYAKASDLYITLPLALLFLIVRYFFELYVATP
LAALLNIKEKTRLRAPPNATLEHFYLTSGKQPKQVEVELLSRQSGLSGRQVERWFRRRRNQDRPSLLKKFREAS
WRFTFYLI
>DN95119_c0_g1_i4lm.810987
MFGEHVLGLTQGSVSDLLNRPKQWGRLSVKGREPYIRMMLWASDELNVESLVHMQKEKRAGKRRHMFITDLDE
LNYEKRDDPEFIPPSMVTTLTKRSRFEPPAQPSNILTLGTNKIRRINLTTREKNRLLEAYNNEPYPSPQTIEGLAQE
LGIRQTTVNNWFHNHRSRSKRTKTAQLAQLPNRKNRKSKSSPQHCDDRDIGSPHSEDDLIDDSMCFEMESPPTL
VPRDSILNSVTYISESIDLSVGRERSVSCSTICSDPPGGSTVCSDPPQSDPSPINSHPVETNPSPREWETSDPSGP
KDNSEITQNIDVSSEVKNTITTSDERAKVPSREPSPLQEPETPPRPNGTPAPPQNLKDSVAIKCEYSEQSCKEPVP
DNSPDVIPVFPSPVKLEADPPSTSGPSPAQPPPTNEYVNLWPSNNRTESTDHVMSPPTPAQKELSNSAEKTALS
SAARLQARVSHTDSGDDWEF*
>DN95135_c0_g1_i2lm.811330
METETLISPPVVSTVTEMSREGADGACTDSDLEGSVSDDLVIDMETNVSTGDAGSKPHRSGSDCNDPPTDEDDS
TSKPPEKIYLAKCETDSGSGGEVRVPEPRIIFELLAKEHRKRRLRTVFSDTQINTLEQAFRFSPYINTEVRVKLSEIT
GLSDNAIQIWFQNRRARQRKTGDPNTTLDSPPQQNNEPCAEPPLQPPVILVPVKLGTVAKQPPNSQPEHASLAM
SPANQTQADSAKVKVPASSSDESNVLKTEASSVSVEESPESTQTGGEEPVTVTATSEELKPLEQNRDPDIEMCP
LCPEKFTCRRLFAGHLEQAHHCLMAIADCLCHQCSENTDPQPPQQDAKQRLSENSPSQHSDADVIGRGDPDYT
PPAPCPTSTGVKRKKRRNSRKSAKTSKRSVAVKQRLKQSKHTGDDVIGSDDVIDRAMTSLVGDSEPSHGGGGD
DNASYEDIRHDVITAASHHVDTSEVVFPSTLTASNSSTPKSAAAGVGCILGTPTSEQAALLARGSPALVLPNNTSQ
SGYVLLPVSFNNGSMTCLAPGGLPGGGQVTIPLSMPTMTSSAVPSQQKIGERWVSQEGTIGGGLTRRSSRSSQ
SVEDRGRNSSPLPTRSDSPSSDDSPHFSDLSQFLNGVNLHPLKPIPSPTPPKVDEKKHMCEICGGLFAMPSGLK
RHMLNHTDTRNYQCKICNKRFKLSHHLDNHLKMHLGLKEACPQCPAKFTQSYALKQHVMRVHHCSRQAKKLRE
VRVNVEKHQCHICNTVLANSTELSTHVSQHVNRKHICDLCGKDFAVGYQLAKHRRQHEKRPFMCIACGDLFGTM
ESLTFHQGFCTDIMERDLKNGEVESGK*
>DN91240_c0_g1_i11lm.568665
MPFFKFQTPNGSIIMEFDENVMVSHVRTAISQVIQMPVTAICLPGTMPVKKLLDSTVMRDIDRKVVLIVESRTQPLP
NDVQHGMKYSPKPQNSIMAEQPGTSLAQLIHHSTPTAGVAVTTAGATVAGAMVQIAPVNRPHPYLQVQSQTTAG
GQQQIPTFIQMPTAATQIVYDPRHIAEAVRMELSRMKITQEVFARAALGKSQGLLSQILKSEGNGQSMQNHLATLV
DFLRKTENERVVLYGQAGCEIVQQRREIARQNAIINGEPYPTGKRRRRTTLSDHAKRVMTTYFMTTSRYLSADAR
MAFSQQLGLSEETVRIFFKNLRANTKNMMKQAETESCSSHYKEELNGDSPEERVGASPLTDSGMPSPSTSYNG
VDLSRSAHGTPPKTIKTENYD*
>DN85043_c0_g1_i1lm.383698
RSRINWDPSAVNILEHHFNTKNRFPNAFERETIAANINGACQ
>DN88028_c0_g1_i1lm.455036
SSGWNTNLTPSHLKKAKLMFFYVRYPNSATLRQHFPDVKFTRHNNAQLIKWFSNFREYFYIQI
>DN74939_c0_g1_i1lm.252753
KRGVLPKQATNIMRSWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQPM
>DN45414_c0_g1_i1lm.105527
KRRRGNLPKDAVNVLRLWLYEHRFNAYPTEQEKVELSEKASLSVLQVCNWFINARRRILPDM

>DN97994_c0_g1_i1|m.1088530
RETKRPSKEMQMTIAQQLGLELSTVSNFFMNARRRSLDKW

	predicted class	PFAM domains
DN63819_c0_g1_i1|m.182148	ANTP	
DN14752_c0_g2_i1|m.34886	ANTP	
DN115451_c0_g1_i1|m.1141018	ANTP	Homez
DN2325_c0_g1_i1|m.5825	ANTP	
DN124767_c0_g1_i1|m.1167403	ANTP	
DN84275_c0_g2_i1|m.368677	ANTP	APC_CDC26, UPF0258
DN3733_c0_g1_i1|m.8775	ANTP	
DN76496_c0_g1_i2|m.266175	ANTP	HTH_23, HTH_29
DN44680_c0_g3_i1|m.102754	ANTP	Homez
DN81549_c2_g2_i6|m.325476	ANTP	Homez
DN90457_c0_g2_i1|m.535164	ANTP	BRF1, Dicty_REP, DUF3373, FimP, Gag_spuma, Hid1, Mucin, Plug, Pou, RED_N, Spt20, Strabismus, TFIIA
DN105712_c0_g1_i1|m.1111542	ANTP	Homez
DN63160_c0_g2_i2|m.177407	ANTP	DDE_Tnp_4, HA
DN131037_c0_g1_i1|m.1186210	ANTP	Homez
DN93010_c0_g1_i1|m.661361	ANTP	
DN101095_c0_g1_i1|m.1098549	ANTP	
DN50727_c0_g1_i1|m.123834	ANTP	
DN92869_c0_g1_i2|m.652722	ANTP	
DN75963_c0_g1_i1|m.260766	ANTP	Homez
DN135951_c0_g1_i1|m.1198937	ANTP	Homez
DN90720_c0_g1_i1|m.547254	ANTP	
DN86948_c1_g1_i1|m.424178	ANTP	Terminase_5
DN69228_c0_g1_i2|m.213920	ANTP	
DN94680_c0_g1_i1|m.774060	ANTP	DUF2638, QLQ
DN41925_c0_g1_i1|m.92873	ANTP	Death
DN122354_c0_g1_i1|m.1160697	ANTP	
DN94983_c0_g2_i4|m.798955	ANTP	Homez
DN84186_c0_g2_i1|m.367331	PRD	OAR
DN48278_c0_g1_i1|m.115742	PRD	hDGE_amylase, HA
DN86933_c0_g1_i1|m.424448	PRD	OAR
DN55770_c0_g1_i2|m.144367	PRD	
DN77439_c0_g1_i2|m.275737	PRD	HTH_23, HTH_29, HTH_Tpn_IS1, PAX
DN86691_c1_g2_i5|m.418460	PRD	
DN49217_c0_g1_i1|m.118984	PRD	CDC45, OAR, Sporozoite_P67
DN88387_c0_g1_i3|m.464016	PRD	HA
DN70414_c0_g3_i2|m.220060	PRD	
DN67189_c0_g1_i1|m.197813	PRD	Hid1
DN14020_c0_g1_i1|m.33082	PRD	
DN12917_c0_g1_i1|m.30558	PRD	Apt1, CDC27, DUF2317, DUF4407, HemX, NARP1, NPR3, TPR_MLP1_2, Vfa1
DN86983_c0_g1_i1|m.424471	PRD	OAR
DN122491_c0_g1_i1|m.1161007	PRD	IBD
DN102158_c0_g1_i1|m.1101472	PRD	
DN122218_c0_g1_i1|m.1160059	PRD	MRP-L20, QLQ
DN83632_c0_g1_i2|m.357974	PRD	
DN63074_c0_g1_i2|m.176628	LIM	Homez, Herpes_capsid
DN91806_c1_g2_i4|m.597628	LIM	LIM, Saccharop_dh
DN95395_c0_g2_i1|m.832504	LIM	LIM, Orbi_VP3
DN11429_c0_g2_i1|m.25212	LIM	LIM, PHF5, zf-dskA_traR
DN81776_c0_g1_i4|m.328858	LIM	LIM, HypA

DN87622_c0_g4_i2|m.442537	LIM	DUF164, LIM, Prok-RING_1
DN87006_c0_g1_i2|m.427067	POU	BRF1, Dicty_REP, DUF3373, FimP, Gag_Spuma, Hid1, Mucin, Plug, Pou, Red_N, Spt20, Strabismus, TFIIA
DN83679_c0_g1_i1|m.358416	POU	Hth_31, Pou
DN93613_c0_g1_i1|m.701698	POU	Hth_3, Pou
DN90583_c0_g1_i1|m.541096	PRD+HNF	DUF3685, HNF-1_N, HTH_19, HTH_3, PCMT
DN63639_c0_g1_i1|m.180328	SINE	Homez
DN23152_c0_g1_i1|m.52491	SINE	CENP-B_N, Homez, HTH_3, TniQ
DN97213_c0_g1_i8|m.1016479	SINE	Homez, HTH_26
DN120045_c0_g1_i1|m.1152712	SINE	CENP-B_N, HTH_3, Homez
DN63799_c0_g1_i1|m.181371	TALE	HA
DN84998_c0_g1_i1|m.382419	TALE	HA
DN64910_c0_g1_i1|m.187887	TALE	
DN79707_c0_g1_i1|m.301745	TALE	
DN68543_c1_g2_i5|m.209442	TALE	DUF1319, HTH_19, HTH_26, HTH_3, PBC, Sec15
DN74939_c0_g1_i1|m.252753	TALE	
DN45414_c0_g1_i1|m.105527	TALE	Homez
DN97994_c0_g1_i1|m.1088530	CUT	
DN93733_c0_g1_i1|m.707538	ZF	HA, MerR_1, zf-C2H2, zf-C2H2_4, zf-C2H2_6, zf-C2H2_jaz, zf-met
DN87293_c2_g2_i1|m.434168	CERS	TRAM_LAG1_CLN8, HTH_23, ESSS
DN112933_c0_g1_i1|m.1132014	Other+CERS	Cers+unassigned
DN88028_c0_g1_i1|m.455036	PROS	HPD
DN95119_c0_g1_i4|m.810987	Other+CUT	CUT, HTH_3, HTH_Crp_2
DN95135_c0_g1_i2|m.811330	Other+PRD+ANTP	FYVE, zf-C2H2, zf-C2H2_4, zf-H2C2_2, zf-C2H2_jaz
DN91240_c0_g1_i11|m.568665	Other+POU+ANTP	CUT
DN85043_c0_g1_i1|m.383698	POU+ANTP+HNF	

6- *Ascoparia sp.*

>Locus_7846.0_Transcript_2/0_Confidence_1_Length_2609|FPKM|3.84|BlastHit|swissprot|Q0J9X2|ROC2_ORYSJ|m.5405
MPAGFMIPARHVPPLMARHTKAGYTTSSGLSLAQPNMAEGQFDMSQTADSEIAKIREEEFESKSGSDNIEGASGEDQDGERR
PRKKRYHRHTQHQIQEMEMFFKECPHPDDKQRKELSRELGLEPLQVKFWFQNKRTQMKTQHERHENTQLRSENEKLRSENM
RYREALNNASCPNCGGPAALGEMSFDEQQLRMENARLREEIDRISGIAAKYVGKPLLSFGPSPLSSLPRSNLDLAVGSYGVQPNI
GPDIYGSSSGGEIGNRSLVGPTEGEKPMVVELAVAAMEELVRMAQLGEPLWSSRPEDGTDILNEEEYIRTFPRGIGPRPYGLKA
EASRETAVVIMNAINMVETLMDVNQWAGMFPGIVSRPVTVDVFSTGVAGNYNGALQVMHAEFQVPSPLVPTRENYFVRYC
KQHSDTIWAVVDVSLDSLRGNSSSVIRCRRRPSGCLIQEMPNSYSKVTWVEHVEADDRAVHHIYRQLVNSGMAFGAKRWIAT
LQRQCERLASVLASNIPSRDLGVIPSPEGRKSILKLAERMVTSFCAGVSASTAHTWTTLSGSGAEDVRVMTRKSVDDPGRPPGII
LSAATSLWLPVPPKKSV*
>Locus_20750.0_Transcript_1/0_Confidence_0_Length_1123|FPKM|71.10|BlastHit|swissprot|Q9YH71|VNT1B_XENLA|m.24127
MVSNNFNVEFLARSCFEEEKLREAERSVPTPPTTPESIQADEYISDCSDEKSSSRKKSRTAFSPSQIYELEKRFRSQKYLAASERGEF
AEKLGLSDTQLKTWFQNRRMKFKRQQQDAENAHLVALYSHCPPTAPQLLHTAPPFYPGTYQTHPMMYNAPYHPMPVYPST
PIATQGLYAPTSPLLPPTGLYPAPPMDNLIPIRPTNLFS*
>Locus_7293.0_Transcript_2/0_Confidence_1_Length_807|FPKM|21.43|m.4837

STQSTGNGTVFAKAAINTCRSTLQVALKPEGAFKQYIPTIMKFLALSREERLIRYRETERLLIDERLEKRLHSKANRSPRPRRNRTLI
TSTHRNIMLQHFLTRGHYLTDFDRKKLSTELHIPELTVRAFFQNLRARVRAKQKSVPKTSLMDAYQSDAR*
>Locus_22263.0_Transcript_1/0_Confidence_0_Length_1118|FPKM|21.33|BlastHit|swissprot|Q00401|HMH
2_DUGTI|m.25649
MLNSTSGTRSTISKVARKRKRRLLFSKRQTFELERRFQTQQYLSAPEREQLANVINLTPTQVKIWFQNHRYKCKRARFDKTIELSQ
NRKYPGLLMDSKNHVSNLMTFRQQLPYVLYPPSWCYWSRFTLSPPFANTDSTLESPTSPACDD*
>Locus_11746.0_Transcript_2/0_Confidence_2_Length_1579|FPKM|8.48|BlastHit|swissprot|Q9ER75|IRX6_
MOUSE|m.7565
STQSTGSSKFFTSILLCVCVNEFTIMTILESVVNNPPVVTQPTSLESRTDSIFPAPSVANMMPLSLPLLPTLHSVPLPTMAPLPPILA
VPLADRTPFPIDRIPIPQMLLGKDLQMERWLQVARQLPYSYPYPMGADFAFDPGLGLRRKNATRETTSALKAWLFEHRKNPYP
TKGEKIYLAMMTKMTLTQVSTWFANARRRLKKENKMTWTPRNRYGDELLKESHAVDANADSDIEHNESEDEDISLSSPDTSE
KLPPASQSSTPPNDKLTSSNDKKKQIWSIAETVASSDTKSNKDEKESSPETSYMTFLEGVYQNSLVSNQNSNSIPLSFPPPPNTYF
NPFLLGSQTVPRFQFPTPPSDAVPTHT*
>Locus_16759.0_Transcript_3/0_Confidence_3_Length_724|FPKM|20.03|BlastHit|swissprot|Q8BG99|PKNX2
_MOUSE|m.12224
QRRVRDQLGVEAQKKGAKRGILPKHATSIMRSWLFQHIVHPYPTEDEKRQIAGQTGLSLLQVNNWFINARRRILQPMLEASNP
DGVSKAKKQKTAGRPAPQRFWPDNLVPISRISGNDDNDCDGDPDSTSQLQSSRPSSHESTDQKEVEIKQQLVQ*
>Locus_4546.0_Transcript_1/0_Confidence_0_Length_637|FPKM|32.75|BlastHit|swissprot|P54366|GSC_DR
OME|m.3021
MTSDQFSIDSLLSSPKSNSCADIPSPMTKAMPEFQPPITLSQPLPLWYPGVPIYPFAQNNIFSPHPTVLHHGNGVFGGVPIPPM
PYPEYLMQSMLLRGRKRKRRHRTIFSEDQLEELEKTFAKTHYPDVVLREQLAIRVDLKEERVEVWF*
>Locus_16350.0_Transcript_2/0_Confidence_5_Length_1221|FPKM|5.20|BlastHit|swissprot|Q9U637|PITX_B
RABE|m.11677
NAEYGDTMASCDTPLLPFSVRNSIATDSKSICDFRKQIKQEVKPVDKGKAVTSTNSPPVGEAEEAKTAEEEGSEEKKKRTRRQRT
HFTSQQLQELEAMFIRNRYPDMSTREEISAWTSLSEPRVRVWFKNRRAKWRKKERGQMADYKTFGGFMSPYDDPTLYHGYT
YNSWAAKVANPLGGKPTFPWSFNPPTPAQGLYLGTPPSSNSMPSSIVPNVNVTPGSTACPYNMASGTQWYDMPSSFGPLRL
KTKQPSPAFTYSPMRQTGMPACQFAAADTAVVG*
>Locus_21083.0_Transcript_1/0_Confidence_0_Length_752|FPKM|25.63|BlastHit|swissprot|P18488|EMS_D
ROME|m.24483
LVLTMMESFIYRSPIQIPPIFHPYYRNICDLPLHPPGAYPNCSPACCSRFSEARFNILPYNQLSYFDKSFLHLQSMKPRRVRTAFTS
QQQEALEKAFIKDQYIIGDERRRLADKLKLSETQIKIWYQNRRTKSKRVIPRRELKEVFPKGKTTNSQMPGSQSEVTSHQSPQVS
TCDDVTR*
>Locus_16822.0_Transcript_2/0_Confidence_2_Length_1384|FPKM|22.14|BlastHit|swissprot|P49640|EVX1_
HUMAN|m.12301
MMAHPPAFIHHPIALMQRPDFPNCRPFIPPPIPGFFHHPPVCHCLPCRRNTSPQLQHTKPISPAANQQVDEQSAEVRLPKSDF
RKRPHSYSTPDSEDESKKALYEKDDPQSRRYRTAFTRDQLMSLEKEFLRENYVSRPKRCELAASLGLAETTIKVWFQNRRMKDK
RQRQALMWPFLDPHYYSYLLSRGLIPPVASIPPPIPLPGIPLPAPSTAAGGSGHFNPASPAIPHQIIRLPWPPTDPGCQPEINLHQ
LPPVFGHHCPSGLPPSNCPACSGLQDELNSRRDNSSSDSKTP*
>Locus_13559.0_Transcript_1/0_Confidence_2_Length_1206|FPKM|43.98|BlastHit|swissprot|O42173|PV1_
XENLA|m.9001
MFVETTASQEGLYRFSPSQVSQLEKFFAEKKYLTAAERSKLATSLRLSDTQLKTWFQQTHEAQTAVTGCGEQPHDVFGQPVLY
NDAAVTSLSYAVTSDTAAILTSDDVQSSTYTPHKFTASAPTSSVCLNTRTDALFSYTSP*
>Locus_22779.0_Transcript_1/0_Confidence_0_Length_837|FPKM|18.81|BlastHit|swissprot|P70397|DLX6_
MOUSE|m.26130
TQSTGKIRKPRTIYSSFQLQILSKRFQRTQYLALPERAELAASLGLTQTQVKIWFQNKRSKFKKVSKQDGEHDTPGSNPSEEPDHT
PPTEEYQSPGWKQTVPYMTDDRPDVKHELPYLTAPEKPEVNYVTAKAEVPYLPPPISLQQYNPWYMHSMPNQETVSSQPAII
T*
>Locus_16759.0_Transcript_2/0_Confidence_4_Length_819|FPKM|15.21|BlastHit|swissprot|Q96KN3|PKNX2
_HUMAN|m.12222
QRRVRGAASNTAHLTIYAAERCFESTDYFRPHSIITNWSHSILTNLDPNESLQPRCCQRRPAAAATTDRKNSKRGILPKHATSVM
RSWLFQHIVHPYPTEDEKRQIAGQTGLTLLQVNNWFINARRRILQPMLESSSPEVLPRSKKAKTMSRPSQQRFWPDSITSINSNI
SNTDSDVDVDDRPCSAASSDDLRSQSEAAKPATSVPELVV*
>Locus_27059.0_Transcript_1/0_Confidence_0_Length_1518|FPKM|3.87|BlastHit|swissprot|P55813|HBX4_E
CHGR|m.28679
MIEVFDLGSMEDKGFVEQKRRLKTPSQVEALENIYAEHKYPTESMKAKLSRELGLSEKQVQRWFRHRRLKDKKGKKEEPDSNE
ELDAGSGSNLQQQQHLVRTELGRSEKKRLLGTSDYPSAVLAAELTDHDMLKRNQRVMEEDPHHAGTVSTSQDTSSLQSKSSYE
MDALRNGNYVEMKGSNSKKSKRQVMHMGMEHTYPPQHLAVATEDQYIAKGLKQAGGSNSKQAKNWNRFKAKDSENHRF

VNNHDYPSAVLAAELTDRELLKGNHDIVEDPYVSLSQERSSLQSGSSYEMEARRLPSKSLNSFEIEARSKRREKEGLHMVVEPIY
WQQAVEHQAISAVKAQLGRLFREDGPILGVEFDPLPPDAFGYR*

7 – Childia submaculatum

>c20992_g1_i4|m.38161
MDQLNLTEAETKLANHIVANFQKADIIKDNSDMFDTMCQDKSFNCRNDNSLDVFEETCQDGKQVKDTNSSPEMENLAQSLTI
DETISQNENFSCIFCGQNFCDTVALYQHKRYDCLQNPEVLIHRANNVNQNNPAFHMTTELFCGDTKRNFERKKRTVYTEKQLE
TLREFYTKQPNPTSDDIERIASLIGLTKRNVQVWFQNKRARDKKSPHERNEDISLFNNTFSDAESSTTEIASSTPIPNDLTVNIPSS
HTTTPLTSKTPKTQKYISPLKLSETPNLTNSSALYTYQWLSMLSFYQALSTQNLLNGGGLNPIVLPQVLQLLKQRMAQQNGINIQ
PVTPMTGSPSTMEIYEEKDPEEPLDLSLPTSQIKVDTSLSDAKPFTPSFINPACVPMDFNKVKPESSPESETSEVSHVSTGEATFAC
NQCSKVFQKQSSLIRHVYEHSGRRPYRCNQCDKAFKHKHHLTEHSRLHSGERPFQCSNCGKRFSHSGSYSQHLSQKRCL*
>c21527_g2_i2|m.41051
SMAPPMNRRCNQTVVEDEIATMTSNSQQHTGSRRNGGLTVDNLADNIMYRLQGKLSDVFCNGGAPLYSACSSGSPQEGDL
WRRFKVNMREASITNEDLLDPELVQAQTFKKAKPESGWMKMLVVSLYRASLLPFIIPWWTSELLPYQNIVLCISWLFALITLSHL
YLFPSCDSPSTTSPKTQSFCNYHMELASPALYFFILTFLYSYISNSRKVNSDDLRVGLRHESEVLKQVDAFNSQIYSNHSRSLNVQA
SHQITTSGRGPVLSGLSHSNLNGRNFSRRRSDRHRVYRDGYIVQNEESRNSWSSQEEHARFPETLERSTRNVRRTRSSPLNVIES
RIEDSIEGTWTQEVPLSPPGSLSEAESDPVSGTELADPSLSPRGSMDGLPVQSIHVVPSQLPNVGNTVQCNEASPEPETESGIVV
TSLPSTSVSSVPIFNLTVPTSQNEILWSTSGHSVTLEPAGPCQMLPVESYSQAQSSRQSDLPRSLDGVSKEKRCVQVLRRRKNVG
RKLKVTDCGPRGLSNNS
>c11585_g1_i1|m.5268
AEYMGIIKNNRHLKLLISSTLYISHGSIVLPMDSILNNIDFRDLTSHNSNASVIHATSTAPSITHLPLQIQPVNSAVAFSSSAFPSISIN
RPLPPSFPTSSPVQSTFNQVPEQINSPVFTSAPSTCLNTVHKSTNPQTAGTNVTTVSLIDPLSHLLPSGAPVSIAAPFQLLNPPTSLI
IPPPTLPLQSCKKERSDAPKIGQISSESEKSDDKIKHERAESLEEENENSATNEDEADKKSKRQRRQRTHFTSQQLQELETLFSRNR
YPDMATREEISAWTNLPEAKVRIWFKNRRAKWRKKERNQLQEYKNAAAFGMNFMPSAAAAAAAYDPHDFYSSQLTAGYNV
YNNWSKMAAAPLTPGKTAACAFPWTFAPGNPNPAASMVPNVNIGASLNPSPPCPYGATGANPQYIYGTRDCPTGVGFGQF
RLKHSPTQSMNAAAMQLGGYATPFRTSASGLPSCQYTNMERPL*
>c11899_g1_i2|m.5600
QKSYVCSEENAIGTCFSESDAESEREESSDLIGQGMLNESSSNGYMGFPHSGGCKLEGPEPRRMLIKDPNGQVKELIFPRGLDL
DRPKRERTSFTPEQLYRLEKEFRSNQYMVGRDRGELAGILGLTETQVKVWFQNRRTKYKRERQREHETKSSGVDNLNGCGNM
LKILEQATGQYYTTPRLYTPKGPGLDSALLIPSSLNQQLQSSMDLSVQNSLLKQISPIYSPYYWVGHS*
>c26298_g1_i1|m.47693
LLYQDSSLASKQTPEVTPVTGPGRGMFVQQMPMQGERGVPNNTAKQNSILSKMINSGSDEEAEEVCDDGEDLDGEGGSFED
GKKHRRNRTTFTTFQLHELERAFEKSHYPDVYSREELALRVNLPEVRVQVWFQNRRAKWRRQEKLEQSQFKLNTDFPLASLRS
HANGKIPGNFEFFP
>c16644_g1_i5|m.18054
AEYMGTPTINSNRYEDGNTANFHLALDPSNLSGQAFEQTPTLNPSNASYRQSTVSSAPAPTNTPPYHTTSHLYAQQCAQPPHQ
NNIISGTHLSAHIPTPNTPVSITCPPNELCSQEKNALYSHPLYPLLELVFEKCELATCSPRDPGSAPGSDVCSSQSFNEDIHIFSKQI
QEAACDKNFYSDAETDTLMLLAIQVLRLHLLELEKVHELCDNFCSRYISCLKNKMPLDLVVEERGESRSGTGRARSPLGNSPGVG
AGAGGLDGETNSADEEYTRAQQQDNQVRNQISPLTPPMYYSNQQSTGCLPVSMIHPHGNLPPLQFPPISGATANSTPNSNSV
HDINSASPSQQTSSPCEIDTSTDSKKENQLDHLIKPKMENISPDSSESMENRSPNPPSHHKQIGKDNHDGVPVTVPITSPSEGEE
DPKKQQKKRGIFPKAATNIMRNWLFQHLTHPYPSEEQKKHLAQETGLTILQVNNWFINARRRIVQPMIDQSNRAGMPNAYV
AATDPSQMGAAFMLDGQQAPYLRHHPMMHAAAAVGLAHGNPVEMAAAFGGYNGAGANGTSNNGGLGSNGTSGAGGL
GVGMGVHPFQHPSQAAMLIPGHPHAAMMMHHPHHAASLSHQHYSHPQSLMDPMTAHVMDSIHT*
>c27478_g1_i1|m.47899
QLQILETTFMINPYPDINQRDILAEQLGLTESRVQVWFQNRRAKTRKFGKRFQNPEANNLNLQKSHCVLPPLTGNTSCFRVGQ
MPPLPELTTTGHINQIGHIPAFGFCPPQSTSMSAYAPIN
>c10150_g1_i1|m.4165
MDKMFENIYKSSQPVRRLLIRDSHGIEKELIFPLALDIDRPKRDRTCFTPEQMSLLEQEFTKCHYVVGKDRSLLANKLNLSEAQVK
VWFQNRRTKFRKGDEKKHSNAENPGNSEEIIESKDDTSLQERPINVVSSIKKPIENQYTDLKTPVDSFHFPMLYNVYYPQLLNAS
KFNYRYRSEDSHGVNNMTYGLVAINATHANKFPPMMSLDSLHQTMLNFG*
>c17743_g1_i2|m.22610
TQSTWAIMNNNNERKQTLIQSLNNRNVKNTSTHLWPSSGENKQQSGHVTSDSTNERSGELGDQKSMNFETAFWLVKTINDE
LLAESANYSVDILKLLLTAQQKVIAKLVIDRQQVRSQINDNSSSENTCPRCECLKHKKTPECSTGNGCNGQNPDLLQTFSIKDDK
DYDIEMTPNDSSSTVEVQVPCDVIGNTGVKDSEVPNTRPVPAIGETNEELSGVPHDPDSNTRLGFPPFDGPLSPSFHQGSNNV
EELNNFAQEFKEHRIKLGYTQSDVGHSLGSLYGSDFSQTTISRFEAFNLSRRNMLKLMPVLKHWLKMATKETDKSKFVPFPGDT

GTGVRMAAEVHMVARKRKRRSIIDMEAKEYLEVMFQGSARPSNELLSQTAELLNMETHTLKVWFSNRRQRYKRGYCIDQTTT
RAEEKEMSQYCNNNGPVELKNDNIQDEDVQNVSTNEIGTCHTTPVNEPLHVQAILSRQPMTEEINNKLISELNSKNLHNETKT
*
>c13477_g1_i2|m.8396
MPGFFMTDILGLDSGGNEQTHPIKQQEITPHQEVEYGEEMLEYDEEEDRTESSTPGLDEEEGRINTKQGDEMDASNYDEMSH
EYDIERNRKKRKRRVLFTKAQTFELERRFRQQRYLSAPERESLAGMIDLTAGQVKIWFQNHRYKCKKSNKERLHTTFQSSPLDM
LSRHNSGTTSKMGPMFPAHPFYTSNGSMQSFNAGNHASLAALWFQRQQNYLAVAAASKLSGNTNPLPMEISNEGSPLHTRK
GSQLAQMISTNSGCIRAGNTPWYSLSLPGTPGNPWQASGNHSNNVLNSNHSSNSLLMKLAGVNPNNAISIAAQNQFLDQLN
PMKYNWLQNIYKTHLALLNSSTNTQMSTSSGSSNQSSPIKPRCEEDNFSPEAENIEMSPPEF
>c2408_g2_i1|m.670
EKETIVGINGKVTLKRKIEGRCNIQDLRSTHKVELNEFTPLSPSSNHVPVYSSPRHSFPLTENLTATHLRKAKLMFFYQRYPNSTTL
RQYFSDVKFNRHNNAQLIKWFSNFREYFYIQIEKFSRDTGIKIKMTERDVSYPMGETGLPKMEMGKSAGKELTPKDLITIDENSE
LFRSLNNHYNKVSSYKPPSCFLHTINATIELFVGAILDGRDVQDSGWKKCIYKIINKHDCEIPDHFKNSYMLTKGSSQVDNQLTPV
VLNEN*
>c15737_g1_i1|m.14698
EYMGSLQLPVSSITKPPAATAMLQELTPTTNLITPNDQITCTPFRENVSSLIKSEINGAGNNIPISKPVPMGETDLDEENDEDNM
YYADQSDSDSQSEPGNGQGMQLTGENQNANDVLNANALLMQQAALYNQLVLQNPAAVSALLSGIAGNPQLMASNQMLA
LAAASGSLGNSIQNDITTDPNQKPGRTKISEEQLRILRQNFDITNSPTEDQYLIISQQTGLPMKVVKHWFRNTLFKERQRDKDSP
YNFNNPPNNPSNNAFNMQKAFLNMQNLMTATAQTQNLSNFANILPGVSVQSDMSPYVSNNSSKRASRTRFNDMQIRALQ
DFFDQNPYPRDDDLENLSRSLNLSPRVVVVWFQNQRQKARRSLEIPNEGQPSQNNPSSINVSTEPSQVQQSVVMLDQSPAPP
LVTTPLLVPPAPLVAPNLNLDATKAVQDVVNFYAGHNLSSLEKVPTITCQEDKIPVIEPQPAPNFEIGNEIAPGSPGNNNVSQSIK
EEPELTINEDEFCYTCNFCTSLLPNEDELAAHQRKHMDEILVYNKNASDINNIQKLLKSGTNQQVTGVNTGGNPMNVPNGTGD
VANQDYFRRMRTTISPEQLEVLYEKYGKNSNPSRKEMENIANSVGLSRRVVQVWYQNTRARQRRGQSRVIGNQIYHCKCPYC
DSICKSKSHLSQHIANCSAARAKITNQPIAPKTETSQQNNSLIPNITMGERININVNHTLSVGNDTTSMFNPMLETRYPTQYPGQ
SNSQKVPNEPSSDEEEAAEVSFYSDASNDEKSQSFSPHGSSQSTPVAITTNPTFTLPNSTFDFLSMIQNAAANGTPNSQIPLLNS
QIPCSSQGAISPDDKKRHRTHISAVQIKTLKSIFNDFKTPSMSECETVGRELGLSRRVIQVWFQNARSKEKKGKSATSPGVEETNS
STNFSNDQLNSSRNCCRFCNNITLPNTSALRDHVFSPTHLAAVRKQTRIKMLSNPKVENDLLYSVNVVSSAASQMDSLASTRFP
TQIAFPSNSGAIMDLSTKAS*
>c12433_g1_i1|m.6367
ILRVAPDLEWHAQCLKCIQCGVNLDEKQTCYLKDGKPYCKADYFRLFSVRCNKCDGIFEKRERVMRAAGDSVYHVECFSCTLCT
KKLLPGDEYGINDGCLYCKEDMLDLMSDKCPSEGIKLEGQSPVATDNNKQNNNSESYLSPSNADVGDSCHSDKVVFSPSDAKS
PSGPDNTGYRMDKSRSHSKESPNGQTGGSGNGGPKPTRIRTVLNEKQLHTLRTCYNANPRPDALMREQLVEMTALSPRVIRV
WFQNKRCKDKKRSIAIKHMHQMQAAAEKNIAGLSGVAMVASSPVRHESSNLDPSTGAAGYPPHPHSHNPYSYRLPPMTGSH
ELDTSYLMQFDNGPESVLSDALASPVSDCSLSLSSE*
>c2827_g1_i1|m.802
NTTPNQSEEPFPMTSLAANTTAKQSEKPFPMTSLAANTTPNQSEEPFPMTSLAANTTAKQSEKPFPMTSLAAYTTANQSEEPIP
IMPFSKLSPSTVPHTFANGIFLPTFAPTFLPQLGSLDRPLTMWLTPPCGIGISTGQQPDMANTGNLGLGVTSWQSAIIRANVSA
SKRRRRHRTIFTEHQLNTLENTFANNRYPGISTREALGKRLSLSEERIEVWFKNRRAKWRKGNKESEKRANNVKCLVR*
>c13944_g1_i1|m.9381
MSASPLKKKQKLVSNFSISSILETCPRAEDVQQKEVNLTENSTNDDEAPRFRTIPNSILFDIGDKMLCENRTLTDGLLMTVPSIKAS
SCIPQFSPKNEEKNHVNDSEEKEENINGKTFDGDDKKECDEEGKREFDEEDDDYENEEDRKKKSRTVFSRNQVGRLETMFDLK
RYLSSNERATLAGDLSLSEQQIKIWFQNRRNKLKRQIKCGAALASPVNAIAPGTVPPPNVNAAVLNTINAMPNSLTNPLLNCVQ
NGALLGSPQTPTSLKITGHHHNNYFSHFLHQTLPSQATPLNLTPSTNPVTANFPN
>c16438_g1_i1|m.17362
RVHGVCETLEESGDFDRLSRFLWSLPPQLLEVCTKNELLLKARATVYFHNGQFRDLYMLLENNRFKKEYHPKLQAMWLEAHYQ
EAERLRGRPLGPVDKYRVRKKYPLPRTIWDGEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQMTGLTPTQVGNWFKNRR
QRDRAAAAKNRGVALNNFSSGDEASPTKLNSLDIDINDEDDDDVEIDNESDSEFESETNGEKEDGMDNSRRIIQTIQLEKNCSQ
ENIRDNEKDCDQNLQPQLNVNHISPKSSQSTHNSPSNQIKSATICSSNVTPPKVHCNQISPHILENKEHIFRKELLPLYGLSPHNM
ANLPMGIPFSAHAAMLQYPGLIAPNTGQPRLPNLTPAAAAHLMMQAQLAAAAAAGRVPAHLFLHPAN
>c13907_g1_i3|m.9294
RRVHGAGACIQNPAASSAAASQFKYSTTAAFQSAAVARQRTFPHPAAGNFFGNPFDDSFFRAETLSAANHQAHIAHKQLVDY
KSHDYFSNVLSDMQNNAARCSQLDLLDQITPSVQFPTSMGTHPATDHQTFHPHHLGFHPGFSTGSMPGLPATANFQISHPE
MEADPRELEAFAEKFKQRRIKLGETQADVGKSLSHLKIPGVGSLSQSTICRFESLTLSHNNMMALKPVLQTWLEEAEAKAKTKM
AEKEGTSLTGLDKKRKRTSIAAPEKRSLEAYFTVQPRPSGEKISAIAEKLELKKNVVRVWFCNQRQKQKRLKFAAFR*
>c17223_g1_i1|m.20502
QSTWGPQNDIITPADYTVKAQYPSVYFTNDYSTNYNHQSDFNTNFYYSNHTLPLHSAIQHAHTQNLPPTIPPSSSPDFSTPNSR
NTQSINSPFSNPPLSPYKLTSSLYPSTSPYGIPYPTSALHNPHPYLFPQTQPNTYHHYYSHPSGLHSYSAPHLAGLANSKTGLSNVT

PASALNINVNCSVEANIETSDHTSTAGTSGHSSESRPNLGNQSIEETRQLFAERTTDNPGQNEDKQKIVSNFGECSPVGSPLDVT
KPGTETVKECTSPDAYSPTNNLLHPTYHPKYNPIPTSIPSLTPAQGTHSSLTHLPGQNPNADPLPYSWMKIKRSQPNLHWNKSH
AGNGMGMSLYSNSLGGPPYTLSNPSQSRGGRTNFTNKQLTELEKEFHYNRYLTRARRIEIANSLSLNETQVKIWFQNRRMKQK
KLVKEGKIV*
>c28994_g1_i1|m.48193
RIAACKKKEDEQMVPKPKQVKKCRLVFTDIQRRTLRQIFHETQRPSKDVQTQIARELSLEVSTVSNFFMNARRRSGDRWKGDD
PKDTKRSSGNESRSVSREMTSSVHHAGAINSLSHSS
>c8723_g2_i1|m.3200
AEYMGDGEETIYCFKEKSRTKLKDSYRKNKYPSPDEKRNLSKNSGLTLTQVSNWFKNRRQRDRTPHKMPPSPKCEIERNIKQEN
ILENGQKDEKRSKIMSKLETMEGAKRQCFGELGLVGVPSLYHGSTTVPYLSGI
>c5411_g1_i1|m.1593
NAEYMGALKPDNHYCTELRILSEESILEGLQGDNVSSFAKEETIKYKPILPKLHQDKERCSRKEIIPITSTKLKHLKSFVMFRGENLN
WLRHQVFSRIARRICAGISQTQLCIMINTKYKDNLTQVQISQLESFEATIERTKLLLSRVDRWYLSEREYTNERQFIVKLTNEYHSK
CANDFHPKSSRCRLHHNVHKLLSNLALMNPTPDMNQLIQISQATQINVKTLKKWFSRRKRRDFKMETMHVNYSTEWNGLLH
AFQRDVSMRKSLARTTRRHRGPNKNTV
>c16553_g1_i2|m.17764
MPQRGKGHSGVNQLGGMYVNGRPLPESTRQQIIELAQKGARPCDISRVLQVSNGCVSKILGRFYETGSILPKAIGGSKPRVATR
EVVRRIAALKRDSPSVFAWEIRDRLVQEGCCNQETVPSVSSINRVLRSLQTEKMIKSPSGQVPEVMQATSAVAAASMYDKLRFI
NPAHWPPAAGSATHPHGYPHAPWYPSHPSTTHLHPANGFQAAAAMLSASISVPSHTDAFGQSALSMYPTGPICTTEGLPEN
GVVDFRTQQTRNNGAIEVSEEKQKVEDEMKGEISPQLKREDIPERSSTGGGSPHSDMSQEVREEEGDGEEDGKKRSQRNRTA
FSQEQLDGLEKEFEKSHYPDGCTRERLSQTTGVAESKIQIWFSNRRAKHRREEKTKSTNSNSLLSGINMNNNNLPQNPKMISET
NSSGNMETNFNIQSFPNPTNLYQPPNTTYFPYNHANSLALSNILQVNQSPVSTCDPTQQISANQISAGHPAFPFTIPLNQEPTA
SPTHVSAPITLPLSPQNNPISQMTVAAAALHYSPTLSNPHHPYYSHQPAYFPSAAHMPPCLTAPASVQLSHPQGAIDTHRTGIS
QTICSTAPPPAPSTVNQNPFWPGHVTMPPSSVYFSQKF*
>c3630_g1_i1|m.1055
RRVHGPDDISNNVSNQTTTPTHMYSTHKQTKPCSKDYNSLEVNRSKEKESTHTQHIPATTPIKPKRHRTRFTPGQLNELERCFN
KTHYPDIFMREELAMRIALTESRVQVWFQNRRAKWKKRKKCNPGMFRNTGGGFLNPGGTNPMTMTPFSQFTSSFAADPFT
SYPFASDPRGWVNQGNPHFPHAFNFPPSFPHRANNIQNFAQFGLDQNVINSTLSHGQARVDPPSGDTSPNFPGIPNSFTTA
MLTQGNATVLEAIPNVPSNEFEDHWKSSGLAALR
>c1431_g1_i2|m.402
PTYYWNPSEAAISPVTSVDHSEIPPPLTRDDINSMLEPVTLIAPSSVTSNTNVLTPLTSPTLAKRRHVRFSSKQLDLLQIQFNICPYT
TERDRRQLSLTTGLDTQQVKIWFQNHRLIVRRRNNGLLPGHNHRRPSNHNQASSAIKQQVDAMAV*
>c26849_g1_i1|m.47814
IYSIQHHQALQQLWLNSHYAEAEKARNKPLGAVGKYRVRRKFPLPRTVWDGEETSYCFKEKSRNVLREYYGHNPYPSPREKRDL
AENTELSVTQESNWFKNRRQRD
>c12068_g1_i2|m.5819
IPCFPSVGRCTRCQKEITDRWYLSTDEGLWHEACLSCCTCNRSLVDDLKCYINQGDPYCNKHYHEKVCKKTFSCRRCAIKLSHTD
YVMKARGTVYHVSCFTCFICGRPFTAGDHIGYTQDDILCLSHVQNANSSDSQHGSSTIANIPNPTLQKIDSSCGDTTPNYEGPP
NCSINIFAQNLPNAFQGCNLNDLITNMPAGFPPHPAVSHHGFSPIDYIKSMVVPGLHASTIKGRSKKKKTTEEVLMYNMSGGA
MLENDLSNAMYCIDTGNGSVRSKRMRTSFKHHQLQAMKQYFNANHNPDAKDLKSLAQKTGLTKRVLQVWFQNARAKYRR
NIIREHKGRKVGRGTEEEDRCGADAEDGGEYQEGKHEDDTCGAEESSDESMMSYSPGVSSTESSNHIDQSSELE*
>c9054_g3_i1|m.3371
EIQSPIFNSNVSSNKQSKNRASKSGLNEKSPNNSSNLTKDSGEDLDTRDIAQRITAELKRYSIPQAVFAEKVLGRSQGTLSDLLRN
PKPWGKLKSGRETFRKMEKWLDEPEFQRMSSLRVAACKKKEIDSVPQCGATGNMTTSVVNVVNQEQKLGTGGTSNHPKKN
RLVFTEMQRKTLFAIFGGNRRPSKDMQVSIARQLGLEHTTVSNFFMNARRRCGDRWKDIENHEMETEIGFKSKCKDNSLKSK
QNTSNGLTIKT*
>c19370_g1_i2|m.29815
MDFSHFGAKCAGCEKGIAPTEVIRRANDHVYHLGCFKCLICNKLLETGQEFFLMENQQLVCKNDYEMAKSKEFEELNGGNKRP
RTTITAKQLDVLKGAYKNSSKPARHVREQLSKDTGLDMRVVQVWFQNRRAKEKRLKKDAGRRLGLSGPHGNVSAAAAQFYSS
TLNQTSGTPTKCRKSKKNSQNQNIIPSMEMPPGLGAHFVGEHEGLSFSNNHGLMHTQALLPYPGHMPQHMFLRDHFPFVPH
GTHNLKPYIPISSGLLSPAESIHMLPPPPPNTHSVAPYPVPYSLPEAVTPPPFPPQEPRTPISLISTPMSHAIAGGLYTPITVPPENP
GKFLENLSGMAPPRTASCTLTPPFSGSDHCNIQSSLHTALTAGRTPGYATVGMLPTLNPISSSTLTTPPIKFEPSFHNMGTPNGQ
NIVPPTAINHQLENLLLW*
>c25162_g1_i1|m.47453
AEYMGIHVARKESKVTRDVVRRTSSASQQCVENQPFPQFSPNHSMISDDLNSVRRYRTAFSKEQQEMLEGEFRKENYVSRPRR
CELASVLGLPETTIKVWFQNRRMKEKRQKQTFAWPPFFDPFYYSYLVSRGFMRDHVLNPGNISGFPQIYPNIFPGQTNECLTFP
GTNNQFTSLLSHHSLQNNTAMQANFSSSPTKALLGGAKSNYLTHSTK

>c15018_g1_i6|m.12359
MALSPKTMASGFSVTDILNPIEESYKRSAMDNALNFATMNYHRQQAVNSGAVPTGAAGAISSPYMQMAAGAGMMGGAAT
CSYWGANVSAASAITDFGNGGFNSTGLVAHPGSAYTDHFGRQASAGSMSNWGYPASTDPRLSFSRLMHTSGGMSMGMAG
MGMGGFGLGDPAKMLTSHRRKRRVLFTQAQVYELERRFKQQKYLSAPEREALAQAINLSPTQVKIWFQNHRYKCKRASKERS
GVVVSTGNNGESGVSGDNTQDMSGKDSSLTIKSEGGKDPDGDSLVDPNESEPVSPQEGDKMQQPMALDYTPTGPPLIGPPG
LEEIPVSFRSCPGIYPNTITDPNLVLFPNGISTRGW*
>c12701_g1_i1|m.6865
GRTCCLPIEGVVSTQSSYPLRVWLEEHYLNPYPTKSEKLMLSIITKMSLTQVSTWFANARRRVKKERKEELEMAEIYRWQFLMP
PNRTTLLTLESNPSLAPSITIPNDYLINSSNLHAKNSYELKKNYMHFKSDDKNTCRKRKLIKQNGVNIKRPSTTNNSGVKDAVHNF
DRKLKEGKIWSLVETATK*
>c14406_g1_i1|m.10632
MISPSNNFKENSQNQKLSPNSAKFSISSLLTHPEDSWKSAGTISYTGHLHVKLTNPYSESANMDNSLEVESLKYQSEDLGKVNKK
RQTSKKSCEKVEENSCKKSESVFMTRKVGRNPRVPFSTTQVSVLEQKYQSKQYLSSLDVSELSSLLVLNETRVKIWFQNRRARAR
KEQEARHTSAGPGPTPTPNTVVNNQLYFTQSNQAAHAWAPALLPNLSGTCQMQIPSSTHHALISQSSVATSSIKPHPNLLSAN
NQLISQINMAALLAKQHLFFPMPINDLPNNMAPPFTTEPIKNCCQSDVPMYSA
>c18358_g1_i9|m.25266
MESNINGIVNENRPNSNETQENALGGATDQLKYTVEQVELLRRLKASGLTAQQIIHGLDSLEKVECSLNNNGSSLETLASFSPW
FLLQQLGAFKNSSQDYLKSLALQQIVALATATSANSGQFGTSPLGLPIQLTNNKPVLTESGISRNQTNNISNKIAELFPYIANGNQ
PVPNATVDPEYQIGTNQSHIKKEPLQTATPAQTILNPQYSISDDNPQLLKLQTMPVSESKEEIKGFMCSHGIKQHQVADRTGVS
QSYISQFLHQGVWMKDSTRTNIFKWYLEEKRKREANNDPETPIVQETPYVITPVEPPKITPHNTAVKKRDRFVWKESSVLLLERF
YAINPYPDDVQREEIARSCNQLNEMSGSYNTDDKKVTAVKVYNWFANRRKEIKRRTAHCMQDGQTNKEINYKSARHIPSIDIS
NYVNELELLDQHQMSVVTAGSLCGSGSLSGDSPVKGDFQEEMEIMDNSIANLIHRKQPEVFRNEQTSPNVSPNNISLCADGDS
ENLVLDLS*
>c14405_g1_i1|m.10631
ITPTSEKFISHSPIKIVSFVSLSKLHQFTIAFLTREIYNAMTIGSSWSSTVPPLLTSTSPPSTLSSAWNHISFPKSPPLPALCQGCGHCI
TDRYILQITEDQCYHESCLRCTQCSCLLDQSCFIRDSRPYCKEHYHSLWGNGKCAGCHGNIPANELVMRVASGQNTINFHVNCF
TCSVCTKLLKTGDKFVLNGDMLICKDDYKHHLNDSFKEDDSLLESKHLNDSSDINRDQSDLLLDEESKTKVEFDEEQDIAQASDS
ENFNETYAESSKGDSDFEADNEDSKNCSSQQKPCRKRVILTTQQRRAFKAAFDLSTKPCRKVREQLARDTGLSVRVVQVWFQN
ERAKVKKLAKRQQQQHQNHLEQGLTSPHMLEDAQFGMFIPNEGYMHAAAGLDPRFPHPGNPALSIMQPDALFHRTLPGEV
FNISTNQQTLNEQDYLPQVSADYPQIIEIQSKIKDEAFVPLKPIAPTSLLHPIDELYSMQCSYFTPS*
>c17704_g1_i1|m.22456
NAEYMGNNHHTSSLNSTGGATANITPSSHNSNGHFNGQHQPFKCSECGLFIKEQSYCRVLDRILHASCLKCAACKIKLDNKCFS
RDGFLFCQNDFYKEAWSRCSGCNKIVGDEELVQRIADYIFHIWCVRCCACGQHLQTGDTVFIQHGMPCAVYCDQDYRTRVIA
NAPDQTSKVSTSGGQWISYSMNNPNHNIPNSLPSSTTIASQAPVSSHNFPPRPSPYTRLPAPATVPNTIPNGHPQPVTYLPSISN
GGSGGGQLSHSSQSAFQPAAAHNSQSAFSGVRPNLHMKQGALSNPNVKHDNSICSEPAVGVGSTLVQGSNPIITPSNTTHIPP
ILGTSGSSAFSPCTPVPSTVQYQSHTQHYTAPQHSAPYIHNEESEKLSDNLSISSETDNQSSPGSMSGHSSLLGPNPGPGPNMH
PVPTQPAPTPRKSGIRPDQSEQLVKFYQTNQRPNKQERHSIAQVVGLEERAVMIWFQNKRSKGKRKELALAAANAAAANAG
NINNQTVASTQQQQIQPINNYTTITHNFTADSGLPSCSADESSVASTVQSNPGTFAGGHPASSDTCVSLDSIGQFSAVISTAPIA
TGDDCPMESNFTMPTTVPFTTQSQPMYTNAAYEQNPLPFVSQNYPTTS*
>c18455_g3_i5|m.25643
MTPVVNPGFHSFPVYLPTSMENVSNNNWTFSNPNTLNQTPQLVNYEPESGQMYPGFDTLRNPGTNTICKPLVTDSMQTMP
VFPANNSWINTHPSPNSKISDLSPQQYWTHSNGYTQNTPRYYSNAPMSTGQGQFNAYFTTEFDSIPDKQNDMFMNNTRQF
GYMTTPDNFPTVVQIPPTMANYEPNQIPNKEGNFPVSNQLQNQPLSVALPQPNFCSIGSKFGSFQPAPAVQPLPDTFPMSHQ
SSMASANQSDWAPNLQPNPSLLHHTSSSQGASIQSEPCISPGPFSPEPNMLNSPLDSSPNSGSVCMEKSGCSSVGLGNTSGLT
STSGGGSQMSWMTRNVSRKKRRPYTKNQTLELEKEFLYNTYITRERRLEIAKNLNLSDRQVKIWFQNRRMKSKKQNMGISSG
QNYIPSIHDGTGYMC*
>c12107_g1_i1|m.5859
AEYMGRTDGNSSKLTVPQAITPQPHANMLSENQNSITRAVTMATNSFKHSRSVPSTASSLFPTVFSFPDNVTSGSVTATNSLSS
ATHSLAGFSPANHGLASLSSANHNLATSAMYWNPALMALQGHRELPSSKQEAKIKGEGVDENGITEKGHHHEPSENECYSEG
EETLRTHGKGTRDGGDNYNEEEMRGRKKRRNRTTFTQVQLEEMERIFLKTHYPDVYVREQLAVKCDLTEARVQVWFQNRRA
KWRKKEKHHPLRSTNTTGTSVGQTPGVVGSVSNNISGTTGAPAVINGTFNADAAMFRPNAVADTIAQHITATWPYAQIHQQ
YSQSLFPATQGYNALTGVAPNYTNPATTFLPNYMIAAQQAYNLPLQSNQNNIHKDINGFDKKLYYDRGLQPLTI*
>c22584_g1_i1|m.46816
NTGAEAILGLRFTNHVNANILYESVGVGINGDKKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALGMSESQVKVWFQNR
RTKWRKKHAAEMASTRFKLNGGGNNCSMGGLDNTQIEDSLVSSNDPFHMLMGQDITVNH
>c13479_g1_i2|m.8400

MFNTSAAMNMGGYTNGMNAGGMSMTGGMQNGGNTGNMAAMGMTGGSMAPLDTSGMAAAAGWYAAANQYTPG
GDPRFATTAVNDPRFAFSNFMRSTAISSMGFPLGFPGADKFFASAAGRRKRRVLFTQSQINELEKRFRQQRYLTAAERENLAH
GINLSPTQVKIWFQNHRYKTKKSRGSQMAANSTVVNVITPTANQMNKESQILESKPFVNNMNDSYKHMSETQIGLNNTYTN
QQNTDPFYVVMDMGIENNDKTHVNMSGLQVTSANNPINFTKQTTFTGPNTQIPNS*
>c13098_g1_i4|m.7582
MKSSFLITDILGQNFSNKTTNSTAIDQPSAESNITEEQVIDRKNSNSHQLLTIQINQEKFRQSTCSFPNTMDSPDSGIDCNSFVSS
ASPANVVMPTATHLVPITRDTLSTSPEAQDSFPSNHDSQNRPFSPGEEASMLEYQKKQRTAFTPDQIYQLERRFSSQKYLSANE
RAEFASKLNLSDTQIKTWFQNRRMKRKRQQQDAENAQLQALYSSVTFFPPTNPHPMLPVRPAELIRPNHAPPNLVICNNPAM
AFTFPNGPNNPQNLAFPQTLHPTSLRLNAYQTPGNFNPGSQPLASIQNPASIKIMPTQTSGLGMFKFV*
>c10242_g1_i1|m.4238
QLATNHLTFPQNPPINTAYIPPHALNLPPTSLPFSTYPTALPSLQNVISGQAANLCQNVYPIYRPKRVRTAFTADQLLKLEDIFQK
NSYLVGHERRKISQQLGLSETQVKVWFQNRRTKQRRGPSVMEEKQSD
>c9603_g1_i1|m.3748
KCELRQRRIRQSVFAREVLGRSQGTLSWLLNHPREWHDLKWGHDSYMRLAHWLNLPDSDKNKLSGTAKTKSTPKGLNGKSQ
ERLVHSVKEMLKAEYSKNMRPSFERITQLSKELNLSTSTVRNFFTNARQRNKMKMLN*
>c12131_g1_i2|m.5901
MEPNSNKFIYSPLLLNSPPFLPVAMEALPSNAQERVFSSLKPYSQEMSLAQILPLHPSVPKPLTQNQPKSTQIPVVPYSSAMHLA
ASARAASMLRAQQEFLLASHQRPNRHPYSPYCPCHNCIQINPHLQGPTPSNGQIQLGQVPIAISQQTNHTPQCYPAPLPNGTSI
PHPMAFISPIFSQDNVDTATLAAAAAGDWAVLGRRKNATRETTAALKTWLQDHIKNPYPTKGEKVMLAVVTKMSLTQVSTW
FANARRRLKKEGKMTWIPRNTYEDEESGEDHELDSELEIENDEHVRTCEKEPDQEGMRTGKITNEGDMIANEIDCIKPLHIEFM
SEVDEGYNEDVTTPKSAINTTLIKANCLDKAGTACVSHKDQNKPALSSPSNCRHLEISASSASHIEKSLVCKKSKFWSIVDTIQINE
PSQTPTSKNFQDL*
>c9479_g1_i1|m.3668
ERDCEETPTNDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLTKWLEEAENNNGC
PSALDKIAAQGRKRKKRTSIEVTIKNAL
>c14657_g1_i1|m.11290
MNPAHLSPNSRNDSDENLTKHFYPEFGQQHPAIPSFGYSTLANLTQTSGSTVHHHGGAPNHFTGAAAAAFNSYVNPGGVLSS
QITAHHAFNPSPASHNPYGQFAADPALLSNAYRTHNMAAASAFPQYIPYPSQLPSPPSLHNAMGGYARAGVFPNTTMRPHL
GYGSNLGEHLLLNERPANIQSDGLGSDKEDMDSGKLSGLGAPMPNGPMSGNTGNNAMGKGNKKLRKPRTIYSSAQIQQLTR
RFQKTQYLALPERADLAASLGLTQTQVKIWFQNKRSKVKKMIKGPNGMPGTGPYNDCNSPEDNDEEEDEDKECQVKAESGN
HGNKDLSWGEGLPNHSAGNVHVTGTMGVNGPSDMSHRYDHMSIKSELNAEGDCPPSSIPMNPMHPNIHNLPPLPHPQPPL
PTEMMYSGGGLMSTPLSIANSHVMNPWFDGKSPVIQANPSALHLQHLNETPPV*
>c21996_g1_i1|m.43490
SQSFVPSILQEIQCRCGVSDFEARNMLIYSQQNVIKHIHRKLEEIISLQTRTKQAVVDENSLTKVVKIPVADLISLENNLSNAANKI
QNPALNVNPLIVLQDDSTVYDPAGMDMNGGYAAVEPATNQQVVIVAPSITTPVMKGPIKFRNVQIETACPVPQDNAQLDLD
NESRGAGLNSQEVETIANNFVTEEFYDDCDKLEELERFAKDFKTRRTHLGFTQADVGMSLGKMHGSDFSQTTISRFEALNMSF
RNMVSLKPFLAQWIQDAERHKNIIEQDITQLSAHDQSRIELFRPNGSKAMSHRKRKRRKVMDAHTRAALEQYFRDKPNPFPD
ETSAMADALNQDRETVRVWFSNRRQKERRDGQGGAIEVTPIMPRSHHPSAQQYTNHSHISSMPGSPQPDFGANISDLSNQS
QDNFPTNEQPDFSLLDSHEQLLPPNIL*
>c9931_g1_i1|m.3974
VKMNLYPPEMSYLKQAHVAAAAAHPYMNHSMLAHHHQMDLLPHPMAMNYPGATHHHLQAAAGVPGRKQRRERTTFTR
NQLEILESLFAKTRYPDIFMREEVANRINLPESRVQVWFKNRRAKCRQQQQQGEKNTTNKVSTNQNVSQHEKRDTTEKDEKS
PSMEPQKPVNQSAERKVSIGGNGTHDTSTIIKKTVSPNFQSPVSLKPPQISPSLSESSLLQKSEPEGLSVAAPLMPQEQTSFLPSV
SSSFVNIWSPAGGAGNATYTDPLNAAAAAALQRPAITGYSQPVTGQPFYPSSYYSFDPTASYAPYFSNAAYQASAGSIAASRYH
NYSSTNETSPATANNLFNSNEIKSELRNTLNDLPSWKSY*
>c12744_g1_i1|m.6950
MNKTYSSNFQQPSIKDFEDCISALDNNDTSQDNKKFKISTAAYEKSTKIQSFHSDFSQPYAGHRNNYDSELMSLLLPFNETNAM
KYQVPVYNTQFNMANTDQYNVAYSPIQLNNFNREGGKNDFQKNFHQNIPDNMVNLPVHMPFGSPEMTYIGETCNVQMGY
PNAYGEINPSMQQFIATNQYLQSHYSPWDESLMGAGDGRWFYNVELDNKKERATFTRSQVWCLEKEFNKHNYLTGERRETL
ADALGLSEHQVKVWFQNRRMKCKRAKPKS
>c11087_g1_i2|m.4846
AVVSTQYSFAYSHVGTKDRFQCPIQPTTFNTSDNYPEAQNLGHNQGLQNMADMTTSETEVSIKPRVSETNSETKTNEEQLESA
NIASHSSRERRQRTQFTSQQLRELEGLFMRNRYPDMNTRETLSKYIGLSETKVRIWFKNRRAKWRKKERHAPFPNQIVDPWKT
NSLLPSATTHFPYHGYETLRHSLIPPPSTFQSDSKPFGHTSGLKHSHPFATYPQSSWNFLSSPLQHQFSTCPPNPFSPHSYSDQVT
SADHALSAFPSTIFTHLPATSNVDQSSHSQTPANSSTNQNSSYFNLDSAHNLCFPSAGR*
>c6188_g1_i1|m.1904

AEYMGTTTQLLPFWAVLVSILTHSSNFPATSHLATTPYPPFPMIQNDSTSEMLEKPVDMKKKRGSLPKDAILIMRKWLELHKYH
AYPTEHEKLELARETNLTPLQVCNWFINARRRILPEMLKNEGNDPSLFTISRKPRANNNNQVSNIKDRKHHIPWQPWKNEPSK
PETSPTVPPYLPKTINIPEHRIPIRFPGITPNIPLLSPYILSKFQPLPSNYNVPVFPNTCNMSQLLWNLGQINSNAIQTKFIKN*
>c16112_g1_i2|m.16143
MMNNLDNNNIHLGMSQTPGQPSTPNSAPGSHMDFKLLHPMSMYPHTPHHTSPGASAPTGHLMQPLPNNIQSQNVKQSPI
GYHHTSTSNVLEMPNTQNTLPDNNQMLGPPSGHSNPSSGTGLTVGPNTPVPPTNSNSGHEPVGNNTPLTTGGSSLGVTDHQ
QFETDKKLIFNHPLFPLLQQLFDKCEQATNSAECSDSRSFDLDIENYIHHHEREKKPFYSENSDSNDLMLRAIQVLRIHLLELEKVS
ELCKDFCQRYIACLKNKMHSDTLLGGGLGPGMSSSGCSLSPQYPGTDNSPTNGMQQHLPTSQQTSLSLPHASSGASHHLGAA
GVTSLPNPGGNPLHPQSAAAAALYHNLHPSMFPPSGPHSFPPVSSAPYPPPGCGPPASGGLNGSSLSAAAAAAHHSAAMLVG
ANHINHMALQQHVNSGNHPLLAHSAMGTGNGSHGMPELNPDGTLKKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKR
QIALQTNLTLLQVNNWFINARRRILQPMLEASNPDSLPKSKKNKSSRPNAANRFWPDNLAQGQIQTCKMESESPTNPGTSPN
SESKLSKEQPHLDSNLDDDDDDLDESDCLPTVNGTNLNSILDSSCDFPDAKLMSVPSPSTLNPRLLKQSSPAAANHATLNDLLPR
HTPSSAIYPPNFSSHLTSVSVTPQSQMLFAESNTSA*
>c18848_g1_i3|m.27379
MSKVKTNSFLMSSILCPEKRSQNKPKADLCLNSNEFLDEHTLRLHREKSKSKIEEPGVENSKEKLESDEDDGEEAEEVGERKKNRT
AFTPSQIFELEKRFRMQRYLSASERGELARNLKLTDTQIKTWFQNRRMKHKRQQVDAENLQLAALYGPLFHQSMTSLLPQAFPI
SPNPNGNPLSIFGPIPNITRTAEEQSNNFPKIPVTNALPQSTPFINYSPSNPQYPGHPLIPVSLISTPVETPFGQLSPLHFNQIQTFL
KGSIQTPGVSTEKGQSDLNPISNHLLTQQLNPLTVGALPISQINQRFHANHDSALIYPMPFANSFQNLSPNTK*

8 – *Convolutriloba macropyga*

>Cmac.rna.tri.29560.1|m.28276
MHTMPHKGHSGVNQLGGVYVNGRPLPESTRSKIVELAHSGARPCDISRLLQVSNGCVSKILGRYYETGTIKPKAIGGSKPRVAT
SEVVSKIADYKTECASIFAWEIRDRLLQDNVCKEDTVPSVSSINRVLRNLSSGDNKNGGHSGATSPGSQSAQPTSVNHSNGNQE
DADSTSSYKIMAHYGSGRPSGNQGHSPLSNAPWGTAGGPGNLKGVPGSASTGMTSPSAFYSYNISPGSQYIAIAAAAAHVAV
QGGFPGSHDKGNDLRVKDDLCDPDRNGVLAMAASEEARMQLKRKLQRNRTSFTPEQIEALEKQFEVTHYPDVFARERLASKIS
LPEARIQVWFSNRRAKWRREEKLRSQKQLQHNPAAHPHSHPHSGNPHSHPHALALAHSTTPPGAQQLVDCVGAQMGMAT
PGAVGGGIGANKLPYGGATHSLGFSAAAGFNHQFQVLANNFYSTLPAAAAAAAYATPFYGGYHCNPPGTGTAGMGFEHPAA
SATRNMYGHPSSAAFIGNLANGLCPSSVSPGSAAMVSNLNPASAVNNIAGSNVNTSDLAIQDITGGGGSRSPPCSSPTTATASF
YDSYFAMTSASSASVAANRCGLSATGAAIGATGEVGSSGTLQQPHPPTADLIAINQHMRSL*
>Cmac.rna.tri.7851.1|m.13583
MIGDSVPQNFSISSFTSSGVAQTTTPSVNELKLKLDEIMDCSLEQSSSISNDCDEHVSSINSVETNKCHFCDEILPDRVSLYQHKRY
NCSKNPEVLLHQQAAYTSSISCQGFAQNQIDDDFSGKLEPGCVGASPGNQPITSVKFAQTSANQNAATPNQQNVERKKRTVY
TDKQLETLRDCYLKQPNPNSDDIEKIANNIGLTKRNVQVWFQNKRARDKKNPGIFDYLDSTISGNSPTNSVGINPQSEIPPSGAA
IKFQSKSPPVEQNEHSPLTPVLSTNQFSISNFSKSESSNVTTPPVKNEGMFGNASSNIGGNNGQNGENLMPALYSYQWLSFLSF
YQAMAASNMLNAGLNPASLLQSQLASLALLNQLKTESSDQNLDSAEKSPSTKPSPVQSETGTEKMDQSELSTEKSSQSDVSIM
EKTGPSECSSEKSEDEPLDLSIVKTEHDSSKTSETSPCSNAASDISAPNSTCTENTSNGTTDSLIMDLSNTLLPLMSPVNGHEATST
SSSPLSVPSGASSFLGGLVNSINAQFGNQEIHVTANRKEGNLGSPSQGMSPMIPSNSTPSPGSGSPSSSSPLVGGENSLHPHPRS
TSPNTSTLLAAACSAAASNSTGPFACTQCKKVFQKQSSLIRHIYEHSGKRPYQCTQCDKAFKHKHHLTEHSRLHSGERPFQCPNC
GKRFSHSGSYSQHLSQKRCL*
>Cmac.rna.tri.33281.1|m.28896
NPFLGCVLPSDPRMRIYRRRKARTVFSDGQLHGLERKFQQQRYLSTPERIEIATQYNLTETQVKTWFQNRRMKQKKIFKVNRCS
SSTSAANEEQDDDDEDSCEETT*
>Cmac.rna.tri.12623.1|m.19772
MASSPISKTLLSNGATENNNLLLKSVSRPTIHDFAALQAALSATSTATNLGKISNPTNLASPVGASVPAKPDVPISIPAPIVTSQLH
MTKQAAQLALSPIAPTITPLGCLPTQSPQMSQVANLAAAAAARTALGQLLKSQQLMGAAHSTTGGTSPIGLGPRPGSGGDGL
GSIPPVSSGISGGGSGVGCQLHLSRGTPCFCSSCALQKSSMHNTLSTLSQSSLAPQQPPIVSSFSSPSALPPPLGPHPGAPPPPTA
GLLPLPLVPNPAAAAAAAAGMLPLPGAAGLLPPVMSAGIYGPPPHSPFFDGGLHPSAADWVALGRRKNATRETTAALKSWLQ
EHLKNPYPTKGEKVMLAIITKMSLTQVSTWFANARRRLKKEGKMTWTPRNSYDDDDDDDMDDDDDVTALDENRSLGSKTSR
QDETPSGAPPKTSDVLRVSVNEDDEGYNDDLFTPKDSSACNNTSSGLSLSGVAGGGGGGGNPVVATAASTYQGLSLLKCKVPP
APFANPITANKITLVPNNCNPNSPFNSPNQTQVAPLPSSVTTPSSTIDLFSDVMNTGSLSTHSPNPSLSLLSQTHKARRRASNGS
KFWSIVDTAAKGEEPISPQADKSPLDLSPQNSDNEATSPKPLTIAVTCSE*
>Cmac.rna.tri.15557.1|m.22506
VLSVKGGGGGGTVGNYPLQQHTVLPQIHHWAGSNPVLGNQSESFEGQEDETSVKRPNSNGDNRYTIEKLLDTSSTSPTSNTTT
PTTKDKQKKTPPHSINSQKPSTNLNNPGSRAKKGKDSQSLSQVQPTGNIPANDLGSFPGGNALFQSQQAAAASLMLTQHAQA
ALMRQWQQSHMINFTQTKDTDHFLDSNLRSRSSTQSTKESPTDSSSLRSLEQFGSSLNTDLTSPLDDQHLNVTVNGGPPPSLLS

SSPVSANPADLNPFGDTPSLSAAGGGGGVGNNKKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALGMSESQVKVWFQ
NRRTKWRKKHAAEVASNRFKMAASMGNNNGSDHTGVGSLANSDGCTLNLSLGGSGDDLDECLYSDSE*
>Cmac.rna.tri.3512.1|m.6793
MTLSPKTTAGFSVTDILNPLEDYSKRSSAAVMDSAALNFASQMGYRAQHTSPHNHQGSNTAASAMNNVNASAITSNPYMAA
QMAAAGNFGAHSMMAAAAGTSAAAAYCNPNAISDISQSFAGHAAAAGLAQYDHFGRHSSMGAANVSGWGYAPSADPRL
SFSRLMHSSNGMGMGMTGSMAGMSMTSFGALGDPSKMLSSQRRKRRVLFTQAQVYELERRFKQQKYLSAPEREALAQGIN
LTPTQVKIWFQNHRYKCKRASKERQSAAAQQDGDKSGNDRQVKLESKDPDTDSLVDPNDSDGAVGSPLNEDDFAQDPVKSQ
LLANYATAAHAQQAGQVQATGYPGQQQTLQVNPGAAALNDMSRCPVIQPTLGAATNPAVYSSDANNLALFQNSLGTRW*
>Cmac.rna.tri.8416.1|m.14439
MVISHLFQPNFSMDFDSFIPPIPPNFTTPGSPALAPPSSQILFPPVNPSAHNLPPPFGCPGIPPPCPKKESIPTPPLGAQQSQCGD
VTSISASTSSSSSSLQSSSSSSASAALSSTAHLADSLNSNNSTINGGRSLDDLSNQSISNKSPFNQDLACDNSENINANSNSSSTHCI
SSPSIGGVGGEKAPKCIGCNQPILDQFILRVAPDLEWHASCLKCNQCATVLDERQTCYLKDGKPYCKLDYIRLFSIRCGKCDQPFE
KTDRVMRTMNKVYHVDCFSCNVCFKQLLPGDEYAIREDLIFCREDMVELMAQSTENHPTDPQNRTCENPTPESRLSDGCSPK
NPNMTKYGSMNNPSGVSPNSTLSLSTTHDLDRDDKIIFSPDDKSSSCKDSLRGEKHRSKDGSKPTRIRTVLNEKQLHTLRTCYNA
NPRPDALMREQLVEMTGLSPRVIRVWFQNKRCKDKKRSIAIKQMHQLQAEKNISGLNGVPMVAGSPMRHDPVLDGQFPPY
PHPPPPFGLPPHHAPHPNEMDPTYMMAFDNGGGPGSVISDGCLGSPTGSDCSLSLSSD*
>Cmac.rna.tri.10190.1|m.16926
MAQFTLKPISSFLQPLRSKYPDTEIQSTKSDNVIVQPKPHVFEHDYDEVSVVASQDEHSYSLMNADRKLQNEIKATRILKNLRPV
KPLTECQLITRIEKKSFNPRVSKAKIYPKVGDVCNFQHDYEMASCLQEVEVQISRLMKARVILNLTQSQFSEYLATNYGVSLQQN
DISKLEKRDLSLKKAQSQLKMLGQLFHDLKDTDWSSIAIEPIPELTRKRTVYPSHYLELFNELFEINAYPHRIELHFLSRFVLTDRERL
REWFSAKRMKMDKKSRLNLKKGNFDLWKEAVYKWDKKWNKVEPIIEEILEGTI*
>Cmac.rna.tri.22367.1|m.26359
MSKKSFLSIEDILSSPSAGQIDNIQTSRMTINDEQLRSPSESNSPLGHNMQNNIVSKLIQLEETKNLHSQFSGINIPGANSALATNL
NALMEAVSAVGPAGICQKLSPVGSDVSPNLTDTACHGEPITPTTADHHLSLLWNSSKTGTGGNPDMTSSLSPERHPNFLLSLQ
RKQQSKRELNSRRQRTSFTAEQTIKLELEFQMNEYITRSRRFELADLLGLSENQIKIWYQNRRAKDKRIEKAHLESNIRRQLNQLA
FGGSNPFGLLPSTANPVSTFAEELHVLDRLVNSESLAAEKFQLMANLQMQSKPTLVQSITANNHMTAKSMLENTLSFL*
>Cmac.rna.tri.14406.1|m.21461
MPPVVLQYPSHNYHSVSSVASQPQMSWAAPHYPTQQQQQLDTTSGTVTNNNNTQQQIAHHFSTHLPTAATAAAAFDPL
TRSAAVAGVNSKSSVNDLALHSANPQLTPVDTMLASPGYAQSFMTSHIDSKLSPFEETQNQQQLYWQANAYANYYNPSMVA
IDGCNRVTAGAFYGPRDASKFMTSHAAGSLYEQQMNTYRNLVGTGSNATNSVIPLNQSFLPLSSNNHQHNNNSLLQPQHHT
PSYDQHQQPTAFDFGANFNSQIQNNPLGAAHATSVNFVSPTGLLGSTNNNNSISQSNQNTLITAVHTNQSNAGTPITPTNQS
ASWPVVSQNATDCASMSVKTCRDSVKSETGSGSPGGPGSLGGGLNNPSSVGSPDSSNESGDNINSPSGASDNISPNPMYSQ
QLASESTNVTGSQQVMSTGHVQGSQSHLMMTNPAGSHTSVPNTGGGVSHSPGWLVRNVSRKKRRPYTKNQTLELEKEFLY
NTYITRERRLEIARSLNLTDRQVKIWFQNRRMKNKKQHSGGPGLGPQMPGGHPPMHLIVPPAPPHHPGHLM*
>Cmac.rna.tri.10176.1|m.16909
MDSFGSHNLDFTDLSSVGINPSIIPNQNNPHNSASVSSPSATTLPPPGSSSTTNSTLPSAPSAPDPQSNNNNSHNSSLVQITHNP
SPSNPNSNLGLHSAAAILPGTSHPHSHLVSNLNPAHISQINPINPLNSLVPGQQRLLHAETHGSLNALTADSLPVHTSLASSTGTLI
GTQNSPQDQLNFQSNFIPSVHGVASSSVAQFSAAAAFLPQSILAQTSVGVVPHHGISGIDKTGQSLDFGTGGLAGIVPGSNNH
LNLLKSTAPLFISTPPSGLAFGTSGGSSCSGDSLKIKRSPGSGETGNLYQNPTNDLSSASQTSTENNLDESCATDCATEPSNNIKSQ
SANHSPADDVMNNDDAIDDEDEDKKNKRTRRQRTHFTSQQLQELETLFARNRYPDMATREEISAWTNLPEAKVRIWFKNRR
AKWRKKERNQLQEYKNAAAFGMNFMPAPYDPHDFYSTQLTAGYYNNWTKMTAAASPLAPGKAPFPWSFNPNPAVAAAG
APSAAAMMNGMTSLNPSPPCPYGSTSATAAAATQYMYSRDATAAAAGGLSQFRLKHSPQSTSLGYGAAFRAPNTLQGACHY
PMDRQL*
>Cmac.rna.tri.21903.1|m.26181
MSCSDPNSPVNLTIQRENTEKTVRNNVPNIGQNPDTNSAVKHAENLRSSLLAENEKSITKSPRERFSIDNLLLKAASFINNSHGD
VTNDATSISSDVRTSLDHVKKIMSSQGCDNVTKLITMAAAGSQSGLLTSASVLSGGQHVNNQCDVRVASNDIRVSASLARDRD
EQMTAMMESDEGSEGDDSSCEEDENDDSCNQSAVGIQSTTGVPASFQTHLSSTAAANFLKSQTLDSLQTAAANKVVVSSVG
DVGGGVISPSTAIEHYEQCLSDLQGANAAGVGVGGRRRKKRTRAAFSHAQVFELERRFAQQKYLSGPERSELAAALKLTETQIKI
WFQNRRYKTKRKQIHGEIIPSQCVGGHLTTAGGVTLESNPHLTHLRKLSAAQLQIALTSQLANNSSMPALNPLSNSITATSQSVV
NPSLIGSAGLDFAGAGSIPAVSGGGNPLFAGDSVAQKTFAQLQLELLRATTMGNNPYLSLLCYSAAAAAASTGILPPGCQPTKA
TSPAGINTSTFPQLQSAAGFASKLAGLSGPNSNNMLSIANLLGNQQTVPTTNPQSFKANCPTTDASSLAKSNQNPINKLNISPA
GDKDFKLAIDQSKAQSDFLLAPVITAKSI*
>Cmac.rna.tri.15185.1|m.22212
MDSVVTKIVDKCCPNLKQSEIHNHPTSTLQPSGSSTSSSSASSGVAPTTTTPASAKQPPHSTNNTSPASSHPHSNHHHQQQQR
YNNNSSKQSTTERTKEGHNHQQDSQSKNVIRASGKMVSTASNTKSSADMYHPMQANILNTNTTVMEPHPSEISPNGGGAN
NRERRQRTQFTSHQLRELESMFARNRYPDMATRERLSAWINLSETKVRIWFKNRRAKWRKKERHTTQQTSAIDSWKPSSLLQ
TTPNVPPFNYPNYDPLRSSLVSAGAPNAFNVDSKTTSPFIAPTNAFGSCSASPFRSYAQPSWNFFSNTLQSQFHAVQPFGGHSY

SDPMLDHHLNPFGHSAATTATGVPPPLGGAGTNDSLLTAAGAAAMNNSNLSQTQAPPNPYIVKQEFWPELDARGATNTIAA
AQNNCFPVTSR*
>Cmac.rna.tri.24033.1|m.26908
MCDFGTVGLGPGAQSNFYFNREQIECIAEVMFQSGDVDKLTRFLWTLPSSELIHGSEVVMKSRAVVAFHTRNFRELYSLLETYT
FSLKNHDLLQDVWYKAHYYEAEKIRGRALGAVDKYRIRRKFPLPKTIWDGEETIYCFKEKSRNKLKESYKKNKYPTPEEKKALAKS
SGLTLTQVSNWFKNRRQRDRTPHKNHDGSDSEDELTIDHKGKFVSLKANSTTTANSNTHPSIYANTAQNHLNTNYSSTNSTSK
QSNNYRSIESNQQKITMPNIFPQSEHKHSSTKNGFKKSPDSQHAKKNLAATFTSGANNVKTGSQNTKPAIRQAPNTNSIANSG
LAPYHALGLFAHNANVPDGTHKLADVSSEHQNHPVAFESIISANTISPYHNAAAYLQSFSKN*
>Cmac.rna.tri.20345.1|m.25571
MAASPIKSFGIKDLLGISEKPTHQLPDKNIQISSPGKISSFSPTKPLKPSVTELDMSLPFSFPGFCTENSSKLDSENLIAHSTANYLHF
HSLLNSIPVQNANNTSSELLSGSYNDLASLMCLLQKSNIPPISSNSSQNVGNLSNAPIPTSSIIPGTPVDVKASTQTSSSNHEQVL
MQQGNVDLWEWSRMIAAIGNQTNLPYQSSGIGQTRPELVPGQVSPASLAEINAAVLLASQGLIPPLGLESHQMIVSPNKQSV
PLSSTPPFASHPHSASFLCSPRDPRLDLLDPVNLQGMNSHHLPTHYHPHLVGGVGVGVPNSSSGSNVGGGILSSDSSSDVSPSG
MGGVMGGITGGGGDKPHKKCRRRHRTIFSSVQVDELEKAFEEAHYPDVLQRETLSAKTELPEDRIQVWFQNRRAKWRKKEKK
WGKASIMAEYGLYGAMVRHSLPLPDAILRSREEDDDDDDEDDDDLDENQSESEDTSNDVTLGNDDSRRCHIDFEGQRSSSGV
TDSGETEDVNKQEKTLSLQKKSPKKKTKKEPIAPWLLGMHNKSLQSLNASKPNKHQSTSRTEDV*
>Cmac.rna.tri.17874.1|m.24213
MVITPGLHDSSIVASGLVQPIQQCSPSYNQSTTGTTSGLSGLNSFGENLSSATFGVATSAAPDYNTSYLASTISPNMPQPTAGLQ
CATATSAVSSAGASAAGLQALQHHSALGNGNSSPYSAYNKDLGAMATAPYSTCLTPPQAGTTSGGSGWLISPESNLSWNATS
FPYQNNPQCLIQSGTFFPNGAGASSIYGAQPAVGMGHPLASSNTWSHLAAAGMHPFNGYMTPSRLDSPPPNVAAPPYPLST
TPTIHSYYAAAAAVHANQSLLYGRTLDETALLDCMERECEETPSNDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQT
TICRFEALQLSFKNMCKLKPLLTKWLDEAENNNGYPSALDKIASQGRKRKKRTSIEVSIKNALEHNFHIQPKPSAHEISALANSLQL
EKEVVRVWFCNRRQKEKRMTGAYLDDPTGQKHPDDLENGVLCDGESSSDDEDLNLREDEQLIEQNPNCNPNNGISSAIDAKL
YQNRNPAFNELNCQFSASEAIPPSLIDPKNEVSLASYCATQGASPHSILGAADKLASMEHQNVAIPSNSTATPLHVNHLNIICPV
PSMDLMKDQPKLSTDTYQQQQHQSAEKERLILQQISEQQSNFNCHQFGYQFST*
>Cmac.rna.tri.35406.1|m.29236
MASSTEDEKWQEFAVHSPDANEELEEPKHGQDNVNESESPLAKEPKEKQSPRFSIDFLLSSHCDKQQLQLQRKTKNDTKEVLF
NPDSSIETQNVFQQSCTSPEPEGTVNQSSDDESDAEDDIDQSDETTHGKPSNHSSSSSATSYCNGGSSSCVGKKHRKARTAFTD
EQLTQLEENFERHKYLSVQDRLDLAASLQLSDTQVKTWYQNRRTKWKRQTAVSLEMLAAAERQHNLCTLQTAALANLHNRH
PHTHPQTHHLPPGIPPLPLPGPLSPHLDHSQHRLDLVNSRSNGPLLLAAAAYNQFSAHAALFSQPTHALHMSQLLSSIVPKGSP
NKTPNSR*
>Cmac.rna.tri.18210.1|m.24417
MDLISDINFASSATLAGDGLVGNAIANNGVNNTAAASPIGAKNANFLNIDHLTSCNTATAAEIPNLGAIATPVSKQDTIPNGNQ
AANSAICDSTRFTNFLMCPDIGSQFSTVAQNSFNPWIPNPGSEKTTPGINFGSSPLQANVSSTALYSNPCAAGSSAKFDFNTPA
FDFNCYNPAAGWAPLDVYSNSNMAASYQSQRNLSATALSACSPAKFAATYPSINVNGIYYPSSYATPYTRYSRNLLSKDGEELD
TRDIAQRITTELKRYSIPQAVFAEKVLCRSQGTLSDLLRNPKPWSKLKSGRETFRRMGKWLEEPEFQRMAALRIAACKKKEEEQH
VIAKPKMSKKSRLVFTDIQRRTLRAIFQETQRPSKESQAQIAKQLGLEVSTVSNFFMNARRRCGDRWKMEDSCAKTSSVVRPV
PTLQHASAINITAANNLNVITGLNCPPIPL*
>Cmac.rna.tri.9717.1|m.16318
SLPSGTPLSLPSGTPQLPQTPISSTSLPLGGIQPPHDFPHLPPNLPPPHSGVGGNPLSAIQAGYNSQPPCSPFNPNHVQPPPQVP
MGFPGSEGIPPPGGYPIPPQSQGDPTALMSHTPTPPKSQKGRPKKKKPPQEDAFQNLYANDFNSAGLHGAHDAVLGLYMAA
TNPAGVPGGAGGLEQQSRTKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRLMMK
NQQNGTNTKFTLDDETPENEDTQNMKISESDDELNDEEEKQNSINNPNSPSAESMSDFGEMLSPDLSPEVTSPTSGYSINQQY
DKSSIDESLSSNSNFSLQDAVLNPKSEPMTAF*
>Cmac.rna.tri.16996.1|m.23638
NRSRKKRRNRTTFTQGQLEEMERIFQKTHYPDVYVREQLAVKCDLTEARVQVWFQNRRAKWRKKEKHQQLRSGFPHPGAHP
GHPHHPITPSTPFNPSAAAAAAVAVDSSSLAAAAAASGRNDVITQQLSSTWPYAASVPNGATAAAAGALSHLGSAAAAAGM
NNPAANPYSSLSAAAAATATNPHQQAAAAASFFPNYIMAAQQVAVAGYGCLPPTTPQHNSTGVGGASSAGVVGTVGQQG
AGAAHNLGSAAAVNGAQFAGSNQAQFPGMSFLGSGAFGHPSLMSAAASSFFSNPNPSHPGSASSAVTASSGSAATSADLFT
PASQMNSLDPRSGFANPFRFAANQQKPSLDTTRW*
>Cmac.rna.tri.16481.1|m.23241
VNNNNNNNLNTMHKMSTENSTFTSPSSHCSEKYSTISPVPNSKSTTPTRGISYSPADNDIIMTSDKQRDECNDEDGAESAGDM
DDDSSRKKNRTAFTPHQIYELEKRFRSQKYLSASERGEFAMSLKLTDTQIKTWFQN
>Cmac.rna.tri.16769.1|m.23465
MQVQRTLELAAALNQHQSNAVAAAAANQHAQLVAHMDKAANVVATGASFCDVTHQQCVGGGGGSHHAKLCRCVSCVAT
AAAVTTAHGNGAAKVLTREEEFILKYGLLANNSGSQTNSATAVKENTYPLRVWLEENRDNPYPTKSEKLMLSIVTKMTLTQVST
WFANARRRLKKERRPEHDFLHDFYPFSLYPPLITAPQFYPPPTAAIPKILPSAAANLIGPQPATVQPIRVLPINHQGLSPTAIHDYL

LSHREALNSLGLPSSVDPGRPKISHHEPHTQLVVNKEQDRIQTLIVNKPTITETTTASPPKPSVLPREHSRFSNISPEVQVIKAEVH
KLPREIITQPQNKSAIKTPIKNIWSIVDTVKTAESK*
>Cmac.rna.tri.13617.1|m.20746
DDDDDDDDDDESGDGSGRKKKTRTVFSRTQVCQLETMFDMKRYLSSSERATLARDLGLSEQQIKIWFQNRRNKLKRQIKTGIE
MAGNISAGTVPPPNVNAALLNSLGSHPALAAPNTVHSQSFLSSSLPQVPFFPGSSPAQANQIFASNLPSSLNFNSGSPTGLTSN
VSQLSSHPIFKTLPPFGNQQPPSFFQQSLANQNRPILTSNQSFNNPFSSTNYNTVAPTSTNNNTFSYSNLQLFKSLSGLV*
>Cmac.rna.tri.22219.1|m.26288
MTDPKLPIGQNPNQFSFLAGIHQLNSLLTNNSTNNSVIPSSTPVTNPNQNPIASVAQFYPSIMGGVSSPRILENAAFNLNAYNL
MFQHHLKFLQQHRGDQSVIERKVAEQKDSAIGEKSCKAFGGVQTKQSKKGATFSKDKNSKCKTEECTEKSSPVSHSETFGFKSE
DKNDGSQLDSNLCAEDLDADMHSDDLCDPNDSKRRRTRTNFTNWQLEQLELAFHRTHYPDVFVRETLAMQLNLMESRVQV
WFQNRRAKWRKMENTRKGPGRPPHNAHPLTCSGEPIINTTTGEPLGPNNNKSQSQNGEGATGRKRNSSTTQSKSNPNNQPI
KSPKLNSREFSGLSFLQPEQDFSSNKVDESKSIVGGVEIANDLKKDERISLFESTVDLITKRCSQNAFKTENDIILPTFLTSPASSPRS
QTNESDDQKKSFRIDSLLDL*
>Cmac.rna.tri.8375.1|m.14395
MKRSFMMVDLLCNNSDSKPETNTKPTPELPGSKLKMPHLDQSVQNLPGVVPSNVGVASNCGLEIVDPVSLFPVTDSLSIQIGD
EIPKNVDNDDEEDRTSDSPDSGVECPSFFEPTTGEEESHDDTNANSEPSLKATRVEPQQSSVANNEIQKIQSTAPISRPASVQSA
KNEDRDMDDDDDKFEGSESYDDFDAQSSGSRDDDSGFHPGGFRSGSSSSKKQRTAFTPDQIHELERRFAQQKYLSASERADF
ASKLKLSDTQIKTWFQNRRMKRKRQQQDIESAQLQALYSSALMMPPPPSHILVPPMGHPNPVAQNPLQNPFQPPANLFSPA
QPITSSASNVPHLTLGLSEHLPTSTGVLATQNQARGPATGDFPQPGAPQKSLFLPALGLSATFGLSNSPAPVGTNATLPNLTSSS
RALELTSQLTSSMPPAMSFSSMGLFGQPKLGL*
>Cmac.rna.tri.20671.1|m.25700
MAAFSIDNLLSKPYEKQKSQSDFETKGPQNSPVADKPSAQPEIPNPTVTLGISTHERSSRAASLLYSATSTLLNPTFHTLYNSASFQ
SPAAPQQFSPYIATATQPRFTSGIGDLPSTWRLKGSVGNTVTSPNASSNSLTSGWSEHDENHVSPIAQSKIKYAMCEQNDCVS
DLGEDEDDLDPENDEDEDEDDDLMNLNSSGNEDKEKTIPEPRRMLIKDTNGQIKELIFPRGLDLDRPKRERTSFTPEQLFRLEKE
FTHNQYMVGRDRGQLATCLGLTETQVKVWFQNRRTKYKREKQREQETKKNDMDTLAACSMFKILEQSTGQYYTASGPRLFA
TPAPGHNSNGAALQNASHCKPATPNGSPVAGVAAGFQPKFPGSAVGHNGPFVMPPSMNPSFSAQMNQLATASSLTKIAHN
GAPPTSNWSMSASFLPQSLYWVGTPN*
>Cmac.rna.tri.31027.1|m.28543
MEHQEEFSLLPTSIVPQLHNEYHIQSLHFPAELNTNSALTLGSDNNPGNNNSDSISITDDFFESLISTSAMQPIVDQTVAEGPVQI
QVQTPIQALEATSDAGQVDQGQPPSVPVNLRQNEANGQDEPAAISVNLKRQEQPQELGLEASPATGNDESDSTDSKPNGNK
RNRTNYTPQQLELLELTYLQNNRYPPAEIREAVATTLGLKDHQVRIWFQNRRAKDKRRSASQIVPCPPTAILQNIANKPPISSSQS
AGTNGIPMNKNIPMNSNNSHHNPIAISIPPVYFPAPNVYDQLQ
>Cmac.rna.tri.22028.1|m.26225
QAAVFSCNPALLLQAGGIAGLGGQNGGVVGNQQMENYSRLLALGMGGTADQTPGFVPPFPPNPLAACSPVDALVAGPAQG
VGNKSRRPRTAFTSQQLLELENQFRKNKYLSRPKRFEVATKLMLTETQVKIWFQNRRMKWKRAKKVLGSSHNHNNNNNNNN
>Cmac.rna.tri.5430.1|m.9933
MQRYPYLRRNVAMKQRKRIYYSPEQKALIWDRYKQGDSLHDIARMFDRFHSSIMPTIYQTGGYRPPERKRHPQSLSLDEREEIS
RCLVSKQSIREIARRLSRAPSTISREIKRNGGLKHYRAVKAEQRALDEALRPKPCKLIGSPELCKLIAVKLKRAWSPQQIAGWLKRQ
YPDNQEMYVSHETIYKTLFIQTRGALKKELQRCLRTKRAVRKSRTSTLKGRGLGQIPDALSISERPASVEDRAVPGHWEGDLICG
SKNSYIATLVERHSRYVMLAKVEDSKSSTVIAALIKQAQKLPSELYKSLTWDRGREIKDHKQFTLATDIKVYLCDPHSPWQRGSN
ENTNRLLRQYFPKSTDLSVHSQQKLSSVARQLNERPRKTLDYETPAQRFNQCVASID*
>Cmac.rna.tri.38352.1|m.29605
MKIKRNNFSIESILASDNKFETSGTSQVDSVIQADLNNAKAAFSRQSSSNVKSRKRSLDAEHESEFDERLSHLAECEQKKSLISSPT
SEPANQKVPNSFYTNYATSESTNDVINDVIAFSEATSAASCRRSRTTYTPWQLHQLESSFEENPYPDVGQRDYLAGLMGVSESR
IQVWFQNRRAKRRKLARVFSCSNDQRTNTNLFTTVTPSIVPTQIAVKLPSGFQSALSGDNSTLFGKQIICSQRIPNNTDTNLQATI
TNESSRRERLPSSAGQMSWTQNLSPEQLSRLHSAAMPNFGYFSNPTAYHQIVGKALVQIAE*
>Cmac.rna.tri.11630.1|m.18623
MPNRFKFQGPNGSIIMEFDDNSTVFQARTLLSQTVDRKVKGLVAVADKLTNGSTNSCSSQSQHSVVRLADNQRMKDVDKRA
VIMMDCEKAVVNYISKVPGSDQTSKQPHNTTESDVNMSIANGDTKTESAAIHLCETESKQIDLTKTQAADSSDKIVSNSSSISSF
LSTAAIAMSGTSPNTTSINSGPSSYSNQNEKVPLGVGLAPLPFNFSPFLLMPAVLKYCIHFEIDKFNINAEIFSHATMGCNAATME
KVLLLPAGEYMDGSGKFLPIVKHFMLQESERAKVYEGVKQQLKAAAAASVTNTLADNITHSAINNNNNQLNPMLKANETLTNS
AIPQMNNNFLLQNSFGSLDNIFTMNKTHGPTMNTLNNLHLLNAAVQRDPLVASLQSSATTNNPIKALDITTLANATNTSSAQN
VSPTSNGSNKSTQNSLSDLAKGGGPNILNQLNQHQQSSTVDSTSGSIKRMRRRTTISRSAKMVMLEWYLSNGRYLTADGRTLL
SSHLGLTEENVRVFFKNLRRLEKKCEESGVSLYDAINRVDNESSALSEDINLLSQQSFNDDEEDSSPTMIKVETLE*
>Cmac.rna.tri.36089.1|m.29329
QPQQQQHHPKKNRLVFTDVQRRTLFAIFRETKRPSKEMQITISRQLGLEQSTVSNFFMNARRRCGDRWKDQDHHVIGQNS
KPIRSKHSLNNEKNKSLMSITPLHASIA*

>Cmac.rna.tri.19752.1|m.25288
MLASSVGNFTDETVADGSLTDGAVKIEAIEKRTNDRGTKPTQERIINETQRTYSISSILGINATAKSIAEIGNFSNAIAEKDQDSLN
KLKTTYQTRHNFTSSSVTLPVDNLFDQKSALKSLFTNSPDQSASGMNEFMVPKALQMSRTFSNLDIKLASEHSGKLLPPDGTEA
MISKERNKSFQQTTSMKSKQSQHSELGLPGHAKTGTNDVRVDEVVKSSDVAAELDNVEEDDDEDVEGDGDEAGKKHRRNRT
TFTTYQLHELERAFEKSHYPDVYSREELATRVSLPEVRVQVWFQNRRAKWRRQEKLEQNQLKLNADFPLASLRAAQNAPIPGM
NKILCSPLDSWLTASAPAAAMASLQQQHNMLTNPALFNLQQAAMAAVNLKSTPEPSSTPNLTTSPNIQGFDLANSTSGGNLF
ATNAISQQQAAYQLLLQSCAPTSFFMPALNPAVNLFPDFGSMDIMTKLPNSTASNSPIPESSPSNEKNKQTTETVEASDKQ*
>Cmac.rna.tri.14860.1|m.21899
MQEITTIHCDSNSNALAGNEPVSYCFACGAEIRDQFVNKVNERCFHLKCLLCHSCHSPLTDKCYFTKDGKVLCKKDFFNQFGAK
CSGCEKGVAPTEVIRRANDHVYHFDCFKCLICGKLLETGEEFFLMENQQLVCKNDYEMAKSKAEFEDLTGSNKRPRTTITAKQLE
VLKAAYKNSSKPARHIREQLSKDTGLDMRVVQVWFQNRRAKEKRLKKDAGRRLAVASNSQFFSPGNGPLTPGSSKKRAKKSPS
NILGQPIEIPQGAAAFPADVGLMASHNNAAAAMQQHGLFPPGYGPQISQPVLFRDNFPIPGMQGIPHGTPPNLKPYMPISTG
LLSPIGDIGSVAASVTASLANAMPPSQHAVVDSRFSQLVNNPGGNLMYPGAIPDNVPTPPVGNLGGSFASPTSSANGGSIAAE
VLASRLANSQMTPPQPTSMYQSVPNASPHIPSPIFPHSLHPGAAPPSSLVFAPNQEPISSLHSALTAGNRTAAANYPPHPLLPG
GCKVEPGSGGPPSNGPSGQIGSGNPPPPPSQLPFLQSAGALQPTAQQLEGILMW*
>Cmac.rna.tri.7258.1|m.12753
MHLDSTNSETIDSMSRESIQTGWIHHLAHNFTNVANNAAAATALATNAAAATAAGTYHPAAGAVTAAGLPALTPSVTNPGH
HAAASLKFPGPTPDHGSPTGLKLHSDHSPIDQQQQNNHHSQFTELRTAAAAQTAQAVASSVDHAAAATAWPASMPHSYLA
VAAASHNQAANMMHPVVSARNSFTGASGATHPNAAWPHPRNSYPTTTQFASAEQVAAAYRTGCNLNIMNAVVAGGAAD
HHSHNSTAAAAAAAMHAMPFYMTAPPNPMAPGAHPAVVHHQNPFVSPHHHPYAAQVAAAEMYSDHTSSTFRYNPTTGK
TRTKDKYRVVYTDRQRAELENEFRSAQYITIRRKSELAMQVGLSERQVKIWFQNRRAKERKVSRKVPGGGNHSSSDVEDSDNEI
EEEDEENNLRSPSKHQSLTSPGTNNTTCHNGVVQLPLIKGEPNRIRSPTNDLYGVKLS*
>Cmac.rna.tri.20114.1|m.25445
MPGFSMDEILGLRSNKTKTASSLTSSVNNVKLRISPTSTSILDPTGNSTIQAGSICSFNDSGFDQEEATSEVGSEEVLDYYEDEDD
DQTQLGDDIEEDDDDDDEDDEDHNGERRSSGISHRVTSSGHKMTGLEGVSSAGAYGRFPTSHADMKHLNSANFESLSDAEKS
SFLSNLIKNPNEAGKQGSPLSGSPGGKGGKRKRRVLFSKSQTYQLESTFRRQRYLSAPEREQLANAINLTPGQVKIWFQNHRYK
CKKSTKDKLSPTMSQFDPLTPQGYCHSGANPYHPLIAPHFSISAAASTGPQGNALNAANFSLMMQQRQQQLAAAASSSNFGK
PSQQAVNEFLRQQSLNALNTRNQQALSLAQSLAAGLGTNPGMKNAWAQHMGMIHSQLQSQSALGASNSGMNPLAYAYL
MASASSQNQQQTQSQIQTSQANSSHLQAQIHKLSLQQQLHELIGMNSKLSAKKSDSPNAVNAEREGISVKLPESVSSQIKPNS
PNVQVTSPNSVSQNGVKDFTRRTLEQEEVA*
>Cmac.rna.tri.34807.1|m.29122
MDFSTCDKPTQEQVSLELENGSKHEESNSCHSPLVVATDLDPDLRPLQVEVPNADFSKSEEQQTTEKEQMKPTSTDANLGQIV
KTNSEPNYRSEPKKAPLSNSQRLETPTKPNNSSSKLPQSSLDPSTLDNFLQNWAMFHHLFKKCSPTSEQNSATMFQNIPENSSA
NPQSARNLAQLLQALNLEIERQRQVKNLADARNALRGFGANYQNQLDDFGWLQHQQHKFCDLLPAQQFAANFRPNSTSAF
PLVPTFPLIPNNPISLMKTNALPSNLMPPQSMNNLGNGHFNPVSIANSLNTVRMNKFMTNEKSSDQKQLHQIQSFKSKITKTT
NKFQMFGQSSNLQPTSSKPKRSCDGYSGFKHVKNPLRTANSSTSGMKRVLREQETAMDCSDTSAGIVHLTENLTATHLRKAKL
MFFYQRYPNSAILRQYFVDVKFNRSNNAQLIKWFSNFREYYYIQIEKHVREALQKHQQLSKDNSSWPSLQFPQSNYTN
>Cmac.rna.tri.16808.1|m.23496
MDHKLKIPESVRRIKITDDNGNEKELVFPLGLDIDRPKRERTCFSAEQLFLLEEEFCKGQYLVGRDRTILAGRLGLTETQIKVWFQ
NRRTKMKRDTSKSSTATSTTATSSTTTTNTSSLEPVKKIISIIIVQFIFIFNHQETQ*
>Cmac.rna.tri.12338.1|m.19457
MINQLSNNFTRPTMFASPAVSVATSNPTSLMHQQPPNLSNNAGATLANCASPDAKYTIEQIELLRRLKNSGLSKIQIVQGLESLE
KIDVNGFTTPTDLSGASKDQEINVNPKANILPNGIGQLGTLNSATTGADYLRTLALQQAFFNAMTNGNFVNGGSHLQQTNIGA
TLAQRNLAAALMCSQGINLPGMGGIHSAAATTNGALTGTTGFHHAASTNAAENSILNQMSHINSNGGGGGGASSPINLKREA
NVGGNGLNNSLGDPSTISDDDEHLVQLQSMSVVDTKEEIKQFMLSHGIKQHQVAEKTGVSQSYISQFLHQGVWMKDSTRTNI
YKWYLDEKRKREANGDLASPVNSSRGNSPVSENMNSSAMTPPGAISSNHTTQGQSQQLPMSSNNSNNSATSPQMQQINN
QNNNNNNNAASGMSSSPMVSNGIAVIAHQTPVVQQVVKKRDRFVWKESCVSLLEKFYCVNPYPDDTKREEIAAACNQLIDST
GQGHDEKMVTAVKVYNWFANRRKEIKRKNHFAEQQAANNLLGGHVGVLPFLLDSKVAINTMNNNQSNNSSKPHLPQQTTS
PSSNAGSASGFGYLDELMEHNRAANLQNQTIVKEELKEDYTTHESALKLCQMMTQNDCDQPHAKKPKISLADQLAEVNNAIS
ELVPGNEENSILDEKDELRNDSAMNENNSTEDELESKQPTFDVLNLTSDRSQND*
>Cmac.rna.tri.66096.1|m.31986
MNTRISHGKQFTSQNETISWSNVYVDNLESSLNEEVNTGGTTVNICRNTEQIIDIPETRAKKKFSFAKYQKQRLLLAFAENSCPSK
ECMHKLAEDLDLSYSSVKNFFLNS
>Cmac.rna.tri.6704.1|m.11891
MLLSPCVRTGRDEQSDISNSLNGLHNNNRSNDSVRINNSTHNHQYTTTGSTSQGEQRREHFNRPLNLDLLGHTNFCSASGQSS
CIDNNANKVDINSSVNKGASTRITSSVNFANDLSDEFGASSSTLSTSYGQELTNAREKLRHRLRDKSSTLLSSDRPSSAGRIQPHN
SAILARRHSSVSPQSSSNPLPFRRNTMSYCSQVSAISLSSTKLTLDNLLDNVISSVQTKLNGTDVSSETQPLLFNNNNNYNTNTSC

GINSSSSACMNQSNLNNNANCYYNANLNGSNSTTTANNNSTAANKPSCIAKPNGAMNSNTNAVLNNNNTGNEDFIQAGDF
LKYKVKLRDATMHQEDLLDPELVQVQTFKKARPESGWASILLNAAIRVAFIPFLYSWWRTEVSRMQYAMLLGAWFMASWNI
KRYFMSQLPIVSNSKFSVYELELLCPSFYLFLLTFTYCHIHNARKRTPVSNNSSSYSTVNTGLHNVSTSPNISRDVDKGNRKNIKQV
FPRRTTRRQPLRSSILTPSVSIGNNRSLIDRNLSGRGSRTQSDLAVDASSIPRLQVIPSRTEVQFLGPNLSPINDQEREENDVESVT
NNEDRDSVSSLEEHVQTTDNTGGDINQARANAVNLQLGQNEFARQTIAERGSYDESQFDGMNGHEDNSAEDTNTTVSRASP
ASQLLDLGSSRDENDLTAVESDHASRIFVATQSSSLDVSLEQGAISIDMPSTSHEPDLDDNQKSKESSGDAVDFLEAMVQSSVS
VFHMSPYNNTNGQFNSVNSSRADTIENRDIIGIGAPSQREETTFSDDLSTHAEIGTSVEVTTAELDSETPPWTSRQLSTQSSSNV
ESKQSQQQLHSTVKQSMSVPLESFYDQSTNYSDPAVMFYTANQTGTDVLGSNTVHEEAKTSDTNIESDIPRPTSISLPLDSFDY
GSRPRSTSTSSLTIGRNRSIRKSVSSVVRLGPSKVTITPSRKRQNSLTSVNSRSSFLHKSIVKYLGDLEQKAREASFKENNAQLSLRD
KFLRFLIACKMKLFGELIESQKSILASPVTPSSVRKTAIGHPPGDDVALGPTRSGGGADGTGRLENGSPSIESEEEEENDHGNILEE
SEEEEDEDVNDGEGQEEETGGSSIEEVEEQNDSANVNLNGENFSTTYDHCPGADEPSHHMYNRSANAASTGPAGTSMSQTP
EKTVKIITWNQMGKCNKANLTAEQVVDMIVTRADSIRNSNFYLAFGVFCAAFISGLPFIYRYFAIRDFIMFEDVAQLIAAADLEK
VAATNFIRMETYLSYFCALSSFMTCTVFFLLLAAAERTYLQRYYYSKYLYRLTSSRKARKCEIPHFRLTKVRNIKMWLYIRFKLKAH
GPQKTVEVIVSNCLLFCILSISCFCFKLLEQHSNANDSAHQTSSSSSNQCSSESAANSPGAGGSGNSNEAKITYPNLASWIAIMW
CLALGSLLLRFITIGTKVNKKYSNTAMLTTERINLSLRMENKPEKREEYFLADRVLKLACRLISEVESPFKVSGLVMNSALYNIMRVI
VLSAVSAVMSEILGFKLKLWKI*
>Cmac.rna.tri.8043.1|m.13845
MNNPRFDEQLPQVPCQYPLTIDSGGGGGGVTTPYDTTSSATPAHLGQYHPGSTTGGGTTPQQPQSNSGSGSIPGYHPGAQP
HPGSVSVPPQPGSSAGVLQTVGGSGHHASGLGGMATTPHGGQDGCVQDKDSVYNHPLFPLLDLIFEKCELATCTPREPGMS
GSDICSSHSFNEDIQVFSKQLFDGMSGEKSSLYSDQDTDSLMLLAIQVLRLHLLELEKVHELCDNFCERYIECLKTKMPMDLIVED
KEGMSTQVSTTTSTTNNSSCLSTTGTQNISVNNSSSSKSNRVPSTDDDQNEFEEFSTAPCNEMGDPEVLSPSLAGFSTSTTTVIP
SQLPVGSKSNGMASSGGDFLPSSFGLSVSTLDQSQSSISAVGNCALNSQLPPVHHLTNPNNQSASTKMSQSNSSILNISSNNAS
ALSNSHTSINGNVSGMSNSKSDSSKYSSSLYGSQKESSKSGKQNLDKNGNTKKSQSQNSSPEISDRSESFSPTSNSTNCVGGTG
NVGSGGGVNNPDSNGGVTSPTGSDEDPKKPQKKRGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFI
NARRRIVQPMIDQSNRAGMPAAYAADASQMGAFMLDGQQAPYLRHHPMMHAAATLAHGGNPGADMAAASGFAAGYP
GAAAAAASGFHHHPSQAAMLIPGHPHAAMMMPPHAAAALHHQHYQQQALMDPMATHVMDSIHS*
>Cmac.rna.tri.18738.1|m.24727
MHSPKIQTSILEPTSESKHEKVTKFSINYLLQKPENEKLGNLDPSPAGNFLNDGVPHFPRAEIFPGSQFLPQLANAHATVGLFRQP
PFSAALGPSLQSLLASASSQQLKRNQVAASVGTDEKLISPPSEWVGKKQQFLSPNKSPNKTLKRSSSADLCLNISASSPDAERHH
HRLEAKDDSRDDLSPNVIMFGHDHKSPSPLGSSNRSSSPIRTGMKLSSPGGGSVTGGSTGTQMHHVTRKLGRNPRVPFSSAQ
VAILEQKYQVKQYLSSIDVAELSALLSLNDSRVKIWFQNRRARARKEQQMGKGGCGGVPVSANPNVTITKHMSASNSTPIVNN
TTSQLHQQLSSQQSSPPKQNLGPDAQGNTAPSWANQTNNNNNNFPPHFSNTNNNNNNLPQIRPQPNIHPAFSPFANFCST
GGNSVTSPFLRPPIGQLLANHNSGAPQL
>Cmac.rna.tri.16817.1|m.23505
MTKGNISIVFPNSNDTNDHQTFKSFKLNLPKFEEAVPKIPSLLNSHQSPLGFDDNDDIDVSPSDLAESNVSSGSSSKNILEPSTKKF
KPQFETCQSCNEEIVDKFLLHVNGKAWHTSCLRCSACMTHLGNQITCFVGKDDQLYCKADYNRYFGVRCAKCQRVVSSTDWV
RRAKQLVYHLACFACNICHRQLSTGEEFALHDHHVLCRQHFMELLHIPGNSNNTNTNTNSGSSVSTSTAGQAGSGGNMAAC
GGIGGTGGVDGGSCDNLSVNGDAQDDLVGGSGGGSSSGDLDCVLGSPSLLHKSKRSRTTFSEDQLKVLQANFNMDSNPDGQ
DLERIASQTGLSKRVIQVWFQNARSRQKKHLSTGATQPGAGVPTGAPAATGTVGNNVNGAAGPLMPVVSGGGGAVYPNGG
TAPKLHHSHPHHHHHHHPHHALSFGPHATPPNFDDVMSPNALTSGIM*
>Cmac.rna.tri.5739.1|m.10428
MFFPANSISLVGTTQISSPSKPDSAGLQSQTQLMDQFHNSNLSKSLVGECKPIMQNTTDQLQAAAAVGLHSQHTGIANNAAG
ANLASAGAVASQGHNDFSKYSASPFQIRQRPAFAPPPTSNFFGHPFDDSFFRADSLTSNHMNHMAPKPMDYKPHEYFSTMM
ADMQNNAARCSQLDLLDQIAPSVQFGGTTDHQTNSFHPATHIGFHPSFGPGNMAAAAGLAGPSTFPQITHVPQPEIEADPRE
LEAFAEKFKQRRIKLGVTQADVGKSLSHLKIPGVGSLSQSTICRFESLTLSHNNMVALKPVLQAWLDEAEAAAKGRLSEKQSSCL
SVVDKKRKRTSIAAPEKRSLEAYFTVQPRPSGEKISAIAEKLDLKKNVVRVWFCNQRQKQKRLKFAAAR*
>Cmac.rna.tri.11838.1|m.18890
MSALTSIALMDYFLGSSSHTSQTSSLLAANLLLNQQQHQTQQQSRLVTSPPAGHVTSLLGGEASPRNQPAAVGGGSSSESALL
GLLASAGIVGGGDVPGSSQLQLGASSLAAAAYHNLTSSSSTGSLHHHHLQPTPSGVTPPHPPIHPMMPTPPHHPPAGTLALLSS
HPVNLAHTYPYHSLPPVGGGHHPPPLPPAVAAHFLSSPAAAMCSALLYPHRSKRIRTAFTPAQLAKLETQFNHCPYVVGHQRR
QISEELGLSETQVKVWFQNRRTKEKRVIHPDDQDNICSQESEDEDEIESRQSSDNEEPITDLPSNSITQPHLLQERDQTSLTNQN
TVSEFDHAPDCEKDKSTKELLKSFQLDAVKSKISVISSIDFGKDSLEKKFGSNLTDS*
>Cmac.rna.tri.20657.1|m.25693
MESFSISNLIKREPVKKPSSEPLDQPVVHSLDKHFEAMQATAADVLGIITDQSELSETRVKRRRSRSSSSDSDFIHPTTIGKSGLTV
DNKDISPTGSTASNASSGSPGVIGSTGASSGTSIGNGANSPGMMGGGDCSLLNGIRRYRTAFSKDQQDILEAEFRKENYVSRPR
RCEMAAALNLPETTIKVWFQNRRMKEKRQKQTLAWPAFFDPYYYGYLVHRGVIPPQPPATTNPGPLPTFDLSNPAAPHPHQP

PTFMPPSVPPLLVPNSAGFKFGHPGVFGHQFPVPGTQAAINSGNVTGLPGLGSLLPLPLNFSGFQSSMNPQTSALGALATLQK
SILTSTQQGLSNISSGPNVNSNAIDAEPNTIVGSKESSETIVEKSPPTNSSVPAVSK*
>Cmac.rna.tri.16684.1|m.23400
QQQQNGNSQQLSQPGSSAASSASAIATTEKQQTVAAGIKVPGSVGEPSATTHHFPRQITSVPSNGFDLAAAVSAARNSQVAA
AAAALQASNSAAAATPHHPHHPAHPHPHTPHPFGSIAQQNPGPGGIPTPGSLTPGHPGLPHPPPPHGPGLFGQPGHPGAAA
AAALFAGHPATAAATGGIHPLLAAAASSGHFSGYHAPDGFGPMDFNMITRRKNATRETTSALKAWLSEHKKNPYPTKGEKIM
LAIITKMTLTQVSTWFANARRRLKKENKMTWAQGSNDDDDDDD
>Cmac.rna.tri.20728.1|m.25727
MTDSVSEKIELESPNLNKPGFSIADILDSKTINIQETLKNLTKDLNIAAAANGERNSKLFESASITSAPSGVVPLSPITTKASQLQSPI
ANPSVALSAFHPLLGYPQAEALLQQSGLLKTLSGGQVAGPSPSAIPPVSGPPSSVPTAALGNPALPPELLFNESTAAAALLRWYS
PWAILSNQGNLGQGAPNPIFQQHLANNLQQLSSGFGSKHANDGIAKLPMIPGFPEMNPSLGSKLVSPATPAFFQPKPPSHIIP
NSNSSSSSTSSASVPFSGSKSSDVSSVVSTISAQNREANRDGIIARVNKPVIRDDNNQATRDSESRDQVSSNNQNGDNKSICGV
DDCDEDFADVCSDLGDDENGGDDSGGEDDGEKRKKKTRTVFSRSQVFQLESTFDVKRYLSSSERANLANSLHLTETQVKIWFQ
NRRNKWKRQIAAEIEAANASAVQVGSNANNAATQRLVQVPLQIYSHHAPGATGPVTHPVLPTPVPSVSSAELQGIKNSLLTSS
AAGGNQAALIGQLQAAAAANYQAGIAAAGSPYYQASSIGAVMAAMARQHQN
>Cmac.rna.tri.9438.1|m.15910
MNRIALIDRAERRRLEKIVQRSKDKRFSRRANAVLLVHKGQSRKHVALLLSAARSSVSRWCKWYEESGIDGLKDAEHGKPAYLP
GDIIVQLLFLLIPWTPQDLGYQRSRWSSELLAKVIQENTGIRVHSSTLRRWMPELGIVWRRAAPTLRIRDPHKEEKLAAIKAALEK
CCAEHPVFYEDEVDIHLNPKIGADWGFRGKQRLVPTPGQNEKYYLAGALNAGTGKVLYVGRHSKCSEIFIRLMEHLRKTYRKAK
TITLIVDNYIIRKSKKTQAWLKQNPKFTLLFQPVYSPWVNKIEKLWHALHETVTRNHRCTEMWELLQRVRRFMEAASPFPGNQ
HGIAKMKQT*
>Cmac.rna.tri.22292.1|m.26333
MSHTLSMAFEEPSNVTHPENPPRSNLKKVKHPKKRASAFTISQILDENDQSEAAEIYPATEHKNTLPTIRTPELESSELEPNQKTN
QLPDKSPKLLEPSIPKVKTPEITVKTNVNDRNVREESKATERDEESLESPDPQQTWELLNTFLRLSQEQTSPDLTTHPDITTESAKR
NIINNNNNINFLDLLEKSSTTNPNQKQTPILPTSFIPPYSIITSPTKAHPNPQIILPPTTFNHPYVGPSQNIPGINLSQILQNPSLLNSL
GQNAHNSQHNQGNPNVHGTQSGVFSSQISNLSPSQSYPNNQTANANPNNYPFLTTQLDFLPGTPEYLKTIELMTSQNQDFG
VMFPGHQHVAGFPHHQQFGGLGFMDPWGLNSALNNRRCRRSRTVFSDSQLKGLEKNFTNTKYLTTNQRNKLAKESGLTQM
QVKTWYQNRRMKWKKQLLQDGATSIPTKPKGRPKKVLEEQEEEEHGD*
>Cmac.rna.tri.12676.1|m.19812
MNLYPSPAEMSYLKQVQQAAAVASQHPAAAYMPTGTAHSMLTAAHPHHPHPHPQMDHLLHSHHHMSMNYPVAAAAHH
QMQAAGAAGIPAGRKQRRERTTFTRTQLEILEALFGKTRYPDIFMREEVANRINLPESRVQVWFQNRRAKWKKRKKCSPSVFR
GGALLTPGSHPMGFGQFAAFGGDPFVCSYNPFSTDPRGWGASGNVGTTPNNSTMASFTFPPTFGNPRHGIATANPINQATA
FGFGVESQHVSANLSGMQTTHPLAPSTPTPVTFPGLNHQAQTFANAMMCASAGGTGAPGTIIETPVSAVAPLNSTSPVPVGA
GGDPASLAALRRKALEQTATMNSFSALSNCGFDLR*
>Cmac.rna.tri.14198.1|m.21255
FAHPPPPLPTLATVQSQQSTQQHVLSSDVTTSTGCSALTSNGNPQGSPLVGIVTSAQGQPQVSMAQSLALPHMNKEAIQQAI
SSSAVAAGMTFESLMAAGAAGGTGPVAGCYPAHPSASSLLAGTMIGGHPGAMAGMTAGGALSILRKKKKKRRRHRTIFTSYQ
IEELEKAFADAHYPDVYAREMLSLKTDLAEDRIQVWFQNRRAKWRKKEKTWGRASIMAEYGLYGAMVRHSLPLPEAILDSAGK
LDQEDVKNSVAPWLLGMHRKSLEAQEALKKEEDDLIVSSDDDDDDLDDVESRNSKISSAKNEIEKLHDKLKLANQSQTRSSLTS
*
>Cmac.rna.tri.8232.1|m.14110
MILNSGEESRSFESPPEFLSRSVPAQRFFDDIQHENSSPISNPGSAPIHEQNPAVNGGVFNPLLPLQCNPVNDETTTPSTPAGIN
TPVATLGVPGLPPSASLLVDPFHFAQNVQLCGGCGEAIIDKYLMNVLDKAWHSSCVRCFDCSAPLTEKCYSRDGRLYCKEDFFR
RFSTKCFGCSEAIMPKELLRKVRKHYFHLSCFVCSLCHKQLATGDELYVTPDNKLICRQDFLTNVVAPNTTATSGGHVIAAPVNL
AEHHTTPNSHTHLHLHPHTPGNSVDGKGDNGGETNSTSTSGCSNIDVTDPVTNSEFHPVNSFQRQNSLDSPQKVKIESNSMEI
SRPEFGNSPSNIFEQNSLPFCQNIINEKTTNTLQNITSNSSNSLTSTNQSISSNNINPVNQHEFCTNHDDKMARVKPTNELFKREE
TDHEASCEMIERNNPSIEQNGNLDEYGSDNCGFNSDDDEKLFDEQDDTGELSDDNKSDYDEADFKNCNSPNKEETEVMGAG
PKAKSQVGSKRRGPRTTIKAKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGARRAAFF
QRRLRGPFEGPGAPGGPHNPGDPFLPHGAIPFFAHQDLTNPDFFPVAPQFSYDFFPGVHTLGQPPNAPPPPPTGGPLGALDA
TEFPMPATLPTPSSVIDPLRIHPDQLGVPGVTGAPFLPATSIHPPSLGGAPNSSISPAGLPPVDDAITGMGSNPGSNQRLILPSPD
IALPPSDRFLSSLTKPRFRNVFEESAAPVW*
>Cmac.rna.tri.62065.1|m.31721
DIRDAQGLVSFPSSNLNETPPGTAASVAGVTAVEKRSRPFGTSSHQTLTRPTAKYGNSSFGPSLGGSGGAGGGPRKKSRYRTTF
TSEQLEEMERAFESAPYPDVFTREELAAKIKLTEARVQVWFQNRRARWRKEKNLRNSSVSVKTGANQDGSSTMASTNQSAAF
GLTDGAMLKPDMLTDLKGSSDCLRESSPNDPTLSNSRPVPRHLPATQYLPASATLTGTPNLHAQNPESYNAAGSSLIQRQQYP
NLNCQPNLYSLSYWYMYQNILQKQKDHRTNSPPSDSMM*
>Cmac.rna.tri.2290.1|m.4767

MKFNSTNASSGNNHCSDVSIAIGASNKKDDYFPTCSNCNLEIKDKYYTKLLDRSWHNSCFVCCACKTKLGSKCFSRDGFIFCQM
DFQTQAWTLCSGCNRSITPDELVQRIHDYIFHVSCLTCSICKAHLQTGDKIFIVHGYPCKVYCHNDYKQYVIGRNSSESAKVMPA
GGQWTSPAVHQAYNQQFPHRQSLPPMAADYGYPQPPSLDNSRGVFRPTPLPATSRNHPQISNVHPQSQPLQPQQQQQPP
PPQAPVTGMSVASPGSNSFPCQPIAQHPPNYKQQPEQPLKQQPPLQHAQPPQINTQSGHPQFHNGHVMQPRYKEEHAMS
PSAPHSARSPNLAGGFEGIQPPFMPQGHHMPVAGNLTNGVTHHGYLQGQTMQQSPAQSSDSNSVDMIEYRQSSQQPLPA
NRALQGSSGVPAASMGNGGGNGVLTNNSSAFSPCPGNNASNVNQQQAQPARGFAPPGTQQQPHQSHLYANGGVPPGVV
AHPKMLPPHIVCDDEKSSDNVSVSSENDLQCSPKSGASMGSSSGGTNSNPNSVSSAANTPPASTQSGGGAATPPTSAKRGPG
NHLKGPQAEFLRSYYLTNQKPSKEEREAIAAEIGVEAKTIATWFQNRRSKEKIKREKAAAAAASAGLDQSASENSQSSSLAHSPS
ASLEPSSHAPPAPLPSSQPGTTANQTSNAPVLTQEAISQSMTNNTNHQANNATASQEHQPASPAPVAHLSESSGYSSYVIKDS
NHPITNDFNMNAQNPQPNETFHHSNFAFAKIESNNEQFQLNQPNFDHNGATGNDQNMEYSTSGHYYVTNDVQQYPTHTA
VPPQNNDASGSNEPYSVVQSAFKSFNTAQPAFTTSSPAQPQFASVVSQNQSSGSTTFVPAQ*
>Cmac.rna.tri.44838.1|m.30329
MSNLYSTEKQSIESDYSDDCDSIINGDEEENLDSDLSEAEQDQEDDERIVKKLKLQRGGDCSQKSQGQKVKRNPFFISELLSEDN
KPKASVYPFAPNCGPFSGGGGVNNIHSGSGGSGSDIVGGGGSVSSSSSASLSFHNFLHKKSRRRRTAFTQAQLNFLERKFRCQK
YLSVADRSDVAELLSLSETQVKTWYQNRRTKWKRQTCADDRLQELENHHCSEPDCPGNLRNNPRPEATTGLP
>Cmac.rna.tri.14812.1|m.21835
MDYNSATSIVNNSSNTGNHYASQISPIDVSNSSITTQQTQQQNNNNNNIPNNSHIVAYPTPKLTPNGTPTPCAQNFQTEQSC
MNELTLNDSINSGLKVEPSAYNDNSQTVHKPILSNFTNSQSYNLSPSYYQTYHNSNSSNSNYNIYTSVNGSAQGSSSFGNFHSST
NPSYNLNSTPSSYLSQQDTVASQPYSKSANYDFGSASAYVTSELLYPSNTILHEYFFPNNPQQYYSYPSGNPNGHAGQGPYANN
VLSNYMAQRNSINAPGSHAHPPTAAVNINVSCNIAGNGPSSLSSEEHSSQQTTPHSGNHSEHTTSPVAKYNAHSSPVYDYMD
ELKQSNFANICNNNPNGLYPTQHSAGTHHDTSPDGTGYPTAIQHLGEPEYMKSSQGLPDVRKMSPGSAPGPLIGGAGDPKTE
MLAATPHPHGHAVGSASTISTANVSPAAAAAAHAASAGAVSAAMSPQFAWMKVRRNQPQHIWTKSASLSQYHASAVGAH
LTAAGCHLGPGVQTRGGRTNFTNKQLTELEKEFHFNRYLTRARRIEIATSLTLNETQVKIWFQNRRMKQKKLLKEGKLS*
>Cmac.rna.tri.15816.1|m.22725
MIDELGFATSLYSATGMQAAATTNGGSLNFPPLNCTTGQIGTANGTRNEISSPLGDLVGNVQNAYSKHSSPQKIKSELSSGGCG
GPVTPDGSSGGLGSGGGPTPARRRHRTTFTQDQLNELEAAFSKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQEKQL
TKAAFVSPP
>Cmac.rna.tri.9888.1|m.16597
QQQQQQQPNVHQFIGTNYGQPAPPPPHLPHQLPCHPDAHFGLSTLAARCEGCGLPIEDRFMLQTTDRQCWHERCLVCAGC
GCKLDQTCFIRNSQPYCRKDYESLISGSCNGCEQKIASSEMVMRVSDQIYHLNCFSCAVCRKVLRTGDQFVLIEEQLYCKTDYEK
QFPQFASAEKQSSAETEASRNLLDSVGDSLSNQNSESSDPTLIDHFFKNENSNHEHENDQLKAENSDNELDSEPKDESDDENDL
ENSKDKNSSIQQKPCRKRVILTTQQRRAFKAAFDLSSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKLAKRQQQQQHN
HLETGLGSPHLLEDGQYSVFIPNEAGFLQPQLDGRFSSHLAQSNHIIHPEGGLYSLPPGLNDLYPVCSGPPIAGQPSENEFINPTS
DFPQILDMKESEIYSSMKLANQIAHPIDKLYSMQNSYFTTS*
>Cmac.rna.tri.31677.1|m.28661
MADMQSFPFNPEQTACICQVMEISNEMKKLEKFMWNIPTMEIYQQNESVLRAKAHLAYHDQQFKEVYRIIQSSNFTIQHHHS
LQQLWLNAHYAEAENSRNKSLGAVGKYRIRRKFPLPRTIWDGEETSYCFKEKSRNMLREYYQHNPYPTPREKRDLAENTDLSVT
QVSNWFKNRRQRDRALENRS*
>Cmac.rna.tri.17874.1|m.24213
MVITPGLHDSSIVASGLVQPIQQCSPSYNQSTTGTTSGLSGLNSFGENLSSATFGVATSAAPDYNTSYLASTISPNMPQPTAGLQ
CATATSAVSSAGASAAGLQALQHHSALGNGNSSPYSAYNKDLGAMATAPYSTCLTPPQAGTTSGGSGWLISPESNLSWNATS
FPYQNNPQCLIQSGTFFPNGAGASSIYGAQPAVGMGHPLASSNTWSHLAAAGMHPFNGYMTPSRLDSPPPNVAAPPYPLST
TPTIHSYYAAAAAVHANQSLLYGRTLDETALLDCMERECEETPSNDDLEQFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQT
TICRFEALQLSFKNMCKLKPLLTKWLDEAENNNGYPSALDKIASQGRKRKKRTSIEVSIKNALEHNFHIQPKPSAHEISALANSLQL
EKEVVRVWFCNRRQKEKRMTGAYLDDPTGQKHPDDLENGVLCDGESSSDDEDLNLREDEQLIEQNPNCNPNNGISSAIDAKL
YQNRNPAFNELNCQFSASEAIPPSLIDPKNEVSLASYCATQGASPHSILGAADKLASMEHQNVAIPSNSTATPLHVNHLNIICPV
PSMDLMKDQPKLSTDTYQQQQHQSAEKERLILQQISEQQSNFNCHQFGYQFST*
>Cmac.rna.tri.3427.1|m.6657
MNGSSAAHLGINNNNNQSPVFDYNAKMMHSISHSLIGSNSGGSGGGHLIAGQHHLSSPINGANTMGGFPQPPHQGGQTTS
PHIPLATNTTATPQNFATSTNNPMKQESSSPLSNSLLTTAQQLHGSDSPNAGAGSIPTLPGASQLSSGGSSSAGIQGSAAGSSQ
ATPVNEVTVAGSNTPLTTGGSSLGVTDHQQFENDKKVIFNHPLFPLLQLLFDKCETATNSAECPSSASFDLDIENFIHLHERDRKP
FYTDNADVNDVMIRAIQVLRIHLLELEKVNELCKDFCQRYIACLKNKMQSDTLLGCGAGGGPGALSPTSGLHHQYPLSDSPNTQ
LQLNCGGPVVNVSHPNAALSAVVTPGAGGAPGPHHSNAPPNGHNTAAAASSAAAAASMHAALYPQLSMFPGMAAAAAAA
STFPGYHSPQGHAPPGTPHGGPHSHGALHNGAAMLSGSTPISQVGLQHPGSVNGGGPPGVSPLHHHAHPGHPHHPHHLLG
PHGGHELNPDGSFKKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQPMLEASNP
ESVSKSKKNKNNRPTAVNRFWPDNLAQGQITNNFKSSNSDDCESPPPGGQHVNSSGGEGGNSTNNNNVSPSNRSGSTPQD

GMLLNGSRENAMNNLQLLNSPSALSDCSDDRSSNEVGMSPPTVGVLNGGPVVQSTGGVVAGGGGGGGGSFLTNGYDAYS
NSSASQSQAASLFAAQTSAI*
>Cmac.rna.tri.8612.1|m.14688
MVTFDLEEEIFKNVDQDPKESVPKLLKLIEEKVNASEAEARNMLIYSQQAVIRSLHHKVGMLIELKQDEKEAFIDLPSLEQKVEIPV
KQLVDLETSLHDVAQNTMQPLNVNPLIVLQPGDFDPSQGYPAIEPSTMQQVLILAPTRATPVAKGPLQLKNPQISVACPVNTDI
DPLNPNAETRSAGLDEKEIDVIAGKYIIPIPASDESSKLEELESFAKDFRTRRIRLGYTQNDVGLSLGKIYSSDFSQTTVSRFEALNLS
FKNMVSLKPFMEKWINDAERHKTLVELDASNLSTEDQKTLEIFNNSTDNLGQRRRKRRTCIDVRTKKALEQYYQDKPHPSNEEC
CAIAEALNQERNTVRIWFGNRRQKDKRLHVNPVIAAPTASFAAHQSSLVNYTANSDLMAMEPTEEEHPMMMVPVSSGGTV
ASCGSSNLVLQPSKSSAFSAPPPPVLHHQQLSATASHIASSADAMSNNSSSFSVAFTHPN*
>Cmac.rna.tri.124781.1|m.34394
SNSARVATSSVLASRKQRSIFSNYQVATLKHEFIQNKFITQAQRRDLSQKLQMSENQVKVWFQNKRAKLRKFRSRQQQQQNIR
SLTRTNTFIESLLNNSV
>Cmac.rna.tri.20322.1|m.25553
MFHLSDQNSFNPWSKIIEQSSNPNILSASYDNSQIDGTNTPLTIPPSNIISNSYKISDIIHQTPALETVGNINQIDNITPSINSNQLN
KQTAETNSNPVGQNLGTNTPPAKVRKSSKRKMSLPQTDFPTPPAVKALQKKIIDKVTGNSNSQIHHQTSQNIKTSPLYTNIMKH
YPENYSLNSQTNLQKDINHLISFNDQPLNYEVKQEVNHQDPMSQFMMYNNQTYYNPNHNQYYFNNQSHAQFMNQYWSQI
NSQQGFGFNTSQNTHAVTAKAEPINIEPTFQESVVPDSQVVPGSAVKASPLDFSLDKRSATFAVPELIKSEAQTLGIACSSGSAD
VLVGDVNVSGQSQSAHCRRLSTNDLSEFLVFDQTGPIESVNRTCEFASDVATEVVASATEPSTTTAVSHGMRPVQTVAPFKND
SAAQSSAHAPPHVLTSKPILGNNSTLVNTTVANMTIPQVPNYQRVQGPQWFQGSRETTGNQFTAGGRNAFQNQGFQSSGY
AHWQQVMQHMHQQQLQGHATNIRHQFPDFYSQGAAHQFNYQTGQHSQQFKPAQQQQQQQVAHPSSDSPAAGTKGK
RIRTMYSKLQTQVLEWHFKQSQYIGLQQRATIASHLNLSQSQVKIWFQNKRAKVRRGDKYVPMPTTSTSNESDDVNTEETNA
TDATPESATIVASEIVESEMVAKENVDMNSTIVALTDIQEASDCHTELQQIEIIEGSADENQTNQIFHDFANYLFPVTTEEQMEQ
QIQEIHPNSLSDGSNITYYSTQELAANDEH*
>Cmac.rna.tri.2258.1|m.4721
MILGGAYDNPQSMDDQHRFALAFLHQQQQQQMNAAAMHSAAGGVDHSGAAGVNMHGGAAVSAAGGWGHLAPQG
GPQSSVGLADATGGQDTLSVATGAGTGGGGGNSAGLLSEETRKRQLQEILQQIMTITEQSLDAAQARKQTLNIHRMRPALFS
VLCEIKEKTGTFLNMRNQNDDDAPDPQIVRLDNMLAAEGVSGEGKSPTGASTTGGAGQPDNTIEHSDYKAKLGQIREIYHTE
MRKYHIACNEFTSHVMNLLREQSRTRPISPKEIDRMVAIIQKKFNAIQMQLKQSTCEAVMILRSRFLDARRKRRNFSKQASEILN
EYFYSHLNNPYPSEEVKEELAKKCGLSVSQISNWFGNKRIRYKKNIDNPGSYNPNDILGPPGSTPGAMGASSSSGPSSGGGASA
SSGGQPVPTGPYSDQDQFSSGQLTPVGSDDTATVAALAQQQAEAHQLFQYPPVPQSADSAMVSSAAAVAAANSSAVVAGS
LFDDPASSASLANWNAAHAAAAAAAASAHQSVQDAIARSANNPSF*
>Cmac.rna.tri.16615.1|m.23347
QQQQQQQAAAVALMQEAALTEAQQQQHQHQQTMQQLVNSRPDLFRSINTANGAAPQIPLPQLNQQLPGHPAPNHPAGL
TAAMLAGGSGVPGSGLGVGIVGPGAGGPGGLGMMELGHLMTSPVHLQNVALWQGALMRASLARRKRRHRTIFTEEQLKEL
EQAFLKCHYPDVVTREAVAHKLSLKEERVEVWFKNRRAKWRKQKKDAAEIRKHSKAAAAALKLRSVSSTSSKDSQNSSSTNLQ
QVSPQSSVSSISSTTNNNNANNGGNSASLTQVP*
>Cmac.rna.tri.6040.1|m.10870
MVDSLSPSPLTSTNTGGNNITSSTNASLTCNSGNSENTSTKIETIALNNETTATNSANETQIMQNSDFQQSHYANGQLNDSTG
DILAPKKIKMEPYETKSFENTANIDSETQQQTMTNGENTEIMQSENCQFSVDVKETDSANCGVCFEGVISFNSEDSTCTLSVRID
SEGIFKVVTANVFTFQDFLHASGNDQLKTFSEELQSISKRDLTSELNFLKQNTDSQKLKSSNLNFMLCFACNTSYSDVIGLSEHGK
LCHNCSFSKDEFSFFDESQCSVVFQPLCSTLENSSIAVLCLDNIREVSLESLKIASRHLIQGNSSGTKSEDDENSSDILTSLFSSESGTT
SNYTNASVDFNSTSNNILSHMAMCHSRNSSKTLKCPKCNWHYKYQQTLDAHMKEKHSESINNCDYCTNNEPHPKLSRGETYS
CGYKPFKCEVCNYSTTTKGNLSIHMQSDKHLNNVQELKRNNGSLEETTNANSATISAVTAQSSANQNAQNLVDTLGGNLQSA
AANFFKCPACDFQTTNAASFKLHVAADHALLSLQTGANSSAATDLASALITQQQQTTNDILQSLILQRQLEAVAAVTTNPEASSL
FVGGLEGILPQINSAAAASASLNAAATTNESGVFSCGICNRFCSDDAHLIQNHALRDRTSMTSQDYSTVNNSTFKCLLCNYSTNL
KANFTLHSQTEKHKQKVQIFNHLKEGNENLESIFAIIHQISPIQITCNACDFQAKSIGQLNAHLTSDSHVLALKLFEIFTSLTSKPGA
KFNCKVCSSGFMTKSEIISHVQQAQHCSKLRSKKLQSGNNSTENSWQDFFEVSSENSMITTDRTSISSSVATKASEIYGEAFNCP
LCAKPSKTRDLLRDHILQDHRVSSDCIEALLAAVKPTMSVKPVARAPRVVKSPVVTSDIVSDSSNVGSSSVMSSALSVASAQSLFS
ANSNVGGLKYRCHRCRSEFNHTLDLYNHVKSCTNYLCSTVPDCKQSFPTHNSLSNHISAHPDQDSSLPCWDPKCGQNFLSEEN
LIQHFFDSTTHLPPGLTCSTCLTTFLSESTFSAHILNSNCASSEQNSRGSQNEFKNDLLDSFPIISQYSTSSLSSRPFQCNNCNFAFIC
QSNLQMHNNTAHKQPNLTGKTGVPNSENQLANSAESSVKTEKPFRCDACDVAYSSESTLKIHLASLLHKETINRQSKLSANQN
TSSSPESNSNEQQSVNPIQSTSEETPGCKLGSGDGLASVFSADNVAGISAGQNLAENLDEAIRNLAQQQQSSNAMNSSLAAM
NLLSNMGLQVSAQDLKMFQDTMAASQQLQSSSLLDVQTQMALLQQQQFLLQHQKLISSLMPFYNNMPQQQIASQAQLTQ
DLLSNSQSLSTNSKLSELFKISISGINGSDEQQSLVEAANALDQINSTSTESDEANTESSIATVDGNQQNPVDSSVSTKTRRSEGSY
SSSPNETVNVLSSKTESSTPIKQSIADVTSLQIAEEQSANINENGLVGNFQDDDDDNMFYSDDFPQSESDSGNNSDSESQSGSS
FAQNNDQLNQQTNAFILQQAALYQQLFMQNPAALSAILPAAALANSNFGSQLFANNPFLSMIASAANGSKSPTTNTTSSTSSS
TETSQKQPRTKISEQQLEILKQNFDISNLPSEERYVAISQQTGLPVKVIKHWFRNTLFKERQKDKDSPYNFNNPPSSSNSANSNSS

STSSATQQSNNAALSLQQAIFNNMQSLLPGANAGTTGQFDPSKFNMGALCSSSSGKPPSPMSIFLQQQNQQHISTKRATRTR
FTDHQIKSLQDFFDQNPYPRDDDLENLSRSLNLSPRVVVVWFQNQRQKARRSMEITTGAETTTGSQSGQSYPNSVSSVGSSA
QDTTSSLGSSSRVDLNPRSTLDLDATKLVQEVVNMVAGNHANQNSYNESVINEHEPEPTTAENFTTVVEPQPAPNYDVLNAV
TSSASTNQETGSGKLRNREDSGDSSKLTIKEDIETNDGGGHSSQDDDEFYYNCSLCPAILASEEELDSHQRAHIEQVALTISGLSA
QQEQQNSQSESCNDLAAMQRMSDKYAALDNVANSNESPNVNIFSDAQSQFDANSAGSQEYFRRMRTTISPDQLEVLYTKYG
ENSNPSRTEMETIANSVGLSRRVVQVWYQNTRARQRRGQCRVIGNTVYHCKCPYCDTICKSKSHLSQHVQKEHSDRAKNAIQI
APSTLDASGNVRSTAPPNVTITTKPSETNGGSGIPNATVGTKTNSANLSNNDPQPLESLVASVSSIVNAASRNFGASLKALGLGD
ESSNDAKSVTSLSDFVNQQNSSIVNVKSFEDSEEVKLHEVSFYSDGSNDDSSALSLTTCNSNNISPTGTSANSDFLSSTAANILGN
ATAAGGGSNPATSAFSLAGTGGMYINDDKKRHRTHLSNVQVRTLKAMFIDFKTPSMGECELIGNELGLSKRVVQVWFQNARS
KEKKGKPIVGCGTDLSSQLGENQVEPGKNSCNYCPELSFPNSSAYRDHLFGSRHVQAVKQSVSAKVREENGYTYAVPVPIGYG
NAVSNNNNYHRNEIMSQINTTGTNNCSAGISTISAPMDLTTKAV*
>Cmac.rna.tri.11709.1|m.18724
MMNSIGICPEATTLLHPISISTTANTNHPKQQQQQMNIQTNNMTIPGNNALKQQNLPLSIGNQHQQQQLAEVSPVSSEVSEP
ATPSSSPAKPNSAGSNSPVKPTSGNNNNSNSAQSVVPKPKRHRTRFTPAQLNELERCFNKTHYPDIFMREELAMRIALTESRVQ
VWFKNRRAKCRQQQQQQQSQGDKNDKGTPSPPDHTASGEMKKDCLSSNDQHAPLSNNRADSPSSLKGTNDGLPVIKHQPS
AVELCDSLLKKSTEGHLAGAEFTTLGSVIKQQPQDLASPSESEVVQGSGLVLGGLTGSKDMSDIASLARSESSPFLPSVSSSFINIW
NPAGAMPNPSAVSGGISGGYTSDTSAAAVGLGRASGIPSYNQTTQNFGYPPAGAAGYYSFDTSYAPYFNNSYPSSAAALSAVS
RYPYSVGGTQQSEGGAGSGPAALFNPISTSAANASAVSDLKQEVIRNQLNDLPPVWKGY*
>Cmac.rna.tri.8779.1|m.14925
MQNIINHSQLSPSNREPHLRGESANDLLMGKTPPGFPYHHPDFPQTAGAFPYGIAGLTGQHHQASTIQHHPSHANPAAGFAP
NPYMQMAAASGHPGNPQHAQNPFHGAHGNPVMNYAMAVANQDPAFLQHHYRSQAHAAAAAGFPGYLASHMYQQQP
GSPQGNPYTGTAAGAFGQANLASRGGAALGYGSPPPNSQITDPSPRGDGALSGEKDIDGESPKNGTTPGGGNGNPNNPSSV
APNSNLSILSVKGGKKLRKPRTIYSTVQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKMKKLMKGNPGGQ
DANFDPAAMSPEDVIKDEDCDDELDQTGGDLNIDQNQGSNDIINATTNPLHQVATNHMTPNVTLANPATPGQPMPHVPPT
IGHPIPVQSPWNMTMSHPGSNPGNPVSQIPSVPNNSVPPTVPQNPMGGYLDSSNIAEMIANSRAANMDPTAPPNIKVEMP
TGLDMTQGSMHSNPLGIRPHVPVMTLNPAAHHDWYSHPKSGLPAHLSHSISQ*
>Cmac.rna.tri.8155.1|m.14011
MHTDPKLGLLLSQAHGTPAAFPPSAAALFAANPMIANHPQLLHQLSAQLAAAQMNNPAGPGHVSGPMHPGAPLPHPALM
GFNPEQVAQVCDTLEESGDFDRLSRFLWSLPPHLLESTMKNESILKARATVYFHNGQFRDLYVLLENNRFKKDYHPKLQAMWL
EAHYQEAERLRGRPLGPVDKYRVRKKYPLPRTIWDGEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGNW
FKNRRQRDRAAAAKNRGTALNNFNSSDNEGSYDKIRSLEIDINDDEDDDDIDLDVSISDSDDEDRENSSDLKKSPSTTLTPDSTS
SKTKLDGSSNKVCKPEPCDLSGAEHTTDSSSALKLKSDNYSPAASTTSEVSPSNSTKSGSSVFPSPTQPNGGNPLSSYSHAAFSAQ
NESLKRELFSLYTAQLGSQHPTGIPQTFPGLHTSASPIHLPNLPPSSLSSAAVAHQQRLLVPPAQLFLQAQLAAAGRLNPSHPGS
NPHSFLNSTNAHSLLRTNQLLISSASLSSSQSKSSSNNSITAVSTTTSMTTSPVSSTVVNPTTTTSPTKPTVFSPVRLCNIKDSK*
>Cmac.rna.tri.25423.1|m.27332
MKKSKFSIADLITPENTEQTIQFLAQRANLLNAINFPPPILGTIPNATQSLVDNLSVKVNNLDSSNLNTCTSLLNDKLLLKTVTNYP
NSNKKPDQSLNESIKPEQSFDTHSALSTKDSEDSSSKNSFRLSVNNHSFQLEENVGESSGKNKFNFSRPARPNQLKVKSESATFT
GKSVKRSLEDDSMLSGGPQEIPDEFQENSSKDESVDMEDVAFSSDLGLLSPGPGSPANSDISDHSPTSPNSATSCSTSAANRRA
RKQGVERKPRQAYSAKQLEKLEEEFKNDKYLSVSKRLELSVALNLSETQIKTWFQNRRTKWKKQLASRVKLAYRQFWQLPQTP
GGIAFPALYNFPTPQQPSLPGGPAHPFASLLPQNGGLLPFGIFNPTKTVQSDATATSLSLQQLAQLVGASNPGSLGNLMTANM
TSLGLNTAQGNTLTNALSFGKLLGANPNVQSRVQDQFGLGSGPDGPTVTASTANFGTDSSDNVSGLIKAENICNDDGDNMQS
NDVNSSNVNLTDDGNDLNVKIGGDVNKKHDFRCSNTAE*
>Cmac.rna.tri.10188.1|m.16925
MDSSGNNWNFNSGMPNFMSNAAAAASAGSMQYTGNPAGGSSFNYPTPSYMQQVAAGMIPSAVNAAMYSSSPGACGG
MSGGSSAGTTATHGSFAAGNSSGIVDAGFPSGASSAIHAFDAFARQQPTASNWYNAANNPGLAFSNLMRSSAAAAVYNPM
TGLALSGFPGNPNDPTKFFGLANRRKRRVLFTQAQIHELERRFRQQRYLTAAERENLAHLINLSPTQVKIWFQNHRYKTKKSKQ
ASPNSSSNSSQGSSSTTQDHSNNQQRNSQQVNNNQQPINNQQYIQQGNKPQIKRQESDIIIKNPKVEENNMSPILNMTLNA
KPENYLNDVIEPNPALPSNVIGAESDGQMDISANSLATANQLRSVAAANHNSLLAFNAENLSTSQYGPNPMF*
>Cmac.rna.tri.21745.1|m.26132
MYENNNNPSPIECSPGSNSQPIPNSLEINPTTNSSHLNPNNSLVSSTSANQSTAQNPSVGGNFYPQNPPSHHGGSVDYNLHVT
SLTVNYPAITAHPPPAGNLPTPAAVQETHQNPLNSLNPTNNLGNISGGSNNPLPLAAAYQSNPYYSANGIAAELGHHYPHPLLP
AAAATLSTYSPYYPSMYHNQSNNTNSLQHGSTNQALQLTANSASVNGTNPVNWHFGFLPTTFDTRTMQSAAANLTHHATSA
GSLKPGAAHVNDLSSAAVQYGQFKFQGSNPVLGSDSLMMVAAGADGGMNSLDLEGVGTESLTTCHEQISPASHVTSNGSK
MFAWMSKRPPTDCKRTRTAYTRFQTLELEKEFHFNRYLTRRRRIEIANLLALTERQIKIWFQNRRMKWKKDNNLKSMSQVDAI
STS*
>Cmac.rna.tri.11982.1|m.19035

MSNLKNEHHHFGLLPTLSQHKFDFGTLVDQANSGKTSLFAPIYQDCSGCGDKICDNSYFSTEDGLWHKRCLTCSVCAANLSDA
QSCYSKYGKLYCKRDYYQEFAGSKNRCANCQSAILPTEMVMRAKHNAYHIECFVCAMCEKLFKPGDQIAFHEDKIYCTLHMTH
LNSSKQTQMQHNQSAENQDAVQQLTKLDLNHPENDLESSAHNHHKALAKTEPTNERRSLQENNDIGHSFVQQNTPVPPDVS
NLITPDQIINNSTTNDPPLSSAAPMINQSPPVMINALPETNPAISTQGAMFQPATPYPTSGGVTMTTNYPSNVVHPAVYSPHS
QPPHPPPPNVLATPAAVALPPPPPPGHFHPHPGFDFLKMMPPSGSAPSSKGRPRKKKHIEDQLMAAYTAHGVLDAEAGLAST
LFGIDAHHRIHHPHHSHITTFSHGHLITTDTSGRSKRMRTSFKHHQLQAMKQYFNANHNPDAKDLKSLAQKTGLTKRVLQVW
FQNARAKFRRNLMREQNQDMNSSLSEKSLDDDLLGDELSRRDLCSMNGSEGGTDSNFGELDESFSDTGDFDDESLMSYSPDS
STESLGGGNGADNVEEEDDIPRE*
>Cmac.rna.tri.15029.1|m.22064
MPHKGHSGVNQLGGVYVNGRPLPDATRQQIIELAQKGARPCDISRVLQVSNGCVSKILGRFYETGSIHPKAIGGSKPRVATTDV
VRRIAALKRESPSVFAWEIRDRLIQDGCCTQENVPSVSSINRVLRSLQSEKLKSPSVTQDMTHPCSVVSMYDKLRLITPHWPPHP
HTGAVAAGHPNPWYNGFAHPSASATHGGPHTMHPNGFSTAAAAAVLSASMGAAIPNHVHHQGHGHDPSFMTHLAAAHQ
NNMSIAGGQNPLNLDTANNMIELTERTMRLSSLSATGGGEGLKGQQISPSDQISLHNGDQQTQITPITDTKSPIANAFQFGNP
NRLLGSQFRDSNETTTQQVARPESHSPTSPENSLGSNEEEDLSSELGGEDGKKGSKSQRNRTAYTVEQIESLEKEFEKSHYPDVY
ARERLSNLTGIPESKIQIWFSNRRAKHRREEKIKTNNNNNNSMDILPSAAHVVTKLEMDITNPTRDNSPPVSALYGGNANVSM
APYFTPYSAHTMHGALSNILQNPAVHHNVPGAHHSLSSAISYNAAAAAQNMPSPNSFTRSATTGQPLPASLPGMSNCSSAAA
SAVPGTSSTLSPMTAAALHYNQSAAALQHNSTGACYYNPQQSYFHAAAAAQAATGTSVVGGGAGGVAGAFPAGGTTATAA
TQPTAHLDHTGTVAGAAANASAQPPNNPAAAQYWGGGLMPTTTRF*

9 – *Diopisthoporus gymnopharyngeus*

>c20548_g1_i1|m.34566
FKGTEQEDIYHTDLVDNCQHQKTNRSTDPIPKAKKHKRRVLFSKSQTCELEKRFVRQKYLSAGEREELAKEIALSATQVKIWFQN
HRYKAKRTGLDNPYPHFGTTANTYNPALEHAQLPTWYNDMAVTFQDQMSSEAYLDLIDHPFTSPLIRPHLEANHRLHLPRCTR
MSPVETHHHSELSETSCGSANLRSGASLNGAVNMRCGANLSSRTNLNNGDTPANLNHGAKFNNDIDLSVGANLSGYMNLSA
HIAALAALHRIAPHAVPPHFGAKSHEELALELLQADNSLYNEDRE
>c13488_g1_i1|m.20348
MSEKSHQISLANGASSNSFLMANLLGEGWSGMQETTTTGGGEMSPDSGLESDSGSYSDIENPEIICQKVCKSSSTLSNIILGEDG
KGKKQRTTFTPRQLRKLERQFDKQKYLSTSERYELADSLKLTDSQIKTWFQNRRMKHKRQQQEQLYPFYSGEGTFPTDVQPPM
FIPITPALSVFNGTAPSNNSPMSYRSDSTRLTSTSSRPTCTSFINKHIGWNQSMRNFIPTTRLGQQMAYSYQQHSLNASLLRSFM
YPERRIKSTQHLPPSTSLISTLPHFQSFPQQLSNMQYRSPLIDQRNQIY*
>c9874_g1_i2|m.8171
QRRVHGAPGAALMDFNTMAVREKVMEMQRLMIPFSTSGRAAFPQKRKKPRTSFTRLQICELEKRFHQQKYLASTERVSVAKN
LKMTDAQVKTWFQNRRTKWRRQTSEEREAERQAASRFMMSIQPDAI*
>c13968_g1_i14|m.23127
NAEYMGCPHNTPSSELTNSYWNKKHEEKFEKVSPMNLQENLSHVKEIRHFSKRDNVPHDFRNALHSQDNVLKRESENTCLKQ
QNEANEPRRMIIKDSNGVTKELVFPRGLDLDRPKRERTSFTVDQLDKLEMAFSANRYMVGHDRAQLADILNLTETQVKVWFQ
NRRTKYKRERTQEAGSDDRVTLDGSGCVFESYNEEHSMSEHQGYTLKGVDISPRPSSKVDPERLALCNNLLEYLEHASALRVLAS
PLPQTGRDQMRSIFPFETETLNGKIVGNNWMKERNLDDDCLKITESEQNSTLFSPSLVSPSPLLPPAGHSREARRDFSRRLSPVA
IMGWQLAGFSPNVISPSH*
>c13263_g1_i1|m.19242
MAFTMQGANEDSVSESMTSNGVLPINYHAHLLTNFLSNAGRAIASECLASSVPSVGTPTQFGAEEMLIQSWLQNCSAAKTPLS
YSVNPLYSSLLHKVRAQSLAQARAKSQQECYPYLDSVLELLNNRPLDYDQQPHQQKAPQTLSNTEASRAPSLLRALPPPHPPCS
SSCSPLTTIEGEDIHTCNKPSPAGDIYNHAVSMGPQMSTPNVYLSMAPTSTCNDYSSGIGAVSSGENSENAGLISLLDKGCDKRR
RTRTNFTLWQLEELEKAFHTSHYPDVFIREALALRLNLVESRVQVWFQNRRAKWRKMENTRKGPGRPPHGAHLLSCSGNPID
EAEMERRQQEKIAKRKLIKKTNGSSPDQSRDSSNAAAIDSQLNVVTGDYESLITGIEPSSGGSLSSATSISDSLTTCSHSSSFTGNLS
LDCDKDCDTTTFWSKLYSEYAEKSSLSYYHTQNMSSELSVPIDLTTQKSEEIMKSSIEWPARLLLLSGASSFPVDSSSCINGTPQIPS
AFSGALLPPMPKLSHATSSFLIKDLIV*
>c5826_g1_i1|m.3314
MVTAVDSSSLIGLPSDLVQSHTMLPPTPALHHTPSSFTSHTVTPPSHPVTLYSQRHQLSLLDSPYFPGVPPSVCASCTKPILDQFI
MRVNPDLEYHTQCLKCVECSMLLDERCTCYIKDGRTLCPTDYLRLYAAKCSKCQRHFGKSDIVMKARNRIYHMECFQCLACSR
QLLPGDEFALRHEGLFCREDNEVLERSYLHQQPKVSPHSPASRHIPSKDMGEDISGLDQSLLTPPSSPSGSMDEKLSDETPSSGG
GKHCKKGGGSGCGSVKDGKPTRVRTVMNEKQLHTLRTCYNANPRPDALMREQLVEMTGLSSRVIRVWFQNKRCKDKKRTIA
MKQIHLLQAEKNINSMSGVPMVASSPVHHNSDRVLLDHSTYPGFLSGYGAENSLMSGAQGELERFMLNFDNLADGSVISDGL
ESPGSDGSFSVMSD*
>c22302_g1_i1|m.35200

KEHLDVISSGKERFESIDRPVHLSIETRHVETHTEDSSSIDSGDDSPKSELLTPANKKHRRNRTTFTTYQLHELERAFDKSHYPDVYS
REDLAARISLPEVRVQVWFQNRRAKWRRQEKLEQSQLKMNTDFPMASLGHQQFNKIPRSMDMLNSVAIGASPLFPVNSQSL
SAATSLIGGLSSQFVDHFSVTGQAASATQQSLAPSLSPSFGASSLLSYLLCCQQLIPTTIDAVSRQTNQGCGTLCATYGHTNASSF
DKHENSIIGGFDENTPKEGLIDKSLNSASNHIYKLDTSQSLGLGETLL*
>c14546_g2_i1|m.26725
MLPERPVVQTSKEPTLLLNSREAVDEFLDKIRVELYRAVDTMTSTASAFSQLTSSELQVLGDAELAERKGVPDASSKTIIATQLQM
ISQLQAQVQRLLDEKQESSTPSNNSLATWGPNGTVNLPQSTLDEFESSLHNSNLPLQLHPFVLVQPTGEGTNEATKLANTNSLP
AVLQPSTMQQVIILRQSQLKPVLTGQLASGQVPTVLQAMPLNHGEAAANSSQADNIQAEMELNLIPSMQPPQEPQQHSLQL
GDISVADEPTQLEELERFAMEFKTRRTKLGFTQGDVGLSLGKMYGSDFSQTTISRFEALNLSFKNMCKLKPVLDRWIEDAERHI
QKLKSGQQHQLMPCVETKPRSFITLRPDDYASGKRRKKRASIESTAKIALEKSFQINPKPSTAEMNVMAENLFMERETVRIWFC
NRRQKAKRLPGDLETTEMEEDPLTSDTGQEEPQRGDVKCSVTSED*
>c14505_g1_i1|m.26402
MPRYKFQTPQGSIIMEFEDTTTVFQVRTLLSQTLKRPAIAIRCPVTNLSSRGDGTSSRREWMSQLLDSLRMKEVDKRTVLIVECE
PEPRNEPDVKLITPSSKNMVPQEDTSISPSESDIEVSEMRVNGQKSYSEMFKNSKPSLANGNEAAPNPQYFSTTEAKQSSGNDC
TKTASSKCTCDKADSQEGLNVAHSVIIPFLSLFGSQTSSAASTHGSTVANHPLPFNPILLHPSVMKYCIHFEIEKFKINPHIFSTAAL
GADSGSLEETLALHPGELLDSSGKFLPIISLLMKSEAERGQIYDELRNKLRASTGVQGYLLQSTPATSVPWISENHSGTSPKPSESK
GLSVDEKSSFPEDLTSSVADSSLKPLMSSTPPPPLYTVEENRELLQQAGGDSLQTSPSTNDLPLIQGTSNGLNHNVQFQQASVA
NALLHQQAIAAMAASSNLNTFLPQATFPGTFTPHQLAAFKQHILQQSMAAGSMSACPPPTLLRVPQENHAPLQQQPLPRVPP
TSAIQDGTNTSHINNIPSLPAETTCSPNVVQTLNPAVTESPLHSARKLGLNHSFVNSSVNQRMRRRTTINKAAKKAMIDWYLAN
GRYLTNEGKIHLSTSLKLTEENVRVFFKNMRRIEKKAEVIGVSLHDMMGVVDLGSEGQQHQQQTSSSTPNWYQQHLSLFSQG
NHLSIGSGDFLSSDRDFKGSQEVSQLMGTDDDELMLEEEEEFSGEDDSSLGVAVGLIKASHSPPTLVVKQEPRLSSMDSTI*
>c20690_g1_i1|m.34655
MLAVEHVMAPTVADDPSIVKPLTPNYLLNHSHDRTALATNHLSMSPWASGAPHCESCALSQQQYSSHAFQDIDAYHEMYSA
MGCVPHTAPGTGYLWPSVYYNYSTQAMMSAQGDPYQQSLLTNKRSSPFISPDFQSSYMYPYSSSLTDAATCQLPSPCSSTANI
AQKSMKPFQNSLENETIVDHDPIIAVPTSHCSPIKVHSPDTVTRLSFTDSQTPTWSRELSRANSEASDDGDSNSSTPIVKEVSSST
SNHDPTPSWLQKNNQRKKRRPYSKNQTLELEKEYAYTAYITRERRLELAEMLNLTDRQVKIWFQNRRMKHKKMVSGGNETPL
LFNHVPVLMALQHMGAEDRYMC*
>c11160_g1_i2|m.11186
TQSTWGYSDMSYLKTAGVGGHATVPHAHHYSAMLASAGHPTHHPGDHLIHHPMAMNYPGSMAHMQASSTMPSRKQRR
ERTTFTRVQLEILEALFAKTRYPDIFMREEVAVKINLPESRVQVWFKNRRAKCRQQQTNNNSSNQGNKHGGNNCQEDSGKSS
PTMGDHEVDVAPEKSLTISTQLSPLAKGVNLIEPIISCAKLSPKSVGETGSPKQSHSPDLLSLAVGHHAAGSYHHPAGSTTAASF
WGYGDPLSSTSRGPGGSSYSHYPSYNPYIPSHPAVAAPHNPYGVVPPPTGPQGAYIDSSCYAPYFSANSYAARMYTSAYHHAV
TPPSSNSGGFDPFVNGRLSADSSKIMDPLSAWKY*
>c11170_g1_i1|m.11217
MACKRKGLSFSISSILEEDFCSSKKIKSIESEYSPIYHNADAMKTKDLISFDWNENYRRQSEKESLSQNESLIVTDSNIQEESNVIGK
EYYSEENICLEEEACLEKNSSLEKGFFLENIIDEEHCIEKDLCLENKTCLEKENSCLDCGSEDFCSEKDEILSEGDEVISGEEEEDEEGR
KKKTRTVFSRTQVQQLEGTFDAKRYLSSTERVTLARSLRLTETQVKIWFQNRRNKLKRQIKTGLENGTFCTGSGPAPPNVNAALL
NLKPICQVSTPYHQSTSKHHPSSPTSAHPAAAAAALAAISAAKTGLQHISPVSQHNNISQHNSLTSQSNNFSPHKPPFQLSSTFT
HNPPSYLDKSSKAPTIALQPLQNIPVSTSSSLPSMAPINSVPLLHNNQPTLPPPIFPQWSQFSQNSYLSSVANANYPFFSGLVPPH
THPSLLVQSSSNSTTPFNHTSTPGVPATARLFTGTTYMSSLSGLV*
>c21845_g1_i1|m.34993
MTNSVKSSFLIDYLTSPLYVQQKGSLSEVTGEEARIHREASQVHTEKVRHKEGRTNFCSIDSMVSDSVTPPISQTNNQETTSTAA
MPSFGQQQGELSMLTADQRASMMAVIIMETGRMMAGGGQVPGYQQTTFMQQQQTQQHSQLIKQGNIMVPPNIWLPSP
NLSLDLSQQVLLQALFSRNLPAPGMGQGHEKDYPEPNEGQDDGEDDDDDDVIGDKEQVKYPGKALSQSFEINKVSPRPDIAS
GTMIPTSIGPLSNPDAQQHVADPSSQAGPQLLPSTNNEVANNTNWQVPNQEQHFLPFTTQALLALQRRRRKDTRHRRQRTT
FSNEQTLQLELEYQRSEYVTRPRRYELADGLSLTETQIKIWFQNRRAKDKRIEKAQIDQQFRCAAAVASLGSLVPGQTLGLVAAQ
DTVETSQSGAMGVSYESMGLGFLGQLALSSQHANVLLPPALMGTTYTNSITPTAEVMATASPCSMYYHGNCNSKPVNQAAP
QQLGEVAKKTYKLDLSANSGASSSTGKLLTDRQDKSPQVA*
>c13539_g1_i1|m.20611
ILSEKKELVVKDHFASLTQDTFIAPESTSLELLKFLERYSSDHLNLSSMVNIKKALYDRLSASNRFLSTDAIPPVGELRVNMPSSTPE
QLVVNNLPRDPALTLPLNTDLPTSSEPPRENPVTCNGCGLPVYDRYLLHVLDKPWHTSCVRCFDCCRQLSERCFSRDNMLFCR
DDFYRRYSTRCHACANPILPNQLLRQIKHASYHVACFVCSMCGQQLSTGDQLYLTMDGRFICREDYTTVKNEVIDSENIPSSEG
VDMNSPQMSMYSTGGGPTTLPFSGGQAPPSALQFGMQSLEHLNFADNAMSLALLMPPNDNSLPPRSEELTPQMRPLSRAT
SLGDEEDEELVDSDCEDKLIMGDEDSDDGESSQMEDDKDDMKKENDKMSDISKSGGSEGGNGGKGSGTPGSKRRGPRTTIK
AKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGARRAAFFHRRMRSFDRPGAAVPDLL
HGPFFTQGGGNPEFFSMPPHGQAFGSPFDFYSAPHPLLTAAPALGAGMVEHSGNHLDFGTMPVDGSIPFDQLAHRQPNHP
VSIMDDSYMSAAGEFHMLTSNPPVSPTYSQSPNTTPHTSSNVIDSSTLW*

>c18308_g1_i1|m.33922
ETIKVDATTIKIKRKRGNLPKDSVQVLKDWLRDHRFNAYPSESQKLELSKAANLTVLQVCNWFINARRRLLPEIIKRDGQDPSLFT
ISRKSGRTALAVDQAGRSQIANQYSSRQSPKNHHFTINAILSQINLNKSPSRINEIPSK
>c14210_g1_i1|m.24563
MAFRGGISEFKSIDQLLRQPNEEEAQGMTVGTVPPSLRPEEATNNYAIYQALGLLSSTPAPSSLLLQAALTLPLLTVPPELVQHVN
RLLGAENKFTAEHHLSTTCQFRETDEQNCDVSSTNYHKPIHRPLNHSLSPPPRASSPAEKELVQRDTSPVQLQSPPSATINQTHN
NNAVCNTVHKKTRYRTTFSSHQLEQMEKVFSETMYPDVFLREDLARRLHLTEARVQVWFQNRRAKWRKLDKLTVHHHDAKL
YESEGPGITFPAQTASNLLEEIMDPSCENNEESCLIEHEAGVKSPRLRRSVCLTTGRYLNNTGSEQESGCSAVEFPMGVGPLFPTT
TSRAISAGDDTEFNETDRINSQEHPMIMPLSSHQTSDSNAHQMRSWYIDYWTATLNQHLLAATINSKNNSPHLAKDRFGRET
SSLGIKHFN*
>c10072_g1_i1|m.8594
MKIEENLRVSDNNSGFAPFSPFVFSNEKKEDFLQYLINRKGNLPVDESYRSGGSSWMTSHNIEELPKIIPGNTLSAYLMATSEAS
QRLAAFNTSPPPHHSLFLPPHPSLVSPQPLGSLTNHAGLQNTRWLMDSNLDQNILTHDETMTQRSGYETPPLHTHFPAGCGK
PGSRCKPEQDSPAETDAGQPSQSLLGSLVVLCTGCERRITEKFYLAVDGSPWHPACLQCCECGNMLEKHSSCFMRQGNIYCKD
DYFRKFVHQPCARCLGPLKPRELVMRARGALYHPSCFACSHCAKILLTGEEFGMRDGHLFCRLHFQLHEGLPPTNSLSPSTMPP
PFLFAGNDPGSFPPSKFHTNNEDAQLPGMPPTSAIPSVPAKLSKGRPRKRKVSQQVPPPPCDIISSDPYLNPYPPGITDLQPSDP
MGLFMNGGTTLEKQQQRNAAAAAAHAKRVRTSFKNHQLRVLKQYFLLNHNPDAKELKQLSQKTSLSKRVLQVWFQNARAK
YRRTLVKNQNGSKEENSKFDLDEASQSPGGTMRGEDTYSGSEDEGLEDYSPRSHDNEDEEMEEDGGCITPPISGREKGVNCG
NEDFSSQFLIPGDCF*
>c744_g2_i1|m.371
VMNEKQLHTLRTCYHANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIAMKQVHHMQAEKKIHHMNGVPLVV
SSPIHQGLGGDVLQFNSGTKVPQNPGDLLTSAHQGDFDQLMLGFDNISDGSVMSDLMESPRSDEDDDTASFSLPSN*
>c1980_g1_i1|m.1071
GFSDYINSRLSSFQNKVDNEPILLLASLRRSKQTPLGYHSLVKDSKDPLPFREATLNDELIDAELIDISTFKKAPPNSGWASVIWNG
ILRALFLPFYWSWWISVISVRQLILFILTWQLLVFGCLALVCGEPHVSRLLSEGAVLDCELLAPLFYLLTLSVLYTHIHSMTKPKKSTP
GTAAAHHEQEDKRPLDATHQPKASSAPTNGHVSYEYSCKEDQRQRAAGGASLSALARKKRPPRQTLAPRSHESCGTALPRRT
NSDEEVTSKSRRVMDSNGLSVSSNGCACNHTPQNQGAECMCESSSSSSFDATGRLLLNSIRYRSQHETEEVTSASEDRQCSSSL
LSSEPLSLIGSQSTDTVLGHEETFGKEDCAGWRELFENQVE
>c11791_g1_i3|m.13048
MMEGAVEDGKQILGSGGTNEVLGELTSPVKSSNMLTSSLKETLVCGGCGNIIHDRFYYQTGDWQWHQQCLKCSECAINLSGD
HSCFTKSGRVFCKQDYQRKYLKRTCGRCDNVITSKELVMRARHLVFHLNCFSCGYCHRMLCSGEQMGVRGENIYCKPHLYMQ
EDVTAAFKFQPGIECCDLGGRPSNLSIKYPLGGLNYSPMILPPPPHSKDFYLPPSAVTPTVPNPLQQIPLRSSTPTDLFHSTQPASH
SQPSTPTSLPPSSQAHNNNNSGVGSEAFQTSIPPVPSPVKTSRHKKKKQNESSKHIHLMPSMDDAALTFGMSGFVEMDLTLYG
GLERSHQQGVSAGYSSRTKRMRTSFKHNQLQAMKQYFNTNHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKYRRNLMR
EQILSVRKDKEKKLLGEEILAGGDGRGGPTPHPQGSPFLHHIGGEGVPPHGGYIDDDQLSLDCFYSEDDLEGSLTSYSPESSEDQ
ANNSEENEGQQ*
>c14264_g1_i2|m.24819
RRVHGLLSPVTSNEPPSLNLFTKSNNNLPAQHNESPRPYQESTTCASTPTPLPVISTNDPSKVTNVGSRKRASGEEKTNFNISNFL
SVDESQNDTQAVQVFNLKDGISNSDETVIKEYGSSEPLTNDLKVEQHETSDCVDAITPTSGYSSSESLKSSTTQEEGCDGSGSGSS
GGDPASPSSKRAKRKRRTTFSEAQLSVLRICHYQANPSTEEIEQLSDYLQLPKRNIQVWFQNRRARDKKQLKPQDNDQVGDSL
AFPGNNLLEMFKSVAASHQPHSPSTPEANPATTPAFPKKGSSMISESSDSPSLTPRDSFLLTPKSDSAAPIPDNMLQHSASVHVP
SFPQFPEDTQVLNDASPFTNNLNPASLHQRLLLSSYLVGLMQASLKSQGGQVHSSAESRAAKSEGISPAEEKDNTSLSSSSDDA
GFDEQPLDLTVKGTCATMSVEQINCSSCVPNSLLLSMPQQPNSLPPINFMTPALSSPHTPTLFSISPEGTSPHMEDGCGQEEGS
YDCHFCRKTFLKQSSLQRHVYEHSGKRPHTCEVCGKSFKHKHHLTEHGRLHSGEKPYQCNNCAKRFSHSGSYSQHLSQKRCLSK
DFKSDLHPSTPQSDVSMSIDTSLCDLSPL*
>c18414_g1_i1|m.33965
DHSAPTLCTVYNWFKDFSDGKRDSFEDLPKSGRPRSSTDDSCKAQVIQALEEDSFISLRQIAAFIDSNKDSVSRMLHEMNYHRIA
GMWVPHNLSHDNKLQRVECAQNLITVLTTISDIGNVYAVLDESWVYFDGSTHGRKHAWVPPKGDVPQITRFSPMTNKKTML
SVCFSPNGRFLVATTRHGQAINSEAFVAFLSRLL
>c11381_g1_i1|m.11848
MNLYGDIPVAYLKSNPGGGPTAGMPGHHMYSAMLSSNSHSTHMDQLIQQSMNYSGSLAHMQAVAAASAAPSRKQRRERT
TFTKAQLEVLEALFAKTRYPDIFMREEMAAKINLAESRVQVWFKNRRAKCRQQSQTNNNSADNQDNKDDGVKCQEDTKKVN
PIDQNIHSDKRISVADQLSPLSNGGNTIERVTCGSGKLSPQSTIDPDSPKQQSLMLHNHIDLLPHLPQPPAYPVGPATGQLWGS
YEDPLSTNGYNPYISHTTYVPAATQNPYMVPPDCGYIDSAGYSPYFNSYTTRMYNSAYHTVTPPGGNNAPFDPSLIGGSLDPTK
TAMENLSAWRY*
>c13168_g1_i2|m.18733

MNQSPLSGENYLDQQQQAAQKMHSLFTTSQDPLRPSSITYPGDFKGLSGGPGSMATPGSYADFSRYTAAPSMDLSHLQQHG
YSAAFGAAAASNPYNQMMYHAMASSHPTSQQVMHHHRNNMSHLSAMSAAASSNHHGDVPSPPPPEYFVQTAPQYYHHP
YAAAFGPPGMAHHIPHGYHTAAGGMSLTGSGYPSHHHLHTPSPTPSLPPSMSTCQDINHHLKREQYSPPPHEYELSKETFHHL
SNSQISPDSNLSKKTNKGGKKLRKARTIYTSLQLQQLNNQFQRTQYLALPERAELAAVLGLTQTQVKIWFQNKRSKVKKVMKGS
GVKASGGITSPLMPYQTVMPPPPPQQPHQQSVAMSLPLS*
>c10980_g1_i1|m.10673
MTTTSFTTFTVEAILSEGPLKSYTNATESRRQGDLATSPLLDAQFTNQVKKAEDEGGLREESGCGKPRRTRTAFTYEQLVALENK
FQHTRYLSVCERLHLALTLHLTETQVKIWFQNRRTKWKKQNIGNESTQEDFESKGSLFASPQFAAEQSGGDKHFPEASPFQDV
AMTTSAGLLEPLVKPFFQPSATSLLFPMLSQQLLVKIASSPLFAQFRATP*
>c2555_g1_i1|m.1383
PTCQIATSSTNQLTGYPPQGGSSLLPPHLNGSSPHFLPTIISHGGPLPHMMQQTMGSYHQTSVVPYNPPHNQIIQGSTSLLQLD
DSLMANHLLNQPEDDQSYKKCKRGVLPKQATQIMRSWLFQHIVHPYPTEDEKRQIAIQANLTLLQVNNWFINARRRILQPML
DASNPDAGCKSKKTKPPGRPNATNRFWPESLAQGQICKVEASTPSPPDMNGNPNNCSS
>c11214_g1_i1|m.11379
MKVHNQQLPYSCYRGGGGSLSHQQVYQHLRDLASLHPPTKSNLSPADELERRSEKQALLLAEMFPTGLLPLPPSVTGAPVSTSS
TSPPTPISLLQQANAAAAAAAALPGGASHLLNFSPEQISVVCDTLEDSGDVERLARFLWSLPMNPVTCEVLNKNESVLRSRATV
AFHQGNFRELYHLLENHRFNKDAHPKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPRTIWDGEQKTHCFKERTRGL
LREWYLQDPYPNPAKKRELAGATGLTPTQVGNWFKNRRQRDRAAAAKNRMNGTMDGGSSCDSTSGDRLKSLGDDEDLLSD
DDDEDQFDEDPRGLEEDSPPIHSSPSSSSGNNLLGLRLESGFPHNFYQSPPDNPSKTHPSPPPTLNRNFPPQFDSSRGTSLVNHF
LMDALRSSTSIPPFPHLGSTPGVKSSLHSTSGLNLSHAPLITSSPHIKAPSSSYAGLSPDTQHYNPFLKLSLPPLPPPPGNPNFAYFL
QAAANAQFTQLLQLGMAGNQFLRPLQQSPMTQSTVPQSPIPQTTVPSKLSSPVETDSKEGETASPSKNLHKPTVFSPVTLCNIK
DTD*
>c8918_g1_i1|m.6553
MVSERSHLHACLMREERAISSSLYSPFSLQPNYCPYQFYANPPLFLHTHNFPHSEVQSLSYPILELMPNGIRNSSNVPPKLIEVQG
VKNIQGTLGTPRPHCPPPLIMPLCSPLQPHSSLCEAKTHRPPISVSNMCHSRPEYSEQPHHHMPLHVTGTLLSPANDFPGHSHQ
RHVSRHHLFNDNSSQIFSKVKRCRTSFTPRQLLGLEKRFVSCPYIFGWHRAMLSQELGLSESQIKVWFQNRRSKLKKSPLFERG*
>c12381_g1_i3|m.15122
MFSIASILQQDEGAAKKLDRAELSSGSGDCSPSKEKLHDFSFFPWMDGKYHENTFDAIPSAGNNFDTTPAALFNSNVINHAFLA
VAAANSRNNQAELQTQNADQKVFPQLGTHLNGLSSQQALLMGLINPWFLPLQQVSHPHHAAWMQAMSLPPSSRAFPGDC
SSVTSHPTSGMTSESETARMPTILNKTDQLLVDTSALEFSVPKPTPSGETPLQQFLQSRHASSVGGEPYFTHSKHSPSSPIREDIT
ASSVHESMLMEEQISDDEMGGDDDGDGQLISRKKKTRTVFSRGQVFQLEATFDSKRYLSSSERAGLALSLHLTETQVKIWFQN
RRNKWKRQLAAELDAANMSASSVVVGGNPSQRAATSLLRASLPLMYRSSSSPPIVKSNNSLLNLMPPPNHIGFIQREAGNSNS
SSVLPSNLLQYRIAKRLQNNDSSVSEAMKVIKGRSEDENKDTRSALFQNYLGLNLPHFQPPKVCSDLVTQSSEISHEDFSPSAEQS
SMTNGITGMLSASDVANMFAAYCQHQQPANRSAALALANQLALAQQQLQENLPTQSRQLTKQSY*
>c11413_g1_i1|m.11951
MVNIASSPTMGSPLVDNNHKTYQAPCLDITSSHEQFCHEPPTSAAVTYLGHMPMVATGYLDSHLAQLYGGEAPGKGFSVNN
NQLCADFYKLDDQDTQFVDTNYPLPFIRGAPPLLDPSGVPTEQSNKALSSPASLLCAGEQPPPSCYQPPPTSTNTSTVTSLHTAC
YPSQIFQIPQKINPYQQHSTSLYGFPMAGFQSSADVQNGTYQPLPGDFYPAGHVPPAGYGCELQENGYYGASDCDYPAVQP
MPFNTPLQAQYPFSDRIDNQCSVRRKNASRETTSALKAWLHEHRKNPYPTKGEKIMLAIISKMTLTQVSTWFANARRRLKKEN
KMTWMPRNKSSSSNPSLIKTQKSTKLDVNSNDKLSADDEELTCKTDTPSMGVTEDCAENVPSDESDVDEDDLDNKEYSTD
>c18933_g1_i1|m.34095
SPSSQVPPVMSGLTQGLMGAGGGEQGSAGGIMLLRKRKKRTRAAFSHAQVFELEKRFAQQKYLSGPERGEVASSLKLTETQV
KIWFQNRRYKTKRRHTPT
>c15419_g1_i1|m.32970
CICQVLEQSSQVQKLERFLQALPVISATPEIRRQENVLRAQAFVAFHRGEFKELYKLVESFDFSAEHHSVLQALWLNAHYKEAEK
TRNKPLGAVGKYRIRRKFPLPRTIWDGEETSYCFKERSRNVLRDYYTHNPYPSPREKRELAGATELTVVQVSNWFKNRRQRDRA
LESRSPSSASSSPTSQVIISALDSHGSSIQQKAVFLHSNTSWAFSSTPSMVSVTPSSEPGTPGSSKTFVAPATRPPGSSLKLSSNTG
ETFFDPATLQLNGNNNPVFSTAAIASQMSDCVKQEPTVGSLFVSGMSCDLDCVDHSDVLGKVVPRRIIRLPTSAQSTASAHQM
MNPMMSAAVKTSPHIKSSPTKKQRSCQVSSASSISPTPLEMCTLLEELDSVGDMNDMLNEDKTSVLIGLSDPNLPISSIPTSSISE
DYFLKEDNSKLVEDFLALQQKVQVKNLTQNKLKGNNDKKKPPVGLNSNGQQMWPNCNNFASNKKVITQKIQSPTTTSSSSVL
TMDLGELLDLSGFVAPSSSSTTTNNNNICSLNFGSSILDSFDNDAMIGGRGGSSLKGSSLVTSPNPFSASEMLQQAALQSTLAS
MLVT*
>c13813_g7_i1|m.22216
RYEDTTISTQWSSGSVPPHLYEEYCRGSGPPTTYAVPQQHFLDLPQTSDCNDLSIQDKELIYNHPLFPLFELIFEKCELATCTPREA
GMACADVCSSESFTEDLSVFSKQLLAAPFDRNCTSNHETDHLMLLAIQVLRLHLLELEKVHELCDNFCQRYITCLKSKMPMDLV
VSEEKDSKLLRTASDEEDQTSPSPVVNSPVARVPSLTPTTHQHQQVVASIPPTMQPPPPPPSTHTLDFDHVLNPFGGHHTQGS
APAGMYMPPPTHHPAYFPSHLMGQVPLMVPPTSHFMNHQLMQAPLMDQKQIIIDHSPTSPYSHQFNGAVSERESCCVEAS

SPTPPANEAENSPDDGSNTNEVKKSSNSQQQKKRGIFPKAATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFI
NARRRIVQPMIDQSNRAAGGYGPEAGHMGFMLDAAAAAGHHGYLRPDMAAFAGPPAPPSYYHTHHQLIPGHSVGGAMV
MSQSPVTAAAHHHFGNGSMQLMDP
>c10631_g1_i1|m.9746
MTYFSSVKHQQPNPLGQSYNFSPNYYNNPPTYQLAQEPKHLPHLNDYDGHLNSDSFVDATSMAYNHPKHIDQSHQHHPEYF
HMVHSPSSLNFISQTLGSADHHHSMEDQQPLSSHLALAMHQQQQQQNYYHHHHSAQSLYGMSTHPYYMSYYPQQSNQEL
ASLTNAMTTANPPSSTLIHHDPSSQLLLLHQPSISPDTSPLDVKSSKNGACSVKTGNAQNFSPIAKAEIADSDSVDESHISMGNR
NGSQMFPWMKRSHTPTGTESKRTRTAYTRHQTLELEKEFYSNRYLTRRRRIEIASQLGLTERQIKIWFQNRRMKWKKDNNLKS
MSQVDAMMI*
>c6050_g1_i2|m.3465
RRVHGGPSTFVKQQIETGSRRAGRTRFTDYQLKSLQDFFEKNAYPRDDDLEKLSAKLKLSPRVVVVWFQNQRQKARRSIEIQKS
GLSVSTSDPYQFSYDDYDASPTDSNEIPSMVEKDYECQPCQARFNSVEEVILHQNTDCPAASKVSIVAPESQLDAAALPLEPNTF
ENSTNGEQMSEEQADDTNKLSTLSDYEDTEEIVYTCNWCTSLLTSEEEYECHQKDHLDELLKAAALATSTPFQSSLSLSCEVKSEV
DASVEDTEDAINNANELNEEPKPLTIVEDECIDSSKQDETVLSISESTTSANSVPFNTLADEPATTLDTS
>c2526_g1_i1|m.1363
MNDPQFVSYSEELGRSSKRCFDLRNFIPSHLVYQHPDMEQSFPGSPLRTSFLSPPSTTTNNHNTSDHHLPSYIEGEDWPIGHTM
VPTTNQHIFNLSENYGITQISSPLSSSLEDVSSTRGGEGFSQHTSLFSHESNSTKNDTSRLLTSPSNNMAKDLPLTSTKQIESPVAN
EEEEDEESLNKSRRQRTHFTSHQLQELEGLFLTNRYPDLAKREQLCAWTQLPEPKVRIWFKNRRAKWRKRERQQLHEYKTFME
TAAVMSGDSQFGYSAYNGWTNRMSGITKYAPEPFPWSSLGVPSTQGYSIYGAPYVMPHSGVDINSCPNGEYLPSLAEPFSQQ
SLTFADAGFKIAYSPPLSAPSSQLNSLNYMPATRHTNRSIGLTNQIAQNTSSLLVT*
>c9332_g1_i1|m.7210
MLNTKDWWSLVNQHYQKQIADGAEESQNTSVKLQLTRSKNEVEETPLLPAQPTSVPTLMHGGFPDYRISSLLGHTQQQHQN
RSNPIQEMVGKRSSDDSTFSQYNPYKVASSIHLEHPSQSVINPFSFWAPHVDPSLSPINIINNNNYNMSNAHYQTTTALGGGGE
QRHVRLSKDMTSEGATFHPTAQGMLSTETLHGGTESCYINGALNLVKPPGHQSPKTVSGVESPFPLSSSEGAAAPPDQAPLST
ASTMLASQLAQYQQAASMMLAQHWHFTTNKQDGNSQSAPTFPLKTKQHQLYGAKDMLCTDINHHAEGVEVRGDETASDN
VISASSPHKGAQGGMGFSPTTNEHAHNGDTVPGPQGDRKKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALSMSESQV
KVWFQNRRTKWRKKNAAEVAASRLKGEISPDEDGC*
>c12955_g2_i1|m.17715
SIFAWEIRDRLLQDDICKEDTIPSVSSINRVLRNLTNDKGNTINSAASFLSAGCLGDGNITSSAAASYDKFRFLNTQAWSAAAGSF
YASYGQYMPNLAISGSLAGDDKKFDLLPEATASEEARMQLKRKLQRNRTSFSPEQIEALEKEFERTHYPDVFARERLAAKIDLPE
ARIQVWFSNRRAKWRREEKMRSQKQQSHSHGSGSNDQSPRSRLHSSGSHTSPTTNQFNAAAFANNFYSPFQYSANTLSTYG
YPSAVVDGYSSTFQHFASQGYTGCSMFTHNPYAPGTGTNTMSPMMMIGQSTNPVGAVKPRFVDASTPPRFGPTVGGIDSLS
AANYLTTGFSVNAGREGSPWHPGLESNSKQLSDQIPSRDLPPASSSPDPATPPPHHHSLDHSTTLGSMSAAALGLTCGNPLIPE
QVNPLAPPMMSPTTHLMYQHPQPHPQCTSTQQDSSFSIQQSNGIQCPNPGLFSAAMASSGATRCSGAATQIPDFTQYWTQ
AARI*
>c14140_g1_i1|m.24180
MAVETNLDFKTAKTFKSNILQQPSQDNLHMSIDSHSQKNPLIPSAAAVMFSASTGILPLGSFLSFNPSPMPSNLAGVFSQNAKP
MSHSYAIAPLKSFGNIGMSSFHQNLTEKRSAHEANTDCSCTTCVFANSEVSKPNFLKLQSANLEFQSMPPKSLNSNVENPVEAE
SELLHSDNFNANQHLEAKGFASHQWSRIVDSAAFGYPMINPMFRQKDVHSSIPPPVSSTLTPGLQQQLQFSQLYLANLKATAG
LAGHLSTSQPTPSTLSTSLPNQLLMPMPLAAFPAQVLRVQGLLGPNPLGFPQDGLYPSEWGPVGRRKNATRETTAALKNWLG
EHLKNPYPTKGEKVMLAIITKMTLTQVSTWFANARRRLKKEGKMTWTPRNTYDDEEDEGADEVGEDLSTLERDSFPDDDSSH
FKWMSRHENVEKSSSRKRKIGSLQKGIWSIVDVATKEQAINKDNLLVDPKHSSPYQNSNYIEEETNQLEEDEDDGHLDVDVVD
LKEEVMHHSLDASSSVSVVRSTETLTSSELLPNYSPIKSLVSPGKGKIKNSEEFQQQPISGSLSPYIAQSKPQNPHDFFSQKKLPVG
CHPMMNLPYIGIQTQSFPMQLGHQAFSANNFDIELKHTSQASNLTAS*
>c14906_g4_i1|m.29310
MTEGGLAKLSATGRGGLAKLSATGRGGFSVKDLLGITENEMVDQSGTMVDNRYEDDEDCDTYIDCCEEDDDDDGDESQKPA
ANQEGAGVNGKIKKRKRRVLFSKAQTYELERRFRQQRYLSAPEREHLANLIKLTPTQVKIWFQNHRYKCKKSTADSPPPPSLGSR
RTGAGGSSSSPTGPGSLHQARRYHQGIPLLLQDEGMSTAMRRGESVSKGGCPPPPLLPATSLSLQSLGLHSSPGQNATAAALSS
QFASLVAAAAMRLPHLIPMSTAAAPPMSSSGMMTHHPPFCFNPSSAATALTNSMLAARVDLLMKATLQQSLAANDQRKLLQ
FGAKEEQRKLLHIGSKDQEGSLRLEPNSPINLSDKKLI*
>c3972_g1_i1|m.2070
MAALTSTDCLGLQGKPRNSEPDLSPDHLAQLHQHPTTLPHEEGSTPAPNSFCDPFSAERKTLYDHSRLDPTKMEFFYSGGHLST
GAGVGNVPVAGSVGAEYPLAANGAMSQLLCGRADMGMLLGHSGDHAIALATSLGGHVTTSLGSHLADSLNHASRGSLGVT
DMLREDEIEDHKMNSDKIQRHLSGEDIDGNLLSRETSPSSLLVNCCPETEVNSRGGGIDATPTSHSGKLGTKSKRHRTRFNPGQ
LNELERCFTKTHYPDIFMREELALRINLTESRVQVSFKIKEPNMTRSIKR*
>c13888_g1_i1|m.22661

MVTQMPSVSASGGASSLLMPHMEPQQFQQAAMMGAFRSPLTLPIPGCFPQRGGPPTISSSSQQQQLFPPSSGTPHPALLAA
IMAASGGLTAPFGRPPLIGGQYSHALPGIPPGHTGFLTAGPHFSEAAIHFQEQFNAGLPAGMVELNALARRKNATRETTSALKA
WLNEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKVTWSVNSGPGEGGDHNNNEDSDDDDELEGEDDDLE
ENNRDVKHMASELNESGIHNDSDDDIDVESQERLVLDGHETKTVKENDEAEKIQIRQETKCENSIPPPSGFLVFQQTPACENVQ
PHLSLLSMLGGGGGSSGFPFLAQDASAAMGKFWSPYTLVSPTQQKQLLETGHTASSPSKCCPEGENGRSTDSSKSRVELVSPV
EMSRTSSKLTLEGSYLRNKFTGAMRKPLKNKIWSLADIATSDAKKVEDTSLVPSSLAPVSPLTEASTLPQPAVTPNYLFSAPLLWQ
SPNIQNCLRTSSSVDSLHSPAFDSNLNLPLSTTPQVSSPAGTQPFHVRDSLSHFQATLFPPRSMHISNTLTPTRTPLLMLPTNSKA
AKQTSLLMPVGGINSQ*
>c15126_g1_i5|m.31113
MESAAASSVFYPNCAPSAHSAAAHAAHAAAAAVAQLDLRDIIQHPHHLAAASHHVASVQQSPGQSVMGVMSGSSPNNFCH
TQSPFFTHSGIGTFGSVSGNDAMTLPSSSSAGAGMVVPTYPQTAHMGAHPFTHRVSYQPHNHASQLFSTHPYPSPIQPQMG
GLAHPFIPHHYSLPSQAKMTDLGSHPLDILQHTAIGASVGGSEDLIKLEARSSPGTSPSRGSRIEQINGRHAASTSSTNHDEADED
EEEDEKGNKRHRRQRTHFTSQQLQELEALFARNRYPDMATREEISVWTNLPEAKVRIWFKNRRAKWRKKERNQLQEYKNAA
AFGMNFMAAASPYDPATAADLYSSQIGYSAYNGWASKLPPLNKHSTSFPWTTLGVAGQPGTGSGSYNPYAMNAAAAAGM
PGGASGAAMMGGYGMSTRGDCTSVAPGGSGPFGQSSARLKFPTSHTANTQSSGGQLLTNPIAAATGIYAATSRSNSDATGSS
SAFVSPVNNNSTLAAAGSLLYEHRTIPG*
>c5392_g1_i1|m.3056
MDVLSVNHYSSIPNEAISYCPNEKEGRLENVIPKFRTDRSEDICSIPVPAHNSLFLKGEEEHSLLTRAESHSSVIKSIDTISANPHGT
QNKTIYNKSNLWINPEKYKDVQYNEVKNTSEENSFKSAHNGVVATGQYENSDTESRLAVTTNTRSTKISEQLTVVTPAMTNCY
PDQSFHYNYNPLDNFQTSSEYFMPQNEAYYHNYAVIPNSNELMEQSANVVYYGRFTPPNSGEYKLIPGRNDMYDQAQFYNY
EHQNDIKASTMNLLNRGIEHTNNIYFSRTGSKYPNISYPFSSNNFEHNTVPRFGNQHSTVANGHGAYFRTWPNAEYSGGNCPT
NLKLGEERCRSEAYQYFFHQRLNLPSSSKDMEYSADLQGYNPSLHDVHGKMQTFNGPYSPSSSEESSEFDGVEKPNHDTCMY
GQNFLNQKNYQTIAFPQQDTVFYNFPENFEKTKLLPEKYFKQYEDPDFTKTMAKNLDDWQYNYEDKFDNVHPYARFPQFHSD
YRNIVAGTKYDLKAGPPCGQDKSRGDSKIMSFPTVGMMGDNQLNPSANSPGEEDSSCDSPLAADKGSLKMEPLRAKAVESA
HSSVPESDNQPLTYNWMKIKRNQPQLASQQSSLDVSSPLTSRAVTAAISLAPFSSGGLGGQSRCGRTNFTNKQLTELEKEFHFN
RYLTRARRIEIASSLALNETQVKIWFQNRRMKQKKLVKEGKLAGCQ*
>c10215_g1_i1|m.8895
QRRVHGNPPISGEVFSAASQKADSLFATPDQFRAFPDFKTLPSYAAELSRYAAGAPPMDISHLQQNGYSTLGANHYNQMMYA
MASSGGGGAPHQAQSHRYNHLSSGMPPHYPSESVGTGMPPHEYFSGAHQFYPHPHPYAFSSATGHGVTPGVHPMTSYHHS
GPLGLPLGYHHHMTTPSPPPSASGSLPQHLKVEYSSPPPEFDKLQAAIPRTHSPDLDSSHNSKKTPQGGKKLRKPRTIYTSVQLQ
HLNSQFHRTQYLALPERAELAAALGLTQSQVKIWFQNRRSKVKKVIKQNTPKPNSCSTVTTSIISAAESPYHHHRDESNIPPQQL
*
>c7806_g2_i1|m.5135
HQSVVAMAAAFRHLYTPKPYAAPRSKDFLLFGESDNVISQLAQNSPLKGPLSTSNIPPSEIMQSPEKKVRLDLEEPCCKDPRSTT
KTHTAQAHRSTNEKNMMSELNSVRDKLESLDQDTSQSMPDSAKKELCGILASLNDMTSNKILLDEKLDESPACGEIQDSKVDV
CQDNLCHRSSYNKPASAVKQVQDLAECKDVLTGAISTSWTSPLDSSEAKFYYPSILQHTGSTLPPTGQPMSLSYSTPHINEEPGG
TNSLLLDLRQKLLERSYLSGTSQGLKADKPNVFTFTDASSVNSPAVSSSSRTAAGNSFPQPLSSSLQHFFKERVCGPRTKVTEEGG
NLKRSGHETLTANHLRKAKLIFFYQRYPNSTTLRQLYSDVQFNRHSNAQLIKWFSNFREYFYIQIEKFVRQAISEGWLERGGNSTS
HLLITESSELYRTLNIHYNRTNFKVPDHFLVAVNTTVQEFLNSILNKRDLEASWKKQIYKVINKLDMPIPEHFKLANFMETLE*
>c5831_g1_i1|m.3320
QVGYDSQRKKRAEHQAPLDLSGKASLPGLEHNFKMQLDSFSTISPSMVPIAAGLPILLPQQQAFLLKNLYEQQHLAAMQFGSQ
QPFIHLQTAQIPAVNQQLHFIENPLHSVSSQSSIRSNDSNVQPNKISTSENQQMNFQQMGSESYNPYPSPGDSGIESFLQLESP
TNRHHLSPPLSTHNTPKSTDSNVQPVSPKTPREPETESSQQLYFSPDSFTHSFIDQQHRSPSSYHPAAAAAAALLARAEFAARLM
ERQERGYLVSPGDGCKSRRPRTAFTSQQLLELENQFKRNKYLSRPKRFEVATKLMLTETQVKIWFQNRRMKWKRARKSTTPSSI
ATMPSRNSLESDAERKSTVAQEHKLPEQFVSLSTTQPLTAAMTRSYPALPQLNFADQPHGTGVRSTDALCGIVTAY*
>c4068_g1_i1|m.2121
SGRETFRKMGKWLNDPEFQRMSLLRLAACKKKEMECALFKPHMPKKSRLVFTDIQRRTLFAIFRETKRPSKEMQTTISQQLGLE
PSTVSNFFMNARRRCGDRWKDFDEMMASKMRGHSVGTHSSKGGSGRGGAKGKNNHSTSRAASLVGQHIHPLSHHHQFD
MIDPRNHTDSSNLEPHSIHDNSMMSAAMHHQLGTSPNSIFSQSLMMSH*
>c7095_g1_i1|m.4340
MTVATSLPISSIDSSSAIRNVDSIRRTFGPSPAQQVPFNNSFDSSYSAPALTFANSLQPVPGMAQGLPICEGCCQPIEERYFLQVQ
SRTWHEHCLKCESCGIQLDHSCFFKAGKPFCKLDYERLYSGKCNGCLERIPPTELVMRTPSGGAVYHLNCFACVICHQRLLKGDE
FIVKDNFIYCKIDFEKEFPSIFELLRRDDTSMTYGCENNSHLHNATAPSCAQPASQLAKFGTNISPNLFTSSASNNCTNFAACLSGT
SDVTSPQSSSEDLGMKENNSCQADSDLEDYDEDACSEAGDASGGIGGKEFIKGGEEGQVGNPGSQQKPCRKRVILTTQQRRD
FKAAFDISSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKLAKRQQQPSQTQMDHNGLSPQHRLDDGSFGMFVGEPF
PPHMDARFLGHPLMADPNFPPHMYPPNGGDFFMHPFCGGAMPPSMRLQHMSSPPHDLIMNSTDPNSSSTQQQQQEDSL
DIPQLLPVSSSLLHTPPSSEELHILSSPPLMGNNNSTESINSNPIDKLYSMQSSYFTQSL*

>c1539_g1_i1|m.833
IQTTMHGRGLLSSSICYSLPDAPWNRYCMLETGASSGSILRPWDMNQLTTNSENDDTRLRFWTHTPSLSSSDQEYKKRFQTQK
LLELHKNGLIDPNHQIPAISTLQSQQNVNTKTSIIIQSKKKEKIKTSFMISNLLKASVSNSLKAVPLSPQREVESESERELLSVDSEKD
WKEAEITKKVINFTKSTLEEEADLDSEKSSKVNSSMEYPVLSCSSNNVCRTTMLDVSSEVKPDFETKLATQPTDILHNCPSSVDTF
YLKDRLEPKSFQYNYYQNPSLAGFPHISPLTSTSAKLNTSPFNQTKLQEKKLAAFHQPLTFNPEPYGYLQKQQFMVGMVTQPHY
QYSMQLHQFSQYQGPGALSFQLYPSGEEEEDSQKRCRRSRTVFSDSQLNSLEKSFDSKKYLITPERIQLAKQMGLNQTQIKTWY
QNRRMKWKKQLIQQGATDIPTKPKGRPKKVALTDS*
>c6340_g1_i1|m.3704
DSKSTPQENLLSSNYFQNRPTSSSWSTLSHATCCDPGMDYFNSPSMYDQRLRATSGAAFSYLNCCPGYPTTPTYSGRLRDFAA
RYETLGCSTPLPSSNNTSSTSIADCSPALASFQQARIEQPYRGFQPYMHAATSPASTDSLQQVLQPYQPYHHQFSHRNFSSAST
GFPPYTYSPSSYVQYQDSDTASTPCKSLLGDDASVIEVETAEIAQRVMTELKRYSIPQSVFAEKVLGRSQGTLSDLLRNPKPWDKL
KSGRETFRRMAQWLEEPELQRMAALRFAISKKKEEDHSDLPQERGVKKSRLVFSDTQRRTLRAIFRESQRPSRELQGQIARQLN
LAPGTVANFFMNARRRCSDRWKKSCNDSELDSTNVDSDERYCNSKQFKVDFTESPLYGELFTG*
>c11370_g1_i6|m.11816
MAGCEGNSDNKCHPSSIISDTNPKDYTDGIPFEGLIDILDAGTAAVKVKFDDEHFDDISVNGLVYCLKECSTFPEFLHIQQTNICDI
SFISNKDKHTFFYLCFSCKICFKEADKFFNHSDQKHNIAFSQEDQKLIINNNCSLILYLTNSQQVLPFLLQFDSLSPLENEADDGLLS
PPFDSTTSNENAQSLNSSNDTPSSPINFFPLKEESNEESNYLTYSDKISKSEVVEEDEDAATLETSDYSELSDLNHVIDSVTAAASQI
CADDFHLDQNVFPNSLLSGGPICGDPGYHANGQSSADCPKCEMIIGSTRNYQTSQLTLSHSRNSCKTLKCPKCNWHYKYQQTL
EAHMKEKHPGTGNSCTYCSRTEPHPKLARGETYSCGYKPFRCDVCNYSTTTKGNLSIHMQSDKHLNNMQELRKTGQCEGNAP
HEQVASGSNLSISSQPDSPNSVSSSIAQTPAMTTSLSKNGKPLWRCDICDYETNVARNLKIHSGSEKHAHNMISMQKLSSNAQS
QPLLSPNSLQINHSSISPTMTSSDHSSQQEQCVATLTQKPQPLPLPAQSSPINAFQTQHAMQQQQTEALLSLLQSVATNNGVE
SNPVFNFFTTGIVNPTLTESTSTASSMNATLALLSQLQQQQAQSQQFAYSSVPTSNDIFTCRLCNKFSSNDLALVQAHALRDRT
QTLSNDCLSFDGINFKCELCGYSTNLKANFTLHSQTEKHKQKVQIFNHLNEGDENLSSILNLVHQVSPVQIECAACSFLTKSIGQ
MNAHVTSEPHLYKQKLFDFFSSFELPERKNAHFLCKVCDIKTASKSEMIVHITSALHCGKLKERQFQGLPVDSWNEICEMVVAE
DLAAGSSSSTEIPQSSQDDDLVTSTISNNVATNSQLTTSPSQSYPLINPVSTSYLLPTSPVEDTTVGETFSCPLCAKIRGFRSRDILRE
HLLSDHHVSSECVDSLMVAVKPTMSVKPIARPRAATTTTACTIGNSQSSVITMQSSTSDESLSLPTTFINTVLPAVSEISMQTAFI
NDVFRCHRCHSSFSLFQDLTSHLTFCLKIPCPLPGCSHNSNTQEELTNHVNESHSFNSQELPARYLCWKKGCNQYFRSEEVLLKH
FSEIHSTSLSGSFSTLFDANSASISNTSTSTSLQASAPLKLQSSVIECIACKKSFNSALEMQVHLLTGKCQPQSSSSQPSTSHRDVND
SLPSVDNMDNTSDQDDMMNTPNLAAMYYSSDERPWKCHPCKMAFTNKSYLQIHQQTVLHDGVDSVGDVSTNSESQAISAS
TSSTTLLSHTKPYKCVMCDVSYSEKSTLDIHLSSVLHQSKVLATFQNSMTNNFNTNPSLSSLQLGLGYLQPMDFSSLGLGFEQLD
QQQRHLLQSLDFHTQEQLQLGLLQQKQQQLLLQQQQLIPSLFPLYSMGSALCPCVKCGQYFASQEILLSHQLICNLSVKHDSNT
STSVNQLLSPKTNDIKRKISTDNESCYTFFRCQLCNASLADKEFAQSHLATVHNLHVKVEEIESHFAKLASESEADSNIKRLKLEPE
DNTSCTSQTTPTEHVDNLVSTSQNQAADNYSTDDLLLSSGTDDSDPNPSSISSDWDQFTNRMAVAATINFLQQSMQQSQQA
SLLTPQQTALLLQVQNPAVASLLPQSLINTSLTSSINSDVPMDSIQESSQSSSTTTPDLQALFSDLHKRPRTRITDGQLKILKENFDI
SNSPSDEQTAAMSRETGLPSKVIKHWFRNTLFKERQKNKDSPYNFNNPPTDPYTSAAAQIAAATAAVYQQQTNSQQIENENK
SLITATFLDSPKQTIAASLPSCSFSQTPMIEILPASQGPSTFV

10 - *Diopisthoporus longitubus*

>c26501_g1_i1|m.12425
RVHGDHSFPNTFQATFIPATCVTFNLNNNNQAGDDLPPSASSTSCFGSGAASAGNPSTDQHLFPEYNSTSNNSFVASSSYSDEL
NTIYPNFSCPPGVGIAAASGSGFYACKQEDSVKRDFGQDNSFYNQQISPELQLSTNKIIPEYSAISRVNGYLVATTDLNHHHRLP
MPCGYGGQSVCDNLVYQYPSATHAATGYVDYGFQTVSGKSNSVKKTPTQAGSGSKSKRQKRLAVGSGALNQDEEEKESEEHS
QQVYHGESMDTREIASKVMGELKKYSIPQAIFADKILGRSQGTLSDLLRNPKPWSQLKSGRETFRKMGKWLQESEFCRMSLLKI
NKGLEPESFSKPKHHHQHLAHLMKTTSMHCISGKGGVDRIRKKIRLVFTDDQRNSLFRVFNETRRPSKDMQSYVARQLGLRMS
TVSNFFMNARRRCVGYADDTTQNKMKDKQCKATRKRIKDDFHSCSIDDEY*
>c32472_g1_i1|m.27509
RVHGDILLRFDFDSTYKFQTKFDHMTTLGSNPGLGSRPSVPLITPAALQGFAMQTGMNPFAQMVSEHHSPTKSNKPFRFDVF
PNTIVKPEMASLIHPHLAATASLGFLGTKHQPQQASLQQHHQHTHQIEAANNNQPISADHISSVCDALEDTGDIDRLARFLWS
LPPNMPMKAENFGKLEAILRAKAIVAFHRADFRELYHLLETHRFSKDAHPKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKK
YPLPRTIWDGEQKTHCFKERTRGLLREWYLQDPYPNPTKKRELAQATGLTPTQVGNWFKNRRQRDRAAAAKNRFNSTIEHVG
EGRDRLKSFGSDDDDDMLSDNESQSIRTDSPSISPTDQSDLSFDHEHGLSPTSSRKVLNYFPVERYLPQPPQAQPSGANFGLLQ
AALAQQHLQHLLTLQCKAAAAAPHMAPFGNFTPAHLTVNAVTKFRPHEKSTTDASEQLEKKQKTNTVFSPVALCNIKDS*
>c19828_g1_i2|m.6621
AEYMGVINTSFLFRNQPKMMPVSVGSKKLPFSINSLIQNEEDPAEITSAANTFLKSMTETEIVPEQHHITGSSSFREKSKQQIEIE
NDLQSRRYRTAFSKEQQELLETEFKKENYVSRPRRCELASDLNLPETTIKVWFQNRRMKEKRQRQSILWPCLDLAYQHHYYSY

MLAKSLFHQQQHQQQNSPGFGSIAYQHLLTGMGSNHSVSEESTMSGTTPLSPDLQKRSNNKCWLPSSSQKEVLHIPRENEVS
TCSISEVSPHAEDSK*
>c5059_g1_i1|m.1297
QVVALCKQPPIRFNNFSNANMPQKCTFTPEERRMMCQMVEVEGLSRRAVALRFECSDSVVRSTLKRYQETGSFEDKPRTGRP
PKLSARSKARVVRACKANPHESARAIREELEVDGFPKCSIWLIRR
>c20323_g1_i1|m.6852
VIIIIKVGTLATKQNQMENFFDVIRHPQFASPASLFPTHIPPPPVLSLISPHHLSSPQKLSPAAANQDAIERSKSFEYLKKLYQSKEW
LEKTKFQCRCLMCVGRVITPEEQFFIKYGSLPQNHDHTSAHVLEHNNSIHCTHRSPSTLNELNKLDDELDENCRLKECSYPLRM
WLDEHRGNPYPTKSEKMMLSIITNMTITQVSTWFANARRRLKKERRSEFEYAELFHYPFAIQKSHLASRESGTDEMTARDLQTF
KTSESLLNLKEKAEPTTKNRNKPFWSLADTALN*
>c35288_g1_i6|m.41368
MLSVILPATVVTSDHTHSSRRFDSLLQLVGMDTPSAKRLIGGTFTIDAILGGAPTGGDTFAGDSMDRNSEEELQRITDLPTNCQ
DLNDVDESETNRRDGKSRDRKRKLGRNPRVPFSSCQVAALESRYSTKQYLSSLDVAELSRKLQLSESRVKIWFQNRRARSRREIL
SGSASTGDGAIPDVSCGDEHTECGTNSIAPSVIRAPEEKSLDLMSRQLLSYRPHSYAIQQPSPTKPAEDQANYFLPAAPFLNQVP
PPPQPSFASSSASFAGNYLSRLWGDNEQNMARIPPPNATLYYPIGRPNFVFHK*
>c28863_g1_i2|m.16620
QRRVHGDQFFDKSLSQEKREHNTADKVCPAIDRFTPMETDSQLYEGLCIETDYNSNGHRDDKNNFRRFACDECRKSFKFKHHL
KEHKRIHTGEKPYECSNCHKRFSHSGSYSSHTTGRKCFPVFNSRGIAGKQTGHCKKPTPKIPFYSPNLLNVQTGVPNNNTATNIP
QMNTPEPPLNALMLSLCKRSNFEWSNNSAFTIQTCRLCGLPFANLFLLHHHQLYVCPVARLLMGGLINQLPPNSALKPLSALIT
HQTSPNDLKEMCETQNTSDNIVITKETEPLISTDLTNLQGPAEDSSPNSDYESASRKKRTVFTESQLQLLKEVYCEEVNPSSDFIEN
LSKQISLPKRNIQVWFQNRRARDKKQLKSEGVRLNDLVIEGSADLNTSHENETVLGQISVEENRQHQPQKRRHNVVDILSSSDS
SEQPLDLRKDLDPVPLSSTRAQSLSIS*
>c18786_g2_i1|m.6070
PLGAVGKYRIRRKFPLPRTIWDGEETSYCFKERSRGVLREYYRRNPYPSPREKRELANGTDLTVVQVSNWFKNRRQRDRALENR
SPSQCRAGISSTQTSSNGYQSNTSPPSSNDSECEKVNSVSSSTGHFGPTTPAVSQPLPRRLPGSSLRLARIKQPHSEHQNHSEHI
DRQMTRLIRIPTTQMMQANDDLQSLIKSESAMSKSPTRPNCLDLSRSSQQTPYSHQPCSDSSSLYMSMLMEQDDGADAELV
DHEALIHEFLAEKMLEE
>c26821_g2_i1|m.12879
AEYMGYLVTEHSTTPPVNNERFNSTLCHPENIKCEEPERANLPTDASSPTKETLLKFGASSPSQRPETTPASQPNSDPIYPQIQHQ
ISANLIHKINLYNMLRMMPNQLKSKESSLQQTPLNMDTHHRDTSPIGKNLLSWHRFNAVGSAKKLHQHHAMAVVETNGGEQ
QLRPEMVTVNDFSAKYVHTSSADDESLLAKRRRTRTNFTSYQLEELETAFLTGHYPDVFVRESLAMRLSLMESRVQVWFQNR
>c37098_g1_i1|m.53235
MPPVFNSLVHSATHLADMSIPWSSNHIAPNNPASAGISPFYFGNLSQNSLHSMNHSIQTSGYHALDSKFPMTDEEHPNHFW
HNCYHSDQYGNSPRYNLHATSSNHPFYPISSATHEKLTHFQPPPPLQSYEGQIMPGRSPFIPTNQMYDEMTSYPLTKFKEATTG
WHHHGSKIMGESFTTSPNCSEEEEGQASPSSGRLCTGVTAAEVDHHSTSSLTPHSLLSSNSEIPTMPLVVAHPSSQHQEQHPW
LQRNGARKKRRPYTKNQTLELEKEFVYTAYITRERRQELAESLHLSDRQVKIWFQNRRMKNKKISCNGAGIMGSQQQNLNTLA
GNGLMVQNMIQI*
>c27878_g1_i1|m.14600
MEKLSLSKNVESNTKVVKIRDANGNIKELIVPRALDLDRPKRDRTQFTSSQLDALESSFRKRAYMIGEDRRELADELGLSQIQVKV
WFQNRRTKSRKMQAIPTSRTTINCQTEAEMNCSDSTLNRSFTSDKCSFVDKIVSLQGPIENDVSKIRTNTPHKYHQGGAFPIQD
DDIY
>c31343_g1_i2|m.23216
MPRFKFQTPNGTMRLELDESLPIAQVKYLLSQSLGRKVIAIRNRPANQIVTKTQSITISDEMRMRDLNQTSCLIVQCEKVNQPNE
LPSTKSPPSSHNVSRGDSNNFAKLSQQHASKNLSPETTYLPFFASMCQDQGGASSSPSFGEDNKLQLPFNPLLLNPAVLKYAIHF
EVDQFRINSQIFCKAVLNTTDVKVLENILDLAPGEFLDSTGKFLPVIKFLATKSVATRKEIYERIRSDLEKSTQPSTSETRTPPMTVEP
APNPQTPVSSDTQLSSVVSNTMRSVLAQSIAQQWLALQQGAAAAAISATNNSQQSQHCANTQHSLVSTTNSRMPMTCSPFK
NSPTFVVSNSSLPPAPFPDNFETPLLNGEGRPPGHSCNNDLMVKTSATPSQNFDWSQRMRRRTTISKNAKRFMYEWYISNGR
YVTNEGRIEMAQKLSLTEENVRVFFKNMRRLEKKGEHAGIPLSAMMSLDSEPGEGSSSLMMMEHEEDEFICLDTPSEENNLDI
KLEIN*
>c36403_g1_i2|m.48279
MSVDHQLDELSPPLVDAAEEEENEEDQIYEAVDKFVAKVRDELYSFVDKVRESELCPEQSLPSSSHFDSGSILNKVDPEFQAIQTK
LRTLVNSQDKCMRMTDKDLVQIETALESSIPQTSLQLHPLIVMSKSSSSNLPKFLQPSSMQQVVLVEPAQTEPVIKAGSDATYLT
TKTGPVTVTPLKVTRIPQPSEPIPATQRNDSPQIFSEEPLQLTQLEAFSKDFKMRRTRLGFTQGDVGLSLGKMYGSDFSQTTISRF
EALNLSFKNMCKLKPVLERWIEDAESHLTLKAPTPTVEYDAATTPIPNNISQHGVTRQRKRKKRITIDSTMKIALERSYALNEKPT
ADEIAVLAHGLGIHFETIRIWFSNRRQRTKRFAPLDHMDDVSCFPTNTDDVNELQNSTVS*
>c35149_g1_i1|m.40575

QRRVHGEKQIATDQILLHNLFCLMMIAPTAQSQQPQLSSCARSMDVHFRTNPLAKMLEGFRPEGYPLHNNNHWPNACQFEY
SSAPPTSPNEYAMGQTSCPPAMHQPLPAAEKANKLTNTTLCTGCGGAIFTQFVLRVSDLSSNHAPTVYHPECLKCVECHVGLT
ESCFVRSGAVYCRNDYNRLFQSTPTSNSPSPSSRTCSKCGLGFKPTQSVMKLSQSDYYHYPSCFQCVVCQASLKTGDQVAIRKA
GLVCTKDLDYFDSATAAKPAGNISSKPLPEEEACSRENKMLQQHFNSPTSSDLNDDECLFDDELDMDEDDSAVTFSGTSHHKN
KKSKHKEGGKPTRVRTVLNEKQLLTLRTCYTANPRPDALMREQLVEMTNLSPRVIRVWFQNKRCKDKKRCIALKQLQQLQAEK
NVNTMSGVALVAESPIRHKQQMDTYKFSPKGHQLDDSYMLSFDSSLDSLESDSLENGAKTLSQDLLQFS*
>c34463_g1_i1|m.36624
MVVWQDSSLDQPPMNNQSNEGFTLNGFRNASGVVTPTHVPPSSTPLENGPSIQLPGSNTNIAFTNVDADTRKQQLQEILQQ
VLSITEETLDEAQERKHTLNCHRMKPALFSVLCEMKEKKGTYLNMRNAQEEETPDSQLVRLDNMLDAEGVSTSDKCSPASIPG
GGPGQPDNVIEHNDYKAKLMQIRQIYNRELEKYSQACQEFTTHVKNLLTDQSRTRPISPKEIDRMVKIIEKKFQSIQVQLKQSTC
EAVMILRSRFLDARRKRRNFSKQASEILNEYFYSHLKNPYPPEDVKEALAKQCGLSVSQISNWFGNKRIRYKKNIDAEASGVYVD
DNATEGGMGSSGADLFGQIDSFDGQHQDGLLTPLDQSNYNQQGAFQNDAGVSQPKMMSSVGGGRPSSTSALNQHLVSSA
SIRPASAEVYSAAGLPFNNAA*
>c29779_g1_i1|m.18825
NAEYMGTSNSKRLFNYFKTLSPEVQTLLCAMSNNESFSEYKMLIPKENFSTLNSAGANRSNFIDGTPRTLHSHIHAIDQNSNAM
MKKKRKRGNLPKTSVRILRRWLQDHRFNAYPTESEKALLAETTNLSILQVCNWFINARRRILPEMIKREQGSDSLNFALSRNLDS
NTVSTARSFPTEKSSPQITRSLRSVSSPIFSFDLHHVNRMLSHGEALFEGGSRSANLLRNSTTSSSSSPTLVNYPTPLILPTTSCSPPT
DVQTKESSTTSKESCGGATVAKMRTGVETTLHSDTCKSSPWQPWVDHEGPSASKPLTPLVDEILPKPECIECSLANQISQFLHLN
STLNALYPVSKLIPSPATTHSD*
>c34016_g3_i1|m.34415
MQTQTMDANVDIKHLMPDFYYQQAASPYQQQNGSRDGTTSGFFPSNVIDCIQAWQNQYYNEEFQTTNEQTLSPADFYPPV
HQTSWHHSNMAQRQLWDRYERQRDILAEEDYSPTYDEDHRESLPETESERFNNYPPPKIECLQMDKSEINESDDLTKSTTEPQ
QKSDSKQTYSWMKVRRSHQQAATNQGEEDRVKLNKSKNNDLVHSNHKQNNNRTMPLANSSANQQQQHSLANQMRGGR
TNFTNKQLTELEKEFHFNRYLTRARRIEIAALLHLNETQVKIWFQNRRMKQKKLVKEGKLIVDKY*
>c33794_g1_i1|m.33297
MTIEGLRWPDHLNNNITNFKSLPPDPLRLPPFIQCETNAAFAPQAPYQMGAPPAFHHMTQQPEFACNIQNFQFPFREICTGCT
LPIMERFLLRVGDGTWHEGCLKCTACKCLLDNTCFIRLGKPYCRFDYDRLYAGKCSGCLKTISPDELVTRVSGGIVYHVSCFECVV
CKNVLKHGDEFVLNDELVYCKNDFEKFMARRNLSGGPAAALTMGDIPNNLMSEHHHHHPSQHKDVMHLDIPSTPKNIKSQP
NSNSGESGKENSSASPADESLGEYDLDLMMDDDEDEDEKQSSQPQKPCRKRVILTTQQRRAFKATFDISSKPCRKVREQLARE
TGLSVRVVQVWFQNERAKVKKLAKRQHHHHQNSVGGGNAGQGGPFEGLSPDSRSELQGTFLYMNEAGFPQAPPSNQLDP
RFSNMILPPPHMQATGGGGDFFLPGAGRRSQPSGQMFRHSIPGVDNVMRSSPQFTPNMDQCPNGISVETPRSPADLHNSD
ELDKLYSIPASPLPYFPDEHAKTIS*
>c13792_g1_i1|m.3393
VFSQTTICRFEALQLSFKNMCKLRPLLRKWLDEAENNNGCPGALDKIASQGRKRKKRTSIEVSIKGALEGHFGLQPKPTAAEIMA
LANSLQLEKEVVRVWFCNRRQKEKRMTGCPLGSPGAINQTTSDDEDD
>c33160_g1_i3|m.30481
MSSPNEQRHDCVNEHDEDVLRIGSQEHEDSPQRSQSQNFLPDKQAHPPFPSQPIHHPFTNGEAALPHQIPYYPNHLDAINSSP
YVNIGSFWAQPDPAWFNSTDSRMCRRSRTVFSDQQLIILEKSFESHRYLTNGLRLKLSTELGLTTNQIKTWYQNRRMKWKKQL
IDAGESNIPTKSKGRPRKDTEGEVILTL*
>c29014_g1_i4|m.17033
RVHGDSFTLTMIKTARAAMEFSHAETERAAMFSLSNLMANAGCPQTILDHQQQSVKLPAILPTAFGNENNTSLALLYQQAYLD
QYLKSCYDNLALNMYPQNQREKSALLFNPPVNSHLTGLMHPENPNDERGMKKSKRIRTAFTNAQICRLEDEFEQCQYVVGRE
RTRMAEELCLSETQIKVWFQNRRTKHKRSVHSRELPGTSSDDENLDSSSNGGFVSLSRNLPNMALMSADNGGDNHRLGRF*
>c33581_g1_i1|m.32277
MLNPNANFLLPKQDDSGTLDNSCIENGGKCTTLMSIGSVGGYPYLSGGNGSSTNHYQQQCKYASTAAAQNTTTHAGSLFATP
PHGVGGGGAATTGHFDGTFFRPDFFMDACNKAAQQFGKTDVISPYVRSGHASNVSPFAVPHQNSTSAGGFCDHLFSFDHLS
SNPFTQQAMLTNEMIAGSQMSHAFQQHHHQPVNQLLHVSSPSTLSSFSHISSSTPQSNYGNATGSHVTVAALASTNALAAH
HLASNAFSADVSQSNPRDLEAFAERFKQRRIKLGVTQADVGKSLQHLKIAGVGSLSQSTICRFESLTLSHNNMVALKPVLEAWL
EEAEIAMATNRLCMVPNNAATVAAQNMAALTGGDISSSMHQHMHVSGDRQSVKLEPGMTISCCGEPGGILSDIEHKRKRTSI
AAPEKRSLEAYFAVQPRPSGEKISAIAEKLDLRKNVVRVWFCNQRQKQKRLKFSAIRT*
>c14975_g1_i1|m.3975
MAKSEQFRITNLLQLNDEGTKREEQQEKISSAAAAVVASTLAERRSCYLARARSQKRMRVLFSQKQVDALEDVYMERSYLSSSE
RQALANAIALTPSQVKIWFQNRRYKEKRNKHDNKVINLEETVRQLELEQLQRKQQQQHSEQHSSQQAMWQQYVCCNASAA
NPMHYLNVLKHMSSMQCPPPPPPQISVSENNNYKPSVYHYQH*
>c34582_g1_i1|m.37287
PFENKLQVVCGFQSPHTTLYPIRVAEMNNQAQFPHHDIRLLSQNTDSFQDGGMLSSYPRCSSCEREILEGNYLSVLEKAWHMG
CIRCSGCQCQLDIKCYTRDGRIYCQNDFFSYAWCQCSACGNSINSDDLVHQVQDFVFHTACVACHLCPHESRLLGPGDRIYIVH

NPRQILCTKHYYQLYVEQKTMEEILNPPQSKPPVLLYPEQQSEHFYPPQNDPIHQIAQPTFDDTDEEAAAVSTPAMMPPSPAQ
SPIPHDEVSSSDQPLLQTEPAAVIPTFDDQDVNNNSSSANEHQSENLTANLTPMLVDRKKKRKRPKKSPQVANETPGENHPNE
SRTDPSPKKKETGTSPHSLKRSHLSKSQVLVLREAFEVNPRPTREERDKIKAECGVENRVIQVWFQNRRSKQRKNPDNKMSPLK
ETNLKSPTPEGCKSPVIQAEEELHGADWSPPQVTFE*
>c30745_g1_i1|m.21253
MQLDILNRGGNEAFTIERLLCSEAIKDSHHHRFNEDTRPSVGASVSVPISDFELKLPNHFKSPRATQIGTLHHPAPPRIPIHHASP
HSSVRNKKKPRTAFTSDQILHLERRFRNQKYLSASERLDFAMAISLSEIQVKTWFQNRRMKHKRQLQDNENAQFAALIAAYSPA
TKTPEVISTEFALWKQHVTKQSAIKPPFHLPMSTLGVPNEFACPPQLVVNPFTILHTLTPKRH*
>c35965_g1_i2|m.45595
MWDKSLNDTCSWFQNKVNWTAPNSQNSPASATKYSEQRWGETTSLRSDDLLDIKLVEISAFKKAHPVVRGGWASVFCATCL
RTAVAPAYFRWWSEEASCNSHACLVVFSWFLAFIPLAVNYELFGQSCKLLIVGQMEPSEIFAPPLYLFLITLLFTHIASTQKVKPGR
ASSTTNNVVIGNNSLSTACSSFERSSTSHSSNLINSSNSGLDRFNFGGNSSRHSTSPSSKYRRVVEDFAEVDERSEPNLSATRSFSF
ESAAEVLSEPSLAHCSADGSGLAADAASAKHSTHHAFGGTLRRRKNHIVNKEGSHTSRAAKRLRKPSLEQQASFSQPPSFPPLH
RLNENLTERQFSTLFDKPLSLWERICNRLITIWMNIKYYTLLIFSCGYYRNRSVLTHIASNASRQSSNDVNFNWDANNRASPKAE
PLHKKASAHTIIGDGEDSEEDDEEEAVNLSSEYEEFEQNSDNELPAHATGSSSSPGGSPAKSPQHPHNEKTINSANDDDVIVGES
GKIRSGRRVDFLNNLKGLECEQMRANSPPSYERSTQLSEHGRFAGIPSKRSLTPPIQQPPATVSALVGFPNDPKKTELTAEQICNL
VVEHSHNAVRYSPFYFCVGLALSAVLPCVPLYYKVNSNGTANGTQFCQMEQPITSGFNELCTLRSDSMSCESYVKYTRHIVSLTS
LLASFIPLCSFFFLLSVAERTYSQRYYYSKFLYQMISSREARKYNLPHFRLNKVKNIKMWLVIRSLLKKQGPRRPVEVIVSDSFLLLV
MLLSHLCLQRLFHHTDYKSDWIFTQLMSESSQTLLIVWILALWLYILRSITLGSKVNKRYSNKATLIIERINLSLRAEKKPAKREELW
LADHVLQLAVELIGEIESPPFKISGLVMNSALYNFTKVIVASAFSAVISDVFGFKLKLWKLL*
>c29539_g1_i1|m.18213
MLPNIVPITFYPPPTNNGNMQAGSSSNTCGWPPSNQLISDCSNYTQFVTNSQYGYPTQLHPANMTSGGIYGSANVSTTATHP
AFAANNHQFANSAGNQSILLHNTSNHHPNFDSYQVLFGPSSQTNHFNTTGGIFPAAAGIKHDDFSALSLHSSLDASDNMLND
ATSNIKDIKQETAKLDSSDEKEDSFKDEDSDDKRPRRQRTHFTSNQLQELESLFARNRYPDMATREEMSNWTNLPESKVRIWF
KNRRAKWRKKERNLLQGYQNTEAAFRAAAGFATSYDLYTSPLTGYSAYNTWASKMAASVVQPNICGSKTPTSSSGVGTGGAN
STCFPWGGFNTQTTNNQTYNMATYNNATTSYNNTNDGPSSNKSSSPNEVLTAATTPNQISTTKTKLW*
>c26109_g1_i2|m.11867
TQSTWGRIGISFKDEPSIHVPANNHPLASFVQSSQCNPSVAPPSIQFSNQNCLNTNQLEILAQITNARQIAAGLAANLAPKVSTA
SPIGGFPIAQQAPGVFLRPNANFNVPNVAAPYGLANGCRPPFPPGIDFMHHPDYCYPGMDMNAALARRKNATRETTSALKC
WLNDHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKMNGNLDMLDTDFKEDDDDLTPNDDMSYLSPSPCPP
YDFEPTNERPKLTRDSPNHIWSLAELATSPNCETNQTTTPVTPPPNESCHAINFSSQLRFLRDQQHASHQLQLASNMLLAFKPD
SNLLMSKSFVDNPKRLGNNS*
>c29396_g1_i2|m.17852
STQSTYLLIRHLFCSKPPYTTYVPYSKMSDSFALKTSEPTSHYSIENLLGLDNVQVEITGDKLLNADEYELTSPNDGKKHRRNRTTF
TTYQLHQLERAFEKSHYPDVYSREELANRVQLPEVRVQVWFQNRRAKWRRQEKLDESQSGDSNSATGFEAIAGNGGLGYNA
DFPLESLKQNALKRQQQLLLNEALAAAPQYSSDQMLVASYVHSLLAPFLTNNLECANYPLASGDQTKTTYELAENNLTNDPRN
NREIKSASPFPSLQMESMQK*
>c27275_g1_i1|m.13532
MNQQHNHRFKSIEDLLKEPSPVPLGNEDFKAEDAARLYYNNVLTNLLSRTTTPTIHYATSPIVTSSNIHRDCIASSARHLDLDLAIN
AKKKSRYRTTFSAQQLEEMENKFLQTPYPDMFAREELASKLCLSEARVQVWFQNRRAKWRKKQNITSPSANITTANHSNDKH
LLTTSAIMKYYTEISSNKTTSGAFDPSTLPQIYLNQLHSLLHNENNCH*
>c33358_g1_i8|m.31322
QRRVHGRPFSFNISCHKSQPAMITPGRILNAAITSTPRRSPMQGHFMSTNLRNMTNIDSSTFFPAIKRPFTFPAVDENLMENKK
QRIQNGMQAFLPIHNPFFPTAIRSELNVSTTGDTSLEHDEIFSPVDEDDSSPSAANAAPNCGHCQKAITTPSLLQVNDQYWHIG
CLKCSICSIDMSSQKSCYLKNGKLCCKDDYESQNFAVCTRCDRPVSPQDWVRRARSLVFHLACFACDVCARQMSTGEEFALQE
SSKILCKQHYLELMGETWPVPKENEQVTYADDGPDSSNEKISPGDLSLDGSSSQNGKSGNNKRSRTTFSEDQLRVLQASFNSDS
NPDGADLERIGELTGLSKRVIQVWFQNARSRQKKHQPHHGSINSPYGSNISSYLNESFHSTSPPSRREL*
>c33358_g1_i3|m.31315
GINAEFALQESSKILCKQHYLELMGETWPVPKENEQVTYADDGPDSSNEKISPGDLSLDGSSSQNGKSGNNKRSRTTFSEDQLR
VLQASFNSDSNPDGADLERIGELTGLSKRVIQVWFQNARSRQKKHQPHHGSINSPYGSNISSYLNESFHSTSPPSRREL*
>c2160_g1_i1|m.439
MELQSSKPPAEHPCSEDDMDEAEMVAGDPKSQSNQIFAWMKRNHPTNESKRTRTAYTRYQTLELEKEFHFNRYLTRRRRIEIA
GLLGLSERQIKIWFQNRRMKWKKDNNLKRMSQVDI
>c30021_g1_i2|m.19472
INAEDMGRIKMSFSADLTRDKEAMDKSTEGRKPVKSFTIDNLLSPDKEDDLKLKCTPPSIASFTTSKATDLFPMPTKCLDRQFLH
NYYEHHPTQCFGRLPLLPHTQSASAAGRLASPGNQSKKKIRKARTAFTDDQLAELERSFDQHKYLSVQDRLELAGRLRLNDTQV

KTWYQNRRTKWKRQASVSLEILAAEALSVTSRSMYSFERGTELGLRYLESMHRNMITRNKPNESSHHKQSINPTVSATEHTILS
PIASKNGQMLNAGFTSRNYYLPPIDKKIPATTLL*
>c32992_g1_i1|m.29659
QMAEDINQNLFMKDETENLCAGCGQQIFDRYMMRALDRSWHSNCIQCCDCRQQLADRCYTRDGLILCKEDFTRRYSTRCYG
CGDPIMPNQLLRRVRQVAYHITCLTCTACFRLLSTGDQVYMTRDNQIICHNDYMLQNENSPDRENLKPSPQPNMSPFDTFQS
ANEKTLIHQDNEALKMKFEASQFAHVVEMGGCETIQHEMQNSSKLDEEEDMLEEEMIEFSDEEDEGDSDDSTESNEYKGKCT
EGNKLDLDKLQSGDGMGKCNPAGSKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKER
RMKQLSVLGARRAAFFHRRLRSFDHHRDLLHQQQLPFFQHAMVAHPGGEMFPQAPNAFEMYGGNPSGPGQMQLPTNRF
MLPPQLSQQQATSADEFEEQMRFHSNKLEEASRSEEMQAQQPSFPDSSQTSDSLLSTAANNNHHQHQILDSSLW*
>c36492_g1_i2|m.48837
MNAAGTGGSTQLGGEQHLIISNGGGGEDAALRQAAGIFDPTSSAKQIEAFGDLYVDSSSASSFRTNSSGYSSGSSCSSVYGAAD
VLPQDGLPPPHAVAASSSATSYGVPSYDQFKDDIYEHELVPLLHVLFEKCELATCTPRESGRENMDVCSKESIKEDILDCMQQIG
QSGEKMRPFNNDLDTLMLGALENLCIHLCEIQKVHELCDDFCERYKGNLQTTMPRDLVVDEKSASASPTPPPMGITGNGHQS
NETASGAGGATAMEEHGGRRIPPGGSPKSPKAVSFYDDEAPPPVVPNSPQQHQHHHFHHLTSPAAPTAPTAAAAYLHHQHA
ADASATAAMNHHHMQPLQPAQLFSPYLAAAQVQGMQPAASHNLQHYGVPTQFYHPSMFTTPQQTNQPSNHYQHIQNSD
LVYPYPLLDCHQTQMATPGSLAPGGGKYSSSPPPPSAIAAAQLQHSIENHHHHRMSPPSPLNRIGLKLGRCVDMSPTSRCSPP
DDEDDAMSETSSSAATDMLHQVAADSGAAPPSQHASGGAKKPQKKRGIFPKAATNIMRAWLFQHLNHPYPSEEQKKHLAQ
ETGLTILQVNNWFINARRRIVQPMIDQSNRAASGFMQASAAAEHNFGIDTQQQSNYFRNTAGMISEMGSFGGYAFGPPSSA
AAAAQSAPADMLPHQQHNHTAAASLMSLFPSNHYNNQVSNSNLMEPINTLDIQT*
>c34423_g1_i5|m.36442
RVHGIFNELSARLSPAMMQVQKRCGFMVEDLIAPTYKDNDEPLDFTNHHSRTAAAAALIFASPHKPFPCNPNGSMLLESPSPC
KKRKRRVLFSKTQTYELEKRFQRDKYLSAPEREILAASISLTPTQVKIWFQNHRYKCKKSTTEDGEEANSDSGEDFAPSPRKSHQH
AVTEYGYHHHANSRAVTDLEFKSLQQQQRTSHSVVNNTNKVTSATALAVQQLLRLKPYDNGNNSWWRSQLSLVALHQQQQ
QRKLEFNIGTSSNKHNLHHQHLLQKQKIESHQRDSSHIKD*
>c37377_g1_i1|m.55305
AEYMGNADSILQHLSSRHSYEISPTEKQYIDQSDCSIALYPGVSNEVAVIMFWIENVINEEKDYNDEPTQCGEMTPDDISSNTTL
DNKTAKNSTNHVRNSCKTLKCPKCNWHYKYQQTLEAHMKEKHGSSLNPGCLHCLRGEAHPRLSRGESYVCGYKPFRCDICNY
STTTKGNLSIHNQSDKHLNNVQEMSKIPSSTTQPLILEETIRMEEQVIASPTKVNGSSSKPPTTKSVWRCDACDYETSIPRNLKIH
NSSEKHAHNALSISLQQNKPQIQEVSRDQTMLHTEEVGGKLYSCLLCNQYSSEDLAMMEAHAMVDRTGTASLDCLVYDGASF
KCQLCNYTTGLKANFTLHSQTEKHKQKIQIYNHMKEGPSNLPSSLMAMASRMSSIHIQCRACGVTAKSIGQLHAHATSEIHQL
NAKLFEFSNRFGDGAKFECRICAITLTSKEMIAHVSSSQHCELLKQKHNTSSLSIKSWQDLFQVLPVEQETFEVRKDSDVVIGETF
PCPLCSKGRSFKSRELLREHLANSHKVSLECIDSIMMVVKPTMSVKPIARPVLEDAAPNLPFKCHRCHAKFGDVRELATHVSTCI
RAPCPLVGCEFSGRDDQIMNHIWTDHDGQGEVAMKFSCWKSDCTMGRFRTEEALATHFAEIHSEYENEVTSLNSSISNAISCE
NCCVRFGSVVDLQAHVLAERCIANTDANRQPPANQQSGSKPYSCSICDVKYAEKSTLDIHLNSVLHQSKVLENKSSNLTDLKAE
QNSSEMFQQTVNSSFQPFDFQSQQQLQLSILQQQQQFLIQQQKLMSSLLPFYAQQQFNGGAGGGVTLCPKCTQILPNAEALL
THTCPEEEPKTNSNQSADPISVTRSEQTGGESECKQMLEHLVGEEGDESQQSQFVCNKCQSSFLGRNAIKSHLEEAHKIAVTLE
NIDKYACEIPKAASSTSVSPPSAKKAKLEEPFSASANPNAWDLAQLFQQSVLQQQQQQQAQAQQQHRVSSSLIPTLAYLQNAA
AASLYTQEQHQQQQEEDSTSAGYKLEDYFDFADSGGKRMRTKITDSQLAILKRNFDLSNTPSEEEIERISIQTELPAKVVKHWYR
NTLFKERQKNKDSPYNFSNPPASSSNATISNYLTQGNKASLADYSMVAAQVAAATAIAYHNQQQESTSSCSTSNSSQQNEMA
APGIVSSLSHGGVDADFDSYADSDQTSDFSVATAQDMANEAAVCNSRRAGRTRFSDFQIKSLQEFFDKNAYPRDDDLERLSS
MLRLSPRVVVVWFQNQRQKARRSLEAMSSSSSTNYASQTNN
>c31776_g1_i2|m.24804
HNFTPVQFQLNENIIRTCFSCCGAIRDSHVYVTPLDQRCWHGACLKCAECKSELSNDLSCFAKEGKIYCRADYLRLFGQQRACSR
CGLSISPHELVMKTLQKLVFHINCFSCMTCSVLFVPGDHIGVRGGQIFCKNHWDKATLPPAFPEVLDMRPSLEDLQKKSMSPST
TQMPPFYTPLPQREATPRTPTSTTAAGQFESDQETPSSVKVPKPPKAKKKKSTGSAVLNHNTSSNSTNSSTNTLQMVNNCQSIS
DQGFTNIECEASYSLSGGSLRTKRMRTSFKQNQLLSMKQYFNSNHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKYRRNLI
REQVQSNRKDKEALTILDGELKAGMHEDVVEGGFTDDDDEEDCEELIAGGSLTSYSPLSSDEFLNSD*
>c14788_g1_i1|m.3892
LIRDSNGGVKELLFPQALDLDRPKRDRTNFSQPQLDKLEEAFRSNRYTIGKDRKILAVNLGLSETQVKVWFQNRRTKYKKDELTP
DISAKESSLNAYLDDLTTKDYKSPAQELEIASNRILGDICSFNTFKSPLQLHL*
>c2747_g1_i1|m.593
RRPPATSIQEGNNMWINNNGMDINLLSLYYCHNLIPNNSSSLADHASSGGGGWGNAAAPSSDPVYHGVNNGYMDQIRELQ
HSFASGRPFGNLSANLAIYGNMAGCTQNPLGGRRNRTAFTSHQLLNLEREFDNNKYLTRLRRIEIATNLDLTEKQIKIWFQNRR
VKDKKRLVIPMQPPPHANQRHLHRQQQEIIKRHRLMNNEHDDVSGMKEEEDDDDEEQESTSLAAANFAANSTMILNGELAD
DDEMTISSRRLLLDNCHLQQIFARGREDNLMPAATCTSMMTQQLMLQSPDSPLNLKQTVGSTGGVCVNFQD*
>c30900_g1_i2|m.21728

MNYGQQQANLWSFVQNYNQTLTKQGMQFKGVAPHNNSTAPDLVCRVGDYSINRLLGRNETEAKCAPFAKTYLLQQPSQSS
NYDKLSKITEAIPSTENLMIDRSELPLLEESDGINKSRLCADRKHTRPTFSGQQIFALEKTFEQTKYLAGPERARLAYALSMSESQV
KVWFQNRRTKWRKKTAADVAAQRICNEFEQQQQNQQPNNDISVQPTSPYNYSSSQNILDHFDE*
>c7930_g1_i2|m.2049
SKPKRFRTRFTPSQLNELERCFGKTHYPDIFMREELAMKILLTESRVQVWFQNRRAKWKKRKRTSTTNHSGSLGGGHYHSVTG
AAAGHRSRENTTPLFGEQPNSFFPHQGIFIST*
>c18082_g1_i1|m.5652
QRRVHGVPNFVEGSAPSEFGDFQPSNENIPATGGYLIPQQQSQPSGEEMQPFGNKLCYYDTTLPPQYQPLNYGVGVRSQAD
QISKYTSGLVDFHQPFTNYPDLSTVTAAMYQNPNFYHRLPNEMMSSYAAVANHCSSQLLGVTTTETPGYLFPPNYSNVVEPYP
NPYPHLTPEDDQQSHRLEEDEGSPTQMLLHHHHHHQAGVEKLTPNKRSQPGPKSYNKKLRKPRTIYTSQQLQQLNVEFQKTQ
YLALPERAHLAAILGLTQTQVKIWFQNRRSKDKKGIKHPGTASAIMYMQPTATAQTQNPMYSA
>c30188_g1_i1|m.19846
MMSRSDRCDENSNLKRPSPIVDLSRHFTSTPAAAQVANPYHTEHISAILSPYSQLLAQMLINNQQHLQKFGPMEWDNDGDRL
FDYGLHNEKLNNRADFLMEDEALYQTKVEPREFSPETQHNGEFNCDALRQERPNSPNMAAKMRNFNRRRHRTIFGTEQVDE
LEKTFEQAHYPDVVTREKLSARTGLPEDRIQVWFQNRRAKWRKTEKCWGKSSIMAEYGLYGAMVRHSLPLPAAILKAATKQA
RREKKCLSSPDIEGTNDHISHKRQKLTAADQDDSESNEEDCEEGDDESIAPWLLGMHKKTLESLSREQQLQSE*
>c22519_g1_i2|m.8420
IGPTGVVRRASDHVYHIACFVCLVCQERLETGEQFFLMDNHKLVCKQDYEIAKSKDFTELESSNGSNKRPRTTIIAKQLDFLKTSY
KTSPKPARHVREHLARQTGLEMRVVQVWFQNRRAKEKRLKKDANRRLGNGSQIYQMIPSNVSKSSSKRSKKIKTESKPTKIEPD
DDEQNLTAAIEQCSNVADKLRDQDENKSMLN
>c28127_g1_i2|m.15108
MTIEVSDNKRSKIFSINDILSGTNRVTPDDSPTNFSEPTEISPDEKCICIEGEAEQSKSDDQNKKSRTVFSRSQVNHLEAIFESKRYL
SSAERVGLSKQLHLTETQIKIWFQNRRNKMKRQFKCMEGSTIAGGFNSAPPNVNAQQMIALSPNSYLGKDPHNLDYFANRNL
QAAANGAVITALGQTSTPGIHGPTAVPSFVHMPPSGLHGTTNASYTSGIAPPSHFNPQAVTMQLAAPRLPMPNNYFLVGHCA
SPGGLPLLMGTNSHSAPADFSTSKMYKLQATKL*
>c52679_g1_i1|m.65772
MALSPKTTSFSVTNILSPVEEQQQALLAAVEKFHDGSYGSQRTSNGLPTMVETAETGIERATLTCNDFGYPSSGASTDTLLAPPS
SQYFPSYMFQTATAATGVGTNGSATAITGGYPQQSDSLAMAPPQSAPSFHVSPSSSTTTNYIPPQQQTGIMNMSAAEHLANS
RSWGCYTANDPRLSWIGFPGYGFQDASKILPQNIQRRKRRVLFSQHQVFELEKRFKSQKYLSAPERETLAQMINLSPTQVKIWF
QNHRYKCKRASKEKHQQQHQSRAAAAAHHHHHPHQLHHQQQAVLQHSGHEAAGGAILFKTEDEEDESEEEDDLAHEMCD
LSSGDKMSGMPYLTHHQHHTMADSQQQHLLLHQPHVLQQQQSLHDNLIQQQQQMAHHQHLLQQQAMTDSTSASCYGA
GIPPVVPGLNLSAAAAESLAAFQQSSVSCFGQRVW*
>c23969_g2_i1|m.9598
LPSSTNQTAHGFDVVPKEHINGFANYAPANWNQCGQEDVIIQPTAISSNMQSVGTDIQYYVPAYFNEQYSDEYLNGEEAIYAP
GNLKFDQLATATNMAHWSIDDNQSFGKIAYNTPLQPNYPFPQRLEQPCHIRRKNATREATAPLKCWLLEHRNNPYPTKGEKIS
LAIITKMTLTQVSTWFANARRRLKKENKMTWIPRNKSTTSAHTKSRTSSEQNCHNQDETSNSYDCLVSNHGLESD
>c28601_g1_i2|m.16051
MAANEKNLEDILGSHQFQAPSAAASMQRPYPIFATNPWIQQGLTPNIGNPAQKKRPTLKSSNGEMDAQQTGGAEKLKQMM
GHSDKEFAHNIGMLRGEKPEDGNESQSASETESICGEDDDDDDKCDEQGRKKKTRTVFSRSQVYQLESTFDVKRYLSSSERASL
ANSLHLTETQVKIWFQNRRNKWKRQIAAELEVSSVPPSAGSILGTLNQGRNTNNARNSSNGGSRTSNPNALIYHPIHGLSHFSP
NIHGGRNHRVPVVPPLYHHHSQLNKDSIDKRAAGCGSGSFFEQQSLLSQQLAAAISSTQAVPQPASSAAAAGAAMFGTVPSC
PVDNHTSESHETD*
>c27522_g1_i3|m.13930
TQSTWDMHQFPDLLGCSQMRPAVIAPAQQHTSEDAFSSLGNHSINAHLLGALRKDSLIGNNNMSILEKENSLHSEMPSHHDD
HQLKVNCWSPMMQHTECHGCGGEIIDRYYLQVTDMTWHTSCLRCRDCQQPMEGQSSCFAKDGFIYCKSDYVKRYIHQSCA
RCLNLIEPAEFVMRARNHIFHTDCFSCMTCGKLLITGDEFGMHGSLIYCQHHWLTDSTSLKMNPHSPTGQPACSPFGIKTSPGA
THSNQPVSGGNPALLVSPHFYGLPPNFPPTSTTIAQQMQQESELLSSSLQLMMGQKVSAGKGRPRKKKLLPPKLSDDAYLNPY
QNDCPNSSGDGALYTQSELSAAEQAQNRSKRVRTSFKNNQLKVLKQYFTLNHNPDAKELKQLSQKTGLSKRVLQVWFQNAR
AKYRRSSSKSGMGDCDSIDMIGKGDDFSPSQSELDEGFSPGSHSNDSLPDSPQNMTDDQDMKMFMPVAAAMGGDIVNHT
NFYIPS*
>c31844_g1_i1|m.25019
MNLYNEHMNSYFKGVNPPTASSLLSGTQSPYSAHYSQPFSASDQLLHPITYPALAQQMIGIQPSSTHGSHYPHGAGRKQRRER
TTFTRAQLEILEGLFAKTRYPDIFMREEVASKIALPESRVQVWFKNRRAKCRQQQTSAGGGAKTSPNSSDEKSATSLHLHQSKQ
QKIDVVTSPPDNASSPYKENSSPLGVDDDSPKSADGAGGILPSANAVLYSTYPGQQGYAGSAQSSGFLWNPYQFEVASNQSR
NFYNQSLQIPAPVSASYYPQQQQQPAASASSYHALSPTLSSYEAQSAGGGASYAPYFPNTSYMTAARMAYTQNTGAAADLFP
QNDLKQGGNNEQNNLATWKY*
>c14425_g1_i2|m.3666

NKNSLSFIGHYFGFNFDLKMIFPCLYCSERFDHVNALGMHIAECHPIEWSGLQRNMMNAQLSAAMVNAMSQQQTTTAQQQ
LESTLGSPGVMENCVTRLYNQLWAQQRVIGQLDLALREERERSASLEASLIALKKETSTHGVIQPPQHNDETPACGDELIQTFQ
ATFPTKQESRTNRRRIFKGVNRSCRRDMRDPRKCRVHKVYHAFENAATSNFGPTFGGKFNVEERAKLDESIQSKASNKDQEAS
SPFSALMTQHSPQSEVEESPTTDECHDLLIDLSNNGNKYDAENSSKISERSFSPGGEQNATTNLLSQKPRNRTRITENDLTYLQSY
FCKFQNPDKDDLQMISLDTHLPIRVIQVWFQNARRLSRNTCPHCGMVCSKNDTFLQHVQLCGQQTASDIEVV*
>c34897_g1_i3|m.39047
MPQKADEAVNQHHQYQQISASAGHSGVNQLGGMYVNGRPLPDTTRQRIVELAHTGARPCDISRVLQVSNGCVSKILGRFYE
TGSIKPKAIGGSKPRVATQGVVNKVAAYKRESPSIFAWEIRDKLVTDGACSAETVPSVSSINRVLRALQTDKLKMCGPGGMDA
QHHVAATMYDKFRLLHQQGSTATGQIAAANGFNQNQAACWAATHPWYNSAPCNQSVVSGQQMLHQSLMAVAAANQYP
GGETATNLLKMEGKDHFLSSILNDQPELAPKCEPSPLANPLNQHPNPSQHQQHTNHVSSTVEHSDNECDDEKKKAPVAERNR
TAFNQDQVEIMEKEFEISHYPDVYAREKLAQKTGLPESKIQVWFSNRRAKYRREEKIRTGSVTDELLAEEYHHHAISEQQHPNGF
VSSPHQTVPVQTAAVLYPSHHPAQQQQQTNYSGCYISFPTYSPYSGVNHLPTSAQPSSSAANNQQYCQNTTGYLSSAHGTLP
NNMQSVTAAADPLSSMQSYWQQRF*
>c31026_g1_i2|m.22145
MFTLPQFNPDSQNLSSVNMISEYPTAITSAGAKTGVEWKDALSRVGNQPQEVDQHDNMAAACSAQPNPWAAAYTSVMM
QRPDARILTDLYLKSPFHSPGLLSPAPWSPNKSALFPFSAAPICEAGSYVKTNEHFNQLRCEGFLQQNTLGTDASSFTSKDPQLLS
EDDEFGLSDGHLDRKKRRNRTTFTGPQLDEMERIFQKTHYPDVYVREQLALRCDLTEARVQVWFQNRRAKWRKKEKFQQQA
QMHTATSNNLFQCSPKRCGEVNANTSLSKITHEQNQFQNNTISMDNCPSISSFGNQFSSHSQNLESGTGGLAFGSRNSMVAN
DSCSEAAVSSSTAITNPSILFSAMTNCWQRNQNSFASLMNNVISPVQMGHFIANSQQFLSPSKES*
>c35630_g1_i1|m.43504
MHSWFWNERFWLPHNITWADLKNRHSQNAGVKSFDAANGLAMAIPLAVLIFGLRILFEYAVAKPLARRLTSKFAKANISKEFL
AHPNPILEKVYKSITREPDFDRLTGLSKQLDWSEIQVKSWFACRRARDKPSHAWKFADSAWRFIAHTSLAGWGILLLLPEHFRN
NLALCWTKYPDQAINSSLYKYYLFELSLHTSLLLVQVFTSHKRWIAADRRALFGALAYRSLLTSLIAASYIVRFIRIGTLALLCHDTSE
LMRHAHRLHRLLAPPTCTEKQLCITFWPFAIGYTLVRLVYFPFVVLRACLIDCALALGFFPAYFAMNCCLLAAQSAHVYWLAAIC
STALRLDSGLGNLLRPEHGDLLYIHNLHSIRQQNGVRPLNGCVNAKPSAAGDVAAAKSNGFGPHHSSYSNGSVVTTSGGRGK*
>c32768_g1_i1|m.28682
RVHGVVNLISQTMNSSQNLTIFPVAQTVRPKVQNKQSPSFLMKDILELPSQNSSPLLPIPDQSFGSTMEETSSEEEYETDEMGK
SSGKKNRTAFTPNQIYELEERFTLQRYLSASERADLAKKLTLTDTQVKTWFQNRRMKHKRQQMDRENSRLSEWFYHTAAGVS
QPQNPLLNVIGPQPQGPPQQHPLNQFIHNPFQPFMRPVLANLNTLDWMQTLKTVSGLGNLRGGGLPLVLPGANPASIGDQL
REDAPRLSTAPSCGDSGKPQLIVPLIHQAEISATTPSSEMAVSPVSASDVSPIIRHSSATPDDSA*
>c34667_g1_i7|m.37783
QSTWGVLGSHNAASIDSFFAPTQTFSSLNTSNGYTNYPQCLFRYPGSTVPGLAPPSVPPPTQFPNSPLTESLLEQANSYQSSKQL
QYQHQDQHQVHRESAPWIDYATFPHQEESQWDSSGCKYDSCSPQATDKEGKQFEQLASSSKDHSHTEYKQFEQLASSSKDH
SHTEYKQFEQLASSSKDHSNTEWSHSPKALPVATEEPSVLITPYYGNYQASGGGGSFNSLFYNPPDVNATGGSNNKLDFLTPPT
FDSGFGVQAPNFGNNTNGTYPNYYGSSQFLTPNNYTGSSSFYPSTFGCVTNSETTDPHHQDPSTPSFGASCGPTSYFTSPTCFN
GFQNSKRDFEVDERTNFSAAASSSFSARALLAKEVEEFETVEIAKKVAAELKRYSIPQAIFAERVLCRSQGTLSDLLRNPKPWLKLK
SGRETFRRMAKWLEEPEIQRMSSLRLAACKKKHSPREEESTFLRDKKSRLVFTELQRRTLRSIFNETQRPSKEMQAQIAMQMGL
EVATVGNFFMNARRRCSTERWKQGLMTNKAKNNAIKNSMELPPNSAPRYS*

11- *Eumecynostomum macrobursalium*

>c10092_g1_i1|m.13285
MDNRPKTFLSIDDILKNSNYATSGDNQTKNSSKDTFASKQHPSNKADAEIVNINGAINNFPILFNLYNNSNILRPNNPSLIPLHN
GFNFGANRKIKKESASRRQRTSFSADQTVKLEKEFQANEYITRSRRFELAECLGLSENQIKIWYQNRRAKDKRIEKAHFEVEYRRR
LNTVGSFLQNDSNCSTVLAHDILKATVIAAHHTKTHNQLLSSPMNFPKRS*
>c6557_g1_i1|m.7161
SSGINAEIFAHSTFGCDANTMEKKLVLAPGEFMDSSGKFLPIVTHFMKNAEEREKIYQSTKAMLSLCRPTTQTSIPKVAPQISVPT
VHNPPQPIPVTVTQTPDSLRQQTVVPTGTNHVSIPQSMVDPVTAQKVRDCFERAQEESRARRQAAQIRARENKLMQASNSYF
QVSGPTMVQMTPATSGEQHQIEQLNCIPSMRNTSPQQQQQSSRQSSAIGYEGFQQAQFLDPLHIQATQSQQHRPSSSNKQ
NSSSDQTGGENRRIRRRTTISRDAKKLMVDWYLNNGRYLTSDARQALAKLLNLTEENVRVFFKNLRRLEKKIGEEGGNLMDAV
NCPLDEQSCNETIQLLDDHSDETESPTLVKLENTDDSS*
>c1545_g1_i1|m.1138
VVSTQSPLGPVDKYRVRKKYPLPRTIWDGEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGNWFKNRRQR
DRAAAAKNRGSALNFSDEDSSRDKLRSFDDDDDDLDDLDLHGSDSEDENHDGDKSDRISKSPISDDASKKHDNNNKSSDPEKH
SSSTAKSLPSELGLKIPVTPSKPLSSKDTTPTSFVSPQSNDSTSPISIENRKFPLMKDNTEQLLSAAFGSSMYGAHNSANMAAAFS

AAAAQMMQQGHPANAMAMTHPGQTPQRLMLPQHALLLQAQAQAAAAQAAGRMGSNPCNVKLSPFSIPQSHPNVPITSP
LSSQNIHQFLRPNPLLLQSLAAATSVPSLTTTTSVTSPSTSPTKPSVFSPVTLCNIKDSK*
>c4091_g1_i1|m.3409
NAEYMGDMNGSSSDDGSTNSSMYCSTNQFPTPPSSTKNNPVISYCEGCHQAIEDKYFLQTTDYQFWHENCLKCRACGCYLD
QTCYLKNSQPYCKRDYEALISGKCSGCFEKIPSTELVMKVADNSYHLHCFTCSICHVVLHTGDQFVPVGNQIFCKIDFEKQYPEFL
QDDDSNSKPNQEDSKSKIKEEHSMDEEKCDDDQEDEALQNGSDNEADSPKDDNDSDDENDGDSSKDKNSLQQKPCRKRVIL
TTQQRRAFKAAFDMSSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKLAKRQQHQHHVDQTMTSPHMLHDDGQFG
MFISNDPYLHPAALDSRFSQQLAASQLAAVAASGESMFHHHINGGEMYQICGGGGPPNSQHETTVEYTNQDFLPSHMDMK
DEMYSSIKLPNSLIPNHPIDKLYSMQTQYFTPS*
>c6994_g1_i1|m.8060
QQQQQQQVPTSIAHSLGHVHGHGQCSCVACARPNSSNGLLTTTLQSNARSQLPLAQQPPSNLASALASGAVNPQLPAAAVA
QLLATVGANNNVQGPHPGLNLPTPAPLISFLPMMQHNQPGLPPHPAFFAQHSPFGHDGSFQPAAGVHGDWNSLGRRKNA
TRETTAALKSWLQEHMKNPYPTKGEKVMLAIITKMSLTQVSTWFANARRRLKKEGKMTWTPRNSFDDDEDDEDDDDDVTE
DDVMLLDKARANQGEDDDDKEDRLLIDVSVTEPPIKQLKTGRPHLNGDTERHEWCHEDDEGYNDDLFTPKAISPTKNAVTKM
EVPTLVSTPEETSKSPKFEEHQITKTKQANIKPKFWSIVDLTSKPVAEIKSRKSSLDNPPETIHRTSSTQLSTMAH*
>c1383_g1_i1|m.1020
DDDDDGDTDSISIDGSERGRKKRRNRTTFTQMQLEEMERIFQKTHYPDVYVREQLAVKCDLTEARVQVWFQNRRAKWRKKE
KHQQMRSGFPTTNNFPNEANALRQQNDPIAQHLTGSWAYGNNPVGNNTSNFLGTNPADLNATVTGRMNNATMNGYSSN
PASLFPNYMLAAQQAAAGYSCLPGTTNPGNLGTVNSQSSAMNLSSPGSATNAFYSNQAHLSGMPFFNGASFNHPSFMTAGS
FFNPGNTSAADLFNQQTPFRLNRMKDKTAMDQLINGVGVVGHPQANAFPSHLIASAAVAGNYK*
>c2923_g1_i1|m.2211
RRVHGGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPMIDQSNRAGMPANYTAADA
AQMGAFMLDGQQSPYLRHHPMVHAAAAAASLAHGAGAPNGLDANGAFGGYGAGVNGGHNFHHHHPSQAAMLIPGHPH
AAMMMQPHALHHQHYHHGQALMDPMAAHVMDTIHS*
>c7970_g1_i2|m.10743
MQEITTILCDSNSKALAGVEPTSYCYGCGSEIRDQFVNKVNDRCFHLKCLQCNVCRVPLNEKCYFSKDGQVLCKDDFFMQFGA
KCSGCEKGIAPTEVIRRATDHVYHLDCFKCLICAKLLETGEEFFLMENHQLVCKNDYEMAKSKEYDDLSNANKRPRTTITAKQLE
VLKCAYKTSSKPARHIREQLSKDTGLDMRVVQVWFQNRRAKEKRLKKDAGRRAAALNGASQYFASQPGQIQNAVTAPSAKSK
RAKKNTGLASSQASPGANVSNNGFMSTCPINLDMNACVAISENGAYPSDLDMANMMQPVQAAGLQHHNLFSMYNPHNM
HPQPLFFRDAPGIPIPNGSGPHSLTGAHLKPYMPMTSGLLSPVTDASLHGMIENRPYQLPGAPFPSLGVTQSTPGVYNLNEPPT
SPVGSTPGSHQFGSGSSSENPARIPTSIQLTPPHSSIYASMPIVSHPSFMHPPPTMNPSTLTGSHAFPHEQISSLHSALTGRSSGY
PFMTPPKFEPQTPSSSGSSGVNPPAYPPTSSSQLENFLIW*
>c4999_g1_i2|m.4553
DNKPFDGSEERRISESRKDEGAITGGRESPQLSPASVGSHDDVNGLDDSKKKNQRNRTAYSQEQLEALEKEFEKSHYPDVYARE
RLAATTEIPEAKIQIWFSNRRAKHRREEKMKSTNNLDVLQSNHFGAAKLEHLDAETAAAASFYQNAAYFPYTQHMTGANPLSL
TNFLTQSTQNNGVTNHGVVQNHTNPYGASSLASPSCFPRMPTTATSGQHSGPTGLNHMTAAALQYPNVSQQAAYYHTQPN
YFQAAAAAQVAGFPAGGSLENHGSTSSMTGQYWPTAAAAATGARF*
>c2432_g1_i1|m.1818
MIMNNLHGQQLSATTSVTSSSSKSSSSAAMTTANDFLNKRSYYGVPDNFDSITAAATGATGSGSYASDRSISPSTGGYTTSSYL
NPYINQFGMYPNNTTNMMGHQDNTAAAAAAAAYRYMPPSAAAAAALVNLPVNYFSAYAAAAMAGNGANATQTPPQPN
NDQTSTPPPPPALLSQTMSATASGPQSAQQPSMNDYDVMARAYQHHNLLQQHQQHQSSSLQQSSNNHQQHSSSSSAAAA
AAHVAMQAASVCKQQSQQNNNHSLYSSTTPSSMMTTMNTFNQQHHPHLQHHPHDSNHLQTQQHNHQQQYHPFHSHSA
AQSLYNHPHFQQHPGGPFSATTATAAATMMDHINNMNHMASMTAAAAAAAGFTGAASAALPTIPNLNQQHNFSYNLDK
DYEVDESPKGGVIGGTNVNGNLIGADVNHVINGNTQTTGKNGKKLRKPRTIYSTIQIQQLTRRFHRTQYLALPERAELAATLGLT
QTQVKIWFQNKRSKMKKIIKHGHVTGGIGEDEQNLQEFDIHGNLIENFKKKKKK
>c7609_g1_i1|m.9612
RRVHGEILKLVDTSAQEAVPKLVELITKKCSATDLEARNMLLFSQQTVIHQLHKNIAEILALKKEEAPATIDINSLTSNITIPVRDLVS
LETNLQNVALNSLQPLNLNPLIVLQPTSNSDDGYNAIQPHNMQQVLVVAPSTTTPVMKGPIKLSNPQVEVACPVGGDIDPEGI
NLTRGAGLDSADINIMTKRYAPASNAVDESSLLEELESFAKSFRARRIKLGFTQGDVGLSLGKMYGSDFSQTTISRFEALNLSFKN
MVNLKPFLDRWIDDAERNRSLLENENEVARMTLEEHKRASVFNTSLDYTTGHRKRKRRCALDAHTKKALEIYFTDNQHPRPEE
LKSLSDNLNQDRETLRIWFSNRRQKERKHENALSSASAAFSNNATEHDQPDTPISQNGSEYESSHFVS*

12- *Meara stichopi*

>Msti.rna.tri.9640.1|m.11387
MDDPSRMLHQAVSSACNLLPPQYINPDMQGQLQPPDPYSSNRKQHIGDPGDILHQIMSITDQSLDEAQAKKAALNNHRMK
PALFDILCEIKQKHGTSVLSMRQSSQEEDSSDPQLVRLDNMLIAEGVAGPEKGPGTGGPGQPDNAIEHSDYREKLNQIRMIYH

QELDKYETACNEFTTHVMNLLKEQSRTRPISAKEIDRMVSIIQKKFNSIQIQLKQSTCEAVMILRSRFLDARRKRRNFNKNATEVL
NEYFYSHLSNPYPSEEVKEELAKKCGLTMSQVSNWFGNKRIRYKKNIGKSQEEANMYITKTVMKPGSPAHSSPTSQTSLSDSPN
HSFSITPTSLSGPNHPADVYPVPSANGYAEQESIRQMINPNSVYPDAYELQHHQYPGEGPPNPDMSRFSTLHR*
>Msti.rna.tri.31629.1|m.25643
MYHFEDVDKSLAYDHPADQTDCQQLTVKLTDPVTNTTASFTLEQIDCICESLRQLNNINKLSRFLWTLPPGNIVQTSESVLKCRA
VVSFQRNNFRELYTILENHTFESQNHSLLQTIWYKAHYGEAQKVRGRPLGAVDKYRIRRKFPLPKTIWDGEETVYCFKEKSRRAL
KDCYKRNRYPTPDEKRNLSRMTGLSLTQVSNWFKNRRQRDRNPKESDEEA
>Msti.rna.tri.33513.1|m.26058
SPTRSLEMCERISDDDAPTSDELESFSKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFEALQLSFKNMCKLKPLLTKWLE
EAESNNGSPTSVDKITAQGRKRKKRTSIEVTIKNSLEQAFLKQPKPAAQEIILLANSLQLEKEVVRVWFCNRRQKEKRMTPQEEG
KDGSKEERDSDTSSPVSSRSC*
>Msti.rna.tri.24223.1|m.23298
MTGSSSSSVHPATNHAHAAPTKLVYSFPVGGGSAFPEPVISGPRTSHAPYIGNGTAATSVHVEAPAPSVPLPHRSTTLAQPVLS
VVPPRFVNSSPIQISLQAMSSTATITSHIINGQQVLSIQGTGGNQQILTLPTTGLQPGVPVTVQVLTIPAPSTNASQVPAPGSAN
KPLGTRETTLQSQSRGSNCQPVCLSDTFPANNTITEKSAGVSSLINPGMVRASAPITGEPLSSDNKLTILRRPTGVDNPLDSDDFS
GNPNTILNTDATTVDGINLEDIKQFARQFKMRRLALGLTQTQLGHAMSVTQGPAYSQSAICRFEKLDITPKSAQKIKPVLEQW
MKDAEEKYSNGNTNLSLPGSSSDKKRKRRTSFTPSALDILNLFFEKTTHPTGTQMTELSRQLNYEREVVRVWFCNKRQALKNTI
KRIRMEPEDGLLEHRLPRMGFPHNAELIGQSDIECINNISQPS*
>Msti.rna.tri.24957.1|m.23593
MLAAGIMDFSLYASMREPQLIDQMAPTDTTFHPGNLSHIAQCAGCEKPIIDRFVLRVLDKTWHGKCLRCYDCNSVLNEKCFTR
NGNVYCKSDFFGRYGARCAGCEKGIPPTEIIRRAQDKIYHLNCFSCVICRRQLATGDEFYLMDDNKLMCKSDYELAKQRECSEN
PDLASGQKRPRTTITAKQLETLKLAYNNSPKPARHVREQLSHETGLDMRVVQVWFQNRRAKEKRLKKDAGRGRWGQGSFSA
SNGSSKSGVARRMRSSRDQSSETSDSPSPSSCLSDALQNNNNDLNHQQKLTPEPISPLYPQQPPPSEVHYPPPPPPMMLHRGP
PIFVSAYPDYPSMMEPPPPRMWDEHPTHCAY*
>Msti.rna.tri.13947.1|m.15971
MSEKLSFGIDRLLEKPHENDHNVSSTRFRRENMTAVSSVDAVATDLYSGQDLFRQWTVQYNAWYASTHRQVNHSYESCIKSIA
KGIPIGNGGVAKHSCWYCQNIPDNSTPGRMRTVFSDTQRENLEAVFANKKYLTKRERQDLSKVQNLTNEQIKIWFQNRRMK
WRQTRGKTNMTS*
>Msti.rna.tri.15677.1|m.17623
MTESSSVAMETHSVTKLPLSDHVSVDSGEDKVEHFRGDIVTQPDGSGYVVETQKTASSRIPTIESYKIVCSNDDKENKEENCTTT
VPKLNFVKMGNSTQHRVLMCFTCKLTYKDNHQFMVHAKSAHNIELTEREAQLLGDETVEAVIQLRNGNDQVMFLEKDQVNS
NSYMSMMANNFVPSPFMNKYSYMPSGEVNDIQNEITEKNSSLNALLSNETERNDGDGVAFSSSSSQGSTFSLKSDMFPSMIG
LELHDPAKCSSSSVNCPKCDMILGSSRSLGGHMTMMHSRNSCKTLKCPKCNWHYKYQQTLDAHMKEKHPESNTNCSYCKRN
QPHPRLARGETYSCGYKPFRCDICNYSTTTKGNLSIHMQSDKHLNNVQEMQKSGEAAMMAAHQQSMSNDQNNKNAAHKS
MWRCDVCDYETNVGRNLRIHLGSEKHTHNILAQQQQAIPGMPISESEIIHSSFPMPATGQPVDKGELSAVFSEFQPGVQVVD
GGYVINPDMVSPPEINESEVDEPKDMFSCSICNNFSTSQIDTLSNHIHKDRSNLLDNDWFVSMNENIQCKLCKYSTTLKANFQL
HLKTDKHIQKMSLVNHIKEGGKHNEWRLKYLNLANPEQVKCTACDFYSNNISKLIAHVDSSRHVLGANIFHTVEEKRRAIGPELK
VYICKPCNFRSRFKLALLQHMRTMKHLKTVPLLNSFTEHGDLDTMLDVAIDTEGIKEKRARESEEDGFTNAKKLCQSEVEYKLRS
CPYCQYSTVESERMNAHLMTQHSIAQQSISLRCPLCQQKFPSRPMLETHLMEIHSVTPDCVLKLVQNVQVEDANESSLTMVSK
TNESSDAVDSGPDCIDRATLDVKSANELDPDQYRCQKCHLIFNDIDTLYTHQNEVCLLDLQQTPRGPGYLCWKKGCNQYFKSS
AALQMHFKEIHAKRQQLSTVELGEYRFKCFECSLAFKSKEKLQVHAKYHMIRAATTCTLCSKTFRTLGALQKHCDTIHSDLGVAN
TDHQSEHSATDANDEKADAADDSKLPCDMPLDLSQSGSSNVDDQLNTSAICAKHYDDPSYPLKCHPCKKAFTKSSYLSNHLGS
TYHKVAETSPPDPTLEKFNDPNRPFKCQVCHDSFTQKNILLMHYNSVSHLHKLKKVQEEKEHSASNVQSTISDDQKPYKCNVCK
VSYSQASTLDIHMRSVLHHTKAAKEVKPSSSLKLEDAMPQLYPYNPVDLSSIYGQLFGAVPGSSTIGLMSGFPFSFPTFPFNPMD
AMKFANPTTMAMPDPASVFNPPMFPSYTVVPQPEVSIPSLFSCQKCTSVFSNQESLTQHQQIHCYYGDMSAQLAAYAASATI
SSNITTSGGFTSESTSSSLSSVTTPVAVQSAPISAISSVLNVKPVRIGSMKAIQRQLLQNIGYELAMQFTENSQNVPLTDLAETLSK
SAKIDTSFFKCKECNKVFSGIWVIKAHLEEVHNIFVPSDDVEDAADSFSRNFDKKYSLPDESVISDSSSEMFAKESDASVIQQQSV
LLARHLAQSSFIPMLMPSIPPNMLGMSVPGPISSTSTMPPPPSVLPQSYVSSTSSISTIPSRSSGDVPTKRPRTRITDGQLKILRAH
FDINNSPSEDQVNKMAAETGLPMKVIKHWFRNTLFKERQRNKDSPYNFNNPPSTILLESIPEREEDQTNQVVEKPVDIVKTEVV
DHLEGQDNPVKGADATPLVMPYTPVPSDSVSVPFMFPNLGSDYNQTTTEVDFSSFYPRYTPATGLPLFSGTTSPSLSPSLMSSS
QHQDMPSRYSDLQRRACRTRFTDPQIKCLQDFFDNNAYPKDDDLDVLSRDLSLNPRVIVVWFQNQRQKARKSFENIPATDDD
GSSRYIKTPGQNYQCKKCMIVCQRMIDLIKHQKESCYSDELSQPSQDMSCGSKSPESVESTEVYSSQHSSTGMQYPCSKCPSVF
NMLESWQEHQQEHLMSPSFTSYMPSLVDNNYPGFMYMPPYTIASPFAGKLATTTIPTTSFCLPNFTLTTSDSTPVTMTCSVSS
HQNNPVPILLPKTESMVTSEDSNGTIPQKDRRLRTTITPEQLEILFEKYNENSNPSRRMLDLISKDVALKKRVVQVWFQNTRAR
QRKGLQRSVGPILCHRRCPYCRAMFKAQSALNAHIKTKHWDQIVESKQKGLPFPDGILESDLGNNVDNDKHKSENAQTGQFY
ATKSLKALLNNQQQKDASNTPDAESDNCEGVDDYDDDQLVIAEDDLDDSSQESMDDEEYDDSQVTFDRSGKVTVSYTVPVD
DEYNVESSCAEKVNNGTSSEILDLSRGSDGLTTPQEMYSNIYKNFEHVLSGTTMPAVDVTETRSSTPLQMPSASPRLSSSYSSDD

IGSETISYDDNLVNEMDPLSPNKGKDSKRYRTQLSHVQLKILKQCFKDYKTPTMQECEILGRTINLARRVIQVWFQNARAKDKK
NKLTYAKHFGMVTEVDLPRDVQCKLCRVKYSIHVTGRDHMFTQKHVEKMKEVYSKDPKPDSLETVSQSSLANRPLLHIRPAVS
TASSNSSLPQSRGLPKSYAFQKMDDGVKQETSTPESMAGITPPYPMAFPMMISPFMMANPYAHIPTSNSIAYPSSFNGKSDSL
CKFLCNKCRTLFDTRELVDVHQKTVCHVDSSNGETDTVTKVSAQQLVDWMSVSSVGDLNVAPPQRECGSTLTPANDS*
>Msti.rna.tri.19313.1|m.20556
MANKSFFIESILGLNQVSMEIPERTGPQGDPSLPLDVYDYSPDLGLKAKTRYRTTFSSWQLHELERAFENAPYPDSIMRNDIANR
MSLSESRVQVWFQNRRAKWRKQKPTKQRSSTELSLLYHDLLVKTARQTPECTVKQNVAMKDDIHRKKNPLTLSHKIRRDTLQC
WPKSAYTLGTNSGLFLHKYMAEELSDDEVLEENTMSPC*
>Msti.rna.tri.28093.1|m.24660
MSVDNVQDESDTSYGVPSTYNPTYVPNLNFVSVPTFERPDGEAGKELRRGKSPEVKVEPRCPVAQDGTHRQVLMCGFCSEAI
TDKCCLLVSGIKCHEQCVRCSECRVSLNYSCFIRENIIYCGRHFRQRFMKKCQQCEEYIQPDEMVMRTKGFVFHHYCFYCTLCFK
QLETGDEYALVKNQVYCRGDLAKVVQHQERNTSGQLSDKEDLEPSPRKKRAALSKEQKESLRHLFNSAPKPSKDEKDLLSLQM
DLPVKIIQVWFQNERAKVRKNCLRKAAKQNSASKIISCSHGSEGQAHHSNQFMNVVRGPLSASEKLSKMAGIICTLNV*
>Msti.rna.tri.18552.1|m.19998
MQEYGSEIPRNDPQTFSTSVGRATPHGINDILSMSRDTKQMDPIGIFLARHSYALQNYHLKEDLPMELSNQNHLDARVMLPSL
LHSFQHQGKQLIPEMMGTVMTGFDASVELGRKKHSRPTFSGHQIYILEKTFEQTKYLAGPERARLAYSLGMSEGQVKVWFQN
RRTKWRKRHAQEMSVAKLKHDSWDETSDEIDHQLFEEPNFQLSIQAVDLTPSEVHASPVNLPDPGSSLPVSQTILPTPYFFNKH
DRSRKLRDYLINTSSR*
>Msti.rna.tri.10027.1|m.11835
MIPETERCHSIDDEEENLTNDVIMKLPKKRRNRTTFTSLQLDELERVFQKTHYPDVYCREQLALRCDLSEARVQVWFQNRRAK
WRKKERYHQSSLSTSEPPPLLPRDAYHHQWGYPLPPYYPTYSFPPLTPLDYPDFRSNSIAALRLKAREHAVALMTPLMSPPR*
>Msti.rna.tri.31571.1|m.25629
MLDHSCHGNHRVMNTRKRRTFLIEDILTEDKLNDEIRVQDQRPRTELSGEVFGDHQRCTEAPRVPLLEQPSSIHHPLAVPLQVS
SPQQQCIPFPEHFQLGGPPICLPFDLGFPPTPFSAQMGSSLIGHHPMILPIPGGLPFYSSFMNSADLTSKHCRRRKARTVFSDSQ
LGGLEKRFDGQRYLSTPERIELATALNLSETQVKTWFQNRRMKHKKLVKKSDNDYNAPMVNCDEATS*
>Msti.rna.tri.18009.1|m.19608
MTKHDPEAVTDGTKVPKLEPFDDSGRFQRCVFQLKQDKICSEDKSVCAGCHRKISDRYYLYALNKQWHLHCLKCSECKEPLDTE
LTCFAKDGNIYCKADYFRRYGAKRCSQCHLGIGSRELIMRARDHIYHLHCFMCSTCQKTLSTGEEFGMRDSSLYCRPHYEMALR
DEFAQQNGLFSLNGTPPFFSGNGVQKGRPRKRKHAIPRDDCPQTFSMSGLDQNQNPDSIDNPYHSLSPGTQNGRNKRVRTSF
KHNQLRIMKSYFNLNHNPDAKDLKQLAQKTGLSKRVLQVWFQNARAKHRRLMLKQSSSTPDQDGNAKLMSGGESDNPGLD
SLSPMSSCHSSMDQSEDSPNNLTDLVSHGEHPDTYP*
>Msti.rna.tri.25446.1|m.23807
MTDSNSSKFCYPLFTELKQASYPSVEPIKEDSKLVSDPGSDNEDKSGKSKRHRTRFTPAQLNELERCFTKTHYPDIFMREELAIRIG
LTESRVQVWFQNRRAKWKKRKKTSNVFRTPGALLSSHGLGSLSGFSDSIYPFHNNENRWSTGLSGSYGKQDHFPSFPSHHMA
QSFPTSNSIGFHSGQHPSGFSSFGMANNNNGNAPEILQGNIYHSGGSSHFDFRNTLSPTCNVSMNPPQPGNMATMRSAGFT
PGINQGQILGGETEEAMWRGTSIASLRKRAIEHAATFSGCQLAGGPAPVDFR*
>Msti.rna.tri.28807.1|m.24905
MEYTAGIHPVYQFGEAFDNLYGMEDAQPDYGGELYMGTAEEFAVKLDRGLITPEGSNQQVYANHQLCSGIEVYPSADSFTSSE
ELCRLNDILNSSGDSHFESQNSSCEMIISDVADVENFSVSPEDSVPASPTPDIPETLFGDVDYPGIYNGITYPVRTFENCLPHEPVQ
EHILGQDRDGTQMANHMEEDRMPFDMVHPLFPIPAPLVNNYLNWHAEMMMGPNFIPAPPPPELPGKKAVRKRNHRKSSE
LCTSLTTKKQKTTYTPFQLAILEQNYVHNPYPDLHERSTIGEMTCIDRDRIQVWYQNRRARTRRKNPSCPESILVSPRLRKSPEYS
KFLINHGPLTRGILPSKFGHPTQQPSNMPQVSPPFFVPLRVNLPQDTPHSMPIMPFEHATEGLLKVSAHQPSIWMSDTSGPST
PLSPASTTSGFVSEEAPHISNIADTYRQLCSADLPSEEFSRIATRCQETQTSTAGVPYQSLNELVRMANFSCGLFENSSNGTKTEE
RLLSQELAPSSSEMSLAVALKTENVKQDLMIPSCLYSDSAMELNNNVIDMTVNSSISGSLHQAVSEAGIHCPGLLCDASAKLQAI
ARQVQTDATSI*
>Msti.rna.tri.37013.1|m.26767
MKAVSSFLITDLLKETLPASNGCTTTATSCWCSYSCPRSFYFAHGPEISPPPLGPVLKPFPLRPTTPLPVMLRKNGRISRKRRSRTT
FTVEQLEQLETYFVKGHYPDIHLKESLSSRLSISEDRIQVWFQNRRARWRREENTWGESTPMAEYGLYGAMVRHSLPLPENILK
KPDINHNSPTSWLLGMNSKVKVDLERTAVSMMHPQLNLELQTQGYGAST*
>Msti.rna.tri.2383.2|m.2106
MEFFTYPPYMPNDYTNYLAEAYKNLLPSTIHSPPLHESGIERKKRRNKRPRVDVGLSRNTYGQNLTLSKARACIMNQKSSPQPIQ
DGPQDLSVKRDSDDRLAQTSVEEISPKQELGLNNNDMQEGQIQTNSNLSNYGDEDYIIDDSEDIDDDPESVKRMRLSSIVSGM
KKKMEVYPVRRSNNNQGELCSSEQCQVEKEEFKLQIVKIKDQLNTLEDTFTKLYGDTSALSGEDRFKPTADVLVQLLNRAISPVV
ERIVKRYLPDVNPEVGQLPGSPHAINNGMNEYSMLPRHMIPPIPPSLFPYGNLQTSVPLPQPLVLIDQQYSSPSTSPGYKDVSYS
MSDNADDVRMKVTEESPMNSSWNNSLTPNHLKKAKLMFFYTRYPSSATLKLHFSDVSFTRHNNAQLIKWFSNFREYFYIQIEK
CARQAISEGVKSAEEIRITEDSELYRTLNLHYNKSNQFQVPSQFLDTVRLTLKEFFNAVHLNRDTDPSWKKQIYKVINKLDIQIPES
FRSLTFIEDLERNVFC*

>Msti.rna.tri.8696.1|m.10233
MVDQTTAEGIAMRKVVDQYWFWKVDFWLPPNQSSAFEKMRLLECPRNNFPKVSDFATKVLPAVVLLVVVRLLLTQFILKPW
ARYLRIKDLTKKSPLGQIDVEYNERILEDSYLKIMSAPTESEIEELAKKLGITDIEVKQWFRLRRKLKNSLPSTEVFYLTFWKFLCVTP
LTVYGFVTLFGRQWFGWKFIKSVYYTYPIKTMDEFMYNYIVAEMAYWITNLIFLVRFRFYRGFYKALLINIFVLSLIIPMFVLNFNQ
VLALMLLCHDLAMPLMYLGKLFHYASLHGQGGILTALHYGLFIWSRLWIYTRYLVPLATYAYVELFKVSGTLFSWTPWVSERW
MPLPMQPHAWQFGVIAVFVLLAMHVYHTLKVTFLLVKLAFTVRQELSQETVKKGS*
>Msti.rna.tri.24424.1|m.23385
FVELLPDKMIADQVKAETRLSASSPVCLGNASSFTKYQNEGNASPVKPAVHTIDAILGIQHKPQDRPDDGTNYRSGGPDREDLR
SESDDDEEEDSHRKKHRRNRTTFTTYQLHELERAFEKSHYPDVYAREEIAIKISLPEVRVQVWFQNRRAKWRRQEKMDAQTTY
KLQEFSPFRGPGFSTPVILDSFLPHLGSPQTTPLQALPGFLPMPNYSFVPLGGMMPTGYSHHMFPGQGLQAFPPAMHLLPTDK
LTSTNSVTNSSPKSEESDSKNSSIVTLRMKAKEHAFGLPEPAKSWSIS*
>Msti.rna.tri.13119.1|m.15087
MTVSPKHTTPFSVTDILNPLEDNYNRRMDMNLAGSAGLGYRTQQPGAATMMNMAAANSAYHAAVHGMGSALTHPCNGT
SMNDLGHYNPYGERAYGTGWYGTSADPRLSTEYFSRLMHPSSTGMSGMGFSEYNKMLPGQQRRKRRVLFTQAQVYELERR
FKQQKYLSAPEREQLAAMIHLTPTQVKIWFQNHRYKMKRQTKEKQTQDGQSNQQSQQHSPRRVAVPLLVKDGKPCSVDNN
NTSPNLPLSSGGPEHHRSQNIPSLLPSASSLCNDRSSCSLNSMAGYDTNSLTTNPYLLHSRAW*
>Msti.rna.tri.38330.1|m.26975
ERYTTKSMNPDLGSSVKSKLCGKRESIVNIFDSLPISLLSGLLKENVNVQAKEINKKDIIDEYDLVFELLRLRYDSIFQEFPSKKSNKQ
LKFLKKVKEITSFPSRCTKYDISYITQLEYKSISIWFQNNRQRGRLTGGE
>Msti.rna.tri.27494.1|m.24471
MDMYYSESCKYGFTQEQVACVCEVLQQSNKTERLASFLWSLPACERLNSNESVLTAKATVAFRRCNFGELYRILEMNNFSLHN
HNKLQELWLAAHYMEAEKLRGRPLGAVGKYRIRRKHPLPRTIWDGEEMSYCFKEKSRNALRDWYVHNPYPSPREKRELAEST
GLTTTQVSNWFKNRRQRDRAAETGKDLYHKT*
>Msti.rna.tri.23494.1|m.22969
MSIFETQDRSMAPLEVKPILPHPLLKFPFSNRPAFFDAKLFGLPLDNRHLEHNALSPLNSLAFSMDFPKPLMKGMERILQRDAFP
YYMPTDMASFDTSLGLRRKNATRETTSALKAWLYEHKKNPYPTKAEKIMLAIITKMTLTQVSTWFANARRRLKKENKMTWTP
KNRYGDNITFNSDDEGESGCLTLTEDKFINNTNAVIQTNTKKTSKIWSIADMVTSPEKQSPESQYKLWLNGAYENSLMDASSSS
TSSSFTHFLQVFVPPSHLKNNSNPSAVSGHSEPAIEQ*
>Msti.rna.tri.9892.1|m.11686
MAYLKAPYGVNGIGLSSSVDLLHPAMGYTSQYSALPTGFTNTGASPGSGGTPRKQRRERTTFTRAQLEVLETLFAKTRYPDIFM
REEVAMKINLPESRVQVWFKNRRAKCRQQQQTTNNSGTPLAVTAKKPSSTPKKHRSSTSPNMPSVDEDPSPKQEAMVNDFA
SKMSLPPASSASSLWNPATIPGDYFTQNSCMQRSMYSNQGQCYPQSYSSNPPYFDTASSYIAPHHQYMGYQTRYAPDPSQC
AVADYPKVEQTSHWKFQAL*
>Msti.rna.tri.16907.1|m.18732
PKHMFGNNFPSRSEPVMNSTLDHTSREQVSPANSDSRSEESIERGPPVRNGNFSYNFHVDNQGPAQMEPCQPQLHSGTWL
NKEDIDDDDLELRDLLRLPVQDVIAELKNFLIVNKITQNIVAEQANVSQGYISQFFCRGVALRASTQLVLYKWFIAERRNVINMG
YSRFAADTARNSPRCAPRRRERFVWRDECMPVLDKYFQENQYPDDRKREEIAQSCNSILEIPGAVLPPDRLVTPSKVYNWFVN
KRNNYKLKMRQMAWEMRDQASMAMNNGQQSRLTVAEMINRNRTGPEVAPGPGQSHHQLHHDPGFHCNAPLDYATDK
QTTGPLSAGGSLPFPAPDKDQTTPLDFAVEKVAVNHAIMAISAHSPPPDDDEEIMPEDDEEEILDSPILKQEVVEVL*
>Msti.rna.tri.13930.1|m.15953
MGNPSFSIDSLLKQGPTIPTQQLIVPHSRGLPPHVLTALPPLHQPVMPSFPMYYPPHPFYSGYAGTYSDYLMHSAFLRNPKRKR
RHRTIFTEDQLEELEKTFYKTHYPDVVLREQLALKVDLKEERVEVWFKNRRAKWRKLKREEEEKKKQDKSTCSSSNTNDVTTERT
GSSSGQSVSAGSEDDEEMVDIEK*
>Msti.rna.tri.12933.1|m.14928
MTSLVESQLPPSPDLDAKGINFLEYCNSQTTSFTPPPLLHNGGQCQYQSPPALYNTGQPHYQDMYPRGPQYGSPHFVQRPNV
YQSFPYNPFTSPTMSPNMNITNINHSYHYMPIQAPPSATQARDEMSYSTNCTTSTQAAMLRYPFGPPPGETLIDKLSDPLQSDL
PRLPGKGKKMRKPRTIYTSLQLQILNKRFHDTQYLALPERAELAASLGLTQTQVKIWFQNKRSKYKKILKTSSGGQQSGMTTPPS
PAVPAQWQDSKSSMANHPHLTDYRNYSSWYSHQTEHIQSPEPLLT*
>Msti.rna.tri.21139.1|m.21727
MTMGDIVPELQRFDRTHMKNRHFMCVGCGNQILDSYVLRVAPDMEWHASCLKCSECQCYLNESSTCFVRNDKTLCKRDYYIL
FGKKCAKCNLRFSKTDLVMRTSVSKIYHLDCFRCTACSRQLITGDKFVVREDGLFCEEDHQVLERAQSDEVQNNNNNNNPAST
KNKPNSKPSSNHRKSSSGEHKPTRVRTVLNEKQLQTLRSCYSANPRPDALMREQLTEMTNLSGRVIRVWFQNKRCKDKKRSM
LINNQMENGQKQSINGVPMVAGSPVRHDTSMQGNSVDVQSYQPWGAPPLETYMNGISGDVHPDYIIYDNMHPSSLSDISS
DICDSLSSPPSNGYMLNNMNECSPPLT*
>Msti.rna.tri.10800.1|m.12637
MQCQIRGIRVHECFRLSSAHFVYMLPYINQSMSSSPKCNFVITRPSFPKSTSVNRAATPERKYGQYLASPNRSRKNSSGSSETLST
SPSNTGQRTYFDVFANKSSKTFGNVDHLPEVNPYPKYRLLPDGEDDYRTPPGSTPDIFYSFPDSYNISINGSVDSNLNSLVPPPTN

CSVRSTSPIEINRSFMPKYPKLFGHGSPRVSPKGSPSCVETSPNSSKSLSVDSRQRKILEAAYNNCRYPSFDTVGSLSTEAGLSRAD
VHSWFDHKITEWRRKEGFPSGSSTIFER*
>Msti.rna.tri.40660.1|m.27355
RQYRPLDIPGHPWTLLCAALCHSVSFGELSTLPIGVQSADMETSYRAPLPRDILINSGNCYPYLLPPCLNPEIMATSRPLISSQETS
FQMHNWGKQRRNRTAFTTQQLLHLEEIFRRTQYPDVMTREQLSICTSLPEARVQVWFK
>Msti.rna.tri.20454.1|m.21273
MLLFPMFHVGSGGFPGGGALPGGAGLPVPPPPLPTLNFTPEQVASVCQTLEDSGDIERLARFLWSLPLNPGALENLSQHEAILR
ARAIVAFHQGNFRELYHLLENFRYTKDSHGKLQAMWLEAHYQEAEKLRGRPLGPVDKYRVRKKFPLPPTIWDGEQKTHCFKER
TRNLLREWYLQDPYPNPQKKRELAHATGLTPTQVGNWFKNRRQRDRAAAAKNRTQSGCRSETPDLSIEPSDSPESMRELL*
>Msti.rna.tri.34952.1|m.26351
PPHPPPHPPLALPHFATHLPSPDAMDMGLLLKSYQLYPPSWLNCRTMSKISPLLPLSLNASNGDSVLPTSTSVPQKDVSPLPSD
MHSPGEESIDAEEESVDEEDGKDDQDKKDRRKKKTRTVFSRHQVYRLETMFDCKKYLSSTERSALASTLHLTETQVKIWFQNRR
K
>Msti.rna.tri.15130.1|m.17123
MGKGHSGVNQLGGVYVNGRPLPDGTRQKIVELAHSGARPCDISRILQVSNGCVSKILSRYYETGSIRPKAIGGSKPRVATQDVV
KRVSQFKRECPSIFAWEIRDRLLSEATCNPENVPSVSSINRVLRSLSQDKGQLHSESMYDKLRLFNTQTWHGSWYPSALPPAM
GPTSHYHGQHPTMMPGAILDSKIKDEYKLGLDTGTRSSDNESQDSEEARMRLKRKLQRNRTSFTPEQIESLEKEFERTHYPDVF
AREKLAEKITLAEARIQVWFSNRRAKWRREEKLRSQIRQGDGPASGNAVPSLNGPGASAPPTFNPGFYSQSAAAMSSPYGDYA
SLTHLPPYSMGPSSQQPNLNYGYNCRSYADPLSGYRPMSSAPQALGHQNSATTVNSSTGLISPGISVPVQVPGASHPLIDIAHH
NAQYWASRI*
>Msti.rna.tri.18958.1|m.20294
MVDIQNATAAAQNMSLEDGETEEMASGDNEHNGPLYTVKRRANLPKSAIDVLKQWLYNHRYNAYPTSEEKLKLALETNLTPL
QVCNWFINARRRYLPELIRKDGHDPSAYTLSRQQKKNLTEGAKDNRSCSEPPLLTFAQDCVTRRASDTFVDYHDPITHQYGFPA
NDQFPLYPVMSYKDKQCNVYSTTFNQPSTDCPCLDCLAYGTADYPAYPAQHCNTAAPPKWYCDCAECVSFPTKTDKDIPGYT
KSGFRPPLPPHMEAGY*
>Msti.rna.tri.21384.1|m.21884
MLNFSIENLLGRSSRELHPNCGHEQKQTPSQHLPEQATENTSSPSTPQSTFYVPPTTAATRYLPWLTPQKDWPLPNYACCYPM
YGVPGQSEFLRLSGMKIRRPRTAFTSQQLLELEKQFRESKYLSRPKRFEVASKLMLTETQVKIWFQNRRMKWKRLRRGDKQFN
QSHAEKSAWEKKIITT*
>Msti.rna.tri.23168.1|m.22788
LANSMQPMQSLANIVQPMQFSLYPGQPIGLQSVLGDSANIARTIQTPMATMMPQIAMQAQSQTFTLLQNMGVPGILAAPN
LTTPKTDMDAQTLMKLPNYPGIQTCQQIPVIAQAGSSPQFIALSQPGVLAAVGAQELSQLQQSLSQQNVNLQQYVFLQPAPM
QNMQTKNSLSPSLVPATQSQLLANQQTATSLAVATSLAVATSLAVTSPIPSYVTAAPVVTAADSNTPAEAVNQLLKYGLALNPS
LLTVVPLSAYLQTISSQSSTTPAGDCATPTTVPVTSKIVNGGQALSAQPSVHQSSLQTEASQLEDLEQFAKTFKQKRIKLGFTQGD
VGLAMGKLYGSDFSQTTISRFEALNLSFKNMCKLKPLLEKWLEDADAAVTVLPGAVMSTATSTMMGITSTELITSRRRKKRTSID
NNIKISLERDFLQNPKPTSEDIHSLAESLNMEKEVIRVWFCNRRQKEKRINPSSMSYINSQPRLYTAAPSETPPSNQGELPIESSL*
>Msti.rna.tri.25124.1|m.23672
MTDAAYSASDPAHASLSQSPVYTPSLSNADYQGSSLYHCQQQYMPKIDYQQPGYCGQWSCPNEQYIQDIKPFTAATREDSCS
GGTIAYKPPYYFSGHYPWNGTSGGGAYVDTTANPLGTYGASNYGSYPAYHMSSPYSVNGHGLHNTMPVVYPHSTDPNDPDH
LQDEEAPSSDDLEVFAKQFKQRRIKLGFTQADVGLALGTLYGNVFSQTTICRFESLQLSFKNMCKLKPLLAKWLEEAEANNTSPT
SIDKMAAQGRKRKKRTSIEVSIKGALEGHFYKQHKPAAQEIGLLAESLQLDKEVVRVWFCNRRQKEKRMTPQFSEEENGEPSEE
GNVTSANRSYCENPKKSFPTLENNNLKKISAQYSYDKPDSKEQESVMKDFRPPNFDASKLNYSLYDNKENFQNSYSQSDSSIYT
QLTRSNLHSSGSLLSQYQLGCRSHFQKTVITTNPSVLDNGLISYAYSGHAPLSPIDDKNGHHLDFVDQARAKPAIMDS*
>Msti.rna.tri.15192.1|m.17189
MLDQYSRMAEKQSLKESRRFYIENLLSKKQEKRSIVHHRPKETTDKKAPIRPAVSHQYQPFKAVLKPKCSKQAAGNQPAHTGRS
QGIDIPGHPSRGARRSTNESVQPSLCQQRLDTALELRYRHNQVYDVGVLKSYKPGVGIRQNYDSNGRVPELYGVGGAILGRFE
VGTVPREALDTLPPEVTDRLITCNHRAGPPRNYHNYNARPCPQFGFHPSFPLADYHQHRMDELTLAAQNNRLQMARPHLQP
VQQNPEQYNHGWLPHCSPPVDNRPMPVVVAQTCYHGQHVSGCESQNTKLCMEPIPYKRDLPLWSSNYMNSWRTSSKRKG
GQVRFTTDQTIGLEKVFMDRKYLSPPERKKVAKNLKLSERQIKTWFQNRRAKWRRVCVESGTPRHDLLPSPQSSTSNGESDVEI
CE*
>Msti.rna.tri.11227.1|m.13086
MTITGMEVPTASDPGQFLDYKREVCYGCQQYIDDKYILKVNEKSWHEQCMVCAECRTPLSDSCYVKDRSVYCKDDYERLYGAR
CHVCQEAIPPRELVMKATGYVYHLKCFACVVCGERLRQGDEFILKDNRLLCKTDFEKEEAEEKSREGQHMQLPIATSTLGVASSI
LPLNSSPVSSNLVATPNMPVLPLDQTLMMQAQQQPMNGGHPLMQSIKPGIAAMPISNKKNQIDKDGSRKRVVLTPEQRKIF
QQAFDMSPKPSRKVREQLASECTVSMRFIQVWFQNQRAKMKKLAAKGESVPGGIVGIDDSTDKMMDDKLNDSGEHDSLME
GGGLNDDDDDRSLDMEDSVDKSDKGFPLVSSPVGGSTTLFPPPNPLGQLPPLGSPVETKPHQESGQD*
>Msti.rna.tri.11875.1|m.13825

MIGGINIDVPPAKTMSFASAYRYEEPSGIHFGATSMDPSNFVASDLARSYPNYNHWTKESPVNSSNYVHQSGYPSSSPGSVLPE
VMKRDKEVVLSHPLYPLLNIIFNKCELATCTPRDSGASEDVCSSVSFNEDVQAILSTKGHSGKQSFLTNDPETDGLMLQAIQVLR
CHLIELEKVHELCDNFCQRYIHCLKGKMPMDLVAEDKEMLGIKQRLGSGSSLEDSTYLESSPNIYTDLSTPTQHTYSPPDVADNL
STCSTDTLGGASIQPLTNLTGVQAQPQCSIYDPLSQPHVQLNQQVTHISSNSEHDRCLSNSASPFENNNPSTPSDNSVNNNNN
LMTAMSPTDADEDKKPQVTGQQKKRGIFPKAATNTMRAWLFQHLSHPYPSEEQKKQLAGDTGLTILQVNNWFINARRRIVQ
PMIDQSNRAGMPPGAYSPEGQHMWMDANMHMRHPSVHPAMASHDMMMSSQGYHHQLHYPVHSQAMLIPGHPHAM
MMAHHPGMHHQSIPSIQGHTQTPGLMHHHHSDHPVMNIHAT*
>Msti.rna.tri.1979.2|m.1761
MKQQFCAGCGKSITDQYLMNVVDRTWHSHCVRCSACECLLSEKCYWREGKLFCKRDFIRCFGTKCMGCKQGILSTDIVRKAR
NLVFHIHCCNCICCEKQLGTGEEMYLVGEDKVYCKVDYFNWKDGFDDEEDEDDVLDKDLDMSLEQENMSPKEHSNELSNQDI
YNNNNDIKEEEKSGCEDDDMDDSKSIKDDESKGNGQKRRGPRTTIKAKQLETLKAAFNATPKPTRHIREQLAQETGLNMRVIQ
VWFQNRRSKERRMKQMRDLGVHRAAYLRHSRRMRALGAPHLDQADYYGQPPYFDFFQPQHMNMPSHMEPSPDFMPPG
GGIPPLMVNNDGYLTSQDMSPDPDSPDSYAPSRQRLMPPHFSNGIMQQSHELCDPVW*
>Msti.rna.tri.13662.1|m.15709
MSSFSIEFLLKSHVSVEKKIVPGVITPPTSPEFDISRNDFVQEECFQLSDEDFLKEELELDESGRKLNKKKTRTAFSPYQVCELEKKFR
DQKYLAASERGDFAVKLGLSDTQVKTWFQNRRMKYKRQQQDIENQSMATMYYTQLLNCSRPMVPPSVATIPALSGHHSTGV
PLPAHLIQPMINPPMCPPGACGCLHCPPSVSMPIQSQGLAPVMTFPSMPGLPPVLPYHANQLMPYQHHHIAVHPC*
>Msti.rna.tri.29890.1|m.25206
NAWRPSQPFGLNNQMYSFDTANMTTQLPAYQGPQYTYDAQEFTPGMRRKNATRETTSTLKAWLYEHRKNPYPTKGEKIML
AIITKMTLTQVSTWFANARRRLKKENKMTWTPRNRSGQNQGKDGIGSDDENNDQDNNNDKEKDSLDSSSLKVDLEEKDIDK
NPRIWSLAQTATSAESCSLSYTTLLNRSAYSEQPPYMDPYRNSYFPPATYPSSGQTIYDQYYPPVTQDTPPSTLCHYRSPHMAEQ
TALDSYTTYPQPPTSLPLGPYIASYGSSVQESLSEGSE*
>Msti.rna.tri.1286.2|m.1185
MLPVYAPPLSVTNCSSSSPTTSMKQILDTDQMDNDKQAVLRHPLLPLLTLLFDKCEKATNTTECPTSSSFDVDIETFVRERERDG
KPFLMQNEELDNLMVRAIQVMRIHLLELEKVNELCKDFCNRYIACLKSKMNSENLLRNDAFHNYGQGIQSSPSPQDLGVHTNP
ILYQPVTMVTPQGAFHLPHQANHLIHGSTPLSQIGTGQGPIRTPANHHHLLLDPSTAEDMKKIGHLSAVNAKRGILPKQATAVL
KSWLFQHLVHPYPTEDEKRQLAGQTGLTLLQVNNWFINARRRILQPMLEATNPEIAPKTKKAKSCSRPAMQRFWPETLVNISG
DPAKDDDSPGFPKDIKRPNSASSVSSHLSVTNNNDQHSIHSSPYHEISTLH*
>Msti.rna.tri.34005.1|m.26153
MLYQMNSSMNLGGDLSSPGVNASSNLQRSPYYPAATGYHKSEPSYTALLDTGLTLTQPMAAVSSESDYPSPNLGSNWQYPYF
SSHQQHSPPTDRKPWPYDISNSLNQIAVNSTQYPVFDSQPTPNYMPPSTNTPSWMSDPSSRYSSSLPLYLGGSTHFLGSYAGG
VSQRYPMMDTRDGSTTYRVNPLTGKTRTKDKYRVVYSDRQRAELENEFRINQYITIKRKTEIATGVGLSERQVKIWFQNRRAKE
RKVARKYTK*
>Msti.rna.tri.16216.1|m.18152
MYSGSVMNTQQHFGMHPAGFATTKYGPPMHRSPDPIRRHFTSTPPVRISDNSYSYMHSVQGNFFSSFDDNFLARAEALAAV
DACSKMQPIKEDPGKYSGFLNGPSHSGSRCVQEPAFGIPNHMGMDASQLGQHVMAHHPPSGFHHPATHGLPHHQIASAM
GHTDMDTDPRELERFAEYFKQRRIKLGVTQADVGRALANLNIPGVGSLSQSTICRFESLTLSHNNMVALKPILHTWLEDAEKA
MRENNNNSLLAEKKRKRTSIATPERRSLEAYFSVQPRPSSEKIAQIADKLELKKNVVRVWFCNQRQKQKRLKFSAIR*
>Msti.rna.tri.15735.1|m.17683
VKTKEHDSYNVELHQPTNAHSAAKIKPGVKTNEHTLDKTELHSRSREHATDEDELRLREIDLRHSDKSKLQMKPSDLQMPENEL
LSVIKERVFEKNRSRSRTNDCGFEKSENRSSYNDRILDKSDSRSGYNEFLLDKTDVHQKGSEQSLEKPTLRSKTNDSIFDKNVSKLF
KDDSFSFDSEPILMNDPKKRGKKAVSVYAPVTSPPLLLLAMTSSAKPSVASGSMPSLAMSTAITSGTFSFNLTSQSVIQSNVITKP
NQMQSLIADAPRMTASANQQNYLGAGFGGVSSHLSGHLLNPPMMSHIVGPRTANITATITSLPGPSVITSVMLTPPLSSKPVA
ETHLSGASPGLSPSISPAAFSAPTTDEMSMWPAKKKLLKRSTDDIALKAALDHAELGSVETQLGLQALDFSKESASSPSVKKSPQ
RGVKGKAGRGRTKASLKQMPDDLLLPGSTFNCVAEGGLDMRYKKLEGHTGIDVSSPYLQPQSTPVIEFTKPHDTPTSPSLSLTT
VSCYNPVTPIMSSSLASYMDLPIMAPFPMPLSMSSMTTQPSVITSLSDTSMYHTSPGHVITSKSHPQTEEPLLAQQNTMPPAD
LQKRTQNAPMAFPQLAVITNEERGVHAGPEQTQKFFQAVLTRGKEVGIPHAITAIKRLGKDDCAKRGKPTPKGRGRSTCKSLKF
GTSMAKPLDGPEAAKIKRKLNPRARNVDVYEFKDDDFDSFVMNPISTSHEVPTAPFIASTNTGLMNVSETATLPGDGKLPSSTA
ISSSTDLATTTSSIPESVCTTLKPSPHHLVKTKAALKQTVPEFDKILTEYCTPDELVAAANDRGVTAMEPKVGKSPVTRGRGRVAA
TVKVLRKGLRSGLRSSLRLNPSLSEEDESSAADVYEPLSDSDDIPLKLSAKTLSSRAEQDSNQTSTDLGKSVATLMPTDVSSVHPSE
GVLEDHGFATKKGNSVGQSEKSVHQTDKAAHQGDKSVIQNVKVVNQGDKSVNQICKIVNESDKSVNRNKRAVDKDDKVDN
QNDKAVSQNGEAMNTIVKALKLSIKTVNQSDKTTVDADNGDVGLTYTPIRVVTNSTPTTNTDDKSMCDDSKLHADPSKKSPIE
EIENEPILSRIGQGGSVAAVSPPTNHPKKRKVPVRKAMMKLRKSSLDESGDGSEKIEGSAEPAKKDDTCELDQPTLSGDPGGAV
PDFVQPRNPPTKRKLKGKRGRGRKKIRPNSPLVAPSRANSLADVSDVDTMRKSGLNSVSGSVEGKLIDSELPASTFDTAIDDSSK
NDAEKIDESITGLDNNDSSDVDEVTASKVDADNVVEEPKPDDELRKGKLVLRVNANVLEKLHPNLATNSLKSSLPKTETDTLTVK
MDAIAQTQPAMLLKSLYPLVPPTPEYETAYETETDMEMQREKFSCIFTSGINLRIRRVGTPINSDEEDDFLDAPAMVYSEPSVPN
PKVFSIGGLNNAVKSTSKNGDEDLSKGDSKSFCRTSPVASNPMMLLHAKNILAGSSAPATNTPVTIGVAVEEKSALKSGVPVKM

FKCCKCEYKSVECGDLVKHFDLIHVKNDPILIGKLPELNASLSETVVVKDTVMSTEVIPVVDAVVSNSKENVIAKEKELIAAKEAFD
VKHPTIPKELRCPICKQTLPSESLAKSHFQKSHNIKVDSDECKAVLTGNELEDAYKLRQIEAETDVESLSKALNISKMNVRLCFYQK
RKHDKSIVLSSESQSESISNNSKTKESPNSFKKPELTPSTSKPKSKKGKKGRKTIEDPSKKFTPYQLSVLRDFSDQSLWWKRGRITEI
CEITKLTSRDVEDWFVLQRRKVNSTKDVAEEPPSHTDTVSDGDAKQTETTEDGIDHIERARLRELDREHKTWVGLYLESSYQKN
KTLSNQETDVISKVSGLSPHTIEYWFQKRQNRDTELISQFSKSPTGKAKQLKGKPMSEDSGPTLLGDHTYTTPTLKDTAMATKVK
VLLSKDAILNATAAMLSKVETKSTSRAESPLFSAEVSDESNDATSSSVSGDSSNVAPFRVQSDETFLSRGMKFAKTAINRKAAHS
PVPAEEPGTMPIEEKVASPRSGHKEPRKSGLNRRRLSSVESTIVDWSVVTQDFVDNHGNSILTCAYCCLALANRESYRMHLMEY
HQNRLIFLCSMCSNKEATCLYTAKQHLLDKHKVCESGFKRWILTDLVDSDYTCPECKLKCRSPSLLRDHILKEHRSNKLFQCSLCS
PWQEALYYVDVYKHIMDLHRLPEATVVSSMITPAESLLDLTHSAMANNELPMVAQKIRPSTELVDATFATLLEDDALVQKSDFS
PSVSPISIPDLSLFDDLPLCFPDDMPQHCAKQDNSSLILEVNEEQMERNTHSAGSDIGVAPDKLRISKVEAAILAGISLAARYDVG
QHESPLSREYAVKGFVNELIGDVGDKMSRKEELMDCRKQCVHFLSWVYSNVETRLQTPAQRFVNQILSNVDDILMKRYTNIK
WHREFKMSCETHEFVGETDCQMQTHHPIAITSLNKKSCSDHNRSDGEIDLFTMASKYDLNVFECGTCYDVFPKATLLESHLEYC
RPQAGGQSSMSVYVPDSFVRSYITSKLGDLSSTGDFTCSICAAAFHGNLVVMLTHLSTHTRPVFTLHPARNPRVSSNTIRHLDYN
PWLKTSNQEIDLHTVEKEEAWRRQNQEEEARLTRLANRKGKRKKKPTRAYVPSKKKKVNLDISVDNEVCRVRQVSSADKSEQA
VDPLLLNMNQSMHRLMTVAPMSAQPKKKVDNMLWSPSKMVTRTSVQQRQQRHEDSPRKRMLLSTTTHSKHKDSLREVVE
DPTKTSRFPEKYDSMSIHLAHSDYIILLSLIREKSDKSFCKSKLDLLDKLQVGKSREACPGDGCTIAQCPCCKVCTLPYSAVRLLYLEL
CTTKKTKMFVYVYFSHYVCYYCNVKLTDDDKLLSHLETHMDITCLDSDGAGREKPDPSDCRSSP*
>Msti.rna.tri.20131.1|m.21074
SGHYDHYTDDIYSYNHYTIKPRKRTSKQQREVLEKVFATTMKPDTALRKRLAENLTMTPREVQIWFQNRRAKEKKLAKKSPNED
SPIPKNQKHTQQFNNNWNTNEQNDIEISPIALNNMSSKFIQSDSAYNFNHHPPTFREWQNTHMNENHNIGTDVNNQHCG
WEGEDEDDYFRN*
>Msti.rna.tri.6770.1|m.7876
MFQEVFQETAGIMNGDYEVTKPRQRKARIHFTDQQQRILEEKFQDTMYPDYPMRKFLSDKLAIEENRIQVWFQNRRAREKKK
PKEYRRLSWDPEIKDRQIDSTPCCESDGLQKHNKLFYHRPQQTLKDLLTPNGAKRSRMTELVNCPKKACHGRIWNSEPNINPN
SPQAPGLTPQSVMRKWANRTLNKERKNVKSAPEFWFNDQNIPFEEGFQISKQGYRATSILKQELSQNSPSEFPPVILRGDRVPL
TPVSTPSSVQSNISATVNEEIKKELDSVLTASFKEKVTLNNDSPTCYSSKDEGYSSCRSSKDEGGIYSPALETPYSSLHNTSSAYCKN
VYPDHGHFGRSRDRIKHETRPSCLYEKRSIAKSPGVPFDFPQIRLQDLDMIIPD*
>Msti.rna.tri.20808.1|m.21491
MSCSFSIDSLLKTERTKDAVPEAEPAIHSESCLGGNCLLCQRLDKNSLLIDPRRGFISRHNTMTLHPSLVCPLRKHKANRKPRTPFT
TQQLLALERKFKEKQYLSIAERAEFSDNLKLTETQVKIWFQNRRAKQKRIQEAEMERVKLQARILQGHLIGRGYGYQQSPNIFSL
GDTFSQWGRKF*
>Msti.rna.tri.21711.1|m.22061
MPQSKGHSGVNQLGGMYVNGRPLPDDTRRKIVELAHSGARPCDISRVLQVSNGCVSKILGRYYETGNIRPKIIGGSKPRVATPE
IVEQIAEYKKENPSIFAWEIKDKLCVDGICDDDTVPSVSSINRVLRNFSLPKLADFYGRLQSKIQEKYSENIWPSKPACQKRQLSRA
DKPVAMTKIDEEVARLQLKKKLQRNRTSFTNKQIESLEEEFERTHYPDVYAREQLSFTTSLAESKIQVWFSNRRAKWRREEKCQ
DRVSNPESDSQKPISAGNFNSATQFDYPELYYHGYPCSQPLTPPASSVDLTAELFQAPEPSPVPPYAPFSQTPIASAVPYPTQSYP
LSGHYAHLPQYPYVQPETQTYSTLLPAQSSSGMLSTYIMPNSSCLGNSPNQYAAAMYRRYGMDYHPNGVK*
>Msti.rna.tri.17312.1|m.19074
MTLAMNWNSSYLEGLPDPLNRFQDKPEKDPRQLFAAYNNGQHFLPYTMYGTTDESCMYYPVEQAFKQKSYLYNGEGPQFQ
NNDHQLAGSVPYNTSLKHSSNDNRLSTSVHVHSQADKTLKYYPPTTISMTTAPDEPTYPAGHLDYTDVNTTEDRKPPGGNSW
LNRNASRKKRRPYTKYQTLELEKEFLFNTYITRERRADIANALNLSDRQVKIWFQNRRMKMKKMNQQQPSHDPYQNCVTPM
*
>Msti.rna.tri.18865.1|m.20221
MMAFPAPILPVSSPVIGQVIPGRLFDLNVASPYHTLCPPLPLPHELGYHPYILPAMGDSREQLLYTDSRYGVDGWMPSGIRRKN
ATRETTNALKAWLSDHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKENKGEWIDQEDDDDEVEKYDVENTMKN
DAKRVTSKPGDDVETRTGCKSRVEGISEKKTKIWSLAETATSDKPPKQNDEHFKSEAIVVSGHNTKLLSADGQFPVPGQPSTTP
GDYDHYSSSLFLQHNRYPPLATASSLPYITSLTGNLLNVNRSTLHPDSRMYPTNSLLAPSQSMMVP*
>Msti.rna.tri.19684.1|m.20788
MILTMDKPGPALQAAKETCYGCDSPIDDRFLLRVAGNSWHEACLQCCMCDIQLFKSCFMREGKLYCKVDYDKVYGVKCLTCA
ECIPSQELVMRAGGYVFHLECFQCSCCGVRLQKGDEFVINVGSLYCRTDYEKLLLDDDSDADNGCNSKENMILQDGEFTFDGD
KSEDGKPEGKGLKGGGKQSNRKRVILTTHQRRTFRAAFDISQKPCRKVKERLAVETGLPVRIVQVWFQNERAKVKKMARRQA
QEQQDGGRRRARGRGETYCDEDLYRERVPDHLIGPMSDIPPEHFAFPPPIRGPPVIWDPEFYPNYPEDPAFQGSPGVINPMDK
LNSMQNSYFTASD*
>Msti.rna.tri.17700.1|m.19381
MVVSSASSLMDRNKLFSVDSIVGSNDQYRPGVLELPQDRCPNLTRGCSSPQCCYSHMSLQHMVPFQYCLTTGLSLSPMMLPH
SHSPIHRLATDFPVNFERTPVVPSQFSSDECSNIVPSSLIPTSFIPTRPLTPLCYHTEPHDGNPVYMQRMGKSKRLRTAFTSEQLT
HLEQEFEKNPYVIGDERRQLAASLKLSETKVKVWFQNKRTKHKRQKMIDGTIEQSNSAGTHSS*

>Msti.rna.tri.1108.1|m.1029
MLFYSRMDSDNDSTGNMSPDSMSDSSSPQIALSSQSEACSSCGDLIMDRWLLRVGDRNWHSECLSCTICGQGLGAHTSCYYR
DNQVFCKLDYVRLFGAKCAKCCRQISSTDWVRRAKSNVYHLACFACDTCQRQLSTGEEFALQDDRVLCKNHYIEMLEGEIGEE
YKTELSKLWVSEAKVSETSVEDNPSTDQNHNTKGTSSHRPKRVRTTFTEDQLQILQANFNLDSNPDGQDLERIAQITGLSKRVT
QVWFQNSRARQKKYSQTGLKTDQNSKQRCPDCGLYYCGNLCIGKRNDLTQLMSPS*
>Msti.rna.tri.14107.1|m.16124
MELVTRVSDVLWLEQFQLPPLGAWWTETRDDPGSAIYLPRAIDLAWALVLAVVISALRRGFKLQVAEPLARWRGIQPPKTVYV
KPDQVLEAVYTNNANPPHRHLMELSKKTDMSTAEIANWFRKRTKKDKISKLDKLSLYLFKSVFHVVMIVWGVRVVLKENYFPD
TRNCWTNMPRTHIPPSVFWLYMVQLGYYLQELPGHFYESKRKDFLMMLAHHVITALLITFSWLCNYVRVGVLVMAVHDPMD
PFAYLAKTCHYLGMRRAARALLVGLIVVFFITRMFIYPYYVINSTFFEYAEMANASIVMAPSTWGRPNSLSLFTTAWGFFNILLTC
LLIFNFIWGIAFLKMFVVMVVSPEKLQDVRSATESSA*
>Msti.rna.tri.22006.1|m.22189
MEFGLRQNVIQSPRQLDWRGMLPHPYLPPIPMFYHPPQMPALALRDDRPFPHLPKKHENEAKHSHFRIENLDDKSSKRKRTA
FTSEQLIELEKEFNSKKYLSVGERCHLASILKLSESQIKIWFQNRRAKWKRLKAGCFGLDANQHIVVPIPIHVNRLSIARELGYPFHS
GPPALRPDIISPRAALPGTTDRPFR*
>Msti.rna.tri.14816.1|m.16832
MEGWLNSGIRKKHAARDTTKSLKAWLSSHKKNPYPTKGEKIMLSFMSKMTLTQVSTWFANARRRLKKERIEWKVVSEESAEIG
DSEDKSRIEYDATIQNSAKKICRNVDQKKKIWSLAELATSGKLTLAEKTRLEDSDRFKVENDCTLKRNCTSPPAAFSSSLGVPSLAA
SSSESLTSLPDFAPS*
>Msti.rna.tri.11416.1|m.13280
MDFFSRKFASEQFWFPKNVTWDSFKNKDDSTSYPESSSLLLAVPLAALMYILRLLVERFIFGPICEIYKVKNSRGTKPEKNTILEKV
FYSVSKFPDDNRIRGLSKQLDWSESKVQSWFRRRRNQDKPSPAQKFNESGWRFLYYLSATVFGSYIAITEGHLFNTRDCWIGYP
FQSVTSQMWWYYVLQLSCYYALIVSQFSDIKRKDFVIMFTHHIITILLIYISFMINFLRIGNLVFVVHDASDILLELAKLLRYVGAKS
VCTVVMVMFAATFFVTRLYIFPFWIIYSVHFESWEIIGPHTCWYLTNSLLWCLVVLHVVWSYMIVKVIVNALKGQLEDSRSDTE
REVSDVGFTGNGSTTKRKD*
>Msti.rna.tri.24090.1|m.23206
LSSVNPSYYQPSNVFRHLGSGMDHAAISLDSYHGEDHQNCKLMDHFNSNIAFRPQSFGMGHLQRGFVASPPKPILSSEKPLED
MGAESKSKDHLEVKKEISPGELSNDDRKRPRRQRTHFTSQQLQDLESMFLRNRYPDMATREEISAWTNLSESRVRVWFKNRR
AKWRKKERNQFCTDYKGAFSSPQFGSFVYPHEDSFYAGPYATYNSWYSKVSSSFNKPTFPWTPSQQICLSTPTPSNTAYPPAPA
STQYYRDCAPHALRLNKAKLPTATSSYGGIPSPGLTGSGSLFSQCQYTDRVNCKIDEFKHF*
>Msti.rna.tri.20507.1|m.21302
MYTMDTFSDLYSGHHLHFNADLHGQRESMKYDLSTSTEMSKPSSTRATFLDHSDYRMDPYYPHPAAGYGMQPGTSGYGTLP
AFQSLPPISTVTEKLLIPNYDMMSHHQQTPNHHYSTTAIQRSTTFAATNPHHRGPTSAFDSTQVAMARLGQPPPSHMMGLN
GFHHGRSDRSSHLQINSNSYKDAEEVNTRDVAQQVMAELKRYSIPQAVFAEKVLCRSQGTLSDLLRNPKPWSKLKSGRETFRR
MWKWIQEPEFQRMSALRLAVQSSSHSACKKKEEEQMTPKNQGPRKPRLVFTDLQRRTLHAIFKETQRPSKDMQIQISQQLGL
NLSTVSNFFMNARRRSIDKWKEENLQKLT*
>Msti.rna.tri.29223.1|m.25016
ALGRGLLHGPRNDDRITEDSQYRFSWLQCTRYKPPKIPRAKKQSDRKRKLGRNPRIPFSQEQVAMLEARFRCAQYLSSMDVVD
ISIKLNLTENRVKIWFQNRRARERREKASKKLAYNGLQSAIGQYPRSDVSSS*

13 – *Sterreria sp*

>c23422_g1_i1|m.5547
MACEFVTMDGEKTDNNFVTRGYNINSCFPHPSPHSLYKQPPYYFSGYIPPLARLHPYYSSKVVGATSEQSLNTSSNHWRPHET
QESARPSECSPSPKSQTCSSPADKYDYKLSYYGDGKVPAYQNYPPYYPFSQNPLNQCLPVDHSRKDGTESLSSPDQSGESILDDS
YLDDSDRTRRKRTAFTSEQLMQLEKEFNNKKYLSVVERSHLAKNLQLSESQIKVWFQNRTAKWKRLKARSLGSSIMSGGQKIVV
PIPVHVNRVGATNSIQQAPSGPTTLHLSASHWSL*

Supplementary Figure S4

Generic homeobox Sequence:

```
             RKRGRQTYTRYQTLELEKEFHF----NRYLTRRRRIEIAHALC-----------LTERQIKIWFQNRRMKWKKEN
```

ANTP class HOXL subclass

Cdx family

```
CDX1       Hs KDKY.VV..DH.R........Y----S..I.I..KS.L.AN.G-----------.....V........A.ER.V. NP_001795
CDX2       Hs KDKY.VV..DH.R........Y----S..I.I..KA.L.AT.G-----------.S...V........A.ER.I. NP_001256
CDX4       Hs KEKY.VV..DH.R........C----...I.IQ.KS.L.VN.G-----------.S...V........A.ER.MI NP_005184
CDX        Bf KDKY.VVYSDH.R.......YS----NK.I.I.RKVQL.NE.G-----------.S...V........A.QR.MA
cad        Dm KDKY.VV..DF.R......YCT----S..I.I..KS.L.QT.S-----------.S...V........A.ER.Q. NP_476954
pal-1      Ce  DKY.MV.SDY.R.......HT----SPFI.SD.KSQLSTM.S-----------.....I........A.DRRDK
Cdx        Cg  DKY.VV.SDH.R........Y----S..I.I..KA.L.QN.A-----------.S...V........A.ER.
Cdx        Ip KDKY.VV..DR.RA...N..RS----AQ.I.I..KS.L.MQVG-----------.S...V........A.ER.VS DN44102_c0_g2_i1|m.23116
Cdx        Sr KDKY.VV..DR.RA...N..RS----AQ.I.I..KS.L.MQVG-----------.S...V........A.ER.VS DN125595_c2_g1_i1|m.86363
Cdx        Hm KDKY.VV..DR.RA...N..RS----.Q.I.I..KT.L.SLVG-----------.S...V........A.ER.VA DN14515_c0_g2_i1|m.3895
Cdx        Nw KDKY.VV.SDR.RA...S..RC----.Q.I.IK.KT.L.GMVG-----------.S...V........A.ER.VA 49193.0
Cdx        Xb KDKY.VV..DR.RA...N..RC----.Q.I.I..KT...G.IG-----------.S...V........A.ER.VA DN63819
NVHD065    Nv .S.A.TA..AS.Q.......LY----S..I..T..K.L.NT.D-----------.S.KH...........K..TD DQ301955
```

Evx family

```
EVX1       Hs MR.Y.TAF..E.IAR.....YR----EN.VS.P..C.L.A..N-----------.P.TT..V........D.RQK NP_001980
EVX2       Hs VR.Y.TAF..E.IAR.....YR----EN.VS.P..C.L.A..N-----------.P.TT..V........D.RQR AAA52414
EVXa       Bf VR.Y.TAF..E.LAR.....YR----EN.VS.P..C.L.AQ.N-----------.P.TT..V........D.RQR
EVXb       Bf VR.Y.TAFS.E.TAR.....HR----DN..S.P..C.L.A..N-----------.P.TT..V........E.RQR
eve        Dm VR.Y.TAF..D.LGR.....YK----EN.VS.P..C.L.AQ.N-----------.P.ST..V........D.RQR NP_523670
Vab-7      Ce MR.Y.TAFS.E.IGR..R..AK----EN.VS.KT.G.L.AE.N-----------.P.GT..V........D.RQR
Evx1       Cg  R.Y.TAF.KE.LNR.....LK----EN.VS.PK.C.L.AQ.N-----------.S.ST..V........D.R
Evx2       Cg  R.Y.TAF..E.IGR.....YK----EN.VS.P..C.L.KL.N-----------.P.NT..V........D.R
Evx        Ip MR.Y.TAFSKH.QDS..M..RK----EN.VS.P..C.L.AS.N-----------.P.TT..V........E.RQR DN12279_c0_g2_i1|m.4929
Evx        Sr IR.Y.TAFSKD.QDI..A..RK----EN.VS.P..C.M.A..N-----------.P.TT..V........E.RQK DN124972_c0_g3_i4|m.83074
Evx        Hm LR.Y.TAF.KE.QDK..A..KK----EN.IS.PK.C.I.SE.N-----------.P.TT..V........E.RQR DN58284_c0_g1_i1|m.41535
Evx a      Nw QR.Y.TAFS.Q.LQR..R.FVK----EN.VS.PK.C.L.SS.G-----------.P.TT..V........D.RQR 48679.0
Evx b      Nw QR.Y.TAFS.Q.LQR..R.FMK----EN.VS.PK.C.L.AT.G-----------.A.TT..V........D.RQR 36518.0
Evx        Xb                  K.YTR----EN.VS.P..C.L.AQ.N-----------LP.TT..V........D.RQR DN14752
EVX (109)  Nv TR.Y.TAF..E.LKR.....MR----EN.VS.T..C.L.N..N-----------.S.TT...........S.RRR DQ206338
```

Gbx family

```
GBX1       Hs SR.R.TAF.SE.L........C----KK..SLTE.SQ.....K-----------.S.V.V........A...RI
GBX2       Hs NR.R.TAF.SE.L........C----KK..SLTE.SQ.....K-----------.S.VQV........A...RVK
```

```
GBX        Bf TR.R.TAF.SE.L........S----KK..SLTE.SQ.....K-----------.S.V.V........A...RVK
Unpg       Dm SR.R.TAF.SE.L....R...A----KK..SLTE.SQ..TS.K-----------.S.V.V........A...RVK
Gbx        Cg  R.R.TAF.SE.L........S----KK..SLTE.SH...N.K-----------.S.V.V........A...R
Gbx        Nw YR.K.TSFSSS.LR....H.NR----KK..SVPD.TLL.NI.H-----------..QP...V......A...RIK 20062
Gbx        Xb SR.R.TAY.SE.L.Q..R...G----KK..SLTE.SH..KN.K-----------...G.V........A...RVK DN90720
GBX (018)  Nv .R.K.TAF.SK.L.Q..R...N----KK.VSLEE.SV..TN.N-----------...V.V........A...RVR DQ206305
```

Gsx family

```
GSH1       Hs S..M.TAF.ST.L....R..AS----.M..S.L......TY.N-----------.S.K.V........V.H...G NP_663632
GSH2       Hs G..M.TAF.ST.L....R..SS----.M..S.L......TY.N-----------.S.K.V........V.H...G NP_573574
GSX        Bf SR.M.TAFSST.L....R..AS----.M..S.L......TF.N-----------.S.K.V........V.H...A
ind        Dm S..I.TAF.ST.L....R..SH----.A..S.L......NR.R-----------.S.K.V........V.Q.KGG NP_996087
Gsx        Cg  K.I.TAF.ST.L....R..NS----.M..S.L......TY.N-----------.S.K.V........V.F..
Gsx        Ip SR.T.TAFSSS.LMC..S..ES----SN..T.I......NK.H-----------...K..........V.N..GA DN47321_c0_g1_i1
Gsx        Sr ER.S.TAFSSV.LIQ..R..DL----.N....M......NR.N-----------...K.V........V.C..QL DN87690_c0_g1_i1
Gsx        Hm AR.S.TAF.SN.LVM.....NN----SK....L...D..TS.K-----------...K..........V.D..TK DN24302_c0_g1_i1|m.7955
Gsx        Xb AK.T.TAFSST.L.N.....SR----.S....I....L.QS.D-----------.S.K.V........V.Q..LL DN115451
GSX (130)  Nv S..I.TA..SM.L.......SQ----....S.L...Q..AL.D-----------.S.K.V........V....DK AAD39349
```

antHox family (Hox1)

```
HOXA1      Hs PNAV.TNF.TK.LT........----.K....A..V...AS.Q-----------.N.T.V..........Q..RE NP_005513
HOXB1      Hs PSGL.TNF.TR.LT........----.K..S.A..V...AT.E-----------.N.T.V..........Q..RE NP_002135
HOXD1      Hs SSAI.TNFSTK.LT........----.K....A......NC.H-----------.NDT.V..........Q..RE NP_078777
HOX1       Bf PNN..TNF.TK.LT.......Y----.K....A..V...AA.N-----------.N.T.V..........Q..RE
lab        Dm NNS..TNF.NK.LT........----......A......NT.Q-----------.N.T.V..........Q..RV NP_476613
ceh-13     Ce NGTN.TNF.TH.LT.......T----AK.VN.T..T...SN.K-----------.Q.A.V..........E..RE
Hox1       Cg  NM..TNF.NK.LT........----.K....A......AA.G-----------.N.T.V..........Q..
antHox     Ip TRG..TNF.NK.LT........----......A......SS.N-----------.N.T.V..........Q..LV DN75362_c0_g1_i3|m.221830
antHox     Hm TRG..TNF.NK.LT........----......A......AS.N-----------.N.T.V..........Q..LL DN54918_c0_g1_i1|m.41056
antHox     Nw TRG..TNF.TK.LT......LY----.K....A..V...AS.N-----------.N.N.V..........Q..LV 35873.0
antHox     Xb SRG..TNF.NK.QT........----......A......AS.S-----------.N.T.V..........Q..LL (Hox 1) DN2325
anthox6_12 Nv SQKK.F.F.QR.LV........----SK....T......TT.K-----------...M.............R.F DQ206301
NVHD060    Nv SDKN.TI.STR.LV.......Y----....C.P......QS.E-----------...K.V............... DQ206321
```

centHox family (Hox4/5)

```
HOXA5      Hs G..A.TA..............----................-----------.S.................D. NP_061975
HOXB5      Hs G..A.TA..............----................-----------.S.................D. NP_002138
HOXC5      Hs G..S.TS..............----............NN..-----------.N................DS NP_061826
           Bf N..T.TA..............----................-----------...................
Scr        Dm T..Q.TS..............----................-----------...................H NP_996164
lin-39     Ce E..Q.TA...N.V........T----HK....K....V..S.M-----------.....V..........H....
```

```
HoxC5      Cg  ..S.TS...H.........Y----.K.............N----------...................  XP_011424787.1
centHox    Ip C..T.TA...F..........----............NL.A-----------.S................D. DN31614/17434
centHox    Sr C..T.TA...F..........----............NL.A-----------...................D. DN118344_c1_g1_i1|m.54580
centHox    Hm S..T.TA...H..........----............TL.S-----------.S................D. DN27605_c0_g1_i1|m.9399
centHox3   Xb S..T.TA...Q.V.......Y----.K..T.......N..L-----------................Y.... DN124767
centHox2   Xb R...................Y----.Q..........TG.G-----------................F.... (HoxM2) DN84275
centHox1   Xb CR.K.SS.D.L.C........----....SK...SQ.SKH.G-----------.....V.V........H..DR (HoxM1) DN76496
```

postHox family (Hox9-13(15))

```
HOXA9      Hs TRKK.CP..KH........L.----.M....D..Y.V.RL.N-----------.....V..........M..I. NP_002133
HOXB9      Hs SRKK.CP..K.........L.----.M....D..H.V.RL.N-----------.S...V..........M..M. NP_076922
HOXC9      Hs TRKK.CP..K.........L.----.M....D..Y.V.RV.N-----------.....V..........M..M. NP_008828
HOXD9      Hs TRKK.CP..K.........L.----.M....D..Y.V.RI.N-----------.....V..........M..MS NP_055028
HOXA10     Hs GRKK.CP..KH........L.----.M....E..L..SRSVH-----------..D..V..........L..M. NP_061824
HOXC10     Hs GRKK.CP..KH........L.----.M....E..L..SKTIN-----------..D..V..........L..M. NP_059105
HOXD10     Hs GRKK.CP..KH........L.----.M....E..L..SKSVN-----------..D..V..........L..MS NP_002139
HOXA11     Hs TRKK.CP..K..IR...R..F.----SV.INKEK.LQLSRM.N-----------..D..V..........E..I. NP_005514
HOXC11     Hs TRKK.CP.SKF.IR...R..F.----.V.INKEK.LQLSRM.N-----------..D..V..........E..L. NP_055027
HOXD11     Hs SRKK.CP..K..IR...R..F.----.V.INKEK.LQLSRM.N-----------..D..V..........E..LN NP_067015
HOXC12     Hs SRKK.KP.SKL.LA...G..LV----.EFI..Q..R.LSDR.N-----------.SDQ.V..........K.RLL NP_776272
HOXD12     Hs ARKK.KP..KQ.IA...N..LV----.EFIN.QK.K.LSNR.N-----------.SDQ.V..........K.RVV NP_067016
HOXA13     Hs GRKK.VP..KV.LK...R.YAT----.KFI.KDK.RR.SATTN-----------.S...VT.......V.E..VI NP_000513
HOXB13     Hs GRKK.IP.SKG.LR...R.YAA----.KFI.KDK.RK.SA.TS-----------.S....T.......V.E..VL NP_006352
HOXC13     Hs GRKK.VP..KV.LK.....YAA----SKFI.KEK.RR.SATTN-----------.S...VT.......V.E..VV NP_059106
HOXD13     Hs GRKK.VP..KL.LK...N.YAI----.KFINKDK.RR.SA.TN-----------.S...VT.......V.D..IV NP_000514
Hox9       Bf SRKK.CP...F.........LY----.M....E..Y..SQHVN-----------.....V..........M..MS
Hox10      Bf GRKK.CP..K..I.......L.----.M.VS.E..Q..SRHVN-----------.SD..V..........M.RM.
Hox11      Bf TRKK.CP..K.........L.----.MFV..E..Q...RQ.N-----------..D..V..........M.RMK
Hox12      Bf SRKK.CP.SKV.L.......LY----.M.I..EQ.G...RKVN-----------..D..V..........M.RMK
Hox13      Bf GRKK.CP..K..LSV..Q.YIQ----.R.VS.ET.L.LSQR.N-----------..D..V..........Q.RLE
Hox14      Bf VRPK.RP.SK..LN...N.YVQ----.Q.IS.DK.LQLSQK.N-----------.....V........I.Q..LD
Hox15      Bf TRKK.RP.SKP.LAL..D.YAS----QKF..KEK.K..SESSS-----------.S...VM.........E..LA
Abd-B      Dm VRKK.KP.SKF.........L.----.A.VSKQK.W.L.RN.Q-----------.....V..........N..NS NP_996220
egl-5      Ce S.K.....Q....SV..AK.QQ----SS.VSKKQ.E.LRLQTQ-----------..D..I..........A...K
nob-1      Ce  .KK..P.KKD.ISR..Y.YSV----.Q...NK..S.LSAQ.M-----------.D.K.V.V........D..LR
php-3      Ce MRKK.KP..KA.........LY----.T.VSKQK.W.L.KY.H-----------.....V..........D..QK
Post1      Cg  RKR.RPYSKF.IA...R.YNN----ST.ISKS..W.LSQLIN-----------.S............IKA..
Post2      Cg -RKK.KP......MV..N..LN----SS.I..Q..W.ISCK.Q-----------.S...V.V........R..
postHox    Ip SRKK.RP..KT.........LY----.T.I..E..L...RS.S-----------..D..V..........N..QM DN66164_c4_g4_i4|m.87486
postHox    Sr SRKK.RP..KN.........L.----.T.I..E..L...RS.N-----------..D..V..........N..QH DN123370_c1_g1_i1|m.74729
```

```
postHox     Hm TRKK.RP..KT.........LY----.T.I..E..LQ..KS.N-----------..D..V.........M..I. DN3061_c0_g2_i1|m.947
postHox2    Nw SRKK.RP..K.........L.----.T.I..E..S..SR..N-----------.SD..V.........M..MT 54694.0
postHox1    Nw  RKK.RP..K.........L.----.V.I..E..S..SRS.N-----------..D..V....           10725.0
postHox     Xb  RKK.RP..KF........L.----.T.I..E..LEISR..N-----------..D..V.........M..M. (HoxP) DN3733
```

Meox/Mox family

```
MEOX1       Hs ARKE.TAF.KE.LR...A..AH----HN....L..Y...VN.D-----------.S...V.V..........RVK
MEOX2       Hs PRKE.TAF.KE.IR...A..AH----HN....L..Y...VN.D-----------.....V.V..........RVK
            Bf PRKE.TAF.KQ.IM...N..RH----HN....L..Y...VK.D-----------.....V.V..........RTK
Btn         Dm NRKE.TAFSKT.LKQ..A..CY----SN....L..Y...V..E-----------.....V.V........C.RIK
Meox        Cg  RKE.TAF.KH.IR......AV----HN....L..Y...V..D-----------.....VKV..........R
Mox         Ip    E.TAF.KF.VHA.....TL----HN.VN.P..QL..RK.C-----------.S...V.V..........RTK DN54162_c1_g1_i1
Mox         Nw KRKD.TAFSKF.IEQ..S..ET----YN....L..F.LSKK.S-----------.S...V.V..........RVK 58015.0
Mox         Xb PRKE.TAF.KH.IR...R..TK----HN....L..Y...VS.E-----------.S...V.V..........RVK DN86948
MOXa (034)  Nv KRKE.TAF.TH.LR...N..TR----.N....L..Y...VS.D-----------.....V.V..........RVK DQ206313
MOXb (110)  Nv KRKE.TAF.KH.IQ...N..QR----.N....L..Y...VS.D-----------.....V.V..........RVK DQ206339
MOXc (019)  Nv KRKE.TAFSKH.LQ...N..IR----.N....L..Y...VS.D-----------.....V.V..........RVK DQ206306
MOXd (119)  Nv KRKE.TVFSK..LT...R..CR----.N....L..Y...VS.D-----------.G...V.V..........RGR DQ206342
```

Mnx / HlxB9 family

```
MNX/HLXB9   Hs CR.P.TAF.SQ.L....HQ.K.----.K..S.PK.F.V.TS.M-----------...T.V...........RSK NP_005506
Mnxa        Bf TR.P.TAF.SQ.L.....Y.KE----.K..S.PK.F.V.TA.M-----------...T.V...........RSK
Mnxb        Bf TR.P.TAF.SQ.L.....H.KE----.K..S.PK.F.V.TS.M-----------...T.V...........RSK
exex        Dm TR.P.TAF.SQ.L.....Q.KQ----.K..S.PK.F.V.SG.M-----------.S.T.V...........RSK NP_648164
ceh-12      Ce MR.P.TAFSSE.LVQ...Q.SD----....S.P..YQL.QQ.S-----------.S.T............N.RCP
Mnx1        Cg                      M----.K..S.PK.F.V.TS.M-----------...T.V...........R
Mnx         Ip SR.P.TAF.                     .S.PK.F.V.TK.L-----------...T.V...........RSK DN98046/DN86741
Mnx         Sr CR.P.TAF.SQ.L....NQ.RK----.K..S.PK.F.V.TK.M-----------...T.V...........RAK DN123752_c0_g1_i1|m.76614
Mnx         Hm SR.P.TAF.SQ.L....NQ.KR----.K..S.PK.F.V.TK.M-----------...T.V...........RAR DN4748_c0_g1_i1|m.1255
Mnx         Nw SR.P.TSYSSK.IKI..EH.VL----....S.PK.F.V.TS.N-----------...T.V..........L.RSE 58481.0
HLXB9-069   Nv HR.P.TAFSSH.L.A..RQ.QL----HK....PQ.Y.L.TS.M-----------...T.V...........RC. DQ206324
```

Ro family

```
ro          Bf QR.Q.T.F.TE..MK..L.YNR----TE.I..P..F.L.EM.N-----------.S.T..........A.D.RIE
ro          Dm QR.Q.T.FSTE...R..V...R----.E.IS.S..F.L.ET.R-----------...T..........A.D.RIE NP_733173
R06F6.6     Ce SR.K.T.FSPE.ATR..A.YIG----DS.MA.EK.HLL.QS.K-----------.S.N.V.T......A.D.RDR
Ro          Cg  R.Q.T.FSSD..VK..I.YSR----TE.I..S..Y.LAEL.C-----------...N..........A.E.R
Ro          Ip SR.Q.TSF.PD...Q..LQYER----SE.I..SK.F.L.ES.G-----------.S.G.....Y....A.D.RIE DN62850_c0_g1_i3|m.66453
Ro          Sr GR.Q.TSFSAE..IK..L..QM----.E.I..S..F.L.DT.G-----------.S.N.....Y....A.D.RIE DN119820_c0_g1_i1|m.59912
Ro a        Hm NR.E.T.F.SD.SVK..I.YQR----SE.I..S..FTL.EM.K-----------MS.N..........A.D.RIE DN18872_c0_g1_i1|m.5077
Ro b        Hm ER.Q.T.F.TK.ANI..S.YQR----RE..SKQQ.QAL.TS.Q-----------.S.T..........A.D.RTG DN12978_c0_g1_i1|m.3456
```

```
Ro          Nw NR.Q.T.FSSE...S..V..QR----SE.IS.SK.Y...EH.H-----------...N.V........A.D.RME 6133.0
RO (126)    Nv PR.Q.T.F.SE...K..L...Q----.E.I..S..F.L.AC.N-----------...T.V........A.D.RLE DQ206293
```

ANTP class NKL subclass

Abox family

```
ABOX        Bf DRKP..A.SSR.LER..E..KK----DK..SVSK.V.L.ES.E-----------...I.V.T......T....QT
CG34031     Dm DRKP..A.SAS.LER..N..NL----DK..SVSK.V.LSKS.S-----------...V.V.T......T....QL
ceh-19      Ce ERKP..A.SAR.LDR..T..QT----DK..SVNK.IQLSQT.N-----------...T...T......T....QL
Abox        Ip                  E..KK----DK..SVSK.L.LSV..S-----------...T...T......T....QV DN10641_c0_g1_i1|m.4062
Abox        Hm ERKP..A.STK.LEK..E..KK----DK..SVS..L.LSV..E-----------...T...T......T....QM DN29024_c0_g1_i1|m.10351
Abox        Sr DRKP..AYSAK.LEK..E..KN----DK..SVSK.L.LSV..N-----------.S.T...T......T....QL DN112643_c0_g1_i1|m.40300
```

Barhl family

```
BARHL1      Hs PRKA.TAF.DH.LAQ..RS.ER----QK..SVQD.M.L.AS.N-----------..DT.V.T.Y....T...RQT NP_064448
BARHL2      Hs PRKA.TAFSDH.LNQ..RS.ER----QK..SVQD.MDL.A..N-----------..DT.V.T.Y....T...RQT NP_064447
Barh        Bf PRKA.TAF.DH.LAQ..RS.ER----QK..SVQD.M.L.AS.N-----------..DT.V.T.Y....T...RQN
B-H1        Dm QRKA.TAF.DH.LQT...S.ER----QK..SVQE.Q.L..K.D-----------.SDC.V.T.Y....T...RQT NP_523387
B-H2        Dm QRKA.TAF.DH.LQT...S.ER----QK..SVQD.M.L.NK.E-----------.SDC.V.T.Y....T...RQT NP_523386
Ceh-30      Ce ARKA.TIF.DK.LQ...NT.EK----QK..SVQD.MDL..RMG-----------..DT.V.T.Y....T...RQA
Ceh-31      Ce ARKA.TIF.DK.LQ...NT.EK----QK..SVQD.M.L..RMG-----------..DT.V.T.Y....T...RQA
BarH1       Cg  RKA.TAF.DH.LNS...S.ER----QK..SVQD.M.L.SK.N-----------..DTQV.T.Y....T...R
BarH2       Cg  RKA.TAF.DH.LNC...S.ER----QK..SVQD.M.L.AK.N-----------..DT.V.T.Y...SHDLLK
BarH3       Cg  RKA.TAF.DQ.LNN...H.AR----QK..SVQD.M.L.AK.G-----------.SDT.V.T.Y....T...R
BarHl       Ip MRKA.TAF.DE.LVK..QN.ER----HK..SVQD.LDL.A..D-----------..DT.V.T.Y....T...RQA DN36044_c0_g2_i1|m.16521
BarHl       Sr HRKA.TAF.DE.LAQ..QN.ER----HK..SVQD.LDL.AS.Q-----------.SDT.V.T.Y....T...RQT DN152752_c0_g1_i1|m.134562
BarHl       Hm TRKA.TAF.EE.LSK..AS.ER----QK..SVQD.LDL.SS.D-----------.SDT.V.T.Y....T...RQA DN68850_c0_g1_i1|m.42447
BarHl       Xb PRKA.TAF.DD.LAR..TS.ER----QK..SVQD.MDL.AN.N-----------..DT.V.T.Y... T...RQT DN44680
```

Bari family

```
Bari        Bf QR.R.TVF.QS.LRA..RK.RC----QK..SVSE.A.V.R..G-----------.S.T...T.Y....T...R.T
CG11085     Dm PR.R.TAF.HA.LAY..RK.RC----QK..SVAD.SDV.ET.N-----------.S.T.V.T.Y....T...RQN
Bari        Ip TR.R.TAF.QA.LNF..RK.RC----QK..SVAD.SDV.EM.R-----------.S.T.V.T.Y....T...RQT DN60676_c0_g1_i2|m.56432
Bari        Sr SVS.SISVPPSSSIPSPLALQN----FLNKKN...SDV.EL..-----------.S.T.V.T.Y....T...RQT DN100159_c0_g1_i1
Bari        Hm PR.R.TAF.QA.LNF..RK.RC----QK..SVAD.N.V.TQ.A-----------.S.T.V.T.Y....T...RQT DN83548_c0_g1_i1|m.43646
```

Barx family

```
BARX1       Hs GR.S.TVF.EL.LMG...R.EK----QK..STPD..DL.ES.G-----------.SQL.V.T.Y.........IV NP_067545
BARX2       Hs PR.S.TIF.EL.LMG...K.QK----QK..STPD.LDL.QS.G-----------..QL.V.T.Y.........MV NP_003649
            Bf CR.S.TVF.EL.L.G...R.EK----QK..STPD.VEL.ES.G-----------.SQL.V.T.Y.........MV
Barx        Cg  R.S.TVF.EL.LIG...R.ES----QK..STPD..DL.ES.G-----------.SQV.V.T.Y.........
Barx a      Ip YR.N.SVFSEK.LAD...W.VA----EK..TASL.VK..SQCG-----------...M.V.T.Y.........QM DN76364_c0_g1_i1|m.230241
```

```
Barx b     Ip CR.S.TVF.EA.LAN...Q.GN----TK..STAQ.SR..KSSG-----------..QM.V.T.Y.........QL DN67947_c0_g2_i1|m.104238
Barx       Sr CR.S.TVFSDQ.LKG...N.CS----TK...TMQ.NRL.KESG-----------..QM.V.T.Y.........QL DN80548_c0_g1_i1|m.15026
Barx       Hm CR.S.TVF.DS.LIG.D.S.KH----SK..NTPE.NRL.KH.G-----------..QT.V.T.Y.........QL DN34181_c0_g1_i1|m.14867
Barx       Nw CR.T.TVF.EL.M.G...R.DM----QK..STPD.LGLSDS.G-----------..QL.V.T.Y.........EI 44999.0
Barx       Xb CR.S.TVF.ES.LVG..RR.AD----QK..STPD..TL.DQ.R-----------.SQI.V.T.Y.........QV DN81549
041        Nv FR.I.TVF..D.LQK..DR.KH----DK..STQD.N.F.RS.Q-----------..PL.L.T.Y.........DM (this study)
```

Bsh/Bsx family

```
LOC390259  Hs .RKA.TVFSDS.LSG...R.EI----Q...STPE.V.L.T..S-----------.S.T.V.T........H..QL XP_372433
bsx        Bf .RKA.TVFSDS.LNG...R.ES----QK..STPE.MDL.A..N-----------...T.V.T........H..LQ
bsh        Dm .RKA.TVFSDP.LSG...R.EG----Q...STPE.V.L.T..G-----------.S.T.V.T........H..QL NP_995728
tab-1      Ce .RKA.TVFSDQ.LQG..RR.ES----Q...STPE...L.N..N-----------.S.T.V.T........H..VV
Bsx        Cg  RKA.TVFSDQ.LNG...R.EA----Q...STPE.V.L.NQ.S-----------.S.T.V.T........H..
Bsx        Ip .RKA.TVFSDG.LHG..RR.ES----Q...STPE..DV.S.FN-----------...T.V-intron         DN57889_c1_g2_i1
Bsx        Sr .RKA.TVFSDG.LHG..RK.QQ----Q...STPE.....TQYN-----------...T.V.T........Q..IF DN106437_c0_g1_i1|m.31009
Bsx        Hm .RKA.TVFSDS.LQG..ER.KD----QK..STPE...M.TL..-----------...T.V.T........D..CC DN1275_c0_g1_i1|m.391
```

Dbx family

```
DBX1       Hs GMLR.AVFSDV.RKA...M.QK----QK.ISKPD.KKL.AK.G-----------.KDS.V...........RNSK
DBX2       Hs GILR.AVFSED.RKA...M.QK----QK.ISKTD.KKL.IN.G-----------.K.S.V...........RNSK
DBX        Bf GMLR.AVFSDF.RKS...M.QK----QK.ISKQD.RKL.EK.G-----------.KDS.V...........RNSK
CG12361    Dm GMMR.AVFSDS.RKG...R.QQ----QK.ISKPD.KKL.ER.G-----------.KDS.V...........RNSK
Dbx        Cg  MLR.AVFSDM.RKG...M.QK----QK.ISKPD.KKL.AK.G-----------.KDS.V...........RN
Dbx        Nw GMLR.AVFSDA.RRG..RK.RE----LK.ISKPD.K.L.NK.D-----------.RDS.V...........RNSK 49786
Dbx        Nw WN.V.TVF.NA.R.G..EQ.RE----QK.VSKPD.K.L.GK.G-----------.SDS.V.V.........RNSL 29958.0
```

Dlx family

```
DLX1       Hs IRKP.TI.SSL.LQA.NRR.QQ----TQ..ALPE.A.L.AS.G-----------..QT.V......K.S.F..LM NP_835221
DLX2       Hs VRKP.TI.SSF.LAA.QRR.QK----TQ..ALPE.A.L.AS.G-----------..QT.V........S.F..MW NP_004396
DLX3       Hs VRKP.TI.SS..LAA.QRR.QK----AQ..ALPE.A.L.AQ.G-----------..QT.V........S.F..LY NP_005211
DLX4       Hs LRKP.TI.SSL.LQH.NQR.QH----TQ..ALPE.AQL.AQ.G-----------..QT.V......K.S.Y..LL NP_001925
DLX5       Hs VRKP.TI.SSF.LAA.QRR.QK----TQ..ALPE.A.L.AS.G-----------..QT.V......K.S.I..IM NP_005212
DLX6       Hs IRKP.TI.SSL.LQA.NHR.QQ----TQ..ALPE.A.L.AS.G-----------..QT.V......K.S.F..LL XP_376652
           Bf MRKP.TI.TSF.LQQ.NRR.QR----TQ..ALPE.A.L.AQ.G-----------..QT.V........S.Y..LM
Dll        Dm MRKP.TI.SSL.LQQ.NRR.QR----TQ..ALPE.A.L.AS.G-----------..QT.V........S.Y..MM NP_523857
Ceh-43     Ce MRKP.TI.NSS.LQM.Q.K.QK----TQ..ALPD.AAL..E.G-----------.SQT.V........S.Q..QK
Dll        Cg   KP.TI.SSL.LQQ.NRR.QR----TQ..ALPE.A.L.AS.G-----------..QT.V........S.A..   Cgi_10016653
Dlx        Ip LRKP.TI.STV.IQQ.TRR.QR----TQ..ALPE.A.L.A..G-----------..QT.V......K.S.M..II DN61228_c2_g1_i1|m.58904
Dlx        Sr MRKP.TI.STI.IQQ.TRR.QR----TQ..ALPE.A.L.A..G-----------..QT.V......K.S.M..LM DN123563_c1_g2_i1|m.75836
Dlx a      Hm IRKP.TI.STM.LQQ.NRR.QR----TQ..ALPE.A.L.A..G-----------..QT.V......K.S.V..IL DN12216_c0_g1_i1|m.3286
Dlx b      Hm IRKP.TI.SST.LQQ.NRRYNT----TQ..ALPE.ADL.AT.G-----------..QT.V......K.S.M..IM DN12035_c0_g1_i1|m.3241
```

```
Dlx a      Nw LRKP.TIYSS..LSM.N.R.QG----TQ..ALPE.A.L.AT.G-----------..QT.V........S.F..LL 50937.0
Dlx b      Nw MRKP.TIYSSL.LQI.N.R.QQ----TQ..ALPE.A.L.AS.G-----------..QT.V..               54562.0
Dlx c      Nw IRKP.TIYSS..LSM.N.R.QG----TQ..ALPE.A.L.AS.G-----------..QT.V........S.F..LM 35817.0
Dlx        Xb MRKP.TI..SL.LQQ.N.R.AR----TQ..ALPE.A.L.AS.G-----------..QT.V........S.Y..VL DN90457
DLX 095    Nv IRKP.TI.SSF.LR..N.R.IK----TQ..ALPE.ADL.AY.G-----------..QT.V........S.F..TL DQ206283
```

Emx family

```
EMX1       Hs P..I.TAFSPS.L.R..RA.EK----.H.VVGAE.KQL.GS.S-----------.S.T.V.V......T.Y.RQK NP_004088
EMX2       Hs P..I.TAFSPS.L.R..HA.EK----.H.VVGAE.KQL..S.S-----------...T.V.V......T.F.RQK NP_004089
Emxa       Bf P..V.TAF.PT.L.R..HA.EK----.H.VVGQE.KQL.QQ.T-----------.S.T.V.V......T.Y.RDQ
Emxb       Bf P..I.TAF.PS.L.R..HA.EK----.H.VVGQE.KHL.QS.S-----------...T.V.V......T.H.REQ
Emxc       Bf P..I.TAF.PS.L.R..HA.EK----.H.VVGQE.KHL.QS.S-----------...T.V.V......T.Y.RDQ
ems        Dm P..I.TAFSPS.L.K..HA.ES----.Q.VVGAE.KAL.QN.N-----------.S.T.V.V......T.H.RMQ NP_731868
E5         Dm P..V.TAFSPT.L.K..HA.EG----.H.VVGAE.KQL.QG.S-----------...T.V.V......T.H.RMQ NP_524825
ceh-2      Ce N..I.TAFSAS.LIQ...A.EG----.H.VVGNE.KQL.AK.S-----------...T.V.V......T.H.RVR
Emx1       Cg  ..I.TAFSPT.L.Q..HA.EK----.H.VVGQE.K.L.LK.G-----------.S.T.V.V......T.Y.R
Emx2       Cg  ..I.TAFSPS.L.K...V.ER----.H.VVGQE.KDL.TN.N-----------.S.T.V.V......T.Y.R
Emx3       Cg  ..I.TAFSPS.L.Q...A.EK----SH.VVGQE.KDL.SE.Q-----------...T.V.V......T.H.R
Emx        Ip C..V.TAF.AT.LAM..RH.LQ----SP.VVGSQ.RKLSSD.G-----------.S.T.V.V......T.A.RCI DN43104_c0_g1_i1|m.22332
Emx        Sr S..I.TAF.PA.LAK..TQ.AR----SP.VVGHQ.RQ.SEE.G-----------...T.V.V......T.E.RVL DN118899_c0_g1_i1|m.56533
Emx a      Nw P..I.TAF.SE.LQR..N..TS----.P.IVGEE.RQF.AS.R-----------.S.T.V.V.Y....T.D.R.I 17210.0
Emx b      Nw P..I.TAF.TE.LQR..D..IA----.P.IVGEE.RRF.AN.N-----------...T.V.V.Y....T.D.R.T 44961.0
Emx        Xb P..V.TAFSPSQL.R..H..EK----NH.VVGEE.KQL.NV.N-----------.S.T.V.V......T.Y.RQK DN105712
EMXa 013   Nv P..I.TAF.PT.L.H..NA.EK----.H.IVGTE.KQL.SY.N-----------.S.T...V......T...RQQ DQ206302
EMXb 104   Nv VR.V.TAF.PF.L.C..TS.DK----.H.VVGTE.KQL.SY.K-----------.S.T.V.V......T...RQA DQ206286
```

Hhex family

```
Hhex       Hs R.G.QVRFSND..I....K.ET----QK..SPPE.KRL.KM.Q-----------.S..QV.T......A..RRLK
Hhex       Bf R.G.QVRFSND...E...K.ES----QK..SPPE.KRL.KS.Q-----------.....V.T......A..RRLK
Hhex       Dm R.G.QIRF.SQ..KN..AR.AS----SK..SPEE.RHL.LQ.K-----------..D.QV.T......A..RRAN
Pha-2      Ce R.G.QIRF.NE..DA..HK.DS----HK..SPQE.KKL.KS.S-----------.S..QV.T......A..RRVR
Hhex       Cg  .G.QVRFSND..V....K.ES----HK..SPPE.KRL.KT.Q-----------....QV.T......A..RR
Hhex       Xb R.G.QVRFSNE...K...T..K----QK..SPSE.KKI.KS.Q-----------.....V.T......A..RRLE DN94983
Hhex 048   Nv K.G.QVRFSNE..M....I.ET----QK..SPPE.KQLSKV.G-----------.S...V.T......A..RRFK DQ206317
```

Msx family

```
MSX1       Hs NRKP.TPF.TA.L.A..RK.RQ----KQ..SIAE.A.FSSS.S-----------...T.V........A.A.RLQ NP_002439
MSX2       Hs NRKP.TPF.TS.L.A..RK.RQ----KQ..SIAE.A.FSSS.N-----------...T.V.......SA.A.RLQ NP_002440
Msx        Bf NRKP.TPF.TQ.L.A..RK.RQ----KQ..SIAE.A.FSAS.N-----------...T.V........A.A.RLQ
Dr         Dm NRKP.TPF.TQ.L.S...K.RE----KQ..SIAE.A.FSSS.R-----------...T.V........A.A.RLQ NP_477324
Vab-15     Ce NRKP.TPFSTQ.LIS..RK.QS----KQ..SIAE.A.FSAS.Q-----------...T.V........A.S.RLQ
```

```
Msx         Cg  RKP.TPF.TS.L.A..RK.RQ----KQ..SIAE.A.FSAS.N-----------...T.V........A.A.R
Msx         Xb NRKP.TPF.TS.L.A..RK.RH----KQ..SIAE.A.FSSS.S-----------...T.V........A.A.RLQ DN63160
MSX (142)   Nv NRKP.TPF.TS.L.A..RK.RQ----KQ..SIAE.A.FSAS.N-----------...T.V........A.A.RLH DQ206296
```

Msxlx family

```
MSX1        Bf GRNP.VPFSAH.VAT..SK.RR----TH..SSVDVS.LSV..N-----------.S.NRV........ARERR.R
CG15696     Dm GRLP.IPF.PQ.LQA..NAYKE----SN..SAEDANKL.DS.E-----------..NTRV........ARERR.K
Msxlx       Cg   NP.VPF.QH.VAV..HK.RR----TH..SSMDVA.LST..S-----------...NRVSSVSA            (end ?)
Msxlx       Ip GRNP.VPFSST.VAV..QKYQH----KQ..SSIDVA.LSSL.N-----------.NDSRV........A.A.K.Q DN65413_c0_g2_i2|m.82199
Msxlx       Sr GRNP.VPFSSA.VAI..QKYQV----KQ..SSIDVA.LSTM.S-----------.NDSRV........ARAR..Q DN124792_c0_g1_i1|m.82241
Msxlx       Hm SRNP.IPF.SN.VAL..QT.QK----KQ..SSMDVSDLSNS.R-----------.SDTR.........ARARRDG DN104433_c0_g1_i1|m.45397
Msxlx       Nw GRNP.VPF.QT.VDH..LK.RE----AK..SSLDVAD.STH.N-----------..DTRV........ARERREK 4231
MSXLXa046   Nv SGNP.VPFSTS.LM...HKYVE----T...SGPEVA.L.NG.R-----------.S.HRV........ARE..AK DQ206269
MSXLXb066   Nv DRNP.IPFSAS.LAA..AK.LD----TH..SSVEVRDLSSS.S-----------V..HRV........ARE..TK DQ206275
```

NK-1/slouch family

```
NKX1-1      Hs PR.A.TAF.YE.LVA..NK.KA----T...SVCE.LNL.LS.S-----------...T.V........T....Q.
NKX1-2      Hs PR.A.TAF.YE.LVA..NK.RA----T...SVCE.LNL.LS.S-----------...T.V........T....Q. XP_372331
Nkx1a       Bf PR.A.TAF.YE.LVA..NK.KQ----T...SVCE.LNL.LS.S-----------...T.V........T....Q.
Nkx1b       Bf PR.V.TAF.YE.LVA..NK.KQ----T...SVCE.LNL.LS.S-----------...T.V........T....Q.
slou        Dm PR.A.TAF.YE.LVS..NK.KT----T...SVCE.LNL.LS.S-----------...T.V........T....Q. NP_476657
ceh-1       Ce MR.A.TAF.YE.LVA..NK.KT----S...SVVE.LNL.IQ.Q-----------.S.T.V........T....H.
Nk1         Cg  R.A.TAF.YE.LVA..NK.KT----T...SVCE.LNL.LS.N-----------...T.V........T....
NK1         Ip .R.S.TAF.YE.LVA..NK.RA----S...SVCE.LN..L..N-----------...T.V........T.F..Q. DN40909_c0_g1_i1|m.20738
NK1         Sr   .S.TAF.YE.LVA..NK.RC----S...SVCE.LN..L..N-----------...T.V........T.F..Q. DN56103_c0_g1_i1|m.8196
NK1         Xb                 ..NK.KL----T...SVCE.LNL.LS.S-----------...N.V........T....   DN131037
SLOU (091)  Nv PR.A.TAF.YE.LVA..NK.KS----T...SVCE.LNL.LS.G-----------...T.V........TNGRSRT DQ206255
```

NK-2.1 family

```
NKX2-1      Hs .RKR.VLFSQA.VY...RR.KQ----QK..SAPE.EHL.SMIH-----------..PT.V......H.Y.M.RQA NP_003308
NKX2-4      Hs .RKR.VLFSQA.VY...RR.KQ----QK..SAPE.EHL.SMIH-----------..PT.V......H.Y.M.RQA
Nkx2-1      Bf .RKR.VLFSQA.VY...RR.KQ----QK..SAPE.EHL.QLIN-----------..PT.V......H.Y.C.RQD
scro        Dm .RKR.VLF.QA.VY...RR.KQ----QR..SAPE.EHL.SLIH-----------..PT.V......H.Y.C.RQA
ceh-24      Ce .RKR.VLFSQA.VY...RR.KQ----AK..TAPE.EQL.NSIR-----------..PT.V......H.Y.C.RQE
ceh-27      Ce .RKR.VLFSPQ.VHV..RK.QI----....SAAD.ENL.KSIN----- -----.SAT.V......Q.Y.C.RQE
Nkx2_1      Cg .RKR.VLFSQA.VY...RR.KQ----QK..SAPE.EHL.SMIN-----------..PT.V......H.Y.C.R
Nkx2.1 a    Ip .RKR.VLF.QA.IY...RR.KQ----QK...AAE.ENL..VIN-----------.SPT.V......H.Y.T..SK DN64840_c0_g1_i1|m.78270
Nkx2.1 a    Sr .RKR.VLF.QA.IH...RR.KQ----QK...AAE.ESL..LIN-----------.SPT.V......H.Y.T..SK DN122139_c1_g1_i1|m.69070
Nkx2.1 b    Ip  RKR.VLF.QA.VY...RR.KA----QK..SAPE.ETL..SIG-----------..PT.V.....          DN66164_c5_g1_i1|m.87491
Nkx2.1 c    Ip .RKR.VLF.QA.VY...RR.KT----QK..SAPE.ETL...IN-----------..PT.V......H.Y.C.RAS DN66164_c5_g2_i1|m.87495
Nkx2.1 b    Sr .RKR.VLF.QA.VY...RR.KQ----QK..SAPE.EAL.QGIN-----------..PT.V......H.Y.C.RAS DN126831_c2_g4_i1|m.94319
```

```
Nkx2.1 a   Hm .RKR.VLF.QA.VY...RR.KQ----Q...SAPE.ENL.QLIN-----------..PT.V......H.Y.Y.RAS DN44693_c0_g1_i1|m.40284
Nkx2.1 b   Hm .RKR.VLF.QA.VY...RR.KQ----Q...SAPE.ETL..LIN-----------..PT.V......H.Y.C.RAS DN15813_c0_g1_i1|m.4392
Nkx2.1 c   Hm PRKR.ILF.QA.VY...RR.KM----QK..NAPE.ESL.QLIN-----------..PT.V......H.Y.C.RAA DN36484_c0_g2_i1|m.17692
Nkx2.1     Nw .RKR.VLF.QA.VY...RR.KQ----QK..SAPE.EQL.AMIH-----------..PT.V......H.Y.M.RQA 45472.0
Nkx2.1     Xb KRKR.VLF.QA.VY...RR.KQ----QK..SAPE.EHL.SLIN-----------..PT.V......H.Y.C.RAA 93010
```

Nk-2.2 family

```
NKX2-2     Hs KRKR.VLFSKA..Y...RR.RQ----Q...SAPE.EHL.SLIR-----------..PT.V......H.Y.M.RAR NP_002500
NKX2-8     Hs ..KR.VLFSKA......RR.RQ----Q...SAPE.EQL.SL.R-----------..PT.V......H.Y.L.RAR NP_055175
Nkx2-2     Bf KRKR.VLFSKA..Y...RR.RQ----Q...SAPE.EHL.RLIR-----------..PT.V......H.Y.C.RAQ
vnd        Dm KRKR.VLF.KA..Y...RR.RQ----Q...SAPE.EHL.SLIR-----------..PT.V......H.Y.T.RAQ NP_476786
ceh-22     Ce KRKR.VLF.KA..Y...RR.RS----QK..SAPE.EAL.MQIR-----------..PT.V......H.Y.T..SH
Nk2_2      Cg  RKR.VLFSKA..Y...RR.RQ----Q...SAPE.EHL.SIIR-----------..PV.V......H.Y.L.R
Nkx2.2     Ip KRKR.VLF.KN..Y...RR.RQ----Q...SAPE.EAL.AMID-----------..PG.V......H.Y.C..S. DN66907_c1_g1_i2|m.94022
Nkx2.2     Sr KRKR.VLFSKS..FQ..SR.RQ----Q...SAPE.EQL.N.IN-----------..PG.V......H.Y.C..SS DN89648_c0_g1_i2|m.18777
Nkx2.2     Hm KRKR.VLFSKA.TF...RR.RQ----Q...SAPE.EHL.NLIK-----------..PT.V......H.Y.C..NH DN90735_c0_g1_i1|m.44111
Nkx2.2     Nw KRKR.VLFSKQ..Y...RR.RQ----Q...SAPE.EHL.SIIN-----------..PT.V...... 20731.0
Nkx2.2     Xb            ..Y...RR.RQ----Q...SAPE.EHL.SIIR-----------.SPT.VEV....H.Y.L.RSR DN101095

NK2c (051) Nv KRKR.ILF.KS.IY...RR.RY----Q...SAPE.EQLGRMIN-----------..PT.V......H.Y.H..QA DQ206270
NK2d (103) Nv KRKR.VLF.KA..Y...RR.RQ----Q...SAPE.EQL.RIIN-----------..PT.V......H.Y.F..Q. DQ206285
NK2a (057) Nv KRKR.ILF.KA..FI..RR.TQ----Q...SAPE.E.L.RIAN-----------..PA.V......H.Y.YR.QM DQ206320
NK2b (025) Nv K.KR.ILF.KS.IF...RR.RQ----QK..SANE.EQL.RIID-----------..PT.V......H.Y.F..QI DQ206262
NK2e (026) Nv KRKQ.VLFSKV.RKQ...K.LE----KP..SYTE.DQL.R.VN-----------..AK.V.V....K.Y.T.ERE DQ206310
```

Nk-3 family

```
NKX3-1     Hs Q..S.AAFSHT.VI...RK.SH----QK..SAPE.AHL.KN.K-----------...T.V........Y.T.RKQ NP_006158
NKX3-2     Hs K..S.AAFSHA.VF...RR.NH----Q...SGPE.ADL.AS.K-----------...T.V........Y.T.RRQ NP_001180
Nkx3       Bf K..S.AAFSHA.VF...RR.SH----Q...SGPE.ADL.A..K-----------...T.V........Y.T.RRQ
bap        Dm K..S.AAFSHA.VF...RR.AQ----Q...SGPE.S.M.KS.R-----------...T.V........Y.T.RKQ NP_732637
Nk3        Cg  ..S.AAFSHA.VF...RR.RH----Q...SGPE.ADL.N..K-----------...T.I........Y.T.R
Nk3        Ip   .T.AAFSHA.VF...RR.AQ----QK..SGPE.ADL.N..K-----------...T..........Y.T.RKQ DN64224_c0_g1_i1
Nk3        Sr K..T.AAFSHA.VF...RR.AQ----QK..SGPE.S.L.S..K-----------...T..........Y.T.RKQ DN121921_c0_g2_i1|m.68018
Nk3        Hm K..T.AAFSHA.VF...RR.TQ----Q...SGPE.S.V.SN.K-----------...T.V........Y.T.RKQ DN79735_c0_g1_i1|m.43259
NK3 (098)  Nv ...S.AAFSHQ.VF...RR.SH----QK..SGPE.ADL.A..K-----------...T.V........Y.T.RRQ DQ206336
```

Nk-4 family

```
NKX2-3     Hs .RKP.VLFSQA.VF...RR.KQ----Q...SAPE.EHL.SS.K-----------..ST.V........Y.C.RQR
NKX2-5     Hs .RKP.VLFSQA.VY...RR.KQ----Q...SAPE.DQL.SV.K-----------..ST.V........Y.C.RQR
NKX2-6     Hs .RKP.VLFSQA.VLA..RR.KQ----Q...SAPE.EHL.S..Q-----------..ST.V........Y.C.RQR
csx        Bf .RKP.VLFSQA.VF...RR.KQ----Q...SAPE.EQL.QM.K-----------..ST.V........Y.C.RQR
```

```
Tin         Dm KRKP.VLFSQA.VL...CR.RL----KK...GAE.EI..QK.N-----------.SAT.V........Y.S.RGD
ceh-28      Ce KRKP.VLF.QH.VN...ER.KK----Q..V.ATE.E.L.QC.G-----------..AT.V........Y.C.RLA
Nk4         Cg  RKP.VLFSQA.VF...RR.KQ----Q...SAPE.EQL.NL.K-----------..ST.V........Y.C.R
Nk4         Hm ..KQ.ILFHHR.IEQ..DF.AA----QP.ISAIE.QCLSQKTG-----------..NN.V........Y.E.RLK DN91365_c0_g1_i1|m.44211
Nk4         Nw  RKP.ILFNHS.IFQ..LR.KH----Q....AND.R.LSDQ.N-----------..QN..........Y.S.RQE 15302
Nk4         Xb .RKP.VLF.QA.VY...RR.KQ----Q...SAPE.EHL.HL.K-----------LSST.V........Y.C.RQR DN50727
NK4         Nv .RKP.VLFSQA.VY...RR.KG----QK..SAPE.DHL.SL.K-----------..PN.V......K.Y.C..QA Q0N4L5
```

Nk5/Hmx family

```
HMX1        Hs K.KT.TVFS.S.VFQ..ST.DL----K...STAE.AGL.AS.Q-----------...T.V........N...RHV NP_061815
HMX2        Hs K.KT.TVFS.S.VYQ..ST.DM----K...SSSE.ACL.SS.Q-----------...T.V.T......N...RQL NP_005510
LOC340784   Hs K.KT.TVFS.S.VFQ..ST.DM----K...SSSE.AGL.AS.H-----------...T.V........N...RQL XP_291716
Hmx         Bf K.KT.TVFS.Q.VFQ..ST.DM----K...SSSE.AGL.AS.H-----------...T.V........N...RQL
hmx         Dm K.KT.TVFS.A.VFQ..ST.DL----K...SSSE.AGL.AS.R-----------...T.V........N...RQL NP_524951
mls-2       Ce K.KT.TVFS.S.VSQ..MM.EC----K...SSQE.SNL.QK.H-----------...T.V........N.F.RQA
Hmx1/nk5    Cg  .KT.TVFS.H.IFY..SA.DA----K...SSTE.S...ST.N-----------...T.V.V......N...S
Hmx2/nk5    Cg  .KT.TVFS.S.VFQ..ST.DM----K...SSSE.AGL.AS.Q-----------...T.V........N...R
Nk5         Ip R.KT.TVFS.N.VCQ..SM.DM----K...SSSE.TGI.RD.G-----------.S.Q.I........N.L.RQI DN15610_c0_g1_i1|m.6431
Nk5 a       Sr K.KT.TVFS.T.VCQ..NM.DL----K...SSSE.ATL.RD.T-----------.S.Q.I........N.L.RQI DN110321_c0_g1_i1|m.36339
Nk5 a       Hm K.KS.TVFS.S.VCQ..GT.DQ----K...SSAE.STL.RS.K-----------...T.V........N.L.RQI DN8401_c0_g1_i1|m.2232
Nk5 b       Sr K.KT.TVFS.S.VFQ..ST.DV----K...SSSE.ANL.NS.H-----------...T.V........N...RQI DN123682_c0_g2_i1|m.76247
Nk5 b       Hm K.KT.TVFS.N.VFQ..TT.DL----K...SSSE.ASL.AT.S-----------...T.V........N...RQL DN43016_c0_g2_i1|m.36485
Nk5         Nw K.KT.TVFS.H.VYQ..TM.DC----KK..SSSE.AGL.AT.H-----------...T.V........N.L.RQF 36577.0
Nk5         Xb K.KT.TVFS.S.VFQ..ST.DM----KR..SSSE.AGL.AS.H-----------...T.V........N...RQL DN92869
HMX (116)   Nv K.KT.TVFS.S.VYQ..ST.DM----K...SSSE.AGL.SQ.H-----------...T.V........N...RQI DQ206341
```

Nk-6 family

```
NKX6-1      Hs ..HT.P.FSGQ.IFA...T.EQ----TK..AGPE.ARL.YS.G-----------M..S.V.V......T..R.KH NP_006159
NKX6-2      Hs K.HS.P.FSGQ.IFA...T.EQ----TK..AGPE.ARL.YS.G-----------M..S.V.V......T..R.RH NP_796374
NKX6-3      Hs K.HT.P.FTGH.IFA...T.EQ----TK..AGPE.ARL.YS.G-----------M..S.V.V......T..R.KS
Nkx6        Bf ..HT.P.FSGQ.IFA...T.EQ----TK..AGPE.ARL.YS.G-----------M..S.V.V......T..R.KH
HGTX        Dm K.HT.P.FSGQ.IFA...T.EQ----TK..AGPE.AKL.Y..G-----------MS.S.V.V......T..R.RH NP_652614
Cog-1       Ce K.QS.P.FTGH.IYQ..RK.EQ----TK..AGAD.AQL.QE.N-----------MS.S.V.V......T..R.KE
Nk6         Cg  .HT.P.FSGQ.IFA...T.EQ----TK..AGPE.ARL.Y..G-----------MS.S.V.V......T..R.
Nk6         Ip K.HT.P.FSGH.IFA...T.EQ----TK..AGPE.ARL.F..G-----------MS.S.V.V......T..R.KH DN69723_c1_g1_i4|m.125267
Nk6         Sr K.HT.P.FSGH.IFA...T.EQ----TK..AGPE.ARL.F..G-----------M..S.V.V......T..R.KH DN130725_c3_g1_i1|m.123269
Nk6         Hm K.HT.P.FSGH.IFA...T.EQ----TK..AGPE.ARL.Y..G-----------MS.S.V.V......T..R.KH DN10228_c0_g1_i1|m.2704
Nk6         Nw K.HT.P.FSGH.IFA...T.EQ----TK..AGPE.GRL.YT.G-----------MS.S.V.V......T..R.RA 41714.0
Nk6         Xb ..HT.P.FSGH.IFA...T.EQ----TK..AGPE.RRL.FS.G-----------M..S.V.V......T..R.KH DN75963
NK6 (112)   Nv ..HT.P.FSGH.IFA...T.EQ----TK..AGPE.ARL.YS.G-----------M..S.V.V......T..R.RH DQ206340
```

Nk7 family

```
           Bf K.KA.T.F.GR.IF....Q.EV----KK..SASE.ADL.AL.N-----------VTDT.V.......T....QD
Nk7.1      Dm K.KA.T.F.GR.IF....M.EN----KK..SASE.T.M.KL.M-----------V..T.V.......T....QD
ceh-9      Ce R.KA.T.FSGK.VF....Q.EA----KK..SSSD.S.L.KR.D-----------V..T.V.......T....IE
Nk7        Cg  .KA.T.F.GR.IF....Q.EQ----KK..SSAE.A.M.TL.N-----------V..T.V.......T....
Nk7        Xb R.KA.TVF.GR.IF...RQ.EA----KK..SAHE.A.M.AL.N-----------V.DQ.V.......T...RLD DN94680
NK7 (070)  Nv K.KA.T.F.GR.IF....Q.EA----KK...ATE.SDM.SL.N-----------V..T.V.......T....QE DQ206276
```

Tlx family

```
TCL3       Hs K.KP.TSF..L.IC....R..R----QK..ASAE.AAL.K..K-----------M.DA.V.T......T..RRQT NP_005512
TLX2       Hs ..KP.TSFS.S.V....RR.LR----QK..ASAE.AAL.K..R-----------M.DA.V.T......T..RRQT NP_001525
TLX3       Hs ..KP.TSFS.V.IC....R..R----QK..ASAE.AAL.KS.K-----------M.DA.V.T......T..RRQT NP_066305
           Bf ..KP.TSF..L.IC....R..K----QK..ASAE.AQL.TQ.K-----------M.DA.V.T......T..RRQT
C15        Dm ..KP.TSF..I.VA....R..K----QK..ASAE.AAL.RG.K-----------M.DA.V.T......T..RRQT NP_476873
Tlx1       Cg  .KP.TSFS.MTII....R.DR----QK..ASSE.TSL.KS.K-----------ISDS.V.T......T..RR
Tlx2       Cg  .KP.TSFS.L.IM....R..R----QK..ASAE.STL.K..K-----------M.DA.V.T......T..RR
Tlx        Ip ..KP.TSF..L.IC....R..Q----QK..ASGE.ATL.SR.N-----------M.DS.V.T......T..RRQS DN68209_c2_g2_i2|m.107169
Tlx        Sr ..KP.TSF..L.IC....R.HQ----QK..ASGE.ATL.SK.N-----------M.DS.V.T......T..RRQS DN74225_c0_g2_i1|m.12757
Tlx        Hm ..KP.TSF..L.IC....R..L----QK..ASTE.ASL.KS.K-----------M.DA.V.T......T..RRQT DN40476_c0_g2_i2|m.25716
Tlx        Nw ..KP.TSF..V.VG....R.IV----QK..ASAE.NSL.KG.G-----------M.DA.V.T......T..RRQT 33755.0
Tlx        Xb  .KP.TSF..L.IM....R..K----QK..ASAE.SSL.KD.K-----------M.DA.V.T......T..RR DN135951
NVHD004    Nv K.KP.TAF.ES.IS....R.QS----QK..GSKE.S.L.GT.G-----------..DT.V                DQ206223
NVHD024    Nv A.KP.TAF.ES.IS....R.QV----TEIPREQE.S.L.GT.G-----------..DT.V.T.......L.RQR
```

Vax family

```
VAX1       Hs P..T.TSF.AE.LYR..M..QR----CQ.VVG.E.T.L.RQ.N-----------.S.T.V.V......T.Q..DQ
VAX2       Hs P..T.TSF.AE.LYR..M..QR----CQ.VVG.E.T.L.RQ.N-----------.S.T.V.V......T.Q..DQ NP_036608
           Bf P..A.TAF.AD.LKD..RAFQG----SQ.VVGQD.T.L.RQ.R-----------...T.V.V......T.Y.R.K
Ceh-5      Ce P..P.TVF.DE.LEK..ESFNT----SG..SGST.AKL.ES.G-----------.SDN.V.V......T.Q..ID
Vax        Cg  ..A.TTFSPE.LYQ.....QR----.Q..VGKE.T.L.KL.K-----------.S.T.NNITFE
Vax a      Ip P..E.TSF.PE.LFR.....RA----.Q.MVG.D.GDL.AS.G-----------.S.T.V.V......T.Y.R.K DN60553_c0_g2_i1|m.55905
Vax a      Sr P..E.TSF.PE.LFR.....SL----.Q.MVG.D.GQL.TC.G-----------...T.V.V......T.Y.R.K DN116214_c0_g1_i1|m.48271
Vax        Hm P..E.TSF.PQ.LYQ......K----.Q.MIG.D.A.L.VL.K-----------...T.V.V......T.Y.R.K DN58320_c0_g1_i1|m.41539
Vax b      Ip P..E.TSF.AQ.LFL..E..LK----AQ..VG.D.T.L.KK.N-----------.S.T...V......T.H.R.K DN89774_c0_g1_i1|m.233752
Vax b      Sr P..E.TCFSAE.LFL..E..T.----GQ.MVG.D.TLL.GR.G-----------...T...V......T.Q.R.S DN109741_c1_g1_i1|m.35411
Vax a      Nw P..Q.TSF.AE.LYI.....RR----.Q.MVGKD.GLL.ES.D-----------...T.V.V......T.Y.R.K 49354.0
Vax b      Nw  ..Q.TCFSAE.LYV..QE.IK----NQ.MVGKD.TLL.QS.T-----------.S.T.V*                20893.0
VAX (072)  Nv P..S.TSF.PA.LDR..D..RV----DM.VVGLK.MKL.ND.N-----------.S...V.V.......Y.R.R DQ206277
NVHD010    Nv ...A.TAFSAE.LKK...R.QA----.H.IVGEE.QK..KD.D-----------.S.A.V.V......T.F.RDQ DQ206300
```

Novel Xenacoelomorpha family

```
           Ip K.KN.TAF.PG.IF....R.RV----Q...SASE.G.F.SH.N-----------..DT...T.......HKRQQ DN56009_c0_g1_i5|m.42659
           Sr ..KN.TAF.PH.IY....R.RS----QK..SASE.G.F.LS.K-----------..DT...T.......H.RQQ DN126921_c0_g1_i2|m.94661
           Hm K.KN.TAF.PG.IC...NR.RL----QQ..STTE.G.F.SR.K-----------..DT.V.T.......L.RQQ DN3930_c0_g1_i1|m.1095
           Ip S.KQ.TAF.PD.VY...RR.SQ----QK..SASE.A.F.SK.Q-----------.SDT...T.......R.RQQ DN60800_c0_g1_i2|m.56990
           Sr C.KQ.TAF.PD.IH...RR.AQ----QK..SASE.ADF.NK.K-----------.SDT...T.......R.RQQ DN130717_c3_g1_i1|m.123804
           Hm K.KQ.TAF.PE.VF....R.QI----QK...ASE.SDFSDS.K-----------.SDT.V.T.......H.RQQ DN20211_c0_g1_i1|m.5588
50700.0    Nw N.KS.TSF.AC.VVK..QK.QS----QK..AAVE.GEF.EK.G-----------.SDT.V.T.......H.RVQ
36321.0    Nw G.KS.TSF.PV.V.K..HK.QT----QK..AAVE.GEF.EK.G-----------..DT.V.T.......H.RVQ
DN122354   Xb K.KT.TAFSPS.IY....R.RL----QK..AASE.GEF.DK.S-----------..DT.V.T.......Y.RQH
```

ANTP CLASS Novel/ambiguous family membership

```
           Ip                                .A......SS.N-----------.N.T.V.........Q..LV DN65240_c0_g1_i5
           Ip KRQG.TVF.KQ.LEI.NGH.NH----RQ..SLAE.AQLSEQ.C-----------.SQK.V.....M.S.A.RVM DN26203_c0_g1_i1|m.10708
           Nw  SHQDHVENQGYQPI..SK.AI----NN.ISGSEVE.LSNR.G-----------I.DI..RD..A...RRGQFSQ 45900.0
           Nw LRIRTPITG.HRASMFQHFLTK----SQ...PVDCSNMSCN.G-----------IP.PA.RVFF..LRARVRAR  31941.1
           Xb G..M.TVF.YK.LQK...V.QD----QQ.MVGED.CRL.DS.G-----------.S.T.VRV......I..R..T DN41925
           Xb S..R.TSFSQE.IRL...K.EE----KQ...AKE.DD..CD.D-----------.KAS.VIT....K.S.IRR.T DN69228
```

PRD CLASS

Alx 3/4 (CART1) family

```
ALX1/CART1 Hs KR.H.T.F.SL.LE....V.QK----TH.PDVYV.EQL.LRTE-----------...ARVQV......A..R.RE NP_008913
ALX3       Hs KR.N.T.FSTF.LE....V.QK----TH.PDVYA.EQL.LRTD-----------...ARVQV......A..R.RE NP_006483
ALX4       Hs KR.N.T.F.S..LE....V.QK----TH.PDVYA.EQL.MRTD-----------...ARVQV......A..R.RE NP_068745
           Bf KR.N.T.F.SF.LE.M.RV.QK----TH.PDVYM.EQL.LRAD-----------...ARVQV......A..R.KE
Alx        Hm KR.N.T.F.NL.LE.M.KI.QK----TH.PDVYV.EQL.NRCE-----------...ARVQV......A..R.RE DN38072_c0_g2_i1|m.20141
Alx        Sr KR.N.T.F.QG.LE.M.RI.QK----TH.PDVYV.EQL.LKCD-----------...ARVQV......A..R.KE DN122029_c0_g1_i1|m.68599
Alx        Ip KR.N.T.F.QA.LE.M.RI.QK----TH.PDVYV.EQL.VKCD-----------...ARVQV......A..R.KE DN38710_c0_g1_i2|m.18965
Alx a      Nw KR.N.T.F.S..LD...RI.QK----TH.PDVYC.EQL.ARCN-----------...ARVQV......A..R.KE 55290.0
Alx b      Nw KR.N.T.F.S..LD...RI.QK----TH.PDVYC.EQL.VRCT-----------...ARVQV......A..R.KE 40829.0
Alx        Xb KR.N.T.F.SF.LD...RV.QK----TH.PDVYL.EQL.LRAD-----------...ARVQV......A..R.KE DN84186
CART1 120  Nv KR.N.T.F.A..LE.M.RV.QK----TH.PDVYT.EQL.LRCA-----------...ARVQV......A..R.RE DQ206256
```

Drgx family

```
DRGX       Hs QR.N.T.F.LQ.LEA..AV.AQ----TH.PDVFT.E.L.MKIN-----------...ARVQV......A..R.TE
Drgx       Bf QR.N.T.F.LH.LEQ..AV.AQ----TH.PDVFT.E.L.MKIN-----------...ARVQV......A..R.QE
CG34340    Dm QR.N.T.F.LQ.LEE..TA.AQ----TH.PDVFT.EDL.MKIN-----------...ARVQV......A..R.AE
Drgx       Cg  R.N.T.F.LQ.LEE..KA.AQ----TH.PDVFT.EDL.MRIN-----------...ARVQV......A..R.
DRGX       Xb QR.N.T.F.LH.LEQ..TA.AQ----TH.PDVFT.E.L.LKIN-----------...ARVQV......A..R.HE DN102158
```

Gsc family

```
GSC        Hs KR.H.TIF.DE.LEA..NL.QE----TK.PDVGT.EQL.RKVH-----------.R.EKVEV..K...A..RRQK NP_776248
GSCL       Hs TR.H.TIFSEE.LQA..AL.VQ----.Q.PDVST.ERL.GRIR-----------.R.ERVEV..K...A..RHQK NP_005306
Gsc        Bf KR.H.TIF.EE.LEL..KT.EK----TH.PDVLL.E.L.MKVE-----------.K.ERVEV..K...A..R.QQ
Gsc        Dm KR.H.TIF.EE.LEQ..AT.DK----TH.PDVVL.EQL.LKVD-----------.K.ERVEV..K...A..R.QK NP_476949
Ceh-45     Ce KR.H.TIFSEE.LNI..TT.ST----TH.PDATT.E.L.VQCS-----------.K.ERVEV..K...A.ER.QK
Gsc1       Cg                                ML.E.L.MKVD-----------.K.ERVEV..K...A..R.
Gsc2       Cg  R.H.TIF.EE.LE....T.NK----TH.PDVLL.E.L.MKID-----------.K.ERVEV..K...A..R.
GSC        Ip KR.H.TIF.EE.LA...AV.VR----SH.PDVLA.EKV.TS.G-----------.K.ERVEV..K...A..R.QK DN67235_c0_g4_i4
GSC        Sr KR.H.TIFSEE.LK...EA.LK----CH.PDVVT.EGV.QK.S-----------.K.ERVEV..K...A..R.QK DN120054_c0_g1_i1|m.60873
GSC        Hm TR.H.TIF.EE.LE...KT.SL----TH.PDVQT.EQL.AKAD-----------.K.ERVEV..K...A..R.QK DN27253_c0_g1_i2|m.9227
GSC        Nw KR.H.TIFSEE.LD...TT.AK----TH.PDVLL.EQL.LRVE-----------.K.ERVEV..K...A..R.YK 40214.0
GSC        Xb KR.H.TIF.EE.LE...RT.EK----TH.PDVLL.EQL.VKVD-----------.KEERVEV..K...A..R.QK DN48278
GSC (068)  Nv KR.H.TIF.EE.-....TT.QK----TH.PDVLL.E.L.MKVD-----------.K.ERVEV..K...A..R.QK DQ206230
```

Hbn family

```
Hbn        Dm VR.S.T.F.TF.LHQ..RA.EK----TQ.PDVFT.EDL.MR.D-----------.S.ARVQV......A..R.RE
Hbn        Cg  R.S.T.F.T..LHQ..RA.EK----TQ.PDVFT.E.L.MR.D-----------.S.ARVQV......A..R.
HBN        Xb VR.S.T.F.TF.LHH..RA.EK----SH.PDVFT.E.L.AS.H-----------.S.ARVQV......A..RKTE DN83632
HBN (079)  Nv VR.S.T.F.T..LHQ..RA.EK----TQ.PDVFT.E.L.LR.D-----------.S.ARVQV......A..R.RE DQ206231
```

Otp family

```
OTP        Hs Q..H.TRF.PA.LN...RS.AK----TH.PDIFM.E.L.LRIG-----------...SRVQV......A....RK NP_115485
           Bf Q..H.TRF.PA.LQ...RS.SK----TH.PDIFM.E.L.LRIG-----------...SRVQV......A....RK
otp        Dm Q..H.TRF.PA.LN...RC.SK----TH.PDIFM.E...MRIG-----------...SRVQV......A....RK NP_995909
Otp        Cg  ..H.TRF.PA.LN...RS.AK----TH.PDIFM.E.L.LRIG-----------...SRVQV......A....
OTP        Ip A..H.TRF.PA.LN...RC.NK----TH.PDIFM.E.L.MRIA-----------...SRVQV......A....RK DN57662_c0_g1_i1|m.46698
OTP        Sr P..H.TRF.PA.LN...RC.NK----TH.PDIFM.E.L.MRIA-----------...SRVQV......A....RK DN116171_c0_g1_i2|m.48160
OTP        Hm P..H.TRFSPN.LN...RC.GK----TH.PDIFM.E.M.LRIG-----------...SRVQV......A....RK DN10420_c0_g1_i1|m.2841
OTP        Nw P..H.TRF.PA.LN...RC.TK----TH.PDIFM.E.L.VRIG-----------...SRVQV......A....RK 53544.0
OTP        Xb P..H.TRF.PA.LN...RS.NK----TH.PDIFM.E.L.MRIG-----------...SRVQV......A....RK DN86933
OTP (047)  Nv Q..H.TRF.PA.LN...RC.AR----TH.PDVFM.E.L.ARIG-----------...SRVQV......A....RK DQ206247
```

Otx family

```
OTX1       Hs QR.E.T.F..S.LDV..AL.AK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQQQ NP_055377
OTX2       Hs QR.E.T.F..A.LDV..AL.AK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQQQ NP_068374
CRX        Hs QR.E.T.F..S.LE...AL.AK----TQ.PDVYA.E.V.LKIN-----------.P.SRVQV..K...A.CRQQR NP_000545
otx        Bf QR.E.T.F..A.LDV..AL.AK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQQA
oc         Dm QR.E.T.F..A.LDV..AL.GK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQQL NP_511091
ceh-36     Ce GR.E.TSFN.G.LDQ...V.RE----TQ.PDVHR.EAL.K.IN--------LPDGRVQV.TV..K...A.DRNNK
ceh-37     Ce NR.E.T.YS.Q.LEI..TL.NE----TQ.PDVFA.ERV.DQIR-----------.Q.SR.QV..K...A.YRLQE
```

```
ttx-1       Ce QR.E.T.F..N.LEI..SY.VK----T..PDIFM.EDM.HKIQ-----------.P.SRVQV..K...A.ARQQK
Otx         Cg  R.E.T.F..A.LDI..SL.QK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQ
OTX         Ip QR.E.T.F..T.LEI..QL.SK----T..PDIFM.E.V.NRIN-----------.P.SRVQV..K...A.CRQQQ DN73075_c6_g1_i5|m.177200
OTX         Sr QR.E.T.F..T.LEI..AL.SK----T..PDIFM.E.V.NRIN-----------.P.SRVQV..K...A.CRQQQ DN123029_c0_g3_i3|m.73095
OTX a       Hm QR.E.T.F..A.LEI..SL.SK----T..PDIFM.E.V.MKIN-----------.P.SRVQV..K...A.CRQQQ DN24659_c0_g2_i1|m.8154
OTX b       Hm PR.E.T.FSKN.VLY..EL.TK----T..PDVFM.Q...DK.H-----------...SRVQV..K...A.SRQQT DN56217_c0_g1_i1|m.41307
OTX         Nw QR.E.T.F..A.LEV..SL.TK----T..PDIFM.E.V.MKIN-----------.P.SRVQV..K...A.CRQQQ 48073.0
OTX         Xb QR.E.T.F..A.LEV..AL.SK----T..PDIFM.E.V.LKIN-----------.P.SRVQV..K...A.CRQQA DN55770
OTXa 114    Nv QR.E.T.F.KN.LEV..EL.AK----T..PDIFM.E.V.IKIN-----------.P.SRVQV..K...A.TRQLE DQ206236
OTXb 123    Nv QR.E.T.F.KN.LEI..EL.AK----T..PDIFM.E.V.IKIN-----------.P.SRVQV..K...A.ARQQA DQ206257
OTXc 078    Nv QR.E.T.F.KN.LEI..EL.AK----T..PDIFM.E.V.IKIN-----------.P.SRVQV..K...A.LRQLS DQ206252
```

Pax 4/6 family

```
PAX4        Hs GH.N.TIFSPS.AEA.....QR----GQ.PDSVA.GKL.T.TS-----------.P.DTVRV..S...A..RRQE NP_006184
PAX6        Hs LQ.N.TSF.QE.IEA.....ER----TH.PDVFA.ERL.AKID-----------.P.AR.QV..S...A..RR.E NP_001595
            Bf LQ.N.TSF.QE.IEA.....ER----TH.PDVFA.ERL.AKID-----------.P.ARIQV..S...A..RR.E
eyg         Dm FR.N.T.FSPE.LE......DK----SH.PCVST.ERLSSRTS-----------.S.ARVQV..S...A..RRH NP_524042
toe         Dm FR.N.T.FSPE.LD......DK----SH.PCVNT.EKL.ARTA-----------.S.ARVQV..S...A..RRHQ NP_524041
toy         Dm LQ.N.TSFSNE.IDS.....ER----TH.PDVFA.ERL.DKIG-----------.P.AR.QV..S...A..RR.E NP_524638
ey          Dm LQ.N.TSF.ND.IDS.....ER----TH.PDVFA.ERL.GKIG-----------.P.AR.QV..S...A..RR.E NP_726607
vab-3       Ce LQ.N.TSF.QV.IES.....ER----TH.PDVFA.ERL.QKIQ-----------.P.AR.QV..S...A..RR.E
eyg-1       Ce .RLS.STF.NE.LQS..EV.LR----DP.PSPNE.ADLVKRTN-----------.P.AR.QV..S...A..R.TN
Pax4_6      Cg  R.NRTTF.SQ.LDL..QS.QK----AH.PGVQT.E.L.TKTN-----------.S.ARVQV..S...A..RR
Pax6        Cg  Q.N.TSF.AA.IEA.....ER----TH.PDVFA.ERLSQKID-----------.P.AR.QV..S...A..RR
Pax4/6 a    Ip NQ.N.TAYSQN.LEL.....EK----SH.PDVYA.ERLSEATD-----------IP.SK.Q...S...A.HRR.E DN54186_c0_g1_i2|m.38019
Pax4/6 a    Sr SQ.N.TA..SE.VDA.....EK----SH.PDVYA.ERLSTLTG-----------IP.TK.Q...S...A.HRR.E DN127496_c0_g2_i1|m.98660
Pax4/6 a    Hm FQ.N.TSF.QS.IDA.....ER----TH.PDVYA.ERLSESTS-----------.A.NK.QV..S...A..RR.E DN37527
Pax4/6 b    Ip LQ.N.TSF.PE.IDA...Q.EV----TH.PDVFA.ERL.GKIA-----------.P.AR.QV..S...A..RR.E DN12334_c0_g1_i1|m.4957
Pax4/6 b    Sr LQ.N.TSF.PE.IEA...Q.EV----TH.PDVFA.ERL.GKIN-----------.P.AR.QV..S...A..RR.E DN116853_c0_g1_i1|m.50040
Pax4/6 b    Hm LQ.N.TSF.QE.IDN..L..DR----TH.PDIFA.EKLSEKIK-----------.A.AR.QV..S...A..RR.E DN15356_c0_g1_i1|m.4176
PAX4/6 a    Nw LQ.N.TSFSVE.IDA..ST.NE----SH.PDSNE.ERL.EKVG-----------IP.AK.QV..S...AR.RR.D 24274.0
PAX4/6 b    Nw LQ.N.TSF.PE.IDA.....ER----TH.PDVFA.EKL.EKIG-----------.P.AR.QV..S...A..RR.E 49939.0 also 46535.0
PAX4/6 a    Xb LQ.N.TSF.PE.IDS.....ER----TH.PDVFA.ERL.KDID-----------.P.AR.QV..S...A..RR.E DN77439
PAX4/6 b    Xb LQ.N.TSF.PE.LES.....ER----TH.PDVFA.ERL.AKID-----------.P.AR.QV..S...A..RR.E DN86691
PAX4/6 022  Nv LR.N.T.F.PD.LEM.....EK----SH.PDVAT.E.L.NKID-----------MS.ARVQV..S...A..RRHQ DQ206243
NVHD148     Nv SR.Q.TNF.DE.IEK...V.EK----TH.PDVFT.E.L.QQVN-----------.S.AR.QV.YS...A..R.EG DQ206260
```

Pitx family

```
PITX1       Hs QR.Q.THF.SQ.LQ...AT.QR----...PDMSM.E...VWTN-----------...PRVRV..K...A..R.RE NP_002644
PITX2       Hs QR.Q.THF.SQ.LQQ..AT.QR----...PDMST.E...VWTN-----------...ARVRV..K...A..R.RE NP_000316
PITX3       Hs QR.Q.THF.SQ.LQ...AT.QR----...PDMST.E...VWTN-----------...ARVRV..K...A..R.RE NP_005020
```

```
pitx        Bf QR.Q.THF.SQ.LQ...AS.AR----...PDMAT.E...AWTN-----------...ARVRV..K...A..R.RE
ptx1        Dm QR.Q.THF.SQ.LQ...HT.SR----...PDMST.E...MWTN-----------...ARVRV..K...A..R.RE NP_996314
unc-30      Ce  R.Q.THF.SH.LT...NW.SR----...PDMAC.E...VWIS-----------...PRVRV..K...A..R.RE
Pitx        Cg  R.Q.THF.SQ.LQ...AT.AR----...PDMAT.E...AWTN-----------...ARVRV..K...A..R.
Pitx a      Ip AR.Q.THF.SQ.LQ...TL.SR----...PDMAT.E..SAWTN-----------.P.AKVR...K...A..R.KE DN51842_c2_g1_i1|m.33653
Pitx a      Sr TR.Q.THF.SQ.LQ...TL.AR----...PDMAT.E..SAWTN-----------.P.AKVR...K...A..R.KE DN122619_c1_g1_i1|m.71282
Pitx a      Hm QR.Q.THF.SQ.LT...TL.SR----...PDMAT.ED.SAYIL-----------.P.SKVR...K...A..R.KE DN26989_c0_g2_i1|m.9094
Pitx b      Ip ER.Q.TQF.SH.LR...SM.AR----...PDMAT.ERLSAWVQ-----------.S.TKVR...K...A..R.KE DN68900_c3_g1_i2|m.114481
Pitx b      Sr ER.Q.TQF.SH.LR...SM.AR----...PDMAT.ERLSAWIS-----------.S.TKVR...K...A..R.KE DN126115_c0_g4_i7|m.89715
Pitx b      Hm ER.Q.TQF.SH.LR...TM.AR----...PDMAS.ERLSAWIS-----------.S.SKVR...K...A..R.KR DN29252_c0_g1_i2|m.10492
Pitx        Nw PR.Q.THF.SQ.LQ...SM.TR----...PDMAT.E...AWTG-----------...SRVRV..K...A..R.KE 27707.0
Pitx        Xb QR.Q.THF.SQ.LQ...AT.AR----...PDMAT.E..SVWTN-----------...ARVRV..K...A..R.KE DN49217
PTX (105)   Nv QR.Q.THF.SF.LQQ..GT.GR----...PDMQM.E...LYTN-----------...ARVRV..K...A..R.KE DQ206235
```

Prop family

```
PROP1       Hs .R.H.T.FSPV.LEQ..SA.GR----.Q.PDIWA.ESL.RDTG-----------.S.AR.QV......A.QR.QE NP_006252 1
            Bf .R.H.T.F.QE.LAQ..EA.SK----SH.PDIYV.E.L.RATK-----------.N.AR.QV......A.YR.QE
CG32532     Dm .R.H.T.F.QE.LA...AA.AK----SH.PDIYC.E.L.RTTK-----------.N.AR.QV......A.YR.QE NP_608318
unc-42      Ce .R.H.T.F.QE.LQ..DAA.QK----SH.PDIYV.E.L.RITK-----------.N.AR.QV......A.HR.HE
Prop        Cg  R.H.T.F.QE.LQ...NA.AK----SH.PDIYC.E.L.RITK-----------.N.AR.QV......A.YR.
Prop        Ip .R.H.T.F.QE.LN...AA.AK----SH.PDIYV.E.L.RITK-----------.N.AR.QV......A.YR.QE DN60252_c0_g2_i3|m.55039
Prop        Sr .R.H.T.F.QD.LN...AA.SK----SH.PDIYV.E.L.RITK-----------.N.AR.QV......A.YR.QE DN118878_c0_g1_i4|m.56464
Prop        Hm .R.H.T.FNQD.LN...AA.SK----SH.PDIYV.E.L.RITK-----------.N.AR.QV......A.YR.HE DN23062_c0_g1_i1|m.7235
Prop        Xb .R.H.T.F.QE.LN...AA.AK----SH.PDIYV.E.L.RITK-----------.N.AR.QV......A.YR.QE DN88387
```

Repo family

```
            Bf KT.Y.T.FST..LE...KA.ER----AP.PDVFA.E.L.MR.N-----------.S.ARVQV......A..R.RE
repo        Dm K.KT.T.F.A..LE...RA.ER----AP.PDVFA.E.L.IK.N-----------.S.SRVQV......A..R.HE NP_477026
Repo        Cg  P.H.T.FSPF.LK.M.KA.RK----AP.PDVMT.E.L.RR.A-----------.N.SRVQ.......A..R.
Repo        Ip KS.Y.T.FSAE.LEQM.RA.QL----AP.PDVFT.E.L.ARVG-----------...ARVQV......AR.R..R DN64595_c0_g1_i1|m.76632
Repo        Sr KS.Y.T.F.SE.LDQM.RA.EC----AP.PDVFT.E.L.NRIK-----------...ARVQV......AR.R..R DN108331_c0_g1_i2|m.33300
Repo        Hm KS.Y.T.F.VF.LES..KA.ER----AA.PDVFA.E.L.RRIN-----------.S.AR.QV......A..R.DK DN41482_c1_g1_i2|m.28797
Repo        Nw KP.H.T.F.SW.LQ.M.RA.ER----AP.PDSPM.EQL.INIG-----------..DARVQV......A..R.QT 18856.0
Repo        Xb KT.Y.T.F.AW.LE.M.RA.ER----AP.PDVFQ.E.L.MKIS-----------.S.ARVQV......A..R.LK DN70414
REPO118     Nv KT.Y.T.FSQ..IE...RA.DK----AP.PDVFA.E.L.AK.G-----------...AR.QV......A..R.RE DQ206237
```

Rax family

```
RX/RAX      Hs HR.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.GKVN-----------.P.VRVQV......A..RRQE NP_038463
MGC15631    Hs HR.N.T.F.T..LHQ..RA.EA----SH.PDVYS.E.L.AKVH-----------.P.VRVQV......A..RRQE NP_116142
rax         Bf HR.N.T.F.TF.LH...RA.EK----SH.PDVYS.E.L.MKIN-----------.P.VRVQV......A..RRQE
```

```
Rx          Dm HR.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.MKVN-----------.P.VRVQV......A..RRQE NP_726006
Ceh-8       Ce QR.N.T.F.TF.LHA..AA.DK----TH.PDVYA.ETL.AKVQ-----------.P.VRVQV......A.FRRQE
Rax         Cg  R.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.MKID-----------.P.VRVQV......A..RR
Rax         Ip HR.N.T.F.TF.LH...RA.EK----SH.PDVYS.E.L.ARVN-----------.P.VRVQV......A..RRQE DN19265_c0_g1_i1|m.7660
Rax         Sr HR.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.SRVN-----------.P.VRVQV......A..RRQE DN126115_c0_g3_i1|m.89710
Rax         Hm HR.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.TRVQ-----------.P.VRVQV......A..RRQE DN20170_c0_g1_i1|m.5558
Rax         Nw HR.N.T.F.T..LH...RA.EK----SH.PDVYA.E.I.LKIN-----------.P.VRVQV......A..RRQE 29086.0 also 43623.0
Rx          Xb HR.N.T.F.T..LH...RA.EK----SH.PDVYS.E.L.MKVA-----------.P.VRVQV......A..RRQE DN67189 (end from DN56207)
RX (129)    Nv LR.N.T.F.TF.LH...RA.EK----SH.PDVYT.E.L.LKIS-----------.P.VRVQV......A..RRQE DQ206258
```

Shox (OG12) family

```
SHOX        Hs QR.S.TNF.LE.LN...RL.DE----TH.PDAFM.E.LSQR.G-----------.S.ARVQV......A.CR.QE NP_000442
SHOX20.61   Hs QR.S.TNF.LE.LN...RL.DE----TH.PDAFM.E.LSQR.G-----------.S.ARVQV......A.CR.QE NP_003021
shox        Bf QR.S.TNF.LE.LQ...RL.DE----TH.PDAFM.E.LSQR.G-----------.S.ARVQV......A.CR.QE
CG34367     Dm QR.S.TNF.LD.LN...RL.EE----TH.PDAFM.E.LSQR.G-----------.S.ARVQV......A.CR.HE
Shox        Cg  R.S.TNF.LE.LN...RL.DE----TH.PDAFM.E.LSQR.G-----------.S.ARVQV......A.CR.
Shox        Hm C..N.TQFSTF.IKR..KL.EE----TH.PDSYM.E.LGKR.N-----------.S.SR.QV......A.CR.YD DN27245_c0_g1_i1|m.9220
Shox        Sr Y..T.TQFSED.ICA...L.EE----SH.PDSSA.N.LVA.TG-----------..DSK.QV......A.AR..K DN120882_c0_g2_i7|m.63873
Shox        Xb QR.T.TNF.LE.LE....L.DE----TH.PDAFM.E.LSS.LG-----------.SEARVQV......A.C     DN14020
```

Uncx family

```
UNCX        Hs .R.T.TNFTGW.LE....A.NE----SH.PDVFM.EAL.LR.D-----------.V.SRVQV......A..R.KE
uncxA       Bf .R.T.TNFNSW.LE...RA.QE----SH.PDVFM.EAL.LR.D-----------.V.SRVQV......A..R.KE
uncxB       Bf QR.S.ANYSQW.LE...KA.ET----TQ.PDIFM.EAL.LR.D-----------.I.ARVQV......A.LRRQL
uncxC       Bf .M.V.TNFTSW.LEA..RA.ET----TH.PDIFM.EAL.LR.D-----------.M.SRVQV......A..RKQE
unc-4       Dm .R.S.TNFNSW.LE...RA.SA----SH.PDIFM.EAL.MR.D-----------.K.SRVAV......A.VR.RE NP_573242
OdsH        Dm KR...TNFNSW.LR...RV.QG----SH.PDIFM.EAL.TK.D-----------.M.GR.AV......A..R.QE NP_523389
unc-4       Ce .R.T.TNFSGW.LE...SA.EA----SH.PDVFM.EAL.MR.D-----------.L.SRVQV......A..R.RE
Uncx        Cg. R.T.TNF.GW.LE...RA.QD----SH.PDVFM.EAL.LR.D-----------.V.SRVQV......A..R.
Uncx        Ip .R.T.TNF.NW.LEQ..LA.HR----TH.PDVFV.ETL.LQ.N-----------.M.SRVQV......A..R    DN58162_c0_g1_i2|m.48062
Uncx        Sr .R.T.TNF.NW.LEQ..MA.HR----TH.PDVFV.ETL.MQ.N-----------.M.SRVQV......A..R.ME DN113036_c0_g1_i1|m.41100
Uncx        Hm .R.T.TNF.SL.LE...RA.AA----SH.PDVFM.ESLGQR.S-----------.L.SRVQV......A..R.IE DN31526_c0_g1_i1|m.12245
Uncx        Xb .R.T.TNFNSW.LE....A.LV----SH.PDIFM.EAL.MR.D-----------.I.SRVQV......A..R.KE DN12917
UNC4 059    Nv .M.V.TNFSPW.LE...HA.ET----TH.PDVFM.EAL.MR.D-----------...ARVQV......A..R.RE DQ206229
```

Vsx/Chx10 family

```
CHX10       Hs KR.H.TVF.AH.LE....A.SE----AH.PDVYA.EML.VKTE-----------.P.DR.QV......A..R.RE NP_878314
VSX1        Hs KR.H.TIF.AH.LE....A.SE----AH.PDVYA.EML.VKTE-----------.P.DR.QV......A..R.RE NP_055403
vsx         Bf KR.H.TIF.S..LE....A.NE----AH.PDVYA.EML.MKTD-----------.P.DR.QV......A..R.KE
4136/vsx1   Dm .RH..TIF.SS.LE....A.KE----AH.PDVSA.ELLSMKTG-----------.A.DR.QV.Y....A..R.TE NP_572232
15782/vsx2  Dm .RHS.TIF.S..LEK..EA.KE----AH.PDVYA.EMLSLKTE-----------.P.DR.QV......A..R.TE NP_572230
```

```
ceh-10      Ce KR.H.TIF.Q..ID....A.QD----SH.PDIYA.EVL.GKTE-----------.Q.DR.QV......A..R.TE
dsc-1       Ce .R.F.TNF.EL.STF..DS.KE----SH.PDHKA.KYM.DF.K-----------IP.DR.TV......A..RRKE
Vsx1        Cg  R.H.TIF.S..VE....T.KE----AH.PDLYA.EVL.LKID-----------.P.DR.QV......A..R.
Vsx2        Cg  R.H.TIF.S..LE....A.KD----AH.PDVYA.EVL.LKTS-----------.P.DR.QV......A..R.
Vsx a       Ip .R.H.TIF.S..IE....A.KE----AH.PDVYA.ELLSIKTD-----------.P.DR.QV......A..R.KE DN31321_c0_g1_i1|m.13615
Vsx b       Ip .R.H.TIF.NS.VE....A.EE----AH.PDVYQ.EILSLKTD-----------.P.DR.QV......A..R.KE DN66046_c0_g1_i4|m.86993
Vsx a       Sr .R.H.TIF.S..IE....A.GD----AH.PDVYA.EMLSLKTD-----------.P.DR.QV......A..R.KE DN98097_c0_g1_i1|m.23524
Vsx b       Sr .R.H.TIFSVS.VE...RA.ED----AH.PDVVQ.ETLSNKTE----------.P.DR.QV......A..R.KE DN111202_c0_g2_i1|m.37759
Vsx a       Hm .R.H.TIF.SL.ID....A.SE----AH.PDVYA.EMLSLKTD-----------.P.DR.QV......A..R.KE DN7688_c0_g1_i1|m.2075
Vsx b       Hm .R.H.TIF.NE.VD....I.QT----AH.PDVNA.EMLSLKTE-----------.P.DR.QV......A..R.KE DN80881_c0_g1_i1|m.43410
Vsx         Nw KR.H.TIF.AL.LE...AA.GD----AH.PDVYA.ELLSIKTD-----------.P.DR.QV......A..R.KE 21928.0
Vsx         Xb RR.H.TIF.S..LEQ...A.TE----AH.PDVYA.EMLSLKTD-----------.P.DR.QV......A..R.KE DN86983
CEH10085    Nv KR.H.TIF.N..LE....A.KE----SH.PDVYA.ENLSLKID-----------.P.DR.QV......A..R.KE DQ206232
```

PRD CLASS / Family membership ambiguous

```
            Ip PQILPLHMLTDAEKT..RS.MR----DHNPADYMYEIMAEEIG-----------WNLN.VKQ.YDRKKAT.RSNE DN73593_c0_g2_i3
            Ip PR.S.T.F.AF.LHQ..SI.LS----N..PDVA..DQL.AM.G-----------...SR.QV......A.RR.AE DN71583_c0_g1_i1|m.152092
            Sr CR.S.T...PW.LHQ..SS.ED----.P.PDVGQ.DYL.GVMG-----------V..SR.QV......A.RR.LA DN112801_c0_g1_i2|m.40593
            Hm KP.V.TFF.SD.LRN..KA.SE----NA.PDVKM.A..CSR.K-----------.R.S.VL.....K.ARN..LK DN35675_c0_g1_i1|m.16721
HD1         Hm                      K----DQFPSKESKKTLSKK.N-----------.DYSSVDV....K.ARTNSKI DN34694_c0_g1_i1|m.15382
HD2         Hm GRKA.TSITPF.QDQ..AA.RL----.PFPLWSDKDY..SE.D-----------.D.SVVQV......ARS.RKL DN34694_c0_g1_i1|m.15382
            Hm QR.N.T.F.NY.LNT..CA.RK----TH.PDVFT.E.L.KKVE-----------...ARVQV......A..R.GE DN36971_c0_g1_i2|m.18384
            Nw K.REKKVFDPM.IAL..HNYME----IR.PDLQ..K.L.AVVG-----------ID.KR.QV......ARTRRKH 54213.0
            Nw K..VKKVFEPG.IAL..HNYIL----.Q.PSPQG.KALSNEVG-----------ID.KRIQV......ARTR.KH 47651.0
            Xb  R.N.MSFQPE.RAI..EAYRV----.K.PDILH.AWISQMTN-----------IPDLR.Q......A.SR.RE DN122491
            Xb QR.N.T.F.PQ.LE...QL.SK----TH.PDIFV.EDL.MRIN-----------.S.ARVQV... DN122218
```

LIM CLASS

Apterous / Lhx2/9 family

```
LHX2        Hs T..M.TSFKHH.LRTMKSY.AI----.HNPDAKDLKQL.QKTG-----------..K.VLQV....A.A.FRRNL NP_004780
LHX9        Hs T..M.TSFKHH.LRTMKSY.AI----.HNPDAKDLKQL.QKTG-----------..K.VLQV....A.A.FRRNL NP_001014434
Lhx2/9a     Bf T..I.TSFKHH.LRT.KSY.AI----.HNPDSKDLQQL.QKTG-----------..K.VLQV....A.A.HRRNA
Lhx2/9b     Bf Q..M.TSFKHH.LRAMKSY.AL----.HNPDAKDLKQL.QKTG-----------..K.VLQV....A.A.YRRNL
ap          Dm T..M.TSFKHH.LRTMKSY.AI----.HNPDAKDLKQLSQKTG-----------.PK.VLQV....A.A..RRMM NP_724428
ttx-3       Ce S..M.TSFKHH.LRAMKTY.AL----.HNPDAKDLKQL.AKTN-----------..K.VLQV....A.A.YRREL
Lhx2_9      Cg  ..V.TSFKHH.LRTMKSY.AL----.HNPDAKDLKQL.QKTG-----------.SK.VLQV....A.A.YRR
Lhx2/9 a    Ip A..V.TSFKNH.LRV.KQY.QL----.HNPDAKELKQLSQKTG-----------.SK.VLQV....A.A.YRRMM DN67846_c0_g1_i1 .2
Lhx2/9 a    Sr T..V.TSFKNH.LRV.KQY.QL----.HNPDAKELKQLSQKTG-----------.SK.VLQV....A.A.YRRLM DN125451_c0_g1_i1|m.85840
Lhx2/9 a    Hm N..V.TSFKNH.LRV.KQY.QL----.HNPDAKELKQLSQKTG-----------.SK.VLQV....A.A.YRRNV DN36226_c0_g1_i1|m.17408
Lhx2/9 b    Ip S..M.TSFKHH.LQAMKVY.NA----.HNPDAKDLKQL.TKTG-----------..K.VLQV....A.A.FRRNL DN67846_c0_g4_i1|m.102863 .1
Lhx2/9 b    Sr S..M.TSFKHH.LQAMKQY.NA----.HNPDAKDLKSL.QKTG-----------..K.VLQV....A.A.FRRNL DN128068_c1_g1_i1|m.102330
```

```
Lhx2/9 b    Hm S..M.TSFKNN.LVQMKQY.TM----.HNPDAKDLKQLSHKTG-----------..K.VLQV....S.A.YRRNM DN9855_c0_g2_i1|m.2527
Lhx2/9 a    Nw T..V.TSFKHH.LRTMKSY.NL----.HNPDAKDLKQLSQKTC-----------.SK.VLQV....A.A.YRRLM 51594.0
Lhx2/9 b    Nw T..V.TSFKNP.LRVLKQY.QL----.HNPDAKELKQLSQKTS-----------.SK.VLQV....A.A.YRRQV 54455.0
Lhx2/9      Xb Q..M.TSFKHH.LRAMKSY.SI----.HNPDAKDLKQL.QKTG-----------..K.VLQV....A.A.FRRNM DN63074
```

Islet family

```
ISL1        Hs TT.V.TVLNEK.LHT.RTCYAA----.PRPDALMKEQLVEMTG-----------.SP.V.RV....K.C.D..RS NP_002193
ISL2        Hs TT.V.TVLNEK.LHT.RTCYAA----.PRPDALMKEQLVEMTS-----------.SP.V.RV....K.C.D..KS NP_665804
isl         Bf PT.V.TVLNEK.LHT.RTCYAA----.PRPDALMKEQLVEMTG-----------.SP.V.RV....K.C.D..KS
tup         Dm PT.V.TVLNEK.LHT.RTCYNA----.PRPDALMKEQLVEMTS-----------.SP.V.RV....K.C.D..KT NP_476774
lim-7       Ce AT.V.TVLNEN.LKI.RDCYSI----.SRPDATLKERLVEMTG-----------.SA.V.RV....K.C.D..RQ
Islet       Cg  T.V.TVLNEK.LHT.RTCYNA----.PRPDALMKEQLVEMTG-----------.SP.V.RV....K.C.D..
Islet       Ip PT.I.TVLNEK.LHT.RTCYNA----.PRPDALM.EQLVEMTG-----------.SP.V.RV....K.C.D..RS DN56056_c0_g1_i1|m.42738
Islet       Sr PT.I.TVLNEK.LHT.RTCYNA----.PRPDALM.EQLVEMTG-----------.SP.V.RV....K.C.D..RS DN125302_c3_g2_i1|m.85233
Islet a     Hm PT.V.TVLNEK.LHT.RTCYNA----.PRPDALMKEQLTEMTA-----------.SP.V.RV....K.C.D..RS DN27964_c0_g1_i1|m.9540
Islet b     Hm SG.V.TVLNEK.LST.RAYYHA----.PRPDAPM.EQLVEMTS-----------.SH.V.RV....K.C.D..RQ DN14479_c0_g1_i1|m.3858
Islet a     Nw PT.V.TVLNEK.LST.RTCYAA----.ARPDALM.EQLSDMTG-----------.SG.V.RV....K.C.D..RS 45303.0
Islet b     Nw PT.V.TVLNEK.LQT.RQCYAA----.SRPDALM.EQLSDMTS-----------.SG.V.RV....K.C.D..RS 22501.1
Islet       Xb TT.V.TVLNEK.LHT.RTCYAA----NPRPDALMKEQLVEMTA-----------.SP.V.RV....K.C.D..KS DN91806
ISL (028)   Nv PT.V.TVLNEK.LHT.RTCYNA----.PRPDAMMKEQLVEMTG-----------.SP.V.RV....K.C.D..KA DQ206225
```

Lhx 1/5 family

```
LIM1/LHX1   Hs .RGP.T.IKAK.LET.KAA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK NP_005559
LHX5        Hs .RGP.T.IKAK.LET.KAA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK NP_071758
Lhx1/5      Bf .RGP.T.IKAK.LET.KAA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK
lim1        Dm .RGP.T.IKAK.LEV.KTA.NQ----TPKP..HI.EQL.KETG-----------.PM.V.QV....K.S.ERRMK NP_572505
lin-11      Ce .RGP.T.IKAK.LET.KNA.AA----TPKPT.HI.EQL.AETG-----------.NM.V.QV......S.ERRMK
mec-3       Ce .RGP.T.IKQN.LDV.NEM.SN----TPKPSKHA.AKL.LETG-----------.SM.V.QV......S.ERRLK
Lhx1/5      Cg  RGP.T.IKAK.LET.KAA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERR
Lhx1/5      Ip .RGP.T.IKAK.LET.KSA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK DN68178_c0_g1_i3|m.106476
Lhx1/5      Sr .RGP.T.IKAK.LET.KSA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK DN118733_c0_g2_i1|m.55910
Lhx1/5      Hm .RGP.T.IKAK.LET.KTA.MN----TPKP..HI.EQL.METG-----------.NM.V.QV......S.ERRMK DN30383_c0_g1_i1|m.11355
Lhx1/5      Nw RRGP.T.IKAK.LDT.KAA.DA----TPKP..HI.EQL.QDTG-----------.NM.V.QV......S.ERRMT 61503.0
Lhx1/5      Xb RRGP.T.IKAK.LET.KAA.AA----TPKP..HI.EQL.QETG-----------.NM.V.QV......S.ERRMK DN95395
LHX1/5 127  Nv .RGP.T.IKAK.LEA.KST.AA----TPKPS.NI.EKL.QETG-----------.NM.V.QV......S.ERRLK DQ206345
```

Lhx 3/4 family

```
LHX3        Hs A..P.T.I.AK.LET.KSAYNT----SPKPA.HV.EQLSSETG-----------.DM.VVQV......A.E.RLK NP_055379
LHX4        Hs A..P.T.I.AK.LET.KNAYKN----SPKPA.HV.EQLSSETG-----------.DM.VVQV......A.E.RLK NP_071758
Lhx3/4      Bf Q..P.T.I.AK.LET.KQAYQN----SPKPA.HV.EQLSQETG-----------.DM.VVQV......A.E.RLK
Lim3        Dm N..P.T.I.AK.LET.KTAYNN----SPKPA.HV.EQLSQDTG-----------.DM.VVQV......A.E.RLK NP_572505
```

```
Ceh-14      Ce N..P.T.ISAKSLET.KQAYQT----SSKPA.HV.EQL.SETG-----------.DM.VVQV......A.E.RLK
Lhx3/4      Cg  ..P.T.I.AK.LEA.KRAYNE----SPKPA.HV.EQLSAETG-----------.DM.VVQV......A.E.R
Lhx3/4      Ip N..P.T.I.AK.LEI.KNAYKS----SSKPA.HI.EQLSKDTG-----------.DM.VVQV......A.E.RLK DN49678_c0_g2_i1|m.29979
Lhx3/4      Sr N..P.T.I.AK.LEV.KSAYKN----SSKPA.HI.EQLSKDTG-----------.DM.VVQV......A.E.RLK DN124757_c1_g2_i2|m.82144
Lhx3/4      Hm N..P.T.INNK.LDV.REAYSK----SAKPA.HI.EDLSKKTD-----------.DM.VVQV......A.E.RLK DN28959_c0_g1_i1|m.10298
Lhx3/4      Nw Q..P.T.I.AK.LET.KTAYSN----SPKPA.HI.EQLSHDTG-----------.DM.VVQV......A.E.RLK 38224.0
Lhx3/4      Xb S..P.T.I.AK.LET.KNAYSN----SPKPA.HV.EQLSNETG-----------.DM.VVQV......A.E.RLK DN11429
055         Nv S..A.TFISND.LAF.KVAYAS----SPKT.L.D.ER..KETG-----------.DM.VVQV......A.D.RLS DQ206345 (newly assigned)
```

Lhx 6/8 family

```
LHX6        Hs A..A.TSF.AE.LQVMQAQ.AQ----DNNPDAQTLQKL.DMTG-----------.SR.V.QV....C.ARH..HT NP_055183
LHX8        Hs A..A.TSF.AD.LQVMQAQ.AQ----DNNPDAQTLQKL.ERTG-----------.SR.V.QV....C.ARH..HV NP_001001933
Lhx6/8      Bf P..V.T.F.EE.LRV.QAN.NI----DSNPDGQDLER..QITG-----------.SK.VTQV....S.ARQ..YG
awh         Dm T..V.T.F.EE.LQV.QAN.QI----DSNPDGQDLER..SVTG-----------.SK.VTQV....S.ARQ..HI NP_523907
lim-4       Ce T..V.T.FAED.LSV.QTY.NR----DSNPDGADLEK..SMTG-----------.SK.VTQV....S.ARQ..WH
Awh1        Cg  ..V.SSF.EE.LQI.QAN.RI----ESNPDSQELNR..ITAG-----------VSR.VAQV....A.ARQ..
Awh2        Cg  ..V.T.F.ED.LQV.QAN.QL----DSNPDGQDLER..QITG-----------.SK.VTQV....S.ARQ..
Lhx6/8      Ip S..S.T.FSED.LKI.QAN.NL----DSNPDGQDLER..QQTG-----------.SK.V.QV....A.SRQ..HM DN27657_c1_g1_i1|m.11322
Lhx6/8      Sr S..S.T.FSED.LKV.QAN.NM----DSNPDGQDLER..SQTG-----------.SK.V.QV....A.SRQ..HL DN123001_c0_g1_i1|m.72616
Lhx6/8      Hm T..S.T.FSEE.LKV.QAN.SL----DSNPDGQDLER..QMTG-----------.SK.V.QV....A.SRQ..HL DN1193_c0_g1_i1|m.357
Lhx6/8      Nw KR.H.TMFHRY.LEQ.KES.KT----.SNPSSIDVEYI.KSVN-----------S.K.IIQV....A.S.QQRRL 51594.0/1
Lhx6/8      Xb A..V.T.F.EE.LQI.QAN.NL----DSNPDGQDLER..QLTG-----------.SK.VTQV....S.ARQ..NN DN81776
LHX6/8 044 Nv A..V.T.F.ED.LQI.QAN.NI----DSNPDGQDLER..QLTG-----------.SK.VTQV....S.ARQ..YG DQ206316
```

Lmx family

```
LMX1A       Hs P..P.TIL.TQ.RRAFKAS.EV----SSKPC.KV.ETL.AETG-----------.SV.VVQV....Q.A.M..LA
LMX1B       Hs P..P.TIL.TQ.RRAFKAS.EV----SSKPC.KV.ETL.AETG-----------.SV.VVQV....Q.A.M..LA NP_002307
            Bf P..P.TIL.TQ.RREFKAL.EV----SPKPC.KV.ETL.AETG-----------.SV.VVQV....Q.A.M..LA
CG32105     Dm P..P.TIL.SQ.RKQFKAS.DQ----SPKPC.KV.EAL.KDTG-----------.SV.VVQV....Q.A.M..IQ NP_729801
CG4328      Dm P..P.TILNTQ.RRAFKAS.EV----SPKPC.KV.ENL.KDTG-----------.SL.IVQV....Q.A.V..IQ NP_648567
lim-6       Ce P..P.TILNAQ.RRQFKTA.ER----SSKPS.KV.EQL.NETG-----------.SV.VVQV....Q.A.I..L.
Lmx1        Cg  ..P.TIL.TS.RRKFKQA.EA----.PKPC.KV.EQL.AETG-----------.TI.VVQV....Q.A.V..
Lmx2        Cg  ..P.TIL.TS.RRKFKSA.EL----.PKPC.KV.EQL.AETG-----------.SV.VVQV....Q.A.V..
Lmx         Ip PC.K.VIL.TQ.RRAFKAA.DL----SSKPC.KV.EQL.RETG-----------.SV.VVQV....E.A.V..LA DN72227_c0_g9_i2|m.162480
Lmx         Hm PC.K.VIL.TQ.RRAFKAA.DV----SSKPC.KV.EQL.RDTG-----------.SV.VVQV....E.A.V..LA DN41603_c0_g1_i1|m.29404
Lmx         Sr PC.K.VIL.TQ.RRAFKAA.DL----SSKPC.KV.EQL.RDTG-----------.SV.VVQV....E.A.V..IG DN127948_c0_g1_i1|m.101565
Lmx a       Nw  RSK.VIL.TT.RRIFRTA.DV----SQKPS.KVKERL.DETG-----------.PV.IVQV....E.A.G..LA 50222.0
Lmx b       Nw VG.K.VIL.HN.RRVLRAA.IV----SQKPS.KVKEDLGGATQ-----------.PV.IVQV....E.A.V..LA 14051.0
Lmx c       Nw  G.K.VLL.NP.RRIFRSA.TI----SDKPC.RVKEGL.MMTQ-----------.PI.IVQV....E.A.MR.LS 7973.0
Lmx d       Nw HNNK.VNLNNM.RHILLAAYET----SHKPT.SMKETL.EQTG-----------.PM.IVQV....E.A.C..MA 48370.0
Lmx e       Nw GS.K.VVL.PE.RKIFQQA.DM----SPKPS.KV.EQL.SECT-----------VSM.FIQV....Q.A.M..LA 52392.0
```

```
Lmx f      Nw GS.K.VVLSPE.RKIFQLA.DA----SAKPS.KV.EDL.NKCS-----------VSM.FIQV....Q.A.M..LA 49053.0
Lmx        Xb PC.K.IIL.TQ.RRIFKAA.EV----SSKPC.KV.ESL.AETG-----------.SV.VVQV....Q.A.V..LA DN87622
```

LIM CLASS / family membership ambiguous

```
           Ip NNNT.KPLKREIAEH.RAY.DK----.SKP.KPE.E.L.NQLG-----------ITE.V.QI......S.DRRVA DN68501_c2_g1_i1|m.110831
           Sr KRGPGNHLKGP.ADY.RSYYLT----.QKPSKEE.DA..AEIG-----------VEAKT.AT......S.E.IRR DN123163_c0_g1_i1|m.73968
           Hm KA.K.TVFTQQ.LAV.QSR.AI----KQILS.QEYANL.DEIG-----------VKIPIVRV....Y.QRS.TNS DN8120_c0_g2_i1|m.2184
```

POU CLASS

Pou2 family

```
POU2F1     Hs .RKK.TSIETNIRVA...S.LE----.QKP.SEEITM..DQ.N-----------MEKEV.RV..C...Q.E.RIN NP_002688
POU2F2     Hs .RKK.TSIETNVRFA...S.LA----.QKP.SEEILL..EQ.H-----------MEKEV.RV..C...Q.E.RIN NP_002689
POU2F3     Hs KRKK.TSIETNIR.T...R.QD----.PKPSSEEISM..EQ.S-----------MEKEVVRV..C...Q.E.RIN NP_055167
           Bf .RKK.TSIETNVRVA...A.IQ----.PKP.SEEIGI..EQ.G-----------MEKEVVRV..C...Q.E.RIN
Nub        Dm .RKK.TSIETTIRGA...A.LA----.QKP.SEEITQL.DR.S-----------MEKEVVRV..C...Q.E.RIN NP_476659
pdm2       Dm .RKK.TSIETTVRTT...A.LM----.CKP.SEEISQLSER.N-----------MDKEV.RV..C...Q.E.RIN NP_723763
ceh-18     Ce .R.K.TNLDMN.RNA.DTF.AL----.PRPDHDKMTD..NS.E-----------.DRDVVRV..C...Q.MRRVD
Pou2       Cg  RKK.TSIETNIRVA...S.QQ----.PKPSSEEITSV.DQ.S-----------MEKEVVRV..C...Q.Q.R
Pou2 a     Ip KRKK.SCIDSNCRET..QL.DQ----.NHP.PDEIVKYSE..N-----------MEPET.R...S...Q.Q.RLA DN55409_c0_g2_i1|m.41119
Pou2 b     Ip KRKR.CVIDAHTKKT..QS.RD----KPHPSPDETNAL.E..G-----------.DRETVR...T...YRERRSE DN18113_c0_g1_i1|m.7284
Pou2       Sr .RKR.TCIDVKTKKA..QFYND----KPHP.HEECCA..E..N-----------QDRNT.R...G...QRD.RLM DN116892_c2_g1_i1|m.50278
Pou2       Hm .RKK.ASIETKSK.A...S.SL----.SKP.AEQINLL.DQ.I-----------MERETVRV..C...Q.E.RVP DN43708_c0_g1_i1|m.40060
Pou2       Xb  RKK.TSIETNVRVA...A.LS----.PKP.SEEIQML.ES.C-----------MEKEVVRV..C...Q.E.R DN87006
```

Pou3 family

```
POU3F1     Hs KRKK.TSIEVGVKGA..SH.LK----CPKPSAHEITGL.DS.Q-----------.EKEVVRV..C...Q.E.RMT NP_002690
POU3F2     Hs KRKK.TSIEVSVKGA..SH.LK----CPKPSAQEITSL.DS.Q-----------.EKEVVRV..C...Q.E.RMT NP_005595
POU3F3     Hs KRKK.TSIEVSVKGA..SH.LK----CPKPSAQEITNL.DS.Q-----------.EKEVVRV..C...Q.E.RMT NP_006227
POU3F4     Hs KRKK.TSIEVSVKGV..TH.LK----CPKPAAQEISSL.DS.Q-----------.EKEVVRV..C...Q.E.RMT NP_000298
POU3       Bf KRKK.TSIEVTVKGA..SH.LK----QPKPSAQEIAQL.DS.Q-----------.EKEVVRV..C...Q.E.RMT
POU3L      Bf KRKK.TSIEVSVKGA..TF.YK----QPKPSAIEISQLSEG.N-----------.DKEVVRV..C...Q.ERRMS
vvl        Dm KRKK.TSIEVSVKGA..QH..K----QPKPSAQEITSL.DS.Q-----------.EKEVVRV..C...Q.E.RMT NP_523948
ceh-6      Ce KRKK.TSIEVNVKSR..FH.QS----.QKPNAQEITQV.ME.Q-----------.EKEVVRV..C...Q.E.RIA
Pou3       Cg  RKK.TSIEVTVKGA..NH.MK----QPKPAAQEISQL.EQ.Q-----------.EKEVVRV..C...Q.E.R
Pou3       Hm KRKK.TSIEVGIKGA..NH..L----QPKPAAPEITGL.DS.Q-----------.EKEVVRV..C...Q.E.RMN DN20967_c0_g1_i1
Pou3       Sr KRKK.TSIEVSIKNA..HN..I----QPKPSAHEISAL.NS.Q-----------.EKEVVRV..C...Q.E.RMT DN123138_c1_g2_i1|m.73697
Pou3       Ip KRKK.TSIEVTIKGA..QH..L----H                                                 DN88733_c0_g1_i1
Pou3       Nw KRKK.TSIEMSVKGA..NH.MK----QAKPCAPEIQAL.DS.Q-----------.DKEVVRV..C...Q.E.RMT 27996.0
Pou3       Xb KRKK.TSIEVTVKGV..SH.QK----QPKPSAQDITNL.ES.Q-----------.EKEVVRV..C...Q. DN83679
POU3a125   Nv KRKK.TSIEVAVKGA..NH.CK----SPKP.AQEISAL.EN.G-----------.DKEVVRV..C...Q.E.RMT DQ206344
```

```
POU3b075   Nv KRKR.TIIEKNVKGV..NH.EK----MPRPSTSDISSL.ES.G-----------.DREVVRV..C...Q.ERRVS DQ206327
```

Pou4 family

```
POU4F1     Hs K..K.TSIAAPEKRS..AY.AV----QPRPSSEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RMK NP_006228
POU4F2     Hs K..K.TSIAAPEKRS..AY.AI----QPRPSSEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RMK NP_004566
POU4F3     Hs ...K.TSIAAPEKRS..AY.AI----QPRPSSEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RMK NP_002691
POU4       Bf K..K.TSIAAPEKRS..AY.AV----QPRPSSEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RMK
acj6       Dm K..K.TSIAAPEKRS..AY.AV----QPRPSGEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RIV NP_727845
unc-86     Ce K..K.TSIAAPEKRE..QF.KQ----QPRPSGERIAS..DR.D-----------.KKNVVRV..C.Q.Q.Q.RDF
Pou4       Cg  ..K.TSIAAPEKRS..AY.AV----QPRPSGEKIAQ..EK.D-----------.KKNVVRV..C.Q.Q.Q.R
Pou4       Ip K..K.TSIAAPEKRS..AY.TV----QPRPSGEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RLK DN75504_c0_g2_i7|m.225572
Pou4       Sr K..K.TSIAAPEKRS..AY.TV----QPRPSGEKISA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RLK DN125354_c0_g1_i1|m.85350
Pou4       Hm K..K.TSIAAPEKRS..AY.AV----QPRPSGEKIAA..EK.D-----------.KKNVVRV..C.Q.Q.Q.RLK DN7261_c0_g1_i1|m.1968
Pou4       Nw R..K.TSIASPEKRS..AY.AV----QPRPSSEKIGQ..DK.D-----------.KKNVVRV..C.Q.Q.Q.RLK 9311.1
POU4(124)  Nv K..K.TSIGAAEKRS..AY.AM----.PRPSSDKIAS..EK.D-----------.SKNVVRV..C.Q.Q.K.RMK DQ206292
```

Pou6 family

```
POU6F1     Hs KRKR.TSF.PQAIEA.NAY.EK----.PLP.GQEIT...KE.N-----------YDREVVRV..C...QTL.NTS NP_002693
POU6F2     Hs KRKR.TSF.PQALEI.NAH.EK----.THPSGQEMT...EK.N-----------YDREVVRV..C.K.QAL.NTI NP_009183
           Bf KRKR.TSF.PQALEV.NAH.EK----.THPSGAEMTAL.EK.N-----------YDREV.RV..C.K.QAL.NTI
pdm3
/CG11641   Dm KRKR.TSF.PQALEL.NAH.ER----.THPSGTEITGL..Q.G-----------YEREV.R...C.K.QAL.NTV NP_610377
Pou6       Cg  RKR.TSF.PQALEV.NNF.ER----.THPSGAEMTELSDK.N-----------YDREV.RV..C.K.QAL.N
Pou6       Hm SRKK.TQFSAET.IF.NDY.EK----ITHPRGFQINE..NT.Q-----------.DREVVRV..C.K.QLM.SKD DN37913_c0_g2_i2|m.19890
Pou6       Nw KRKR.TSFPPAAIDI.NSY.LK----.SHPSGHEFTQLSRL.E-----------HDREV.RV..C.K.QAMRHYA 52441.0
Pou6       Xb  RKR.TSF.PAALEI.NVH.EK----.THPSGSEMTE..SK.N-----------YDREVVRV..C.K.QAL. DN93613
NVHD084    Nv KRKR.TSFSNEA.RL.ISH.EQ----.PKPSSSEIAQ..SK.G-----------.EPVTVRV..C..KQML.RMA DQ206280
```

Pou family amb

```
           Nw .RKR.TSLNPDVREK.EAH.EV----.SRPGANQIFV.SEH.S-----------.DREVVRV..C.RRQ.I.RVN 53834.0
                                                                                          46372 added aa in end
```

HNF CLASS

Hmbox family

```
HMBOX1     Hs    .RGS.F.WRKECLAVM.SY.NE.Q.PDEAK.E...N.CN-----AVIQKPGKKLSDLERV.SLKVYN..A...KEI.RRA
HMBOX1A    Bf    .RGH.FNWPEACITIM..Y.EE.Q.PDEKK.E..TN.CN-----SVIQKPGVELPAQMLVNSARVYN..A...KDV.RRH
HMBOX1B    Bf    .RGSKFHWPPSAVMIV..Y.QQ.P.P.EGE.D...N.CN-----SVLQKPGEDLPSYKLVSPVRVQT..I...REA.IKQ
Hmbx-1     Ce    QR.E.YVFRPILIRM..SF.TQTPFPDLP..V...N.CN---HVLKMDKKGVGLMPKEVVSPQVVSN..A.K.KELRRR
Hmbox      Ip    K..D.FVWKENCVAL...YYEA.P.PDDT..E...A.CN------QLNGNSGSGQEERVV.AVKVYN..A...KEV.RK. DN65023_c0_g1_i1|m.79542
Hmbox      Sr    K..D.FVWKESCVCL...FY.V.P.PDDAK.E...Q.CN------QLIDSTGQGHDEKLV.AVKVYN..A...KEI.RKN DN121592_c0_g1_i4|m.66335
```

```
Hmbox       Hm    K..D.FVWKDSCVSI..QFYQE.S.PDDLK.E...Q.CN-----NALELQGQFLNDEKLV.SVKVYN..A...KEI.RKQ DN37226_c0_g1_i1|m.18864
Hmbox       Nw       E.FVWREECMPV.DKF.NENQYPDDRKREDIASACN-GVLEAPGVHIPGERLVTPAKVPAKVYN..V...NSY.IKM 48953.0
Hmbox       Xb        RS.INWDPSAVNI.EHH.NTK--------..FPNAFE.ET..ANINGACHSKGRKV.AHH.QT..N...N.RTGVS DN90583 (Hnf?)
```

SINE CLASS

Six 1/2 family

```
SIX1        Hs GEETSYCFKEKSRGV.REWYAH----.P.PSP.EKR.L.E.TG-----------..TT.VSN..K...QRDRAAE NP_005973
SIX2        Hs GEETSYCFKEKSRSV.REWYAH----.P.PSP.EKR.LTE.TG-----------..TT.VSN..K...QRDRAAE NP_058628
SIX1/2      Bf GEETSYCFKEKSRGV.REWYAH----.P.PSP.EKR.L.E.TG-----------..TT.VSN..K...QRDRAAE
so          Dm GEETSYCFKEKSRSV.RDWYSH----.P.PSP.EKRDL.E.TG-----------..TT.VSN..K...QRDRAAE NP_476733
ceh-33      Ce GEETSYCFRDKSRVL.RDWYCR----.S.PSP.EKR.L.EKTH-----------..VT.VSN..K...QRDRAGV
ceh-34      Ce GEETNYCFKSKSRNV.RDAYKK----CQ.PSVEDKRRL.QQTE-----------.SII.VSN..K.K.QRERAAG
Six1_2      Cg  EETSYCFKEKSRTI.REWYQH----.P.PSP.EKREL.EATG-----------..TT.VSN..K...QRDRA
Six1/2      Ip GEETSYCFKEKSRNM.RDYYGH----.P.PSP.EKRDL.EATD-----------.SIT.VSN..K...QRDRALE DN45560_c0_g1_i1|m.24666
Six1/2      Sr GEETSYCFKEKSRNM.REYYQH----.P.PSP.EKRDL.ENTD-----------.SVT.VSN..K...QRDRALE DN87552_c0_g1_i1|m.17761
Six1/2      Hm GEETSYCFKEKSRNV.RDYYLH----.P.PSP.EKRDL.EATD-----------..II.VSN..K...QRDRALE DN32866_c0_g1_i1|m.13539
Six1/2      Xb GEETSYCFKEKSRTV.REWYSH----.P.PSP.EKRELSDGTG-----------..VT.VSN..K...QRDRAAE DN63639
SIX1/2073   Nv GEETSYCFKEKSRNI.REWYSH----.P.PSP.EKR.L.EGTG-----------..TT.VSN..K...QRDRAAE DQ206326
```

Six 3/6 family

```
SIX3        Hs GEQKTHCFKERTRSL.REWYLQ----DP.PNPSKKR.L.Q.TG-----------..PT.VGN..K...QRDRAAA NP_005404
SIX6        Hs GEQKTHCFKERTRNL.REWYLQ----DP.PNPSKKR.L.Q.TG-----------..PT.VGN..K...QRDRAAA NP_031400
Six3/6      Bf GEQKTHCFKERTRSL.REWYLQ----DP.PNPQKKR.L.Q.TG-----------..PT.VGN..K...QRDRAAA
optix       Dm GEQKTHCFKERTRSL.REWYLQ----DP.PNPTKKR.L.K.TG-----------.NPT.VGN..K...QRDRAAA NP_524695
ceh-32      Ce GEQKTHCFKERTRSL.REWYLK----DP.PNPPKKK.L.N.TG-----------..QM.VGN..K...QRDRAAA
Six3_6      Cg  EQKTHCFKERTRNL.REWYLQ----DP.PNPTKKR.L.S.TG-----------..PT.VGN..K...QRDRA
Six3/6      Ip GEQKTHCFKERTRGL.REYYLT----DP.PNPNKKK.L.QLTG-----------..PT.VGN..K...QRDRAAA DN62049_c0_g2_i3|m.62452
SixX3/6     Sr GEQKTHCFKERTRGL.REYYLT----DP.PNPNKKK.L.QLTG-----------..PT.VGN..K...QRDRAAA DN126162_c1_g1_i3|m.89990
Six3/6      Hm GEQKTHCFKERTRGL.REWYLQ----DP.PNPAKKR.L...TG-----------..PT.VGN..K...QRDRAAA DN28007_c0_g1_i1|m.9588
Six3/6 a    Nw GEQKTHCFKERTRNL.REWYLQ----DP.PNPQKKR.L.A.TG-----------..PT.VGN..K...QRDRAAA 36924.0
Six3/6 b    Nw GEQKTHCFKERTRNL.REWYLQ----DP.PNPQKKR.L.G.TG-----------..PT.VGN..K...QRDRAAA 42137.0
Six3/6 a    Xb GEQKTHCFKERTRSL.REWYLQ----DP.PNPTKKR.L.Q.TG-----------..PT.VGN..K...QRDRAAA DN23152
Six3/6 b    Xb EAQTTRLFQKNTLSV.KDFYQS----DP.PNLNKIKDL.R.TR-----------.SPI.IRT..K...QRDRIAA DN97213
SIX3/6128   Nv GEQKTHCFKERTRSL.REWYLQ----DP.PNPTKKR.L.Q.TG-----------..PT.VGN..K...QRDRAAA DQ206294
```

Six 4/5 family

```
SIX4        Hs GEETVYCFKEKSRNA.KELYKQ----...PSPAEKRHL.KITG-----------.SLT.VSN..K...QRDRNPS NP_059116
SIX5        Hs GEETVYCFKERSRAA.KACYRG----...P.PDEKRRL.TLTG-----------.SLT.VSN..K...QRDRTGA NP_787071
SIX4/5      Bf GEETVYCFKEKARQA.KEMYNN----...P.PDEKRNL.KKTG-----------..LT.VSN..K...QRDRTPS
six4        Dm GEETVYCFKEKSRNA.KDCYLT----...P.PDEKKTL.KKTG-----------..LT.VSN..K...QRDRTPQ NP_649256
unc-39      Ce GEEIVYSFKDSSRKF.KQFFRN---VSE.P.QEQKR..SRATG-----------.KIV.ISN..K...QRDKSN
```

```
Six4_5      Cg  EETIYCFKEKSRQS.KECYKN----.R.P.PDEKRNL.KKTG-----------..LT.VSN..K...QRDRT
Six4/5      Ip GEETVYCFKEKSRAK.KDSYRR----.K.PSPEEKRSL.KSSG-----------..LT.VSN..K...QRDRTPH DN26501_c1_g1_i1
Six4/5      Sr GEETIYCFKEKSRNK.KESYKK----.K.P.PEEKKAL.KNSG-----------..LT.VSN..K...QRDRTPH DN118860_c2_g1_i2|m.56407
Six4/5      Hm GEETVYCFKEKARRK.KESYKK----.K.P.PEEKKSL.KASG-----------..LT.VSN..K...QRDRAPT DN66848_c0_g1_i1|m.42105
Six4/5      Xb  EETVYCFKEKSRQA.KECYRR----.R.P.PEEKRSL.KMTG-----------..LT.VSN..K...QRDRAPK DN120045
SIX4/5a030 Nv GEETVYCFKEKARAA.KDCYEQ----.K.P.PQEKRL..KQTN-----------..LK.VSN..K...QRGRIPS DQ206264
SIX4/5b061 Nv GEKTIYFFKEKVRTV.KECYEH----KK.P.LKEKRV..TQTN-----------..L..VRN..R...HRDRISS DQ206322
```

TALE CLASS

Irx family

```
IRX1        Hs DPGRPKNA.RES.ST.KAWLNEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_077313
IRX2        Hs DPAY.KNA..DA.AT.KAWLNEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_150366
IRX3        Hs DPSRPKNA.RES.ST.KAWLNEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_077312
IRX4        Hs SGTR.KNA..ET.ST.KAWLQEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_057442
IRX5        Hs DPAY.KNA.RDA.AT.KAWLNEHRK-...P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_005844
IRX6        Hs GAGR.KNA..ET.ST.KAWLNEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_077311
IRXA        Bf DGAR.KNA..ET.AT.KAWLMEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL....
IRXB1       Bf DGTR.KNA..DA.ST.KAWLNEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL....
IRXB2       Bf DGTP.KNA..DA.ST.KAWLNEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL....
ara         Dm LAAR.KNA..ES.AT.KAWLNEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_524045
caup        Dm LAAR.KNA..ES.AT.KAWLSEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_524046
mirr        Dm NGAR.KNA..ET.ST.KAWLNEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... NP_524047
irx-1       Ce DGIKKRNA..EA.AP.KDWLHSHRK-.P.PSKADKVML.VGTG-----------M.LT.VST..A.A.RRL....
Irx1        Cg  AARRKNA..ET.SA.KAWLYEHRK-.PYP.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL..
Irx2        Cg  NARRKNA..ET.NT.KAWLYEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL..
Irx3        Cg  GGRRKNA..ET.AP.KAWLKEHMK-.P.P.KAEKVML.IVTK-----------M.LT.IST..A.A.RRL..
Irx4        Cg  AARRKNA..ET.NT.KAWLYEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL..
Irx a       Ip SSIR.KNAS.ET.AA.KAWLYEHRN-.P.P.KGEK.ML.IITK-----------M.LT..ST..A.A.RRL.... DN44510_c0_g1_i1|m.23508
Irx a       Sr SSMR.KNAS.ET.AA.KTWLFEHRK-.P.P.KGEK.ML.IITK-----------MSLT..ST..A.A.RRL.... DN126498_c0_g1_i4|m.91973
Irx a       Hm CNIR.KNTS.ET.AA.KSWLYEHRN-.P.P.KGEKVML.IITK-----------M.LT.VST..A.A.RRL.... DN21447_c0_g3_i1|m.6190
Irx a       Nw SGMRRKNSS.ET.QY.KSWLKHHQK-.P.P.KGEK.ML.ILTK-----------M.LT.VST..A.S.RRI.     26582.1
Irx a       Xb GMMRKKSA..ET.NA.KAWLREHRR-.P.P.KSEK.ML.IITK-----------M.LT.VST..A.A.RRL. DN63799
Irx b       Ip MITR.KNA..ET.SA.KAWLNDHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... DN66612_c1_g4_i2|m.91737
Irx b       Sr MITR.KNA..ET.SA.KAWLSEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... DN127996_c1_g1_i3|m.101906
Irx b       Hm GIAR.KNA..ET.SA.KAWLSEHKK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL.... DN36375_c0_g1_i1
            (sequencing issue, remove one A from a polyA stretch)
Irx c       Ip SSPASPSAVKES.YP.RVWMEENRA-.P.P.KNQKLIL.LVTK-----------..LT.VST..A.A.RRL...R DN73792_c1_g1_i1
Irx c       Sr SSPNSATAVKEN.YP.RVWLEENRD-.P.P.KSEKLMLSIVTK-----------M.LT.VST..A.A.RRL...R DN123214_c0_g1_i1
Irx d       Sr ALGR.KNA..ET.AA.KSWLQDHLK-.P.P.KGEKVML.IITK-----------MSLT.VST..A.A.RRL...G DN126207_c1_g2_i3|m.90240
Irx d       Hm PLAR.KNA..ET.AA.KTWLNDHLK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL...G DN24763_c0_g1_i1|m.8213
Irx d       Ip                           .P.P.KGEKVML.IITK-----------MSLT.VST..A.A.RRL...G DN71570_c5_g7_i5
```

```
Irx b       Nw VGCRRKNA..ET.SA.KTWLADHRK-.P.P.KAEK.ML.VITK-----------M.LT.VST..A.A.RRL. 10005.0
Irx b       Xb  GVARRKNA.ET.SA.KAWLHEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL. DN84998
Irx c       Nw LGLRRKNA..ET.SP.KAWLYEHRK-.P.P.KAEK.ML.IITK-----------M.LT.VST..A.A.RRL. 28871.0
Irx d       Nw NGIRRKNA..ET.ST.KAWLYEHRK-.P.P.KGEK.ML.VITK-----------M.LT.VST..A.A.RRL. 36182.0
Irx e       Nw PGIRRKNA..ET.ST.KAWLYEHRK-.P.P.KGEK.ML.IITK-----------M.LT.VST..A.A.RRL. 48316.0
IRX (108)   Nv AGAR.KNA..ET.ST.KAWLFEHRK-.P.P.KGEK.ML.ILTK-----------M.LT.VST..A.A.RRL.... DQ206288
```

Meis family

```
MEIS1       Hs .HKK.GIFPKVA.NIMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM NP_002389
MEIS2       Hs .QKK.GIFPKVA.NIMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM NP_002390
MEIS3       Hs .NKK.GIFPKVA.NIMRAWLFQHLS-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM NP_001009813
Meis        Bf .QKK.GIFPKVA.NIMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM
hth         Dm NQKK.GIFPKVA.NI.RAWLFQHLT-HP.PSEDQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM NP_476577
unc-62      Ce KSRG.GIFPKQA.NR.RQWLFQNLT-HP.PSEEQKKQL.KETG-----------..IL.VNN..I.A.RRIVQPM
Meis        Cg                IMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQ
Meis        Ip PQKK.GIFPKTA.NIMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.V                DN70650_c1_g5_i7
ext above   Ip                             P.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM DN69333_c0_g4_i5
Meis        Sr PQKK.GIFPKTA.NIMRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM DN129314_c1_g2_i3|m.111748
Meis a      Hm PQKK.GIFPKAA.NIMRTWLFTNLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM DN34684_c0_g1_i2|m.15334
Meis b      Hm PQKK.GIFPKAA.NIMRSWLFTNLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIVQPM DN37189_c0_g1_i1|m.18754
Meis a      Xb HQKK.GIFPKSA.NVMRAWLFQHLS-HP.PSEEQKKQL.QETG-----------..IL.VNN..I.A.RRIV    DN64910
Meis b      Xb YQKK.GIFPKSA.NILRAWLFQHLT-HP.PSEEQKKQL.QDTG-----------..IL.VNN..I.A.RRIV    DN79707
MEIS (107)  Nv SQKK.GIFPKAA.NIMKAWLFQHLT-HP.PSEEQKRSL.QETG-----------..IL.VNN..I.A.RRIVQPM DQ206227
```

Pbx family

```
PBX1        Hs AR.K.RNFNKQA.EI.NEY.YSHLS-.P.PSEEAKE.L.KKCG-----------I.VS.VSN..G.K.IRY..NI NP_002576
PBX2        Hs AR.K.RNFSKQA.EV.NEY.YSHLS-.P.PSEEAKE.L.KKCG-----------I.VS.VSN..G.K.IRY..NI NP_002577
PBX3        Hs AR.K.RNFSKQA.EI.NEY.YSHLS-.P.PSEEAKE.L.KKCS-----------I.VS.VSN..G.K.IRY..NI NP_006186
PBX4        Hs AR.K.RNFSKQA.EV.NEY.YSHLN-.P.PSEEAKE.L.RKGG-----------..IS.VSN..G.K.IRY..NM NP_079521
PBX         Bf NR.K.RNFSKQA.EV.NEY.YSHLS-.P.PSEEAKE.L.RKCG-----------I.VS.VSN..G.K.IRF..NI
exd         Dm AR.K.RNFSKQASEI.NEY.YSHLS-.P.PSEEAKE.L.RKCG-----------I.VS.VSN..G.K.IRY..NI NP_523360
Ceh-20      Ce AR.K.RNFSKQA.EV.NEY.YGHLS-.P.PSEEAKEDL.RQCN-----------I.VS.VSN..G.K.IRY..NM
Ceh-40      Ce AR.K.RNFSKTS.EI.NEY.LANIN-HP.PSEEVKQAL.MQCN-----------ISVA.VSN..G.K.IRY..TM
Ceh-60      Ce QG.K.RNFDKNT.DI.QNW..DHRQ-.P.PSDQEKA.L.KQCN-----------IKIS.VNN..G.Q.IRT.QQA
Pbx         Cg  R.K.RNFSKQA.EL.NEY.YSHLS-.P.PSEEAKE.L.RKCG-----------I.VSQVSN..G.K.IRY..
Pbx         Ip AR.K.RNFSKQASEI.NEY.YSHLN-.P.PSEEVKE.L.KKCC-----------.SVS..SN..G.K.IRY..NI DN70276_c2_g3_i3|m.132878
Pbx         Sr AR.K.RNFSKQASEI.NEY.YSHLN-.P.PSEEVKE.L.KKCG-----------.SVS..SN..G.K.IRY..NI DN127821_c2_g3_i2|m.101192
Pbx         Hm AR.K.RNFSKQASEI.NEY.YSHLC-.P.PSEEVKE.L.KKCC-----------.SVS..SN..G.K.IRY..NI DN21081_c0_g1_i1|m.5973
Pbx a       Nw AR.K.RNFNKTA.EV.NEY.YSHLS-.P.PSEEVKEDL.KKCN-----------LSIS.VSN..G.K.IRY..NI 52332.1 A
Pbx b       Nw AR.K.RNFNKTA.EV.NEY.YSHLS-.P.PSEEVKEDL.KKCG-----------LSIS.VSN..G.K.IRY. 52332.1 B
Pbx c       Nw AR.K.RNFNKNA.EF.NEY.YSHLN-.P.PSEETKE.L.KKCG-----------LSIS.VSN..G.K.IRY. 52537.0
```

```
Pbx        Xb AR.K.RNFNTNA.QI.NEY.YSHLS-.P.PSEEVKE.L.RKCG-----------..VS..SN..G.K.IRY. DN68543
PBX (040)  Nv YR.K.RNFSKQA.EI.NEY.YSHLS-.P.PSEEAKE.L.RKCN-----------ISVA..SN..G.K.IRY..NI DQ206245
```

Pknox family

```
PKNOX1     Hs S.NK.GVLPKHA.NVMRSWLFQHIG-HP.P.EDEKKQ..AQTN-----------..LL.VNN..I.A.RRILQPM NP_004562
PKNOX2     Hs S.NK.GVLPKHA.NIMRSWLFQHLM-HP.P.EDEKRQ..AQTN-----------..LL.VNN..I.A.RRILQPM
PKNOX      Bf .NGK.GILPKQA.DVMRSWLFQHIV-HP.P.EDEKRA..NQTN-----------..LL.VNN..I.A.RRILQPM
Pknox      Cg  .SK.GVLPKHA.QVMKSWLFQHIV-HP.P.EDEKRQ..NQTN-----------..LL.VNN..I.A.RRILQ
Pknox      Ip K.TK.GVLPKHA.QIMRQWLFQHIV-HP.P.EDEKRQ..VQTN-----------..LL.VNN..I.A.RRILQPM DN69528_c1_g2_i3|m.122462
Pknox      Hm K.TK.GVLPKQA.QIMRSWLFQHIV-HP.P.EDEKRQL.QQTS-----------..LL.VNN..I.A.RRILQPM DN37578_c0_g1_i2|m.19358
Pknox      Sr K.TK.GVLPKHA.QIMRQWLFQHIV-HP.P.EDEKRQ..VQTN-----------..LL.VNN..I.A.RRILQPM DN128964_c2_g7_i4|m.109271
Pknox      Nw KVSK.GILPKHA.SVMRSWLFQHIV-HP.P.EDEKRQL.GQTG-----------..LL.VNN..I.A.RRILQPM 45317.0 added aa
Pknox      Xb    K.GVLPKQA.NIMRSWLFQHIV-HP.P.EDEKRQ..VQTN-----------..LL.VNN..I.A.RRILQPM DN74939
NVHD036    Nv R.SK.GVLPKQA.SIM.TWL.QHIM-HP.P.EDEKRS..QQTN-----------..IL.VNN..I.A.RRILQPM DQ301954
NVHD143    Nv ARKR.GKLPESARTR.MTWLFAHSN-RP.P.EEEKSNL.TITG-----------..PI..NN..S.A.RRIL.KA DQ301952
```

Tgif family

```
TGIF1      Hs KR.R.GNLPKESVQI.RDWLYEHRY-.A.PSEQEKALLSQQTH-----------.STL.VCN..I.A.RRLLPDM NP_003235
TGIF2      Hs KRKR.GNLPKESVKI.RDWLYLHRY-.A.PSEQEKLSLSGQTN-----------.SVL..CN..I.A.RRLLPDM
TGIF2LX    Hs K.KRKGNLPAESVKI.RDWMYKHRF-KA.PSEEEKQMLSEKTN-----------.SLL..SN..I.A.RRILPDM
TGIF2LY    Hs K.KRKGNLPAESVKI.RDWMYKHRF-KA.PSEEEKQMLSEKTN-----------.SLLR.SN..I.A.RRILPDM
TGIF       Bf ARKR.GNLPKEAVQI.RAWLYDHRY-.A.P.DAEKLDL.REAG-----------.TVL.VCN..I.A.RRILPDM
achi       Dm LRKR.GNLPKTSVKI.KRWLYEHRY-.A.PSDAEKFTLSQEAN-----------..VL.VCN..I.A.RRILP.M NP_725182
vis        Dm LRKR.GNLPKSSVKI.KRWLYHRY.-.A.PSDAEKFTLSQEAN-----------..VL.VCN..I.A.RRILP.M NP_523714
Tgif       Cg  R.R.GNLPKEAVRI.KTWLYEHRY-.A.PSDQEKVYLSS.TN-----------..VL.VCN..I.A.RRILP
Tgif       Ip E.KK.GSLPKEAICIMRSWLEQHRY-HA.P.EQEKLSL.QQTN-----------..PL.VCN..I.A.RRVLPEI DN71480_c1_g4_i4
Tgif       Sr L.KQ.GTLPKEAVDILRSWLELHRY-HA.P.EQEK.SLSE.TN-----------..PL.VCN..I.A.RRILPAL DN142180_c0_g1_i1|m.133360
Tgif       Hm M.KK.GNLPKESVEVMK.WLNEHRY-.A.PNETEKNQL.EKAN-----------..VL.VCN..I.A.RRILPEM DN12757_c0_g1_i1|m.3416
Tgif       Xb  ..R.GNLPKDAVNV.RLWLYEHRF-.A.P.EQEKVELSEKAS-----------.SVL.VCN..I.A.RRILPDM DN45414
TGIF (149) Nv P..R.GNLPKDSVNV.RLWLWEHRF-.A.PSEAEKQYLSK.AN-----------.SVL.VCN..I.A.RRILPDM DQ206298
```

TALE CLASS / Family membership ambiguous

```
           Sr SEARSASFNDNALEV.KEWLEGHLD-HP.IE.SDILK.QTRTS-----------.SRK..MG.CT.I.RRRLTII DN106703_c0_g1_i2
           Ip SR.RSKDQIEED.VPLRHWLLLHSG-HPFP..VEKKRL.RECG-----------.QLK.VED..L.T.RRIL.HH DN68679_c0_g2_i1
```

CUT CLASS

Onecut family

```
ONECUT1    Hs P.KP.LVF.DV.RRT.HAI.KE----.KRPSKELQ.T.SQQ..-----------.ELSTVSNF.M.A.RRSLDKW NP_004489
```

```
ONECUT2     Hs Q.KS.LVF.DL.RRT.FAI.KE----.KRPSKEMQ.T.SQQ.G-----------.ELTTVSNF.M.A.RRSLEKW NP_004843
Loc390874   Hs P.KQ.LVF.DL.RRT.IAI.KE----.KRPSKEMQVT.SQQ.G-----------.ELNTVSNF.M.A.RRCMNRW XP_372702
ONECUT      Bf P.KP.LVF.DL.RRT.HAI.KE----.KRPSKEMQAQ.AKQ.G-----------.DLSTVCNF.M.A.RRSQDKW
onecut      Dm P.KP.LVF.DL.RRT.QAI.KE----TKRPSKEMQVT..RQ.G-----------.EPTTVGNF.M.A.RRSMDKW NP_524842
ceh-21      Ce V.VI.LTF.ET.LKS.QKS.QQ----.HRPT.EMRQKLSAT.E-----------.DFSTVGNF.M.S.RRLRIDQ
ceh-38      Ce S.RP.LVF.DI.KRT.QAI.KE----TQRPS.EMQQT.AEH.R-----------.DLSTVANF.M.A.RRSRLGG
ceh-39      Ce  RIR.FTF.QT.LDS.HTV.QQ----QDRPN.EMQQALSAT.K-----------.NRSTVGNF.M.A.RRLPKAA
ceh-41      Ce P.KTKSPF.EHEIAVMMAL.EI----.KSPNHEEVQKLAVQ.N-----------.GYRSVANF.M.K.A.ER.KL
ceh-48      Ce P.KT.LVFSDI.RRT.QAI.RE----TKRPS.EMQIT.SQQ.N-----------.DPTTVANF.M.A.RRGHDLK
ceh-49      Ce TP.A.CLLSQD.KSQ.SIF.ET----.PRPDSLEMKQLGST.N-----------.CKSTIINY.T.M.RRKATSS
Onecut      Cg  .KP.LVF.DI.RRT.HAI.KE----TKRPSKEMQAT.AQQ.G-----------.EVTTVANF.M.A.RRSLD
Onecut a    Ip Q.KN.LVF.DI.RRT.FAI.RE----TKRPSKEMQLQ.SRQ.G-----------.EPTTVSNF.M.A.RRCGDRW DN36933_c0_g1_i1|m.17321
Onecut b    Ip M.KS.LVF.DL.RRT.RAI.QE----SQRPSKESQAQ..KQ.G-----------.EVTTVSNF.M.A.RRCGDRW DN31212_c0_g1_i1|m.13546
Onecut a    Sr T.KN.LVF.DL.RRT.FAI.RE----TKRPSKEMQ.T.SRQ.G-----------.EQSTVSNF.M.A.RRCSDRW DN125608_c1_g1_i1|m.86741
Onecut b    Sr S.KS.LVF.DI.RRT.RAI.QE----TQRPSKESQVQ.GKQ.G-----------.EVSTVSNF.M.A.RRCGDRW DN117560_c0_g1_i1|m.52153
Onecut      Hm P.KI.LVF.DI.RRT.YAI.RE----TKRPSKEMQQT.GQQ.G-----------.DPTTVSNF.M.A.RRCMGDR DN40002_c0_g1_i1|m.24477
Onecut      Nw   KQ.LVF.DL.RRT.HAI.KE----TQRPSKEMQLQ..SQ.G-----------.QLTTVSNF.M.A.RRSVDKW 46342.0
Onecut      Xb                      RE----TKRPSKEMQMT..QQ.G-----------.ELSTVSNF.M.A.RRSLDKW DN97994
            Nv S.KI.RQF.ET.KRT.FAI.KE----TKKPTKDMQKAV.EE.G-----------.DLNSVANF.M.A.RRHTDA  XP_001637067.1
```

CUT class ambiguous

```
            Nw S..L.YSLDDDSKRFLLSVYQH----VTHPSQDTISRL.TH.N-----------MQPATVYN......QRF.VKD 35933
            Nw LRARTVISQKDRAIMFQHFLTR----.Q...ANELLHLSRS.D-----------IPQPT.R.F...LRARIRART 53646.0
            Xb N.IRRINL.TREKNR.LEAYNN----EP.PSPQTIEGL.QELG-----------IRQTTVNN..H.H.SRS.RTK DN95119
            Xb RR.RTTLSDHAKRVMTTYFMTT----SR.LSADA.MAFSQQLG-----------.S.ETVR.F.K.L.ANT.NM  DN91240
```

ZF CLASS

Zfhx family

```
ZFHX2 I     Hs .RFS.TKF.EF..QA.QSF.ET----SA.PKDGEVERL.SL.G-----------LAS.VVVV....A.Q.AR.NA
ZFHX2 II    Hs D..L.T.ILPE.LEI.YRWYMQ----DSNP..KMLDC.SEEVG-----------.KK.VVQV....T.ARER.GQ
ZFHX2 III   Hs QR.Y.TQMSSL.LKIMKACYEA----Y.TP.MQECEVLGEEIG-----------.PK.V.QV....A.A.E..AK NP_207646
ZFHX3 I     Hs N..P.TRI.DD.LRV.RQY.DI----.NSPSEEQIK.M.DKSG-----------.PQKV..H..R.TLF.ERQR.
ZFHX3 II    Hs KRSS.TRF.D..LRV.QDF.DA----.A.PKDDEFEQLSNL.N-----------.PT.V.VV....A.Q.AR.NY
ZFHX3 III   Hs D..L.T.I.PE.LEI.YQKYLL----DSNP..KMLDH...EVG-----------.KK.VVQV....T.ARER.GQ
ZFHX3 IV    Hs Q..F.TQM.NL.LKV.KSC.ND----Y.TP.MLECEVLGNDIG-----------.PK.VVQV....A.A.E..SK
ZFHX4 I     Hs F..P.TRI.DD.LKI.RAY.DI----.NSPSEEQIQ.M.EKSG-----------.SQKV..H..R.TLF.ERQR.
ZFHX4 II    Hs KRSS.TRF.D..LRV.QDF.DT----.A.PKDDEIEQLSTV.N-----------.PT.V.VV....A.Q.AR.SY
ZFHX4 III   Hs D..L.T.I.PE.LEI.YEKYLL----DSNP..KMLDHI.REVG-----------.KK.VVQV....T.ARER.GQ
ZFHX4 IV    Hs H..F.TQMSNL.LKV.KAC.SD----Y.TP.MQECEMLGNEIG-----------.PK.VVQV....A.A.E..FK
```

```
ZFHX I      Bf G..P.TRI.DE.LKI.RAH.DI----.NSPSEDQIL.MSEKSG-----------.PTKV..H..R.TLF.ERQRN
ZFHX II     Bf RRAG.TRF.DY.IKV.QEF.EQ----.A.PKDDDLEHLSKI.N-----------.SP.V.VV....A.Q.AR.SF
ZFHX III    Bf N..L.T.I.PE.LEI.YQKYQI----DCNP..KMLDL..REVG-----------.KK.VVQV....T.ARER.GQ
ZFHX IV     Bf G..Y.TQMSNL.LKM.KSA.TE----YKTP.MQECEILGNEIG-----------.PK.VVQV....A.A.D..SK
zfh2 I      Dm N..L.T.ILPE.LNF.YECYQS----ESNPS.KMLE..SKKVN-----------.KK.VVQV....S.A.D..SR NP_524623
zfh2 II     Dm KRAN.TRF.D..IKV.QEF.EN----.S.PKDSDLEYLSKL.L-----------.SP.V.VV....A.QKQR.IY
zfh2 III    Dm N..L.T.ILPE.LNF.YECYQS----ESNPS.KMLE..SKKVN-----------.KK.VVQV....S.A.DK.SR
ZC123.3 I   Ce A..A.TRI.DD.LKV.RQY.NI----.NSPSEAQIKEMSQKAS-----------.P.KV..H..R.TLFKERQRD
ZC123.3 II  Ce RRAN.TRF.DF.LRT.QQF.DK----QAYPKDDDLEALSKK.Q-----------.SP.V.VV....A.Q.AR.IY
ZC123.3 3   Ce N..F.THL.PM.VQMMKSV.NE----YKTPSMAECELLGKEVG-----------.HK.VVQV....A.A.ER.TR
Zfhx        Cg Q..A.TRINDE.LKI.RAH.DI----.NSPSEEQINAMSEQSG-----------.PQKV.KH..R.TLF.ERQRN
            Cg KRAN.TRF.DY.IKV.QEY.EQ----.A.PKDDELDHLSQM.N-----------.SP.V.VV....A.Q.AR.IY
            Cg D..L.T.ILPE.LDY.YQKYQI----DCNPS.KQLETISAEVG-----------.KK.VVQV....T.ARER.GQ
            Cg G..Y.TQM.SL.VRIMKSI.ID----YKTP.MAECEMLGREIG-----------.PK.VVQV....A.A.E..AK
3550-3700   Ip SKQP.TKISDE.LAI.KQN.DI----SNSPSEEQYLAISVQTG-----------.PMKV..H..R.TLF.ERQKD DN75125_c1_g6_i2|m.217565
3950-4050   Ip KRAT.TRF.DH.IKA.QDF.DQ----.P.PRDDDLETLSKN.N-----------.SP.VVVV....Q.Q.ARRSL DN75125_c1_g6_i2|m.217565
4750-4940   Ip FR.M.T.ISPD.LEV.YEKYGE----.SNPS.SEMEA.SA.VG-----------.SR.VVQV.Y..T.ARQRRGQ DN75125_c1_g6_i2|m.217565
5540-5620   Ip ...H.TNLSML.IKT.KSI.AE----YKTP.MTECET.GSE.N-----------.SK.VVQV....A.S.E..IQ DN75125_c1_g6_i2|m.217565
4784-4963   Sr QKQP.TKISEP.LEI.KQN.DI----SNLPSEEKY.ALSQQTG-----------.PVKV..H..R.TLF.ERQKD DN127982_c1_g1_i6|m.101793
5225-5404   Sr KRAT.TRF.DH.IKS.QDF.DQ----.P.PRDDDLENLSKS.S-----------.SP.VVVV....Q.Q.ARRSM DN127982_c1_g1_i6|m.101793
6167-6346   Sr FR.M.T.ISPD.LEI.YTKYGE----.SNPS.TEMET.ASSVG-----------.SR.VVQV.Y..T.ARQRRGQ DN127982_c1_g1_i6|m.101793
6977-7156   Sr K..H.THLSNV.VRT.KAM.LD----FKTPSMGECEQ.G.E.G-----------.SK.VVQV....A.S.E..GK DN127982_c1_g1_i6|m.101793
2704-2883   Hm QK.P.TRISDE.LKI.KDH.DL----SNSPTDSEVDQ..IKTG-----------.QGKV..H..R.TLF.ERQK. DN33495_c0_g1_i4|m.14054
3046-3225   Hm .RAS.TRFSDF.LKS.QEF.ER----.A.PRDEDLEKLSKS.N-----------.SP.VVVV....Q.Q.ARRNI DN33495_c0_g1_i4|m.14054
3628-3807   Hm DKRI.TTISPA.LEV.YSSYQE----DSNPS.QMLE..CAKVG-----------.KR.VVQV....T.ARQR.GA DN33495_c0_g1_i4|m.14054
4459-4638   Hm .R.Y.THLSVT.IKIFKVI.AD----YKTPTMSDCEVLGSQ.D-----------.PK.VVQV....A.S.E..AR DN33495_c0_g1_i4|m.14054
I           Nw  K.P.TRITDD.LKV.RTH.DI----.NSPNDDQIAN.SQETV-----------.PVKV.KH..R.TLF.ERQRN 60532.0
II          Nw KRAG.TRFTDW.IQK.QEF.EI----.A.PKDDDLDTLSKDLS-----------.NP.V.VV....Q.Q.AR.SF 64721.0
III         Nw DR.L.TSITPK.LNT.YKW.KE----.SNPS.KMLEI.SVDVS-----------.KK.VVQV....T.ARQR.AS 50181.0
IV          Nw D..Y.TQLSHL.IRV.KACYVD----YKTPTMNECE.LGKTIG-----------.AK.VVQV....A.A.D..SK 50181.0 is one protein.
I           Xb TK.P.TRITDE.LKI.RAY.DI----.NSPSEEQII.MSEKSG-----------.PQKV.KH..R.TLF.ERQRN DN93733
II          Xb RRAG.TRFTDF.IKR.QEF.DT----.AYPKDDDLDRLSKELN-----------.NP.V.VV....A.Q.AR.SF DN93733
III         Xb DR.L.TTITPE.LEI.YEKYQE----.SNPS..MLDI.SREVG-----------.KK.VVQV....T.ARQR.GQ DN93733
IV          Xb G..F.TQMSNT.LKV.KHC.TE----Y.TPTMQECELLGSCIG-----------.QK.VVQV....A.A.E..SK DN93733
```

ZF CLASS / Family membership ambiguous

```
            Sr ERKK.TV..DK.LET.REYYVK----QPNPNSDDIEK..NNIG-----------..KRNVQV....K.ARD..NP DN130571_c0_g1_i2|m.121942
            Ip ERKK.TV..DQ.LDT.KEYYIR----QPNP.SDDIER..SVIG-----------..K.NVQV....K.ARD..NP DN55547_c0_g2_i6|m.41438
            Hm QRKK.TVF.DF.QKM.RELYKI----.SSPTSEEISKL.DNIG-----------.SKRSIQV....K.ARD..SR DN34160_c0_g1_i1|m.14868

            Ip .H.T.TKLPAAELAL.K.IYET----.PSPSKKQFA.L.RRTS-----------.PIKVLKV....T.RLTRNTC DN61481_c0_g1_i1|m.60009
```

```
          Sr .H.T.TKL.QSEVVL.KSIYEN----TPSPSK.QFE.L.NSTG-----------.PTKVL.V....T.RLRRNTC DN108290_c0_g1_i1|m.33236
          Hm .H.V.TRL.AQDL.Y.HNMYRR----TANPS.RLLA..SNHIG-----------.PL.V.QV....T.RLM.NTC DN42050_c0_g2_i1|m.31057
```

Cers CLASS

```
LASS2      Hs NIKEKTRLRAPPNAT..HFYLTS---GKQPKQVEVELLSRQSG-----------.SG..VER..RR..NQDRPSL NM_022075
LASS3      Hs FGIKETVRKVTPNTV..NF.KHS---T.QPLQTDIYGL.KKCN-----------.....VER..RS..NQERPSR NP_849164
LASS4      Hs GVRDQTRRQVKPNAT...H.LTE---GHRPKEPQLSLL.AQCG-----------..LQ.TQR..RR..NQDRPQL
LASS5      Hs GIEDSGP.QAQPNAI...V.ISI---TK.PDKK.LEGLSKQ.D-----------WNV.K.QC..RH..NQD.PPT NP_671723
LASS6      Hs NIQANGPQIAPPNAI...V.TAI---TKHPDEK.LEGLSKQ.D-----------WDVRSIQR..RQ..NQE.PST
CERS       Bf LGIPGERKYAEPNAI...V.TSI---TKNPDEK.LQGL.KQ.D-----------WS...VQR..RR..NQDRPTL
Lag1/schlk Dm GI.SSRPKKAANVPI...TYAK----STR.DKKKLVPLSKQTD-----------MS..E.ER.WRL..AQD.PST NP_652526
Cers       Cg   IKDGRPKTVQPISALENSYRI---KKFPSEDTVKGLEKQTD-----------M..RE.QR..RM..NQERP
Cers       Ip GISP.GPRKPVHNAI..RV.HTI---TQKPDEK.LKGLEKQ.D-----------WSPY..RQ..AR..AQIRPTK DN63567_c1_g1_i1|m.70349
Cers       Sr GICP.GARKPYPNAI.ERV.LTI---TQRPDEKHLKGLEKQ.DC----------WNQYKIKQ..AL..AQLRPSK DN124898_c2_g4_i2
Cers       Hm VLPST.QP.PQPNAI.DKVYTSI---ARDPDAV.LAGL.KQ.D-----------WSIR.VQV..IA..ALDRPSH DN30308_c0_g1_i1|m.11388
Cers a     Nw IRSSSNQLRPIDNLV.ERVYLSV---SRTPDDS.LKGLSKQ.D-----------WS.KQVEA..HL..AQDRAS  51994.0
Cers b     Nw  IKAKKPSNGGLINPVLEEAYGK---CTRPNREAIKGM.KKAD-----------MEADRVER..RR..QQDQP   44536.0
Cers c     Nw   RSSNHLRAVDNLV.ERVYASV---SRNPDGN.IKGLSKQ.D-----------WN.KQVQA..HL..AQDRSS  40648.1
Cers d     Nw  QIKPKQRNSKSIHP.LEQVYKD---KSSPPDKVIE.L.KK.D-----------SEPPK.RR..SR.HRHDQP   62117.0
Cers a     Xb       RLRAPPNAT.EHFYLTS---GKQPKQVEVELLSRQSG-----------LSGRQVER..RR..NQDRPSL DN112933
Cers b     Xb GLKIPARH.PEDNTI.EKVYRSV---SKQPDAK.IEGL.KQ.D-----------WSERQVQR..RR..NSDRPS  DN87293
           Nv  GVKDKKINFTPNPFCEKVYQTI---.KFPPSD.ILGLSKQIG-----------W.T.EVER..RH...QS.PSL XP_001635937.1
```

Pros Class

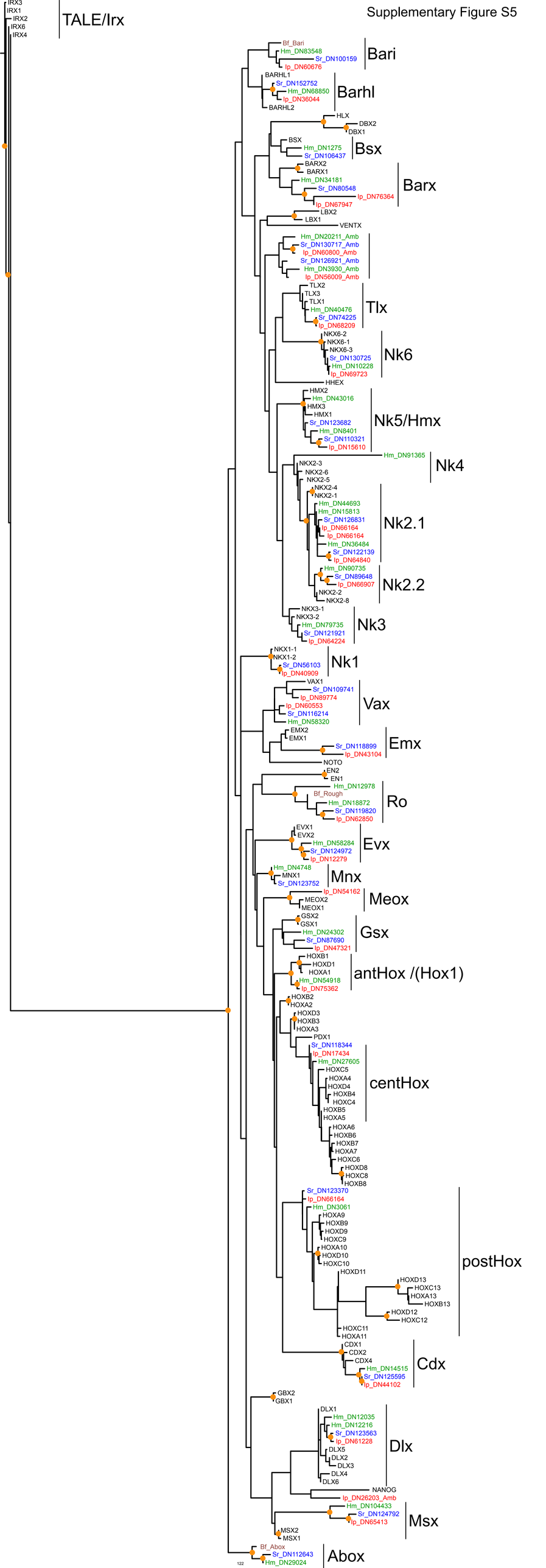
```
PROX1      Hs GSAMQEGLSPNHLKKAKLM.FY----T..PSSNMLKTYFSDVKFNR--------CITS.LIK..S.F.EFYYIQM NP_002754
PROX2      Hs LVHIQEGLNPGHLKKAKLM.FF----T..PSSNLLKVYFPDVQFNR--------CITS.MIK..S.F.EFYYIQM
PROX       Bf SPPHPGSL.PTHLKKAKLM.FY----T..PSSSTLKQHFPDVKFNR--------CITS.LIK..S.F.EFYYIQM
Pros       Dm TPLHSSTL.PMHLRKAKLM.FW----V..PSSAVLKMYFPDIKFNK--------NNTA.LVK..S.F.EFYYIQM FBpp0293145
Ceh-26     Ce GGGSSSML.PMHLRKAKLM.FY----T..PNSNLLKSYFPDIRFNK--------NNTA.LVK..S.F.EFYYNQM
Pros X1    Cg LDHWNGTL.PMHLRKAKLM.FY----V..PSSAILKVYFPDVKFNK--------NNTA.LVK..S.F.EFYYIQM
Prox       Ip AVGSTDNL.ANHLRKAKLM.FF----Q..P.SSTLRQYFTDVKFNR--------HNNA.LIK..S.F.EYYYIQI DN52339_c0_g1_i2
Prox       Hm HSLAAE.L.ANHLKKAKLI.FY----Q..PNPSTLRMYFPDVRFNR--------QNNG.LIK..S.F.EYFYNQI DN24356_c0_g1_i1
Prox       Sr VVQMTENL.ANHLRKAKLM.FY----Q..PNSAILRQFFVDVKFNR--------SNNA.LIK..S.F.EYYYIQI DN113132_c0_g1_i1
Prox a     Nw NQGWATNL.PNHLKKAKLM.FY----T..PSSAVLRLHFPDVRFSR--------HNNA.LIK..S.F.EFFYLQI 44308.0
Prox b     Nw ASFWTSNL.PEHLKKAKLM.FY----T..P.SAMLRINFPDVRFTR--------HINA.LIK..S.F.EYFYIQI 51275.0
Prox       Xb SSGWNTNL.PSHLKKAKLM.FY----V..PNSATLRQHFPDVKFTR--------HNNA.LIK..S.F.EYFYIQI DN88028
```

Ambiguous class membership

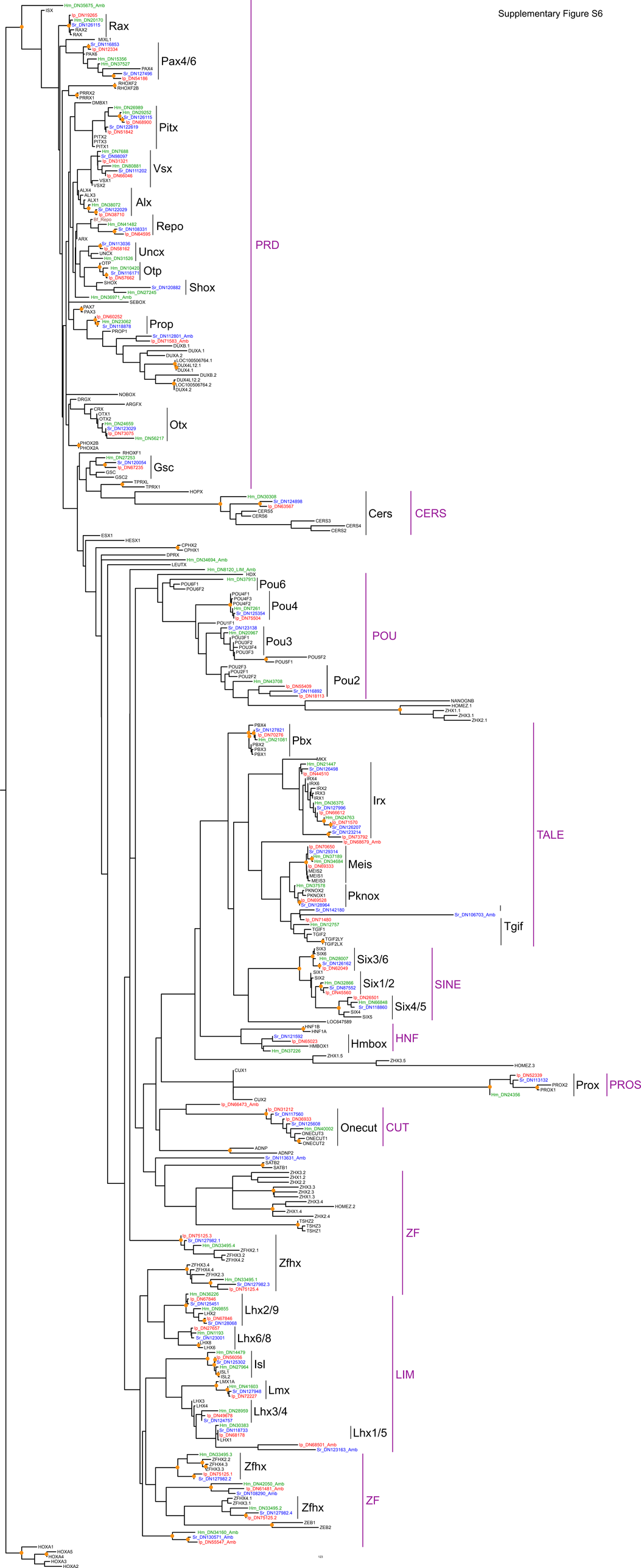
```
Sr MR.RTTISRSAKQVMF.WYLSN----G....ADG.SLLSSH.G-----------...ENVRVF.K.L.RLE..CE DN122932_c0_g2_i2|m.72688
Ip AR.RTTISKYSKKVMV.WYLNN----G....TEG.QLLSNH.G-----------...ENVRVF.K.L.RIE..GE DN65824_c0_g1_i1|m.85443
Hm IR.RTTIPKFAKKVMIDWYLGN----G....TEG.QLL.SQ.Q-----------...ENVRVF.            DN36591_c0_g1_i1|m.17917

Sr SS.H.CLFNSH.INL.KSL.EV----.AFPKKAELSIL.NH.R-----------KDRESLRK..DK..LNFRRQG DN113631_c0_g1_i3|m.42241
Ip MQKRREILPVQKKEH.IKV.NV----...PYPPQ.V.L.RE.K-----------.RQETVKV..R...YRL.HTK DN29418_c1_g1_i1|m.12421
Sr A..RKHILKFE.VSI..NYFHI----DSSWKMATVEKA.AE.G-----------.PVKKVYK.GYD.KHVSRNLG DN132449_c0_g1_i1|m.132246
Ip ...A.KSF.LQ.KSV.ST..SA----.QRPS.H.ISQL.SE.N-----------..SASVANF.A.S.QRN.HIS DN66473_c2_g1_i2|m.89822
Nw PTTVNLNLDSE.LSV..RAYKW----.KCPSAQEMVD.SNNTD-----------.STSYV.A..E..LSE.RRRE 39214.0
Nw RRHRTTFSEHAKSVLTQYFMTK----TH..SPTE.MSL.NQVN-----------.PVETVR.Y.K.LRS.N.NF 32393.0
Nw PSTAEMELSSH.LSV..SAYKL----.KCPDTDELSILSEKAN-----------MSTSLV.T.YE.KLSE.RRRE 44118.0
Xb KR.LRTVFSDT.INT..QA.RF----SP.IN.EV.VKLSEITG-----------.SDNA.Q.......ARQR.TG DN95135
Xb     RSRINWDPSAVNILEHHFNTKNRFPNAFERETIAANINGACQ                              DN85043
```

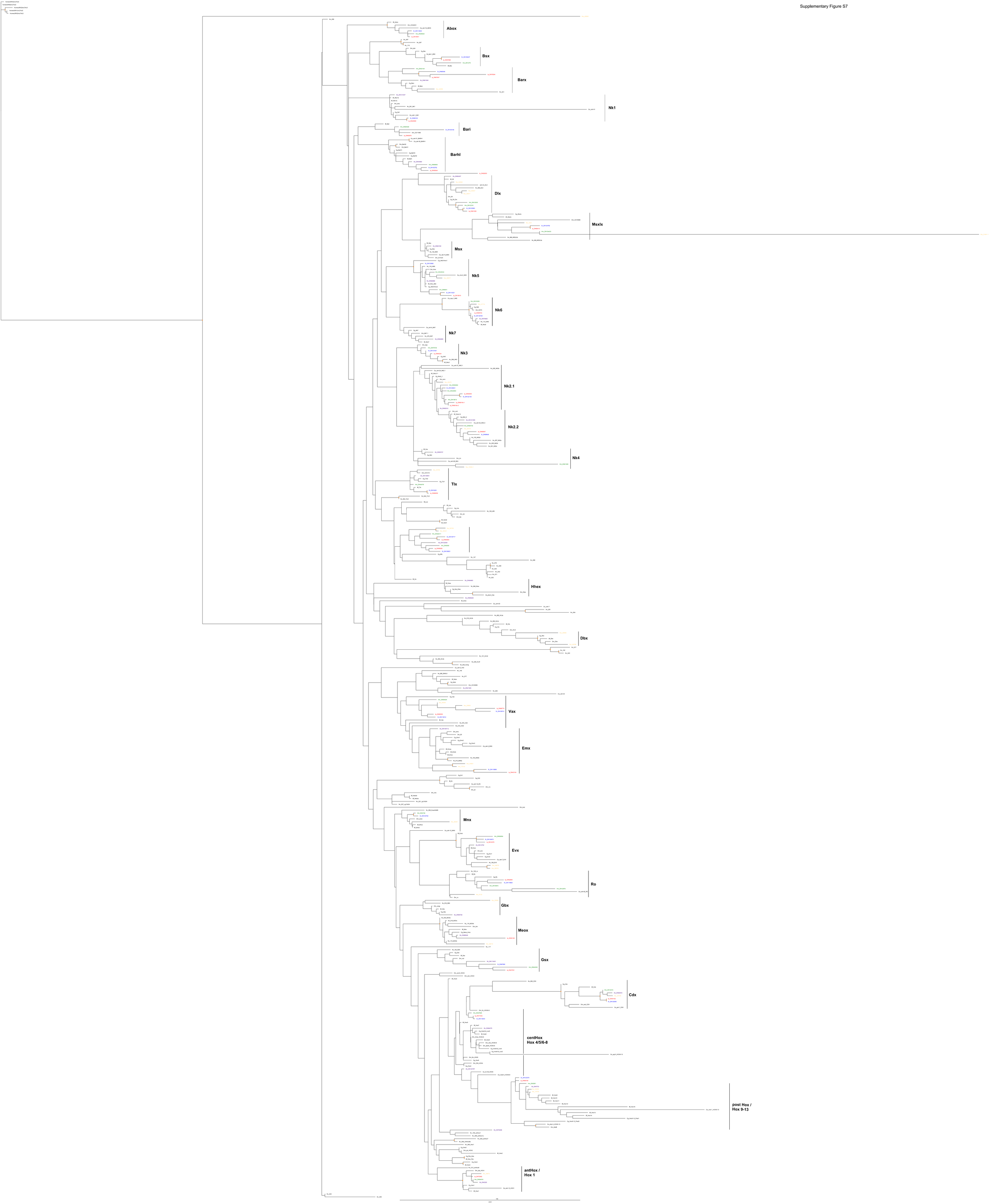
Supplementary Figure S5



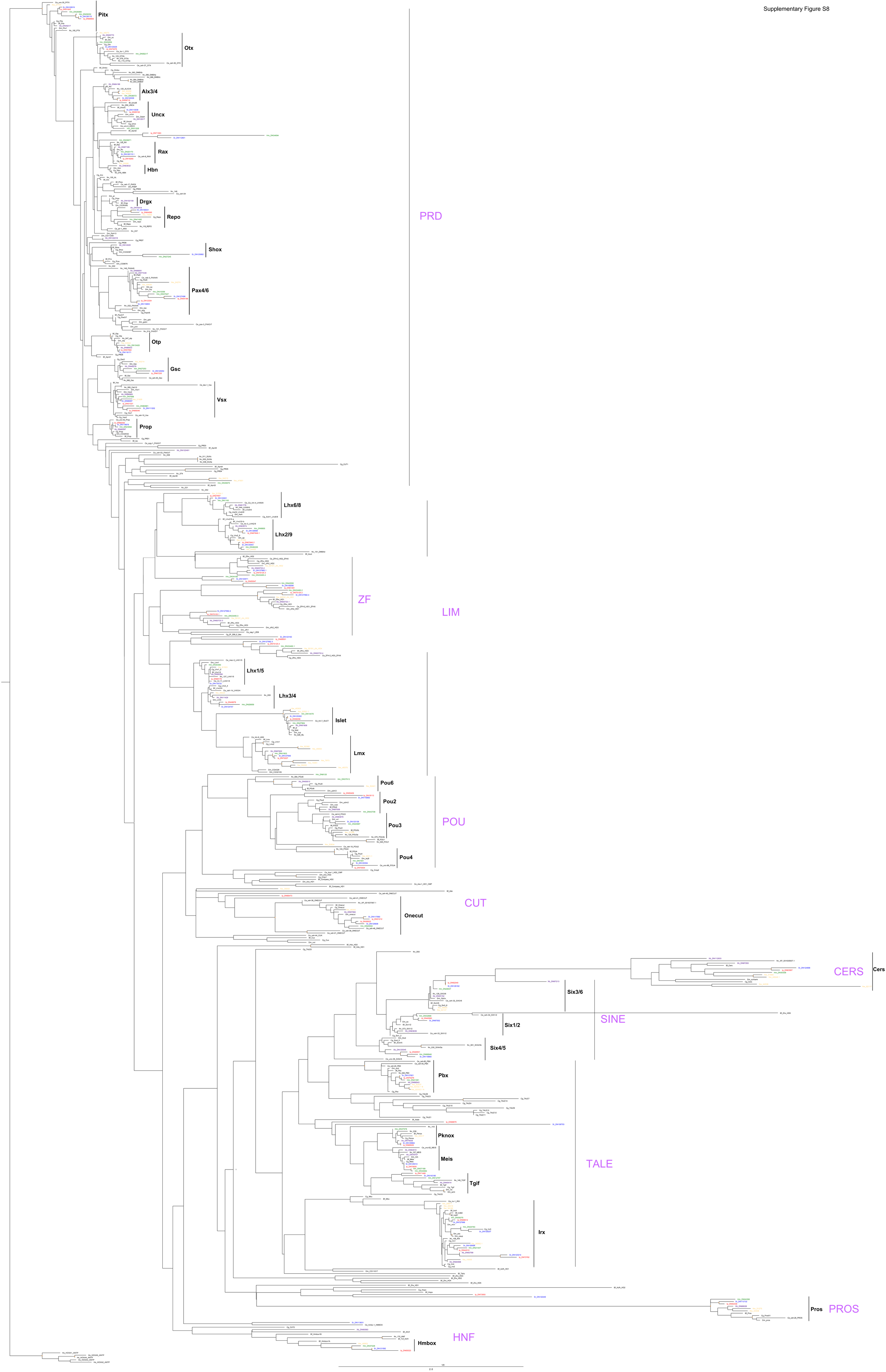
Supplementary Figure S6



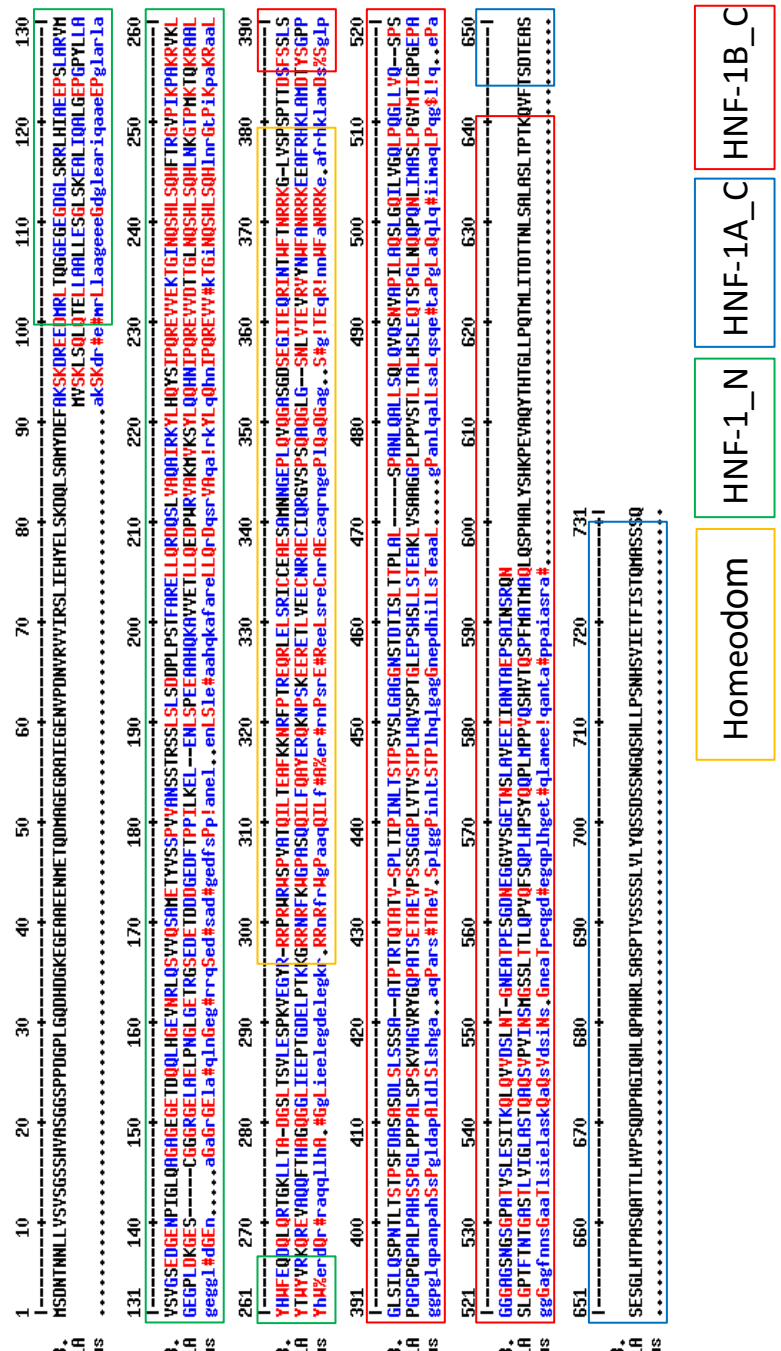
Supplementary Figure S7



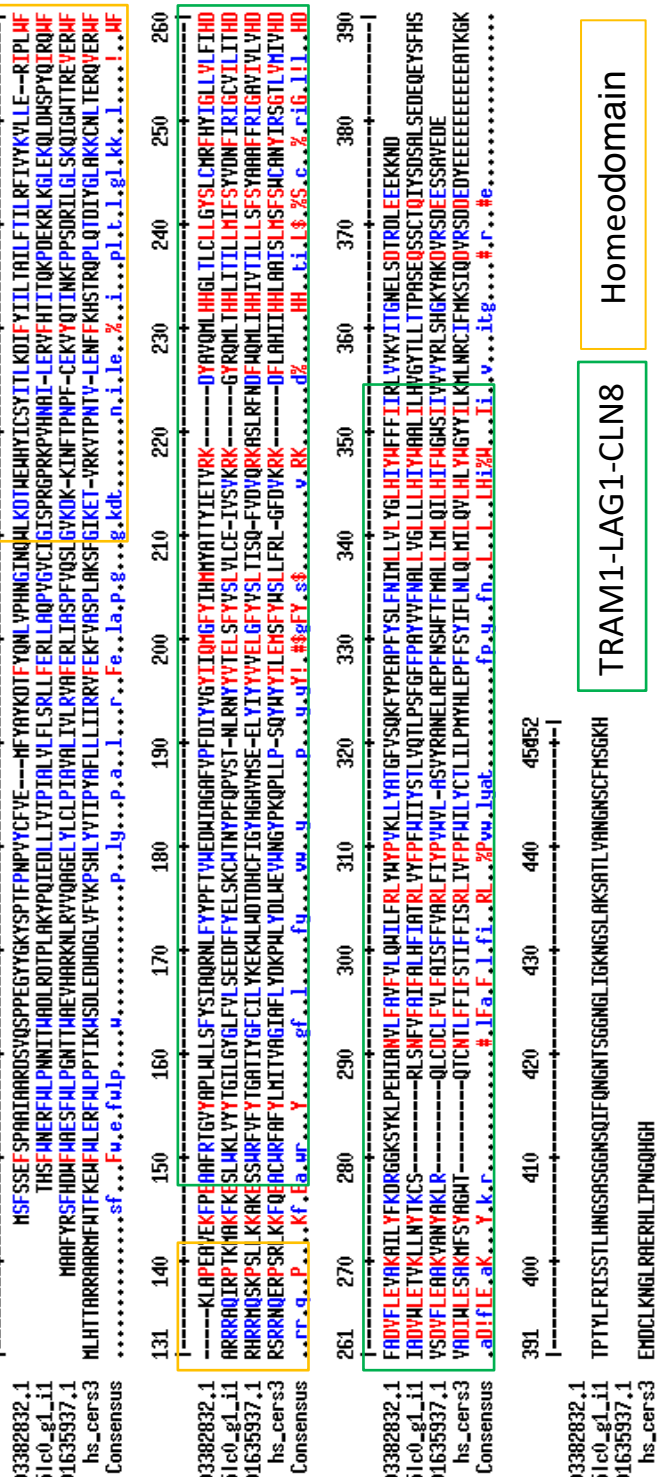
Supplementary Figure S8



Suppl. Figure S9: Human HNF1A vs Aq XP_003390473.1



Suppl. Figure S10: CERS class genes of Aq, Ip, Nv, Hs



Supplementary Table S11

ANTP HD sequences

```
>Ce_pal-1_CDX
ADKYRMVYSDYQRLELEKEFHTSPFITSDRKSQLSTMLSLTERQIKIWFQNRRAKDRRDK
>Ce_vab-7_EVX
MRRYRTAFSREQIGRLEREFAKENYVSRKTRGELAAELNLPEGTIKVWFQNRRMKDKRQR
>Ce_ceh-13_HOX1
NGTNRTNFTTHQLTELEKEFHTAKYVNRTRRTEIASNLKLQEAQVKIWFQNRRMKEKKRE
>Ce_lin-39_HOX5
EKRQRTAYTRNQVLELEKEFHTHKYLTRKRRIEVAHSLMLTERQVKIWFQNRRMKHKKEN
>Ce_mab-5_HOX6-8
SKRTRQTYSRSQTLELEKEFHYHKYLTRKRRQEISETLHLTERQVKIWFQNRRMKHKKEA
>Ce_nob-1_HOX9-13
GKKKRQPYKKDQISRLEYEYSVNQYLTNKRRSELSAQLMLDEKQVKVWFQVCYRQYSAES
>Ce_php-3_HOX9-13
MRKKRKPYTKAQTLELEKEFLYNTYVSKQKRWELAKYLHLTERQVKIWFQNRRMKDKKQK
>Ce_egl-5_HOX9-13
SKKGRQTYQRYQTSVLEAKFQQSSYVSKKQREELRLQTQLTDRQIKIWFQNRRMKAKKEK
>Ce_ceh-12_MNX
MRRPRTAFSSEQLVQLEKQFSDNRYLSRPRRYQLAQQLSLSETQIKIWFQNRRMKNKRCP
>Ce_ceh-62_RO
SRRKRTTFSPEQATRLEAEYIGDSYMAREKRHLLAQSLKLSENQVKTWFQNRRAKDKRDR
>Ce_ceh-19_ABOX
ERKPRQAYSARQLDRLETEFQTDKYLSVNKRIQLSQTLNLTETQIKTWFQNRRTKWKKQL
>Ce_ceh-30_BARH1
SRKARTIFTDKQLQELENTFEKQKYLSVQDRMDLAHRMGLTDTQVKTWYQNRRTKWKRQA
>Ce_ceh-31_BARH1
ARKARTIFTDKQLQELENTFEKQKYLSVQDRMELAHRMGLTDTQVKTWYQNRRTKWKRQA
>Ce_tab-1_BSX
RRKARTVFSDQQLQGLERRFESQRYLSTPERIELANALNLSETQVKTWFQNRRMKHKKVV
>ceh-43_DLX
MRKPRTIYNSSQLQMLQKKFQKTQYLALPDRAALAHELGLSQTQVKIWFQNRRSKQKKQK
>Ce_ceh-2_EMX
NKRIRTAFSASQLIQLEKAFEGNHYVVGNERKQLAAKLSLTETQVKVWFQNRRTKHKRVR
>Ce_vab-15_MSX
NRKPRTPFSTQQLISLERKFQSKQYLSIAERAEFSASLQLTETQVKIWFQNRRAKSKRLQ
>Ce_ceh-1_NK1
MRRARTAFTYEQLVALENKFKTSRYLSVVERLNLAIQLQLSETQVKIWFQNRRTKWKKHN
>Ce_ceh-24_NK2.1
RRKRRVLFSQAQVYELERRFKQAKYLTAPEREQLANSIRLTPTQVKIWFQNHRYKCKRQE
>Ce_ceh-27_NK2.1
RRKRRVLFSPQQVHVLERKFQINRYLSAADRENLAKSINLSATQVKIWFQNQRYKCKRQE
>Ce_ceh-22_NK2.2
KRKRRVLFTKAQTYELERRFRSQKYLSAPEREALAMQIRLTPTQVKIWFQNHRYKTKKSH
>Ce_ceh-28_NK4
KRKPRVLFTQHQVNELEERFKKQRYVTATEREELAQCLGLTATQVKIWFQNRRYKCKRLA
>Ce_mls-2_NK5
KKKTRTVFSRSQVSQLEMMFECKRYLSSQERSNLAQKLHLTETQVKIWFQNRRNKFKRQA
>Ce_cog-1_NK6
KKQSRPTFTGHQIYQLERKFEQTKYLAGADRAQLAQELNMSESQVKVWFQNRRTKWRKKE
>Ce_ceh-9_NK7
RKKARTTFSGKQVFELEKQFEAKKYLSSSDRSELAKRLDVTETQVKIWFQNRRTKWKKIE
>Ce_ceh-5_VAX
PKRPRTVFTDEQLEKLEESFNTSGYLSGSTRAKLAESLGLSDNQVKVWFQNRRTKQKKID
>Ce_ceh-51
RRGARTPFSDSQLYALRTRFEQCDTIKVDERRKLGAVIGLSPEQIKIWFQNRRFKLRKEK
```

```
>Ce_ceh-7
IPRRRTTFTVEQLYLLEMYFAQSQYVGCDERERLARILSLDEYQVKIWFQNRRIRMRREA
>Ce_ceh-16_EN
EKRPRTAFTGDQLDRLKTEFRESRYLTEKRRQELAHELGLNESQIKIWFQNKRAKLKKST
>Ce_ceh-23
HRKARTIYGTTQTQQLEDMFKGQMYVVGAERENLAQRLGLSPSQVRIWFQNRRSKHRRKQ
>Ce_pha-2_HEX
RKGGQIRFTNEQTDALEHKFDSHKYLSPQERKKLAKSLSLSERQVKTWFQNRRAKWRRVR
>Ce_ceh-63
RPTKRTTFTSEQVTLLELEFAKNEYICKDRRGELAQTIELTECQVKTWFQNRRTKKRRCT
>Dm_lab
NNSGRTNFTNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKKRV
>Dm_pb
PRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKHKRQT
>Dm_zen2
SKRSRTAFSSLQLIELEREFHLNKYLARTRRIEISQRLALTERQVKIWFQNRRMKLKKST
>Dm_zen
LKRSRTAFTSVQLVELENEFKSNMYLYRTRRIEIAQRLSLCERQVKIWFQNRRMKFKKDI
>Dm_bcd
PRRTRTTFTSSQIAELEQHFLQGRYLTAPRLADLSAKLALGTAQVKIWFKNRRRRHKIQS
>Dm_Dfd
PKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDN
>Dm_Scr
TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEH
>Dm_ftz
SKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKKDR
>Dm_Antp
RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
>Dm_Ubx
RRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLKKEI
>Dm_abdA
RRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEL
>Dm_AbdB
VRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLQLTERQVKIWFQNRRMKNKKNS
>Dm_eve
VRRYRTAFTRDQLGRLEKEFYKENYVSRPRRCELAAQLNLPESTIKVWFQNRRMKDKRQR
>Dm_ind
SKRIRTAFTSTQLLELEREFSHNAYLSRLRRIEIANRLRLSEKQVKIWFQNRRVKQKKGG
>Dm_cad
KDKYRVVYTDFQRLELEKEYCTSRYITIRRKSELAQTLSLSERQVKIWFQNRRAKERKQN
>Dm_btn
NRKERTAFSKTQLKQLEAEFCYSNYLTRLRRYEIAVALELTERQVKVWFQNRRMKCKRIK
>Dm_ro
QRRQRTTFSTEQTLRLEVEFHRNEYISRSRRFELAETLRLTETQIKIWFQNRRAKDKRIE
>Dm_Hmx
KKKTRTVFSRAQVFQLESTFDLKRYLSSSERAGLAASLRLTETQVKIWFQNRRNKWKRQL
>Dm_slou
PRRARTAFTYEQLVSLENKFKTTRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQN
>Dm_Cll/C15
RKKPRTSFTRIQVAELEKRFHKQKYLASAERAALARGLKMTDAQVKTWFQNRRTKWRRQT
>Dm_lbe
KRKSRTAFTNHQIFELEKRFLYQKYLSPADRDEIAASLGLSNAQVITWFQNRRAKQKRDI
>Dm_lbl
KRKSRTAFTNQQIFELEKRFLYQKYLSPADRDEIAGGLGLSNAQVITWFQNRRAKLKRDM
>Dm_bap
KKRSRAAFSHAQVFELERRFAQQRYLSGPERSEMAKSLRLTETQVKIWFQNRRYKTKRKQ
```

```
>Dm_tin
KRKPRVLFSQAQVLELECRFRLKKYLTGAEREIIAQKLNLSATQVKIWFQNRRYKSKRGD
>Dm_Dr/msh
NRKPRTPFTTQQLLSLEKKFREKQYLSIAERAEFSSSLRLTETQVKIWFQNRRAKAKRLQ
>Dm_BarH1
QRKARTAFTDHQLQTLEKSFERQKYLSVQERQELAHKLDLSDCQVKTWYQNRRTKWKRQT
>Dm_BarH2
QRKARTAFTDHQLQTLEKSFERQKYLSVQDRMELANKLELSDCQVKTWYQNRRTKWKRQT
>Dm_bsh
RRKARTVFSDPQLSGLEKRFEGQRYLSTPERVELATALGLSETQVKTWFQNRRMKHKKQL
>Dm_CG11085
PRRRRTAFTHAQLAYLERKFRCQKYLSVADRSDVAETLNLSETQVKTWYQNRRTKWKRQN
>Dm_vnd
KRKRRVLFTKAQTYELERRFRQQRYLSAPEREHLASLIRLTPTQVKIWFQNHRYKTKRAQ
>Dm_scro
RRKRRVLFTQAQVYELERRFKQQRYLSAPEREHLASLIHLTPTQVKIWFQNHRYKCKRQA
>Dm_Dbx
GMMRRAVFSDSQRKGLEKRFQQQKYISKPDRKKLAERLGLKDSQVKIWFQNRRMKWRNSK
>Dm_lms
KKRPRTAFSAAQIKALETEFERGKYLSVAKRTALAKQLQLTETQIKIWFQNRRTKWKRKY
>Dm_CG15696
GRLPRIPFTPQQLQALENAYKESNYLSAEDANKLADSLELTNTRVKIWFQNRRARERREK
>Dm_CG18599
NKRVRTIFTPEQLECLEAEFERQQYMVGPERLYLAHTLKLTEAQVKVWFQNRRIKWRKHH
>Dm_CG34031
DRKPRQAYSASQLERLENEFNLDKYLSVSKRVELSKSLSLTEVQVKTWFQNRRTKWKKQL
>Dm_HHEX
RKGGQIRFTSQQTKNLEARFASSKYLSPEERRHLALQLKLTDRQVKTWFQNRRAKWRRAN
>Dm_Dll
MRKPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQVKIWFQNRRSKYKKMM
>Dm_E5
PKRVRTAFSPTQLLKLEHAFEGNHYVVGAERKQLAQGLSLTETQVKVWFQNRRTKHKRMQ
>Dm_ems
PKRIRTAFSPSQLLKLEHAFESNQYVVGAERKALAQNLNLSETQVKVWFQNRRTKHKRMQ
>Dm_exex
TRRPRTAFTSQQLLELEKQFKQNKYLSRPKRFEVASGLMLSETQVKIWFQNRRMKWKRSK
>Dm_HGTX
KKHTRPTFSGQQIFALEKTFEQTKYLAGPERAKLAYALGMSESQVKVWFQNRRTKWRKRH
>Dm_H2.0
RSWSRAVFSNLQRKGLEIQFQQQKYITKPDRRKLAARLNLTDAQVKVWFQNRRMKWRHTR
>Dm_unpg
SRRRRTAFTSEQLLELEREFHAKKYLSLTERSQIATSLKLSEVQVKIWFQNRRAKWKRVK
>Dm_NK7.1
KKKARTTFTGRQIFELEKMFENKKYLSASERTEMAKLLMVTETQVKIWFQNRRTKWKKQD
>Dm_en
EKRPRTAFSSEQLARLKREFNENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKST
>Dm_inv
DKRPRTAFSGTQLARLKHEFNENRYLTEKRRQQLSGELGLNEAQIKIWFQNKRAKLKKSS
>Cg_Cdx
DKYRVVYSDHQRLELEKEFHYSRYITIRRKAELAQNLALSERQVKIWFQNRRAKERK
>Cg_Evx1
RRYRTAFTKEQLNRLEKEFLKENYVSRPKRCELAAQLNLSESTIKVWFQNRRMKDKR
>Cg_Evx2
RRYRTAFTREQIGRLEKEFYKENYVSRPRRCELAKLLNLPENTIKVWFQNRRMKDKR
>Cg_Gbx
RRRRTAFTSEQLLELEKEFHSKKYLSLTERSHIAHNLKLSEVQVKIWFQNRRAKWKR
```

```
>Cg_Gsx
KRIRTAFTSTQLLELEREFNSNMYLSRLRRIEIATYLNLSEKQVKIWFQNRRVKFKK
>Cg_Hox1
NMGRTNFTNKQLTELEKEFHFNKYLTRARRIEIAAALGLNETQVKIWFQNRRMKQKK
>Cg_Hox2
RRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIAASLDLTERQVKVWFQNRRMKYKR
>Cg_Hox3
KRARTAYTSAQLVELEKEFHFNRYLCRPRRIEMAALLSLTERQIKIWFQNRRMKFKK
>Cg_Hox4
KRNRTAYTRHQILELEKEFHFNRYLTRRRRIEIAHTLCLSERQIKIWFQNRRMKWKKE
>Cg_Hox5
KRSRTSYTRHQTLELEKEFHYNKYLTRRRRIEIAHALNLTERQIKIWFQNRRMKWKKE
>Cg_Hox6-8_Lox2
RRGRQTYTRFQTLELEKEFKFNRYLTRRRRIELSHMLCLTERQIKIWFQNRRMKEKK
>Cg_Hox6-8_Lox4
RRGRQTYSRYQTLELEKEFQFNHYLTRKRRIEVAHSLCLTERQIKIWFQNRRMKLKK
>Cg_Hox6-8_Lox5
KRTRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHLLGLTERQIKIWFQNRRMKWKK
>Cg_Hox9-13_Post1
RKRRRPYSKFQIAELEREYNNSTYISKSRRWELSQLINLSERQIKIWFQNRRIKAKK
>Cg_Hox9-13_Post2
RKKRKPYTRYQTMVLENEFLNSSYITRQKRWEISCKLQLSERQVKVWFQNRRMKRKK
>Cg_Meox_Hrox
RKERTAFTKHQIRELEKEFAVHNYLTRLRRYEIAVALDLTERQVKVWFQNRRMKWKR
>Cg_Pdx_Xlox
KRTRTAYTRGQLLELEKEFHFNKYISRPRRIELAAMLNLTERHIKIWFQNRRMKWKK
>Cg_BarH1
RKARTAFTDHQLNSLEKSFERQKYLSVQDRMELASKLNLTDTQVKTWYQNRRTKWKR
>Cg_BarH2
RKARTAFTDHQLNCLEKSFERQKYLSVQDRMELAAKLNLTDTQVKTWYQNRSHDLLK
>Cg_BarH3
RKARTAFTDQQLNNLEKHFARQKYLSVQDRMELAAKLGLSDTQVKTWYQNRRTKWKR
>Cg_Barx
RRSRTVFTELQLIGLEKRFESQKYLSTPDRIDLAESLGLSQVQVKTWYQNRRMKWKK
>Cg_Bsx
RKARTVFSDQQLNGLEKRFEAQRYLSTPERVELANQLSLSETQVKTWFQNRRMKHKK
>Cg_Dbx
MLRRAVFSDMQRKGLEKMFQKQKYISKPDRKKLAAKLGLKDSQVKIWFQNRRMKWRN
>Cg_Dll_Dlx
KPRTIYSSLQLQQLNRRFQRTQYLALPERAELAASLGLTQTQSVKIWFQNRRSKAKK
>Cg_Emx1
KRIRTAFSPTQLLQLEHAFEKNHYVVGQERKELALKLGLSETQVKVWFQNRRTKYKR
>Cg_Emx2
KRIRTAFSPSQLLKLEKVFERNHYVVGQERKDLATNLNLSETQVKVWFQNRRTKYKR
>Cg_Emx3
KRIRTAFSPSQLLQLEKAFEKSHYVVGQERKDLASELQLTETQVKVWFQNRRTKHKR
>Cg_En1
KRPRTAFTTDQLQRLKSEFEENRYLTEKRRLELSEELKLSESQIKIWFQNKRAKIKK
>Cg_En2
KRPRTAFSNEQLQRLKVEFEKCQYLTEQRRLELAKTLNLSEGQIKIWFQNKRAKVKK
>Cg_Hex_HHEX
KGGQVRFSNDQTVELEKKFESHKYLSPPERKRLAKTLQLTERQVKTWFQNRRAKWRR
>Cg_Hlx
SWSRAVFSNLQRKGLEKRFEVQKYVTKPDRRQLAAMLGLTDAQVKVWFQNRRMKWRH
>Cg_Lbx
RKSRTAFTNQQIYELEKRFLYQKYLTPADRDEIAQNLGLSNAQVITWFQNRRAKLKR
```

```
>Cg_Msx
RKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKR
>Cg_Msxlx
NPRVPFTQHQVAVLEHKFRRTHYLSSMDVAELSTALSLTENRVSSVSA
>Cg_Nk1
RRARTAFTYEQLVALENKFKTTRYLSVCERLNLALSLNLTETQVKIWFQNRRTKWKK
>Cg_Nkx2_1
RKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMINLTPTQVKIWFQNHRYKCKR
>Cg_Nk2_2
RKRRVLFSKAQTYELERRFRQQRYLSAPEREHLASIIRLTPVQVKIWFQNHRYKLKR
>Cg_Nk3
KRSRAAFSHAQVFELERRFRHQRYLSGPERADLANALKLTETQIKIWFQNRRYKTKR
>Cg_Nk4
RKPRVLFSQAQVFELERRFKQQRYLSAPEREQLANLLKLTSTQVKIWFQNRRYKCKR
>Cg_Nk5/Hmx1
KKTRTVFSRHQIFYLESAFDAKRYLSSTERSEIASTLNLTETQVKVWFQNRRNKWKS
>Cg_Nk5/Hmx2
KKTRTVFSRSQVFQLESTFDMKRYLSSSERAGLAASLQLTETQVKIWFQNRRNKWKR
>Cg_Nk6
KHTRPTFSGQQIFALEKTFEQTKYLAGPERARLAYALGMSESQVKVWFQNRRTKWRK
>Cg_Nk7
KKARTTFTGRQIFELEKQFEQKKYLSSAERAEMATLLNVTETQVKIWFQNRRTKWKK
>Cg_Noto
KRVRTIFTPEQLERLEAEFEKQQYMVGTERYYLACSLNLTEAQVKVWFQNRRIKWRK
>Cg_Ro
RRQRTTFSSDQTVKLEIEYSRTEYITRSRRYELAELLCLTENQIKIWFQNRRAKEKR
>Cg_Tlx1
KKPRTSFSRMTIIELEKRFDRQKYLASSERTSLAKSLKISDSQVKTWFQNRRTKWRR
>Cg_Tlx2
KKPRTSFSRLQIMELEKRFHRQKYLASAERSTLAKALKMTDAQVKTWFQNRRTKWRR
>Cg_Vax
KRARTTFSPEQLYQLEKEFQRNQYLVGKERTELAKLLKLSETQNNITFE
>Cg_NKL
KKARTAFSTEQIHDLEKRYQAQKYLPANERQALAEKLGLSDQQVKTWFQNRRMKEKR
>Bf_Abox
DRKPRQAYSSRQLERLEEEFKKDKYLSVSKRVELAESLELTEIQVKTWFQNRRTKWKKQT
>Bf_Ankx
RRRQRRLFSGHQVMELETRFLEKNYLSRIERICLANALNLSETQVKTWFQNRRTKARRKH
>Bf_Barh
PRKARTAFTDHQLAQLERSFERQKYLSVQDRMELAASLNLTDTQVKTWYQNRRTKWKRQN
>Bf_Bari
QRRRRTVFTQSQLRALERKFRCQKYLSVSERAEVARALGLSETQIKTWYQNRRTKWKRET
>Bf_Barx
CRRSRTVFTELQLLGLEKRFEKQKYLSTPDRVELAESLGLSQLQVKTWYQNRRMKWKKMV
>Bf_Bsx
RRKARTVFSDSQLNGLEKRFESQKYLSTPERMDLAAALNLTETQVKTWFQNRRMKHKKLQ
>Bf_Cdx
KDKYRVVYSDHQRLELEKEFYSNKYITIKRKVQLANELGLSERQVKIWFQNRRAKQRKMA
>Bf_Csx
RRKPRVLFSQAQVFELERRFKQQRYLSAPEREQLAQMLKLTSTQVKIWFQNRRYKCKRQR
>Bf_Dbx
GMLRRAVFSDFQRKSLEKMFQKQKYISKQDRRKLAEKLGLKDSQVKIWFQNRRMKWRNSK
>Bf_Dll
MRKPRTIYTSFQLQQLNRRFQRTQYLALPERAELAAQLGLTQTQVKIWFQNRRSKYKKLM
>Bf_Emxa
PKRVRTAFTPTQLLRLEHAFEKNHYVVGQERKQLAQQLTLSETQVKVWFQNRRTKYKRDQ
```

```
>Bf_Emxb
PKRIRTAFTPSQLLRLEHAFEKNHYVVGQERKHLAQSLSLTETQVKVWFQNRRTKHKREQ
>Bf_Emxc
PKRIRTAFTPSQLLRLEHAFEKNHYVVGQERKHLAQSLSLTETQVKVWFQNRRTKYKRDQ
>Bf_En
EKRPRTAFTSEQLQRLKKEFQENRYLTEQRRQDLARELKLNESQIKIWFQNKRAKIKKAA
>Bf_Evxa
VRRYRTAFTREQLARLEKEFYRENYVSRPRRCELAAQLNLPETTIKVWFQNRRMKDKRQR
>Bf_Evxb
VRRYRTAFSREQTARLEKEFHRDNYLSRPRRCELAAALNLPETTIKVWFQNRRMKEKRQR
>Bf_Gbx
TRRRRTAFTSEQLLELEKEFHSKKYLSLTERSQIAHALKLSEVQVKIWFQNRRAKWKRVK
>Bf_Hhex
RKGGQVRFSNDQTLELEKKFESQKYLSPPERKRLAKSLQLTERQVKTWFQNRRAKWRRLK
>Bf_Hlx
RTWSRAVFSNLQRKGLEKRFDIQKYVTKPDRKQLASMLGLTDAQVKVWFQNRRMKWRNAM
>Bf_Gsx
SRRMRTAFSSTQLLELEREFASNMYLSRLRRIEIATFLNLSEKQVKIWFQNRRVKHKKEA
>Bf_Hmx_Nk5
KKKTRTVFSRQQVFQLESTFDMKRYLSSSERAGLAASLHLTETQVKIWFQNRRNKWKRQL
>Bf_Hox1
PNNGRTNFTTKQLTELEKEFHYNKYLTRARRVEIAAALNLNETQVKIWFQNRRMKQKKRE
>Bf_Hox10
GRKKRCPYTKYQILELEKEFLFNMYVSRERRQEISRHVNLSDRQVKIWFQNRRMKMKRMN
>Bf_Hox11
TRKKRCPYTKYQTLELEKEFLFNMFVTRERRQEIARQLNLTDRQVKIWFQNRRMKMKRMK
>Bf_Hox12
SRKKRCPYSKVQLLELEKEFLYNMYITREQRGEIARKVNLTDRQVKIWFQNRRMKMKRMK
>Bf_Hox13
GRKKRCPYTKYQLSVLEQEYIQNRYVSRETRLELSQRLNLTDRQVKIWFQNRRMKQKRLE
>Bf_Hox14
VRPKRRPYSKYQLNELENEYVQNQYISRDKRLQLSQKLNLTERQVKIWFQNRRIKQKKLD
>Bf_Hox15
TRKKRRPYSKPQLALLEDEYASQKFLTKEKRKEISESSSLSERQVMIWFQNRRMKEKKLA
>Bf_Hox2
SRRLRTVFTNTQLLELEKEFHYNKYVCKPRRKEIASYLDLNERQVKIWFQNRRMRQKRRD
>Bf_Hox3
GKRARTAYTSAQLVELEKEFHFNRYLCRPRRVEMAAMLNLTERQIKIWFQNRRMKYKKEQ
>Bf_Hox4
TKRSRTAYTRQQVLELEKEFHFNRYLTRRRRIEIAHSLGLTERQIKIWFQNRRMKWKKDN
>Bf_Hox5
NKRTRTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
>Bf_Hox6
KKRGRQTYTRYQTLELEKEFHFNKYLTRKRRIEIAHLLGLTERQIKIWFQNRRMKWKKEN
>Bf_Hox7
RKRGRQTYTRYQTLELEKEFHFNKYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
>Bf_Hox8
RRRGRQTYSRYQTLELEKEFHFNKYLTRRRRIEIAHALGLTERQIKIWFQNRRMKLKKEA
>Bf_Hox9
SRKKRCPYTRFQTLELEKEFLYNMYLTRERRYEISQHVNLTERQVKIWFQNRRMKMKKMS
>Bf_Hx
KNKARTAFAGYQVFQLEKRFQTQKYLAAAERQELAQRIGLTDTQVKTWFQNRRMKWKRQQ
>Bf_Lbx
RRKSRTAFTNHQIYELEKRFLYQKYLSPADRDQIAQSLGLTNAQVITWFQNRRAKLKRDL
>Bf_Lcx
SKKLRTAFSSHQVHELETRFSTQKYLSASDREELSQALDLTDAQVKTWFQNRRMKWKRQV
```

```
>Bf_Mnxa
TRRPRTAFTSQQLLELEKYFKENKYLSRPKRFEVATALMLTETQVKIWFQNRRMKWKRSK
>Bf_Mnxb
TRRPRTAFTSQQLLELEKHFKENKYLSRPKRFEVATSLMLTETQVKIWFQNRRMKWKRSK
>Bf_Mox
PRKERTAFTKQQIMELENEFRHHNYLTRLRRYEIAVKLDLTERQVKVWFQNRRMKWKRTK
>Bf_Msx
NRKPRTPFTTQQLLALERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKRLQ
>Bf_Msxlx
GRNPRVPFSAHQVATLESKFRRTHYLSSVDVSELSVALNLSENRVKIWFQNRRARERRER
>Bf_Nedxa
KKRPRTAFTAEQIKELEGEFQKNKYLSVTKRLELSNQLKLTETQIKIWFQNRRTKWKRKF
>Bf_Nedxb
KKRPRTAFTAEQIKELEGEFQKNKYLSVTKRLELSNQLKLTETQIKIWFQNRRTKWKRKY
>Bf_Nkx1a
PRRARTAFTYEQLVALENKFKQTRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQN
>Bf_Nkx1b
PRRVRTAFTYEQLVALENKFKQTRYLSVCERLNLALSLSLTETQVKIWFQNRRTKWKKQN
>Bf_Nkx2-1
RRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLAQLINLTPTQVKIWFQNHRYKCKRQD
>Bf_Nkx2-2
KRKRRVLFSKAQTYELERRFRQQRYLSAPEREHLARLIRLTPTQVKIWFQNHRYKCKRAQ
>Bf_Nkx3
KKRSRAAFSHAQVFELERRFSHQRYLSGPERADLAAALKLTETQVKIWFQNRRYKTKRRQ
>Bf_Nkx6
RKHTRPTFSGQQIFALEKTFEQTKYLAGPERARLAYSLGMTESQVKVWFQNRRTKWRKKH
>Bf_Nkx7
KKKARTTFTGRQIFELEKQFEVKKYLSASERADLAALLNVTDTQVKIWFQNRRTKWKKQD
>Bf_Noto
GKRIRTIFTPEQLERLEREFSRQQYMVGSERYHLAASLNLTEAQVKVWFQNRRIKWRRQN
>Bf_Tlx
RKKPRTSFTRLQICELEKRFHKQKYLASAERAQLATQLKMTDAQVKTWFQNRRTKWRRQT
>Bf_Vax
PKRARTAFTADQLKDLERAFQGSQYVVGQDRTELARQLRLTETQVKVWFQNRRTKYKREK
>Bf_Vent1
GRKARTAFTTEQVMALEERFRLQKYLSAADRETLAKATGLTDEQVKTWFQNRRMKLKRQQ
>Bf_Vent2
GRKARTAFTTEQVMALEERFRLQKYLSAADRETLAKATGLTDEQVKTWFQNRRMKLKRQQ
>Bf_Ro
QRRQRTTFTTEQTMKLELEYNRTEYITRPRRFELAEMLNLSETQIKIWFQNRRAKDKRIE
>Bf_Xlox_Pdx
NKRTRTAYTRGQLLELEKEFHFNKYISRPRRIELAAMLNLTERHIKIWFQNRRMKWKKEQ
>Nv_065_CDX
RSRARTAYTASQQLELEKEFLYSRYITRTRRKELANTLDLSEKHIKIWFQNRRMKKKKTD
>Nv_109_EVX
TRRYRTAFTREQLKRLEKEFMRENYVSRTRRCELANALNLSETTIKIWFQNRRMKSKRRR
>Nv_069_ExexHlxB9
HRRPRTAFSSHQLLALERQFQLHKYLTRPQRYELATSLMLTETQVKIWFQNRRMKWKRCN
>Nv_018_GBX
RRRKRTAFTSKQLLQLEREFHNKKYVSLEERSVIATNLNLTEVQVKIWFQNRRAKWKRVR
>Nv_130_GSX
SKRIRTAYTSMQLLELEKEFSQNRYLSRLRRIQIAALLDLSEKQVKIWFQNRRVKWKKDK
>Nv_012_anthox6
SQKKRFTFTQRQLVELEKEFHFSKYLTRTRRIEIATTLKLTEMQIKIWFQNRRMKWKREF
>Nv_060_Hox1
SDKNRTIYSTRQLVELEKEFHYNRYLCRPRRIEIAQSLELTEKQVKIWFQNRRMKWKKEN
```

>Nv_029_anthox7
TKRYRTSYTNRQLLELEKEFHYNKYLCGTRRRELANAMKLTERQVKVWFQNRRMKLKKDE
>Nv_006_Anthox8a
SKRHRTSYTNKQLLELEKEFHFNKYLCSSRRREISKALQLTERQVKIWFQNRRMKWKKDE
>Nv_034_MOXa
KRKERTAFTTHQLRELENEFTRNNYLTRLRRYEIAVSLDLTERQVKVWFQNRRMKWKRVK
>Nv_110_MOXb
KRKERTAFTKHQIQELENEFQRNNYLTRLRRYEIAVSLDLTERQVKVWFQNRRMKWKRVK
>Nv_019_MOXc
KRKERTAFSKHQLQELENEFIRNNYLTRLRRYEIAVSLDLTERQVKVWFQNRRMKWKRVK
>Nv_119_MOXd
KRKERTVFSKYQLTELEREFCRNNYLTRLRRYEIAVSLDLGERQVKVWFQNRRMKWKRGR
>Nv_126_ro
PRRQRTTFTSEQTLKLELEFHQNEYITRSRRFELAACLNLTETQVKIWFQNRRAKDKRLE
>Nv_099_anthox1a
KHRKRMAYTRIQLLELEKEFHFTRYLTKERRTEMARMLDLTERQVKIWFQNRRMKWKKDN
>Nv_106_anthox1
GKRKRTAYTRKQLLELEKEFHFNHFLTKERRSEMASQLNLTERQVKIWFQNRRMKWKKCN
>Nv_117
NHAPRSAFTTVQQLEIEKEFLYDHYVSRVRRIEIVMALDLSEKQVRTWFQNRRMKLKREA
>Nv_021_cg13424
KKRTRTAFTNTQIRELEAEFQKNKYLTITRRAELANSLELSDTQIKIWFQNRRTKLKRKL
>Nv_027_cg13424
KKRPRTAFTNEQIKDLEAEFQKSKYLSVSRRMELANSLSLTETQIKIWFQNRRTKWKRKM
>Nv_095_DLX
IRKPRTIYSSFQLRELNKRFIKTQYLALPERADLAAYLGLTQTQVKIWFQNRRSKFKKTL
>Nv_013_EMXa
PKRIRTAFTPTQLLHLENAFEKNHYIVGTERKQLASYLNLSETQIKVWFQNRRTKWKRQQ
>Nv_104_EMXb
VRRVRTAFTPFQLLCLETSFDKNHYVVGTERKQLASYLKLSETQVKVWFQNRRTKWKRQA
>Nv_080_EMXLX
PKRIRTIFTPEQLERLEKEFERQQYMVGAERHYLAASLNLTETQVKVWFQNRRIKWRKQK
>Nv_077
PKRVRTIFSPDQLDRLEREFDNQQYIVGTERFYLATELGLTETQVKVWFQNRRIKWRKQN
>Nv_048_HHEX
KKGGQVRFSNEQTMELEKIFETQKYLSPPERKQLSKVLGLSERQVKTWFQNRRAKWRRFK
>Nv_082_HLXa
PPRPRPVFSHFQRQHLENYFLSKKYLNKTERLKIAEALGLSENQVKIWFQNRRVKWRAEV
>Nv_015_HLXb
RPWSRPVFTQLQRRGLEKRFQVQKYITKHDRYQLAAMLGLSETQVKVWFQNRRTKWRQDM
>Nv_092_HLXc
RPWTRAVFSNLQRKGLEKRFQVQKYLTKADRHQLASMLGLTDNQVKVWFQNRRMKWRQDA
>Nv_121_HLXd
KQKQRPLFGQKQIQRLEEEFVCEKYISKSRRSALAAEIDLTDAQVRTWFQNRRTRWRKEE
>Nv_094_HLXe
KPKQRPLFSHNQIQRLEKEFKEEKYLTESKRAELSKDLGMTETQVKTWFQNRRTKWRKEL
>Nv_049_HLXf
PTKLRPLFSKTQLSRLEKEFSGNKYLTESKREQLSKDLGMTETQVKTWFQNRRTKWRKKK
>Nv_035_HLXg
RPKLRPLFSKTQLSRLEKEFEEKRYLTETKRAELSKDLEMTETQVKTWFQNRRTKWRKEK
>Nv_116_HMX
KKKTRTVFSRSQVYQLESTFDMKRYLSSSERAGLASQLHLTETQVKIWFQNRRNKWKRQI
>Nv_122_LBX
RRKPRTCFTNRQIIELERRFMYQKYLSPSDRDDIAMALGLPGAQIITWFQNRRAKMRRDV
>Nv_142_MSX
NRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSASLNLTETQVKIWFQNRRAKAKRLH

>Nv_046_MSXLXa
SGNPRVPFSTSQLMELEHKYVETRYLSGPEVAELANGLRLSEHRVKIWFQNRRAREKKAK
>Nv_066_MSXLXb
DRNPRIPFSASQLAALEAKFLDTHYLSSVEVRDLSSSLSVTEHRVKIWFQNRRAREKKTK
>Nv_091_NK1
PRRARTAFTYEQLVALENKFKSTRYLSVCERLNLALSLGLTETQVKIWFQNRRTNGRSRT
>Nv_057_NK2a
KRKRRILFTKAQTFILERRFTQQRYLSAPEREELARIANLTPAQVKIWFQNHRYKYRKQM
>Nv_025_NK2b
KKKRRILFTKSQIFELERRFRQQKYLSANEREQLARIIDLTPTQVKIWFQNHRYKFKKQI
>Nv_051_NK2c
KRKRRILFTKSQIYELERRFRYQRYLSAPEREQLGRMINLTPTQVKIWFQNHRYKHKKQA
>Nv_103_NK2d
KRKRRVLFTKAQTYELERRFRQQRYLSAPEREQLARIINLTPTQVKIWFQNHRYKFKKQN
>Nv_026_NK2e
KRKQRVLFSKVQRKQLEKKFLEKPYLSYTERDQLARAVNLTAKQVKVWFQNKRYKTKERE
>Nv_098_NK3
RKRSRAAFSHQQVFELERRFSHQKYLSGPERADLAAALKLTETQVKIWFQNRRYKTKRRQ
>Nv_112_NK6
RKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAYSLGMTESQVKVWFQNRRTKWRKRH
>Nv_070_NK7
KKKARTTFTGRQIFELEKQFEAKKYLTATERSDMASLLNVTETQVKIWFQNRRTKWKKQE
>Nv_004_TLXl
KKKPRTAFTESQISELEKRFQSQKYLGSKERSELAGTLGLTDTQV
>Nv_024_TLXl
AKKPRTAFTESQISELEKRFQVTEIPREQERSELAGTLGLTDTQVKTWFQNRRMKLKRQR
>Nv_072_VAX
PKRSRTSFTPAQLDRLEDEFRVDMYVVGLKRMKLANDLNLSERQVKVWFQNRRMKYKRER
>Nv_010_VAXl
RKRARTAFSAEQLKKLEKRFQANHYIVGEERQKIAKDLDLSEAQVKVWFQNRRTKFKRDQ
>Nv_067
CRKSRTVFTDLQLRVLEKTFSEQKYLDSTNRAKLAQILGLNEAQVKTWFQNRRMKWKKKN
>Nv_097
CRKSRTVFTDLQLRVLEKTFSEQKYLDTSSRSKLAQTLGLNETQVKTWFQNRRMKWKKET
>Nv_115
CRKSRTVFTDLQLRVLEKTFSEQKYLDTSSRAKLAQTLGLNETQVKTWFQNRRMKWKKES
>Nv_009
TRPRRSFFSADQVNQLERVFSAQQYVSAKERAEIAETFNMTDEQVKNWFQNRRMKRKRKR
>Nv_017
ARNKRPVFHPDVVAQLEQVFDERRYISADQRFALAKELNMTEEQVKSWFHNKRTAMTKKL
>Nv_023
TKPRRSFFSADQVNQLERVFTAQQYVSAKERAEIAETFNMTDDQVKNWFQNRRMKRKRKR
>Nv_032
TKPRRSFFSADQVNQLERVFAAQQYVSAKERAEIAETLNMTDDQVKIWFQNRRMKRKRKR
>Nv_042
AKPRRSFFSADQVSQLERVFAARQYVSAKERAEIAETLNMTDDQVKIWFQNRRVKRKRKR
>Nv_043
QPRKRPVFHPDVVARLEQFFAKRRYINAEQQYELAKEINMTEEQIKSWLHNKRTAMKRKL
>Nv_071
TKPRRSFFSADQVNQLERVFAAHQYVSANERAEIAETLNMSDDQVKIWFQNRRMKRKRKR
>Nv_076
TKPRRSFFSADQVNQLERVFTAQQYVSAKERAEIAETFNMTDDQVKNWFQNRRMKRKRKR
>Nv_096
RTTTRSFFTEHQVACLQRVFANRRYVCSSERQKLAKELNMTDQQVKTWFQNRRMKVKRKR
>Nv_102
QPRKRPVFHPDVVARLEQFFAKRRYINAEQRYALAKEINMTEEQIKSWFHNKRTAMKRKL

>Nv_147
PKKNRSSFSAEQVFRLEKTFELQQYLGTKERQQLALALNMTDNQVKTWFQNRRMKQKRKR
>Nv_033
SRRRRTAFTSSQLKSLEEKFQEKKYLTISERNSLAKSMHLTNTQVKTWFQNRRTKWKKQM
>Nv_050
KRRKRTAFTSFQLKCLEDKFKFSKYLTIAERDMMARSLQLTNRQIKTWFQNRRTKWKREN
>Nv_054
KKRHRSHFSQLQLQYLDAIFARQHYLSRDERTVLAGALDMTELQVRNWFQNKRYQRRNRE
>Nv_081
NPRTRTHFTERQLKYLETYYSNGRYLSRDERTVLAQALEMTELQVRNWFQNRRYQRKQKQ
>Nv_056
SRRKRTAFTSSQLKYLEEKFQEKKYLTISERNELAKSMYLSDTQVKTWFQNRRTKWKKQM
>Nv_041
FRRIRTVFTRDQLQKLEDRFKHDKYLSTQDRNEFARSLQLTPLQLKTWYQNRRMKWKKDM
>Nv_083
ARRGRTVFSAQQLQVLERVFAGSQYIVGSQRKFLASQLRLSETQVRVWFQNRRIKWRKQM
>Nv_145
KRRKRTIFTSEQLRRLEEEFDKHQYLVGTERQNLAKSLNLSETQVKIWFQNRRIKWRKDN
>Nw_49193
KDKYRVVYSDRQRAELESEFRCNQYITIKRKTELAGMVGLSERQVKIWFQNRRAKERKVA
>Nw_48679
QRRYRTAFSRQQLQRLEREFVKENYVSRPKRCELASSLGLPETTIKVWFQNRRMKDKRQR
>Nw_36518
QRRYRTAFSRQQLQRLEREFMKENYVSRPKRCELAATLGLAETTIKVWFQNRRMKDKRQR
>Nw_35873
TRGGRTNFTTKQLTELEKEFLYNKYLTRARRVEIAASLNLNENQVKIWFQNRRMKQKKLV
>Nw_54694
SRKKRRPYTKYQTLELEKEFLFNTYITRERRSEISRALNLSDRQVKIWFQNRRMKMKKMT
>Nw_10725
RKKRRPYTKYQTLELEKEFLFNVYITRERRSEISRSLNLTDRQVKIWFQ
>Nw_58015
KRKDRTAFSKFQIEQLESEFETYNYLTRLRRFELSKKLSLSERQVKVWFQNRRMKWKRVK
>Nw_58481
SRRPRTSYSSKQIKILEEHFVLNRYLSRPKRFEVATSLNLTETQVKIWFQNRRMKLKRSE
>Nw_6133
NRRQRTTFSSEQTLSLEVEFQRSEYISRSKRYEIAEHLHLTENQVKIWFQNRRAKDKRME
>Nw_44999
CRRTRTVFTELQMLGLEKRFDMQKYLSTPDRLGLSDSLGLTQLQVKTWYQNRRMKWKKEI
>Nw_50937
LRKPRTIYSSYQLSMLNKRFQGTQYLALPERAELAATLGLTQTQVKIWFQNRRSKFKKLL
>Nw_54562
MRKPRTIYSSLQLQILNKRFQQTQYLALPERAELAASLGLTQTQVKI
>Nw_35817
IRKPRTIYSSYQLSMLNKRFQGTQYLALPERAELAASLGLTQTQVKIWFQNRRSKFKKLM
>Nw_17210
PKRIRTAFTSEQLQRLENEFTSNPYIVGEERRQFAASLRLSETQVKVWYQNRRTKDKREI
>Nw_44961
PKRIRTAFTTEQLQRLEDEFIANPYIVGEERRRFAANLNLTETQVKVWYQNRRTKDKRET
>Nw_45472
RRKRRVLFTQAQVYELERRFKQQKYLSAPEREQLAAMIHLTPTQVKIWFQNHRYKMKRQA
>Nw_20731
KRKRRVLFSKQQTYELERRFRQQRYLSAPEREHLASIINLTPTQVKIWFQN
>Nw_36577
KKKTRTVFSRHQVYQLETMFDCKKYLSSSERAGLAATLHLTETQVKIWFQNRRNKLKRQF
>Nw_41714
KKHTRPTFSGHQIFALEKTFEQTKYLAGPERGRLAYTLGMSESQVKVWFQNRRTKWRKRA

>Nw_33755
RKKPRTSFTRVQVGELEKRFIVQKYLASAERNSLAKGLGMTDAQVKTWFQNRRTKWRRQT
>Nw_49354
PKRQRTSFTAEQLYILEKEFRRNQYMVGKDRGLLAESLDLTETQVKVWFQNRRTKYKREK
>Nw_4231
GRNPRVPFTQTQVDHLELKFREAKYLSSLDVADISTHLNLTDTRVKIWFQNRRARERREK
>Nw_15302.1
RKPRILFNHSQIFQLELRFKHQRYLTANDRRELSDQLNLTQNQIKIWFQNRRYKSKRQE
>Nw_20062
YRRKRTSFSSSQLRELEKHFNRKKYLSVPDRTLLANILHLTQPQIKVWFQNRRAKWKRIK
>Nw_49786
GMLRRAVFSDAQRRGLERKFRELKYISKPDRKELANKLDLRDSQVKIWFQNRRMKWRNSK
>Nw_29958
WNRVRTVFTNAQRLGLEEQFREQKYVSKPDRKELAGKLGLSDSQVKVWFQNRRMKWRNSL
>Nw_31941.1
RLRIRTPITGRHRASMFQHFLTKSQYLTPVDCSNMSCNLGIPEPAIRVFFQNLRARVRAR
>Nw_36321
GKKSRTSFTPVQVLKLEHKFQTQKYLAAVERGEFAEKLGLTDTQVKTWFQNRRMKHKRVQ
>Nw_45900
TSHQDHVENQGYQPILESKFAINNYISGSEVEELSNRLGITDIQIRDWFANRRRRGQFSQ
>Nw_50700
NKKSRTSFTACQVVKLEQKFQSQKYLAAVERGEFAEKLGLSDTQVKTWFQNRRMKHKRVQ
>Nw_20893
KRQRTCFSAEQLYVLEQEFIKNQYMVGKDRTLLAQSLTLSETQV
>Xb_DN63819
KDKYRVVYTDRQRAELENEFRCNQYITIRRKTEIAGAIGLSERQVKIWFQNRRAKERKVA
>Xb_DN14752
KEYTRENYVSRPRRCELAAQLNLPETTIKVWFQNRRMKDKRQR
>Xb_DN90720
SRRRRTAYTSEQLLQLEREFHGKKYLSLTERSHIAKNLKLTEGQVKIWFQNRRAKWKRVK
>Xb_DN115451
AKRTRTAFSSTQLLNLEKEFSRNSYLTRIRRIELAQSLDLSEKQVKIWFQNRRVKQKKLL
>Xb_DN2325
SRGGRTNFTNKQQTELEKEFHFNRYLTRARRIEIAASLSLNETQVKIWFQNRRMKQKKLL
>Xb_DN124767
SKRTRTAYTRQQVLELEKEFHYNKYLTRRRRIEIANALLLTERQIKIWFQNRRMKYKKEN
>Xb_DN76496
CRRKRSSYDRLQCLELEKEFHFNRYLSKRRRSQISKHLGLTERQVKVWFQNRRMKHKKDR
>Xb_DN84275
RKRGRQTYTRYQTLELEKEFHYNQYLTRRRRIEIATGLGLTERQIKIWFQNRRMKFKKEN
>Xb_DN3733
RKKRRPYTKFQTLELEKEFLFNTYITRERRLEISRALNLTDRQVKIWFQNRRMKMKKMNS
>Xb_DN86948
PRKERTAFTKHQIRELEREFTKHNYLTRLRRYEIAVSLELSERQVKVWFQNRRMKWKRVK
>Xb_DN44680
PRKARTAFTDDQLARLETSFERQKYLSVQDRMDLAANLNLTDTQVKTWYQNRRTKWKRQT
>Xb_DN81549
CRRSRTVFTESQLVGLERRFADQKYLSTPDRITLADQLRLSQIQVKTWYQNRRMKWKKQV
>Xb_DN90457
MRKPRTIYTSLQLQQLNKRFARTQYLALPERAELAASLGLTQTQVKIWFQNRRSKYKKVL
>Xb_DN105712
PKRVRTAFSPSQLLRLEHEFEKNHYVVGEERKQLANVLNLSETQVKVWFQNRRTKYKRQK
>Xb_DN63160
NRKPRTPFTTSQLLALERKFRHKQYLSIAERAEFSSSLSLTETQVKIWFQNRRAKAKRLQ
>Xb_DN131037
LENKFKLTRYLSVCERLNLALSLSLTENQVKIWFQNRRTKWKK

>Xb_DN93010
KRKRRVLFTQAQVYELERRFKQQKYLSAPEREHLASLINLTPTQVKIWFQNHRYKCKRAA
>Xb_DN101095
QTYELERRFRQQRYLSAPEREHLASIIRLSPTQVEVWFQNHRYKLKRSR
>Xb_DN50727
RRKPRVLFTQAQVYELERRFKQQRYLSAPEREHLAHLLKLSSTQVKIWFQNRRYKCKRQR
>Xb_DN92869
KKKTRTVFSRSQVFQLESTFDMKRYLSSSERAGLAASLHLTETQVKIWFQNRRNKWKRQL
>Xb_DN75963
RKHTRPTFSGHQIFALEKTFEQTKYLAGPERRRLAFSLGMTESQVKVWFQNRRTKWRKKH
>Xb_DN135951
KKPRTSFTRLQIMELEKRFHKQKYLASAERSSLAKDLKMTDAQVKTWFQNRRTKWRR
>Xb_DN69228
SKRRRTSFSQEQIRLLEKKFEEKQYLTAKERDDIACDLDLKASQVITWFQNKRSKIRRET
>Xb_DN94680
RKKARTVFTGRQIFELERQFEAKKYLSAHERAEMAALLNVTDQQVKIWFQNRRTKWKRLD
>Xb_DN41925
GKRMRTVFTYKQLQKLEKVFQDQQYMVGEDRCRLADSLGLSETQVRVWFQNRRIKWRKET
>Xb_DN122354
KKKTRTAFSPSQIYELEKRFRLQKYLAASERGEFADKLSLTDTQVKTWFQNRRMKYKRQH
>Xb_DN94983
RKGGQVRFSNEQTLKLEKTFHKQKYLSPSERKKIAKSLQLTERQVKTWFQNRRAKWRRLE
>IP_DN44102
KDKYRVVYTDRQRAELENEFRSAQYITIRRKSELAMQVGLSERQVKIWFQNRRAKERKVS
>IP_DN12279
MRRYRTAFSKHQQDSLEMEFRKENYVSRPRRCELAASLNLPETTIKVWFQNRRMKEKRQR
>IP_DN47321
SRRTRTAFSSSQLMCLESEFESSNYLTRIRRIEIANKLHLTEKQIKIWFQNRRVKNKKGA
>IP_DN75362
TRGGRTNFTNKQLTELEKEFHFNRYLTRARRIEIASSLNLNETQVKIWFQNRRMKQKKLV
>IP_DN17434
CKRTRTAYTRFQTLELEKEFHFNRYLTRRRRIEIANLLALSERQIKIWFQNRRMKWKKDN
>IP_DN66164
SRKKRRPYTKTQTLELEKEFLYNTYITRERRLEIARSLSLTDRQVKIWFQNRRMKNKKQM
>IP_DN54162
ERTAFTKFQVHALEKEFTLHNYVNRPRRQLIARKLGLSERQVKVWFQNRRMKWKRTK
>IP_DN62850
SRRQRTSFTPDQTLQLELQYERSEYITRSKRFELAESLGLSEGQIKIWYQNRRAKDKRIE
>IP_DN10641
EEFKKDKYLSVSKRLELSVALSLTETQIKTWFQNRRTKWKKQV
>IP_DN36044
MRKARTAFTDEQLVKLEQNFERHKYLSVQDRLDLAAALDLTDTQVKTWYQNRRTKWKRQA
>IP_DN60676
TRRRRTAFTQAQLNFLERKFRCQKYLSVADRSDVAEMLRLSETQVKTWYQNRRTKWKRQT
>IP_DN67947
CRRSRTVFTEAQLANLEKQFGNTKYLSTAQRSRIAKSSGLTQMQVKTWYQNRRMKWKKQL
>IP_DN76364
YRRNRSVFSEKQLADLEKWFVAEKYLTASLRVKIASQCGLTEMQVKTWYQNRRMKWKKQM
>IP_DN57889
RRKARTVFSDGQLHGLERRFESQRYLSTPERIDVASAFNLTETQV
>IP_DN61228
LRKPRTIYSTVQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKMKKII
>IP_DN26203
KRQGRTVFTKQQLEILNGHFNHRQYLSLAERAQLSEQLCLSQKQVKIWFQNMRSKAKRVM
>IP_DN43104
CKRVRTAFTATQLAMLERHFLQSPYVVGSQRRKLSSDLGLSETQVKVWFQNRRTKAKRCI

>IP_DN60553
PKRERTSFTPEQLFRLEKEFRANQYMVGRDRGDLAASLGLSETQVKVWFQNRRTKYKREK
>IP_DN65413
GRNPRVPFSSTQVAVLEQKYQHKQYLSSIDVAELSSLLNLNDSRVKIWFQNRRARARKEQ
>IP_DN40909
RRRSRTAFTYEQLVALENKFRASRYLSVCERLNIALALNLTETQVKIWFQNRRTKFKKQN
>IP_DN64840
RRKRRVLFTQAQIYELERRFKQQKYLTAAERENLAHVINLSPTQVKIWFQNHRYKTKKSK
>IP_DN66164.1
RRKRRVLFTQAQVYELERRFKAQKYLSAPERETLAHSIGLTPTQVKIWFQN
>IP_DN66164.2
RRKRRVLFTQAQVYELERRFKTQKYLSAPERETLAHAINLTPTQVKIWFQNHRYKCKRAS
>IP_DN66907
KRKRRVLFTKNQTYELERRFRQQRYLSAPEREALAAMIDLTPGQVKIWFQNHRYKCKKSN
>IP_DN64224
RTSAAFSHAQVFELERRFAQQKYLSGPERADLANALKLTETQIKIWFQNRRYKTKRKQ
>IP_DN15610
RKKTRTVFSRNQVCQLESMFDMKRYLSSSERTGIARDLGLSEQQIKIWFQNRRNKLKRQI
>IP_DN69723
KKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALGMSESQVKVWFQNRRTKWRKKH
>IP_DN68209
RKKPRTSFTRLQICELEKRFHQQKYLASGERATLASRLNMTDSQVKTWFQNRRTKWRRQS
>IP_DN89774
PKRERTSFTAQQLFLLEEEFLKAQYLVGRDRTELAKKLNLSETQIKVWFQNRRTKHKREK
>IP_DN56009
KKKNRTAFTPGQIFELEKRFRVQKYLSASERGEFASHLNLTDTQIKTWFQNRRMKHKRQQ
>IP_DN60800
SKKQRTAFTPDQVYELERRFSQQKYLSASERAEFASKLQLSDTQIKTWFQNRRMKRKRQQ
>SR_DN125595
KDKYRVVYTDRQRAELENEFRSAQYITIRRKSELAMQVGLSERQVKIWFQNRRAKERKVS
>SR_DN124972
IRRYRTAFSKDQQDILEAEFRKENYVSRPRRCEMAAALNLPETTIKVWFQNRRMKEKRQK
>SR_DN87690
ERRSRTAFSSVQLIQLEREFDLNNYLTRMRRIEIANRLNLTEKQVKIWFQNRRVKCKKQL
>SR_DN118344
CKRTRTAYTRFQTLELEKEFHFNRYLTRRRRIEIANLLALTERQIKIWFQNRRMKWKKDN
>SR_DN123370
SRKKRRPYTKNQTLELEKEFLFNTYITRERRLEIARSLNLTDRQVKIWFQNRRMKNKKQH
>SR_DN123752
CRRPRTAFTSQQLLELENQFRKNKYLSRPKRFEVATKLMLTETQVKIWFQNRRMKWKRAK
>SR_DN119820
GRRQRTSFSAEQTIKLELEFQMNEYITRSRRFELADTLGLSENQIKIWYQNRRAKDKRIE
>SR_DN112643
DRKPRQAYSAKQLEKLEEEFKNDKYLSVSKRLELSVALSLSETQIKTWFQNRRTKWKKQL
>SR_DN152752
HRKARTAFTDEQLAQLEQNFERHKYLSVQDRLDLAASLQLSDTQVKTWYQNRRTKWKRQT
>SR_DN100159
SVSGSISVPPSSSIPSPLALQNFLNKKNRRRSDVAELLCLSETQVKTWYQNRRTKWKRQT
>SR_DN80548
CRRSRTVFSDQQLKGLEKNFCSTKYLTTMQRNRLAKESGLTQMQVKTWYQNRRMKWKKQL
>SR_DN106437
RRKARTVFSDGQLHGLERKFQQQRYLSTPERIEIATQYNLTETQVKTWFQNRRMKQKKIF
>SR_DN123563
MRKPRTIYSTIQIQQLTRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKMKKLM
>SR_DN118899
SKRIRTAFTPAQLAKLETQFARSPYVVGHQRRQISEELGLTETQVKVWFQNRRTKEKRVL

>SR_DN116214
PKRERTSFTPEQLFRLEKEFSLNQYMVGRDRGQLATCLGLTETQVKVWFQNRRTKYKREK
>SR_DN124792
GRNPRVPFSSAQVAILEQKYQVKQYLSSIDVAELSTMLSLNDSRVKIWFQNRRARARKEQ
>SR_DN56103
RSRTAFTYEQLVALENKFRCSRYLSVCERLNIALALNLTETQVKIWFQNRRTKFKKQN
>SR_DN122139
RRKRRVLFTQAQIHELERRFKQQKYLTAAERESLAHLINLSPTQVKIWFQNHRYKTKKSK
>SR_DN126831
RRKRRVLFTQAQVYELERRFKQQKYLSAPEREALAQGINLTPTQVKIWFQNHRYKCKRAS
>SR_DN89648
KRKRRVLFSKSQTFQLESRFRQQRYLSAPEREQLANAINLTPGQVKIWFQNHRYKCKKSS
>SR_DN121921
KKRTRAAFSHAQVFELERRFAQQKYLSGPERSELASALKLTETQIKIWFQNRRYKTKRKQ
>SR_DN123682
KKKTRTVFSRSQVFQLESTFDVKRYLSSSERANLANSLHLTETQVKIWFQNRRNKWKRQI
>SR_DN110321
KKKTRTVFSRTQVCQLENMFDLKRYLSSSERATLARDLTLSEQQIKIWFQNRRNKLKRQI
>SR_DN130725
KKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAFALGMTESQVKVWFQNRRTKWRKKH
>SR_DN74225
RKKPRTSFTRLQICELEKRFHQQKYLASGERATLASKLNMTDSQVKTWFQNRRTKWRRQS
>SR_DN109741
PKRERTCFSAEQLFLLEEEFTRGQYMVGRDRTLLAGRLGLTETQIKVWFQNRRTKQKRES
>SR_DN130717
CKKQRTAFTPDQIHELERRFAQQKYLSASERADFANKLKLSDTQIKTWFQNRRMKRKRQQ
>SR_DN126921
RKKNRTAFTPHQIYELEKRFRSQKYLSASERGEFALSLKLTDTQIKTWFQNRRMKHKRQQ
>HM_DN14515
KDKYRVVYTDRQRAELENEFRSNQYITIRRKTELASLVGLSERQVKIWFQNRRAKERKVA
>HM_DN58284
LRRYRTAFTKEQQDKLEAEFKKENYISRPKRCEIASELNLPETTIKVWFQNRRMKEKRQR
>HM_DN24302
ARRSRTAFTSNQLVMLEKEFNNSKYLTRLRRIDIATSLKLTEKQIKIWFQNRRVKDKKTK
>HM_DN54918
TRGGRTNFTNKQLTELEKEFHFNRYLTRARRIEIAASLNLNETQVKIWFQNRRMKQKKLL
>HM_DN27605
SKRTRTAYTRHQTLELEKEFHFNRYLTRRRRIEIATLLSLSERQIKIWFQNRRMKWKKDN
>HM_DN3061
TRKKRRPYTKTQTLELEKEFLYNTYITRERRLQIAKSLNLTDRQVKIWFQNRRMKMKKIN
>HM_DN4748
SRRPRTAFTSQQLLELENQFKRNKYLSRPKRFEVATKLMLTETQVKIWFQNRRMKWKRAR
>HM_DN18872
NRRERTTFTSDQSVKLEIEYQRSEYITRSRRFTLAEMLKMSENQIKIWFQNRRAKDKRIE
>HM_DN12978
ERRQRTTFTTKQANILESEYQRREYLSKQQRQALATSLQLSETQIKIWFQNRRAKDKRTG
>HM_DN29024
ERKPRQAYSTKQLEKLEEEFKKDKYLSVSKRLELSVALELTETQIKTWFQNRRTKWKKQM
>HM_DN68850
TRKARTAFTEEQLSKLEASFERQKYLSVQDRLDLASSLDLSDTQVKTWYQNRRTKWKRQA
>HM_DN83548
PRRRRTAFTQAQLNFLERKFRCQKYLSVADRNEVATQLALSETQVKTWYQNRRTKWKRQT
>HM_DN34181
CRRSRTVFTDSQLIGLDKSFKHSKYLNTPERNRLAKHLGLTQTQVKTWYQNRRMKWKKQL
>HM_DN1275
RRKARTVFSDSQLQGLEERFKDQKYLSTPERIEMATLLCLTETQVKTWFQNRRMKDKKCC

>HM_DN12035
IRKPRTIYSSTQLQQLNRRYNTTQYLALPERADLAATLGLTQTQVKIWFQNKRSKMKKIM
>HM_DN12216
IRKPRTIYSTMQLQQLNRRFQRTQYLALPERAELAAALGLTQTQVKIWFQNKRSKVKKIL
>HM_DN58320
PKRERTSFTPQQLYQLEKEFHKNQYMIGRDRAELAVLLKLTETQVKVWFQNRRTKYKREK
>HM_DN104433
SRNPRIPFTSNQVALLEQTFQKKQYLSSMDVSDLSNSLRLSDTRIKIWFQNRRARARRDG
>HM_DN36484
PRKRRILFTQAQVYELERRFKMQKYLNAPERESLAQLINLTPTQVKIWFQNHRYKCKRAA
>HM_DN15813
RRKRRVLFTQAQVYELERRFKQQRYLSAPERETLAHLINLTPTQVKIWFQNHRYKCKRAS
>HM_DN44693
RRKRRVLFTQAQVYELERRFKQQRYLSAPERENLAQLINLTPTQVKIWFQNHRYKYKRAS
>HM_DN90735
KRKRRVLFSKAQTFELERRFRQQRYLSAPEREHLANLIKLTPTQVKIWFQNHRYKCKKNH
>HM_DN79735
KKRTRAAFSHAQVFELERRFTQQRYLSGPERSEVASNLKLTETQVKIWFQNRRYKTKRKQ
>HM_DN43016
KKKTRTVFSRNQVFQLETTFDLKRYLSSSERASLAATLSLTETQVKIWFQNRRNKWKRQL
>HM_DN8401
KKKSRTVFSRSQVCQLEGTFDQKRYLSSAERSTLARSLKLTETQVKIWFQNRRNKLKRQI
>HM_DN10228
KKHTRPTFSGHQIFALEKTFEQTKYLAGPERARLAYALGMSESQVKVWFQNRRTKWRKKH
>HM_DN40476
RKKPRTSFTRLQICELEKRFHLQKYLASTERASLAKSLKMTDAQVKTWFQNRRTKWRRQT
>HM_DN3930
KKKNRTAFTPGQICELENRFRLQQYLSTTERGEFASRLKLTDTQVKTWFQNRRMKLKRQQ
>HM_DN20211
KKKQRTAFTPEQVFELEKRFQIQKYLTASERSDFSDSLKLSDTQVKTWFQNRRMKHKRQQ
>HM_DN91365
RKKQRILFHHRQIEQLEDFFAAQPYISAIERQCLSQKTGLTNNQVKIWFQNRRYKEKRLK

Non ANTP HD sequences

>Ce_alr-1_ARX
QRRYRTTFSAFQLDELEKVFARTHYPDVFTREELATRVQLTEARVQVWFQNRRAKYRKQE
>Ce_ceh-45_GSC
KRRHRTIFSEEQLNILETTFSTTHYPDATTREELAVQCSLKEERVEVWFKNRRAKERKQK
>Ce_ceh-37_OTX
NRRERTTYSRQQLEILETLFNETQYPDVFARERVADQIRLQESRIQVWFKNRRAKYRLQE
>Ce_ttx-1_OTX
QRRERTTFTRNQLEILESYFVKTRYPDIFMREDMAHKIQLPESRVQVWFKNRRAKARQQK
>Ce_ceh-36_OTX
GRRERTSFNRGQLDQLEKVFRETQYPDVHRREALAKAINLPDGRVQVITVWFKNRRAKDRNNK
>Ce_vab-3_PAX4/6
LQRNRTSFTQVQIESLEKEFERTHYPDVFARERLAQKIQLPEARIQVWFSNRRAKWRREE
>Ce_eyg-1_PAX3/7
RRLSRSTFTNEQLQSLEEVFLRDPYPSPNERADLVKRTNLPEARIQVWFSNRRAKWRKTN
>Ce_ceh-53_PAX3/7
VRRLRTAFSENQLELLEEAFLKCQYPDVQQRETLGKQTELAEARIQVWFKNRRAKARKRQ
>Ce_pax-3_PAX3/7
SRRNRTSFTAEQLDVLENAFRADTYPHANARESISKETGLSEEKIMTWFSNRRARCRKNM
>Ce_unc-30_PITX

PRRQRTHFTSHQLTELENWFSRNRYPDMACREEIAVWISLTEPRVRVWFKNRRAKWRKRE
>Ce_unc-42_PROP
RRRHRTTFTQEQLQELDAAFQKSHYPDIYVREELARITKLNEARIQVWFQNRRAKHRKHE
>Ce_ceh-8_RAX
QRRNRTTFTTFQLHALEAAFDKTHYPDVYARETLAAKVQLPEVRVQVWFQNRRAKFRRQE
>Ce_unc-4_UNCX
RRRTRTNFSGWQLEELESAFEASHYPDVFMREALAMRLDLLESRVQVWFQNRRAKWRKRE
>Ce_ceh-10_VSX
KRRHRTIFTQYQIDELEKAFQDSHYPDIYAREVLAGKTELQEDRIQVWFQNRRAKWRKTE
>Ce_dsc-1_VSX
RRRFRTNFTELQSTFLEDSFKESHYPDHKAKKYMADFLKIPEDRITVWFQNRRAKWRRKE
>Ce_ceh-17_PHOX
QRRIRTTFTSGQLKELERSFCETHYPDIYTREEIAMRIDLTEARVQVWFQNRRAKYRKQE
>Ce_ceh-54
KKRNRITFDANQIDEMEKVFAENQYPDTMSREKLANKIQLHEERVQIWFQNRRAKYRREQ
>Ce_ceh-14_LHX3/4
NKRPRTTISAKSLETLKQAYQTSSKPARHVREQLASETGLDMRVVQVWFQNRRAKEKRLK
>Ce_lim-6_LMX
PKRPRTILNAQQRRQFKTAFERSSKPSRKVREQLANETGLSVRVVQVWFQNQRAKIKKLN
>Ce_lin-11_LHX1/5
RRGPRTTIKAKQLETLKNAFAATPKPTRHIREQLAAETGLNMRVIQVWFQNRRSKERRMK
>Ce_mec-3_LHX1/5
RRGPRTTIKQNQLDVLNEMFSNTPKPSKHARAKLALETGLSMRVIQVWFQNRRSKERRLK
>Ce_ttx-3_LHX2/9
SKRMRTSFKHHQLRAMKTYFALNHNPDAKDLKQLAAKTNLTKRVLQVWFQNARAKYRREL
>Ce_Ce_lim-4_LHX6/8
TKRVRTTFAEDQLSVLQTYFNRDSNPDGADLEKIASMTGLSKRVTQVWFQNSRARQKKWH
>Ce_lim-7_ISLET
ATRVRTVLNENQLKILRDCYSINSRPDATLKERLVEMTGLSARVIRVWFQNKRCKDKKRQ
>Ce_ceh-6_POU3
KRKKRTSIEVNVKSRLEFHFQSNQKPNAQEITQVAMELQLEKEVVRVWFCNRRQKEKRIA
>Ce_unc-86_POU4
KKRKRTSIAAPEKRELEQFFKQQPRPSGERIASIADRLDLKKNVVRVWFCNQRQKQKRDF
>Ce_ceh-18_POU2
RRRKRTNLDMNQRNALDTFFALNPRPDHDKMTDIANSLELDRDVVRVWFCNRRQKMRRVD
>Ce_hmbx-1_HMBOX
QRRERYVFRPILIRMLESFFTQTPFPDLPRRVEIANACNHVLKMDKKGVGLMPKEVVSPQVVSNWFAN
KRKELRRR
>Ce_ceh-32_SIX3/6
GEQKTHCFKERTRSLLREWYLKDPYPNPPKKKELANATGLTQMQVGNWFKNRRQRDRAAA
>Ce_ceh-33_SIX1/2
GEETSYCFRDKSRVLLRDWYCRNSYPSPREKRELAEKTHLTVTQVSNWFKNRRQRDRAGV
>Ce_ceh-34_SIX1/2
GEETNYCFKSKSRNVLRDAYKKCQYPSVEDKRRLAQQTELSIIQVSNWFKNKRQRERAAG
>Ce_unc-39_SIX4/5
GEEIVYSFKDSSRKFLKQFFRNVSEYPTQEQKREISRATGLKIVQISNWFKNRRQRDKSNN
>Ce_ceh-20_PBX
ARRKRRNFSKQATEVLNEYFYGHLSNPYPSEEAKEDLARQCNITVSQVSNWFGNKRIRYKKNM
>Ce_ceh-40_PBX
ARRKRRNFSKTSTEILNEYFLANINHPYPSEEVKQALAMQCNISVAQVSNWFGNKRIRYKKTM
>Ce_ceh-60_PBX
QGRKRRNFDKNTTDILQNWFHDHRQNPYPSDQEKAELAKQCNIKISQVNNWFGNQRIRTKQQA
>Ce_unc-62_MEIS
KRKVPKVFSKEAITKFRAWLFHNLTHPYPSEEQKKQLAKETGLTILQVNNWFINARRRIVQPM
>Ce_irx-1_IRX
DGIKKRNATREATAPLKDWLHSHRKNPYPSKADKVMLAVGTGMTLTQVSTWFANARRRLKKEN

>Ce_ceh-21_ONECUT
VKVIRLTFTETQLKSLQKSFQQNHRPTREMRQKLSATLELDFSTVGNFFMNSRRRLRIDQ
>Ce_ceh-38_ONECUT
SKRPRLVFTDIQKRTLQAIFKETQRPSREMQQTIAEHLRLDLSTVANFFMNARRRSRLGG
>Ce_ceh-39_ONECUT
RIRRFTFTQTQLDSLHTVFQQQDRPNREMQQALSATLKLNRSTVGNFFMNARRRLPKAAP
>Ce_ceh-41_ONECUT
PKKTKSPFTEHEIAVMMALFEINKSPNHEEVQKLAVQLNLGYRSVANFFMNKRAKERKKL
>Ce_ceh-44_CUX
KPIQRTVITDYQKDTLRFVFVNEQHPSNELCEQISLKLDMSLRTVQNWFHNHRTRSKARE
>Ce_ceh-48_ONECUT
PKKTRLVFSDIQRRTLQAIFRETKRPSREMQITISQQLNLDPTTVANFFMNARRRGHDLK
>Ce_ceh-49_ONECUT
TPRARCLLSQDQKSQLSIFFETNPRPDSLEMKQLGSTLNLCKSTIINYFTNMRRRKATSS
>Ce_dve-1_HD1_CMP
PTQVQRFNTLYEIPRLDKWFKSDANPSKQKMNNYLSQLNQSPFRKNNSKISYQQICNWFTQKRSSSRT
LA
>Ce_dve-1_HD2_CMP
AARSRLMFDPLTELPVLEKWFEENPHPTWMQIDQYTQCLNNCAYRENYPHISQHNVKIWFKNRRAKCK
RLL
>Ce_zag-1_ZEB
PLRSRSFLNDSQVAVLQNHFKRNPFPSKYELSAVAEQIGVNKRVVQVWFQNTRAKERRSN
>Ce_ZFH-2_HD1_ZFHX
AKRARTRITDDQLKVLRQYFNINNSPSEAQIKEMSQKASLPEKVIKHWFRNTLFKERQRD
>Ce_ZFH-2_HD2_ZFHX
RRANRTRFTDFQLRTLQQFFDKQAYPKDDDLEALSKKLQLSPRVIVVWFQNARQKARKIY
>Ce_ZFH-2_HD3_ZFHX
NKRFRTHLTPMQVQMMKSVFNEYKTPSMAECELLGKEVGLHKRVVQVWFQNARAKERKTR
>Ce_ceh-26_PROS
GGGSSSMLTPMHLRKAKLMFFYTRYPNSNLLKSYFPDIRFNKNNTAQLVKWFSNFREFYYNQMEKFAR
QALAEGITDRNDIFV
>Dm_ey
LQRNRTSFTNDQIDSLEKEFERTHYPDVFARERLAGKIGLPEARIQVWFSNRRAKWRREE
>Dm_toy
LQRNRTSFSNEQIDSLEKEFERTHYPDVFARERLADKIGLPEARIQVWFSNRRAKWRREE
>Dm_eyg
FRRNRTTFSPEQLEELEKEFDKSHYPCVSTRERLSSRTSLSEARVQVWFSNRRAKWRRHQ
>Dm_toe
FRRNRTTFSPEQLDELEKEFDKSHYPCVNTREKLAARTALSEARVQVWFSNRRAKWRRHQ
>Dm_gsb
QRRSRTTFSNDQIDALERIFARTQYPDVYTREELAQSTGLTEARVQVWFSNRRARLRKQL
>Dm_gsbn
QRRSRTTFTAEQLEALERAFSRTQYPDVYTREELAQTTALTEARIQVWFSNRRARLRKHS
>Dm_prd
QRRCRTTFSASQLDELERAFERTQYPDIYTREELAQRTNLTEARIQVWFSNRRARLRKQH
>Dm_al
QRRYRTTFTSFQLEELEKAFSRTHYPDVFTREELAMKIGLTEARIQVWFQNRRAKWRKQE
>Dm_CG11294
QRRNRTTFTPQQLQELEALFQKTHYPDVFLREEVALRISLSEARVQVWFQNRRAKWRKQA
>Dm_CG34340
QRRNRTTFTLQQLEELETAFAQTHYPDVFTREDLAMKINLTEARVQVWFQNRRAKWRKAE
>Dm_CG34367
QRRSRTNFTLDQLNELERLFEETHYPDAFMREELSQRLGLSEARVQVWFQNRRAKCRKHE
>Dm_CG9876
PRRNRTTFSSAQLTALEKVFERTHYPDAFVREELATKVHLSEARVQVWFQNRRAKFRRNE
>Dm_CG32532

RRRHRTTFTQEQLAELEAAFAKSHYPDIYCREELARTTKLNEARIQVWFQNRRAKYRKQE
>Dm_Gsc
KRRHRTIFTEEQLEQLEATFDKTHYPDVVLREQLALKVDLKEERVEVWFKNRRAKWRKQK
>Dm_Vsx1
RRHGRTIFTSSQLEELEKAFKEAHYPDVSARELLSMKTGLAEDRIQVWYQNRRAKWRKTE
>Dm_Vsx2
RRHSRTIFTSYQLEKLEEAFKEAHYPDVYAREMLSLKTELPEDRIQVWFQNRRAKWRKTE
>Dm_unc4
RRRSRTNFNSWQLEELERAFSASHYPDIFMREALAMRLDLKESRVAVWFQNRRAKVRKRE
>Dm_oc
QRRERTTFTRAQLDVLEALFGKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCRQQL
>Dm_OdsH
KRRGRTNFNSWQLRELERVFQGSHYPDIFMREALATKLDLMEGRIAVWFQNRRAKWRKQE
>Dm_hbn
VRRSRTTFTTFQLHQLERAFEKTQYPDVFTREDLAMRLDLSEARVQVWFQNRRAKWRKRE
>Dm_otp
QKRHRTRFTPAQLNELERCFSKTHYPDIFMREEIAMRIGLTESRVQVWFQNRRAKWKKRK
>Dm_PHDP
QRRIRTTFTSNQLNELEKIFLETHYPDIYTREEIASKLHLTEARVQVWFQNRRAKFRKQE
>Dm_Pph13
QRRYRTTFNTLQLQELERAFQRTHYPDVFFREELAVRIDLTEARVQVWFQNRRAKWRKQE
>Dm_Ptx1
QRRQRTHFTSQQLQELEHTFSRNRYPDMSTREEIAMWTNLTEARVRVWFKNRRAKWRKRE
>Dm_repo
KKKTRTTFTAYQLEELERAFERAPYPDVFAREELAIKLNLSESRVQVWFQNRRAKWRKHE
>Dm_Rx
HRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAMKVNLPEVRVQVWFQNRRAKWRRQE
>Dm_tup
PTRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTSLSPRVIRVWFQNKRCKDKKKT
>Dm_ap
TKRMRTSFKHHQLRTMKSYFAINHNPDAKDLKQLSQKTGLPKRVLQVWFQNARAKWRRMM
>Dm_Awh
TKRVRTTFTEEQLQVLQANFQIDSNPDGQDLERIASVTGLSKRVTQVWFQNSRARQKKHI
>Dm_Lim1
RRGPRTTIKAKQLEVLKTAFNQTPKPTRHIREQLAKETGLPMRVIQVWFQNKRSKERRMK
>Dm_Lim3
NKRPRTTITAKQLETLKTAYNNSPKPARHVREQLSQDTGLDMRVVQVWFQNRRAKEKRLK
>Dm_CG32105
PKRPRTILTSQQRKQFKASFDQSPKPCRKVREALAKDTGLSVRVVQVWFQNQRAKMKKIQ
>Dm_CG4328
PKRPRTILNTQQRRAFKASFEVSPKPCRKVRENLAKDTGLSLRIVQVWFQNQRAKVKKIQ
>Dm_zfh1
KVRVRTAINEEQQQQLKQHYSLNARPSRDEFRMIAARLQLDPRVVQVWFQNNRSRERKMQ
>Dm_zfh2_HD1
QKRARTRITDDQLKILRAHFDINNSPSEESIMEMSQKANLPMKVVKHWFRNTLFKERQRN
>Dm_zfh2_HD2
KRANRTRFTDYQIKVLQEFFENNSYPKDSDLEYLSKLLLLSPRVIVVWFQNARQKQRKIY
>Dm_zfh2_HD3
NKRLRTTILPEQLNFLYECYQSESNPSRKMLEEISKKVNLKKRVVQVWFQNSRAKDKKSR
>Dm_cut
SKKQRVLFSEEQKEALRLAFALDPYPNVGTIEFLANELGLATRTITNWFHNHRMRLKQQV
>Dm_onecut
PKKPRLVFTDLQRRTLQAIFKETKRPSKEMQVTIARQLGLEPTTVGNFFMNARRRSMDKW
>Dm_dve_HD1
KTRMRTSFDPEMELPKLQKWFADNPHPSRQQIQTYVVQLNALESRRGRKPLDVNNVVYWFKNARAAQK
RAE

```
>Dm_dve_HD2
RKRNRTFIDPVTEVPKLEQWFAMNTHPSHNLILKYTEDLNTMPYRQKFPRLESKNVQFWFKNRRAKCK
RLK
>Dm_nub
RRKKRTSIETTIRGALEKAFLANQKPTSEEITQLADRLSMEKEVVRVWFCNRRQKEKRIN
>Dm_pdm2
RRKKRTSIETTVRTTLEKAFLMNCKPTSEEISQLSERLNMDKEVIRVWFCNRRQKEKRIN
>Dm_vvl
KRKKRTSIEVSVKGALEQHFHKQPKPSAQEITSLADSLQLEKEVVRVWFCNRRQKEKRMT
>Dm_acj6
GEKKRTSIAAPEKRSLEAYFAVQPRPSGEKIAAIAEKLDLKKNVVRVWFCNQRQKQKRIV
>Dm_pdm3
KRKRRTSFTPQALELLNAHFERNTHPSGTEITGLAHQLGYEREVIRIWFCNKRQALKNTV
>Dm_schlank
GIRSSRPKKAANVPILEKTYAKSTRLDKKKLVPLSKQTDMSEREIERWWRLRRAQDKPST
>Dm_pros
TPLHSSTLTPMHLRKAKLMFFWVRYPSSAVLKMYFPDIKFNKNNTAQLVKWFSNFREFYYIQM
>Dm_so
GEETSYCFKEKSRSVLRDWYSHNPYPSPREKRDLAEATGLTTTQVSNWFKNRRQRDRAAE
>Dm_Optix
GEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAKATGLNPTQVGNWFKNRRQRDRAAA
>Dm_Six4
GEETVYCFKEKSRNALKDCYLTNRYPTPDEKKTLAKKTGLTLTQVSNWFKNRRQRDRTPQ
>Dm_CG11617
SRATKRLFTPDIKRMLKDWLIRRRENPYPSREEKKQLAAETGLTYTQICNWFANWRRKLKNSE
>Dm_ara
LAARRKNATRESTATLKAWLNEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Dm_caup
LAARRKNATRESTATLKAWLSEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Dm_mirr
NGARRKNATRETTSTLKAWLNEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Dm_vis
LRKRRGNLPKSSVKILKRWLYEHRYNAYPSDAEKFTLSQEANLTVLQVCNWFINARRRILPEM
>Dm_achi
LRKRRGNLPKTSVKILKRWLYEHRYNAYPSDAEKFTLSQEANLTVLQVCNWFINARRRILPEM
>Dm_hth
NQKKRGIFPKVATNILRAWLFQHLTHPYPSEDQKKQLAQDTGLTILQVNNWFINARRRIVQPM
>Dm_exd
ARRKRRNFSKQASEILNEYFYSHLSNPYPSEEAKEELARKCGITVSQVSNWFGNKRIRYKKNI
>Cg_Arx
RRYRTTFTSFQLEELERAFQKTHYPDVFMREELAMRIDLTEARVQVWFQNRRAKWRK
>Cg_Dmbx
RRCRTAFTNQQLSTLEKTFAKTHYPDVVMRERLAMMTNLPEARIQVWFKNRRAKYRK
>Cg_Drgx
RRNRTTFTLQQLEELEKAFAQTHYPDVFTREDLAMRINLTEARVQVWFQNRRAKWRK
>Cg_Gsc2
RRHRTIFTEEQLEELEKTFNKTHYPDVLLREELAMKIDLKEERVEVWFKNRRAKWRK
>Cg_Hopx
TRLPKIKEDQERVLQSNFSSNRNPTDLDQTLIAAEAGLSEEDVKLWYQHRLACWRQ
>Cg_Hbn
RRSRTTFTTYQLHQLERAFEKTQYPDVFTREELAMRLDLSEARVQVWFQNRRAKWRK
>Cg_Otp
KRHRTRFTPAQLNELERSFAKTHYPDIFMREELALRIGLTESRVQVWFQNRRAKWKK
>Cg_Otx
RRERTTFTRAQLDILESLFQKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCRQ
>Cg_Pax3_7
```

RRSRTTFTAEQLEELERAFERTHYPDIYTREELAQRTKLTEARVQVWFSNRRARWRK
>Cg_Pax4_6
RRNRTTFTSQQLDLLEQSFQKAHYPGVQTREELATKTNLSEARVQVWFSNRRAKWRR
>Cg_Pax6
QRNRTSFTAAQIEALEKEFERTHYPDVFARERLSQKIDLPEARIQVWFSNRRAKWRR
>Cg_Pitx
RRQRTHFTSQQLQELEATFARNRYPDMATREEIAAWTNLTEARVRVWFKNRRAKWRK
>Cg_Prop
RRHRTTFTQEQLQELENAFAKSHYPDIYCREELARITKLNEARIQVWFQNRRAKYRK
>Cg_Prxx
RRNRTTFNSNQLAALERVFERTHYPDAFVREELARRVNLSEARVQVWFQNRRAKFRR
>Cg_Rax
RRNRTTFTTYQLHELERAFEKSHYPDVYSREELAMKIDLPEVRVQVWFQNRRAKWRR
>Cg_Repo
PRHRTTFSPFQLKEMEKAFRKAPYPDVMTREELARRLALNESRVQIWFQNRRAKWRK
>Cg_Shox
RRSRTNFTLEQLNELERLFDETHYPDAFMREELSQRLGLSEARVQVWFQNRRAKCRK
>Cg_Uncx
RRTRTNFTGWQLEELERAFQDSHYPDVFMREALALRLDLVESRVQVWFQNRRAKWRK
>Cg_Vsx1
RRHRTIFTSYQVEELEKTFKEAHYPDLYAREVLALKIDLPEDRIQVWFQNRRAKWRK
>Cg_Vsx2
RRHRTIFTSYQLEELEKAFKDAHYPDVYAREVLALKTSLPEDRIQVWFQNRRAKWRK
>Cg_PRD1
RRFRTTFTADQLRCLENVFRLTHYPDVNSREELSQKTGLPEARVQIWFQNRRAKWRK
>Cg_PRD3
KKTRTNYTPVQVQTLEKENPYPDAEKMEQLSKELDIPENKLKVWFQNKRARWRR
>Cg_PRD4
TKPRTKYSPEQLELLEGAFLDSQYPDTETLEDLAERLDVSTDKVSVWFQNRRSKFKR
>Cg_PRD5
TRSRIKYRQDQLETLEESFRSHQYPDSESIENLAEKVGVSSERVAIWFQNRRAKFKR
>Cg_PRD6
KRNRTTFSTKQLQELEKVFRKTHYPDVFLREKLASKIKLPESRIQVWFQNRRAKWRK
>Cg_PRD7
QRRQRTTFTPYQITALEDLFSRTHYPDIFVREELALQINLCEARIQSKVAEG
>Cg_PRD8
RRNRTTFTSDQLRELEAVFQHTHYPDCTLREQIAEKVDLTEARVQVWFQNRRAKWRK
>Cg_PRD9
RRFRTTFSKGQLSELENIFLQNHYPDVFQRELLAMKLDLTESRVQVWFQNRRAKWRK
>Cg_Islet
TRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKK
>Cg_Lhx1_5
RGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERR
>Cg_Lhx2_9
KRVRTSFKHHQLRTMKSYFALNHNPDAKDLKQLAQKTGLSKRVLQVWFQNARAKYRR
>Cg_Lhx3_4
KRPRTTITAKQLEALKRAYNESPKPARHVREQLSAETGLDMRVVQVWFQNRRAKEKR
>Cg_Awh1_Lhx6/8
KRVRSSFTEEQLQILQANFRIESNPDSQELNRIAITAGVSRRVAQVWFQNARARQKK
>Cg_Awh2_Lhx6/8
KRVRTTFTEDQLQVLQANFQLDSNPDGQDLERIAQITGLSKRVTQVWFQNSRARQKK
>Cg_Lmx1
KRPRTILTTSQRRKFKQAFEANPKPCRKVREQLAAETGLTIRVVQVWFQNQRAKVKK
>Cg_Lmx2
KRPRTILTTSQRRKFKSAFELNPKPCRKVREQLAAETGLSVRVVQVWFQNQRAKVKK
>Cg_Pou2

RKKRTSIETNIRVALEKSFQQNPKPSSEEITSVADQLSMEKEVVRVWFCNRRQKQKR
>Cg_Pou3
RKKRTSIEVTVKGALENHFMKQPKPAAQEISQLAEQLQLEKEVVRVWFCNRRQKEKR
>Cg_Pou4
KRKRTSIAAPEKRSLEAYFAVQPRPSGEKIAQIAEKLDLKKNVVRVWFCNQRQKQKR
>Cg_Pou6
RKRRTSFTPQALEVLNNFFERNTHPSGAEMTELSDKLNYDREVIRVWFCNKRQALKN
>Cg_Six1_2
EETSYCFKEKSRTILREWYQHNPYPSPREKRELAEATGLTTTQVSNWFKNRRQRDRA
>Cg_Six3_6
EQKTHCFKERTRNLLREWYLQDPYPNPTKKRELASATGLTPTQVGNWFKNRRQRDRA
>Cg_Six4_5
EETIYCFKEKSRQSLKECYKNNRYPTPDEKRNLAKKTGLTLTQVSNWFKNRRQRDRT
>Cg_Irx1
AARRKNATRETTSALKAWLYEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKK
>Cg_Irx2
NARRKNATRETTNTLKAWLYEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKK
>Cg_Irx3
GGRRKNATRETTAPLKAWLKEHMKNPYPTKAEKVMLAIVTKMTLTQISTWFANARRRLKK
>Cg_Irx4
AARRKNATRETTNTLKAWLYEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKK
>Cg_Meis
IMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQ
>Cg_Mkx
MTRPLKQWLCKNRDNPYPSKLEKDILARESQMTNTQVSNWFANARRRLKN
>Cg_Pbx
RRKRRNFSKQATELLNEYFYSHLSNPYPSEEAKEELARKCGITVSQVSNWFGNKRIRYKK
>Cg_Pknox
KSKRGVLPKHATQVMKSWLFQHIVHPYPTEDEKRQIANQTNLTLLQVNNWFINARRRILQ
>Cg_Tgif
RRRRGNLPKEAVRILKTWLYEHRYNAYPSDQEKVYLSSATNLTVLQVCNWFINARRRILP
>Cg_TALE1
RPRSRQLNSKATSIMSHWFEKNIDHPYPSDEQKEQLAREGGITVAQVKAWFANKRNRTSN
>Cg_TALE2
SKKHPSLPKEAVAIMLEWLRQHKDNPYPNDDEKAMLIKQTGLTINQINYWFTNARRRILP
>Cg_TALE3
QRHNQPLNLKAVRIMTEWYDRHEENPYPSKSEKEIMAKEGGISVTQVKSWFANKRNRSNN
>Cg_TALE4
RKQRASFPKKAILIMEDWYEKNVSHPYPDSITVELIAIQGGITGEQVKKWFGNKRNRSNN
>Cg_TALE5
TLKEWFSRHESNPYPSKQQREELCEETGLTDYQVKTWFSNARRKVKH
>Cg_TALE6
ARKTRLLPKRSVKILDGWFTDNISNPYPSRDQTIRLALDCGLTVEQVRKWFANKRNRSRN
>Cg_TALE7
ALNNWYQSHINYPYPTDQEVQELVTLSGLSRKQVKKWMANKRVRCFN
>Cg_TALE9
LKNRTLLTKKAVDMMEEWYLSNLDHPYPCHKVVQSLAVFGNIREEQVRKWFANKRTRQGR
>Cg_TALE10
IKPRPLLSKRAVQLMEKWYDNHLEHPYPNIDTIEQLATTGNITPEQVKKWFANKRNRSNN
>Cg_TALE11
LRSRPSLSQKAVQLMEEWYYSHRDHPYPPHHIIQDLARKGGVREEQVKKWFSNKRNRSRR
>Cg_TALE12
IRSRPSLSQRAVRLMEEWYYSHRDHPYPPQHIIQDLARRGGVKEEQVKKWFSNKRNRSRR
>Cg_TALE13
LRSRPSLSQKAVQLMEEWYYSHRDNPYPPSHIIQDLARKGGVREEQVKKWFSNKRNRSRR
>Cg_TALE14

ILSKWYNEHIRYPYPSDLEVEQLAEQTELSPQQIKKWMANKRVRNYN
>Cg_Cmp1
RRTRVFIDPMSEIPKLEKWFAEDTHPSSYMIDKYCEELNKSEYRQKFPKLEAKNVQLWFKNHRAKVKR
>Cg_Cmp2
KDSRLTFHPQHEIPQMRKWFHSCKTPSDEKLKFFAAELNKGHVRQERPNVTVAKLKIWWKNERQREKR
>Cg_Cux
KRPRVFFTEEQKDKLRMAYNQDPYPNQNTIEALANELNVGVKTVINWFHNHRMRAKQ
>Cg_Onecut
KKPRLVFTDIQRRTLHAIFKETKRPSKEMQATIAQQLGLEVTTVANFFMNARRRSLD
>Cg_CUT1
GSGRRGKFTLFQLMVLCRFFEEDQNPSLNTRMLLAEKLLVSLERITIWFQNQRARGFP
>Cg_CUT2
RYSTFSMFALDILKEFYHINGFPSKSERTKICEIIGETEKRVQNWFKGQRKVDRRRD
>Cg_ProsX1
LDHWNGTLTPMHLRKAKLMFFYVRYPSSAILKVYFPDVKFNKNNTAQLVKWFSNFREFYYIQM
>Cg_ZF_ZEB_E_Box
KFRMRTHISEEQLAVLKSHYKENPRPRKFELIRIGKEIGKEKRVVQVWFQNMRARDRRQG
>Cg_Zfhx_HD1
QKRARTRINDEQLKILRAHFDINNSPSEEQINAMSEQSGLPQKVIKHWFRNTLFKERQRN
>Cg_Zfhx_HD2
KRANRTRFTDYQIKVLQEYFEQNAYPKDDELDHLSQMLNLSPRVIVVWFQNARQKARKIY
>Cg_Zfhx_HD3
DKRLRTTILPEQLDYLYQKYQIDCNPSRKQLETISAEVGLKKRVVQVWFQNTRARERKGQ
>Cg_Zfhx_HD4
GKRYRTQMTSLQVRIMKSIFIDYKTPTMAECEMLGREIGLPKRVVQVWFQNARAKEKKAK
>Cg_Cers
IKDGRPKTVQPISALENSYRIKKFPSEDTVKGLEKQTDMTEREIQRWFRMRRNQERP
>Bf_Alx
KRRNRTTFTSFQLEEMERVFQKTHYPDVYMREQLALRADLTEARVQVWFQNRRAKWRKKE
>Bf_Aprd1
PKRNRTTFSTRQLQELERAFRQTHYPDIFMREELAMRVNLPESRVQVWFQNRRAKWRKRE
>Bf_Aprd2
PRRNRTTFTTFQIHELELAYDRTPYPDLSMREELANKLELTEARIQVWFQNRRAKTRRQE
>Bf_Aprd3
KRRARTVFSAQQLKRLEEEYKKCSYPDKERRQTVAEATGLTEGCVRIWYQNRRAKDHRLK
>Bf_Aprd4
GRRLRINFSAQQQSTLEAMFAQNRYPDADTRDRLAQSLGVAESRVIVWFQNKRARSRRQD
>Bf_Aprd5
AKKTRTTFKYHQVSALEQVFAMTHYPDNDTLDWLTTTLQLPEAKIKVWFQNKRARWKKQR
>Bf_Aprd6
RRRCRTKYSKEQLMAMEELFMKNRYPDIVEREDLAERIGLTEDRIQVWFQNRRSRWRARE
>Bf_Arx
QRRYRTTFTSYQLEELERAFAKTHYPDVFTREELAMRVDLTEARVQVWFQNRRAKWRKRE
>Bf_Dmbx
QRRSRTAFTSQQLAALEKCFQKTHYPDVVMRERLAMCTNLPESRVQVWFKNRRAKWRKRQ
>Bf_Drgx
QRRNRTTFTLHQLEQLEAVFAQTHYPDVFTREELAMKINLTEARVQVWFQNRRAKWRKQE
>Bf_Gsc
KRRHRTIFTEEQLELLEKTFEKTHYPDVLLREELAMKVELKEERVEVWFKNRRAKWRKQQ
>Bf_Hopx
PPGQMSRISDDQERILEYNFGHLTKTPDGITLEIIAAESGLSVEETAKWFQHRYALWRQKE
>Bf_Otp
QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELALRIGLTESRVQVWFQNRRAKWKKRK
>Bf_Otx
QRRERTTFTRAQLDVLEALFAKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCRQQA
>Bf_Pax3/7

QRRSRTTFTPEQLEELEKAFERTHYPDIYTREELAQRTKLTEARVQVWFSNRRARWRKQQ
>Bf_Pax6_
LQRNRTSFTQEQIEALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREE
>Bf_Isx
MRRVRTTFSAEQLKALEDVFKVTHYPDVHTREQLARKTKLPEARVQIWFQNRRAKWRKYE
>Bf_Phox
QRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARVQVWFQNRRAKFRKQE
>Bf_Pitx
QRRQRTHFTSQQLQELEASFARNRYPDMATREEIAAWTNLTEARVRVWFKNRRAKWRKRE
>Bf_Prrx
QRRNRTTFTSQQLQALEKVFERTHYPDAFVREELARRVNLSEARVQVWFQNRRAKFRRNE
>Bf_Rax
HRRNRTTFTTFQLHELERAFEKSHYPDVYSREELAMKINLPEVRVQVWFQNRRAKWRRQE
>Bf_Repo
KTRYRTTFSTYQLEELEKAFERAPYPDVFAREELAMRLNLSEARVQVWFQNRRAKWRKRE
>Bf_Shox
QRRSRTNFTLEQLQELERLFDETHYPDAFMREELSQRLGLSEARVQVWFQNRRAKCRKQE
>Bf_Prop
RRRHRTTFTQEQLAQLEEAFSKSHYPDIYVREELARATKLNEARIQVWFQNRRAKYRKQE
>Bf_UncxA
RRRTRTNFNSWQLEELERAFQESHYPDVFMREALALRLDLVESRVQVWFQNRRAKWRKKE
>Bf_UncxB
QRRSRANYSQWQLEELEKAFETTQYPDIFMREALALRLDLIEARVQVWFQNRRAKLRRQL
>Bf_UncxC
RMRVRTNFTSWQLEALERAFETTHYPDIFMREALALRLDLMESRVQVWFQNRRAKWRKQE
>Bf_Vsx
KRRHRTIFTSYQLEELEKAFNEAHYPDVYAREMLAMKTDLPEDRIQVWFQNRRAKWRKKE
>Bf_Isl
PTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKS
>Bf_Lhx1/5
RRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMK
>Bf_Lhx2/9-a
TKRIRTSFKHHQLRTLKSYFAINHNPDSKDLQQLAQKTGLTKRVLQVWFQNARAKHRRNA
>Bf_Lhx2/9-b
QKRMRTSFKHHQLRAMKSYFALNHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKYRRNL
>Bf_Lhx3/4
QKRPRTTITAKQLETLKQAYQNSPKPARHVREQLSQETGLDMRVVQVWFQNRRAKEKRLK
>Bf_Lhx6/8
PKRVRTTFTEEQLRVLQANFNIDSNPDGQDLERIAQITGLSKRVTQVWFQNSRARQKKYG
>Bf_Lmx
PKRPRTILTTQQRREFKALFEVSPKPCRKVRETLAAETGLSVRVVQVWFQNQRAKMKKLA
>Bf_POU1
KRKRRTTIGLSAKEALERHFLKQNKPSSQDIVRVAESLRLDKEVVRVWFCNRRQREKRVK
>Bf_POU2
RRKKRTSIETNVRVALEKAFIQNPKPTSEEIGIIAEQLGMEKEVVRVWFCNRRQKEKRIN
>Bf_POU3
KRKKRTSIEVTVKGALESHFLKQPKPSAQEIAQLADSLQLEKEVVRVWFCNRRQKEKRMT
>Bf_POU3L
KRKKRTSIEVSVKGALETFFYKQPKPSAIEISQLSEGLNLDKEVVRVWFCNRRQKERRMS
>Bf_POU4
KKRKRTSIAAPEKRSLEAYFAVQPRPSSEKIAAIAEKLDLKKNVVRVWFCNQRQKQKRMK
>Bf_POU6
KRKRRTSFTPQALEVLNAHFEKNTHPSGAEMTALAEKLNYDREVIRVWFCNKRQALKNTI
>Bf_Hdx_HD1
PFQERQAFSNKQLETLKRYYENGMNSQRLANREKMLACATETGLEFDTVRNWVGNYRRKLRLDA
>Bf_Hdx_HD2

GPFERTVFTEEQVRLLTEHYENGMNCQAKSCLDKMLAVAKELNLEFDVVRNWVGNQRRKRRMEA
>Bf_Tcf_Hnf1
ARRNRFKWGPASQRILYEAYQRQRNPSKEEREALVEECNRAECLQRGVSPSQASGLGSNLVTEVRVYN
WFANRRKEEAFRH
>Bf_Ahnf
PRKRRFTWPIEALNILEKFYETNPYPSVHDYQIIIASLEPVGCSPTPDQVRFWYQNRRRKDESSV
>Bf_Hmbox1A
RRGHRFNWPEACITIMEKYFEENQYPDEKKREEITNACNSVIQKPGVELPAQMLVNSARVYNWFANRR
KDVKRRH
>Bf_Hmbox1B
RRGSKFHWPPSAVMIVEKYFQQNPYPTEGERDEIANACNSVLQKPGEDLPSYKLVSPVRVQTWFINRR
REAKIKQ
>Bf_Six1/2
GEETSYCFKEKSRGVLREWYAHNPYPSPREKRELAEATGLTTTQVSNWFKNRRQRDRAAE
>Bf_Six3/6
GEQKTHCFKERTRSLLREWYLQDPYPNPQKKRELAQATGLTPTQVGNWFKNRRQRDRAAA
>Bf_Six4/5
GEETVYCFKEKARQALKEMYNNNRYPTPDEKRNLAKKTGLTLTQVSNWFKNRRQRDRTPS
>Bf_Atale
HGRKMLTLPREAKKKLQAWFLDHLYHPYPSGAEKYDLAKLTDLTVKQVSTWFQNTRRRLHQQQ
>Bf_IrxA
DGARRKNATRETTATLKAWLMEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Bf_IrxB1
DGTRRKNATRDATSTLKAWLNEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Bf_IrxB2
DGTPRKNATRDATSTLKAWLNEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>Bf_Meis
RQKKRGIFPKVATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPM
>Bf_Mkx
RVRHKRQVMQDMARPLKQWLIKHRDNPYPTKTEKILLALTSQMTLVQVSNWFANARRRLKNTV
>Bf_Pbx
NRRKRRNFSKQATEVLNEYFYSHLSNPYPSEEAKEELARKCGITVSQVSNWFGNKRIRFKKNI
>Bf_Pknox
RNGKRGILPKQATDVMRSWLFQHIVHPYPTEDEKRAIANQTNLTLLQVNNWFINARRRILQPM
>Bf_Tgif
ARKRRGNLPKEAVQILRAWLYDHRYNAYPTDAEKLDLAREAGLTVLQVCNWFINARRRILPDM
>Bf_Onecut
PKKPRLVFTDLQRRTLHAIFKENKRPSKEMQAQIAKQLGLDLSTVCNFFMNARRRSQDKW
>Bf_Cux
HKKQRVVLSPEEKEALRKAYEQEPYPSPSTIEYLAAKLNLRPCTVTNWFHNYRSRLRRGS
>Bf_Acut
QCEEPTRFTLSQQNVLQELFSRRKYPTESEIKSLAEELALSQRVVSTWFQNRRCQYRNKY
>Bf_Compass_HD1
KRFRTTFSLEHEIPKLMAWFREDPRPDPMKINRYLQELNNSEVRKHQAPLRYQAVHIWFKNARAKYKK
EG
>Bf_Compass_HD2
TRRRLIINPVSELPRLQTWFNVNPKPSRSEMKKMAEELNASEWRRNQPFELKSVAIWFKNARAKVAKK
E
>Bf_Azfh_HD1
FMPANETTLKEWEHLLKASYACNQYPSREEIQRLAMETGQEFALIRTWFAYERQNQKLGP
>Bf_Azfh_HD2
PRNVYYKQSTDHTAYMEKELQKNPKPTVQDMSRIADKTGLSRGLVIYWFKKRLNLAGLNL
>Bf_Tshz
RRGRHSSWNPVHQMILQAYIVSNMDTAEGRPRSPPPDISHDDKVGIAHATGLTMTAINHWVANVRYRL
RHAG
>Bf_Zeb

KAKSELDLSSEHLTVLKAYYALHAQPSPEELIKISSVVGLPKDFVESWFANMRERNEPSG
>Bf_Zfhx_HD1
GKRPRTRITDEQLKILRAHFDINNSPSEDQILEMSEKSGLPTKVIKHWFRNTLFKERQRN
>Bf_Zfhx_HD2
RRAGRTRFTDYQIKVLQEFFEQNAYPKDDDLEHLSKILNLSPRVIVVWFQNARQKARKSF
>Bf_Zfhx_HD3
NKRLRTTITPEQLEILYQKYQIDCNPTRKMLDLIAREVGLKKRVVQVWFQNTRARERKGQ
>Bf_Zfhx_HD4
GKRYRTQMSNLQLKMLKSAFTEYKTPTMQECEILGNEIGLPKRVVQVWFQNARAKDKKSK
>Bf_Zhx_HD1
ATISVSNADSHSDAVLREYFNKFPYFSEKELEEIVDKTGMTEDEVRTWYTAQRQQLSVNW
>Bf_Zhx_HD2
GEGGNWKKTPEQLQLLKRAFGKTQFPTPLEVDKLVRATGLTSGVVKKWFSDTRYHVKHSK
>Bf_Zhx_HD3
EWLERKGKSEYQLDLLKINFQQNPYPSEFDVDKLREETKMTAREIGRWFNDCRREKGLPQ
>Bf_Zhx_HD4
GRVKRRFKTEYQLRLLKESFQRCPHPDDLEIERLMRLTRLSKQDIRNYFSDKRYDVFKRN
>Bf_Zhx_HD5
DGKIDGKKTAFQLDVIKASFSLNQRPGDTEVARINKLTKLSRKEILCWFGDTRYAFKHGK
>Bf_Zhx_HD6
MNPPKPLTNVEKLKMLTEQYERCPFILYHDAEELAKEIDLRVEQVLSWFERKREEEKKDT
>Bf_Cers
LGIPGERKYAEPNAILEKVFTSITKNPDEKRLQGLAKQLDWSERQVQRWFRRRRNQDRPTL
>Bf_Prox
SPPHPGSLTPTHLKKAKLMFFYTRYPSSSTLKQHFPDVKFNRCITSQLIKWFSNFREFYYIQM
>Nv_170_HNF
SRRNRFKWGPASTNILYQSYEQQRNPSKEEREALVEACNRAECEQRGFNLVTESRVYNWFANRRKEET
FRM
>Nv_028_ISL
PTRVRTVLNEKQLHTLRTCYNANPRPDAMMKEQLVEMTGLSPRVIRVWFQNKRCKDKKKA
>Nv_127_LHX1/5
RRGPRTTIKAKQLEALKSTFAATPKPSRNIREKLAQETGLNMRVIQVWFQNRRSKERRLK
>Nv_044_LHX6/8
AKRVRTTFTEDQLQILQANFNIDSNPDGQDLERIAQLTGLSKRVTQVWFQNSRARQKKYG
>Nv_055
SKRARTFISNDQLAFLKVAYASSPKTTLRDRERIAKETGLDMRVVQVWFQNRRAKDKRLS
>Nv_020_POU1
RRKRRTTIGLAAKEALENHFMKQTKPSSPEIVRIADGLRLDKEVVRVWFCNRRQREKRVK
>Nv_125_POU3a
KRKKRTSIEVAVKGALENHFCKSPKPTAQEISALAENLGLDKEVVRVWFCNRRQKEKRMT
>Nv_075_POU3b
KRKRRTIIEKNVKGVLENHFEKMPRPSTSDISSLAESLGLDREVVRVWFCNRRQKERRVS
>Nv_124_POU4
KKRKRTSIGAAEKRSLEAYFAMNPRPSSDKIASIAEKLDLSKNVVRVWFCNQRQKKKRMK
>Nv_084_POU6
KRKRRTSFSNEALRLLISHFEQNPKPSSSEIAQIASKLGLEPVTVRVWFCNRKQMLKRMA
>Nv_120_ALX3/4
KRRNRTTFTAYQLEEMERVFQKTHYPDVYTREQLALRCALTEARVQVWFQNRRAKWRKRE
>Nv_100_AL
QRRYRTTFTSYQLEELERAFAKTHYPDVFTREALAVKIDLTEARVQVWFQNRRAKWRKRE
>Nv_085_CEH10
KRRHRTIFTNYQLEELEKAFKESHYPDVYARENLSLKIDLPEDRIQVWFQNRRAKWRKKE
>Nv_090_DMBXa
IRRTRTTFNQFQLDTLERAFSRTHYPDVLLREQLAVYTNLPESRIQVWFKNRRAKYRKSK
>Nv_063_DMBXb
SRRTRTAFTHQQLTALEKVFSKTHYPDVEVREQLATSTNLQEARIQVWFKNRRAKYRKDQ

```
>Nv_086_DMBXc
PRRMRTCFTPYQLQVLENTFCNTHYPDVMLREQLASYVNLPEARIQVWFKNRRAKHRKND
>Nv_151_DMBXd
SQITRTRFTPYQTQVLNETFSSSAYIDASTCGQLARLLGISSRSIQVWFKNKRYKLRIQA
>Nv_064_DMBXe
HRRTRTAFTHQQLQILESTFSKTHYPDVVMREQLAAYINIPESRIQVWFKNRRAKYRKQV
>Nv_003_DMBXf
HRRTRTAFTHQQLQILESTFSKTHYPDVVMREQLAAYINIPESRIQVWFKNRRAKYRKQV
>Nv_038_DUXa
HRRQRTFFSKEQTVILEGAFQYERFPGIQIREKLARELDIDESRIQVWFQNRRSRQNRQK
>Nv_005_DUXb
HRRQRTFFSKEQTVILEGAFQYERFPGIQIREKLARELDIDESRIQVWFQNRRSRQNRQK
>Nv_011_DUXc
HRRQRTFFSKEQTVILEGAFQYERFPGIQIREKLARELDIDESRIQVWFQNRRSRQNRQK
>Nv_068_GSC
KRRHRTIFTEEQLELLETTFQKTHYPDVLLREELAMKVDLKEERVEVWFKNRRAKWRKQK
>Nv_079_HBN
VRRSRTTFTTYQLHQLERAFEKTQYPDVFTREELALRLDLSEARVQVWFQNRRAKWRKRE
>Nv_047_OTP
QKRHRTRFTPAQLNELERCFARTHYPDVFMREELAARIGLTESRVQVWFQNRRAKWKKRK
>Nv_114_OTXa
QRRERTTFTKNQLEVLEELFAKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKTRQLE
>Nv_123_OTXb
QRRERTTFTKNQLEILEELFAKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKARQQA
>Nv_078_OTXc
QRRERTTFTKNQLEILEELFAKTRYPDIFMREEVAIKINLPESRVQVWFKNRRAKLRQLS
>Nv_014_PAX3/7
QRRSRTKFTSKQVDELEKAFLKTQYPDVYTREELAQRLNLTEARVQVWFSNRRARLRKKK
>Nv_101_PAX3/7
PRRSRTRFTVSQTDELERAFRKTHYPDIYAREELAQRLGLSEARVQVWFSNRRARLRKER
>Nv_022_PAX4/6
LRRNRTTFTPDQLEMLEKEFEKSHYPDVATREELANKIDMSEARVQVWFSNRRAKWRRHQ
>Nv_148_PAX4/6
SRRQRTNFTDEQIEKLEKVFEKTHYPDVFTREELAQQVNLSEARIQVWYSNRRAKWRKEG
>Nv_105_PTX
QRRQRTHFTSFQLQQLEGTFGRNRYPDMQMREEIALYTNLTEARVRVWFKNRRAKWRKKE
>Nv_118_REPO
KTRYRTTFSQYQIEELERAFDKAPYPDVFAREELAAKLGLTEARIQVWFQNRRAKWRKRE
>Nv_129_RX
LRRNRTTFTTFQLHELERAFEKSHYPDVYTREELALKISLPEVRVQVWFQNRRAKWRRQE
>Nv_059_UNC4
RMRVRTNFSPWQLEELEHAFETTHYPDVFMREALAMRLDLTEARVQVWFQNRRAKWRKRE
>Nv_002
QRRNRTTFTKQQLQELEKVFEKKHYPDIALREELAAKINISEARIQV
>Nv_031
RRRCRTVFSTEQLAILEEGFQKQHFPDNKLRQAIATRAGLPEDRVQVWFQNRRAKEKRLL
>Nv_037
KRRHRTTFSKLQLDSLEEAFSRSQYPDVFTREQLAKRINLNEARVQVWFQNRRAKHRKQE
>Nv_052
IKRKRRNLTKRQQKILETAYSTIKYPTIEDRQRLETSTQLSEDRIQVWFQNRRAKDRRLQ
>Nv_058
HRRKRTTFTHDQLQLMEAYFHNNRYPGIEQREGLAEKIQVSESRLQVWFQNRRSKWRKHQ
>Nv_074
QRRNRTKFTAYQLEQLEDAYQKAKYPDVQARETLAQRLGVAESRVQVWFSNRRSKGKRKE
>Nv_146
GKRYRATFDKAQIFQMERVFLLNHYPDVAARSELSRRTGLSESQVQIWFQNRRAKWRKQQ
```

```
>Nv_073_SIX1/2
GEETSYCFKEKSRNILREWYSHNPYPSPREKRELAEGTGLTTTQVSNWFKNRRQRDRAAE
>Nv_128_SIX3/6
GEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQVGNWFKNRRQRDRAAA
>Nv_030_SIX4/5a
GEETVYCFKEKARAALKDCYEQNKYPTPQEKRLIAKQTNLTLKQVSNWFKNRRQRGRIPS
>Nv_061_SIX4/5b
GEKTIYFFKEKVRTVLKECYEHKKYPTLKEKRVIATQTNLTLRQVRNWFRNRRHRDRISS
>Nv_093
GEETTYCFKEKSRSILNKAYVDSPYPTPREKHELAKMTDLTVTQVSNWFKNKRQRVRAAE
>Nv_108_IRX
AGARRKNATRETTSTLKAWLFEHRKNPYPTKGEKIMLAILTKMTLTQVSTWFANARRRLKKEN
>Nv_107_MEIS
SQKKRGIFPKAATNIMKAWLFQHLTHPYPSEEQKRSLAQETGLTILQVNNWFINARRRIVQPM
>Nv_040_PBX
YRRKRRNFSKQATEILNEYFYSHLSNPYPSEEAKEELARKCNISVAQISNWFGNKRIRYKKNI
>Nv_149_TGIF
PKRRRGNLPKDSVNVLRLWLWEHRFNAYPSEAEKQYLSKAANLSVLQVCNWFINARRRILPDM
>Nv_036
SKKSKRGVLPKQATSIMKTWLFQHIMHPYPTEDEKRSIAQQTNLTILQVNNWFINARRRILQPM
>Nv_143
ARKRRGKLPESARTRLMTWLFAHSNRPYPTEEEKSNLATITGLTPIQINNWFSNARRRILKKA
>Nv_XP_001637067.1
SKKIRRQFTETQKRTLFAIFKETKKPTKDMQKAVAEELGLDLNSVANFFMNARRRHTDA
>Nv_XP_001635937.1
GVKDKKINFTPNPFCEKVYQTINKFPPSDRILGLSKQIGWTTREVERWFRHRRMQSKPSL
>Nw_55290
KRRNRTTFTSYQLDELERIFQKTHYPDVYCREQLAARCNLTEARVQVWFQNRRAKWRKKE
>Nw_40829
KRRNRTTFTSYQLDELERIFQKTHYPDVYCREQLAVRCTLTEARVQVWFQNRRAKWRKKE
>Nw_40214
KRRHRTIFSEEQLDELETTFAKTHYPDVLLREQLALRVELKEERVEVWFKNRRAKWRKYK
>Nw_53544
PKRHRTRFTPAQLNELERCFTKTHYPDIFMREELAVRIGLTESRVQVWFQNRRAKWKKRK
>Nw_48073
QRRERTTFTRAQLEVLESLFTKTRYPDIFMREEVAMKINLPESRVQVWFKNRRAKCRQQQ
>Nw_49939
LQRNRTSFTPEQIDALEKEFERTHYPDVFAREKLAEKIGLPEARIQVWFSNRRAKWRREE
>Nw_24274
LQRNRTSFSVEQIDALESTFNESHYPDSNERERLAEKVGIPEAKIQVWFSNRRARWRRED
>Nw_27707
PRRQRTHFTSQQLQELESMFTRNRYPDMATREEIAAWTGLTESRVRVWFKNRRAKWRKKE
>Nw_29086
HRRNRTTFTTYQLHELERAFEKSHYPDVYAREEIALKINLPEVRVQVWFQNRRAKWRRQE
>Nw_21928
KRRHRTIFTALQLEELEAAFGDAHYPDVYARELLSIKTDLPEDRIQVWFQNRRAKWRKKE
>Nw_54213
KKREKKVFDPMQIALLEHNYMEIRYPDLQRRKELAAVVGIDEKRIQVWFQNRRARTRRKH
>Nw_18856
KPRHRTTFTSWQLQEMERAFERAPYPDSPMREQLAINIGLTDARVQVWFQNRRAKWRKQT
>Nw_47651
KKRVKKVFEPGQIALLEHNYILNQYPSPQGRKALSNEVGIDEKRIQVWFQNRRARTRKKH
>Nw_54455
TKRVRTSFKNPQLRVLKQYFQLNHNPDAKELKQLSQKTSLSKRVLQVWFQNARAKYRRQV
>Nw_51594
TKRVRTSFKHHQLRTMKSYFNLNHNPDAKDLKQLSQKTCLSKRVLQVWFQNARAKYRRLM
```

>Nw_51594
KRRHRTMFHRYQLEQLKESFKTNSNPSSIDVEYIAKSVNSTKRIIQVWFQNARSKQQRRL
>Nw_45303
PTRVRTVLNEKQLSTLRTCYAANARPDALMREQLSDMTGLSGRVIRVWFQNKRCKDKKRS
>Nw_22501.1
PTRVRTVLNEKQLQTLRQCYAANSRPDALMREQLSDMTSLSGRVIRVWFQNKRCKDKKRS
>Nw_61503
RRGPRTTIKAKQLDTLKAAFDATPKPTRHIREQLAQDTGLNMRVIQVWFQNRRSKERRMT
>Nw_38224
QKRPRTTITAKQLETLKTAYSNSPKPARHIREQLSHDTGLDMRVVQVWFQNRRAKEKRLK
>Nw_50222
RSKRVILTTTQRRIFRTAFDVSQKPSRKVKERLADETGLPVRIVQVWFQNERAKGKKLAK
>Nw_14051
VGRKRVILTHNQRRVLRAAFIVSQKPSRKVKEDLGGATQLPVRIVQVWFQNERAKVKKLA
>Nw_49053
GSRKRVVLSPEQRKIFQLAFDASAKPSRKVREDLANKCSVSMRFIQVWFQNQRAKMKKLA
>Nw_7973
GRKRVLLTNPQRRIFRSAFTISDKPCRRVKEGLAMMTQLPIRIVQVWFQNERAKMRKLSM
>Nw_48370
HNNKRVNLNNMQRHILLAAYETSHKPTRSMKETLAEQTGLPMRIVQVWFQNERAKCKKMA
>Nw_52392
GSRKRVVLTPEQRKIFQQAFDMSPKPSRKVREQLASECTVSMRFIQVWFQNQRAKMKKLA
>Nw_27996
KRKKRTSIEMSVKGALENHFMKQAKPCAPEIQALADSLQLDKEVVRVWFCNRRQKEKRMT
>Nw_9311.1
RKRKRTSIASPEKRSLEAYFAVQPRPSSEKIGQIADKLDLKKNVVRVWFCNQRQKQKRLK
>Nw_52441
KRKRRTSFPPAAIDILNSYFLKNSHPSGHEFTQLSRLLEHDREVIRVWFCNKRQAMRHYA
>Nw_53834
RRKRRTSLNPDVREKLEAHFEVNSRPGANQIFVISEHLSLDREVVRVWFCNRRQQKIKRVN
>Nw_48953
RRRERFVWREECMPVLDKFFNENQYPDDRKREDIASACNGVLEAPGVHIPGERLVTPAKVPAKVYNWF
VNRRNSYKIKMRQM
>Nw_36924
GEQKTHCFKERTRNLLREWYLQDPYPNPQKKRELAAATGLTPTQVGNWFKNRRQRDRAAA
>Nw_42137
GEQKTHCFKERTRNLLREWYLQDPYPNPQKKRELAGATGLTPTQVGNWFKNRRQRDRAAA
>Nw_28871
LGLRRKNATRETTSPLKAWLYEHRKNPYPTKAEKIMLAIITKMTLTQVSTWFANARRRLK
>Nw_10005
VGCRRKNATRETTSALKTWLADHRKNPYPTKAEKIMLAVITKMTLTQVSTWFANARRRLK
>Nw_26582.1
SGMRRKNSSRETTQYLKSWLKHHQKNPYPTKGEKIMLAILTKMTLTQVSTWFANSRRRIK
>Nw_36182
NGIRRKNATRETTSTLKAWLYEHRKNPYPTKGEKIMLAVITKMTLTQVSTWFANARRRLK
>Nw_48316
PGIRRKNATRETTSTLKAWLYEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLK
>Nw_52332.1
ARRKRRNFNKTATEVLNEYFYSHLSNPYPSEEVKEDLAKKCNLSISQVSNWFGNKRIRYK
>Nw_52332.1
ARRKRRNFNKTATEVLNEYFYSHLSNPYPSEEVKEDLAKKCGLSISQVSNWFGNKRIRYK
>Nw_52537
ARRKRRNFNKNATEFLNEYFYSHLNNPYPSEETKEELAKKCGLSISQVSNWFGNKRIRYK
>Nw_45317
KVSKRGILPKHATSVMRSWLFQHIVHPYPTEDEKRQLAGQTGLTLLQVNNWFINARRRILQPM
>Nw_46342

KQRLVFTDLQRRTLHAIFKETQRPSKEMQLQIASQLGLQLTTVSNFFMNARRRSVDKWKD
>Nw_35933
SKRLRYSLDDDSKRFLLSVYQHVTHPSQDTISRLATHLNMQPATVYNWFQNRRQRFKVKD
>Nw_60532_zfx_HD1
KRPRTRITDDQLKVLRTHFDINNSPNDDQIANISQETVLPVKVIKHWFRNTLFKERQRNK
>Nw_64721_zfx_HD2
KRAGRTRFTDWQIQKLQEFFEINAYPKDDDLDTLSKDLSLNPRVIVVWFQNQRQKARKSF
>Nw_50181_zfx_HD3
DRRLRTSITPKQLNTLYKWFKENSNPSRKMLEIISVDVSLKKRVVQVWFQNTRARQRKAS
>Nw_50181_zfx_HD4
DKRYRTQLSHLQIRVLKACYVDYKTPTMNECEELGKTIGLAKRVVQVWFQNARAKDKKSK
>Nw_51994
IRSSSNQLRPIDNLVLERVYLSVSRTPDDSRLKGLSKQLDWSEKQVEAWFHLRRAQDRAS
>Nw_44536
IKAKKPSNGGLINPVLEEAYGKCTRPNREAIKGMAKKADMEADRVERWFRRRRQQDQP
>Nw_40648.1
GIRSSNHLRAVDNLVLERVYASVSRNPDGNRIKGLSKQLDWNEKQVQAWFHLRRAQDRSS
>Nw_62117
QIKPKQRNSKSIHPLLEQVYKDKSSPPDKVIEELAKKLDSEPPKIRRWFSRRHRHDQP
>Nw_44308
NQGWATNLTPNHLKKAKLMFFYTRYPSSAVLRLHFPDVRFSRHNNAQLIKWFSNFREFFYLQI
>Nw_51275
ASFWTSNLTPEHLKKAKLMFFYTRYPTSAMLRINFPDVRFTRHINAQLIKWFSNFREYFYIQI
>Xb_DN84186
KRRNRTTFTSFQLDELERVFQKTHYPDVYLREQLALRADLTEARVQVWFQNRRAKWRKKE
>Xb_DN48278
KRRHRTIFTEEQLEELERTFEKTHYPDVLLREQLAVKVDLKEERVEVWFKNRRAKWRKQK
>Xb_DN86933
PKRHRTRFTPAQLNELERSFNKTHYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRK
>Xb_DN55770
QRRERTTFTRAQLEVLEALFSKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCRQQA
>Xb_DN77439
LQRNRTSFTPEQIDSLEKEFERTHYPDVFARERLAKDIDLPEARIQVWFSNRRAKWRREE
>Xb_DN86691
LQRNRTSFTPEQLESLEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREE
>Xb_DN49217
QRRQRTHFTSQQLQELEATFARNRYPDMATREEISVWTNLTEARVRVWFKNRRAKWRKKE
>Xb_DN88387
RRRHRTTFTQEQLNELEAAFAKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQE
>Xb_DN70414
KTRYRTTFTAWQLEEMERAFERAPYPDVFQREELAMKISLSEARVQVWFQNRRAKWRKLK
>Xb_DN67189
HRRNRTTFTTYQLHELERAFEKSHYPDVYSREELAMKVALPEVRVQVWFQNRRAKWRRQE
>Xb_DN14020
QRRTRTNFTLEQLEELEKLFDETHYPDAFMREELSSRLGLSEARVQVWFQNRRAKC
>Xb_DN12917
RRRTRTNFNSWQLEELEKAFLVSHYPDIFMREALAMRLDLIESRVQVWFQNRRAKWRKKE
>Xb_DN86983
RRRHRTIFTSYQLEQLEKAFTEAHYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKKE
>Xb_DN122491
RRNRMSFQPEQRAILEEAYRVNKYPDILHRAWISQMTNIPDLRIQIWFQNRRAKSRKREF
>Xb_DN102158
QRRNRTTFTLHQLEQLETAFAQTHYPDVFTREELALKINLTEARVQVWFQNRRAKWRKHE
>Xb_DN122218
QRRNRTTFTPQQLEELEQLFSKTHYPDIFVREDLAMRINLSEARVQVWFQ
>Xb_DN83632

VRRSRTTFTTFQLHHLERAFEKSHYPDVFTREELAASLHLSEARVQVWFQNRRAKWRKTE
>Xb_DN63074
QKRMRTSFKHHQLRAMKSYFSINHNPDAKDLKQLAQKTGLTKRVLQVWFQNARAKFRRNM
>Xb_DN91806
TTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTALSPRVIRVWFQNKRCKDKKKS
>Xb_DN95395
RRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMK
>Xb_DN11429
SKRPRTTITAKQLETLKNAYSNSPKPARHVREQLSNETGLDMRVVQVWFQNRRAKEKRLK
>Xb_DN81776
AKRVRTTFTEEQLQILQANFNLDSNPDGQDLERIAQLTGLSKRVTQVWFQNSRARQKKNN
>Xb_DN87622
PCRKRIILTTQQRRIFKAAFEVSSKPCRKVRESLAAETGLSVRVVQVWFQNQRAKVKKLA
>Xb_DN87006
RKKRTSIETNVRVALEKAFLSNPKPTSEEIQMLAESLCMEKEVVRVWFCNRRQKEKR
>Xb_DN83679
KRKKRTSIEVTVKGVLESHFQKQPKPSAQDITNLAESLQLEKEVVRVWFCNRRQK
>Xb_DN93613
RKRRTSFTPAALEILNVHFEKNTHPSGSEMTEIASKLNYDREVVRVWFCNKRQALK
>Xb_DN90583
RSRINWDPSAVNILEHHFNTKNRFPNAFERETIAANINGACHSKGRKVTAHHIQTWFNNRRNKRTGVS
>Xb_DN63639
GEETSYCFKEKSRTVLREWYSHNPYPSPREKRELSDGTGLTVTQVSNWFKNRRQRDRAAE
>Xb_DN23152
GEQKTHCFKERTRSLLREWYLQDPYPNPTKKRELAQATGLTPTQVGNWFKNRRQRDRAAA
>Xb_DN97213
EAQTTRLFQKNTLSVLKDFYQSDPYPNLNKIKDLARATRLSPIQIRTWFKNRRQRDRIAA
>Xb_DN120045
EETVYCFKEKSRQALKECYRRNRYPTPEEKRSLAKMTGLTLTQVSNWFKNRRQRDRAPK
>Xb_DN63799
GMMRKKSATRETTNALKAWLREHRRNPYPTKSEKIMLAIITKMTLTQVSTWFANARRRLK
>Xb_DN84998
GVARRKNATRETTSALKAWLHEHRKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLK
>Xb_DN64910
HQKKRGIFPKSATNVMRAWLFQHLSHPYPSEEQKKQLAQETGLTILQVNNWFINARRRIV
>Xb_DN79707
YQKKRGIFPKSATNILRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIV
>Xb_DN68543
ARRKRRNFNTNATQILNEYFYSHLSNPYPSEEVKEELARKCGLTVSQISNWFGNKRIRYK
>Xb_DN74939
KRGVLPKQATNIMRSWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQPM
>Xb_DN45414
KRRRGNLPKDAVNVLRLWLYEHRFNAYPTEQEKVELSEKASLSVLQVCNWFINARRRILPDM
>Xb_DN97994
RETKRPSKEMQMTIAQQLGLELSTVSNFFMNARRRSLDKW
>Xb_DN93733.1
TKRPRTRITDEQLKILRAYFDINNSPSEEQIIEMSEKSGLPQKVIKHWFRNTLFKERQRN
>Xb_DN93733.2
RRAGRTRFTDFQIKRLQEFFDTNAYPKDDDLDRLSKELNLNPRVIVVWFQNARQKARKSF
>Xb_DN93733.3
DRRLRTTITPEQLEILYEKYQENSNPSRRMLDIISREVGLKKRVVQVWFQNTRARQRKGQ
>Xb_DN93733.4
GKRFRTQMSNTQLKVLKHCFTEYRTPTMQECELLGSCIGLQKRVVQVWFQNARAKEKKSK
>Xb_DN87293
GLKIPARHTPEDNTILEKVYRSVSKQPDAKRIEGLAKQLDWSERQVQRWFRRRRNSDRPS
>Xb_DN112933

RLRAPPNATLEHFYLTSGKQPKQVEVELLSRQSGLSGRQVERWFRRRRNQDRPSL
>Xb_DN88028
SSGWNTNLTPSHLKKAKLMFFYVRYPNSATLRQHFPDVKFTRHNNAQLIKWFSNFREYFYIQI
>IP_DN38710
KRRNRTTFTQAQLEEMERIFQKTHYPDVYVREQLAVKCDLTEARVQVWFQNRRAKWRKKE
>IP_DN67235
KRRHRTIFTEEQLAELEAVFVRSHYPDVLAREKVATSLGLKEERVEVWFKNRRAKWRKQK
>IP_DN57662
AKRHRTRFTPAQLNELERCFNKTHYPDIFMREELAMRIALTESRVQVWFQNRRAKWKKRK
>IP_DN73075
QRRERTTFTRTQLEILEQLFSKTRYPDIFMREEVANRINLPESRVQVWFKNRRAKCRQQQ
>IP_DN54186
NQRNRTAYSQNQLELLEKEFEKSHYPDVYARERLSEATDIPESKIQIWFSNRRAKHRREE
>IP_DN12334
LQRNRTSFTPEQIDALEKQFEVTHYPDVFARERLAGKIALPEARIQVWFSNRRAKWRREE
>IP_DN51842
ARRQRTHFTSQQLQELETLFSRNRYPDMATREEISAWTNLPEAKVRIWFKNRRAKWRKKE
>IP_DN68900
ERRQRTQFTSHQLRELESMFARNRYPDMATRERLSAWVQLSETKVRIWFKNRRAKWRKKE
>IP_DN60252
RRRHRTTFTQEQLNELEAAFAKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQE
>IP_DN64595
KSRYRTTFSAEQLEQMERAFQLAPYPDVFTREELAARVGLTEARVQVWFQNRRARWRKER
>IP_DN19265
HRRNRTTFTTFQLHELERAFEKSHYPDVYSREELAARVNLPEVRVQVWFQNRRAKWRRQE
>IP_DN58162
RRRTRTNFTNWQLEQLELAFHRTHYPDVFVRETLALQLNLMESRVQVWFQNRRAKWR
>IP_DN66046
RRRHRTIFTNSQVEELEKAFEEAHYPDVYQREILSLKTDLPEDRIQVWFQNRRAKWRKKE
>IP_DN31321
RRRHRTIFTSYQIEELEKAFKEAHYPDVYARELLSIKTDLPEDRIQVWFQNRRAKWRKKE
>IP_DN73593
PQILPLHMLTDAEKTLERSFMRDHNPADYMYEIMAEEIGWNLNQVKQWYDRKKATWRSNE
>IP_DN71583
PRRSRTTFTAFQLHQLESIFLSNPYPDVARRDQLAAMLGLTESRIQVWFQNRRAKRRKAE

>IP_DN67846.1
SKRMRTSFKHHQLQAMKVYFNANHNPDAKDLKQLATKTGLTKRVLQVWFQNARAKFRRNL
>IP_DN67846.2
AKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRMM
>IP_DN56056
PTRIRTVLNEKQLHTLRTCYNANPRPDALMREQLVEMTGLSPRVIRVWFQNKRCKDKKRS
>IP_DN68501
NNNTRKPLKREIAEHLRAYFDKNSKPTKPEREELANQLGITERVIQIWFQNRRSKDRRVA
>IP_DN68178
RRGPRTTIKAKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMK
>IP_DN49678
NKRPRTTITAKQLEILKNAYKSSSKPARHIREQLSKDTGLDMRVVQVWFQNRRAKEKRLK
>IP_DN27657
SKRSRTTFSEDQLKILQANFNLDSNPDGQDLERIAQQTGLSKRVIQVWFQNARSRQKKHM
>IP_DN72227
PCRKRVILTTQQRRAFKAAFDLSSKPCRKVREQLARETGLSVRVVQVWFQNERAKVKKLA
>IP_DN75504
KKRKRTSIAAPEKRSLEAYFTVQPRPSGEKIAAIAEKLDLKKNVVRVWFCNQRQKQKRLK
>IP_DN18113
KRKRRCVIDAHTKKTLEQSFRDKPHPSPDETNALAEALGLDRETVRIWFTNRRYRERRSE

>IP_DN55409
KRKKRSCIDSNCRETLEQLFDQNNHPTPDEIVKYSEALNMEPETIRIWFSNRRQKQKRLA
>IP_DN65023
KKRDRFVWKENCVALLEKYYEANPYPDDTRREEIAAACNQLNGNSGSGQEERVVTAVKVYNWFANRRKEVKRKN
>IP_DN45560
GEETSYCFKEKSRNMLRDYYGHNPYPSPREKRDLAEATDLSITQVSNWFKNRRQRDRALE
>IP_DN26501
GEETVYCFKEKSRAKLKDSYRRNKYPSPEEKRSLAKSSGLTLTQVSNWFKNRRQRDRTPH
>IP_DN62049
GEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGNWFKNRRQRDRAAA
>IP_DN44510
SSIRRKNASRETTAALKAWLYEHRNNPYPTKGEKIMLAIITKMTLTQISTWFANARRRLKKEN
>IP_DN73792
SSPASPSAVKESTYPLRVWMEENRANPYPTKNQKLILALVTKLTLTQVSTWFANARRRLKKER
>IP_DN66612
MITRRKNATRETTSALKAWLNDHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>IP_DN70650
PQKKRGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQV
>IP_DN70276
ARRKRRNFSKQASEILNEYFYSHLNNPYPSEEVKEELAKKCCLSVSQISNWFGNKRIRYKKNI
>IP_DN69528
KKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQPM
>IP_DN71480
EKKKRGSLPKEAICIMRSWLEQHRYHAYPTEQEKLSLAQQTNLTPLQVCNWFINARRRVLPEI
>IP_DN68679
SRRRSKDQIEEDTVPLRHWLLLHSGHPFPTRVEKKRLARECGLQLKQVEDWFLNTRRRILKHH
>IP_DN31212
MKKSRLVFTDLQRRTLRAIFQESQRPSKESQAQIAKQLGLEVTTVSNFFMNARRRCGDRW
>IP_DN36933
QKKNRLVFTDIQRRTLFAIFRETKRPSKEMQLQISRQLGLEPTTVSNFFMNARRRCGDRW
>IP_DN75125.2
SKQPRTKISDEQLAILKQNFDISNSPSEEQYLAISVQTGLPMKVIKHWFRNTLFKERQKD
>IP_DN75125.1
FRRMRTTISPDQLEVLYEKYGENSNPSRSEMEAISAAVGLSRRVVQVWYQNTRARQRRGQ
>IP_DN75125.4
RKRHRTNLSMLQIKTLKSIFAEYKTPTMTECETIGSELNLSKRVVQVWFQNARSKEKKIQ
>IP_DN75125.3
KRATRTRFTDHQIKALQDFFDQNPYPRDDDLETLSKNLNLSPRVVVVWFQNQRQKARRSL
>IP_DN61481
RHRTRTKLPAAELALLKKIYETNPSPSKKQFAELARRTSLPIKVLKVWFQNTRRLTRNTC
>IP_DN55547
ERKKRTVYTDQQLDTLKEYYIRQPNPTSDDIERIASVIGLTKRNVQVWFQNKRARDKKNP
>IP_DN63567
GISPRGPRKPVHNAILERVFHTITQKPDEKRLKGLEKQLDWSPYQIRQWFARRRAQIRPTK
>IP_DN52339
AVGSTDNLTANHLRKAKLMFFFQRYPTSSTLRQYFTDVKFNRHNNAQLIKWFSNFREYYYIQI
>IP_DN66473
RKRARKSFTLQQKSVLSTEFSANQRPSRHRISQLASELNLTSASVANFFANSRQRNKHIS
>SR_DN122029
KRRNRTTFTQGQLEEMERIFQKTHYPDVYVREQLALKCDLTEARVQVWFQNRRAKWRKKE
>SR_DN120054
KRRHRTIFSEEQLKELEEAFLKCHYPDVVTREGVAQKLSLKEERVEVWFKNRRAKWRKQK
>SR_DN116171
PKRHRTRFTPAQLNELERCFNKTHYPDIFMREELAMRIALTESRVQVWFQNRRAKWKKRK
>SR_DN123029
QRRERTTFTRTQLEILEALFSKTRYPDIFMREEVANRINLPESRVQVWFKNRRAKCRQQQ

>SR_DN127496
SQRNRTAYTSEQVDALEKEFEKSHYPDVYARERLSTLTGIPETKIQIWFSNRRAKHRREE
>SR_DN116853
LQRNRTSFTPEQIEALEKQFEVTHYPDVFARERLAGKINLPEARIQVWFSNRRAKWRREE
>SR_DN122619
TRRQRTHFTSQQLQELETLFARNRYPDMATREEISAWTNLPEAKVRIWFKNRRAKWRKKE
>SR_DN126115
ERRQRTQFTSHQLRELESMFARNRYPDMATRERLSAWISLSETKVRIWFKNRRAKWRKKE
>SR_DN118878
RRRHRTTFTQDQLNELEAAFSKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKQE
>SR_DN108331
KSRYRTTFTSEQLDQMERAFECAPYPDVFTREELANRIKLTEARVQVWFQNRRARWRKER
>SR_DN126115
HRRNRTTFTTYQLHELERAFEKSHYPDVYSREELASRVNLPEVRVQVWFQNRRAKWRRQE
>SR_DN120882
YKRTRTQFSEDQICALEKLFEESHYPDSSARNELVAATGLTDSKIQVWFQNRRAKARKEK
>SR_DN113036
RRRTRTNFTNWQLEQLEMAFHRTHYPDVFVRETLAMQLNLMESRVQVWFQNRRAKWRKME
>SR_DN111202
RRRHRTIFSVSQVEELERAFEDAHYPDVVQRETLSNKTELPEDRIQVWFQNRRAKWRKKE
>SR_DN98097
RRRHRTIFTSYQIEELEKAFGDAHYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKKE
>SR_DN112801
CRRSRTTYTPWQLHQLESSFEDNPYPDVGQRDYLAGVMGVTESRIQVWFQNRRAKRRKLA
>SR_DN125451
TKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRLM
>SR_DN128068
SKRMRTSFKHHQLQAMKQYFNANHNPDAKDLKSLAQKTGLTKRVLQVWFQNARAKFRRNL
>SR_DN125302
PTRIRTVLNEKQLHTLRTCYNANPRPDALMREQLVEMTGLSPRVIRVWFQNKRCKDKKRS
>SR_DN118733
RRGPRTTIKAKQLETLKSAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMK
>SR_DN123163
KRGPGNHLKGPQADYLRSYYLTNQKPSKEERDAIAAEIGVEAKTIATWFQNRRSKEKIRR
>SR_DN124757
NKRPRTTITAKQLEVLKSAYKNSSKPARHIREQLSKDTGLDMRVVQVWFQNRRAKEKRLK
>SR_DN123001
SKRSRTTFSEDQLKVLQANFNMDSNPDGQDLERIASQTGLSKRVIQVWFQNARSRQKKHL
>SR_DN127948
PCRKRVILTTQQRRAFKAAFDLSSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKIG
>SR_DN116892
RRKRRTCIDVKTKKALEQFYNDKPHPTHEECCAIAEALNQDRNTIRIWFGNRRQRDKRLM
>SR_DN123138
KRKKRTSIEVSIKNALEHNFHIQPKPSAHEISALANSLQLEKEVVRVWFCNRRQKEKRMT
>SR_DN125354
KKRKRTSIAAPEKRSLEAYFTVQPRPSGEKISAIAEKLDLKKNVVRVWFCNQRQKQKRLK
>SR_DN121592
KKRDRFVWKESCVCLLEKFYHVNPYPDDAKREEIAQACNQLIDSTGQGHDEKLVTAVKVYNWFANRRKEIKRKN
>SR_DN87552
GEETSYCFKEKSRNMLREYYQHNPYPSPREKRDLAENTDLSVTQVSNWFKNRRQRDRALE
>SR_DN118860
GEETIYCFKEKSRNKLKESYKKNKYPTPEEKKALAKNSGLTLTQVSNWFKNRRQRDRTPH
>SR_DN126162
GEQKTHCFKERTRGLLREYYLTDPYPNPNKKKELAQLTGLTPTQVGNWFKNRRQRDRAAA
>SR_DN123214
SSPNSATAVKENTYPLRVWLEENRDNPYPTKSEKLMLSIVTKMTLTQVSTWFANARRRLKKER

>SR_DN126207
ALGRRKNATRETTAALKSWLQDHLKNPYPTKGEKVMLAIITKMSLTQVSTWFANARRRLKKEG
>SR_DN127996
MITRRKNATRETTSALKAWLSEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>SR_DN126498
SSMRRENASRETTAALKTWLFEHRKNPYPTKGEKIMLAIITKMSLTQISTWFANARRRLKKEN
>SR_DN129314
PQKKRGIFPKTATNIMRAWLFQHLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPM
>SR_DN127821
ARRKRRNFSKQASEILNEYFYSHLNNPYPSEEVKEELAKKCGLSVSQISNWFGNKRIRYKKNI
>SR_DN128964
KKTKRGVLPKHATQIMRQWLFQHIVHPYPTEDEKRQIAVQTNLTLLQVNNWFINARRRILQPM
>SR_DN142180
LKKQRGTLPKEAVDILRSWLELHRYHAYPTEQEKISLSEATNLTPLQVCNWFINARRRILPAL
>SR_DN106703
RSASFNDNALEVLKEWLEGHLDHPYIERSDILKIQTRTSLSRKQIMGWCTNIRRRRLTII
>SR_DN125608
TKKNRLVFTDLQRRTLFAIFRETKRPSKEMQITISRQLGLEQSTVSNFFMNARRRCSDRW
>SR_DN117560
SKKSRLVFTDIQRRTLRAIFQETQRPSKESQVQIGKQLGLEVSTVSNFFMNARRRCGDRW
>SR_DN127982.1
KRATRTRFTDHQIKSLQDFFDQNPYPRDDDLENLSKSLSLSPRVVVVWFQNQRQKARRSM
>SR_DN127982.4
QKQPRTKISEPQLEILKQNFDISNLPSEEKYIALSQQTGLPVKVIKHWFRNTLFKERQKDK
>SR_DN127982.3
KKRHRTHLSNVQVRTLKAMFLDFKTPSMGECEQIGHELGLSKRVVQVWFQNARSKEKKGK
>SR_DN127982.2
FRRMRTTISPDQLEILYTKYGENSNPSRTEMETIASSVGLSRRVVQVWYQNTRARQRRGQ
>SR_DN130571
ERKKRTVYTDKQLETLREYYVKQPNPNSDDIEKIANNIGLTKRNVQVWFQNKRARDKKNP
>SR_DN108290
RHRTRTKLTQSEVVLLKSIYENTPSPSKRQFEELANSTGLPTKVLKVWFQNTRRLRRNTC
>SR_DN124898
GICPRGARKPYPNAILERVFLTITQRPDEKHLKGLEKQLDCWNQYKIKQWFALRRAQLRPSK
>SR_DN113132
VVQMTENLTANHLRKAKLMFFYQRYPNSAILRQFFVDVKFNRSNNAQLIKWFSNFREYYYIQI
>SR_DN113631
SSRHRCLFNSHQINLLKSLFEVNAFPKKAELSILANHLRKDRESLRKWFDKRRLNFRRQG
>SR_DN132449
KRRKHILKFEQVSILENYFHIDSSWKMATVEKAAAELGLPVKKVYKWGYDRKHVSRNL
>HM_DN38072
KRRNRTTFTNLQLEEMEKIFQKTHYPDVYVREQLANRCELTEARVQVWFQNRRAKWRKRE
>HM_DN36971
QRRNRTTFTNYQLNTLECAFRKTHYPDVFTREELAKKVELTEARVQVWFQNRRAKWRKGE
>HM_DN27253
TRRHRTIFTEEQLEELEKTFSLTHYPDVQTREQLAAKADLKEERVEVWFKNRRAKWRKQK
>HM_DN10420
PKRHRTRFSPNQLNELERCFGKTHYPDIFMREEMALRIGLTESRVQVWFQNRRAKWKKRK
>HM_DN56217
PRRERTTFSKNQVLYLEELFTKTRYPDVFMRQEIADKLHLTESRVQVWFKNRRAKSRQQT
>HM_DN24659
QRRERTTFTRAQLEILESLFSKTRYPDIFMREEVAMKINLPESRVQVWFKNRRAKCRQQQ
>HM_DN15356
LQRNRTSFTQEQIDNLELEFDRTHYPDIFAREKLSEKIKLAEARIQVWFSNRRAKWRREE
>HM_DN37527
FQRNRTSFTQSQIDALEKEFERTHYPDVYARERLSESTSLAENKIQVWFSNRRAKWRREE

>HM_DN26989
QRRQRTHFTSQQLTELETLFSRNRYPDMATREDISAYILLPESKVRIWFKNRRAKWRKKE
>HM_DN29252
ERRQRTQFTSHQLRELETMFARNRYPDMASRERLSAWISLSESKVRIWFKNRRAKWRKKE
>HM_DN23062
RRRHRTTFNQDQLNELEAAFSKSHYPDIYVREELARITKLNEARIQVWFQNRRAKYRKHE
>HM_DN41482
KSRYRTTFTVFQLESLEKAFERAAYPDVFAREELARRINLSEARIQVWFQNRRAKWRKDK
>HM_DN20170
HRRNRTTFTTYQLHELERAFEKSHYPDVYSREELATRVQLPEVRVQVWFQNRRAKWRRQE
>HM_DN27245
CKRNRTQFSTFQIKRLEKLFEETHYPDSYMREELGKRLNLSESRIQVWFQNRRAKCRKYD
>HM_DN31526
RRRTRTNFTSLQLEELERAFAASHYPDVFMRESLGQRLSLLESRVQVWFQNRRAKWRKIE
>HM_DN7688
RRRHRTIFTSLQIDELEKAFSEAHYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKKE
>HM_DN80881
RRRHRTIFTNEQVDELEKIFQTAHYPDVNAREMLSLKTELPEDRIQVWFQNRRAKWRKKE
>HM_DN35675
KPRVRTFFTSDQLRNLEKAFSENAYPDVKMRAEICSRLKLRESQVLIWFQNKRARNKKLK
>HM_DN34694
GRKARTSITPFQQDQLEAAFRLNPFPLWSDKDYIASELDLDYSVVQVWFQNKRARSKRKL
>HM_DN9855
SKRMRTSFKNNQLVQMKQYFTMNHNPDAKDLKQLSHKTGLTKRVLQVWFQNSRAKYRRNM
>HM_DN36226
NKRVRTSFKNHQLRVLKQYFQLNHNPDAKELKQLSQKTGLSKRVLQVWFQNARAKYRRNV
>HM_DN27964
PTRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLTEMTALSPRVIRVWFQNKRCKDKKRS
>HM_DN14479
SGRVRTVLNEKQLSTLRAYYHANPRPDAPMREQLVEMTSLSHRVIRVWFQNKRCKDKKRQ
>HM_DN30383
RRGPRTTIKAKQLETLKTAFMNTPKPTRHIREQLAMETGLNMRVIQVWFQNRRSKERRMK
>HM_DN28959
NKRPRTTINNKQLDVLREAYSKSAKPARHIREDLSKKTDLDMRVVQVWFQNRRAKEKRLK
>HM_DN1193
TKRSRTTFSEEQLKVLQANFSLDSNPDGQDLERIAQMTGLSKRVIQVWFQNARSRQKKHL
>HM_DN41603
PCRKRVILTTQQRRAFKAAFDVSSKPCRKVREQLARDTGLSVRVVQVWFQNERAKVKKLA
>HM_DN8120
KARKRTVFTQQQLAVLQSRFAIKQILSRQEYANLADEIGVKIPIVRVWFQNYRQRSKTNS
>HM_DN20967
KRKKRTSIEVGIKGALENHFHLQPKPAAPEITGLADSLQLEKEVVRVWFCNRRQKEKRMN
>HM_DN43708
RRKKRASIETKSKLALEKSFSLNSKPTAEQINLLADQLIMERETVRVWFCNRRQKEKRVP
>HM_DN7261
KKRKRTSIAAPEKRSLEAYFAVQPRPSGEKIAAIAEKLDLKKNVVRVWFCNQRQKQKRLK
>HM_DN37913
SRKKRTQFSAETTIFLNDYFEKITHPRGFQINEIANTLQLDREVVRVWFCNKRQLMKSKD
>HM_DN37226
KKRDRFVWKDSCVSILEQFYQENSYPDDLKREEIAQACNNALELQGQFLNDEKLVTSVKVYNWFANRRKEIKRK
Q
>HM_DN66848
GEETVYCFKEKARRKLKESYKKNKYPTPEEKKSLAKASGLTLTQVSNWFKNRRQRDRAPT
>HM_DN32866
GEETSYCFKEKSRNVLRDYYLHNPYPSPREKRDLAEATDLTIIQVSNWFKNRRQRDRALE
>HM_DN28007

GEQKTHCFKERTRGLLREWYLQDPYPNPAKKRELAHATGLTPTQVGNWFKNRRQRDRAAA
>HM_DN21447
CNIRRKNTSRETTAALKSWLYEHRNNPYPTKGEKVMLAIITKMTLTQVSTWFANARRRLKKEN
>HM_DN24763
PLARRKNATRETTAALKTWLNDHLKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEG
>HM_DN36375
GIARRKNATRETTSALKAWLSEHKKNPYPTKGEKIMLAIITKMTLTQVSTWFANARRRLKKEN
>HM_DN37189
PQKKRGIFPKAATNIMRSWLFTNLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPM
>HM_DN34684
PQKKRGIFPKAATNIMRTWLFTNLTHPYPSEEQKKQLAQDTGLTILQVNNWFINARRRIVQPM
>HM_DN21081
ARRKRRNFSKQASEILNEYFYSHLCNPYPSEEVKEDLAKKCCLSVSQISNWFGNKRIRYKKNI
>HM_DN37578
KKTKRGVLPKQATQIMRSWLFQHIVHPYPTEDEKRQLAQQTSLTLLQVNNWFINARRRILQPM
>HM_DN12757
MKKKRGNLPKESVEVMKKWLNEHRYNAYPNETEKNQLAEKANLTVLQVCNWFINARRRILPEM
>HM_DN40002
PKKIRLVFTDIQRRTLYAIFRETKRPSKEMQQTIGQQLGLDPTTVSNFFMNARRRCMGDR
>HM_DN33495.2
QKRPRTRISDEQLKILKDHFDLSNSPTDSEVDQIAIKTGLQGKVIKHWFRNTLFKERQKN
>HM_DN33495.4
RRASRTRFSDFQLKSLQEFFERNAYPRDEDLEKLSKSLNLSPRVVVVWFQNQRQKARRNI
>HM_DN33495.1
RRRYRTHLSVTQIKIFKVIFADYKTPTMSDCEVLGSQLDLPKRVVQVWFQNARSKEKKAR
>HM_DN33495.3
DKRIRTTISPAQLEVLYSSYQEDSNPSRQMLEEICAKVGLKRRVVQVWFQNTRARQRKGA
>HM_DN42050
RHRVRTRLTAQDLLYLHNMYRRTANPSRRLLAEISNHIGLPLRVIQVWFQNTRRLMKNTC
>HM_DN34160
QRKKRTVFTDFQQKMLRELYKINSSPTSEEISKLADNIGLSKRSIQVWFQNKRARDKKSR
>HM_DN30308
VLPSTQQPTPQPNAILDKVYTSIARDPDAVRLAGLAKQLDWSIRQVQVWFIARRALDRPSH
>Hm_DN24356
1HSLAAETLTANHLKKAKLIFFYQRYPNPSTLRMYFPDVRFNRQNNGQLIKWFSNFREYFYNQI