

Objective Sequence-Based Subfamily Classification of Mouse

Homeodomains Reflects their *in vitro*

DNA Binding Preferences

Miguel A. Santos¹, Andrei L. Turinsky¹, Serene Ong¹, Jennifer Tsai^{1,3}, Mike F. Berger^{4,5},
Gwenael Badis⁸, Shaheynoor Talukder⁸, Andrew R. Gehrke⁴, Martha L. Bulyk^{4,6,9},
Timothy R. Hughes^{7,8} and Shoshana J. Wodak^{1,2,3#}

¹ Molecular Structure and Function Program, Hospital for Sick Children, 555
University Avenue, Toronto ON, Canada M5G 1X8

² Department of Biochemistry University of Toronto, 1 Kings College Circle, Toronto ON,
Canada M5S 1A8

³ Department of Molecular Genetics, University of Toronto, 1 Kings College Circle,
Toronto ON, Canada M5S 1A8

⁴ Division of Genetics, Department of Medicine, and ⁵ Department of Pathology, Brigham
and Women's Hospital and Harvard Medical School, Boston, MA 02115.

⁶ Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA
02138.

⁷ Banting and Best Department of Medical Research, ⁸ Department of Medical Genetics
and Microbiology, University of Toronto, Toronto, ON M4T 2J4

⁹ Harvard/MIT Division of Health Sciences and Technology (HST), Harvard Medical
School, Boston, MA 02115.

Supplementary Figure S1

Figure S1: Multiple sequence alignment of the 381 homeodomains from different organisms added to mH178 to build the H559 dataset.

	10	20	30	40	50	60	70	80	90
CCE07325_1::Cyprinus	FTKEQIRELESEFAHH	NYL	TRLR	RYEIAVN	LDLTERQVK	VWFQNR	RMKWK	KR	
SPE28572_1::Strongylocentrotus	FTKEQIRELENEFNHH	NYL	TRLR	RYEIAVT	LNLTERQVK	VWFQNR	RMKWK	KR	
ACF06894_1::Allium	LTAEQVNMLEKSFEESE	NKL	EPER	KGELAKK	LGLQPRQVA	VWFQNR	RRARWK		
AMF13795_1::Ambystoma	YSRYQTLLELEKEFLFN	PYL	TRKR	RIEVSHA	LGLTERQVK	IWFQNR	RMKWK	K	
AMF15085_1::Ambystoma	YTGRQLAELEKEFRTN	RYI	TSGR	KEELVAS	LGLTNRQVK	IWFQNR	RAKEKR		
DIP01793_1::Brugia_malayi	YTKYQTLLELEKEFLYN	TYV	SKOK	RYELAKN	LYLSERQVK	IWFQNR	RMKDKK		
SPE11478_1::Strongylocentrotus	FSNDQTMLEKKEFENQ	KYL	SPPE	RKKLAKV	LQLSERQVK	TWFQNR	RAKWR		
GHP09300_1 Vitis_vinifera	LTSEQVYLLKESFEAE	NKL	EPER	KSOLAKK	LGLQPRQVA	VWFQNR	RRARWK		
LUP01625_1::Lotus_japonicus	LTAEQVNLLEKSFEEAE	NKL	EPER	KTELAKK	LGLQPRQVA	VWFQNR	RRARCK		
RPP01355_1::Medicago_truncatula	LSVEQVHFLEKSFEE	NKL	EPER	KTRLAKE	LGLQPRQVA	IWFQNR	RRARWK		
SSD00939_1::Saccharum	LTPEQVHLLERSFEE	NKL	EPER	KTELARK	LGLQPRQVA	VWFQNR	RRARWK		
BMD56378_1::Bombyx	FTSEQLLELEREFHAK	KYL	SLTE	RSQIAAA	LKLSEQVK	IWFQNR	RAKWR		
BMD57173_1::Bombyx	FTDHQLQTLKESFERQ	KYL	SVQD	RMELAAK	LGLTDIQVK	TWYQNR	RTKWK		
CCE05499_1::Cyprinus	YTKHRTLLELEKEFLFN	MYL	TPER	RLEIXKS	INLTDQVK	IWFQNR	RMKLLK		
CSH13076_1::Ciona	FTHEQVRQLELDFSEN	HYL	TRLR	RYELSLK	LSLTERQIK	VWFQNR	RMKLLK		
HRP00014_1::Halocynthia_oretzi	FTPEQLELEREFLLKQ	QYM	VGTE	RFXLAKE	LNLGEAQVK	VWFQNR	RRIKWR		
SSF11272_1::Salmo	YSRAQLLELEKEFLFN	KYI	SRPR	RYELATT	LKLTERHIK	IWFQNR	RMKWK		
SSF44037_1::Salmo	YSRAQLLELEKEFLFN	KYI	SRPR	RYELATT	LNLTERHIK	IWFQNR	RMKWK		
AOP00376_1::Populus_trichocarpa	LSGEQVRLEKKNFEVE	NKL	EPER	KARLAQE	LGLQPRQVA	VWFQNR	RRARWK		
GRD05488_1::Gossypium	LTVDQVQFLEKSFEEV	NKL	EPER	KTQLAKE	LGLQPRQVA	IWFQNR	RRARWK		
HAD32146_1::Helianthus	LSTEQVRMLERSFEE	NKL	EPER	KTELAKK	LGLQPRQVA	VWFQNR	RRARW		
JRP01502_1 Arabidopsis_thaliana	LTTEQVHLLKESFEAE	NKL	EPDR	KTVLAKK	LGLQPRQVA	VWFQNR	RRARWK		
PAD00461_1::Prunus	LSVEQVKALEKKNFEVE	NKL	EPER	KVQLAQE	LGLQPRQVA	VWFQNR	RRARWK		
PPG03907_1::Physcomitrella	LSLEQVRSLEARNFEVE	NKL	EPER	KMQLAKE	LGLQPRQVA	VWFQNR	RRARWK		
AMF15721_1::Ambystoma	YTRQOVLELEKEFHYN	RYL	TRRR	RIEIAHS	LCLSERQIK	IWFQNR	RMKWK		
CSH18746_1::Ciona	YTKFQLAELEREFETAN	EFI	SREM	REEIATR	VGLNDRQVK	IWFQNR	RMKCKR		
IPD19513_1::Ictalurus	YTRQOVLELEKEFHFN	RYL	TRRR	RIEIAHT	LCLSERQIK	IWFQNR	RMKWK		
SPE27756_1::Strongylocentrotus	YSKLQIYELEKEFETT	MYL	TRDR	RSKLSQA	LDLTERQVK	IWFQNR	RMKMK		
SRP01584_1::Oryzias_latipes	FTTHQLTELEKEYYTS	KYL	DRSR	RREIAKQ	LALNETQVK	IWFQNR	RMKEK		
1b72a::HUMAN	LRTNFTTRQLTELEKEFHFN	KYL	SRAR	RVEIAAT	LELNETQVK	IWFQNR	RMKQK		
ACF04417_1::Allium	LTIEQVRTLEKSFEEVG	NKL	EPER	KMQLARA	LGLQPRQVA	IWFQNR	RRARWK		
ACF08359_1::Allium	LSVTQVQFLEKSFEEV	NKL	EPER	KVQLAKE	LGLQPRQVA	IWFQNR	RRARWK		
BNP01381_1 Arabidopsis_thaliana	LTSQVHLLKESFETE	NKL	EPER	KTQLANK	LGLQPRQVA	VWFQNR	RRARWK		
BVP01068_1::Ricinus_communis	LSVEQVQFLEKSFEEV	NKL	EPER	KVQLAKD	LGLQPRQVA	IWFQNR	RRARWK		
CSE02860_1::Citrus	LTVDQVQFLEKSFEEV	NKL	EPER	KIQLAKD	LGLQPRQVA	IWFQNR	RRARWK		
CSE16705_1::Citrus	LTSDQLESLEERSFQEE	IKL	DPDR	KMKLARE	LGLQPRQIA	VWFQNR	RRARWK		
GRD03861_1::Gossypium	LSVDQVKALEKKNFEVE	NKL	EPER	KVQLAQE	LGLQPRQVA	VWFQNR	RRARWK		
GRD06591_1::Gossypium	LTVDQIQFLEKSFEEVD	NKL	EPER	KIQLAKD	LGLQPRQVA	IWFQNR	RRARWK		
JRP02068_1::Picea_sitchensis	LTADQVQFLEKSFEEVD	NKL	EPER	KSQLAKD	LGLQPRQVA	IWFQNR	RRARWK		
MSE04106_1::Medicago	LSXDQVQFLEKSFEEED	SKL	EPER	KTKLAKD	LGLQPRQVA	IWFQNR	RRARWK		
PED06136_1::Populus	LTVDQVQFLEKSFEELE	NKL	EPER	KIQLAKD	LGLQPRQVA	IWFQNR	RRARWK		
PTE03183_1::Poncirus	LSVDQVKALEKKNFEVD	NKL	EPER	KVQLAQE	LGLQPRQVA	VWFQNR	RRARWK		
SMP00439_1 Vitis_vinifera	LTPEQVHMLEKSLEAE	NKL	EPER	KTQLAKK	LNLQPRQVA	VWFQNR	RRARWK		

O42365::Danio_rerio	RRLR	TAYTNTQLLELEKEEFHFN	KYL	CRPR	RVEIAAL	LDLTERQVK	VWFQNR	RRMKHKR
P09022::MOUSE	RTNF	TTKQLLELEKEEFHFN	KYL	TRAR	RVEIAAS	LQLNETQVK	IWFQNR	RMKQKK
AMF10171_1::Ambystoma		YTDHQRLELEKEEFHYS	RYI	TIRR	KAELAMA	LGLSERQVK	IWFQNR	RAKERK
BFP09676_1::Branchiostoma_florida		YSDHQRLELEKEEFYSN	KYI	TIKR	KVQLANE	LGLSERQVK	IWFQNR	RAKQRK
BMD62527_1::Bombyx		FTGTQLLELEREFMSN	MYL	SRLR	RIEIASR	LKLSEKQVK	IWFQNR	KVKLXK
CVP02277_1::Saccoglossus_kowalevskii		FSDQQLNGLEKRFEAQ	RYL	STPE	RVELANQ	LSLSETQVK	TWFQNR	RMKHKK
GAP01770_1::Ornithorhynchus_anatinus		FTNHQIYELEKRFYLYQ	KYL	SPAD	RDQIAQQ	LGLTNAQVI	TWFQNR	RAKLKR
HCG02656_1::Homalodisca		YTDHQRLELEKEEFHYS	RYI	TIRR	KAELATN	LGLSERQVK	IWFQNR	RAKERK
IPD17949_1::Ictalurus		FTHLQVLELEKKSFRQ	RYL	SAPF	RAHLASA	LRLTETQVK	IWFQNR	RYKTKR
SPE19093_1::Strongylocentrotus		FSGRQIFLELEKQFEVK	KYL	SASE	RAELASL	LNVTDQVK	IWFQNR	RTKWKK
lpufA::MOUSE	RKKRC	PTYKHQTLLEKEEFLFN	MYL	TRDR	RYEVARL	LNLTERQVK	IWFQNR	RMKMKK
ACF08945_1::Allium		LSITQVQFLEKSFEVE	NKL	EPER	KVQLAKE	IGLQPRQVA	IWFQNR	RARWK
BRE05356_1::Brassicarapa		LSVVQVKALEKNFEID	NKL	XPER	KVKLAQE	LGLQPRQVA	IWFQNR	RRARWK
BVP03504_1 Vitis_vinifera		LSVDQVKALERNFEVE	NKL	EPER	KVKLAQE	LGLQPRQVA	VWFQNR	RRARWK
GAD16259_1::Gossypium		LTQEQVDLLELNFGNE	HKL	ESER	KDRLEASE	LGLDPRQVP	VWFQNR	RRARWK
GGP01641_1::Glycine_max		LTPEQVRSLETSEFESE	NKL	EPDR	KLKLAQQ	LGLQPRQVA	VWFQNR	RRARWK
GRD00608_1::Gossypium		LSVDQVKALEKNFEVE	NKL	EPDR	KSKLAQE	LGLQPRQVA	VWFQNR	RRARWK
GRD07392_1::Gossypium		LTVDQVQFLEKSFEAE	NKL	EPDR	KVQLAKD	LGLQSRQVA	IWFQNR	RRARWK
JRP00287_1::Vitis_vinifera		LSVDQVKALEKNFEVE	NKL	EPER	KVKLAQE	LGLQPRQVA	VWFQNR	RIARWK
LUP01153_1 Ricinus_communis		FTDEQIKSLESIFESE	TRL	EPK	KVQLAKE	LGLQPRQVA	IWFQNR	KRARWK
PPE01139_1::Prunus		LTVDQVQFLEKSFDFME	NKL	EPER	KILLAKD	LGLQPRQVA	IWFQNR	RRARWK
PSE12995_1::Picea		LSLQVRSLEKTFEVE	NKL	EPER	KLQLAQE	LGLQPRQIA	VWFQNR	RRARWK
RPP00443_1::Phaseolus_vulgaris		LSVDQVKALEKNFEVE	NKL	EPDR	KVKLAQE	LGLQPRQVA	VWFQNR	RRARWK
SMP04043_1::Catharanthus_roseus		LTAEQVQFLEKSFAVE	NKL	EPDR	KNELAKK	LGLQPRQVA	IWFQNR	RRARSK
TVD00827_1::Triphysaria		LNNEQVRTLEKNFELG	NKL	EPER	KIELARA	LGLQPRQIA	IWFQNR	RRARWK
1b8iA::Drosophila_melanogaster	RQTY	TRYQTLELEKEFHTN	HYL	TRRR	RIEMAHA	LSLTERQIK	IWFQNR	RMKCLKK
BME03454_1::Boophilus		FTTQQLLALERKFRVK	QYL	SIAE	RAEFSSS	LNLTERQVK	IWFQNR	RAKEKR
MCP00368_1::Priapulus_caudatus		YGKGQTDLETEYCTS	PYV	TKQR	RYELSRK	LGLTERQVK	IWFQNR	RMKTKK
SPE27538_1::Strongylocentrotus		FSRSQVQLESTFEVK	RYL	SSSE	RAGLAAN	LHLTERQVK	IWFQNR	RNKWKR
SPE27764_1::Strongylocentrotus		FTREQIGLEKEFARE	NYV	SRPK	RCELATA	LNLPTTIK	VWFQNR	RMKDKR
SSF01710_1::Salmo		FTELQLMGLEKRFKQ	KYL	STPD	RIDLAEK	LDSLQIVK	TWYQNR	RMKWKK
SSF11513_1::Salmo		YTDHQRLELEKEEFHYS	RYI	TIRR	KAELATS	LGLSERQVK	IWFQNR	RAKERK
TCF01014_1::Tribolium		YTSACLVELEREFHHG	KYL	SRPR	RIQIAEN	LNLSEKQIK	IWFQNR	RMKHKK
1hom::Drosophila_melanogaster	KRGR	QTYTRYQTLELEKEEFHFN	RYL	TRRR	RIEIAHA	LCLTERQIK	IWFQNR	RMKWKK
O43365::HUMAN	KRAR	TAYTSAQLVELEKEEFHFN	RYL	CRPR	RVEMANL	LNLTERQIK	IWFQNR	RMKYKK
P04476::XENLA	KRGR	QTYTRYQTLELEKEEFHFN	RYL	TRRR	RIEIAHT	LCLTERQIK	IWFQNR	RMKWKK
P06798::MOUSE	KRSR	TAYTRQQVLELEKEEFHFN	RYL	TRRR	RIEIAHT	LCLSERQVK	IWFQNR	RMKWKK
P09023::MOUSE	RRGR	QTYTRYQTLELEKEEFHYN	RYL	TRRR	RIEIAHA	LCLTERQIK	IWFQNR	RMKWKK
P09024::MOUSE	KRGR	QTYTRYQTLELEKEEFHYN	RYL	TRRR	RIEIAHT	LCLTERQIK	IWFQNR	RMKWKK
P14150::SCHAM	KRPRT	AFSGQLARLKHEFTEN	RYL	TERR	RQELARE	LGLNEAQIK	IWFQNR	KRAKIKK
Q05640::Artemia_franciscana	KRPRT	AFTAELSLRLKHEFNEN	RYL	TERR	RQDLARE	LGLHENQIK	IWFQNR	RAKLKK
AAF03294_1::Acorus		LSVEQVRILEKNFELG	NKL	EPER	KMQLARA	LGLQPRQIA	IWFQNR	RRARWK
BNP04867_1 Arabidopsis_thaliana		LSDEQVRMLEMSFGLG	HKL	ESER	KNRLASE	LGLDSRQVA	VWFQNR	RRARWK
CND18854_1::Filobasidiellaneoformans		TTPEQLKLVLEFWYDIN	PKP	DNQL	REQLAAQ	LGMTKRNQV	VWFQNR	RAKMK
CRP01704_1::Ceratopteris_richardii		LTAEQVNFLETSEFSMD	LKL	EPER	KAHLAKQ	LGIQPRQVA	IWFQNR	RRARWK
CRP03497_1::Ceratopteris_richardii		LTAEQVNFLEMSFNID	LKL	EPER	KALLAKK	LGIQPRQVA	IWFQNR	RRARWK

CSE16783_1::Citrus	LTAEQAELEEMHFGNE	HKL	ESDR	KDKLAAE	LGLDPRQVA	VWFQNRARDK
GHP07947_1::Ricinus_communis	FSDEQIKSLELMFESE	TRL	EPGK	KLEVAKE	LGLHPRQVA	IWFQNKRARWK
GRD02024_1::Gossypium	LTATQVEFLERSFEVE	NKL	ESDR	KLRLAKE	LGLQPRQVA	IWFQNRARSK
GRD28327_1::Gossypium	FSDEQIRLLESIFESE	TKL	EPRK	KMQLARE	LGLQPRQVA	IWFQNRARWK
HAD01492_1::Helianthus	LTTDQVQFLEKSF DEN	NKL	EPER	KVHLAKE	LNLQPRQVA	IWFQNR
PPE01722_1::Prunus	FSDEQIRLSLESLF ESE	SRL	EPRK	KMQLAKE	LGLQPRQVA	IWFQNKRARWK
PPG05037_1::Physcomitrella	LSKEQSALLEESFKEH	STL	NPQK	KNALAKQ	LGLRPRQVE	VWFQNRARTK
PSE02887_1::Picea	LSKEQSALLEESFKEN	SSL	NPQK	KQALAKR	LNLRPRQVE	VWFQNRARTK
PSE04668_1::Picea	LSKEQSALLEESFLEH	STL	NPQK	KNALAKE	LNLQPRQVE	VWFQNRARTK
ZEP07385_1 Arabidopsis_thaliana	LTKEQSALFLEDSFKEH	STL	NPQK	KQALAKQ	LNLRPRQVE	VWFQNRARTK
3hddA::Drosophila_melanogaster	RTAFSSQLARLKR EFNEN	RYL	TERR	RQQLSSE	LGLNEAQIK	IWFQNKRAKIKK
AME09631_1::Apis	FNSWQLEELERAF LSS	HYP	DVFM	REALAVR	LELKE SRVA	VWFQNRRAKWRK
AMF10710_1::Ambystoma	YTDHQRLLELEKEFHYN	RYI	TITR	KAQLAAN	LRLSERQIK	IWFQNRRAKERK
ASP02539_1 Brugia_malayi	FTTFQLHELEQAF EKC	HYP	DVYA	RELLAQK	VKLPEVRVQ	VWFQNRRAKWRR
ASP08402_1::Caenorhabditis_briggsae_AF16	YTDYQRLLELEKEFRIT	QFI	NSER	KSQLSSE	LQLTERQIK	IWFQNRRAKDRR
ASP20371_1::Caenorhabditis_elegans	FTPTQADTLEKEYLTD	QYM	PRTR	RILIAES	LGLSEGQVK	TWFQNRRAKEKR
AYP03228_1::Brugia_malayi	FSAYQLDELEKVFART	HYP	DVFT	REELAQR	VTLTEARVQ	VWFQNRRAKFRK
BFP10457_1::Branchiostoma_floridae	YSQWQLELEKAFETT	QYP	DIFM	REALALR	LDLIEARVQ	VWFQNRRAKLRR
CSG02378_1::Callinectes	FSQAQVYELERRFKQQ	RYL	SAP E	REHLAGL	LKLTSTQVK	IWFQNRRYKCKR
ECD01305_1::Equus	FSHTQVXELERKFXRQ	KYL	SAP E	RAHLAXN	LKLT EXQVK	IWFQNRRYKTKR
IFP07589_1::Danio_erio	FSKAQTYELERRFRQQ	RYL	SAQE	REQLAHL	LRLTPTQVK	IWFQNHRYKMKR
IPD18117_1::Ictalurus	FSQAQVLELERRFKQQ	RYL	SAP E	REHLANT	LKLTSTQVK	IWFQNRRYKCKR
P31535::MYXGL	ADQLARLRAEFQAN	RYL	TEER	RQNLARE	LSLNEAQIK	IWFQNKRAKIKK
SPE27868_1::Strongylocentrotus	FTTFQLHQLERAFDMT	QYP	DVFM	REELALR	LDLSESRVQ	VWFQNRRAKWRK
SSF27381_1::Salmo	FPPSQVELEKVFLET	HYP	DVHI	RDKLASR	LQLTEGRVQ	IWFQNRRAKWRK
SSF54974_1::Salmo	FSKRQIFQLESTFDMK	RYL	SSAE	RACCLASS	LQLTETQVK	IWFQNRNKLKR
SSP03100_1::Brugia_malayi	FSRHQVSLQEMTFDMK	RYL	SSQE	RAHLASN	LQLTETQVK	IWFQNRNKKWR
TCP00488_1::Caenorhabditis_elegans	FTPAQADTLEKEYLTD	QYM	PRTR	RILIAES	LGLNEGQVK	TWFQNRRAKEKR
WBP00763_1 Brugia_malayi	FTQSQVNELEERFKLQ	RYV	NAAE	RERLAVT	LGLTSTQVK	IWFQNRRYKCKR
O09113::MOUSE	KRHRTRFTPAQLNELERSFAKT	HYP	DIFM	REELALR	IGLTESRVQ	VWFQNRRAKWKK
O13074::FUGRU	KRSRTAYTRQQVLELEKEFHYN	RYL	TRRR	RVEIAHT	LCLSERQIK	IWFQNRMMKWKK
O42115::Danio_erio	RRYRTTFTSYQLEELERAFQKT	HYP	DVFT	REELAMR	LDLTEARVQ	VWFQNRRAKWRK
O42368::Danio_erio	KRARTAYTSAQLVLELEKEHFHN	RYL	CRPR	RVEMANL	LNLSESRQIK	IWFQNRMMKYKK
O57374::Danio_erio	KRSRTAYTRQQVLELEKEHFHN	RYL	TRRR	RIESAHT	LSLSESRQIK	IWFQNRMMKWKK
P02832::XENLA	RRGRQIYSRYQTLELEKEHFHN	RYL	TRRR	RIEIANA	LCLTERQIK	IWFQNRMMKWKK
P07548::Drosophila_melanogaster	KRQRTAYTRHQIILELEKEFHYN	RYL	TRRR	RIEIAHT	LVLSESRQIK	IWFQNRMMKWKK
P09013::Danio_erio	KRARTAYTRYQTLELEKEHFHN	RYL	TRRR	RIEIAHA	LCLSERQIK	IWFQNRMMKWKK
P09016::HUMAN	KRSRTAYTRQQVLELEKEHFHN	RYL	TRRR	RIEIAHT	LCLSERQIK	IWFQNRMMKWKK
P09017::HUMAN	KRSRAAYTRQQVLELEKEFHYN	RYL	TRRR	RIEIAHS	LCLSERQIK	IWFQNRMMKWKK
P09019::XENLA	KRARTAYTRYQTLELEKEHFHN	RYL	TRRR	RIEIAHT	LCLSERQIK	IWFQNRMMKWKK
P09020::XENLA	KRSRTSYTRYQTLELEKEHFHN	RYL	TRRR	RIEIANN	LCLNERQIK	IWFQNRMMKWKK
P23397::HELTR	KRPRTAFTGDLARLKR EFS EN	KYL	TEQR	RTCLAKE	LNLNESQIK	IWFQNKRAKMKK
P31533::Danio_erio	KRPRTAFTAELQRLKNEFQNN	RYL	TEQR	RQALAQE	LGLNESQIK	IWFQNKRAKIKK
P34326::Caenorhabditis_elegans	KRPRTAFTGDLRLKTEFRES	RYL	TEKR	RQELAHE	LGLNESQIK	IWFQNKRAKLLK
P52730::XENLA	KRPRTAFTAELQRLKA E FQTN	RYL	TEQR	RQSLAQE	LGLNESQIK	IWFQNKRAKIKK
AOP03183_1 Vitis_vinifera	LSKEQSRLL EESFRQN	HTL	NPTQ	KEALASR	LRLKPRQVE	VWFQNRARTK

BNP04177_1|Arabidopsis_thaliana -----FSDEQIKSLEMMFSE-----TKL-----EPRK-----KVQLARE-----LGLQPRQVA-IWFQNKRARWK-----
 BNP18445_1|Arabidopsis_thaliana -----LTSGQLASLERSFQED-----IKL-----DSDR-----KLKLSRE-----LGLQPRQIA-VWFQNRRARWK-----
 CJP00479_1::Picea_sitchensis -----LSKEQSALLEESFREH-----STL-----NPKQ-----KNALAKQ-----LNLRRPQVE-VWFQNRARTK-----
 CSE03377_1::Citrus -----FSDEQIRSLLELMFENE-----TRL-----EPRK-----KLQLAKE-----LGWQPRQVA-IWFQNKRARWK-----
 CSE04725_1::Citrus -----LSKDQSAILEESFKEH-----NTL-----NPKQ-----KLALAKQ-----LGLRPRQVE-VWFQNRARTK-----
 CSE14614_1::Citrus -----LSKEQSAFLEESFKEH-----NTL-----NPKQ-----KLALAKQ-----LNLRRPQVE-VWFQNRARTK-----
 GAD15382_1::Gossypium -----LSKDQSAILEENFKEH-----NTL-----NPKQ-----KLALAKQ-----LGLRPRQVE-VWFQNRARTK-----
 GRD08825_1::Gossypium -----LSKDQSAILEESFKEH-----NTL-----NPKQ-----KMALAKQ-----LGLRPRQVE-VWFQNRARTK-----
 GRD15275_1::Gossypium -----FSDEQIKSLELMFSE-----TRL-----EPRK-----KLQVAKE-----LGLQPRQVA-IWFQNKRARWK-----
 GRD29631_1::Gossypium -----LSKDQSAILEECFKEH-----NTL-----NPKQ-----KLALAKQ-----LGLRPRQVE-VWFQNRARTK-----
 PED03264_1::Populus -----FSDEQIKSLETMFSE-----TRL-----EPRK-----KMQLARE-----LGLQPRQVA-IWFQNKRARWK-----
 PED07486_1::Populus -----LSKDQSAFLEESFKEH-----NTL-----TPKQ-----KLALAKE-----LNLRRPQVE-VWFQNRARTK-----
 RPP00177_1|Glycine_max -----LSKDQSIILEESFKEH-----NTL-----NPKQ-----KSALAKQ-----LGLRARQVE-VWFQNRARTK-----
 SSF35481_1::Salmo -----FTDHQLAQLERSFERQ-----KYL-----SVQD-----RME LAAS-----LNLTDTVQVK-TWYQNR-----
 1jggA|Drosophila_melanogaster -----RYRTAFTRDQLGRLEKEFYKE-----NYV-----SRPR-----RCELAAQ-----LNLPESTIK-VWFQNRMKDKR-----
 CCE05724_1::Cyprinus -----FTTSQLLVLERKFLQK-----QYL-----SIAE-----RAEFNS-----LNLTTETQVK-IWFQNRTRAKAKR-----
 P31537::XENLA -----AEQLQRLKAQEFQAN-----RYI-----TEQR-----RQSLAQE-----LSLNEQIK-IWFQNKRAKIKK-----
 PBP05845_1::Anopheles_gambiae_str._PEST -----FSKSQTFELERRFKQA-----RYL-----SAPE-----REHLASM-----INLTPQVK-IWFQNHRYKTKR-----
 SSF22556_1::Salmo -----FSRVQICELEKRFHRQ-----KYL-----ASAE-----RATLAKS-----LKMTDAQVK-TWFQNRRTKWR-----
 O15266::HUMAN -----RRSRTNFTLEQLNELERLFD-----HYP-----DAFM-----REELSQR-----LGLSEARVQ-VWFQNRRAKCRK-----
 O35137::MOUSE -----RRNRFTFTSYQLEELKVFQKT-----HYP-----DVYA-----REQLAMR-----TDLTEARVQ-VWFQNRRAKWRK-----
 P09076::Apis -----KRPRTAFAEQQLARLKREFAEN-----RYL-----TERR-----RQQLSRD-----LGLTEAQIK-IWFQNKRAKIKK-----
 P09145::DROVI -----KRPRTAFSSEQLARLKREFNEN-----RYL-----TERR-----RQQLSSE-----LGLNEAQIK-IWFQNKRAKIKK-----
 P19622::HUMAN -----KRPRTAFTAQQLQRLKAQFQTN-----RYL-----TEQR-----RQSLAQE-----LSLNEQIK-IWFQNKRAKIKK-----
 P27609::BOMMO -----KRPRTAFSGAQLARLKHEFAEN-----RYL-----TERR-----RQSLAAE-----LGLAEAQIK-IWFQNKRAKIKK-----
 P27610::BOMMO -----KRPRTAFSGPQLARLKHEFAEN-----RYL-----TERR-----RQSLAAE-----LGLAEAQIK-IWFQNKRAKIKK-----
 AAF02115_1::Acorus -----LSKDQSAVLEESFKEH-----STL-----NPKQ-----KLALAKQ-----LNLRRPQVE-VWFQNRARTK-----
 ACF09585_1::Allium -----LSKDQSRLLLEEFKEH-----STL-----NPKQ-----KQALAKH-----LNLQPRQVE-VWFQNRARTK-----
 ASP07310_1::Caenorhabditis_elegans -----YTRNQVLELEKEFHFN-----KYL-----TRKR-----RIEIAHS-----LMLTEHQVK-IWFQ-----
 GRD07374_1::Gossypium -----LTKDQSALLEESFKQH-----STL-----NPKQ-----KQALAKQ-----LNLRRPQVE-VWFQNRARTK-----
 GRD21285_1::Gossypium -----LSKEQSLVLEETFKEH-----STL-----NPKQ-----KLALAMQ-----LNLRRPQVE-VWFQNRARTK-----
 HAD09917_1::Helianthus -----LSKEQSAYLEETFKEH-----NTL-----NPKQ-----KLALAEQ-----LHLRPRQVE-VWFQNRARTK-----
 JRP01754_1|Populus_trichocarpa -----LTKEQSALLEESFKQH-----STL-----NPKQ-----KQALARQ-----LNLRRPQVE-VWFQNRARTK-----
 MCP03763_1::Nasonia_vitripennis -----FSDQQLQGLEQRFNGQ-----KYL-----STPE-----RISLAES-----LHLSSETQVK-TWFQNRMK-----
 PGD09699_1::Picea -----LSKDQSSLLEESFREH-----SAL-----SPKH-----KSALACK-----LNLQPRQVE-VWFQNRARTK-----
 PPE02653_1::Prunus -----LSKEQSATLEDSFREH-----TTL-----NPKQ-----KQDLARK-----LNLRRPQVE-VWFQNRARTK-----
 RHD00015_1::Rosa -----LSKDQSAILEESFKDH-----NTL-----NPKQ-----KLALAKQ-----LGLRPRQVE-VWFQNRARTK-----
 lig7A::MOUSE -----RKPRTPFTTAQLLALERKFRQK-----QYL-----SIAE-----RAEFSS-----LSLTTETQVK-IWFQNRRAKAKR-----
 O02491::Anopheles_gambiae_str._PEST -----KRPRTAFSNAQLQRLKNEFNEN-----RYL-----TEKR-----RQTLSE-----LGLNEAQIK-IWFQNKRAKIKK-----
 O35602::MOUSE -----RRNRFTFTTYQLHELERAFEKS-----HYP-----DVYS-----REELAGK-----VNLPEVRVQ-VWFQNRRAKWR-----
 O35690::MOUSE -----RRIRTTFTSAQLKELERVFAET-----HYP-----DIYT-----REELALK-----IDLTEARVQ-VWFQNRRAKFRK-----
 O42250::Danio_erio -----RRHRVTFVTSQLEELKAFNEA-----HYP-----DVYA-----REMLAMK-----TELPEDRIQ-VWFQNRRAKWRK-----
 O42358::Danio_erio -----RRNRFTFTTYQLHELERAFEKS-----HYP-----DVYS-----REELALK-----VNLPEVRVQ-VWFQNRRAKWR-----
 O42567::XENLA -----RRNRFTFTTYQLHELERAFEKS-----HYP-----DVYS-----REELAMK-----VNLPEVRVQ-VWFQNRRAKWR-----
 O70137::MOUSE -----RRNRFTFTTYQLHELERAFEKS-----HYP-----DVYA-----REQLALR-----TDLTEARVQ-VWFQNRRAKWRK-----
 P05527::Drosophila_melanogaster -----KRPRTAFSGTQLARLKHEFNEN-----RYL-----TEKR-----RQQLSGE-----LGLNEAQIK-IWFQNKRAKIKK-----

P09075::Apis -----KRPRTAFSGEQ LARLKREFAE N-----RYL-----TERR-----RQQLSRD-----LGLNEA QIK-IWFQNKRAKIKK-----
 P31538::XENLA -----KRPRTAFTA EQ LRLKAEFQAN-----RYI-----TEQR-----RQTLAQE-----LSLNESQIK-IWFQNKRAKIKK-----
 P52729::XENLA -----KRPRTAFTADQLRLKAEFQTN-----RYL-----TEQR-----RQSLAQE-----LSLNESQIK-IWFQNKRAKIKK-----
 Q04896::Danio_rerio -----KRPRTAFTA EQ LRLKAEFQTS-----RYI-----TEQR-----RQALARE-----LGLNESQIK-IWFQNKRAKIKK-----
 Q05916::CHICK -----KRPRTAFTA EQ LRLKAEFQAN-----RYI-----TEQR-----RQSLAQE-----LSLNESRVK-IWFQNKRAKIKK-----
 AMF10979_1::Ambystoma -----FSKAQTLQLERRFRQQ-----RYL-----SAPE-----RDHLAHL-----LHLTPTQVK-IWFQNHRYKMKR-----
 AYP02578_1::Caenorhabditis_elegans -----FTQEQLAELDSAFQKS-----HYP-----DIYV-----REELARI-----TKLNEARIQ-VWFQNRRAKHRK-----
 CSE04691_1::Citrus -----FSDEQIRLLESIFES E-----TKL-----EPRK-----KMQVATE-----LGLQPRQVA-IWFQNKRARWK-----
 MIP05372_1::Convolutriloba_longifissura -----FSQQQVCELEKMFQRK-----KYL-----NAPE-----RESLAQA-----IGLKPTQVK-IWFQNHRYKCKR-----
 PPP02701_1::Pristionchus_pacificus -----YSDIYIRLELEKEFH MN-----QFI-----NADR-----KADLATK-----LNLTERQIK-IWFQNRRAKRR-----
 SPE35398_1::Strongylocentrotus -----FSKAQTYELERRFRQQ-----RYL-----SAPE-----REHLASI-----IRLSPTQVK-IWFQNHRYKLR-----
 SRP04407_1::Caenorhabditis_elegans -----FTQEQLAELDNAFQKS-----HYP-----DIYV-----REELARI-----TKLNEARIQ-VWFQNRRAKHRK-----
 SSF10042_1::Salmo -----FTSAQLEVLERFFQES-----QYP-----DIHS-----RELLASQ-----TQLSEARVQ-IWFQNRVVKWRK-----
 O73917::Oryzias_latipes -----RNRRTSFTQE QIEALEKEFERT-----HYP-----DVFA-----RERLAAK-----IDLPEARIQ-VWFSNRRAKWR-----
 ACF08172_1::Allium -----LSKDQSALEESFKGH-----NTL-----NPKQ-----KQALAKQ-----LNLRPRQVE-VWFQNRRAKTK-----
 BNP12498_1|Arabidopsis_thaliana -----FSEEQIKSLEMMFSE-----TRL-----EPRK-----KVQLARG-----LSLQPRQVA-IWFQNKRARWK-----
 BVP05281_1|Vitis_vinifera -----LTKDQSALEESFKQQ-----STL-----NPKQ-----KQALADR-----LNLRPRQVE-VWFQNRRAKTK-----
 CSE21246_1::Citrus -----LTKEQSALEESFKQH-----STL-----NPKQ-----KQALARQ-----LNLRPRQVE-VWFPNRRARTK-----
 CSE25906_1::Citrus -----LTQDQVRLEETCFNAN-----QKL-----QVDR-----KLELARR-----LGLPPRQIA-VWYQNRRAEK-----
 GHP06126_1::Gossypium_arboreum -----TADQIREMEALFKES-----PHP-----DEKQ-----RQQLSKQ-----LGLAPRQVK-FWFQNRRTQIK-----
 GMD03350_1::Glossinamorsitans -----FSSFQRKLEIQFQQQ-----KYI-----TKPD-----RRKLAAR-----LNLTAQVK-VWFQNRMKWR-----
 HAD05132_1::Helianthus -----LSKEQSAYLEETFKEH-----NTL-----NPKQ-----KLALAEQ-----LHLRPRQVE-VWFQNRRAH-----
 HVP89597_1::Triticum_aestivum -----LSKDQAAVLEECFKTH-----STL-----NPKQ-----KRALANR-----LGLRPRQVE-VWFQNRRAKTK-----
 INP02125_1::Populus_trichocarpa -----LTKQQSIVLEDSFKQH-----TTL-----NSKQ-----KQELARR-----LNLRPRQVE-VWFQNRRAKTK-----
 TMD04719_1::Triticum -----LTAEQAALLEKSFRAH-----NVL-----SHGE-----KHDLAEQ-----LGLKPSKWK-VWFQNRRAKTK-----
 lftz::Drosophila_melanogaster -----KRTRTYTRYQTLLELEKEHF N-----RYI-----TRRR-----RIDIANA-----LSLSERQIK-IWFQNRMKSKK-----
 lnk2P|Drosophila_melanogaster -----RKRRVLFTKAQTYELERRFRQQ-----RYL-----SAPE-----REHLASL-----IRLTPQVK-IWFQNHRYKTKR-----
 O18381::Drosophila_melanogaster -----RRQRTHFTSQQLQELEHTFSRN-----RYP-----DMST-----REIAMW-----TNLTEARVR-VWFQNRRAKWRK-----
 O35160::MOUSE -----RRQRTHFTSQQLQELEATFORN-----RYP-----DMST-----REIAVW-----TNLTEARVR-VWFQNRRAKWRK-----
 O42477::Danio_rerio -----RRHRTIFTSYELEELEKAFNEA-----HYP-----DVYA-----REMLAMK-----TELPEDRIQ-VWFQNRRAKWRK-----
 P09532::TRIGR -----KRPRTAFSASQLRLKQEFQQS-----NYL-----TEQR-----RRSLAKE-----LTLSERQIK-IWFQNKRAKIKK-----
 BMD62598_1::Bombyx -----FTPQQLSELES LFQKT-----HYP-----DVFL-----REEVALR-----ISLSEARVQ-VWXQRRRAKWRK-----
 HJP02417_1::Paracoccidioides_brasiliensis_Pi -----ATQDQLTLEMEFNKNP-----TP-----TASV-----RDRI AEE-----INMTERS VQ-IWFQNRRAKIK-----
 MIP04433_1::Ovis_aries -----FTQNQVTRLEYMFSIK-----HYL-----SAQE-----REQISSE-----IGLKPNQVK-IWFQNHRYKIKR-----
 MIP04826_1::Rattus_norvegicus -----FTQNQVTRLEYMFSMK-----QYL-----SAQE-----REQISSE-----IGLKPNQVK-IWFQNHRYKIKR-----
 SSF08630_1::Salmo -----FSQAQVYELERRFKQQ-----KYL-----SAPE-----REHLASM-----IHLSPQVK-IWFQNHRYKMKR-----
 AMF17136_1::Ambystoma -----YSKGQLRELEKEYASS-----KFI-----TKDR-----RRQIATD-----TNLSERQIT-IWFQNRVKEK-----
 BNP16314_1::Arabidopsis_thaliana -----HTTDQIRHMEALFKET-----PHP-----DEKQ-----RQQLSKQ-----LGLAPRQVK-FWFQNRRTQIK-----
 CND11316_1::Filobasidiellaneoformans -----TNDVQLAMLSDFVQRT-----QYP-----STEE-----RDELARQ-----LGMTSRSVQ-IWFQNRRAVK-----
 IPD11218_1::Ictalurus -----FTKEHLELLRMAFNVD-----PYP-----GISV-----RESLSQA-----TGLPESRIQ-VWFQNKRAR-----
 SPE28460_1::Strongylocentrotus -----FTRAQLDVLET LFSRT-----RYP-----DIFM-----REEVAMK-----INLPESRVQ-VWFKNRRACKR-----
 PCE03163_1::Phaseolus -----KTPFQLETLEKAYAVD-----NYP-----SETM-----RGELSEK-----LGLSDRQLQ-MWFCHRR LKDK-----
 PGD06527_1::Picea -----KTPQQVEGLESFYAEH-----KYP-----SEAM-----KAQLSEE-----LGLTEKQVQ-GWFC HRR LKDK-----
 PTP01283_1::Caenorhabditis_elegans -----FNDTQLDELEKCFKMC-----QYP-----DVSL-----REKLSKE-----INLPEARIQ-VWFKNRC AKHRR-----
 O15499::HUMAN -----RRHRTIFSEEQLQALEALFVQN-----QYP-----DVST-----RERLAGR-----IRLREERVE-VWFKNRRAKWR-----
 O54751::MOUSE -----RRERTTFTRSQLEELEALFAKT-----QYP-----DVYA-----REEVALK-----INLPESRVQ-VWFKNRRACKR-----

EHP01599_1::Postia_placenta_Mad_698_R	IPKHALQTLQVFKDD	KFP	SVET	RKNLAAE	LRVTPRQVQ	VWFQNKQR
GAP02607_1::Oryzias_latipes	FSESQMSALVQRFVQ	RYL	TPAE	MKNLAKM	TGLTXQVQK	TWFQNR
INP04052_1::Mirabilis_jalapa	FSNEQVKSLLETIFKLE	TKL	ETKK	KLQVARD	LGLQPRQVA	IWFQNKARWK
OOP03311_1::Caenorhabditis_elegans	FNRRQQLVLETLFEAT	QYP	DVFT	REKVAEQ	IQLQESRIQ	VWFKNRRAKHR
SMP06236_1 Vitis_vinifera	LNQEQVRLLEASFDAG	KKL	EPER	KFQLARD	LGVPPRQIA	IWYQNKRARWK
SRD01034_1::Stevia	FTDKQISFLEYMFETQ	SRP	ELRM	KHQLAHK	LGLHPRQVA	IWFQNKRARSK
1fjl1::Drosophila_melanogaster	RRSRRTTFASQLDELERAFERT	QYP	DIYT	REELAQR	TNLTEARIQ	VWFQNRRLRLRK
1ftt::Rattus_norvegicus	RKRRVLFSAQVYELERRFKQQ	KYL	SAPE	REHLASM	IHLTPTQVK	IWFQNHRYKMKR
BMD20907_1::Bombyx	FTGDQQLRLEQLTLEKT	QYI	NGTD	RRELAQK	WGIGEKGIK	IWFQNRMMKNKR
BMD49473_1::Bombyx	FTTEQINYLENEFKKS	HYI	SAVQ	RKEIANI	VNVPKVIK	IWFQNRMRREKK
CPG01541_1::Coccidioides	LTKEQVETLEAQFRAQP	KP	TSNV	KRQLAMQ	TNLTLPRVA	NWFQNRRAKEK
O43316::HUMAN	RRRTTFSPSQAEALEKEFQRG	QYP	DSVA	RGKLATA	TSLPEDTVR	VWFSNRRAKWRR
ACF11510_1::Allium	YTPEQVEALERVYSEC	PKP	SSIR	RQQLIRECPIL	SNIEPKQIK	VWFQNRRCREK
CSE21677_1::Citrus	YTPEQVEALERLYHEC	PKP	SSMR	RQQLIRECPIL	SNIEPKQIK	VWFQNRRCREK
IPD21151_1::Ictalurus	FSKQATEVLNEYFYSHLSNPYP		SEEA	KEELAKQ	CGITVSQVS	NWFGNKRIRYKK
PCE03158_1::Phaseolus	EQVEALERLYHEC	PKP	SSLR	RQQLKKECPIL	CNIEPKQIK	VWFQNRRCREK
PPD21046_1::Pongo	FNKQVTEILNEYFYSHLSNPYP		SEEA	KEELAKK	CGITVSQVS	NWFGNKRIRYKK
P36200::XENLA	KRPRTTITAKQLETLKNAYNNS	PKP	ARHV	REQLSSE	TGLDMRVVQ	VWFQNRRAKEKR
P50481::MOUSE	KRPRTTITAKQLETLKSAYNTS	PKP	ARHV	REQLSSE	TGLDMRVVQ	VWFQNRRAKEKR
INP10932_1::Petroselinum_crispum	FGEDAIAKRLNEAFKEN	HYP	KRNV	KESLARE	LGLTLRQVD	KWFGNSR
TVE09682_1::Trichomonas	FNKEQKIKLERIFKIN	PTP	KIRQ	RDEIARE	LNIPLKSVT	YWFQNR
BNP10524_1::Arabidopsis_thaliana	KTPFQLQTLLEEVYAE	TYP	SEAT	RAELSEK	LDLSRQLQ	MWFCHRRLKDKK
CSE06316_1::Citrus	KTPAQVMALEKIFYNEH	KYP	TEEM	KSQVAEQ	IGLTEKQVS	GWFCHRRLKEKR
PBD14388_1::Paracoccidioides	ATQDQLATLEMEFNKNP	TP	TAAV	REERIAEE	INMTERSQV	IWFQNR
SPE07312_1::Strongylocentrotus	FTRNQVYSMERRFDQQ	RYL	TSVE	RKEFSNS	IGLDDHHVK	IWFQNRSSKLLK
TDP00719_1::Caenorhabditis_briggsae_AF16	FTEAQSLLLEAFQES	HYP	DQTA	KKDMAEK	LDIPEDRIT	VWFQNRRAKWRR
WBP00326_1 Brugia_malayi	ISAKSLETLKQAYQAS	SKP	ARHV	REQLAAD	TGLDMRVVQ	VWFQNRRAKEKR
ACF08938_1::Allium	TAEQVEALERVYAEC	PKP	SSMR	RQQLVRDCPIL	SNIEPKQIK	VWFQNRRCREK
ACF10646_1::Allium	YTPEQVEALERVYSEC	PKP	SSIR	RQQLIRECPIL	SNIEPKQIK	VWFQNRRCREK
BVP09757_1 Vitis_vinifera	YTPEQVEALERLYHDC	PKP	SSLR	RQQLIRECPIL	SNIEPKQIK	VWFQNRRCREK
GMD02533_1::Glossinamorsitans	FSKQASEILNEYFYSHLSNPYP		SEEA	KEELARK	CGITVSQVS	NWFGNKRIRYKK
1lfuP::MOUSE	RRKRRNFNKQATEILNEYFYSH	PYP	SEEA	KEELAKK	SGITVSQVS	NWFGNKRIRYKK
HGP09547_1::Tetraodon_nigroviridis	ISSDQLDKLENVFEIR	KHL	NASE	QGRLGKA	IGLTEEQVN	EWFEQRNRWR
Q62231::MOUSE	FKEKSRGVLREWYAHN	PYP	SPRE	KRELAAE	TGLTTTQVS	NWFKNRRQRDR
Q9NPC8::HUMAN	FKEKSRVLEREWYAHN	PYP	SPRE	KRELAAE	TGLTTTQVS	NWFKNRRQRDR
1akhB::Saccharomyces_cerevisiae	RGHRFTKENVRILESWFAKNIENPYL		DTKG	LENLMKN	TSLSRIQIK	NWVSNRRRKEKT
PBD05144_1::Paracoccidioides	LTKEQVDTLAQFQHP	KP	NSNV	KRQLATQ	TNLSLPRVA	NWFQNRRAKAK
SPE02053_1::Strongylocentrotus	FNLAQINALERIFLDV	EYP	DGYL	KAKLANR	LEVDENTVQ	IWFQNRRAKKR
1b72B::HUMAN	RKRRNFNKQATEILNEYFYSHLSNPYP		SEEA	KEELAKK	CGITVSQVS	NWFGNKRIRYKK
1b8iB Drosophila_melanogaster	RRNFSKQASEILNEYFYSHLSNPYP		SEEA	KEELARK	CGITVSQVS	NWFGNKRIRYKK
O88609::MOUSE	KRPRTILTQQRRFAKASFEVS	SKP	CRKV	RETLAAE	TGLSVRVVQ	VWFQNRRAKMKK
P20154::Caenorhabditis_elegans	RGPRTTIKAKQLETLKNAFAAT	PKP	TRHI	REQLAAE	TGLNMRVIQ	VWFQNRSSKERR
P20271::Caenorhabditis_elegans	KRPRTTISAKSLETLKQAYQTS	SKP	ARHV	REQLASE	TGLDMRVVQ	VWFQNRRAKEKR
P37137::XENLA	RGPRTTIKAKQLETLKAAFIAT	PKP	TRHI	REQLAQE	TGLNMRVIQ	VWFQNRSSKERR
P48742::HUMAN	RGPRTTIKAKQLETLKAAFAAT	PKP	TRHI	REQLAQE	TGLNMRVIQ	VWFQNRSSKERR
P528::Danio_rerio	RGPRTTIKAKQLETLKAAFAAT	PKP	TRHI	REQLAQE	TGLNMRVIQ	VWFQNRSSKERR

MIP05948_1::Rattus_norvegicus -----FSEEQICYLEDFFGNTC--HYP-----DSYQ-----KEEIIARR-----LNITTDRIIT-VWFQNRRSKFRK-----
Q27350::Drosophila_melanogaster -----FKEKRSRSLRDWYSHN--PYP-----SPRE-----KRDLEAE-----TGLTTTQVS-NWFKNRRQRDR-----
P09088::Caenorhabditis_elegans -----RGPRTTIKQNQLDVLNEMFSNT--PKP-----SKHA-----RAKLALAE-----TGLSMRVIQ-VWFQNRRSKERR-----
P34764::Caenorhabditis_briggsae_AF16 -----RGPRTTIRQNQLDVLNEMFSNT--PKP-----SKHA-----RAKLALAE-----TGLSMRVIQ-VWFQNRRSKERR-----
AME08554_1::Apis -----FSEEQKEALRLAFAL--DPYP-----NVAT-----IEFLAGE-----LAL-SSRTITNWFHHRMRLK-----
O73708::Danio_rerio -----FKERTSLLREWYLD--PYP-----NPSK-----KRELAQA-----TGLTPTQVG-NWFKNRRQRDR-----
O73709::Danio_rerio -----FKERTRGLLREWYLD--PYP-----NPSK-----KRELAQA-----TGLTPTQVG-NWFKNRRQRDR-----
O73916::Oryzias_latipes -----FKERTRGLLREWYLD--PYP-----NPGK-----KRELAHA-----TGLTPTQVG-NWFKNRRQRDR-----
O93282::Danio_rerio -----FKERTSLLREWYLD--PYP-----NPSR-----KRHLAQA-----TGLTPTQVG-NWFKNRRQRDR-----
O95475::HUMAN -----FKERTRNLLREWYLD--PYP-----NPSK-----KRELAQA-----TGLTPTQVG-NWFKNRRQRDR-----
PTE05656_1::Poncirus -----TPLQAKALKFYSEE--KYP-----TKRE-----MEGLAAA-----LDLTYKQVR-TWIEKRRRDK-----
Q23175::Caenorhabditis_elegans -----FKERTSLLREWYLD--PYP-----NPPK-----KKELANA-----TGLTQMVG-NWFKNRRQRDR-----
Q5TYZ2::Danio_rerio -----FKERTRHLLREWYLD--PYP-----NPSK-----KRELAQA-----TGLTPTQVG-NWFKNRRQRDR-----
Q94166::Caenorhabditis_elegans -----FRDKSRVLLRDWYCRN--SYP-----SPRE-----KRELAEK-----THLVTQVS-NWFKNRRQRDR-----
AME10017_1::Apis -----FKHHQLRTMKSIFYAIN--HNP-----DAKD-----LKQLSQK-----TGLPKRVLQ-VWFQNAKAWRR-----
HGP07414_1|Caenorhabditis_briggsae_AF16 -----LNENQLRILKQTYQGN--QRP-----DTNT-----KEQLVKM-----TGLNARVIR-VWFQNKCKDKK-----
Q9UIU6::HUMAN -----FKEKSRNALKELYKQN--RYP-----SPAЕ-----KRHLAKI-----TGLSLTQVS-NWFKNRRQRDRN-----
P34765::CAEVU -----RGPRTTIKQNQLDVLNEMFSNT--PKP-----SKHA-----RAKKALE-----TGLSMRVIQ-VWFQNRRSKERR-----
luhsA::MOUSE -----AATMTEDQVEILEYFNKVN--KHP-----DPTT-----LCLIAAE-----AGLTEEQTQ-KWFKQRLAEWRR-----
P10180::Drosophila_melanogaster -----KKQRVLFSEEQKEALRLAFAL--DPYP-----NVGT-----IEFLANE-----LGL-ATRTITNWFHHRMRLK-----
lakhA|Saccharomyces_cerevisiae -----ISPQARAFLEEVFRRK--QSL-----NSKE-----KEEVAKK-----CGITPLQVR-VWFINKMRS-----
CCE07653_1::Cyprinus -----LEGTVRSALSYFVK--CPKP-----NTLE-----ITHISDD-----LGL-ERDVVRVWFENRRQKGR-----
ECD00006_1::Equus -----IENRVRGLENMFLQ--CPKP-----TLQQ-----ISHIAQQ-----LGL-EKDVRVWFENRRQKGR-----
GAP00667_1::Oryzias_latipes -----LEGAVRSALSYFIK--CPKP-----NTQE-----ITHISDD-----LGL-ERDVVRVWFENRRQKGR-----
GRD08656_1::Gossypium -----PTPLQLQILENIYEQG--TGT-----PSKQKIKEIASELAQH-----GQISETNVY-NWFQNRARSKR-----
HAD01349_1::Helianthus -----PTPVQLQILERLFEQG--NGT-----PSKQKIKEITSELSQH-----GQISETNVY-NWFQNRARSKR-----
GAP02710_1|Tetraodon_nigroviridis -----KTKEQLDLVKQHFRLC--QWP-----KSED-----YTELVKL-----TNLPRADVI-QWFGDTRYAVKN-----
HSP00500_1::Caenorhabditis_briggsae_AF16 -----LNENQLRILKQTYQGN--QRP-----DTNT-----KEQLVEM-----TGLNARVIR-VWFQNKCKDKK-----
O88706::MOUSE -----KRARTSF TAEQLQVMAQFAQD--NNP-----DAQT-----LQKLADM-----TGLSRVVIQ-VWFQNCRARHKK-----
P50458::HUMAN -----KRMRTSFKHHQLRTMKSIFYAIN--HNP-----DAKD-----LKQLAQK-----TGLTKRVLQ-VWFQNAKAFRR-----
PPD17649_1::Pongo -----SHEQLSALKGSFCRN--QFP-----GQSE-----VEHLTKV-----TGLSTREVR-KWFSDRRYHCRR-----
PYD04598_1::Porphyra -----TESQRRLLSATFAQN--PYP-----DVTT-----KNLLAEQ-----LGVNRPVVS-KWFQHRQRARR-----
ECD02963_1::Equus -----TEDQVEILEYFNKVN--KHP-----DPTT-----LCLIAAE-----AGLSEETQ-KWFKQRLAQWRR-----
IPD13553_1::Ictalurus -----IRVSVKGALESHFLK--CPKT-----SAQE-----ISTLADT-----LQLGEGRVVRVWVWENRRQKE-----
PTJ02837_1::Pantroglodytes -----TEDQVEILEYFNKVD--KHP-----DSTT-----LCLIAAE-----AGLSEETQ-KWFKQRLAKWRR-----
P13528::Caenorhabditis_elegans -----RKKRTSIAAPEKRELEQFFKQ--QPRP-----SGER-----IASIADR-----LDL-KKNVVRVWFENRRQKQKR-----
P16241::Drosophila_melanogaster -----RKKRTSIEVSVKGALEQHFHK--QPKP-----SAQE-----ITSLADS-----LQL-EKEVVRVWFENRRQKEKR-----
P20263::MOUSE -----RKKRTSIENRVRWSLETMFLK--CPKP-----SLQQ-----ITHIANQ-----LGL-EKDVRVWFENRRQKGR-----
P20264::HUMAN -----RKKRTSIEVSVKGALESHFLK--CPKP-----SAQE-----ITNLADS-----LQL-EKEVVRVWFENRRQKEKR-----
P20265::HUMAN -----RKKRTSIEVSVKGALESHFLK--CPKP-----SAQE-----ITSLADS-----LQL-EKEVVRVWFENRRQKEKR-----
P20267::Rattus -----RKKRTSIEVGVKGALESHFLK--CPKP-----SAHE-----ITGLADS-----LQL-EKEVVRVWFENRRQKEKR-----
P31362::MOUSE -----RKKRTSIETNIRLTLEKRFQD--NPKP-----SSEE-----ISMIAEQ-----LSM-EKEVVRVWFENRRQKEKR-----
P31363::XENLA -----RKKRTSIEVGVKGALENHFLK--CPKP-----SAHE-----ITSLADS-----LQL-EKEVVRVWFENRRQKEKR-----
CSE07550_1::Citrus -----PTPVQLQILESIFDQG--TGT-----PSKQKIKEITVELSQH-----GQISETNVY-NWFQNRARSKR-----
GRD14308_1::Gossypium -----PTPVQLQILERIFDQG--TGT-----PSKQKIKEITSELSQH-----GQISETNVY-NWFQNRARSKR-----
P50480::Rattus -----RVRTVLNEKQLHLTRLTCYAA--PRP-----DALM-----KEQLVEM-----TGLSPRVIR-VWFQNKCKDKK-----

P53407::Danio_rerio -----RVRTLNEKQLHLRLTCYNAN--PRP-----DALM-----KEQLVEM-----TGLSPRVIR-VWFQNKRCCKDKK-----
ACF08129_1::Allium -----YTPDXNQVLEEFYCRN--PHP-----DAND-----RKQLGKALG-----FTDDRIRK-YWFQNRRAVDRR-----
MIP03533_1|Brugia_malayi -----LDTAQKLSLDTFRFRI--DPRP-----DNAR-----MVEIATL-----LDL-DHDVVRVWFQNRQKLRK-----
PPD00896_1::Pongo -----FPKVATNIMRAWLFWHLTHPYP-----SEEQ-----KKQLAQD-----TGLTILQVN-NWFINARRR-----
SPE20651_1::Strongylocentrotus -----FLKSATNIMRAWLFWHLTHPYP-----SEEQ-----KKQLAQD-----TGLTILQVN-NWFINARRR-----
SSF51768_1::Salmo -----IDTNIRVALEKSFLQ--NQKP-----SSDE-----ISLIADQ-----LNM-EKEVIRVWFQNRQKEKR-----
Q8SIX5::HUMAN -----FKERSRAALKACYRGN--RYP-----TPDE-----KRRLATL-----TGLSLTQVS-NWFKNRQRDR-----
SPE28381_1::Strongylocentrotus -----FKERTSLLREWYLD--PYP-----NPTK-----KRELAQA-----TGLTPTQVG-NWFKN-----
O35652::MOUSE -----KRARTSFTADQLQVMQAQFAQD--NNP-----DAQT-----LQKLAER-----TGLSRRVIQ-VWFQNCRAHKK-----
P29673::Drosophila_melanogaster -----KRMRTSFKHHQLRMTKSYFAIN--HNP-----DAKD-----LKQLSQK-----TGLPKRVLQ-VWFQNAKAWRR-----
CND11255_1::Filobasidiellaneoformans -----FTKRELEALEVLWSIA--KSP-----SKYE-----RQLRG--AWL--G-VKTKHIT-VWFQKRQEEKR-----
PPG46627_1::Physcomitrella -----ASQTEVLERAYAVE--KYP-----SEAT-----RQKLVDR-----LDLSDKQLQ-IWFTHRRYKDRR-----
1au7A::Rattus_norvegicus -----KRRRTTISIAAKDALERHFGE--HMKP-----SSQE-----IMRMAEE-----LNL-EKEVVRVWFQNRQREKR-----
1bw5::Rattus_norvegicus -----TTRVRTVLEKQLHLRLTCYAN--PRP-----DALM-----KEQLVEM-----TGLSPRVIR-VWFQNKRCCKDKK-----
DJP03285_1::Girardia_tigrina -----IEANVKSILESSFMK--LSKP-----SAQD-----ISSLAEK-----LSL-EKEVVRVWFQNRQ-----
HPP00771_1::Strongylocentrotus_purpuratus -----FTDLQRRTLHAIFKE--NKRK-----SKEM-----QITIAQQ-----LGL-ELSTVSNFFMNARRR-----
1cqtB::HUMAN -----RKKRTSIEETNIRVALEKSFLQ--NQKP-----TSEE-----ITMIADQ-----LNM-EKEVIRVWFQNRQKEKR-----
O97552::BOVIN -----KRRRTSIEENRVRGNLESFMLQ--CPKP-----TLQQ-----ISHIAQQ-----LGL-EKDVRVWFQNRQKQKR-----
P10036::BOVIN -----KRRRTTISIAAKDALERHFGE--QNKP-----SSQE-----ILRMAEE-----LNL-EKEVVRVWFQNRQREKR-----
P10037::Rattus -----KRRRTTISIAAKDALERHFGE--HMKP-----SSQE-----IMRMAEE-----LNL-EKEVVRVWFQNRQREKR-----
P14859::HUMAN -----RKKRTSIEETNIRVALEKSFLQ--NQKP-----TSEE-----ITMIADQ-----LNM-EKEVIRVWFQNRQKEKR-----
P25425::MOUSE -----RKKRTSIEETNIRVALEKSFME--NQKP-----TSED-----ITLIAEQ-----LNM-EKEVIRVWFQNRQKEKR-----
P28069::HUMAN -----KRRRTTISIAAKDALERHFGE--QNKP-----SSQE-----IMRMAEE-----LNL-EKEVVRVWFQNRQREKR-----
P31364::XENLA -----RKKRTSIEVSVKGVLETHFLK--CPKP-----AALE-----ITSLADS-----LQL-EKEVVRVWFQNRQKEKR-----
ENP07708_1::Aspergillus_nidulans_FGSC_A4 -----YSPEDYAILAEYQRN--PKP-----DKIS-----RASIVSRVS-----LGEKEVQ-IWFQNRQNDRR-----
P24350::Drosophila_melanogaster -----KKRTSIAAPEKRSLEAYFAV--QPRP-----SGEK-----IAAIAEK-----LDL-KKNVVRVWFQNRQKQKR-----
AME03197_1::Apis -----LTTDQEAVLQEQFNRP--RAP-----HTAD-----IVLLAAE-----TGLSEADVE-AWYSIRLAQWRK-----
P09086::HUMAN -----RKKRTSIEETNIRVALEKSFLQ--NQKP-----TSEE-----ILLIAEQ-----LHM-EKEVIRVWFQNRQKEKR-----
P17208::MOUSE -----KRRRTSIAAPEKRSLEAYFAV--QPRP-----SSEK-----IAAIAEK-----LDL-KKNVVRVWFQNRQKQKR-----
P20268::Caenorhabditis_elegans -----RKKRTSIEVNVKSLRLEPHFQS--NQKP-----NAQE-----ITQVAME-----LQL-EKEVVRVWFQNRQKEKR-----
P39881::CANFA -----KKPRVVLAPPEEKEALKRAYQQ--KPYK-----SPKT-----IEELATQ-----LNL-KTSTVINWFHNYRSRIRR-----
21fb|Rattus_rattus -----RRNRFKWPASQQLFQAYERQ--KNP-----SKEE-----RETLVEE-----CNVTEVRVY-NWFANRRKEE-----
Q5TV95::Anopheles_gambiae_str._PEST PRTIWDGEQKTHCFKERTSLLREWYLD--PYP-----NPTK-----KRELAQA-----TGLTPTQVG-NWFKNRQRDR-----
CPG05237_1::Coccidioides -----LPKPTTDILRAWFYEHLDHPYP-----SEQD-----KQMFMT-----TGLTISQIS-NWFINARRRH-----
TVE05086_1::Trichomonas -----FSDSQRSILMHWLKNHQSHPYP-----TSSE-----KQELIEK-----TGLNRDQIN-VWFTNNVRH-----
CCD00425_1::Conidiobolus -----GNSWQINRLFEEFFERC--PRP-----TRAQ-----VHSLSVE-----LEMPKISIR-IWFQNRHSKQ-----
Q94165::Caenorhabditis_elegans -----FKSKSRNVLRLDAYKKC--QYP-----SVED-----KRRLAQQ-----TELSIQVS-NWFKNKQRER-----
Q9NJB5::Drosophila_melanogaster -----KKPRLVFTDLQRRTLQAIFKE--TKRP-----SKEM-----QVTIARQ-----LGL-EPTTVGNFFMNARRR-----
Q22811::Caenorhabditis_elegans -----RLTFTETQLKSLQKSFQQ--NHRP-----TREM-----RQKLSAT-----LEL-DFSTVGNFFMNSRRLR-----
P39880::HUMAN -----KKPRVVLAPPEEKEALKRAYQQ--KPYK-----SPKT-----IEDLATQ-----LNL-KTSTVINWFHNYRSRIRR-----
AYP01413_1|Caenorhabditis_briggsae_AF16 -----FTDIQRRTLQAIFKE--TKRP-----SREM-----QLTISQQ-----LGL-DPTTVANFFMNARRR-----
HCP08366_1|Caenorhabditis_elegans -----FTDIQKRTLQAIFKE--TQRP-----SREM-----QQTIAEH-----LRL-DLSTVANFFMNARRSR-----
SSP00594_1|Brugia_malayi -----FSDVQKRTLQAIFKE--TERP-----SKEM-----QQTIAEH-----LGL-DPSTVSNYFNMARRSR-----
1wjhA::Drosophila_melanogaster -----RKTKEQLAILKSFLLQC--QWA-----RRED-----YQKLEQI-----TGLPRPEII-QWFGDTRYALKHGQLKW-----
1s7eA::MOUSE -----KKPRLVFTDVQRRTLHAIFKE--NKRK-----SKEL-----QITISQQ-----LGL-ELSTVSNFFMNAR-----
O60422|HUMAN -----KKQRLVFTDLQRRTLHAIFKE--NKRK-----SKEM-----QVTISQQ-----LGL-ELNLTVSNFFMNARRC-----

O95948::HUMAN -----KKSRLVFTDLQRRTLFAIFKE--NKRPSKEM-----QITISQQ-----LGL-ELTTVSNFFMNARRR-----
 P70512::Rattus -----KKPRLVFTDVQRRTLHAIFKE--NKRPSKEL-----QITISQQ-----LGL-ELSTVSNFFMNARRR-----
 Q5MD20::Danio_erio -----KKPRLVFTDVQRRTLHAIFKE--NKRPSKEL-----QITISQQ-----LGL-ELATVSNFFMNGRRR-----
 O14529::HUMAN -----KKPRVVLAPEEKEALRKAYQL--EPYPSQQT-----IELLSFQ-----LNL-KTNTVINWFHNYRSRMRR-----
 P20266::Rattus -----KKRRTSIAAPEKRSLEAYFAV--QPRPSSEK-----IAATAEK-----LDL-KKNVVRVWFCN-----
 Q19720::Caenorhabditis_elegans -----KRPRLVFTDIQKRTLQAIFKE--TQRPSREM-----QQTIAEH-----LRL-DLSTVANFFMNARRRSR-----
 lwi3A::HUMAN -----RTKISLEALGILQSFIDVG--LYPDQEA-----IHTLSAQ-----LDLPKHTII-KFFQNRQYHVK-----
 P70298::MOUSE -----KKPRVVLAPAEKEALRKAYQL--EPYPSQQT-----IELLSFQ-----LNL-KTNTVINWFHNYRSRMRR-----
 Q22812::Caenorhabditis_elegans -----RIRRFFTQTQLDSLHTVVFQ--QDRPSNREM-----QQALSAT-----LKL-NRSTVGNFFMNARRRLPK-----
 Q61JC4::Caenorhabditis_briggsae_AF16 -----RTVITDYQKDVLRVVFVN--ELHPTNEM-----IEQIATK-----LEM-SLRTVQNWFFHNHRTRSK-----
 AMF20019_1::Ambystoma -----ISVEALGILQSFIDVG--LYPDEEA-----IHTLSAQ-----LDLPKYTII-KFFQNRQY-----
 Q01826::HUMAN -----RPRTKISVEALGILQSFIDVG--LYPDEEA-----IQTLAQ-----LDLPKYTII-KFFQNRQYLLK-----
 O45080::Caenorhabditis_elegans -----KKTRLVFSDIQRRTLQAIFRE--TKRPSREM-----QITISQQ-----LNL-DPTTVANFFMNARRR-----
 O16218::Caenorhabditis_elegans -----RARCLLSQDQKSQLSIFFET--NPRPSDSE-----MKQLGST-----LNL-CKSTIINYFTNMRRR-----
 HJP06140_1::Gibberella_zeae_PH1 -----FAPHVEKRLEEYYQKQ--PFPNDSE-----SAFLASNLG-----IEPFHLG-LWFHRRER-----
 HGP08389_1|Brugia_malayi -----FRPVLLRVLETYFQKC--PFPDIAKRVEIANACN-----AHLQVDKRGVQLMPKEVVTPQVIA-NWFANKRREMRR-----

VNE blosum62

