Discovery notes

Open Access

Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'–5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch Dapeng Zhang¹, Huiling Xiong¹, Jufang Shan², Xuhua Xia¹ and Vance L Trudeau^{*1}

Address: ¹Centre for Advanced Research in Environmental Genomics (CAREG), Department of Biology, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada and ²Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, NY 10021, USA

Email: Dapeng Zhang - dzhan033@uottawa.ca; Huiling Xiong - hxion102@uottawa.ca; Jufang Shan - jus2013@med.cornell.edu; Xuhua Xia - xuhua.xia@uottawa.ca; Vance L Trudeau* - trudeauv@uottawa.ca

* Corresponding author

Published: 25 November 2008

Biology Direct 2008, 3:48 doi:10.1186/1745-6150-3-48

This article is available from: http://www.biology-direct.com/content/3/1/48

© 2008 Zhang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 18 September 2008 Accepted: 25 November 2008

Abstract

Maelstrom (MAEL) plays a crucial role in a recently-discovered piRNA pathway; however its specific function remains unknown. Here a novel MAEL-specific domain characterized by a set of conserved residues (Glu-His-His-Cys, EHHCHC) was identified in a broad range of species including vertebrates, sea squirts, insects, nematodes, and protists. It exhibits ancient lineage-specific expansions in several species, however, appears to be lost in all examined teleost fish species. Functional involvement of MAEL domains in DNA- and RNA-related processes was further revealed by its association with HMG, SR-25-like and HDAC interact domains. A distant similarity to the DnaQ-H 3'-5' exonuclease family with the RNase H fold was discovered based on the evidence that all MAEL domains adopt the canonical RNase H fold; and several protist MAEL domains contain the conserved 3'-5' exonuclease active site residues (Asp-Glu-Asp-His-Asp, DEDHD). This evolutionary link together with structural examinations leads to a hypothesis that MAEL domains may have a potential nuclease activity or RNA-binding ability that may be implicated in piRNA biogenesis. The observed transition of two sets of characteristic residues between the ancestral DnaQ-H and the descendent MAEL domains may suggest a new mode for protein function evolution called "active site switch", in which the protist MAEL homologues are the likely evolutionary intermediates due to harboring the specific characteristics of both 3'-5' exonuclease and MAEL domains.

Reviewers

This article was reviewed by L Aravind, Wing-Cheong Wong and Frank Eisenhaber. For the full reviews, please go to the Reviewers' Comments section.

Background

Germline cells among different species are characterized by the presence of a morphologically unique organelle called the germ plasm (also referred to as nuage, polar granules or mitochondrial cloud) [1,2]. This organelle has been considered the determinant of germline development. Very recently a germ plasm-specific small RNA pathway has been identified, in which a new type of small RNAs called PIWI-interacting RNAs (piRNAs) or repeatassociated small interfering RNAs (rasiRNAs) play a role in ensuring the genomic stability of germline cells by silencing certain endogenous genetic elements such as retrotransposons and repetitive sequences [3-8]. Different from short interfering RNAs (siRNAs) and microRNAs which are usually 21-22 nt long, piRNAs or rasiRNAs have longer nucleotide composition (26-31 nt) and 2'Omethyl modification in 3' end. Many germ plasm proteins are functionally important in piRNAs synthesis and function, including PIWI proteins (PIWI, Aubergine and AGO3) [4,9,10], VASA [11], MAEL [11], SPN-E [12,13], Oskar [14], Tudor domain proteins [15], Armitage [13], Krimper [11], Cutoff [16], Dead end [17] and Zucchini and Squash [18]. Their loss-of-function mutations commonly cause a huge reduction in the amount of piRNAs or rasiRNAs and an increase in transcript level of transposable elements in the germline cells [11,19,20] as well as the spindle-class gene phenotypes: failure in establishing anterior/posterior polarity in early oocytes, disrupted asymmetric subcellular mRNA localization of Oskar, Gurken and Biocoid, ectopic expression of Oskar and Gurken, failure to proceed to the karyosome stage [8,11,13,21].

The molecular functions of most germ plasm proteins in the piRNA pathway have been assigned based on domain examination, biochemical and genetic characterizations. For instance, PIWI proteins contain the PAZ and PIWI domains, which contribute to recognition of singlestranded RNA [22] and sequence-specific endonucleolytic cleavage of target nucleotide [23,24], respectively. VASA, SPN-E and Armitage share DEAD RNA helicase domains, which provide helicase activities for piRNA production or retrotransposon silencing [13,25]. Zucchini and Squash are putative nucleases, which are believed to be involved in piRNA maturation [18]. Other Dead end, Krimper and Tudor proteins, contain RNA binding domains RRM [26] or Tudor [27] which may facilitate the assembly of multiprotein RNA-induced silencing complex (RISC) and targeting substrate RNA recognition during cleavage. In contrast, although many studies including specific knockouts, protein interaction and cellular distribution experiments have been conducted, the definitive function of MAEL in piRNA pathway remains unknown. MAEL was initially identified in a genetic loss-of-function Drosophila mutant, whose germline cells exhibit incorrect posterior

localizations of several transcripts (i.e., Gurken, Oskar and Bicoid) [12]. It is a germ plasm-specific protein with all spindle-class gene phenotypes [12,13,21,28] and directly involved in the piRNA pathway [11,29]. The correct location of either SPN-E, VASA, Aubergine, Tudor or Krimper in germ plasm determines the location of MAEL [11], which in turn delineates the location of Dicer and Argonaute2 [21]. MAEL can shuttle between germ plasm and the nucleus [21]. Direct interaction between MAEL and chromatin remodeling proteins SNF5/INI1 and SIN3B during heterochromatin formation has also been demonstrated [30]. Therefore, MAEL is the only known protein connecting germ plasm and piRNA pathway to chromatin remodeling, a process required for piRNA-initiated genome transposon silencing [31]. In the present study, we were motivated to understand the putative function of MAEL using combined bioinformatic strategies including extensive homologous sequence mining, phylogenetic analysis, domain architecture, protein fold recognition, and structure modeling.

Results

A conserved MAEL-specific domain and its unique lineagespecific evolutionary expansion and loss

Domain annotation showed that mouse MAEL protein contains a HMG domain in its N-terminal segment, which is a DNA-binding module in many non-histone components and transcriptional regulators [32]. However, no domain information could be assigned for the C-terminal segments of MAEL proteins (240 amino acids long). We conducted homologous sequence searching for this region using PSI-BLAST against the NCBI NR database. Many unique homologues were identified in a broad range of species from veterbrates, echinoderms, insects, nematodes, to the protists (Entamoeba histolytica, Entamoeba dispar, and Trypanosoma brucei). We also examined NCBI nucleotide and Ensembl genome databases and identified eight other homologues in insects and urochordates (Ciona intestinalis and Ciona savignyi). Three more protist homologues were obtained through searching GeneDB database. A multiple sequence alignment was built for all the retrieved sequences (additional file 1) and a condensed one is shown in Figure 1. Although the overall sequence identity is very low, the conservation is apparent across all these MAEL homologues. Six residues Glu-His-His-Cys-His-Cys (EHHCHC) are highly conserved, suggesting that they may contribute to MAEL-specific activity. Thus the C-terminal segment appears to define a novel MAEL-specific domain that we now refer to as the MAEL domain.

For the majority of species, only one copy of MAEL domain exists. However, there are multiple MAEL homologues in several other species; for instance, two copies are found in sea squirts (*C. intestinalis* and *C. savignyi*) and

>Am_110759058									FLES (6) FPPLFTAKD-LSPVV
>Ag_118793711									FLLQ(3)MPLLFTDET-DVPRV
>Aa_108883695	PKE <mark>L</mark> EKLE	FYFIS <mark>F</mark> AY <mark>FC</mark> V <mark>T</mark> (1	L)GGT <mark>YI</mark> P <mark>AE</mark> M	G <mark>LVRY</mark> S <mark>E</mark> KDGV <mark>M</mark> D	KLHM <mark>FI</mark> DPGKLP-LG	MAYDA <mark>K</mark> QHSESDH) <mark>L</mark> PIPPDAK <mark>G</mark> EKE	-NDE <mark>II</mark> LK <mark>L</mark> FS	FLSQ(3)MPPLFTETN-DIRMV
>Aa 108875394	NNALEKLE	VFFMSCNYFCKT (1	l) tea <mark>fv</mark> p <mark>aei</mark>	ALIKYNELGVLD	KL <mark>HELI</mark> NPVRLP-LG	LAHEALTYSEQTH	E <mark>L</mark> PTPPNAM <mark>GET</mark> E	-FYT <mark>VLQKI</mark> LS	FTDY (6) KLAIMTDAK-EVPVI
>Gm 78540983		IAVNYFTKT (2	2) GNV <mark>YI</mark> PAEL	SVCEYSKOGVNR	I <mark>FHTLIN</mark> PGTNV-YG	HOYEAOHHSETTH	I <mark>L</mark> PLPPNAM <mark>G</mark> DEN	- <mark>lgtiynev</mark> lk	FLGA (3) YPPLYTVRE-NIHIV
>Dm 21429066	SHDLENAK	EVEVAENYETKA (2) TDV <mark>VVPAE</mark> F	AACEYSKEGTRS	TYSTMIDPGOTI-FG	OGSDALLHSSTTH	PLPPNAL GEKN	-MTKLYRNTVD	YLSK(6)TLVVFTPAE-NITMV
>Ci 23575304									FTGS (5) LLCMEN-EIRKN
>Ci 23576040									FIRQ (7) LPPIYCRMS-EVRKN
>Cs1_in_Ensembl	NMDVKQEH	rillsrQSLIEL (2	2) EEGILFCEL	ICVDIISGGIQN	IIWHQIIIDPGAIK-PV	LMSEVKFFRERING	21SKDCG-LAKSL	- I I AMWREL VA	FIRG(8)IPPIYSRMS-EIKKN
>Cs2_in_Ensembl									FVED (6) LPP <mark>LLC</mark> MEN-EIRKN
>Gg_118083700									FAQP (4) WPRF <mark>YC</mark> KSD-DRFRI
>Xt_118404620									FVCP (4) VVP <mark>VYT</mark> KAN-DIYRV
>Md_126306192	DQA <mark>V</mark> LGSF	FYFLNIFS <mark>HG</mark> EL (4	1)EQR <mark>FL</mark> PCEI	G <mark>CIKYS</mark> QEGIVA	E <mark>FHRFI</mark> DPGEVP-RG	FRFHC <mark>Q</mark> AASDASHI	TPISNFHS <mark>G</mark> DD-	-YAVVLQNLYR	FISP(4)WPPVYCKSD-DRYRV
>Hs 20306906	DOALLGGI	FYFLNIFSHGEL (4	1) EOR <mark>FL</mark> PCEI	GCVKYSOEGIMA	D <mark>FHSFI</mark> NPGEIP-RG	FRFHC <mark>OAA</mark> SDSSHI	TPISNFERGHN-	-OATVLONLYR	FIHP (4) WPPIYCKSD-DRTRV
>Ce 3873737									RVE(44)RFILVLOSE-LDLMV
>Cb_39591559									RVE(44)RFILVLQSE-LDLMV
>Eh 67477376 C									YLNS (3) KKLLVLKDE-TIGGD
									YIEQ(3) LAFFVSKEE-SLAGD
>Eh_67476664									
>Eh_67466465									FIK(10)LPI <mark>IIC</mark> TPFISS
>Eh_67484628									FIWT(9)VPI <mark>VIC</mark> LSFRRG
>Eh_67477376_N	EII <mark>T</mark> NDIP	VHFVNFEYCARA	MKM <mark>TY</mark> PIEL	G <mark>I</mark> CT <mark>Y</mark> K <u>MSE</u> FK-L <mark>L</mark> G	E <mark>FHQLI</mark> YCDIS	EFINK <mark>K</mark> QL-TNHH(G <mark>l</mark> dsksq-f <mark>lrk</mark> e	- <mark>Y</mark> KN <mark>IV</mark> NE <mark>L</mark> MK	YLKS(4)IYC <mark>IV</mark> KRKE-IRNPS
>Ed_167389979_C	EIN <mark>I</mark> FEFP	IHFFDFEFSSSK	TDG <mark>VI</mark> P <mark>LE</mark> L	G <mark>I</mark> ST <mark>Y</mark> Y I NESK-E <mark>I</mark> N	IF <mark>YH</mark> T <mark>LI</mark> KPS	HYNSSFERAVGVH	G <mark>I</mark> DLHMN <mark>Y</mark> TQS	- <mark>Y</mark> SEIINGLTN	YLNS (3) KKLLVLKDE-TIGGD
>Ed 167384018	TQTLENGV	FHFYDFEY <mark>AAQF</mark>	SEK <mark>IF</mark> PIEI	G <mark>ISSYS</mark> KENK-E <mark>I</mark> N	IS <mark>YH</mark> KLLYPG	KEKNVEARTOMIH	DARDP-RLEON	- <mark>Y</mark> SL <mark>VCIEL</mark> IK	YIEQ (3) LAFFVSKEE-SLAGD
>Ed 167391142									FIK(11)LPIVICTPFISS
>Ed 167395021	GEONIKOP	FFFTDFCINVTT	-DETCURTET	TOPR-T DRSAFRVK	PETOTINEW/PSO	VILRAKOHADEEH	TTOFNNDSDDTN		FIWT (9) VPIVICLSFRRG
>Ed 167389979 N									YLKS (4) ICCIVKRKE-TRNPS
>Tv_tviv441h03_q1k_3									FLSCKQVVLVNKGTLMD(4)
>Tb_71748114									FLSCERV <mark>VLI</mark> NKGSLMD(4)
									FLSYECV <mark>VFV</mark> NKDSLMD(4)
>Lb_LbrM35_V2_6220	AGD <mark>W</mark> SNIT	FVA <mark>V</mark> DTEA <mark>YA</mark> VM (1	L)HSVP <mark>L</mark> PAEY	A <mark>F</mark> LP(6)STSSAV <mark>L</mark> S	SPLHF <mark>F</mark> CHPGN(4)AE	N <mark>EENVL</mark> YNCLNT H I	JPYHSATF <mark>L</mark> TDN	if <mark>y</mark> dka <mark>v</mark> lvdrQ	FVRNPSV <mark>ILI</mark> SK(6)TLMD(4)
Consensus/80%	h	bbbhshhh.h.	phhPhEl	tha*Lppsb.	.aHpblpss	bb.psb.hspH	lsshpps	.blh.pl.p	alplhhps
>am 110759058	FSILTKMI	DASNEST (1) DET	VSTEALECA-	PNAAVO (A) PSTDIT	AFNFESKD <mark>EL</mark> C(1)T	PGLE DE KILD (WARTICDY	CEVINIK IECVHREKETDELO
>Am_110759058									CCEYLNIKLIEGVHREKETPFLQ
>Ag_118793711	ES <mark>ML</mark> EHII	SDHLSEIE <mark>L</mark> R <mark>I</mark>	<mark>IC</mark> P <mark>LAELF</mark> FR-	L <mark>KQ</mark> NVEL(6)TFPSVY <mark>I</mark>	AQQIITKD <mark>VY</mark> D (1) T	KGIS <mark>CD<mark>YH</mark>EEKD (:</mark>) VL <mark>YC</mark> P <mark>LSRCIF</mark>	WAYIISDN	CCQDMGIEPIPGKHVPLNANTNP
>Ag_118793711 >Aa_108883695	ES <mark>MLEHII</mark> EN <mark>ILKGII</mark>	SDHLSEIE <mark>LRI</mark> NQGSMDE (1) T <mark>L</mark> LV	ICP <mark>LAE</mark> LFFR- /CPLSELFYQ-	L <mark>KQNVEL (6) TFPSVY</mark> I LKRATES (6) TFPSVH <mark>I</mark>	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKD <mark>VY</mark> E(1)T	KGIS <mark>ODYH</mark> EEKD (: KDIS <mark>OE</mark> FHEDQG (:) VL <mark>YCP</mark> LSRCIF) GK <mark>YC</mark> PLSRCVF	WAYIISDN WAYIISDS	CCQDMGIEPIPGKHVPLNANTNP CCLDLSIE <mark>M</mark> KPGRHLPMNADTAL
>Ag_118793711 >Aa_108883695 >Aa_108875394	ES <mark>ML</mark> EHIL ENILKGIL ES <mark>LL</mark> SQLN	SDHLSEIE <mark>LRI</mark> NQGSMDE (1) TLL <mark>N</mark> DDVKLEYQ <mark>FLN</mark>	ICP <mark>LAE</mark> LFFR- /CPLSELFYQ- /IP <mark>L</mark> GE <mark>FF</mark> FH-	LKQNVEL (6) TFPSVY <mark>I</mark> LKRATES (6) TFPSVHI LKRATEK (6) TFPTKT <mark>V</mark>	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKD <mark>VY</mark> E(1)T ADILLKKD <mark>AY</mark> E(1)T	KGISCD <mark>YHEE</mark> KD (KDISCEFHEDQG (SGIACD <mark>FHE</mark> KLG (l)VL <mark>YOPL</mark> SRCIF l)GKYOPLSRCVF l)QR <mark>FCAL</mark> SKVVF	WAYIISDN WAYIISDS WS <mark>YII</mark> SDN	CCQDMGIEPIPGKHV=LNANTNP CCLDLSIEMKPGRHL=MNADTAL CCLDLSIDLIAGRHL=SNADTTL
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983	ES <mark>MLEHIL</mark> ENILKGIL ESLLSQLN VS <mark>VLDFL</mark> K	SDHLSEIELRI NQGSMDE(1)TLLA DDVKLEYQFLA SDIRASN(1)TLNA	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV-	LKQNVEL (6) TFPSVY <mark>I</mark> LKRATES (6) TFPSVHI LKRATEK (6) TFPTKT <mark>V</mark> MKESTCE (5) KPKSFH <mark>I</mark>	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFERD <mark>CF</mark> E(1)Q	KGISCD <mark>YHEEKD(</mark> KDISCEFHEDQG(SGIACDFHEKLG(NGIACQFHEDKD(L) VLYOPLSRCIF L) GKYOPLSRCVF L) QRFOALSKVVF L) SK <mark>YOTQ</mark> SIVTF	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY	CCQDMGIEPIPGKHV ⁻ LNANTNP CCLDLSIEMKPGRHL ⁻ MNADTAL CCLDLSIDLIAGRHL ⁻ SNADTTL MCRD <mark>I</mark> AVPLIAGRHI ⁻ QNTNLEA
>Ag_118793711 >Aa_108883695 >Aa_108875394	ES <mark>MLEHIL</mark> ENILKGIL ESLLSQLN VS <mark>VLDFL</mark> K	SDHLSEIELRI NQGSMDE(1)TLLA DDVKLEYQFLA SDIRASN(1)TLNA	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV-	LKQNVEL (6) TFPSVY <mark>I</mark> LKRATES (6) TFPSVHI LKRATEK (6) TFPTKT <mark>V</mark> MKESTCE (5) KPKSFH <mark>I</mark>	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFERD <mark>CF</mark> E(1)Q	KGISCD <mark>YHEEKD(</mark> KDISCEFHEDQG(SGIACDFHEKLG(NGIACQFHEDKD(L) VLYOPLSRCIF L) GKYOPLSRCVF L) QRFOALSKVVF L) SK <mark>YOTQ</mark> SIVTF	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY	CCQDMGIEPIPGKHV=LNANTNP CCLDLSIEMKPGRHL=MNADTAL CCLDLSIDLIAGRHL=SNADTTL
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983	ES <mark>MLEHIL</mark> ENILKGIL ESLLSQLN VS <mark>VLDFL</mark> K KS <mark>CF</mark> RYLE	SDHLSEIELRI NQGSMDE (1) TLLV DDVKLEYQFLV SDIRASN (1) TLNV CDDDFRD (3) KIQV	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV- /FDIQYLLFI-	LKQNVEL (6) TFPSVYI LKRATES (6) TFPSVHI LKRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI LKKEVMN (6) EKINKFA	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKDVYE(1)T 'ADILLKKDAYE(1)T TDAHFERD <mark>CF</mark> E(1)Q TDAFFKKD <mark>FF</mark> E(1)T	KGISOD <mark>YHEEKD</mark> (KDISOEFHEDQG(SGIAODFHEKLG(NGIAOQFHEDKD(AGIAOQ <mark>YHED</mark> ND(L) VL <mark>YOPLSRCIF L) GKYOPLSRCVF L) QRFOALSKVVF L) SKYOT<mark>O</mark>SIVTF L) TK<mark>YOTO</mark>SMVTF</mark>	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY WAYTFTDF	CCQDMGIEPIPGKHV ⁻ LNANTNP CCLDLSIEMKPGRHL ⁻ MNADTAL CCLDLSIDLIAGRHL ⁻ SNADTTL MCRD <mark>I</mark> AVPLIAGRHI ⁻ QNTNLEA
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066	ES <mark>MLEHIL ENILKGIL</mark> ESLLSQLN VS <mark>VLDFL</mark> K KS <mark>CFRYLE</mark> CF <mark>CL</mark> HWLA	SDHLSEIELRI NQGSMDE(1)TLL DDVKLEYQFL SDIRASN(1)TLN CDDDFRD(3)KIQ EK-AYNNHFE	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV- /FDIQYLLFI- /YGFESITSA-	LKQNVEL (6) TFPSVYI LKRATES (6) TFPSVHI LKRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI LKKEVMN (6) EKINKFA LYGYTEPVCLSPVL	AQQIITKD <mark>VY</mark> D(1)T AQAIIQKDVYE(1)T 'ADILLKKDAYE(1)T TDAHFERD <mark>CF</mark> E(1)Q TDAFFKKDFFE(1)T .ITAGCNSS <mark>MF</mark> D(1)E	KGISOD <mark>YHEEKD(</mark> KDISOEFHEDQG(SGIAOFHEKLG(NGIAOFHECKD(AGIAOYHEDND(LGIKOEYHEEIE) VLYOPLSRCIF) GKYOPLSRCVF) QRFOALSKVVF) SKYOTQSIVTF) TKYOTQSMVTF CTCOTLLTVKF	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY WAYTFTDF CCYWMTNT	CCQDMGIEPIPGKHVPLNANTNP CCLDLSIEMKPGRHLPMNADTAL CCLDLSIDLIAGRHLPSNADTTL MCRDIAVPLIAGRHLPQNTNLEA MCGDLAITVQPGKHIPAQTKPNY
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23576040	ES <mark>MLEHII</mark> ENILKGII ESLLSQLN VS <mark>VLDFL</mark> K KSCFRYLE CFCLHWLA QFCLNWLS	SDHLSEIELRI NQGSMDE(1)TLLY DDVKLEYQFLY SDIRASN(1)TLN CODDFRD(3)KIQY EK-AYNNHEY QNAQMAN(1)LSKI	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV- /FDIQYLLFI- /YGFESITSA- .HELEHLARE-	LKQNVEL (6) TFPSVYI LKRATES (6) TFPSVHI LKRATEK (6) TFPTKTV WKESTCE (5) KPKSFHI LKKEVMN (6) EKINKFA YGYTEPVCLSPVI LICNAVNTKFALTS	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFERDCFE(1)T IDAFFKKDFFE(1)T ITAGCNSSMFD(1)E IEDGFNSSMWE(1)E	KGIS DYHEEKD(: KDIS EFHEDQG(: SGIA DFHEKLG(: NGIA OFHEDKD(: AGIA OYHEDND(: LGIK EYHEEIE- PGIK RYHEDVD-	L) VLY PLSRCIF L) GKY PLSRCVF L) QRF ALSKVVF L) SKY TQSIVTF L) TKY TQSMVTF CTC TLLTVKF CYF TMLTIKK	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY WAYTFTDF CCYWMTNT CCFWMSEV	CCDDMGIEPIPGKHV LNANTNP CCDD.SIENKEGRHL MNADTAL CCDD.SIENKEGRHL MNADTAL MCCD AVPLIAGRHI QNTNLEA MCCD AITVQFGKHI AQTKPNY FSGI NVDITES-HL VKSQSF LAPY GFELTRN-HL EESDEPL
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575304 >Ci_23576040 >Csi_in_Ensembl	ES <mark>MLEHTL ENILKGTL ESLLSQIN VS<mark>VLDFLK</mark> KSCFRYLE CFCLHWLA QFCLNWLS VFCLDWLA</mark>	SDHLSEIELRI NQGSMDE (1) TLL DDVKLEYQFL SDIRASN (1) TLN CCDDDFRD (3) KIQV EK-AYNNHFE QNAQMAN (1) LCK QNAGMAN (1) LCK	ICPLAELFFR- ICPLSELFYQ- IIPLGEFFFH- IYPIQYLFYV- IFDIQYLLFI- IYGFESITSA- SHELEHLARE- IFELEFLVSE-	KQNVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVI ICNAVNTKFALTS ISNAMQTKTPLSS	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T ITAGCNSSMPD(1)E IEDGFNSSMWE(1)E IESGFTSSMWE(1)E	KGIS DYHEEKD (KDIS EFHEDQG (SGIA DFHEKLG (NGIA OFHEDKD (AGIA OYHEDND (LGIK EYHEEIE PGIK RYHEDVD IGIK EYHEEID	L) VLY PL SRCIF L) GKY PL SRCVF L) QRF AL SKVVF L) SKY TO SIVTF L) TKY TO SMVTF CTC TLLTVKF CYF TMLTIKK CYF TMLTIKK	WAYIISDN WAYIISDS WSYIISDN WAYTFIDY CCYMMINT CCFWMSEV CCFWISEV	CCODGIEPIPCKHV LNANTNP CCLD SIEKFORHL MNADTAL CCLD SIEKFORHL MNADTAL MCRD AVP_IAGRHL SNADTTL MCRD AVPUPCKHI SQTKPNY FSGINVDITES-HL VKSQSSF LAPY GFELTKN-HL SESDEPL LSPI HFFJ TEN-HL SESDEPL LSPI HFFJ TEN-HL SEQADCF
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575040 >Cs1_in_Ensembl >Cs2_in_Ensembl	ES <mark>ML</mark> EHIL ENILKGIL ESLISQIN VSVIDFIK KSCFRYLE CFCIHWLA QFCINWLS VFCIDWLA SFCLQWLA	SDHLSEIELR NQGSMDE(1)TLL DDVKLEYQFL SDIRASN(1)TLN CDDDFRD(3)KIQV EK-AYNNHEE QNAQMAN(1)LSKI QNAGMAN(1)LCKK NKGGYPNLER	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV- /FDIQYLLFIV- /YGFESITSA- .HELEHLARE- /FELEFLVSE- /YGFESMASA-	KQNVEL (6) TFPSVHI KRATES (6) TFPSVHI KRATEK (6) TFPTKTV WKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKFPLSS	AQQIITKDVYC(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFERDCFE(1)Q IDAFFKKDFFE(1)T ITAGFSSMFD(1)E IEDGFNSSMWE(1)E IESGFTSSMWE(1)E IQSGCNSSMFD(1)E	KGIS DYHEEKD (KDIS EFHEDQG (SGIA DFHEKLG (NGIA OFHEDKD (AGIA OYHEDND (LGIK EYHEEIE PGIK RYHEDVD IGIK EYHEEID AGIK YFHEENE	L) VLY PLSRCIF CRF ALSKVVF CRF ALSKVVF CSIVTF TCSIVTF CTCTLLTVKF -CTC TLLTVKF -CYF TMLTIKK -CYF TMLTIKK -CTC ALLTVKF	WAYIISDN WAYIISDS WSYIISDN WAYTFTDF CCYWMTNT CCFWMSEV CCFWMSEV CAYWMSDA	CCDDMGIEPIPGKHVSLMANTNP CCLDSIENKPGRHLMNADTAL CCLDSIENLAGRHLSNADTTL MCRDAVPILAGRHISNADTTL AGGDAITVQPGKHISNATKPNY FSGIVNDITES-HLVKSQSSF LAPYGFETRN-HLSESDEPL LSPIYHFPTEN-HLSEQADCF ELGVYDVITES-HFVKSEASF
>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23576040 >Csl_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl	ESMLEHIL ENILKGIL ESLLSQIN VSVIDFIK KSCFRYLE CFCIHWLA QFCINWIS VFCIDWLA SFCLQWLA SWCLERMA	SDHLSEIELRI NQGSMDE(1)TLL DDVKLEYQELV SDIRASN(1)TLN CDDDFRD(3)KIQV EK-AYNNHEN QNAQMAN(1)LSKI QNAGMAN(1)LSKI NKGGYPNLERK SIAGVDSPEI	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFFH- /YPIQYLFYV- /FDIQYLLFI- /YGFESITSA- .HELEHLARE- /YGFESITSA- /YGFESMASA- LITVEDLVIK-	KQNVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YOYTEPVCLSPVL ISNAMQTKTPLSS YANAQRTKTPLSS YANAQRTKLSLSL YQKKYHKEPSKTW	AQQIITKDVYC(1)T AQAIIQKDVYC(1)T ADAILLKKDAYC(1)T TDAHFERDCFE(1)Q ITDAFFKKDFFE(1)T ITAGCNSSMFD(1)E IEDGFNSSMWC(1)E IESGFTSSMWC(1)E IQSGCNSSMFD(1)E VSRELDVVLWD(1)S	KGIS DYHEEKD (KDIS EFHEDQG (SGIA DFHEKLG (NGIA OFHEDKD (AGIA OYHEDND (LGIK EYHEEIE- PGIK RYHEDVD- IGIK EYHEEID- AGIK YFHEENE- SNTR EWHEEND-	1) VLY PLSRCIF 2) GKY PLSRCVF 3) QRF ALSKVVF 3) SKY TOSIVTF 5) TKY TOSMVTF CTC TLLTVKF CYF TMLTIKK CYF TMLTIKK CTC ALLTVKF CTC ALLTVKF ILC ALASCKF	WAYIISDN WAYIISDS WSYIISDN WAYTFTDF CCYWMTNT CCFWMSEV CCFWISEV CAYWMSDA IAYCISKS	CCODGIEPIPCKHV LNANTNP CCLD SIEKEVERLUNANTAL CCLD SIEKEVERLUNANTAL MCRD AVPLIAGRHL SNADTIL MCRD AVPUPCKHI AQTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFELTRN-HL SEQADCF LSPI HFPLTEN-HL SEQADCF FLGV DYD ITES-HF VKSEASF LAGVY GVSL TAA-HL PKDCVSN
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23576040 >Csl_in_Ensembl >Cs2_in_Ensembl >Gg_118083700 >Xt_118404620	ESMLEHIL ENILKGIL ESLLSQLN VSVLDFLK KSCFRYLE CFCLHWLA QFCLNWLS VFCLDWLA SWCLERMA DWCLQWLA	SDHLSEIEIR NQCSMDE (1) TILN DDVKLEYQEL SDIRASN (1) TINN CDDDFRD (3) KIQ (NAQMAN (1) SK QNAQMAN (1) LSK NKGGYPNLER NKGGYPNHER NKAGMENHER	ICPLAELFFR- /CPLSELFYQ- /IPLGEFFHY- /FDIQYLFYV- /FDIQYLLFI- HELEHLARE- HELEFLVSE- /YGFESMASA- LITVEDLVIK- /QEVETLIK-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTFALTS ISNAMQTFPLASE YANAQRTPLALSE YQKKYHKEPSKTW YQDKLQEPSPRFT	AQQIITKDVYC (1) T AQAIIQKDVYE (1) T ADILLKKDAYE (1) T TDAHFERDCFE (1) Q TDAFFKKDFFE (1) Q ITAGCNSSMFD (1) E IEEOGFNSSMWE (1) E IESGFTSSMWE (1) E IVSRELDVULWD (1) S VSRLLDVULWD (1) S	KGIS DYHEEKD (KDIS CFHEDQG (SGIA DFHEDQG (NGIA OFHEDKD (AGIA OFHEDKD (LGIK QYHEDUD (PGIK RYHEDVD- IGIK SYHEDID- AGIK YFHEDND- SNTR CHEEND- SNTR CHEEND-	L) VLY PLSRCIP L) GKY PLSRCVP L) QRF ALSKVVF L) SKY TOSIVTF L) TKY TOSMVTF CTC TLLTVKF CYF TMLTIKF CYF TMLTIKF CYF TMLTIKF CYF ALLSKKF ILC ALASCKF	WAYIISDN WAYIISDS WSYIISDN WAYTFJDF CCYWMINT CCFWMSEV CAYWMSDA IAYCISKA IAYCISKA	CCODGIEPIPCKWSLANATNP CCLD_SIEMKPGRHLMNADTAL CCLD_SIEMKPGRHLMNADTAL CCLD_SIEMKPGRHLMNATTL MCRD_AVP_IAGRHLSNADTTL MCRD_AVPUPCKHISQTKPN FSGINVDITES-HLVKSQSSF LAPYGFLTRN-HLSESSPL LSPIHFPLTEN-HLSESSPL LSGVGVSTAA-HLPKDCVSN LASVGVTTPA-HLPKDCVSN LASVGTTPA-HLPKSRN
>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575040 >Cs1_in_Ensembl >Cs2_	ESMLEHIL ENILKGIL ESLLSQLN VSVLDFIK KSCFRYLE CFCLHWLA QFCLNWLS VFCLQWLA SFCLQWLA SWCLERMA DWCLQWLA NWCLKHMA	SDHLSETELR NQGSMDE (1) TILI DDVLEYQFL SDIRASN (1) TIN CDDDFRD (3) KIQ EK-AYNNHER QNAQMAN (1) LSK QNAGMAN (1) LSK SIAGVDSFLE NKAGMENHER KKLETRQE	ICPLAELFFR- //PLGEFFFH- //PDIQYLFYV- /FDIQYLFYV- HELEHLARE- HELEHLARE- /YGFESMASA- LIVEDLVVC- LIVEDLVVG-	KQNVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATEK (6) TFPSTKT KREVM (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKFPLSS YANAQRTKPLSS YQKRYHKEPSKTW YQQKLQEEPSRFT YQQKLQKEPSRFT	AQUITHOVYD(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFEKDCFE(1)Q ITDAFFKKDFE(1)T ITDAFFKKDFE(1)T IEDGFNSSMFE(1)E IEDGFNSSMFE(1)E IESGFNSSMFE(1)E IQSGCNSSMFD(1)S VSRLLDVVLWD(1)S	KGIS DYHEEKD (KDIS EFHEDG (SGIA DFHEKLG (NGIA OFHEKLG (NGIA OFHEKLG (LGIK EYHEELE - PGIK RYHEDVD - IGIK EYHEELD - AGIK YFHENE - SNTR EYHEEND - SNTR KHEEND -) VLY PLSRCIF) GKY PLSRCVF) QRF ALSKVVF) J SKY TQSIVTF) TKY TQSVTF -CTC TLLTVKF -CYF TMLTIKF -CYF TMLTIKF -ILC ALASCKF -MWC ALASCKF -ILF ALAVCKF	WAYIISDN WAYIISDS WSYIISDN WAYTFTDF CCYWMINT CCFWMSEV CCFWISEV CAYWMSDA IAYCISKS IAYCISKS IAYCISNS	CCDC GIEPIPCKHV LNANTNP CCLD SIEKPCRLINNADTAL CCLD SIELIAGRHL SNADTIL MCRD AVP IAGRHI ONTNLEA MCCD AIVOPCKHI AQTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFELTRN-HL SESDEPL LSPI HFP ITSN-HL SESDEPL LSPI HFP ITSN-HL SEQADCF FLGV DYD ITSS-HF VKSEASF LAGV GVS ITAS-HL NPERSRN LASV GVT ITA-HL NPERSRN LASV GVT ITA-HL NPERSRN LATL GFE TEA-HV CQVYRA
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575304 >Csl_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs1_18083700 >Xt_118404620 >Md_126306192 >Hs_20306906	ESMLEHDI ENTIKGI ESLISQLN VSVLDFIK KSCFRYLE CFCLMWLA QFCLMWLA SFCLQWLA SWCLERMA DWCLCKHA NWCLKHMA	SDHLSEIELR NQCSMDE (1) TILN DDVKLEYQFL SDIRASN (1) TIN CDDDFRD (3) KIQ (NAQMAN (1) LSK QNAQMAN (1) LSK QNAQMAN (1) LSK NKGGYPNHER NKAGENHER KKLETRQELEI KKASEIRQDQ	LCPLAELFFR- VCPLSELFYQ- VIPLGEFFH- VFDIQYLLFV- VFDIQYLLFV- VTGESITSA- HELEHLARE- VTGELFLVSE- LFUSE- LTVEDLVIX- LTVEDLVVG- LTVEDLVVG- LTVEDLVVG-	KONVEL (6) TFPSVYT KRATES (6) TFPSVHI KRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKFPLTS YANAQRTFLALSL YQKKHKEPSKTW YQQKLQKEPSKTW YQQKLHKEPSKTW	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T AQILLKKDAYE(1)T TDAHFKRDFEF(1)T TDAFFKRDFEF(1)T LTAGFKKDFEF(1)T LEDGFNSSMWE(1)E LESGFTSSMWE(1)E LQSGCNSSMFD(1)E VSRELLDVLWD(1)S VSRELLDVLWD(1)S VSRELLDVLMD(1)S	KGIS DYHEEKD (KDIS EFHEDQG (SGIA DFHEDKG (NGIA OFHEDKD (AGIA OFHEDKD (AGIA OFHEDKD (LGK EYHEDVD- IGIK EYHEDVD- IGIK EYHEDVD- AGIK YFHEDND- SNTR KWHEND- SNTR KWHEDND- SNTR KWHEDND-) VLY PLSCIF) GKY PLSCVF) QRF ALSKVVF) XKY TQSIVTF) XKY TQSIVTF -CTC TLLTVKF -CYF TMLTIKF -CYF TMLTIKF -CYF ALLTVKF -ILC ALSCKF -HWC ALSCKF -ILF ALAVCKF	MAYIISDN WAYIISDS WSYIISDN WGYMFSDY CCYWMNTT CCFWMSEV CCFWMSEV CAYWMSDA IAYCISKS IAYCISKS IAYCISNS IAYCISNS	CCODEGIEPIPCHWSLNANTNP CCLD_SIENKPGRHLMNADTAL CCLD_SIENKPGRHLMNADTAL MCRD_AVP_IAGRHLSNADTTL MCRD_AVPUPCHISONTRPAY FSGINVDITES-HLSVKSSSF LAPY GFE_TRN-HLSSSOPPL LSPIHFPLTEN-HLSSSOPPL LSPIHFPLTEN-HLSSSOPPL LAGY GVSLTAN-HLSKDCVSN LASV GVT_TPA-HLSNPCSSN LGTLGIP_TEA-HVSLOPYSS LATLGIP_TEA-HVSLOPYSS
>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575040 >Cs1_in_Ensembl >Cs2_	ESMLEHII ENTLKGI SSLISQIN VSVLDFLK KSCFRYLE CFCLHWLS VFCLDWLA SFCLQULA SWCLERVA DWCLQWLA NWCLKHAA NWCLKHAA	SDHLSEIELR NQGSMDE (1) TLLX DDVKLEYQFL SDIRASN (1) TLNN CDDDFRD (3) KIQ EK-AYNNHE QNAQMAN (1) LSK NKGGYPNLF S IAGVDSPLE NKAGMENHFR KKLETRQELG KASEIRQELG NNVGFHY (6) PNC	ICPLAELFFR- CPLSELFYQ- IPLCFFFH- YPLQYLFYV- YFDLQYLFI- YGFESITSA- HELEHLARE- YELFSWSA- ITVEDLVIC- LTVEDLVIC- LTVEDLVIC- LTVEDLVVC- LTVEDLVVC-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPSKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA JGYTEPVCLSFVL ICNAVNTKFALTS ISNAMQTKFPLSS YANAQRTFPLALSI YQKKYHKEPSKTW YQQKLHKEPSKTW YQQKLHKEPSKTW SDINNEKIDVET	AQQIITKDVY0(1)T AQAIIQKDYYE(1)T ADILLKKDAYE(1)T TDAHFKKDFFE(1)T ITDAFFKKDFFE(1)T ITDAFFKKDFFE(1)T ITDAFFKKDFWE(1)E IESGFTSSWWE(1)E IQSGCNSSMFD(1)E VSRELDVVLWD(1)S VSRLEDVVQVD(1)S VSRLEDVVQVD(1)S VRSLLDVSSWD(1)S	KGIS DYHEEKD(KDIS EFHELQG(SGIA DFHEKLG(NGIA OFHEKLG(NGIA OFHEND(LGIK EYHEEND(LGIK EYHEEDD) AGIK YHEETD- AGIK YFHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- TDFH AR SEPK-) VLY PLSRCIF) GKY PLSRCVF) QRF ALSKVVF) SKY TQSIVTF -CTC TLLTVKF -CTF TMLTIKK -CTF ALTVKF -ILC ALASCKF -ILC ALASCKF -ILF ALAVCKF -SNF ALSVCF	WAYIISDN WAYIISDS WGYMFSDY WGYMFSDY CCYWMINT CCYWMINT CCFWMSEV CAYWMSDA IAYCISKA IAYCISKA IAYCISNS TCCIVYHVIGS	CCODMGIEPIPCHWS LANATTNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL MCRD AVPLIAGRHL GNADTTL MCCD AITVQPCHI GNTKPNY FGINNDITES-HL VKSQSSF LAPYGFE TRN-HL EESDEPL LSPIHFP: TEN-HL SEQADCF FLGV DYDITES-HF VKSEASF LAGWGVS TAA-HL PKDCVSN LASVGVT TPA-HL PKDCVSN LGSVGTTPA-HL PKDCVSN LGTLGIPITEA-HV LQDYEAS LATL GIQITEA-HV LQDYEAS
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575304 >Csl_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs1_18083700 >Xt_118404620 >Md_126306192 >Hs_20306906	ESMLEHII ENTLKGI SSLISQIN VSVLDFLK KSCFRYLE CFCLHWLS VFCLDWLA SFCLQULA SWCLERVA DWCLQWLA NWCLKHAA NWCLKHAA	SDHLSEIELR NQGSMDE (1) TLLX DDVKLEYQFL SDIRASN (1) TLNN CDDDFRD (3) KIQ EK-AYNNHE QNAQMAN (1) LSK NKGGYPNLF S IAGVDSPLE NKAGMENHFR KKLETRQELG KASEIRQELG NNVGFHY (6) PNC	ICPLAELFFR- CPLSELFYQ- IPLCFFFH- YPLQYLFYV- YFDLQYLFI- YGFESITSA- HELEHLARE- YELFSWSA- ITVEDLVIC- LTVEDLVIC- LTVEDLVIC- LTVEDLVVC- LTVEDLVVC-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPSKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA JGYTEPVCLSFVL ICNAVNTKFALTS ISNAMQTKFPLSS YANAQRTFPLALSI YQKKYHKEPSKTW YQQKLHKEPSKTW YQQKLHKEPSKTW SDINNEKIDVET	AQQIITKDVY0(1)T AQAIIQKDYYE(1)T ADILLKKDAYE(1)T TDAHFKKDFFE(1)T ITDAFFKKDFFE(1)T ITDAFFKKDFFE(1)T ITDAFFKKDFWE(1)E IESGFTSSWWE(1)E IQSGCNSSMFD(1)E VSRELDVVLWD(1)S VSRLEDVVQVD(1)S VSRLEDVVQVD(1)S VRSLLDVSSWD(1)S	KGIS DYHEEKD(KDIS EFHELQG(SGIA DFHEKLG(NGIA OFHEKLG(NGIA OFHEND(LGIK EYHEEND(LGIK EYHEEDD) AGIK YHEETD- AGIK YFHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- SNTR WHEEND- TDFH AR SEPK-) VLY PLSRCIF) GKY PLSRCVF) QRF ALSKVVF) SKY TQSIVTF -CTC TLLTVKF -CTF TMLTIKK -CTF ALTVKF -ILC ALASCKF -ILC ALASCKF -ILF ALAVCKF -SNF ALSVCF	WAYIISDN WAYIISDS WGYMFSDY WGYMFSDY CCYWMINT CCFWMSEV CCFWMSEV CAYWMSDA IAYCISKA IAYCISKA IAYCISNS TCCIVYHVIGS	CCODEGIEPIPCHWSLNANTNP CCLD_SIENKPGRHLMNADTAL CCLD_SIENKPGRHLMNADTAL CCLD_SIENKPGRHIMNATAL MCRD_AIVPLIAGRHIMNATAL MCCD_AIVVDCRHIAQTKPNY FSGINVDITES-HLWKSQSSF LAPY GFE_TRN-HLWSQSSF LAPY GFE_TRN-HLWSQSF LGVVDYDITES-HFWKSEASF LAGV GVSLTAA-HLWSCASSF LAGV GVT_TRA-HLWSEASF LGVVGVT_TPA-HLWPERSRN LGTLGIPLTEA-HVSLQDYEAS LATLGIPLTEA-HVSLQDYEAS
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_1_n_Ensemb1 >Cs2_in_Ensemb1 >Gg_118083700 >Xt_118404620 >Md_126306192 >Hs_20306906 >Ce_3873737	ESMLEHI ESLLSQLN USVLDFTK KSCFRTK CFCLHWLA QFCLNWLS SFCLQWLA SWCLERWA DWCLQWLA NWCLKHMA NWCLKHMA DWCLKHMA ES-MKHLA	SDHLSETELR NQGSMDE(1) TILL DDVKLEYQFI SDIRASN(1) TIM SDIRASN(1) TIM (CDDDFR0(3) KIQ EK-AYNNHEC (NAQMAN(1) LSK (NAGMAN(1) LSK (NAGMAN(1) LSK SIAGVDSPLE NKAGMENELE KASEIRQDLQI NNVGFHY(6) QNN KTUGFNY(6) QNN	CPLAELFFR- VCPLSELFYQ- JPLCEFFH- VPLQVLFYV- JFDLQVLFYV- VFDLQVLFYV- JFDLQVLFV FELEFLNER- LFELFLNER- LFVEDLVK- LTVEDLVVG- VLFVEDLVVG- VLFAEVEA-	KONVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATES (6) TFPSVHI KESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ISNAMQTKIPLSS YANAQRTFLALSL YANAQRTEPLSS YQKLYKEPSKTW YQQKLKEPSKTW YQQKFLKEPSKTW SDIMEKUDVET SDIMEKUDVET	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T ADILLKKDAYE(1)T TDAHFKRDCFE(1)Q [IDAFFKKDFFE(1)T [IDAFFKKDFFE(1)T [IDAFFKKDFFE(1)T [IDAFFKKDFFE(1)T [IDAFFKLDFV(1)T VSRELDVLMP(1)S VSRELDVLMP(1)S VSRELDVLMP(1)S ITRSLLDVAMD(1)S MRWFSLLCQK(12)G	KGIS DY HERKO (KGIS DY HERKG (SGIA DF HERKG (NGIA OF HERKG (NGIA OF HERKD (LGIK CY HERDO (LGIK CY HERDO (LGIK CY HERDO (GIK CY HERDO (SNTR VF) ULY PLSCIF) GKY PLSCVF U) GKY ALSVVF U) GKY ALSVVF U) GKY TOSVVF I) TKY TOSVVF U) TKY TOSVVF U) TKY TOSVVF U) TKY TOSVVF U) TASSA U) TA	WAYIISDN WAYIISDS WSYIISDN WGYMFSDY CCYWMINT CCFWISEV CAYWMSDA IAYCISKA IAYCISKA IAYCISKS ICCIYHVIGS TCCIYHVIGS	CCODMGIEPIPCHWS LANATINP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL MCRD AVPLIAGRHL GNADTTL MCCD AITVQPCHI GNTKENY FGINNDITES-HL VKSQSSF LAPYGFE TRN-HL EESDEPL LSPIHFPJTTEN-HL SEQADCF FLGV DYDITES-HF VKSEASF LAGWGVS TAA-HL PKDCVSN LASVGVT TFA-HL PKDCVSN LGSUGIPTEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_2357504 >Ci_23576040 >Csl_in_Ensembl >Cs2_in_Ensembl >Gg_118083700 >Xt_118404620 >Md_126306192 >Hs_20306906 >Ce_3873737 >Cb_39591559 >Eh_67477376_C	ESMLEHI ENLIKGI ESLISOIN VSVLDELK KSCFRILE CFCLHWLA QFCINWLS SFCLOWLA SFCLOWLA SWCLFRMA DWCLKHMA NWCLKHMA DS-MKHLA ES-MKHLA VLCFNNLF	SDHLSEIELR NQGSMDE (1) T LLX DDVKLEYQFLA SDIRASN (1) T LM CODDFRE (2) KIQ EK-AYNNHE QNAQMAN (1) LSK NKGGYPNLER NKAGMENHER NKAGMENFLE NKAGMENFLE NKAGMENFLE NKAGFLRQFLO NUVGFHY (6) PNC KTVGFNY (6) QNN GYIQQSI (2) DY II	CPLAELFER CPLSELFYQ- IPLCEFFH- YPIQYLFYV- FPIQYLFYV- FPIQYLFIX- HELEHLARE- FELEFLVSE- UVGESMASA- LIVEDLVK- QEVETLIIK- LIVEDLVK- LIVEDLVK- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- THHFLLEY-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKTPLSS YANAQRTFLALSL YQKKYHKEPSKTW YQQKLQEEPSRPT YQQKLQEEPSRPT YQQKLHKEPSKTW SDIMMEKIDVET SDIMGEKVDMET ATMKGF(1) KAKECGI	AQQIITKDVYD(1)T AQAIIQKDYYE(1)T AQILLKKDAYE(1)T TDAHFKRDFEF(1)T TDAFFKRDFEF(1)T ITDAFFKRDFEF(1)T IEDGFTSSMWE(1)E IEDGGTSSMWE(1)E IEDGGTSSMWE(1)E USRELDVVLWD(1)S VSRLLDVVLWD(1)S VSRLLDVSMWD(1)S IRSLLDAKMD(1)S MRWFSLLGGK(12)G MRWFSLLGK(12)G	KGIS DYHEEKD(KDIS EF EDQG(SGIA DF EKLG(NGIA OF EKKD(LGIK EY EEND FGIK AY EEND AGIK YF EEND AGIK YF EEND SNTR KHEEND SNTR KHEEND SNTR KHEEND SNTR KHEEND TDFH AR SEPK -CCK TY QLLA() ULY PLEACH) GKY PLEACVF) GKY ALEXVIP () GKY ALEXVIP () GKY ALEXVIP () GKY ALEXVIP () GKY ()	WAYIISDN WAYIISDS WGYMFSDY CCYWMINT CCFWMSEV CCFWMSEV CAYWMSDA IAYCISKA IAYCISKA IAYCISKS ICCIIYHVIGS CCIIYHVIGS CCIIYHVIGS	CCODEGIEPIPCKWS LNANTNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHI MNADTAL MCRD AVE IAGRHL MNTLEA MCCD AIVOPCKHI AQTKPNY FSGI NVD TES-HL VKSQSSF LAPY GFE INN-HL SESOPPL LSPI HFF: INN-HL SEQADCF FLGV DYD TES-HF VKSEASF LAGV GVT TRA-HL PRECVSN LASV GVT TRA-HL PRECVSN LASV GVT TRA-HL NPERSN LGTL GIP TEA-HV LQDYEAS LATL GIP. TEA-HV LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQNSSSNSM
<pre>>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs1_in_Ensemb1 >Cs2_in_Ensemb1 >Cs2_in_Ensemb1 >Cs2_in& >Md_126306192 >Hs_2036906 >Cc_387373 >Cb_39591559 >Eh_674778664</pre>	ESMLEHT ESLISOLN VSVLDFLK KSCFRYLE CFCLHMLA OFCLMML3 SFCLOMLA SFCLOMLA SFCLOMLA NWCLENNA NWCLKHA NWCLKHA DS-MKHLA ES-MKHLA VLCFNNLE	SDHLSETELR MQGSMEC 1]TLL DDVKLEYQELA SDIRASN(1)TLN SDIRASN(1)TLN (CDDDFRE(3)KIQ CODDFRE(3)KIQ (NAQMAN(1)LSK QNAQMAN(1)LSK QNAQMAN(1)LSK NKGGYPNELE NKAGYPNELE KASEIRQELE KASEIRQELE KASEIRQELE KASEIRQELE KASEIRQ	CPLAELFFR- VCPLSELFYQ- VCPLSELFYQ- VFDLQYLFYV- VFDLQYLFYV- VFDLQYLFFN- VFDLFHLARE- VFELFLVSE- VELFLVSE- VGEVETLIK- LTVEDLVVG- LTVEDLVVG- LTVEDLVVG- VIVEAFVEA- VIVEAFVEA- FTTHFLLFY- ITHQLFDY-	KONVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATEK (6) TFPSTKT MKESTCE (5) KFKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSFVL ICNAVNTKFALTS ISNAMQTKFPLSS YANAQRTFLALSI YQCKLYKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW SDIMKEKIDVET SDIMKEKIDVET SDIMKEKIDVET ATMKGF (1) KAKECGI MCSIGHI (1) EKSSFI	AQUITIKDVYD(1)T AQNIIGKVYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T TTAGCNSSMFD(1)E LESGENSSMFD(1)E LESGENSSMF(1)E LESGENSSMF(1)E USSELDVVLND(1)S VSRELDVVLND(1)S VSRELDVVLND(1)S USSELLDVAMD(1)S HRWFSLLGQK(12)G MRWFSLLGQK(12)G IMNDYKQEXI-	KGIS DY HERKD (KGIS DY HERKG (SGIA DF HERKG (SGIA OF HERKG (AGIA CY HERKD (LGIK EY HER DOVD - LGIK EY HERD - SNTR WHERD - SNTR WHERD - SNTR WHERD - SNTR WHERD - SNTR WHERD - TDFH AR SEPK - TDFH AR SEPK - - DFK TY QLLA (- ER EY KKIN () ULY PLSECT) GKY PLSECT) GKY PLSECT) SKY TGSTVTH GKY TGSTVTH -CTC TLLTKK -CYF TLLTKK -CYF TLLTKK -CYF TLLTKK -CYF TLLASCK -TLC ALSCK -TLC ALSCK -TLF ALSCK -SNF ASVIVG -SNF ASVIC -SNF ASVIC -S	MAYIISDN WAYIISDN WSVIISDN WGYMFSDY CCYWMINT CCFWMSEV CAYWMSDA IAYCISKS IAYCISKS IAYCISNS IAYCISNS ICCIYYNVIGS TALA-ILI SLM-ELI	CCODEGIEPIPCHW9 LNANTNP CCLD SIE KPCRLI NNADTAL CCLD SIE KPCRLI NNADTAL MCRD AVP IAGRHI CNTNLEA MCCD AIVOPCKHI AGTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFE TRN-HL SESDEPL LSPI HFP IEN-HL SEQADCF FLGV DVD IES-HF VKSQSSF LAGV GVI TRA-HL PKDCVSN LASV GVI TRA-HL PKDCVSN LASV GVI TRA-HL PKDCSN LATL GIQ TEA-HV LQDYEAS FFRK HLK IFTAHQ SSSNSM FFRK HLK IFTAHQ SSSNSM FFRK HLK IFTAHQNSSSNSM LKD GYE IKNNDKF FCVVHKN
>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575304 >Cs_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs2_in_Ensembl >Cs_108083700 >Xt_118404620 >Md_126306192 >Hs_20306906 >Ce_3873737 >Cb_39591559 >Eh_67477376_C >Eh_6746664 >Eh_6746664	ESMLEHTI ENTIKGII ESLISOIN VSVLDFIK KSCFRYLE CFCLHWIA QFCLNWIA SVCLERVA DWCLOWIA NWCLKHNA DWCLOWIA DCS-MKHLA VLCFNNLF KKCIDELF	SDHLSEIELR NQGSMDE (1) TILL DDVKLEYQFL SDIRASN (1) TIM CODDFRE (2) KIQ (CDDFFR (2) KIQ (CAQMAN (1) LSK QNAGMAN (1) LSK QNAGMAN (1) LSK SIAGVDS	CPLAELFFR- VPLSELFYO- VIPLSELFYO- VIPLSELFYO- VPDLYLLFI- VPSLYLLFI- VPSLYLLFI- VPSLYLVSE- VPSLYVSE- VPSLYVSE- VPSLYVSE- VPSLYVSE- VSLYLVSE- VSLYLVSE- VIVLSEVEA- VIVL	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKTV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKFALTS YANAQRTFLALSL YQVKYHKEPSKTW YQQKLKEPSKTW YQQKLHKEPSKTW SDIMEKUPVET SDIMECKUPVET ATMKGF (1) KAKECGI WCSIQHI (1) EKSSTI MRFKEIJPNTM	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T AQILQKDVYE(1)T TDAHFKRDFE(1)T TDAFFKRDFE(1)T ITDAFFKRDFE(1)T IEDGFNSSMWE(1)E IEDGFNSSMWE(1)E IESGFTSSMWE(1)E IESGFTSSMWE(1)E ISGLIDVUWD(1)S VSRLLDVUWD(1)S VSRLLDVUWD(1)S ISRLLDVSMWD(1)S MRWF5LLGQK(12)G IMRWF5LLGQK(12)G IMRWF5LLGQK(12)G IMRWFSLLGV(12)G IMRWFVYQLKI NHIFKQLCA IYNFY-KP(C3)N	KGIS DYHEEKD(KDIS EFEDQG(SGIA DFHEKLG(NGIA OFHEKLG(NGIA OFHEKDG(LGIK CYHEEDDO LGIK CYHEEDDO LGIK CYHEEDDO LGIK CYHEEDDO SNTR CHEENDO SNTR CHEENDO SNTR CHEENDO SNTR CHEENDO SNTR CHEENDO TDFH AR SEPK- -CDK TYHOLLA(ER CYHKIN() ULY PLSCIF) GKY PLSCIF) GKY DLSCVF) GRF ALSKVVF U SKY TO SIVIF -CTC TILIVK: -CYF THLTIK: -CYF THLTIK: -CYF THLTIK: -CYF THLTIK: -CYF ALLIVK: -TIC ALSCK: -HIC ALSCK: -HIC ALSCK: -SYF ALVIVG -SYF ASVIVG -SYF ASVIVG -SYF ASVIVG -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF ALSCK: -SYF ASVIVG -SYF	WAYIISDN WAYIISDN WSYNISDN WSYNFSDY CCYWMSEV CCFWMSEV CCFWMSEV CAYWMSDA IAYCISKS IAYCISKS IAYCISKS IAYCISKS ICCIIYNVISS IALA-ILI TSIM-ELI TSIM-ELI LASTUCU	CCODGIEPIPCKHV LNANTNP CCLD SIE KPGRHL MNADTAL CCLD SIE KPGRHL MNADTAL CCLD SIE KPGRHL MNADTAL MCRD AVP IAGRHL NATTL FSGINVDITES-HL VKSQSSF LAPY GFELTRN-HL EESDEPL LSPI HFPLTEN-HL SEQADCF FLGV DYDITES-HF VKSEASF LACV GVI TAA-HL PKCOVSN LASV GVI TPA-HL NPERSRN LGTL GIP TFA-HV LQDYEAS LATL GID TFA-HV LQDYEAS LATL GID TFA-HV LQDYEAS FFRR HLK IPTAHOSSSNSM ILKD GYE KNNDKF FCVVHKN CMKS GAT ISSDI SVKFVKS YKIT KSKF VAS-M
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_1_a576040 >Cs_1_n_Ensembl >Gg_118083700 >Xt_118404620 >Md_126306192 >Hs_20306906 >Ce_3873737 >Cb_39591559 >Eh_67476664 >Eh_67486665 >Eh_67484628	ESMLEHTI ENTIKGII ESLISQIN VSVLDFIK KSCFRYLE CFCLHWLA QFCLMWLA SFCLQWLA SFCLQWLA DWCLQWLA DWCLQWLA DWCLKHDA DG-MKHLA DS-MKHLA DS-MKHLA KKCIDEIF KKCIDEIF VQCVELA	SDHLSEIELR NQGSMDE (1) T LLX DDVKLEYQELA SDIRASN (1) T LNX CDDDFRC 13 K 1Q EK-AYNNHEN QNAQMAN (1) LSK NKGGYENLER NKGGYENELE SIAGVDSFLEI NKAGMENHER KKLETRQELE KASEIRQDLQ NVGFHY (6) PNC KTVGFNY (6) PNC KTVGFNY (6) PNC SQAKVSD (5) FNT TKAGLYD (6) FNH	CPLAELFFR- CPLSELFYQ- JPLCEFFFH- YPLQYLFY- YGESITSA- HELEHLARE- FELFLVSE- YGESMASA- LIVEDLVYG- QEVETLIK- LIVEDLVYG- LIVEDLVVG- VIVEAFVEA- VIVEAFVEA- TIHIFLEY- FIHIQLFDY- FIHIQLFDY- HFELFVKF-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKIV MKESTCE (5) KPKSFHI KKEVUN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTFALTS ISNAMQTFPLASE YANAQRTFLALSE YQQKLQEEPSRFW YQQKLHKEPSKTW YQQKHKEPSKTW YQQKHKEPSKTW SDIMNEKIDVET SDIMEKIDVET SDIMEKIDVET MATMKGF(1) KAKECGI WSIQHI(1) EKSSFIL NRFKEIIPNTNA FYESQGKVVPISA	AQQIITKDVY0(1)T AQAIIQKDYE(1)T AQAIIQKDYE(1)T TDAHFKRDFE(1)T TDAFFKRDFEF(1)T TDAFFKRDFEF(1)T IEDGFTSSMWE(1)E IEDGGTSSWWE(1)E USSELDVVLWD(1)S VSSLLDVVLWD(1)S VSSLLDVVLWD(1)S MRWFSLLGQK(12)G MRWFSLGQK(12)G MRWFSLGQX(12)G	KGIS DYHEEKD(KDIS EFHEDQG(SGIA DFHEKLG(NGIA OFHEKKG(NGIA OFHEKKG(LGIK EYHEEND LGIK EYHEEND SITR KYHEEND SNTR KYHEEND SNTR KYHEEND SNTR KYHEEND SNTR KYHEEND SNTR KYHEEND SNTR KYHEEND SNTR KYHEEND CHAR SEPK TDFH SR SEPK TDFH SR SEPK CK TYHOLLA(ER EYHKIN ()) ULY PLSCIF) GKY PLSCVFW) GRY ALSXVFW D GRY ALSXVFW -CTC TLLTVKF -CTC TLLTVKF -CTC TLLTVKF -CTC ALLIVKF -CTC ALLIVKF -CTC ALLIVKF -ILC ALASCKF -ILF ALASCKF -ILF ALAVCKF -SNF ASVIGE -SNF ASVIE -SNF ASVIE -SNF ASVIE -SNF	NAYIISDN WAYIISDN WGYNISDN CCYWMINT CCFWMEV CCFWMEV CAYMMSDA IAYCISKS IAYCISKS IAYCISKS IAYCISNS TCCIYHVIGS TCCIYHVIGS TCCIYHVIGS TCLIYHVIGS TLA-ILI ISLM-ELI LASTLCL VG	CCODGIEPIPCKHV LANATNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL MCRD AVP_IAGRHL GNATTL MCRD AVPUPCKHI SQTKPNY PSGI NVD ITES-HL VKSQSSF LAPY GFE ITN-HL SESDEPL LSPI HFPL ITN-HL SEQADCF FLGV DYD ITES-HE VKSEASF LAGV GVS ITAA-HL PKDCVSN LASV GVI TPA-HL PKDCVSN LASV GVI TPA-HL PKDCVSN LGTL GIP ITEA-HV LQDYEAS LATL GID ITEA-HV LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQSSSNSM LLKD GYE KNNDKF FCVVHKN GMKS GAT IGSDTL SVKFVKS KITI SKI VAS SQVGSNV
<pre>>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >ci_23575304 >ci_23575304 >cs2_in_Ensemb1 >cs2_in_Ens</pre>	ESMLEHTI ENTIKGI ENTIKGI SULSON VSVLDFIK KSCERYIE OFCLIMWLS VFCLMWLS SWCLENIA SWCLENIA DWCLKHIA NWCLKHIA NWCLKHIA NWCLKHIA DS-MKHLA VLCFNNLE KKCIDEF VQCVEFLA VQCVEFLA VCCTENLA	SDHLSEIELR NQGSMDE(1) TILL NDGVKLEYQFIN SDIRASN(1) TIMN (CDDDFR0(3) KIQ (CDDFR0(3) KIQ (CDDFR0(3) KIQ (CNAQMAN(1) LCK (NKGGYPNLFN (NKGGYQNLFN (C) KKLETRQFLE KASEIRQFLE KASEIRQFLE KASEIRQFLQ NNVGFHY(6) QNN GYIQQSI(2) DY SQAKVSD(5) FNT TKAGLUD(6) FNT TKAGLUD(6) FNT	CPLARLFFR- VPLSEFY- VIPLSEFY- VIPLSVLFT- VPISVLFT- VPISVLFT- VPISVLFT- VPISTSMASA- HELEHARE- FELFIVSE- VFESMASA- LIVEDLVIK- VIVEAFVEA- LIVEDLVVG- LIVEDLVVG- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VITHIGLEDV- HESVDFVEC- LIPEAFVEF	KONVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATEK (6) TFPSVHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ISNAWQTKPLSS ISNAWQTKPLSS YANAQRTFLALSL YQKKJHKEPSKTW YQCKLQKEPSRPT YQQKLLKEPSKTW SDIMEKUDVET SDIMGEKUDVET SDIMGEKUDVET ATMKGF (1) KAKECGI WCSIQHI(1) EKSSFIL NRFKEIVPNTA YESGGKVVPISA C-SFYN(2) INNFSS	AQQIITKDVYD(1)T AQAIIQKDVYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T TDAFFKKDFFE(1)T ITAGCNSSMP(1)E IEDGFNSSMWE(1)E IEDGFNSSMWE(1)E IEDGSNSSMP(1)E VSRELDVLMD(1)S VSRELDVLMD(1)S VSRELDVLMD(1)S VSRELDVLMD(1)S IRSLLDVAMD(1)S IRSLLDVAMD(1)S IRSLLDVAMD(1)S IRSLLCQK(12)G IMNDYKD[KI- MFKF5LLCGC- TYNFY-KPLVC(3)N IFKCL-CPLLS(3)E	KGIS DY HERKD(KDIS EF EDQG(SGIA DF HERLG(NGIA CF EDKD(LGIK CF EDND(LGIK CF EDND(LGIK CF EEND- GGIK CF EEND- GGIK CF EEND- SNTR KF EEND- SNTR KF EEND- SNTR KF EEND- SNTR KF EEND- TDFH AR SEPK- -CDK TF QLLA(NDFK DF HKSNG(TYYK) F SNIK (NDFK DF HKSNG() ULY PLSCH) GKY PLSCH) GRF ALSKVVF USRF ALSKVVF USRF ALSKVVF SKV	NAY115DN WAY115DN WSYNF5DY CCWMF5DY CCWMF5V CCWMF5V CCWMF5V CCWMF5V CCWMF5V CAYWM5DA IAYC15KA IAYC15KA IAYC15KA IAYC15KA	CCODEGIEPIPCKHV LANATTNP CCLD SIE KFORHL WNADTAL CCLD SIE KFORHL WNADTAL MCRD AVP IAGRHI CNTNLEA MCCD AIVOPCKHI AQTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFE ITN-HL SEQADCF LGV DYD ITES-HF VKSEASF LACV GVI TRA-HL PKCVSN LASV GVI TRA-HL PKCVSN LASV GVI TRA-HL PKCVSN LASV GVI TRA-HL PLCVSN LASV GVI TRA-HL VLQDYEAS LATL GIQ.TEA-HV LQDYEAS FFRR HLK IFTAHQSSSNSM LKD GYE KNNDKF FCVVHKN GVKS GATI IGSDTI SVKFVKS YKTI SKI VASM SQVQSMV ELTCNF-N VNQFNIL
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_1ar_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_2in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Eh_67477376_N >Ed_167389979_C	ESMLEHTI ENTIKGII ESLISQIN VSVIJEFIK KSCFRYLE CFCLHWIA SFCLOWIA SWCLERVA DWCLQWIA NWCLKHAA DWCLQWIA DWCLQWIA DS-MKHLA ES-MKHLA VLCFNNLF KKCIDELF VQCVEFIA VQCIENLA VLCFNNLF	SDHLSEIELR NQGSMDE (1) T LL DDVKLEYQFL SDIRASN (1) T LM SDIRASN (1) T LM CODDFRE (2) KIQ EK-AYNNHE NKAGYNHE NKAGYPNLER NKAGYPNLER NKAGYPNFLE NKAGYPNFLE NKAGYN(6) PNC KTVGFNY (6) PNC KTVGFNY (6) PNC KTVGFNY (6) PNC SQAVXD (5) FNT TKAGLYD (6) FMH QLANECF (1) NFK	CPLAELFFR- VPLSELFYO- VPLSELFYO- VPLSELFYO VPLSULFT- VSESITSA- HELEHLARE- VFELFUSE- VSESMASA- LTVEDLVIG- LTVEDLVIG- UTVEDLVIG- VIVEAFVEAFVE- VIESEVI- VIESEVI- VIESEVI- VIESEVI-	KONVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATEK (6) TFPTKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKTPLSS YANAQRTFLALSL YQUKLQEEPSRFT YQQKLQEEPSRFT YQQKLQKEPSKTW YQQKLKEPSKTW SDIMEKIDVET ISDIMGEKVDVET ATMKGF(1) KAKECGI WCSIQHI(1) EKSSFIL MSFKEIIPNTMA FYESQGKVVPISA C-SFYN(2) IENPFSL ATMKGF(3) KAKEYGI	AQQIITKDVYD(1)T AQAIIQKDYYE(1)T AQAIIQKDYE(1)T TDAHFKRDFEF(1)T TDAFFKRDFEF(1)T ITDAFFKRDFEF(1)T IEDGFNSSMFE(1)E IEDGFNSSMFE(1)E IEDGFNSSMFE(1)E IEDGFNSSMFE(1)E IESGFTSSMFE(1)E IESGFTSSMFE(1)E VSRELLDVLMD(1)S VSRELLDVLMD(1)S VSRELLDVLMD(1)S MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G IMNDYKQC[KI INHFYKQLECA INHFYKQLECA IMNFYKCL-CPLLS(3)E IMNOYKQLKI	KGIS DY HEEKD (KDIS EF EDQG (SGIA DF EKLG (NGIA OF EKLG (NGIA OF EKLG (LGIK EY EELD- PGIK RY EELD- AGIK YF EELD- AGIK YF EEND- SNTR CM EEND- SNTR CM EEND- SNTR CM EEND- SNTR CM EEND- CDK TY QLLA (ER CY KKIN (NDFK DF KSNG (TKYK SF SNIK (EK CI KEIM SNIK () ULY PLSCIF) GKY PLSTOFF) GRY PLSTOFF) SRY TO STUTT -CTC TLLTWE -CTC TLLTWE -CYT TMLTIKK -CYT TMLTIKK -CTC ALLTWE -CYT TMLTIKK -CTC ALLTWE -TLC ALSSCK -TLF ALAYCKK -SRF ALAYCK -SRF ALAYCK -SRF ALAYCK -SRF ALAYCK -SRF ALAYCK -SRF ALYCG -SRF ALYCG -SRF ALYCG -SRF ALYCG -SRF ALYCG -SRF ALYCG -SRF ALYCK -SRF ALCONK -SRF ALCON	WAY115DN WAY15DS WGYMF5DY CCYWMINT CCFW15EV CAYWM5DV IAYC15KS IAYC15KS IAYC15KS IAYC15KS IAYC15KS ICC11YHV16S TCC11YHV16S TCC11YHV16S TALA-1L1 ISSMF5L LYG SABS[FV TALA-1L	CCODEGIEPIPCKHV LANATNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHI MNADTAL MCRD AVPLIAGRHI MNTLA MCCD AIVUPCKHI AQTKPNY FSGINVDITES-HL VKSQSSF LAPY GFELTRN-HL EESDEPL LSPIHFPLTEN-HL SEQADCF FLGVUPDITES-HF VKSEASF LAGV GVT TRA-HL NPERSRN LASV GVT TRA-HL NPERSRN LGTL GIPLTEA-HV LQDYEAS LATL GIPLTEA-HV LQDYEAS LATL GIPLTEA-HV LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSNSVQ MKS GATIGSDTL SVKFVKS WTI SKI VASM SQVQSMV
<pre>>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_i_a576040 >Cs_i_n_Ensemb1 >Cs2_in_Ensemb1 >Cs2_in_Ensemb1 >ds_126306192 >Hs_20306906 >Ce_387373 >Cb_39591559 >Eh_67477376_C >Eh_67476664 >Eh_67466455 >Eh_67464628 >Eh_67477376_N >Ed_167384018</pre>	ESMLEHT ENTIKGI ESLISOLN VSVLDFIK KSCFRYLE CFCLHWLS VFCLHWLS SWCLENN SWCLENN SWCLENN DWCLWLA NWCLKHA NWCLKHA DS-KKLA ES-KKLA VLCFNNLE YECFVKLE VCCENNLS KKCIDELF	SDHLSETELR MQGSMEC 1]TLL DDVKLEYQELA SDIRASN(1)TLN SDIRASN(1)TLN (CDDDFR0[3]KIQ EK-AYNNHEE (NAGMAN(1)LCK NKGGYPNLFR (NKLETRQLGR KASEIRQ	CPLAELFFR (PLSEFY) (PLSEFY) (PLSEFY) (PLSEFY) (PLSEF) (PLSES)	KQNVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATES (6) TFPSVIT KREXTE (5) TFPSKIT KEXEVM (6) EKINKFA IGYTEPVCLSFVL ICNAVWTKFALTS ISNAMQTKFPLSS ISNAMQTKFPLSS ISNAMQTKFPLSS YANAQRFEPSKTA YQCKLHKEPSKTA YQCKLHKEPSKTA SDIMGEKIDVET SDIMGEKIDVET SDIMGEKVPUES ATMGG (1) KAKECGI NGSIQHI(1) EKSSFIL NFFKEIIPNTNA FYESQCKVPISA ATMGG (3) KAKEYGI ATMGG (3) KAKEYGI	AQUITREDVYD(1)T AQUAIIGKUVYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T TTAGREKDFFE(1)T TITAGCNSSMFD(1)E TEGGENSSMFD(1)E TEGGENSSMFD(1)E USSELDVVIND(1)S VSRLLDVVIND(1)S VSRLLDVVMD(1)S VSRLLDVAMD(1)S MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQC MANDYKQLKI MRFKQLGCA	KGIS DY HERDO KIJS ZHEDOG (SGIA DFHEKLG (NGIA OFHENDG (AGIA OFHENDG (LGIK YHEEDOD IGIK YHEEDOD IGIK YHEEDOD IGIK YHEEDOD SNTR WHEEND SNTR WHEEND SNTR WHEEND SNTR WHEEND SNTR WHEEND TDHHAR SENN TDHAR SENN TDH AR SEPK TDH AR SEPK -CDK TY QLLA (NDFK DFHKIN (NDFK DFHKIN (SNTK YHEEND -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHKIN (SNTR YHEAN)) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY TO STUTT GKY TO STUTT GKY TO STUTT - CTC THLTWK - CYF THLTIKK - SNF AAVIVG - SNF AAUNG - SNF AAUN	NAYIISDN NAYIISDN NSYIISDN NGYMFSDY CCYWNSDY CCYWNSDA IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISS TGCIYHVIGS TGCIYHVIGS TGCIYHVIGS TGLA SANSIFU SANSIFV TALA-TLL SLM-ELI	CCDC GIEPIPCKHV LNANTNP CCLD SIE KPCRLI WANTAL CCLD SIELIAGRHL SNADTIL MCCD AVP IAGRHI (NTNLEA MCCD AIVOPCKHI ACTR FSGI NVDITES-HL VKSQSSF LAPY GFE ITN-HL SEODEPL LSPI HFP TEN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LGV DYD TES-HF VKSEASF LAGV GVI TEA-HL NPERSRN LGTL GIO TEA-HU LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ YKTI SKI VASSQVQSMV
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_1ar_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_2in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Eh_67477376_N >Ed_167389979_C	ESMLEHT ENTIKGI ESLISOLN VSVLDFIK KSCFRYLE CFCLHWLS VFCLHWLS SWCLENN SWCLENN SWCLENN DWCLWLA NWCLKHA NWCLKHA DS-KKLA ES-KKLA VLCFNNLE YECFVKLE VCCENNLS KKCIDELF	SDHLSETELR MQGSMEC 1]TLL DDVKLEYQELA SDIRASN(1)TLN SDIRASN(1)TLN (CDDDFR0[3]KIQ EK-AYNNHEE (NAGMAN(1)LCK NKGGYPNLFR (NKLETRQLGR KASEIRQ	CPLAELFFR (PLSEFY) (PLSEFY) (PLSEFY) (PLSEFY) (PLSEF) (PLSES)	KQNVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATES (6) TFPSVIT KREXTE (5) TFPSKIT KEXEVM (6) EKINKFA IGYTEPVCLSFVL ICNAVWTKFALTS ISNAMQTKFPLSS ISNAMQTKFPLSS ISNAMQTKFPLSS YANAQRFEPSKTA YQCKLHKEPSKTA YQCKLHKEPSKTA SDIMGEKIDVET SDIMGEKIDVET SDIMGEKVPUES ATMGG (1) KAKECGI NGSIQHI(1) EKSSFIL NFFKEIIPNTNA FYESQCKVPISA ATMGG (3) KAKEYGI ATMGG (3) KAKEYGI	AQUITREDVYD(1)T AQUAIIGKUVYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T TTAGREKDFFE(1)T TITAGCNSSMFD(1)E TEGGENSSMFD(1)E TEGGENSSMFD(1)E USSELDVVIND(1)S VSRLLDVVIND(1)S VSRLLDVVMD(1)S VSRLLDVAMD(1)S MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQK(12)G MRWFSLGQC MANDYKQLKI MRFKQLGCA	KGIS DY HERDO KIJS ZHEDOG (SGIA DFHEKLG (NGIA OFHENDG (AGIA OFHENDG (LGIK YHEEDOD IGIK YHEEDOD IGIK YHEEDOD IGIK YHEEDOD SNTR WHEEND SNTR WHEEND SNTR WHEEND SNTR WHEEND SNTR WHEEND TDHHAR SENN TDHAR SENN TDH AR SEPK TDH AR SEPK -CDK TY QLLA (NDFK DFHKIN (NDFK DFHKIN (SNTK YHEEND -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHEEND) -CH YHKIN (SNTR YHKIN (SNTR YHEAN)) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY TO STUTT GKY TO STUTT GKY TO STUTT - CTC THLTWK - CYF THLTIKK - SNF AAVIVG - SNF AAUNG - SNF AAUN	NAYIISDN NAYIISDN NSYIISDN NGYMFSDY CCYWNSDY CCYWNSDA IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISKS IAYCISS TGCIYHVIGS TGCIYHVIGS TGCIYHVIGS TGLA SANSIFU SANSIFV TALA-TLL SLM-ELI	CCODEGIEPIPCKHV LANATNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHI MNADTAL MCRD AVPLIAGRHI MNTLA MCCD AIVUPCKHI AQTKPNY FSGINVDITES-HL VKSQSSF LAPY GFELTRN-HL EESDEPL LSPIHFPLTEN-HL SEQADCF FLGVUPDITES-HF VKSEASF LAGV GVT TRA-HL NPERSRN LASV GVT TRA-HL NPERSRN LGTL GIPLTEA-HV LQDYEAS LATL GIPLTEA-HV LQDYEAS LATL GIPLTEA-HV LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSNSVQ MKS GATIGSDTL SVKFVKS WTI SKI VASM SQVQSMV
<pre>>Ag_118793711 >Aa_10883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_i_a576040 >Cs_i_n_Ensemb1 >Cs2_in_Ensemb1 >Cs2_in_Ensemb1 >ds_126306192 >Hs_20306906 >Ce_387373 >Cb_39591559 >Eh_67477376_C >Eh_67476664 >Eh_67466455 >Eh_67464628 >Eh_67477376_N >Ed_167384018</pre>	ESHLEHTI ENTIKGI ESLISOIN VSVIDFIK KSCERYIE CFCLHWIA SCIONIA SVCLENIA DWCLOWIA DWCLOWIA DWCLOWIA DWCLOWIA DWCLOWIA CHCANNIF KKGIDEI VCCIENIA VCCIENIA VCCIENIA	SDHLSEIELR NQGSMDE (1) TILL DDVKLEYQFL SDIRASN (1) TIM SDIRASN (1) TIM (2000FR0 (3) KIQ EK-AYNNHEE (NAGMAN (1) LSK (NAGMAN (1) LSK (NAGMAN (1) LSK (NKGEYNERE (KASEIRQELEI KASEIRQELEI (KASEIRQELEI (KASEIRQELEI (KASEIRQELEI (KASEIRQELEI (SQAVSD (5) FNT TKAGLYD (6) FMT TKAGLYD (6) FMT TKAGLYD (6) FMT (1) NKK GYIQQSI (2) QI LRGNUFI (2) QI	CPLAELFFR- VPLSELFYO- VIPLSELFYO- VIPLSELFYO- VPLSUTLFI- VPSDIVLFI- VPSDIVLFI- VPSDIVLFI- VPSDIVLFI- VPSDIVS- VPSDIVSDIV VPSDIVSDIV VPSDIVSDIV VPSDIVSDIV VPSDIVSDIV VIVEAFVEA- VIEAFVEA- VIVEAFVEA-	KONVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATEK (6) TFPTKIV KESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKTPLSS YANAQRTFLALSL YQKKYHKEPSKTW YQQKLHKEPSKTW YQQKLHKEPSKTW SDIMEKUDVET SDIMEKUPSTI ATMKGF (1) KAKECGI MCSIGHI (1) EKSSFIL NFFKEIIPNITA VCSYGL-SFYL ZOSTIG KKEYGI (3) EKSSFIL NFKEIIPNITA NFKEIIPNITA NFKEIIPNITA	AQQIITKDVYD(1)T AQQIIQKDVYE(1)T AQDILQKDAYE(1)T TDAHFKRDFE(1)T TDAFFKRDFE(1)T ITAGCNSSMP(1)E IESGFTSSSMP(1)E IESGFTSSSMP(1)E IESGFTSSMP(1)E IESGFTSSMP(1)E IESGFTSSMP(1)E ISSGTLVUMD(1)S VSRLLDVVMD(1)S VSRLLDVVMD(1)S VSRLLDVVMD(1)S IRSLLDVAMD(1)S IMRWFSLLGQK(12)G IMNFSLLGQLCA IMNFYRUF(2)S IMSVFTHEIGT IMSVFTHEIGT IMNDVYKQLKI IMNDVYKQLKI IMNDVYKQLKI IMNDVYKQLKI IMNDVYKQLKI IMNDVYKQLKI IMNDVYKQLKI	KGIS DY HEEKD (KGIS DY HEEKD (KGIA DF HEKLG (KGIA OF HEKLG (KGIA OF HEKLG (LGIK CY HEEND (LGIK CY HEEND - IGIK CY HEEND - SNTR KHEEND - SNTR KHEEND - SNTR KHEEND - SNTR KHEEND - TDFH AR SEPK - -CDK TY QLLA () ULY PLSCIF) GKY PLSCVF) GRF ALSKVVF U ORF ALSKVVF U STVT GSTVF - CTS TNSTVF - CYF TNLTIKK - CYF TNLTIKK - CYF TNLTIKK - CYF TNLTIKK - TLF ALAVCK - TLF ALAVCK - SNF ALAVCK - SNF ALVIVG XYH ALSDAR TFC SKTNSEY - SKG ALDSVKI XYH ALSDAR	WAYIISDN WAYIISDN WGYMFSDY CCVWMINT CCVWMINT CCFWMSEV CCFWMSEV CCFWMSEV CCFWMSEV IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN SIM-ELI TALA-TLL TSLM-ELI TSLM-ELI	CCDC GIEPIPCKHV LNANTNP CCLD SIE KPCRLI WANTAL CCLD SIELIAGRHL SNADTIL MCCD AVP IAGRHI (NTNLEA MCCD AIVOPCKHI ACTR FSGI NVDITES-HL VKSQSSF LAPY GFE ITN-HL SEODEPL LSPI HFP TEN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LGV DYD TES-HF VKSEASF LAGV GVI TEA-HL NPERSRN LGTL GIO TEA-HU LQDYEAS FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ FFRR HLK IPTAHQ SSSNSVQ YKTI SKI VASSQVQSMV
<pre>>Ag_118793711 >Aa_10883695 >Aa_10883695 >Dm_21429066 >Ci_23575040 >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_1in_Ensembl >Cs_2in_Ensembl >Cs_2in_Ensembl >Cs_37373 >Cb_39591559 >Eh_67477376_C >Eh_67477376_N >Eh_67477376_N >Ed_167389979_C >Ed_167395021</pre>	ESMLEHT ENTLKGI ESLISOLN VSVLDFIK KSCERYLE CFCLHWLS SFCLWLS SFCLWLS SWCLERVA DWCLKHA NWCLKHA NWCLKHA DS-MKHLA DS-MKHLA DS-MKHLA VLCFNNLF VCCIENLA VLCFNNLF VLCFNNLF VLCENNLF VCCIENLA	SDHLSETELR MQGSMEC 1 TLL DDVKLEY	CPLAELFFR (PLSEFYO) (PLSEFYO) (PLSEFYO) (PLSEFYO) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLS) (PLSE (PLSE)	KONVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATES (6) TFPSVIT KREVM (6) TFPSKTV MESTCE (5) KFKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSFVL ICNAVMTKFALSS YANAQRTKFPLSS YANAQRTKFPLSS YANAQRTKFPLSS YANAQRKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKUPMET ATMKGF (1) KAKECGI MCSIGHI (1) EKSSFI ATMKGF (3) KAKEYGI MCSIGHI (3) EKPSFI ATMKGFIIPNTNA FVESQC	AQQIITKEDVYD (1) T AQAIIQKUYE(1) T TDAHFERDCEE (1) Q TDAFFKKDFEF (1) T TDAFFKKDFEF (1) T ITAGCNSSMFD (1) E IESGENTSSMFE (1) E USSELDVVLMD (1) S VSRLLDVVLMD (1) S VSRLLDVVMD (1) S VSRLLDVAMD (1) S MRMF5LLGQK (12) G MRMF5LLGQK (12) G MRMF5LLGQK (12) G IIRSLLDVAMD (1) S IIRSLLDVAMD (1) S IIRSLLDVAMD (1) S IIRSLLGQK (12) G MRMF5LLGQK (12) G MRMF5LLGQK (12) G MRMF5LLGQK (12) G MRMF5LLGQK (12) G MRMF5LLGQK (12) G MRMF5LGQK (12) G MRMF5LGQK (12) G MRMF5LGQK (12) G IINSLDYKGLK (12) G MRM5VKGLK (12) G IINSLDYKGLK (12) G MRM5VKGLK (12) G	KGISDYHEEKD(KGISDYHEKGG(SGIADFHEKLG(SGIACFEENG(KGIACFEENG(KGIACFEEND(LGIKGYHEEDU GGIKGYHEEDU AGIKGYHEEDU AGIKGYHEEDU AGIKGYHEEDU AGIKGYHEEDU AGIKGYHEEDU AGIKGYHEEDU SNTRCHEEND SNTRC) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY TO STVTH CST TO STVTH -CTC TLLTKK -CYT TLLTKK -CYT TLLTKK -CYT TLLTKK -CYT TLLTKK -CYT TLLTKK -CTC ALLTVK -TLC ALASCK -TLC ALASCK -TLC ALASCK -TLF ALAVCK -SKF ASVIVG -SKF ASVIVG -SKF ASVIVG -SKF ASVIVG -SKF ASVIVG -SKF ASVIVG -SKF ASVIVG -SKF ALSCK -SKF ASVIVG -SKF ALSCK -SKF ASVIVG -SKF ALSCK -SKF ASVIVG -SKF ALSCK -SKF ASVIVG -SKF ASVIVG -S	NAYIISDN	CCDD GIEPIPCKHV LNANTNP CCDD SIE KPCRLI MNADTAL CCLD SIELAGRHL SNADTIL MCRD AVP IAGRHI ONTNLEA MCCD AIVOPCKHI AQTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFE TRN-HL SESDEPL LSPI HFP TEN-HL SEOADCF FLGV DVD TES-HF VKSGSSF LAGV GVI TRA-HL PRDCVSN LASV GVI TRA-HL PRDCVSN LASV GVI TRA-HL PRDESSN LATL GIO TEA-HV LQDYEAS FFRR HLK IPTAHONSSSFNSM LIKD GYE KNNDKF FCVVHKN MKS GAT IGSDTI SVKFVKS YKTI SKI FVAS
<pre>>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >ci_23575304 >ci_23575304 >cs2_in_Ensemb1 >cs1 >cs2_in_Ensemb1 >cs2_in_Ensemb1 >cs1 >cs2_in_Ensemb1 >cs2_in_Ensemb1 >cs2_in_Ensemb1 >cs1 >cs1 >cs2_in_Ensemb1 >cs1 >cs2_in_Ensemb1 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs1 >cs2 >cs2 >cs1 >cs2 >cs2 >cs2 >cs2 >cs2 >cs2 >cs2 >cs2</pre>	ESMLEHTI ENTIKGI ESLISOIN VSVLDFIK KSCERYIE OFCLHWLA SPCLMUA SPCLMUA SPCLMUA DVCLMUA NWCLKHMA NWCLKHMA NWCLKHMA NWCLKHMA VLCFNNLE SS-MKHLA VLCFNNLE VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA VCCVEFLA	SDHLSEIELR NQGSMDE(1) TILL NDVKLEYQFIN SDIRASN(1) TIMN (CDDDFR0(3) KIQ (CDDDFR0(3) KIQ (CDDFR0(3) KIQ (CDDFR0(3) KIQ (NAQMAN(1) LCK (NAGMAN(1) LCK (1) LCK	CPLARLFFR- VPLSEFY- VIPLSEFY- VIPLSFFH- VPICVLFT- VPICVLFT- RELFIVSE- VFESTISA- RELFIVSE- VISESMASA- LITEDLVKC- LITEDLVKC- LITEDLVKC- LITEDLVKC- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VITHICLEDV- VITHICLEDV- VITHICLEDV- VITHICLEDV- VITHICLEDV- VITHICLEDV- VITHICLEDV- VITHICLEDV-	KONVEL (6) TFPSVYI KRATES (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPSTHI KEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKPALTS YANAQRTKPLSS YANAQRTFLALSL YQKK1HKEPSKTW YQCKLQEEPSRPT YQQKLLKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW SDIMEKEPSKTW CSIQH(1) KAKECGI MRFKEIIPNTNA YESGGKVVPISA C-SFYN(2) INSPFSL NRFKEIIPNTNA YESGGRVVPISA C-SFYN(2) INSPFSL C-SFYN(2) INSPFSL C-SFYN(2) INSPFSL	AQ0IITROVYD(1)T AQ0IIQK0VYE(1)T TDAHFERDCFE(1)Q TDAFFKKDFFE(1)T TDAFFKKDFFE(1)T ITAGCNSSMP(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E IESGFTSSNME(1)E INSKELLCVLME(1)E INSKSTFIEGT INNDYKDLKI IMFFKDLECA INFFKLECA INFFKLECA INFFKLECA INFFKLECA INFFKLECA INFFKLECA INFFKLECA INFFKLECA ISSNME	KGIS DY HEEKD (KGIS DY HEEKD (SGIA DF HEKLG (NGIA CF HEKLG (NGIA CF HEND (LGIK CY HEELD - AGIK YF HEELD - AGIK YF HEEND - NITR KY HEEND - SNTR KY HEEND - SNTR KY HEEND - SNTR KY HEEND - TDFH AR SEPK - -CDK TY QLLA (NDFK DF HKSNG (TYYK SF SNIKKI) (NDFK DF HKSNG (TYYK SF SNIKKI) (NNFK DF HKSNG (TYYK SF SNIKKI) (NNFK DF HKSNG (TYYK SF ANIKKI) (SNFK SF ANIKKI) (S) ULY PLSCH) GKY PLSCH) GKY PLSCH) GKY SKUP) SKY TOSTUT SKY 	NAY115DN WAY115DN WAY15DN WGYMF5DY CCWM5EV CCWM5EV CCWM5EV CCWM5EV CCWM5EV CAYWM5DA IAYC15KA IAYC15KA IAYC15KA IAYC15KA	CCODEGIEPIPCKHV LANATTNP CCLD SIENKPCRHL WNADTAL CCLD SIENKPCRHL WNADTAL CCLD SIENKPCRHL WNADTAL MCRD AVP_IAGRHL SNADTIL MCRD AVPUPCKHI AQTKPNY FSGI NVDITES-HL VKSQSSF LAPY GFE ITN-HL SEQADCF LGV UYD ITS-HF VKSEASF LACV GVI TRA-HL PKECVSN LASV GVI TRA-HL PKECVSN LASV GVI TRA-HL PKERSRN LGTL GIQ ITEA-HV LQDYEAS LATL GIQ.TEA-HV LQDYEAS FFRR HLK IFTAHQSSSNSM LKD GYE KNNDKF FCVVHKN GWKS GATI IGSDI SVKFVKS YKTI SKI VASM SQVQSMV
>Ag_118793711 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs_2ir_Ensembl >Cs_ir_Ensembl >Cs_ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Cs_2ir_Ensembl >Est_67477376_C >Est_67477376_N >Ed_167389979_C >Ed_167389079_N >Tv_tviv41h03_q1k_3	ESMLEHTI ENTIKGI ESLISOLN VSVLDFIK KSCFRYLE CFCLHWLS VFCLMVLS SWCLENA SWCLENA DWCLWLA NWCLKHA NWCLKHA NWCLKHA NWCLKHA VLCFNNLF VCCIENLA YECFVKLF KKCIDEIF KKCIDEIF KKCIDEIF VCCIENLA YECFVKLF KKCIDEIF	SDHLSETELR MQGSMEC 1 TLL DDVKLEYOF LA SDIRASN (1) TLN CDDDFRO [3 KI Q CDDDFRO [3 KI Q CDDDFRO [3 KI Q CNAQMAN (1) LSK ONAQMAN (1) LSK NKGGYPNHER SI AGVDS	CPLAELFFR (PLAELFYO) (PLSEFYO)	KQNVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATES (6) TFPSVIT KREVM (6) TFPSKTV MESTCE (5) KPKSFHI KKEVMN (6) EXINKFA YGYTEPVCLSPVL ICNAVMTKFPLSS ISNAMQTKFPLSS YANAQRTKFPLSS YANAQRTKFPLSS YANAQRKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW SDIMMEKEPSKTW CSIQHI(1) EKSSFIL NRFKEIIPNTNA FYESQCKVVPISA ATMKGF (1) KAKEYGI MSFFEIIPNTNA FYESQCRVVPISA C-SFYN (2) IENFFIL NRFKEIIPNTNA FYESQC-RVVPISA	AQUITHEDVYD(1)T AQUAIIGMVYE(1)T TDAHFERDCFE(1)Q TDAFFKHDFFE(1)T TTAGENSMFP(1)E IESGENSSMFP(1)E IESGENSSMFP(1)E USSELSVULMD(1)S VSRLLDVVMD(1)S VSRLLDVVMD(1)S VSRLLDVMMP(1)S IRSLLLDVAMP(1)S IRSLLLDVAMP(1)S IRSLLDVAMP(1)S IRSFLLGGK(12)G MRMFFSLLGGK(12)G MRMFSLLGCK(12)G IMNDYKGLS(12)G INNDYKGLS(12)G MRMFSLLGCK(12)G MRMFSLLGCK(12)G MRMFSLLGCK(12)G MRMFSLLGCK(12)G MRMFSLLGCK(12)G MRMFSLLGCK(12)G INNDYKGLS(12)G MRMFSLGCK(12)G INNDYKGLS(12)G INNT INNT INNT INNT INNT INNT INNT INN	KGIS DY HERKO KIJS CHENOG (SGIA DFHEKLG (NGIA CFHENKIG (NGIA CFHENKIG (LGIK CFHENKIG (POINT AND) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY TO STVTH CST TO STVTH -CTC TLLTKK -CYT TLLTKK -CYT TLLTKK -CYT TLLTKK -CYT TLLTKK -CTC ALLTVK -TLC ALASCKK -TLC ALASCKK -TLF ALAVCK -TLF ALAVCK -SKF ASVIVG SFC COSVAN -SKF ALSCK -SKF ASVIVG) RYH GLODAR) TFC SKTNSE) TFC CSNVA -SKF ALDEVKT) TFC CSNVA -SKF ALDEVKT) TFC CSNVA -SKF ALDEVKT) TFC CSNVA -SKF ALDEVKT) SFC CSNVA -SKF ALDEVKT) SFC CSNVA -SKF ALDEVKT -SKF A	NAYIISDN	CCDD GIEPIPCKHV LNANTNP CCLD SIEKPCRLI MNADTAL CCLD SIELAGRHL SNADTIL MCRD AVP IAGRHI ONTNLEA MCCD AIVPCRHI AQTKPNY FSGINVDITES-HL VKSQSSF LAPYGFE TRN-HL SEODEPL LSPI HFP TEN-HL SEODEPL LSPI HFP TEN-HL SEODEPL LGV DYD TES-HF VKSASF LAGV GVI TRA-HL PRECVSN LASV GVI TRA-HL PRERSRN LGTL GIPL TRA-HL PRERSRN LGTL GIPL TRA-HL ODYEAS FFRR HLK IPTAHONSSENSM FFRR HLK IPTAHONSSENSM FFRR HLK IPTAHONSSENSM LIKD GYE KNNDKLF CVVHKN MKS GAT IGSETL SVRFVKS YKTI SKIFVAS SQVQSMV
<pre>>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs2_in_Ensembl >Cs2_in_</pre>	ESMLEHI ENTIKGI ESLISOLN VSVLDFIK KSCFRYLE CFCLHWLA SFCLOWLA SFCLOWLA SFCLOWLA SFCLOWLA DWCLKHA NWCLKHA NWCLKHA NWCLKHA NWCLKHA NWCLKHA VLCFNNLE S-MKHLA VLCFNNLE VQCVEFLA	SDHLSEIELR MQGSMEC1171LL DDVKLEYQE1A SDIRASN(1)71M SDIRASN(1)71M (CDDDFR0(3)KIQ EK-AYNNHEC (NAQMAN(1)LSK NKGGYPNLFR NKAGENC	CPLARLFFR VPLSEFYO- VIPLSEFYO- VIPLSFTS- KELEHLARE- KELEHLARE- VFELEFLVSE- VFSESTISA- LITVEDLVK- VFSESTISA- UTVEDLVK- UVAFVEA- TITVEDLVVG- VIVAFVEA- TITVEDLVVG- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VIVAFVEA- VITTITULFDV- MESVDFVEC- MLAFDEFVF- VISFUVAFF- VISFEVF	KQNVEL (6) TFPSVYI KRATES (6) TFPSVHI KRATES (6) TFPSVHI KESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSFVL ICNAVWTKFALTS ISNAMQTKTPLSS YANAQRTKIPLSS YANAQRTKIPLSS YANAQRTKIPLSS YANAQRTELALSL YQCKLHKEPSKTW YQCKFLKEPSKTW YQQKFLKEPSKTW YQQKFLKEPSKTW YQQKFLKEPSKTW SDIMEKIDVET SDIMEKUDVET SDIMEKUPSTA ATMKGF (1) KAKECGI WCSIQHI (1) EKSSFL ATMKGF (3) KAKEYGI WCSIQHI (3) EKSFL ATMKGF (3) KAKEYGI WCSIQHI (3) EKSFL NCSIQHI (3) EKSFL NCSIQHI (3) EKSFL C-SFYN (2) ISPYSL FSD	AQUITREDVYD(1)T AQUITREDVYD(1)T TDAHFERDCFF(1)Q TDAFFKRDFFF(1)T TDAFFKRDFFF(1)T ITAGCNSSMP(1)E EESGFTSSMME(1)E EESGFTSSMME(1)E IESGFTSSMME(1)E USSELDVLMD(1)S VSRLLDVVQD(1)S VSRLLDVVQD(1)S VSRLLDVLMD(1)S WRWFSLLCQK(12)G MRWFSLLCQK(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ(12)G INNFYGLECQ INFY	KGIS DY HEEKD (KGIS DY HEEKD (SGIA DF HELG (NGIA OF EDND (LGIK CY EELD (LGIK CY EELD (LGIK CY EELD (LGIK CY EELD (AGIK CY EELD (SNTR CK EEND (CK CK CK (CK CK CK CK (CK CK CK CK CK (CK CK CK CK CK CK CK CK CK (CK CK C) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY ALSEVVF) JKY TO STUTT -CTC TLLTVK -CYF TLLTVK -SVF ALSON -SVF ALSON	WAY115DN WAY115DN WAY15DN WCYMF5DY CCYWNTNT CCYWNSDA CCYWNSDA IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IAYC15XS IASTICDL	CQD.GIEPIPCKHV LANATNP CCLD SIE KEVERLI WANTAL CCLD SIE KEVERLI WANTAL CCLD SIE KEVERLI WANTAL MCRD AVP_IAGRHI (NTNLEA MCCD AIVOPCKHI AQTKPNY FSGI NVD.TES-HL VKSQSSF LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GVI TRA-HL PAEOUSN LASU GVI TRA-HL PAEOUSN LASU GVI TRA-HL VLEASSF LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS FRR HLK IFTAHOSSSNSM LKD GYE KNNDKF FCVVHKN GVKS GATI IGSDTI SVKFVKS YKTI SKI VASM SQVQSW EITCFN VNQENIL LKNK KISKEIDQL ILVFINK LKSK GYR INSEL FOUNDEN KKS GATI GSETI SVRFVKS YKTI SKI VASV PQVQSWV
<pre>>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Ci_23575304 >Cs2_in_Ensemb1 >Cs2_in_Ens</pre>	ESMLEHTI ENTIKGI ESLISON VSVLDFIK KSCERYIE CFCLHWA SCLOWA SSCLOWA SSCLOWA SSCLOWA NWCLCWA DGLOWA NWCLKHA DG-MKHLA VLCENNIE KKCIDEF VCCIENIA VCCIENIA VCCIENIA VCCIENIA VCCIENIA VCCIENIA VCCIENIA VCCIENIA VCCIENIA SECOVEFA	SDHLSEIELR NQGSMDE (1) TILL DDVKLEYQFL SDIRASN (1) TILN CCDDFRE (2) KIQ (CCDDFRE (2) KIQ (CAGMAN (1) LSK (CAGMAN (1) LSK (1) CAGMAN (1) LSK (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1) CAGMAN (1)	CPLAELFFR- VPLSELFYO- VPLSELFYO- VPLSELFYO- VPLSITSA- HELENLARE- FELFIVSE- VYGESMASA- LITVEDLVIK- LITVEDLVIK- LITVEDLVIK- UVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIVEAFVEA- VIESELYL- VESDEVVEF- VIESELYL- VESELYL- SEDVVWEF- FDIEVVEL-	KONVEL (6) TFPSVYI KRATEK (6) TFPSVHI KRATEK (6) TFPTKIV MKESTCE (5) KPKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSPVL ICNAVNTKFALTS ISNAMQTKFPLTS YANAQRTFLALSL YQKKYHKEPSKTW YQQKLHKEPSKTW YQQKLHKEPSKTW YQQKLHKEPSKTW SDIMEKUDVET SDIMEKUPKTT ATMKGF (1) KAKECGI MCSIGHI (1) EKSSTIL ATMKGF (3) KAKEYGI ATMKGF (3) KAKEYGI ACSIGHI (3) EKPSTIL ATMKGF (3) KAKEYGI C-SFYN (2) IESPSTIL ATMKGF (2) IESPSTIL ATMKFF (2	AQQIITKDVYD(1)T AQQIIQKDYE(1)T AQDILKKDAYE(1)T TDAHFKRDFE(1)T TDAFFKRDFE(1)T ITAGCNSMPD(1)E IESGFTSSMME(1)E IESGFTSSMME(1)E IESGFTSSMME(1)E IESGFTSSMME(1)E IESGFTSSMME(1)E ISSGTSSME(1)E ISSGTSME(1)E MRWFSLLGVKD(1)S MRWFSLLGVK(12)G IMRWFSLLGV(12)G IMRWFSLLGV(12)G IMRWFSLLGV(12)G IMRWFSLLGV(12)G IMRWFSLGV(12)G IMRWFSLGVC(3)N IFRCL-CFLS(3)E IMSVS-THEGE IMMIFKQLECA IMMIFKQLECA IMMIFKGLEGS IDVKKSYDNLKSL IEMSST-FKGGAZET	KGIS DY HEEKD (KGIS DY HEEKD (KGIA DF HEKLG (NGIA OF HEKLG (NGIA OF HEKLG (NGIA OF HECND (LGIK CY HEELD - AGIK Y EELD - AGIK Y EELD - AGIK Y EELD - SNTR KHEEND - SNTR KHEEND - SNTR KHEEND - TDFH AR SEPK - -CDK TY QLLA (ER Y KKIN (TKYK SF HSNIK (NFK DF HSNIK (NFK OF HSNIK (NFK NF KIN (ER Y KKIN () ULY PLSCTP) GKY PLSTYP GKY PLSTYP STORE) GKY ALSKVYP I SKY TO STVTP -CYF TNLTIKK -CYF TNLTIKK -CYF TNLTIKK -CYF TNLTIKK -TIC ALSCK -TIC ALSCK -TIC ALSCK -TIC ALSCK -TIF ALAVCK -SNF AAVTVG I FY GLQDAR I FY GLDAR I FY GLD	WAYIISDN WAYIISDN WSYNISDN WSYNFSDY CCVWMSEV CCFWMSEV CCFWMSEV CCFWMSEV IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN IAYCISKN TALA-TLL TSLM-ELI TSLM-ELI TSLM-ELI TSLM-ELI TSLM-ELI TSLM-ELI TSLM-ELI TSANSIC LOTVRPL	CCODGIEPIPCKHV LANATNP CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL CCLD SIENKPGRHL MNADTAL MCCD AIVD/CKHI AQTKPNY FSGINVDITES-HL VKSQSSF LAPY GFE TRN-HL EESDEPL LSPI HFPLTEN-HL EESDEPL LSPI HFPLTEN-HL EESDEPL LGV GVI TES-HF VKSEASF LACV GVI TAA-HL PKERSRN LGTL GIP TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS LATL GIG TEA-HV LQDYEAS LATL GIG TEA-HV LQDYEAS VXII SKIFVAS-VKS YKTI SKIFVAS-VKS YKTI SKIFVAS-VKS YKTI SKIFVAS-VKS YKTI SKIFVAS-V SVRVKS YKTI SKIFVAS-VKS YKTI SKIFVAS-VFVKS YKTI SKIFV
<pre>>Ag_118793711 >Aa_108883695 >Aa_108875394 >Gm_78540983 >Dm_21429066 >Ci_23575304 >Cs2_in_Ensembl >Cs2_in_</pre>	ESMLEHTI ENTIKGI ESLISON VSVLDFIK KSCERYLE CFCLHWLS SPCLWLS SPCLWLS SWCLERNA DWCLKHA NWCLKHA NWCLKHA NWCLKHA NWCLKHA CS-MKHLA DS-MKHLA DS-MKHLA VLCFNNLF VCCIENLA YECFVKLF VCCIENLA YECFVKLF VCCIENLA SKCIDEF VCCIENLA SKCIDEF VCCIENLA SKKIDEF VCCIENLA SKKIDEF SWYFAAR RWYFAAR RWYFAAR	SDHLSETELR MQGSMEC 1 TLL DDVKLEYOFIN SDIRASN (1) TLN CODDFRE (3) KIQ CODDFRE (3) KIQ CODDFRE (3) KIQ (NAQMAN (1) LSK (NAGMAN (1) LSK NKGGYPNHER SIASUDSFIE NKAGMENHER KASEIRQHER KASEIRQHER KASEIRQHER KASEIRQHER KASEIRQ	CPLAELFFR (PLSEFYO) (PLSEFYO) (PLSEFYO) (PLSEFYO) (PLSEFYO) (PLSISISA (PLSEF) (PLSESISA (PLSEF) (PLSESISA (PLSES) (PLS	KQNVEL (6) TFPSVIT KRATES (6) TFPSVIT KRATES (6) TFPSVIT KREATEK (6) TFPTKTV MESTCE (5) KFKSFHI KKEVMN (6) EKINKFA YGYTEPVCLSFVL ICNAVMTKFALTS ISNAMQTKFPLSS YANAQRTFPLSS YANAQRTFPLSS YANAQRTFPLSS YANAQRKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW YQQKLKEPSKTW SDIMGEKVDMET SDIMGEKVDMET ATMKGF (1) KAKECGI WCSIGHI (2) EKSSTI ATMKGF (3) KAKEYGI WCSIGHI (3) EKPSTI CSFYN (2) IENPFSL ATMKGF (2) KAKEYGI WCSIGHI (2) EKPSTI CSFYN (2) IENPFSL CSFYN (2) IENFFSL CSFYN (2) IENFFSL CSFY	AQUITKEDVYD (1) T AQUAITGROVYE (1) T TDAHFERDCEE (1) Q TDAFFEKDEFE (1) T ITDAGROSSMEP (1) E IESGETSSME (1) E IESGETSSME (1) E VSRELDVVLMD (1) S VSRLLDVVMD (1) S VSRLLDVVMD (1) S VSRLLDVAMD (1) S IRSLLDVAMD (1) S INFKCL-ES	KGIS DY HEEKD (KGIS DY HEEKD (SGIA DF HEKLG (SGIA DF HEKLG (SGIA OF HEND (LGIK CY HEEND (LGIK CY HEEND - AGIK YF HEEND - SNTR CW HEND - -CDK T OLLA (-ER CY KIN (NDFK OF KSNG (-IEK OI KEIM - -SNR PT OLLA (-ER CY KKIN (NNFK OF KSNG (-IEK OI KEIM - SNR CY HKIN (NNFK OF KSNG (-IEK OI KEIM - -ER CY KKIN (NNFK OF KSNG (-IEK OI KEIM - -ER CY KKIN (NNFK OF KSNG (-IEK OI KEIM - -ER CY KKIN (NNFK OF KSNG (-IEK OI KEIM - -ER CY KKIN (NNFK OF KSNG (-IEK OI KEIM - -ER CY KKIN (-IEK OI KEIM - -IEK OI KEIM - - SUN -) ULY PLSECT) GKY PLSECT) GKY PLSECT) GKY CSTUT SKY TGSTUT - CTC TLLTKK - CTC TLLTKK - CTC ALLTVK - CTC ALLTVK - TLC ALASCK - MC - MC ALASCK - MC A	NAYIISDN	CQD.GIEPIPCKHV LANATNP CCLD SIE KEVERLI WANTAL CCLD SIE KEVERLI WANTAL CCLD SIE KEVERLI WANTAL MCRD AVP_IAGRHI (NTNLEA MCCD AIVOPCKHI AQTKPNY FSGI NVD.TES-HL VKSQSSF LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GFE ITN-HL SEODEPL LAPY GVI TRA-HL PAEOUSN LASU GVI TRA-HL PAEOUSN LASU GVI TRA-HL VLEASSF LATL GIQ TEA-HV LQDYEAS LATL GIQ TEA-HV LQDYEAS FRR HLK IFTAHOSSSNSM LKD GYE KNNDKF FCVVHKN GVKS GATI IGSDTI SVKFVKS YKTI SKI VASM SQVQSW EITCFN VNQENIL LKNK KISKEIDQL ILVFINK LKSK GYR INSEL FOUNDEN KKS GATI GSETI SVRFVKS YKTI SKI VASV PQVQSWV

Figure I

Multiple sequence alignment of representatives of MAEL domain. The sequences are represented by an abbreviation of species name followed by database entry ID. The homologues of *C. savignyi* identified in Ensembl database are indicated by Cs1 and Cs2. The consensus in 80% of the sequences is shown below the alignment based on default amino acid classes in Chroma. The numbers in bracket are indicative of the excluded residues from sequences. For a complete multiple sequence alignment refer to additional file 1. Species name abbreviations: Aa, Aedes aegypti; Ag, Anopheles gambiae; Am, Apis mellifera; Cb, Caenorhabditis briggsae; Ce, Caenorhabditis elegans; Ci, Ciona intestinalis; Cs,Ciona savignyi; Dm, Drosophila melanogaster; Ed, Entamoeba dispar SAW760; Eh, Entamoeba histolytica; Gg, Gallus gallus; Gm, Glossina morsitans; Hs, Homo sapiens; Md, Monodelphis domestica; Lb, Leishmania braziliensis; Tb, Trypanosoma brucei TREU927; Tr, Trypanosoma congolense; Tv, Trypanosoma vivax; Xt, Xenopus tropicalis.

mosquito (*A. aegypti*), three copies in *Culex pipiens*, and five copies each in amoeba *E. dispar* and *E. histolytica*. Phylogenetic tree construction suggests that multiple MAEL copies are generated from a series of ancient lineage-specific duplication events (Figure 2A). Strikingly, no fish MAEL homologues could be identified. Its absence in teleost fish was confirmed by carefully examining the published whole genome databases in Ensembl for five

different species (*Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes*, *Takifugu rubripes* and *Tetraodon nigroviridis*). It can be inferred that it is the ancestor of the fish lineage after the divergence of teleost and tetrapod lineages that underwent the loss of MAEL domain. The timing of the loss is probably related to the ancient fish-specific genome duplication [33].

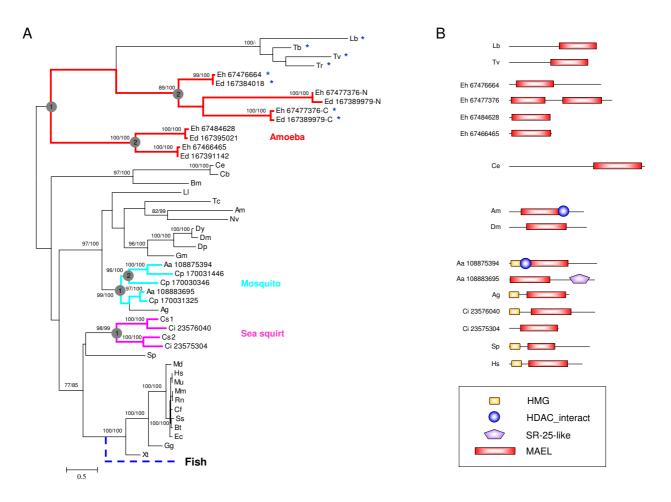


Figure 2

Phylogenetic relationship and domain architectures of MAEL proteins. (A) An unrooted phylogenetic tree was reconstructed using maximum likelihood (ML) analysis and Bayesian analysis. Single MAEL domains are represented by species names. The duplicated ones in the species of Aa, Ci, Cp, Ed, Eh are represented by species names following Genbank ID, whereas Cs domains are represented by Cs1 and Cs2. Branch length is proportional to estimated evolutionary change by PhyML program; the scale bar represents 0.5 substitution per site. Node supporting values greater than 75% from ML bootstrap analyses and Bayesian MCMCMC sampling are shown on the left and on the right of the slash, respectively. Lineage-specific expansions of MAEL domains in amoeba, mosquito and sea squirt are highlighted with different colors (red, turquoise blue and pink) and ancient duplication events were indicated by circled numbers. The loss of MAEL in teleost fish is indicated in blue dashed. Asterisk labeled MAEL domains are the ones containing both conserved EHHCHC and EDDHD residues (see following main text). (B) Domain architectures of representatives of the MAEL proteins were deduced through searching against Pfam and SMART domain databases and drawn approximately to scale. The domains shown are: HDAC_interact, named after Histone deacetylase (HDAC) interacting (SMART: SM00761); HMG, named after High Mobility Group (SMART: SM00398); SR-25-like (DUF1777, Pfam: PF08648). New species name abbreviations: Bm, Brugia malayi; Bt, Bos taurus; Cf, Canis familiaris; Cp, Culex pipiens quinquefasciatus; Dp, Drosophila pseudoobscura; Dy, Drosophila yakuba; Ec, Equus caballus; Lb, Leishmania braziliensis; Ll, Lutzomyia longipalpis; Mm, Mus musculus; Mu, Macaca mulatta; Nv, Nasonia vitripennis; Rn, Rattus norvegicus; Sp, Strongylocentrotus purpuratus; Ss, Sus scrofa; Tc, Tribolium castaneum. Other species name abbreviations refer to Figure 1 legend.

Functional insight from domain architectures

Three other domains are associated with MAEL domains, including HMG (SMART: SM00398), HDAC_interact (SMART: SM00761), and SR-25-like domain (DUF1777, Pfam: PF08648) (Figure 2B). HMG is a common DNA-

binding module in a variety of chromatin-associated proteins and functionally involved in the nucleoprotein complex assembly during genome recombination, initiation of transcription, and DNA repair [32]. The association between MAEL and HMG domains in most species suggests that the MAEL domain may somehow function in a DNA-related process. This functional assignment is also suggested by the association of MAEL domain with HDAC_interact domain in two homologues from mosquitoes (A. aegypti and A. gambiae). The HDAC_interact domain is known to bind to histone deacetylases (HDACs), core enzymes for removing acetyl group from lysine residue of histones during chromatin remodeling process [34]. It has been observed that pairs of interacting domains in one organism may have a fusion homologue composing of these two domains in another organism, known as the rosetta stone protein theory [35]. Mosquito MAELs may be rosetta stone proteins and it can be hypothesized that there are interactions between other MAEL and some HDAC_interact-containing proteins in other species. Indeed, it has been illustrated that mouse MAEL can interact with the SIN3B protein which contains an HDAC_interact domain [30]. The associated SR-25like domain provides another link between the MAEL domain and RNA-related process. The SR-25-like domain is associated with RNA-binding modules, RNA recognition motif (RRM) [26] and PRP38 [36], It is also distantly related to SR-25 domain which may be involved in RNA splicing, as revealed by the SCOOP program [37]. Therefore, domain architecture suggests a potential involvement of MAEL domains in DNA binding, RNA binding and chromatin remodeling.

A distant similarity between MAEL domains and the DnaQ-H 3'–5' exonuclease family with the RNase H fold

We applied a fold recognition strategy to identify remotely related homologues of MAEL domains. The rationale is that in the case of remote homology, conserved protein structural folds can be kept despite limited sequence identity [38]. A meta server was utilized, which assembles various state-of-the-art fold recognition methods and further evaluates modeled structures based on a consensus score computed by a 3D-JURY system [39]. MAEL domains from human, X. tropicalis, Ciona and Drosophila were first used as queries and several structural hits were identified by MetaBasic, ORFeus and BasicDist with consensus scores from 21 to 46. Although these 3D-Jury scores are below the cutoff 50, which corresponds to correct assignment with statistical significance [40], domain and fold examinations showed that all retrieved structures belong to the DnaQ-H 3'-5' exonuclease family with the RNase H fold [41,42]. We extended our search using an ancestral E. histolytica MAEL domain (GI: 67477376, residues 315-532) as a query. Eleven structural hits were identified with high scores around 58-69, and they all belong to the DnaQ-H 3'-5' exonuclease family. Structural fold similarities between DnaQ-H and MAEL domains encouraged us to re-examine this relationship using PSI-BLAST. We noticed that several DnaQ-H exonucleases can be retrieved as insignificant candidates in our initial PSI-

BLAST searching with a profile inclusion expectation (E) value of 0.005. However, when we set inclusion E value at 0.05, significant similarity between the first 100aa segment of MAEL domains and several prokaryotic DnaQ-H exonucleases was achieved in the fourth iteration.

We next examined this homologous relationship by building structure-based multiple sequence alignments for MAEL domains and DnaQ-H 3'-5' exonucleases. Since sequence identity among different DnaQ-H domains is very low, their alignment was first generated based on structural information as assessed by a CE-MC server [43] followed by manual adjustment based on published literature. Thereafter, we combined this alignment with the aligned MAEL domains based on fold recognition results and predicted secondary structures. The final alignment showed the conserved residues among/between two domains and compositions of secondary structures (Figure 3A). It is to be noted (Figure 3A) that equivalents of beta sheet (β) 3 of the RNase H fold in most MAEL domains are predicted to be alpha helix (α). We believe that this is a wrong prediction since in the canonical RNase H fold, β 3 is an edge β strand, which can usually be misidentified as an α helix because of its solvent sequence property [44]. As shown in Figure 3A, the secondary structures of MAEL domains resemble those of DnaQ-H 3'-5' exonucleases; both have a β 1- β 2- β 3- α 1- α 2- β 4- α 3- β 5- α 4- α 5- α 6 composition. More importantly, several ancestral protist MAEL domains also share all the critical DnaQcharacteristic residues (Asp-Glu-Asp-His-Asp, Η DEDHD). These residues are commonly utilized by diverse DnaQ-H 3'-5' exonucleases and interact with two divalent metal ions to form an active site [45-47]. Thus, in contrast to a very low sequence identity (<15%) between MAEL domains and DnaQ-H 3'-5' exonucleases, the similar structural fold and the notable existence of DEDHD residues in protist MAEL domains strongly support a distant evolutionary relationship.

Structural examinations on active sites by DEDHD and EHHCHC residues in MAEL domains

The tertiary structures of protist and chicken MAEL domains were further constructed by comparative modeling. Like DnaQ-H domains (Figure 3B, C), these MAEL domains adopt a similar RNase H structural fold which is characterized by a compact α/β fold with open anti-parallel β sheets in the middle and several α helices surrounded (Figure 3D, E). Moreover, the characteristic DEDHD residues in protist MAEL domains are clustered into a structural core, which resembles active sites of DnaQ-H domains (Figure 3A, B, C). In contrast, most other MAEL domains lack the DnaQ-H specific residues DEDHD. However, they are characterized by another conserved stretch of residues, EHHCHC. During evolution such conservation of MAEL-specific residues may reflect functional

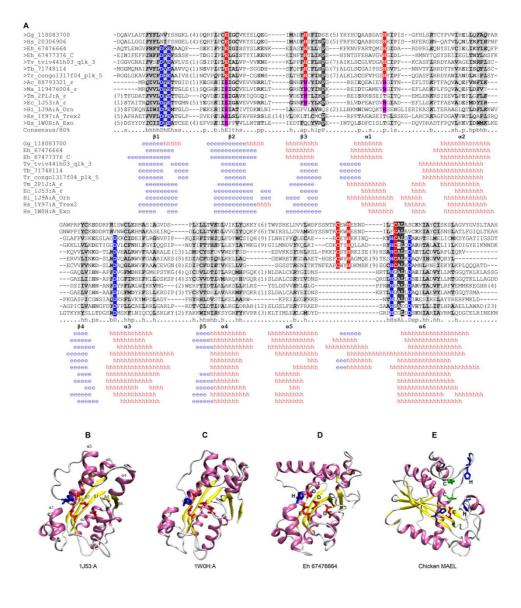


Figure 3

Sequence and structure similarity between MAEL and DnaQ-H domains. (A) Sequence and the secondary structure alignment of MAEL and DnaQ-H domains. Seven DnaQ-H domains are included and five domains have 3-D structures: Thermotaga maritime ε exonuclease (Tm ε , 2PIJ:A), Escherichia coli ε exonuclease (Ec_ ε , IJ53:A), Haemophilus influenzae oligoribonuclease (Hi_Orn, IJ9A:A), human Trex2 exonuclease (Hs_Trex2, IY97:A) and human 3'-5' exoribonuclease (Hs_Exo, IW0H:A). The conserved DnaQ-H specific residues (DEDHD) are highlighted with the blue background; whereas the conserved MAELspecific residues (EHHCHC) are highlighted with the red background. The MAEL-specific residues (EHH) also exist at counterpart positions in some DnaQ-H domains, and were highlighted with pink background. Detailed secondary structures for MAEL domains and DnaQ-H domains are obtained from secondary structure predictions and the 3-D structures, respectively; they are shown below the alignment (h in red, α helix; e in blue, β sheet). The structural sequence alignment was established carefully by hand on the basis of CE-MC results, alignment in fold recognition, literature information, and predicted secondary structures. The numbers in bracket are indicative of the excluded residues from sequences. New species name abbreviations: Tm, Thermotaga maritime; AcAlteromonas macleodii; Mamarine gamma proteobacterium. (B-E) NewCartoon diagrams for DnaQ-H domains (1)53:A and 1W0H:A) and the homology model of two MAEL domains (Eh 67476664 and chicken MAEL). The α helices are shown in pink, β sheets in yellow, and loops in white; Their spatial locations are labeled in 1J53:A. Strictly conserved DnaQ-H active site residues DEDHD of protist MAEL domain (Eh 67476664) are highlighted in licorice drawing with acidic residues (D and E) in red and basic His in blue. The MAEL-specific residues EHHCHC of chicken MAEL domain are highlighted with Glu in red, His in blue and Cys in orange.

contributions most likely to a distinct active site. The spatial locations of EHHCHC residues were then examined in these modeled MAEL structures to check their possibility of forming an active site. Unexpectedly, we found that all MAEL-specific residues have very close spatial locations and they are clustered together at one side of the middle anti-parallel β sheets (Figure 3E). Four residues (EHHchC) can shape a structural core and other two residues may also potentially face down to it with slight structural rearrangements. Similar change of structural conformations of α 5 and α 6 comprising last CHC residues has been observed in crystal structures of DnaQ-H domains (additional file 2). There may exist another possibility that a disulfide bond (-S-S-) is formed between two Cys residues (C178 and C189 in the chicken MAEL domain) because of their close proximity. This is also supported by disulfide bond predictions [48]. Formation of a disulfide bond may facilitate the last His to approach other EHH residues, thus forming an active site with EHHH residues. Therefore, structural examinations suggest that protist MAEL domains with DEDHD residues may form a DnaQ-H active site whereas other MAEL domains with EHHCHC residues may potentially form a new active site based on the canonical RNase H scaffold.

Discussions and conclusion

Functional insight into MAEL in germline piRNA pathway The proposed evolutionary link of MAEL domains to DnaO-H 3'-5' exonuclease with RNase H fold may provide functional clues for MAEL domains. The DnaQ-H 3'-5' exonuclease family, also known as DEDDh exonuclease family or Exonuc_X-T domain (Pfam ID: PF00929), is one member of RNase H fold superfamily (SCOP: 53098) which also includes RNase H, mu transposase, crossover junction resolvase RuvC, and PIWI domain families [24,49-52]. They all share a canonical RNase H fold but contain different active site residues. The DnaQ-H family is characterized by five conserved residues, DEDHD, which form an active site in coordination with divalent metal ions (Figure 3A). Its members contribute to diverse nucleic acid metabolism processes such as replicative proofreading (1J53:A) [47], DNA repair or RNA degradation (exonuclease I and oligoribonuclease) [45,46], and RNA interference (ERI-1) [53]. Although different nucleotide targets (DNA or RNA) or diverse metal ions (Zn2+, Mg^{2+} , or Mn^{2+}) are involved [45-47], their active sites formed by the EDDHD residues delineates a common 3'-5' exonuclease activity. That is, the acidic DEDD together with two metal ions shape a negative pocket, which provides space for accommodating the 3' termini of oligonucleotide (DNA or RNA) chains. Thereafter, the coordinated metal ions and another conserved H are in direct contact with the bound chain, which induces a break of the phosphodiester bond of nucleotide in the 3'-5' direction [46]. Therefore, protist MAEL domains, harboring DnaQ-H specific DEDHD residues and active sites, may also employ a 3'–5' exonuclease activity, although their associated metal ions and nucleotide targets are still unknown.

In contrast to the protist MAEL domains, most recent MAEL domains do not contain the DnaQ-H specific residues but are characterized by the EHHCHC residues. What is the functional contribution of these residues to MAEL domains? Structural observations showed that a structural core can be potentially formed by the MAELspecific residues EHHCHC or EHHH. This may provide a structural basis for an active site. On the one hand, this active site may confer RNA-binding ability for MAEL domains because of the lack of DnaQ-H specific residues. In this way, MAEL may contribute to stabilizing or positioning the RNA substrate in piRNA pathway. On the other hand, MAEL-specific residues and its potential active site may define another nuclease activity. We noticed that although all related families with the RNase H fold have low sequence identities and contain different active site residues, they all have DNA/RNA 3' or 5' enddirected nuclease activities with metal ion coordination in their own active sites [50,51]. For example, RNase H is a non-specific endonuclease whose catalytic activity requires divalent ions (Mg²⁺ or Mn²⁺) and is responsible for the hydrolysis of the RNA in a DNA/RNA duplex [52,54]. In contrast, PIWI domains contribute to 5'-3' exonulcease catalytic activity for the Argonaute family proteins (Slicer) in all types of small RNA pathways (siRNA, miRNA, and piRNA). The activity is achieved by three PIWI active site residues, DDH, in coordination with one divalent ion and used to cleave single-stranded RNA substrate guided by complementary double-stranded small RNAs (piRNA or siRNA) [23,24,55-57]. It seems that the RNase H structural fold is an efficient scaffold from which diverse nuclease families have evolved distinct nuclease activities by developing their own active site residues with metal ion coordination. Therefore, being one member of RNase H superfamily, the MAEL domain may share this characteristic, thus the residues EHHCHC may form an active site with a new nuclease activity. It has been shown in diverse proteins that H, C and E residues often interact with Zn²⁺ [58]. Moreover, the residue composition of EHHH is commonly utilized by several Escherichia coli proteins including ColE7 endonuclease [59], Zinc transport protein ZnuA [60], and Aldolase (1DOS) for their active sites, which also interact with metal ions, especially Zn²⁺ [61],

Experimental evidence have suggested that MAEL may be involved in piRNA biogenesis since its loss-of-function mutant impairs the production of piRNAs or rasiRNAs and increases the transcript level of transposable elements [11]. Different from siRNA and miRNA pathways, piRNAs biogenesis employs a Dicer-independent mechanism [4,10]. A ping-pong model has been recently proposed for this process and it is hypothesized that AGO3 bound to the sense strand of piRNAs catalyzes cleavage of the antisense strand that generates 5' end of antisense piRNAs. The 3' end of the resulting antisense piRNAs is subjected to a 3' cleavage by an unknown endonuclease or exonuclease and a HEN1-processed 3' methylation. Thereafter, the produced antisense piRNAs associate with Aubergine or PIWI and direct cleavage of transposon sequences, which then generates the sense strand piRNAs after 5' cleavage, 3' cleavage and 3' methylation [5,7,8]. This cycling model is not complete since the exonuclease or endonuclease enzyme responsible for the 3' terminal maturation remains uncharacterized [5,7,8]. Thus, because of its evolutionary relationship to 3'-5' DnaQ-H exonuclease and the potential (3'-5' exo-) nuclease activities, MAEL may be the nuclease candidate implicated in the cleavage of the 3' termini. Recently, the nucleases Zucchini and Squash have been proposed as the 3' termini nuclease candidate based on the evidence that they are also located in germ plasm and have a similar mutation phenotype in a loss of transposon silencing [18]. However, MAEL is distinct from those above two nucleases due to its translocation between germ plasm and nucleus and the direct interaction with chromatin remodeling proteins [21,30]. We believe that multiple nucleases are involved in the diverse steps of piRNA pathway in a sequential manner, similar to PIWI family members targeting 5' cleavage of piRNAs [62]; and MAEL is involved in a genomic DNA-related piRNA step, which may include chromatin remodeling process and initial transcriptions of transposon. In this way, MAEL-associated HMG domain or other chromatin remodeling proteins facilitate the access of piRNA complex to the genomic regions where are enriched with transposon sequences. The transposon transcripts undergoing processing interact with the piRNA complex in which PIWI, one RNase H member, generates 5' end of transposon transcripts via a piRNAdirected homologous cleavage whereas MAEL, another RNase H member, contributes to a 3' terminal cleavage of transposon transcripts.

Unique evolutionary characteristics for MAEL domains

Phylogenetic analysis has revealed several unique characteristics of MAEL domains including single-copy status in most species, ancient lineage-specific expansion and the loss in the teleost fish lineage. It has been long recognized that during evolution eukaryotic species have high duplication rates [63] and vertebrates have experienced two or three whole or regional genome duplications [33,64,65], which led to expansions of some domain families. It is of great interest that MAEL domain has escaped the usual duplication potential in most species, especially in vertebrates. It is also possible that the duplicated sequence was lost after duplication. However, it seems that this singlecopy status is commonly inherited by several domains including SANTA domain [66]; an evolutionary selection against domain duplication together with the functional conservation, therefore, should account for the establishment of this status. We did observe MAEL domain expansion in several species. One or two duplication events occur at the ancestor of each lineage before its further divergence (Figure 2A). This ancient lineage-specific expansion may be caused by the release of evolutionary constraints in individual lineages. Thereafter, functional complexity may have arisen, as exemplified by diverse protist MAEL domains with either DEDHC+EHHCHC residues or EHHCHC residues (Figure 2A and legend).

We also observed the loss of MAEL domain in all examined teleost fish species. Gene loss in protein family evolution is well-recognized. The lost member may be functionally replaced by another member of the same family. However MAEL does not belong to this case because of its single-copy nature especially in the vertebrates. What happens in teleost fish germline cells without the MAEL protein? One possibility is that fish have a distinct but functionally similar counterpart, which remains to be characterized. Another possibility is that MAEL loss results in a unique piRNA pathway or a unique developmental morphology in fish germline cells compared to mammals and flies. Indeed, a distinct cellular distribution of Vasa protein, a marker for germline cells, has been observed in fish [10]. Moreover, it seems that although RNAi is evolutionarily conserved among species, individual lineage tends to develop some unique steps for the RNAi pathway, as shown in plant-specific XS domain in post-transcriptional gene silencing [67] and worm-specific Argonaute subfamily [62]. Furthermore, although the evolutionary and functional implications of MAEL loss in the teleost lineage are not yet understood, a practical implication can be hypothesized that fish may be amenable natural MAEL knockout-like models where transgenic insertion of MAEL proteins could be used to as a strategy for studying its function and the germline piRNA pathway.

Active site switch, a novel path towards protein function change

How did MAEL domain evolve from the DnaQ-H domain? Considering the oldest identified MAEL domains are from *Protista* that represents the earliest eukaryotic branches [68], we believe that the first generation of MAEL domains should be traced back to an ancestral eukaryotic or a prokaryotic DnaQ-H domain, from which the MAEL-specific characteristics might have originated. Indeed, the first three MAEL-specific residues EHH are more ancient than others and commonly found in different prokaryotic ϵ exonucleases (Figure 3A). Their spa-

tial locations are also close as shown in 1J53:A [47] (additional file 3), thus providing a substrate for evolving to a mature active site. It can be hypothesized that the DnaQ-H ancestor underwent a gene duplication event (additional file 4) in early Eukaryota or during the divergence of the prokaryotes and eukaryotes, corresponding to the time when small RNA pathways emerged. Thereafter, the duplicated one (MAEL ancestor) obtained a protein motif comprising CHC residues, forming an evolutionary intermediate which has both DEDHD and EHHCHC residues. The original DnaQ-H activity was attained by some ancestral protist MAEL domains. However, driven by relaxed evolutionary constraint associated with functional specification, other MAEL domains generated by further lineage-specific duplications or species speciation (duplication 2) may have lost the original active site with DEDHD residues, but at the same time developed a new active site with EHHCHC residues while keeping RNase H structural scaffold (Figure 4). The diversity of characteristic residues among three Eh MAEL domains (Eh67476664, Eh67477376-C, and Eh67477376-N) in an amoeba duplication branch (node value 89%/100%) supports this evolutionary path (Figure 2A). Compared to two other paralogs (Eh67476664, Eh67477376-C) which have both sets of DEDHD and EHHCHC residues, the Eh67477376-N has lost the DnaQ-H specific residues. Thus, MAEL domains have experienced a transition from DnaQ-H active site residues to MAEL active site residues which, we believe, may represent a novel mode for protein function evolution called the active site switch.

It has been long recognized that although protein superfamilies tend to preserve their structure during evolution, a divergent evolution with functional changes is permitted [38,69-71]. Protein function changes involve diversity or variability in active sites, properties of related residues or their spatial locations, as reviewed by Todd et al. [70]. Several possible mechanisms underlying protein function changes have been proposed including evolutionary optimization via functional residue hopping, independent recruitment of active sites in different lineages, circular permutation, and functional convergence after divergence [70]. Here, MAEL domains undergoing the active site switch provide another mode for protein function change; that is, during evolution new activities can be developed by introducing new active sites based on a preexisting protein scaffold. This evolutionary mode has long been hypothesized based on many in vitro directed evolution studies [72-74]. It has been shown that new activity can be introduced by simultaneous incorporation and adjustment of functional elements through insertion, deletion, and substitution of several active site loops, followed by point mutations to fine-tune the activity [73]. A similar process may have occurred in MAEL domain evolution. In addition, the ancestral protist MAEL domains which harbor the characteristics of both DnaQ-H and MAEL domains, for the first time, illustrate the existence of an evolutionary intermediate during protein function evolution. The identification of such an evolutionary intermediate may facilitate establishing real evolutionary links between protein superfamily members with different catalytic activities, or protein superfamilies which have overall similar structural folds but different functions.

Materials and methods

See additional file 5.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DZ initiated the idea, conducted data analysis and drafted the manuscript. HX and JS were involved in Bayesian phylogenetic tree construction and protein loop modeling, respectively. VT and XX contributed to discussion and revising manuscript. All authors have read and approved the final manuscript.

Reviewers' comments

Reviewer's report 1: L Aravind, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, USA

Zhang et al show that the globular domain found C-terminal to the HMG domain in Maelstrom is a member of the 3'->5' exonuclease superfamily of the RNase H fold. This finding leads to a key functional prediction that might help in understanding the role of this major regulator of gene expression which lies at the interface between the RNAdependent process and chromatin dynamics. The basic relationship proposed here is sound; however, the authors note that the active site of this domain might have drastically been reconfigured in subset of the family with the utilization of an entirely new constellation of residues. This is a rather bold proposal based on homology modeling and the observed conservation. However, it is weakened by the fact that, as observed correctly by the authors, the canonical active site is preserved outside of the animal radiation along with the maelstrom family specific residues preserved in animals. This makes the claim suspect as it would imply that both active sites were simultaneously present in the ancestor. Hence, I strongly recommend that the authors completely rework this section and concede the strong possibility of the absence of nuclease activity in the forms lacking the canonical active site. It is quite possible that at least in animals it is an inactive RNA-binding protein.

Authors' response

Thank you for the invaluable comments. We have revised the whole paper to take into considerations these constructive criticisms. Towards the possible activity (either nuclease or RNAbinding) of MAEL with EHHCHC residues, we now only

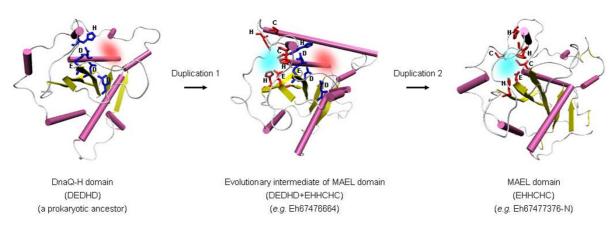


Figure 4

A proposed "active site switch" mode for MAEL domain function evolution. The cartoon drawing of protein structure is shown. The α helices are shown in pink, β sheets in yellow and loops in white. The DnaQ-H specific residues (DEDHD) are highlighted in blue, whereas MAEL specific residues (EHHCHC) are highlighted in red. Red cloud and green cloud indicate DnaQ-H active site and MAEL-specific active site, respectively.

present structural evidence and give discussion based on other evolutionary and functional information. We agree with the reviewer on the question of whether both active sites (DnaQ-H and MAEL-specific activities) can be realized in the same protist MAEL domain. We believe they could since their conservation likely reflects functional contribution; otherwise the conservation of these residues should have been lost during evolution. It seems that because of the structural conformation constraints, the same protist MAEL domain cannot form two active sites at the same time; instead, it may adopt different conformations for each of the different activities.

Further, there are several points in the paper that need to be addressed for it to be suitable for publication.

-Nomenclatural: Currently the authors use the term DNAQ-H for the superfamily. This is confusing as one could imagine that these indeed arose from DNAQ itself. Instead they should use the terminology "3'->5' exonucle-ase superfamily in the RNAse H fold".

-Phylogenetic analysis: The resolution is probably insufficient and the topology of the tree appears suspect as a result. Further some proteins are evolving very rapidly and at different rates and distort topology (e.g. Entamoeba). The tree is not critical for the argument and it is best that it is presented in the additional file.

-Introduction is too long. The authors can very briefly state the importance of Mael and its biology rather than attempting the current detailed description.

-The key functional prediction and sequence/structure analysis can be also briefly presented.

-Rename DUF1777 to something more meaningful.

In conclusion a modified version of the paper, which is suitably condensed and presents the major findings succinctly would be suitable for publication as a discovery note in Biology Direct.

Authors' response

We thank the reviewer for these suggestions. We have revised the introduction, results and discussion sections and transferred the methods section to additional file 5. We considered the proposal to put Fig 2A in the supplement but feel strongly that it should be in the main paper. We discuss MAEL evolution extensively in the main text, so the tree will be important for readers to understand MAEL evolution, especially the transition between DnaQ-H and MAEL residues. We agree that some supporting values are weak, so we only present supporting value greater than 75% for the nodes. We renamed DUF1777 as SR-25-like domain because of its similarity to SR-25 domain family (pfam: PF10500) as revealed by the SCOOP program. For terminology, we used DnaQ-H 3'-5' exonuclease family with the RNase H fold. The reason why we emphasized DnaQ-H (also called DEDDh) is to differentiate another DEDDy family of 3'-5' exonuclease with the RNase H fold, which is characterized by conserved residues DEDYD.

Reviewer's report 2: Wing-Cheong Wong, Frank Eisenhaber, Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR)

In this paper, Dapeng Zhang et. al. attempt to decipher the molecular function of the germ plasm-specific protein, Maelstrom (MAEL) which has been implicated in the piRNA pathway and also in chromatin remodeling from previous experimental studies. The authors conjectured that Maelstrom has nuclease-like activity from three main findings: Firstly, the novel MAEL-specific domain is defined by a set similar sequence segments (related via a few PSI-BLAST searches) with a conserved motif involving residues (Glu-His-His-Cys-His-Cys) from mostly metazoans (except of fish species) and some protists. Some of these protist MAEL sequences also contain the DnaQ-H specific site (Asp-Glu-Asp-His-Asp) that exhibits a 3-5' exonuclease catalytic activity. Therefore, it seems likely that the metazoan MAEL proteins had inherited nucleaselike activity from their protist MAEL ancestors. Secondly, domain architecture analysis of MAEL-related proteins showed the association of the MAEL domain with the HMG (SMART: SM00398), DUF1777 (PFAM: PF08648) and HDAC_interact (SMART: SM00761) for DNA binding, RNA binding and chromatin remodeling respectively. Finally, structural modeling showed that the MAEL-specific domain (Glu-His-His-Cys-His-Cys) in metazoan is able to form a structural core despite the lack of the DnaQ-H active residues. The authors also argued that the residues His, Cys and Glu are the most frequently residues capable of interacting with Zn2+ and also utilized by ColE7 endonuclease, Zinc transport protein ZnuA and Aldolase; analogous to metal ion-binding DnaQ-H.

There are several critical points with this manuscript:

(1) The sequence segment family collection of homologous Maelstrom protein sequences is incomplete. Using the fan-like search methodology as described in Schneider et al. BMC Bioinformatics 2006 v.7, 164), more MAELlike sequences including sequences from *Danio Rerio* (e.g. A2CF13_DANRE, EXOD1_DANRE, EXOD1_DANRE/ Q502M8), oxidoreductases, DNA polymerases III and 3– 5' exonucleases (e.g. Q503G0_DANRE, THEX1_HUMAN;1W0H:A) from numerous species can be found. Thus, there are homologs among fish species.

Authors' response

We thank you for your insights and suggestions. We also appreciate your attempts at retrieving additional sequences using your novel methodology. The sequences you identified all are DnaQ-H domains and some sequences you mentioned like 1W0H:A have been included in our study as representatives of the DnaQ-H domain. As we mentioned in the main text, PSI-BLAST searching with a profile inclusion E value of 0.05 can retrieve several DnaQ-H exonucleases as significant hits. However, they are not included in our initial sequence analysis for MAEL domains since they do not have MAEL specific residues (EHHCHC), and introducing these sequences may dilute conserved characteristics of MAEL domains. We used an E value of 0.005 for PSI-BLAST searches with different MAEL domains as queries. They all retrieved the same set of MAEL domain sequences. We also tried the HHsenser server, another sensitive sequence searching program, which retrieved similar results. We could not detect any fish MAEL domain from protein, nucleotide/EST or even Ensembl genome databases of five fish species. So, we are proposing that the MAEL domain is lost in fish species according to these observations. This discovery should be very interesting to experimental biologists who are working on the piRNA pathway. We agree that other distant homologs exist in fish, such as DnaQ-H members and other RNase H members (like PIWI).

Reviewers' response

The emphasis on a set of conserved positions (yet without a clear functional role) does not make the definition of a domain. Most importantly, the notion of globular domains unifies protein sequence segments having similarity of their fold (and, as a consequence, in their hydrophobic pattern). Besides understanding the types of protein families that are in the vicinity of the starting sequence, the purpose of performing fan-like search is also to determine if the search space of the starting sequence for its orthologous sequences is well sampled. When sequences are been collected, the relationship of orthology or paralogy is not obvious. But eventually, with sufficient sequence collection, sequences from different taxonomic groups will be able to form distinct group of protein families. Finally, with reference to these neighboring protein families, one can then use clustering or phylogenetic methods to determine the orthology coverage of the starting sequence. This has not been done in the work of the authors. We have carried out a full sequence family collection with a fan-like PSI-BLAST search (inclusion value for score matrix of ≤ 0.001 ; e-value for PSI-blast initialization < 0.06), aligned the family and created a phylogenetic tree from hits (with the group of exonucleases represented by the structure 1Y97 as outgroup, see attachment). It looks as if the so-called maelstrom group is surrounded by the bloom syndrome proteins (DNA helicases), DEAD-domain containing RNA helicases followed by bacterial nucleases as next hits. The fish sequences mentioned by us are in the neighboring helicase groups and, apparently, are not nucleases.

Authors' response

We have conducted profile-profile alignments between MAEL, DnaQ (Exonuc X-T, Pfam: PF00929), and DEAD helicase (including bloom syndrome proteins, Pfam: PF00270) domains using the logomat-p program (additional file 6). In contrast to detectable similarity between MAEL and DnaQ domains, no global similarity between MAEL and DEAD helicase can be identified. The similarity between MAEL and DEAD helicase can be identified. The similarity between MAEL and DnaQ domains is shown for the first 100 amino acid segment, also seen in the PSI-BLAST results. The reason why the second half segment does not appear to be homologous is that conserved residues are different (CHC in MAEL and HD inDnaQ) and that no structural fold considerations were made. Therefore, the evolutionary tree inferred from unrelated sequences is not reliable. We do not agree with the assessment of the reviewers. (2) An exhaustive search for homologous sequences across all species is the foremost important task in function annotation transfer via homology. This exhaustive list of the homologous sequences enables one to construct clusters of orthologous and paralogous genes and to group them in a phylogenetic tree. Among orthologous sequences, function annotation transfer is able to hold well especially for one-to-one orthologs, with decreased confidence at greater evolutionary distances. On the opposite end, paralogous sequences are generalized to be functional diversified and specialized. This makes the task of function annotation transfer more complex (see Koonin, 2005, Annu. Rev. Genet., 39, 309-338). In this paper, the exact homology relationships among the collected sequences were not well established and, thus, function annotation transfer in this context is problematic. It appears to us that the exonucleases are in another branch of the tree compared with maelstrom sequences; thus, the predicted function might not be correct.

Authors' response

This paper presents an evolutionary relationship between MAEL domains and DnaQ-H domains with the RNase H fold based on structural fold similarity as well as the evidence that protist MAEL domains have DnaQ-H specific residues. We do agree that a direct function annotation transfer may not be guaranteed based on this evolutionary link because of functional divergences during protein evolution. But considering the general functions in nuclease activities of DnaQ-H family as well as its distantly related RNase-H superfamily members, we predicted that MAEL may have a similar function with either nuclease or RNA-binding activity. We provide a preliminary evolutionary tree between DnaQ-H and MAEL domains in the additional file 4 and combined this with Figure 2A for extensive discussion in the last section. We hope this discovery will facilitate the further investigation on MAEL function.

Reviewers' response

We think that the conclusion about the functional relationship to the DnaQ-H domain is premature in this form. A hit with 3D-jury is, at best indicative. Our family search and the resulting phylogenetic tree (see attachment) bring the maelstrom group equally close to various helicases and nucleases. This more stringent homology search results (inclusion value for score matrix of ≤ 0.001 ; e-value for PSI-blast initialization < 0.06) revealed that the Maelstrom sequences are in close vicinity to a group of Bloom syndrome proteins (belonging to the DNA helicase family), bacteria nucleases and helicases while the exonucleases were not significant enough to be found (consistent with authors' PSI-blast results of insignificant p-value for the exonucleases). A preliminary phylogenetic study (with exonuclease as the out-group) showed that the Maelstrom sequences are most homologous to the Bloom syndrome protein sequences in comparison to the other sequences. DnaQ-H is by far not the closest functionally characterized neighbors. In the absence of further structural and catalytic information of the MAEL motif (Glu-His-His-Cys-His-Cys), the functional evolution relationship between Maelstrom and exonuclease is still unclear except for a potential similarity of fold.

If you do not have an own resource for correct family collection, we strongly suggest the authors to use protein family searcher like HHsenser <u>http://toolkit.tuebingen.mpg.de/hhsenser</u> to collect more homologous sequences to clarify the relationship of Maelstrom to its adjacent protein families.

Authors' response

As we indicated previously, no sequence similarity can be detected between MAEL and DEAD helicase domains. We tried HHsenser to retrieve MAEL homologues sequences, and it generated similar results as PSI-BLAST. We thank reviewers for this suggestion.

(3) Furthermore, the suggestion of nuclease-like activity in metazoan MAEL proteins is weak given that the DnaQ-H active residues were not conserved even if the predicted tertiary structure is correct and, probably, conserved in the family. A structural is only a plausibility argument; it does not prove the conclusion. Doubts are the more appropriate since the homology model involves a translocation/ shifting of the active site.

At the end, a set of sequentially similar sequence segments without any trustworthy molecular function prediction remains. This result is not necessarily demanding another publication.

Authors' response

We present general discoveries about MAEL, its evolutionary link to DnaQ-H domains and structural predictions on active sites. We agree that functional prediction is not the definitive conclusion. However, we believe that our rigorous analysis may give us a strong basis to hypothesize on function. Firstly, the evolutionary link and possible DnaQ-H active site in protist MAEL domains may suggest that protist MAEL domains have a 3'-5' exonuclease activity. Secondly, for the MAEL domains with EHHCHC residues, the high conservation of these residues likely reflects their functional contributions. Structural examinations direct our attention to an active site since these conserved EHHCHC residues are located closely together. We then found other evidence including the property of E, H and C residues to interact with metal ions and general functions of evolutionarily related RNase-H fold families. Although we do not have experimental support, these lines of evidence provide structural, chemical and evolutionary basis for an active site, and thus lead to our hypothesis that it may have nuclease activity or RNA-binding ability. Thirdly, translocation/shifting of the active site is common in evolution of protein families as reviewed by Todd et al. (2002) and Anantharaman et al. (2003). It is also true for the RNase H fold superfamily in which the DnaQ-H family and other families use their own specific residues to form different active sites. Therefore, we believe that the DnaQ-H active site is lost during MAEL evolution and the MAEL domain developed its own active residues. More importantly, we identified some protist domains which have both sets of active residues of DnaQ-H and MAEL domains. They can serve as an evolutionary intermediate during this translocation/shifting, thus suggesting a new mode for protein function evolution.

Reviewers' response

Firstly, the authors utilized the 3D-jury results to indicate that the maelstrom protein segment might confer a similar fold to that of DnaQ-H domain exemplified by pdb 1W0H:A. It appears to us that the evolutionary distance to these exonucleases is considerable and that other groups are much more closely related. We found the Maelstrom sequences to be most homologous to the Bloom syndrome protein sequences. Therefore, the structural fold prediction might not be reliable and, at such evolutionary distances, it would be not surprising if the relative positions of important residues are scrambled. Furthermore, the metazoan Maelstrom proteins have lost these residues that appear indispensable for the nuclease activity. Unless it can proven experimentally, the suggestion that metazoan Maelstroms have nuclease activity seems less plausible, especially given the presence of a more closely related group of Bloom syndrome proteins.

Secondly, Anantharaman et al. state that the presence of a characteristic set of conserved active residues is important for the identification of enzymes in sequence analysis. The set of conserved active residues are typically derived from known set of sequences and structures of related enzymes. For those proteins with preserved structure but varying catalytic residues, the detection of evolutionary relationship is far more difficult. In the case of the Maelstrom, the structure is purely hypothetical and the MAEL motif (Glu-His-His-Cys-His-Cys) has yet to show nuclease-like activity. Therefore, to say that a translocation or shifting of the active nucleatic site has occurred in the Maelstrom in the course of its evolution simply cannot be proven at this point without further experimentation or other type of compelling information. Thus, the molecular function of the maelstrom domain remains unclear and the current stage of research does not justify a report; otherwise, any additional branch of the phylogenetic tree would deserve another article.

Authors' response

Firstly, DEAD Helicase domains belong to the P-loop containing nucleoside triphosphate hydrolases fold, whereas DnaQ-H domains belong to the RNase H fold. It is not possible that a reliable searching with MAEL sequences can retrieve both DnaQ-H domains and DEAD helicase domains. Secondly, since no similarity exists between the MAEL/DnaQ and DEAD domains, it is not reasonable to align them together and infer their evolutionary history. Thirdly, we agree with the reviewer that similar structural fold alone does not provide sufficient evidence of common ancestry [75]. However, significant sequence conservation, structural resemblance and catalytic residue conservation may strongly indicate evolutionary relationship [71,75]. In our study, the proposed evolutionary relationship is established on the basis of three lines of evidence: 1, sequence similarity via PSI-BLAST, which provides the most straightforward evidence of homology [75]; 2, similar structural fold; 3, ancestral protist MAEL domains have DnaQ-H characteristic residues. We thank the reviewers for their efforts.

Additional material

Additional File 1

A complete multiple sequence alignment of MAEL domains. The domain sequences are represented by an abbreviation of species name followed by database ID and domain regions. The consensus in 75% of the sequences is shown below the alignment based on default amino acid classes in Chroma. The numbers in bracket are indicative of the excluded residues from sequences. Species name abbreviations refer to Figure 2 legend.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S1.pdf]

Additional File 2

Structural alignments between five different DnaQ-H domains showing the plasticity of structural conformations of α 5 and α 6. (A) The cartoon structures of four DnaQ-H domains are shown with different colors, in which 2P1J is colored with red. For the structure of 1J53, the NewCartoon diagram is shown with α helices in pink and β sheets in yellow. (B) The structural alignment of 1J53 and 2P1J. The structure of 2P1J is colored in red whereas for 1J53, its α helices are colored in pink and β sheets in yellow. (C) The structural locations of active site residues in both 1J53 and 2P1J domains.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S2.ppt]

Additional File 3

The conservation of MAEL-specific residues in E. coli DnaQ-H domain (1J53:A). The cartoon drawing of protein structure is shown. The α helices are shown in pink, β sheets in yellow and loops in cyan. The DnaQ-H specific residues (DEDHD) are highlighted with acidic residues (D and E) in light red and basic His in light blue, whereas three MAELspecific residues (EHH) are highlighted with acidic Glu in red and basic His in blue.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S3.ppt]

Additional File 4

Evolutionary relationship between DnaQ-H and MAEL domains. Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S4.ppt]

Additional File 5

Materials and methods. Detailed description on materials and methods used in the present study. Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S5.pdf]

Additional File 6

Profile-profile alignment among MAEL, DnaQ (Exonuc X-T, Pfam: PF00929), and DEAD (Pfam: PF00270) domains by the logomat-p program.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1745-6150-3-48-S6.doc]

Acknowledgements

We would like to thank Jimin Pei in University of Texas Southwestern Medical Center for helpful discussion and suggestions. We are also grateful to Lingyan Jiang in SUNY Upstate Medical University for critically reading of the manuscript and Martin Smith in Laval University for his suggestion in sequence searching. This work was supported by the University of Ottawa International Scholarship program (to DZ) and the NSERC Discovery program (to VLT).

References

- Kotaja N, Sassone-Corsi P: The chromatoid body: a germ-cellspecific RNA-processing centre. Nat Rev Mol Cell Biol 2007, 8(1):85-90.
- 2. Ikenishi K: Germ plasm in Caenorhabditis elegans, Drosophila and Xenopus. *Dev Growth Differ* 1998, 40(1):1-10.
- Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T: The small RNA profile during Drosophila melanogaster development. Dev Cell 2003, 5(2):337-350.
- Vagín VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD: A distinct small RNA pathway silences selfish genetic elements in the germline. Science 2006, 313(5785):320-324.
- 5. Hartig JV, Tomari Y, Forstemann K: piRNAs-the ancient hunters of genome invaders. Genes Dev 2007, 21(14):1707-1713.
- Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE: Characterization of the piRNA complex from rat testes. Science 2006, 313(5785):363-367.
- Aravin AA, Hannon GJ, Brennecke J: The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 2007, 318(5851):761-764.
- 8. Klattenhoff C, Theurkauf W: Biogenesis and germline functions of piRNAs. Development 2008, 135(1):3-9.
- Carmell MA, Girard A, Kant HJ van de, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ: MIW12 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 2007, 12(4):503-514.
- Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, Elst H van den, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RH, Hannon GJ, Draper BW, Ketting RF: A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 2007, 129(1):69-82.
- 11. Lim AK, Kai T: Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci USA 2007, 104(16):6714-6719.

- 12. Clegg NJ, Frost DM, Larkin MK, Subrahmanyan L, Bryant Z, Ruohola-Baker H: maelstrom is required for an early step in the establishment of Drosophila oocyte polarity: posterior localization of grk mRNA. Development 1997, 124(22):4661-4671.
- Cook HA, Koppetsch BS, Wu J, Theurkauf WE: The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* 2004, 116(6):817-829.
- Jones JR, Macdonald PM: Oskar controls morphology of polar granules and nuclear bodies in Drosophila. Development 2007, 134(2):233-236.
- Chuma S, Hosokawa M, Kitamura K, Kasai S, Fujioka M, Hiyoshi M, Takamune K, Noce T, Nakatsuji N: Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. Proc Natl Acad Sci USA 2006, 103(43):15894-15899.
- Chen Y, Pane A, Schupbach T: Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Curr Biol 2007, 17(7):637-642.
- Weidinger G, Stebler J, Slanchev K, Dumstrei K, Wise C, Lovell-Badge R, Thisse C, Thisse B, Raz E: dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. *Curr Biol* 2003, 13(16):1429-1434.
- Pane A, Wehr K, Schupbach T: zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell 2007, 12(6):851-862.
- Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC: Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. RNA 2007, 13(11):1911-1922.
- Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA: Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res 2007, 35(16):5430-5438.
- Findley SD, Tamanaha M, Clegg NJ, Ruohola-Baker H: Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development 2003, 130(5):859-871.
- Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L: The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol 2003, 10(12):1026-1032.
- Song JJ, Smith SK, Hannon GJ, Joshua-Tor L: Crystal structure of Argonaute and its implications for RISC slicer activity. Science 2004, 305(5689):1434-1437.
- Parker JS, Roe SM, Barford D: Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J 2004, 23(24):4727-4737.
- Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S: Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell 2006, 125(2):287-300.
- Maris C, Dominguez C, Allain FH: The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 2005, 272(9):2118-2131.
- 27. Ponting CP: Tudor domains in proteins that interact with RNA. Trends Biochem Sci 1997, 22(2):51-52.
- Clegg NJ, Findley SD, Mahowald AP, Ruohola-Baker H: Maelstrom is required to position the MTOC in stage 2-6 Drosophila oocytes. Dev Genes Evol 2001, 211(1):44-48.
- Soper SF, Heijden GW van der, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A: Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev Cell 2008, 15(2):285-297.
- Costa Y, Speed RM, Gautier P, Semple CA, Maratou K, Turner JM, Cooke HJ: Mouse MAELSTROM: the link between meiotic silencing of unsynapsed chromatin and microRNA pathway? Hum Mol Genet 2006, 15(15):2324-2334.
- Robert VJ, Sijen T, van Wolfswinkel J, Plasterk RH: Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. Genes Dev 2005, 19(7):782-787.
- 32. Bianchi ME, Agresti A: **HMG proteins: dynamic players in gene** regulation and differentiation. *Curr Opin Genet Dev* 2005, 15(5):496-506.
- 33. Hoegg S, Meyer A: Hox clusters as models for vertebrate genome evolution. *Trends Genet* 2005, **21(8):**421-424.

- Khochbin S, Verdel A, Lemercier C, Seigneurin-Berny D: Functional significance of histone deacetylase diversity. Curr Opin Genet Dev 2001, 11(2):162-166.
- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: Detecting protein function and protein-protein interactions from genome sequences. Science 1999, 285(5428):751-753.
 Blanton S, Science A, D. Science 1999, 1999.
- Blanton S, Srinivasan A, Rymond BC: PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. Mol Cell Biol 1992, 12(9):3939-3947.
- Bateman A, Finn RD: SCOOP: a simple method for identification of novel protein superfamily relationships. *Bioinformatics* 2007, 23(7):809-814.
- Orengo CÁ, Thornton JM: Protein families and their evolutiona structural perspective. Annu Rev Biochem 2005, 74:867-900.
- Ginalski K, Elofsson A, Fischer D, Rychlewski L: 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003, 19(8):1015-1018.
- Ginalski K, Rychlewski L: Detection of reliable and unexpected protein fold predictions using 3D-Jury. Nucleic Acids Res 2003, 31(13):3291-3292.
- Zuo Y, Deutscher MP: Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res 2001, 29(5):1017-1026.
- Thore S, Mauxion F, Seraphin B, Suck D: X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. EMBO Rep 2003, 4(12):1150-1155.
- Guda C, Lu S, Scheeff ED, Bourne PE, Shindyalov IN: CE-MC: a multiple protein structure alignment server. Nucleic Acids Res 2004:W100-103.
- Siepen JA, Radford SE, Westhead DR: Beta edge strands in protein structure prediction and aggregation. Protein Sci 2003, 12(10):2348-2359.
- Perrino FW, Harvey S, McMillin S, Hollis T: The human TREX2 3' -> 5'-exonuclease structure suggests a mechanism for efficient nonprocessive DNA catalysis. J Biol Chem 2005, 280(15):15212-15218.
- Cheng Ý, Patel DJ: Crystallographic structure of the nuclease domain of 3'hExo, a DEDDh family member, bound to rAMP. J Mol Biol 2004, 343(2):305-312.
- 47. Hamdan S, Carr PD, Brown SE, Ollis DL, Dixon NE: Structural basis for proofreading during replication of the Escherichia coli chromosome. *Structure* 2002, 10(4):535-546.
- Ferre F, Clote P: DiANNA I.I: an extension of the DiANNA web server for ternary cysteine classification. Nucleic Acids Res 2006:W182-185.
- Aravind L, Makarova KS, Koonin EV: SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res 2000, 28(18):3417-3432.
- 50. Rice P, Mizuuchi K: Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* 1995, **82(2)**:209-220.
- Ariyoshi M, Vassylyev DG, Iwasaki H, Nakamura H, Shinagawa H, Morikawa K: Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from E. coli. Cell 1994, 78(6):1063-1072.
- Katayanagi K, Okumura M, Morikawa K: Crystal structure of Escherichia coli RNase HI in complex with Mg2+ at 2.8 A resolution: proof for a single Mg(2+)-binding site. Proteins 1993, 17(4):337-346.
- Kennedy S, Wang D, Ruvkun G: A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature* 2004, 427(6975):645-649.
- Ding J, Das K, Moereels H, Koymans L, Andries K, Janssen PA, Hughes SH, Arnold E: Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. Nat Struct Biol 1995, 2(5):407-415.
- 55. Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L: Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 2005, 12(4):340-349.
- Parker JS, Roe SM, Barford D: Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature 2005, 434(7033):663-666.

- 57. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ: Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature 2005, 434(7033):666-670.
- Tamames B, Sousa SF, Tamames J, Fernandes PA, Ramos MJ: Analysis of zinc-ligand bond lengths in metalloproteins: trends and patterns. Proteins 2007, 69(3):466-475.
 Doudeva LG, Huang H, Hsia KC, Shi Z, Li CL, Shen Y, Cheng YS, Yuan
- Doudeva LG, Huang H, Hsia KĆ, Shi Z, Li CL, Shen Y, Cheng YS, Yuan HS: Crystal structural analysis and metal-dependent stability and activity studies of the ColE7 endonuclease domain in complex with DNA/Zn2+ or inhibitor/Ni2+. Protein Sci 2006, 15(2):269-280.
- Li H, Jogl G: Crystal structure of the zinc-binding transport protein ZnuA from Escherichia coli reveals an unexpected variation in metal coordination. J Mol Biol 2007, 368(5):1358-1366.
- Blom NS, Tetreault S, Coulombe R, Sygusch J: Novel active site in Escherichia coli fructose 1,6-bisphosphate aldolase. Nat Struct Biol 1996, 3(10):856-862.
- Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC: Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 2006, 127(4):747-757.
- Wagner A: Birth and death of duplicated genes in completely sequenced eukaryotes. Trends Genet 2001, 17(5):237-239.
 Panopoulou G, Poustka AJ: Timing and mechanism of ancient
- Panopoulou G, Poustka AJ: Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis. Trends Genet 2005, 21 (10):559-567.
- Evans BJ, Kelley DB, Melnick DJ, Cannatella DC: Evolution of RAG-I in polyploid clawed frogs. Mol Biol Evol 2005, 22(5):1193-1207.
- Zhang D, Martyniuk CJ, Trudeau VL: SANTA domain: a novel conserved protein module in Eukaryota with potential involvement in chromatin regulation. Bioinformatics 2006, 22(20):2459-2462.
- 67. Zhang D, Trudeau VL: The XS domain of a plant specific SGS3 protein adopts a unique RNA recognition motif (RRM) fold. *Cell Cycle* 2008, **7(14):**2268-2270.
- Eichinger L, Noegel AA: Comparative genomics of Dictyostelium discoideum and Entamoeba histolytica. Curr Opin Microbiol 2005, 8(5):606-611.
- Glasner ME, Gerlt JA, Babbitt PC: Evolution of enzyme superfamilies. Curr Opin Chem Biol 2006, 10(5):492-497.
- Todd AE, Orengo CA, Thornton JM: Plasticity of enzyme active sites. Trends Biochem Sci 2002, 27(8):419-426.
- Anantharaman V, Aravind L, Koonin EV: Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* 2003, 7(1):12-20.
- Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS: The 'evolvability' of promiscuous protein functions. Nat Genet 2005, 37(1):73-76.
- Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS: Design and evolution of new catalytic activity with an existing protein scaffold. Science 2006, 311(5760):535-538.
- Patrick WM, Matsumura I: A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J Mol Biol* 2008, 377(2):323-336.
- Kinch LN, Grishin NV: Evolution of protein structures and functions. Curr Opin Struct Biol 2002, 12(3):400-408.

