Supporting information

Trypanosoma brucei harbors a divergent XPB helicase paralog that is specialized in nucleotide excision repair and conserved among kinetoplastid organisms

Nitika Badjatia,[1] Tu N Nguyen,[1] Ju Huck Lee,[1,2] and Arthur Günzl [1,#]

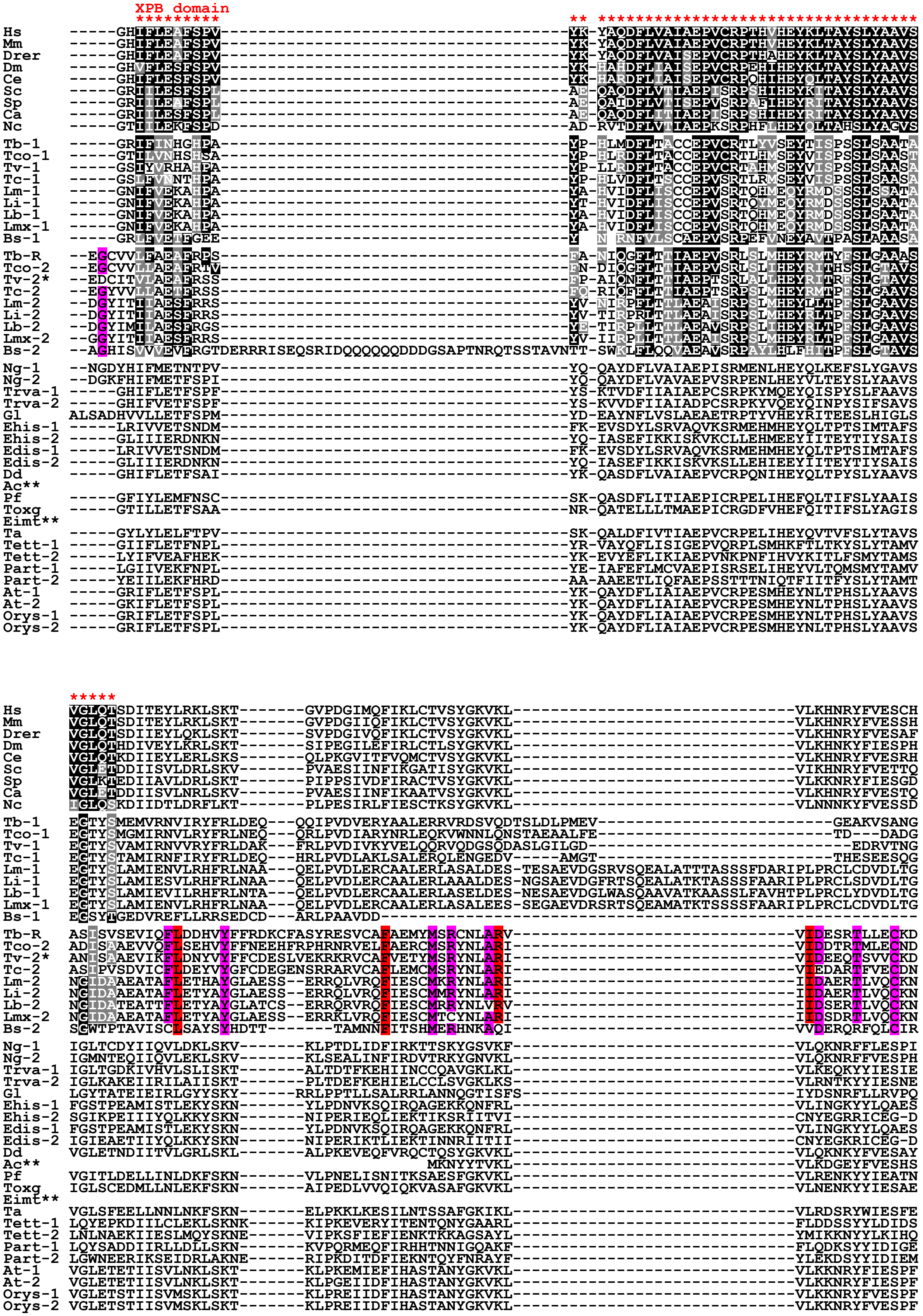
[1] Department of Genetics and Developmental Biology, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030-6403, USA

[2] Current address: Department of Microbiology & Molecular Genetics, University of Texas-Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA

[#] For correspondence, Email: gunzl@uchc.edu; Tel: (860) 679-8878; Fax: (860) 679-8345.

Content

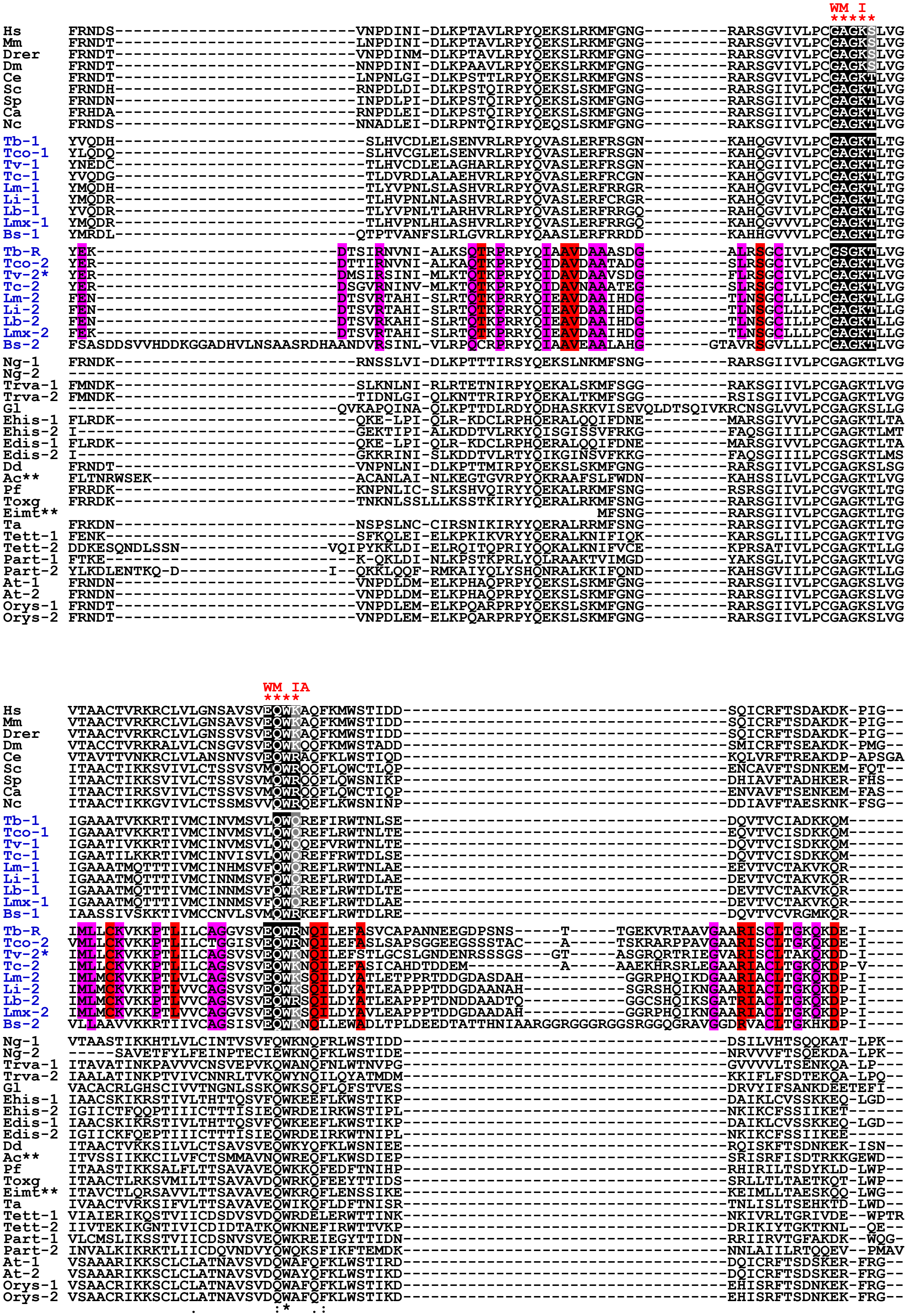
```
CLUSTAL O(1.1.0) multiple sequence alignment

Hs                            MGKRD-----RADRDKKKSRKRHYEDE----------------E-DDEEDAPG--------------
Mm                            MGKRD-----RVDRDKKKSKKRQYEEE----------------E-EDEDDIPG--------------
Drer                          MGRKD-----KSDREKK-SKKRYYEDE----------------E--EDEEVIG--------------
Dm                            MGPPKKSRKDRSGGDKFG-KKRRAEDEAFTQ------------L-VDDNDSLD--------------
Ce                            MATKERKRR-----GKWD----TYKA-----------------E-EAPSLYSG-NA-----------
Sc     MTDVEGYQSKSKGKIFPDMGESFF----SSDEDSP------ATDAEI-DENYDDNRETL---EGRGE-RDTGAMVT-GL--------KKP
Sp                            M----SLKRKNN------AR-EGTPDEDLEEYSDY-SDVDNYGE-EDDDSYKP-AP-----------
Ca                            M----SSRNREP------VNYADLEDENFSDEDVYVSNKTGAKD-NDDDEYNT-PS--------QKS
Nc           MPPKRKAVGAPQAGVAKAGRTSALSTPGP------ATPRSLE-------SSFQS--DGDDD-LNDDDVDE-VI--------KRK
Tb-1
Tco-1
Tv-1
Tc-1
Lm-1
Li-1
Lb-1
Lmx-1
Bs-1
Tb-R
Tco-2
Tv-2*
Tc-2
Lm-2
Li-2
Lb-2
Lmx-2
Bs-2
Ng-1                          MNNS--Q----------------QEETNQTQHS------------NNPNPTNEKNE--------KKR
Ng-2                         MLNNH--L----------------HDQTTG-----------------TSSNRLQSED--------KKR
Trva-1                                          MS-----DAEYHEN---------------LKN-LDVGH-----SRQNRK
Trva-2                                          MT-----EEELDDF---------------KEEQTTQNL-----VKKKRT
Gl
Ehis-1
Ehis-2
Edis-1
Edis-2
Dd                            MSSGDSNL-----KRRRGGNT-GQSSKSYNTWTDYEE------D-LEE-------------------
Ac**
Pf                                             MMN-SLKDSSNYSPNLSHTRGSRKRSD-IFNENYYAKNLKRRKLKAAKKL
Toxg             MEPGAFPCSGTFSSEARSPFRASSPPAAKRFRR-SEPDDEDWDPDTCLYEASQADRESRASRTWPAAQVPRVAPPVALRV
Eimt**
Ta
Tett-1                                                           MA---------DKKVNPKSR------EQSKRL
Tett-2                        MSV----------------------------PVQYIA---------DYSNTPEA-----------RL
Part-1
Part-2
At-1                            MGNGERGRPNKKMKYGGKD-DQKMKNIQNAEDYYD------D-ADE-------------------
At-2                            MGNDERKRPTKKMKYGGKD-DQKMKNIQNVEDYYD------D-ADE-------------------
Orys-1                        MAGGDGDRARAPKRHKSSA---PSRSIDETAELDYTD------D-VDD-------------------
Orys-2                        MAGGDGDRARAPKRHKSSA---PSRSIDETAELDYTD------D-VDD-------------------


Hs     -----------NDPQEAVPS-----------------AAGKQ---VDESGTK-VDEY-----GAKDYRLQMPLKDDHTSRPLWVAP-D--
Mm     -----------NESQEAVPS-----------------AAGKQ---VDESSTK-VDEY-----GAKDYRQQMPLKGDHTSRPLWVAP-D--
Drer   -----------GESQEAVPA-----------------AAGKQ---VDESSTK-LDEY-----GAKDYRLQMLLKNDHSSRPLWVAP-D--
Dm     -----------ATESEGIPG-----------------AASKN-AETNDEQIN-TDEY-----GAKDYRSQMQLRPDHGNRPLWVAP-N--
Ce     -----------DKETSSVPK-----------------AASHNLNGENASSVM-TDEF-----GAKDYRKDMPLKGDFTARPLWVAP-D--
Sc     RKKT-K-SSRHT----AADS---SMNQMDAKDKALLQDTNSDIPADFVPDSV-SGMF-----RSHDFSY-LRLRPDHASRPLWISPSD--
Sp     RIRI---NNNKTKAQT------------------------TTNSNEARQSGI-SAMF-----GQNDFSNLLGLKLDHTARPLWINPID--
Ca     RKST-GNNSKKRKAQESVDSLKQNLDETNYS-DNELIELTPDVPADYIPDAV-SKNF-----GKGDFSY-LKLKPDHFSRPIWISPND--
Nc     EE------VIAREADEFVNTWA--------VDSSRFQA--RQDGGGSHRDGA-TQFFG---SGKRDFSY-LNLKPDHDQKPLWIDPEK--
Tb-1                                                                                    MILGP-G--
Tco-1                                                                                   MILGP-G--
Tv-1                                                                                    MIVGQ-G--
Tc-1                                                                                    MIVGA-N--
Lm-1                                                                                    MNIGD-S--
Li-1                                                                                    MNIGD-S--
Lb-1                                                                                    MNIGD-S--
Lmx-1                                                                                   MNIGD-S--
Bs-1                                                                                    MFLDR-D--
Tb-R                                                                           MQCLFLEERIK--------K--
Tco-2                                                                          MQCLFLEERTA--------K--
Tv-2*                                                                        MPMQCLFLRERTA--------T--
Tc-2                                                                           MQCLFLEERVS--------A--
Lm-2                                  MAH-------------------SL----------SGDARTGEKCLFVESRIE--------S--
Li-2                                  MAH-------------------SL----------SGDARAGEECLFVESRIE--------S--
Lb-2                                  MAY-------------------SL----------NGDALIGEKCLFVESRAQ--------S--
Lmx-2                                 MTH-------------------SL----------SGDARTDEKCLFVESRIE--------S--
Bs-2                                                                       MQQGECCFFLVDCTDKR------L--
Ng-1   RKSETKSKNKKTKYDN------------------IENDTDSQISK-QAEKLA-TDIFKQKEGIQFDH-SSMDLKSDHESRPIWVCSGD--
Ng-2   KKKEKEAKKKKQKYEE------------------ILNDLDNEVAKKAQEKLV-SDIFSQKKDVQFDH-SNLELKPDHKLRPVWVCSGT--
Trva-1 RKSEN-----KIKLEN--------------------ITNEYFNGETSARS------------EVSKSFDDLTFIPGSENRPAIVFP-D--
Trva-2 RSHAP-----KQKLLQ--------------------LEANYFDGETRARS------------NKYKYFKDLYILENSDNRPAIVMP-D--
Gl     ------------------------------------------------------------MSSKDAPLSFEDLKPDHHKRPLSVFPYSTN
Ehis-1 ------------------------MHKK------------RDTSDDSS-----DDSLE---DMKESTDYDMRLKPNHPELPMWVSS-N--
Ehis-2 --------------------------------------------------MK-STIE----N---DDSEEFKLQIINTSNPYWICP-N--
Edis-1 ------------------------MHKK------------RDTSDDSS-----EDSLE---DMKESTDYDMRLKPNHPELPMWVSS-N--
Edis-2 --------------------------------------------------MK-SNIE----NNNDEDLEEYNVEIINTSNPYWVCP-N--
Dd     --------------------------------SGEFNQSIKKTTNTSSATLT-SSEEK---GSLLDYSKRCILKQDNKSRPIWVCP-D--
Ac**
Pf     KDSTNGEFTKKVK-KQ----------LKDY---YEM----RFDKKVINLPFS-TDSINIQQRGFHDYSKDMKLKKNHMNKPLWICS-D--
Toxg   R--EEGMAEKSAGKTK----------DARP---SET----AQGGVYGSLPLT-ADSLSLGVGGFRDFGSKLALKVDHAHRPLWVCP-D--
Eimt**
Ta     ---------------------------MSK---KSI----LVENTWGSLPFT-IDGSEQWIRNFRDYS-NLKLKTNHTARPLWVCP-D--
Tett-1 KS-------NSQKQNQN--G---------KQQKNEEGDDGLFGQKQGKAQVEGN-----QFQQFYDYS-TLELKDDYRERPILICP-D--
Tett-2 KK-------NHQKY------------------------------------------CLYLKPADRDYSSSFEI----QENRGFFPE-T--
Part-1 ---------------------------------------MSFEVKKRKLQTQPPEVQKLLIGDNMDYR-NIEI--VQSNKPLILSP-D--
Part-2 ------------------------------------------------MRIDEPQ--------N-----EQQLKK-TKNSNDTQWK-K--
At-1   --------------------------------DSR------------------DGEGE---EKKRDF-TKLELKPDHGNRPLWACA-D--
At-2   --------------------------------DSR------------------DGEGE---EKRRDF-TDLELKPDHGNRPLWACA-D--
Orys-1 --------------------------------DVR------------------DADRE---VKKRDF-TKLELKPDHANRPLWACA-D--
Orys-2 --------------------------------DVR------------------DADRE---VKKRDF-TKLELKPDHANRPLWACA-D--
```



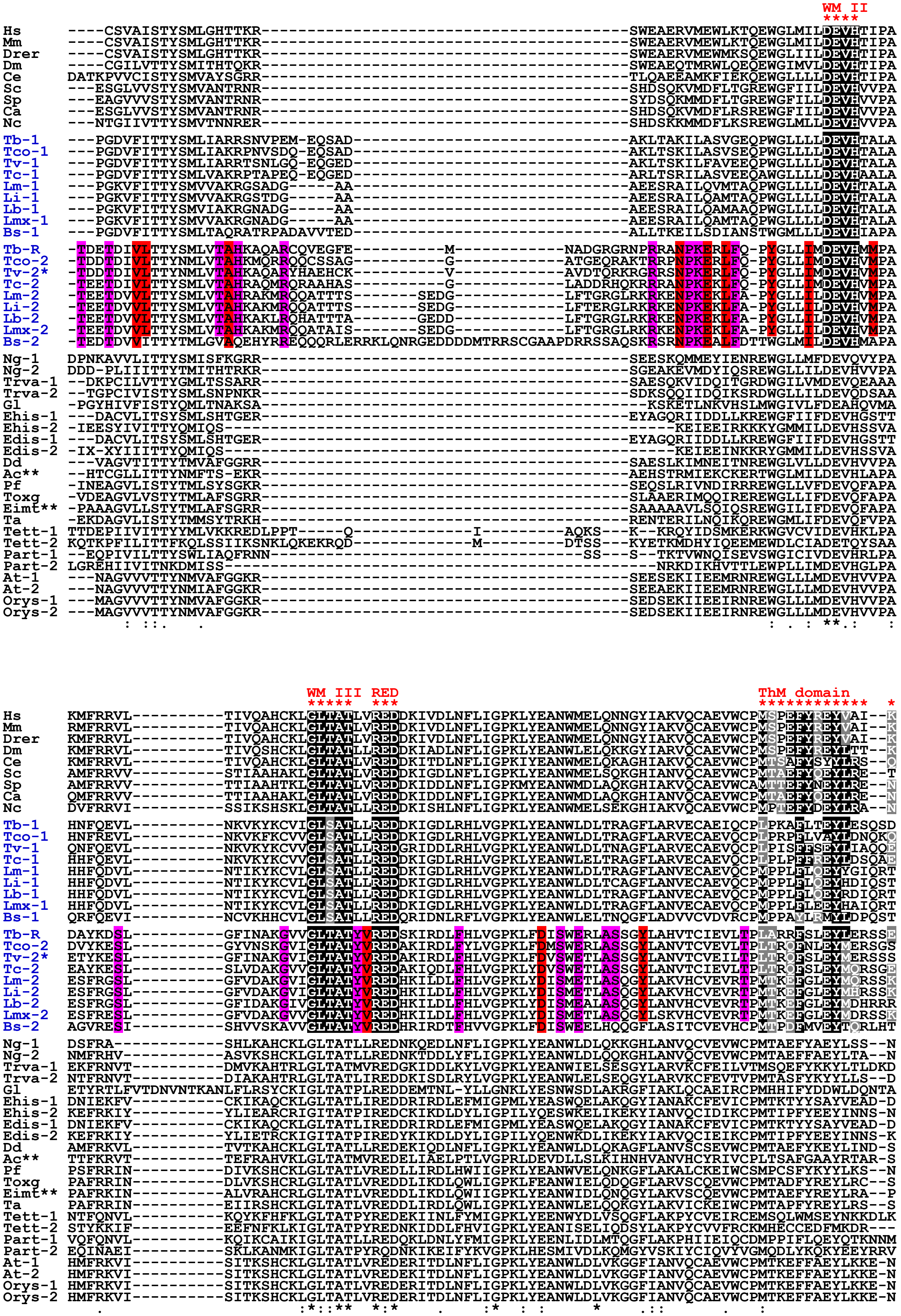
```
              XPB domain
              *********                                     ** **********************************
Hs     -----GHIFLEAFSPV-------------------------------------YK-YAQDFLVAIAEPVCRPTHVHEYKLTAYSLYAAVS
Mm     -----GHIFLEAFSPV-------------------------------------YK-YAQDFLVAIAEPVCRPTHVHEYKLTAYSLYAAVS
Drer   -----GHIFLEAFSPV-------------------------------------YK-YAQDFLVAISEPVCRPTHAHEYKLTAYSLYAAVS
Dm     -----GHVFLESFSPV-------------------------------------YK-HAHDFLIAISEPVCRPEHIHEYKLTAYSLYAAVS
Ce     -----GHIFLESFSPV-------------------------------------YK-HARDFLIAISEPVCRPQHIHEYQLTAYSLYAAVS
Sc     -----GRIILESFSPL-------------------------------------AE-QAQDFLVTIAEPISRPSHIHEYKITAYSLYAAVS
Sp     -----GRIILEAFSPL-------------------------------------AE-QAIDFLVTISEPVSRPAFIHEYRITAYSLYAAVS
Ca     -----GRIILESFSPL-------------------------------------AE-QAQDFLITIAEPISRPSHIHEYRITAYSLYAAVS
Nc     -----GTIILEKFSPD-------------------------------------AD-RVTDFLVTIAEPKSRPHFLHEYQLTAHSLYAGVS

Tb-1   -----GRIFINHGHPA-------------------------------------YP-HLMDFLTACCEPVCRTLYVSEYTISPSSLSAATA
Tco-1  -----GTILVNHSHSA-------------------------------------YP-HLRDFLTACCEPVCRTLHMSEYVISPSSLSAAST
Tv-1   -----GSIYVRHAHPA-------------------------------------YP-LLRDFLTACCEPVCRTAHMSEYVISPSSLSAASA
Tc-1   -----GSLFVNNTHPA-------------------------------------YP-HLVDFLTSCCEPVSRTLRMSEYVISPSSLSAASA
Lm-1   -----GNIFVEKAHPA-------------------------------------YA-HVIDFLISCCEPVSRTQHMEQYRMDSSSLSSATA
Li-1   -----GNIFVEKAHPA-------------------------------------YT-HVIDFLISCCEPVSRTQYMEQYRMDSSSLSAATA
Lb-1   -----GNIFVEKAHPA-------------------------------------YA-HVIDFLISCCEPVSRTQHMEQYRMDSSSLSAATA
Lmx-1  -----GNIFVEKAHPA-------------------------------------YA-HVIDFLISCCEPVSRTQHMEQYRMDSSSLSAATA
Bs-1   -----GRLFVETFGEE-------------------------------------Y  N RNFVLSCAEPVSRPEFVNEYAVTPASLAAASA

Tb-R   --EGCVVLFAEAFRPS-------------------------------------FA-NIQGFLTTIAEPVSRLSLMHEYRMTYFSLGAAAS
Tco-2  --EGCVVLLAEAFRTV-------------------------------------FN-DIQGFLTTIAEPVSRLSLIHEYRITHSSLGTAVS
Tv-2*  --EDCITVLAEAFRSS-------------------------------------FP-AIQNFLTTIAEPTSRLALLHEYRITRFSLGTAVS
Tc-2   --EGYVVLLAETFRSS-------------------------------------FQ-RIQFFLTTIAEPTSRPSLMHEYRMTPFSLGAAVS
Lm-2   --DGYITIIAESFRRS-------------------------------------YV-NIRPFLTTLAEAISRPSLMHEYLLTPFSLGAAVS
Li-2   --DGYITIIAESFRRS-------------------------------------YV-TIRPRLTTLAEAISRPSLMHEYRLTPFSLGAAVS
Lb-2   --DGYIMILAESFRGS-------------------------------------YE-TIRPLLTTLAEAVSRPSLIHEYRLTPFSLGAAVS
Lmx-2  --GGYITIIAESFRRS-------------------------------------YV-IIRPLLTTLAEAISRPSLMHEYRLTPFSLGAAVS
Bs-2   --AGHISVVVEVFRGTDERRRISEQSRIDQQQQQQDDDGSAPTNRQTSSTAVNTT-SWKLFLQQVAEAVSRPAYLHLFHITPFSLGTAVS

Ng-1   --NGDYHIFMETNTPV-------------------------------------YQ-QAYDFLVAIAEPISRMENLHEYQLKEFSLYGAVS
Ng-2   --DGKFHIFMETFSPI-------------------------------------YQ-QAYDFLVAVAEPVSRPENLHEYVLTEYSLYAAVS
Trva-1 -----GHIFLETFSPF-------------------------------------YS-KTVDFIIAIAEPCSRPKYMQEYQISPYSLFAAVS
Trva-2 -----GHIFVETFSPF-------------------------------------YS-KVVDFIIAIADPCSRPKYVQEYQINPYSIFSAVS
Gl     ALSADHVVLLETFSPM-------------------------------------YD-EAYNFLVSLAEAETRPTYVHEYRITEESLHIGLS
Ehis-1 -----LRIVVETSNDM-------------------------------------FK-EVSDYLSRVAQVKSRMEHMHEYQLTPTSIMTAFS
Ehis-2 -----GLIIIERDNKN-------------------------------------YQ-IASEFIKKISKVKCLLEHMEEYIITEYTIYSAIS
Edis-1 -----LRIVVETSNDM-------------------------------------FK-EVSDYLSRVAQVKSRMEHMHEYQLTPTSIMTAFS
Edis-2 -----GLIIIERDNKN-------------------------------------YQ-IASEFIKKISKVKSLLEHIEEYIITEYTIYSAIS
Dd     -----GHIFLETFSAI-------------------------------------YK-QASDFLVAIAEPVCRPQNIHEYQLTPYSLYAAVS
Ac**
Pf     -----GFIYLEMFNSC-------------------------------------SK-QASDFLITIAEPICRPELIHEFQLTIFSLYAAIS
Toxg   -----GTILLETFSAA-------------------------------------NR-QATELLLTMAEPICRGDFVHEFQITIFSLYAGIS
Eimt**
Ta     -----GYLYLELFTPV-------------------------------------SK-QALDFIVTIAEPVCRPELIHEYQVTVFSLYTAVS
Tett-1 -----GIIFLETFNPL-------------------------------------YR-VAYQFLISIGEPVQRPLSMHKFTLTKYSLYTAMV
Tett-2 -----LYIFVEAFHEK-------------------------------------YK-EVYEFLIKIAEPVNKPNFIHVYKITLFSMYTAMS
Part-1 -----LGIIVEKFNPL-------------------------------------YE-IAFEFLMCVAEPISRSELIHEYVLTQMSMYTAMV
Part-2 -----YEIILEKFHRD-------------------------------------AA-AAEETLIQFAEPSSTTTNIQTFIITFYSLYTAMT
At-1   -----GRIFLETFSPL-------------------------------------YK-QAYDFLIAIAEPVCRPESMHEYNLTPHSLYAAVS
At-2   -----GKIFLETFSPL-------------------------------------YK-QAYDFLIAIAEPVCRPESMHEYNLTPHSLYAAVS
Orys-1 -----GRIFLETFSPL-------------------------------------YK-QAYDFLIAIAEPVCRPESMHEYNLTPHSLYAAVS
Orys-2 -----GRIFLETFSPL-------------------------------------YK-QAYDFLIAIAEPVCRPESMHEYNLTPHSLYAAVS


       *****
Hs     VGLQTSDITEYLRKLSKT-------GVPDGIMQFIKLCTVSYGKVKL------------------------------VLKHNRYFVESCH
Mm     VGLQTSDITEYLRKLSKT-------GVPDGIIQFIKLCTVSYGKVKL------------------------------VLKHNRYFVESSH
Drer   VGLQTSDIIEYLQKLSKT-------SVPDGIVQFIKLCTVSYGKVKL------------------------------VLKHNRYFVESAF
Dm     VGLQTHDIVEYLKRLSKT-------SIPEGILEFIRLCTLSYGKVKL------------------------------VLKHNKYFIESPH
Ce     VGLQTKDIIEYLERLSKS-------QLPKGVITFVQMCTVSYGKVKL------------------------------VLKHNRYFVESRH
Sc     VGLETDDIISVLDRLSKV-------PVAESIINFIKGATISYGKVKL------------------------------VIKHNRYFVETTQ
Sp     VGLKTEDIIAVLDRLSKT-------PIPPSIVDFIRACTVSYGKVKL------------------------------VLKKNRYFIESGD
Ca     VGLETDDIISVLNRLSKV-------PVAESIINFIKAATVSYGKVKL------------------------------VLKHNRYFVESTQ
Nc     IGLQSKDIIDTLDRFLKT-------PLPESIRLFIESCTKSYGKVKL------------------------------VLNNNKYFVESSD

Tb-1   EGTYSMEMVRNVIRYFRLDEQ---QQIPVDVERYAALERRVRDSVQDTSLDLPMEV-------------------------GEAKVSANG
Tco-1  EGTYSMGMIRNVLRYFRLNEQ---QRLPVDIARYNRLEQKVWNNLQNSTAEAALFE-------------------------TD---DADG
Tv-1   EGTYSVAMIRNVVRYFRLDAK---FRLPVDIVKYVELQQRVQDGSQDASLGILGD----------------------------EDRVTNG
Tc-1   EGTYSTAMIRNFIRYFRLDEQ---HRLPVDLAKLSALERQLENGEDV-----AMGT-------------------------THESEESQG
Lm-1   EGTYSLAMIENVLRHFRLNAA---QELPVDLERCAALERLASALDES-TESAEVDGSRVSQEALATTTASSSFDARIPLPRCLCDVDLTG
Li-1   EGTYSLAMIESVLRHFRLNAA---QELPVDLERCAALERLAAALDES-NGSAEVDGFRTSQEALATKTASSSFAARIPLPRCLSDVDLTG
Lb-1   EGTYSLAMIEVILRHFRLNTA---QELPVDLERCAALERLASELDES-NESAEVDGLWASQAAVATKAASSLFAVHTPLPRCLYDVDLTG
Lmx-1  EGTYSLAMIENVLRHFRLNAA---QELPVDLERCAALERLASALEES-SEGAEVDRSRTSQEAMATKTSSSSFAARIPLPRCLCDVDLTG
Bs-1   EGSYTGEDVREFLLRRSEDCD---ARLPAAVDD---------------------------------------------------------

Tb-R   ASISVSEVIQFLDDHVYFFRDKCFASYRESVCAFAEMYMSRCNLARV------------------------------VIDESRTLLECKD
Tco-2  ADISAAEVVQFLSEHVYFFNEEHFRPHRNRVELFAERCMSRYNLARI------------------------------VIDDTRTMLECND
Tv-2*  ANISAAEVIKFLDNYVYFFCDESLVEKRKRVCAFVETYMSRYNLARI------------------------------VIDEEQTSVVCKD
Tc-2   ASIPVSDVICFLDEYVYGFCDEGENSRRARVCAFLEMCMSRYNLARI------------------------------VIEDARTFVECDN
Lm-2   NGIDAAEATAFLETHAYGLAESS--ERRQLVRQFIESCMKRYNLARI------------------------------IIDAERTLVQCKN
Li-2   NGIDAAEATAFLETYAYGLAESS--ERRQLVRQFIESCMMRYNLARI------------------------------IIDAERTLVQCKN
Lb-2   NGIDATEATTFLETYAYGLATCS--ERRQRVRQFIESCMRRYNLVRV------------------------------IIDSERTLVQCKN
Lmx-2  NGIDAAEATAFLETYAYGLAESS--ERRKLVRQFIESCMTCYNLARI------------------------------IIDSDRTLVQCKN
Bs-2   SGWTPTAVISCLSAYSYHDTT-------TAMNNFITSHMERHNKAQI------------------------------VVDERQRFQLCIR

Ng-1   IGLTCDYIIQVLDKLSKV-------KLPTDLIDFIRKTTSKYGSVKF------------------------------VLQKNRFFLESPI
Ng-2   IGMNTEQIIQVLEKLSKV-------KLSEALINFIRDVTRKYGNVKL------------------------------VLQKNRFFVESPH
Trva-1 IGLTGDKIVHVLSLISKT-------ALTDTFKEHIINCCQAVGKLKL------------------------------VLKEQKYYIESIE
Trva-2 IGLKAKEIIRILAIISKT-------PLTDEFKEHIELCCLSVGKLKS------------------------------VLRNTKYYIESNE
Gl     LGYTATEIEIRLGYYSKY------RRLPPTLLSALRRLANNQGTISFS-----------------------------IYDSNRFLLRVPQ
Ehis-1 FGSTPEAMISTLEKYSKN-------YLPDNVKSQIRQAGEKKQNFRL------------------------------VLINGKYYLQAES
Ehis-2 SGIKPEIIIYQLKKYSKN-------NIPERIEQLIEKTIKSRIITVI------------------------------CNYEGRRICEG-D
Edis-1 FGSTPEAMISTLEKYSKN-------YLPDNVKSQIRQAGEKKQNFRL------------------------------VLINGKYYLQAES
Edis-2 IGIEAETIIYQLKKYSKN-------NIPERIKTLIEKTINNRIITII------------------------------CNYEGKRICEG-D
Dd     VGLETNDIITVLGRLSKL-------ALPKEVEQFVRQCTQSYGKVKL------------------------------VLQKNKYFVESAY
Ac**                                         MKNYYTVKL------------------------------VLKDGEYFVESHS
Pf     VGITLDELLINLDKFSKN-------VLPNELISNITKSAESFGKVKL------------------------------VLRENKYYIEATN
Toxg   IGLSCEDMLLNLEKFSKN-------AIPEDLVVQIQKVASAFGKVKL------------------------------VLNENKYYIESAE
Eimt**
Ta     VGLSFEELLNNLNKFSKN-------ELPKKLKESILNTSSAFGKIKL------------------------------VLRDSRYWIESFE
Tett-1 LQYEPKDIILCLEKLSKNK------KIPKEVERYITENTQNYGAARL------------------------------FLDDSSYYLDIDS
Tett-2 LNLNAEKIIESLMQYSKNE------VIPKSFIEFIENKTKKAGSAYL------------------------------YMIKKNYYLKIHQ
Part-1 LQYSADDIIRLLDLLSKN-------KVPQRMEQFIRHHTNNIGQAKF------------------------------FLQDKSYYIDIGE
Part-2 LGWNEERIKSEIDRLAKNE------RIPKDITDFIEKNTQYFNRAYF------------------------------YLEKDSYYIDIEM
At-1   VGLETETIISVLNKLSKT-------KLPKEMIEFIHASTANYGKVKL------------------------------VLKKNRYFIESPF
At-2   VGLETETIISVLNKLSKT-------KLPGEIIDFIHASTANYGKVKL------------------------------VLKKNRYFIESPF
Orys-1 VGLETSTIISVMSKLSKT-------KLPREIIDFIHASTANYGKVKL------------------------------VLKKNRYFVESPF
Orys-2 VGLETSTIISVMSKLSKT-------KLPREIIDFIHASTANYGKVKL------------------------------VLKKNRYFVESPF
```

```
Hs      -------------------------PDVIQHLLQDPVIRECRLRNSEGEATELITETFT-------------------------------
Mm      -------------------------PDVIQHLLQDPVIRECRLRNAEGEATELITETFT-------------------------------
Drer    -------------------------PDVIQRLLQDTVIRDCRLRSAEGEETELITETIS-------------------------------
Dm      -------------------------PEVLQKLLKDPVIQKCRLIRSEGEDF--IQGTLD-------------------------------
Ce      -------------------------SDVMQKLLKDSVIQSCILDDRVQPAQQT---ELP-------------------------------
Sc      -------------------------ADILQMLLNDSVIGPLRIDSDHQVQPPEDVLQQQ-------------------------------
Sp      -------------------------ASVLRLLLRDPVIGPLRIDYSTQSSKQKSSKP---------------------------------
Ca      -------------------------ADILQMLLKDPVIGPLRIQSSENIVTSNGLITSA-------------------------------
Nc      -------------------------AELLQKLLRDEVIGKWRVQGSGDITTSYAPTM---------------------------------

Tb-1    DVKAEEG--------------------------------CEEATDELSPLAG--------QVKKEETKEVAAEPRRRFLSL---------
Tco-1   SSDAEQE--------------------------------ECEGTEALSPCGE---FASPKRFKRDPGGGGAAPPKHRVLSL---------
Tv-1    ATLENGG--------------------------------TCHSEPAVKKL--KKEDEEFVPEVKVNSLENTEKPRRRFLSL---------
Tc-1    K--KQQK--------------------------------QEEDEDGLGPKRAKKENGERVSLVKQEATEETAAPSRRFFSL---------
Lm-1    TILASLNRTSADDAPSPAVTAHHVGNDSLETLVKREKVEEEEMGDRLSAGGAAASALKTVSSRSAAPSG---SPTSRLLQLFRPLLPTPA
Li-1    TILASLNRTGADDVASPAVSAPHVSNDSWKTLVKREKVEEEAVGDRLSAGDAAASALETVSSRAAAPSG---SPTSRLLQLFRPLLPTPA
Lb-1    TLLASLNRPSADDAASSTLFSPHMRNDTLKTRIKSEKVEEE-EGELLSAGGEGASALEPVLPRTAAPFPRPRSPATRVLQLFRPSSPAPA
Lmx-1   TILASLNGASADGVTSPAVPTPHVSNDRLKSLVKRENVEEEEVGDRLSAADAAASASETVSSRTAAPSS---SPTSRLLQLFRPLLPTPA
Bs-1    ------------------------------------------------------------------------------------------

Tb-R    -------------------------IVTAKTLLKDRVVRSLCCQPKETVGE---------------------------------------
Tco-2   -------------------------IETANILLRDLTVRSLCAEPPEIVTT---------------------------------------
Tv-2*   -------------------------ASTAKALLEDRVVRSLCAHPPKASRT---------------------------------------
Tc-2    -------------------------VETAETLLLDPVVRSLCAEPPQIVPV---------------------------------------
Lm-2    -------------------------EETAQMLLRDAVIASVAATPLRIFYEESS------------------------------------
Li-2    -------------------------EETAQMLLRDAVIASVAVTPLRIFYEKSS------------------------------------
Lb-2    -------------------------EETAQMLLCDAVITSVAATPPQIFYETSS------------------------------------
Lmx-2   -------------------------EETAQMLLRDAVIASVAATPLRIFYEESS------------------------------------
Bs-2    ------G------------------RRLAEAILRDQRIVSLCAVRDTSVGATAAAT---VRNRQQQT-----------------------

Ng-1    -------------------------LQIIEKLRNDSVIGKAALIPNPNEFEYLDSS----------------------------------
Ng-2    -------------------------LTIMEKLYEDKIISAAAIVPDPNEFPNVDRS----------------------------------
Trva-1  -------------------------EKLISELIKERFFAEKIVSPPSGVMTHPKYP----------------------------------
Trva-2  -------------------------IRRLLELLEIPFFEERHVYPEPDEKT-YKHT----------------------------------
Gl      TLDLTPDIKDLLED-----ASDNVADTLFEELLRKKRNDTAEEL----------------QKRELKTH----------------------
Ehis-1  -------------------------QEQMKKVMEENFKQFY-------------------------------------------------
Ehis-2  -------------------------KKIIEEIK---------------------------------------------------------
Edis-1  -------------------------QEQMKKVMEENFKQFY-------------------------------------------------
Edis-2  -------------------------KKIIEEVK---------------------------------------------------------
Dd      -------------------------PEVLEFLLKDSSIATARIKPTLEESVVDP------------------------------------
Ac**    -------------------------PSILNIIQQR-VLGQVEMS----------------------------------------------
Pf      -------------------------KSELDYLLNNDTIQNARIYSTDNNNDKNMISLYNLNNRHRDNNKNDN------------------
Toxg    -------------------------KKELDYLLSNPLIRSSAVAHRRPE----LAQTPQVPAA---------------------------
Eimt**
Ta      -------------------------RSELDLLLTNQEIKNSRIHSNIWDKQPQLQ-----------------------------------
Tett-1  --------------------------QVWTSIIERNPHIKDLVSQTQFNFDNIINQNKELRD----------------------------
Tett-2  --------------------------EFLEKIIRIDEEPY----------KSIIE-----------------------------------
Part-1  --------------------------TVLNQIMMNLQDKNSL--------QEVIQEEQIIHN----------------------------
Part-2  --------------------------KLLQELKHRKFGQMH-----------------NLT-----------------------------
At-1    -------------------------PEVLKRLLSDDVINRARFSSEPYYG----------------------------------------
At-2    -------------------------PEVLKRLLSDDVINRARFTSEPYYG----------------------------------------
Orys-1  -------------------------PEVLKTLLKDDIICRARISPEDSVG----------------------------------------
Orys-2  -------------------------PEVLKTLLKDDIICRARISPEDSVG----------------------------------------


Hs      --------------------------------------------------------------------SKSAI----------------S
Mm      --------------------------------------------------------------------SKSAI----------------S
Drer    --------------------------------------------------------------------SKSAI----------------S
Dm      --------------------------------------------------------------------GKAIT----------------Q
Ce      --------------------------------------------------------------------AQEKI--------------KFS
Sc      --------------------------------------------------------------------LQQT------------------
Sp      --------------------------------------------------------------------SNEDN--------------VED
Ca      --------------------------------------------------------------------SEPSN--------------IVI
Nc      -----------------------------------------------------------------------GG--------------LVI

Tb-1    ---------------SKRISAKSEPLVTRAVVNT-GALQPLPADLEQMLREEENSSRVRIVLQPCLRPVKRRAV-GGDKQHSD-------
Tco-1   ---------------CKRSVSSAEPFVSRAVINA-DALRPLPDDIEAMLQEEENSSKVRIVLQPCLQSRKLRT---GKFVHSN-------
Tv-1    ---------------VRKPPAPLEPLVARTLVNP-DALLPLPDDLEQMLQDEERSSKVRIVLQRRLQGRQRCSS-------SQ-------
Tc-1    ---------------VKGVTIQPEPLISHALCFP-DALRPLPPDLEQMLLEEERSSRVRIVLQPQLRSFQRQQP-Q----H---------
Lm-1    SAAAPTHSPTTRSQLEQGAPIAARPLISLAVVRPADVRRPLPESLAQMLEEEENASRVQIVLQPRLRRFHQFEASRGASFASTTATSRAT
Li-1    SAAAPIHSPTAGSPLEQGTPIAARPLISLAVVRPADVHRPLPESLAQMLKEEEHASRVQIVLQPRLRRFHQFEASRGASFASTTATSRAT
Lb-1    STAVMLRSPTADSQLKQAAPSAARPLISLAVVRPADVQRPLPELLAQMLKEEEKASRVQIVLQPGLRRFHQFEASRGAFFASTSATCRAT
Lmx-1   SAAAPLHSPTTGSQLEQGAPIAARPLISLAVVRPADVHRPLPELLAQMLKEEENASRVQIVLQPCLRRFHQFQASRGASFASTTATPRAT
Bs-1    -----------------------------------------------MIAVEIMSEKISLTLAPDLRGDFDARSSGGASRS---------

Tb-R    -----------------------DA-----------------------------------------------------------------
Tco-2   -----------------------SA-----------------------------------------------------------------
Tv-2*   -----------------------EA-----------------------------------------------------------------
Tc-2    -----------------------VP-----------------------------------------------------------------
Lm-2    --------------WEE----SPAP-----------------------------------------------------------------
Li-2    --------------WEE----SPAP-----------------------------------------------------------------
Lb-2    --------------WAE----NPAP-----------------------------------------------------------------
Lmx-2   --------------WDE----NPAP-----------------------------------------------------------------
Bs-2    ------STKTSDKKIEE----GDES-------------------L---------SHLYITR---------RII----------------R

Ng-1    ---------------------------------------------------------------------KRFI--------------IKS
Ng-2    ---------------------------------------------------------------------KRFI--------------IRT
Trva-1  -----------------------------------------------------------------------KI--------------M-I
Trva-2  -----------------------------------------------------------------------KF--------------L-K
Gl      ---------------EQGVQS-DQS--------------------------LWYTQSFKIASIPKLKRYFFPL-----------------
Ehis-1  ------------------------------------------------------------------------------------------
Ehis-2  ----------------------------------------------------------------------ECI--------------T--
Edis-1  ------------------------------------------------------------------------------------------
Edis-2  ----------------------------------------------------------------------EFI--------------I--
Dd      ---------------------------------------------------------------------KTGF----------------I
Ac**    -----------------------------------------------------------------------NF----------------S
Pf      -----KNNNDDNNNIDK----TTSN-------------------LLN--EDQKGKDCYVTYEAPVLDTTQLGF--------------KIS
Toxg    ---------------AS----GSSS-------------------LLKSLPSPSNDGLYVIANAPTLDASQLAF--------------QVS
Eimt**
Ta      ----------------------------------------------N--STSNRNNEYVTISVPT---SDFSF--------------VFN
Tett-1  -----------------------------------------------------------IQK--DLQSFFQNIKTENNQQSSKKNQ-KMI
Tett-2  -----------------------------------------------------------MEQNPPQQLLQT-------------------
Part-1  -----------------------------------------------------------IQVNPQLKNFEQQIEL---------------
Part-2  ----------------------------------------------------------------SIENYER-------------------
At-1    ---------------------------------------------------------------------GDGF----------------S
At-2    ---------------------------------------------------------------------GDGF----------------T
Orys-1  ---------------------------------------------------------------------APSF----------------T
Orys-2  ---------------------------------------------------------------------APSF----------------T
```

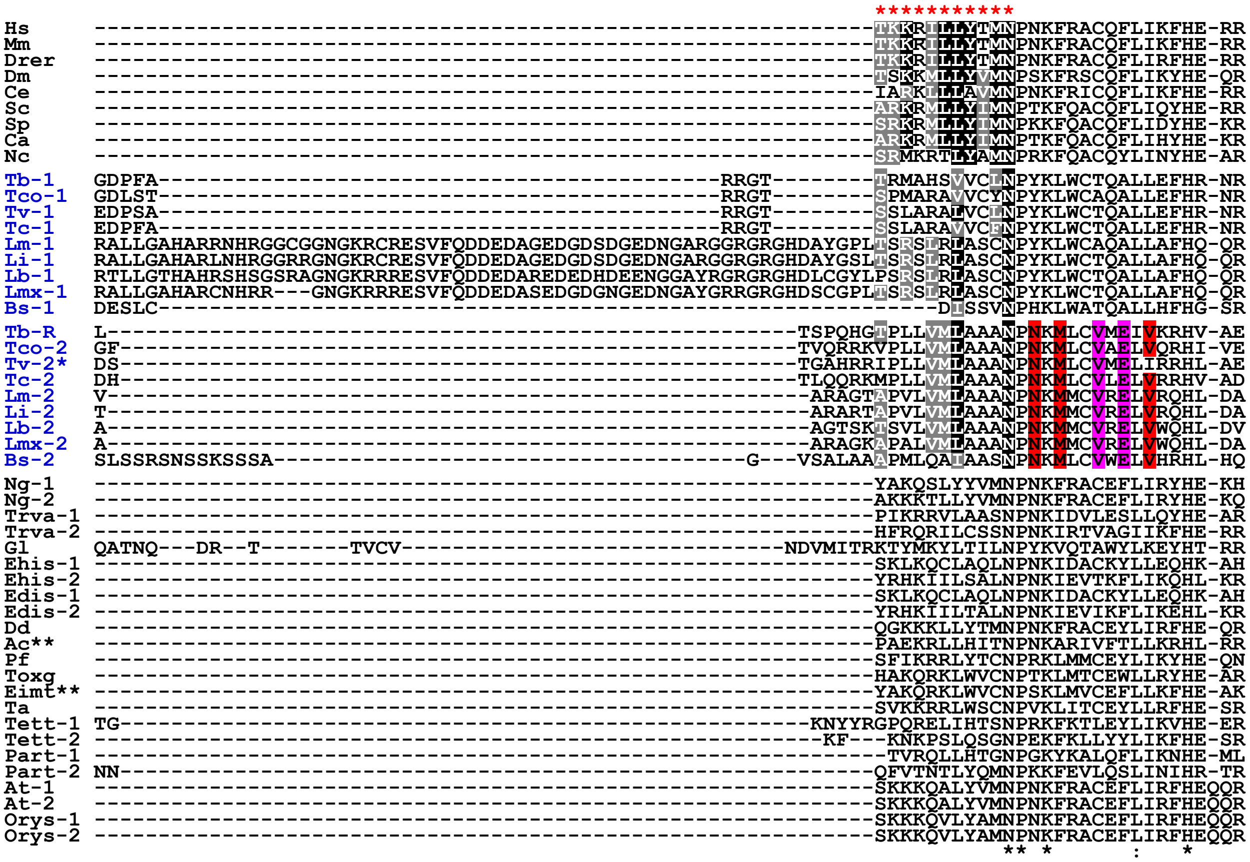
```
Hs     KTA-ESSG-------GP-STSRVT-DPQ----------GKSDIPMDLFDFYE---------QM---------------------------
Mm     KTAAEGSG-------GP-STSQGV-DAQ----------ATSDIPKDLFDFYE---------QM---------------------------
Drer   KSQQDN-G-------GP-SSSQPADGQR----------SGTQVPEDIFSYYE---------QM---------------------------
Dm     FGTKLPPGATD----KPTPDPAAAAGAD----------GTTAVPEDITDFYE---------KI---------------------------
Ce     HGNENEKGVE-----KDGAAADGTAAAA----------TDGKVPADIDEFYG---------KI---------------------------
Sc     ---A-------------G--KPAT----------------NVNPNDVEAVFS---------AVI--------------------------
Sp     KKDI-------------TNDSSKE----------------TAEKSSSDELFS---------AVV--------------------------
Ca     PGTK-------------KNDSTKE----------------DPANNGAEDIFA---------SVV--------------------------
Nc     PGTKDAAGVNQAGLNKSGNKKTAEDG------------AADAAATNEADLYA---------AI---------------------------

Tb-1   ------EQ----------QCQRAEETKLAYFLTSPDRNHMEHLVSRLQDFLVPVLLHGTRRWVVSDVDRGV---------------EERS
Tco-1  ------FN------QLQRHQYQSDENKLAYFLTSPDRQRMEYLVSHLQEYLVPISLHGAIRWIMSDVDRGV---------------EERS
Tv-1   ------AH------NYQKQNESSEEEALAYFLVSHDRQHMEFLVEHLREYIVPVSLYGALRWVISDIDRGV---------------EERS
Tc-1   --------------HSQGQIHHEDEGRLVYFLVSQDRQQMEFLVSRLQEFLVPVSLHGVMRWVLSDVDRGV---------------EEKS
Lm-1   VSKER-NG------AATGALAERDDQQLFYFVQSRERAHLEHVLPALKEFLEPVVICGAERWIVSDVDRTPLGNTKGIGD------AERL
Li-1   VWKER-NG------AAAGALADRDDQQLLYFVQSRQRAHLEHVLPALKEFLEPVVICGAERWILSDVDRSPLGNTKDIGD------AERL
Lb-1   APKGR-NG------AATGTPADRDDQQLFYFVQSRQRVHLELVLPALKEFLKPVVICGVERWILSDVDRSPLMNTSDIKD------AERL
Lmx-1  VSKER-NG------AATGAVADQDDQQLFYFVQSRQRAHLEHVLPALKEFLEPVVICGAERWILSDIDRSPLGDTKGVSD------AERV
Bs-1   --------------NLFSSGVAPSSGNLAYFLQCSQKPLLEAFANEVKEFLDPYLLGGERAFVISDLVVSD---------------APRS

Tb-R   ------------------------------------------------------------------------------------------
Tco-2  ------------------------------------------------------------------------------------------
Tv-2*  ------------------------------------------------------------------------------------------
Tc-2   ------------------------------------------------------------------------------------------
Lm-2   ------------------------------------------------------------------------------------------
Li-2   ------------------------------------------------------------------------------------------
Lb-2   ------------------------------------------------------------------------------------------
Lmx-2  ------------------------------------------------------------------------------------------
Bs-2   APDEEEDE------------------------------------------E----------------------------------SHEMS

Ng-1   KSN--QVEAFQKQTLISTNSNSLSNT-------------MNEQHHVQKEMIG---------AIN--------------------------
Ng-2   RSKQNELEKFQKQALANTS------T-------------VNKGSKIQTEMIR---------AVN--------------------------
Trva-1 VDEKEAD------TIISGV--------------------------GQAAYYG---------SVQ--------------------------
Trva-2 VGKEELE------TIISGI--------------------------GESSYHG---------NVA--------------------------
Gl     ----------------------------------E---ALTSIRKKLKEMRDPLS--GTRSLIT---------------------IEEYL
Ehis-1 EGSAR-------------------------------------------------------------------------------------
Ehis-2 ------------------------------------------------------------------------------------------
Edis-1 EGSAR-------------------------------------------------------------------------------------
Edis-2 ------------------------------------------------------------------------------------------
Dd     INKEVVTGAQISGGLQANQS---------------------LDPVLKNDALS---------NLL--------------------------
Ac**   TTDE--------------------------------------------------------------------------------------
Pf     ESEKQLM---MEE-----------------------------------------------------------------------------
Toxg   ERDEETD---PAGLTGAGKSAEP--G--------------DTAPGAVTAVAG---------NLASPGASGD-GDPSETRKRPREEGRETS
Eimt**
Ta     RNHEAPE---SEE-----------------------------------------------------------------------------
Tett-1 VEDEQDDG--------------------------------------IPDKKN---------KLMKNGDSSI-ANY---------------
Tett-2 ------------------------------------------------------------------------------------------
Part-1 ------------------------------------------------------------------------------------------
Part-2 ------------------------------------------------------------------------------------------
At-1   VGRT-------SGELEAG--------------------------------PG---------ELL--------------------------
At-2   IGKT-------SGELEAG--------------------------------PG---------ELL--------------------------
Orys-1 ISKT-------AGEMASG--------------------------------HE---------DLL--------------------------
Orys-2 ISKT-------AGEMASG--------------------------------HE---------DLL--------------------------


Hs     ----------------------------------DKDEEEE-E-----E--TQTVSFEVKQ---EMIEELQ-KRC-IH--LEYPLLAEYD
Mm     ----------------------------------DKDEEEE-E-----E--TQTVSFEVKQ---EMIEELQ-KRC-IC--LEYPLLAEYD
Drer   ----------------------------------DKEEEEE-E-----E--TQTVSFEIRQ---EMIEELQ-KRC-IQ--LEYPLLAEYD
Dm     ----------------------------------DKEEEDE-D-----EANLKTVSFEVAQ---EKIEVIQ-KRC-IE--IEHPLLAEYD
Ce     ----------------------------------DGDDEEDAE-----IRNLQLLTFEIKQ---ETIETVQ-KRC-IE--LEYPLLAEYD
Sc     ---------------------GG-----------DNEREEE-D-----DDIDAVHSFEIAN---ESVEVVK-KRC-QE--IDYPVLEEYD
Sp     ---------------------GL---------------QEE-E-----DDEDAVHLFEIKH---SSVETIK-KRC-AE--IDYPLLEEYD
Ca     ---------------------GN---------------RDD-D-----DDMDTIHSFEIAH---DSVEIVK-RRC-QD--IEYPVLEEYD
Nc     ------------------------------------NEEDD-E-----DDKDAVHAFEIPE---TAVEIVQ-RRC-LD--LGFPILEEYD

Tb-1   TAESGRAKTLRRLFEAPSSASGRSVASKSLT--NEGANGDGLGGGVGRRCTRIVYKSQVMD---GKMRNVRERLY-KE--LSVRADLFYD
Tco-1  TVESGRASVLRWLFDTPSRLPVFQCHGKEGAGATATPYSTGTGGGLG-RCTRIVYKSQVAD---GKMRNVRERLY-KQ--LGVRADLFYD
Tv-1   VVEDGRARTLRILFDTPSGQKSELQS-----------ANSSRMARSEKKYTRIVYKSQVMK---GRLRNVRERLF-KQ--LGVRVDLSYD
Tc-1   VVESGRAKTLRRLFDTPSVSFASAMD-------DVGSCATAAANSTGRKYSRVVYKSQVKD---GRLRNVRERLF-KE--LGVRADLFYD
Lm-1   ARDAGRPAVLRLLYEPPRVGTQYEKA------------------EATPSASRFLYKCQVRD---GKLREVKECLF-SK--FDIRADCYYD
Li-1   ARDAGRPAVLRLLYESPRAGTRHEKA------------------AATPSASRFLYKCQVRD---GKLREVKECLF-SK--FDIRADCYYD
Lb-1   TRDAGRPAVLRLLYEFPHVGTVSERT------------------TTASTASRFVYKCQVRD---GKLREVKERLF-SK--FGIRADCYYD
Lmx-1  ARDAGRPAVLRLLYEPPRVGTRYEAA------------------AATPSASCFVYKCRVRD---GKLREVKECLF-SK--FDIRADCYYD
Bs-1   ---SGSL--L---------------------------RERRLNGSSGPVTKQMVYKCRIRA---GCQRQIKEALY-QR---SIRVICTYD

Tb-R   -----------------------------------------------------CPHFLLKSR--AAARVVAER-C-VL--LGYPLQQQYE
Tco-2  -----------------------------------------------------RPYSLLRNR--AVTKVVAER-C-AA--LGYPLQQQYD
Tv-2*  -----------------------------------------------------CSF-LLRSR--AAARPVAER-C-VA--LGFPLQQQYD
Tc-2   -----------------------------------------------------HSLLVLRSR--AVARRVAER-C-VL--LGFPLQQQYD
Lm-2   -----------------------------------------------------YPSFLLQSR--AMSKVVAAQ-C-VV--LGLPIQQQYD
Li-2   -----------------------------------------------------YPSFLLQSR--AMSKVVAAQ-C-VV--LGLPIQQQYD
Lb-2   -----------------------------------------------------YPSFLLQSR--AMSKVVAAR-C-MV--LGFPIQQQYD
Lmx-2  -----------------------------------------------------YPSFLLQSR--VMSKVVAAQ-C-VV--LGLPIQQQYD
Bs-2   --SEAKNAAI-----------VA-----------DKEEHDDDG-------ELEEVQVLLQDH--AASRLVSDVLW-AH--HHLPICQVYH

Ng-1   ---------------------HS-----------NYEEEYEDDNDLDSEPHETLYSFEISS---KSIESVK-KQC-IV--IGHPTLDEYD
Ng-2   ---------------------QI-----------GDDEDDDADDDFDIS-----------------------------------------
Trva-1 ---------------------RL-------------TEGFD-EIFE--VEKPKIYRFELKP---ETVRDVR-KHA-VD--NNLFLSDEYD
Trva-2 ---------------------RL-------------TEGID-DLLDTPLEKQTILRFQIKT---ESVREIR-QYA-VD--HNLFISDEYD
Gl     LLDEGIDAVR-------------------------NEQVQRLQSLMAAEINTEAEMCEL--------KKIKEE-----------FAKAY-
Ehis-1 ------------------------------------EVKESDELYRELDLDHHVFIIEVKQ---TSVFKLK-KKC-KK--KKVRVYEEYH
Ehis-2 ---------------------------------------------------INKNEYELK----SNIIVFK-KKC-QE--KGILINEKIE
Edis-1 ------------------------------------EVKESDELYRELDLDHHVFIIEVKQ---TSVFKLK-KKC-KK--KKVRVYEEYH
Edis-2 ---------------------------------------------------INKKEYEIK----NNIIEFK-KKC-QE--KGILINEKIE
Dd     ----------------------------------EEEEEDTVN-----NSDQHFHSFEIDP---QQVEEVK-KRC-IQ--LDYPVLEEYD
Ac**   ------------------------------------------------ETGAQVLVYALPISTVQEVAEIK-KLL-LA--HNHPFIEEFD
Pf     ------------------------------------KKHANLNANENSTNSAEVYSFEVNC---DKIEEVK-QEA-LQ-TMQRPLLMEYD
Toxg   GAEEARGKAQVPL-ASPHASASP-----------TGEDGDSAKRKTMSAPTRQVFSFQVSS---DRVEEVK-RVS-VE-SMHRPLLNEYD
Eimt**
Ta     ------------------------------------TTTAKFAGRSDKTTTRQVFSFEVQQ---EKIEDLK-REA-LQ-TMRRPLVMEYD
Tett-1 --D------------DEDEEDSD-----------DGDDDFVYNFIMNNKTKRSFKRIQI-I---GDHFEVT-KAV-IN--CSVPLIQEYD
Tett-2 ----------------------Q--------------E-------ESRVWNFKFYRIKQ-K---HNLFQIT-EKI-IK--EGISLINEYH
Part-1 --K------------EADIEQMA-----------KEIDTFKDWLAQDKIQTIKFKRFRI-V---GDYFDVA-QAL-IR--SSVPLIQEYD
Part-2 ----------------QD------------------------SLLNDN---KTFVKFKIKA---NEHFSVN-RELVED--FNVPVKQEFD
At-1   -----------------------------------NEAEFAAA-----AEEKETHSFEIDP---AQVENVK-QRC-LPNALNYPMLEEYD
At-2   -----------------------------------NEAELAAA-----AEEKETHSFEIDP---ALVENVK-QRC-LPNALNYPMLEEYD
Orys-1 -----------------------------------DGMELAAA-----TEDKETHSFEIDP---SQVENVK-QRC-LPNALNFPMLEEYD
Orys-2 -----------------------------------DGMELAAA-----TEDKETHSFEIDP---SQVENVK-QRC-LPNALNFPMLEEYD
```



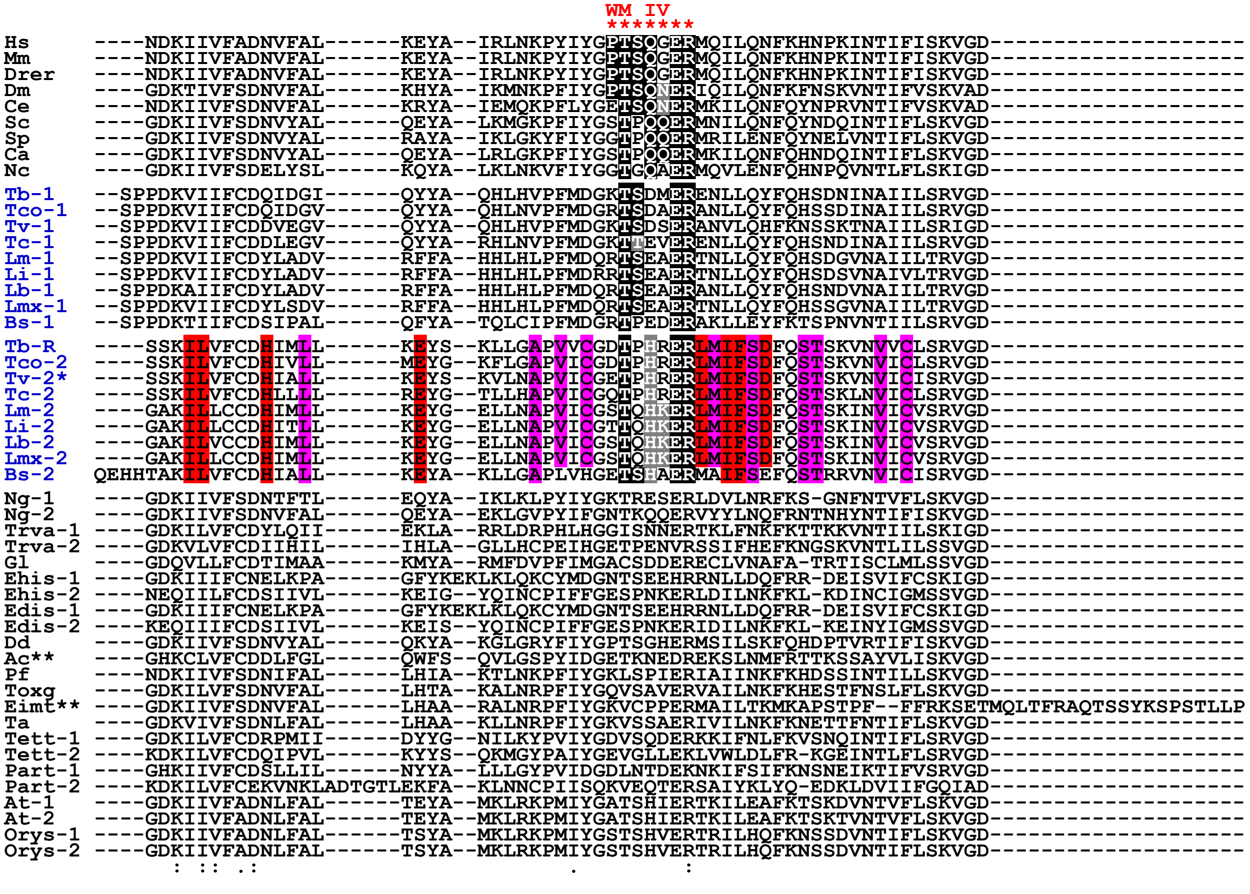
```
                                                                                         WM I
                                                                                         *****
Hs     FRNDS--------------------------VNPDINI-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Mm     FRNDT--------------------------LNPDINI-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Drer   FRNDT--------------------------VNPDINM-DLKPTAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Dm     FRNDT--------------------------NNPDINI-DLKPAAVLRPYQEKSLRKMFGNG---------RARSGVIVLPCGAGKSLVG
Ce     FRNDT--------------------------LNPNLGI-DLKPSTTLRPYQEKSLRKMFGNS---------RARSGVIVLPCGAGKTLVG
Sc     FRNDH--------------------------RNPDLDI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Sp     FRNDN--------------------------INPDLPI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Ca     FRHDA--------------------------RNPDLEI-DLKPSTQIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKTLVG
Nc     FRNDS--------------------------NNADLEI-DLRPNTQIRPYQEQSLSKMFGNG---------RAKSGIIVLPCGAGKTLVG

Tb-1   YVQDH---------------------------SLHVCDLELSENVRLRPYQVASLERFRSGN---------KAHQGVIVLPCGAGKTLTG
Tco-1  YLQDQ---------------------------SLHVCGLELSENVRLRPYQVASLERFRSGN---------KAHQGVIVLPCGAGKTLTG
Tv-1   YNEDC---------------------------TLHVCDLELAGHARLRPYQVASLERFRSGN---------KAHQGIIVLPCGAGKTLTG
Tc-1   YVQDG---------------------------TLDVRDLALAEHVRLRPYQVASLERFRCGN---------KAHQGVIVLPCGAGKTLTG
Lm-1   YMQDH---------------------------TLYVPNLSLASHVRLRPYQVASLERFRRGR---------KAHQGVIVLPCGAGKTLTG
Li-1   YMQDR---------------------------TLHVPNLNLASHVRLRPYQVASLERFCRGR---------KAHQGVIVLPCGAGKTLTG
Lb-1   YVQDR---------------------------TLYVPNLTLARHVRLRPYQVASLERFRRGQ---------RAHQGVIVLPCGAGKTLTG
Lmx-1  YMQDR---------------------------TLHVPNLHLASHVRLRPYQVASLERFRRGQ---------KAHQGVVVLPCGAGKTLTG
Bs-1   YMRDL--------------------------QTPTVANFSLRLGVRLRPYQAASLERFRRDD---------KAHHGVVVLPCGAGKTLTG

Tb-R   YEK--------------------------DTSIRNVNI-ALKSQTRPRPYQIAAVDAAASDG---------ALRSGCIVLPCGSGKTLVG
Tco-2  YER--------------------------DTTIRNVNI-ALKAQTKPRPYQIDAVDAATADG---------SLRSGCIVLPCGAGKTLVG
Tv-2*  YER--------------------------DMSIRSINI-MLKTQTKPRPYQIDAVDAAVSDG---------FLRSGCIVLPCGAGKTLVG
Tc-2   YER--------------------------DSGVRNINV-MLKTQTKPRPYQIDAVNAAATEG---------SLRSGCIVLPCGAGKTLVG
Lm-2   FEN--------------------------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Li-2   FEN--------------------------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Lb-2   FEN--------------------------DTSVRKAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCILLPCGAGKTLLG
Lmx-2  FEK--------------------------DTSVRTAHI-SLRTQTKPRRYQIEAVDAAIHDG---------TLNSGCLLLPCGAGKTLLG
Bs-2   FSASDDSVVHDDKGGADHVLNSAASRDHAANDVRSINL-VLRPQCRPRPYQIAAVEAALAHG-------GTAVRSGVLLLPCGAGKTLVG

Ng-1   FRNDK--------------------------RNSSLVI-DLKPTTTIRSYQEKSLNKMFSNG---------RARSGIIVLPCGAGKTLVG
Ng-2   ------------------------------------------------------------------------------------------
Trva-1 FMNDK--------------------------SLKNLNI-RLRTETNIRPYQEKALSKMFSGG---------RAKSGIIVLPCGAGKTLVG
Trva-2 FMNDK--------------------------TIDNLGI-QLKNTTRIRPYQEKALTKMFSGG---------RSISGIIVLPCGAGKTLVG
Gl     -----------------------------QVKAPQINA-QLKPTTDLRDYQDHASKKVISEVQLDTSQIVKRCNSGLVVLPCGAGKSLLG
Ehis-1 FLRDK--------------------------QKE-LPI-QLR-KDCLRPHQERALQQIFDNE---------MARSGIVVLPCGAGKTLTA
Ehis-2 I------------------------------GEKTIPI-ALKDDTVLRKYQISGISSVFRKG---------FAQSGIIILPCGAGKTLMT
Edis-1 FLRDK--------------------------QKE-LPI-QLR-KDCLRPHQERALQQIFDNE---------MARSGIVVLPCGAGKTLTA
Edis-2 I------------------------------GKKRINI-SLKDDTVLRTYQIKGINSVFKKG---------FAQSGIIILPCGSGKTLMS
Dd     FRNDT--------------------------VNPNLNI-DLKPTTMIRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLSG
Ac**   FLTNRWSEK----------------------ACANLAI-NLKEGTGVRPYQKRAAFSLFWDN---------KAHSSILVLPCGAGKTLVG
Pf     FRRDK--------------------------KNPNLIC-SLKSHVQIRYYQEKALRKMFSNG---------RSRSGIIVLPCGVGKTLTG
Toxg   FRRDK--------------------------TNKNLSSLLLKSSTKIRYYQERALRKMFSNG---------RARSGIIVLPCGAGKTLTG
Eimt**                                                          MFSNG---------RARSGIIVLPCGAGKTLTG
Ta     FRKDN--------------------------NSPSLNC-CIRSNIKIRYYQERALRRMFSNG---------RARSGIIVLPCGAGKTLTG
Tett-1 FENK---------------------------SFKQLEI-ELKPKIKVRYYQERALKNIFIQK---------KARSGLIILPCGAGKTIVG
Tett-2 DDKESQNDLSSN----------------VQIPYKKLDI-ELRQITQPRIYQQKALKNIFVCE---------KPRSATIVLPCGAGKTLLG
Part-1 FTKE---------------------------K-QKLDI-NLKPSTKPRLYQLRAAKTVIMGD---------YAKSGLIVLPCGAGKTLVG
Part-2 YLKDLENTKQ-D----------------I--QKKLQQF-RMKAIYQLYSHQNRALKKIFQND---------KAHSGVIILPCGAGKTLLG
At-1   FRNDN--------------------------VNPDLDM-ELKPHAQPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
At-2   FRNDN--------------------------VNPDLDM-ELKPHAQPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
Orys-1 FRNDT--------------------------VNPDLEM-ELKPQARPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG
Orys-2 FRNDT--------------------------VNPDLEM-ELKPQARPRPYQEKSLSKMFGNG---------RARSGIIVLPCGAGKSLVG


                            WM IA
                            ****
Hs     VTAACTVRKRCLVLGNSAVSVEQWKAQFKMWSTIDD-----------------------------------SQICRFTSDAKDK-PIG--
Mm     VTAACTVRKRCLVLGNSAVSVEQWKAQFKMWSTIDD-----------------------------------SQICRFTSDAKDK-PIG--
Drer   VTAACTVRKRCLVLGNSSVSVEQWKAQFKMWSTIDD-----------------------------------SQICRFTSDAKDK-PIG--
Dm     VTACCTVRKRALVLCNSGVSVEQWKQQFKMWSTADD-----------------------------------SMICRFTSEAKDK-PMG--
Ce     VTAVTTVNKRCLVLANSNVSVEQWRAQFKLWSTIQD-----------------------------------KQLVRFTREAKDP-APSGA
Sc     ITAACTIKKSVIVLCTSSVSVMQWRQQFLQWCTLQP-----------------------------------ENCAVFTSDNKEM-FQT--
Sp     ITAACTIKKSVIVLCTSSVSVMQWRQQFLQWSNIKP-----------------------------------DHIAVFTADHKER-FHS--
Ca     ITAACTIRKSVIVLCTSSVSVMQWRQQFLQWCTIQP-----------------------------------ENVAVFTSENKEM-FAS--
Nc     ITAACTIKKGVIVLCTSSMSVVQWRQEFLKWSNINP-----------------------------------DDIAVFTAESKNK-FSG--

Tb-1   IGAAATVKKRTIVMCINVMSVLQWQREFIRWTNLSE-----------------------------------DQVTVCIADKKQM------
Tco-1  IGAAATVKKRTIVMCINVMSVLQWQREFLRWTNLTE-----------------------------------EQVTVCISDKKQM------
Tv-1   IGAATIVKKRTIVMCINVMSVLQWQQEFVRWTNLTE-----------------------------------DQVTVCISDKKQM------
Tc-1   IGAATILKKRTIVMCINVISVLQWQREFFRWTDLSE-----------------------------------DQVTVCISDKKQM------
Lm-1   IGAAATMQTTTIVMCINHMSVFQWQREFLRWTNLAE-----------------------------------EEVTVCTAKVKQR------
Li-1   IGAAATMQTTTIVMCINNMSVFQWQREFLRWTDLAE-----------------------------------DEVTVCTAKVKQR------
Lb-1   IGAAATMQTTTIVMCINNMSVFQWKREFLRWTDLTE-----------------------------------DEVTVCTAKVKQR------
Lmx-1  IGAAATMQTTTIVMCINNMSVFQWQREFLRWTDLAE-----------------------------------DEVTVCTAKVKQR------
Bs-1   IAASSIVSKKTIVMCCNVLSVMQWRKEFLRWTDLRE-----------------------------------DQVTVCVRGMKQR------

Tb-R   IMLLCKVKKPTLILCAGGVSVEQWRNQILEFASVCAPANNEEGDPSNS-----T-----TGEKVRTAAVGAARISCLTGKQKDE-I----
Tco-2  VMLLCKVKKPTLILCTGGISVEQWRNQILEFASLSAPSGGEEGSSSTAC----A-----TSKRARPPAVGAARISCLTGKQKDE-I----
Tv-2*  IMLLCKVKKPTLILCAGSVSVEQWKNQILEFSTLGCSLGNDENRSSSGS---GT-----ASGRQRTRIEGVARISCLTAKQKDD-I----
Tc-2   IMLLCKVKKPTLILCAGSVSVEQWKNQILEFASICAHDTDDEM----------A-----AAEKHRSRLEGAARIACLTGKQKDP-V----
Lm-2   IMLMCKVKKPTLVLCAGSVSVEQWKSQILDYATLETPPRTDDGDASDAH-----------GGRPHQIKDGAARIACLTGKQKDP-I----
Li-2   IMLMCKVKKPTLVVCAGSVSVEQWKSQILDYATLEAPPPTDDGDAANAH-----------SGRSHQIKNGAARIACLTGKQKDP-I----
Lb-2   IMLMCKVKKPTLVVCAGSVSVEQWRSQILDYATLEAPPPTDNDDAADTQ-----------GGCSHQIKSGATRIACLTGKQKDP-I----
Lmx-2  IMLMCKVKKPTLVVCAGGVSVEQWKSQILDYAVLEAPPPTDDGDAADAH-----------GGRPHQIKNGAARIACLTGKQKDP-I----
Bs-2   VLLAAVVKKRTIIVCAGSISVEQWKNQLLEWADLTPLDEEDTATTHNIAARGGRGGGRGGSRGGQGRAVGGDRVACLTGKHKDP-I----

Ng-1   VTAASTIKKHTLVLCINTVSVFQWKNQFRLWSTIDD-----------------------------------DSILVHTSQQKAT-LPK--
Ng-2   -----SAVETFYLFEINPTECIEWKNQFKLWSTIDE-----------------------------------NRVVVFTSQEKDA-LPK--
Trva-1 ITAVATINKPAVVVCNSVEPVKQWANQFNLWTNVPG-----------------------------------GVVVVLTSENKQA-LP---
Trva-2 IAALATINKPTVIVCNNRLTVKQWYNQILQYATMDM-----------------------------------KKIFLFSDTEKQA-LPQ--
Gl     VACACRLGHSCIVVTNGNLSSKQWKSQFLQFSTVES-----------------------------------DRVYIFSANKDEETEFI--
Ehis-1 IAACSKIKRSTIVLTHTTQSVFQWKEEFLKWSTIKP-----------------------------------DAIKLCVSSKKEQ-LGD--
Ehis-2 IGIICTFQQPTIIICTTTISIEQWRDEIRKWSTIPL-----------------------------------NKIKCFSSIIKET------
Edis-1 IAACSKIKRSTIVLTHTTQSVFQWKEEFLKWSTIKP-----------------------------------DAIKLCVSSKKEQ-LGD--
Edis-2 IGIICKFQEPTIIICTTTISIEQWRDEIRKWTNIPL-----------------------------------NKIKCFSSIIKEE------
Dd     ITAACTVKKSILVLCTSAVSVEQWKYQFKLWSNIEE-----------------------------------RQISKFTSDNKEK-ISN--
Ac**   ITVSSIIKKCILVFCTSMMAVNQWREQFLKWSDIEP-----------------------------------SRISRFISDTRKKGEWD--
Pf     ITAASTIKKSALFLTTSAVAVEQWKKQFEDFTNIHP-----------------------------------RHIRILTSDYKLD-LWP--
Toxg   ITAACTLRKSVMILTTSAVAVDQWRKQFEEYTTIDS-----------------------------------SRLLTLTAETKQT-LWP--
Eimt** ITAVCTLQRSAVVLTTSAVAVEQWKRQFLENSSIKE-----------------------------------KEIMLLTAESKQQ-LWG--
Ta     IVAACTVRKSIFVLTTSAVAVEQWIKQFLDFTNISR-----------------------------------TNLISLTSEHKTD-LWD--
Tett-1 VIAIERIKQSTVIICDSDVSVDQWRDELERWTTINK-----------------------------------NKIVRLTGRIVDE--WPTR
Tett-2 IIVTEKIKGNTIVICDIDTATKQWKNEFIRWTTVKP-----------------------------------DRIKIYTGKTKNL--QE--
Part-1 VLCMSLIKSSTVIICDSNVSVEQWKREIEGYTTIDN-----------------------------------RRIIRVTGFAKDK--WQG-
Part-2 INVALKIKRKTLIICDQVNDVYQWQKSFIKFTEMDK-----------------------------------NNLAIILRTQQEV--PMAV
At-1   VSAAARIKKSCLCLATNAVSVDQWAFQFKLWSTIRD-----------------------------------DQICRFTSDSKER-FRG--
At-2   VSAAARIKKSCLCLATNAVSVDQWAYQFKLWSTIKD-----------------------------------DQICRFTSDSKER-FRG--
Orys-1 VSAACRIKKSCLCLATNAVSVDQWAFQFKLWSTIKD-----------------------------------EHISRFTSDNKEK-FRG--
Orys-2 VSAACRIKKSCLCLATNAVSVDQWAFQFKLWSTIKD-----------------------------------EHISRFTSDNKEK-FRG--
                    .        :*  .:
```



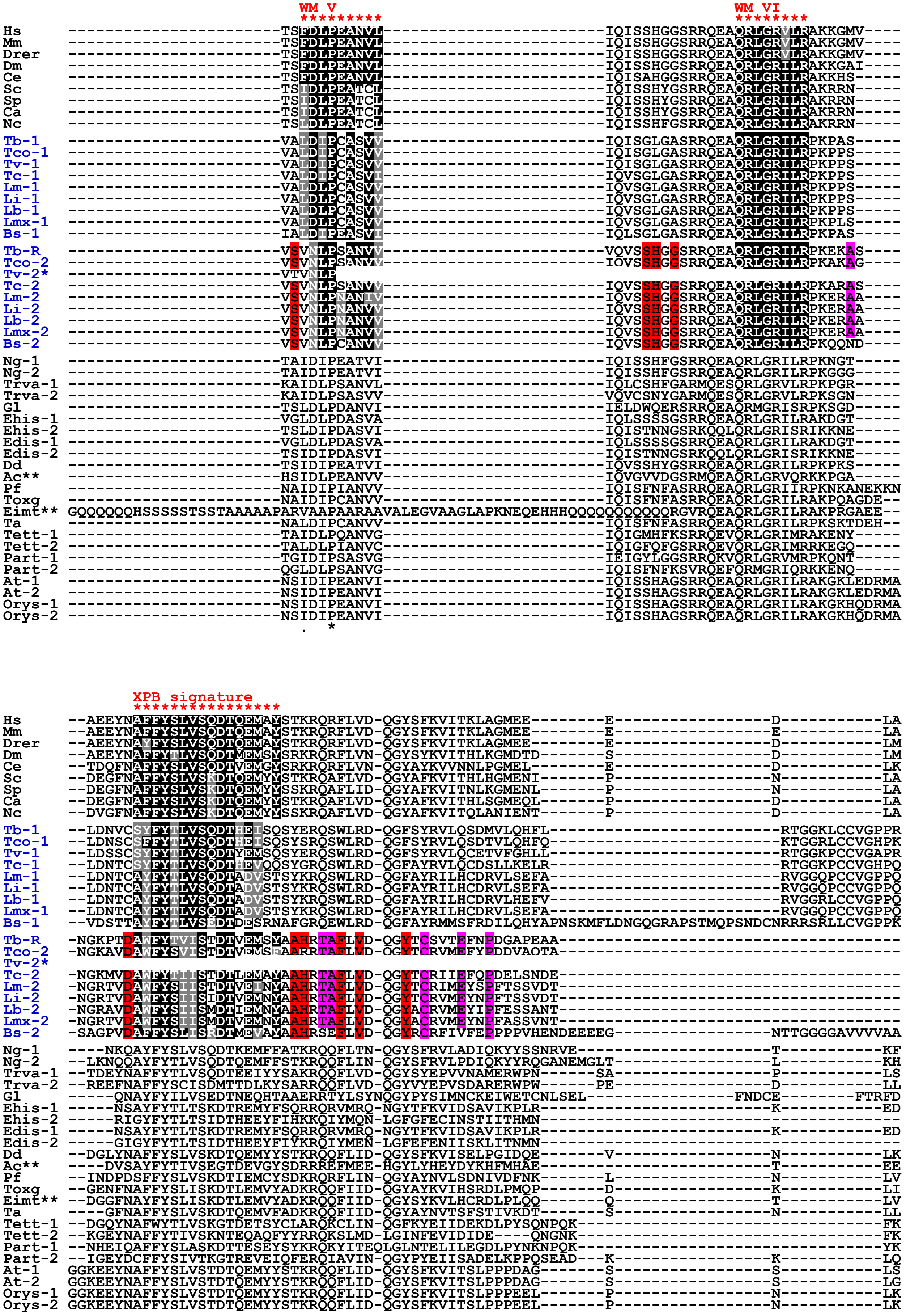
```
                                                                                         WM II
                                                                                         ****
Hs     ----CSVAISTYSMLGHTTKR----------------------------------------SWEAERVMEWLKTQEWGLMILDEVHTIPA
Mm     ----CSVAISTYSMLGHTTKR----------------------------------------SWEAERVMEWLKTQEWGLMILDEVHTIPA
Drer   ----CSVAISTYSMLGHTTKR----------------------------------------SWEAERVMEWMKSQEWGLIILDEVHTIPA
Dm     ----CGILVTTYSMITHTQKR----------------------------------------SWEAEQTMRWLQEQEWGIMVLDEVHTIPA
Ce     DATKPVVCISTYSMVAYSGRR----------------------------------------TLQAEEAMKFIEKQEWGLLLLDEVHTIPA
Sc     ---ESGLVVSTYSMVANTRNR----------------------------------------SHDSQKVMDFLTGREWGFIILDEVHVVPA
Sp     ---EAGVVVSTYSMVANTRNR----------------------------------------SYDSQKMMDFLTGREWGFILLDEVHVVPA
Ca     ---ESGLVVSTYSMVANTRNR----------------------------------------SHDSQKVMDFLRSREWGFIILDEVHVVPA
Nc     ---NTGIIVTTYSMVTNNRER----------------------------------------SHDSKKMMDFLKSREWGLMLLDEVHVVPA

Tb-1   ---PGDVFITTYSMLIARRSNVPEM-EQSAD------------------------------AKLTAKILASVGEQPWGLLLLDEVHTALA
Tco-1  ---PGDVFITTYSMLIAKRPNVSDQ-EQSAD------------------------------AKLTSKILASVSEQPWGLLLLDEVHTALA
Tv-1   ---PGDVFITTYSMLIARRTSNLGQ-EQGED------------------------------AKLTSKILFAVEEQPWGLLLLDEVHTALA
Tc-1   ---PGDVFITTYSMLVAKRPTAPEQ-EQGED------------------------------ARLTSRILASVEEQAWGLLLLDEVHAALA
Lm-1   ---PGKVFITTYSMVVAKRGSADG-----AA------------------------------AEESRAILQVMTAQPWGLLLLDEVHTALA
Li-1   ---PGKVFITTYSMVVAKRGSTDG-----AA------------------------------AEESRAILQAMTAQPWGLLLLDEVHTALA
Lb-1   ---PGKVFITTYSMVIAKRGNADG-----AA------------------------------AEESRAILQAMAAQPWGLLLLDEVHTALA
Lmx-1  ---PGKVFITTYSMVVAKRGNADG-----AA------------------------------AEESRAILQAMTAQPWGLLLLDEVHTALA
Bs-1   ---PGDVFITTYSMLTAQRATRPADAVVTED------------------------------ALLTKEILSDIANSTWGMLLLDEVHIAPA

Tb-R   -TDETDIVLTTYSMLVTAHKAQARCQVEGFE----------M------------NADGRGRNPRRANPKERLFQ-PYGLLIMDEVHMMPA
Tco-2  -TDDTDIVLTTYNMLVTAHKMQRRQQCSSAG----------G------------ATGEQRAKTRRPNPKERLFQ-PYGLLIMDEVHVMPA
Tv-2*  -TDDTDIVLTTYNMLVTAHKAQARYHAEHCK----------V------------AVDTQRKRGRRSNPKERLFQ-PYGLLIMDEVHVMPA
Tc-2   -TEETDIVLTTYSMLVTAHRAQMRQRAAHAS----------G------------LADDRHQKRRRANPKEKLFQ-PYGLLIMDEVHVMPA
Lm-2   -TEETDVVLTTYSMLVTAHRAKMRQQATTTS-------SEDG------------LFTGRGLRKRKENPKEKLFA-PYGLLILDEVHVMPA
Li-2   -TEETDVVLTTYSMLVTAHKAKMRQQATTTS-------SEDG------------LFTERGLRKRKENPKEKLFA-PYGLLILDEVHVMPA
Lb-2   -TEETDVVLTTYSMLVTAHKAKLRQHATTTA-------GEDD------------LFTGRGLRKRKENPKEKLFA-PYGLLILDEVHVMPA
Lmx-2  -TEETDVVLTTYSMLVTAHKAKMRQQATAIS-------SEDG------------LFTGRGLRKRKENPKEKLFA-PYGLLILDEVHVMPA
Bs-2   -TEDTDVVITTYTMLGVAQEHYRREQQQRLERRKLQNRGEDDDDMTRRSCGAAPDRRSSAQSKRSRNPKEALFDTTWGLMILDEVHMAPA

Ng-1   DPNKAVVLITTYSMISFKGRR----------------------------------------SEESKQMMEYIENREWGLLMFDEVQVYPA
Ng-2   DDD-PLIIITTYTMITHTRKR----------------------------------------SGEAKEVMDYIQSREWGLLILDEVHVVPA
Trva-1 --DKPCILVTTYGMLTSSARR----------------------------------------SAESQKVIDQITGRDWGILVMDEVQEAAA
Trva-2 --TGPCIVISTYSMLSNPNKR----------------------------------------SDKSQQIIDQIKSRDWGLLILDEVQDSAA
Gl     -PGYHIVFISTYQMLTNAKSA------------------------------------------KSKETLNKVHSLMWGIVLFDEAHQVMA
Ehis-1 ---DACVLITSYSMLSHTGER----------------------------------------EYAGQRIIDDLLKREWGFIIFDEVHGSTT
Ehis-2 -IEESYIVITTYQMIQS-------------------------------------------------KEIEEIRKKKYGMMILDEVHSSVA
Edis-1 ---DACVLITSYSMLSHTGER----------------------------------------EYAGQRIIDDLLKREWGFIIFDEVHGSTT
Edis-2 -IX-XYIIITTYQMIQS-------------------------------------------------KEIEEINKKRYGMMILDEVHSSVA
Dd     ---VAGVTITTYTMVAFGGRR----------------------------------------SAESLKIMNEITNREWGLVLLDEVHVVPA
Ac**   --HTCGLLITTYNMFTS-EKR----------------------------------------AEHSTRMIEKCKERTWGLMILDEVHLAPA
Pf     -INEAGVLISTYTMLSYSGKR----------------------------------------SEQSLRIVNDIRRREWGLLVFDEVQFAPA
Toxg   -VDEAGVLVSTYTMLAFSGRR----------------------------------------SLAAERIMQQIREREWGLLIFDEVQFAPA
Eimt** -PAAAGVLLSTYTMLAFSGRR----------------------------------------SAAAAAVLSQIQSREWGILIFDEVQFAPA
Ta     -EKDAGVLISTYTMMSYTRKH----------------------------------------RENTERILNQIKQREWGMLIFDEVQFVPA
Tett-1 TTDEPIIVITTYYMLVKKREDLPPT-----Q-------------I---------AQKS---K--KRQYIDSMKERKWGVCVIDEVHKLPA
Tett-2 KQTKPFILITTFKQLSSIIKSNKLQKEKRQD-------------M---------DTSS---KYETKMDHYIQEEMEWDLCIADETQYSAA
Part-1 --EQPIVILTTYSWLIAQFRNN----------------------------------SS---S--TKTVWNQISEVSWGICIVDEVHRLPA
Part-2 LGREHIIVITNKDMISS-----------------------------------------------NRKDIKHVTTLEWPLLIMDEVHGLPA
At-1   ---NAGVVVTTYNMVAFGGKR----------------------------------------SEESEKIIEEMRNREWGLLLMDEVHVVPA
At-2   ---NAGVVVTTYNMIAFGGKR----------------------------------------SEEAEKIIEEMRNREWGLLLMDEVHVVPA
Orys-1 ---MAGVVVTTYNMVAFGGKR----------------------------------------SEDSEKIIEEIRNREWGLLLMDEVHVVPA
Orys-2 ---MAGVVVTTYNMVAFGGKR----------------------------------------SEDSEKIIEEIRNREWGLLLMDEVHVVPA
             : ::.   .                                                             : . : **.:   :


                                 WM III RED                                       ThM domain
                                 *****  ***                                       ************  *
Hs     KMFRRVL----------TIVQAHCKLGLTATLVREDDKIVDLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAI--K
Mm     RMFRRVL----------TIVQAHCKLGLTATLVREDDKIVDLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAI--K
Drer   KMFRRVL----------TIVQAHCKLGLTATLVREDDKIVDLNFLIGPKLYEANWMELQNNGYIAKVQCAEVWCPMSPEFYREYVAI--K
Dm     KMFRRVL----------TIVQSHCKLGLTATLLREDDKIADLNFLIGPKLYEANWLELQKKGYIARVQCAEVWCPMSPEFYREYLTT--K
Ce     KMFRRVL----------TIVQAHCKLGLTATLVREDDKITDLNFLIGPKIYEANWMELQKAGHIAKVQCAEVWCPMTSAFYSYYLRS--Q
Sc     AMFRRVV----------STIAAHAKLGLTATLVREDDKIGDLNFLIGPKLYEANWMELSQKGHIANVQCAEVWCPMTAEFYQEYLRE--T
Sp     AMFRRVV----------TTIAAHTKLGLTATLVREDDKIDDLNFLIGPKMYEANWMDLAQKGHIAKVQCAEVWCAMTTEFYNEYLRE--N
Ca     QMFRRVV----------TTIAAHAKLGLTATLVREDDKIDDLNFLIGPKLYEANWMDLAQKGHIANVQCAEVWCPMTAEFYQEYLRE--N
Nc     DVFRRVI----------SSIKSHSKLGLTATLLREDDKISHLNFLIGPKLYEANWMELSEKGHIAKVQCAEVWCPMPTEFYDEYLRA--N

Tb-1   HNFQEVL----------NKVKYKCVIGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTRAGFLARVECAEIQCPLPKAFLTEYLESQSD
Tco-1  HNFREVL----------NKVKFKCVVGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTRAGFLARVECAEVQCPLPRPFLVAYLDNQKQ
Tv-1   QNFQEVL----------NKVKYKCVVGLSATLLREDGRIDDLRHLVGPKLYEANWLDLTNAGFLARVECAEVQCPLPISFFSEYLIAQQE
Tc-1   HHFQEVL----------NKVKYKCVVGLSATLLREDDKIGDLRHLVGPKLYEANWLELTRAGFLARVECAEVQCPLPLPFFREYLDSQAE
Lm-1   HHFQDVL----------NTIKYKCVLGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTRAGFLANVECAEVQCPMPPLFLQEYYGIQRT
Li-1   HHFQDVL----------NTIKYKCVLGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTCAGFLANVECAEVQCPMPPLFLQEYHDIQST
Lb-1   HHFQDVL----------NTIKYKCVLGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTRAGFLANVECAEVQCPMPPLFLQEYRDIQRT
Lmx-1  HHFQDVL----------NTIKYKCVLGLSATLLREDDKIGDLRHLVGPKLYEANWLDLTRAGFLANVECAEVQCPMPPLFLEEYHAIQRT
Bs-1   QRFQEVI----------NCVKHHCVLGLSATLLREDQRIDNLRFLVGPKLYEANWLELTNQGFLADVVCVDVRCPMPPAYLRMYLDPQST

Tb-R   DAYKDSL----------GFINAKGVVGLTATYVREDSKIRDLFHLVGPKLFDISWERLASSGYLAHVTCIEVLTPLARRFSLEYLERSSE
Tco-2  DVYKESL----------GYVNSKGVIGLTATYVREDAKIRDLFYLVGPKLFDMSWERLASSGYLANVTCVEVLTPLTRQFNLEYMERSGS
Tv-2*  ETYKESL----------GFINAKGVIGLTATYVREDAKIRDLFHLVGPKLFDVSWETLASSGYLANVTCVEILTPLTRQFSLEYMERSSS
Tc-2   EAYKESL----------SLVDAKGVVGLTATYVREDAKIQDLFHLVGPKLYDVSWETLASSGYLANVTCIEVLTPLTRQFSLEYMQRSGE
Lm-2   ESFRGSL----------GFVDAKGVIGLTATYVREDHKILDLFHLVGPKLYDISMETLASQGYLAKVHCVEVRTPMTKEFGLEYMQRSSK
Li-2   ESFRGSL----------GFVDAKGVIGLTATYVREDHKILDLFHLVGPKLYDISMETLASQGYLAKVHCVEVRTPMTKEFGLEYMHRSSK
Lb-2   ESFRGSL----------GFIDAKGIVGLTATYVREDHKILDLFHLVGPKLYDISMEALASQGYLANVHCVEVRTPMTKEFGLEYMDHRRR
Lmx-2  ESFRESL----------GFVDAKGVVGLTATYVREDHKILDLFHLVGPKLYDISMETLASQGYLSKVHCVEVRTPMTKEFGLEYMDRSSK
Bs-2   AGVRESI----------SHVVSKAVVGLTATYVREDHRIRDTFHVVGPKLFDISWEELHQQGFLASITCVEVHCPMTPDFMVEYTQRLHT

Ng-1   DSFRA------------SHLKAHCKLGLTATLLREDNKQEDLNFLIGPKLYEANWLDLQKKGHLANVQCVEVWCPMTAEFYAEYLSS--N
Ng-2   NMFRHV-----------ASVKSHCKLGLTATLLREDNKTDDLYFLIGPKLYEANWLDLQQKGHLANVQCVEVWCPMTAEFFAEYLTA--N
Trva-1 EKFRNVT----------DMVKAHTRLGLTATMVREDGKIDDLKYLVGPKLYEANWIELSESGYLARVKCFEILVTMSQEFYKKYLTLDKD
Trva-2 NTFRNVT----------DIAKAHTRLGLTATLIREDDKISDLRYLVGPKLYEANWLELSEQGYLARVKCFEVTVPMTASFYKYYLLS--D
Gl     ETYRTLFVTDNVNTKANLFLRSYCKIGLTATPLREDDEMTNL-YLLGNKLYESNWSDLAKRGFIAKLQCAEIRCPMHHIFYDDWLDQNTA
Ehis-1 DNIEKFV----------CKIKAQCKLGLTATLIREDDRIRDLEFMIGPMLYEASWQELAKQGYIANAKCFEVICPMTKTYYSAYVEAD-D
Ehis-2 KEFRKIY----------YLIEARCRIGITATPIREDCKIKDLDYLIGPILYQESWKELIKEKYIANVQCIDIKCPMTIPFYEEYINNS-N
Edis-1 DNIEKFV----------CKIKAQCKLGLTATLIREDDRIRDLEFMIGPMLYEASWQELAKQGYIANAKCFEVICPMTKTYYSAYVEAD-D
Edis-2 KEFRKIY----------YLIETRCKIGITATPIREDYKIKDLDYLIGPILYQENWKDLIKEKYIAKVQCIEIKCPMTIPFYEEYINNS-K
Dd     AMFRKVL----------TVTKAHCKLGLTATLLREDEKIQDLNFLIGPKLYEANWLDLQKAGFLANVSCSEVWCPMTAEFYKEYLIND-S
Ac**   TTFKRVT----------TEFRAHVKLGLTATMVREDELIAELPTLVGPRLDEVDLLSLKIHNHVANVHCYRIVCPLTSAFGAAYRTAR-S
Pf     PSFRRIN----------DIVKSHCKLGLTATLVREDLLIRDLHWIIGPKLYEANWVELQNKGFLAKALCKEIWCSMPCSFYKYYLKS--N
Toxg   PAFRRIN----------DLVKSHCRLGLTATLVREDDLIKDLQWLIGPKLFEANWIELQDQGFLARVSCQEVWCPMTADFYREYLRC--S
Eimt** PAFRKIN----------ALVRAHCRLGLTATLVREDDLIKDLQWIIGPKLYEANWIDLQNQGYLAKVSCQEVWCPMTAEFYREYLRA--P
Ta     PAFRRIN----------EIIRSHCKLGLTATLVREDDLIRDLQWLIGPKLYEANWLELQQKGYLAKVICKEIWCPMTAPFYREYLRS--S
Tett-1 NTFQNVL----------KQYKFHFKLGLTATPYREDEKIINLFYMIGPKLYEENWYDLVSQGFLAKPYCVEIRCEMSQLWMSEYNKKDLK
Tett-2 STYKKIF----------EEFNFKLKIGLTATPYREDNKIDDLFHVIGPKLYEANISELIQDSYLAKPYCVVFRCKMHECCEDFMKDR---
Part-1 VQFQNVL----------KQIKCAIKIGLTATLLREDQKLDNLYFMIGPKLYEENLIDLMTQGFLAKPHIIEIQCDMPPIFLQEYQTKNNM
Part-2 EQINAEI----------SKLKANMKIGLTATPYRQDNKIKEIFYKVGPKLHESMIVDLKQMGYVSKIYCIQVYVGMQDLYKQKYEEYRRV
At-1   HMFRKVI----------SITKSHCKLGLTATLVREDERITDLNFLIGPKLYEANWLDLVKGGFIANVQCAEVWCPMTKEFFAEYLKKE-N
At-2   HMFRKVI----------SITKSHCKLGLTATLVREDEKITDLNFLIGPKLYEANWLDLVKGGFIANVQCAEVWCPMTKEFFAEYLKKE-N
Orys-1 HMFRKVI----------SITKSHCKLGLTATLVREDERITDLNFLIGPKLYEANWLDLVKGGFIANVQCAEVWCPMTKEFFAEYLKKE-N
Orys-2 HMFRKVI----------SITKSHCKLGLTATLVREDERITDLNFLIGPKLYEANWLDLVKGGFIANVQCAEVWCPMTKEFFAEYLKKE-N
          .                     :*::**  *:*    .    :*  : :     *    .::      .   :
```



```
                                                                    ***********
Hs     -------------------------------------------------------------TKKRILLYTMNPNKFRACQFLIKFHE-RR
Mm     -------------------------------------------------------------TKKRILLYTMNPNKFRACQFLIKFHE-RR
Drer   -------------------------------------------------------------TKKRILLYTMNPNKFRACQFLIRFHE-RR
Dm     -------------------------------------------------------------TSKKMLLYVMNPSKFRSCQFLIKYHE-QR
Ce     -------------------------------------------------------------IARKLLLAVMNPNKFRICQFLIKFHE-RR
Sc     -------------------------------------------------------------ARKRMLLYIMNPTKFQACQFLIQYHE-RR
Sp     -------------------------------------------------------------SRKRMLLYIMNPKKFQACQFLIDYHE-KR
Ca     -------------------------------------------------------------ARKRMLLYIMNPTKFQACQFLIHYHE-KR
Nc     -------------------------------------------------------------SRMKRTLYAMNPRKFQACQYLINYHE-AR

Tb-1   GDPFA--------------------------------------------RRGT--------TRMAHSVVCLNPYKLWCTQALLEFHR-NR
Tco-1  GDLST--------------------------------------------RRGT--------SPMARAVVCYNPYKLWCAQALLEFHR-NR
Tv-1   EDPSA--------------------------------------------RRGT--------SSLARALVCLNPYKLWCTQALLEFHR-NR
Tc-1   EDPFA--------------------------------------------RRGT--------SSLARAVVCFNPYKLWCTQALLEFHR-NR
Lm-1   RALLGAHARRNHRGGCGGNGKRCRESVFQDDEDAGEDGDSDGEDNGARGGRGRGHDAYGPLTSRSLRLASCNPYKLWCAQALLAFHQ-QR
Li-1   RALLGAHARLNHRGGRRGNGKRCRESVFQDDEDAGEDGDSDGEDNGARGGRGRGHDAYGSLTSRSLRLASCNPYKLWCTQALLAFHQ-QR
Lb-1   RTLLGTHAHRSHSGSRAGNGKRRRESVFQDDEDAREDEDHDEENGGAYRGRGRGHDLCGYLPSRSLRLASCNPYKLWCTQALLAFHQ-QR
Lmx-1  RALLGAHARCNHRR---GNGKRRRESVFQDDEDASEDGDGNGEDNGAYGRRGRGHDSCGPLTSRSLRLASCNPYKLWCTQALLAFHQ-QR
Bs-1   DESLC-------------------------------------------------------------DISSVNPHKLWATQALLHFHG-SR

Tb-R   L------------------------------------------------------TSPQHGTPLLVMLAAANPNKMLCVMEIVKRHV-AE
Tco-2  GF-----------------------------------------------------TVQRRKVPLLVMLAAANPNKMLCVAELVQRHI-VE
Tv-2*  DS-----------------------------------------------------TGAHRRIPLLVMLAAANPNKMLCVMELIRRHL-AE
Tc-2   DH-----------------------------------------------------TLQQRKMPLLVMLAAANPNKMLCVLELVRRHV-AD
Lm-2   V-------------------------------------------------------ARAGTAPVLVMLAAANPNKMMCVRELVRQHL-DA
Li-2   T-------------------------------------------------------ARARTAPVLVMLAAANPNKMMCVRELVRQHL-DA
Lb-2   A-------------------------------------------------------AGTSKTSVLVMLAAANPNKMMCVRELVWQHL-DV
Lmx-2  A-------------------------------------------------------ARAGKAPALVMLAAANPNKMMCVRELVWQHL-DA
Bs-2   SLSSRSNSSKSSSA-------------------------------------G---VSALAAAPMLQAIAASNPNKMLCVWELVHRHL-HQ

Ng-1   -------------------------------------------------------------YAKQSLYYVMNPNKFRACEFLIRYHE-KH
Ng-2   -------------------------------------------------------------AKKKTLLYVMNPNKFRACEFLIRYHE-KQ
Trva-1 -------------------------------------------------------------PIKRRVLAASNPNKIDVLESLLQYHE-AR
Trva-2 -------------------------------------------------------------HFRQRILCSSNPNKIRTVAGIIKFHE-RR
Gl     QATNQ---DR--T-------TVCV------------------------------NDVMITRKTYMKYLTILNPYKVQTAWYLKEYHT-RR
Ehis-1 -------------------------------------------------------------SKLKQCLAQLNPNKIDACKYLLEQHK-AH
Ehis-2 -------------------------------------------------------------YRHKIILSALNPNKIEVTKFLIKQHL-KR
Edis-1 -------------------------------------------------------------SKLKQCLAQLNPNKIDACKYLLEQHK-AH
Edis-2 -------------------------------------------------------------YRHKIILTALNPNKIEVIKFLIKEHL-KR
Dd     -------------------------------------------------------------QGKKKLLYTMNPNKFRACEYLIRFHE-QR
Ac**   -------------------------------------------------------------PAEKRLLHITNPNKARIVFTLLKRHL-RR
Pf     -------------------------------------------------------------SFIKRRLYTCNPRKLMMCEYLIKYHE-QN
Toxg   -------------------------------------------------------------HAKQRKLWVCNPTKLMTCEWLLRYHE-AR
Eimt** -------------------------------------------------------------YAKQRKLWVCNPSKLMVCEFLLKFHE-AK
Ta     -------------------------------------------------------------SVKKRRLWSCNPVKLITCEYLLRFHE-SR
Tett-1 TG------------------------------------------------------KNYYRGPQRELIHTSNPRKFKTLEYLIKVHE-ER
Tett-2 ---------------------------------------------------------KF---KNKPSLQSGNPEKFKLLYYLIKFHE-SR
Part-1 --------------------------------------------------------------TVRQLLHTGNPGKYKALQFLIKNHE-ML
Part-2 NN-----------------------------------------------------------QFVTNTLYQMNPKKFEVLQSLINIHR-TR
At-1   -------------------------------------------------------------SKKKQALYVMNPNKFRACEFLIRFHEQQR
At-2   -------------------------------------------------------------SKKKQALYVMNPNKFRACEFLIRFHEQQR
Orys-1 -------------------------------------------------------------SKKKQVLYAMNPNKFRACEFLIRFHEQQR
Orys-2 -------------------------------------------------------------SKKKQVLYAMNPNKFRACEFLIRFHEQQR
                                                                              ** *      :   *
```



```
                                               WM IV
                                               *******
Hs     ----NDKIIVFADNVFAL------KEYA--IRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGD--------------------
Mm     ----NDKIIVFADNVFAL------KEYA--IRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGD--------------------
Drer   ----NDKIIVFADNVFAL------KEYA--IRLNKPYIYGPTSQGERMQILQNFKHNPKINTIFISKVGD--------------------
Dm     ----GDKTIVFSDNVFAL------KHYA--IKMNKPFIYGPTSQNERIQILQNFKFNSKVNTIFVSKVAD--------------------
Ce     ----NDKIIVFSDNVFAL------KRYA--IEMQKPFLYGETSQNERMKILQNFQYNPRVNTIFVSKVAD--------------------
Sc     ----GDKIIVFSDNVYAL------QEYA--LKMGKPFIYGSTPQQERMNILQNFQYNDQINTIFLSKVGD--------------------
Sp     ----GDKIIVFSDNVYAL------RAYA--IKLGKYFIYGGTPQQERMRILENFQYNELVNTIFLSKVGD--------------------
Ca     ----GDKIIVFSDNVYAL------QEYA--LRLGKPFIYGSTPQQERMKILQNFQHNDQINTIFLSKVGD--------------------
Nc     ----GDKIIVFSDELYSL------KQYA--LKLNKVFIYGGTGQAERMQVLENFQHNPQVNTLFLSKIGD--------------------

Tb-1   --SPPDKVIIFCDQIDGI------QYYA--QHLHVPFMDGKTSDMERENLLQYFQHSDNINAIILSRVGD--------------------
Tco-1  --SPPDKVIIFCDQIDGV------QYYA--QHLNVPFMDGRTSDAERANLLQYFQHSSDINAIILSRVGD--------------------
Tv-1   --SPPDKVIIFCDDVEGV------QYYA--QHLHVPFMDGKTSDSERANVLQHFKNSSKTNAIILSRIGD--------------------
Tc-1   --SPPDKVIIFCDDLEGV------QYYA--RHLNVPFMDGKTTEVERENLLQYFQHSNDINAIILSRVGD--------------------
Lm-1   --SPPDKVIIFCDYLADV------RFFA--HHLHLPFMDQRTSEAERTNLLQYFQHSDGVNAIILTRVGD--------------------
Li-1   --SPPDKVIIFCDYLADV------RFFA--HHLHLPFMDRRTSEAERTNLLQYFQHSDSVNAIVLTRVGD--------------------
Lb-1   --SPPDKAIIFCDYLADV------RFFA--HHLHLPFMDQRTSEAERANLLQYFQHSNDVNAIILTRVGD--------------------
Lmx-1  --SPPDKVIIFCDYLSDV------RFFA--HHLHLPFMDQRTSEAERTNLLQYFQHSSGVNAIILTRVGD--------------------
Bs-1   --SPPDKTIIFCDSIPAL------QFYA--TQLCIPFMDGRTPEDERAKLLEYFKTSPNVNTIILSRVGD--------------------

Tb-R   ----SSKILVFCDHIMLL------KEYS--KLLGAPVVCGDTPHRERLMIFSDFQSTSKVNVVCLSRVGD--------------------
Tco-2  ----SSKILVFCDHIVLL------MEYG--KFLGAPVICGDTPHRERLMIFSDFQSTSKVNVICLSRVGD--------------------
Tv-2*  ----SSKILVFCDHIALL------KEYS--KVLNAPVICGETPHRERLMIFSDFQSTSKVNVICISRVGD--------------------
Tc-2   ----SSKILVFCDHLLLL------REYG--TLLHAPVICGQTPHRERLMIFSDFQSTSKLNVICLSRVGD--------------------
Lm-2   ----GAKILLCCDHIMLL------KEYG--ELLNAPVICGSTQHKERLMIFSDFQSTSKINVICVSRVGD--------------------
Li-2   ----GAKILLCCDHITLL------KEYG--ELLNAPVICGTTQHKERLMIFSDFQSTSKINVICVSRVGD--------------------
Lb-2   ----GAKILVCCDHIMLL------KEYG--ELLNAPVICGSTQHKERLMIFSDFQSTSKINVICVSRVGD--------------------
Lmx-2  ----GAKILLCCDHIMLL------KEYG--ELLNAPVICGSTQHKERLMIFSDFQSTSKINVICVSRVGD--------------------
Bs-2   QEHHTAKILVFCDHIALL------KEYA--KLLGAPLVHGETSHAERMAIFSEFQSTRRVNVICISRVGD--------------------

Ng-1   ----GDKIIVFSDNTFTL------EQYA--IKLKLPYIYGKTRESERLDVLNRFKS-GNFNTVFLSKVGD--------------------
Ng-2   ----GDKIIVFSDNVFAL------QEYA--EKLGVPYIFGNTKQQERVYYLNQFRNTNHYNTIFISKVGD--------------------
Trva-1 ----GDKILVFCDYLQII------EKLA--RRLDRPHLHGGISNNERTKLFNKFKTTKKVNTIILSKIGD--------------------
Trva-2 ----GDKVLVFCDIIHIL------IHLA--GLLHCPEIHGETPENVRSSIFHEFKNGSKVNTLILSSVGD--------------------
Gl     ----GDQVLLFCDTIMAA------KMYA--RMFDVPFIMGACSDDERECLVNAFA-TRTISCLMLSSVGD--------------------
Ehis-1 ----GDKIIIFCNELKPA------GFYKEKLKLQKCYMDGNTSEEHRRNLLDQFRR-DEISVIFCSKIGD--------------------
Ehis-2 ----NEQIILFCDSIIVL------KEIG--YQINCPIFFGESPNKERLDILNKFKL-KDINCIGMSSVGD--------------------
Edis-1 ----GDKIIIFCNELKPA------GFYKEKLKLQKCYMDGNTSEEHRRNLLDQFRR-DEISVIFCSKIGD--------------------
Edis-2 ----KEQIIIFCDSIIVL------KEIS--YQINCPIFFGESPNKERIDILNKFKL-KEINYIGMSSVGD--------------------
Dd     ----GDKIIVFSDNVYAL------QKYA--KGLGRYFIYGPTSGHERMSILSKFQHDPTVRTIFISKVGD--------------------
Ac**   ----GHKCLVFCDDLFGL------QWFS--QVLGSPYIDGETKNEDREKSLNMFRTTKSSAYVLISKVGD--------------------
Pf     ----NDKIIVFSDNIFAL------LHIA--KTLNKPFIYGKLSPIERIAIINKFKHDSSINTILLSKVGD--------------------
Toxg   ----GDKILVFSDNVFAL------LHTA--KALNRPFIYGQVSAVERVAILNKFKHESTFNSLFLSKVGD--------------------
Eimt** ----GDKIIVFSDNVFAL------LHAA--RALNRPFIYGKVCPPERMAILTKMKAPSTPF--FFRKSETMQLTFRAQTSSYKSPSTLLP
Ta     ----GDKVIVFSDNLFAL------LHAA--KLLNRPFIYGKVSSAERIVILNKFKNETTFNTIFLSKVGD--------------------
Tett-1 ----GDKILVFCDRPMII------DYYG--NILKYPVIYGDVSQDERKKIFNLFKVSNQINTIFLSRVGD--------------------
Tett-2 ----KDKILVFCDQIPVL------KYYS--QKMGYPAIYGEVGLLEKLVWLDLFR-KGEINTLFLSRVGD--------------------
Part-1 ----GHKIIVFCDSLLIL------NYYA--LLLGYPVIDGDLNTDEKNKIFSIFKNSNEIKTIFVSRVGD--------------------
Part-2 ----KDKILVFCEKVNKLADTGTLEKFA--KLNNCPIISQKVEQTERSAIYKLYQ-EDKLDVIIFGQIAD--------------------
At-1   ----GDKIIVFADNLFAL------TEYA--MKLRKPMIYGATSHIERTKILEAFKTSKDVNTVFLSKVGD--------------------
At-2   ----GDKIIVFADNLFAL------TEYA--MKLRKPMIYGATSHIERTKILEAFKTSKTVNTVFLSKVGD--------------------
Orys-1 ----GDKIIVFADNLFAL------TSYA--MKLRKPMIYGSTSHVERTRILHQFKNSSDVNTIFLSKVGD--------------------
Orys-2 ----GDKIIVFADNLFAL------TSYA--MKLRKPMIYGSTSHVERTRILHQFKNSSDVNTIFLSKVGD--------------------
             : :: .:                        .        :
```



```
                                WM V                                           WM VI
                                *********                                      ********
Hs     -----------------------TSFDLPEANVL------------------------IQISSHGGSRRQEAQRLGRVLRAKKGMV----
Mm     -----------------------TSFDLPEANVL------------------------IQISSHGGSRRQEAQRLGRVLRAKKGMV----
Drer   -----------------------TSFDLPEANVL------------------------IQISSHGGSRRQEAQRLGRVLRAKKGMV----
Dm     -----------------------TSFDLPEANVL------------------------IQISSHGGSRRQEAQRLGRILRAKKGAI----
Ce     -----------------------TSFDLPEANVL------------------------IQISAHGGSRRQEAQRLGRILRAKKHS-----
Sc     -----------------------TSIDLPEATCL------------------------IQISSHYGSRRQEAQRLGRILRAKRRN-----
Sp     -----------------------TSIDLPEATCL------------------------IQISSHYGSRRQEAQRLGRILRAKRRN-----
Ca     -----------------------TSIDLPEATCL------------------------IQISSHYGSRRQEAQRLGRILRAKRRN-----
Nc     -----------------------TSLDLPEATCL------------------------IQISSHFGSRRQEAQRLGRILRAKRRN-----

Tb-1   -----------------------VALDIPCASVV------------------------IQISGLGASRRQEAQRLGRILRPKPAS-----
Tco-1  -----------------------VALDIPCASVV------------------------IQISGLGASRRQEAQRLGRILRPKPPS-----
Tv-1   -----------------------VALDIPCASVV------------------------IQISGLGASRRQEAQRLGRILRPKPPS-----
Tc-1   -----------------------VALDIPCASVI------------------------IQVSGLGASRRQEAQRLGRILRPKPPS-----
Lm-1   -----------------------VALDLPCASVV------------------------IQVSGLGASRRQEAQRLGRILRPKPPS-----
Li-1   -----------------------VALDLPCASVV------------------------IQVSGLGASRRQEAQRLGRILRPKPPS-----
Lb-1   -----------------------VALDLPCASVV------------------------IQVSGLGASRRQEAQRLGRILRPKPPS-----
Lmx-1  -----------------------VALDLPCASVV------------------------IQVSGLGASRRQEAQRLGRILRPKPLS-----
Bs-1   -----------------------IALDIPEASVI------------------------IQLSGLGASRRQEAQRLGRILRPKPAS-----

Tb-R   -----------------------VSVNLPSANVV------------------------VQVSSHGGSRRQEAQRLGRILRPKEKAS----
Tco-2  -----------------------VSVNLPSANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKAKAG----
Tv-2*  -----------------------VTVNLP
Tc-2   -----------------------VSVNLPSANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKARAS----
Lm-2   -----------------------VSVNLPNANIV------------------------IQVSSHGGSRRQEAQRLGRILRPKERAA----
Li-2   -----------------------VSVNLPNANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKERAA----
Lb-2   -----------------------VSVNLPNANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKERAA----
Lmx-2  -----------------------VSVNLPNANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKERAA----
Bs-2   -----------------------VSVNLPCANVV------------------------IQVSSHGGSRRQEAQRLGRILRPKQQND----

Ng-1   -----------------------TAIDIPEATVI------------------------IQISSHFGSRRQEAQRLGRILRPKNGT-----
Ng-2   -----------------------TAIDIPEATVI------------------------IQISSHFGSRRQEAQRLGRILRPKGGG-----
Trva-1 -----------------------KAIDLPSANVL------------------------IQLCSHFGARMQESQRLGRVLRPKPGR-----
Trva-2 -----------------------KAIDLPSASVV------------------------VQVCSNYGARMQESQRLGRVLRPKSGN-----
Gl     -----------------------TSLDLPDANVI------------------------IELDWQERSRRQEAQRMGRISRPKSGD-----
Ehis-1 -----------------------VGLDLPDASVA------------------------IQLSSSSGSRRQEAQRLGRILRAKDGT-----
Ehis-2 -----------------------TSLDIPDASVI------------------------IQISTNNGSRKQQLQRLGRISRIKKNE-----
Edis-1 -----------------------VGLDLPDASVA------------------------IQLSSSSGSRRQEAQRLGRILRAKDGT-----
Edis-2 -----------------------TSIDIPDASVI------------------------IQISTNNGSRKQQLQRLGRISRIKKNE-----
Dd     -----------------------TSIDIPEATVI------------------------IQVSSHYGSRRQEAQRLGRILRPKPKS-----
Ac**   -----------------------HSIDLPEANVV------------------------IQVGVVDGSRMQEAQRLGRVQRKKPGA-----
Pf     -----------------------NAIDIPIANVV------------------------IQISFNFASRRQEAQRLGRIIRPKNKANEKKN
Toxg   -----------------------NAIDIPCANVV------------------------IQISFNFASRRQEAQRLGRILRAKPQAGDE--
Eimt** GQQQQQQHSSSSSTSSTAAAAAPARVAAPAARAAVALEGVAAGLAPKNEQEHHHQQQQQQQQQQQRGVRQEAQRLGRILRAKPRGAEE--
Ta     -----------------------NALDIPCANVV------------------------IQISFNFASRRQEAQRLGRILRPKSKTDEH--
Tett-1 -----------------------TAIDLPQANVG------------------------IQIGMHFKSRRQEVQRLGRIMRAKENY-----
Tett-2 -----------------------TALDLPIANVC------------------------IQIGFQFGSRRQEVQRLGRIMRRKEGQ-----
Part-1 -----------------------TGIDIPSASVG------------------------IEIGYLGGSRRQKVQRLGRVMRPKQNT-----
Part-2 -----------------------QGLDLPSANVG------------------------IQISFNFKSVRQEFQRMGRIQRKKENQ-----
At-1   -----------------------NSIDIPEANVI------------------------IQISSHAGSRRQEAQRLGRILRAKGKLEDRMA
At-2   -----------------------NSIDIPEANVI------------------------IQISSHAGSRRQEAQRLGRILRAKGKLEDRMA
Orys-1 -----------------------NSIDIPEANVI------------------------IQISSHAGSRRQEAQRLGRILRAKGKHQDRMA
Orys-2 -----------------------NSIDIPEANVI------------------------IQISSHAGSRRQEAQRLGRILRAKGKHQDRMA
                                .  *
```

```
              XPB signature
              ****************
Hs     --AEEYNAFFYSLVSQDTQEMAYSTKRQRFLVD-QGYSFKVITKLAGMEE--------E-----------------D-----------LA
Mm     --AEEYNAFFYSLVSQDTQEMAYSTKRQRFLVD-QGYSFKVITKLAGMEE--------E-----------------E-----------LA
Drer   --AEEYNAYFYSLVSQDTQEMAYSTKRQRFLVD-QGYSFKVITKLAGMEE--------E-----------------D-----------LM
Dm     --AEEYNAFFYTLVSQDTMEMSYSRKRQRFLVN-QGYSYKVITHLKGMDTD-------S-----------------D-----------LM
Ce     --TDQFNAFFYSLVSQDTVEMGYSRKRQRFLVN-QGYAYKVVNNLPGMEL--------E-----------------D-----------LK
Sc     --DEGFNAFFYSLVSKDTQEMYYSTKRQAFLVD-QGYAFKVITHLHGMENI-------P-----------------N-----------LA
Sp     --DEGFNAFFYSLVSKDTQEMYYSSKRQAFLID-QGYAFKVITNLKGMENL-------P-----------------N-----------LA
Ca     --DEGFNAFFYSLVSKDTQEMYYSTKRQAFLVD-QGYAFKVITHLSGMEQL-------P-----------------D-----------LA
Nc     --DVGFNAFFYSLVSKDTQEMYYSSKRQAFLVD-QGYAFKVITQLANIENT-------P-----------------D-----------LA

Tb-1   --LDNVCSYFYTLVSQDTHEISQSYERQSWLRD-QGFSYRVLQSDMVLQHFL-------------------------RTGGKLCCVGPPR
Tco-1  --LDNSCSFFYTLVSQDTHEISQSYSRQSWLRD-QGFSYRVLQSDTVLQHFQ-------------------------KTGGRLCCVGHPK
Tv-1   --LDSSCSYFYTLVSQDTYEMSQSYERQSWLRD-QGFSYRVLQCETVFGHLL-------------------------RTGGKPCCVGAPR
Tc-1   --LDNTCSYFYTLVSQDTHEVQQSYGRQSWLRD-QGFAYRVLQCDSLLKELR-------------------------RTGGKPCCVGHPQ
Lm-1   --LDNTCAYFYTLVSQDTADVSTSYKRQSWLRD-QGFAYRILHCDRVLSEFA-------------------------RVGGQPCCVGPPQ
Li-1   --LDNTCAYFYTLVSQDTADVSTSYKRQSWLRD-QGFAYRILHCDRVLSEFA-------------------------RVGGQPCCVGPPQ
Lb-1   --LDNTCAYFYTLVSQDTADVSTSYKRQSWLRD-QGFAYRILHCDRVLHEFV-------------------------RVGGRLCCVGPPQ
Lmx-1  --LDNTCAYFYTLVSQDTADVSTSYKRQSWLRD-QGFAYRILHCDRVLSEFA-------------------------RVGGQPCCVGPPQ
Bs-1   --VDSTTAYFYTLVSEDTDESRNAFGRQEWLRD-QGFAYRMMSFRDILQHYAPNSKMFLDNGQGRAPSTMQPSNDCNRRRSRLLCVGPPK

Tb-R   -NGKPTDAWFYTVISTDTVEMSYAAHRTAFLVD-QGYTCSVTEFNPDGAPEAA-------------------------------------
Tco-2  -NGKAVDAWFYSVISTDTVEMSEAARRTAFLVD-QGYTCRVMEFYPDDVAQTA-------------------------------------
Tv-2*
Tc-2   -NGKMVDAWFYTIISTDTLEMAYAAHRTAFLVD-QGYTCRIIEFQPDELSNDE-------------------------------------
Lm-2   -NGRTVDAWFYSIISTDTVEINYAAHRTAFLVD-QGYTCRIMEYSPFTSSVDT-------------------------------------
Li-2   -NGRTVDAWFYSIISIDTVEMNYAAHRTAFLVD-QGYTCRVMEYNPFTSSVDT-------------------------------------
Lb-2   -NGRAVDAWFYSIISMDTIEMNYAAHRTAFLVD-QGYACRVMEYIPFESSANT-------------------------------------
Lmx-2  -NGRTVDAWFYSIISMDTVEMNYAAHRTAFLVD-QGYACRVMEYNPFASSVNT-------------------------------------
Bs-2   -SAGPVDAFFYSLISRDTMEVAYAAHRSEFLVD-QGYRCRFIVFEPPPPVHENDEEEEG-----------------NTTGGGGAVVVVAA

Ng-1   ----NKQAYFYSLVSQDTKEMFFATKRQQFLTN-QGYSFRVLADIQKYYSSNRVE---------------------T-----------KF
Ng-2   --LKNQQAYFYTLVSQDTQEMFFSTKRQQFLIN-QGYSFRVLPDIQKYYRQGANEMGLT-----------------L-----------KH
Trva-1 --TDEYNAFFYTLVSQDTEEIYYSAKRQQFLVD-QGYSYEPVVNAMERWPN------SA-----------------P-----------LS
Trva-2 --REEFNAFFYSCISDMTTDLKYSARRQQFLVD-QGYVYEPVSDARERWPW------PE-----------------D-----------LL
Gl     -----QNAYFYILVSEDTNEQHTAAERRTYLSYNQGYPYSIMNCKEIWETCNLSEL----------------FNDCE--------FTRFD
Ehis-1 -----NSAYFYTLTSKDTREMYFSQRRQRVMRQ-NGYTFKVIDSAVIKPLR-------------------------K-----------ED
Ehis-2 -----RIGYFYTLTSIDTHEEYFIHKKQIYMQN-LGFGFECINSTIITHMN---------------------------------------
Edis-1 -----NSAYFYTLTSKDTREMYFSQRRQRVMRQ-NGYTFKVIDSAVIKPLR-------------------------K-----------ED
Edis-2 -----GIGYFYTLTSIDTHEEYFIYKRQIYMEN-LGFEFENIISKLITNMN---------------------------------------
Dd     --DGLYNAFFYSLVSKDTQEMYYSTKRQQFLID-QGYSFKVISELPGIDQE-------V-----------------N-----------LK
Ac**   ----DVSAYFYTIVSEGTDEVGYSDRRREFMEE-HGYLYHEYDYKHFMHAE-------------------------T-----------EE
Pf     --INDPDSFFYSLVSKDTIEMCYSDKRQRFLIN-QGYAYNVLSDNIVDFNK-------L-----------------N-----------LV
Toxg   --GENFNAFFYSLISKDTLEMVYADKRQQFIID-QGYAYKVIHSRDLPMQP-------D-----------------K-----------LI
Eimt** --DGGFNAYFYSLISKDTLEMVYADKRQQFIID-QGYSYKVLHCRDLPLQQ-------Q-----------------T-----------LV
Ta     ----GFNAFFYSLVSKDTQEMVFADKRQQFIID-QGYAYNVTSFSTIVKDT-------S-----------------N-----------LL
Tett-1 --DGQYNAFWYTLVSKGTDETSYCLARQKCLIN-QGFKYEIIDEKDLPYSQNPQK---------------------------------FK
Tett-2 --KGEYNAFFYTIVSKNTEQAQFYYRRQKSLMD-LGINFEVIDIDE----QNGNK---------------------------------FK
Part-1 --NHEIQAFFYSLASKDTTESEYSYKRQKYITEQLGLNTELILEGDLPYNKNPQK---------------------------------YK
Part-2 --IGEYDCFFYSIVTKGTREVEIQFERQIAVIN-QGYPYEIISADELKPPQSEAD---K-----------------K-----------LQ
At-1   GGKEEYNAFFYSLVSTDTQEMYYSTKRQQFLID-QGYSFKVITSLPPPDAG-------S-----------------S-----------LS
At-2   GGKEEYNAFFYSLVSTDTQEMYYSTKRQQFLID-QGYSFKVITSLPPPDAG-------S-----------------S-----------LG
Orys-1 GGKEEYNAFFYSLVSTDTQEMYYSTKRQQFLID-QGYSFKVITSLPPPEEL-------P-----------------N-----------LK
Orys-2 GGKEEYNAFFYSLVSTDTQEMYYSTKRQQFLID-QGYSFKVITSLPPPEEL-------P-----------------N-----------LK
```

```
Hs      FSTK-------EEQQQL---LQKV-LAATDLDA----EEEVV--------A---G-EF----GSRS------------------------
Mm      FSTK-------EEQQQL---LQKV-LAATDLDA----EEEVV--------A---G-EF----GSRS------------------------
Drer    FSTR-------DEQQQL---LQKV-LAASDLDA----EEEVV--------M---G-EV----GGKP------------------------
Dm      YGTQ-------EEQGQL---LQLV-LSASDLDC----EDEKL--------P---G-EP----GYRP------------------------
Ce      LASK-------ESQLQL---LQQV-LATSDADA----EEEDV--------K---E-EL----AD--------------------------
Sc      YASP-------RERREL---LQEV-LLKNEEAA----GIEVG--------DDADN-SV----GRGSNG----------------------
Sp      YASK-------AERLEL---LQEV-LLQNEEAA----DLDDG--------EDT---SF----GSRS------------------------
Ca      YSSA-------RERREL---LQQV-LLKNEDAA----GLEIG--------DDADT-NF----ISKEKR----------------------
Nc      FATA-------QERREL---LQRT-LVDNEKGA----EDDVE--------T--DD-LF----GKVGRG----------------------

Tb-1    WWYE-------CAGPSCD-------------SAVAAKGTYWIP-------------------FSQEAALRM---QSRFV---AGVRGCDL
Tco-1   WWYE-------CTASSSP-------------LSVAAKGSYWTP-------------------FSRMASARL---QSRFV---AGVCSCEL
Tv-1    WWYE-------SDRAADS-------------AAISWKGSYWLP-------------------FSPEAARCI---QRRFI---DGSASCEL
Tc-1    WWYE-------TLERKLP-------------SAVAAKGSIWLP-------------------FSQEASRRM---HRFFV---NGREVCEL
Lm-1    WWYQ-------THNTASATPSSET-AAAVDVGWVHYDGFYWAP-------------------FSADAAALI---EAAFQ---HGRATCIV
Li-1    WWYQ-------THNTASATPSSET-AAVADAGVVHYDGLYWAP-------------------FSVDAAALI---EAAFQ---HGRVTCTL
Lb-1    WWYQ-------THLTVSAAASSEM-EVGAGVGVAHHDGLYWAR-------------------FSAEASALV---EAAFQ---HGRATCTL
Lmx-1   WWYQ-------MHNTASATPSSET-AAVADVGVVHYDGLYWVP-------------------FSADAAALI---EAAFQ---HGRATCTL
Bs-1    WYWL-------SDELESPEFVNAT-H---SSDAENGRIGRWKL-------------------FSRAHSVAI---DEAFQ---RGQLSLPL

Tb-R    ----------------------------VEG-VDDTAPGDVVSIRQAKLRGTFKK-QELKCS-------------VA-----SP-TAQGS
Tco-2   -----------------------------PEVAEEPVMGDVVSIKQERLRGDFSK-QRLKCS-------------SP-----SS-AIRGN
Tv-2*                                                --
Tc-2    ----------------------------TDDVAVAAEAGDAVSLKQEKLRSALRE-KKLQCE-------------SL-----T--VPQGS
Lm-2    ------------------------ALGGVKQQVKKHRIGDVVSIKQESLYETLHQ-SSVLRPRSGE--------CSG-----RD-ADAGD
Li-2    ------------------------ALGGVKQQVKKHRIGDVVSIKQESLYETLHH-SSVLQPRSGE--------RSG-----RG-ADAGD
Lb-2    ------------------------SLGRMKQLAKKERVGDAVSIKQEFLYETVHQ-SSVLWSRAAE--------RSG-----HG-AGADD
Lmx-2   ------------------------ALGGVKEQVKKHRVGDVVSIKQESLYETLHQ-SSVLRPRSGE--------RSG-----CG-ADAGD
Bs-2    YGEA-------SE-KNV---KRETNHGGEHSSPHNGRTGDAATVKRETLMNDVNH-ERRLVFGASETTRIAHANHSTFGGG-GGAPHTSI

Ng-1    MKTK-------EDELEF---LKLV-KNAKDVNG----KLEDI--------PNDDF-QS--------------------------------
Ng-2    MNTK-------KEELDM---LAIV-KKAEDARG----QEEKI--------NVDDE-VE--------------------------------
Trva-1  LDTQ-------EKREDW---LRQM-MDTDDSLG----LDADA--------TVPDE-DRS---FD--------------VPQGGGQR----
Trva-2  SD-E-------KSEEEW---LRKM-LKVEKIQS----EETEY--------ESYSS-D-E---YE-------ESQ-SDYIDTIGKSA----
Gl      TWKKVRIFYDDVRQAAIEQKLSE
Ehis-1  PAKE-------GKDINY---VTEM-LKKKDEKG----VKEVV--------KKTKK-KT
Ehis-2  ---T-------NEQLNL---IKYI-LNESDKRYFNSKSKKES--------IID---F
Edis-1  PAKE-------GKDINY---VTEM-LKKKDEKG----VKEIV--------KKTKK-KTFLN
Edis-2  ---I-------NEQLNL---IKYI-LNESDKRYFNSKSKKEI--------NIN---D-----F
Dd      YSSK-------QDQLDL---LAQV-LGEGEDSG----KNEIL--------EE----DF----DDITRG----------------------
Ac**    FTRD-------SKQDNL---LAAI-RQEMEKNPEGRSEYSEL--------DDKPT-KRETKKKGTTRQRLNK----IVNKRTGGREARE-
Pf      YKNK-------KIQENL---LKCI-LASTDDGN----MDEDD--------DLFED-----QSFKKDNTKVNKTN-SNIL----------L
Toxg    YGDP-------QRQREI---LTDI-LASDDNDK----TLDDD--------EDDPS-RQVLSAVAGDRGRFFRVN-SNLGDGWGGSEECQL
Eimt**  YGDR-------DRQQAL---LRDI-LLSDDLDK----SLDED--------EDEGA-YLPQQQEQQ-------------------------
Ta      YSKP-------SIQQEL---LQDI-ITSAEDVD----DEEE----------VNPT-----DLFTPTNE----SS-LQVS-----------
Tett-1  WMSQ-------LDESDY---LYTL-LMNSEQNQQNGK----MSS------EEESSSGDSSGSDSEESEG---------------------
Tett-2  YDPD-------ILDQKI---IR-----ASDEIIKDVKKDAWFLI------QDQPDKKV-DEQDDD
Part-1  YIKQ-------MNEMDL---LEQI-MLSSNDLTNKVI----DQI------EEEVQ--VVTNTFSQEFGG---------------------
Part-2  LDKI-------VDEM-----------------QRKVL-------------ETCTQ----KGDDDREEDG---------------------
At-1    YHSQ-------EEQLSL---LGKV-MNAGDDLV----GLEQL--------EE----DT----DGMA------------------------
At-2    YHSQ-------EEQLSL---LGKV-LNAGDDMV----GLEQL--------EE----DT----DGKA------------------------
Orys-1  YYTL-------NDQLEL---LAQL-LSARDDMI----GIEHL--------EE----DS----DGKA------------------------
Orys-2  YYTL-------NDQLEL---LAQS-LSARDDMI----GIEHL--------EE----DS----DGKA------------------------


Hs      -------------SQA-SRRFGTM-SSMSGAD-DTVYMEYHSSRSKA--P----------------------------------------
Mm      -------------GQA-SRRCGTM-SSLSGAD-DTVYMEYHSSRSKA--S----------------------------------------
Drer    --------------QF-SRRAGTM-SSMSGAD-DALYMEYQMPRGSKASV----------------------------------------
Dm      ----------SGSGGA-VRRVGGL-SSMSGGD-DAIYYEHRKK------N----------------------------------------
Ce      -----------GTIRI-SRREATM-ASMSGGQ-GAQYHSK-------AKA----------------------------------------
Sc      --------HKRFKSKA-VRGEGSL-SGLAGGE-DMAYMEYSTNKNKELKE----------------------------------------
Sp      --------LSRAPAKA-KRSSGSL-STLAGAD-NMAYVEYNKSANKQLKK----------------------------------------
Ca      ----MRLEQERNGGGA-TYSSGSL-AGLAGGE-DMAYIEYGKNKNKELRE----------------------------------------
Nc      ----RG-GAKGRAAAV-RRMAGTL-GELSGGQ-DMAYIEQNKAANKGLKK----------------------------------------

Tb-1    TATVLR-----DTPRPPE--LKN--MGVEEKW-TVCFSDSCAPETFGTVQLV-----------------EG-------------------
Tco-1   TSSVLG-----DTPRPEG--LRD--SGVEERW-TVRFSDSNAPETFGTVQVS-----------------EG-------------------
Tv-1    DGSVLV-----GTPRPEG--LGG--SSLEDSW-TVRFSDSNAPETFGIVQFV-----------------KE-------------------
Tc-1    DSTVLN-----DTPRTAE--LAV--SGLGEKW-IVRFSASNAPETFGTVQMA-----------------ED-------------------
Lm-1    RGSELG----RGTPRPSSEELGRYPLLSFESW-KVTFSSVDAPQTFGTVLIG-----------------EG-------------------
Li-1    RGGELG----RGTPRPSSEELGRYPLLSSESW-KVTFSSVDAPRTFGTVLIG-----------------EG-------------------
Lb-1    RGSDLG----RDTPRPSSEELGRYPLLSTESW-KVTFSSADAPQTFGTVLIG-----------------EG-------------------
Lmx-1   RGSELG----RGTPLPSSEELGRYPLLSSESW-KVTFSSVDAPQTFGTVLIG-----------------EG-------------------
Bs-1    RVSHAL----PSSSQRGSAALTLR--PGQLTL-RVTFSSADAPITFGTVQCFS--------------VAEG-------------------

Tb-R    ----------VNPRSL-DYQEKLL-CRVVASW-ELDYQN---ATSQQNEPGVANDTATG------------------------LID----
Tco-2   ----------INVRSA-QYQLDLL-ARVVSNW-EMDYHN---AVSTQKGSGKSEGLQRG------------------------DLK----
Tv-2*
Tc-2    ----------VDARKI-EYQLDLL-SRVVSNW-ELEYHS---QGLKQASSGHDGFPPDD------------------------DANVSE-
Lm-2    ----------CRAESL-SYQLKLL-SKVVSSW-EIEFQQ---ECRKRRRTAVRDADDNA------------------------ANETDS-
Li-2    ----------CRTESL-SYQLELL-SKVVSSW-EIEFQQ---ECRKRRRTAVRDADDNA------------------------ANETES-
Lb-2    ----------RRTESL-SYQLGLL-SKVVSSW-EMEFQQ---ECRKRRRTAAEDVAENA------------------------T---DD-
Lmx-2   ----------CRTESL-SYQLELL-SKVVSSW-EIEFQQ---ECRKRRRTAVRDADDNA------------------------ADETGS-
Bs-2    ----------VDVGRR-KWQLRLL-ASMVGRW-ELQYLQ---ATSRGAVSAKTNARRTNQKGNRRNNQDEGNESESSVSSSSGSSDADNG

Ng-1    --------------GS-HSGDTSL-ALLSGGANEVYYHTFEASNSSKK----------------------------------------KK
Ng-2    --------------KS----SSSL-ASI-SGGNEIYYHSYEASTTTSRRRGK------------------------------------AT
Trva-1  -------NQFKAIVK
Trva-2  -------PHFSLLN
Gl
Ehis-1
Ehis-2
Edis-1
Edis-2
Dd      ----AKKSKSSAPTVS-RTTGGST-RALSGGN-DMNYMEYQAPAIYK-------------------------------------------
Ac**
Pf      ------NKKEDSLKKI-DNTTGGL-LKLSSNM-DVTFADKKKIP---------------------------------------------T
Toxg    ESFGESLRGFGDAPHV-QPLQGGL-GRLAGDG-GAAYGARPGGRGGGRSGRS------------------------------------GG
Eimt**  --QQQQQQQHEQQPNV-RVVQGGL-AGISGKP-STHSFASP
Ta      ---------------I-SQKKTGL-ASLSGSL-KNQLQYKHLLPPS-EKPAK------------------------------------KS
Tett-1  --------HFNITTKA-ANPLQNL-ALYKETQ----LNKY
Tett-2
Part-1  --------VYKFQ
Part-2  --------Y
At-1    ------------LQKA-RRSMGSM-SVMSGSK-GMVYMEYNSGRHKSGQQFK------------------------------------KP
At-2    ------------LK-T-RRSMGSM-SAMSGAN-GRVYMEYNSGRQKSGNQSK------------------------------------KP
Orys-1  ------------LMKA-RRSAGSM-SAFSGSG-GMVYMEYSTGKGKAGA-SK------------------------------------KP
Orys-2  ------------LMKA-RRSAGSM-SAFSGSG-GMVYMEYSTGKGKAGA-SK------------------------------------KP
```

```
Hs     --SKHVHPLFKRFRK
Mm     --SKHVHPLFKRFRK
Drer   --GKNIHPLFKRFRK
Dm     --IGSVHPLFKKFRG
Ce     --IAERHPLFKRFRQ
Sc     -----HHPLIRKMYYKNLKK
Sp     -DSKEHHALFRKHLYTKRR
Ca     ----SHHPLIQKMYYKQNKAKK
Nc     -KGAAKA--EQSAFFKKLQREKERSRAL-----------ARQG

Tb-1   ---NP---LLVRRI--------------CC-------GPLAVEHDCLHGG-----EECLQYAVQQMKVMVAKNSKNRIPLTT
Tco-1  ---DP---LAVRRI--------------CC-------GSLLLNHNCLSEE-----EGCLLDLVERMRSAVEKVSKH
Tv-1   ---KP---LFVRRV--------------CC-------GPIGTDHNCLK-------EECLQHALQHAKSVATKRSTR
Tc-1   ---NP---TLVRRI--------------CY-------GPLDAAHDCLKGE-----EECIQYILGQMRYRIKRRSNPR
Lm-1   ---APSSPLRERRV--------------RR-------GCLDRSHRCIPATA-DQESRCISFARNTVLRARDTRGSDAVTEVE
Li-1   ---APSSSLRERRV--------------RR-------GCLDRSHRCIPATA-DQESHCISFVRNTVLRARDTRGSDAVTEAE
Lb-1   ---APSSPLRERRV--------------RR-------GCLDRSHQCISAAA-DHESRCISFARGAVLHGRNTRDSDSVIVAE
Lmx-1  ---APSSPLRERRV--------------RR-------GCLDRSHRCIPATA-DQESRCIRFARNTVLCVRDTRGSGAVTETE
Bs-1   ---GSQQLVGERRI--------------AKHLWNEPLGDTEEDHCCTD------HDECLMWLLQRVGDAKQLHKRGRQD

Tb-R   ---------------KKMKRARDETAEDIKREWNSGAQTTQPRG-----------DFCRLPLQR------LVGANDDVVYHES-
Tco-2  ---------------KEVKAGVSDSPEAIKREWAPRAAAQNQGA-----------ALYRTSLQD------LVGVSDDFIYHEL-
Tv-2*
Tc-2   -------------GRDVMEIVPDSDLAAIKREWAGGTRMDQFQF-----------GHSCTSLQG------LVSVDENFVYHEL
Lm-2   -------------ADDVVEVRPIRTYAAIKSEWRGGADSAA----------------RVTALRD------LVGVDDGFVYHEL
Li-2   -------------ADDVVEARPIRTYAAIKSEWRSGADSAA----------------RVTALRD------LVGVDDGFVYHEL
Lb-2   -------------ADGVTEARSMRTYATIKNEWRVSGDHET----------------RATALHN------LVGVDDGFVYHEL
Lmx-2  -------------ADDVVEARPIRTYAAIKSEWRGGADSAN----------------RVTALRD------LVGVDDGFVYHEL
Bs-2   SDVEIVEPAISKKGNKAQQSTNHRSLIDIKKEYGVTAGVPARGMTVVTTTTLQQQDASMMLAR------FTTSADGIVYHELR

Ng-1   KVAKTVHPLFKKRN
Ng-2   KVPSTVHPLFKTRTQHKD
Trva-1
Trva-2
Gl
Ehis-1
Ehis-2
Edis-1
Edis-2
Dd     -SIPTQHALFKQRAKNKQ
Ac**
Pf     KKFADKHILFRKFLSQNK
Toxg   ADATSMHPLFRGLHGKK
Eimt**
Ta     KSSSEQHPYFKKLYSTK
Tett-1
Tett-2
Part-1
Part-2
At-1   KDPTKRHNLFKKRYV
At-2   KDPTKRHNIFKKRYV
Orys-1 KDPSKRHYLFKKRYQ
Orys-2 KDPSKRHYLFKKRYQ
```

Fig. S1. Multiple sequence alignment of XPB amino acid sequences

The encoded protein sequences of specified *XPB* genes were aligned using the Clustal Omega server of the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=clustalo) at default parameters (Sievers *et al.*, 2011).

The XPB domain, Walker motifs (WM I, WM Ia, WM II-VI), RED residue loop, thumb-like domain (Thm) and XPB signature motif were highlighted in the members of the Opisthokonta supergroup, comprising the eukaryotic model organisms for TFIIH research (top nine sequences). If a position within the Opisthokonta was more than 50% identical or similar, the residue was highlighted in black or gray, respectively, and the highlighting was then applied to the kinetoplastid sequences (blue lettering) to demonstrate the conservation of these domains in the latter.

To detect putative DNA repair-specific sequence conservation in kinetoplastid XPB-R sequences, positions of XPB-R-specific identity (one mismatch allowed) that was not present in any of the kinetoplastid XPB sequences was highlighted in purple. If this position was also conserved in the Opisthokonta model organisms (more than 50% similarity) the identities were highlighted in red.

If a species contained two *XPB* sequences, the one with greater similarity to *T. brucei* XPB received the name affix 1 and the other sequence the affix 2.

Kinetoplastid accession numbers were derived from www.GeneDB.org. Ac, *Acanthamoeba castellanii* (XP_004356478); At-1/At-2, *Arabidopsis thaliana* (Q38861 / NP_568591); Bs-1/Bs-2 *Bodo saltans* (BS16690.1:pep / BS91250.1:pep); Ca, *Candida albicans* (EEQ43396); Ce, *Caenorhabditis elegans* (NP_499487); Dd, *Dictyostelium discoideum* (XP_647819); Dm, *Drosophila melanogaster* (NP_001137931); Drer, *Danio rerio* (Q7ZVV1); Edis-1/Edis-2, *Entamoeba dispar* (EDI_083940 / EDI_292260); Ehis-1/Ehis-2, *Entamoeba histolytica* (XP_649651 / XP_654948); Eimt, *Eimeria tenella* (ETH_00032995); Gl, *Giardia lamblia* (GL50803_16512); Hs, *Homo sapiens* (accession number P19447); Lb-1/Lb-2, *Leishmania braziliensis* (LbrM.29.0600 / LbrM.32.4160); Li-1/Li-2, *Leishmania infantum* (LinJ.29.0610 / LinJ.32.4070); Lm-1/Lm-2, *Leishmania major* (LmjF29.0590 / LmjF32.3920); Lmx-1/Lmx-2, *Leishmania mexicana* (LmxM.08_29.0590 / LmxM.31.3920); Mm, *Mus musculus* (NP_598419); Nc, *Neurospora crassa* (XP_957329); Ng-1/Ng-2, *Naegleria gruberi* (XP_002674455 / XP_002683056); Orys-1/Orys-2, *Oryza sativa* (EEE55219 / EEC71313.1); Part-1/Part-2, *Paramecium tetraurelia* (XP_001448651 / XP_001425941); Pf, *Plasmodium falciparum* (PF10_0369); Sc, *Saccharomyces cerevisiae* (CAY80363); Sp, *Schizosaccharomyces pombe* (NP_593474); Ta, *Theileria annulata* (XP_952567.1); Tb-1/Tb-R, *Trypanosoma brucei* XPB/XPB-R (Tb927.3.5100 / Tb927.11.16270); Tc-1/Tc-2, *Trypanosoma cruzi* (TcCLB.510149.50 / TcCLB.511527.20); Tco-1/Tco-2, *Trypanosoma congolense* (TcIL3000_3_3090 / TcIL3000.11.16180); Tett-1/Tett-2, *Tetrahymena thermophila* (XP_001032673 / XP_001470758); Toxg, *Toxoplasma gondii* (XP_002365647); Trva-1/Trva-2, *Trichomonas vaginalis* (XP_001582312 / XP_001326086); Tv-1/Tv-2, *Trypanosoma vivax* (TvY486_0304470 / TvY486_1117180).

* The *T. vivax* XPB-2 gene appears to be a partial gene, missing sequence for ~220 C-terminal amino acids

** The single *E. tenella* and *A. castellanii XPB* genes miss a large portion of the XPB N-terminus including the XPB domain, suggesting that the reading frames of these genes are incomplete.

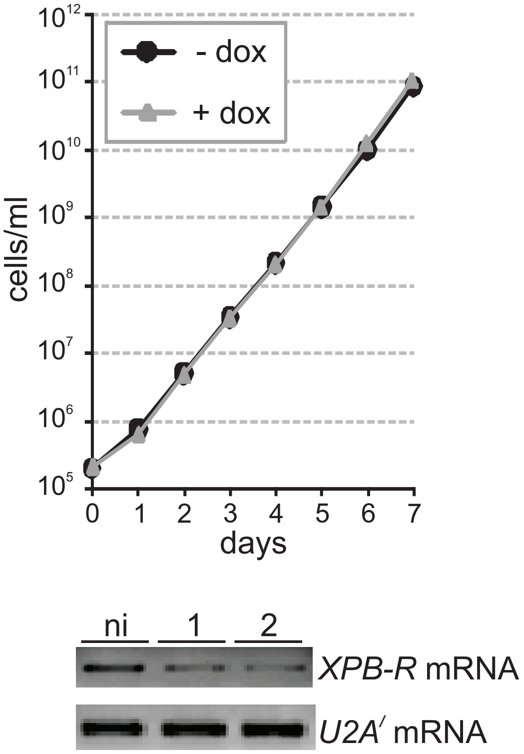


Fig. S2. *XPB-R* is not essential for trypanosome proliferation in culture.

Growth curve of a representative bloodstream form cell line in which doxycycline-induced *XPB-R* dsRNA synthesis led to RNAi-mediated *XPB-R* silencing. The lower panel shows a semi-quantitative PCR analysis of total RNA that was reverse transcribed with oligo-dT and prepared from non-induced cells (ni) and cells that were induced with doxycycline for one or two days. The spliceosomal *U2A′* coding sequence was amplified as a control.

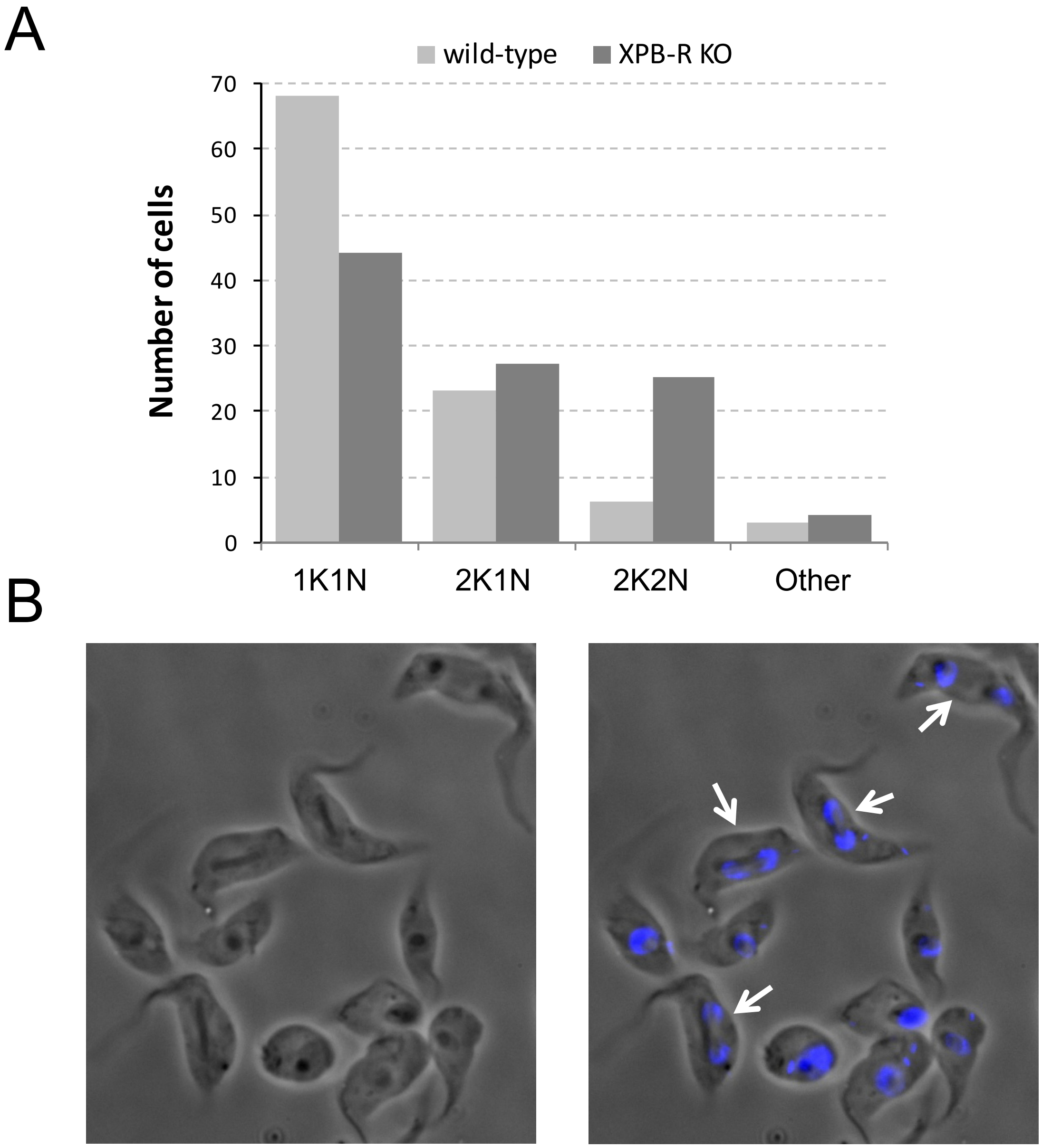
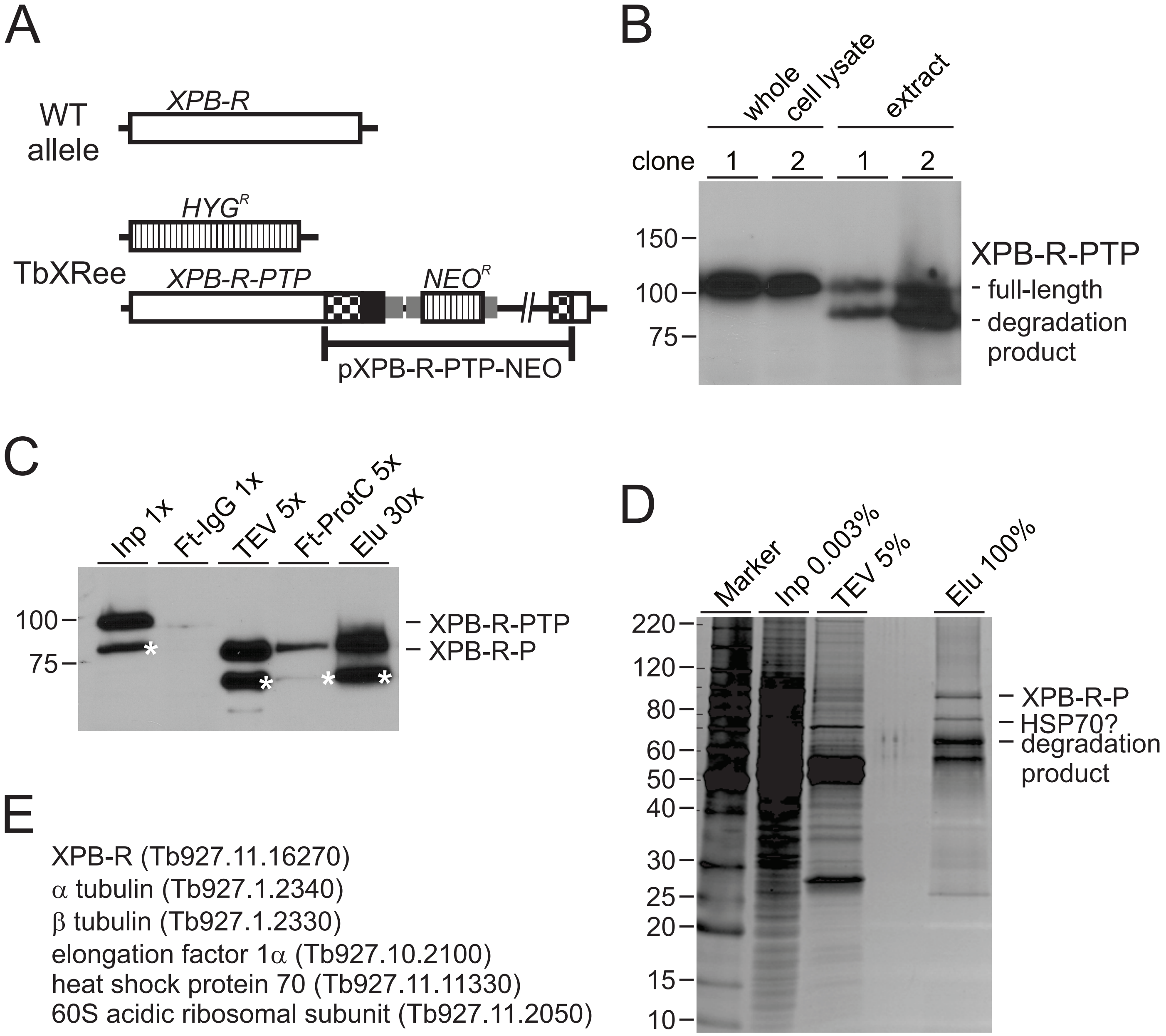


Fig. S3. *XPB-R* knockout appears to affect completion of mitosis

A. Diagram showing the number of wild-type procyclic (n=100) and *XPB-R* knockout cells (XPB-R KO; n= 100) that are in the 1K1N (1 kinetoplast and 1 nucleus), 2K1N, and 2K2N stages of the cell cycle. “Other” combines aberrant cells that have no nuclei (zoids) or more than two nuclei.
B. Image of XPB-R KO cells stained with DAPI showing an accumulation of 2K2N cells (white arrows).



XPB-R (Tb927.11.16270)
α tubulin (Tb927.1.2340)
β tubulin (Tb927.1.2330)
elongation factor 1α (Tb927.10.2100)
heat shock protein 70 (Tb927.11.11330)
60S acidic ribosomal subunit (Tb927.11.2050)

Fig. S4. PTP tagging and tandem affinity purification (TAP) of XPB-R

A. Schematic depiction (not to scale) of a *XPB-R* wild-type (WT) allele and of the *XPB-R* locus of procyclic TbXRee cells that exclusively express XPB-R-PTP and no wild-type, untagged XPB-R. XPB-R, HYG^{R}/NEO^{R} coding regions and PTP tag are indicated by open, striped and black boxes, respectively. Gene flanks for RNA processing signals are drawn as smaller gray boxes.

B. Whole cell lysates and crude, cell-free extracts of two TbXRee cell lines were analyzed by immunoblotting with the monoclonal anti-Protein C epitope antibody HPC4 that recognizes the PTP tag.

XPB-R-PTP is N-terminally degraded upon extract preparation.

C. Immunoblot monitoring of XPB-R-PTP tandem affinity purification. Aliquots of crude extract (Inp), the flowthrough of the IgG affinity chromatography (Ft-IgG), the TEV protease elution (TEV), the flowthrough of the anti-ProtC affinity chromatography (Ft-ProtC), and the final eluate (Elu) were probed with the HPC4 antibody. The x-values indicate relative amounts analyzed. Note that during the TEV protease digest ~19 kDa of protein is removed from XPB-R-PTP resulting in XPB-R-P. Asterisks indicate the XPB-R degradation product.

D. Aliquots of input material (Inp), the TEV protease eluate (TEV) and the final eluate (Elu) were separated on a 10-20% SDS-polyacrylamide gradient gel and stained with sypro ruby. Percentages indicate relative amounts analyzed. Band assignments on the right are guided by apparent protein size.

E. List of proteins that were unambiguously identified by liquid chromatography/tandem mass spectrometry in the final TAP eluate.

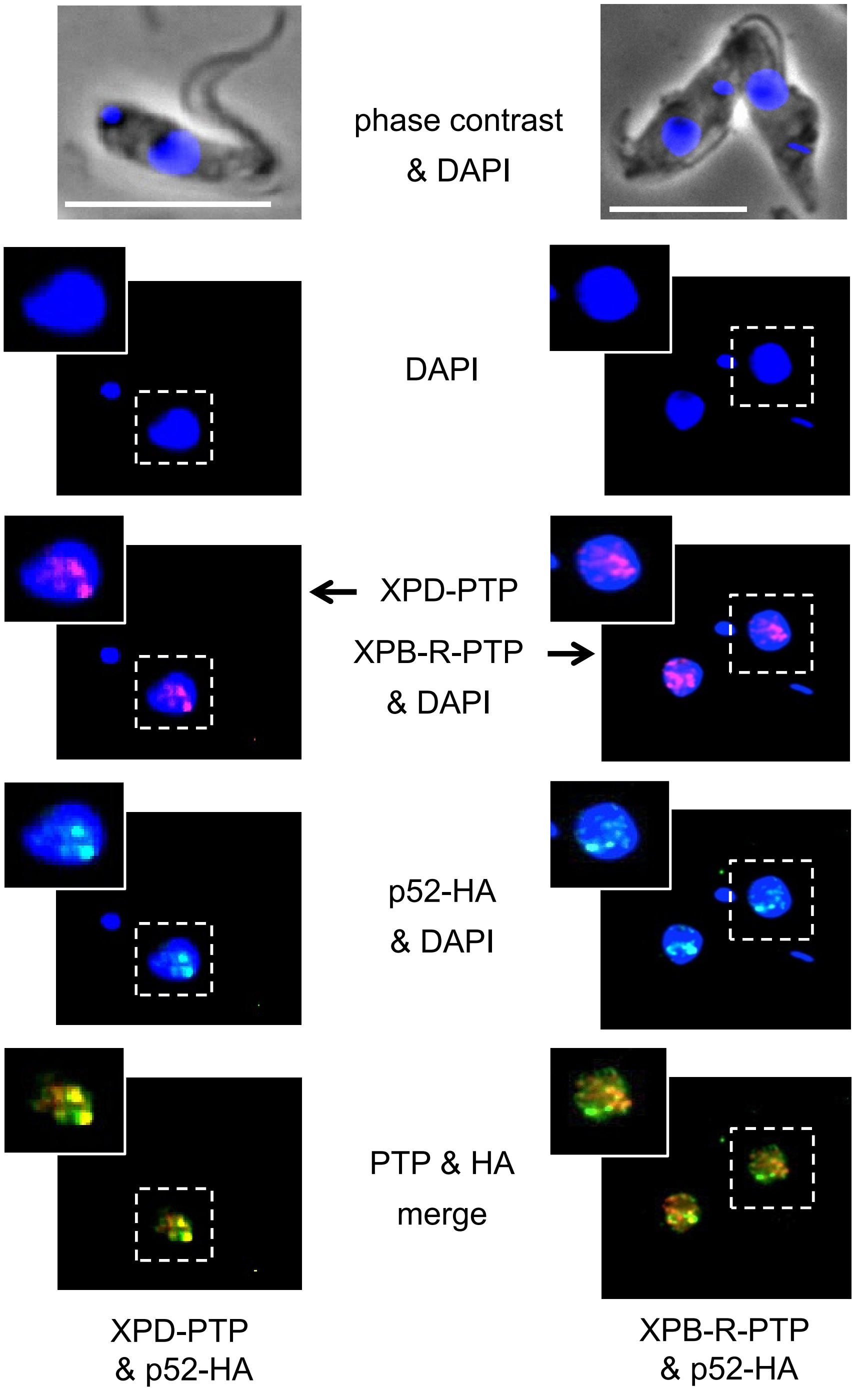


Fig. S5. XPB-R-PTP co-localizes with p52-HA outside putative *SLRNA* expression foci in the nucleus

p52-HA (green) was co-localized with XPD-PTP (red; left panels) or with XPB-R-PTP (red, right panels). White bars represent 10 μm.

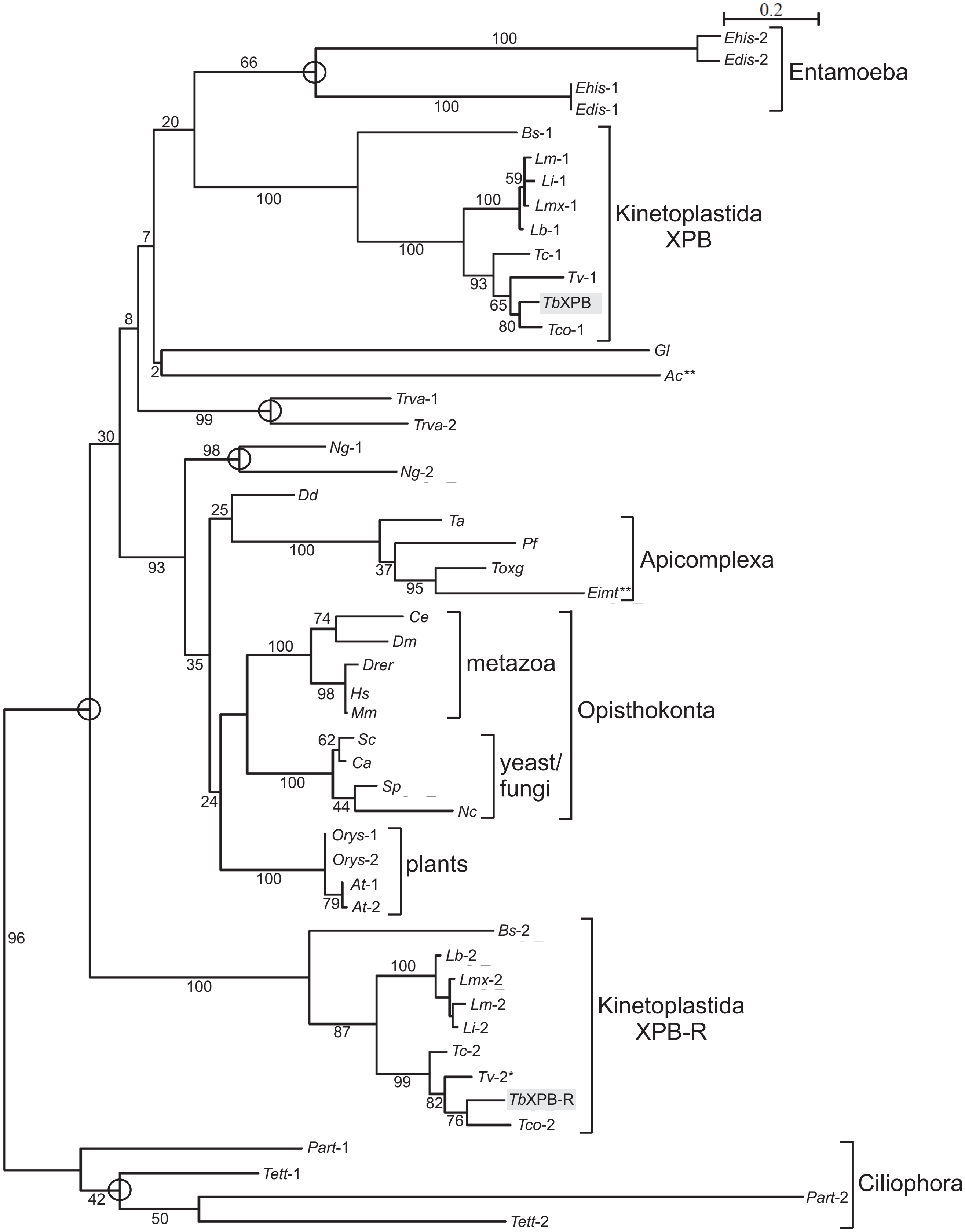

0.2
100
Ehis-2
Edis-2
Entamoeba
66
Ehis-1
Edis-1
100
20
Bs-1
Lm-1
59
Li-1
Lmx-1
100
100
Lb-1
Kinetoplastida XPB
7
100
Tc-1
Tv-1
93
TbXPB
65
80
Tco-1
8
Gl
2
Ac**
Trva-1
99
Trva-2
30
98
Ng-1
Ng-2
Dd
25
Ta
Pf
100
Apicomplexa
Toxg
37
95
Eimt**
93
74
Ce
Dm
100
metazoa
Drer
35
98
Hs
Mm
Opisthokonta
62
Sc
Ca
yeast/ fungi
100
Sp
24
44
Nc
Orys-1
Orys-2
plants
At-1
100
79
At-2
96
Bs-2
100
Lb-2
Lmx-2
100
Lm-2
Kinetoplastida XPB-R
Li-2
87
Tc-2
Tv-2*
99
82
TbXPB-R
76
Tco-2
Part-1
Tett-1
Ciliophora
42
Part-2
50
Tett-2

Fig. S6. Divergent *XPB* paralogs are present in several protistan taxa.

Multiple eukaryotic XPB amino acid sequences were aligned and phylogenetically analyzed using maximum likelihood. Bootstrap values were obtained by sampling a thousand replicates and are indicated as percentages. The scale bar indicates the number of amino acid substitutions per site. Open circles mark the splits of two divergent XPB paralogs within a taxon. If a species contained two *XPB* genes/sequences, the one with greater similarity to *T. brucei* XPB received the name affix 1 and the other sequence the affix 2. Accession numbers are specified in Fig. S1. *Ac, Acanthamoeba castellanii; At*-1/*At*-2, *Arabidopsis thaliana*; *Bs*-1/*Bs*-2, *Bodo saltans*; *Ca, Candida albicans; Ce, Caenorhabditis elegans; Dd*, *Dictyostelium discoideum; Dm, Drosophila melanogaster; Drer, Danio rerio; Eimt, Eimeria tenella; Gl*, *Giardia lamblia*; *Hs, Homo sapiens; Lb*-1/*Lb*-2, *Leishmania braziliensis*; *Li*-1/*Li*-2, *Leishmania infantum*; *Lm*-1/*L*m-2, *Leishmania major*; *Lmx*-1/*Lmx*-2, *Leishmania mexicana*; *Mm, Mus musculus; Nc, Neurospora crassa; Ng*-1/*Ng*-2, *Naegleria gruberi*; *Orys*-1/*Orys*-2, *Oryza sativa*; *Part*-1/*Part*-2, *Paramecium tetraurelia*; *Pf, Plasmodium falciparum; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe; Ta, Theileria annulata*; *Tb*-1/*Tb*-R, *Trypanosoma brucei*; *Tc*-1/*Tc*-2, *Trypanosoma cruzi*; *Tco*-1/*Tco*-2, *Trypanosoma congolense*; *Tett*-1/*Tett*-2, *Tetrahymena thermophila*; *Toxg, Toxoplasma gondii; Trva*-1/*Trva*-2, *Trichomonas vaginalis*; *Tv*-1/*Tv*-2, *Trypanosoma vivax.*

* The *T. vivax XPB-2* gene appears to be a partial gene, missing sequence for ~220 C-terminal amino acids

** The single *E. tenella* and *A. castellanii XPB* genes miss a large portion of the XPB N-terminus including the XPB domain, suggesting that the reading frames of these genes are incomplete.

Method of Phylogenetic Analysis

XPB amino acid sequences from 33 organisms representing the major eukaryotic supergroups (Dacks *et al.*, 2008) were aligned using the multiple sequence alignment tool MUSCLE at the EMBL European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/muscle/). The alignment was uploaded onto the graphical user interface Seaview (Gouy *et al.*, 2010); http://pbil.univ-lyon1.fr/software/seaview.html) to drive the GBlocks program using the less stringent option for the selection of conserved sequence blocks that are likely to be aligned correctly and for the removal of poorly aligned positions (Talavera and Castresana, 2007). The best-fit evolution model for phylogenetic analysis was determined by using the ProtTest Server 2.4 (Abascal *et al.*, 2005); http://darwin.uvigo.es/software/prottest2_server.html). A maximum likelihood tree was generated employing the LG empirical matrix (Le and Gascuel, 2008) with estimated invariable sites, substitution rate categories of 4, estimated gamma distribution and model equilibrium frequencies.

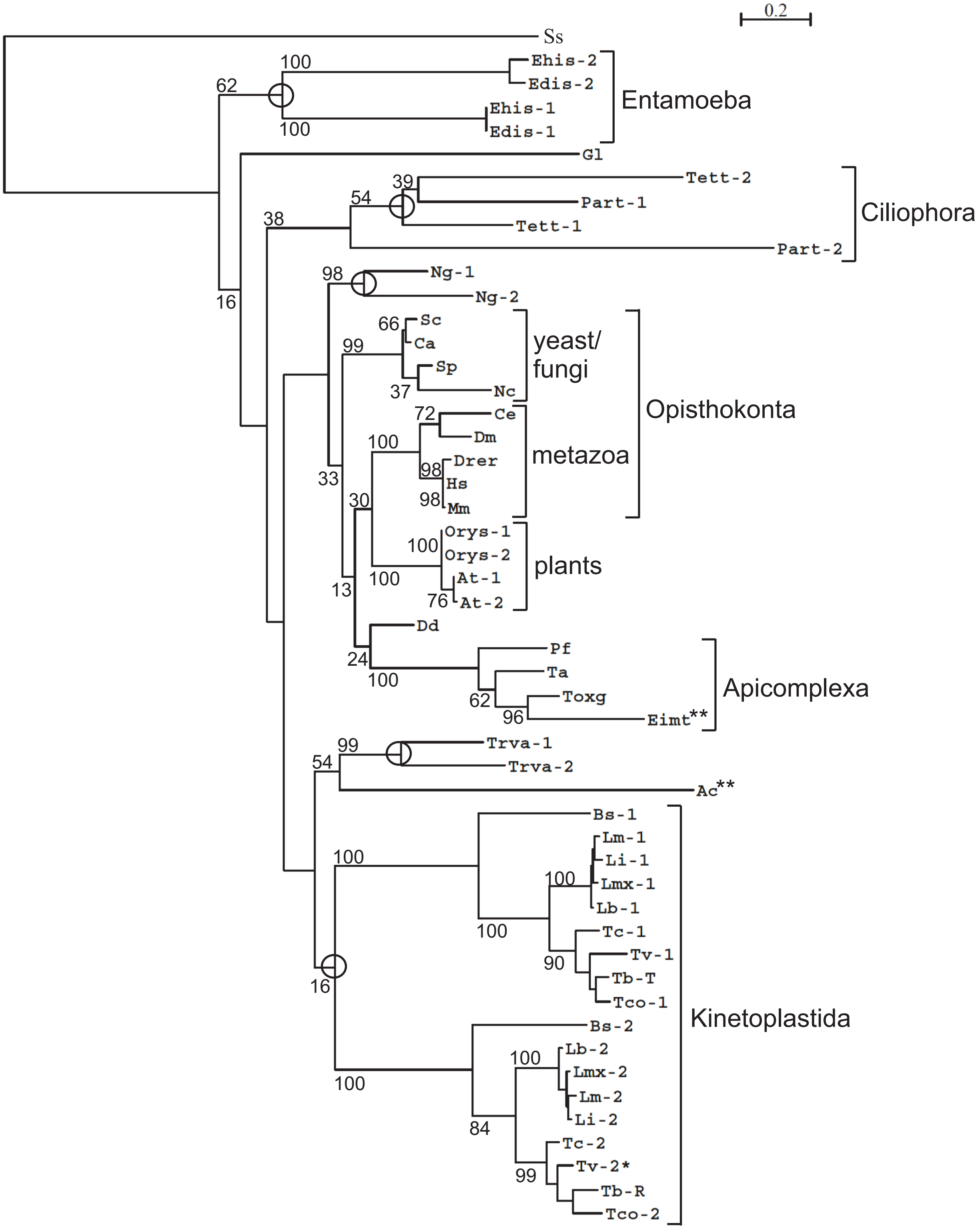


Fig. S7. XPB-derived phylogenetic tree with the archaeon *Sulfolobus solfataricus* as an outgroup

The phylogenetic analysis corresponds to that shown in figure 7 except that the XPB sequence from *Sulfolobus solfataricus* (Ss, accession number NP_342443) was included.

Table S1. Oligonucleotide used in this study

pGene	Sequence	**Usage**
XPB-R	5′-GGAGTGGAACTGCGAGGCCAGTAGTG-3′ / 5′-CCAAATCCACCTGAATACGGCGACG-3′	Sense/antisense oligonucleotides that hybridize in the 5′ and 3′ XPB-R gene flanks, respectively; used to analyze correct integration of transfected DNA constructs that replace the wild-type *XPB-R* alleles.
	5′-TAAGGGCCCAGCACACCGAACTGCATTCCTCG-3′/ 5′-TAACGGCCGCGATGACGTTGTCTATCATGAGAGT-3′	Sense/antisense oligonucleotides that hybridize within the coding region of XPB-R; used for RT-PCR of XPB-R mRNA.
U2A′	5′-TAGGGCCCTCCAAAAACGGCAGTGCCGATACTGC-3′ / 5′-TACGGCCGTGATGTGCGAGTCTTCTTTGTCCCTTTAGAAGC-3′	Sense/antisense oligonucleotides that hybridize within the coding region of U2A′; used for RT-PCR of U2A′ mRNA.
SL RNA	5′-CTACCGACACATTTCTGGC-3′ / 5′-GGTATGAGAAGCTCCCAGTAGCAGC-3′	Sense/antisense oligonucleotides that hybridize in the *SLRNA* promoter region; used for PCR / qPCR amplifications in ChIP experiments.
	5′-ATGGCTTATACGTGCTCGTTTCTCC-3′ / 5′-CACATATAGGCGCTTTAAAGTCTGCT-3′	Sense/antisense oligonucleotides that hybridize in the *SLRNA* intergenic region; used for PCR / qPCR amplifications in ChIP experiments.
α-Tubulin	5′-GTGCATTGAACGTGGATCTG-3′ / 5′-GCCTACCACGAGCAACTCTC-3′	Sense/antisense oligonucleotides that hybridize within the coding region of α-tubulin; used for PCR / qPCR amplifications in ChIP experiments.
p52	5′-TAGGGCCCGCAGCCCCTTCAACAACAGTGG-3′ / 5′- TACGGCCGATCCGGAGTAATGAAATTGGAC-3′	Sense/antisense oligonucleotides that hybridize within the coding region of p52; used for RT-PCR of p52 mRNA.

Supplemental References

Abascal,F., Zardoya,R., and Posada,D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104-2105.

Dacks,J.B., Walker,G., and Field,M.C. (2008) Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int* **57**: 97-104.

Gouy,M., Guindon,S., and Gascuel,O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221-224.

Le,S.Q., and Gascuel,O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307-1320.

Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Talavera,G., and Castresana,J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564-577.