Legends for supplementary data

Supplementary Figure 1. Alignment of different domains. (**A**) HD domain. (**B**) ZLZ domain. The cysteins that could form a zip structure are boxed. (**C**) START-SAD domain. The point mutations are indicated in the figure.

Supplementary Figure 2. (**A**) Expression of proteins used in binding assay. HD (aa 1-115), HD-ZLZ (aa 1-237), START-SAD (aa 238-725).GbML1 indicates the full-length protein. MBP indicates maltose binding protein. All parts of GbML1 are fused in frame to MBP. The purified MBP or MBP fusion proteins are separated by SDS-PAGE gel and stained by Coomassie Blue. Arrow indicated the cloned protein with correct size. (**B**) The purified proteins used in point mutation assay. (**C**) Commassie Blue staining of a native gel that containing the binding mix with GbML1 full-length protein or HD-ZLZ protein. Arrows indicate the binding mix, which is also detected in figure 2, panel D.

Supplementary Figure 3. The property of the sequence in *GbML1* promoter region. The promoter sequence is analyzed using PLACE (http://www.dna.affrc.go.jp/PLACE/, Higo et al., 1999).

Supplementary Figure 4. Yeast transactivation assay of GbML1, GbMYB25 and GbMYB2. (A) GbML1 is a weak activator while GbMYB25 is a strong activator. BD indicates pDEST32 vector; AD indicates pDEST22 vector. Each BD, BD-GbML1, BD-GbMYB25 and BD-GBMYB2 vector is co-transformed with AD vector into yeast strain AH109. The yeast clones grown on selective plates are presented. (B) C2 part of GbMYB25 confers the strong activation activity. (a) Diagram of the three parts of GbMYB25. MYB indicates the N terminal part of GbMYB25 containing the two MYB repeats (aa: 1-112); C1 indicates the domain predicted with similarity of other MYB proteins (aa: 113-170); C2 is the domain at the C terminus with low similarities to other proteins (aa: 171-309). (b) The yeast clones harboring AD vector and BD-MYB/BD-C1/BD-C2 /BD-GbMYB25 grown on selective plates are presented.

Supplementary Figure 5. Alignment of GbMYB25 with GhMYB25. The SNPs are indicated by arrows.

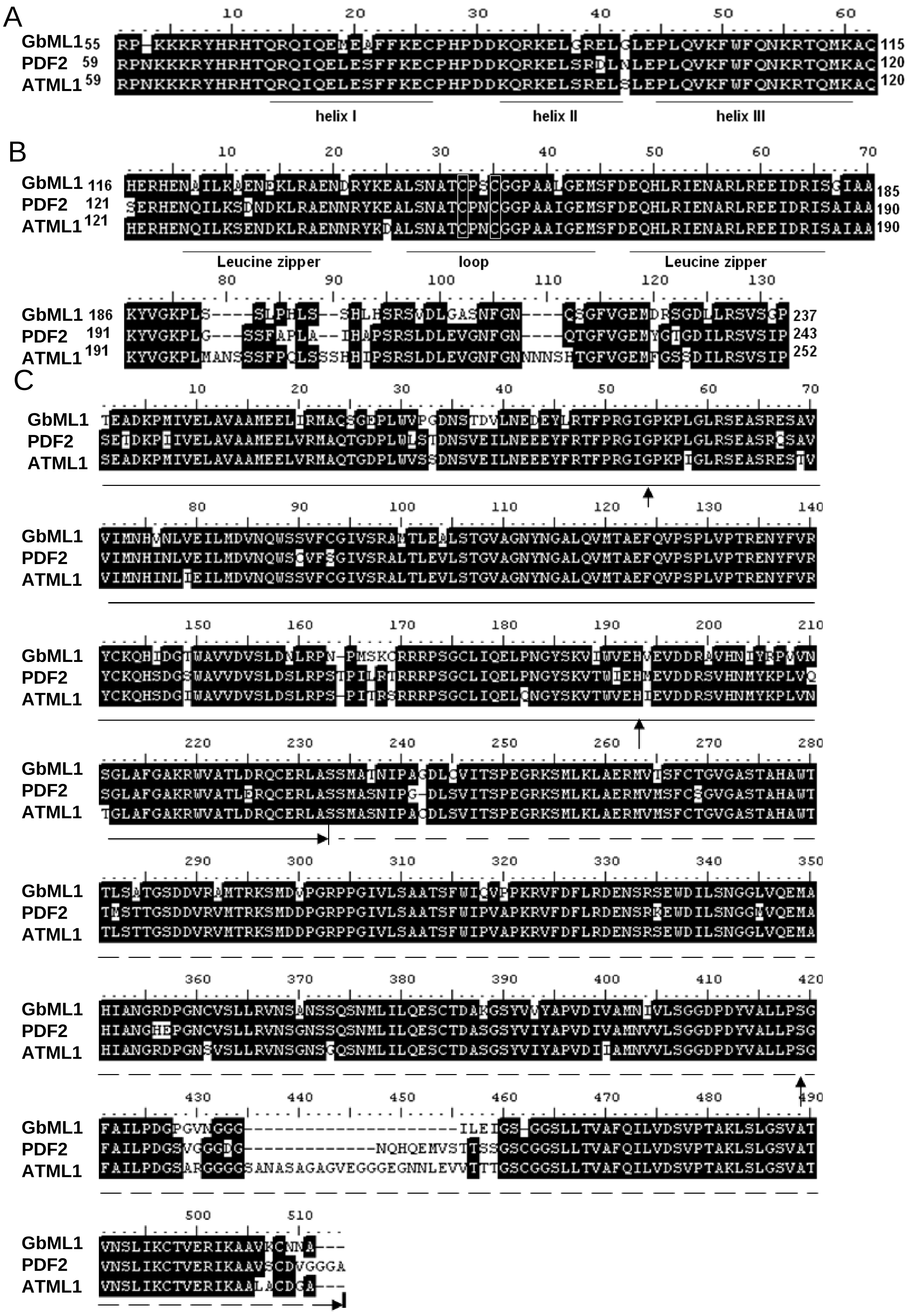
Supplementary Figure 6. Alignment of GbMYB2, GaMYB2 and GhMYB2. The SNPs are indicated by arrows.

Supplementary Figure 7. Alignment that uses to generate the phylogeny tree.

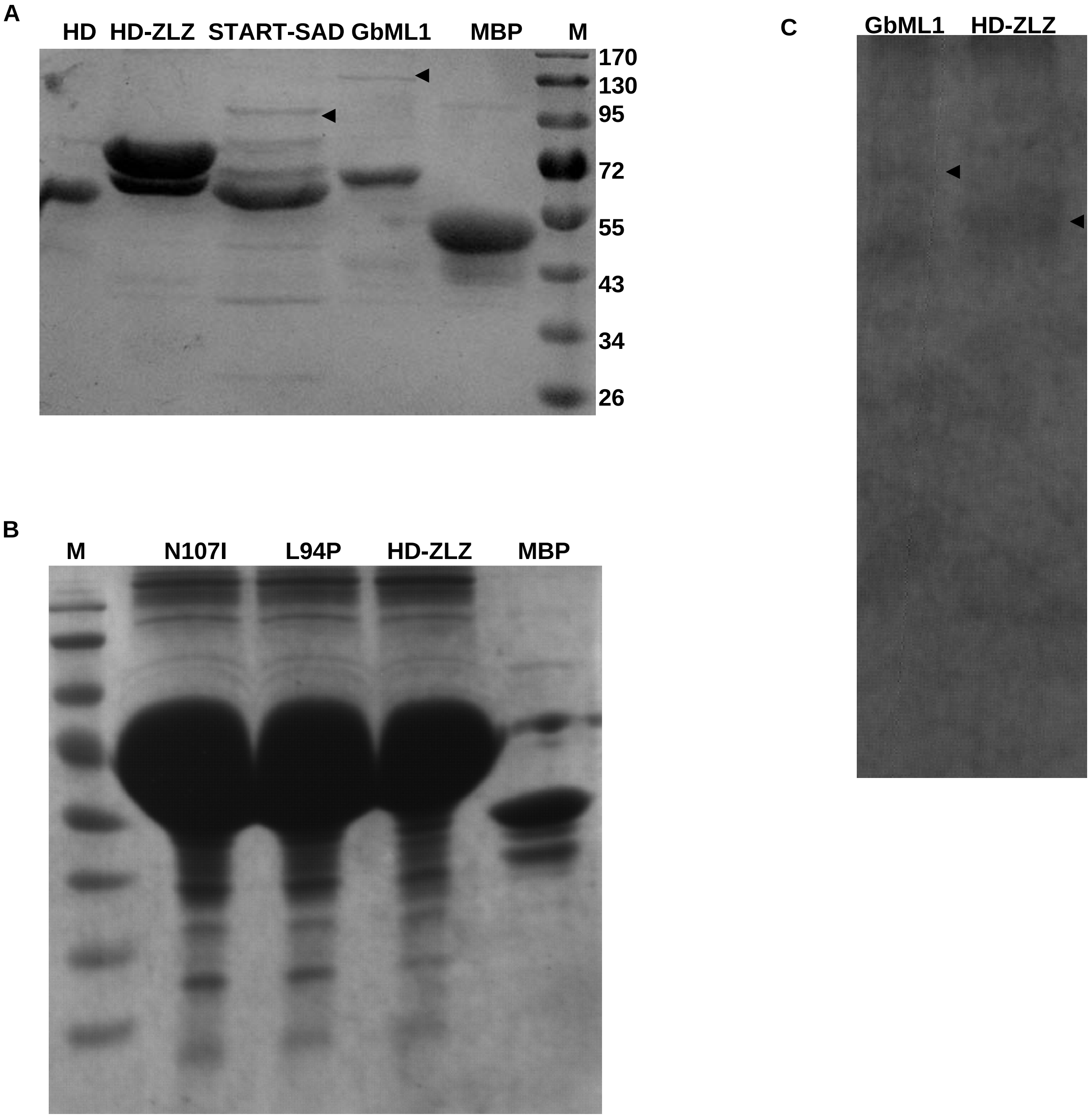
Supplementary Table S1. Primers and probes used in this research.

Reference:

K. Higo, Y. Ugawa, M. Iwamoto and T. Korenaga (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Research* **27**: 297-300



Supplementary Figure 1



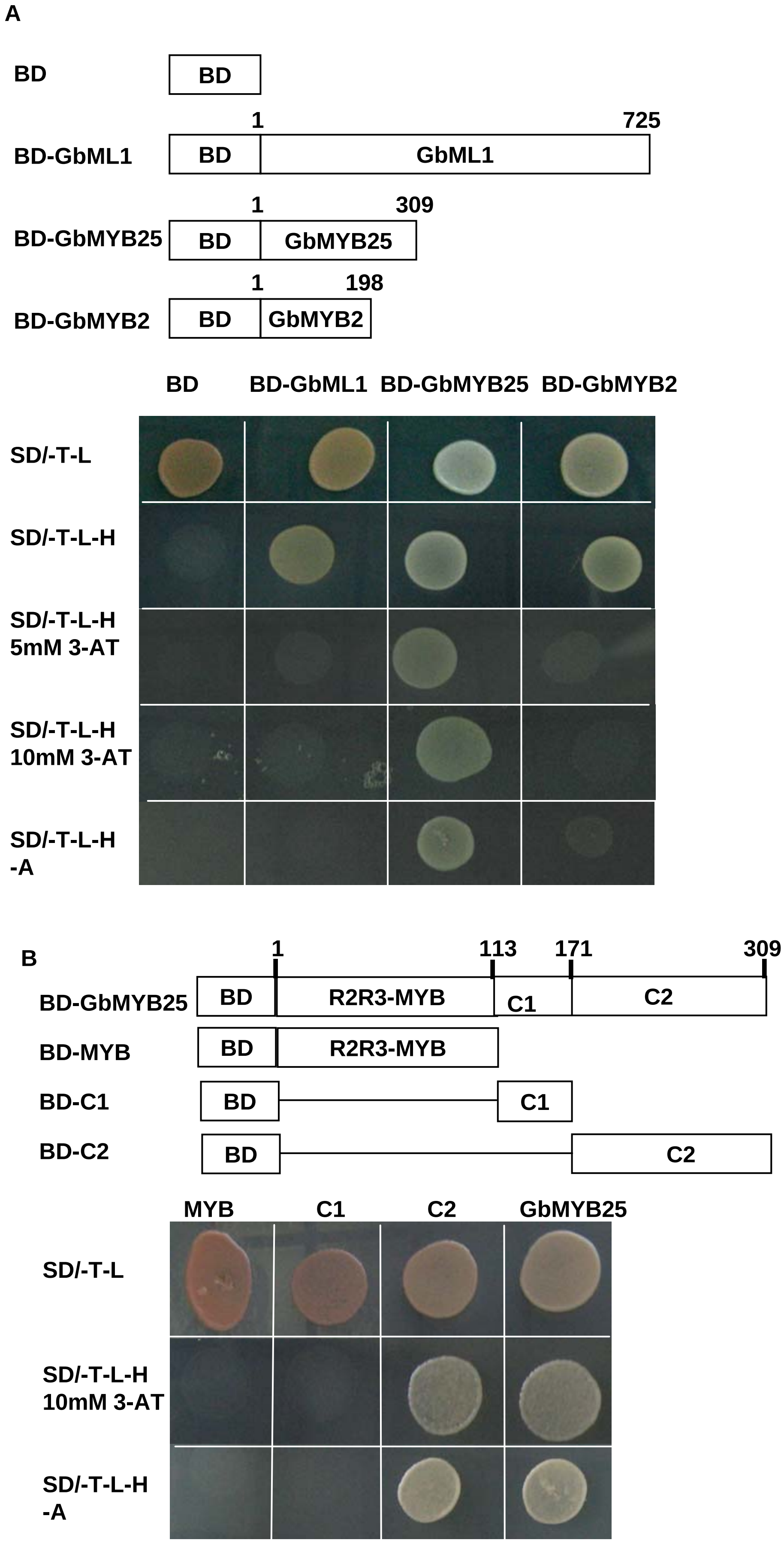
Supplementary figure 2

CTTGACAATTCTCTTCATTTTCCTCACCCCCCCTTGGAACTTCAAACCCCTTTCTCTCTCTCTAGACTTTAACCCTCTTCTCTCTCAAGGTAGGTCTCTGTCAA
ACCAGTGACCCCAAACTTTTAAAAACCACCCTCTTACCTCCTTTCATTAATCTTTTTCAACAATGTTCAAAAGGGCTTTCTTTTGTTTGCCTATACGCCCTGTA
AAAAGTCTCTACTCACGAAACTATCTTTTTCCTGGACCAGCTTGTTTATAGCACCCGATCCGTGTTCACTCCATTTGCCATTCTCTCTTTCATATATTCGCTTA
TAAAGCTTTGCTGAGAATCAAGAGAGAGTGCAAAAAGTGTTTGTAATCTTTGGTTTGTGGGGACTTTCTATCAAAATCAACCTAAATACCACCAATATAGGAA
AGCGGGGCCAAGAATCTAACAAGCAAGAGACTGATATGTGTAGCATGTATTGCTGCTGGTTAGGGTTTTATAGGTCAATGGCTATCTATGTCTCCCAACATG
AGGTTTCTAGTGACCAACTTACCACTGCATTTAAGACAAGAAAATAGTTGCCCTAAGCTCTTAGTCCCCTTCTCCAAATAATCCATAAGGTAATTTTCTTTCCT
CCTTTTTCCCCAACTTTTGCTGCCCATTTTATTTCTCCGAGGTTCTTTCACTTTTCCCATGGGAATAGATTAGGATAAGCTTTCGACTAATCCATCATTTTAAT
CTTCGAAGTATATGCTACCTTTAAATATAGTAGCAGTTGGTTATTAAGATATTCAAAGAAAAATCCTTCTCGATAAGCGAACTAGACCTTCTTTTTCTTTGAAA
GAGAGATGGTATATGGTTTACAGCTCCATTGTTTATTTGCCTTTCAAATGTTTTACGATTCCATTGTCATGCACTGGTTTTGACTTTTTTAACTTTGTACTTACA
TTTTGTGTCTTTCGTTACTGTTCTACAGTTTATTTAACCCTTTTCACCCTTTATTTTAGTTTGCTTGACTTGTGTATGTTCTCTCGGCAGAACCCAGAATATCCA
T*ACCGTCGACATTCGAGCACTTTGTTAAC***ATG**

TGGTTAG :MYBATRD22 TAAATGYA :L1BOXATPDF1 **ATG**: Translation start codon RACE 5' end nucleotides are italic

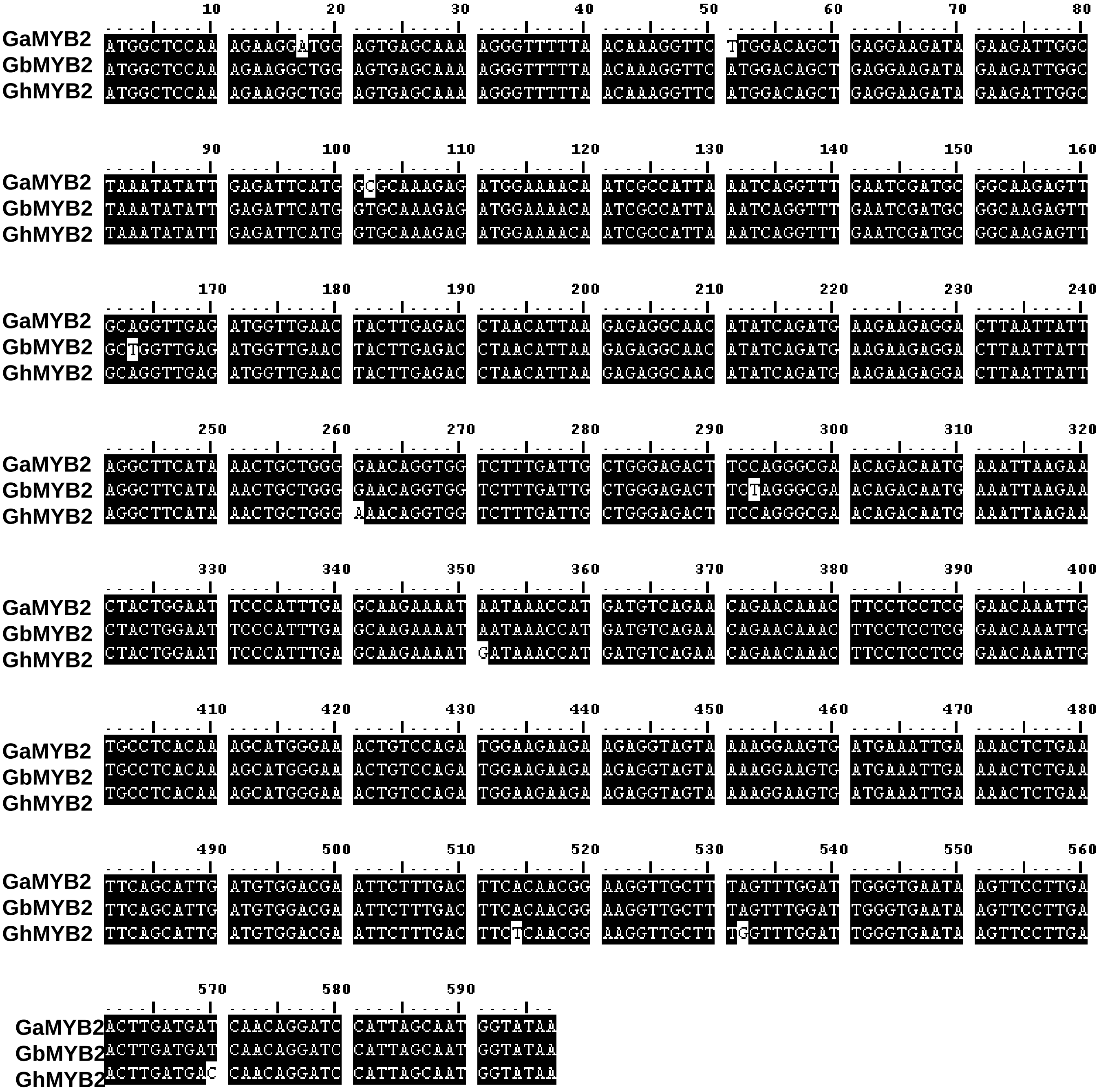
Sequence used in EMSA is
GCATGTATTGCTGCTGGTTAGGGTTTTATAGGTCAATGGCTATCTATGTCTCCCAACATGAGGTTTCTAGTGACCAACTTACCACTGCATTTAAGACAAGAAAATAGTTG

Supplementary Figure 3



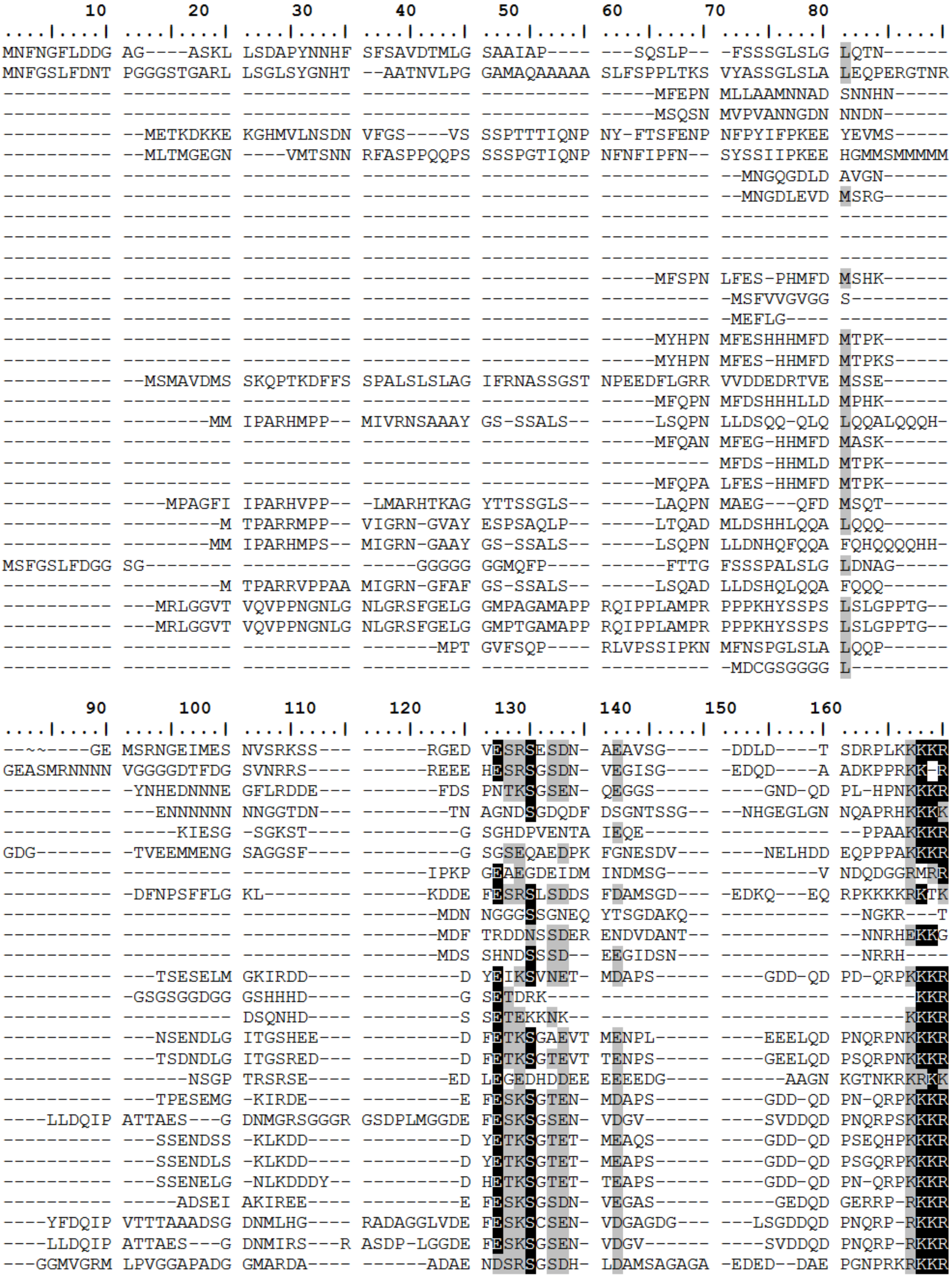
Supplementary Figure 4

Supplementary Figure 5

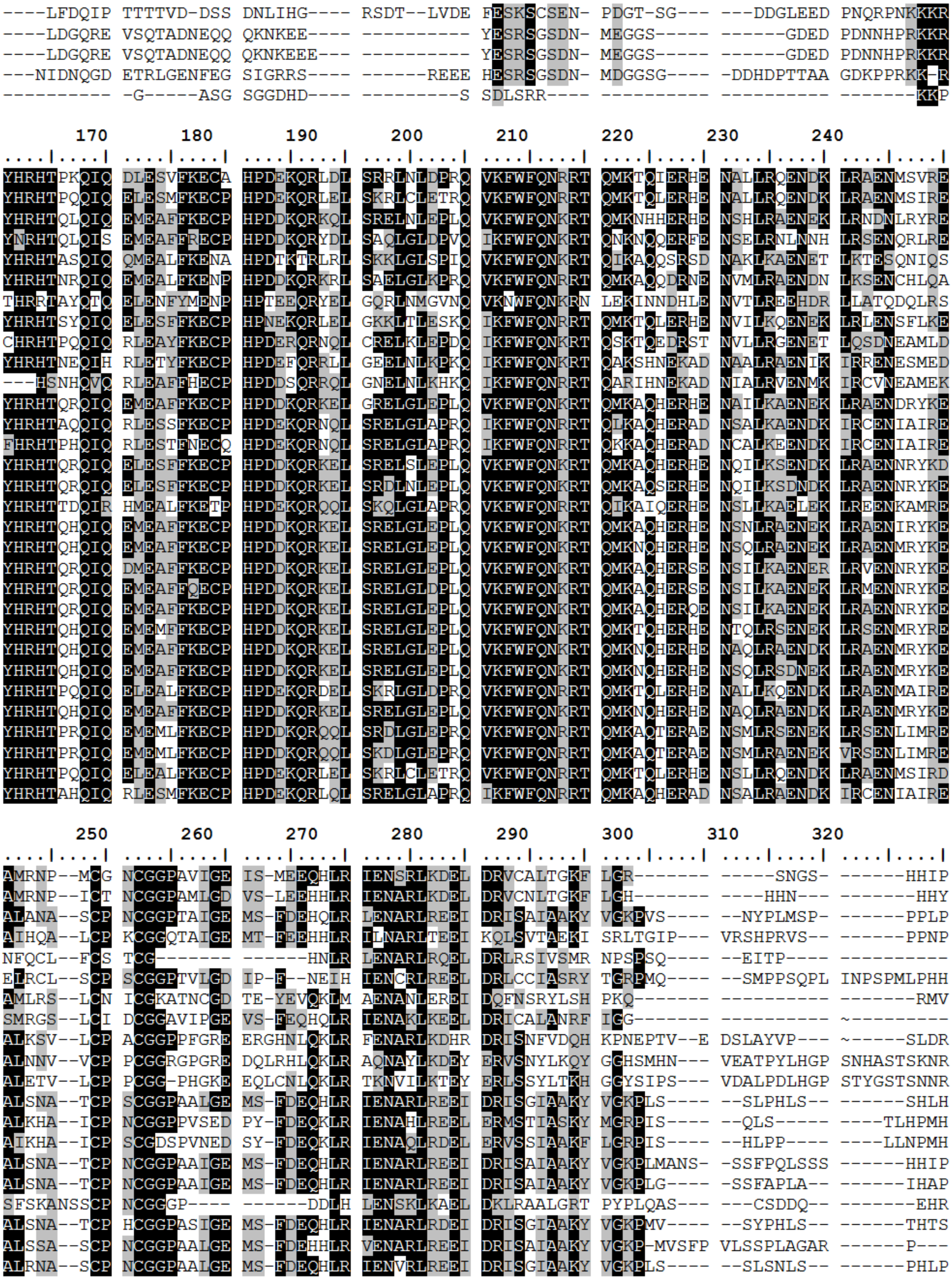


```
                 10         20         30         40         50         60         70         80
GaMYB2  ATGGCTCCAA AGAAGGATGG AGTGAGCAAA AGGGTTTTTA ACAAAGGTTC TTGGACAGCT GAGGAAGATA GAAGATTGGC
GbMYB2  ATGGCTCCAA AGAAGGCTGG AGTGAGCAAA AGGGTTTTTA ACAAAGGTTC ATGGACAGCT GAGGAAGATA GAAGATTGGC
GhMYB2  ATGGCTCCAA AGAAGGCTGG AGTGAGCAAA AGGGTTTTTA ACAAAGGTTC ATGGACAGCT GAGGAAGATA GAAGATTGGC

                 90        100        110        120        130        140        150        160
GaMYB2  TAAATATATT GAGATTCATG GCGCAAAGAG ATGGAAAACA ATCGCCATTA AATCAGGTTT GAATCGATGC GGCAAGAGTT
GbMYB2  TAAATATATT GAGATTCATG GTGCAAAGAG ATGGAAAACA ATCGCCATTA AATCAGGTTT GAATCGATGC GGCAAGAGTT
GhMYB2  TAAATATATT GAGATTCATG GTGCAAAGAG ATGGAAAACA ATCGCCATTA AATCAGGTTT GAATCGATGC GGCAAGAGTT

                170        180        190        200        210        220        230        240
GaMYB2  GCAGGTTGAG ATGGTTGAAC TACTTGAGAC CTAACATTAA GAGAGGCAAC ATATCAGATG AAGAAGAGGA CTTAATTATT
GbMYB2  GCTGGTTGAG ATGGTTGAAC TACTTGAGAC CTAACATTAA GAGAGGCAAC ATATCAGATG AAGAAGAGGA CTTAATTATT
GhMYB2  GCAGGTTGAG ATGGTTGAAC TACTTGAGAC CTAACATTAA GAGAGGCAAC ATATCAGATG AAGAAGAGGA CTTAATTATT

                250        260        270        280        290        300        310        320
GaMYB2  AGGCTTCATA AACTGCTGGG GAACAGGTGG TCTTTGATTG CTGGGAGACT TCCAGGGCGA ACAGACAATG AAATTAAGAA
GbMYB2  AGGCTTCATA AACTGCTGGG GAACAGGTGG TCTTTGATTG CTGGGAGACT TCTAGGGCGA ACAGACAATG AAATTAAGAA
GhMYB2  AGGCTTCATA AACTGCTGGG AAACAGGTGG TCTTTGATTG CTGGGAGACT TCCAGGGCGA ACAGACAATG AAATTAAGAA

                330        340        350        360        370        380        390        400
GaMYB2  CTACTGGAAT TCCCATTTGA GCAAGAAAAT AATAAACCAT GATGTCAGAA CAGAACAAAC TTCCTCCTCG GAACAAATTG
GbMYB2  CTACTGGAAT TCCCATTTGA GCAAGAAAAT AATAAACCAT GATGTCAGAA CAGAACAAAC TTCCTCCTCG GAACAAATTG
GhMYB2  CTACTGGAAT TCCCATTTGA GCAAGAAAAT GATAAACCAT GATGTCAGAA CAGAACAAAC TTCCTCCTCG GAACAAATTG

                410        420        430        440        450        460        470        480
GaMYB2  TGCCTCACAA AGCATGGGAA ACTGTCCAGA TGGAAGAAGA AGAGGTAGTA AAAGGAAGTG ATGAAATTGA AAACTCTGAA
GbMYB2  TGCCTCACAA AGCATGGGAA ACTGTCCAGA TGGAAGAAGA AGAGGTAGTA AAAGGAAGTG ATGAAATTGA AAACTCTGAA
GhMYB2  TGCCTCACAA AGCATGGGAA ACTGTCCAGA TGGAAGAAGA AGAGGTAGTA AAAGGAAGTG ATGAAATTGA AAACTCTGAA

                490        500        510        520        530        540        550        560
GaMYB2  TTCAGCATTG ATGTGGACGA ATTCTTTGAC TTCACAACGG AAGGTTGCTT TAGTTTGGAT TGGGTGAATA AGTTCCTTGA
GbMYB2  TTCAGCATTG ATGTGGACGA ATTCTTTGAC TTCACAACGG AAGGTTGCTT TAGTTTGGAT TGGGTGAATA AGTTCCTTGA
GhMYB2  TTCAGCATTG ATGTGGACGA ATTCTTTGAC TTCTCAACGG AAGGTTGCTT TGGTTTGGAT TGGGTGAATA AGTTCCTTGA

                570        580        590
GaMYB2  ACTTGATGAT CAACAGGATC CATTAGCAAT GGTATAA
GbMYB2  ACTTGATGAT CAACAGGATC CATTAGCAAT GGTATAA
GhMYB2  ACTTGATGAC CAACAGGATC CATTAGCAAT GGTATAA
```

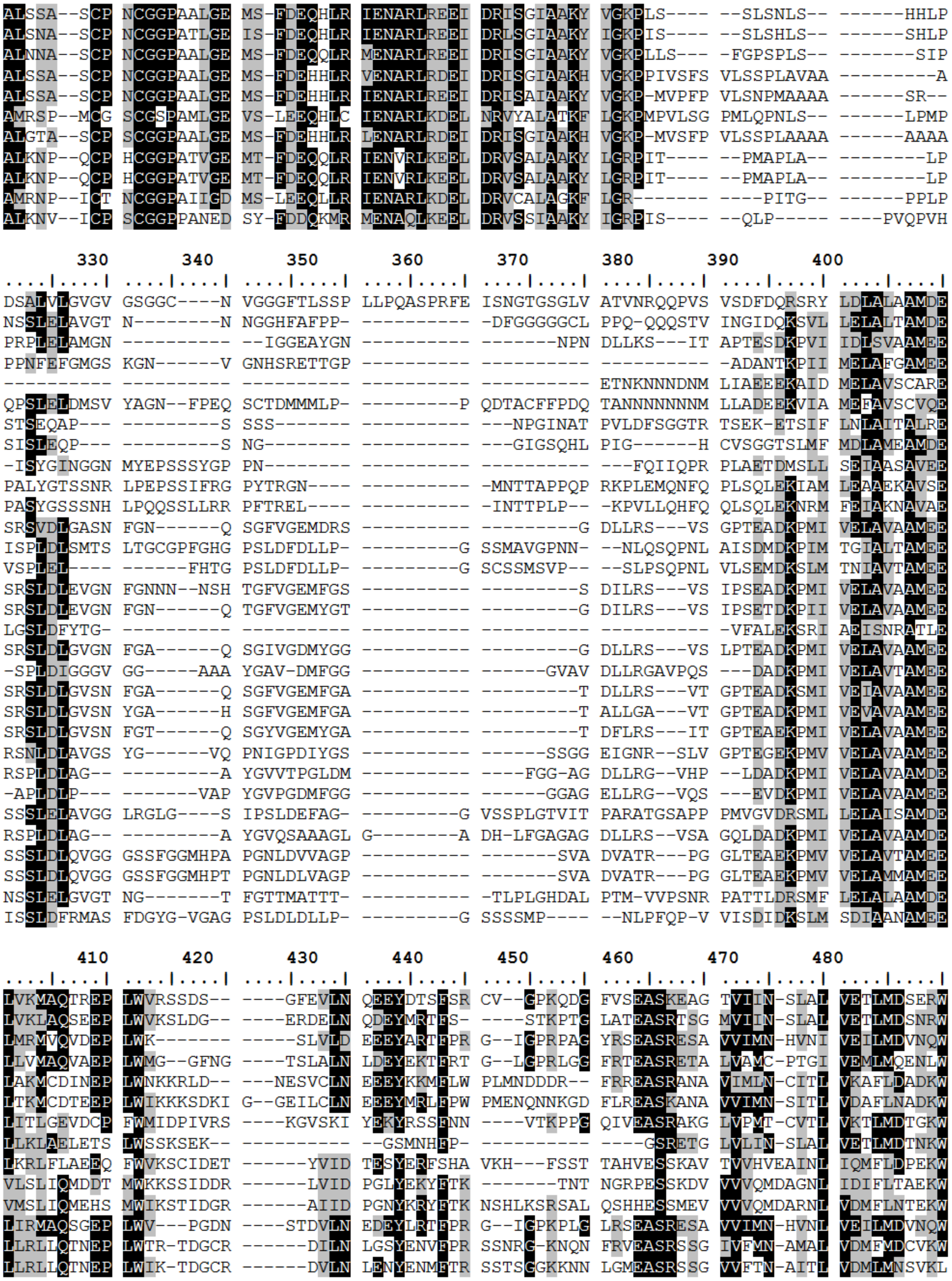
Supplementary Figure 6



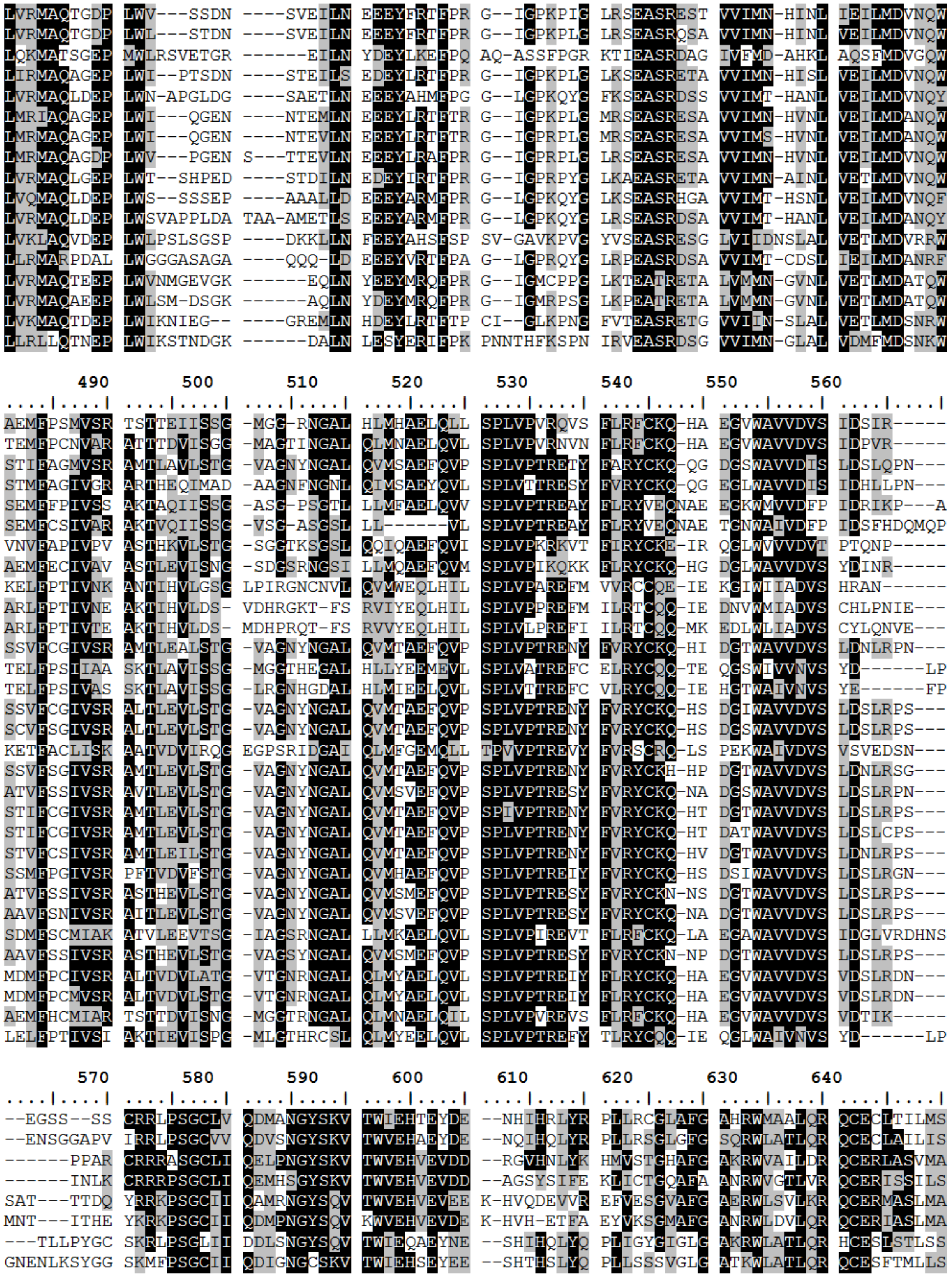
```
        10         20         30         40         50         60         70         80
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MNFNGFLDDG AG----ASKL LSDAPYNNHF SFSAVDTMLG SAAIAP---- ---SQSLP-- -FSSSGLSLG LQTN------
MNFGSLFDNT PGGGSTGARL LSGLSYGNHT --AATNVLPG GAMAQAAAAA SLFSPPLTKS VYASSGLSLA LEQPERGTNR
---------- ---------- ---------- ---------- ---------- -----MFEPN MLLAAMNNAD SNNHN-----
---------- ---------- ---------- ---------- ---------- -----MSQSN MVPVANNGDN NNDN------
---------- --METKDKKE KGHMVLNSDN VFGS----VS SSPTTTIQNP NY-FTSFENP NFPYIFPKEE YEVMS-----
---------- --MLTMGEGN ----VMTSNN RFASPPQQPS SSSPGTIQNP NFNFIPFN-- SYSSIIPKEE HGMMSMMMMM
---------- ---------- ---------- ---------- ---------- ---------- --MNGQGDLD AVGN------
---------- ---------- ---------- ---------- ---------- ---------- --MNGDLEVD MSRG------
---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ---------- -----MFSPN LFES-PHMFD MSHK------
---------- ---------- ---------- ---------- ---------- ---------- -MSFVVGVGG S---------
---------- ---------- ---------- ---------- ---------- ---------- -MEFLG---- ----------
---------- ---------- ---------- ---------- ---------- -----MYHPN MFESHHHMFD MTPK------
---------- ---------- ---------- ---------- ---------- -----MYHPN MFES-HHMFD MTPKS-----
---------- --MSMAVDMS SKQPTKDFFS SPALSLSLAG IFRNASSGST NPEEDFLGRR VVDDEDRTVE MSSE------
---------- ---------- ---------- ---------- ---------- -----MFQPN MFDSHHHLLD MPHK------
---------- --------MM IPARHMPP-- MIVRNSAAAY GS-SSALS-- -----LSQPN LLDSQQ-QLQ LQQALQQQH-
---------- ---------- ---------- ---------- ---------- -----MFQAN MFEG-HHMFD MASK------
---------- ---------- ---------- ---------- ---------- ---------- MFDS-HHMLD MTPK------
---------- ---------- ---------- ---------- ---------- -----MFQPA LFES-HHMFD MTPK------
---------- ----MPAGFI IPARHVPP-- -LMARHTKAG YTTSSGLS-- -----LAQPN MAEG---QFD MSQT------
---------- ---------M TPARRMPP-- VIGRN-GVAY ESPSAQLP-- -----LTQAD MLDSHHLQQA LQQQ------
---------- --------MM IPARHMPS-- MIGRN-GAAY GS-SSALS-- -----LSQPN LLDNHQFQQA FQHQQQHH-
MSFGSLFDGG SG-------- ---------- -----GGGGG GGMQFP---- ------FTTG FSSSPALSLG LDNAG-----
---------- ---------M TPARRVPPAA MIGRN-GFAF GS-SSALS-- -----LSQAD LLDSHQLQQA FQQQ------
---------- ---MRLGGVT VQVPPNGNLG NLGRSFGELG GMPAGAMAPP RQIPPLAMPR PPPKHYSSPS LSLGPPTG--
---------- ---MRLGGVT VQVPPNGNLG NLGRSFGELG GMPTGAMAPP RQIPPLAMPR PPPKHYSSPS LSLGPPTG--
---------- ---------- ---------- -------MPT GVFSQP---- RLVPSSIPKN MFNSPGLSLA LQQP------
---------- ---------- ---------- ---------- ---------- ---------- -MDCGSGGGG L---------

        90        100        110        120        130        140        150        160
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
--~~-----GE MSRNGEIMES NVSRKSS--- ------RGED VESRSESDN- AEAVSG---- -DDLD----T SDRPLKKKKR
GEASMRNNNN VGGGGDTFDG SVNRRS---- ------REEE HESRSGSDN- VEGISG---- -EDQD----A ADKPPRKK-R
---------- -YNHEDNNNE GFLRDDE--- -------FDS PNTKSGSEN- QEGGS----- ----GND-QD PL-HPNKKKR
---------- ---ENNNNNN NNGGTDN--- --------TN AGNDSGDQDF DSGNTSSG-- --NHGEGLGN NQAPRHKKKK
---------- -----KIESG -SGKST---- ---------G SGHDPVENTA IEQE------ ---------- --PPAAKKKR
GDG------- -TVEEMMENG SAGGSF---- ---------G SGSEQAEDPK FGNESDV--- ----NELHDD EQPPPAKKKR
---------- ---------- ---------- ------IPKP GEAEGDEIDM INDMSG---- ---------V NDQDGGRMRR
---------- -DFNPSFFLG KL-------- ------KDDE FESRSLSDDS FDAMSGD--- -EDKQ---EQ RPKKKKRKTK
---------- ---------- ---------- -------MDN NGGGSSGNEQ YTSGDAKQ-- ---------- --NGKR---T
---------- ---------- ---------- -------MDF TRDDNSSDER ENDVDANT-- ---------- --NNRHEKKG
---------- ---------- ---------- -------MDS SHNDSSSD-- EEGIDSN--- ---------- --NRRH----
---------- ---TSESELM GKIRDD---- ---------D YEIKSVNET- MDAPS----- ----GDD-QD PD-QRPKKKR
---------- -GSGSGGDGG GSHHHD---- ---------G SETDRK---- ---------- ---------- -------KKR
---------- ---------- DSQNHD---- ---------S SETEKKNK-- ---------- ---------- ------KKKR
---------- ---NSENDLG ITGSHEE--- ---------D FETKSGAEVT MENPL----- ----EEELQD PNQRPNKKKR
---------- ---TSDNDLG ITGSRED--- ---------D FETKSGTEVT TENPS----- ----GEELQD PSQRPNKKKR
---------- ------NSGP TRSRSE---- --------ED LEGEDHDDEE EEEEDG---- ------AAGN KGTNKRKRKK
---------- ---TPESEMG -KIRDE---- ---------E FESKSGTEN- MDAPS----- ----GDD-QD PN-QRPKKKR
----LLDQIP ATTAES---G DNMGRSGGGR GSDPLMGGDE FESKSGSEN- VDGV------ ----SVDDQD PNQRPSKKKR
---------- ---SSENDSS -KLKDD---- ---------D YETKSGTET- MEAQS----- ----GDD-QD PSEQHPKKKR
---------- ---SSENDLS -KLKDD---- ---------D YETKSGTET- MEAPS----- ----GDD-QD PSGQRPKKKR
---------- ---SSENELG -NLKDDDY-- ---------D HETKSGTET- TEAPS----- ----GDD-QD PN-QRPKKKR
---------- -----ADSEI AKIREE---- ---------E FESKSGSDN- VEGAS----- -----GEDQD GERRP-RKKR
----YFDQIP VTTTAAADSG DNMLHG---- RADAGGLVDE FESKSCSEN- VDGAGDG--- ---LSGDDQD PNQRP-RKKR
----LLDQIP ATTAES---G DNMIRS---R ASDP-LGGDE FESKSGSEN- VDGV------ ----SVDDQD PNQRP-RKKR
---GGMVGRM LPVGGAPADG GMARDA---- ------ADAE NDSRSGSDH- LDAMSAGAGA -EDED--DAE PGNPRKRKKR
```



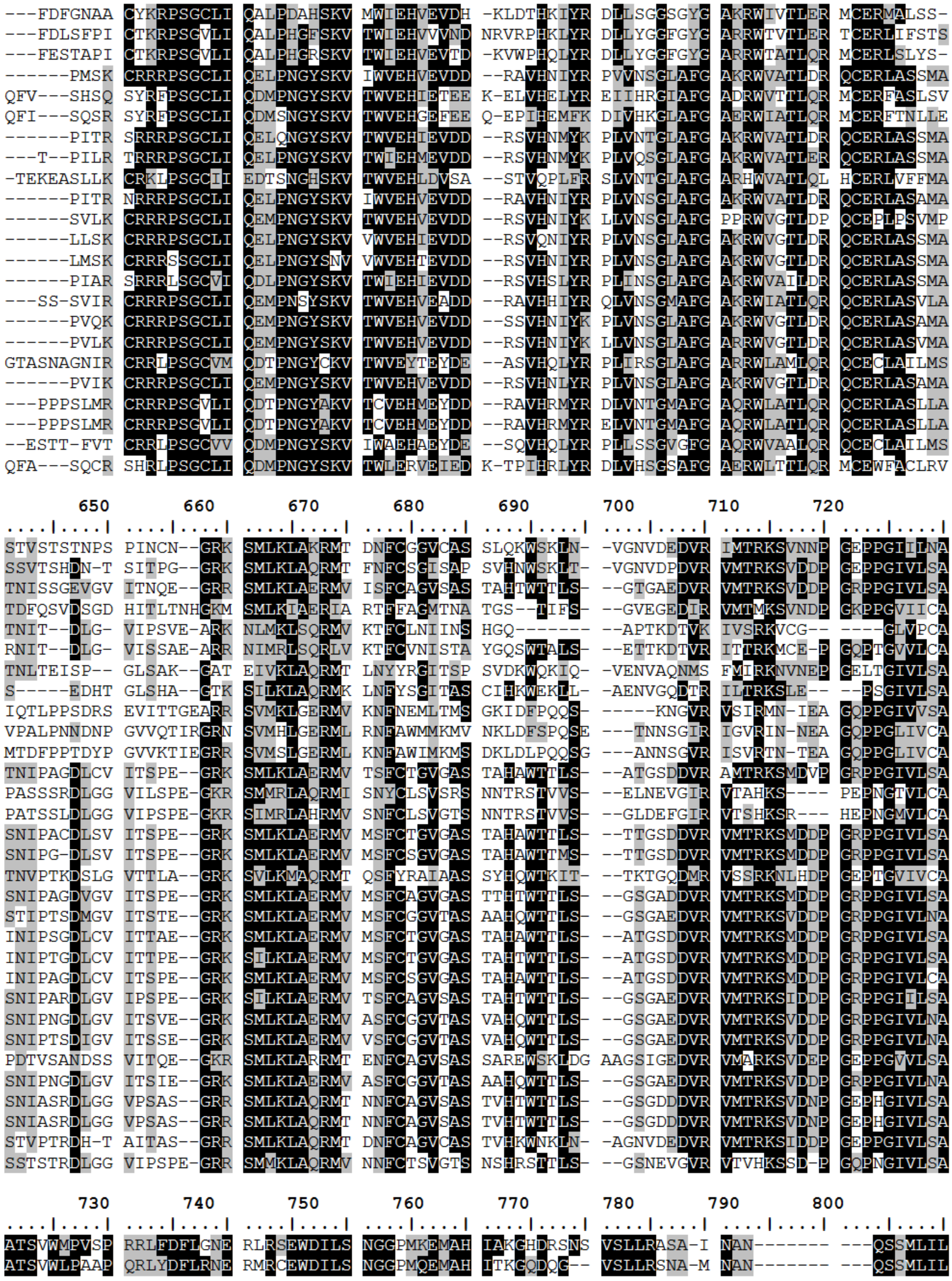
```
----LFDQIP TTTTVD-DSS DNLIHG---- RSDT--LVDE FESKSCSEN- PDGT-SG--- ---DDGLEED PNQRPNKKKR
----LDGQRE VSQTADNEQQ QKNKEE---- ---------- YESRSGSDN- MEGGS----- ------GDED PDNNHPRKKR
----LDGQRE VSQTADNEQQ QKNKEEE--- ---------- YESRSGSDN- MEGGS----- ------GDED PDNNHPRKKR
---NIDNQGD ETRLGENFEG SIGRRS---- ------REEE HESRSGSDN- MDGGSG---- -DDHDPTTAA GDKPPRKK-R
---------- -G-----ASG SGGDHD---- ---------S SDLSRR---- ---------- ---------- -------KKP

                170        180        190        200        210        220        230        240
       ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
YHRHTPKQIQ DLESVFKECA HPDEKQRLDL SRRLNLDPRQ VKFWFQNRRT QMKTQIERHE NALLRQENDK LRAENMSVRE
YHRHTPQQIQ ELESMFKECP HPDEKQRLEL SKRLCLETRQ VKFWFQNRRT QMKTQLERHE NALLRQENDK LRAENMSIRE
YHRHTQLQIQ EMEAFFKECP HPDDKQRKQL SRELNLEPLQ VKFWFQNKRT QMKNHHERHE NSHLRAENEK LRNDNLRYRE
YNRHTQLQIS EMEAFFRECP HPDDKQRYDL SAQLGLDPVQ IKFWFQNKRT QNKNQQERFE NSELRNLNNH LRSENQRLRE
YHRHTASQIQ QMEALFKENA HPDTKTRLRL SKKLGLSPIQ VKFWFQNKRT QIKAQQSRSD NAKLKAENET LKTESQNIQS
YHRHTNRQIQ EMEALFKENP HPDDKQRKRL SAELGLKPRQ VKFWFQNRRT QMKAQQDRNE NVMLRAENDN LKSENCHLQA
THRRTAYQTQ ELENFYMENP HPTEEQRYEL GQRLNMGVNQ VKNWFQNKRN LEKINNDHLE NVTLREEHDR LLATQDQLRS
YHRHTSYQIQ ELESFFKECP HPNEKQRLEL GKKLTLESKQ IKFWFQNRRT QMKTQLERHE NVILKQENEK LRLENSFLKE
CHRHTPQQIQ RLEAYFKECP HPDERQRNQL CRELKLEPDQ IKFWFQNKRT QSKTQEDRST NVLLRGENET LQSDNEAMLD
YHRHTNEQIH RLETYFKECP HPDEFQRRLL GEELNLKPKQ IKFWFQNKRT QAKSHNEKAD NAALRAENIK IRRENESMED
---HSNHQVQ RLEAFFHECP HPDDSQRRQL GNELNLKHKQ IKFWFQNRRT QARIHNEKAD NIALRVENMK IRCVNEAMEK
YHRHTQRQIQ EMEAFFKECP HPDDKQRKEL GRELGLEPLQ VKFWFQNKRT QMKAQHERHE NAILKAENEK LRAENDRYKE
YHRHTAQQIQ RLESSFKECP HPDEKQRNQL SRELGLAPRQ IKFWFQNRRT QLKAQHERAD NSALKAENDK IRCENIAIRE
FHRHTPHQIQ RLESTFNECQ HPDEKQRNQL SRELGLAPRQ IKFWFQNRRT QKKAQHERAD NCALKEENDK IRCENIAIRE
YHRHTQRQIQ ELESFFKECP HPDDKQRKEL SRELSLEPLQ VKFWFQNKRT QMKAQHERHE NQILKSENDK LRAENNRYKD
YHRHTQRQIQ ELESFFKECP HPDDKQRKEL SRDLNLEPLQ VKFWFQNKRT QMKAQSERHE NQILKSDNDK LRAENNRYKE
YHRHTTDQIR HMEALFKETP HPDEKQRQQL SKQLGLAPRQ VKFWFQNRRT QIKAIQERHE NSLLKAELEK LREENKAMRE
YHRHTQHQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKAQHERHE NSNLRAENEK LRAENIRYKE
YHRHTQHQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKNQHERHE NSQLRAENEK LRAENMRYKE
YHRHTQRQIQ DMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKAQHERSE NSILKAENER LRVENNRYKE
YHRHTQRQIQ EMEAFFQECP HPDDKQRKEL SRELGLDPLQ VKFWFQNKRT QMKAQHERSE NSILKAENEK LRMENNRYKE
YHRHTQRQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKAQHERQE NSILKAENEK LRAENNRYKE
YHRHTQHQIQ EMEMFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKTQHERHE NTQLRSENEK LRSENMRYRE
YHRHTQHQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKNQHERHE NAQLRAENDK LRAENMRYKE
YHRHTQHQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKNQHERHE NSQLRSDNEK LRAENMRYKE
YHRHTPQQIQ ELEALFKECP HPDEKQRDEL SKRLGLDPRQ VKFWFQNRRT QMKTQLERHE NALLKQENDK LRAENMAIRE
YHRHTQHQIQ EMEAFFKECP HPDDKQRKEL SRELGLEPLQ VKFWFQNKRT QMKNQHERHE NAQLRAENDK LRAENMRYKE
YHRHTPRQIQ EMEMLFKECP HPDDKQRQQL SRDLGLEPRQ VKFWFQNRRT QMKAQTERAE NSMLRSENEK LRSENLIMRE
YHRHTPRQIQ EMEMLFKECP HPDDKQRQQL SKDLGLEPRQ VKFWFQNRRT QMKAQTERAE NSMLRAENEK VRSENLIMRE
YHRHTPQQIQ ELEALFKECP HPDEKQRLEL SKRLCLETRQ VKFWFQNRRT QMKTQLERHE NSLLRQENDK LRAENMSIRD
YHRHTAHQIQ RLESMFKECP HPDEKQRLQL SRELGLAPRQ IKFWFQNRRT QMKAQHERAD NSALRAENDK IRCENIAIRE

                250        260        270        280        290        300        310        320
       ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AMRNP--MCG NCGGPAVIGE IS-MEEQHLR IENSRLKDEL DRVCALTGKF LGR------- -----SNGS- ------HHIP
AMRNP--ICT NCGGPAMLGD VS-LEEHHLR IENARLKDEL DRVCNLTGKF LGH------- ----HHN--- -------HHY
ALANA--SCP NCGGPTAIGE MS-FDEHQLR LENARLREEI DRISAIAAKY VGKPVS---- --NYPLMSP- ------PPLP
AIHQA--LCP KCGGQTAIGE MT-FEEHHLR ILNARLTEEI KQLSVTAEKI SRLTGIP--- VRSHPRVS-- ------PPNP
NFQCL--FCS TCG------- ------HNLR LENARLRQEL DRLRSIVSMR NPSPSQ---- --EITP---- ----------
ELRCL--SCP SCGGPTVLGD IP-F--NEIH IENCRLREEL DRLCCIASRY TGRPMQ---- --SMPPSQPL INPSPMLPHH
AMLRS--LCN ICGKATNCGD TE-YEVQKLM AENANLEREI DQFNSRYLSH PKQ------- ---------- -------RMV
SMRGS--LCI DCGGAVIPGE VS-FEQHQLR IENAKLKEEL DRICALANRF IGG------- ---------- ~---------
ALKSV--LCP ACGGPPFGRE ERGHNLQKLR FENARLKDHR DRISNFVDQH KPNEPTV--E DSLAYVP--- ~-----SLDR
ALNNV--VCP PCGGRGPGRE DQLRHLQKLR AQNAYLKDEY ERVSNYLKQY GGHSMHN--- VEATPYLHGP SNHASTSKNR
ALETV--LCP PCGG-PHGKE EQLCNLQKLR TKNVILKTEY ERLSSYLTKH GGYSIPS--- VDALPDLHGP STYGSTSNNR
ALSNA--TCP SCGGPAALGE MS-FDEQHLR IENARLREEI DRISGIAAKY VGKPLS---- --SLPHLS-- ------SHLH
ALKHA--ICP NCGGPPVSED PY-FDEQKLR IENAHLREEL ERMSTIASKY MGRPIS---- --QLS----- ----TLHPMH
AIKHA--ICP SCGDSPVNED SY-FDEQKLR IENAQLRDEL ERVSSIAAKF LGRPIS---- --HLPP---- ----LLNPMH
ALSNA--TCP NCGGPAAIGE MS-FDEQHLR IENARLREEI DRISAIAAKY VGKPLMANS- -SSFPQLSSS ------HHIP
ALSNA--TCP NCGGPAAIGE MS-FDEQHLR IENARLREEI DRISAIAAKY VGKPLG---- -SSFAPLA-- ------IHAP
SFSKANSSCP NCGGGP---- ------DDLH LENSKLKAEL DKLRAALGRT PYPLQAS--- ---CSDDQ-- -------EHR
ALSNA--TCP HCGGPASIGE MS-FDEQHLR IENARLRDEI DRISGIAAKY VGKPMV---- --SYPHLS-- ------THTS
ALSSA--SCP NCGGPAALGE MS-FDEHHLR VENARLREEI DRISAIAAKY VGKP-MVSFP VLSSPLAGAR ------P---
ALRNA--SCP NCGGPAALGE MS-FDEQHLR IENVRLREEI DRISGIAAKY VGKPLS---- --SLSNLS-- ------PHLP
```



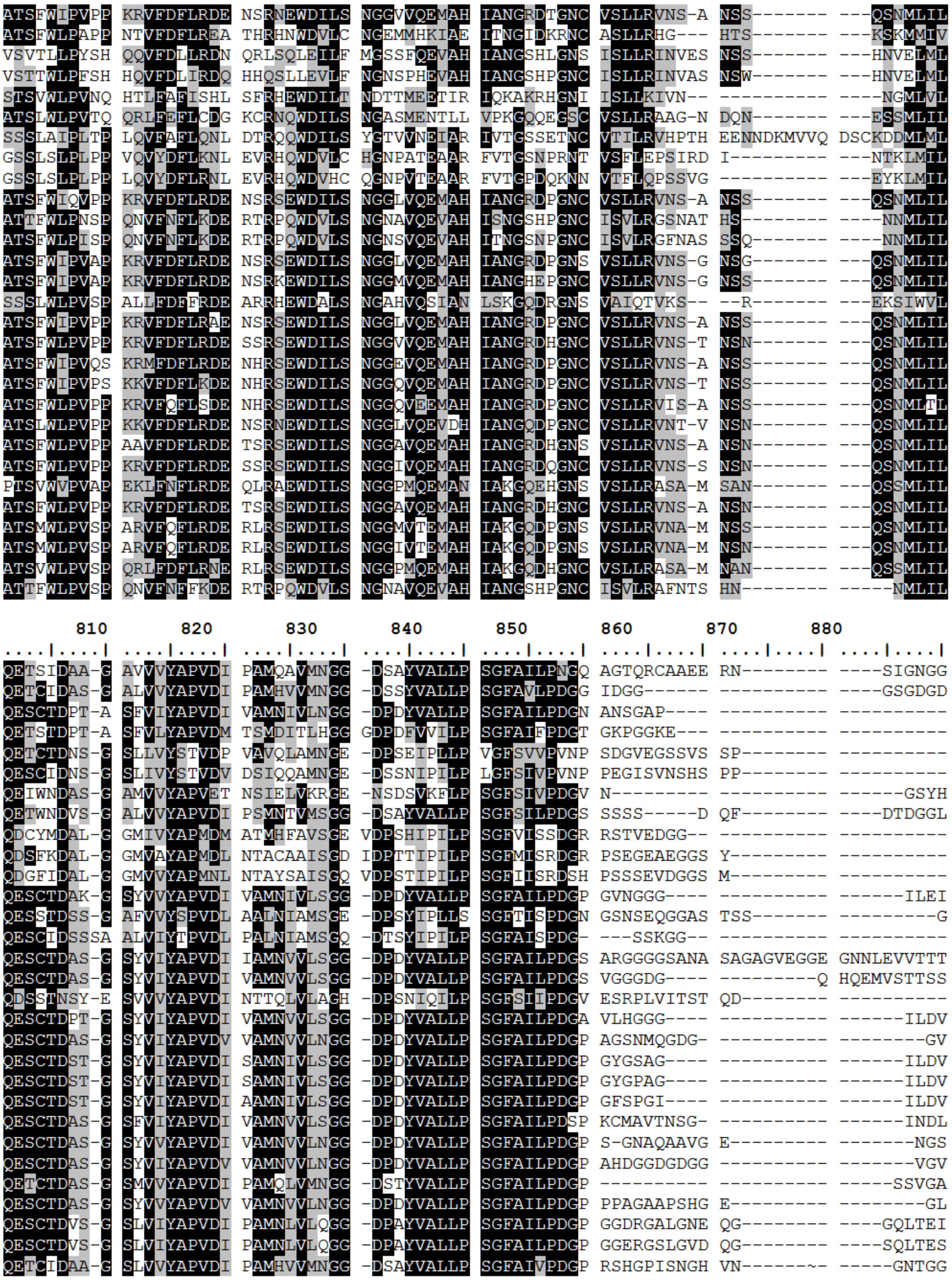
```
ALSSA--SCP NCGGPAALGE MS-FDEQHLR IENARLREEI DRISGIAAKY VGKPLS---- --SLSNLS-- ------HHLP
ALSNA--SCP NCGGPATLGE IS-FDEQHLR IENARLREEI DRLSGIAAKY IGKPIS---- --SLSHLS-- ------SHLP
ALNNA--SCP NCGGPAALGE MS-FDEQQLR MENARLREEI DRISGIAAKY VGKPLLS--- -FGPSPLS-- -------SIP
ALSSA--SCP NCGGPAALGE MS-FDEHHLR VENARLRDEI DRISGIAAKH VGKPPIVSFS VLSSPLAVAA ---------A
ALSSA--SCP NCGGPAALGE MS-FDEHHLR IENARLREEI DRISAIAAKY VGKP-MVPFP VLSNPMAAAA ------SR--
AMRSP--MCG SCGSPAMLGE VS-LEEQHLC IENARLKDEL NRVYALATKF LGKPMPVLSG PMLQPNLS-- ------LPMP
ALGTA--SCP SCGGPAALGE MS-FDEHHLR LENARLRDEI DRISGIAAKH VGKP-MVSFP VLSSPLAAAA ------AAAA
ALKNP--QCP HCGGPATVGE MT-FDEQQLR IENVRLKEEL DRVSALAAKY LGRPIT---- --PMAPLA-- --------LP
ALKNP--QCP HCGGPATVGE MT-FDEQQLR IENVRLKEEL DRVSALAAKY LGRPIT---- --PMAPLA-- --------LP
AMRNP--ICT NCGGPAIIGD MS-LEEQLLR IENARLKDEL DRVCALAGKF LGR------- ----PITG-- ------PPLP
ALKNV--ICP SCGGPPANED SY-FDDQKMR MENAQLKEEL DRVSSIAAKY IGRPIS---- --QLP----- ----PVQPVH

       330        340        350        360        370        380        390        400
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
DSALVLGVGV GSGGC----N VGGGFTLSSP LLPQASPRFE ISNGTGSGLV ATVNRQQPVS VSDFDQRSRY LDLALAAMDE
NSSLELAVGT N--------N NGGHFAFPP- ---------- -DFGGGGGCL PPQ-QQQSTV INGIDQKSVL LELALTAMDE
PRPLELAMGN ---------- --IGGEAYGN ---------- -------NPN DLLKS---IT APTESDKPVI IDLSVAAMEE
PPNFEFGMGS KGN------V GNHSRETTGP ---------- ---------- ---------- -ADANTKPII MELAFGAMEE
---------- ---------- ---------- ---------- ---------- ETNKNNNDNM LIAEEEKAID MELAVSCARE
QPSLELDMSV YAGN--FPEQ SCTDMMMLP- ---------P QDTACFFPDQ TANNNNNNNM LLADEEKVIA MEFAVSCVQE
STSEQAP--- ---------S SSS------- ---------- ---NPGINAT PVLDFSGGTR TSEK-ETSIF LNLAITALRE
SISLEQP--- ---------S NG-------- ---------- ---GIGSQHL PIG------H CVSGGTSLMF MDLAMEAMDE
-ISYGINGGN MYEPSSSYGP PN-------- ---------- ---------- ---FQIIQPR PLAETDMSLL SEIAASAVEE
PALYGTSSNR LPEPSSIFRG PYTRGN---- ---------- -MNTTAPPQP RKPLEMQNFQ PLSQLEKIAM LEAAEKAVSE
PASYGSSSNH LPQQSSLLRR PFTREL---- ---------- -INTTPLP-- -KPVLLQHFQ QLSQLEKNRM FEIAKNAVAE
SRSVDLGASN FGN------Q SGFVGEMDRS ---------- ---------G DLLRS---VS GPTEADKPMI VELAVAAMEE
ISPLDLSMTS LTGCGPFGHG PSLDFDLLP- ---------G SSMAVGPNN- --NLQSQPNL AISDMDKPIM TGIALTAMEE
VSPLEL---- ------FHTG PSLDFDLLP- ---------G SCSSMSVP-- --SLPSQPNL VLSEMDKSLM TNIAVTAMEE
SRSLDLEVGN FGNNN--NSH TGFVGEMFGS ---------- ---------S DILRS---VS IPSEADKPMI VELAVAAMEE
SRSLDLEVGN FGN------Q TGFVGEMYGT ---------- ---------G DILRS---VS IPSETDKPII VELAVAAMEE
LGSLDFYTG- ---------- ---------- ---------- ---------- ---------- -VFALEKSRI AEISNRATLE
SRSLDLGVGN FGA------Q SGIVGDMYGG ---------- ---------G DLLRS---VS LPTEADKPMI VELAVAAMEE
-SPLDIGGGV GG-----AAA YGAV-DMFGG ---------- ------GVAV DLLRGAVPQS ---DADKPMI VELAVTAMEE
SRSLDLGVSN FGA------Q SGFVGEMFGA ---------- ---------T DLLRS---VT GPTEADKSMI VEIAVAAMEE
SRSLDLGVSN YGA------H SGFVGEMFGA ---------- ---------T ALLGA---VT GPTEADKPMI VEVAVAAMEE
SRSLDLGVSN FGT------Q SGYVGEMYGA ---------- ---------T DFLRS---IT GPTEAEKPMI VELAVAAMEE
RSNLDLAVGS YG------VQ PNIGPDIYGS ---------- ------SSGG EIGNR--SLV GPTEGEKPMV VELAVAAMEE
RSPLDLAG-- ---------A YGVVTPGLDM ---------- ----FGG-AG DLLRG--VHP --LDADKPMI VELAVAAMDE
-APLDLP--- -------VAP YGVPGDMFGG ---------- ------GGAG ELLRG--VQS ---EVDKPMI VELAVAAMEE
SSSLELAVGG LRGLG----S IPSLDEFAG- ---------G VSSPLGTVIT PARATGSAPP PMVGVDRSML LELAISAMDE
RSPLDLAG-- ---------A YGVQSAAAGL G--------A DH-LFGAGAG DLLRS--VSA GQLDADKPMI VELAVAAMDE
SSSLDLQVGG GSSFGGMHPA PGNLDVVAGP ---------- -------SVA DVATR---PG GLTEAEKPMV VELAVTAMEE
SSSLDLQVGG GSSFGGMHPT PGNLDLVAGP ---------- -------SVA DVATR---PG GLTEAEKPMV VELAMMAMEE
NSSLELGVGT NG-------T FGTTMATTT- ---------- -TLPLGHDAL PTM-VVPSNR PATTLDRSMF LELALAAMDE
ISSLDFRMAS FDGYG-VGAG PSLDLDLLP- ---------G SSSSMP---- --NLPFQP-V VISDIDKSLM SDIAANAMEE

       410        420        430        440        450        460        470        480
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
LVKMAQTREP LWVRSSDS-- ----GFEVLN QEEYDTSFSR CV--GPKQDG FVSEASKEAG TVIIN-SLAL VETLMDSERW
LVKLAQSEEP LWVKSLDG-- ----ERDELN QDEYMRTFS- ----STKPTG LATEASRTSG MVIIN-SLAL VETLMDSNRW
LMRMVQVDEP LWK------- -----SLVLD EEEYARTFPR G--IGPRPAG YRSEASRESA VVIMN-HVNI VEILMDVNQW
LLVMAQVAEP LWMG--GFNG ----TSLALN LDEYEKTFRT G--LGPRLGG FRTEASRETA LVAMC-PTGI VEMLMQENLW
LAKMCDINEP LWNKKRLD-- ---NESVCLN EEEYKKMFLW PLMNDDDR-- FRREASRANA VIMLN-CITL VKAFLDADKW
LTKMCDTEEP LWIKKKSDKI G--GEILCLN EEEYMRLFPW PMENQNNKGD FLREASKANA VVIMN-SITL VDAFLNADKW
LITLGEVDCP FWMIDPIVRS ----KGVSKI YEKYRSSFNN ----VTKPPG QIVEASRAKG LVPMT-CVTL VKTLMDTGKW
LLKLAELETS LWSSKSEK-- ---------- --GSMNHFP- ---------- ----GSRETG LVLIN-SLAL VETLMDTNKW
LKRLFLAEEQ FWVKSCIDET ------YVID TESYERFSHA VKH---FSST TAHVESSKAV TVVHVEAINL IQMFLDPEKW
VLSLIQMDDT MWKKSSIDDR ------LVID PGLYEKYFTK -------TNT NGRPESSKDV VVVQMDAGNL IDIFLTAEKW
VMSLIQMEHS MWIKSTIDGR ------AIID PGNYKRYFTK NSHLKSRSAL QSHHESSMEV VVVQMDARNL VDMFLNTEKW
LIRMAQSGEP LWV---PGDN ----STDVLN EDEYLRTFPR G--IGPKPLG LRSEASRESA VVIMN-HVNL VEILMDVNQW
LLRLLQTNEP LWTR-TDGCR ------DILN LGSYENVFPR SSNRG-KNQN FRVEASRSSG IVFMN-AMAL VDMFMDCVKW
LLRLLQTNEP LWIK-TDGCR ------DVLN LENYENMFTR SSTSGGKKNN LGMEASRSSG VVFTN-AITL VDMLMNSVKL
```



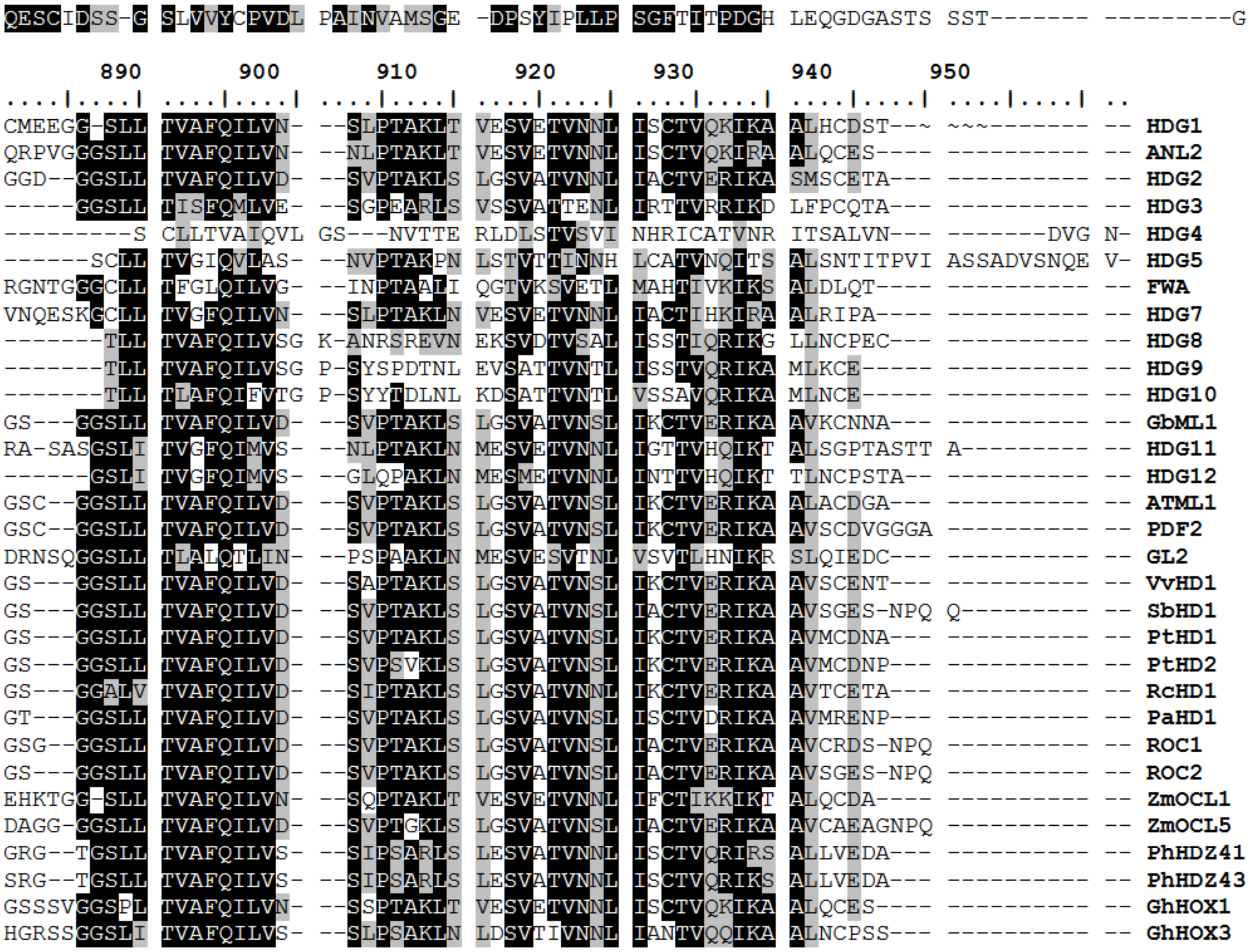
```
LVRMAQTGDP LWV---SSDN ----SVEILN EEEYFRTFPR G--IGPKPIG LRSEASREST VVIMN-HINL IEILMDVNQW
LVRMAQTGDP LWL---STDN ----SVEILN EEEYFRTFPR G--IGPKPLG LRSEASRQSA VVIMN-HINL VEILMDVNQW
LQKMATSGEP MWLRSVETGR ------EILN YDEYLKEFPQ AQ-ASSFPGR KTIEASRDAG IVFMD-AHKL AQSFMDVGQW
LIRMAQAGEP LWI--PTSDN ----STEILS EDEYLRTFPR G--IGPKPLG LKSEASRETA VVIMN-HISL VEILMDVNQW
LVRMAQLDEP LWN-APGLDG ----SAETLN EEEYAHMFPG G--LGPKQYG FKSEASRDSS VVIMT-HANL VEILMDVNQY
LMRIAQAGEP LWI---QGEN ----NTEMLN EEEYLRTFTR G--IGPKPLG MRSEASRESA VVIMN-HVNL VEILMDANQW
LMRMAQAGEP LWI---QGEN ----NTEVLN EEEYLRTFTR G--IGPRPLG MRSEASRESA VVIMS-HVNL VEILMDANQW
LMRMAQAGDP LWV---PGEN S---TTEVLN EEEYLRAFPR G--IGPRPLG LRSEASRESA VVIMN-HVNL VEILMDVNQW
LVRMAQLGEP LWT--SHPED ----STDILN EDEYIRTFPR G--IGPRPYG LKAEASRETA VVIMN-AINL VETLMDVNQW
LVQMAQLDEP LWS--SSSEP ----AAALLD EEEYARMFPR G--LGPKQYG LKSEASRHGA VVIMT-HSNL VEILMDVNQF
LVRMAQLDEP LWSVAPPLDA TAA-AMETLS EEEYARMFPR G--LGPKQYG LRSEASRDSA VVIMT-HANL VEILMDANQY
LVKLAQVDEP LWLPSLSGSP ----DKKLLN FEEYAHSFSP SV-GAVKPVG YVSEASRESG LVIIDNSLAL VETLMDVRRW
LLRMARPDAL LWGGGASAGA ----QQQ-LD EEEYVRTFPA G--LGPRQYG LRPEASRDSA VVIMT-CDSL IEILMDANRF
LVRMAQTEEP LWVNMGEVGK ------EQLN YEEYMRQFPR G--IGMCPPG LKTEATRETA LVMMN-GVNL VETLMDATQW
LVRMAQAEEP LWLSM-DSGK ------AQLN YDEYMRQFPR G--IGMRPSG LKPEATRETA LVMMN-GVNL VETLMDATQW
LVKMAQTDEP LWIKNIEG-- ----GREMLN HDEYLRTFTP CI--GLKPNG FVTEASRETG VVIIN-SLAL VETLMDSNRW
LLRLLQTNEP LWIKSTNDGK ------DALN LESYERIFPK PNNTHFKSPN IRVEASRDSG VVIMN-GLAL VDMFMDSNKW

       490        500        510        520        530        540        550        560
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AEMFPSMVSR TSTTEIISSG -MGG-RNGAL HLMHAELQLL SPLVPVRQVS FLRFCKQ-HA EGVWAVVDVS IDSIR-----
TEMFPCNVAR ATTTDVISGG -MAGTINGAL QLMNAELQVL SPLVPVRNVN FLRFCKQ-HA EGVWAVVDVS IDPVR-----
STIFAGMVSR AMTLAVLSTG -VAGNYNGAL QVMSAEFQVP SPLVPTRETY FARYCKQ-QG DGSWAVVDIS LDSLQPN---
STMFAGIVGR ARTHEQIMAD -AAGNFNGNL QIMSAEYQVL SPLVTTRESY FVRYCKQ-QG EGLWAVVDIS IDHLLPN---
SEMFFPIVSS AKTAQIISSG -ASG-PSGTL LLMFAELQVV SPLVPTREAY FLRYVEQNAE EGKWMVVDFP IDRIKP---A
SEMFCSIVAR AKTVQIISSG -VSG-ASGSL LL------VL SPLVPTREAY FLRYVEQNAE TGNWAIVDFP IDSFHDQMQP
VNVFAPIVPV ASTHKVLSTG -SGGTKSGSL QQIQAEFQVI SPLVPKRKVT FIRYCKE-IR QGLWVVVDVT PTQNP-----
AEMFECIVAV ASTLEVISNG -SDGSRNGSI LLMQAEFQVM SPLVPIKQKK FLRYCKQ-HG DGLWAVVDVS YDINR-----
KELFPTIVNK ANTIHVLGSG LPIRGNCNVL QVMWEQLHIL SPLVPAREFM VVRCCQE-IE KGIWIIADVS HRAN------
ARLFPTIVNE AKTIHVLDS- VDHRGKT-FS RVIYEQLHIL SPLVPPREFM ILRTCQQ-IE DNVWMIADVS CHLPNIE---
ARLFPTIVTE AKTIHVLDS- MDHPRQT-FS RVVYEQLHIL SPLVLPREFI ILRTCQQ-MK EDLWLIADVS CYLQNVE---
SSVFCGIVSR AMTLEALSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKQ-HI DGTWAVVDVS LDNLRPN---
TELFPSIIAA SKTLAVISSG -MGGTHEGAL HLLYEEMEVL SPLVATREFC ELRYCQQ-TE QGSWIVVNVS YD------LP
TELFPSIVAS SKTLAVISSG -LRGNHGDAL HLMIEELQVL SPLVTTREFC VLRYCQQ-IE HGTWAIVNVS YE------FP
SSVFCGIVSR ALTLEVLSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKQ-HS DGIWAVVDVS LDSLRPS---
SCVFSGIVSR ALTLEVLSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKQ-HS DGSWAVVDVS LDSLRPS---
KETFACLISK AATVDVIRQG EGPSRIDGAI QLMFGEMQLL TPVVPTREVY FVRSCRQ-LS PEKWAIVDVS VSVEDSN---
SSVFSGIVSR AMTLEVLSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKH-HP DGTWAVVDVS LDNLRSG---
ATVFSSIVSR AVTLEVLSTG -VAGNYNGAL QVMSVEFQVP SPLVPTRESY FVRYCKQ-NA DGSWAVVDVS LDSLRPN---
STIFCGIVSR AMTLEVLSTG -VAGNYNGAL QVMTAEFQVP SPIVPTRENY FVRYCKQ-HT DGTWAVVDVS LDSLRPS---
STIFCGIVSR AMTLEVLSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKQ-HT DATWAVVDVS LDSLCPS---
STVFCSIVSR AMTLEILSTG -VAGNYNGAL QVMTAEFQVP SPLVPTRENY FVRYCKQ-HV DGTWAVVDVS LDNLRPS---
SSMFPGIVSR PFTVDVFSTG -VAGNYNGAL QVMHAEFQVP SPLVPTREIY FVRYCKQ-HS DSIWAVVDVS LDSLRGN---
ATVFSSIVSR ASTHEVLSTG -VAGNYNGAL QVMSMEFQVP SPLVPTRESY FVRYCKN-NS DGTWAVVDVS LDSLRPS---
AAVFSNIVSR AITLEVLSTG -VAGNYNGAL QVMSVEFQVP SPLVPTRESY FVRYCKQ-NA DGTWAVVDVS LDSLRPS---
SDMFSCMIAK ATVLEEVTSG -IAGSRNGAL LLMKAELQVL SPLVPIREVT FLRFCKQ-LA EGAWAVVDVS IDGLVRDHNS
AAVFSSIVSR ASTHEVLSTG -VAGSYNGAL QVMSMEFQVP SPLVPTRESY FVRYCKN-NP DGTWAVVDVS LDSLRPS---
MDMFPCIVSR ALTVDVLATG -VTGNRNGAL QLMYAELQVL SPLVPTREIY FLRYCKQ-HA EGVWAVVDVS VDSLRDN---
MDMFPCMVSR ALTVDVLSTG -VTGNRNGAL QLMYAELQVL SPLVPTREIY FLRYCKQ-HA EGVWAVVDVS VDSLRDN---
AEMFHCMIAR TSTTDVISNG -MGGTRNGAL QLMNAELQIL SPLVPVREVS FLRFCKQ-HA EGVWAVVDVS VDTIK-----
LELFPTIVSI AKTIEVISPG -MLGTHRCSL QLMYEELQVL SPLVPTREFY TLRYCQQ-IE QGLWAIVNVS YD------LP

       570        580        590        600        610        620        630        640
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
--EGSS--SS CRRLPSGCLV QDMANGYSKV TWIEHTEYDE --NHIHRLYR PLLRCGLAFG AHRWMAALQR QCECLTILMS
--ENSGGAPV IRRLPSGCVV QDVSNGYSKV TWVEHAEYDE --NQIHQLYR PLLRSGLGFG SQRWLATLQR QCECLAILIS
------PPAR CRRRASGCLI QELPNGYSKV TWVEHVEVDD --RGVHNLYK HMVSTGHAFG AKRWVAILDR QCERLASVMA
------INLK CRRRPSGCLI QEMHSGYSKV TWVEHVEVDD --AGSYSIFE KLICTGQAFA ANRWVGTLVR QCERISSILS
SAT---TTDQ YRRKPSGCII QAMRNGYSQV TWVEHVEVEE K-HVQDEVVR EFVESGVAFG AERWLSVLKR QCERMASLMA
MNT---ITHE YKRKPSGCII QDMPNGYSQV KWVEHVEVDE K-HVH-ETFA EYVKSGMAFG ANRWLDVLQR QCERIASLMA
---TLLPYGC SKRLPSGLII DDLSNGYSQV TWIEQAEYNE --SHIHQLYQ PLIGYGIGLG AKRWLATLQR HCESLSTLSS
GNENLKSYGG SKMFPSGCII QDIGNGCSKV TWIEHSEYEE --SHTHSLYQ PLLSSSVGLG ATKWLATLQR QCESFTMLLS
```



```
---FDFGNAA CYKRPSGCLI QALPDAHSKV MWIEHVEVDH -KLDTHKIYR DLLSGGSGYG AKRWIVTLER MCERMALSS-
---FDLSFPI CTKRPSGVLI QALPHGFSKV TWIEHVVVND NRVRPHKLYR DLLYGGFGYG ARRWTVTLER TCERLIFSTS
---FESTAPI CTKRPSGVLI QALPHGRSKV TWIEHVEVTD -KVWPHQLYR DLLYGGFGYG ARRWTATLQR MCERLSLYS-
------PMSK CRRRPSGCLI QELPNGYSKV IWVEHVEVDD --RAVHNIYR PVVNSGLAFG AKRWVATLDR QCERLASSMA
QFV---SHSQ SYRFPSGCLI QDMPNGYSKV TWVEHIETEE K-ELVHELYR EIIHRGIAFG ADRWVTTLQR MCERFASLSV
QFI---SQSR SYRFPSGCLI QDMSNGYSKV TWVEHGEFEE Q-EPIHEMFK DIVHKGLAFG AERWIATLQR MCERFTNLLE
------PITR SRRRPSGCLI QELQNGYSKV TWVEHIEVDD --RSVHNMYK PLVNTGLAFG AKRWVATLDR QCERLASSMA
---T--PILR TRRRPSGCLI QELPNGYSKV TWIEHMEVDD --RSVHNMYK PLVQSGLAFG AKRWVATLER QCERLASSMA
-TEKEASLLK CRKLPSGCII EDTSNGHSKV TWVEHLDVSA --STVQPLFR SLVNTGLAFG ARHWVATLQL HCERLVFFMA
------PITR NRRRPSGCLI QELPNGYSKV IWVEHVEVDD --RAVHNIYR PLVNSGLAFG AKRWVATLDR QCERLASAMA
------SVLK CRRRPSGCLI QEMPNGYSKV TWVEHVEVDD --RSVHNIYK LLVNSGLAFG PPRWVGTLDP QCEPLPSVMP
------LLSK CRRRPSGCLI QELPNGYSKV VWVEHIEVDD --RSVQNIYR PLVNSGLAFG AKRWVGTLDR QCERLASSMA
------LMSK CRRRSSGCLI QELPNGYSNV VWVEHTEVDD --RSVHNIYR PLVNSGLAFG AKRWVGTLDR QCERLASSMA
------PIAR SRRRLSGCVI QDLPNGYSKV TWIEHIEVDD --RSVHSLYR PLINSGLAFG AKRWVAILDR QCERLASSMA
---SS-SVIR CRRRPSGCLI QEMPNSYSKV TWVEHVEADD --RAVHHIYR QLVNSGMAFG AKRWIATLQR QCERLASVLA
------PVQK CRRRPSGCLI QEMPNGYSKV TWVEHVEVDD --SSVHNIYK PLVNSGLAFG AKRWVGTLDR QCERLASAMA
------PVLK CRRRPSGCLI QEMPNGYSKV TWVEHVEVDD --RSVHNIYK LLVNSGLAFG ARRWVGTLDR QCERLASVMA
GTASNAGNIR CRRLPSGCVM QDTPNGYCKV TWVEYTEYDE --ASVHQLYR PLIRSGLAFG ARRWLAMLQR QCECLAILMS
------PVIK CRRRPSGCLI QEMPNGYSKV TWVEHVEVDD --RSVHNLYR PLVNSGLAFG AKRWVGTLDR QCERLASAMA
---PPPSLMR CRRRPSGVLI QDTPNGYAKV TCVEHMEYDD --RAVHRMYR DLVNTGMAFG AQRWLATLQR QCERLASLLA
---PPPSLMR CRRRPSGVLI QDTPNGYAKV TCVEHMEYDD --RAVHRMYR ELVNTGMAFG AQRWLATLQR QCERLASLLA
--ESTT-FVT CRRLPSGCVV QDMPNGYSKV IWAEHAEYDE --SQVHQLYR PLLSSGVGFG AQRWVAALQR QCECLAILMS
QFA---SQCR SHRLPSGCLI QDMPNGYSKV TWLERVEIED K-TPIHRLYR DLVHSGSAFG AERWLTTLQR MCEWFACLRV

       650        660        670        680        690        700        710        720
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
STVSTSTNPS PINCN--GRK SMLKLAKRMT DNFCGGVCAS SLQKWSKLN- -VGNVDEDVR IMTRKSVNNP GEPPGIILNA
SSVTSHDN-T SITPG--GRK SMLKLAQRMT FNFCSGISAP SVHNWSKLT- -VGNVDPDVR VMTRKSVDDP GEPPGIVLSA
TNISSGEVGV ITNQE--GRR SMLKLAERMV ISFCAGVSAS TAHTWTTLS- --GTGAEDVR VMTRKSVDDP GRPPGIVLSA
TDFQSVDSGD HITLTNHGKM SMLKIAERIA RTFFAGMTNA TGS--TIFS- --GVEGEDIR VMTMKSVNDP GKPPGVIICA
TNIT--DLG- VIPSVE-ARK NLMKLSQRMV KTFCLNIINS HGQ------- --APTKDTVK IVSRKVCG-- ----GLVPCA
RNIT--DLG- VISSAE-ARR NIMRLSQRLV KTFCVNISTA YGQSWTALS- --ETTKDTVR ITTRKMCE-P GQPTGVVLCA
TNLTEISP-- GLSAK--GAT EIVKLAQRMT LNYYRGITSP SVDKWQKIQ- -VENVAQNMS FMIRKNVNEP GELTGIVLSA
S-----EDHT GLSHA--GTK SILKLAQRMK LNFYSGITAS CIHKWEKLL- -AENVGQDTR ILTRKSLE-- --PSGIVLSA
IQTLPPSDRS EVITTGEARR SVMKLGERMV KNFNEMLTMS GKIDFPQQS- -----KNGVR VSIRMN-IEA GQPPGIVVSA
VPALPNNDNP GVVQTIRGRN SVMHLGERML RNFAWMMKMV NKLDFSPQSE ---TNNSGIR IGVRIN-NEA GQPPGLIVCA
MTDFPPTDYP GVVKTIEGRR SVMSLGERML KNFAWIMKMS DKLDLPQQSG ---ANNSGVR ISVRTN-TEA GQPPGLIVCA
TNIPAGDLCV ITSPE--GRK SMLKLAERMV TSFCTGVGAS TAHAWTTLS- --ATGSDDVR AMTRKSMDVP GRPPGIVLSA
PASSSRDLGG VILSPE-GKR SMMRLAQRMI SNYCLSVSRS NNTRSTVVS- --ELNEVGIR VTAHKS---- PEPNGTVLCA
PATSSLDLGG VIPSPE-GKR SIMRLAHRMV SNFCLSVGTS NNTRSTVVS- --GLDEFGIR VTSHKSR--- HEPNGMVLCA
SNIPACDLSV ITSPE--GRK SMLKLAERMV MSFCTGVGAS TAHAWTTLS- --TTGSDDVR VMTRKSMDDP GRPPGIVLSA
SNIPG-DLSV ITSPE--GRK SMLKLAERMV MSFCSGVGAS TAHAWTTMS- --TTGSDDVR VMTRKSMDDP GRPPGIVLSA
TNVPTKDSLG VTTLA--GRK SVLKMAQRMT QSFYRAIAAS SYHQWTKIT- --TKTGQDMR VSSRKNLHDP GEPTGVIVCA
SNIPAGDVGV ITSPE--GRK SMLKLAERMV MSFCAGVGAS TTHTWTTLS- --GSGADDVR VMTRKSMDDP GRPPGIVLSA
STIPTSDMGV ITSTE--GRK SMLKLAERMV MSFCGGVTAS AAHQWTTLS- --GSGAEDVR VMTRKSVDDP GRPPGIVLNA
INIPSGDLCV ITTAE--GRK SMLKLAERMV MSFCTGVGAS TAHAWTTLS- --ATGSDDVR VMTRKSMDDP GRPPGIVLSA
INIPTGDLCV ITTPE--GRK SILKLAERMV MSFCTGVGAS TAHTWTTLS- --ATGSDDVR VMTRKSMDDP GRPPGIVLSA
INIPAGDLCV ITSPE--GRK SMLKLAERMV MSFCSGVGAS TAHAWTTLS- --ATGSDDVR VMTRKSMDDP GRPPGIVLCA
SNIPARDLGV IPSPE--GRK SILKLAERMV TSFCAGVSAS TAHTWTTLS- --GSGAEDVR VMTRKSIDDP GRPPGIILSA
SNIPNGDLGV ITSVE--GRK SMLKLAERMV ASFCGGVTAS VAHQWTTLS- --GSGAEDVR VMTRKSVDDP GRPPGIVLNA
SNIPTSDIGV ITSSE--GRK SMLKLAERMV VSFCGGVTAS VAHQWTTLS- --GSGAEDVR VMTRKSVDDP GRPPGIVLNA
PDTVSANDSS VITQE--GKR SMLKLARRMT ENFCAGVSAS SAREWSKLDG AAGSIGEDVR VMARKSVDEP GEPPGVVLSA
SNIPNGDLGV ITSIE--GRK SMLKLAERMV ASFCGGVTAS AAHQWTTLS- --GSGAEDVR VMTRKSVDDP GRPPGIVLNA
SNIASRDLGG VPSAS--GRR SMLKLAQRMT NNFCAGVSAS TVHTWTTLS- --GSGDDDVR VMTRKSVDNP GEPHGIVLSA
SNIASRDLGG VPSAS--GRR SMLKLAQRMT NNFCAGVSAS TVHTWTTLS- --GSGDDDVR VMTRKSVDNP GEPHGIVLSA
STVPTRDH-T AITAS--GRR SMLKLAQRMT DNFCAGVCAS TVHKWNKLN- -AGNVDEDVR VMTRKSIDDP GEPPGIVLSA
SSTSTRDLGG VIPSPE-GRR SMMKLAQRMV NNFCTSVGTS NSHRSTTLS- --GSNEVGVR VTVHKSSD-P GQPNGIVLSA

       730        740        750        760        770        780        790        800
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
ATSVWMPVSP RRLFDFLGNE RLRSEWDILS NGGPMKEMAH IAKGHDRSNS VSLLRASA-I NAN------- ---QSSMLIL
ATSVWLPAAP QRLYDFLRNE RMRCEWDILS NGGPMQEMAH ITKGQDQG-- VSLLRSNA-M NAN------- ---QSSMLIL
```



```
ATSFWIPVPP KRVFDFLRDE NSRNEWDILS NGGVVQEMAH IANGRDTGNC VSLLRVNS-A NSS------- ---QSNMLIL
ATSFWLPAPP NTVFDFLREA THRHNWDVLC NGEMMHKIAE ITNGIDKRNC ASLLRHG--- HTS------- ---KSKMMIV
VSVTLLPYSH QQVFDLLRDN QRLSQLEILF MGSSFQEVAH IANGSHLGNS ISLLRINVES NSS------- ---HNVELML
VSTTWLPFSH HQVFDLIRDQ HHQSLLEVLF NGNSPHEVAH IANGSHPGNC ISLLRINVAS NSW------- ---HNVELML
STSVWLPVNQ HTLFAFISHL SFRHEWDILT NDTTMEETIR IQKAKRHGNI ISLLKIVN-- ---------- ----NGMLVL
ATSLWLPVTQ QRLFEFLCDG KCRNQWDILS NGASMENTLL VPKGQQEGSC VSLLRAAG-N DQN------- ---ESSMLIL
SSSLAIPLTP LQVFAFLQNL DTRQQWDILS YGTVVNEIAR IVTGSSETNC VTILRVHPTH EENNDKMVVQ DSCKDDMLML
GSSLSLPLPP VQVYDFLKNL EVRHQWDVLC HGNPATEAAR FVTGSNPRNT VSFLEPSIRD I--------- ---NTKLMIL
GSSLSLPLPP LQVYDFLRNL EVRHQWDVHC QGNPVTEAAR FVTGPDQKNN VTFLQPSSVG ---------- ---EYKLMIL
ATSFWIQVPP KRVFDFLRDE NSRSEWDILS NGGLVQEMAH IANGRDPGNC VSLLRVNS-A NSS------- ---QSNMLIL
ATTFWLPNSP QNVFNFLKDE RTRPQWDVLS NGNAVQEVAH ISNGSHPGNC ISVLRGSNAT HS-------- ----NNMLIL
ATSFWLPISP QNVFNFLKDE RTRPQWDVLS NGNSVQEVAH ITNGSNPGNC ISVLRGFNAS SSQ------- ----NNMLIL
ATSFWIPVAP KRVFDFLRDE NSRSEWDILS NGGLVQEMAH IANGRDPGNS VSLLRVNS-G NSG------- ---QSNMLIL
ATSFWIPVAP KRVFDFLRDE NSRKEWDILS NGGMVQEMAH IANGHEPGNC VSLLRVNS-G NSS------- ---QSNMLIL
SSSLWLPVSP ALLFDFFRDE ARRHEWDALS NGAHVQSIAN LSKGQDRGNS VAIQTVKS-- --R------- ---EKSIWVL
ATSFWIPVPP KRVFDFLRAE NSRSEWDILS NGGLVQEMAH IANGRDPGNC VSLLRVNS-A NSS------- ---QSNMLIL
ATSFWLPVPP KRVFDFLRDE SSRSEWDILS NGGVVQEMAH IANGRDHGNC VSLLRVNS-T NSN------- ---QSNMLIL
ATSFWIPVQS KRMFDFLRDE NHRSEWDILS NGGEVQEMAH IANGRDPGNC VSLLRVNS-A NSS------- ---QSNMLIL
ATSFWIPVPS KKVFDFLKDE NHRSEWDILS NGGQVQEMAH IANGRDPGNC VSLLRVNS-T NSS------- ---QSNMLIL
ATSFWLPVPP KRVFQFLSDE NHRSEWDILS NGGQVEEMAH IANGRDPGNC VSLLRVIS-A NSS------- ---QSNMLTL
ATSLWLPVPP KKVFDFLRDE NSRNEWDILS NGGLVQEVDH IANGQDPGNC VSLLRVNT-V NSN------- ---QSNMLIL
ATSFWLPVPP AAVFDFLRDE TSRSEWDILS NGGAVQEMAH IANGRDHGNS VSLLRVNS-A NSN------- ---QSNMLIL
ATSFWLPVPP KRVFDFLRDE SSRSEWDILS NGGIVQEMAH IANGRDQGNC VSLLRVNS-S NSN------- ---QSNMLIL
PTSVWVPVAP EKLFNFLRDE QLRAEWDILS NGGPMQEMAN IAKGQEHGNS VSLLRASA-M SAN------- ---QSSMLIL
ATSFWLPVPP KRVFDFLRDE TSRSEWDILS NGGAVQEMAH IANGRDHGNC VSLLRVNS-A NSN------- ---QSNMLIL
ATSMWLPVSP ARVFQFLRDE RLRSEWDILS NGGMVTEMAH IAKGQDPGNS VSLLRVNA-M NSS------- ---QSNMLIL
ATSMWLPVSP ARVFQFLRDE RLRSEWDILS NGGIVTEMAH IAKGQDPGNS VSLLRVNA-M NSN------- ---QSNMLIL
ATSVWLPVSP QRLFDFLRNE RLRSEWDILS NGGPMQEMAH IAKGQDHGNC VSLLRASA-M NAN------- ---QSSMLIL
ATTFWLPVSP QNVFNFFKDE RTRPQWDVLS NGNAVQEVAH IANGSHPGNC ISVLRAFNTS HN-------- -----NMLIL

       810        820        830        840        850        860        870        880
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
QETSIDAA-G AVVVYAPVDI PAMQAVMNGG -DSAYVALLP SGFAILPNGQ AGTQRCAAEE RN-------- ----SIGNGG
QETCIDAS-G ALVVYAPVDI PAMHVVMNGG -DSSYVALLP SGFAVLPDGG IDGG------ ---------- ----GSGDGD
QESCTDPT-A SFVIYAPVDI VAMNIVLNGG -DPDYVALLP SGFAILPDGN ANSGAP---- ---------- ----------
QETSTDPT-A SFVLYAPVDM TSMDITLHGG GDPDFVVILP SGFAIFPDGT GKPGGKE--- ---------- ----------
QETCTDNS-G SLLVYSTVDP VAVQLAMNGE -DPSEIPLLP VGFSVVPVNP SDGVEGSSVS SP-------- ----------
QESCIDNS-G SLIVYSTVDV DSIQQAMNGE -DSSNIPILP LGFSIVPVNP PEGISVNSHS PP-------- ----------
QEIWNDAS-G AMVVYAPVET NSIELVKRGE -NSDSVKFLP SGFSIVPDGV N--------- ---------- ------GSYH
QETWNDVS-G ALVVYAPVDI PSMNTVMSGG -DSAYVALLP SGFSILPDGS SSSS-----D QF-------- ----DTDGGL
QDCYMDAL-G GMIVYAPMDM ATMHFAVSGE VDPSHIPILP SGFVISSDGR RSTVEDGG-- ---------- ----------
QDSFKDAL-G GMVAYAPMDL NTACAAISGD IDPTTIPILP SGFMISRDGR PSEGEAEGGS Y--------- ----------
QDGFIDAL-G GMVVYAPMNL NTAYSAISGQ VDPSTIPILP SGFIISRDSH PSSSEVDGGS M--------- ----------
QESCTDAK-G SYVVYAPVDI VAMNIVLSGG -DPDYVALLP SGFAILPDGP GVNGGG---- ---------- ------ILEI
QESSTDSS-G AFVVYSPVDL AALNIAMSGE -DPSYIPLLS SGFTISPDGN GSNSEQGGAS TSS------- ---------G
QESCIDSSSA ALVIYTPVDL PALNIAMSGQ -DTSYIPILP SGFAISPDG- ---SSKGG-- ---------- ----------
QESCTDAS-G SYVIYAPVDI IAMNVVLSGG -DPDYVALLP SGFAILPDGS ARGGGGSANA SAGAGVEGGE GNNLEVVTTT
QESCTDAS-G SYVIYAPVDI VAMNVVLSGG -DPDYVALLP SGFAILPDGS VGGGDG---- ---------Q HQEMVSTTSS
QDSSTNSY-E SVVVYAPVDI NTTQLVLAGH -DPSNIQILP SGFSIIPDGV ESRPLVITST QD-------- ----------
QESCTDPT-G SYVIYAPVDI VAMNVVLSGG -DPDYVALLP SGFAILPDGA VLHGGG---- ---------- ------ILDV
QESCTDAS-G SYVIYAPVDV VAMNVVLNGG -DPDYVALLP SGFAILPDGP AGSNMQGDG- ---------- --------GV
QESCTDST-G SYVIYAPVDI SAMNIVLSGG -DPDYVALLP SGFAILPDGP GYGSAG---- ---------- ------ILDV
QESCTDST-G SYVIYAPVDI SAMNIVLSGG -DPDYVALLP SGFAILPDGP GYGPAG---- ---------- ------ILDV
QESCTDST-G SYVIYAPVDI AAMNIVLSGG -DPDYVALLP SGFAILPDGP GFSPGI---- ---------- ------ILDV
QESCTDAS-G SFVIYAPVDI VAMNVVLSGG -DPDYVALLP SGFAILPDSP KCMAVTNSG- ---------- ------INDL
QESCTDAS-G SYVVYAPVDI VAMNVVLNGG -DPDYVALLP SGFAILPDGP S-GNAQAAVG E--------- -------NGS
QESCTDAS-G SYVIYAPVDV VAMNVVLNGG -DPDYVALLP SGFAILPDGP AHDGGDGDGG ---------- -------VGV
QETCTDAS-G SMVVYAPVDI PAMQLVMNGG -DSTYVALLP SGFAILPDGP ---------- ---------- -----SSVGA
QESCTDAS-G SYVVYAPVDV VAMNVVLNGG -DPDYVALLP SGFAILPDGP PPAGAAPSHG E--------- --------GL
QESCTDVS-G SLVIYAPVDI PAMNLVLQGG -DPAYVALLP SGFAILPDGP GGDRGALGNE QG-------- ----GQLTEI
QESCTDVS-G SLVIYAPVDI PAMNLVLQGG -DPAYVALLP SGFAILPDGP GGERGSLGVD QG-------- ----SQLTES
QETCIDAA-G SLVVYAPVDI PAMHVVMNGG -DSAYVALLP SGFAIVPDGP RSHGPISNGH VN------~- -----GNTGG
```



```
QESCIDSS-G SLVVYCPVDL PAINVAMSGE -DPSYIPLLP SGFTITPDGH LEQGDGASTS SST------- ---------G

       890       900       910       920       930       940       950
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ..
CMEEGG-SLL TVAFQILVN- --SLPTAKLT VESVETVNNL ISCTVQKIKA ALHCDST--~ ~~~------- -- HDG1
QRPVGGGSLL TVAFQILVN- --NLPTAKLT VESVETVNNL ISCTVQKIRA ALQCES---- ---------- -- ANL2
GGD--GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNNL IACTVERIKA SMSCETA--- ---------- -- HDG2
-----GGSLL TISFQMLVE- --SGPEARLS VSSVATTENL IRTTVRRIKD LFPCQTA--- ---------- -- HDG3
---------S CLLTVAIQVL GS---NVTTE RLDLSTVSVI NHRICATVNR ITSALVN--- -------DVG N- HDG4
------SCLL TVGIQVLAS- --NVPTAKPN LSTVTTINNH LCATVNQITS ALSNTITPVI ASSADVSNQE V- HDG5
RGNTGGGCLL TFGLQILVG- --INPTAALI QGTVKSVETL MAHTIVKIKS ALDLQT---- ---------- -- FWA
VNQESKGCLL TVGFQILVN- --SLPTAKLN VESVETVNNL IACTIHKIRA ALRIPA---- ---------- -- HDG7
-------TLL TVAFQILVSG K-ANRSREVN EKSVDTVSAL ISSTIQRIKG LLNCPEC--- ---------- -- HDG8
-------TLL TVAFQILVSG P-SYSPDTNL EVSATTVNTL ISSTVQRIKA MLKCE----- ---------- -- HDG9
-------TLL TLAFQIFVTG P-SYYTDLNL KDSATTVNTL VSSAVQRIKA MLNCE----- ---------- -- HDG10
GS---GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IKCTVERIKA AVKCNNA--- ---------- -- GbML1
RA-SASGSLI TVGFQIMVS- --NLPTAKLN MESVETVNNL IGTTVHQIKT ALSGPTASTT A--------- -- HDG11
------GSLI TVGFQIMVS- --GLQPAKLN MESMETVNNL INTTVHQIKT TLNCPSTA-- ---------- -- HDG12
GSC--GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IKCTVERIKA ALACDGA--- ---------- -- ATML1
GSC--GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IKCTVERIKA AVSCDVGGGA ---------- -- PDF2
DRNSQGGSLL TLALQTLIN- --PSPAAKLN MESVESVTNL VSVTLHNIKR SLQIEDC--- ---------- -- GL2
GS---GGSLL TVAFQILVD- --SAPTAKLS LGSVATVNSL IKCTVERIKA AVSCENT--- ---------- -- VvHD1
GS---GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IACTVERIKA AVSGES-NPQ Q--------- -- SbHD1
GS---GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IKCTVERIKA AVMCDNA--- ---------- -- PtHD1
GS---GGSLL TVAFQILVD- --SVPSVKLS LGSVATVNSL IKCTVERIKA AVMCDNP--- ---------- -- PtHD2
GS---GGALV TVAFQILVD- --SIPTAKLS LGSVATVNNL IKCTVERIKA AVTCETA--- ---------- -- RcHD1
GT---GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL ISCTVDRIKA AVMRENP--- ---------- -- PaHD1
GSG--GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IACTVERIKA AVCRDS-NPQ ---------- -- ROC1
GS---GGSLL TVAFQILVD- --SVPTAKLS LGSVATVNSL IACTVERIKA AVSGES-NPQ ---------- -- ROC2
EHKTGG-SLL TVAFQILVN- --SQPTAKLT VESVETVNNL IFCTIKKIKT ALQCDA---- ---------- -- ZmOCL1
DAGG-GGSLL TVAFQILVD- --SVPTGKLS LGSVATVNSL IACTVERIKA AVCAEAGNPQ ---------- -- ZmOCL5
GRG--TGSLL TVAFQILVS- --SIPSARLS LESVATVNNL ISCTVQRIRS ALLVEDA--- ---------- -- PhHDZ41
SRG--TGSLL TVAFQILVS- --SIPSARLS LESVATVNNL ISCTVQRIKS ALLVEDA--- ---------- -- PhHDZ43
GSSSVGGSPL TVAFQILVN- --SSPTAKLT VESVETVNNL ISCTVQKIKA ALQCES---- ---------- -- GhHOX1
HGRSSGGSLI TVAFQILVS- --SLPSAKLN LDSVTIVNNL IANTVQQIKA ALNCPSS--- ---------- -- GhHOX3
```

Supplementary Figure 7