A

```
1    GGC ACG AGG CGA AGT TCG ATT TTT TCG ATT CGA GGG AAA ACT TCT GAA ATT AGT
57   AGA CAT AAA ATG ACG GAA GAA GCG ACG AAC GGC GAA ATC ACG AAC ACT GAT CAG
                 M  T  E  E  A  T  N  G  E  I  T  N  T  D  Q   15

111  CGA CAG CCC AAG AAC ATG GAT TTA AGC AGC GCT ATA GCA ACT GCC GTG AAT CAT
      R  Q  P  K  N  M  D  L  S  S  A  I  A  T  A  V  N  H   33

165  GTT TTG GAC GGA TAC GAC TGG TCT TTG ATT CCG CTG CCT GTG AGA GTC AAC GGT
      V  L  D  G  Y  D  W  S  L  I  P  L  P  V  R  V  N  G   51

219  GGG CAC AAA CAT AAA CCA CAC GTT AAA AGA CCC ATG AAC GCG TTT ATG GTT TGG
      G  H  K  H  K  P  H  V  K  R  P  M  N  A  F  M  V  W   69

273  GCA CAA GCT GCT AGG AGG AAA CTA GCC GAT CAA TAT CCT CAT CTG CAC AAT GCA
      A  Q  A  A  R  R  K  L  A  D  Q  Y  P  H  L  H  N  A   87

327  GAA CTT AGC AAG ACG CTG GGA AAA CTG TGG AAA ATG CTT AAA GAT GCG GAG AAG
      E  L  S  K  T  L  G  K  L  W  K  M  L  K  D  A  E  K  105

381  AAA CCT TTC ATC GAA GAA GCC GAA AGG CTC AGG CTG AAG CAC AAG AGA GAG CAC
      K  P  F  I  E  E  A  E  R  L  R  L  K  H  K  R  E  H  123

435  CCA GAT TAC AAA TAC CAA CCG AGG AGA AAG AAA CAG AAA GGA AAC GGC GGA CCA
      P  D  Y  K  Y  Q  P  R  R  K  K  Q  K  G  N  G  G  P  141

489  GAT CAA CCA GAG GCA ACT ATC TCG GCT GAT GAT CTT TTG AAA GTA CTG AAA GGG
      D  Q  P  E  A  T  I  S  A  D  D  L  L  K  V  L  K  G  159

543  GAT CCA CAG GTG GTT GGA AAA TCG ATG AGC CAG GAT AGC AGC TGC GCC AGC CCT
      D  P  Q  V  V  G  K  S  M  S  Q  D  S  S  C  A  S  P  177

597  GAC AGC AGT TTG AGC GAC GAG GAA AGC AGC CCT GGG TCA AGC AGC CCG GCT ACG
      D  S  S  L  S  D  E  E  S  S  P  G  S  S  S  P  A  T  195

651  TCG CCT CCA ACA ACT CCA ACA GCT ACG ATC AAA GTT GAA GGG GGA TCA AAA TGC
      S  P  P  T  T  P  T  A  T  I  K  V  E  G  G  S  K  C  213

705  TCT GAG CCG AGT GTT CCG GAT CCG GAC ATT CAA GCG CAG TCG TCT CCG ACA TTT
      S  E  P  S  V  P  D  P  D  I  Q  A  Q  S  S  P  T  F  231

759  CCA CTC AAG AAA TCA GAA AAT AGC ACT GTT CCT TCT ACG ACA AAC AAT AGC GGC
      P  L  K  K  S  E  N  S  T  V  P  S  T  T  N  N  S  G  249

813  AAT AGC GCG ATC GAT TTC CAC GTT GAA ATG GGC GAC CTA ATG GTT GAT ACG ACA
      N  S  A  I  D  F  H  V  E  M  G  D  L  M  V  D  T  T  267

867  GAA TTT GAC CAG TAC CTT CAC TCC TAC ACA CAG CCC TTA CAG ATG CCC TCA TCG
      E  F  D  Q  Y  L  H  S  Y  T  Q  P  L  Q  M  P  S  S  285

921  CCG CCC TCT GTC ACT TAC ACC TCA CAC GGC GGC GTT TAC ACA CCA ACT ACT CAC
      P  P  S  V  T  Y  T  S  H  G  G  V  Y  T  P  T  T  H  303

975  GGC GTT TCA AGG TTT GCG CCT AAC AAT AAC CCT GTG ACA TCG AGT TAT ACG GAG
      G  V  S  R  F  A  P  N  N  N  P  V  T  S  S  Y  T  E  321

1029 TTT ATG GAG CAA CTT CAA AAA CTG CCA TCG TCC TAC CCA CCG AAT CAG GTT GCC
      F  M  E  Q  L  Q  K  L  P  S  S  Y  P  P  N  Q  V  A  339

1083 CCA TCA GCC TTG CAC CAA AGA AAC CAA TCA GAA AAT GGT TCT TTT CCA TTT TCA
      P  S  A  L  H  Q  R  N  Q  S  E  N  G  S  F  P  F  S  357

1137 TTC AAC GAG CCT GCC GAC GTA GCA AAC GGT AGC AGT CCT CGA TCC TTT GGA ATC
      F  N  E  P  A  D  V  A  N  G  S  S  P  R  S  F  G  I  375

1191 TCC AGT GTG CCT TCC CTC GCG ATT CCC GCA TCT TCG CCA GCC AAC GGA ACA CAG
      S  S  V  P  S  L  A  I  P  A  S  S  P  A  N  G  T  Q  393

1245 AAC CCG CAA AGC TCC CCC TCA AGA CTT CAC ACT CTC ATG TGG AAG TAA GCT TTT
      N  P  Q  S  S  P  S  R  L  H  T  L  M  W  K  *   408

1299 GCA ATA ATC TCG AAC ATT TCG GAG AAT TTT GCT CCT ATG GGG CAC GAA TTT TAA
1353 CTT GTT GCT TTC AGC AGC TTT TTA AGT TGC TGG CTA AAA CAA TAC AGA ACT TTA
1407 TCT TGT TTC TGA TTT TTA CAG AAA GTT GAA GTT CGA CAC GCC ATA GTT TAT AAC
1461 AGC GAC GTT TTC CAA AAA TTC CAT GTT TTG GTG CGC CGC TCG AGT CAA ATT TCA
1515 TGC TGA CGT TTA TTG TAC ACA TCA CGA AAG AGT AAA ACA CTC ACT CAC TTG ATA
1569 TTC AGA TCT TTT TAT TTT TTG AGT CGT TGT GAG TAA CGT TTG GAA ATT TTT CGC
1623 CAT GTT GAG GTT TAG AAA ATT GTT GGC AAA TGT TGC TTT TTG CGA GAG AGT TAA
1677 GTT AGC AAG CTG ATG CTC ATT AGT TAC TGC TTA GCT TAT GTA ATT TCT AAT TCA
1731 TGA GCT TGC TTC CCA AAA GAA AAT GTG ATT TAT CAC CCT TAG ATC CCC ATG TAC
1785 ATA ATT AAT GTG AAT AGG GAT GAT AGC TTT TGA AAA ATC GTA GCT TTA GAA GTA
1839 AAA AAT TTT TAC TGA TTC AAT TTT GCC AAA CCT CGA ATT TCT GTG TCC TGG CGT
1893 TTA TAT TTT TTG GCA AAA TAA ATA AAA CTA CTT TTG TGC GTG GGA AAA AAA AAA
1947 AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA 3'
```

B

```
AmSoxE1    1 ----------------------------------------------------------------------------------------------------
NvSoxE1    1 ----------------------------------------------------------------------------------------------------
DmSoxE     1 ----------------------------------------------------------------------------------------------------
CiSoxE     1 MNNDMTLLSGHKQTMFNQQSGSQDKWSSVSPNSPDKDIMEHLRKGTDFSCFASSNDVEFSSGVEGKENGSKGKSGGPFSPEDLFKPNQDENSTSTSEIFS
HsSox8     1 ----------------------------------------------------------------------------------------------------
HsSox9     1 ----------------------------------------------------------------------------------------------------
FrSox9     1 ----------------------------------------------------------------------------------------------------
DrSox9     1 ----------------------------------------------------------------------------------------------------
MmSox10    1 ----------------------------------------------------------------------------------------------------

AmSoxE1    1 ----------------------------------------------MTEEATNGEITNTDQRQPKNMDLSSAMATVNHVLDGYDWSLIPPPVRVNG-
NvSoxE1    1 ----------------------------------------MDKKVTEQQVQAVLGLDTDGSQVRNHQLSNAMASAVNHVLDGYDWSLIPLPVRVNG-
DmSoxE     1 ----------------------------------MSDSSSSNCSKDRAKPVETLVLANYALKAEQKKAQGGGGRKEDEHTTAVMKVLDGYDWLVQASEAAPH-
CiSoxE   101 MSSPESLSNFCNVKSAVAVARAAATLLESKSSSEEENYDESLMNTGSSARSASPGTNDDLSDRDSNFEKDDMSKDMKDAVSQVLKGYDWLVPMPVRMNG-
HsSox8     1 --MLDMSEARSQPPCSPSGTASSMSHVEDSDSDAPPSPAGSEGLGRAG-----VAVGGARGDPAEAADERFPACMRDAVSQVLKGYDWSLVPMPVRGGG-G
HsSox9     1 MNLLDPFMKMTDEQEKGLSGAPSPTMSEDSAGSPCPSGSGSDTENTRPQENTFPKGEPDLK--KESEEDKFPVCMREAVSQVLKGYDWELVPMPVRVNG-
FrSox9     1 MNLLDPYLKMTEEQEKCHSDAPSPSMSEDSAGSPCPSGSGSDTENTRPSDNHLLLGPDYK--KENEEEKFPVCMREAVSQVLKGYDWELVPMPVRVNG-
DrSox9     1 MNLLDPYLKMTDEQEKCLSDAPSPSMSEDSAGSPCPSASGSDTENTRPAENSLLAADGTLGDFKKDEEDKFPVCMREAVSQVLKGYDWELVPMPVRVNG-
MmSox10    1 ---MAEEQDLSEVELSPVGSEEPRCLSPGSAPSLGPDGGGGGGSGLRASPGPGELGKVKKEQQDGEADDDKFPVCMREAVSQVLKGYDWELVPMPVRVNG-

                                   ***          *          *          *                                   i
AmSoxE1   52 -GHGHGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWKMLKDASKKPFIEEAERLRLKHKREHPDYKYQPRRKKQKGNGGPDQPAATISA
NvSoxE1   57 -IKTQKPHVKRPMNAFMVWAQAIRRKLADQYPHLHNAELSKTLGKLWKMLNDSEKKPFIEEAERLRIKHKREHPDYKYQPRKKKQKGNGGDAGDATISA
DmSoxE    71 --DRKGEHEKRPMNAFMVWSEQYPHLDNSELSHGLGKLWSNLKSDKKPFVDHAERLRMTHKQEHPDYKYQPREKAREVLPSQQSGDGGSSPS
CiSoxE   200 -SQKGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLSETEKKPFVDEAERLRIKHKKDHPDYKYQPRRKSSKTASGVGEGTGQAQ
HsSox8    94 GALAGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLSESSKKPFVEEAERLRVQHKKDHPDYKYQPRRKSAKAGHSDSGSGAELG
HsSox9    98 -SGAGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESEKKPFVEEAERLRVQHKKDHPDYKYQPRRKSVKRGQALAAPTTEQT
FrSox9    97 -SNAGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNEVEKKPFVEEAERLRVQHKKDHPDYKYQPRRKSVKNGQNDPEVG-EQT
DrSox9   100 -SSAGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNEVEKKPFVEEAERLRVQHKKDHPDYKYQPRRKSVKNGQSESDGSEQT
MmSox10   97 -ASKGKPHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESDMKPFIEEAERLRMQHKKDHPDYKYQPRRKNGKAAQGEAECPGGPA

                                                                HMG domain
AmSoxE1  151 DDLLKVFKGDPQVVGKSMQDSSCASPDSSL------------------SDGESSFGSSSPATSPFPTPTAGIKVEGGSKCSPSVPDPDLQAGSSPTF
NvSoxE1  156 DDLLKVFKGDSKLVPNNGDASASCASPESVS-----------------DGEVSG-SEECSVPSFPTPTAVPVKNEDVKNDALASAQPGFPSCE----
DmSoxE   169 PEMTLSATMGSSGKPRSSNSNGQRRAGKGNAAADLGSCASTISHANVGSNSSIVPSNEAFKSLNSACAASLMEQSLIEYGLDFPCSTASSMSSLSPPAT
CiSoxE   299 NQSLKQQSGKVRKQDSQSMDECQGVQQALVANPISGKQQSKQQLSSHHSPQSVCHSHSNQSPPHQGSIHNIYEVMHREQRGYHDSPDATVPCSPPTAINM
HsSox8   194 PHPGGGAVYKAEAGLGDGHHHG-------------------------------------DHTGGHGPPTPTHPHTHLQQAGAKPELKLEGRRPVD----
HsSox9   197 HISPNAIFKALQADSPHSHGG-------------------------------------MGHVHSHGEHGQSGQPTPTPTPTPTPVQPGKADLKEGRPLPE-G----
FrSox9   195 HISPNAIFKALQQADSPAS---------------------------------------LGHVHSHGDHGQSGQPTPTPTPTPTDLVSSKADLKREGRPMQEGT----
DrSox9   199 HISPNAIFKALQQADSPAS---------------------------------------MGHVHSHSEHGQSGQPTPTPTPTPKTDTQPGKADLKREARPLQENT----
MmSox10  196 EQGGAAAIQAHYKSAHLDHRHPEEGS-------------------------------PMSDGNHEHPGQSHGPTPTPTPTPKTPLQSGKADPKRDGRSLGE------

                                                                    ii
AmSoxE1  232 PLKKSENSTVPSTTNNSCN----------------------SAIDFHVEMGDLMVDTTSFDQYLH-----------------------------SYTQP
NvSoxE1  231 ---KKDDSNSHAIDFDVSD-------------------------LTTDLMAMG---DVDSTSFDQYLP-----------------------------TSQA
DmSoxE   269 PYNVAPSNAKASAANNFSLLLR-------------------QLSEPVANAGDGYGVLLEAGREYVAIGEVNYQGQSAGVQSGAEGGGAGQEMDFLENINGY
CiSoxE   399 DESKKPHSTLHSRNSSFGSKQQRGSSIDLTSECSSHSMHGVMDTNTQDIMAKPGFHVTTPSQVMPGACNPAVVRQHEEAFGYSCMGEPHAKQKRCNFTDT
HsSox8   254 ----SGRQNIDFSNVDISE---------------------LSSEVMGTMD--AFDVHSFDQYLPLGGPAPPEPG-----------------QAGGA
HsSox9   263 ----GRQPPIDFSDVDIGE---------------------LSSHVISNIE--TFDVMSFDQYLPPNGHPGVPATHG-QVTYTGSYGISSTAATPASAGH
FrSox9   261 ----SRQLNIDFGAVDIGE---------------------LSSEVISNMG--SFDVDSFDQYLPPHSHAGVSGAPQAGYTGSYGISSSSVSQAAGVGAQ
DrSox9   265 ----GRPLSINFQDVDIGE---------------------LSSDVIE----TFDVNSFDQYLPPNGHQNAP-----------------YAGGYA
MmSox10  266 ----GGKPHIDFGNVDIGE---------------------ISHEVMSNME--TFDVTSFDQYLPPNGHPGHVGSYS----------------AAGHGLG

                                                                iii
AmSoxE1  280 LQWPSSPPSVTYTNHG-GVYTPTTHG-------------------------------------------------------------VSKFAPNNNPVTPSWK
NvSoxE1  273 LLPSTLTKAINTTQINTQSLSNSRFT-------------------------------------------------------------TSQAVQSFPLPSPK
DmSoxE   351 GGYTGSRVSYPAYSYPANGGHFATEEQQQQ-------------------------------------------------QALQASEALNYKPAAAIDPKE
CiSoxE   499 SQNMPSPVDCSVQQNPMFSPNKQGTPQYRPSSWVEGYETGVMTEASVSPNASMDQQFSQPENFSYDSVTSACSPLQSSSVTSQSGEFNTYSMNQSSPHSC
HsSox8   307 YFHAGASPVWAHKSAPSASASPTETG-----------------------------------------------------------PPRPHKTEQPSPGPH
HsSox9   334 VWTSKQQAPPPPQQPPQAPASQAPPQPQAPPQPQAPTLTTLSSEPQGSRTHITEQLSPHH
FrSox9   333 AWMSKQQQQQQQQQQHSLTALSGGGE-----------------------------------------------------------QSQQGQQRPAQIKTEQLSPGMK
DrSox9   313 AWMTKFQNGSPQSSQLTPLNP-------------------------------------------------------------AEPDQPRTTHIKTEQLSPGMK
MmSox10  322 SAIAVASGHSAWIRPPGVALPTVSP-----------------------------------------------------------PGVDAKAQVKTETTGPQGPPH

AmSoxE1  320 TPFMEQLQKLPS--SSYPPQVAQS-----------------------------------------------------------------------
NvSoxE1  314 RPFMVQLQKLPSEGSFPPRVAQS-----------------------------------------------------------------------
DmSoxE   403 IDQYFMDQMLPMTQHHHPHTHPLHHP-----------------------------------------------------------------------
CiSoxE   599 AELPVKKEQISPIQQHFFPHVPPATRFQCSELHKAADVAPQPYQAYFDANIQRAPMEDTMVDQRADALYHHNAYDHSGPFPNMHGRFDACSAKQQEMNLD
HsSox8   349 GEQP---RGSPDYGSGQSSATPA-------------------------------------------------------------------------
HsSox9   408 SQQQQHSPQIAYSPFNLPHYSHS-------------------------------------------------------------------------
FrSox9   381 SQQQ-GSPQHVTYGSFNLQHYSTS-------------------------------------------------------------------------
DrSox9   355 NQQQ-GSPQHISYGSFNVQHLQHYS-------------------------------------------------------------------------
MmSox10  370 TPQP--STSQIAYPLSLPHYGSA---------------------------------------------------------------------------

AmSoxE1  342 ----------------------------ALHQRNQPNGSFPFSFNEPADVANGSPRSFGISSVFSLAIPASSPNGPQNQSSPRLHTLMQK---
NvSoxE1  338 ----------------------------ATMQQRNQPDSNSFFPFSESEVNVCSG------LAATQPAFLSPSTSGQLSSSNSGRTHTLIQK---
DmSoxE   430 -----------------------LHHSPPLNSSASLSSACSASSQQPVAEYYEHLGYSPAASSASQNPNFGPQQPYANGANSMQPTLGDPAPQQELQSQQQEQ
CiSoxE   699 SPTRCSPDPPRPFEHAAYDSNITNPLQRRFSLPLIPQMQQQGGANPAYRHFGHSHNQSLAHHYPKPNERQQLYSAPEEFSTPHNQHYNMAQQQQMNPLPS
HsSox8   371 ----------------------APAGPPAGSMQDYGS---LQASSYYGAYPGYAPGLYTYPCFHSH-RPPYAMPLLNGLAIPPAHSP---TSHMDQP-
HsSox9   432 ----------------------YPPITRSQYDYTD--HQNSSSYYSHAAGGGTGLYTFTYMNSAQRPMYTPIMDTSGYGSIPQT-HSPQHWEQP-
FrSox9   404 -----------------------SYPSMTRAQYDYGD-HQGGANSYYSHAAGGGGLYTFSYMNSQRPMYTPIADNAGYGSVPQT-HSPQHWEQQP
DrSox9   379 -----------------------TSFPSITAQYDYGDSHQGGASSYYTHAGGGSGLYTFSYMSSQRPMYTPIADSPGYGSIPQNHSPQHWDQQP
MmSox10  392 -----------------------FPSISRPPDYGD--HQPSGPYYGHAG-QAGGLYTAFSYMGPSQRPLYTAISDPG--GSGPGS-HSPTHWEQP-

AmSoxE1      ------------------
NvSoxE1      ------------------
DmSoxE   511 QHQNPSQHHLWGTYTYVNP
CiSoxE   799 TSAEVFSPPH---------
HsSox8   439 VYTTLTRP----------
HsSox9   502 VYTQLTRP----------
FrSox9   477 IYTQLSRP----------
DrSox9   455 VYTQLSRP----------
MmSox10  459 VYTTLSRP----------
```