



Input formats to BLAST

Multiple sequence alignment

The `-in_msa psiblast` option provides a way to jump start psiblast from a master-slave multiple sequence alignment computed outside psiblast. The multiple sequence alignment must contain the query sequence as one of its sequences, but it need not be the first sequence. The multiple sequence alignment must be specified in a format that is derived from Clustal, but without some headers and trailers (see example below).

The rules are also described by the following words. Suppose the multiple sequence alignment has N sequences. It may be presented in one or more blocks, where each block presents a range of columns from the multiple sequence alignment. E.g., the first block might have columns 1-60, the second block might have columns 61-95, the third block might have columns 96-128. Each block should have N rows, one row per sequence. The sequences should be in the same order in every block. Blocks are separated by one or more black lines. Within a block there are no blank lines, and each line consists of one sequence identifier followed by some whitespace followed by characters (and gaps) for that sequence in the multiple sequence alignment. In each column, all letters must be in upper case, or all letters must be in lower case.

```
# Example multiple sequence alignment file
align1
-----
26SPS9_Hs      IHAAEEKDWKTAYSIFYEAFEGYdsidspkaitslkymllckimlntpedvqalvsgkla
F57B9_Ce      LHAADEKDFKTAFSYFYEAFEGYdsvdekvsaltalkymllckvmlldlpdevnslsaki
YDL097c_Sc    ILHCEDKDYKTAFSYFFESFESYhnlthnsyekacqvlkymllskimlnliddvkninl
YMJ5_Ce       LYSAEERDYKTSFSYFYEAFEGFasigdkinatsalkymilckimlneteqlagllaake
FUS6_ARATH    KNYIRTRDYCTTTKHI IHMCMNailvsiemgqfthvtsyvnkaeqnpetlepvmnaklrc
COS41.8_Ci    SLDYKCLKTYLTIARLYLEDEDPVqaemyinrasllqnetadeqlqihykvcyarvldyrr
644879        KCYSRARDYCTSAKHVINMCLNVikvsvylqnvshvlsyvaskaestpeiaeqrgerdsqt
YPR108w_Sc    IHCLAVRNFKEAAKLLVDSLATFtsieltsyesiatyasvtglftlertdlkskvidspe
eif-3p110_Hs SKAMKMGDWKTCHSFI INEKMNGkw-----
T23D8.4_Ce    SKAMLNGDWKKCQDYIVNDKMNQkw-----
YD95_Sp       IYLMSIRNFSGAADLLDCMSTFsstellppyydvrvyavisgaisldrdrvdktkivdspe
KIAA0107_Hs   LYCVAIRDFKQAAELFLDTVSTFtsyelmdyktfvtytvvyvmialerpdldrekvikgae
F49C12.8_Hs   LYRMSVRDFAGAADLFLEAVPTFgsyelmtyenlilytvittttfaldrpdlrtkvircne
Int-6_Mm      KFQYECGNYSGAAEYLYFFRVLVpatdrnalsslwglaseilmqnvdaamedltrlket

26SPS9_Hs      lryagrqtealkcvaqasknrsladfekaltdy-----
F57B9_Ce      alkyngsdldamkaiaaaaqkrslkdfqvafgsf-----
YDL097c_Sc    akytketyqsrqidamkavaeaynnrslldfntalkqy-----
YMJ5_Ce       ivayqkspriiairsmadafrkrslkdfvkalaeh-----
FUS6_ARATH    asglahlelkkyklaarkfldvnpelgnsyeneviapqdiatyggllcalasfdrselkqkv
COS41.8_Ci    kfleaagrynelysksaiheteqtkalekalncailapagqqrsmrlatlfkdercqllp
644879        qailtklkcaaglaelaarkykqaakclllasfdhcdfpellspsnvaiyggllcalatfd
YPR108w_Sc    llslistaalqsisstislyasdyasyfpyllety-----
```

```

eif-3p110_Hs -----
T23D8.4_Ce -----
YD95_Sp      vlavlpqnesmssleacinslylcdysgffrtladve-----
KIAA0107_Hs  ilevlhslpavrqylfslyecrysvffqslavv-----
F49C12.8_Hs  vqeqltggglngtlipvreylesyydchyrffiqlaale-----
Int-6_Mm     idnnsvssplqslqqrwtlihwslfvffnhpkgrdniidlflyqpqylnaiqtmcpphilr

26SPS9_Hs    -----
F57B9_Ce     -----
YDL097c_Sc   -----
YMJ5_Ce      -----
FUS6_ARATH   idninfrrnflvlpdvrelindfyssryascleylasl-----
COS41.8_Ci   sfgilekmfldriiksdemeeefar-----
644879       rqelqrnvissssfklflelepqvrddiifkfyekyasclkmldem-----
YPR108w_Sc   -----
eif-3p110_Hs -----
T23D8.4_Ce   -----
YD95_Sp      -----
KIAA0107_Hs  -----
F49C12.8_Hs  -----
Int-6_Mm     ylttavitnkdvrrkrqvlkdlvkviqqesytykdpitefveclyvnfdfdgaqkklrec

26SPS9_Hs    RAELRDDPIIISTHLAKLYDNLLLEQNLIRVIEPFSRVQIEHISSLIKLSKADVERKLSQMI
F57B9_Ce     PQELQMDPVVRKHFHSLSERMLEKDLCRIIEPYSFVQIEHVAQQIGIDRSKVEKKLSQMI
YDL097c_Sc   EKELMGDELTRSHFNALYDTLLESNLCKIIEPFECVEISHISKIIGLDTQQVEGKLSQMI
YMJ5_Ce      KIELVEDKVVAVHSQNLERNMLEKEISRVIIEPYSEIELSYIARVIGMTVPPVERAIARMI
FUS6_ARATH   KSNLLLDIHLHDHVDTLYDQIRKKALIQYTLPFVSVDLSRMADAFKTSVSGLEKELEALI
COS41.8_Ci   QLMPHQKAITADGSNILHRAVTEHNLLSASKLYNNIRFTELGALLEIPHQMAEKVASQMI
644879       KDNLLLDMYLAPHVRTLYTQIRNRALIQYFSPYVSADMHRMAAAFNTTVAALEDELTLQLI
YPR108w_Sc   ANVLIIPCKYLNHRHADFFVREMRRKVYAQLLESYKTLSLKSMASAFGVSVAFLDNDLGKFI
eif-3p110_Hs DLFPEADKVRTMLVRKIQEESLRTYLFTYSSVYDSISMETLSDMFELDLPTVHSIISKMI
T23D8.4_Ce   NLFHNAETVKGMVVRRIQEESLRTYLLTYSTVYATVSLKKLADLFELSKKDVHSIISKMI
YD95_Sp      VNHLKCDQFLVAHYRYVREMRRRAYAQLLESYRALSIDSMAASFGVSVDYIDRDLASFI
KIAA0107_Hs  EQEMKKDWLFAPHYRYVREMRIHAYSQLESYRSLTLGYMAEAFGVGVFIDQELSRFI
F49C12.8_Hs  SERFKFDRYLSPHFNYYSRGMHRHAYEQFLTYPYKTVRIDMMAKDFGVSRAFIDRELHRLI
Int-6_Mm     ESVLVNDFFLVACLEDFIENARLFIFETFCRIHQCISINMLADKLNMTPEEAERWIVNLI

26SPS9_Hs    LDKKFHGIIDQEGVLIIFDEPP
F57B9_Ce     LDQKLSGSLDQEGMLIVFEIAV
YDL097c_Sc   LDKIFYGVLDQGNGLVYVYETPN
YMJ5_Ce      LDKKLMGSIDQHGDTVVVYPKAD
FUS6_ARATH   TDNQIQARIDSHNKILYARHADQ
COS41.8_Ci   CESRMKGHIDQIDGIVFFERRET
644879       LEGLISARVDSHSHKILYARDVDQ
YPR108w_Sc   PNKQLNCVIDRVNGIVETNRPDN
eif-3p110_Hs INEELMASLDQPTQTVVMHRTEP
T23D8.4_Ce   IQEELSATLDEPTDCLIMHRVEP
YD95_Sp      PDNKLNCVIDRVNGVVFTNRPDE
KIAA0107_Hs  AAGRLHCKIDKVNEIVETNRPDS
F49C12.8_Hs  ATGQLQCRIDAVNGVIEVNRHDS
Int-6_Mm     RNARLDAKIDSKLGHVVMGNNAV

```