1) Discuss the following: Why does using protein sequences result in more accurate sequence alignments? Think about biological reasons, statistical reasons, and computational reasons.

2) Download the sequences you will use in the exercises (the **msa.txt** file).

   a) Can you figure out the origins of these sequences from the identification line?

   b) Use **T-Coffee** and **Clustal-O** to perform and compare the MSA for the given sequences (http://www.ebi.ac.uk/Tools/msa/). Compare the results.

      - What are the main differences? (Show some examples with screenshots)

      - Why do you think these differences occur? (No need to explain the algorithmic differences, just try to explain the main difference between the two tools).

3) Answer the following questions about Clustal-O.

   a) What type of MSA method does Clustal-O use?

   b) What is the substitution matrix used in Clustal-O?

   c) What are the default gap opening and extension penalties for Clustal-O?

   d) What do the (*), (:) and (.) symbols at the alignment positions indicate? What does it mean if there are no symbols?

4) Submit the same set of sequences (upload or paste) in the **msa.txt** file to **MEME Suite** with default parameters.

    a) How many motifs are these sequences sharing? Is that the maximum number of motifs you can find from these alignments?

    b) Observe the **1) seq logo, 2) the alignment** and **3) the graphical view** of the motifs. What is the specific purpose of each representation?

Protein sequence alignments are more accurate than DNA seq. alignments, both theoretically and practically. By the means of theory, proteins are made up 20 amino acids whereas DNA letters include only 4 bases, therefore signal-to-n-ratio is much better in proteins than DNA (or RNA) sequences. In addition, due to the possibility of too many DNA mutations, there is a significant workload / computational process timing difference between protein and DNA sequences when statistically compared. Also, proteins can be observed empirically, and matrices that are empirical substitution matrices like BLOSUM62 can be used on proteins. Also, since mutations are high in DNA, phylogenetic patterns are more conserved in proteins, since proteins are directly represent the phenotype of the organisms or the gene. Therefore, the mutations in the gene affect the organism's survive and fitness less than the mutations in the amino acid sequence. Amino Acid sequences, any deletions / insertions / substitutions, etc. directly affect the protein shape, therefore affect the protein function. And changes in the protein function is directly changes the survive or fitness states of the organism. Since protein sequences are more conserved comparing to the DNA sequences, and DNA has higher evolving rates than proteins, converting DNA sequence to the protein sequence and then making alignment is more common technique. Also, protein sequence alignments are more beneficial in terms of the seeing amino acid replacements which is affecting the physical and chemical properties of the proteins. Substitution matrices contains empirically derived scores from amino acid sequences, and aligning non-identical amino acids are more accurate in this techniques.

sequences indicate from which organism the sequences were originally obtained

>Curvularia (P49053) Vanadium chloroperoxidase
MGSVTPIPLPKIDEPEEYNTNYILFWNHVGLELNRVTHTVGGPLTGPPLSARALGMLHLAIHDAYFSICP
PTDFTTFLSPDTENAAYRLPSPNGANDARQAVAGAALKMLSSLYMKPVEQPNPNPGANISDNAYAQLGLV
LDRSVLEAPGGVDRESASFMFGEDVADVFFALLNDPRGASQEGYHPTPGRYKFDDEPTHPVVLIPVDPNN
PNGPKMPFRQYHAPFYGKTTKRFATQSEHFLADPPGLRSNADETAEYDDAVRVAIAMGGAQALNSTKRSP
WQTAQGLYWAYDGSNLIGTPPRFYNQIVRRIAVTYKKEEDLANSEVNNADFARLFALVDVACTDAGIFSW
KEKWEFEFWRPLSGVRDDGRPDHGDPFWLTLGAPATNTNDIPFKPPFPAYPSGHATFGGAVFQMVRRYYN
GRVGTWKDDEPDNIAIDMMISEELNGVNRDLRQPYDPTAPIEDQPGIVRTRIVRHFDSAWELMFENAISR
IFLGVHWRFDAAAARDILIPTTTKDVYAVDNNGATVFQNVEDIRYTTRGTREDPEGLFPIGGVPLGIEIA
DEIFNNGLKPTPPEIQPMPQETPVQKPVGQQPVKGMWEEEQAPVVKEAP

>Embellisia (emb|CAA72344.1) vanadium chloroperoxidase
MTIDFTPVELPVVEEDAEYNWNYILFWNNVGLELNRVTHTFGALKAGPPLSPRALGMLQLAVHDAYFAIH
PSAGFTTFLTPGAEDGAYRLPDPSYAKDARQAVAGAAIAMLSKLYMKPKVVPRSPISHNAYAQLQHVLDI
SVTKAPAACDPASSSFIFGKAVATAVFDLLFHKEGADQSGYSPKPGPFKFNDEPTHPVELIPVDANIPDG
DKMPRRQYHAPYYGETAKRFGTQTEHMLADPPGIRCAGEVAEYDDAIREVYAMGGAPGLNTTKRTPHQTV
QGMFWAYDGPKLIGTPPRLYNQIVRKIAVTYKKDNDLVNSEVNNADFARLLALVNVAMTDAGIFAWKEKW
EFEFWRPLSGVRDDVLRDPEGKASTAAIHSGLASAPQLQNSDEAPFKPPFPAYPSGHATFGAAAFQMVRK
YYNGRLGKWATTSRDTIAVEMFVSEELNGVSRDLSNPYDPKRPITDQPGIVPTRMPRRFSSCWEMMFENA
VSRIFLGVHWRFDAAAGQDILIPTTKKDVYAVDDKGAALFKNVEDIRYKTKGTRKGHKGLLPIGGVPLGI
EIANEIYNNKLSPTPPGEQPMPQPPQHQGPPRKKGELAEAKDEEQAPMMDVAP

>Drechslera (emb|CAA72008.1) vanadium chloroperoxidase
EFWRPLSGVRDDGRPDHGDPFWLTLGAPATNTNDIPFKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTW

Clustal-O is for medium or large alignments. Clustal-O uses guide tress and Hidden Markov Models to generate alignments. The default transition matrix is Gonnet, gap opening penalty is 6 bits, gap extension is 1 bit (EBI).It uses Progressive alignment construction method to make MSA. *=conserved residue, :=conservations between groups with strong similar properties, .=conservation btw groups with weak similar properties, symbols represent the related sequences are compared and if there is no symbol this means there is no consensus sequence

Clustal Omega

Input form   Web services   Help & Documentation   Bioinformatics Tools FAQ

💬 Fee

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-I20211205-091907-0811-35690280-p2m

Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Download Alignment File | Show Colors

```
CLUSTAL O(1.2.4) multiple sequence alignment


Embellisia       ------------------------------------------------------------   0
Curvularia       ------------------------------------------------------------   0
Drechslera       ------------------------------------------------------------   0
```

```
CLUSTAL O(1.2.4) multiple sequence alignment


Embellisia      ------------------------------------------------------------    0
Curvularia      ------------------------------------------------------------    0
Drechslera      ------------------------------------------------------------    0
Nostoc          ------------------------------------------------------------    0
Corallina       ------------------------------------------------------------    0
Deinococcus     ------------------------------------------------------------    0
Ascophyllum     ------------------------------------------------------------    0
Fucus           MLCHAADTTRGSPMPDTGVLRLLTSEQRAKGWRRQLEGEKSLGFHPSETPYIKYLEGSET   60


Embellisia      ------------------------------------------------------------    0
Curvularia      ------------------------------------------------------------    0
Drechslera      ------------------------------------------------------------    0
Nostoc          ------------------------------------------------------------    0
Corallina       ---MGIPAD---------------------------------------------------    6
Deinococcus     ------------------------------------------------------------    0
Ascophyllum     ------------------------------------------------------------    0
Fucus           WKKVKLPTDGISASKILGKIMARVRIATALAVVLAAPCLAFDEVTASGVFPEEHKHTGEG   120


Embellisia      ------------------------------------------------------------    0
Curvularia      ------------------------------------------------------------    0
Drechslera      ------------------------------------------------------------    0
Nostoc          ------------------------------------------------------------    0
Corallina       ---------------NLQSRAKASFDTRVSAAELAL-ARGVVPSLANGEELLYRNPDPE   49
Deinococcus     ------------------------------------------------------------    0
Ascophyllum     ---QTCSTSDDADDPTPPNERDDEAFASRVAAAKRELEGTGTV-------------CQIN   44
Fucus           RHLQTCTNSDDALDPTAPNRRDNVAFASRRDAARRERDGTGTV-------------CQIT   167
```

we can see the number of hits in Clustal but not in t-coffe.

```
Embellisia    -----MTIDFTPVELPVVEEDAEYNWNYILFWNNVGLE-----LNRVT-HTFGALKAGPP    49
Curvularia    ------MGSVTPIPLPKIDEPEEYNTNYILFWNHVGLE-----LNRVT-HTVGGPLTGPP    48
Drechslera    ------------------------------------------------------------    0
Nostoc        ----------------------MPDQVINWN------------NVYLQ-TIRSNGGAP     23
Corallina     NGDPSFIVSFT-KGLPHDDNGAIIDPDDFLAFVRAINSGDEKEIADLTLGPARDPDTGLP   108
Deinococcus   ------------------------------------------------------------    0
Ascophyllum   NGETDLAAKFH-KSLPHDDLGQV-DADAFAALEDCILNGDLSICEDVPVGNSEGD-----    97
Fucus         NGETDLATMFH-KSLPHDELGQV-TADDFAILEDCILNGDFSICEDVPA----GD-----   216


Embellisia    L-SPRALGMLQLAVHDAYFAIHPSAGFTTFLTPGAEDGAYRLPDPSYAKDARQAVAGAAI   108
Curvularia    L-SARALGMLHLAIHDAYFSICPPTDFTTFLSPDTENAAYRLPSPNGANDARQAVAGAAL   107
Drechslera    ------------------------------------------------------------    0
Nostoc        TWISRTGAILHSAIYDA---VNS-------IEKKYNPYLEIIPANPGASP-------EAA    66
Corallina     IWRSDLANSLELEVRGW---ENSSAGLTFDLEGPDAQSIAMPPAPVLTSP-------ELI   158
Deinococcus   ------------------------------------------------------------    0
Ascophyllum   ------------PVGRL---VNPTAAFAIDISGPAFSATTIPPVPTLPSP-------ELA   135
Fucus         ------------PAGRL---VNPTAAFAIDISGPAFSATTIPPVPTLSSP-------ELA   254


Embellisia    AMLSKLYMKPK--VV----PRSPISHNAYAQLQH--VLDISVTKAP-------------A   147
Curvularia    KMLSSLYMKPV--EQPNPNPGANISDNAYAQLGL--VLDRSVLEAP-------------G   150
Drechslera    ------------------------------------------------------------    0
Nostoc        AIYAAYT--VLTSDIVYPNANFPKSK----------AKNNSFFET-----------ERDK   103
Corallina     AEIAELYLMALGREIEFSEFDSPKNA-EYIQFAIDQLNGLEWFNTPAMLGDPPAEIRRRR   217
Deinococcus   ------------------------------------------------------------    0
Ascophyllum   AQLAEVYWMALARDVPFMQYGTDDIT-VTAA---ANLAGMEGFPNLD-----AVSIGSDG   186
Fucus         AQLAELYWMALARDVPFMQYGTDEIT-TTAA---ANLAGMGGFPNLD-----AVSIGSDG   305
```

```
Embellisia    AC---DPASS---SFIFGKAVATAVFDLLFHKEGADQSGYSP--------KPGPFKFNDE    193
Curvularia    GV---DRESA---SFMFGEDVADVFFALLNDPRGASQEGYHP--------TPGRYKFDDE    196
Drechslera    ----------------------------------------------------------    0
Nostoc        AIEELVSSGVSVQSIGDGKELGIAAAQAIL--QNRQADGYND--NTPYTPGNQPGDWRP-    158
Corallina     GE---VTVGNLFRGILPGSEVGPYLSQYII--VGSKQIGSATGGNKTLVSPNAADEFDGE    272
Deinococcus   ----------------------------------------------------------    0
Ascophyllum   TV---DPLSQLFRATFVGVETGPFISQLLV--NSF------------------------    216
Fucus         TV---DPFSQLFRATFVGVETGPFVSQLLV--NSF------------------------    335


Embellisia    PTHPVELIPVDA----NIPDGDKMPRRQYHAPYYGETAKRFGTQTEHMLADPPGIRCA-G    248
Curvularia    PTHPVVLIPVDP----NNPNGPKMPFRQYHAPFYGKTTKRFATQSEHFLADPPGLRSNAD    252
Drechslera    ----------------------------------------------------------    0
Nostoc        -T--GSSAP-------VTPNWGKVK------TFSK----API--KQFRPTRPAGFKTKKA    196
Corallina     IA--YGSITISQRVRIATPGRDFMT------DLKV-----FL--DVQDAADFRGFESYEP    317
Deinococcus   ----------------------------------------------------------    0
Ascophyllum   -T--IDSITVEPKQETFAPDVNYMV------DFDE-----WL--NIQNGGPPAGPELLDD    260
Fucus         -T--IDAITVEPKQETFAPDLNYMV------DFDE-----WL--NIQNGGPPAGPEELDE    379


Embellisia    E--VAEYDDAIREVYAMGGAPGLNTTKRTPHQTVQGMFWAYDGPKLIGTPP--RLYNQIV    304
Curvularia    E--TAEYDDAVRVAIAMGGAQALNSTKRSPWQTAQGLYWAYDGSNLIGTPP--RFYNQIV    308
Drechslera    ----------------------------------------------------------    0
Nostoc        LLASLEYAAQVNEVKRLG-ADNS--TERTQEQTDIALFWANDLDGTYKPPGHLFSITQIV    253
Corallina     G---ARLIRTIRDLATWVHFDAL--YEAYLNAC--LILLA--N-RVPFDPNIPFQQEDKL    367
Deinococcus   ----------------------------------------MTP--L----------GLWIEEA    11
Ascophyllum   E---LRFVRNARDLARVTFTDNI--NTEAYRGA--LILLG--L-DAFNRAGVNGPFIDID    310
Fucus         E---LRFIRNARDLARVSFVDNI--NTEAYRGS--LILLE--L-GAFSRPGINGPFIDSD    429


Embellisia    RKIAVTYKKDNDLVNSEVNNADFARLLALVNVAMTDAGIFAWKEKWEFE-----------    353
Curvularia    RRIAVTYKKEEDLANSEVNNADFARLFALVDVACTDAGIFSWKEKWEFE-----------    357
Drechslera    --------------------------------------------------E-----------    1
Nostoc        S-------------KLKGLSFYENARLFALVGLGLGDAGILAWDAKYNTD-----------    290
Corallina     D-------------NQDVFVNFGDAHVLSLVTEVATRALKAVRYQKFNIHRRLRPEATGGL    415
Deinococcus   L-------------RLGEQARLGGSDLAQVLAAT-AVAGHDAFISCWQ-------------    45
Ascophyllum   R-------------QAGFVNFGISHYFRLIGAA-ELAQRSSWYQKWQVHRFARPEALGGT    356
Fucus         R-------------QAGFVNFGTSHYFRLIGAA-ELAQRASCYQKWQVHRFARPEALGGT    475
```

```
Embellisia    --------------------FWRPLSGVRDDVLRDPEGKASTAAIHSG--LASAPQLQN    390
Curvularia    --------------------FWRPLSGVRDDGRPDHGDPF-------W--LTLGAPATN    387
Drechslera    --------------------FWRPLSGVRDDGRPDHGDPF-------W--LTLGAPATN     31
Nostoc        ------------------LDLWRPESAIQL--------AHTDGNPGTVADPTWRPLSPN    323
Corallina     ISVNKIAAEKGESVFPEVDLAVEELEDILEKAEISNRKQNIADGDPDPD----PSFLL-P    470
Deinococcus   ---------------GKFEYNVAR--------------------PQSW----MDHVQPG     65
Ascophyllum   LHLTIKGELN-----ADFDLSLLENAELLKRVAAIN-----AAQNPNNE----VTYLL-P    401
Fucus         LHNTIAGDLD-----ADFDISLLENDELLKRVAEIN-----AAQNPNNE----VTYLL-P    520
                                                       .  .

Embellisia    SDEAPFKPPFPAYPSGHATFGAAAFQMVRKYYNGRLGKWATTSRDTIAVEMFVSEELNGV    450
Curvularia    TNDIPFKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPDNIAIDMMISEELNGV    447
Drechslera    TNDIPFKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPDNIAIDMMISEELNGL     91
Nostoc        PDGTRFSPPFPAYISGHATFGAIHAGILRNFFGTDNVTFTAT------------------    365
Corallina     QAFAEGSPFHPSYGSGHAVVAGACVTILKAFFDSNFQIDQVF------------------    512
Deinococcus   WAPSLPTPPFPSYPSGHATVSGAAAEVLAQFFPLQA-----------------------    101
Ascophyllum   QAIQEGSPTHPSYPSGHATQNGAFATVLKALIGLDRGGDCYP------------------    443
Fucus         QAIQVGSPTHPSYPSGHATQNGAFATVLKALIGLDRGGECFP------------------    562
                   .*  .*:* ****.    .       ::

Embellisia    SRDLSNPYDPKRPITDQPGIVPTRM----PRRFSSCWEMMFENAVSRIFLGVHWRFDAAA    506
Curvularia    NRDLRQPYDPTAPIEDQPGIVRTRI----VRHFDSAWELMFENAISRIFLGVHWRFDAAA    503
Drechslera    NRDLRQPYDPTAPIEDQPGIVRTRI----VRHFDSAWEMMFENAISRIFLGVHWRFDAAA    147
Nostoc        ---------SEDPSAR--GANGIRVT---RT-FNSFSAAALENGRSRVYLGVHYQWDADA    410
Corallina     ---------EVDKDED--KLVKSSFK-GTLTVAGELNKLADNIAIGRNMAGVHYFSDQFE    560
Deinococcus   ------------------------------RQLRRDARDAAFSRVVGGIHWGVDGVA    128
Ascophyllum   --------DPVYPDDD--GLKLIDFRGSCLTFEGEINKLAVNVAFGRQMLGIHYRFDGIQ    493
Fucus         --------NPVFPSDD--GLELINFEGACLTYEGEINKLAVNVAFGRQMLGIHYRFDGIQ    612
                                                 *   *  *    *
```

```
Embellisia    SRDLSNPYDPKRPITDQPGIVPTRM----PRRFSSCWEMMFENAVSRIFLGVHWRFDAAA    506
Curvularia    NRDLRQPYDPTAPIEDQPGIVRTRI----VRHFDSAWELMFENAISRIFLGVHWRFDAAA    503
Drechslera    NRDLRQPYDPTAPIEDQPGIVRTRI----VRHFDSAWEMMFENAISRIFLGVHWRFDAAA    147
Nostoc        ---------SEDPSAR--GANGIRVT---RT-FNSFSAAALENGRSRVYLGVHYQWDADA    410
Corallina     ---------EVDKDED--KLVKSSFK-GTLTVAGELNKLADNIAIGRNMAGVHYFSDQFE    560
Deinococcus   -----------------------------RQLRRDARDAAFSRVVGGIHWGVDGVA       128
Ascophyllum   --------DPVYPDDD--GLKLIDFRGSCLTFEGEINKLAVNVAFGRQMLGIHYRFDGIQ    493
Fucus         --------NPVFPSDD--GLELINFEGACLTYEGEINKLAVNVAFGRQMLGIHYRFDGIQ    612
                            .           :  . .*    *:*:   *

Embellisia    GQDIL----IPTTKKDVYAVDDKGAALFKNVEDIRYKTKGTRKGHKGLLPIGGV------    556
Curvularia    ARDIL----IPTTTKDVYAVDNNGATVFQNVEDIRYTTRGTREDPEGLFPIGGV------    553
Drechslera    ARDIL----IPTTTKDVYAVDNNGATVFQNVEDVRYSTKGTREGREGLFPIGGV------    197
Nostoc        AYVSGTKLADFVSE-NLLTPSVRC-----------------------------------    433
Corallina     SILLGEQVAIGILEEQSLTYGENF---FFNLP--KFDGTTIQI-----------------    598
Deinococcus   GLDVGQRVARALLEKRP-----------------------------------------    145
Ascophyllum   GLLLGETITVRTLHQELMTFAEES---TFEFR--LFTGEVIKLFQDGTFTIDGFKCPGLV    548
Fucus         GLLLGETITVRTLHQELMTFAEEA---TFEFR--LFTGEVIKLFQDGTFSIDGDMCSGLV    667
                .

Embellisia    PLGIEIANEIYNNKLSPTPPGEQPMPQPPQHQGPPRKKGELAEAKDEEQAPMMDVAP     613
Curvularia    PLGIEIADEIFNNGLKPTPPEIQPMPQETPVQKPVGQQ-PVKGMWEEEQAPVVKEAP     609
Drechslera    PLGIEIADEIFNNGLRPTPPELQPMPQETPVQKP------VEGMWEEEQAPIVNEAP     248
Nostoc        --------------------------------------------------------     433
Corallina     --------------------------------------------------------     598
Deinococcus   --------------------------------------------------------     145
Ascophyllum   YTGVENCVS-----------------------------------------------     557
Fucus         YTGVADCQA-----------------------------------------------     676
```

PLEASE NOTE: Showing colors on large alignments is slow.

Results for job clustalo-I20211205-091907-0811-35690280-p2m

Input Sequences

clustalo-I20211205-091907-0811-35690280-p2m.input

Tool Output

clustalo-I20211205-091907-0811-35690280-p2m.output

Alignment in CLUSTAL format with base/residue numbering

clustalo-I20211205-091907-0811-35690280-p2m.clustal_num

Guide Tree

clustalo-I20211205-091907-0811-35690280-p2m.dnd

Phylogenetic Tree

clustalo-I20211205-091907-0811-35690280-p2m.ph

Percent Identity Matrix

clustalo-I20211205-091907-0811-35690280-p2m.pim

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-I20211205-091907-0811-35690280-p2m

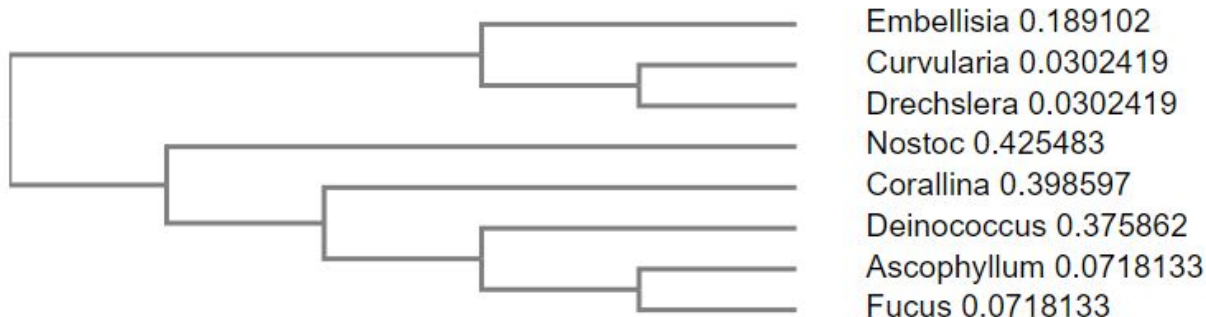| Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details |

Download Guide Tree Data

Phylogram

Branch length: ● Cladogram   ○ Real

Embellisia 0.189102
Curvularia 0.0302419
Drechslera 0.0302419
Nostoc 0.425483
Corallina 0.398597
Deinococcus 0.375862
Ascophyllum 0.0718133
Fucus 0.0718133

Guide Tree

```
(
(
Embellisia:0.189102

,
(
Curvularia:0.0302419

,
Drechslera:0.0302419
):0.15886
):0.240292

,
(
Nostoc:0.425483

,
(
Corallina:0.398597

,
(
Deinococcus:0.375862

,
(
Ascophyllum:0.0718133

,
Fucus:0.0718133
):0.304049
):0.0227354
):0.0268853
):0.00391147
)
;
```

Download Phylogenetic Tree Data

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ● Cladogram ○ Real
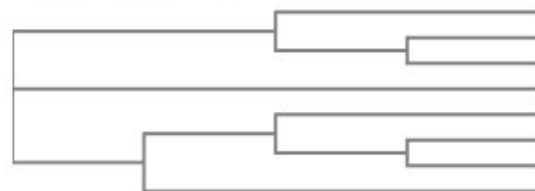
Embellisia 0.1744
Curvularia 0.03365
Drechslera 0.01473
Nostoc 0.35951
Corallina 0.34994
Ascophyllum 0.06329
Fucus 0.06871
Deinococcus 0.34398

Tree Data

```
(
(
Embellisia:0.17440,
(
Curvularia:0.03365,
Drechslera:0.01473)
:0.13964)
:0.20044,
Nostoc:0.35951,
(
(
Corallina:0.34994,
(
Ascophyllum:0.06329,
Fucus:0.06871)
:0.25749)
:0.06830,
Deinococcus:0.34398)
:0.04462);
```

Results for job clustalo-I20211205-091907-0811-35690280-p2m

Alignments	Result Summary	Guide Tree	Phylogenetic Tree	Results Viewers	Submission Details

Program	Number of Sequences	Launched Date
clustalo	8	Sun, Dec 05, 2021 at 09:19:08
Version		**End Date**
1.2.4		Sun, Dec 05, 2021 at 09:19:10

Input Sequences

clustalo-I20211205-091907-0811-35690280-p2m.input

Output Result

clustalo-I20211205-091907-0811-35690280-p2m.output

Command

```
$APPBIN/clustal-omega-1.2.4/bin/clustalo --infile clustalo-I20211205-091907-0811-35690280-p2m.sequence --threads 8 --MAC-
RAM 8000 --verbose --guidetree-out clustalo-I20211205-091907-0811-35690280-p2m.dnd --outfmt clustal --resno --outfile
clustalo-I20211205-091907-0811-35690280-p2m.clustal_num --output-order tree-order --seqtype protein
```

Input Parameters

Output guide tree

true

Output distance matrix

false

input parameters are more higher in clustal because it based on HMM but input parameters are low in t-coffe since it based on heuristic algorithms. it affects the general output and phylogeny of the sequences. Also guide trees are different. T-coffee can be detect patterns whereas Clustal can be detect conserved areas more efficiently.

true

Output distance matrix
false

Dealign input sequences
false

mBed-like clustering guide tree
true

mBed-like clustering iteration
true

Number of iterations
0

Maximum guide tree iterations
-1

Maximum HMM iterations
-1

Output alignment format
clustal_num

Output order
aligned

Sequence Type
protein

T-Coffee is used for small alignments. In T-coffee, comparison is made by creating two libraries for both global (with ClustalW) and local (with Lalign) and uses heuristic algorithms on this libraries. T-coffee motifs are good at identifying.

T-Coffee

Input form   Web services   Help & Documentation   Bioinformatics Tools FAQ   💬 Feedback

Tools > Multiple Sequence Alignment > T-Coffee

Results for job tcoffee-I20211205-092003-0880-78434929-p1m

Alignments   Result Summary   Guide Tree   Phylogenetic Tree   Results Viewers   Submission Details

Download Alignment File   Show Colors

```
CLUSTAL W (1.83) multiple sequence alignment

Ascophyllum      QTCS-----------------T-----------------------
Corallina        MGIP-----------------A-----------------------
Curvularia       MGSV-TPIPLPKIDEPEEYNTNYILFWNHVGLELNRVTHTVGGPLTGPPL
Deinococcus      MTPL----GLWI---------------------------------
Drechslera       --------------------------------------------
Embellisia       MTIDFTPVELPVVEEDAEYNWNYILFWNNVGLELNRVTHTFGALKAGPPL
Fucus            MLCH-----------------A-----------------------
Nostoc           MDD                  QVINVNNVVIOTIRS          NCGARTW
```

```
CLUSTAL W (1.83) multiple sequence alignment

Ascophyllum    QTCS------------------T------------------------
Corallina      MGIP------------------A------------------------
Curvularia     MGSV-TPIPLPKIDEPEEYNTNYILFWNHVGLELNRVTHTVGGPLTGPPL
Deinococcus    MTPL----GLWI-----------------------------------
Drechslera     -----------------------------------------------
Embellisia     MTIDFTPVELPVVEEDAEYNWNYILFWNNVGLELNRVTHTFGALKAGPPL
Fucus          MLCH------------------A------------------------
Nostoc         MPD-----------------QVINWNNVYLQTIRS------NGGAPTW


Ascophyllum    -----------------------------------------------
Corallina      -----------------------------------------------
Curvularia     SARALGMLHLAIHDAYFSICPPTDFTTFLSPDTENAAYRLP-SPNGANDA
Deinococcus    -----------------------------------------------
Drechslera     -----------------------------------------------
Embellisia     SPRALGMLQLAVHDAYFAIHPSAGFTTFLTPGAEDGAYRLP-DPSYAKDA
Fucus          ---------------------ADT-TRGSPMPDTGVLRLLTSEQRAKGW
Nostoc         ISRTGAILHSAIYDAVNSIEKKYNP--YLE--------IIP-ANPGASPE


Ascophyllum    -----------------------------------------------
Corallina      -----------------------------------------------
Curvularia     RQAVAGA-----------------------------------------
Deinococcus    -----------------------------------------------
Drechslera     -----------------------------------------------
Embellisia     RQAVAGA-----------------------------------------
Fucus          RRQLEGEKSLGFHPSETPYIKYLEGSETWKKVKLPTDGISASKILGKIMA
Nostoc         AAAIYAA-----------------------------------------


Ascophyllum    -----------------------------------------------
Corallina      -----------------------------------------------
Curvularia     -------------ALKMLSSLYMKPVEQPNPNPGAN--------------
Deinococcus    ----------------------------E-------------------
Drechslera     -----------------------------------------------
Embellisia     -------------AIAMLSKLYMKPKVVP----RSP--------------
Fucus          RVRIATALAVVLAAPCLAFDEVTASGVFP----EEHKHTGEGRHLQTCTN
Nostoc         -------------YTVLTSDIVYPNANFP----KSK--------------
```

the order of the sequences (organisms) is different than culstal

```
Ascophyllum    SDDADDPTPPNERDDEAFASRV--AAAKRELEGTGTVCQI-NNG------
Corallina      DN-------LQSRAKASFDTRV--SAAELAL-ARGVVPSL-ANGEELLYR
Curvularia     -----------ISDNAYAQLGLVLDRSVLEA-PGGVDRES--AS------
Deinococcus    --------------------------------------------------
Drechslera     --------------------------------------------------
Embellisia     -----------ISHNAYAQLQHVLDISVTKA-PAACDPAS--SS------
Fucus          SDDALDPTAPNRRDNVAFASRR--DAARRERDGTGTVCQI-TNG------
Nostoc         -----------AKNNSFFETER--DKAIEELVSSGVSVQSIGDG------


Ascophyllum    -------ETDLAAKFHKSLPHDDLGQ-VDADAFAALEDCILNGDLSICED
Corallina      NPDPENGDPSFIVSFTKGLPHDDNGAIIDPDDFLAFVRAINSGDEKEIAD
Curvularia     ----------------------FMFGE-DVAD--VFFALLNDPRGASQEGY
Deinococcus    --------------------------------------------------
Drechslera     --------------------------------------------------
Embellisia     --------------------FIFGK-AVAT--AVFDLLFHKEGADQSGY
Fucus          -------ETDLATMFHKSLPHDELGQ-VTADDFAILEDCILNGDFSICED
Nostoc         --------------------KELGI-AAAQ--AILQNRQADGYNDNTPY


Ascophyllum    --VPVGNSEGDP------------------VGRLVNPTAAFAIDISGPAF
Corallina      --LTLGP-ARDPDTGLPIWRSDLANSLELEVRGWENSSAGLTFDLEGPDA
Curvularia     HPTPG-RYKFD----------------------DE-----------PTH
Deinococcus    --------------------------------------------------
Drechslera     --------------------------------------------------
Embellisia     SPKPG-PFKFN----------------------DE-----------PTH
Fucus          --VPAG----DP------------------AGRLVNPTAAFAIDISGPAF
Nostoc         --TPG-NQPGD----------------------WR-----------PTG


Ascophyllum    SAT-TIPPVPTLPSPELAAQLAEVYWMALARDVPFMQY-------------
Corallina      QSI-AMPPAPVLTSPELIAEIAELYLMALGREIEFSEFDSPKNAEYIQ-F
Curvularia     PVV-LIPVDPN---------------NPNGPKMPFRQYHAPFYGKTTKRF
Deinococcus    -----------------------------------------------EALRL
Drechslera     --------------------------------------------------
Embellisia     PVE-LIPVDAN---------------IPDGDKMPRRQYHAPYYGETAKRF
Fucus          SAT-TIPPVPTLSSPELAAQLAELYWMALARDVPFMQY-------------
Nostoc         SSAPVTPNWGK---------------VKTFSKAPIKQFRP----------
```

```
Ascophyllum    GTDDITVTAAANLAGMEGFPNLDAVSIGSDGTVDPLSQLFRATFV-GVET
Corallina      AIDQLNGLEWFNTPAMLGDPPAEIRRRRGEVTV---GNLFRGILP-GSEV
Curvularia     A---------TQSEHFLADPPG--------------------LRSNADET
Deinococcus    G-------------------------------------------------
Drechslera     --------------------------------------------------
Embellisia     G---------TQTEHMLADPPG--------------------IRC-AGEV
Fucus          GTDEITTTAAANLAGMGGFPNLDAVSIGSDGTVDPFSQLFRATFV-GVET
Nostoc         ----------------TRPAG--------------------FKTKKALL


Ascophyllum    --GPFISQLLV-N------------------------SFTIDSITVEPK
Corallina      --GPYLSQYIIVGSKQIGSATGGNKTLVSPNAADEFDGEIAYGSITISQR
Curvularia     --AEYDD--AVRVAIAMGGAQALNSTKRSPWQTAQ---------------
Deinococcus    -------------EQARL--------------------------------
Drechslera     --------------------------------------------------
Embellisia     --AEYDD--AIREVYAMGGAPGLNTTKRTPHQTVQ---------------
Fucus          --GPFVSQLLV-N------------------------SFTIDAITVEPK
Nostoc         ASLEYAA--QVNEVKRLGA---DNSTERTQEQTDI---------------


Ascophyllum    QETFAPDVNYMVDFDEWLNIQNGGPPAGPELLDDELRFVRNARDLARVTF
Corallina      VRIATPGRDFMTDLKVFLDVQDAADFRGFESYEPGARLIRTIRDLATWVH
Curvularia     ---------------GLYW----AY--DGSNLIGTPPRFYN----------
Deinococcus    --------------------------------------------------
Drechslera     --------------------------------------------------
Embellisia     ---------------GMFW----AY--DGPKLIGTPPRLYN----------
Fucus          QETFAPDLNYMVDFDEWLNIQNGGPPAGPEELDEELRFIRNARDLARVSF
Nostoc         ---------------ALFW----AN--DLDGTYKPPGHLFSI---------


Ascophyllum    TDNINTEAYRGA-LILLGLDAFNRAGVNGPFID---IDRQAGFVNFGISH
Corallina      FDALY-EAYLNACLILLANRVPFDP--NIPFQQEDKLDNQDVFVNFGDAH
Curvularia     -------------------QIVRRI--AVTYKKEEDLANSEV-NNADFAR
Deinococcus    --------------------------------------------GGSDLAQ
Drechslera     --------------------------------------------------
Embellisia     -------------------QIVRKI--AVTYKKDNDLVNSEV-NNADFAR
Fucus          VDNINTEAYRGS-LILLELGAFSRPGINGPFID---SDRQAGFVNFGTSH
Nostoc         -------------------TQIVSKL--KG------------L-SFYENAR
```

```
Ascophyllum    YFRLIG-AAELAQRSSWYQKWQ--VHRFARPEALGGTLHL-TIKGE----
Corallina      VLSLVTEVATRALKAVRYQKFN--IHRRLRPEATGGLISVNKIAAEKGES
Curvularia     LFALVDVACTDAGIFSWKEKWE--F-EFWRPLSGV--R------------
Deinococcus    VLAATAVAGHDAFISCWQGKFE--Y-NVARPQSWM---------------
Drechslera     --------------------------EFWRPLSGV--R------------
Embellisia     LLALVNVAMTDAGIFAWKEKWE--F-EFWRPLSGV--R------------
Fucus          YFRLIG-AAELAQRASCYQKWQ--VHRFARPEALGGTLHN-TIAGD----
Nostoc         LFALVGLGLGDAGILAWDAKYNTDL-DLWRPESAI--Q------------
                                                      ** :

Ascophyllum    LNADFDLSLLENAELLK------RVAAINAAQNPNN----EVTYLLPQAI
Corallina      VFPEVDLAVEELEDILEKAEISNRKQNI-ADGDPDP----DPSFLLPQAF
Curvularia     ------------DDGR------PD-HGD-----PFWLTLGAPA-TN-TND
Deinococcus    ----------------------D-HVQ-----PGW----AP--------
Drechslera     ------------DDGR------PD-HGD-----PFWLTLGAPA-TN-TND
Embellisia     ------------DDVLRDPEGKASTAAI-----HSGLAS-APQLQN-SDE
Fucus          LDADFDISLLENDELLK------RVAEINAAQNPNN----EVTYLLPQAI
Nostoc         ---------LAHTDGN------PGTVAD-----PTW----RPLSPN-PDG

Ascophyllum    QEGSPTHPSYPSGHATQNGAFATVLKALIGLD------------------
Corallina      AEGSPFHPSYGSGHAVVAGACVTILKAFFDSN------------------
Curvularia     IPFKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPDNIAIDMMI
Deinococcus    SLPTPPFPSYPSGHATVSGAAAEVLAQFFPLQARQLR-------------
Drechslera     IPFKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPDNIAIDMMI
Embellisia     APFKPPFPAYPSGHATFGAAAFQMVRKYYNGRLGKWATTSRDTIAVEMFV
Fucus          QVGSPTHPSYPSGHATQNGAFATVLKALIGLD------------------
Nostoc         TRFSPPFPAYISGHATFGAIHAGILRNFFGTDNVTFTATSE---------
                .* .*:* ****.  .      ::

Ascophyllum    --------RGGDCYPDPVYPDDDGLKLIDFRGSCLTFEGEINKLAVNVAF
Corallina      --------FQIDQVF-EVDKDEDKLVKSSFKGT-LTVAGELNKLADNIAI
Curvularia     SEELNGVNRDLRQPYDPTAPIEDQPGIVRTRI-VRHFD-SAWELMFENAI
Deinococcus    -------------------------------------------RDARDAAF
Drechslera     SEELNGLNRDLRQPYDPTAPIEDQPGIVRTRI-VRHFD-SAWEMMFENAI
Embellisia     SEELNGVSRDLSNPYDPKRPITDQPGIVPTRM-PRRFS-SCWEMMFENAV
Fucus          --------RGGECFPNPVFPSDDGLELINFEGACLTYEGEINKLAVNVAF
Nostoc         ---------------DPSAR-----GANGIRV-TRTFN-SFSAAALENGR
                                                        : .
```

```
Ascophyllum    GRQMLGIHYRFDGIQGLLLGETITVRTLHQELMTFAEESTFEFRL-----
Corallina      GRNMAGVHYFSDQFESILLGEQVAIGILEEQSLTYGENFFFNLPK-----
Curvularia     SRIFLGVHWRFDAAAARDI----LIPTTTKDVYAVDNNGATVFQNVEDIR
Deinococcus    SRVVGGIHWGVDGVAGLDV-------------------------------
Drechslera     SRIFLGVHWRFDAAAARDI----LIPTTTKDVYAVDNNGATVFQNVEDVR
Embellisia     SRIFLGVHWRFDAAAGQDI----LIPTTKKDVYAVDDKGAALFKNVEDIR
Fucus          GRQMLGIHYRFDGIQGLLLGETITVRTLHQELMTFAEEATFEFRL-----
Nostoc         SRVYLGVHYQWDADAAYVS-------------------------------
               .*    *:*:  *    .

Ascophyllum    FTGEVIKLFQDGTFTIDGFKCPGLVYTGVENCV-----------------
Corallina      FDGTTIQ-------------------------------------------
Curvularia     YTTRGTREDPEGLFPIGG------VPLGIEIADEIFNNGLKPTPPEIQPM
Deinococcus    --------------------------GQRVARALL---------------
Drechslera     YSTKGTREGREGLFPIGG------VPLGIEIADEIFNNGLRPTPPELQPM
Embellisia     YKTKGTRKGHKGLLPIGG------VPLGIEIANEIYNNKLSPTPPGEQPM
Fucus          FTGEVIKLFQDGTFSIDGDMCSGLVYTGVADCQ-----------------
Nostoc         -------------------------GTKLADFVSENLLTPSVR-----

Ascophyllum    -------------------------------S
Corallina      -------------------------------I
Curvularia     PQETPVQKPVGQQ-PVKGMWEEEQAPVVKEAP
Deinococcus    ---------------------------EKRP
Drechslera     PQETPVQKP------VEGMWEEEQAPIVNEAP
Embellisia     PQPPQHQGPPRKKGELAEAKDEEQAPMMDVAP
Fucus          -------------------------------A
Nostoc         -------------------------------C
```

PLEASE NOTE: Showing colors on large alignments is slow.

T-Coffee

Tools > Multiple Sequence Alignment > T-Coffee

Results for job tcoffee-I20211205-092003-0880-78434929-p1m

Alignments | **Result Summary** | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Input Sequences

tcoffee-I20211205-092003-0880-78434929-p1m.input

Tool Output

tcoffee-I20211205-092003-0880-78434929-p1m.output

Alignment in CLUSTAL format

tcoffee-I20211205-092003-0880-78434929-p1m.clustalw

Guide Tree

tcoffee-I20211205-092003-0880-78434929-p1m.dnd

Phylogenetic Tree

tcoffee-I20211205-092003-0880-78434929-p1m.ph

Percent Identity Matrix

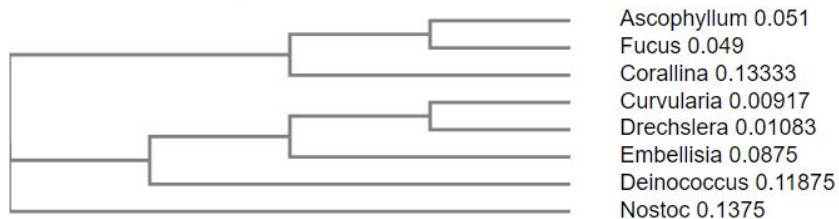tcoffee-I20211205-092003-0880-78434929-p1m.pim

Tools > Multiple Sequence Alignment > T-Coffee

Results for job tcoffee-I20211205-092003-0880-78434929-p1m

| Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details |

| Download Guide Tree Data |

Phylogram

Branch length: ● Cladogram ○ Real

Ascophyllum 0.051
Fucus 0.049
Corallina 0.13333
Curvularia 0.00917
Drechslera 0.01083
Embellisia 0.0875
Deinococcus 0.11875
Nostoc 0.1375

Guide Tree

(((Ascophyllum:0.05100,Fucus:0.04900):0.07167,Corallina:0.13333):0.00750,(((Curvularia:0.00917,Drechslera:0.01083):0.08250,Embellisia:0.08750):0.02875,Deinococcus:0.11875):0.00750,Nostoc:0.13750);

Tools > Multiple Sequence Alignment > T-Coffee

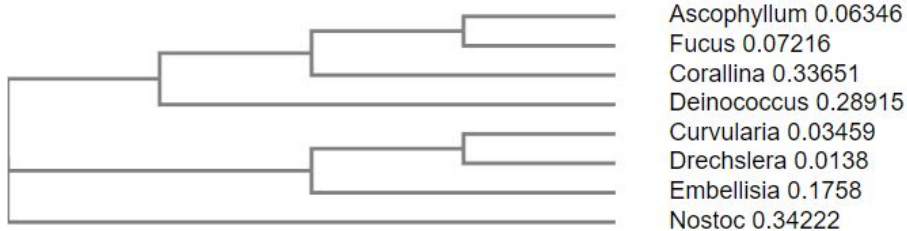# Results for job tcoffee-I20211205-092003-0880-78434929-p1m

Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Download Phylogenetic Tree Data

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ● Cladogram   ○ Real

Ascophyllum 0.06346
Fucus 0.07216
Corallina 0.33651
Deinococcus 0.28915
Curvularia 0.03459
Drechslera 0.0138
Embellisia 0.1758
Nostoc 0.34222

guide trees and phylogenetic trees are different in t-coffee and clustal!!
because the approaches of to align the seqeunces are different &&
T-coffee motifs are good at identifying.

Tree Data

```
(
(
(
(
Ascophyllum:0.06346,
Fucus:0.07216)
:0.23498,
Corallina:0.33651)
:0.10449,
Deinococcus:0.28915)
:0.04356,
(
(
Curvularia:0.03459,
Drechslera:0.01380)
:0.13184,
Embellisia:0.17580)
:0.16521,
Nostoc:0.34222);
```

| Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details |

Program

T-COFFEE

Version

13.41.0.28bdc39

Number of Sequences

8

Launched Date

Sun, Dec 05, 2021 at 09:11:42

End Date

Sun, Dec 05, 2021 at 09:11:48

Input Sequences

tcoffee-I20211205-092003-0880-78434929-p1m.input

Output Result

tcoffee-I20211205-092003-0880-78434929-p1m.output

Command

```
$APPBIN/tcoffee:13.45.0.4846264 /t_coffee -in tcoffee-I20211205-092003-0880-78434929-p1m.sequence -case=upper -n_core=8 -
output=clustalw,msf,phylip,score_html,fasta -outorder=aligned -type=protein; echo ' '
```

Input Parameters

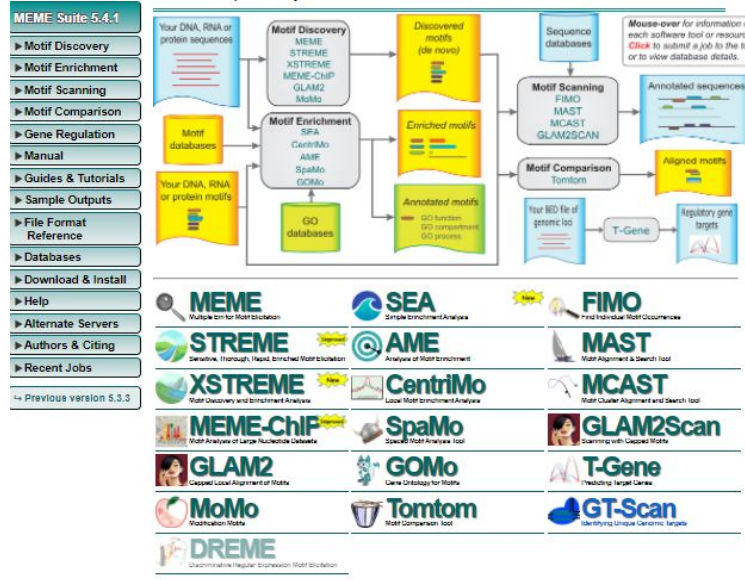**Alignment format**

clustalw

Matrix

Output Order

aligned

Sequence Type

protein

The MEME Suite

Motif-based sequence analysis tools

MEME Suite 5.4.1

- ▶ Motif Discovery
- ▶ Motif Enrichment
- ▶ Motif Scanning
- ▶ Motif Comparison
- ▶ Gene Regulation
- ▶ Manual
- ▶ Guides & Tutorials
- ▶ Sample Outputs
- ▶ File Format Reference
- ▶ Databases
- ▶ Download & Install
- ▶ Help
- ▶ Alternate Servers
- ▶ Authors & Citing
- ▶ Recent Jobs

↳ Previous version 5.3.3

MEME — Multiple Em for Motif Elicitation

SEA — Simple Enrichment Analysis

FIMO — Find Individual Motif Occurrences

STREME — Sensitive, Thorough, Rapid, Enriched Motif Elicitation

AME — Analysis of Motif Enrichment

MAST — Motif Alignment & Search Tool

XSTREME — Motif Discovery and Enrichment Analysis

CentriMo — Local Motif Enrichment Analysis

MCAST — Motif Cluster Alignment and Search Tool

MEME-ChIP — Motif Analysis of Large Nucleotide Datasets

SpaMo — Spaced Motif Analysis Tool

GLAM2Scan — Scanning with Gapped Motifs

GLAM2 — Gapped Local Alignment of Motifs

GOMo — Gene Ontology for Motifs

T-Gene — Predicting Target Genes

MoMo — Modification Motifs

Tomtom — Motif Comparison Tool

GT-Scan — Identifying Unique Genomic Targets

DREME — Discriminative Regular Expression Motif Elicitation

Maintenance and development of the MEME Suite is funded by grant R01 GM103544 from the National Institutes of Health.

Developed and maintained by

University of Nevada, Reno   UNIVERSITY OF WASHINGTON   THE UNIVERSITY OF QUEENSLAND AUSTRALIA   UCSD

MEME
Multiple Em for Motif Elicitation

MEME Suite 5.4.1

▶Motif Discovery
▶Motif Enrichment
▶Motif Scanning
▶Motif Comparison
▶Gene Regulation
▶Manual
▶Guides & Tutorials
▶Sample Outputs
▶File Format Reference
▶Databases
▶Download & Install
▶Help
▶Alternate Servers
▶Authors & Citing
▶Recent Jobs

↪ Previous version 5.3.3

Please wait. Your MEME job is now queued awaiting available resources. You may bookmark this page or use the **Recent Jobs** menu at the left to access your job's results.

Job Details ...

Results

- (Primary) Sequences

Status Messages

- Parsing arguments
- Arguments ok
- Starting meme
  meme msa.txt -protein -oc . -nostatus -time 14400 -mod zoops -nmotifs 3 -minw 6 -maxw 50 -objfun classic -markov_order 0

MEME
Multiple Em for Motif Elicitation

For further information on how to interpret these results please access https://meme-suite.org/meme/doc/meme.html.
To get a copy of the MEME software please access https://meme-suite.org.

If you use MEME in your research, please cite the following paper:
Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biol* AAAI Press, Menlo Park, California, 1994. [full text]

DISCOVERED MOTIFS | **MOTIF LOCATIONS** | **INPUTS & SETTINGS** | **PROGRAM INFORMATION** | **RESULTS IN TEXT FORMAT** ? | **RESULTS IN XML FORMAT** ?

DISCOVERED MOTIFS

	Logo ?	E-value ?	Sites ?	Width ?	More ?	Submit/Download ?
1.		1.8e-071	8	40	↧	⟶
2.		1.1e-046	7	27	↧	⟶
3.		3.9e-026	3	39	↧	⟶

DISCOVERED MOTIFS

1.

E-value: 1.8e-071 ⯑   Site Count: 8 ⯑   Width: 40 ⯑

Log Likelihood Ratio: 628 ⯑   **Information Content:** 120.3 ⯑   **Relative Entropy:** 113.2 ⯑   **Bayes Threshold:** 8.79726 ⯑

Name	Start	p-value	Sites
3. Drechslera	37	2.59e-44	APATNTNDIP FKPPF AYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPD NIAIDMMISE
1. Curvularia	393	2.59e-44	APATNTNDIP FKPPF AYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPD NIAIDMMISE
2. Embellisia	396	9.84e-38	PQLQNSDEAP FKPPF AYPSGHATFGAAAFQMVRKYYNGRLGKWATTSRD TIAVEMFVSE
6. Ascophyllum	407	1.34e-33	TYLLPQAIQE GSPTHPSYPSGHATQNGAFATVLKALIGLDRGGDCYPDPV YPDDDGLKLI
8. Fucus	526	4.72e-33	TYLLPQAIQV GSPTHPSYPSGHATQNGAFATVLKALIGLDRGGECFPNPV FPSDDGLELI
4. Nostoc	329	1.87e-28	PLSPNPDGTR FSPPF AYISGHATFGAIPAGILRNFFGTDNVFTATSED PSARGANGIR
5. Deinococcus	71	9.48e-28	HVQPGWAPSL PTPPFPSYPSGHATVSGAAAEVLAQFFPLQARQLRRDARD AAFSRVVGGI
7. Corallina	476	1.57e-27	SFLLPQAFAE GSPFHFSYGSGHAVVAGACVTILKAFFDSNFQIDQVFEVD KDEDKLVKSS

2.

E-value: 1.1e-046 ⯑   Site Count: 7 ⯑   Width: 27 ⯑

Log Likelihood Ratio: 409 ⯑   **Information Content:** 86.5 ⯑   **Relative Entropy:** 84.4 ⯑   **Bayes Threshold:** 8.87541 ⯑

Name	Start	p-value	Sites
2. Embellisia	481	2.28e-32	VPTRMPRRFS SCWEMMFENAVSRIFLGVIWRFDAAAG QDILIPTTKK
3. Drechslera	122	2.03e-31	VRTRIVRHFD SAWEMMFENAISRIFLGVIWRFDAAAA RDILIPTTTK
1. Curvularia	478	3.54e-31	VRTRIVRHFD SAWELMFENAISRIFLGVIWRFDAAAA RDILIPTTTK
8. Fucus	587	2.85e-26	FEGACLTYEG EINKLAVNVAFGRQMLGIYRFDGIQG LLLGETITVR
6. Ascophyllum	468	2.85e-26	FRGSCLTFEG EINKLAVNVAFGRQMLGIYRFDGIQG LLLGETITVR
7. Corallina	535	9.78e-20	SFKGTLTVAG ELNKLADNIAIGRNMAGVYYFSDQFES ILLGEQVAIG
9. Nostoc	385	4.02e-19	NGIRVTRTFN SFSAAALENGRSRVYLGVQWDADAA YVSGTKLADF

A motif is a sequence pattern that occurs repeatedly in a group of related sequences and here is the motif patterns of the alignments
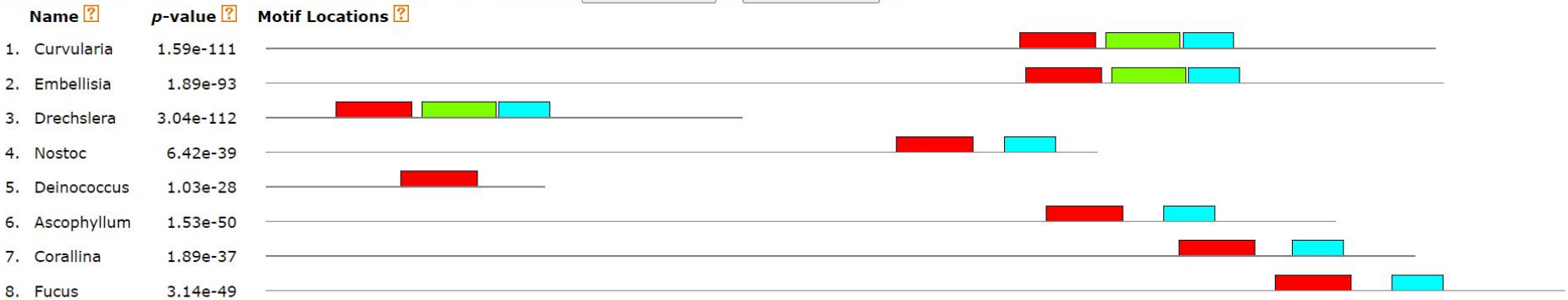

3.
E-value: 3.9e-026 ?   Site Count: 3 ?   Width: 39 ?

Log Likelihood Ratio: 317 ?   Information Content: 156.6 ?   Relative Entropy: 152.3 ?   Bayes Threshold: 9.87619 ?

Name ?	Start ?	p-value ?	Sites ?
1. Curvularia	438	2.57e-50	DDEPDNIAID MMISEELNGVNRDLRQ YDPTAPIEDQPGIVRTRIVRHF DSAWELMFEN
3. Drechslera	82	1.65e-49	DDEPDNIAID MMISEELNGLNRDLRQ YDPTAPIEDQPGIVRTRIVRHF DSAWEMMFEN
2. Embellisia	441	1.73e-37	TTSRDTIAVE MFVSEELNGVSRDLSN YDPKRPITDQPGIVPTRMPRRF SSCWEMMFEN

the location of the motifs that are found in the sequences with p values (significance values)

MOTIF LOCATIONS

○ Only Motif Sites [?] ○ Motif Sites+Scanned Sites [?] ○ All Sequences [?]   [Download PDF] [?]   [Download SVG] [?]

Name [?]	p-value [?]	Motif Locations [?]
1. Curvularia	1.59e-111	
2. Embellisia	1.89e-93	
3. Drechslera	3.04e-112	
4. Nostoc	6.42e-39	
5. Deinococcus	1.03e-28	
6. Ascophyllum	1.53e-50	
7. Corallina	1.89e-37	
8. Fucus	3.14e-49	

INPUTS & SETTINGS

Sequences

Role	Source	Alphabet	Sequence Count	Total Size
Primary Sequences	msa.txt	Protein	8	3879

Background Model

Source: built from the (primary) sequences

Order: 0

	Name	Freq.	Bg.
A	Alanine	0.104	0.104
C	Cysteine	0.00722	0.00722
D	Aspartic acid	0.0693	0.0694
E	Glutamic acid	0.0611	0.0611
F	Phenylalanine	0.0534	0.0534
G	Glycine	0.0809	0.081
H	Histidine	0.0183	0.0183
I	Isoleucine	0.0526	0.0526
K	Lysine	0.0338	0.0338
L	Leucine	0.082	0.082
M	Methionine	0.0162	0.0162
N	Asparagine	0.048	0.048
P	Proline	0.0755	0.0755
Q	Glutamine	0.0351	0.0351
R	Arginine	0.0518	0.0518
S	Serine	0.0498	0.0498
T	Threonine	0.0596	0.0596
V	Valine	0.0606	0.0606
W	Tryptophan	0.0137	0.0137
Y	Tyrosine	0.0276	0.0276

max. number of motifs are 3

Other Settings

Motif Site Distribution	ZOOPS: Zero or one site per sequence
Objective Function	E-value of product of p-values
Starting Point Function	E-value of product of p-values
Site Strand Handling	This alphabet only has one strand
Maximum Number of Motifs	3
Motif E-value Threshold	no limit
Minimum Motif Width	6
Maximum Motif Width	50
Minimum Sites per Motif	2
Maximum Sites per Motif	8

Show Advanced Settings

MEME version
5.4.1 (Release date: Sat Aug 21 19:23:23 2021 -0700)

Reference
Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California, 1994. [full text]

Command line
```
meme msa.txt -protein -oc . -nostatus -time 14400 -mod zoops -nmotifs 3 -minw 6 -maxw 50 -objfun classic -markov_order 0
```

References

Wernersson, R., & Pedersen, A. G. (2003). RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic acids research, 31(13), 3537–3539. https://doi.org/10.1093/nar/gkg609

https://www.ebi.ac.uk/seqdb/confluence/display/THD/Help+-+Clustal+Omega+FAQ#HelpClustalOmegaFAQ-WhatisClustalOmega?

https://user.ceng.metu.edu.tr/~tcan/ceng465/Schedule/MSAComparison.pdf