

muscle

-in <inputfile>, 输入文件

-out <outfile>, 输出文件

-html, 输出HTML格式

- making a multiple sequence alignment(MSA)

多重序列比对

```
muscle -in seqs.fa -out seqs.afa
```

输出clustalw格式(比fasta格式更容易读)

```
muscle -in seqs.fa -clw
```

```
MUSCLE (3.8) multiple sequence alignment

gi|3650339|emb|AJ011501.1|:123-1      ATGATGAAGCGCAATATTCTGGCAGTGGTGATCCCTGCCCTGCTGGTAGCCGGTGCAGCC
gi|409684750|gb|JX291114.1|:c112    ---ATGAAAGTTAAAGTACTGTCCCTCCTGGTACCGGCTCTGCTGGTAGCAGGCGCAGCA
          *****      * * * * * * * * * * * * * * * * * * * * * *

gi|3650339|emb|AJ011501.1|:123-1      AACGCTGCAGAAATCTATAACAAAAACGGCAACAACTGGACTTCTATGGAAAAATGGTC
gi|409684750|gb|JX291114.1|:c112    AATGCGGCTGAAATTTATAACAAAGACGGCAACAAATTAGACCTGTACGGTAAAATTGAC
          ** * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

输出GCG MSF格式

```
muscle -in seqs.fa -msf
```

```
MSF: 1156  Type: N  Check: 0000  ..

Name: gi|3650339|emb|AJ011501.1|:123-1202  Len: 1156  Check: 9121  Weight: 0.5
Name: gi|409684750|gb|JX291114.1|:c1125-1  Len: 1156  Check: 1552  Weight: 0.5

//

gi|3650339|emb|AJ011501.1|:123-1202      ATGATGAAGC  GCAATATTCT  GGCAGTGGTG  ATCCCTGCCC  TGCTGGTAGC
gi|409684750|gb|JX291114.1|:c1125-1    ...ATGAAAG  TTAAAGTACT  GTCCCTCCTG  GTACCGGCTC  TGCTGGTAGC

gi|3650339|emb|AJ011501.1|:123-1202      CGGTGCAGCC  AACGCTGCAG  AAATCTATAA  CAAAAACGGC  AACAACTGG
gi|409684750|gb|JX291114.1|:c1125-1    AGGCGCAGCA  AATGCGGCTG  AAATTTATAA  CAAAGACGGC  AACAAATTAG
```

clustalo

使用Clustal Omega程序用于蛋白序列比对

-i, --in, --infile={<file>, -} 多重序列文件输入

-o, --out, --outfile={file, -}多重序列比对输出

--hmm-in=<file>, 指定输入为HMMs文件

-t, --seqtype= {Protein, RNA, DNA}, 指定输入序列类型

--outfmt={a2m=fa, clu[stal], msf, phy[lip], st[ockholm], vie: MSA}, 输出我呢就爱你格式, 默认为fasta

```
clustalo -i seqs.prot.fa --outfmt clu
```

```
CLUSTAL O(1.2.3) multiple sequence alignment
```

```
gi|3650339|emb|AJ011501.1|:123-1202      MMKRNILAVVIPALLVAGAANA AEIYNKNGNKLD FYGKMVGEHVWTTNGDTSSDDTTYAR
gi|409684750|gb|JX291114.1|:c1125-1      -MKVKVLSLLVPALLVAGAANA AEIYNKDG NKLDLYGKIDGLHYFSDDK-SVDGDQTYMR
      ** :::::::::::*****:*****:***: * * :: : ..* ** *

gi|3650339|emb|AJ011501.1|:123-1202      IGLKGETQINDQLIGYGQWEYNMDASNVEGS-QTTKTRLAFAGLKAGEYGSFDYGRNYGA
gi|409684750|gb|JX291114.1|:c1125-1      VGVKGETQINDQLTGYGQWEYNVQANNTES SSDQAWTRLAFAGLKFGDAGSFDYGRNYGV
      :*:***** *****:*.*.*. * : : ***** *: *****.
```

```
clustalo -i seqs.prot.fa --outfmt msf
```

```
Name: gi_3650339_emb_AJ011501_1__123-1202 Len: 381 Check: 8282 Weight: -1.00
Name: gi_409684750_gb_JX291114_1__c1125-1 Len: 381 Check: 7628 Weight: -1.00

//

              1                               50
gi_3650339_emb_AJ011501_1__123-1202 MMKRNILAVV IPALLVAGAA NAAEIYNKNG NKLD FYGKMV GEHVWTTNGD
gi_409684750_gb_JX291114_1__c1125-1 ~MKVKVLSLL VPALLVAGAA NAAEIYNKDG NKLDLYGKID GLHYFSDDK.

              51                               100
gi_3650339_emb_AJ011501_1__123-1202 TSSDDTTYAR IGLKGETQIN DQLIGYGQWE YNMDASNVEG S.QTTKTRLA
gi_409684750_gb_JX291114_1__c1125-1 SVDGDQTYMR VGVKGETQIN DQLTGYGQWE YNVQANNTES SSDQAWTRLA
```

```
clustalo -i seqs.prot.fa --outfmt vie
```

```
>gi|3650339|emb|AJ011501.1|:123-1202 Klebsiella pneumoniae (strain K1755) ompK35 gene
MMKRNILAVVIPALLVAGAANA AEIYNKNGNKLD FYGKMVGEHVWTTNGDTSSDDTTYARIGLKGETQINDQLIGYGQWEYNMDASNVEGS-QTTKTRLAFAGLKAGEYGSFDYGRNYGAIYDVEAATDMLVEWGGDGWNYT-DNYMTGRTNGVATYRNSDFFGLVDGLSFALQYQGKNDH
-----DRAIRKQNGDGFSTAATYAFDNGIALSAGYSSSNRSVDQ--KADGNGDKAEANATSAKYDANNIYAAMYSQTYNMTPEE-----DNHFAGKTQNF EAVVQYQDFGLRPSIGYVQTKGKDLQSRAGFSGGDADLVKYIEVG TWYFFNKNMNVYAAYKFNQLDNDYT
KAAGVATDDQAAVGIVYQF
>gi|409684750|gb|JX291114.1|:c1125-1 Klebsiella pneumoniae strain H5062105 OmpK36 porin (ompk36) gene, complete cds
-MKVKVL SLLVPALLVAGAANA AEIYNKDG NKLDLYGKIDGLHYFSDDK-SVDGDQTYMRVGVKGETQINDQLTGYGQWEYNVQANNTES SSDQANTR LAFAGLKFGDAGSFDYGRNYGVVYDVTSWTDVLP EFGGDDGT YGSDNFLQSRANGVATYRNSDFFGLVDGLNFALQYQGKNGS
VSSEGTSP TNNRGALKQNGDGF GTS LTYDIYDGISAGFAYSHSKRNGDQNR LDKGRGDAET YTGGLKYDANNIYLATQYTQTYNATRFSGNGESDSISGFANKAQNF EAVVQYQDFGLRPSVAYLQSKGKDIEG-----YGDQDLLKYVDVGATYFFNKNMSTYVDYKINLLDENDFT
RRAGISTDDVVALGLVYQF
```

指定hmm输入，尚未理解，略！！

```
clustalo -i seqs.prot.fa --outfmt sto > seqs.prot.sto
```

```
hmmbuild seqs.hmms seqs.prot.sto
```

```
clustalo -i seqs.prot.fa --hmm-in=seqs.hmms
```

```
>gi|3650339|emb|AJ011501.1|:123-1202 Klebsiella pneumoniae (strain K1755) ompK35 gene
MMKRNILAVVIPALLVAGAANA AEIYNKNGNKLD FYGKMVGEHVWTTNGDTSSDDTTYAR
IGLKGETQINDQLIGYGQWEYNMDASNVEGS-QTTKTRLAFAGLKAGEYGSFDYGRNYGA
IYDVEAATDMLVEWGGDGWNYT-DNYMTGRTNGVATYRNSDFFGLVDGLSFALQYQGKND
H-----DRAIRKQNGDGFSTAATYAFDNGIALSAGYSSSNRSVDQK--ADGNGD
KAEAWATSAKYDANNIYAAMYSQTYNMTPEE-----DNHFAGKTQNF EAVVQYQDFD
GLRPSIGYVQTKGKDLQSRAGFSGGDADLVKYIEVG TWYFFNKNMNVYAAYKFNQLDND
YTAKAGVATDDQAAVGIVYQF
>gi|409684750|gb|JX291114.1|:c1125-1 Klebsiella pneumoniae strain H5062105 OmpK36 porin (ompk36) gene, complete cds
-MKVKVL SLLVPALLVAGAANA AEIYNKDG NKLDLYGKIDGLHYFSDDK-SVDGDQTYMR
VGVKGETQINDQLTGYGQWEYNVQANNTES SSDQAWTRLAFAGLKFGDAGSFDYGRNYGV
VYDVTSWTDVLP EFGGDDGT YGSDNFLQSRANGVATYRNSDFFGLVDGLNFALQYQGKNG
SVSSEGTSP TNNRGALKQNGDGF GTS LTYDIYDGISAGFAYSHSKRNGDQNR LDKGRGD
NAET YTGGLKYDANNIYLATQYTQTYNATRFSGNGESDSISGFANKAQNF EAVVQYQDFD
GLRPSVAYLQSKGKDIEG-----YGDQDLLKYVDVGATYFFNKNMSTYVDYKINLLDEND
FTDRAGISTDDVVALGLVYQF
```