

实验三：两条序列比对与多序列比对

实验目的：

学会使用 MegAlign, ClustalX 和 MUSCLE 进行两条序列和多条序列比对分析

实验内容：

双序列比对是使两条序列产生最高相似性得分的序列排列方式和空格插入方式。两条序列比对是生物信息学最基础的研究手段。第一次实验我们用 dotplot 方法直观地认识了两条序列比对。但是 dotplot 仅仅是展示了两条序列中所有可能的配对，并不是真正意义上的序列比对。这里介绍进行两条序列比对的软件-MegAlign。

多序列比对是将多条序列同时比对，使尽可能多的相同（或相似）字符出现在同一列中。多序列比对的目的是发现多条序列的共性。如果说序列两两比对主要用于建立两条序列的**同源关系，从而推测它们的结构和功能**，那么，同时比对多条序列对于研究**分子结构、功能及进化关系更为有用**。多序列比对对于**系统发育分析、蛋白质家族成员鉴定、蛋白质结构预测、保守模块的搜寻等具有非常重要的作用**。我们这节课主要学习多条序列比对的软件-ClustalX, MUSCLE。

一、MegAlign

DNASTAR 公司的 Lasergene 软件包是一个比较全面的生物信息学软件，它包含了 7 个模块。其中 MegAlign 可进行两条或多条序列比对分析。

1. 两条序列比对

1.1 安装程序

解压 DNASTAR Lasergene 软件压缩包，双击 Lasergene710WinInstall.exe 文件，按照默认路径安装软件到自己电脑上。

1.2 载入序列

a. 点击 **开始—程序—Lasergene—MegAlign**，打开软件。

我们首先用演示序列（demo sequence）学习软件的使用。演示序列所在位置：**C:\Program files\DNASTAR\Lasergene\ Demo Megalign\ Histone Sequences**。

b. 点击主菜单 **File—Enter sequence—选择序列所在文件夹**，选择序列 **tethis21.seq** 和 **tethis22.seq**，点击 **Add**，这两条序列将出现在右侧 **selected sequences** 框中（Figure 3.1），选择完毕点击 **Done** 回到程序页面。

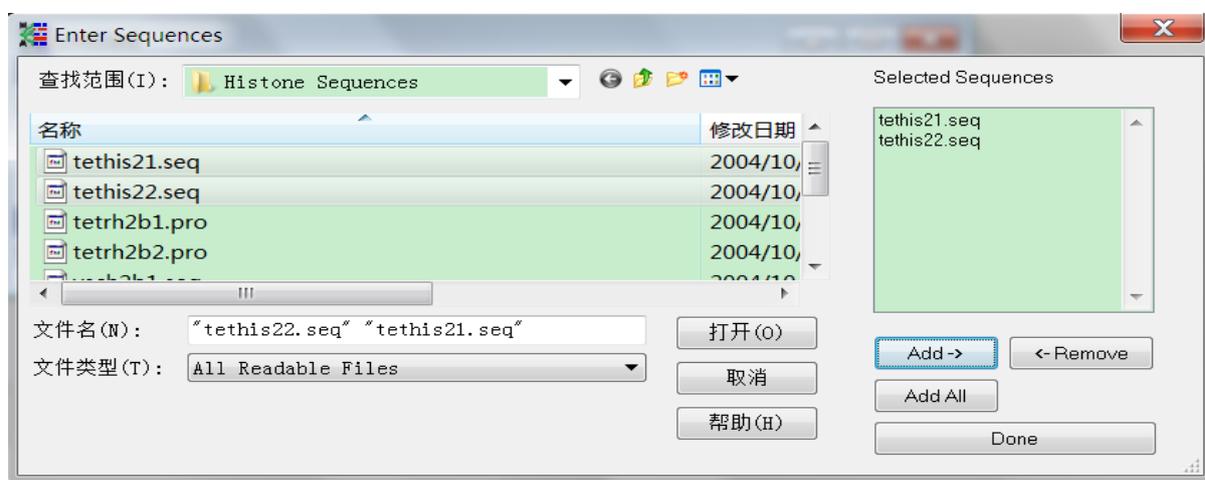


Figure 3.1 载入序列

此时程序窗口分为三部分，最左侧较窄的是 sequence name，中间显示的是序列起始位置，最右侧显示序列末尾部分，可以通过拖动窗口底部滚动条，查看序列其它部分（Figure 3.2）。若想改变字体显示方式，点击主菜单 OPTIONS，选择 Font 改变字体，选择 Size 改变字号大小。若要移除序列，选中 sequence name 的序列名，右击，选 clear。

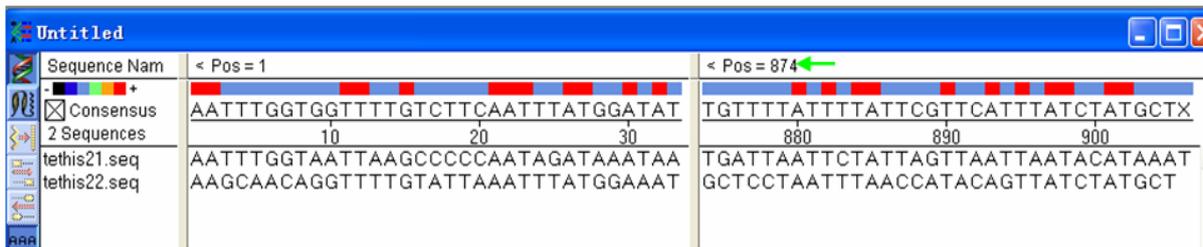


Figure 3.2 载入序列后（注意标注的绿色箭头，即为坐标位置）

1.3 设定序列比对位置

MegAlign 允许使用者选择序列的一部分进行比对分析，例如，可以根据 GenBank 格式的序列中 Features 部分关于编码区（CDS）位置的描述，设定只对此编码区进行分析。

a. 点击最左侧 Sequence Name 框中的第一条序列 tethis，然后选择主菜单 OPTIONS—Set sequence limits—**from feature table**。（Figure 3.3）此时根据 feature 内容，出现四个可以选择的片段，第一个为全长，从序列起始到末尾（1-906），其它三个则只包括序列的一部分，选择最后一个 Histone H2B-1—CDS，点击 Change the Reset，点击 OK，同样对第二条序列进行上述操作，回到主界面工作区，此时窗口中的序列起始和终止位置已经发生了变化。（Figure 3.4）

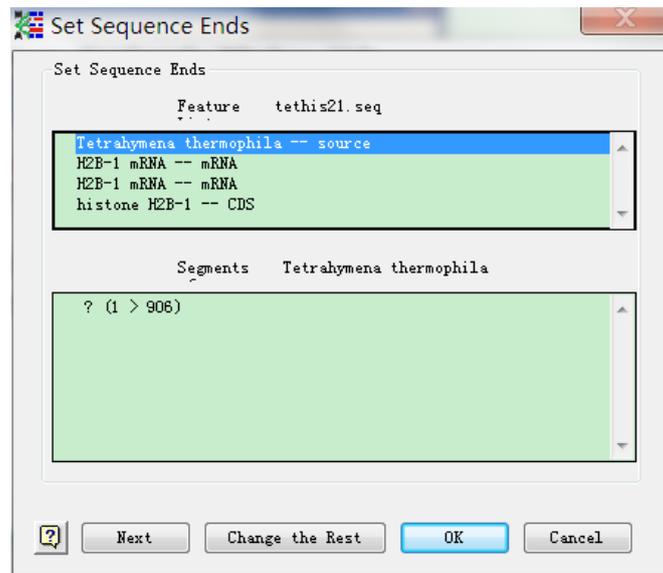


Figure 3.3 利用 Feature Table 选择序列特定部分

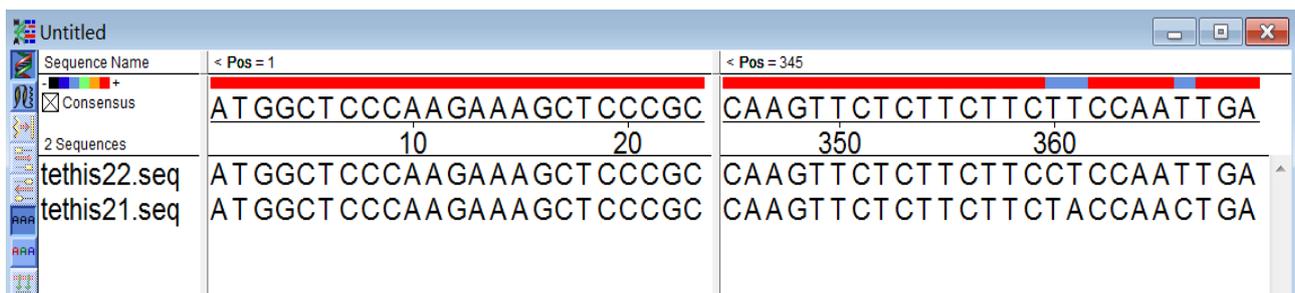


Figure 3.4 选择序列特定部分

b. 我们还可以通过设定序列坐标进行部分序列比对，首先选定序列，选择主菜单

OPTIONS—Set sequence limits—by coordinates, 输入起始和终止位置坐标来选择部分序列进行分析。

注意: 只有 genbank 格式的序列才可以 Set sequence limits from feature table, fasta 格式的序列因为没有 feature 那一项内容, 只可以 Set sequence limits by coordinates。

1.4 进行两条序列比对

如果输入两条序列后不设置序列起始和终止位置, 默认是全长序列进行比对。

按住Shift选择序列tethis21和tethis22, 然后点击主菜单Align-One pair, 由于目前输入的是核酸序列, 此时有两个选项, Wilbur-Lipman Method和Martiner NW Method。如果输入的是蛋白质序列, 这两个选项将是灰色, 只能用Lipman-Pearson Method进行比对。Wilbur-Lipman Method是一种以word为单位的 (word-based) 启发式局部比对方法; Martiner NW Method是一种改进了的全局动态规划算法。Lipman-Pearson Method是序列相似度搜索软件Fasta的比对算法, 也是一种以word为单位的快速启发式算法。选择其中一个, 出现比对参数设定窗口 (Figure 3.5), 选择默认参数不做更改, 直接点击OK即可。

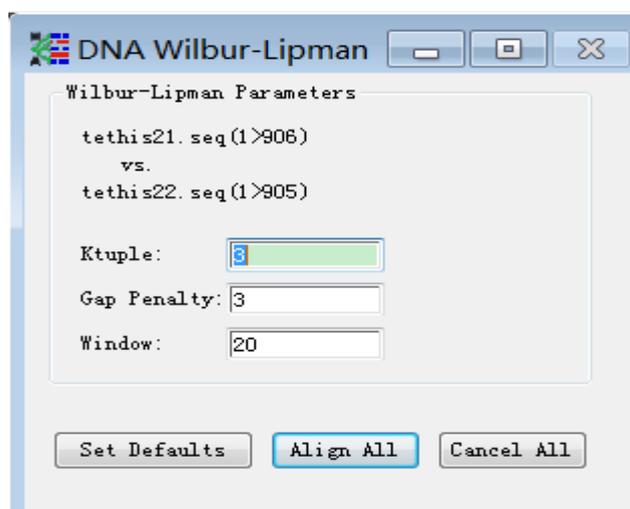


Figure 3.5 Wilbur-Lipman 比对方法参数设定

这时出现一个新窗口, 即为比对结果。可以选择OPTION-size, 放大字号观察比对结果。可以看到在窗口上部显示的是**比对方法名称**, **所用参数**, 两条**序列各自的起止位置**, **相似度值**, **比对结果中空位数目**, **长度和一致序列的长度**。随后就是比对结果部分, 其中第一行是第一条序列, 它上面**的v70**是标尺, 其中的“V”的位置对应的是第一条序列的第70个核苷酸所在位置; 第三行是第二条序列, 它下方的数字同样对应该序列位置坐标; 中间那行是根据两条序列比对结果中匹配部分推断出来的一致序列 (consensus sequence), 错配或空位显示为空白 (Figure 3.6)。

```
          v70          v80          v90          v100          v110
TTATCCAATCAGAACGCAGAAAAAGAAGATTATCCAATCAGATGCTTTTT
TTATCCAATCA A C CAGAAA AG      TA      AAT          TTT
TTATCCAATCAAATCACAGAAATAGGTAGATACAGAAT-----TTTA
          ^100          ^110          ^120          ^130
          v120          v130          v140          v150          v160
TAATAGAAGATACAGATAACTGGAAGAGATAGAAAAAAGAAAAGGATAA
ATA  AGA                                AAG AA      AA
AGATATTAGA-----AAGTAATTTTAAA
          ^140                                ^150
```

Figure 3.6 Wilbur-Lipman 方法比对结果

设置比对结果显示方式： 点击比对结果窗口最左侧的  按钮，出现 Alignment View Options 窗口，可以选择匹配，错配和一致序列的字符颜色和其它显示选项。**推荐使用设置：** 选择 match 为红色，mismatch 为绿色，consensus 为蓝色，并选择 show identities as vertical bars（一致序列显示为竖线），则得到 Figure 3.7。还可以尝试选中或不选 show header, show ruler, show names, show contest 四个选项，看看显示结果有何变化。

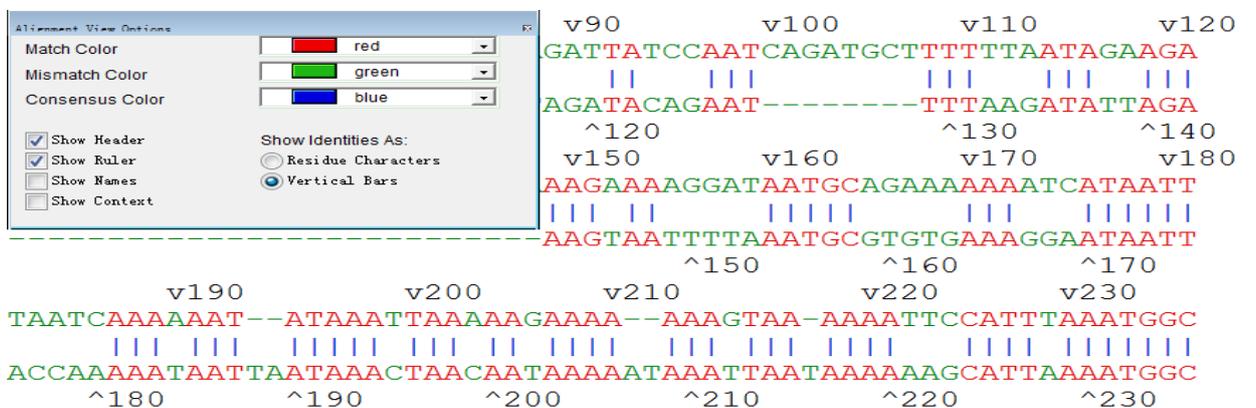


Figure 3.7 Alignment View Options

TIP: MegAlign 分析自己下载的序列时要注意序列扩展名

如果是从 NCBI 直接下载的 fasta 格式文件，可以象上面一样，用 enter sequence 直接将序列读入程序。但是如果序列文件是复制粘贴到 txt 文档中的，MegAlign 程序是无法识别扩展名为 txt 的文件。此时可将每条序列文件（fasta 或 genbank 格式皆可）扩展名改为 MegAlign 可以识别的类型(核酸序列为 seq, 蛋白质序列为 pro), 即可从 File-Enter sequence 载入。

更改文件扩展名的方法： 找到你要更改扩展名的文件，将 .txt 改为 .seq 或 .pro，此时会弹窗口，提示“如果改变文件扩展名，可能会导致文件不可用。确实要更改吗？”选择“是”，文件图标会变成 MegAlign 特定图标，说明修改成功。若扩展名自动隐藏，打开文件夹，点击窗口上的主菜单 工具—文件夹选项，在打开的页面选择选项卡查看，去掉“隐藏已知文件类型的扩展名”前面的对勾，确定退出。然后再用上述方法更改扩展名。

2. 多序列比对

2.1 载入序列

进行多条序列比对的演示序列(demo sequence)在 c:\program files\dnastar\ lasergene\ demo megalign\ Calmodulin Sequences\ 文件夹里。

点击主菜单 File-Enter Sequence-根据路径到达 Calmodulin Sequences 文件夹，点击 Add All，此时 14 条序列全都出现在右侧的 selected sequences 框中，点击 Done，回到主程序工作区。(Figure 3.8) 这是来自 14 个物种的钙调蛋白。

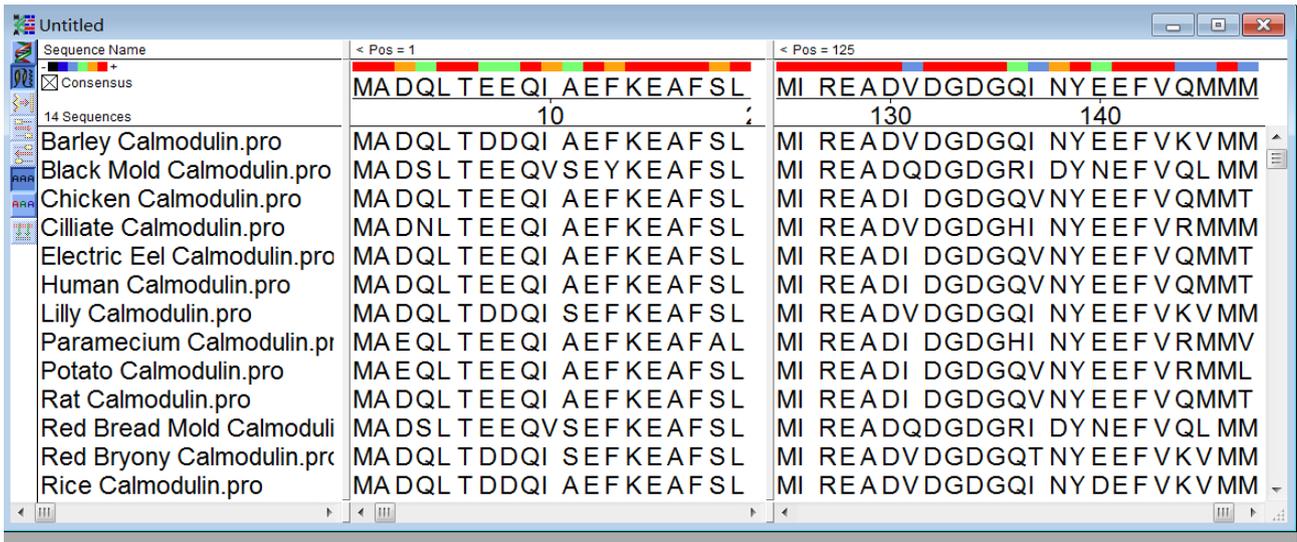


Figure 3.8 载入 14 条序列

2.2 序列比对

第一步，选择比对所用的打分矩阵。点击主菜单 Align—Set residue Weight Table，由于钙调蛋白比较保守，我们选择 PAM100 作为打分矩阵，点击 OK 结束设定(Figure 3.9)。

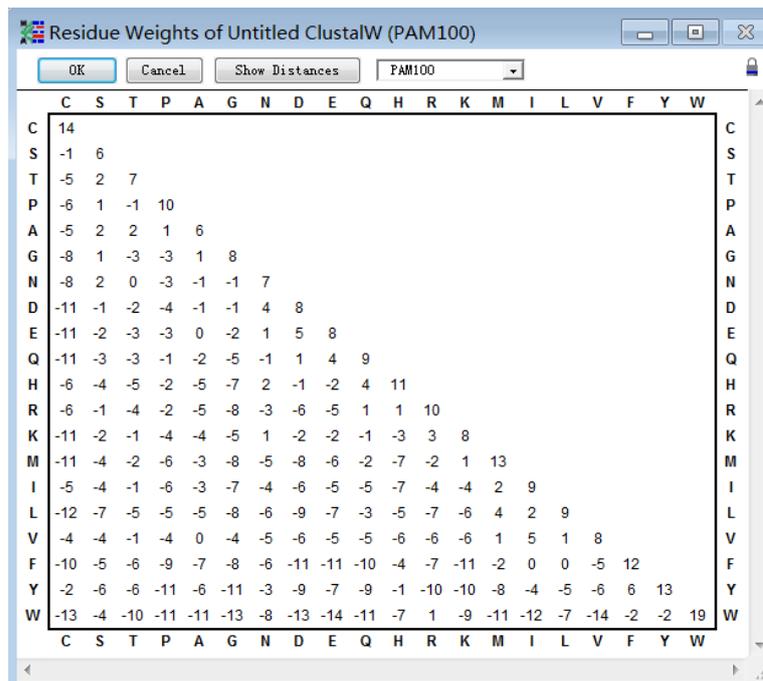


Figure 3.9 选择打分矩阵

此时还可以通过点击 Align-Method Parameters 设定比对所用的其它参数。打开的新窗口中包含三个选项卡，Jotun Hein、Clustal V 和 Clustal W，对应程序中多条序列比对可用的三种算法。推荐大家不做修改，使用默认参数即可。

第二步，比对。点击 Align-by Clustal V Method，此时出现窗口显示比对进度，比对结束后，回到原来工作窗口，显示比对结果。注意序列上方彩色条块，颜色代表对应列中相似程度，相似度由低到高，依次以深蓝、浅蓝、绿、黄、桔、红几种颜色代表。(Figure 3.10)

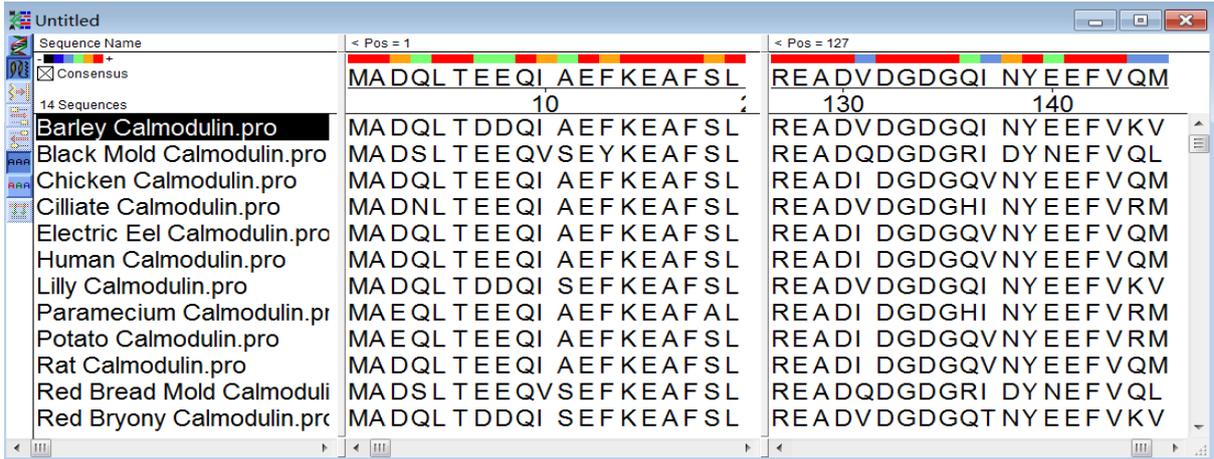


Figure 3.10 比对后结果

2.3 查看比对结果

此时可以通过几种方式观察比对结果。

- 点击 View-Sequence Distances 出现新窗口，显示两两序列 percent identity(上半部分)和 divergence (下半部分)。

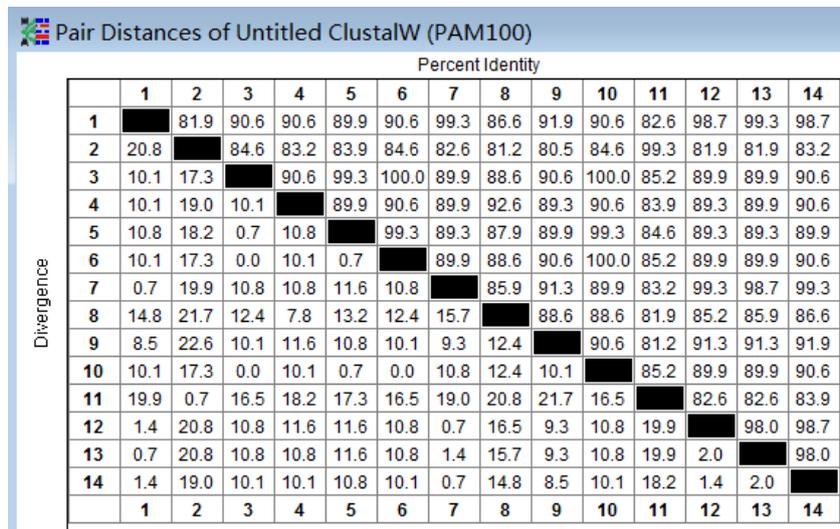


Figure 3.11 比对结果——一致度 (identity)

- 点击 View-Residue Substitutions 出现新窗口，显示比对中所有替换的类型和数目。

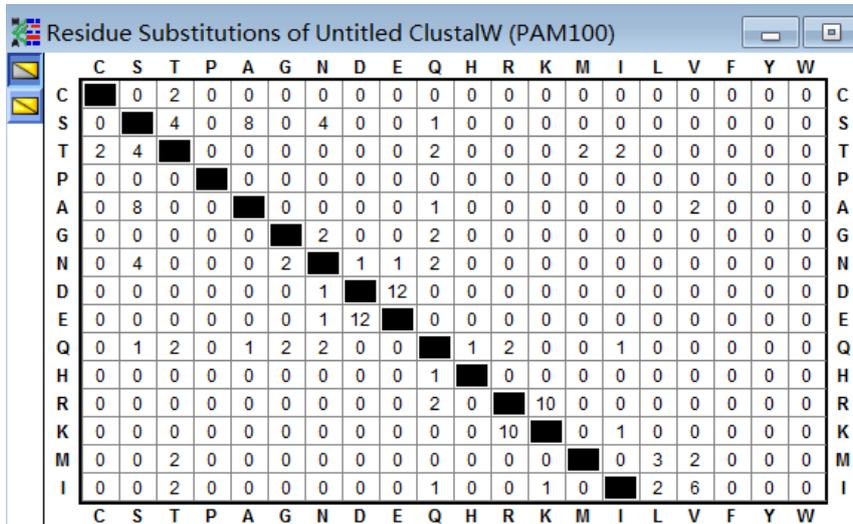


Figure 3.12 比对结果——替换情况

c. 点击 View-Phylogenetic Tree 出现新窗口，显示根据 14 条序列比对结果构建出的进化树。

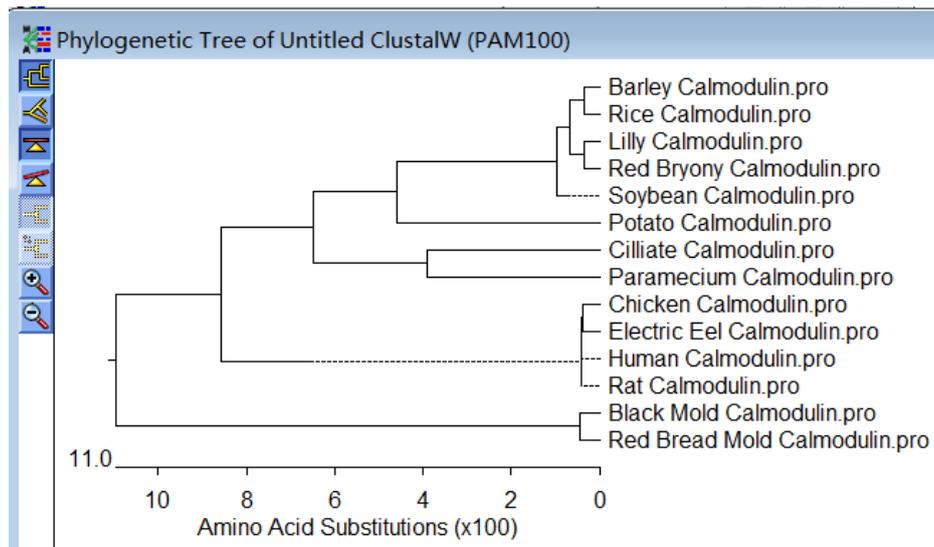


Figure 3.13 比对结果—进化树

d. 点击 View-Alignment Reports 出现新窗口，显示比对结果报告。点击 OPTIONS-Alignment report contents，选中 show consensus strength，其它不变，点击 OK。在序列上方出现条块，显示每一列序列的相似程度。

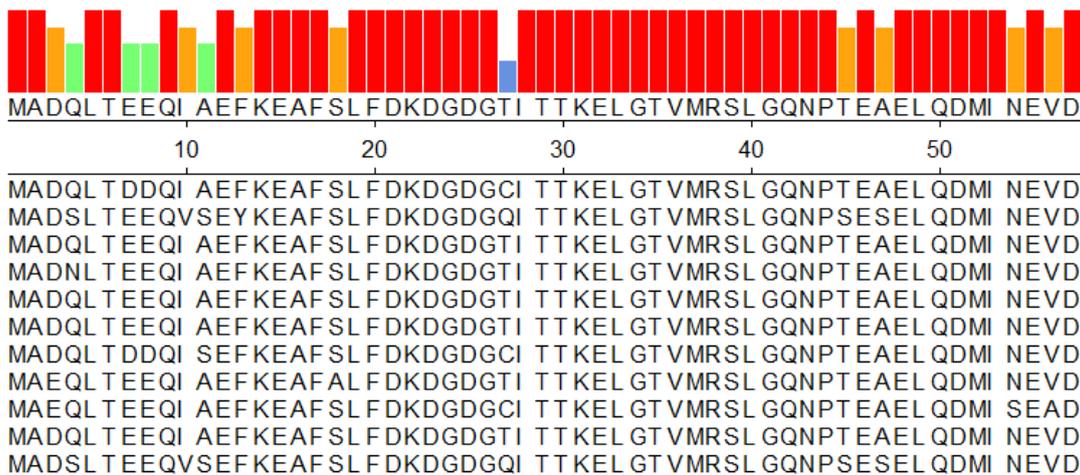


Figure 3.14 选择 show consensus strength 显示结果

设置比对结果显示方式：突出显示匹配或错配的氨基酸。点击 OPTIONS-New Decorations，在 alignment decoration name 框里输入 shade disagreements(自己定义名字)，选择 decoration parameters 为 shade—residues differing from—the consensus，此时下方出现新的选项，选择对选定字符突出显示的颜色，选择完毕，点击 OK，则与 majority 序列不同的字符将突出显示。(Figure 3.15)

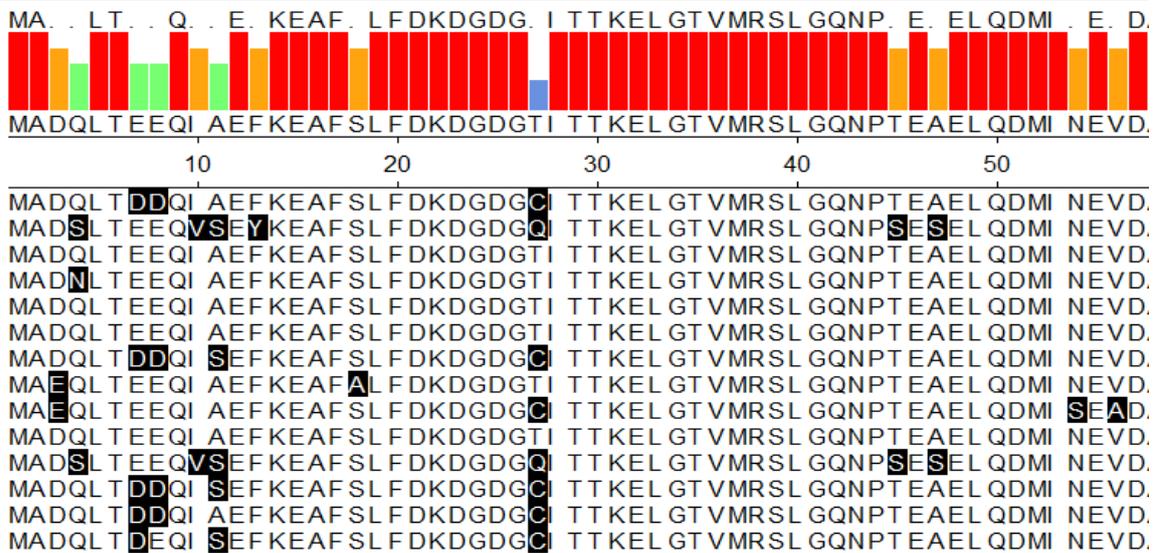
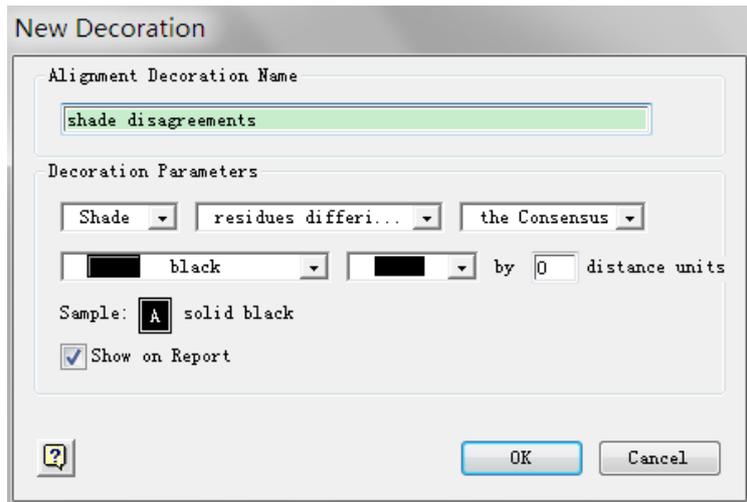


Figure 3.15 修改 alignment report 显示模式

二、Clustalx

<http://www.clustal.org/>

Clustal 是一种利用渐近法 (progressive alignment) 进行多条序列比对的软件。即从多条序列中最相似 (距离最近) 的两条序列开始比对, 按照各个序列在进化树上的位置, 由近及远的将其它序列依次加入到最终的比对结果。(Figure 3.16)

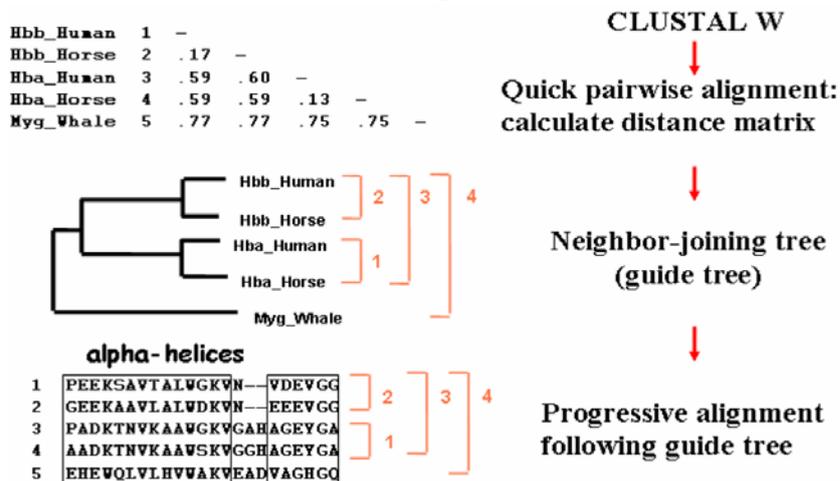


Figure 3.16 clustal 算法

Clustal 软件有两个版本，其中 clustalw 采用命令行的形式在 DOS 下运行的。Clustalx 是可视化界面的程序，我们今天学习 Clustalx 的使用。

2.1 安装 clustalx

下载 clustalx 软件，按照默认安装到自己的电脑上。

2.2 准备要比对的序列

将上节课搜索到的同源核酸fasta文件，全部粘贴到一个文本文件中，所有的蛋白质序列存入另一个文本文件。注意序列的登录号最好是以NM、NP、NR开头，不要使用NC、NT或NW开头的序列，因为全基因组序列太长，分析起来速度非常慢。

TIP:可以在fasta序列“>”之后加上物种名称，加空位，方便看树时了解进化关系。

2.3 载入序列

点击 开始—程序—clustalX2—clustalX2。

点主菜单 File，选择 Load Sequence—选择刚保存的序列文件，点打开。

注意: ClustalX 程序无法识别汉字、带空位的文件夹名，如 my document。不要将序列文件保存在桌面上或带汉字的文件夹中，推荐保存在 D 盘根目录下。

载入序列后在左侧窗口里是 fasta 格式序列的标识号，取自序列第一行“>”后的字符。

(Figure 3.17)

TIP: 如果每条序列单独保存为一个文件，可以使用 File—Append sequence 选项将序列一条条添加进来。

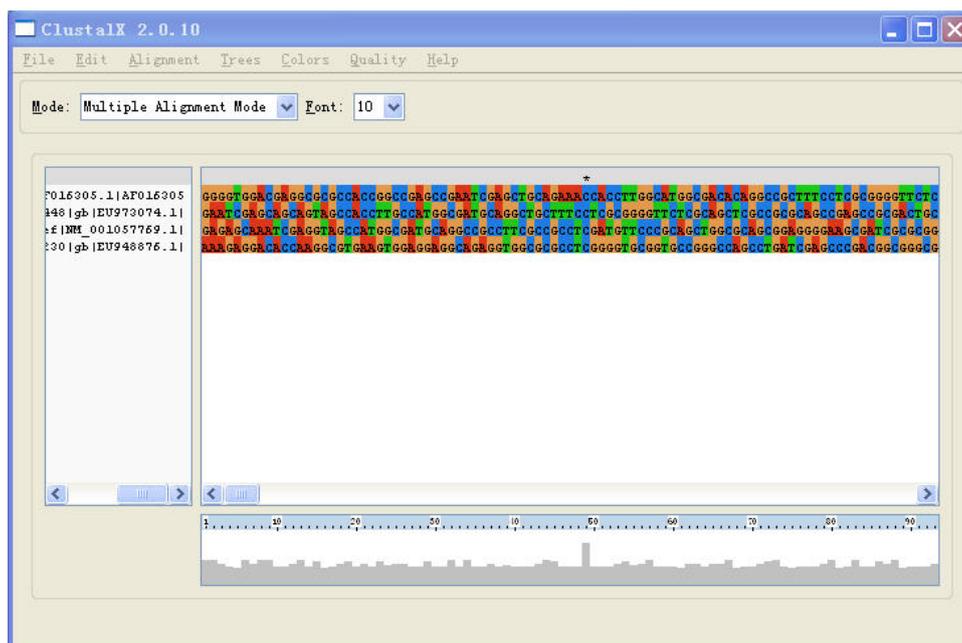


Figure 3.17 载入序列

2.4 比对参数的选择

比对前要先设置两条序列比对的参数和多条序列比对的参数。

a. 两条序列比对的参数

点击 Alignment 菜单，选择 Alignment Parameters, 再选择 Pairwise Alignment Parameters, 如 Figure 3.18. 首先可以选择比对的效果，是 slow/accurate 还是 fast/approximate。第一种模式采用的是动态规划算法进行比对的，第二种模式采用的是启发式的算法。除非序列非常长，一般采用第一种模式。可以选择空位罚分系统，DNA 或蛋白质替换矩阵，也可以自己上传某个替换矩阵进行比对。

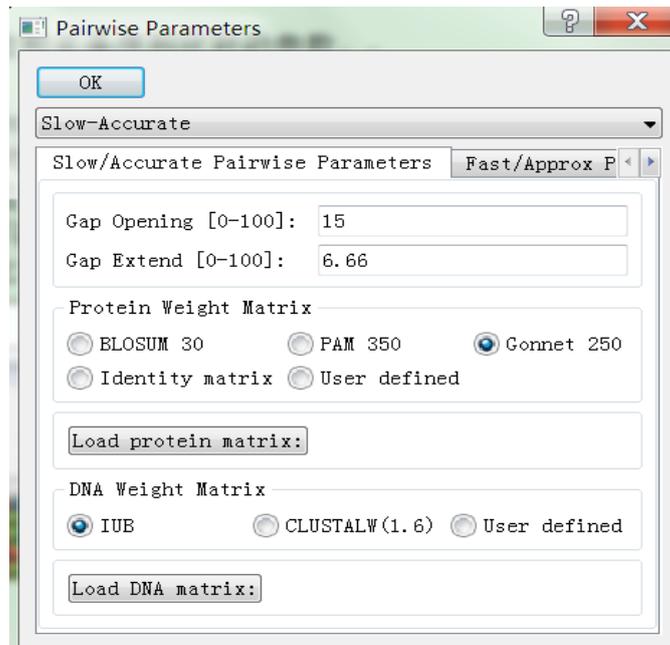


Figure 3.18 Pairwise Alignment Parameters

b. **多条序列** 比对参数

点击 Alignment 菜单，选择 Alignment Parameters, 再选择 Multiple Alignment Parameters, 如 Figure 3.19.

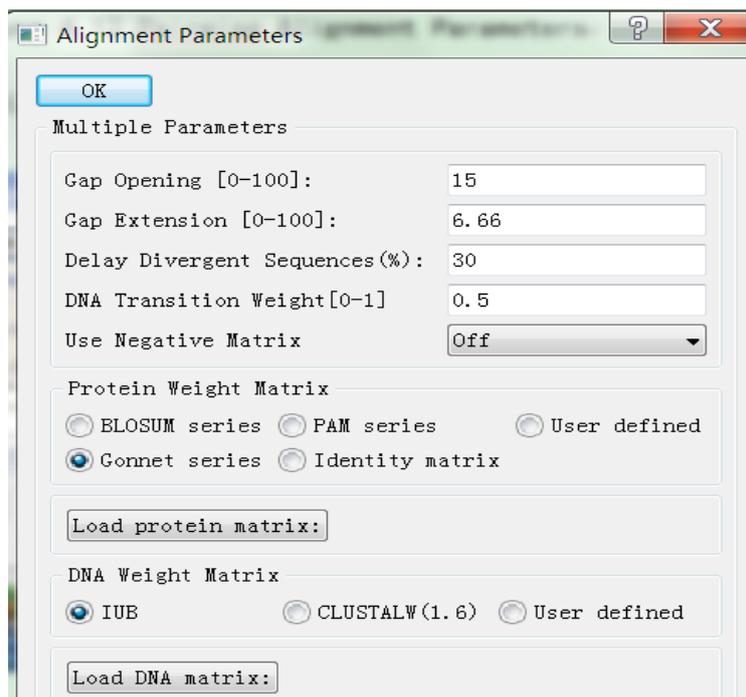


Figure 3.19 Multiple Alignment Parameters

Delay divergent sequence 是指当两条序列的差异大于某个值（百分比）时，这两条序列的比对将推迟进行，程序先比对相似序列，对于相似度不够高的序列，晚些时候进行比对，加入到最终的多条序列比对结果时也要迟些。**DNA transition Weight** 等于 0 的时候，程序将转换当作错配（mismatch）看待，等于 1 的时候，将转换和颠换同等看待。当参与比对的序列差异较大时，DNA transition Weight 应该选择的小些（接近 0），如果参与比对的序列差异较小时，DNA transition Weight 可选择的大些（接近 1）。

2.5 更改输出格式

点击 Alignment 菜单，选择 Output Format Options, 页面如 Figure 3.20 。

默认的是输出 clustal format, 如果需要其它格式, 可在复选框里打勾。如 PHYLIP 格式是利用 PHYLIP 软件进行建树时, 需要输入的格式 (这里两种格式都选上, 以备下节课构建系统发育树使用)。

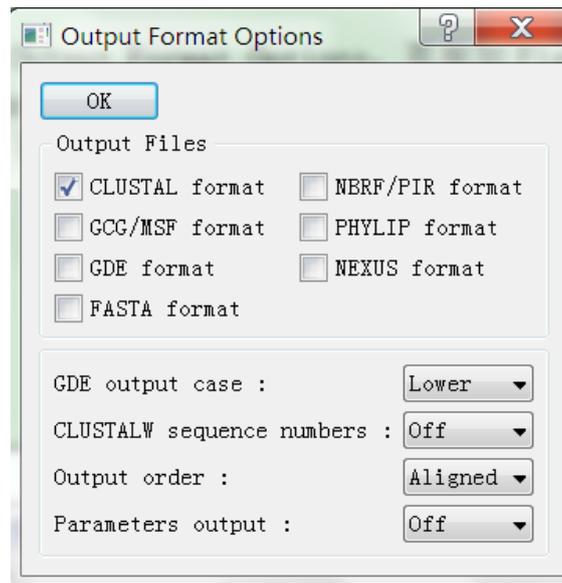


Figure 3.20 输出格式选项

2.6 进行比对

点击 Alignment 菜单, 选择 Do Complete Alignment. 此时出现一个对话框, 提示比对结果保存的位置, 上一步选择了多少种输出格式, 这里就需要给出多少个文件的路径。选择好了点 OK 即可。

要得到理想的比对结果, 你可能需要选择不同的参数, 进行多次比对, 最后再对各种比对结果进行分析, 选择哪个是最合理的结果 (result making biological sense)。

比对结束后生成的 aln 文件是多条序列比对的结果, 可以用写字板打开浏览 (Figure 3.21)。在某一列比对结果下方如果出现 “*”, 说明这列是完全匹配。生成的 dnd 文件是比对过程中利用 NJ 方法生成的进化树, 可以用 treeview 程序浏览。treeview 软件安装和使用见第 2.8 部分内容。

```

human      AAGCATCCCAGAAGACTGCGCCATGGGGC----TCAGCGACGGGGAATGG---CAGTTGG
Troglodyte TGACATCCCAGGCCA-TGGGCAGGAAGTCCTCATCAGGTAAAAGGAAGAGATTCCATTGC
canis      TGTCTTCTTCAGCC--TGTACCATGGGGC----TCAGCGACGGGGAATGG---CAGTTGG
macaca     -----
mouse     TGTCCTCTTTAGAAG--CCACCATGGGGC----TCAGTGATGGGGAGTGG---CAGCTGG

human      TGCTGA-ACGTCTGGGGGAAGGTGGAGGCTGACATCCCAGGCCATGGGCAGGAAGTCCTC
Troglodyte CCCTGCCACCCACACCCTAAGATCAAGGGTG----TTCAGCTGCAAGGTGGAAAGTTTGC
canis      TGCTGA-ACATCTGGGGGAAGGTGGAGACTGACCTGGCGGGCCATGGGCAGGAGGTCCTC
macaca     -----ATGGCAGTTGGTGCTGAACGTC-TGGG--GGAAGTCCTC
mouse     TGCTGA-ATGTCTGGGGGAAGGTGGAGGCCGACCTTGCTGGCCATGGACAGGAAGTCCTC
                *      *                      *  *  *  *  *

human      A--TCAGGCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACC
Troglodyte ACGTGGGGCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACC
canis      A--TCAGGCTCTTTAAGAACCACCCGAGACCCTGGATAAGTTTCGACAAGTTCAAGCACC
macaca     A--TCAGGCTCTTTAAGGGTCACCCGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACC
mouse     A--TCGGTCTGTTTAAGACTCACCCGAGACCCTGGATAAGTTTGACAAGTTCAAGAACT
                *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
    
```

Figure 3.21 生成的 aln 文件

2.7 迭代比对

可以采用迭代选项，多次迭代来寻找最佳比对结果。

点击 Alignment 菜单，选择 iteration, 选择 iterate each alignment step 或 iterate final alignment.

然后再点击 Alignment 菜单，选择 Do Complete Alignment 进行比对，即可达到迭代的效果，将没有利用迭代比对得到的结果与迭代后的结果进行比较，看是否存在差异。

其它不详之处请参考 clustalx.pdf 文件。

2.8 Treeview

下载地址: <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>

Treeview 是一个专门绘制和浏览进化树的软件。Clustalx 产生的进化树(即后缀为 dnd 文件)，可以通过 treeview 软件浏览。

双击 treeview_setup.exe 文件按照默认将程序安装到电脑上。

双击后缀为 dnd 的文件，选择 treeview 程序打开即可。也可以打开 treeview 软件，将 dnd 文件拖放到 treeview 软件窗口里打开。

TIP: 由于电脑中没有安装打印机导致 treevie 无法正常显示解决方案。

我的电脑右击—管理—服务和应用程序—服务—printspooler-点右键启动。

如果仍有问题，点击开始—设备和打印机—选添加打印机—按照向导提示任意添加一个打印机驱动即可。

三、 MUSCLE

MUSCLE 是一款非常好用的多序列比对软件，它的速度不输于 Clustal, 但精确度要高的多。

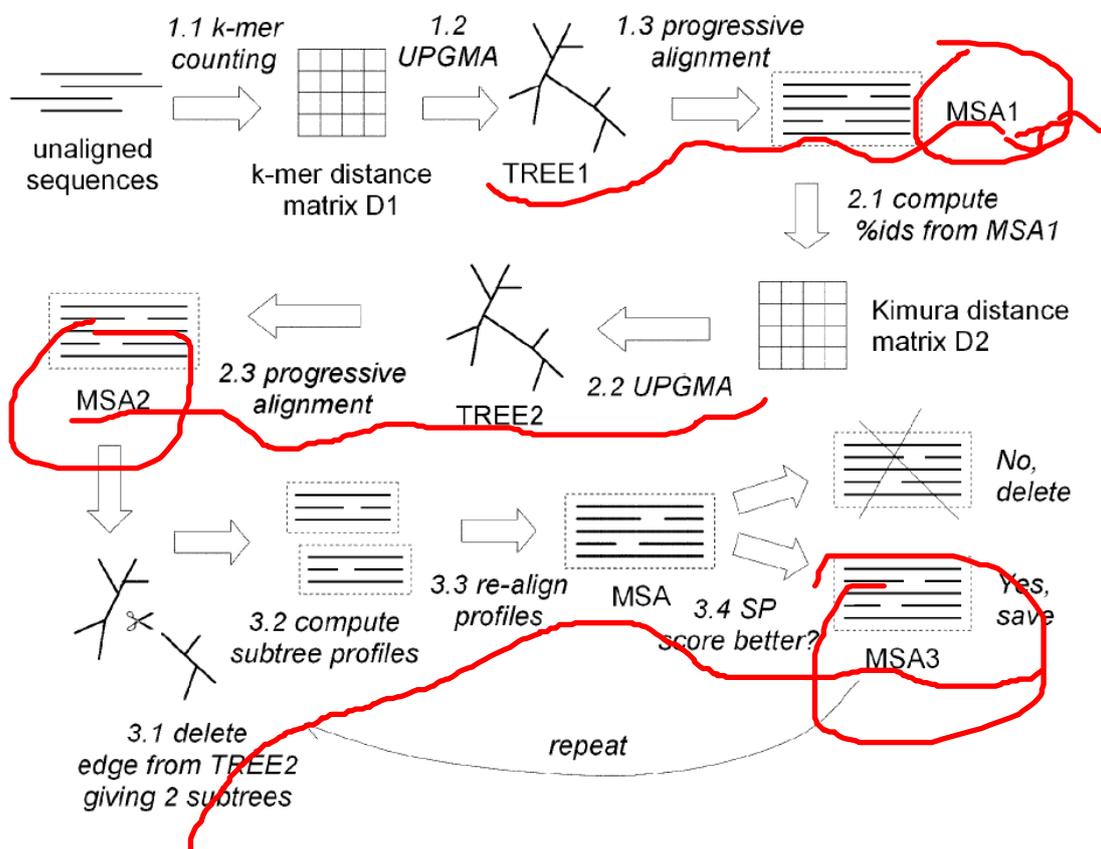


Figure 3.22 MUSCLE 算法

MUSCLE 的算法分为三步：首先计算两两序列共有的短片段（k-mer）数量，以此为基础构建初步引导树（TREE1），参照引导树，采用渐进算法得到多序列比对结果 MSA1；然后，根据 MSA1 计算两两序列的距离，根据新的距离矩阵，构建更精确的引导树（TREE2），比较 TREE2 和 TREE1，将发生变动部分的序列重新比对，得到新的多序列比对结果，重复前面的过程，即根据多序列比对结果构建距离矩阵，计算新的引导树，比较新树与旧树差异，重新比对部分序列，得到新的多序列比对结果，等到树型稳定或迭代次数超过一个值，这一迭代即可终止；第三步才是真正意义上的迭代，以引导树为基础将序列分为两组，分别比后再比对得到所有序列的比对结果，如果新的比对方式使得分增加就保留，反之则抛弃，这样不断分组，比对，评估，直到比对得分收敛或迭代次数达到一定值。MUSCLE 是一款 DOS 环境的程序，EBI 还提供了网页版的 MUSCLE。

3.1 网页版 MUSCLE: <http://www.ebi.ac.uk/Tools/msa/muscle/>

The screenshot displays the MUSCLE web interface, organized into three distinct steps:

- STEP 1 - Enter your input sequences**: This section contains a large text input area for pasting sequences. Below it, there is a file upload option with a "Browse..." button. Chinese annotations include "粘贴多条序列，或直接上传序列文件。" (Paste multiple sequences, or directly upload sequence files.)
- STEP 2 - Set your Parameters**: This section allows users to configure output settings. It includes dropdown menus for "OUTPUT FORMAT" (set to ClustalW), "OUTPUT TREE" (set to From second iteration), and "OUTPUT ORDER" (set to aligned). Chinese annotations explain: "参数选择部分" (Parameter selection part), "序列比对结果输出格式" (Sequence alignment result output format), and "输出进化树的格式" (Output phylogenetic tree format).
- STEP 3 - Submit your job**: This section features a checkbox for email notifications ("Be notified by email (Tick this box if you want to be notified by email when the results are available)") with the Chinese note "可将输出结果发送到你的邮箱" (Can send output results to your mailbox). A red "Submit" button is present, accompanied by the Chinese text "提交比对任务" (Submit alignment task).

Figure 3.23 MUSCLE @ EBI

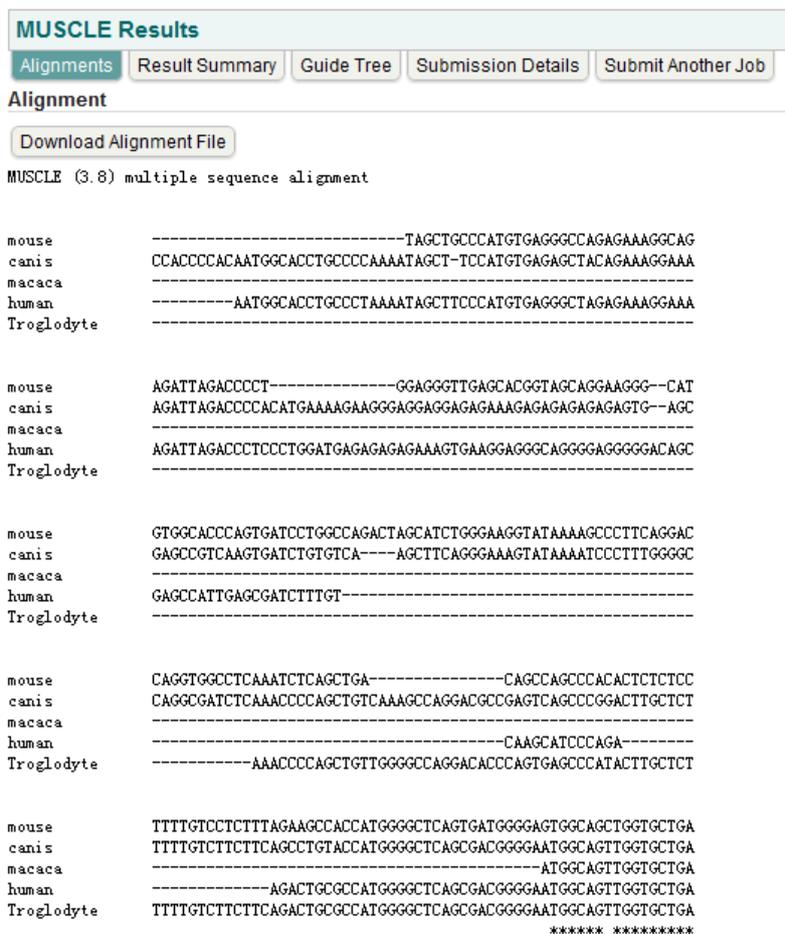


Figure 3.24 MUSCLE 比对结果页面

3.2 DOS 版 MUSCLE

主页左侧一栏 Download 可下载 DOS 版程序，DOS 版本的也很容易使用。

点击开始菜单—运行—输入 CMD—利用 DOS 命令 (cd) 到达 MUSCLE 所在文件夹。—在 DOS 窗口中输入 “MUSCLE -in protein.txt -out output.txt -clw” 点回车即可。这里解释一下上面参数的意义：

- in 后面是待比对的序列文件名(所有待比对 fasta 序列放在一个文件中)，可自己定义。
- out 后面是输出文件名，同样可以自定。
- clw 输出格式类似 clustal 程序，方便我们与 clustal 方法的结果作比较。

运行结束打开输出文件 (output.txt) 即可查看。

```

C:\Windows\system32\cmd.exe

D:\>MUSCLE -in protein.txt -out output.txt -clw

MUSCLE v3.8.31 by Robert C. Edgar

http://www.drive5.com/muscle
This software is donated to the public domain.
Please cite: Edgar, R.C. Nucleic Acids Res 32(5), 1792-97.

protein 5 seqs, max length 154, avg length 139
00:00:00      4 MB(0%)  Iter   1  100.00%  K-mer dist pass 1
00:00:00      4 MB(0%)  Iter   1  100.00%  K-mer dist pass 2
00:00:00      6 MB(0%)  Iter   1  100.00%  Align node
00:00:00      6 MB(0%)  Iter   1  100.00%  Root alignment
00:00:00      6 MB(0%)  Iter   2  100.00%  Root alignment
00:00:00      6 MB(0%)  Iter   3  100.00%  Refine biparts
00:00:00      6 MB(0%)  Iter   4  100.00%  Refine biparts
00:00:00      6 MB(0%)  Iter   5  100.00%  Refine biparts
00:00:00      6 MB(0%)  Iter   6  100.00%  Refine biparts
00:00:00      6 MB(0%)  Iter   7  100.00%  Refine biparts

```

Figure 3.25 4.19 DOS 版 MUSCLE 运行页面

```

MUSCLE (3.8) multiple sequence alignment

mouse      MGLSDGEWQLVLNVWGKVEADLAGHGQEVLIIGLFKTHPETLDKFDKFKNLKSEEDMKGSE
canis      MGLSDGEWQLVLNIWGVETDLAGHGQEVLIIRLFKKNHPETLDKFDKFKHLKTEDEMKGSE
macaca     -----MAVGAERLG---EVLIRLFKGGHPETLEKFDKFKHLKSEDEMKASE
human      MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIIRLFKGGHPETLEKFDKFKHLKSEDEMKASE
troglodyte -----MKASE
                                         **.*

mouse      DLKKHGCTVLTALGTILKKKGQHAAEIQPLAQSHATKHKIPVKYLEFISEIIIEVLKRRH
canis      DLKKHGNTVLTALGGILKKKGHHEAELKPLAQSHATKHKIPVKYLEFISDAIIQVLQSKH
macaca     DLKKHGTVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLELISESIIQVLQSKH
human      DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH
troglodyte DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLHSKH
*****  *****  *****.*  **.:*****.***:  **.*:..*

mouse      SGDFGADAQGAMSKALELFRNDIAAKYKELGFQG
canis      SGDFHADTEAAMKALELFRNDIAAKYKELGFQG
macaca     PGDFGADAQGAMNKALELFRNDMAAKYKELGFQG
human      PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
troglodyte PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
.***  **.:..*.*****.*:.*:*****

```

Figure 3.26 MUSCLE 运行结果

作业

1. 从上节课搜索到的同源序列中选择两条序列，使用 MegAlign 进行全长比对，比对结果设置为红色突出显示匹配字符。再选择序列部分区域进行比对（通过 feature table 选择（需 Genbank 格式序列）或通过坐标 coordinate 选择（Fasta 格式序列即可）），设置比对结果为匹配字符绿色，错配字符红色，用蓝色竖线表示一致序列，不显示标尺。
2. 利用 MegAlign 对之前搜索到的同源核酸和蛋白质序列进行多序列比对分析。比对结果报告（alignment report）以条状图显示每一列序列的相似程度，并以红色突出显示错配字符。
3. 利用 ClustalX 对上题的序列进行多序列比对分析。说明你的参数如何设置，解释比对产生的进化树（dnd 文件），并对比对结果进行分析（序列之间相似度关系，是否存在保守位点及其所在位置等）。
4. 利用 MUSCLE 将上题中的序列进行多序列比对，比对结果以 Clustalw 格式输出。
5. 比较 MegAlign, ClustalX 和 MUSCLE 三种方法的结果是否存在差异？