

多序列比对-工具篇

为什么进行多序列比对

- ❑ 序列保守 意味着 潜在的 功能保守
 - ✿ 不同物种中的同源基因，功能保守，序列相似性较高
 - ✿ 通过多条序列的比较，发现保守与变异的部分
- ❑ 构建分子进化树的必须步骤
- ❑ 比较基因组学研究的基础
- ❑ 两类：全局或局部的多序列比对
- ❑ 本章：全局多序列比对

全局多序列比对



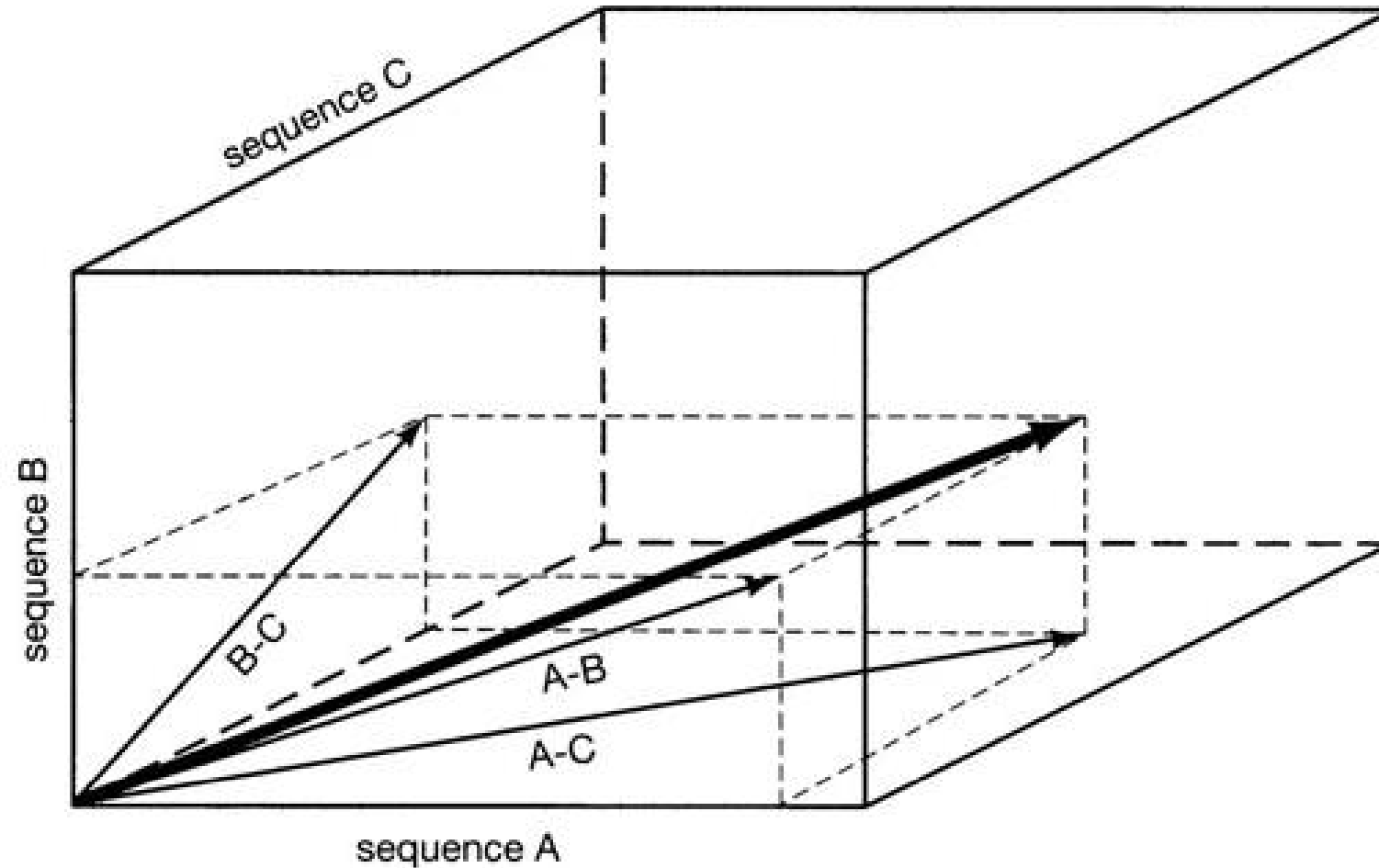
□ 蛋白激酶PKA家族的多序列比对结果（部分）

DmPka-C2	100	RFPFLIYLV DSTKCFDYLYLILPLVIEGELFSYHRRWRKFM EKH R FYNA VFLALEYMHKMHLMYRD	168
DmCG12069	102	MTFP TVLIASYKDFDSLVLPLIGGELFTYHRKVRKFTEK R FYNA VFLALEYLHHCSLLYRD	170
ScTPK2	125	VEHPFLIRMWSTF DARRIFMVMDYIEGELFSLLRKS RFPFPVAKFYAEVILALEYLH H I IYRD	193
ScTPK1	142	VTHPFLIRMWSTF DARRIFMIMDYIEGELFSLLRKS RFPFPVAKFYAEVCLALEYLH SKDI IYRD	210
ScTPK3	143	VSHPFILIRMWSTF DARRIFMVMDYIEGELFSLLRKS RFPFPVAKFYAEVCLALEYLH SKDI IYRD	211
Cekin-1	136	IDFPFLVM TFSFKDMSLYMVLEFISGEMFSHLRRIRRFSEPHSRFYNA IVLAFEYLHSLDLIYRD	204
DmPka-C1	101	IFFPFLVSLRYHFKDMSLYMVLEYVPSGEMFSHLRKVRRFSEPHSRFYNA IVLAFEYLHSLDLIYRD	169
HsPKACg	99	IDFPFLVKLF SFSKDN SYLYLVMEYVPSGEMFSRL RVRRFSEPHCFYNA VVLAV IYHSLDLIHRD	167
HsPKACa	99	VFPFLVKLEFSFSKDN SYLYMVMEYVPSGEMFSHLRRIRRFSEPH R FYNA IVLTFFEYLHSLDLIYRD	167
HsPKACb	99	VFPFLVRLEYAFKDN SYLYMVMEYVPSGEMFSHLRRIRRFSEPH R FYNA IVLTFFEYLHSLDLIYRD	167
DmPKA-C3	329	IRHPFVISLEWSTKDDSYLYMIFDYVCGELFTYLR AGKFTS TSYFYAEIVSALEYLHSL IYRD	397
HsPRKX	104	VSHPFILIRLFWTWHERFLYMLMEYVPSGELFSYLR RGRFSSTTCLFYSAEIIICAEYLHLSKEIVYRD	172
HsPRKY	104	VSHPFILIRLFWTWHEERFLYMLMEYVPSGELFSYLR RGRFSSTTCLFYSAEIIICAEYLHLSKEIVYRD	172

Made by GENEDOC

<http://genedoc.software.informer.com/>

动态规划算法：全空间



<http://www.ncbi.nlm.nih.gov/CBBresearch/Schaffer/msa.html>

注意



- 最优的多序列比对，其两两序列之间的比对 不一定最优

V	S	N	-	S
-	S	N	A	-
-	-	-	A	S

最优的多序列比对

-	S	N	A	-
-	-	-	A	S

非最优的双序列比对

多序列比对的计算方法



- ❑ 渐进方法: **Progressive methods**
- ❑ 迭代方法: **Iterative refinement**
- ❑ 部分有向图算法
- ❑ 隐马尔科夫模型: **HMM profile-profile**
- ❑ 整合算法: **Meta-methods**
- ❑ 结构特征

渐进方法

Progressive methods



- ❑ 渐进方法: Pairwise alignment
- ❑ ClustalW/X: "Classic Clustal"
 - ✿ <http://www.clustal.org/>
 - ✿ <http://www.clustal.org/clustal2/>
- ❑ T-Coffee
 - ✿ <http://tcoffee.org/>
 - ✿ <http://tcoffee.crg.cat/apps/tcoffee/all.html>

ClustalW/X



- ❑ Clustal: 1988年开发
- ❑ ClustalW: 1994年, Julie D. Thompson 等人改进、开发
- ❑ ClustalX: 1997年, 图形化软件

Table 1
Multiple Alignment Methods

Method (Developer)	Algorithm	Matrix*	Indels	Limits ^b	Assumptions ^c	Features ^d	Data Type ^e
Global:							
AMULT (G. Barton)	NW	Any	C		Y, S	R, SE	P
ASSEMBLE (M. Vingron)	Dot matrix NW	Log odds	I+E		Y, S		P
CLUSTAL V (D. Higgins)	WL	Any	I+E			I	P, N
DFALIGN (D.-F. Feng)	NW	Log odds	C	UP	Y, E, O		P
GENALIGN ^f (H. Martinez)	CW, NW	UM	I+E			SE	P, N
MSA (S. Altschul)	CL	PAM250	I+E	ROS	N	B, FA	P
MULTAL (W. Taylor)	NW	UM, PAM250	C		S	AP, FA	P
MWT (J. Kececioglu)	maximum weight trace	Any	C	ROS	N		P
TULLA (S. Subbiah)	NW	Any	RGW	10 sequences	S	R, SE	P
Local:							
MACAW (G. Schuler)	SW	PAM250		DOS	Y	SE, FA, MD	P
PIMA (P. Smith)	SW	AACH	I+E		Y	MD	P
PRALIGN (M. Waterman)	CW	PAM250	I+E*		Y	MD, MC	P, N ^h

ClustalW/X: 计算过程



- ❑ 将所有序列两两比对，计算进化距离（差异）矩阵
- ❑ 使用邻接法（**neighbor-joining**）构建指导树（**guide tree**）
- ❑ 将进化距离最近的两条序列用全局动态规划算法进行比对
- ❑ “渐进”地加上其他序列

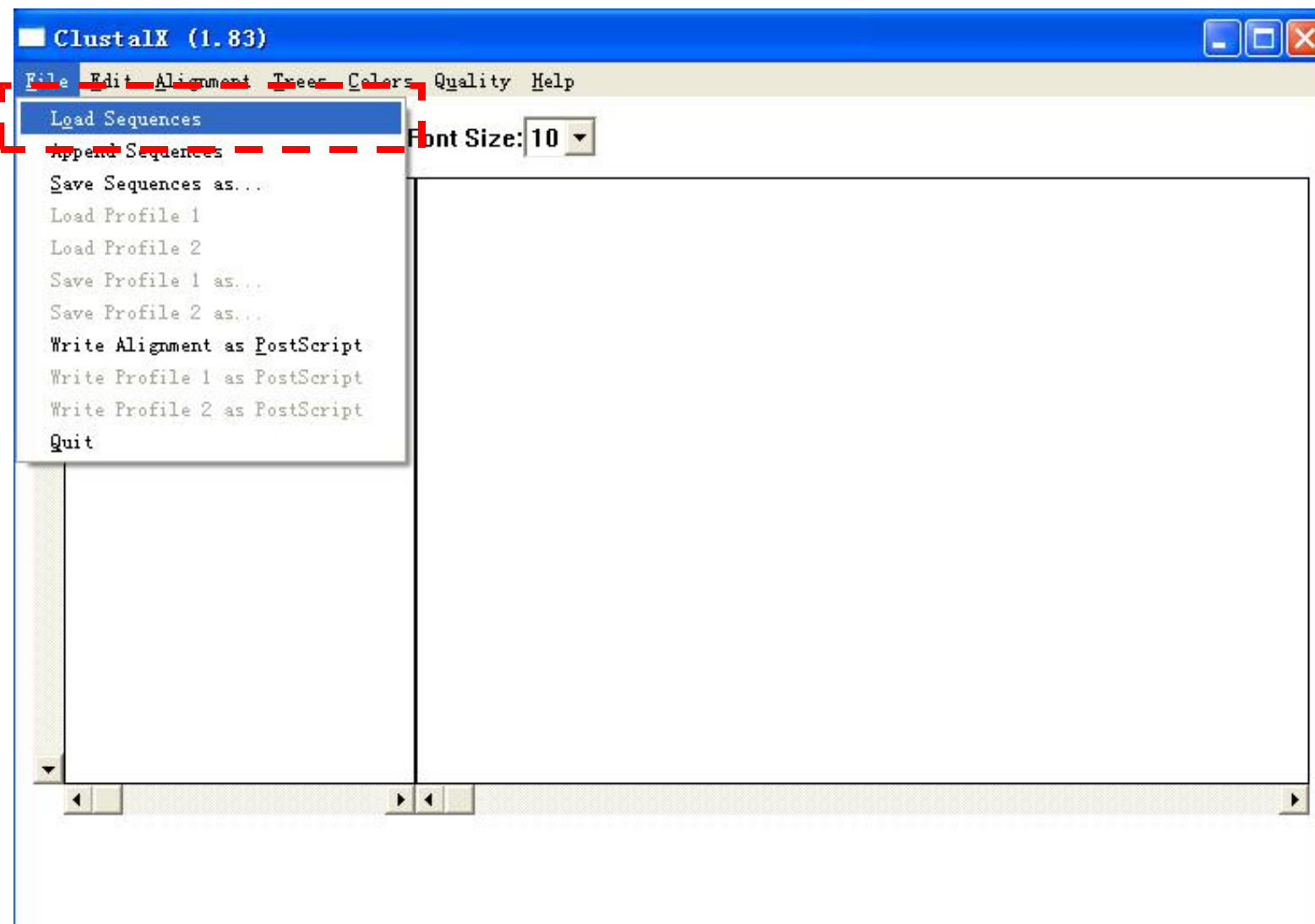
ClustalX: 使用指南



□ FASTA序列格式，多序列

```
>ScTPK1
MSTEEQNGGGQKSLDDRQGEESQKGETSERETTATESGNESEKSVKEGGETQEKPKQPHV
TYYNEEQYKQFIAQARVTSGKYSLQDFQILRTLGTGSFGRVHLIRSRHNGRYYAMKVLKK
EIVVRLKQVEHTNDERLMLSIVTHPFIIRMWGTFFQDAQQIFMIMDYIEGGELFSLLRKSQ
RFPNPVAKFYAAEVCLALEYLHSDKIIYRDLKPENILLDKNGHIKITDFGFAKYVPDVTY
TLCGTPDYIAPEVVSTKPYNKSIDWWSFGILYEMLAGYTPFYDSNTMKTYEKILNAELR
FPPFFNEDVKDLLSRLITRDLSQLGNLQNGTEDVKNHPWFKEVVWEKLLSRNIETPYEP
PIQQQGQDTSQFDKYPEEDINYGVGQGEDPYADLFRDF
>ScTPK2
MEFVAERAQPVGQTIQQQNVNTYGGVLPQPHHDLQQRQQQQQQRQHQQLLTSQLPQKSLV
SKGKYTLHDFQIMRTLGTGSFGRVHLVRSVHNGRYAIAKVLKKQQVVKMKQVEHTNDERR
MLKLVEHPFLIRMWGTFFQDARNIFMVMDYIEGGELFSLLRKSQRFPNPVAKFYAAEVILA
LEYLHAHNIIYRDLKPENILLDRNGHIKITDFGFAKEVQVTWTLCGTPDYIAPEVITTK
PYNKSVDWWSLGVLIYEMLAGYTPFYDTPMKTYEKILQGKVVPYPPYFHPDVVDLLSKLI
TADLTRRIGNLQSGSRDIKAHPWFSEVVWERLLAKDIETPYEPPITSGIGDTSLFDQYPE
EQLDYGIQGDDPYAEYFQDF
>ScTPK3
MYVDPMNNNEIRKLSITAKTETTPDNVGQDIPVNAHSVHEECSSNTPVEINGRNSGKLKE
EASAGICLVKKPMLQYRDTSGKYSLSDFQILRTLGTGSFGRVHLIRSNHNGRFYALKTLK
KHTIVKLKQVEHTNDERMLSIVSHPFIIRMWGTFFQDSQQVFMVMDYIEGGELFSLLRKS
QRFPNPVAKFYAAEVCLALEYLHSDKIIYRDLKPENILLDKNGHIKITDFGFAKYVPDVT
YTLCGTPDYIAPEVVSTKPYNKSVDWWSFGVLIYEMLAGYTPFYNSNTMKTYENILNAEL
KFPPFFHPDAQDILLKKLITRDLSERLGNLQNGSEDVKNHPWFNEVIWEKLLARYIETPYE
PPIQQQGQDTSQFDRYPEEEFNIGIYGEDPYMDLMKEF
>Cekin-1
MPTRLDIVGNLQFSSSTDNGDEDQEADVTAFCVLPSPSSFSKLSILDDPVEDFKEFLDKA
REDFKQRWENPAQNTACLDDFDRIKTLGTGSFGRVMLVKHKQSGNYAMKILDKQKVVKL
KQVEHTLNEKRIHQAIQDFPFLVNMFTSFKDNSNLYMVLEFISGGEMFSLHRRIGRFSEPH
SRFYAAQIVLAFEYLHSLDLIYRDLKPENLLIDSTGYLKITDFGFAKRVKGRTWTLCGTP
EYLAPEIILSKGYNKAVDWWALGVLIYEMAAGYPPFFADQPIQIYEKIVSGKVKFSPSHFS
NELKDLLKNLLQVDLTKRYGNLKNGVADIKNHKWFGSTDWIAIYQKKITPPSFSKGESNG
RLFALYPRVDGPADTRHFVEEVQEPTEFVIAATPQLEELFVEF
```

导入序列文件



执行比对



ClustalX (1.83)

File Edit Alignment Trees Colors Quality Help

Multiple

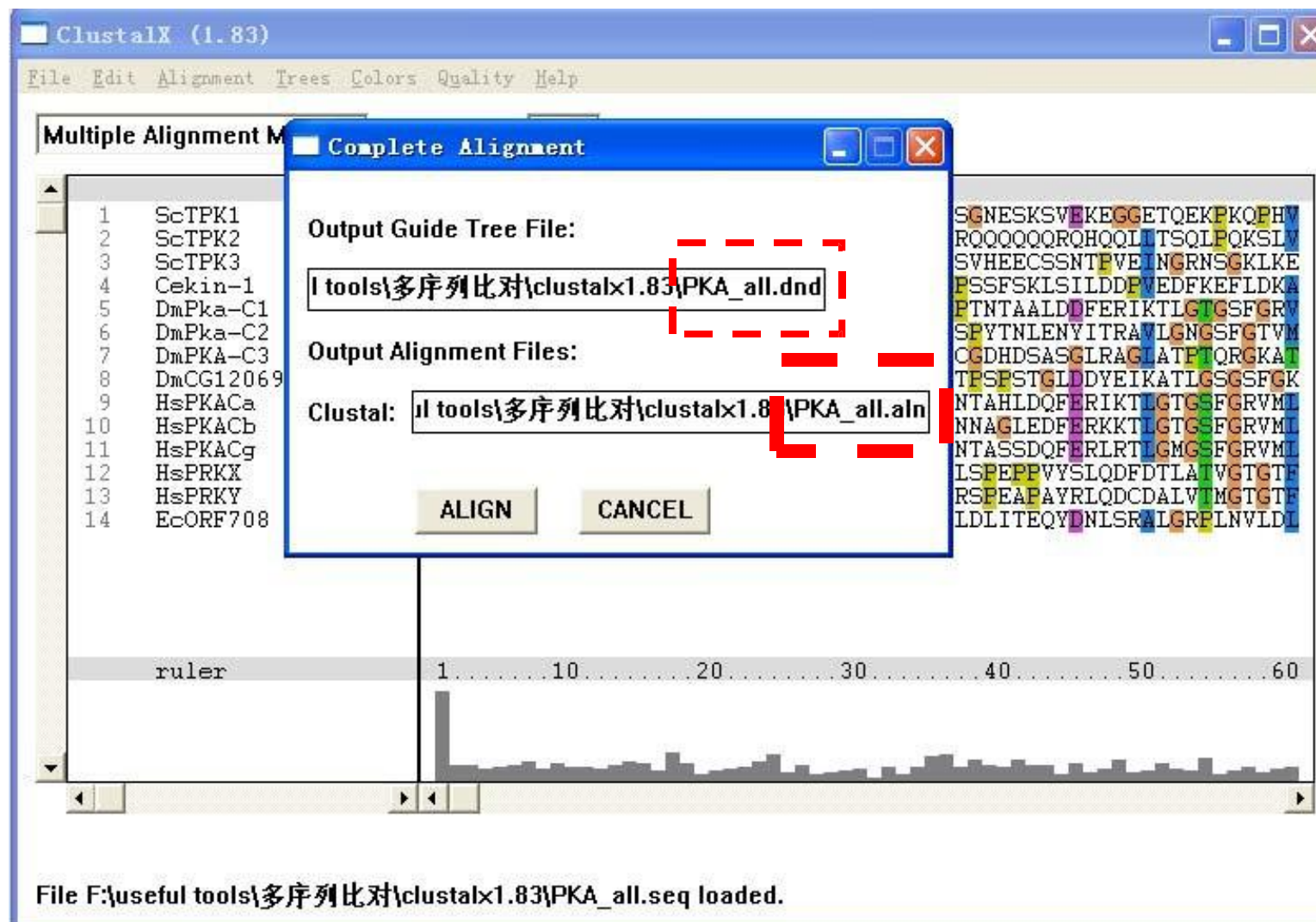
- Do Complete Alignment
- Produce Guide Tree Only
- Do Alignment from Guide Tree
- Realign Selected Sequences
- Realign Selected Residue Range
- Align Profile 2 to Profile 1
- Align Profiles from Guide Trees
- Align Sequences to Profile 1
- Align Sequences to Profile 1 from Tree
- Alignment Parameters
- Save Log File
- Output Format Options

1 DDROGEESQKGETSERETTATESGNESKSVKEGGETQEKPKQPHV
2 IQQNVNTYGGVLOPHHDLOQRQQQQQQRHQQLITSQLPKSLV
3 SITAKTETTFDNGQDIFVNAHSVHEECSSNTEVEINGRNSGKLKE
4 SSTDNGDEDQEADVTA CFVLPSPSSFSKLSILDDFVEDFKEFLDKA
5 AETVKEFLEQAKEEFEDKWRNPTNTAALDDFERIKTLGTGSGFRV
6 DYNVILDNMSREFEERNHQTQSPYTNLENYITRAVLGNGSFGTVM
7 CRVSASVFANFGGLYSSWKLICGDHDSASGLRAGLATFTORGKAT
8 KVDYILILDKLREDFNKKFATNTFSPSTGLDDYEIKATLGSGSFGK
9 SVKEFLAKAKEDFLKKWESPAQNTAHLDOFERIKTLGTGSGFRVML
10 SVKEFLAKAKEDFLKKWENPTQNNAGLEDFERKKTILGTGSGFRVML
11 SVNEFLAKARGDELYRWGNPAQNTASSDQFERLRTLGMGSFGRVML
12 DSRKVAEETFDGAPALCPSPEALSPEPPVYSLQDFDTLATVGTGTF
13 NSREVTEDAADWAPALCPSPEARSPEAPAYRLQDCDALVMGTGTF
14 IYQTI FGHPEWDGDAARDCNQRLDLITEQYDNL SRALGRPLNVLDL

ruler 1 10 20 30 40 50 60

File F:\useful tools\多序列比对\clustalx1.83\PKA_all.seq loaded.

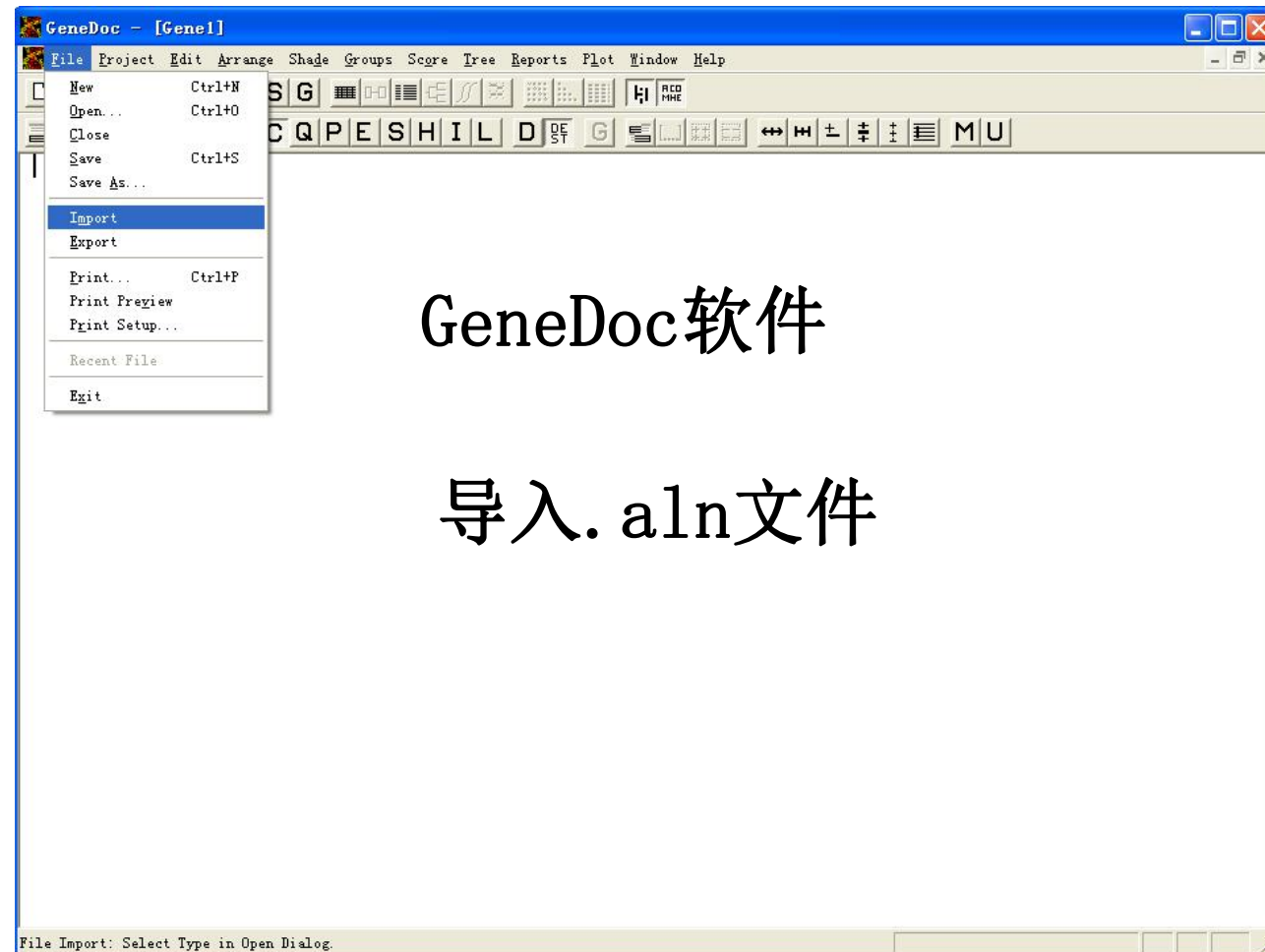
文件导出



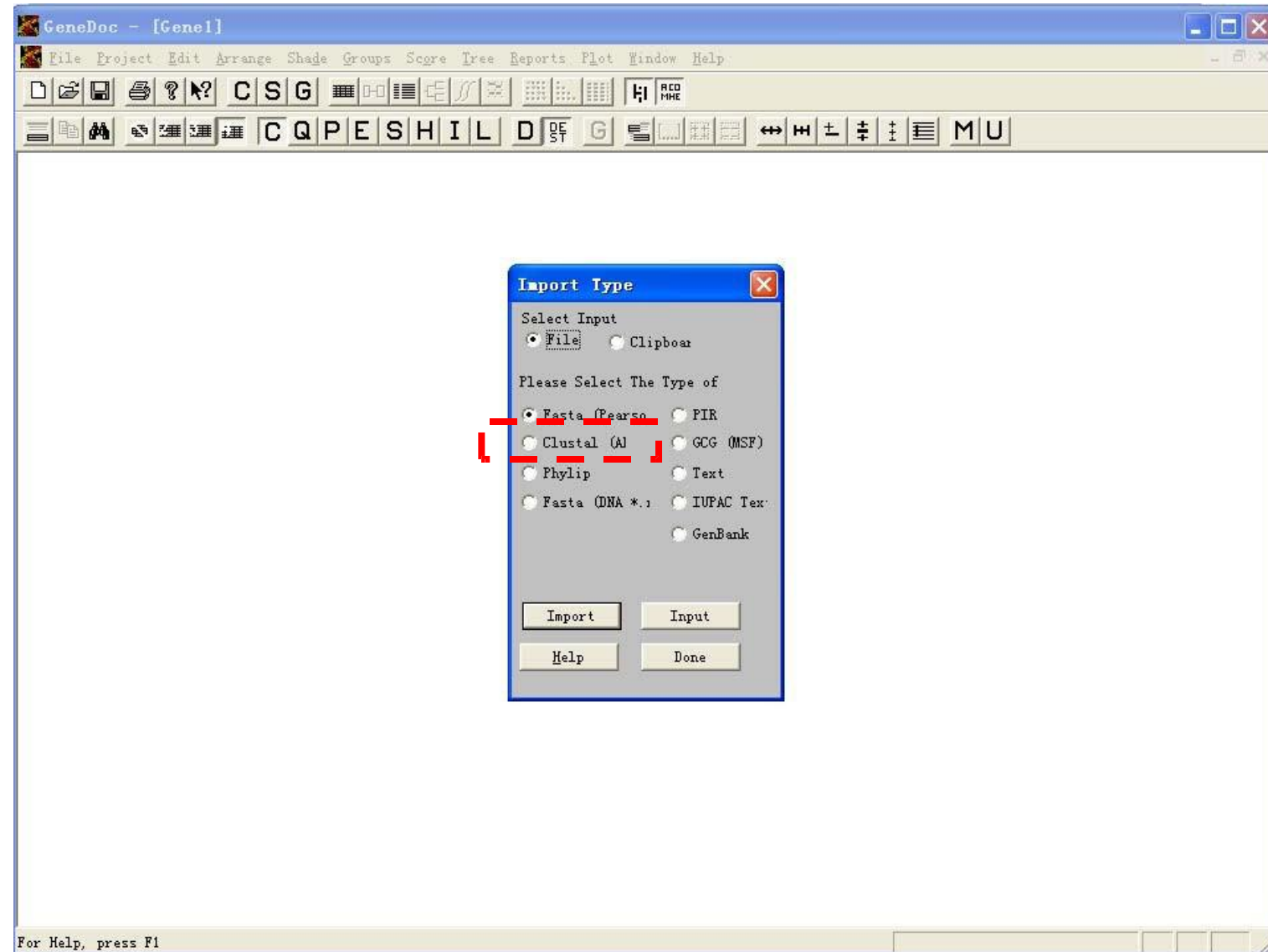
多序列比对：结果处理



□ GeneDoc, BioEdit等软件



选择文件格式



成功导入文件



GeneDoc - [Gene1]

File Project Edit Arrange Shade Groups Score Tree Reports Plot Window Help

C S G C Q P E S H I L D G M U

hspk1 : ---ADCFSEAK-----SEAK-----AKRQDCDA-LVINGIGTFGRVHLVRLAKH : 74

DmPKA-C3 : ADDATHDSSESIEEDDGNEDDDEDDDESESSSVQTAAGVRKYHLDYQI-IKTVGTGTFGRVCLCRDRISEKY : 299

EcORF708 : --VSQPDPRELIEQCAFYRLIGFBDTHLSPVPRPMYLVSNHRVLLINDENQPFQHWQNPYAGAGLAHKRRRYE : 244

t g g 5g v 6

320 340 360

DmPka-C2 : YAAKMMSKEDLVRL-----KQVAHVHNEKHVLAARFPFLIYLV DSTKC-FDYLYLILPLVNGGELEFSY : 133

DmCG12069 : YASKQLSKDQIVKT-----KQVSHVMSEKNVLRSMFTFNTVNLIASYKD-FDSLYLVPLIGGGELEFTY : 135

HsPKACa : YAMKILDQKQVVKL-----KQIEHTLNEKRILQAVNFPFLVKLEFSFKD-NSNLYMVMEXVPGGEMESH : 132

HsPKACb : YAMKILDQKQVVKL-----KQIEHTLNEKRILQAVNFPFLVRLLEYAFKD-NSNLYMVMEXVPGGEMESH : 132

HsPKACg : YAMKILNKQKQVVKM-----KQVEHILNEKRILQAIQFPFLVKLQFSFKD-NSYLYLVMEYVPGGEMFSR : 132

DmPka-C1 : YAMKILDQKQVVKL-----KQVEHTLNEKRILQAIQFPFLVSLRYHFKD-NSNLYMVLEYVPGGEMESH : 134

Cekin-1 : YAMKILDQKQVVKL-----KQVEHTLNEKRILQAIQFPFLVNMFTFSFKD-NSNLYMVLEETISGGEMESH : 169

ScTPK1 : YAMKVLKKEIVRL-----KQVEHTNDERLMLSIVTHPFIIRMWGTFQD-AQQIFMIMDYIEGGELESL : 175

ScTPK3 : YALKTLKKHTIVKL-----KQVEHTNDERRMLSIVSHPFPIIRMWGTFQD-SQQVFVMMDYIEGGELESL : 176

ScTPK2 : YAIKVLKQKQVVKM-----KQVEHTNDERRMLKLVEHPFLIRMWGTFQD-ARNIFVMMDYIEGGELESL : 158

HsPRKX : FALKVMSIPDVIRL-----KQEQHVHNEKSVLKEVSHPFILRLFWTWHD-ERFLYMLMEYVPGGELESY : 137

HsPRKY : FALKVMSIPDVIRL-----KQEQHVHNEKSVLKEVSHPFILRLFWTWHE-ERFLYMLMEYVPGGELESY : 137

DmPKA-C3 : CAMKILAMTEVIRL-----KQIEHVKNERNILREIRHPFVISLEWSTKD-DSNLYMIFDYVCGGELETY : 362

EcORF708 : FGEDYVCKFFYYDMPHGILTAEESSQNKHELHNEIKFLTQPPAGFDAPAVLAHGENAQSGWLVMEKLEPG-RLLSD : 318

a k 6 k 4q h E L pf d 566 6 Gge6f3

380 400 420 440

DmPka-C2 : HRRVRKFNEKHARFYAAQVALALEYMHKMHLMYRDLKPENILLDQRGYIKITDFG-ETKRVDGRTSTLCGTF--- : 204

DmCG12069 : HRKVRKFTEKQARFYAAQVELALEYLHCSLLYRDLKPENIMMDKNGYLKVTDFG-FAKKVETRTMTLCGTF--- : 206

HsPKACa : LRRIGRFSEPHARFYAAQIVLTFEYLHSLDLIYRDLKPENILLIDQCCYIQVTDG-FAKRVKGRNTWTLGTF--- : 203

HsPKACb : LRRIGRFSEPHARFYAAQIVLTFEYLHSLDLIYRDLKPENILLIDHQQYIQVTDG-FAKRVKGRNTWTLGTF--- : 203

HsPKACg : LQRVGRFSEPHACFYAAQVLAQVYLHSLDLIHRDLKPENILLIDQGYLQVTDG-FAKRVKGRNTWTLGTF--- : 203

DmPka-C1 : LRKVGRFSEPHSRFYAAQIVLAQVYLHSLDLIYRDLKPENILLIDSGYLKVTDFG-FAKRVKGRNTWTLGTF--- : 205

Cekin-1 : LRRIGRFSEPHSRFYAAQIVLAQVYLHSLDLIYRDLKPENILLIDSTGYLKITDFG-FAKRVKGRNTWTLGTF--- : 240

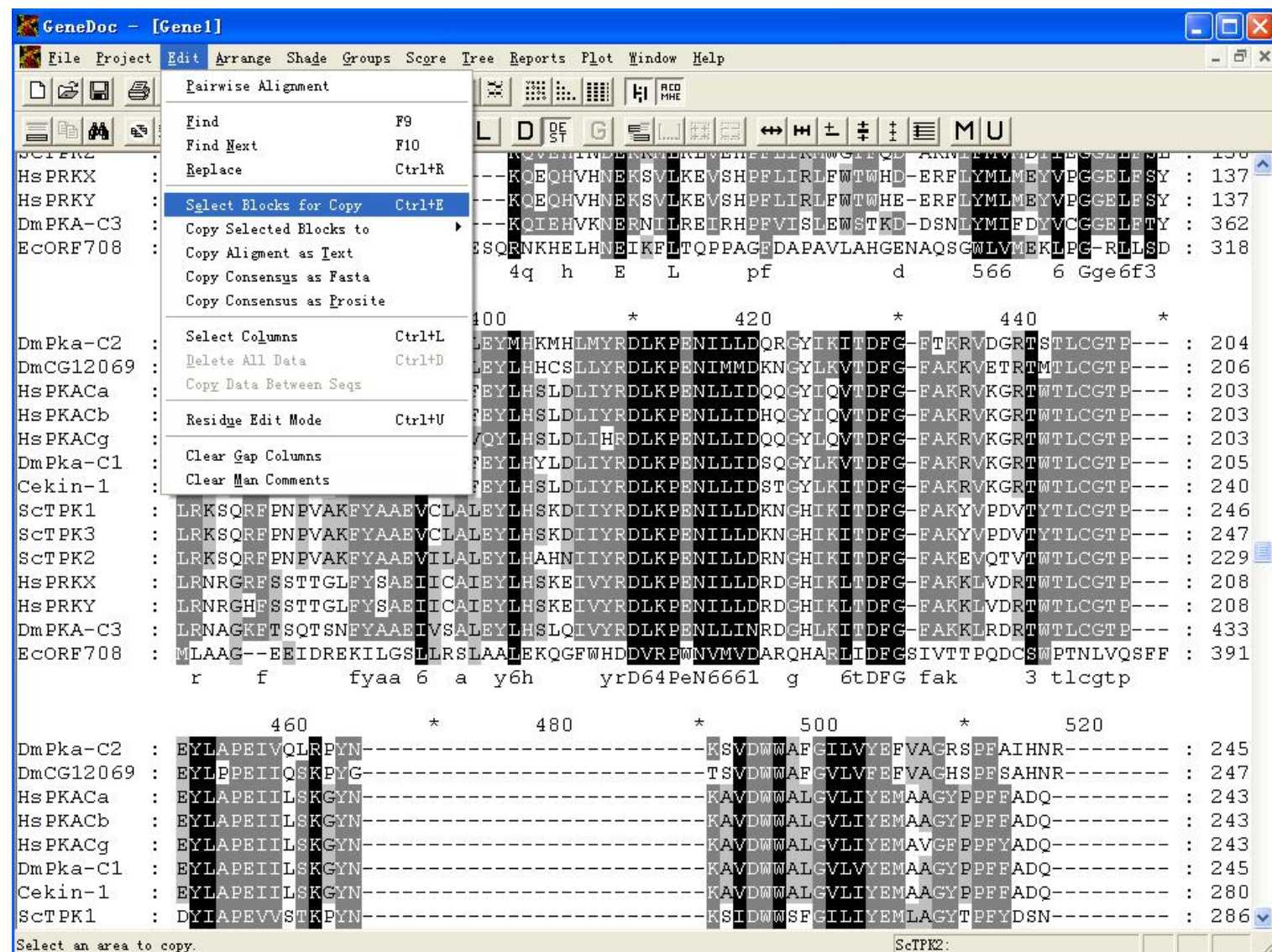
ScTPK1 : LRKSQRFPNPVAKFYAAEVCLALEYLHSDKIYRDLKPENILLDKNGHIKITDFG-FAKYVPDVTYTLGTF--- : 246

ScTPK3 : LRKSQRFPNPVAKFYAAEVCLALEYLHSDKIYRDLKPENILLDKNGHIKITDFG-FAKYVPDVTYTLGTF--- : 247

ScTPK2 : LRKSQRFPNPVAKFYAAEVILALEYLHAHNIYRDLKPENILLDRNGHIKITDFG-FAKEVQTVNTWTLGTF--- : 229

For Help, press F1

选择需要拷贝的行



比对结果的美化和后处理



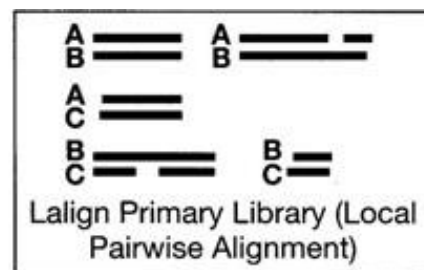
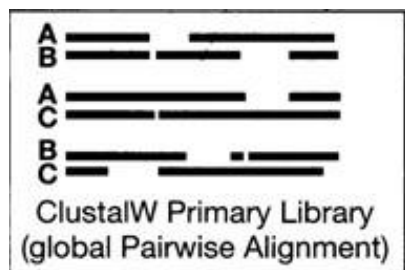
Cekin-2	SGGRRTGISAE	89
ScBCY1	NAQRRTISVSGE	149
DmPka-R2	ASSRRKSVFAE	88
HsPRKAR2A	N--RRVSVCAE	103

Cekin-2	DYFGEIALLLDRPRAATVVAKTH	329
ScBCY1	DYFGEVALINDLPRQATVTATKR	386
DmPka-R2	QYFGELALVTHRPRAASVYATGG	329
HsPRKAR2A	QYFGELALVTNKPRAASAYAVGD	356

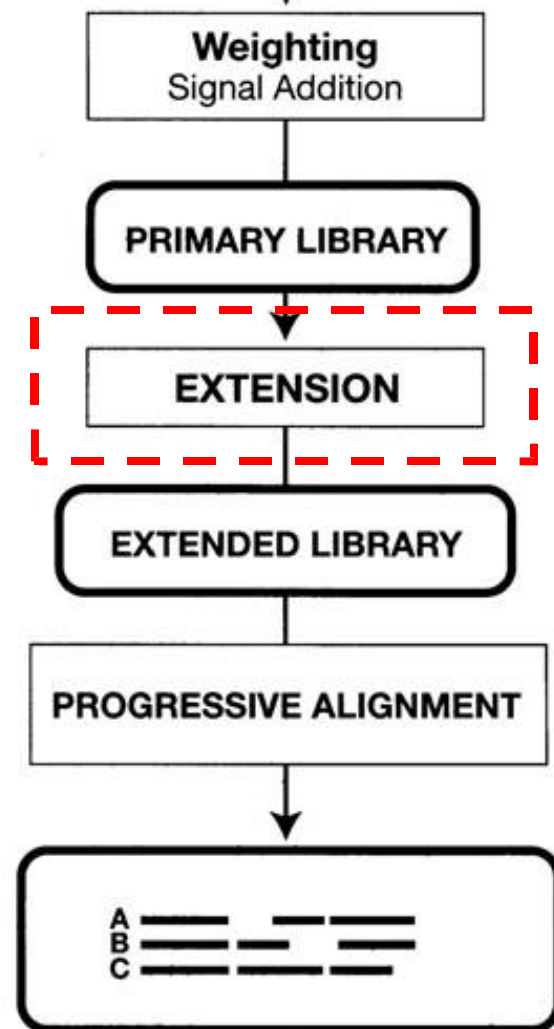
T-Coffee



- ❑ 采用**Clustal**程序计算两两序列之间的全局 最优比对结果
- ❑ 采用**LALIGN**程序计算两两序列之间的局 部最优比对的结 果
 - ✿ <https://www.ebi.ac.uk/Tools/psa/lalign/>
- ❑ 设计加权系统，综合考虑上述两两部分结 果，构建指导库
- ❑ 采用渐进算法，得到最终的结果



同时进行全局和局部的
双序列比对



对以上打分的结果设计
权重系统，找到序列中
最保守的部分

渐进比对，基于上述计
算得到的指导库
(**primary library**)

渐进方法存在的问题



- 通常使用启发式策略选择初始序列
 - ✿ 最终结果可能受初始选定的序列的影响
- 距离最近的，有两组序列**AB**和**CD**，哪组最先比对？两种方案：
 - ✿ **A.** 分别、同时比对。究竟应以**AB**为准，加入 **CD**，然后再加上其他序列，还是以**CD**为准？ 结果可能出入很大
 - ✿ **B.** 随机挑选一组作为基准
- 当序列之间差异较大时，上述问题更加明显



□ 三条序列:

Seq1: ARKCV

Seq2: ARCV

Seq3: AKCV

□ 若Seq1, 2先比对, 再
加入Seq3:

A R K C V

A R - C V

A-KCV

□ Seq1, 3先比对, 再加
入Seq2:

A R K C V

A - R C V

A-KCV

□ Seq2, 3先比对, 再加
入Seq1:

A R K C V

A R - C V

AK-CV

迭代算法

迭代算法



- ❑ 部分解决渐进算法存在的问题，主要是 ClustalW/X 存在的问题

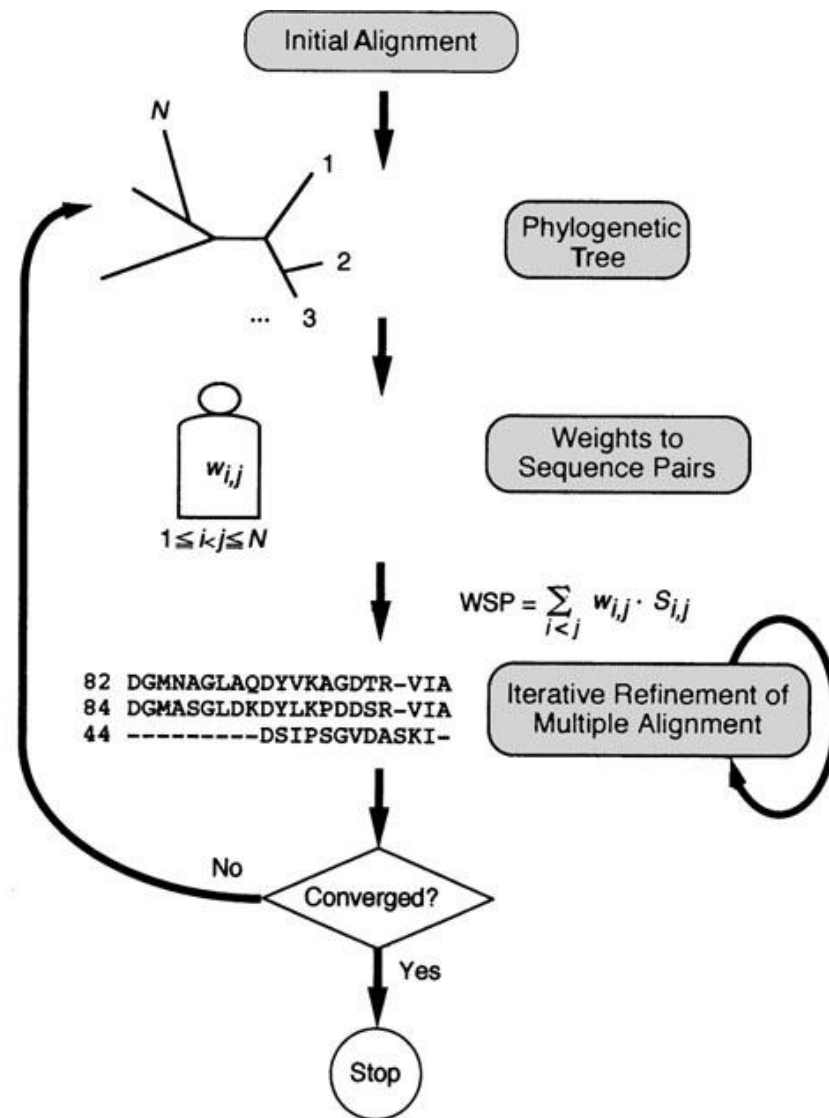
- ❑ PRRP/PRRN

 - ✿ <https://www.genome.jp/tools-bin/prn>

- ❑ DIALIGN

 - ✿ <http://dialign.gobics.de/>

PRRP/PRRN



1. 先用“渐进”算法进行多序列比对
2. 基于多序列比对的结果构建进化树
3. 重新计算序列之间的进化距离，再用“渐进”算法进行多序列比对
4. 重复上述步骤，直到结果不再发生改变为止

DIALIGN



- 对所有序列进行两两之间的局部最优比对
- 找到所有能够匹配的部分**M1**；将重叠的、前后一致的（**consistency**）匹配部分连接起来为**M2**
- 将剩下的未比对的序列重新比对，再发现能够匹配的部分，构成新**M1**，将一致的部分构成**M2**
- 重复上述步骤，直到结果收敛

有向图算法

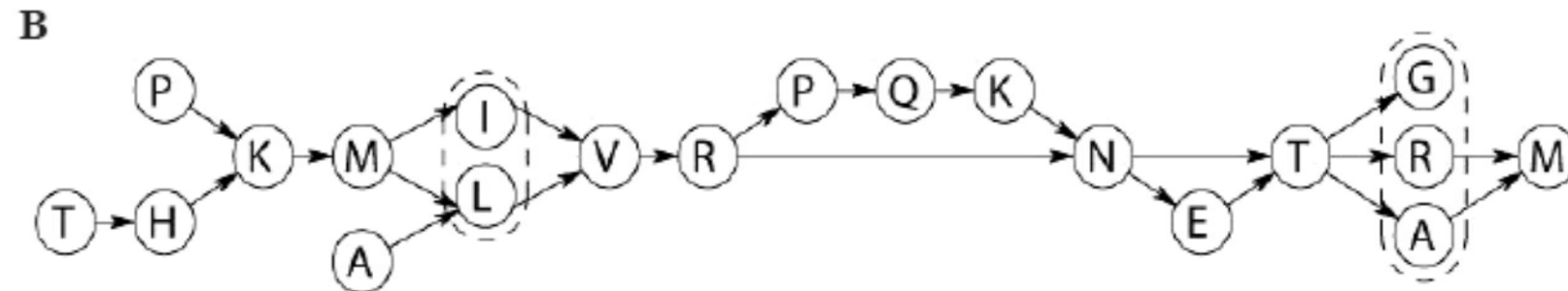
部分有向图算法：POA



- ❑ <https://simpsonlab.github.io/2015/05/01/understanding-poa/>
- ❑ <https://sourceforge.net/projects/poamsa/>

A

.	.	P	K	M	.	I	V	R	P	Q	K	N	E	T	G	.
.	A	L	V	R	P	Q	K	N	.	T	R	M
T	H	.	K	M	.	L	V	R	.	.	.	N	E	T	A	M



隐马尔可夫以及整合算法

隐马尔科夫模型: ProbCons



- ❑ <http://probcons.stanford.edu/>
- ❑ 主要改进:
- ❑ 所有序列的两两比对, 通过**profile HMM**的方法进行双序列比对
- ❑ 将渐进算法与迭代算法整合

整合算法MUSCLE



- 算法分为三个部分，每个部分相对独立
- **Draft progressive:**
 - ✿ (1) 对两条序列，计算距离采用**k-mer**的思想；
 - ✿ (2) 用**UPGMA**算法构建引导树
 - ✿ (3) 使用渐进算法进行多序列比对
- 优点：两条序列之间的距离不采用动态规划算法进行比对，节省时间

MUSCLE (2)



❑ Improved progressive:

- ✿ (1) 基于 k -mer 得到的树可能会产生次优结果， 因此，采用 Kimura 距离的方法对 k -mer 产生 的树重新计算距离矩阵
- ✿ (2) 重新用 **UPGMA** 构建进化树
- ✿ (3) 使用渐进算法进行多序列比对

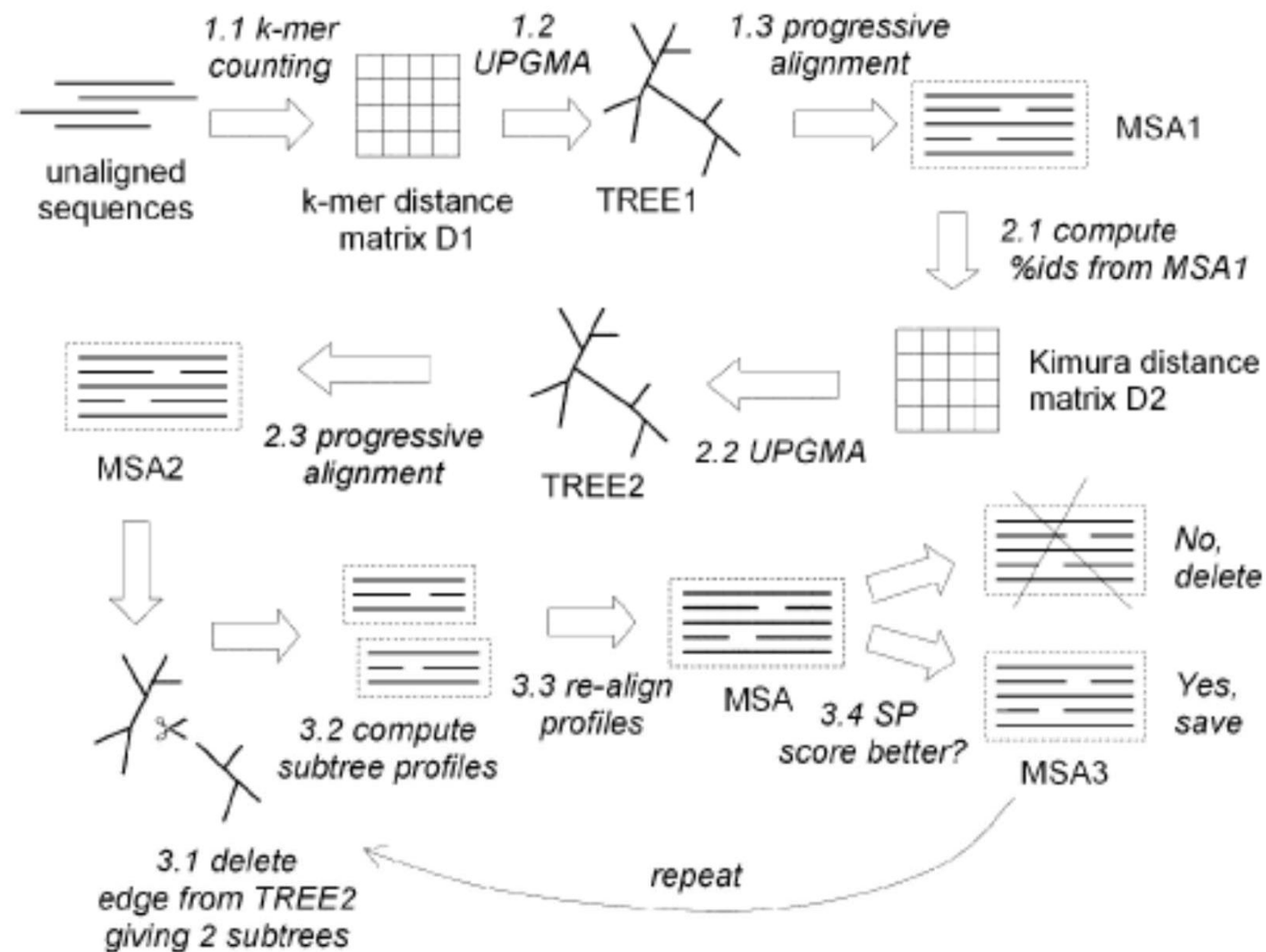
MUSCLE (3)



□ Refinement:

- ✿ (1) 随机从进化树上挑出一条边，删除
- ✿ (2) 得到两组树，对每组树，计算**profile**
- ✿ (3) 将两组**profile**进行比对
- ✿ (4) 如果最终得分提高，保留结果，否则丢弃

MUSCLE的算法流程





MUSCLE: 使用指南

[Home](#) [Software](#) [Services](#) [About](#) [Contact](#)

MUSCLE

MUSCLE has been cited by
34,183 papers
[Google scholar](#)
Last updated 10 Feb 2020

[Downloads](#)

[Documentation](#)

[Support](#)

USEARCH

Ultra-fast sequence analysis



10 - 1,250x BLAST
1 - 1,000x CD-HIT

Popular multiple alignment software

MUSCLE is one of the most widely-used methods in biology. On average, MUSCLE is cited by ten new papers every day.

Fast, accurate and easy to use

MUSCLE is one of the best-performing multiple alignment programs according to published benchmark tests, with accuracy and speed that are consistently better than CLUSTALW. MUSCLE can align hundreds of sequences in seconds. Most users learn everything they need to know about MUSCLE in a few minutes—only a handful of command-line options are needed to perform common alignment tasks.

Papers

There are two papers. The first (NAR) introduced the algorithm, and is the primary citation if you use the program. The second (BMC Bioinformatics) gives more technical details, including descriptions of non-default options.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput *Nucleic Acids Res.* **32**(5):1792-1797 [[Link to PubMed](#)].

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity *BMC Bioinformatics*, **(5)** 113 [[Link to PubMed](#)].

<http://www.drive5.com/muscle/>

MUSCLE: 使用说明



```
C:\ 命令提示符
E:\muscle3.6>muscle -in PKA.seq -out PKA.aln -clw
MUSCLE v3.6 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.

PKA 14 seqs, max length 708, avg length 403
00:00:00      2 MB(3%) Iter  1 100.00% K-mer dist pass 1
00:00:00      2 MB(3%) Iter  1 100.00% K-mer dist pass 2
00:00:01      7 MB(11%) Iter  1 100.00% Align node
00:00:01      7 MB(11%) Iter  1 100.00% Root alignment
00:00:01      7 MB(11%) Iter  2 100.00% Refine tree
00:00:01      7 MB(11%) Iter  2 100.00% Root alignment
00:00:01      7 MB(11%) Iter  2 100.00% Root alignment
00:00:01      7 MB(11%) Iter  3 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  4 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  5 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  5 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  6 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  7 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  8 100.00% Refine biparts
00:00:01      7 MB(11%) Iter  9 100.00% Refine biparts
```



Clustal Omega



❑ 算法原理类似MUSCLE


⚙ <http://www.clustal.org/omega/>

⚙ <https://www.ebi.ac.uk/Tools/msa/clustalo/>



Clustal Omega

"The last alignment program you'll ever need"



- Home
- Webservers
- Download
- Documentation
- Contact
- News

Introduction

Clustal Omega is the latest addition to the Clustal family. It offers a significant increase in scalability over previous versions, allowing hundreds of thousands of sequences to be aligned in only a few hours. It will also make use of multiple processors, where present. In addition, the quality of alignments is superior to previous versions, as measured by a range of popular benchmarks.

Please note that Clustal Omega is currently a command line-only tool.

A full description of the algorithms used by Clustal Omega is available in the Molecular Systems Biology paper [Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega](#). Latest additions to Clustal Omega are described in [Clustal Omega for making accurate alignments of many protein sciences](#)

Webservers

Download Clustal Omega

其他多序列比对工具



□ MAFFT: 渐进 & 迭代

✿ <https://www.genome.jp/tools-bin/mafft>

□ T-Coffee (M-coffee): 整合其他工具的输出结果

✿ <http://tcoffee.crg.cat/apps/tcoffee/all.html>

□ Multiple Sequence Alignment

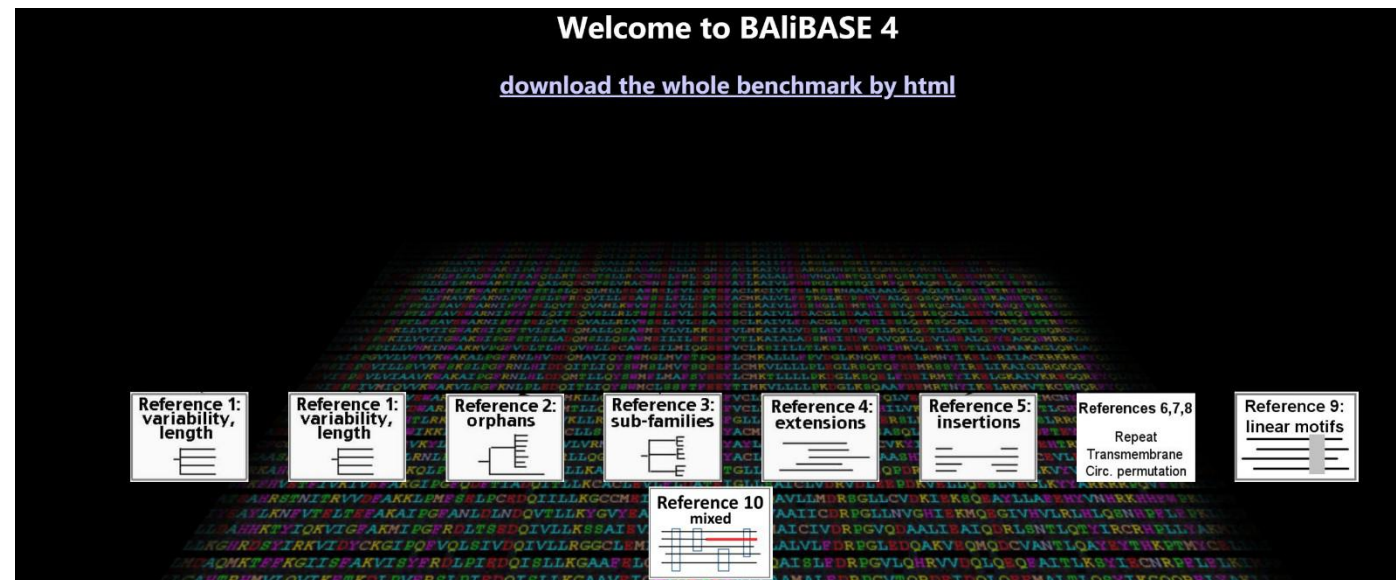
✿ <https://www.ebi.ac.uk/Tools/msa/>

多序列比对：性能检验



- ❑ **BAliBASE**: 基于蛋白质三级结构，将同一家族的蛋白质序列进行多序列比较
- ❑ 多序列比对工具的性能检验：能否与BAliBASE中的比对结果相吻合

🌸 <http://www.lbgi.fr/balibase/>



性能比较



- ❑ **ClustalW/X**: 最经典、最被广泛接受的工具
- ❑ **MUSCLE**: 最流行的多序列比对工具
- ❑ **Clustal Omega**: 类似**MUSCLE**
- ❑ **T-Coffee**: 序列相似性高时最准确
- ❑ **DIALIGN**: 序列相似性低时较准确
- ❑ **POA**: 性能接近**T-Coffee**和**DIALIGN**, 速度最快（目前主要用于三代测序数据分析）

运算时间比较

