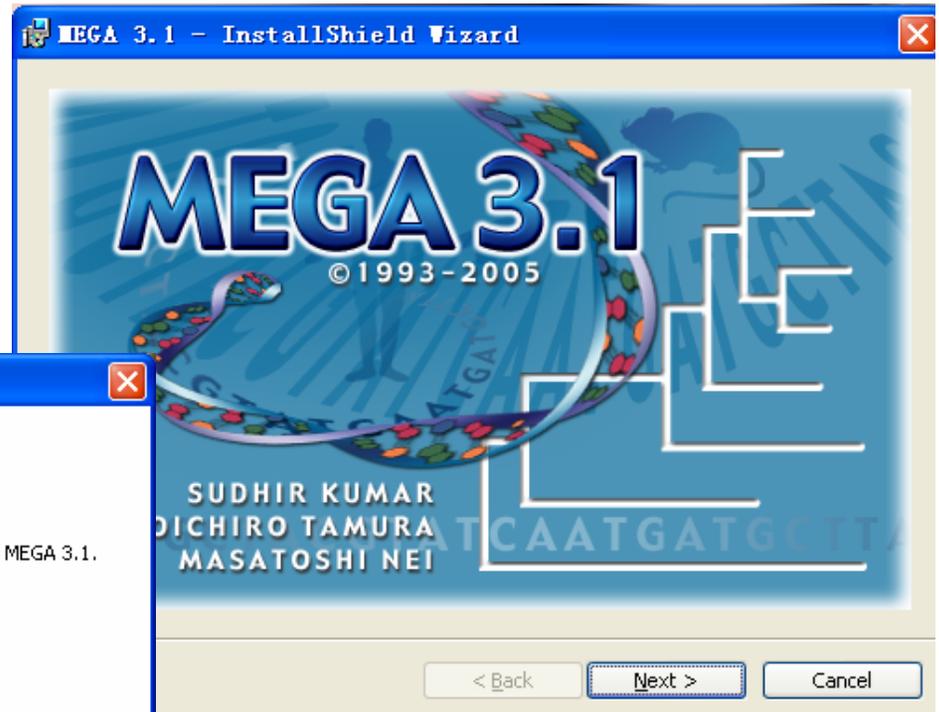


# MEGA软件的使用初探

胡亚平

82101061040

# 一、MEGA的安装和主操作界面



- MEGA软件的安装过程与其他在windows下使用的软件一样，比较方便。



- MEGA软件同样具有纤细的使用说明，我们可以按照使用说明一步步进行操作。本次作业主要使用它的序列比对和构建系统发育树的功能。

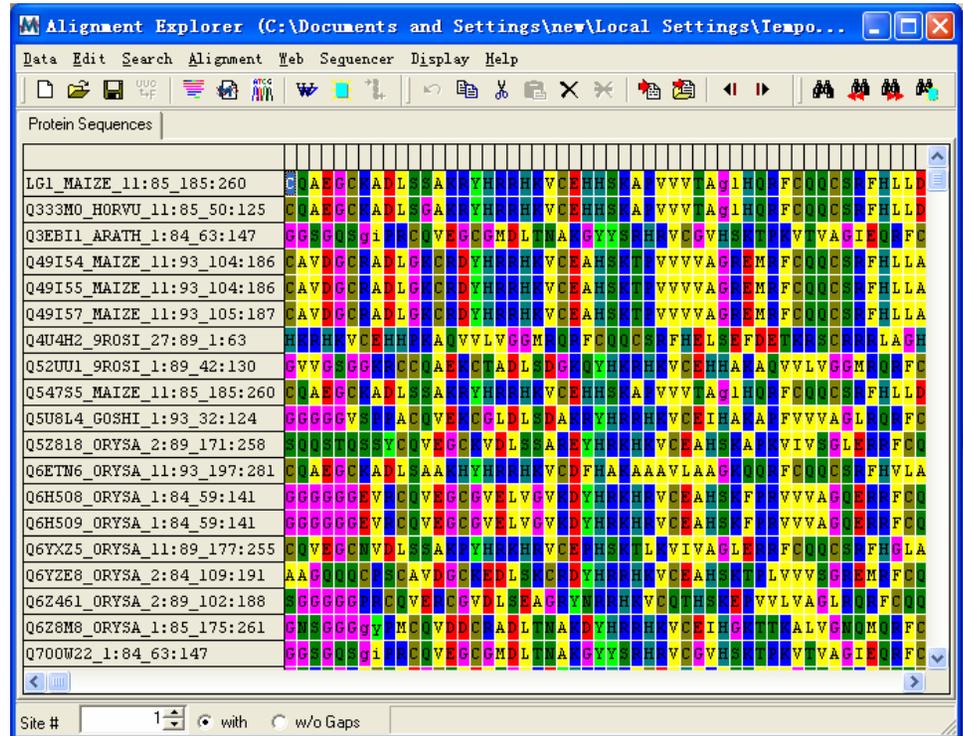
## 二、利用MEGA进行序列比对

Outline  
Lectures  
**Samples**  
Manuals  
References  
Exercises  
Homework  
Examples  
Exam  
Download

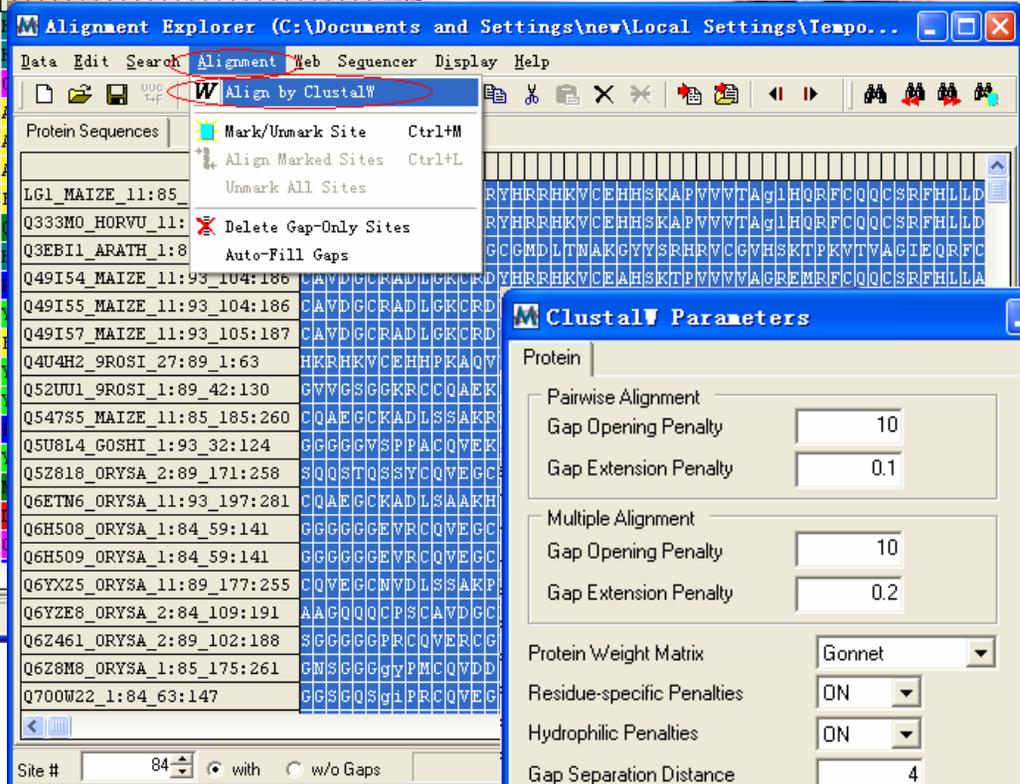
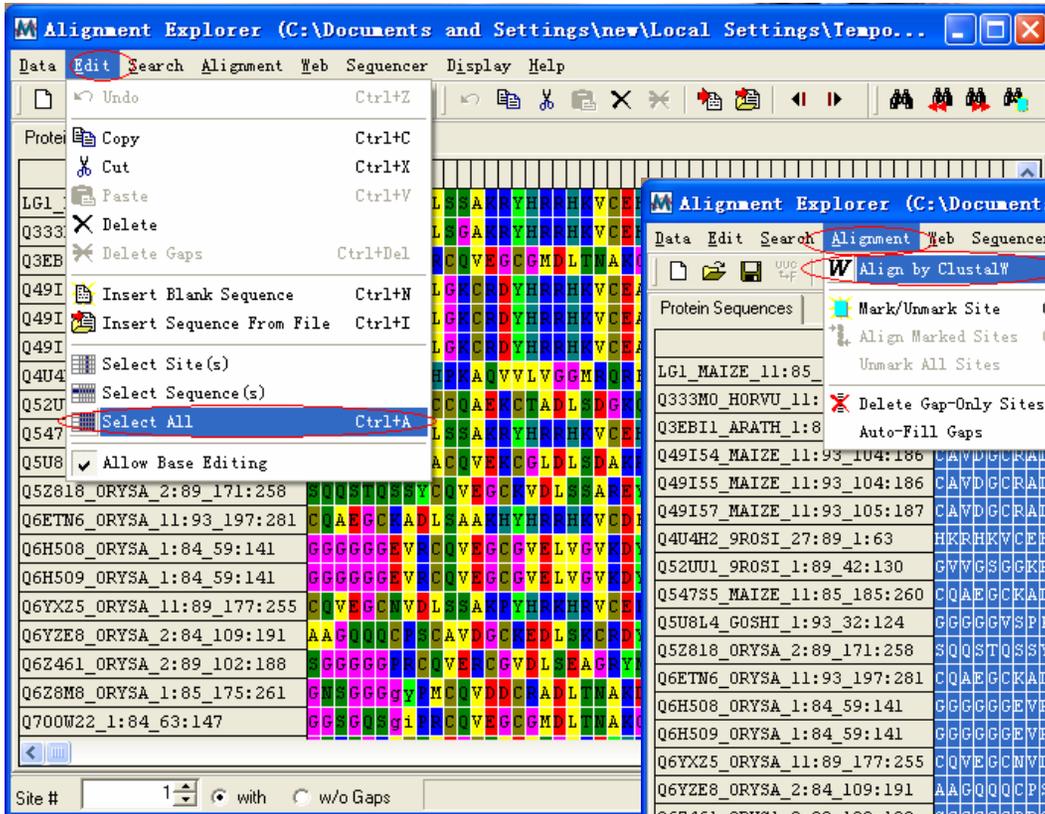
Notice (9-Dec-2006)

this week, we will learn how to use SRS efficiently. We will also do hands-on practice to predict functional domains for a set of sequences such as the 3P transcription factor family.

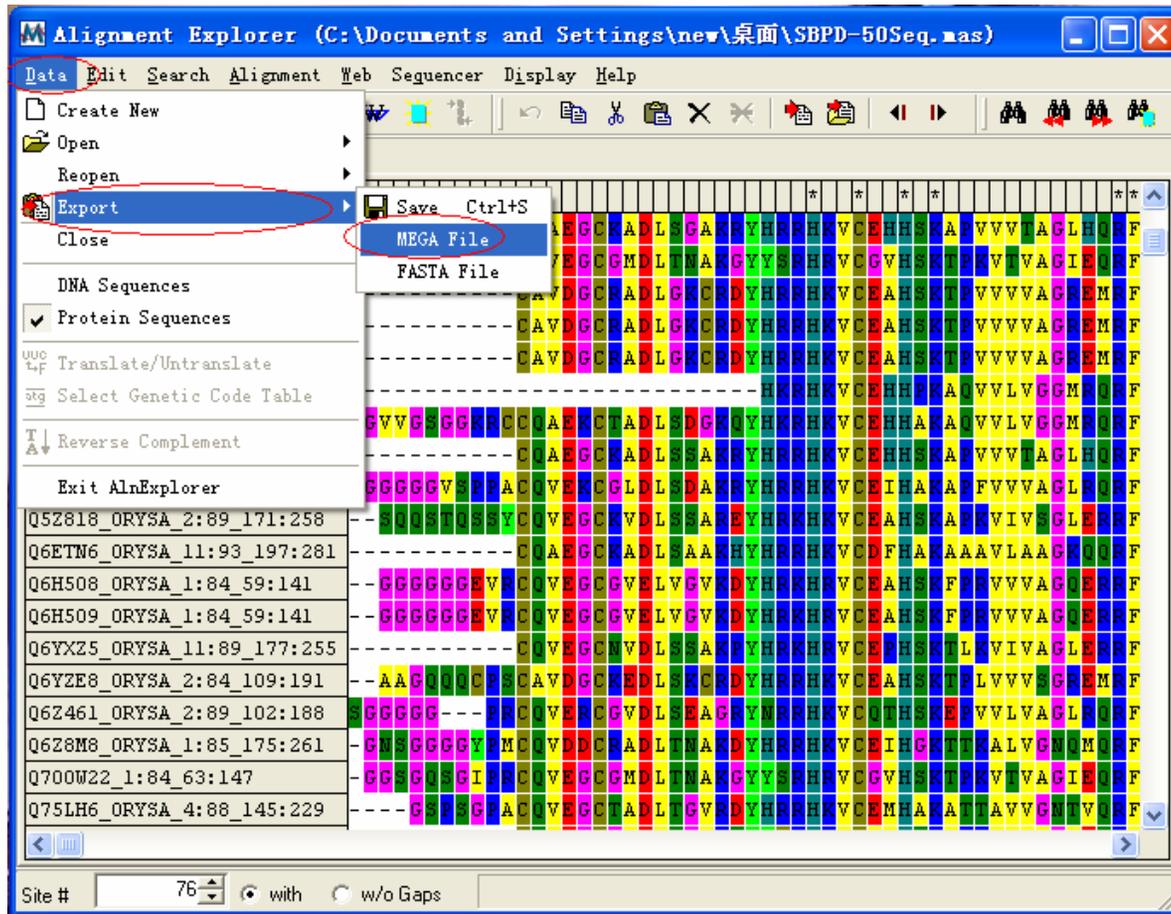
- [MDR2](#) - Multi-dr
- [MDR3](#) - Multi-dr
- Protein sequences
  - [PPF-1](#) - Pisum sa
  - [OsHT](#) - Rice hist
  - [AtLHT1](#) - Arabid
  - [1UL5\\_ARATH](#) - C
  - [SPL7\\_ARATH](#) - A
  - [SPL8\\_ARATH](#) - A
  - [CEA\\_HUMAN](#) - H
  - [CEA\\_MOUSE](#) - M
  - [MDR2\\_FUGU](#) - M
  - [MDR3\\_FUGU](#) - M
  - [Q57997](#) - ATP bi
  - [NP\\_149073](#) - Cyl
  - [XP\\_372459](#) - Olf
  - [HBA\\_Human](#) - H
  - [LGB2\\_LUPLU](#) - Y
  - [GST7\\_CAEEL](#) - N
- 3D structures
  - [3INS](#) [PDB, Cn3]
  - [1A4F](#) [PDB, Cn3]
  - [1QK6](#) [PDB, Cn3]
  - [1L6Z](#) [PDB, Cn3]
  - [1REI](#) [PDB, Cn3]
  - [1VCU](#) [PDB, Cn3]
- Phylogeny data
  - [6DNA](#) [Fasta, Ph
  - sequences for te
  - [LHT-9SEQ](#) [Fasta
  - sequences of Ly:
  - [SBPD-10Seq](#) - DI
  - [SBPD-Rice](#) - N-te
  - [SBPD-50Seq](#) - DI
  - [HBA-13Seq](#) - Hei
  - [HBA-15Seq](#) - Hei
  - [SARS20](#) - 30 site
  - [NDHF12](#) - Inform
  - (FASTA)



- 在ABC的网站中可以直接用MEGA打开后缀为\*.fasta格式的序列。如果我们使用自己的序列，可以建一个TXT的文档，输入fasta格式的一系列序列，保存时命名为\*.fasta文件就可以进行被MEGA所正确识别。我们先选取SBPD-50Seq进行练习。



- 序列被正确识别后，我们可以利用MEGA进行序列比对，先edit/select all选中要比对的序列，然后在alignment/align by clustalw进行比对，在弹出的界面可以更改参数，点击OK得到结果。



- 得到的序列比对结果可以保存为**MEG**格式的文件，以便进行后续的建立系统发育树等操作。以\*.fasta格式保存的文件进行序列比对的过程也基本相同。

### 三、利用MEGA在NCBI等网站下载并保存数据

The image displays two overlapping software windows. The foreground window is the 'MEGA Web Browser (Entrez Protein)'. Its address bar contains the URL <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein>. The search bar is set to 'Protein' and contains the query 'mnd1'. The 'Go' button is highlighted. Below the search bar, there is a yellow information box about Entrez protein entries and a 'Human Genome' section. The background window is 'Alignment Explorer (C:\Documents and Settings\new\桌面\SBPD-50Seq.mas)'. It shows a list of protein sequences on the left, including Q333M0\_HORVU\_11:85\_50:125, Q3EBI1\_ARATH\_1:84\_63:147, Q49I54\_MAIZE\_11:93\_104:186, Q49I55\_MAIZE\_11:93\_104:186, and Q49I57\_MAIZE\_11:93\_105:187. A 'Web Sequencer' menu is open, showing options like 'Query Genbank' and 'Do BLAST Search'. The main area of the Alignment Explorer shows a multi-sequence alignment of amino acid residues, with gaps indicated by dashes.

MEGA可以作为浏览器查看NCBI等网站的页面，可以在其中查找感兴趣的基因序列并保存在MEGA中

MEGA Web Browser (Entrez Protein)

Data Edit Option View Links Go Help

Address <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=protein>

Links [NCBI](#)

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search Protein for mnd1 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

All: 95 Bacteria: 2 RefSeq: 42 Related Structures: 49

Items 1 - 20 of 95 Page 1 of 5 Next

1: [AAQ24512](#) Reports BLink, Conserved Domains, Links  
Mnd1 [Giardia intestinalis]  
gi|33667824|gb|AAQ24512.1|[33667824]

2: [ABD91830](#) Reports BLink, Conserved Domains, Links  
mnd1 [Acanthamoeba castellanii]  
gi|90192335|gb|ABD91830.1|[90192335]

3: [NP\\_084073](#) Reports BLink, Conserved Domains, Links  
GAJ protein [Mus musculus]  
gi|37591185|ref|NP\_084073.1|[37591185]

4: [NP\\_115493](#) Reports BLink, Conserved Domains, Links  
GAJ protein [Homo sapiens]  
gi|14149769|ref|NP\_115493.1|[14149769]

5: [ABB73190](#) Reports BLink, Conserved Domains, Links  
meiotic nuclear division 1-like protein [Arabidopsis thaliana]  
gi|82408770|gb|ABB73190.1|[82408770]

MEGA Web Browser (NCBI Sequence Viewer v2.0)

Data Edit Option View Links Go Help

Address <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&val=33667824>

Links [NCBI](#)

NCBI Protein

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Protein for Go Clear

Limits Preview/Index History Clipboard Details

Display GenPept Show 5 Send to

Range: from begin to end Features:  CDD + Refresh

1: [AAQ24512](#) Reports [Mnd1](#) [Giardia int... [gi:33667824] BLink, Conserved Domains, Links

Features Sequence

LOCUS AAQ24512 203 aa linear INV 08-FEB-2005

DEFINITION Mnd1 [Giardia intestinalis].

ACCESSION AAQ24512

VERSION AAQ24512.1 GI:33667824

DBSOURCE accession [AY295092.1](#)

KEYWORDS .

SOURCE Giardia intestinalis (Giardia lamblia)

ORGANISM [Giardia intestinalis](#)

Eukaryota; Diplomonadida; Hexamitidae; Giardiainae; Giardia.

REFERENCE 1 (residues 1 to 203)

AUTHORS Ramesh, M.A., Malik, S.B. and Logsdon, J.M. Jr.

TITLE A phylogenomic inventory of meiotic genes; evidence for sex in giardia and an early eukaryotic origin of meiosis

JOURNAL Curr. Biol. 15 (2), 185-191 (2005)

PUBMED [15668177](#)

REFERENCE 2 (residues 1 to 203)

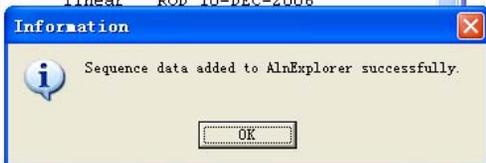
AUTHORS Ramesh, M.A. and Logsdon, J.M. Jr.

TITLE Direct Submission

JOURNAL Submitted (28-APR-2003) Biology, Emory University, 1510 Clifton Road, Atlanta, GA 30322, USA

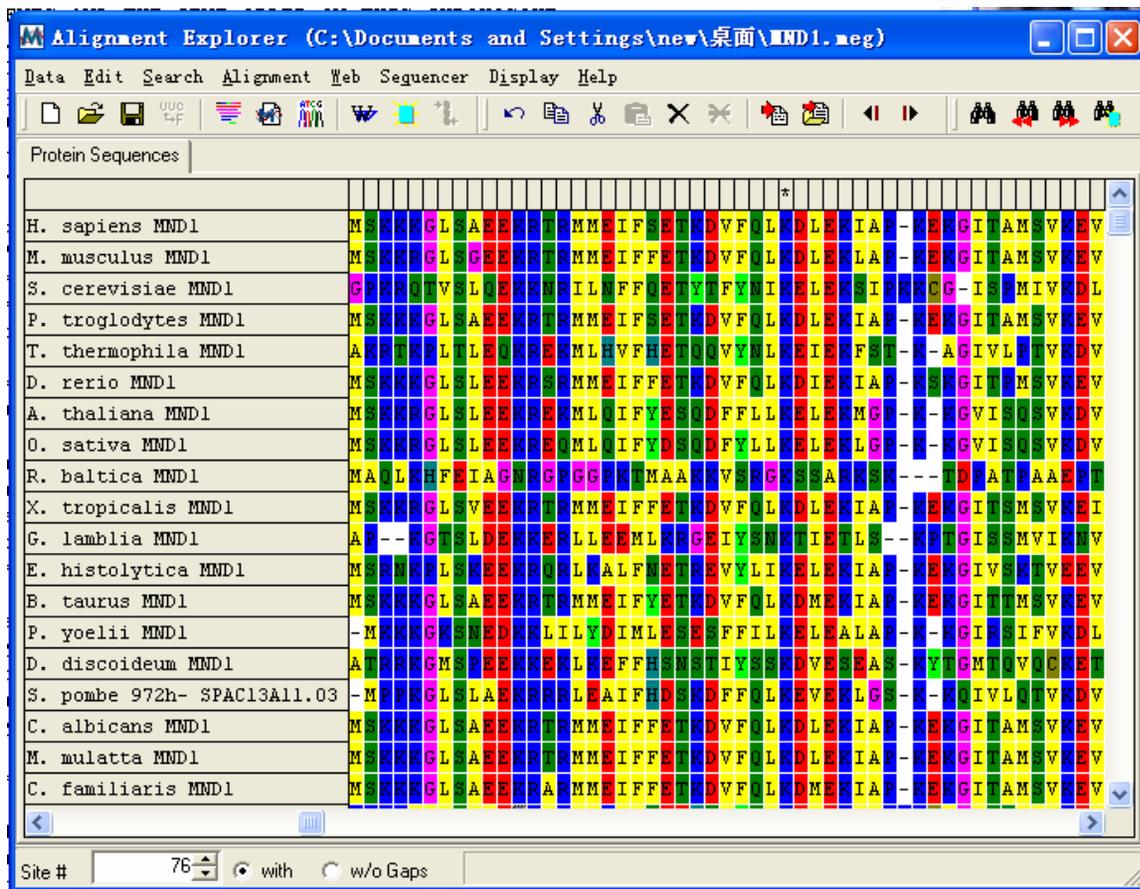
COMMENT Method: conceptual translation supplied by author.

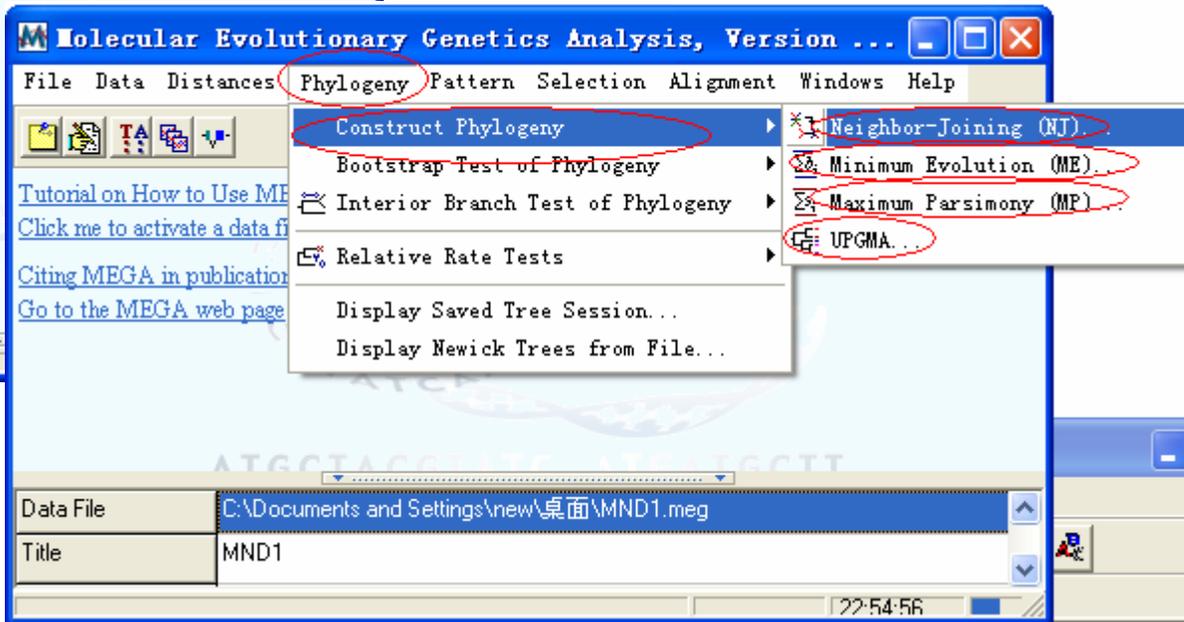
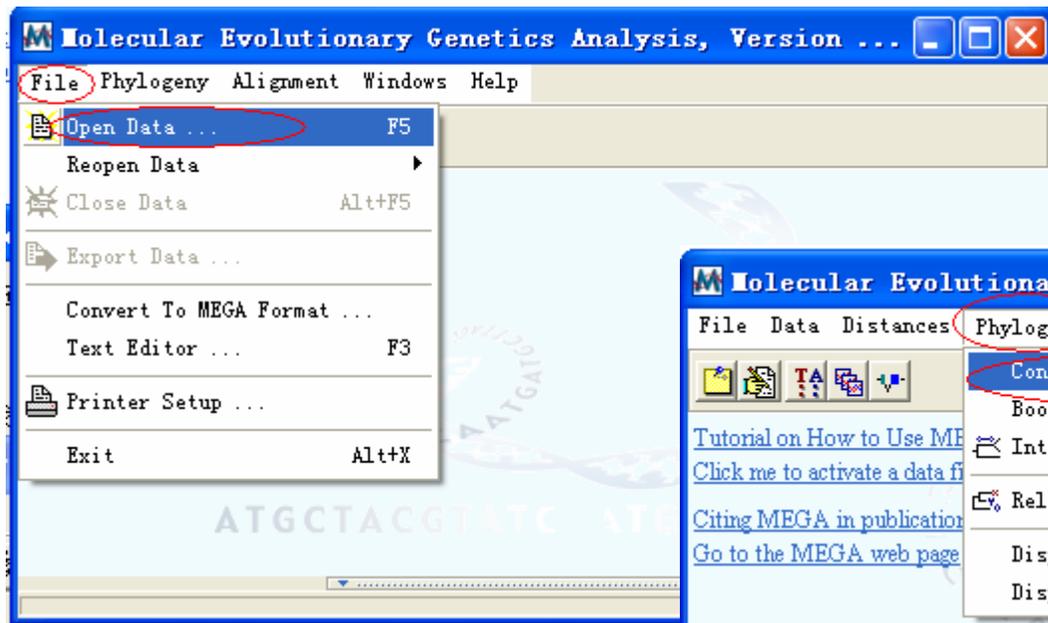
- 在得到需要的序列后可以点击add to alignment保存序列。



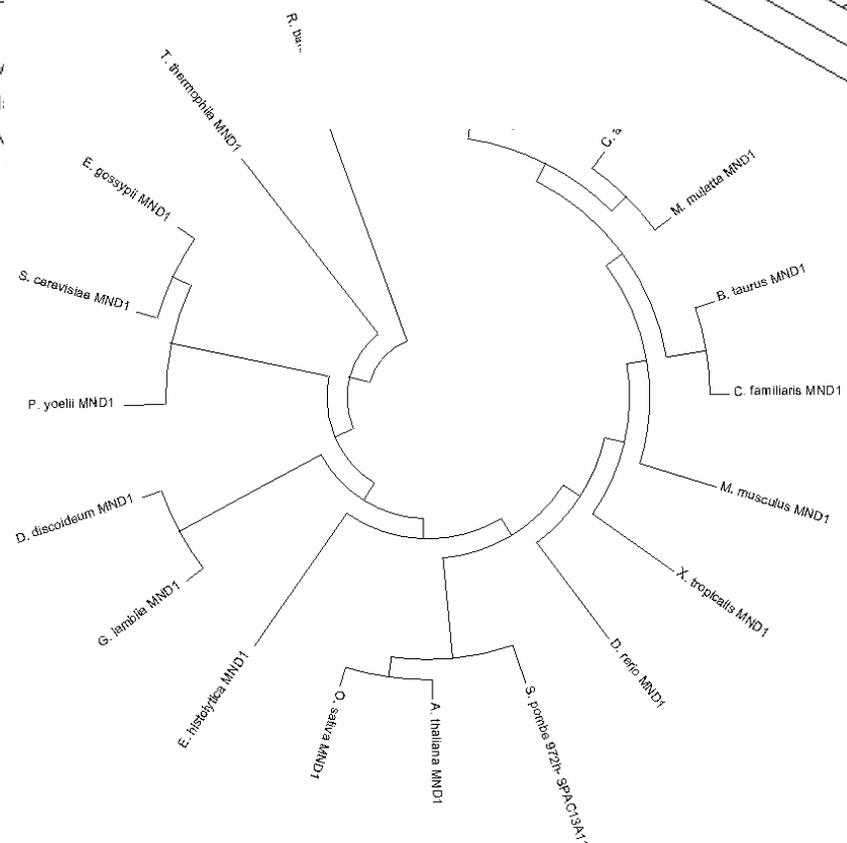
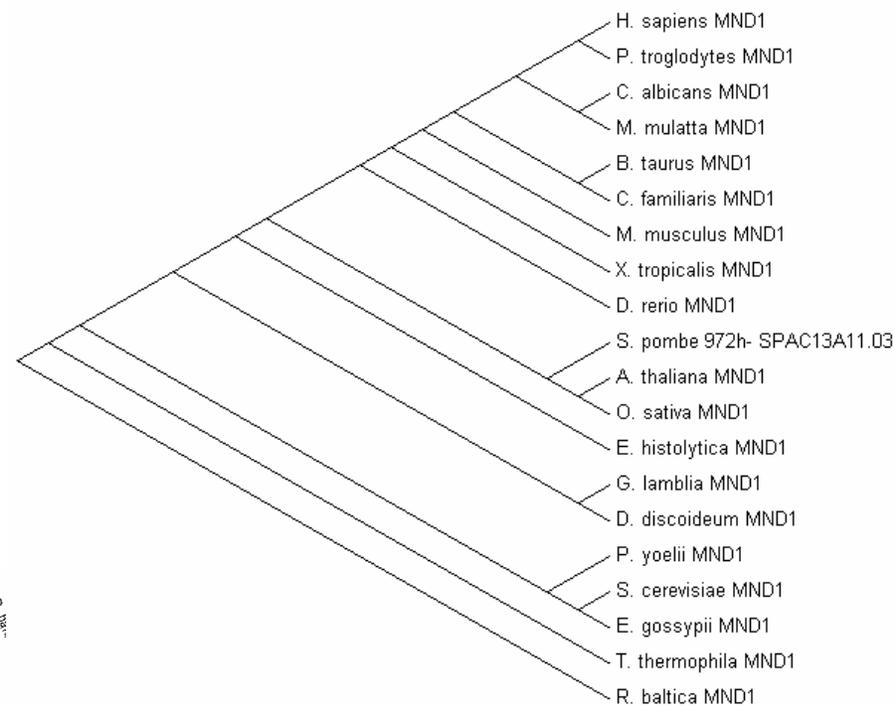
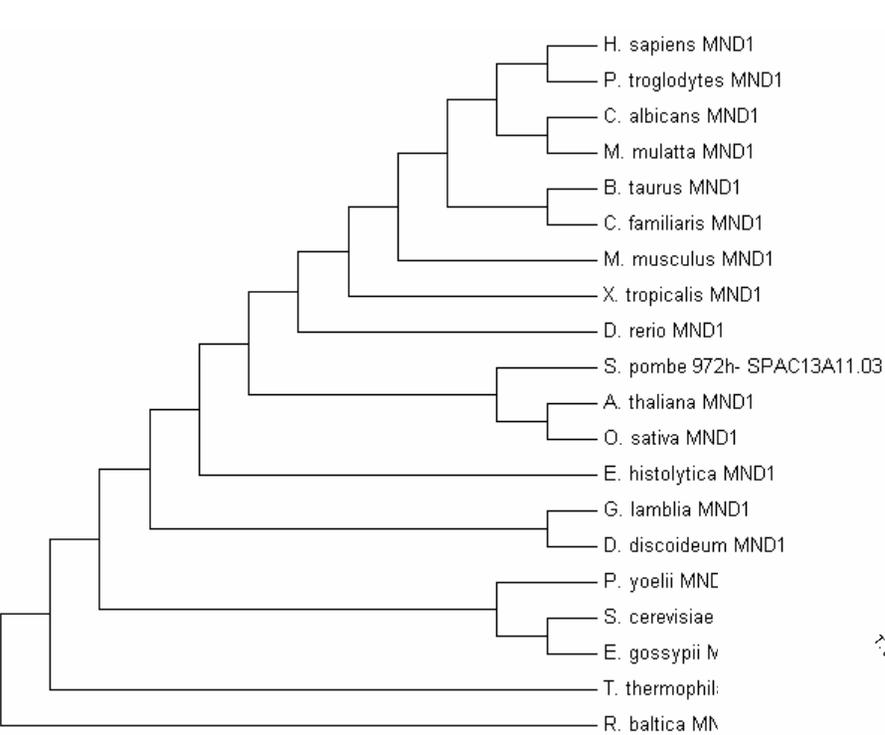
raniata; Vertebrata; Eucariotomi;  
 lires; Glires; Rodentia;  
 ; Murinae; Mus.

- 保存的序列可以进行比对后保存为 **\*.MEG**，以便进行后续操作，也可以保存为 **\*.fasta** 格式的文件。





- \*.MEG格式的文件可以用来构建系统发育树，我可以构建四种类型的树，分别为 neighbor joining、maximum parsimony、minimum evolution和UPGMA四种算法。



我用neighbor joining  
 法构建的减数分裂基  
 因mnd1在不同物种中  
 同源基因的系统发育  
 树。可以用不同的模  
 式显示。

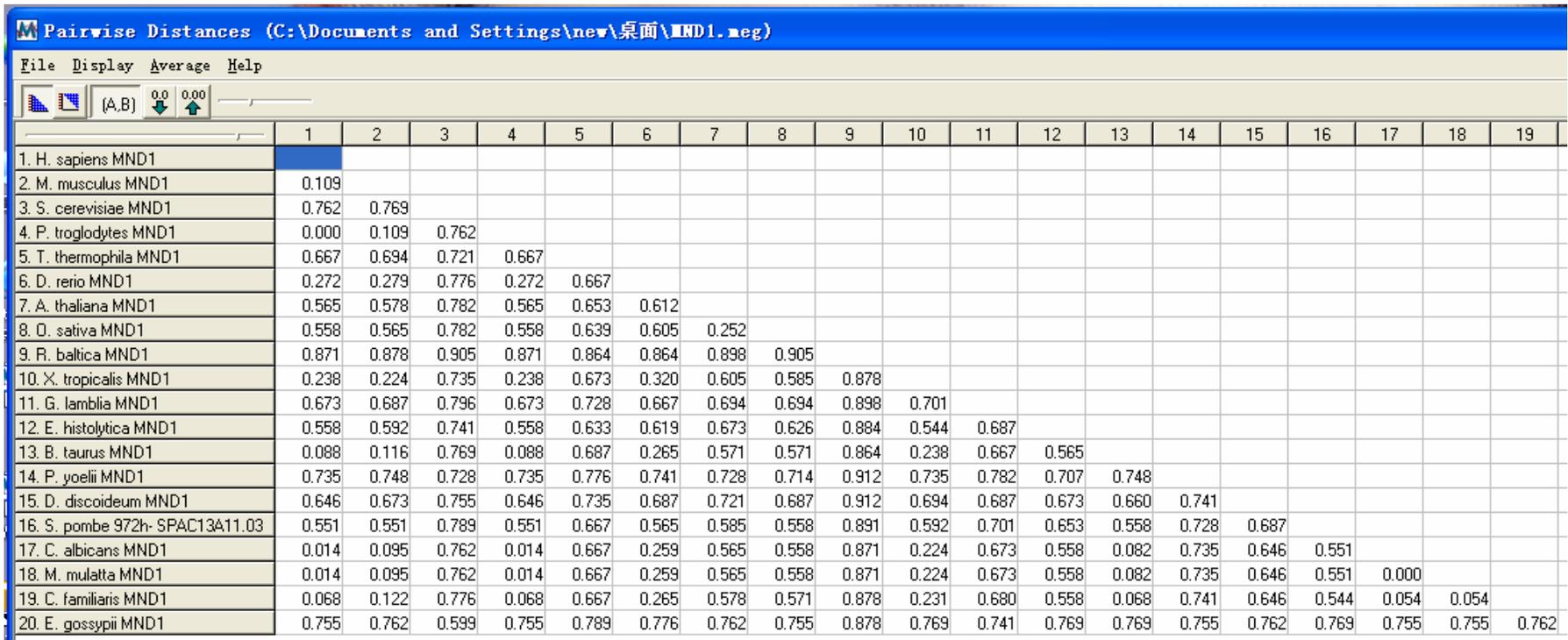
# 四、用MEGA估算进化距离

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. H. sapiens MND1																				
2. M. musculus MND1	0.115																			
3. S. cerevisiae MND1	1.435	1.464																		
4. P. troglodytes MND1	0.000	0.115	1.435																	
5. T. thermophila MND1	1.099	1.184	1.277	1.099																
6. D. rerio MND1	0.318	0.327	1.494	0.318	1.099															
7. A. thaliana MND1	0.832	0.863	1.525	0.832	1.059	0.947														
8. D. sativa MND1	0.816	0.832	1.525	0.816	1.020	0.930	0.290													
9. R. baltica MND1	2.046	2.100	2.351	2.046	1.995	1.995	2.282	2.351												
10. X. tropicalis MND1	0.272	0.254	1.327	0.272	1.119	0.385	0.930	0.880	2.100											
11. G. lamblia MND1	1.119	1.162	1.589	1.119	1.302	1.099	1.184	1.184	2.282	1.206										
12. E. histolytica MND1	0.816	0.896	1.353	0.816	1.001	0.965	1.119	0.983	2.157	0.786	1.162									
13. B. taurus MND1	0.093	0.123	1.464	0.093	1.162	0.308	0.847	0.847	1.995	0.272	1.099	0.832								
14. P. yoelii MND1	1.327	1.380	1.302	1.327	1.494	1.353	1.302	1.253	2.425	1.327	1.525	1.229	1.380							
15. D. discoideum MND1	1.039	1.119	1.407	1.039	1.327	1.162	1.277	1.162	2.425	1.184	1.162	1.119	1.078	1.353						
16. S. pombe 972h- SPAC13A11.03	0.801	0.801	1.556	0.801	1.099	0.832	0.880	0.816	2.218	0.896	1.206	1.059	0.816	1.302	1.162					
17. C. albicans MND1	0.014	0.100	1.435	0.014	1.099	0.299	0.832	0.816	2.046	0.254	1.119	0.816	0.085	1.327	1.039	0.801				
18. M. mulatta MND1	0.014	0.100	1.435	0.014	1.099	0.299	0.832	0.816	2.046	0.254	1.119	0.816	0.085	1.327	1.039	0.801	0.000			
19. C. familiaris MND1	0.070	0.131	1.494	0.070	1.099	0.308	0.863	0.847	2.100	0.263	1.140	0.816	0.070	1.353	1.039	0.786	0.056	0.056		
20. E. gossypii MND1	1.407	1.435	0.913	1.407	1.556	1.494	1.435	1.407	2.100	1.464	1.353	1.464	1.464	1.407	1.435	1.464	1.407	1.407	1.407	1.435

- 打开前面保存的\*.meg文件，点击Distance|Compute Pairwise command (F7) 显示操作参数；设定好Models等参数后点击Compute启动计算，得到如上图的结果。

# MEGA也可以用来显示不同序列

## 氨基酸组成比例的差异



Pairwise Distances (C:\Documents and Settings\new\桌面\MND1.meg)

File Display Average Help

(A,B) 0.0 0.00

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. H. sapiens MND1																			
2. M. musculus MND1	0.109																		
3. S. cerevisiae MND1	0.762	0.769																	
4. P. troglodytes MND1	0.000	0.109	0.762																
5. T. thermophila MND1	0.667	0.694	0.721	0.667															
6. D. rerio MND1	0.272	0.279	0.776	0.272	0.667														
7. A. thaliana MND1	0.565	0.578	0.782	0.565	0.653	0.612													
8. D. sativa MND1	0.558	0.565	0.782	0.558	0.639	0.605	0.252												
9. R. baltica MND1	0.871	0.878	0.905	0.871	0.864	0.864	0.898	0.905											
10. X. tropicalis MND1	0.238	0.224	0.735	0.238	0.673	0.320	0.605	0.585	0.878										
11. G. lamblia MND1	0.673	0.687	0.796	0.673	0.728	0.667	0.694	0.694	0.898	0.701									
12. E. histolytica MND1	0.558	0.592	0.741	0.558	0.633	0.619	0.673	0.626	0.884	0.544	0.687								
13. B. taurus MND1	0.088	0.116	0.769	0.088	0.687	0.265	0.571	0.571	0.864	0.238	0.667	0.565							
14. P. yoelii MND1	0.735	0.748	0.728	0.735	0.776	0.741	0.728	0.714	0.912	0.735	0.782	0.707	0.748						
15. D. discoideum MND1	0.646	0.673	0.755	0.646	0.735	0.687	0.721	0.687	0.912	0.694	0.687	0.673	0.660	0.741					
16. S. pombe 972h- SPAC13A11.03	0.551	0.551	0.789	0.551	0.667	0.565	0.585	0.558	0.891	0.592	0.701	0.653	0.558	0.728	0.687				
17. C. albicans MND1	0.014	0.095	0.762	0.014	0.667	0.259	0.565	0.558	0.871	0.224	0.673	0.558	0.082	0.735	0.646	0.551			
18. M. mulatta MND1	0.014	0.095	0.762	0.014	0.667	0.259	0.565	0.558	0.871	0.224	0.673	0.558	0.082	0.735	0.646	0.551	0.000		
19. C. familiaris MND1	0.068	0.122	0.776	0.068	0.667	0.265	0.578	0.571	0.878	0.231	0.680	0.558	0.068	0.741	0.646	0.544	0.054	0.054	
20. E. gossypii MND1	0.755	0.762	0.599	0.755	0.789	0.776	0.762	0.755	0.878	0.769	0.741	0.769	0.769	0.755	0.762	0.769	0.755	0.755	0.762

- MEGA还可以计算氨基酸比例的差异，具体操作如下：打开前面保存的\*.meg文件，点击Distance / Compute Pairwise command (F7) 显示操作参数；设定好Models为Amino Acid / p-distance等参数后点击Compute启动计算，得到如上图的结果。