

# 多序列比对与分子进化树

日期: 2022-11-9

实验者: 生信 2001 张子栋

[MarkdownNotes/软件第3次作业.md at main · Bluuur/MarkdownNotes \(github.com\)](#)

[生物信息学原理/软件第3次作业.md · blur/MarkdownNotes - 码云 - 开源中国 \(gitee.com\)](#)

## 实验目的

1. 掌握至少一种常用多序列比对软件的使用, 并理解关键参数的含义, 熟悉常用多序列比对文件格式, 了解不同多序列比对软件的比对性能差异。
2. 了解比对编辑与美化软件的使用
3. 会使用至少一种软件构建系统发育树, 并理解关键参数的含义, 熟悉进化树的不同存储格式和结果展示方式, 了解不同算法构建进化树的差异。

## 实验内容

1. 常用多序列比对软件进行DNA或蛋白质多序列比对的使用
  - Clustal在线与本地版
  - T-coffee在线版
  - MUSCLE 在线版
2. 使用比对编辑美化软件的使用 (选做)
  - Jalview
  - ESPript
3. 分子进化与系统发育树软件的使用
  - MEGA
4. 进化树的编辑与美化 (选做)

## 实验步骤

1. 在基因重组人胰岛素面市之前, 糖尿病患者所需胰岛素主要来自屠宰场的动物胰脏。请分析来源自猪、牛和羊的胰岛素哪一种最适于使用, 说明理由。四种蛋白的 NCBI 注册号分别是 `AAA59172` (人), `AAQ00954` (猪), `AAA30722` (牛) 和 `P01318` (羊)。

使用软件: MEGA 11.0.13

### 下载数据

[insulin \[Homo sapiens\] - Protein - NCBI \(nih.gov\)](#)

[preproinsulin \[Sus scrofa\] - Protein - NCBI \(nih.gov\)](#)

[insulin precursor \[Bos taurus\] - Protein - NCBI \(nih.gov\)](#)

[RecName: Full=Insulin; Contains: RecName: Full=Insulin B chain; Contai - Protein - NCBI \(nih.gov\)](#)

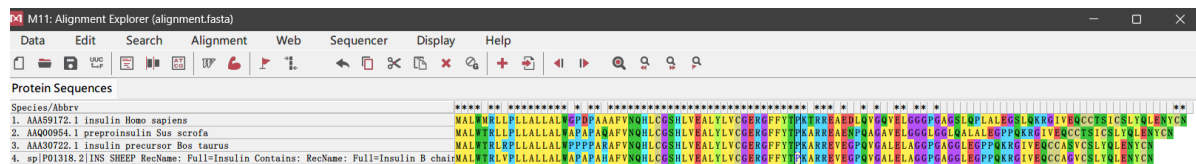
```
1 >AAA59172.1 insulin [Homo sapiens]
2 MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGG
```

```

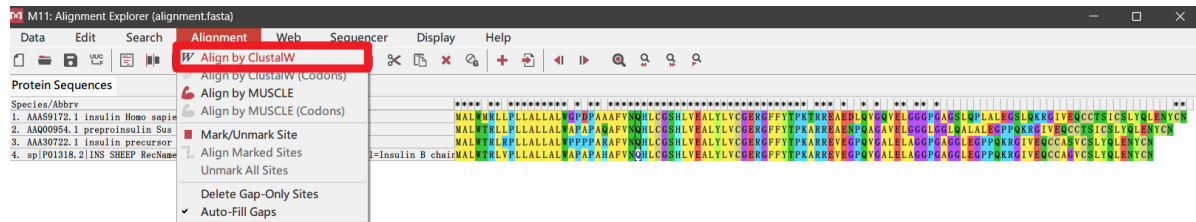
3  GPGAGSLQPLALEGSLQKRGIVEQCCTSIICSLYQLENYCN
4
5  >AAQ00954.1 preproinsulin [Sus scrofa]
6  MALWTRLPLLLALLALWAPAPAAQAFVNQHLGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGG
7  GLGGLQALALEGPPQKRGIVEQCCTSIICSLYQLENYCN
8
9  >AAA30722.1 insulin precursor [Bos taurus]
10 MALWTRLRPLLLALLALWPPPPARAFVNQHLGSHLVEALYLVCGERGFFYTPKARREVEGPQVGALELAG
11 GPGAGGLEGPPQKRGIVEQCCASVCSLYQLENYCN
12
13 >sp|P01318.2|INS_SHEEP RecName: Full=Insulin; Contains: RecName:
    Full=Insulin B chain; Contains: RecName: Full=Insulin A chain; Flags:
    Precursor
14 MALWTRLVPLLLALLALWAPAPAHAFVNQHLGSHLVEALYLVCGERGFFYTPKARREVEGPQVGALELAG
15 GPGAGGLEGPPQKRGIVEQCCAGVCSLYQLENYCN

```

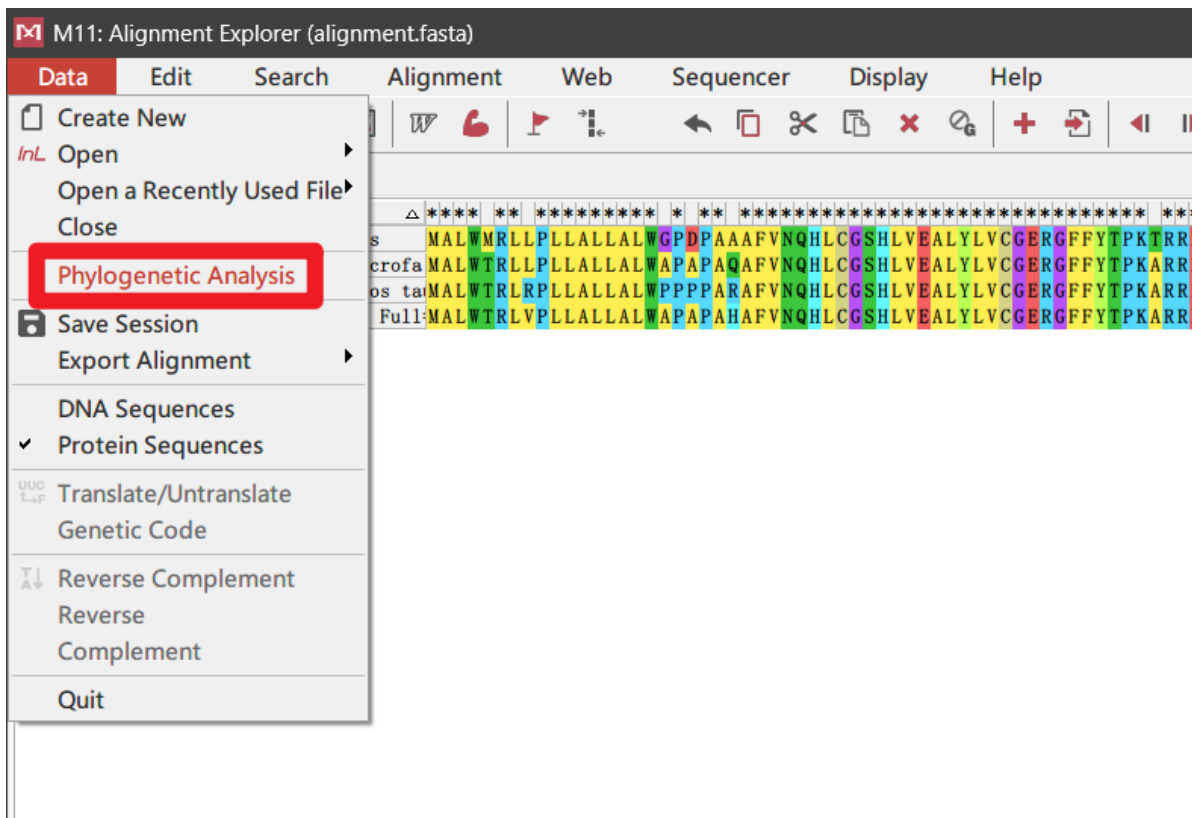
- 导入数据到 MEGA



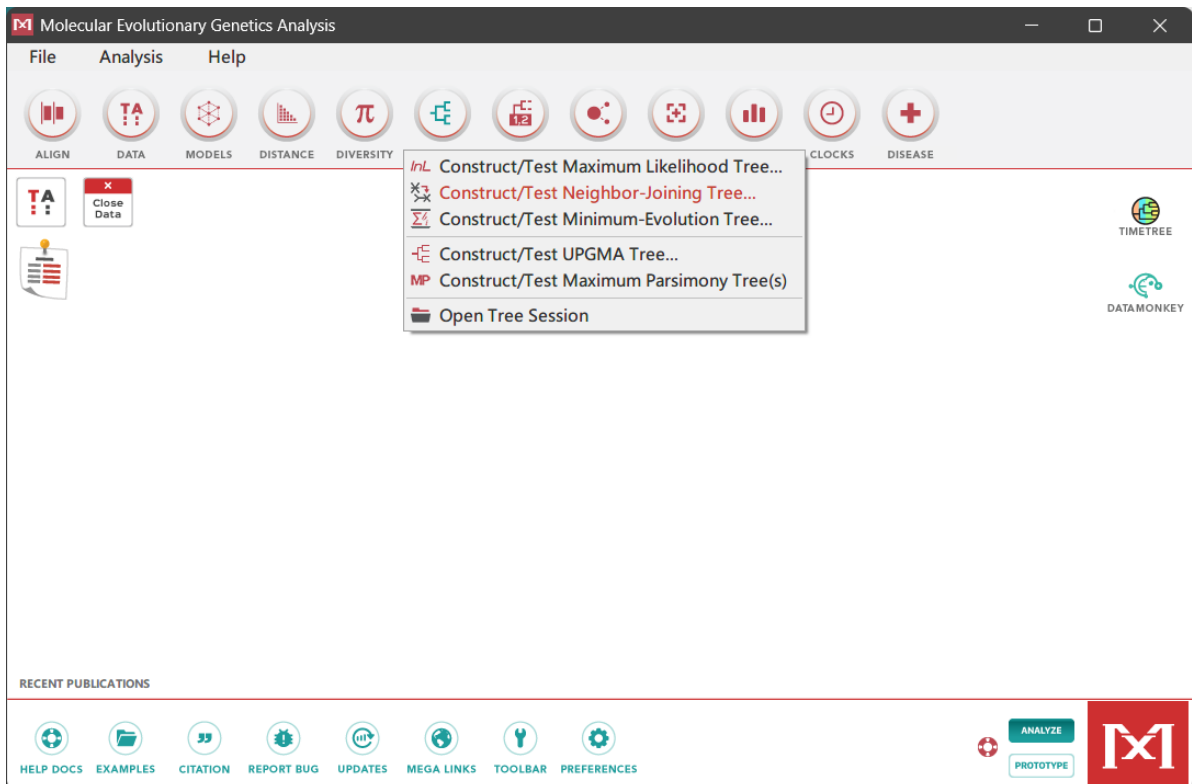
- ClustaW 序列比对

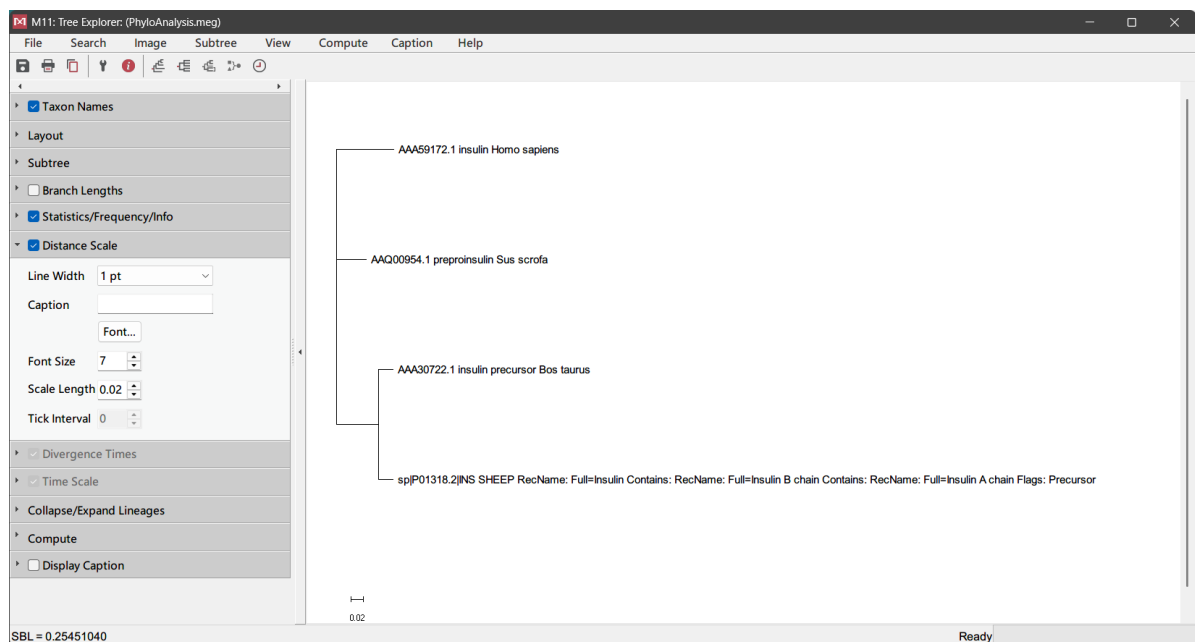


- 数据分析

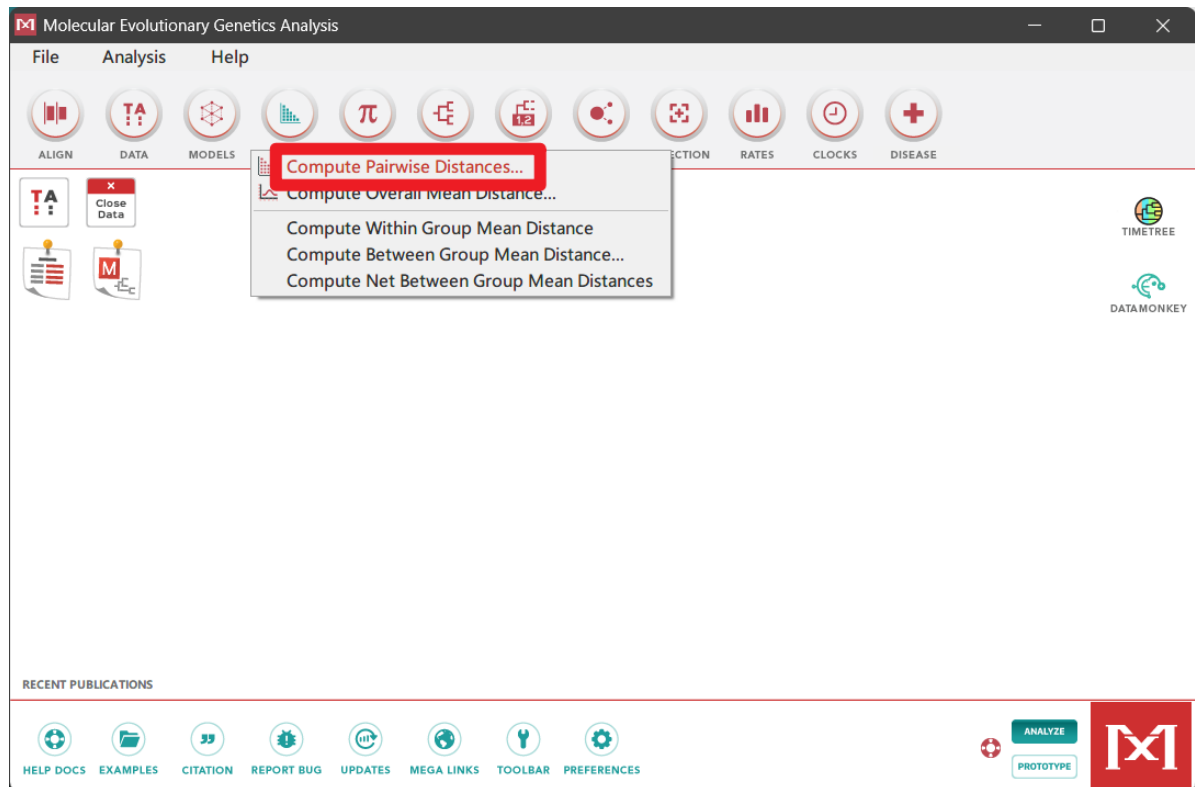


- 建树 (NJ 法)
  - 均选择默认选项





- 构建距离矩阵



M11: Pairwise Distances (PhyloAnalysis.meg)

File Display Average Caption Help

	1	2	3	4
1. AAA59172.1 insulin Homo sapiens				
2. AAQ00954.1 preproinsulin Sus scrofa	0.1388			
3. AAA30722.1 insulin precursor Bos taurus	0.1766	0.1446		
4. spP01318.2 INS SHEEP RecName: Full=Insulin	0.1881	0.1335	0.0488	

[2,1] (AAQ00954.1 preproinsulin Sus scrofa-AAA59172.1 insulin Homo sapiens) / Amino: Poisson

- 由以上结果可以得出，猪胰岛素与人胰岛素序列遗传距离更近，相似性更好。

1. Keratin 是一种微管蛋白，有 type I 和 type II 两种类型，在染色体上成簇分布，对上皮细胞的正常结构十分重要。请根据人类 type II keratin 2p 蛋白序列（NCBI 注册号：CAD91891）对 NCBI Homo sapiens RefSeq protein 序列数据库进行 BLASTP 检索，根据检索结果，下载人基因组中 E-value<1e-200 的匹配的 RefSeq 序列，按照 Total Score 的顺序由大到小重新命名（比如 1-10）后进行分子进化分析。

1. 任选一种工具，进行多序列比对，并根据比对结果列出距离矩阵
2. 任选一种工具，构建分子进化树并对各分支的可靠性进行评估。

- 获取序列

Protein BLAST: search protein da x +

https://blast.ncbi.nlm.nih.gov/Blast.cgi...

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. more...

Reset page

Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

CAD91891

Query subrange ?

From

To

Or, upload file

选择文件 未选择文件 ?

Job Title

CAD91891:keratin 2p [Homo sapiens]

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Databases

☒ Standard databases (nr etc.): **New** ☐ Experimental databases

Try experimental clustered nr database ?

For more info see What is clustered nr?

Compare

☐ Select to compare standard and experimental database ?

Standard

Database

RefSeq Select proteins (refseq\_select)

Organism

Optional

human (taxid:9606) ☐ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. ?

Exclude

Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm ?

BLAST

Search database RefSeq Select proteins (refseq\_select) using Blastp (protein-protein BLAST)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦

sign

NCBI BlastCAD91891:keratin 2p x +

https://blast.ncbi.nlm.nih.gov/Blast.cgi

An official website of the United States government [Here's how you know](#)

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Log in

BLAST® » blastp suite » results for RID-RT1S0G67016

Home Recent Results Saved Strategies Help

[Edit Search](#) Save Search Search Summary

How to read this report? BLAST Help Videos Back to Traditional Results Page

Your search is limited to records that include: human (taxid:9606)

Job Title CAD91891:keratin 2p [Homo sapiens]

RID RT1S0G67016 Search expires on 11-23 15:23 pm Download All

Program BLASTP Citation

Database refseq\_select\_prot See details

Query ID CAD91891.1

Description keratin 2p [Homo sapiens]

Molecule type amino acid

Query Length 638

Other reports Distance tree of results Multiple alignment MSA viewer

**Filter Results**

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity E value Query Coverage

to to 1e-200 to

Filter Reset

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 64 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
keratin, type II cytoskeletal 2 oral [Homo sapiens]	Homo sapiens	1256	1256	100%	0.0	100.00%	638	NP_056932.2
keratin, type II cytoskeletal 3 [Homo sapiens]	Homo sapiens	627	627	54%	0.0	86.78%	628	NP_476429.2
keratin, type II cytoskeletal 75 [Homo sapiens]	Homo sapiens	587	587	79%	0.0	62.65%	551	NP_004684.2
keratin, type II cytoskeletal 6B [Homo sapiens]	Homo sapiens	586	586	79%	0.0	64.90%	564	NP_005546.2
keratin, type II cytoskeletal 6C [Homo sapiens]	Homo sapiens	584	584	79%	0.0	64.71%	564	NP_775109.2
keratin, type II cytoskeletal 6A [Homo sapiens]	Homo sapiens	584	584	72%	0.0	68.48%	564	NP_005545.1
keratin, type II cytoskeletal 5 [Homo sapiens]	Homo sapiens	565	565	83%	0.0	61.17%	590	NP_000415.2
keratin, type II cytoskeletal 79 [Homo sapiens]	Homo sapiens	563	563	73%	0.0	62.87%	535	NP_787028.1
keratin, type II cytoskeletal 73 [Homo sapiens]	Homo sapiens	562	562	88%	0.0	57.09%	540	NP_778238.1
keratin, type II cytoskeletal 4 [Homo sapiens]	Homo sapiens	557	557	73%	0.0	60.79%	520	NP_002263.3
keratin, type II cytoskeletal 71 [Homo sapiens]	Homo sapiens	553	553	91%	0.0	55.67%	523	NP_258259.1
keratin, type II cytoskeletal 1 [Homo sapiens]	Homo sapiens	535	535	57%	0.0	68.94%	644	NP_006112.3

- 下载符合条件的序列

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 15 sequences selected GenPept FASTA (complete sequence) Hits Multiple alignment MSA Viewer

FASTA (aligned sequences)

GenBank (complete sequence)

Hit Table (text)

Hit Table (CSV)

Text

Descriptions Table (CSV)

XML

ASN.1

Description	Scientific Name	Query Cover	E value	Per. Ident	Acc. Len	Accession
keratin, type II cytoskeletal 2 oral [Homo sapiens]	Homo sapiens	100%	0.0	100.00%	638	NP_056932.2
keratin, type II cytoskeletal 3 [Homo sapiens]	Homo sapiens	54%	0.0	86.78%	628	NP_476429.2
keratin, type II cytoskeletal 75 [Homo sapiens]	Homo sapiens	79%	0.0	62.65%	551	NP_004684.2
keratin, type II cytoskeletal 6B [Homo sapiens]	Homo sapiens	79%	0.0	64.90%	564	NP_005546.2
keratin, type II cytoskeletal 6C [Homo sapiens]	Homo sapiens	79%	0.0	64.71%	564	NP_775109.2
keratin, type II cytoskeletal 6A [Homo sapiens]	Homo sapiens	72%	0.0	68.48%	564	NP_005545.1
keratin, type II cytoskeletal 5 [Homo sapiens]	Homo sapiens	83%	0.0	61.17%	590	NP_000415.2
keratin, type II cytoskeletal 79 [Homo sapiens]	Homo sapiens	73%	0.0	62.87%	535	NP_787028.1
keratin, type II cytoskeletal 73 [Homo sapiens]	Homo sapiens	88%	0.0	57.09%	540	NP_778238.1
keratin, type II cytoskeletal 4 [Homo sapiens]	Homo sapiens	73%	0.0	60.79%	520	NP_002263.3
keratin, type II cytoskeletal 71 [Homo sapiens]	Homo sapiens	91%	0.0	55.67%	523	NP_258259.1
keratin, type II cytoskeletal 1 [Homo sapiens]	Homo sapiens	57%	0.0	68.94%	644	NP_006112.3
keratin, type II cytoskeletal 74 [Homo sapiens]	Homo sapiens	78%	0.0	58.60%	529	NP_778223.2
keratin, type II cytoskeletal 72 isoform 1 [Homo sapiens]	Homo sapiens	71%	0.0	62.69%	511	NP_542785.1
keratin, type II cytoskeletal 1b [Homo sapiens]	Homo sapiens	71%	0.0	61.40%	578	NP_778253.2

```
1 >NP_056932.2 keratin, type II cytoskeletal 2 oral [Homo sapiens]
2 MNRQVCKKSFSGRSQGFSGRSASVVGSSRMSCVARSGGAGGGACGFRSGAGSFGSRSLYNLGSNKSISISVAAGS
  SRAGG
3 FGGGRSSCGFAGGYGGGFGGSYGGGFGGGRGVGSGFGGAGGFGGAGGFGGPGVFGGPGSFGGPGGFGPGGFGPGGI
  QEIV
```

4 NQSLQLPLNVEIDPQIGQVKAQEREQIKTLNNKFASFIDKVRFLEQQNKVLETKWELLQQQTTSGPSLSLEPCFE  
SYISF  
5 LCKQLDSL LGERNLEGE LKSMQDLVEDFKKKYEDEINKRTAAENEFVGLKKDVDAAFMNKVELQAKVDSL TDEV  
SFLRT  
6 LYEMELSQM SHASDTSVVLSMDNNRCLDLGSIIAEVRAQYEEIAQRSKSEAEALYQTKLGELQTTAGRHGDDL R  
NTKSE  
7 IMELNRM IQR LRAE IENVKKQANLQTAIAEAEQRGEMALKDANAKLQDLQTALQKAKDDLARLLRDYQELMNVK  
LALDV  
8 EIATYRK LLEGE ECRMSGECQSAVCISVVSNTSTSGSSGSSRGVFGGVSGSGSGGYKGGSSSSSSSGYGVSGGS  
GSGYG  
9 GVSSGSTGGRSSGSYQSSSSGSR LGAGSISVSHSGMGSSSGSIQTSGSGYKSGGGGSTSIRFSQTTSSSQHS  
STK  
10 >NP\_476429.2 keratin, type II cytoskeletal 3 [Homo sapiens]  
11 MSRQASKTSGGGSQGFSGRSAVVSGSSRMSCVAHSGGAGGGAYGFRSGAGGFGSRSLYNLGGNKSSISISVAAGGS  
RAGGF  
12 GGGRSSCAFAGGYGGGFGSGYGGGFGGGFGGGRGMGGGFGGAGGFGGAGGFGGAGGFGGPGGFGGSGGFGGPGSL  
GSPGG  
13 FGPGGFPGGIQEVTINQSLQLPLNVEIDPQIGQVKAQEREQIKTLNNKFASFIDKVRFLEQQNKVLETKWNL LQQ  
QGTSS  
14 ISGTNNLEPLFENHINYLRSYLDN ILGERGR LDSELKNMEDLVEDFKKKYEDEINKRTAAENEFVTLKKDVDSAY  
MNKVE  
15 LQAKVDALIDEIDFLRTLYDAELS QMQSHISDTSVVLSMDNNRSLDLSIIAEVRAQYEDIAQRSKAEAEALYQT  
KL GEL  
16 QTTAGRHGDDL RNTKSEIIELNRM IQR LRAE IEGVKKQANLQTAIAEAEQH GEMALKDANAKLQELQAALQKAK  
DDLAR  
17 LLRDYQELMNVKLALDVEIATYRK LLEGE EYRMSGECPSAVSISVVSSTTSASAGGYGGGYGGMGGLGGGFS  
AGGGS  
18 GSGFGRGGGGIGGGFGGGSSGFSGGSGFGSISGARYGVSGGGFSSASNRGGSIKFSQSSQSSQRYSR  
19 >NP\_004684.2 keratin, type II cytoskeletal 75 [Homo sapiens]  
20 MSRQSSITFQSGSRRGFSTTSAITPAAGRSRFSVSVARSAAGSGGLGRISSAGASFGSRSLYNLGGAKRV SING  
CGSSC  
21 RSGFGRASNRFGVN SGFGYGGGVGGGFSGPSFPVCPPGGIQEVTVNQSL LTPHLQIDPTIQRVRAEEREQIKT  
LNNKF  
22 ASFIDKVRFLEQQNKVLETKWALLQE QGSRTVRQNLEPLFDSYTS ELRRQLESITTERGRLEAELRNMQDVVEDF  
KVR YE  
23 DEINKRTAAENEFVALKKDVDAAYMNKVELEAKVKSLPEEINFIHVSFDAELS QLQTQVGDTSVVLSMDNNRNL D  
LDSII  
24 AEVKAQYEDIANRSRAEAE SWYQTKYEELQVTAGRHGDDL RNTKQEISEMNRM IQR LRAEIDS VKKQCSSLQTAI  
ADAEQ  
25 RGELALKDARAKLVDLEEALQKAKQDMARLLREYQELMNIKLALDVEIATYRK LLEGE ECR LSGEGVSPVNISVV  
TSTLS  
26 SGYSGSSIGGGLGLGGSGYSFTTSGGHS LGAGLGGSGFSATSNRGLGGSGSSVKFVSTTSSSQKSYTH  
27 >NP\_005546.2 keratin, type II cytoskeletal 6B [Homo sapiens]  
28 MASTSTTIRSHSSSRGFSANSARLP GVSRS GFSSISVSRSGGLGGACGGAGFGSRSLYGLGGSKRISIGGG  
SCAIS  
29 GGYGSRAGGSYFGGAGSGFGFGGGAGIGFGLGGGAGLAGGFGGPGFPVCPPGGIQEVTVNQSL LTPNLQIDPA  
IQRVR  
30 AEEREQIKTLNNKFASFIDKVRFLEQQNKVLDTKWTL LQE QGTKTVRQNLEPLFEQYINNLRRLDNIVGERGR L  
DSELR  
31 NMQDLVEDLKNKYEDEINKRTAAENEFVTLKKDVDAAYMNKVELQAKADTLTDEINFLRALYDAELS QMQTHISD  
TSVVL  
32 SMDNNRNL DLSIIAEVKAQYEEIAQRSRAEAE SWYQTKYEELQITAGRHGDDL RNTKQEIAEINRM IQR L RSEI  
DHVKK  
33 QCANLQAAIADAEQRGEMALKDAKNKLEGLE DALQKAKQDLARLLKEYQELMNVKLALDVEIATYRK LLEGE ECR  
LN GEG

34 VGQVNISVVQSTVSSGYGGASGVGSLGLGGGSSSYSGSLGVGGGFSSSSGRATGGGLSSVGGGSSTIKYTTTS  
SSSRK  
35 SYKH  
36 >NP\_775109.2 keratin, type II cytoskeletal 6C [Homo sapiens]  
37 MASTSTTIRSHSSRRGFSANSARLPGVSRSGFSSISVSRSGSLGGACGGAGFGSRSLYGLGGSKRISIGGG  
SCAIS  
38 GGYGSRAGGSYGFGGAGSGFGFGGAGIGFGLGGGAGLAGGFGGPGFPVCPGGIQEVTVNQSLLTPLNLQIDPA  
IQRVR  
39 AEEREQIKTLNNKFASFIDKVRFLEQQNKVLDTKWTLTLLQEQGKTVRQNLEPLFEQYINNLRRLQDSIVGERGRL  
DSELR  
40 NMQDLVEDLKNKYEDEINKRTAAENEFVTLKKDVDAAYMKNKVELQAKADTLTDEINFLRALYDAELSQMQTHISD  
TSVVL  
41 SMDNNRNLDLDSIIAEVKAQYEEIAQRSRAEAESWYQTKYEELQVTAGRHGDDLNTKQIEAEINRMIQRLRSEI  
DHVKK  
42 QCASLQAAIADAEQRGEMALKDAKNKLEGLEDALQKAKQDLARLLKEYQELMNVKLALDVEIATYRKLLGEEECR  
LNNEG  
43 VGQVNVSVVQSTISSGYGGASGVGSLGLGGGSSSYSGSLGIGGGFSSSSGRAIGGGLSSVGGGSSTIKYTTTS  
SSSRK  
44 SYKH  
45 >NP\_005545.1 keratin, type II cytoskeletal 6A [Homo sapiens]  
46 MASTSTTIRSHSSRRGFSANSARLPGVSRSGFSSSVSRSGSLGGACGGAGFGSRSLYGLGGSKRISIGGG  
SCAIS  
47 GGYGSRAGGSYGFGGAGSGFGFGGAGIGFGLGGGAGLAGGFGGPGFPVCPGGIQEVTVNQSLLTPLNLQIDPT  
IQRVR  
48 AEEREQIKTLNNKFASFIDKVRFLEQQNKVLETKWTLTLLQEQGKTVRQNLEPLFEQYINNLRRLQDSIVGERGRL  
DSELR  
49 GMQDLVEDFKNKYEDEINKRTAAENEFVTLKKDVDAAYMKNKVELQAKADTLTDEINFLRALYDAELSQMQTHISD  
TSVVL  
50 SMDNNRNLDLDSIIAEVKAQYEEIAQRSRAEAESWYQTKYEELQVTAGRHGDDLNTKQIEAEINRMIQRLRSEI  
DHVKK  
51 QCANLQAAIADAEQRGEMALKDAKNKLEGLEDALQKAKQDLARLLKEYQELMNVKLALDVEIATYRKLLGEEECR  
LNNEG  
52 VGQVNISVVQSTVSSGYGGASGVGSLGLGGGSSSYSGSLGVGGGFSSSSGRAIGGGLSSVGGGSSTIKYTTTS  
SSSRK  
53 SYKH  
54 >NP\_000415.2 keratin, type II cytoskeletal 5 [Homo sapiens]  
55 MSRQSSVSFRSGGSRSFSTASAITPSVSRSTFTSVSRSGGGGGGFGRVSLAGACGVGGYGSRLYNLGGSKRIS  
ISTSG  
56 GSFRNRFAGAGGGYGFGGGAGSGFGFGGAGGGFGLGGGAGFGGGFGGPGFPVCPGGIQEVTVNQSLLTPLNL  
QIDPS  
57 IQRVRTEEREQIKTLNNKFASFIDKVRFLEQQNKVLDTKWTLTLLQEQGKTVRQNLEPLFEQYINNLRRLQDSIVG  
ERGR  
58 DSELRNMQDLVEDFKNKYEDEINKRTAAENEFVTLKKDVDAAYMKNKVELEAKVDALMDEINFMKMFFDAELSQM  
QTHVSD  
59 TSVVLSMDNNRNLDLDSIIAEVKAQYEEIANRSRTEAESWYQTKYEELQVTAGRHGDDLNTKHEISEMNRMIQR  
LRAEI  
60 DNVKKQCANLQNAIADAEQRGELALKDARNKLAEEELQKAKQDMARLLREYQELMNTKLALDVEIATYRKLL  
GEECR  
61 LSGEGVGPVNISVVTSSVSSGYSGSGYGGGLGGGLGGGLGGLAGGSSGSYSSSSGGVGLGGGLSVGGSGFSA  
SSGRG  
62 LGVGFSGGGSSSVKFVSTTSSSRKSFKS  
63 >NP\_787028.1 keratin, type II cytoskeletal 79 [Homo sapiens]  
64 MRSSVSRQTYSTKGGFSSNSASGGSGSQARTSFSSVTVSRSSGSGGAHCGPGTGGFGSRSLYNLGGHKSISVSV  
AGGAL



65 LGRALGGFGFGSRAFMGQAGRQTFGPACPPGGIQEVTVNQSLLTPLHVEIDPEIQRVRTQEREQIKTLNNKFAS  
FIDKV  
66 RFLEQQNKVLETKWALLQEQGQNLGVTRNNLEPLFEAYLGSMRSTLDRLQSERGRLDSELNRNVQDLVEDFKNKYE  
DEINK  
67 HTAAENEFVVLKKDVAAYMGRMDLHGKVGTLTQEIDFLQQLYEMELSQVQTHVSNTNVVLSMDNNRNLDLDSII  
AEVKA  
68 QYELIAQRSRAEAEAWYQTKYEELQVTAGKHGDNLRDTKNEIAELTRTIQRLQGEADAAKKQCQQLQTAIAEAEQ  
RGELA  
69 LKDAQKKLGDLDVALHQAkedLTRLRDYQELMNVKLALDVEIATYRKLESEESRMSGECPSAVSISVTGNSTT  
VCGGG  
70 AASFGGGISLGGSGGATKGGFSTNVGYSTVKGGPVSAGTSILRKTTT VKTSSQRY  
71 >NP\_778238.1 keratin, type II cytoskeletal 73 [Homo sapiens]  
72 MSRQFTYKSGAAAKGGFSGCSAVLSGGSSSSYRAGGKGLSGGFSSRSLYSLGGARSISFNVASGSGWAGGYGFR  
GRASG  
73 FAGSMFGSVALGSVCPSLCPGGIHQVTINKSLLAPLNVELDPEIQKVRAQEREQIKVLNNKFASFIDKVRFLEQ  
QNQVL  
74 ETKWELLQQLDLNNCKNNLEPILEGYISNLRKQLETLSGDRVRDSELRSVREVVEDYKKRYEEEEINKRTTAENE  
FVVLK  
75 KDVDAAYTSKVELQAKVDALDGEIKFFKCLYEGETAQIQSHISDTSIILSMDNNRNLDLDSIIAEVRAQYEEIAR  
KSKAE  
76 AEALYQTKFQELQLAAGRHGDDLKHTKNEISELTRLIQRLRSEIESVKKQCANLETAIADAEQRGDCALKDARAK  
LDELE  
77 GALQQAKEELARMLREYQELLSVKLSLDIEIATYRKLEGEECRMSGEYTNVSISVINSSMAGMAGTGAGFGFS  
NAGTY  
78 GYWPSVSGGYSMPLPGGCVTSGGNCSPRGEARTRLGSASEFRDSQGKTALSSPTKKTMR  
79 >NP\_002263.3 keratin, type II cytoskeletal 4 [Homo sapiens]  
80 MIARQQCVRGGPRGFSCGSAIVGGGKRGAFSSVSMGGAGRCSSGGFGSRSLYNLRGNKSISMSVAGSRQGACFG  
GAGGF  
81 GTGGFGGGFGGSFSGKGGPGFPVCPAGGIQEV TINQSLLTPLHVEIDPEIQKV RTEEREQIKLLNNKFASFIDKV  
QFLEQ  
82 QNKVLETKWNLLQQTTTTSSKNLEPLFETYLSVL RKQLDTLGNDKGRLQSELKTMQDSVEDFKTKYEEEEINKRT  
AAEND  
83 FVVLKKDVAAYLNKVELEAKVDSLND EINF LKVL YDAELS QMQTHVSDTSVVL SMDNNRNLDLDSIIAEVRAQY  
EEIAQ  
84 RSKAEAEALYQTKVQQLQISVDQHGDNLKNTKSEIAELNRM IQLRAE IENIKKQCQTLQVSVADAEQRGENALK  
DAHSK  
85 RVELEAALQQAKEELARMLREYQELMSVKLALDIEIATYRKLEGE EYRMSGECQSAVSISVVGSTSTGGISGG  
LGSGS  
86 GFGLSSGFGSGSGSGFGGSGVSGSSSSKIISTTTLNKRR  
87 >NP\_258259.1 keratin, type II cytoskeletal 71 [Homo sapiens]  
88 MSRQFTCKSGAAAKGGFSGCSAVLSGGSSSSFRAGSKGLSGGFGSRSLYSLGGVRSLNVASGSGKSGGYGFRGR  
ASGFA  
89 GSMFGSVALGPVCPTVCPGGIHQVT VNESLLAPLNVELDPEIQKVRAQEREQIKALNNKFASFIDKVRFLEQQN  
QVLET  
90 KWELLQQLDLNNCKNNLEPILEGYISNLRKQLETLSGDRVRDSELNRNVRDVVEDYKKRYEEEEINKRTAAENEFV  
LLKKD  
91 VDAAYANKVELQAKVESMDQEIKFFRCLFEAEITQIQSHISDMSVILSMDNNRNLDLDSIIDEVRTQYEEIALKS  
KAEAE  
92 ALYQTKFQELQLAAGRHGDDLKNTKNEISELTRLIQIRSEIENVKKQASNLETAIADAEQRGDNALKDARAKLD  
ELEGA  
93 LHQAKEELARMLREYQELMSLKLALDMEIATYRKLESEECRMSG EFPSPVSISISSTSGGSVYGFPRPSMVSGG  
YVANS  
94 SNCISGVCSVRGGEGRSRGSANDYKDTLGKGSSLSAPSKKTSR  
95 >NP\_006112.3 keratin, type II cytoskeletal 1 [Homo sapiens]

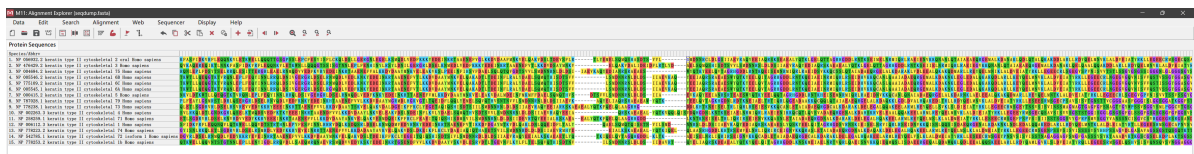
96 MSRQFSSRSgyrsgggfssgsagiinyqrRTTSSSTRRSgggggrfSSCGGGGSFGAGGGFGSRSLVNLGGSKS  
ISISV  
97 ARGGGRGSFGGGYGGGGFGGGGFGGGGIGGGFGGFGSGGGFGGGGFGGGGYGGGYGPVCPGGIQEV  
TINQS  
98 LLQPLNVEIDPEIQKVKSREREQIKSLNNQFASFIDKVRFLEQQNQVLQTKWELLQQVDTSTRTHNLEPYFESFI  
NNLRR  
99 RVDQLKSDQSRLDSELKNMQDMVEDYRNKYEDEINKRTNAENEFVTIKKDVGAYMTKVDLQAKLDNLQEQIDFL  
TALYQ  
100 AELSQMQTQISETNVILSMDNNRSLDLDIIAEVKAQYEDIAQKSKAEAESLYQSKYEELQITAGRHGDSVRNSK  
IEISE  
101 LNRVIQRLRSEIDNVKKQISNLQQSISDAEQRGENALKDAKNKLNLDLEDALQQAKEDLARLLRDYQELMNTKLAL  
DLEIA  
102 TYRTLLEGEESRMSGECAPNVSVSVSTSHTTISGGGSRGGGGGGYSGGSSYSGGGSYSGGGGGGGGRGSYSGS  
GSSYG  
103 SGGGSYSGGGGGGHGSYSGSSSGGYRGGSGGGGGSSGGRGSGGGSSGGSIGGRGSSSGVKSSGGSSSVKFV  
STTYS  
104 GVTR  
105 >NP\_778223.2 keratin, type II cytoskeletal 74 [Homo sapiens]  
106 MSRQLNIKSSGDKGNFSVHSAVVPRKAVGSLASYCAAGRAGAGFGSRSLYSLGNNRRISFNVAGGGVRAGGYGF  
RPGSG  
107 YGGGRASGFAGSMFGSVALGPACLSVCPGGIHQVTVNKSL LAPLNVELDPEIQKVRQEREQIKVLNDKFASFI  
DKVRF  
108 LEQQNQVLETKWELLQQDLNNCKKNLEPILEGYISNLRKQLETLSGDRVRLDSEL RSMRDLVEDYKKRYEVEIN  
RRTTA  
109 ENEFVVLKKDADAAYAVKVELQAKVDSL DKEIKFLKCLYDAEIAQIQTHASETSVILSMDNNRDL DLDIIAEVR  
MHYEE  
110 IALKSKAEAEALYQTKIQELQLAASRHGDDLKHTRSEMVELNRLIQRIRCEIGNVKKQRASLETAIADAEQRGDN  
ALKDA  
111 QAKLDELEGALHQAKEELARMLREYQELMSLKLALDMEIATYRKLLGEEECRMSENPSSVSISVISSSSYHH  
PSSAG  
112 VDLGASAVAGSSGSTQSGQTKTTEARGGDLKDTQGKSTPASIPARKATR  
113 >NP\_542785.1 keratin, type II cytoskeletal 72 isoform 1 [Homo sapiens]  
114 MSRQLTHFPRGERLGFSGCSAVLSGGIGSSSASFARVRKGSASFGSKSL SCLGGSRLALSAAARRGGGRLGGFV  
GTAFG  
115 SAGLGPKCPSVCPGGIPQVTVNKSL LAPLNVEMDPEIQRVRAQEREQIKALNNKFASFIDKVRFLEQQNQVLET  
KWNLL  
116 QQDLNNCRKNLEPIYEGYISNLQKQLEMLSGDGVRLDSELNMQDLVEDYKKRYEVEINRRTAAENEFVVLKKD  
VDAAY  
117 MNKVELQAKVDSLTD EIKFFKCLYEGEITQIQSHISDTSIVLSMDNNRDL DLDIIAEVRAQYEEIALKSKAEAE  
TLYQT  
118 KIQELQVTAGQHGGDLKLTAEISELNRLIQRIRSEIGNVKKQCADLETAIADAEQRGDCALKDARAKLDELEGA  
LHQAQ  
119 EELARMLREYQELVSLKLALDMEIATYRKLL EEECRMSEYPNVSISVISSTNAGAGGAGFSMGFGASSSY  
KTAAA  
120 DVKTKGSCGSELKDPLAKTSGSSCATKKASR  
121 >NP\_778253.2 keratin, type II cytoskeletal 1b [Homo sapiens]  
122 MSHQFSSQSASFSSMRVYSTSSSAGSGGGSPAVGSVCYARGRCGGGGYGIHGRGFGSRSLYNLGGSRISISINLM  
GRSTS  
123 GFCQGGGVGGFGGGRGFGVGSTAGGFGGGGFGGAGFGT SNFGLGGFGPYCPGGIQEVTINQSLLEPLHLEVDP  
EIQRI  
124 KTQEREQIMVLNNKFASFIDKVRFLEQQNQVLQTKWELLQQVNTSTGTNNLEPLLENYIGDLRRQVDLLSAEQMR  
QNAEV  
125 RSMQDVVEDYKSKYEDEINKRTGSENFVVLKKDVDAAYVSKVDLESRVDTLTGEVNFLKYLFLTELSQVQTHIS  
DTNVI

```
126  LSMDNNRSLDLDSIIDAVRTQYELIAQRSKDEAEALYQTKYQELQITAGRHGDDLKNSKMEIAELNRTVQRLQAE
    ISNVK
127  KQIEQMQLISDAEERGEQALQDAWQKLQDLEEALQQSKEELARLLRDYQAMLGVKLSLDVEIATYRQLLEGEES
    RMSGE
128  LQSHVSVISVQNSQVSVNGGAGGGGSYGGGYGGGSGGGYGGGRSYRGGGARGRSGGGYGSGCGGGGGSYGGSGRS
    GRGSS
129  RVQIIQTSTNTSHRRILE
130
```

- 在 MEGA 中处理数据



- 使用 ClustaW 进行序列比对



- 分析数据并构建距离矩阵

M11: Pairwise Distances (PhyloAnalysis.meg)															
File	Display	Average	Caption	Help											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. NP 056932.2 keratin type II cytoskeletal 2 oral Homo sapiens															
2. NP 476429.2 keratin type II cytoskeletal 3 Homo sapiens	1.762														
3. NP 004684.2 keratin type II cytoskeletal 75 Homo sapiens	2.2101	2.4451													
4. NP 005546.2 keratin type II cytoskeletal 68 Homo sapiens	2.0920	2.2711	2.1700												
5. NP 775109.2 keratin type II cytoskeletal 6C Homo sapiens	2.0920	2.2540	2.1942	0.0125											
6. NP 005545.1 keratin type II cytoskeletal 6A Homo sapiens	2.1066	2.2711	2.1700	0.0143	0.0161										
7. NP 000415.2 keratin type II cytoskeletal 5 Homo sapiens	1.9476	1.9911	1.9898	2.0495	2.0495	2.0495									
8. NP 787028.1 keratin type II cytoskeletal 79 Homo sapiens	2.4962	2.4246	2.0244	2.1199	2.1199	2.0851	2.4072								
9. NP 778238.1 keratin type II cytoskeletal 73 Homo sapiens	2.2200	2.3197	2.1607	2.4207	2.4662	2.4662	2.4756	2.5314							
10. NP 002263.3 keratin type II cytoskeletal 4 Homo sapiens	2.2680	2.6826	2.0873	2.2031	2.2031	2.2031	2.3862	1.9776	2.6430						
11. NP 258259.1 keratin type II cytoskeletal 71 Homo sapiens	2.1915	2.5287	2.1652	2.8430	2.8102	2.8102	2.9326	2.3788	1.9644	2.2887					
12. NP 006112.3 keratin type II cytoskeletal 1 Homo sapiens	1.6173	1.7572	2.0736	1.9495	1.9495	1.9495	1.7681	2.4343	2.2674	2.2829	2.3341				
13. NP 778223.2 keratin type II cytoskeletal 74 Homo sapiens	2.3495	2.1102	2.1934	2.4405	2.4624	2.4405	2.5707	2.3514	2.3788	2.1209	2.3770	2.3860			
14. NP 542785.1 keratin type II cytoskeletal 72 isoform 1 Homo sapiens	2.3026	2.2967	2.3145	2.5336	2.5649	2.5336	2.6489	2.6205	2.0637	2.3106	2.2152	2.4483	2.1681		
15. NP 778253.2 keratin type II cytoskeletal 1b Homo sapiens	2.0351	2.0386	2.2214	1.7403	1.7403	1.7403	1.8528	2.3840	2.3429	2.3730	2.5084	2.0011	2.3960	2.7041	

- 构建进化树 (NJ 法)

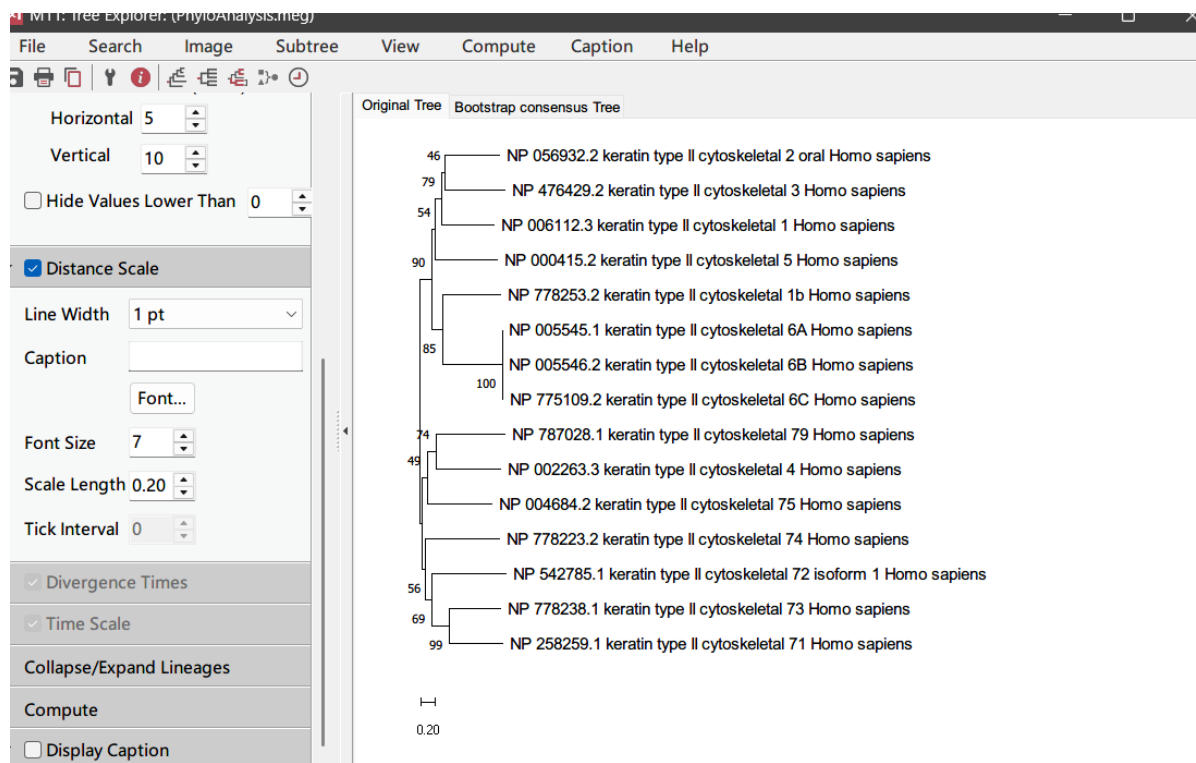
- 使用自展法

## M11: Analysis Preferences

## Phylogeny Reconstruction

Option	Setting
<b>ANALYSIS</b>	
Scope	→ <i>All Selected Taxa</i>
Statistical Method	→ <i>Neighbor-joining</i>
<b>PHYLOGENY TEST</b>	
Test of Phylogeny	→ Bootstrap method
No. of Bootstrap Replications	→ 1000
<b>SUBSTITUTION MODEL</b>	
Substitutions Type	→ <i>Amino acid</i>
Model/Method	→ <i>Poisson model</i>
<b>RATES AND PATTERNS</b>	
Rates among Sites	→ <i>Uniform Rates</i>
Gamma Parameter	→ <i>Not Applicable</i>
Pattern among Lineages	→ <i>Same (Homogeneous)</i>
<b>DATA SUBSET TO USE</b>	
Gaps/Missing Data Treatment	→ <i>Pairwise deletion</i>
Site Coverage Cutoff (%)	→ <i>Not Applicable</i>
<b>SYSTEM RESOURCE USAGE</b>	
Number of Threads	→ 8

ⓘ Help ✕ Cancel ✓ OK



## 讨论

在本次实验中熟悉掌握了几种多序列比对的工具。