
“实用生物信息技术”课程小组讨论总结报告

组次：G5 组长：韩佳凝 执笔：赵宏芳

1. 时间

2026.4.20

2. 方式

线上腾讯会议讨论

3. 主题

课后对构建系统发育树的复习总结

4. 总结内容

(1) MEGA12 与 MEGA7

MEGA12 (Molecular Evolutionary Genetics Analysis) 是一款跨平台、免费的分子进化与系统发育分析软件，支持 Windows、macOS、Linux 系统。它集成序列比对、进化树构建、模型选择、分子钟分析等功能，可处理 DNA、蛋白序列。采用 ClustalW 等算法比对，提供邻接法、最大似然法等建树方法，支持自举检验评估可靠。

下载路径在 abc 网站上

常用系统发生分析软件

1. Molecular Evolutionary Genetics Analysis (MEGA) [Windows用户](#)
2. Data Analysis and Molecular Biology and Evolution (DAMBE)
3. Phylogeny Inference (Phylip)

[点击该链接即可下载](#)

MEGA7 是免费的分子进化与系统发育分析软件，2016 年发布，适配 64 位系统，支持 Windows、macOS、Linux。内置 ClustalW 等比对工具，可处理 DNA / 蛋白序列。提供邻接法、最大似然法等建树算法，支持自举检验，新增基因复制预测与分化时间分析功能。

MEGA7 (2016) 与 MEGA12 (2024) 均为分子进化分析软件，核心功能一致，支持序列比对、进化树构建与自举检验；MEGA12 在性能、功能、兼容性上全面升级，计算速度更快、适配更大数据集，新增分化时间校准、分支稳定性分析等高级功能，界面更友好，更适合现代高通量进化研究，而 MEGA7 功能基础、运算较慢，适合常规小样本分析。

本节课和本次讨论主要聚焦于 MEGA7 与 MEGA12 的系统发生树构建。

(2) 几种主要的方法及其优缺点

a. 邻接法 (Neighbor-joining)

首先把整体上（例如在 DNA 序列上）最相似的物种相连接，然后将这样的聚簇与下一个与它们最相似的物种相连接以形成更大的簇，重复这一过程直到所有物种都连接到了一个簇。其适用范围为远缘序列，进化距离不大，信息位点少的短序列。优点：假设少，树的构建相对准确，计算速度快，只得到一棵树，可以分析较多序列，运行速度优于最大简约法。缺点：序

列上所有位点等同对待，且所分析的序列进化距离不能太大。

b. 最大简约法 (Maximum parsimony)

根据序列的多重比对结果,对所有可能正确的拓扑结构进行计算并挑选出所需替代数最少的拓扑结构作为最优树,即能够利用最少的步骤去解释多重比对中的碱基差异。适用范围:近缘序列,物种序列数目 ≤ 12 。优点:善于分析某些特殊的分子数据如插入、缺失等序列。缺点:只适用于数据中的序列数目 ≤ 12 ,存在较多回复突变或平行突变时,结果较差,对变异大的序列进行分析时易产生错误。

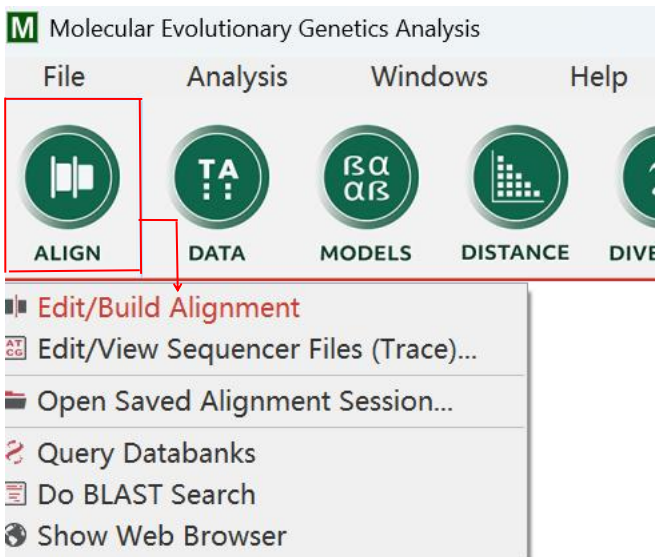
c. 最大似然法 (Maximum Likelihood)

基于数据本身推断或估计 DNA 序列的进化模型,然后考察所有可能的树中的哪一棵树能使所观察到的特征状态进化的可能性(即似然性)达到最大。

总结:邻接法 (NJ) 是 MEGA 中常用的距离法建树算法,先计算序列间遗传距离,再迭代连接最近的序列,逐步构建无根进化树,运算快、适合大数据,MEGA 中可搭配自举检验评估树的可靠性。最大似然法 (ML) 是 MEGA 中常用的概率型建树算法,基于序列进化模型,计算所有可能进化树的最大似然值,择优构建最优树;精度高、可靠性强,可搭配自举检验评估分支支持度,适合解析序列真实进化关系。

(3) 操作实践

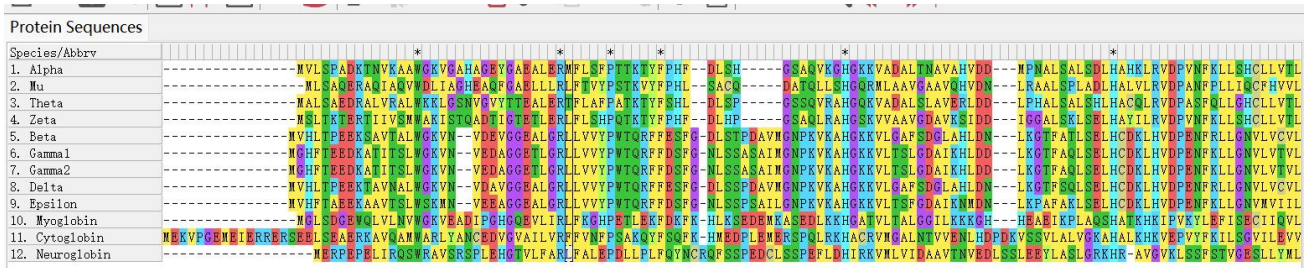
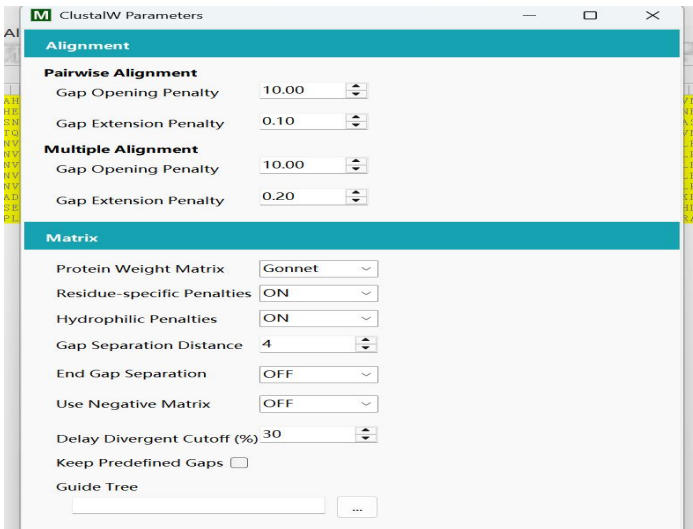
a. 以人类 12 个珠蛋白蛋白序列,采用 ClustalW 比对、邻接法 + 100 次自举构建进化树为例;



粘入人的珠蛋白家族 12 个蛋白质序列,删除最后一行;



Alignment → Align by ClustalW → OK (默认参数)



Data → Phylogenetic Analysis → Phylogeny → Construct/Test Neighbor-Joining Tree → OK

注：本张图片中，不同颜色通常对应不同化学性质的氨基酸，如绿色代表极性/不带电荷的氨基酸，红色代表带正电荷的碱性氨基酸等。序列上方的星号标记的是完全保守位点，也就是所有序列在这个位置的氨基酸都一模一样，说明这个位点在进化中被严格保留，意味着该位点对蛋白的结构或者功能至关重要。序列中间的短横线是空位，代表序列在这个位置发生了插入或缺失突变。下文相似图片同样。

Molecular Evolutionary Genetics Analysis

File Analysis Windows Help

ALIGN DATA MODELS DISTANCE DIVERSITY PHYLOGENY USER TREE ANCESTORS SELECTION RAT

Close Data

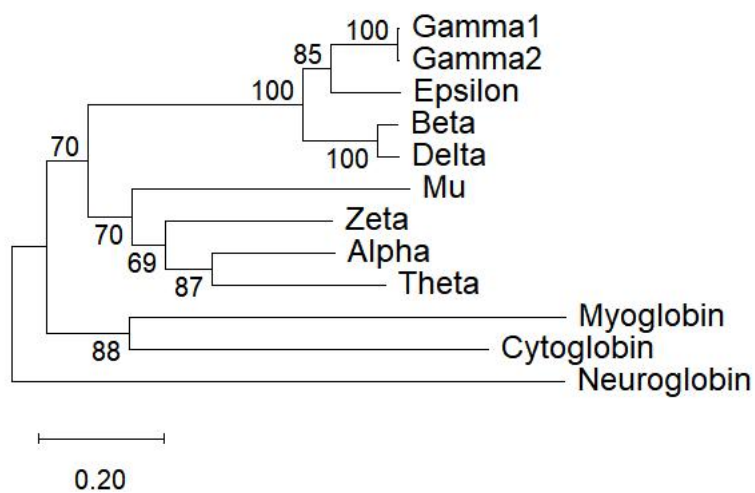
- Construct/Test Maximum Likelihood Tree...
- Construct/Test Neighbor-Joining Tree...
- Construct/Test Minimum-Evolution Tree...
- Construct/Test UPGMA Tree...
- Construct/Test Maximum Parsimony Tree(s)
- Open Tree Session

M12: Analysis Preferences

NJ Phylogeny Reconstruction

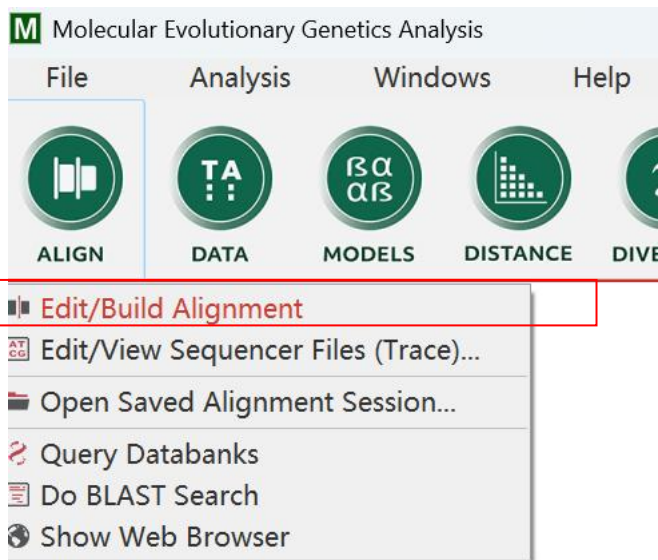
Option	Setting
ANALYSIS	
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny	Bootstrap method
Bootstrap Replicates	100
SUBSTITUTION MODEL	
Substitutions Type	Amino acid
Model/Method	Poisson model
RATES AND PATTERNS	
Rates among Sites	Uniform Rates
Pattern among Lineages	Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data	Pairwise deletion
SYSTEM RESOURCE USAGE	
Number of Threads	8

Help Cancel Ok



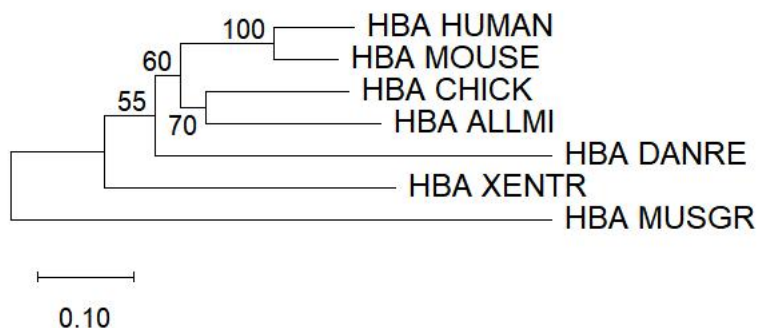
结果显示该家族分为基部类群（神经珠蛋白、胞珠蛋白、肌红蛋白）与两大功能分支： α -珠蛋白簇（Zeta、Alpha、Theta）和 β -珠蛋白簇（Gamma/2、Epsilon、Beta、Delta），Mu 珠蛋白为独立旁支；多数分支自举支持度较高，清晰反映了珠蛋白家族通过多次基因复制事件分化、形成不同时空表达亚型的进化历程。

b. 以 7 种脊椎动物 α 血红蛋白蛋白序列为例，构建进化树并验证物种亲缘关系；



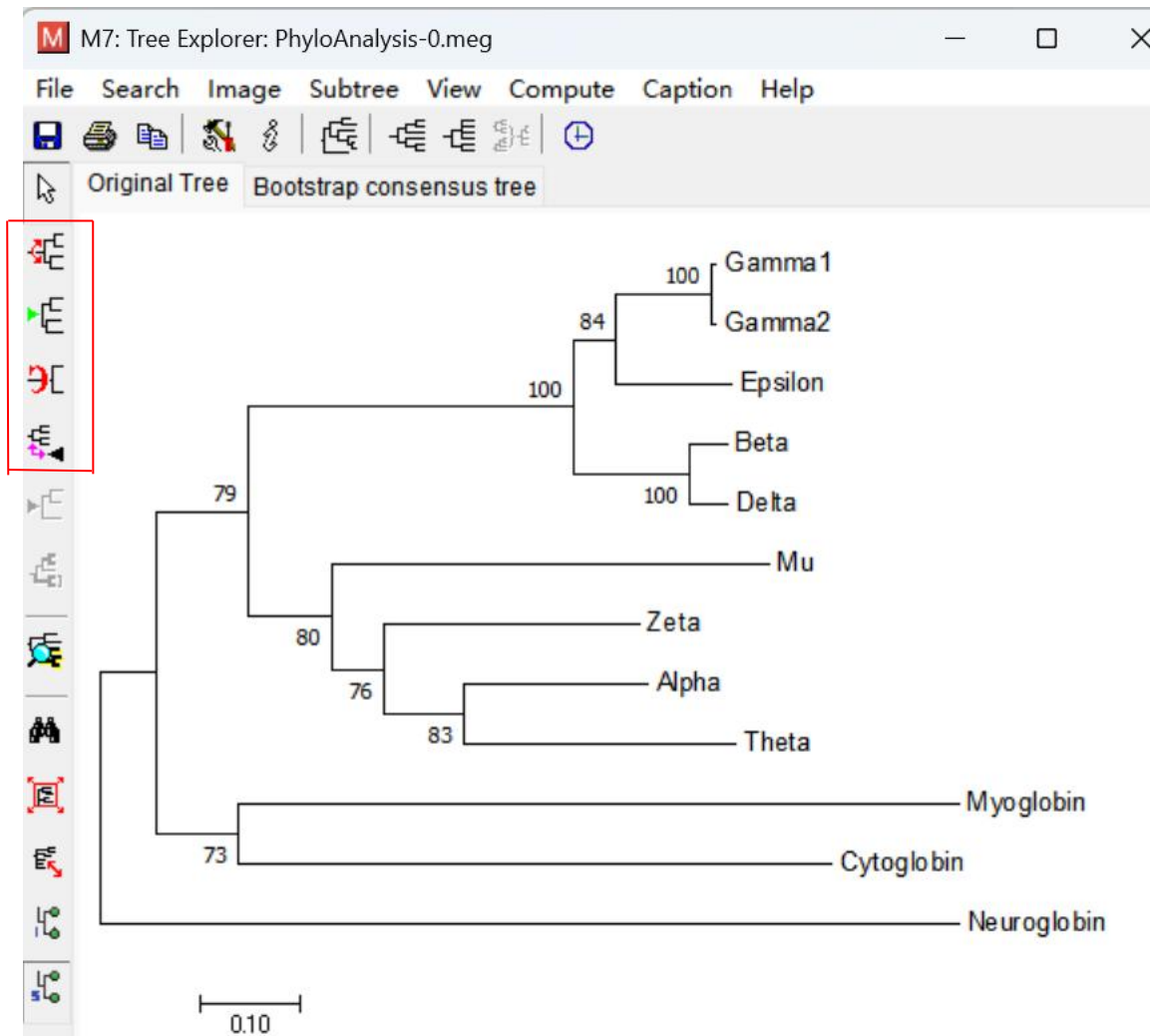
Alignment → Align by ClustalW → OK（默认参数）Data → Phylogenetic Analysis → Phylogeny → Construct/Test Neighbor-Joining Tree → OK

Species/Abbrv	Sequence
1. HBA HUMAN	MVLSPADKTNVRAAGKVGAGHAGCYGABALERMFLSPFTKTYPPHFDLSHGSAQVKGCHGKKVADALNVAHVDDMPNALGALSOLHAKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDRFLASVSTIV
2. HBA MOUSE	MVLSGDEKSNIKAAWGTGGHAGCYGABALERMFLSPFTKTYPPHFDVSHGSAQVKGHGGKVAALASAGHLDDLPGALGALSOLHAKLRVDPVNFKLLSHCLLVTLAHHHPADFTPAVHASLDRFLASVSTIV
3. HBA CHICK	MVLSAADKNNVKGIFETKAGHAEYGAELERMFTYPPKTYPPHFDLSHGSAQVKGHGGKVAALASAGHLDDLPGALGALSOLHAKLRVDPVNFKLLGCFLLVVAIHHHPAALTPAVHASLDRFLASVSTIV
4. HBA ALLMI	MVLSMEDKSNVKAIVKAGHLEYGAEELERMFCAYPQTKTYPPHFDLSHGSAQVKGHGGKVAALASAGHLDDLPGALGALSOLHAKLRVDPVNFKLLGCFLLVVAIHHHPAALTPAVHASLDRFLASVSTIV
5. HBA XENTR	MVLSLADDKHKIKAIWPSVAAGCYGABALERMFLSPFTKTYPPHFDLSHGSAQVKGHGGKVAALASAGHLDDLPGALGALSOLHAKLRVDPVNFKLLGCFLLVVAIHHHPAALTPAVHASLDRFLASVSTIV
6. HBA DANRE	MVLSLTDKAVVKAIWAKTSPKADETGABALARMFLVYPTKTYPPHFDLSPGSGPVYKHGKKEINGAVGCAIKIDLDLVGCLAALSLHAFKLRVDPANFKILSHNVIVVIAMLFPADFTPEVHVSDKFFLNAL
7. HBA MUSGR	MAFLACDRGKIKIAQVLAQPEAYGAECLARLRFVTHPGCKYTPPEYKDYGAAGAKVQVHGKVIKAVVKAABHVDDLHSHLELALHGGKLLVDPGNFPMISECTIVTLAHLTFPDPDTHCAVDKLLSATCQEL



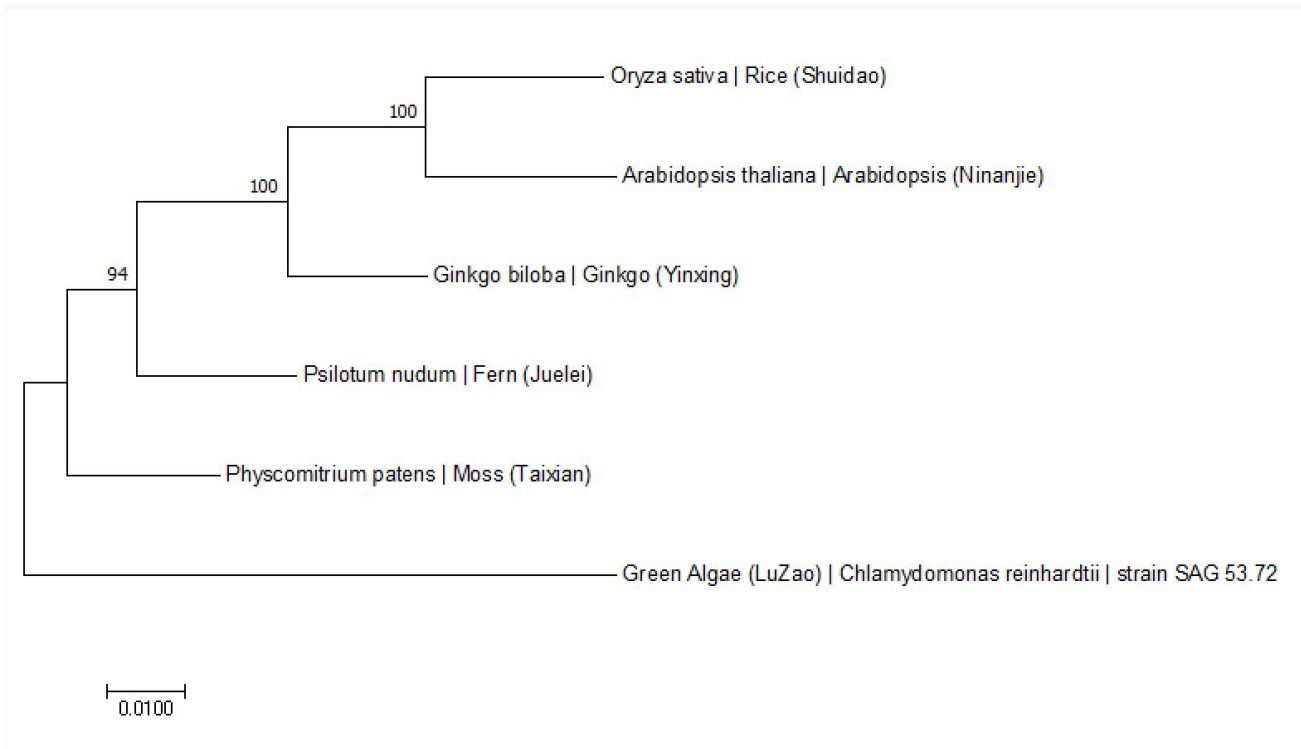
树中哺乳类（人、小鼠）聚为高支持度分支，与鸟类（鸡、鹌鹑）形成次一级分支，再依次与鱼类、两栖类、七鳃鳗序列聚类，整体拓扑符合脊椎动物由低等到高等的系统演化关系，清晰反映了物种分化过程中 α -血红蛋白基因的进化历程。

c. 有根树无根树

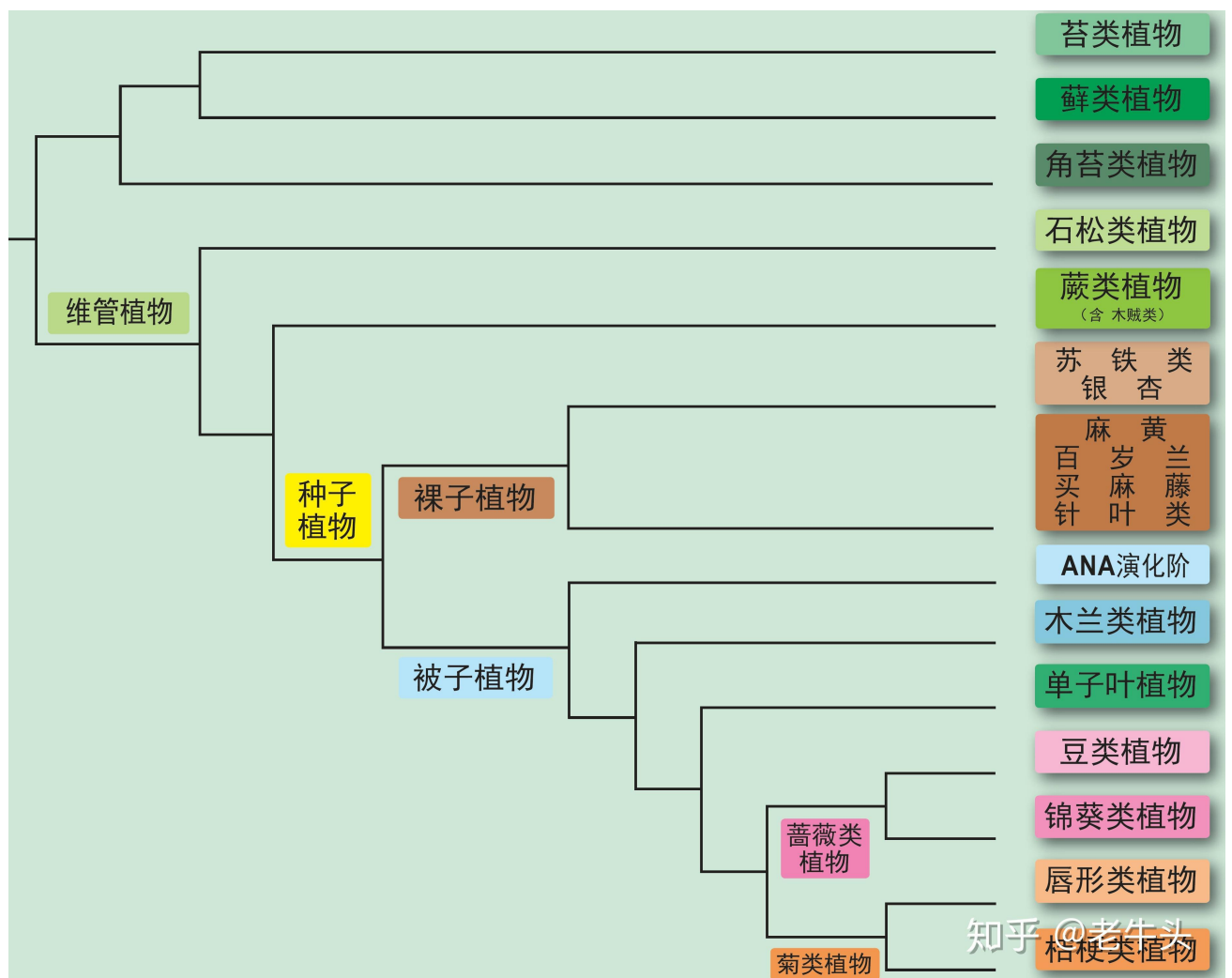


MEGA7 的 Tree Explorer 左侧工具栏集成了进化树的核心操作工具，其中箭头图标用于选择和编辑分支；根树 / 无根树工具可切换树的拓扑结构，根树通过指定外类群明确进化方向，无根树则仅展示序列间的亲缘关系；Subtree 相关工具支持分支折叠、旋转、重新排列与修剪，可调整树的显示结构；分支颜色 / 高亮工具可标记特定类群；缩放工具用于调整视图大小；导出 / 保存工具支持将进化树图片或数据文件输出。

d. 以 7 个代表性植物 18S rRNA 为例，构建系统发生树



这是绿色植物分类谱系分析所构建的系统发生树：



和我们自己构建的系统发生树相比：

绿藻 (*Chlamydomonas reinhardtii*)：绿藻门，但分类谱系中没有该分支。

苔藓 (*Physcomitrium patens*)：藓类植物，最先分化出来的陆生植物。

松叶蕨 (*Psilotum nudum*)：蕨类植物，属于维管植物，比苔藓更高级。

银杏 (*Ginkgo biloba*)：裸子植物，属于种子植物，比蕨类更高级。

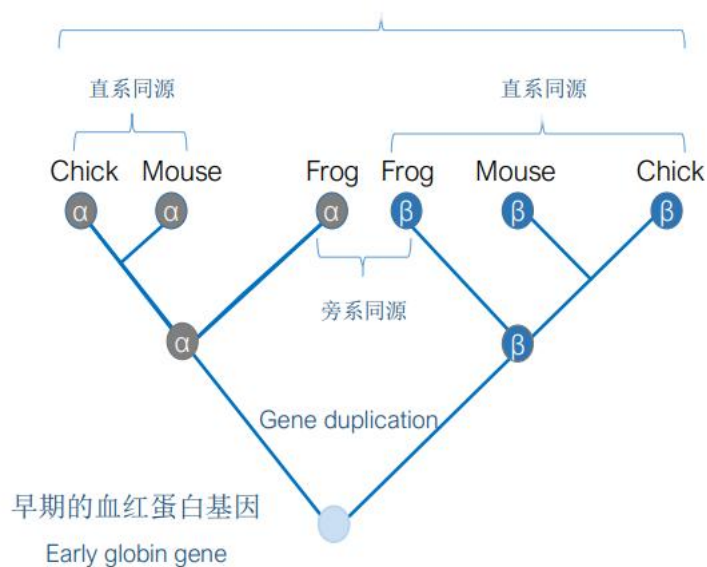
水稻 (*Oryza sativa*)：被子植物中的单子叶植物。

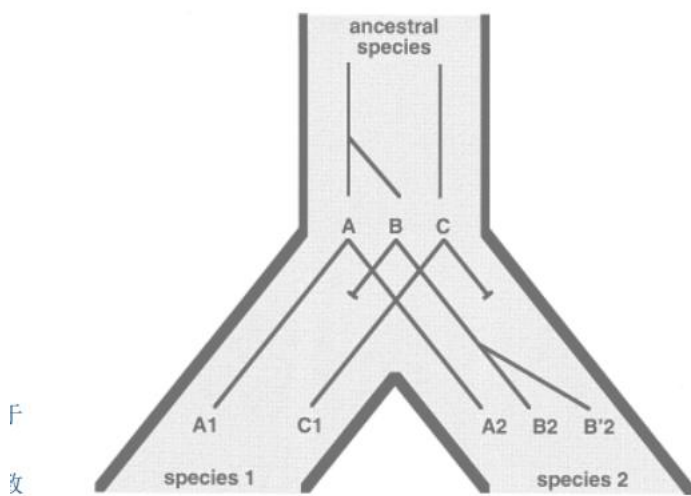
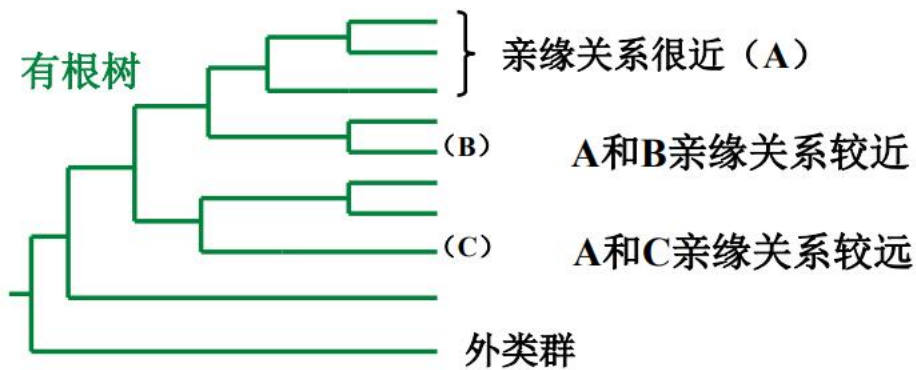
拟南芥 (*Arabidopsis thaliana*)：被子植物中的双子叶植物。

该 18S rRNA 系统发生树所呈现的演化关系（绿藻最早分化，依次为苔藓、蕨类、裸子植物、被子植物，且被子植物中单子叶与双子叶分开）与经典分类谱系基本一致，能够正确反映 6 种植物的系统发生关系。

(4) 阅读文献获得的相关知识

直系同源 (orthology) 直系同源的序列因物种形成 (speciation) 而被区分开 (separated)：若一个基因原先存在于某个物种，而该物种分化为了两个物种，那么新物种中的基因是直系同源的；旁系同源 (paralogy) 旁系同源的序列因基因复制 (gene duplication) 而被区分开 (separated)：若生物体中的某个基因被复制了，那么两个副本序列就是旁系同源的。





基因树由单个基因序列构建，反映该基因的演化历程；物种树则基于全基因组或多基因数据整合构建，体现物种真实的分化历史。进化分支树仅展示类群间的拓扑关系，不包含分支长度；系统发生树则同时呈现拓扑结构与分支长度，分支长度代表序列间的进化差异。无根树仅显示类群间的亲缘关系，无法体现演化方向；有根树通过引入外类群等方式确定根节点，明确从祖先到后代的演化方向。分支是连接节点的线段，代表类群的进化路径；节点是分支的交汇点，代表一次分化事件。内部节点位于树的中间位置，代表已灭绝的祖先类群；外部节点即叶节点，位于树的末端，代表现存的研究类群，其中根节点是有根树的起始节点，代表所有类群的共同祖先。

(5) 问题

不同系统版本（Windows/macOS/Linux）的 MEGA 存在兼容性差异，部分功能在 macOS/Linux 平台易出现报错；长时间运行大型分析时，软件偶发崩溃。