
“实用生物信息技术”课程小组讨论总结报告

组：G4 次：R4 组长：高嘉禾 执笔：高嘉禾

1. 时间：2026年4月29日 18:30
2. 方式：线上（腾讯会议，微信）
3. 主题：复习系统发育树构建相关内容及完成课后练习
4. 内容：
 - A. 复习巩固课堂内容
 - B. 完成*课后练习5
 - C. 小组成员交流学习心得与问题

A. 复习巩固课堂内容

复习分子演化和系统发育树基础概念：

(1) 物种分化和分子演化

- 物种分化：指一个祖先物种通过遗传隔离和表型变异，逐渐分裂为两个或多个独立物种的过程，是生物多样性的主要来源。
- 分子演化：研究生物大分子（如DNA、RNA、蛋白质）在进化过程中的序列、结构及功能变化规律。通过比较不同物种的同源分子序列，可以推断演化历史、选择压力及亲缘关系。

(2) 分支图和系统树

- 分支图：一种仅表示分类单元之间相对亲疏关系的树状图，分支长度无实际意义，只关注拓扑结构。常用于表述假定的演化关系。
- 系统树：即系统发生树，一种反映生物或基因演化历史的树状图，其分支长度通常代表演化时间或遗传距离，既能展示拓扑关系，也能体现演化速率或时间尺度。

(3) 物种树和基因树

- 物种树：描述物种之间分化与亲缘关系的系统发生树，每个分支代表一个物种谱系，其分支点对应物种形成事件。
- 基因树：基于特定同源基因序列构建的系统发生树，反映该基因在物种间的演化历程。由于不完全谱系分选、基因水平转移或基因重复等因素，基因树可能与物种树不一致。

(4) 有根树和无根树

- 有根树：包含一个特定根节点的系统发生树，根节点指明了演化方向，可通过外群或分子钟假设确定根的位置。
- 无根树：仅表示分类单元之间的相对亲缘关系，不指明演化方向，其拓扑结构可以通过加入外群或指定根节点转化为有根树。

(5) 二叉树与多歧树

- 二叉树：每个内部节点恰好有两个分支的系统发生树。在理想的分叉式物种形成假设下，二叉树是标准演化模型。
- 多歧树：存在至少一个内部节点具有三个或以上直接分支的系统发生树。通常用于表示演化关系不明确或同时发生辐射分化的情况。

(6) 外部节点和内部节点

- 外部节点：指叶节点，位于系统发生树末端的节点，代表当前研究中所涉及的实际分类单元。
- 内部节点：位于树分支内部的节点，代表假定的祖先，每个内部节点对应一次分化事件。

(7) 内部节点和根节点

- 内部节点：见上（6）中的定义。在系统树中，除叶节点外的所有节点均为内部节点。
- 根节点：有根树中的内部节点，代表所有分类单元的最后共同祖先，是演化时间的起点，确定了树的方向。

(8) 系统发生树稳定性检验

定义：通过统计学或重采样方法评估系统发生树拓扑结构可靠性的过程。常用方法包括：

- 自举法 (Bootstrap)：通过重复采样数据集构建多棵树，统计分支出现频率。
- 似然比检验 (Likelihood Ratio Test)：比较不同树模型的支持度。
- 贝叶斯后验概率 (Bayesian Posterior Probability)：根据贝叶斯方法估计树分支的可信度。

B. 完成课外练习 5

*课外练习 5

1. 参阅 ABC 网站有关资料，查阅相关文献，说明以下基本概念

(1) 分子演化和系统发生

- 分子演化：研究生物大分子（如 DNA、RNA、蛋白质）在进化过程中的序列、结构及功能变化规律。通过比较不同物种的同源分子序列，可以推断演化历史、选择压力及亲缘关系。

-
- 系统发生 (Phylogeny) : 指一个物种或生物类群的形成和演化历史, 关心的是不同物种之间由共同祖先分化而来的亲缘关系和演化时序。

(2) 序列相似性 (Similarity) 和序列同源性 (Homology)

- 序列相似性: 是一个量化概念, 通常用两个序列比对中相同或相似残基所占的比例来表示。较高的相似性往往是同源性存在的证据, 但也可能由趋同进化造成。
- 序列同源性: 是一个定性概念, 指两个序列拥有共同的祖先, 要么同源, 要么非同源, 不存在百分比同源性。同源序列通常具有较高的相似性, 但高相似性并不必然等价于同源性。

(3) 直系同源 (Ortholog) 和旁系同源 (Paralog)

- 直系同源: 指分布于不同物种的基因, 起源于物种形成事件, 是祖先基因在不同后代物种中垂直传递的拷贝。直系同源基因通常功能保守。
- 旁系同源: 起源于基因复制事件, 当一个祖先基因在基因组内发生复制产生多个拷贝时形成。旁系同源基因可能出现功能分化, 既可位于同一物种, 也可分布在不同物种中。

(4) 核苷酸替换模型和氨基酸替换模型

- 核苷酸替换模型: 描述 DNA 序列中碱基随时间发生替换的概率, 通过设定不同碱基替换类型的相对速率来准确估计遗传距离。
- 氨基酸替换模型: 描述蛋白质序列中氨基酸的变化概率, 通过分析大量比对数据估计不同氨基酸之间的替换偏好, 为系统发生分析提供替换矩阵。

(5) 突变速率和分子钟

- 突变速率: 指单位时间内 DNA 序列发生替换的频率, 因物种、基因组区域等因素而异。
- 分子钟: 是一个理论模型, 假设对于中性序列, 替换速率在不同谱系间相对恒定, 因此两个谱系间的序列差异量可用来推算它们的分化时间。

(6) 进化分支树 (Cladogram) 和系统发生树 (Phylogram): 同 A (2)

(7) 基因树和物种树: 同 A (3)

(8) 无根树和有根树: 同 A (4)

(9) 分支和节点

- 分支: 是树中连接两个节点的线段, 代表两个分类单元之间的演化路径。
- 节点: 代表一个分类单元或一次演化事件, 每个分支的两端各有一个节点。

(10) 内部节点和外部节点: 同 A (6)

(11) 根节点和叶节点

- 根节点：是有根树中最特殊的一个内部节点，代表整棵树所有分类单元的终极共同祖先。
- 叶节点：与外部节点同义，位于树的分支顶端，代表正在研究的生物体、基因或序列。

(12) 距离法和位点法

- 距离法：首先通过序列比对计算所有两两序列之间的遗传距离，形成距离矩阵，再基于该矩阵构建树。距离法计算速度快，但可能损失位点信息。
- 位点法：将多序列比对中每个位置的碱基或氨基酸特征作为离散特征，直接评估不同树形结构对这些特征的解释程度，从中选出最优树。

(13) 最大简约法和最大似然法

- 最大简约法：依据简约原则，在所有可能的树中选择能够解释序列差异所需总演化步骤最少的树作为最优树。
- 最大似然法：建立在特定的演化替换模型之上，计算每棵候选树产生实际观测数据的概率，选择概率最高的树，统计基础扎实但计算量较大。

2. 参阅 ABC 网站中有关资料，查阅相关文献，回答以下问题

(1) 构建系统发生树的基本步骤：

- 选择序列数据：收集目标基因的核苷酸或氨基酸序列。
- 序列比对：将所有序列按照进化对应关系排列。
- 模型选择：确定进化替代模型（如 JC、K2P、GTR、WAG、JTT 等）。
- 建树方法选择：选择适当算法（邻接法、最大似然法、最大简约法、贝叶斯法）。
- 构建系统发育树：运行建树程序生成系统发育树。
- 验证与优化：使用 Bootstrap、后根化、模型对比等评估树的可靠性。

(2) 构建系统发生树时选择核苷酸序列或氨基酸序列的原则：

- 氨基酸序列适合远缘进化分析：氨基酸更保守，可减少同义替代的噪声，适用于物种分化较远或深系谱关系研究。
- 核苷酸序列适合近缘物种或基因内部比较：含有三倍体信号，可捕捉更多微小变异，可用于检测选择压力和碱基替代模式。
- 一般原则：若序列差异大，使用氨基酸序列；若关系较近且需要分析同义/非同义替代，使用核苷酸序列；若对比多个基因/全基因组，可两者一起分析。

(3) 利用自举法 (Bootstrap) 检验系统发生树稳定性的原理:

Bootstrap 是一种统计重抽样方法, 用于评估树中每个分支节点的支持度: 从比对矩阵中对列位点进行有放回重抽样, 每次生成一个新比对数据集, 然后构建树, 统计每个节点在所有重抽样树中出现的频率。Bootstrap 值即节点出现的百分比。

原理: 如果某个分支在不同随机抽样数据中频繁出现, 则该分支稳定、可信。通常认为:

- $\geq 90\%$: 高度支撑
- 70 - 89%: 中等支撑
- $< 50\%$: 支撑弱, 不可靠

(4) 确定无根树根节点的方法:

- 外群定根 (Outgroup rooting): 选一个与研究组关系较远、明确先分化的物种/序列作为外群, 将树根定在外群与内群之间。最常用、最可靠。
- 中间点定根 (Midpoint rooting): 在无根树中找两端最长距离的中点作为根。
- 已知时间/化石定根: 利用分子钟或化石记录提前确定特定节点位置。

(5) 如何通过所构建的系统发生树判断“先有物种”还是“先有基因”:

系统发育树只能反映谱系分化顺序, 必须结合时间信息。才能判断物种与基因谁先出现。

在系统发育树中可以简单判断: 若一个基因在多个物种中都有, 那么这个基因可能在物种分化之前就存在, 说明“先有基因”; 若一个基因只在特定物种中出现, 且属于物种最近分化的支系, 则可能是后出现的, 说明“先有物种”。

(6) 不同建树方法的基本原理和特点:

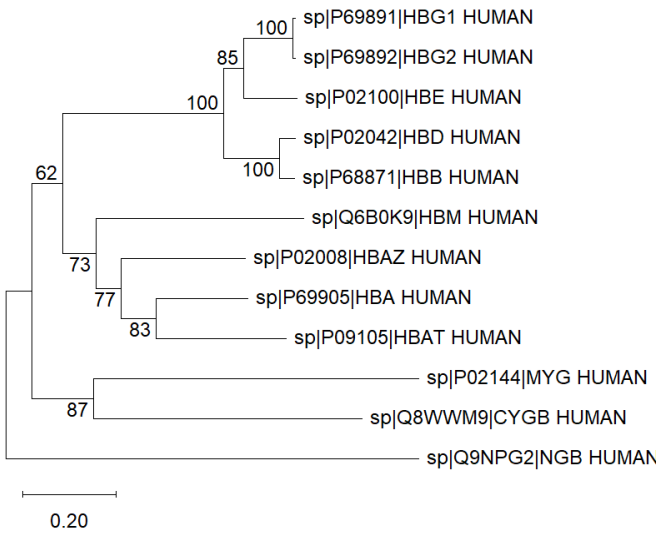
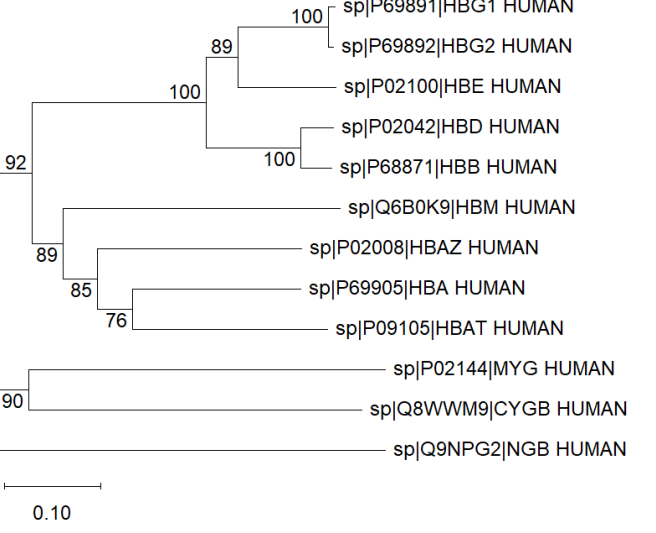
方法	基本原理	优点	缺点
邻接法 (NJ)	基于距离矩阵最小化 (距离聚合)	快速、简单	准确性一般
最大简约法 (MP)	选择最少变异次数的树	直观、参数简单	对长枝吸引敏感
最大似然法 (ML)	在进化模型下找最可能产生数据的树	准确性高、统计基础扎实	计算量大
贝叶斯推断 (BI)	结合先验和似然概率	支持度更直观、高效 MCMC	需设先验、计算资源需求 高
距离法变种 (UPGMA)	用平均距离构建树	快	要求分子钟假设

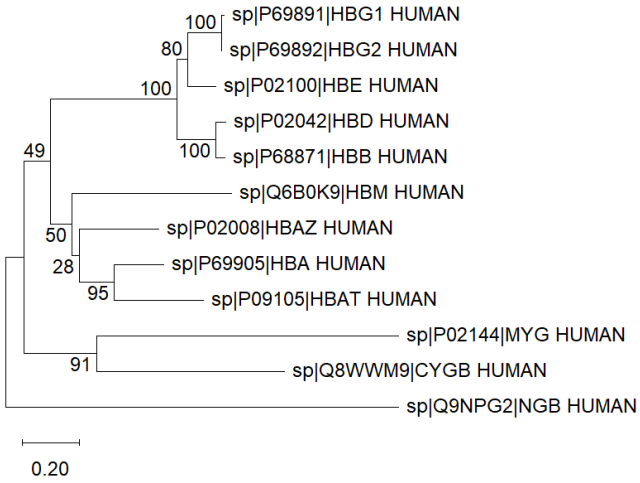
3. 人珠蛋白基因家族系统发生树实例

(1) 以人珠蛋白基因家族 12 个成员蛋白质序列，用 MEGA 邻接法构建系统发生树；选择不同氨基酸替换模型 (Substitution Model)，比较所构建的系统发生树的拓扑结构和稳定性值 (Bootstrap value)，说明不同替换模型对结果的影响。

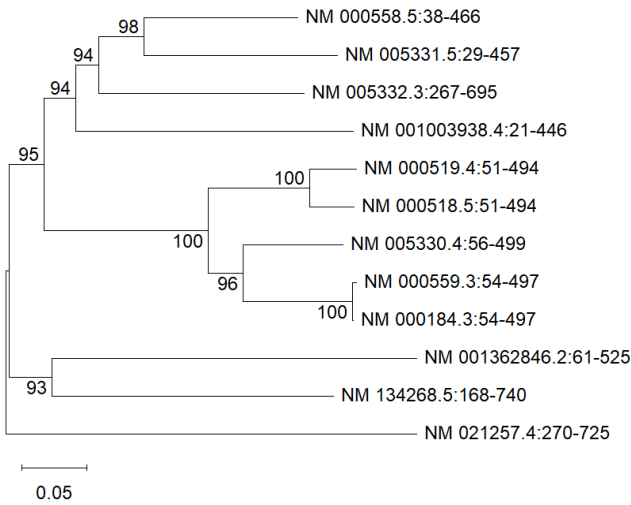
Bootstrap value: 衡量系统发生树中某个分支 (节点) 可信度的统计指标，范围通常为 0~100 (或 0~1)。数值越高，该分支越稳定、可靠。一般认为：

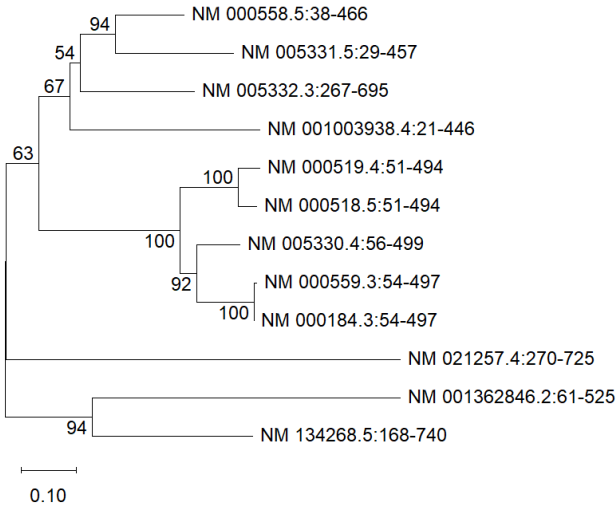
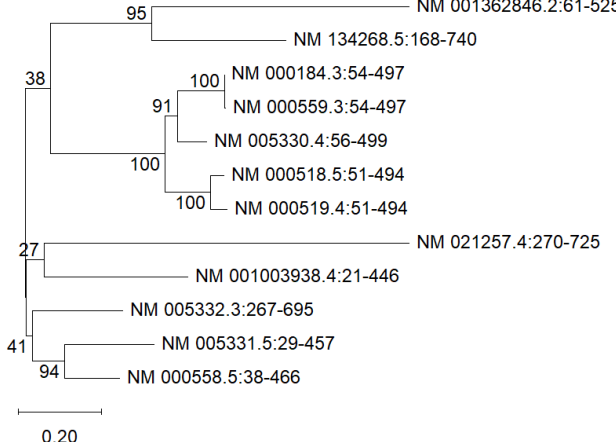
- $\geq 70\%$: 可信度较高，拓扑结构可靠。
- 50%~70%: 可信度中等，需谨慎解释。
- $< 50\%$: 不可信，该分支可能不稳定，不宜作为结论依据。

Substitution Model	Phylogeny	Difference
p-distance		<p>拓扑结构: 完全一致，均表现为：HBG1, HBG2, HBE 与 HBD, HBB, HBM 聚在一起；HBAZ, HBA, HBAT 一起；MYG, CYGB, NGB 独立成支。 拓扑结构对替换模型不敏感,珠蛋白演化关系的核心结论可靠。</p> <p>稳定性值:</p>
Poisson		<p>listance: 所有节点的 bootstrap 值 $\geq 76\%$，表为“高稳定性”。但由模型简单，未校正多重替换和回复突变，在列差异较大时会高估持率，导致产生假阳分支；</p> <p>isson: 部分节点降至等水平，但仍高于 %。校正了部分多重替，但忽略氨基酸替换好，结果介于两者之；</p>

<p>Jones-Taylor-Thornton (JTT)</p>		<p>JTT: 多个节点支持率极低，揭示这些分支实际上不可靠。基于经验替换矩阵，真实反映演化过程，暴露信号微弱的不稳定分支。 稳定性值对模型敏感，简单模型（p-distance）给出虚假的高支持率；复杂模型（JTT）给出真实但偏低的支持率。</p>
<ol style="list-style-type: none"> 1. 用 Uniprot 下载的人珠蛋白家族的 12 个成员蛋白序列，使用 MEGA12 构建系统发生树； 2. Substitution Model 包括 No. of differences, p-distance, poisson model, equal input model, Dayhoff model, Jones-Taylor-Thornton。选取了 3 种进行比较，默认自举法重复数 100。 		

(2) 以人珠蛋白基因家族 12 个成员编码区序列，用 MEGA 邻接法构建系统发生树；，选择不同核苷酸替换模型，比较所构建的系统发生树的拓扑结构和稳定性值（Bootstrap value），说明不同替换模型对结果的影响。

Substitution Model	Phylogeny	Difference
<p>p-distance</p>		<p>拓扑结构： 完全一致，均表现为： HBG1, HBG2, HBE 与 HBD, HBB, HBM 聚在一起； HBAZ, HBA, HBAT 聚在一起； MYG, CYGB, NGB 作为外群。 拓扑结构对核苷酸替换模型不敏感，珠蛋白演化关系的核心结论可靠。</p> <p>稳定性值： p-distance: 最简单，对隐藏</p>

<p>Tajima-Nei</p>		<p>替换完全未校正，导致分支长度被压缩，Bootstrap 值系统偏高。容易误将远缘且速率相近的序列聚在一起，产生虚假的高置信度。</p> <p>Tajima-Nei: 中等复杂度，考虑了碱基组成偏倚和转换/颠换差异，但仍未充分考虑所有替换类型，Bootstrap 值较低，提示部分分支信号不强。</p> <p>Maximum Composite Likelihood: 较复杂，近似最大似然的距离校正，同时估计多个替换参数，更符合分子演化实际。Bootstrap 值能真实反映数据：只有强演化信号的分支保持高支持，而内部细分化分支支持率极低，说明这些关系在核苷酸水平上不稳定。</p> <p>稳定性值对模型选择极为敏感，简单模型给出虚假的高支持，复杂模型暴露真实的不确定性。</p>
<p>Maximum Composite Likelihood</p>		
<ol style="list-style-type: none"> 1. 使用 NCBI 下载的人珠蛋白家族的 12 个成员 CDS 序列，使用 MEGA12 构建系统发生树； 2. Substitution Model 包括 No. of differences, p-distance, Jukes-Cantor model, Kimura 2-parameter model, Tajima-Nei model, Tamura 3-parameter model, Tamura-Nei model, Maximum Composite Likelihood, LogDet (Tamura-Kumar)。选取了 3 种进行比较，默认自举法重复数 100。 		

(3) 根据所构建的系统发生树，参阅 Burmester 和 Hardison 论文，说明人珠蛋白基因家族 12 个成员之间的演化关系。

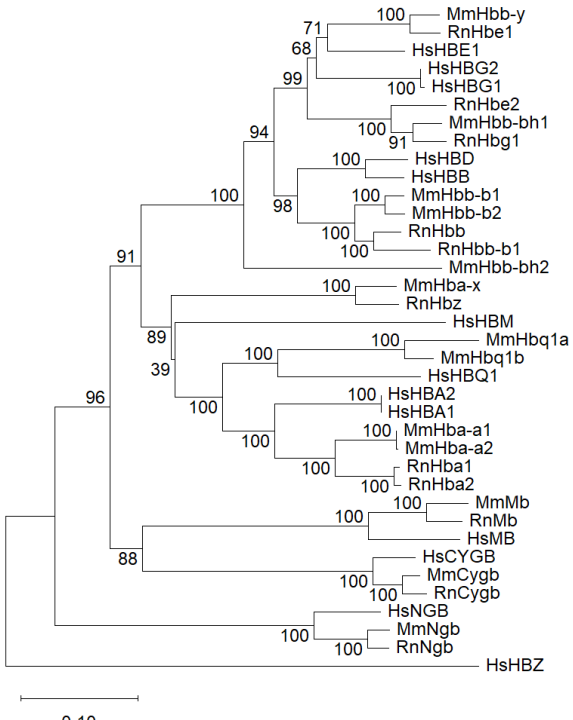
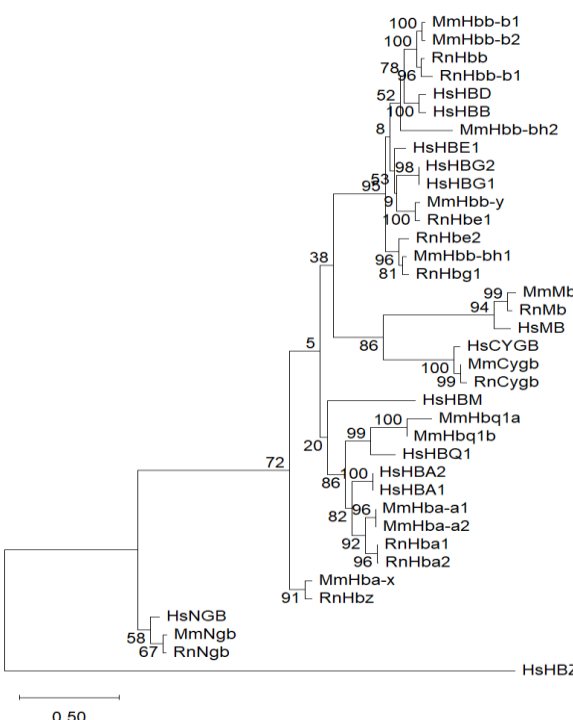
最早分化出肌红蛋白 (MYG) 和脑/胞红蛋白 (NGB、CYGB) 分支。

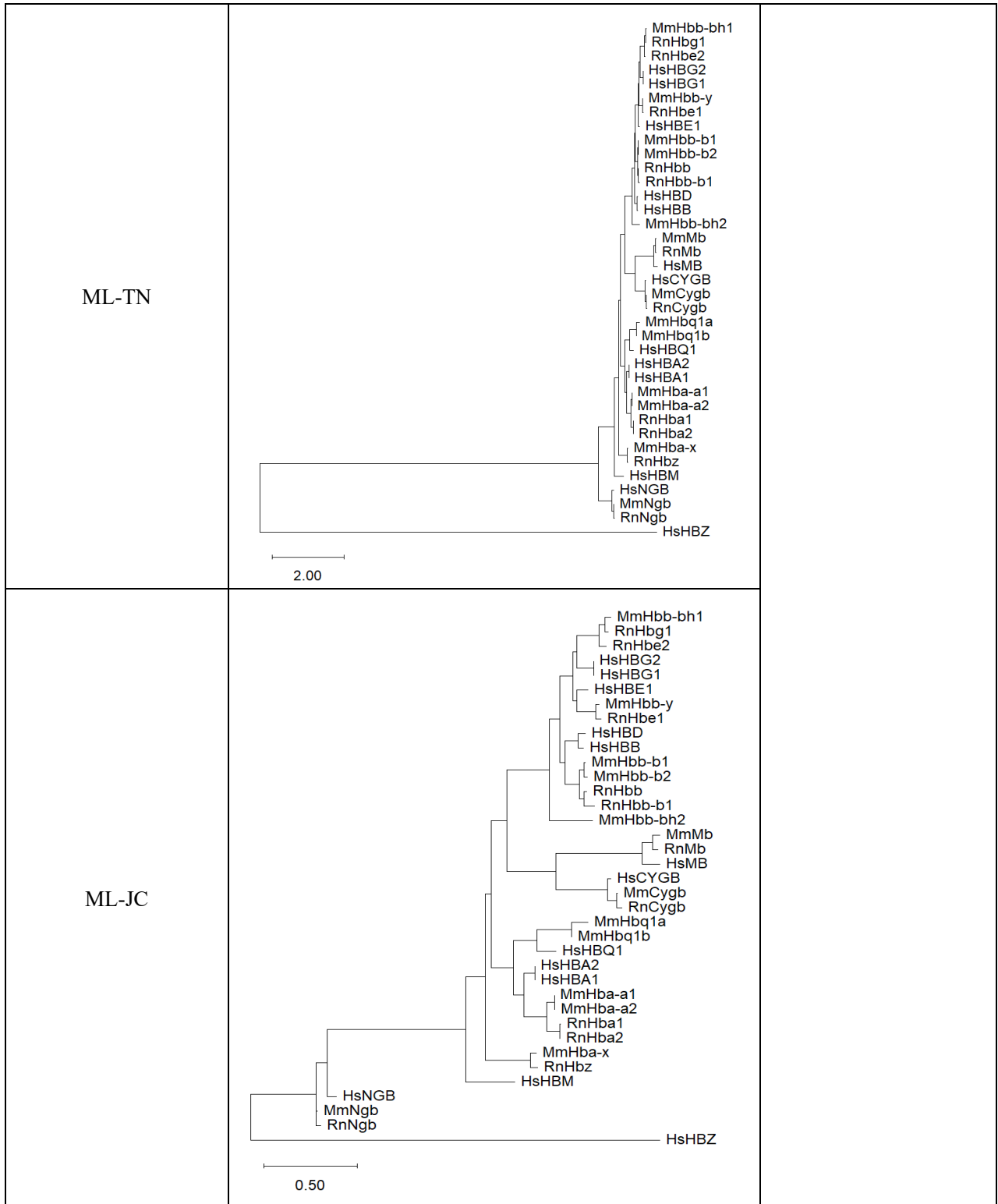
随后形成 alpha-珠蛋白簇 (HBAZ、HBA、HBAT) 和 beta-珠蛋白簇 (HBE、HBG1、HBG2、HBD、HBB、HBM)。

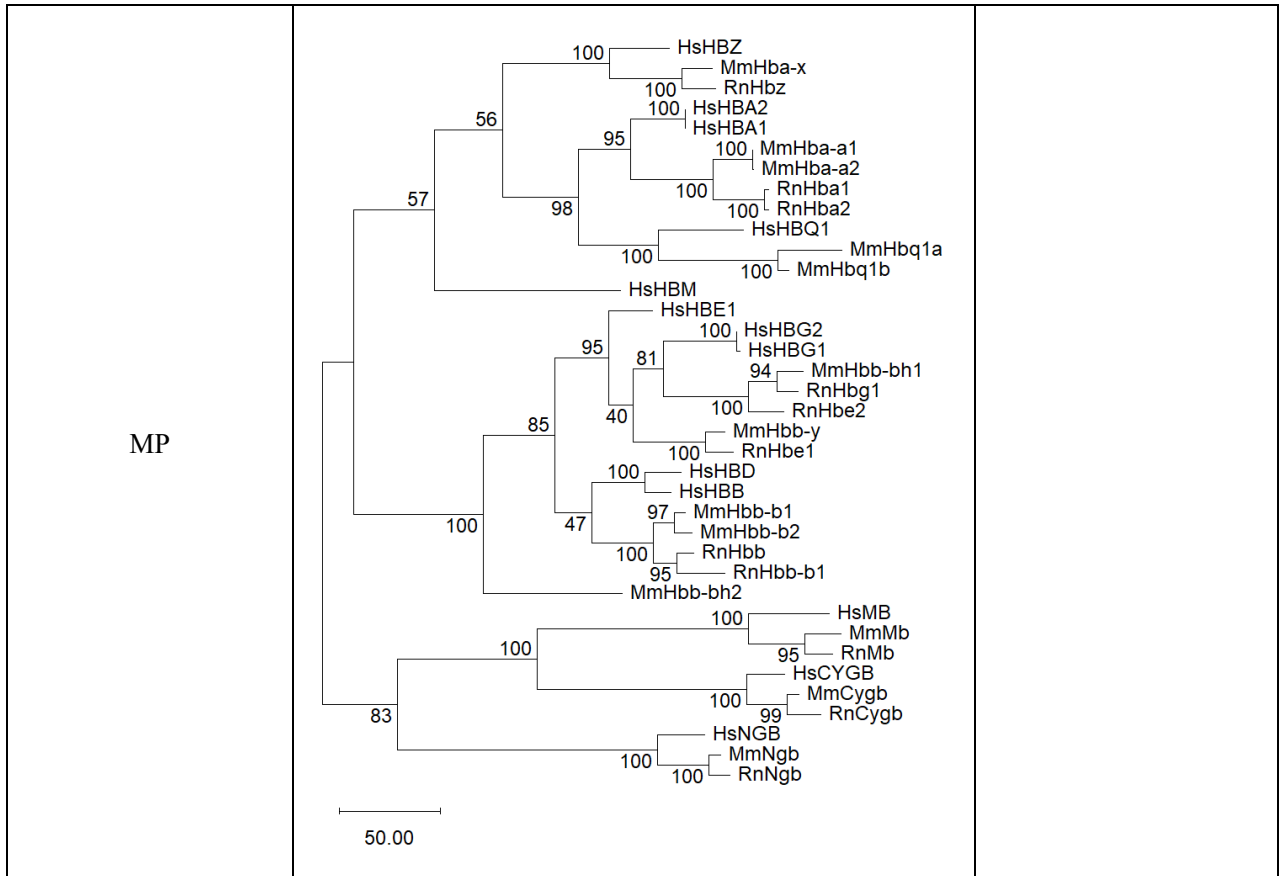
在 alpha 簇和 beta 簇内部，进一步通过串联重复产生了胚胎 (zeta、epsilon)、胎儿 (gamma) 和成人 (alpha、beta、delta) 等成员。

4. 人、小鼠和大鼠三个物种珠蛋白家族系统发生树实例

(1) 以人、小鼠和大鼠三个物种珠蛋白家族 37 个成员编码区序列，采用邻接法、最大简约法和最大似然法构建系统发育树，选择适当的替换模型和参数，比较采用不同方法、不同模型和不同参数时所构建的系统发生树的拓扑结构和稳定性值。

Method - Model	Phylogeny	Difference
<p>NJ- p-distance</p>		<p>拓扑结构: 大部分基因簇(如 MB、HB、HBA/HBA2 等)在五种方法中都能聚成一致的分支。说明这些基因的系统关系比较稳。</p> <p>分支长度: NJ 的两张图长度差异较大 (pdistance vs mcl), 说明距离矩阵和模型选择会影响分支尺度。</p> <p>支持度 (Bootstrap): ML 和 MP 树上有标注数字(一般是 bootstrap 值, 0-100), 可以看到: >90 表示这些分支非常稳定; <50 说明存在不确定性。</p>
<p>NJ-MCL</p>		<p>模型影响: ML 的 JC 和 TN 模型差异主要在一些小分支的分化上; JC 是最简单的替代模型, 假设所有碱基替代率相同; TN (Tamura-Nei) 考虑了不同碱基频率和转换/颠换差异, 更符合真实序列进化。</p>





1. 使用 ABC 教学网站下载的人、小鼠和大鼠的珠蛋白家族的 37 个成员氨基酸序列，使用 MEGA12 构建系统发生树；
2. 使用最大简约法，以及在邻接法和最大似然法中分别选取的 2 种模型，共 5 种模型建树进行比较，默认自举法重复数 100。

(2) 根据上述人、小鼠和大鼠三个物种珠蛋白家族 37 个成员编码区序列系统发生树，参阅相关文献，说明珠蛋白基因家族的起源和演化。

起源：珠蛋白基因家族源自早期脊椎动物之前的单一祖先基因，通过基因重复和片段复制形成多个亚家族，包括 α -血红蛋白、 β -血红蛋白、肌红蛋白（MB）、脑血红蛋白（NGB）等。

功能分化：不同成员逐渐分化出特定功能：MB 储存和释放肌肉中的氧，HB 在红细胞中运输氧，NGB 供神经组织氧，CYGB 可能参与抗氧化。核心功能位点高度保守，非核心区域则产生多样性以适应不同组织环境。

系统发育与演化：人、小鼠、大鼠 37 个珠蛋白基因编码区序列的系统发育树显示，MB 和 NGB 等早期分化， α -和 β -血红蛋白形成独立簇，分支支持值高，说明关系稳定。整体演化体现了基因复制驱动功能分化的模式，为理解脊椎动物氧运输适应性和分子进化提供了依据。

5. 问题

Q: 使用 MEGA 序列比对时, “Alignment” 何时使用 “codons” 选项?

A: “Codons” 选项的意义: 保证比对时不会破坏阅读框 (reading frame), 即保证每个三联密码子仍然对应原来的氨基酸。由于蛋白序列已经是翻译后的氨基酸序列, 不再有三联密码子概念, 所以不需要使用 “codons” 选项。

- Align by MUSCLE / ClustalW: 直接对蛋白质序列或核苷酸序列做比对。
- Align by MUSCLE (Codons) / ClustalW (Codons): 适用于核酸序列 (DNA/RNA) 的比对, 按三联密码子 (codon) 进行。

必须用 codon 对齐的场景: 比对核酸序列, 想保留翻译后的蛋白质信息; 构建核酸系统树时, 需要考虑同义/非同义替换; 需要计算 dN/dS (非同义/同义替换率) 或做分子进化分析。

小组个人总结

G4A-高嘉禾:

由于之前其他课使用过 MEGA7, 所以我对于 MEGA 软件构建系统发育树的流程并不陌生, 但之前并没有系统的学习关于系统发育树以及 MEGA 这个软件的使用。通过课堂上的实操练习, 我学会了并且理解了如何使用 MEGA 软件进行序列比对与构建系统发育树。对于这部分的学习, 我认为理解系统发育树的核心信息是非常关键, 这些信息能够揭示基因家族的演化关系和物种间亲缘关系, 从而指导育种策略, 进行功能基因分析。

在建树时, 对序列的选择十分重要, 对于远缘物种或不同家族的基因, 使用蛋白质序列能够更准确地显示功能保守性; 对于近缘物种或同一基因家族内的基因, 使用核酸 CDS 序列进行比对能反映微小变异和同义/非同义替换。

通过练习邻接法 (NJ)、最大似然法 (ML) 等方法, 我掌握了不同方法的适用场景及特点。另外在比对序列时, “Codons” 选项用于保证比对时不破坏阅读框, 从而保持每个三联密码子对应其原始氨基酸。对于蛋白质序列, 不需要使用 “Codons” 选项, 而是直接使用 Align by MUSCLE / ClustalW 进行比对; 对于核酸序列, 若要保留翻译后的蛋白信息、在构建系统树时考虑同义与非同义替换或要计算 dN/dS 比率, 则使用 Align by MUSCLE (Codons) / ClustalW (Codons) 进行三联密码子的对齐。

系统发育树的拓扑结构直接反映了不同品种或个体间的分化顺序与亲缘远近。在分析我国地方猪种 (如民猪、金华猪、太湖猪) 与欧洲商业猪种 (如长白猪、大白猪) 的遗传关系时, 树的拓扑结构可以揭示独立成支的地方猪种和与外来猪种存在基因交流的猪种。

树的分支长度代表遗传差异大小, 长度越短, 说明品种间遗传背景越接近。通过 Bootstrap 检验判断树中节点的可靠性, 明确了 $\geq 70\%$ 的节点为可靠分支, 可以通过分支长度评估某些关键性育种相关基因是否经历过复制或保守演化。同时也了解了不同的模型构建出来的树的稳定性不同, 简单模型可能高估分支支持, 复杂模型能真实反映分支稳定性。

通过本次学习，我已经能够使用 MEGA 进行序列比对、建树和结果分析，能基本准确解读系统发育树的拓扑结构、分支长度，从中提取实际信息，并且根据自己的需求美化构建的系统发育树。这部分的学习不仅提升了我对 MEGA 的操作能力，更强化了我理解树形信息、判断基因演化与物种关系的专业能力，为未来的研究中使用基因组和转录组数据进行系统分析奠定了基础。

G4B-赵雪倩：

本次关于系统发育树构建的系统学习，对我而言具有重要的理论意义和实践价值。抗菌肽作为生物体内天然免疫的关键效应分子，其基因的演化、复制、丢失以及在不同物种中的分布模式，往往需要借助系统发育分析来揭示。通过本次以人珠蛋白家族、脊椎动物 α 血红蛋白以及植物 18S rRNA 等为例的系统发育树构建训练，我基本掌握了 MEGA 软件的基本操作。

首先，在序列比对与建树方法的选择上，抗菌肽通常序列较短，且不同家族间保守性差异大。对于同一家族内亲缘关系较近的抗菌肽（如某些防御素家族），优先使用核酸 CDS 序列进行比对和建树，并选择 Align by ClustalW 模式，充分考虑密码子的兼并性和偏好性，能够更准确地反映同源关系。而对于不同家族的抗菌肽，或跨远缘物种的比较，则需要使用蛋白质序列构建系统发育树，因为氨基酸的保守性更能体现功能上的相似性。报告中提到的“亲缘关系远用蛋白质、亲缘关系近用核酸”这一原则，将直接指导我后续对抗菌肽家族演化历史的分析。

其次，在构建方法上，我掌握了邻接法（NJ）的基本操作与参数设置，特别是自举法（Bootstrap）重复次数和分支可信度的判断标准。报告中强调 Bootstrap 值 $\geq 70\%$ 的节点才认为可信，这一标准对于抗菌肽研究中判断基因复制事件发生的时间、区分直系同源与旁系同源至关重要。例如，当我在不同物种中鉴定到新的抗菌肽序列时，只有构建出高支持的发育树，才能可靠地判断这些序列是来源于共同祖先的直系同源（功能可能保守），还是物种特异的基因复制产生的旁系同源（功能可能分化）。报告中关于“先有基因、后有物种”与“先有物种、后有基因”的讨论，让我深刻认识到基因树和物种树的区别——抗菌肽基因家族往往经历频繁的基因复制和丢失，因此不能简单地将抗菌肽的基因树等同于物种树。

此外，通过分析“7 个代表性脊椎动物 α 血红蛋白系统发生树”这一案例，我获得了重要的启发。报告中指出， α 血红蛋白构建的树在某些分支上无法正确反映物种的系统发生关系，原因是基因序列的差异受到环境适应、基因丢失或功能转换的影响。这一发现对抗菌肽研究极具警示意义：抗菌肽作为免疫相关基因，往往受到强烈的选择压力，趋同进化、适应性进化和基因丢失等现象可能更为常见。因此，在利用抗菌肽序列构建发育树时，不能简单地将其结果直接解释为物种间的演化关系，而应结合其他保守基因（如 18S rRNA）构建的物种树进行对比分析，区分“基因演化”与“物种演化”的不同信号。

最后，报告中关于 18S rRNA 作为分子标记的系统发育分析，也为我提供了研究抗菌肽宿主物种背景的方法工具。在挖掘新抗菌肽时，准确鉴定宿主物种是前提，而 18S rRNA 是真核生物物种鉴定的常用分子标

记。掌握利用 18S rRNA 构建系统发育树的方法，可以帮助我在发现新序列时快速确认其分类地位，并为后续的抗菌肽功能注释提供可靠的物种背景。

综上所述，本次系统发育树构建的学习，为我从事抗菌肽的演化分析、功能预测和物种鉴定提供了坚实的生物信息学基础。我将把所学技能应用于双斑萤叶甲致死基因挖掘和桃柱螺中肠微生物分析等课题中，对抗菌肽基因家族进行系统发育分析，探索其演化规律与功能分化的关系。

G4C-张俊鑫：

这段时间在课程中学习了系统发生树的构建与相关分析的基础理论和核心操作。在学习过程中，我首先理清了系统发生树的基本原理、拓扑结构、分支长度、置信度等系统发生树核心概念的含义，不再是看着树形结构不知道怎么阅读与理解，而是能理解每个结构对应的物种进化亲缘关系的逻辑。同时，我掌握了建树的完整学习流程，从序列比对、数据筛选，到模型选择、初步构树，一步步吃透了背后的理论依据。我也明白了不同构树算法的区别和适用场景。通过这次学习，我学到了种质资源进化分析的又一个理论基础与常用技术，建立了系统的进化发育分析思维。也意识到生物信息分析严谨性的重要性，每一步学习积累的理论知识，都为我后续的专业学习打下了坚实基础。后续我也会继续巩固知识点，熟练掌握相关理论逻辑，精进自身专业能力。