

密码子使用偏好性的跨物种比较分析

1. 引言

密码子是 mRNA 上三个相邻的负责编码特定氨基酸的核苷酸组合。编码相同氨基酸的不同密码子称为同义密码子。然而，不同物种会非随机地偏好使用某些同义密码子，这种现象称为密码子使用偏好性（Codon Usage Bias）。不同物种具有一致且特征性的密码子偏好模式。密码子使用偏好性与基因表达水平、基因组组成、自然选择和进化压力、突变压力、基因的位置和功能以及环境等因素有关，不同物种的密码子使用偏好性模式反映了它们在进化过程中对各种内外部因素的适应性调整。哺乳动物，如人类和小鼠，通常表现出高度保守的密码子优化策略，以适应复杂的多细胞调控需求；而单细胞真核生物，如酵母则可能通过更灵活的密码子使用适应快速变化的环境压力。因此，比较不同物种的密码子偏好模式，有助于揭示进化过程中的保守机制和适应性分化。

本研究选取人类（*Homo sapiens*）、小鼠（*Mus musculus*）和酿酒酵母（*Saccharomyces cerevisiae*）作为研究对象，通过多维度可视化比较这三个物种的同义密码子使用特征，分析其偏好模式的保守性与差异性。通过柱状图展示亮氨酸（L）的密码子使用比例的物种间差异，通过散点图揭示人类与小鼠密码子频率的高度相关性，通过热图比较不同氨基酸密码子偏好的熵值分布，通过箱线图分析各物种密码子使用频率的整体分布特征。

2. 方法

2.1. 数据来源

本研究使用的人类、小鼠和酿酒酵母密码子频率数据来源为 Kazusa 密码子数据库（<https://www.kazusa.or.jp/codon/>）。

2.2. 分析方法

采用四种统计图表进行多角度分析：使用柱状图比较特定氨基酸（如亮氨酸）的同义密码子使用比例；使用散点图展示人类与小鼠密码子频率的线性关系；使

用热图呈现各氨基酸密码子偏好的熵值分布；使用箱线图分析密码子频率的整体分布差异。

2.3. Python 工具包

pandas: 数据清洗、表格操作与统计分析

numpy: 支持数值计算（如熵值计算）

matplotlib: 生成基础图形（柱状图、散点图、箱线图）

seaborn: 优化图形样式（热图、回归线、颜色映射）

openpyxl: 读取 Excel 原始数据

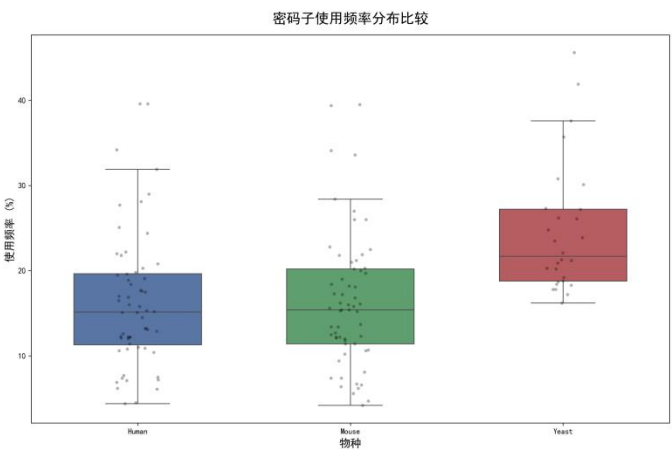
2.4. 代码复现

GitHub 链接: <https://github.com/LyuXYue/lvxiaoye-12433463/tree/main>

3. 结果

3.1 密码子使用频率分布比较

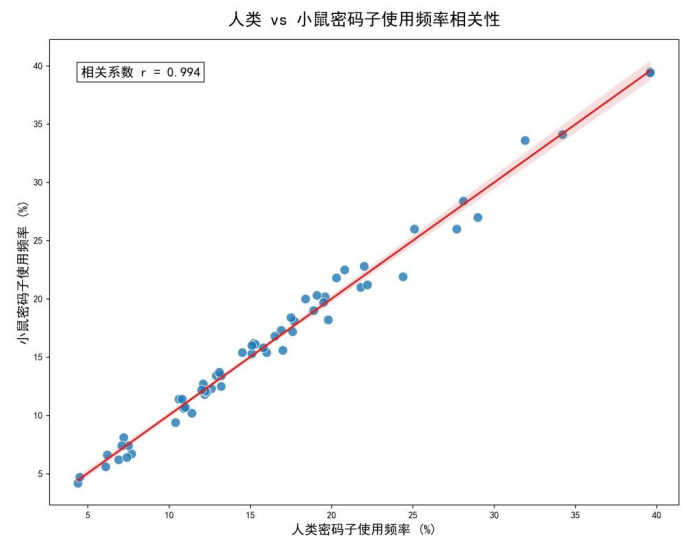
人类密码子使用频率的中位数约为 20%，分布范围较广，从 10%到 40%；小鼠：密码子使用频率的中位数略低于人类，分布范围也较广，但整体频率略低于人类；酵母密码子使用频率的中位数最高，分布范围从 15%到 35%，显示出酵母在密码子使用上的集中性。



3.2 人类 vs 小鼠密码子使用频率相关性

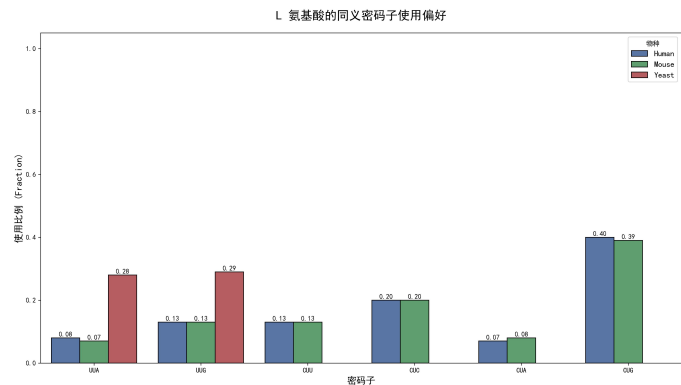
相关系数 $r=0.994$ 表明人类和小鼠在密码子使用频率上具有非常高的相关

性，这可能反映了两者在进化上的亲缘关系以及相近的生物学需求。



3.3 亮氨酸 (L) 同义密码子使用偏好

UUA 在酵母使用频率最高 (0.28)，人类 (0.08) 和酵母 (0.07) 中使用频率较低；UUG 在酵母使用频率最高 (0.29)，在人类和小鼠中较低且相同，均为 0.13；CUU 在人类和小鼠中使用频率相同，均为 0.13；CUC 在人类和小鼠中使用频率也相同，均为 0.2；CUA 在人类和小鼠中使用频率相近，在小鼠中略高，为 0.08，在人类中为 0.07；CUG：在人类和小鼠使用频率最高，分别为 0.40 和 0.39，小鼠较低，为 0.13。



3.4 不同物种氨基酸密码子偏好熵值比较

熵值是衡量密码子使用偏好性的指标，值越高表示偏好性越强。对于精氨酸 (R)，人类和小鼠的熵值最高 (2.51)，酵母较低 (0.99)，表明人类和小鼠在

精氨酸密码子使用上表现出更强的偏好性。而对于丝氨酸（S），酵母的熵值最高（2.47），人类和小鼠较低（2.46 和 1.45），表明酵母在丝氨酸密码子使用上表现出更强的偏好性。其他大多数氨基酸在人类和小鼠中的熵值接近 1，表明密码子使用偏好性较为均衡；而酵母的熵值在某些氨基酸上显著不同，显示出酵母在这些氨基酸密码子使用上的特定偏好。



4. 分析与讨论

4.1 结果分析

相似性：人类和小鼠在密码子使用频率上表现出高度相关性，这可能反映了两者在进化上的亲缘关系以及相似的生物学需求。人类和小鼠在大多数氨基酸的密码子使用上表现出较为均衡的偏好性，这可能与两者复杂的生理功能和环境适应性有关。

差异性：酵母在某些氨基酸（如精氨酸和丝氨酸）的密码子使用上表现出更强的偏好性，这可能与酵母独特的生物学特性和环境适应性有关。酵母的密码子使用频率分布较为集中，这可能反映了酵母在特定环境条件下对翻译效率的优化需求。

不同物种的密码子使用偏好性受到多种因素的影响，包括进化关系、生物学需求和环境适应性。人类和小鼠在密码子使用上表现出高度相关性，而酵母则在某些氨基酸的密码子使用上表现出更强的偏好性。了解不同物种的密码子使用偏好性有助于我们更好地理解基因表达调控机制和物种适应性进化。

4.2 研究局限性

- (1) 数据来源限制：本研究依赖于特定数据库中的密码子使用数据，可能存在数据不完整或更新不及时的问题。不同数据库可能采用不同的方法来计算密码子使用频率，这可能影响结果的一致性。
- (2) 物种代表性：本研究仅分析了人类、小鼠和酿酒酵母三种物种，可能无法全面代表所有真核生物或原核生物的密码子使用偏好性。不同物种可能因为其独特的生物学特性和进化历史而展现出不同的密码子使用模式。
- (3) 环境和条件影响：密码子使用偏好性可能受到环境条件的影响，如温度、营养状态等，本研究未能考虑这些环境因素对密码子使用的影响，这可能限制了对密码子使用偏好性全面理解。

4.3 未来研究方向

未来的研究可以通过扩大物种范围，包括更多的真核生物和原核生物，来更全面地理解不同生物类群的密码子使用偏好性。同时，考虑环境条件对密码子使用的影响，结合基因的功能注释和表达数据，将有助于揭示密码子使用偏好性与基因功能和表达水平之间的关系。此外，采用实验验证，如密码子替换实验，可以进一步确认计算分析结果的准确性，而深入的跨物种比较分析将有助于揭示不同物种间密码子使用偏好性的相似性和差异性，并探讨其进化意义。这些研究方向将有助于我们更全面、深入地理解密码子使用偏好性及其对生物学功能的影响。