

常见问题解析

1. 如何避免混淆相似性与同源性？

高相似度（如90%）并不等同于同源性，可能是由于因趋同进化或短片段匹配导致误判。可参考三项指标Per. Ident项>70% + Query Cover项>90%以及E值<1e-20来进行判断。或采用InterProScan分析保守域，避免仅依赖序列相似度。

2. 如何减少覆盖度（Query Cover）引起的误判？

低覆盖表明仅匹配局部碱基序列区域，无法代表全长基因的功能（如仅匹配蛋白结构域而非整体）。可根据匹配的Accession号下载数据库中的全长序列重新进行比对分析来验证。或通过CD-Search验证功能域的完整性。

参考文献

- [1] Dudnyk K, Cai D, Shi C, et al. Sequence basis of transcription initiation in the human genome [J]. *Science (New York, NY)*, 2024, 384(6694): eadj0116.
- [2] Guillotin B, Rahni R, Passalacqua M, et al. A pan-grass transcriptome reveals patterns of cellular divergence in crops [J]. *Nature*, 2023, 617(7962): 785-791.