

# 两个**独立**计算pipeline（管道）的结果说明

**pipeline1流程：** 去除接头引物和宿主污染→cleaned reads→**比对到病毒蛋白库**→**根据e值进行整理**

**lower\_1e-6文件夹：** e值低于1e-6的**命中reads**（一般假阳性较低，和已知病毒蛋白相似性高，重点看这个）；

lower\_1e-3文件夹：**命中reads**满足 “ $1e-6 < e\text{值} < 1e-3$ ” ；

lower\_1e-1文件夹：**命中reads**满足 “ $1e-3 < e\text{值} < 1e-1$ ” 。

**注：**该管道会提取出命中到每个参考毒株的**reads**序列，存放在每个类别下的**reads\_out**文件夹里

---

**pipeline2流程：** 去除接头引物和宿主污染→cleaned reads→**拼接成contig**→**比对到病毒蛋白库**→**根据e值进行整理**

**lower\_1e-6文件夹：** e值低于1e-6的**命中contigs**（一般假阳性较低，和已知病毒蛋白相似性高，重点看这个）；

lower\_1e-3文件夹：**命中contigs**满足 “ $1e-6 < e\text{值} < 1e-3$ ” ；

lower\_1e-1文件夹：**命中contigs**满足 “ $1e-3 < e\text{值} < 1e-1$ ” 。

**注：**该管道会提取出命中到每个参考毒株的**contigs**序列，存放在每个类别下的**contig\_out**文件夹里

## 特别注意 (Important Notes)

metav is not an entirely new method for virus identification. As a pipeline that calls existing software, it aims to reduce the complexity of switching between tools. The identification results from metav are preliminary. During the BLAST step, it is difficult to avoid false positives hits, especially with reads. Therefore, the final output of metav requires further evaluation. For instance, applying additional filters based on the number of hit reads or performing alignment against reference genomes can help reduce false positives. Additionally, a more stringent e-value threshold can be set for BLAST.

metav并非全新的病毒鉴定方法，作为一个调用现有软件的工具，其初衷是简化多软件切换的繁琐操作。metav得到的鉴定结果是初步的。由于在BLAST步骤，假阳性命中难以完全避免，特别是对于短序列reads，因此还需要对metav最后的输出结果进一步评估。例如，根据命中reads数量设置阈值再次进行过滤，或者比对到参考基因组，均可帮助降低假阳性。此外，也可以为BLAST设置更严格的e值阈值。