

练习题：

scaffold: plum_0630.scafSeq.FG 和 gene 注释结果(Prunus_mume_scaffold.gff)下操作：

1.

- a) 统计 scaffold.fa 染色体基本信息，输出的内容与格式如下：

Name	Total Length	Effective Length	N Length	GC Length	GC Rate (%)
scaffold.fa					

结果输出成文本文件。

- b) 计算 scaffold.fa 的 N50 , N90; (N50 定义 google)

- c) 以 250bp 为窗口，无重叠的沿序列滑动，计算 GC 含量，用 R 或者 SVG 做出 GC 含量分布曲线

d) 根据序列 scaffold.fa 以及基因注释文件 scaffold.glimmer.gff，提取 scaffold.fa 上的所有的基因的 CDS 序列，结果以 fa 格式输出；(gff 文件的格式 google 一下就知道每列意义)

- e) 将提取的 CDS 序列翻译成氨基酸序列，结果以 fa 格式输出。(三联体密码子对应的氨基酸自己建 hash)

计算 scaffold.fa 的 N50 ; (N50 定义 google)