练习题：

scaffold：plum\_0630.scafSeq.FG和gene注释结果(Prunus\_mume\_scaffold.gff)下操作：

1.

1. 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。

1. 计算scaffold.fa的N50 ，N90；（N50定义google）
2. 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）