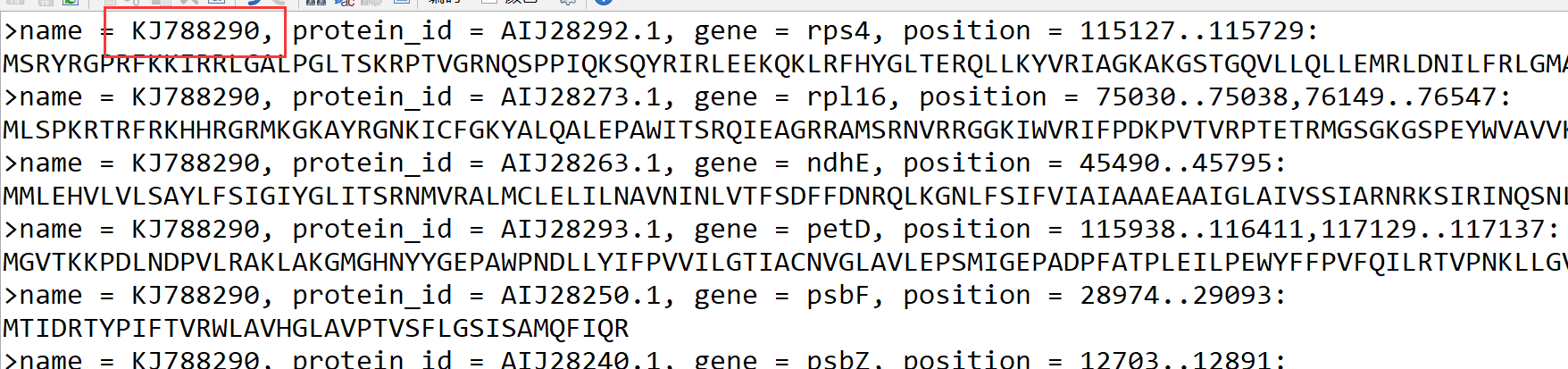
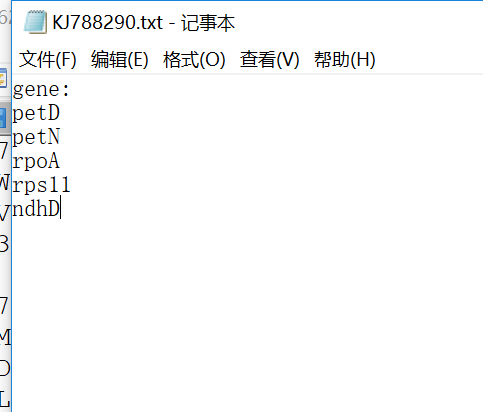
1. 这一步操作的数据都在文件夹13\_data里面。

提取每个文件中的gene名称，每个物种分别存为一个文件，格式可以是txt

例如：以文件KJ88290为例，在它的.fasta文件中如下图所示：

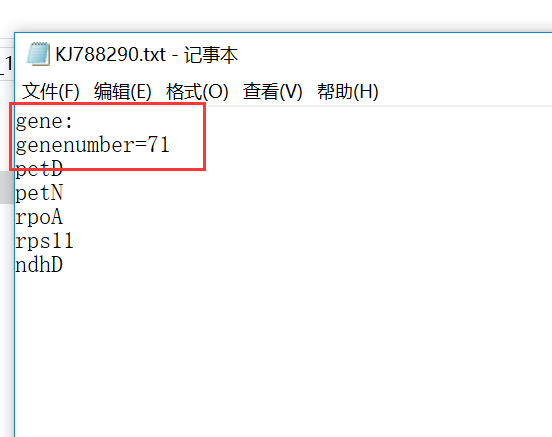


需要将名字为gene的字符提出来：例如：



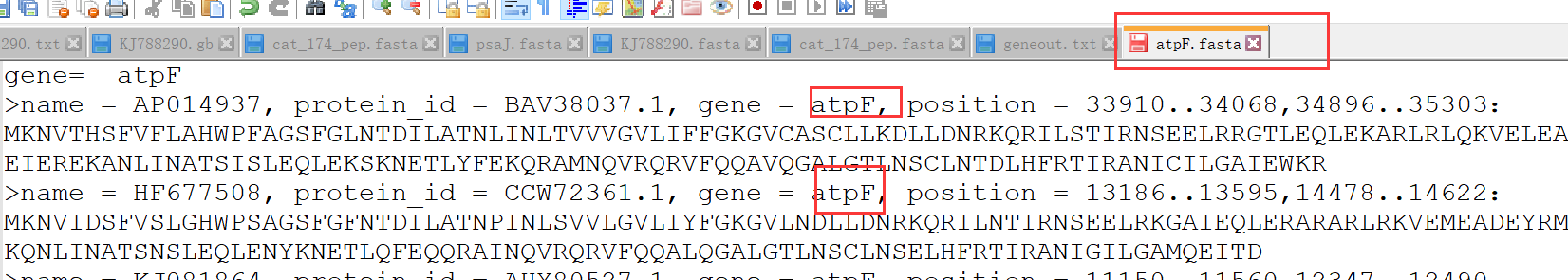
以此类推，找出文件KJ88290中所有gene名字，注意gene不要重复的，如果有重复的，只输出一个，也就是说输出的gene每个都是唯一的，并将基因个数也统计出来，最终单独存为KJ88290.TXT。

最终格式：这个71数字是假设的，这个数字要根据找出来的然后统计。



1. 找出每个基因在所有物种中出现的次数。这一步操作的数据在cat\_174\_pep.fasta里面。

在cat\_174\_pep.fasta 文件中共有174个物种，将每个基因对应的物种名字输出，例如：



以基因的名字命名的.fasta文件，也就是说在这个cat\_174\_pep.fasta找出所有不重复基因的对应的name物种然后输出，注意每个基因名字输出一个文件，然后把这个文件里面不同的name名字的个数统计后写在gene后面。假设atpF这个基因文件里面有171个不同的物种名字，那么它这个文件最终的输出格式应该是：

