20220831

Question1：

基因注释cazy和vfdb，软件是否能够自己跑出来？还是需要手动使用数据库注释？

Answer1：

如果是使用默认数据库进行注释，软件中会自动生成相关gene family信息，可以通过R进行数据库匹配确定

Question2：

如何将自有菌株（自己试验所用的，数据库中没有的菌株）加入数据库中进行分析？

Answer2：

可以先进行泛基因组分析，利用矩阵将自有菌株与数据库中菌株进行AJI（Average jacard index）计算比对，若AJI>=99，则重建过程中自有菌株会与数据库中现有菌株合并，因此没有必要重建。如果AJI<99，可以考虑重建。

**翻译**

Question2：

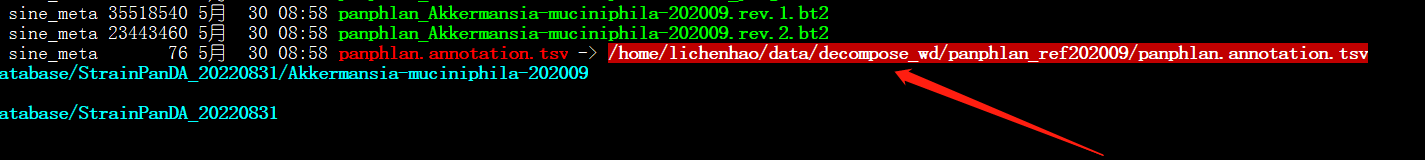
How to analysis your own strains (strains used in your own experiments that are not in the database) using StrainPanDA?

Answer2：

First, you can compare the pangenome of the references in the database with your own strain, and calculate AJI (Average Jacard Index). If AJI >=99, your self-owned strain will be merged with the existing strains in the database during the reconstruction process, therefore database reconstruction is meaningless. If AJI < 99, you may consider rebuilding the database. Or without regeneration, you can compare the predicted gene profile of strains with you own strain to see their similarity.

Question3：

部分数据库中存在额外的软连接文件，是否会影响分析结果？



Answer3：

对使用没有影响，只是多余的文件，在后续的更新中会进行清理。

Question4：

软件使用过程中产生的.counts.csv文件是reads计数，还是tpm值？



Answer4：

.counts.csv文件这个不是StrainPanDA输出的文件，而是panphlan生成的文件，其可作为StrainPanDA的输入。

Question5：

网页和wget下载数据库的速度较慢，是否有快一点的方式？

Answer5：

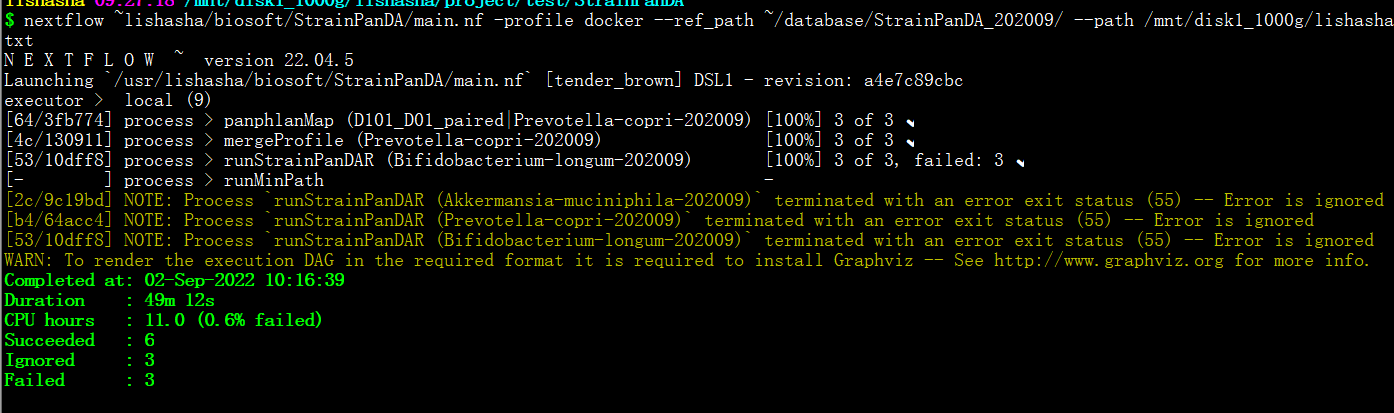
（暂无回答，待师兄补充）

整体来说，你后面其实可以直接在这个文件上增加（已经加了的可以删掉，没有加的先留着，免得问的人多了的时候还得重新写）

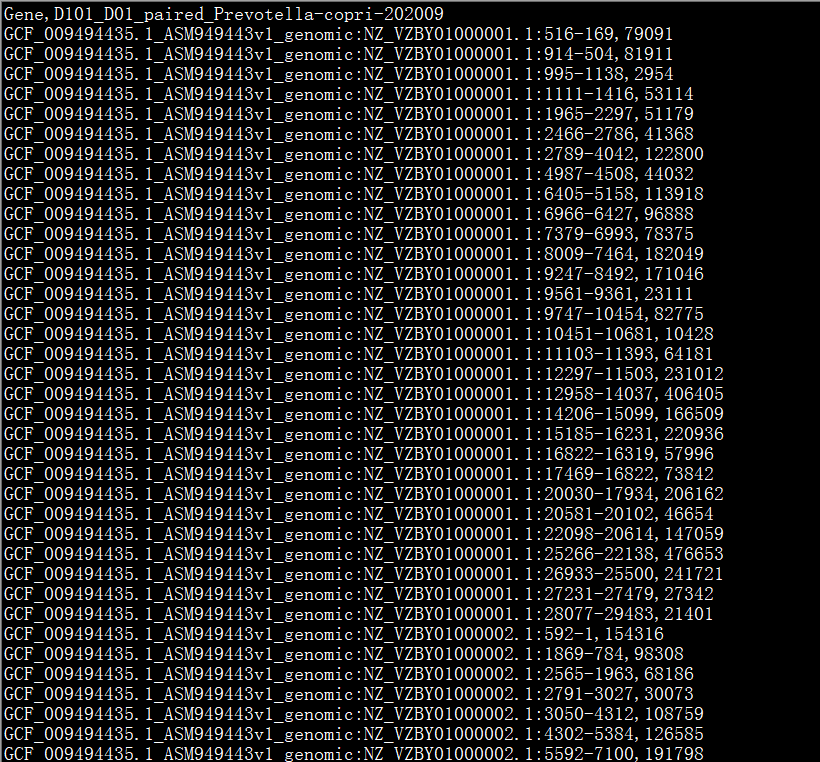
20220902

Question1：

运行程序**第三步**的时候出现失败，截图如下



打开任意 .counts. csv文件，可见如下截图情况



Answer1：

此处用户只使用了1个样品即运行StrainPanDA，正常情况下StrainPanDA最少需要5个样品进行分析，我们推荐20个样品进行分析以获得最佳效果。若是测试StrainPanDA的使用，推荐使用教程提供的数据进行测试

Question2：

是否后续会更新已经建好的数据库？

Answer2：

取决于使用者的需求，可能后期会不定期更新

20220905

Question1：

.conunts.csv表格中存在的样本，在菌株成分表中不存在，是什么原因？

Answer1：

StrainPanDA有对样品进行过滤的指标，如果样品在该物种的覆盖度不足，则该样品会被过滤掉，具体的过滤规则为：如果检测到的基因家族数目低于0.9 × gmin  (gmin为所有参考基因组的基因家族数目中最小的值)，该样本被过滤掉。因此，在.conunts.csv表格中一般会有所有样本（但是如果该样品没这个菌种，也可能会缺失），在菌株成分表中某些样品是可能会不存在的。

**翻译**

Question1:

Why was part of the input sample missing from the final analysis?

Answer1:

StrainPanDA has an index for filtering samples, and if the sample has insufficient coverage of the species, the sample will be filtered out. Specific filtering rules are as follows: **Samples were filtered out if the number of gene families detected was below 0.9 × gmin (gmin is the minimum number of gene families found in all reference genomes).** Therefore, the .counts.csv table will generally have all samples (but may be missing if the sample does not have the strain), and some samples may not be present in the strain composition list.