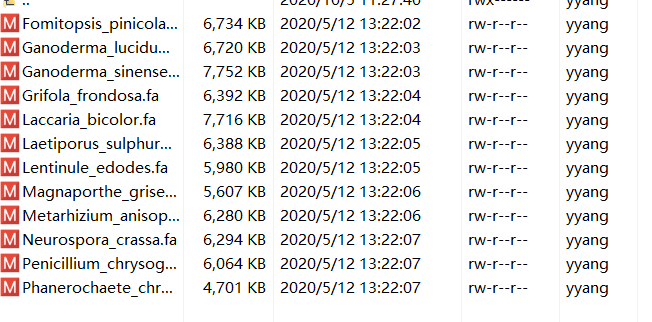
1. 下载相关物种的蛋白质（注意蛋白质序列应该先放到tbtools下改一下基因的名字方便后面排序）
2. 

运行 orthofinder

orthofinder -f Mycoplasma/ -t 2

建文件夹align以及文件夹zli

1. 比对处理

for i in {3086..4305};do muscle -in OG000${i}.fa -out /public/home/yyang/compara/OrthoFinder/Results\_Oct05/Single\_Copy\_Orthologue\_Sequences/align/OG000${i}.fas;done

1. Gblocks它可以消除DNA或蛋白质序列中排列不一致的位置和不同的区域

for i in {3086..4305};do /public/home/yyang/compara/Gblocks\_0.91b/Gblocks OG000${i}.fas -t=p;done

1. linux脚本处理序列

for i in {3086..4305};do seqkit sort --quiet -n -i OG000${i}.fas-gb > /public/home/yyang/compara/OrthoFinder/Results\_Oct05/Single\_Copy\_Orthologue\_Sequences/zli/OG000${i}.fasta;done（排序）

find . -name "\*" -type f -size 0c | xargs -n 1 rm -f 出去0kb文件

1. 建1到10十个文件夹
2. 把一个物种的序列放到一起（因为上一步排序过了）

for i in {3086..4305};do awk -v RS='>' 'NR>1{i++}i>=1&&i<=1{print ">"$0}' OG000${i}.fasta|sed '/^$/d'>/public/home/yyang/compara/OrthoFinder/Results\_Oct05/Single\_Copy\_Orthologue\_Sequences/1/OG000${i}.fa;done >=1 <=1 从一重复到十

1. 合并

建立文件夹last，用hb.sh脚本进行合并

然后grep "^>" -v chrun.fasta | awk '{ ORS = ""; $1 = $1; print $0}' > chrUn.fasta 去掉名字合并为一条大序列，你自己需要添加名字。数据处理结束

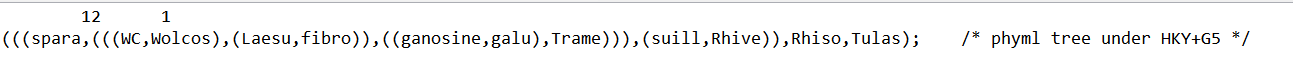
1. 进化树的构建

把得到的序列先用tbtools合并，合并为一个大文件。raxml和mega都行,raxml比较认可

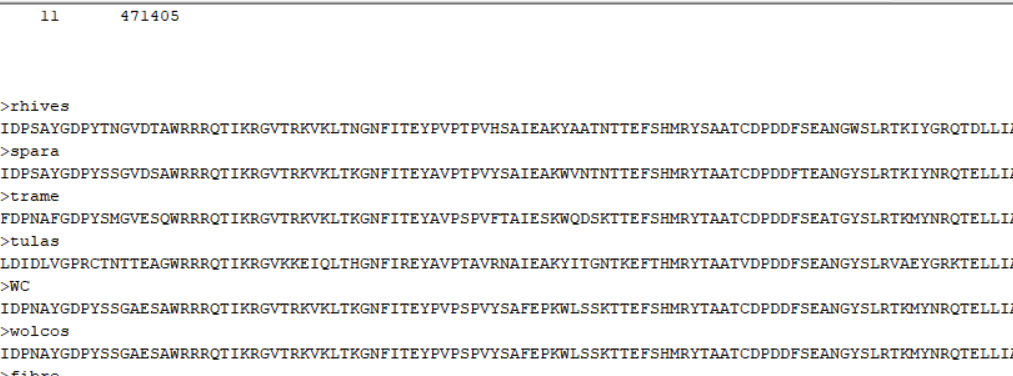
raxmlHPC-PTHREADS-SSE3 \  
-s /public/home/yyang/compara/OrthoFinder/Results\_Oct05/Single\_Copy\_Orthologue\_Sequences/zli/OG0003086.fasta \  
-T 12 \  
-n tre \  
-m PROTGAMMALGX \  
-N 1000 \  
-f a \  
-x 12345 \  
-p 12345 选besttree作为进化树。

1. 时间进化树因为上面第一步我就没做那个是后面补的，所以用公司的数据吧。软件用r8s以及paml里面的mcmctree。

Paml 软件十分讲究格式 fasta蛋白序列我们要以



进化树我们要以



格式。。。上面说反了

然后有可能会出现报错

1601967062(1)

这个是因为软件只能识别2叉树，但是这个进化树不是二叉



我们可以修改进化树或者去掉相应的蛋白序列。

10.1 r8s构建进化树。这个我记得安装复杂，就在自己虚拟机上运行.

r8s构建时间进化树分三步，第一步是准备，第二步运行，第三步提取

<https://blog.csdn.net/u012110870/article/details/102804514> 具体看这个人写的。。

重点整理一下作图

F:\D盘微信文件\进化树美化itol\exercise 里面有研究过的脚本

因为好像字体是改不了的，更多细节在AI上面改或者PPT.

WGD全基因组复制

RuntimeError: Click will abort further execution because Python 3 was configured to use ASCII as encoding for the environment. Consult https://click.palletsprojects.com/python3/ for mitigation steps.

This system lists a couple of UTF-8 supporting locales that you can pick from. The following suitable locales were discovered

报错： 这个报错不要怕，重新运行一次就行了

export LD\_LIBRARY\_PATH=/public/home/yyang/miniconda3/envs/wgd/lib ：开始先输入这个.

grep -vE "^[[:blank:]]\*$" WC.gff3 >111.gff3 ：去除文件空格行

安装很复杂 记得matplotlib版本是3.0.5 参考文章：<https://zhuanlan.zhihu.com/p/123875815>

第一步：wgd mcl -n 12 --cds --mcl -s zzz.fasta -o zzz.out

第二步：wgd ksd --n\_threads 6 /public/home/yyang/WGD/zzz.out/zzz.fasta.blast.tsv.mcl zzz.fasta

第三步：wgd mix /public/home/yyang/WGD/wgd\_ksd/zzz.fasta.ks.tsv

mummer2circos -l -r genomes/NZ\_CP008827.fna -q genomes/\*fna

多基因组圈图 ： conda activate mummer2circos