**Runjelly\_v1.3.1**

**Overview**

Runjell\_v1.3.1是pbjelly的优化版。

* 在运行pbjelly之前用minimap2进行过滤，去除与补洞无关的reads,比对速度可以提升8-10倍。
* Pbjelly使用的blasr1.3会出现内存回收不完全而导致的内存飙升问题，例如比对一个15G的fastq，mapping内存峰值可以达到80G,如果资源不够会出现任务被杀的情况；runjelly的比对软件使用blasr5.3，没有这个问题。Blasr5.3的实际补洞效果也略好于blasr1.3。
* 可以自动投递任务。

**Installation**

集群路径：

阿里云：*/ALBNAS12/Plant/Project/WORK/PBjelly/software/tar/runjelly\_v1.3.1.tar.gz*

将软件包拷贝到流程目录下，解包：

$ tar -zxvf runjelly\_v1.3.1.tar.gz

完成后，进入目录，运行setup.sh:

$ sh setup.sh

将检查pip list中是否包含networkx1.1。（若没有将尝试通过pip安装networkx1.1,安装路径在runjelly\_v1.3.1/pythonlib中）

**Requirements**

* Python 2.7+
* Python package: networkx 1.1+ (networkx 2.0及以上版本会报错）
* blasr5, bedtools，minimap2,libhdf5 (这些软件在软件包中已包含）

**Usage**

1. 准备好reads .fastq文件和reference .fasta(<4G)文件。

如果是bam格式，用prepare.sh脚本（此脚本需要samtools）转换为fastq格式：

$ path\_to\_bin\_dir/prepare.sh input\_bam\_dir Out\_fastq\_dir Size Opts

Input\_bam\_dir: 所有bam文件所在目录路径

Out\_fastq\_dir: 保存输出fastq文件的目录路径

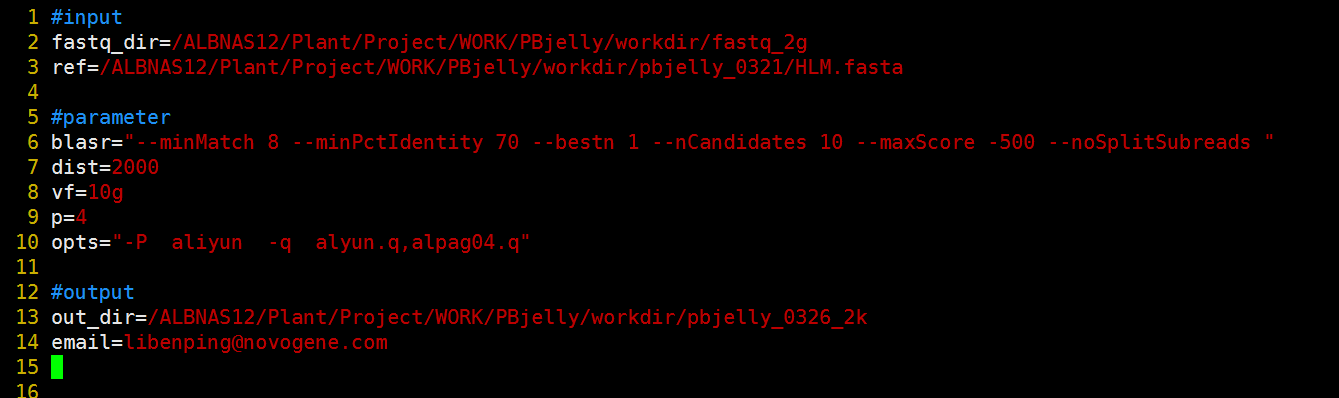
Size:每个fastq文件的大小，单位为G,建议设置为2或者3

Opts：集群队列信息，例如-P aliyun -q alyun.q,alpag01.q

如果已经有fastq文件，fastq文件的大小最好相接近，这样比对所需时间相近，也方便分配内存大小和线程数；可以用fastq\_split.sh脚本将一个目录下的fastq文件分割成大小相同的N份(以vf=1g,p=1投递即可）：

$ path\_to\_bin\_dir/fastq\_split.sh input\_fastq\_dir output\_fastq\_dir N

1. 建立pbjelly输出文件夹, 将软件包目录下的example.cfg文件拷贝到输出文件夹下，修改其内容：



fastq\_dir: fastq文件存放目录。

ref: reference文件路径。

blasr:blasr 参数，一般不需要修改。

dist:minimap2过滤时挑选gap周围多少bp的reads,默认为2000。dist值越小，挑选的reads越少，比对速度相应也越快，但补洞指标可能会因此降低。

vf:每个任务投递的内存，一般设置为10g即可。

p:每个任务投递的线程数。

opts:其他投递参数，例如队列，没有可以不填。

out\_dir:输出路径。

email:任务完成后发送一封邮件到指定邮箱，可以不填(阿里云测试无法发送邮件）。

1. 在输出文件夹下运行：

$ path\_to\_bin\_dir/runjelly.sh example.cfg

**Tips**

1、如果reference>4G,由于sawriter的内存限制，将无法生成.sa文件。请用bin目录下的fasta-splitter.pl将fasta文件切成小于4G的若干份，分别跑runjelly，跑完后将jelly.out.fasta合并。

2、如果需要重新运行程序，在输出目录下运行:

$ sh path\_to\_bin\_dir/cleanup.sh

即可停止运行，并且删除除了\*.cfg以外的所有文件。

如果只想停止而不删除文件，可以在输出目录下运行：

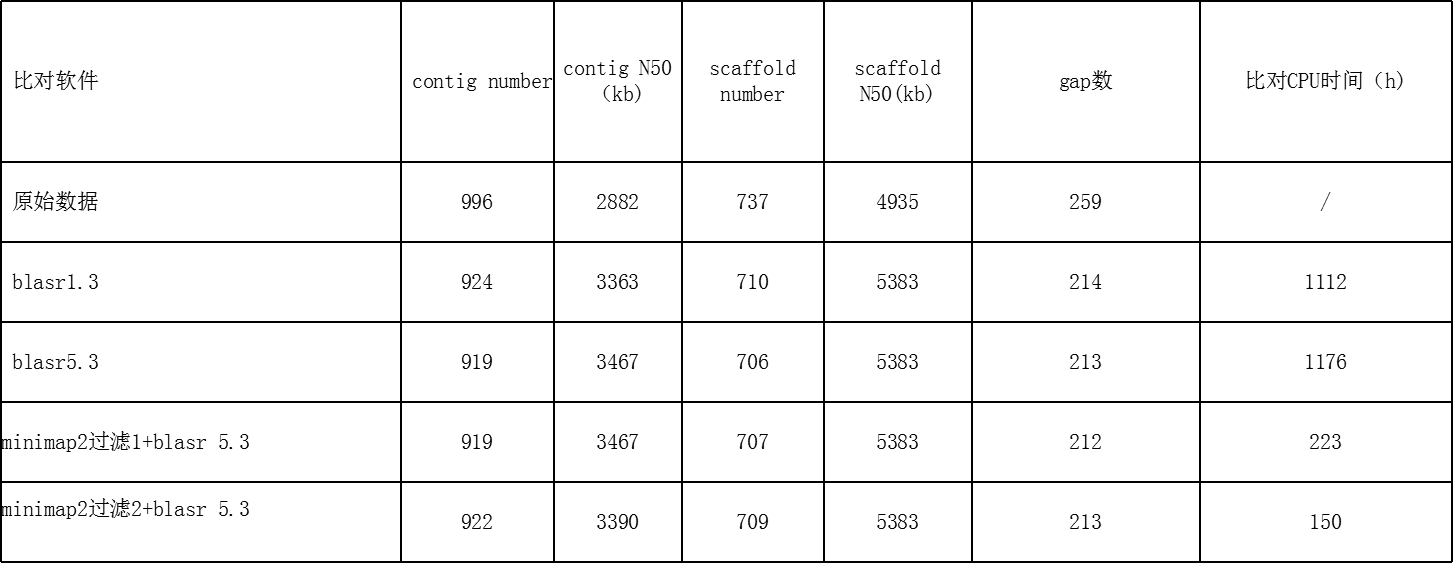
$ sh ./scripts/stop\_jelly.sh

**Contact**

[如有任何问题，请联系zhouyiqi@novogene.com](mailto:如有任何问题，请联系zhouyiqi@novogene.com)

**测试结果**：

山梨(640M基因组，120G fastq):



过滤1：过滤范围2000bp

过滤2: 过滤范围1500bp

黄梁木（700M基因组，174G fastq）

