BSA和遗传进化流程说明文档

1. **BSA流程说明文档**

BSA（Bulked Segrega-on Analysis）：分离体分组混合分析法，在基因组层面对家系材料进行性状定位，将目标性状定位到染色体的一段区间上，并注释该区段基因，找到与目标性状相关联的基因。

* 1. **线下BSA项目基本流程**

将变异检测得到的vcf文件根据不同子代类型进行亲本标记开发，并利用SNP-index滑窗法、SNP-index loess拟合曲线法、ED法、G统计四个方法进行候选区域的定位，将得到的候选区域内的基因型进行GO、KEGG富集分析（图1）。通过该流程得到01.vcftable等6个结果目录（图1）。

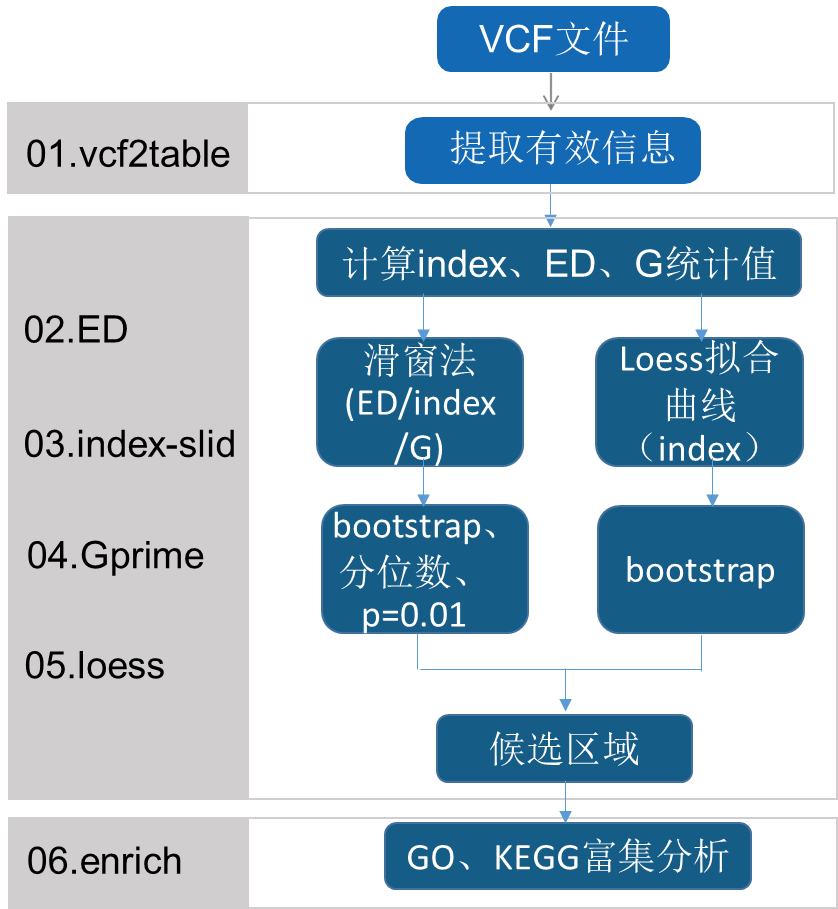


图1 BSA线下基本流程图

* 1. **流程代码说明**
     1. **BSA流程**

**第一步：准备文件**

1. **workflow\_results**

云平台上变异检测结果，下载到工作目录下

1. **group.list**

亲本和子代混池对应文件，其中P1为野生型，P2为突变，B1为野生型性状子代，B2为突变性状子代，P1-B1,P2-B2性状是一致的，需要和老师对接清楚。后两列分别是过滤最大深度和最小深度，其中亲本是10~100之间，子代混池是10~1000之间。在工作目录下新建data文件夹，将该文件放到data文件夹下。

图形用户界面

中度可信度描述已自动生成

1. chr.list

染色体文件

文本

描述已自动生成 图示

中度可信度描述已自动生成

**第二步跑BSA流程：**

在工作目录下运行：

bash /mnt/ilustre/users/jing.wang/script/BSA/run\_BSA.sh

该脚本会生成准备文件go.result、kegg.result、chr.list并放入工作目录下的data文件夹下，然后运行BSA nextflow脚本。运行BSA nextflow使用的是默认参数，可根据结果进行参数调整，如果报错，可以根据报错信息去相应的工作目录重跑报错命令，再返回工作目录，利用-resume 参数重跑BSA nextflow，这样就会在当前断掉的位置继续往前跑，不影响前面跑的结果。

文本

中度可信度描述已自动生成

* + 1. **BSA结果报告和readme生成流程**

**第一步准备文件：**

1. **project.info.xls**

按照以下格式填写相应信息，放入工作目录下的data文件夹中

文本

描述已自动生成

1. **project.txt**

填写合同编号、项目编号、客户姓名、物种名、物种拉丁学名、样本数量，将文件放入工作目录下的data文件夹中



1. **result.info.xls**

按照下列格式填写相应信息，只改动第二列，将文件放入工作目录下的data文件夹中

图形用户界面, 文本

描述已自动生成

1. 工作目录下需要有变异检测结果目录workflow\_results和BSA结果目录result

**第二步：跑结果报告生成脚本**

在工作目录下运行：

bash /mnt/ilustre/users/jing.wang/script/BSA/run\_BSA\_report.sh



最后得到data\_release和report结果目录，其中report是生成report结果，data\_release是需要下载下来给老师的结果，包括BSA 6个结果子目录、report文档以及readme文档。