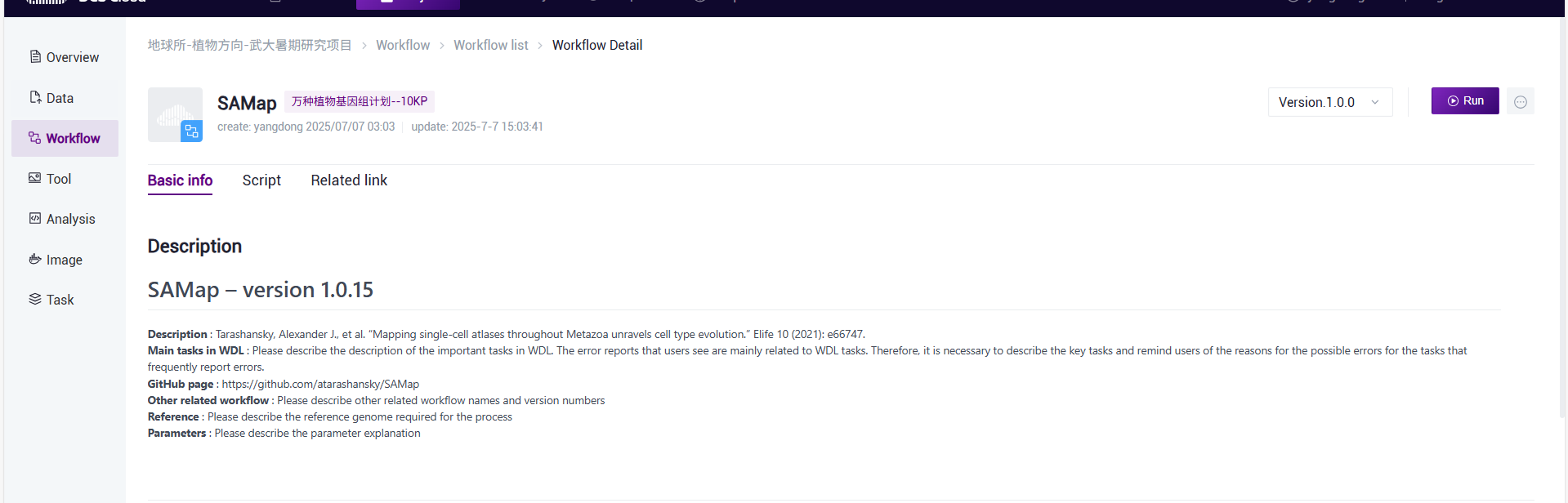
**使用singleR来注释存在的困难？**

用singleR要求参考数据集和待注释的数据集的基因名要有较高的common基因，这种情况是基因命名体系不一致导致的。如果要继续使用singleR就需要统一基因名，做基因的同源比对，用到blast等比对软件。

**使用SAMap做棉花数据注释**

SAMap这个软件和这个流程内置了blast步骤，可以较易解决这个问题



数据准备：

自测的棉花数据的蛋白质文件：<http://magen.whu.edu.cn/downloads/genomeInfo/gosArb/CGP/IGIA/cds.faa>

试试在云平台个性分析用wget命令下载

参考数据集的蛋白质序列：需要找一下文章使用的蛋白质序列，然后做下载

文章相关信息（项目号：PRJNA600131；文章：[Single‐cell RNA‐seq reveals fate determination control of an individual fibre cell initiation in cotton (Gossypium hirsutum) - PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC9674311/) <https://pmc.ncbi.nlm.nih.gov/articles/PMC9674311/）>

参数说明：

rawh5ads：待注释数据的h5ad文件，已被注释数据文件的h5ad文件(保守的注释信息储存键为celltype)

proteins: 待注释数据物种的蛋白质fasta文件，已被注释数据物种文件的fasta文件

注意：蛋白质序列里面的序列名一定要和单细胞数据的基因名对应，存在后缀等不一致情况需要做处理

names：两个字符，且首字母一定要大写，即一大一小结构

nonrawlst：提供一个文本文件，里面随意有几个字符即可

sx: true

main\_mem: 使用的内存

例子：

