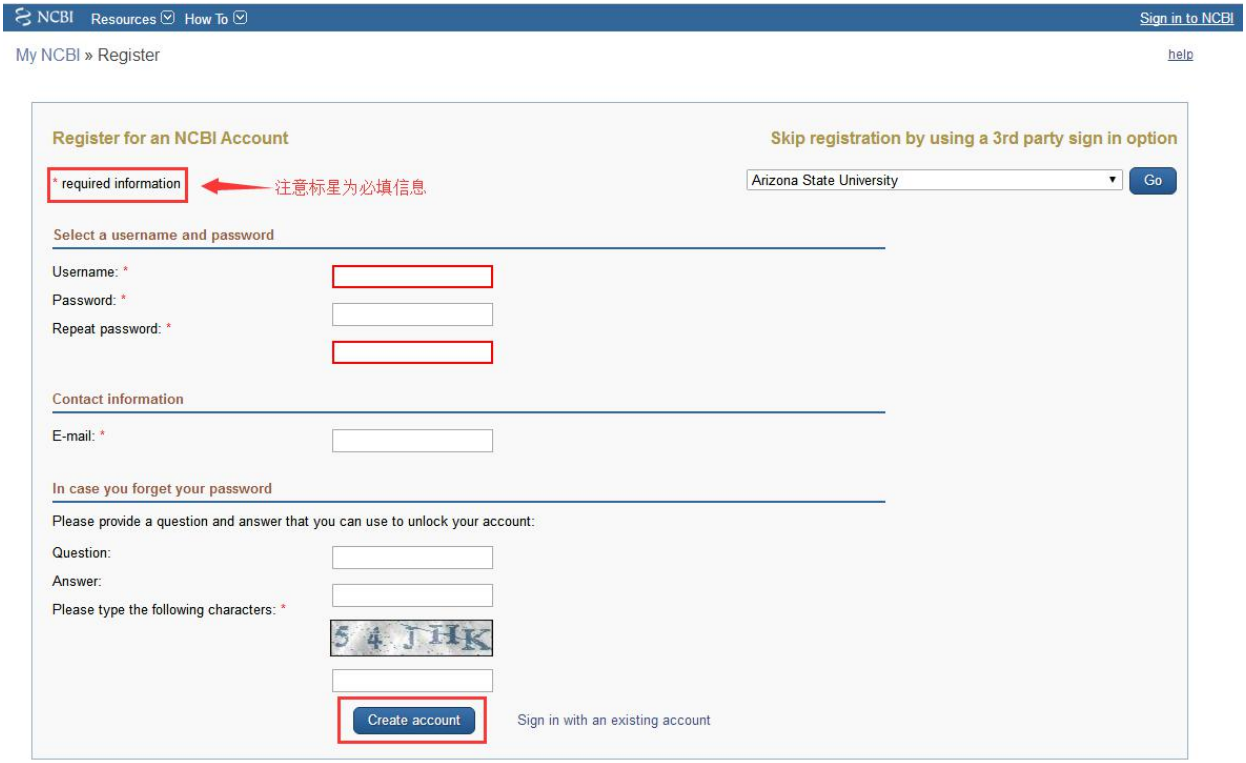
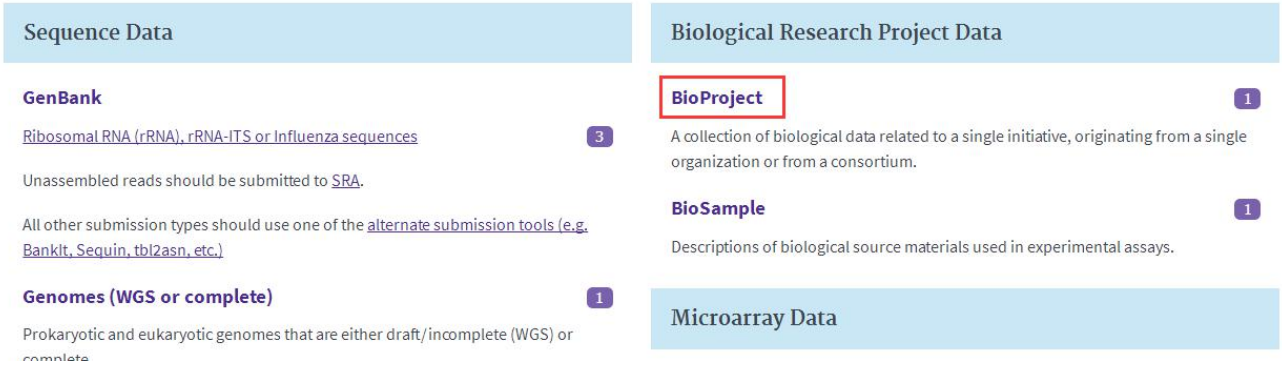
# 微生物基因组数据上传指南

## BioProject ID 的获得

1、在 NCBI 主页右上角（https://www.ncbi.nlm.nih.gov/）My NCBI 登录系统中创建新的账号（已有就不需要），点击 Register for an account 创建账号：

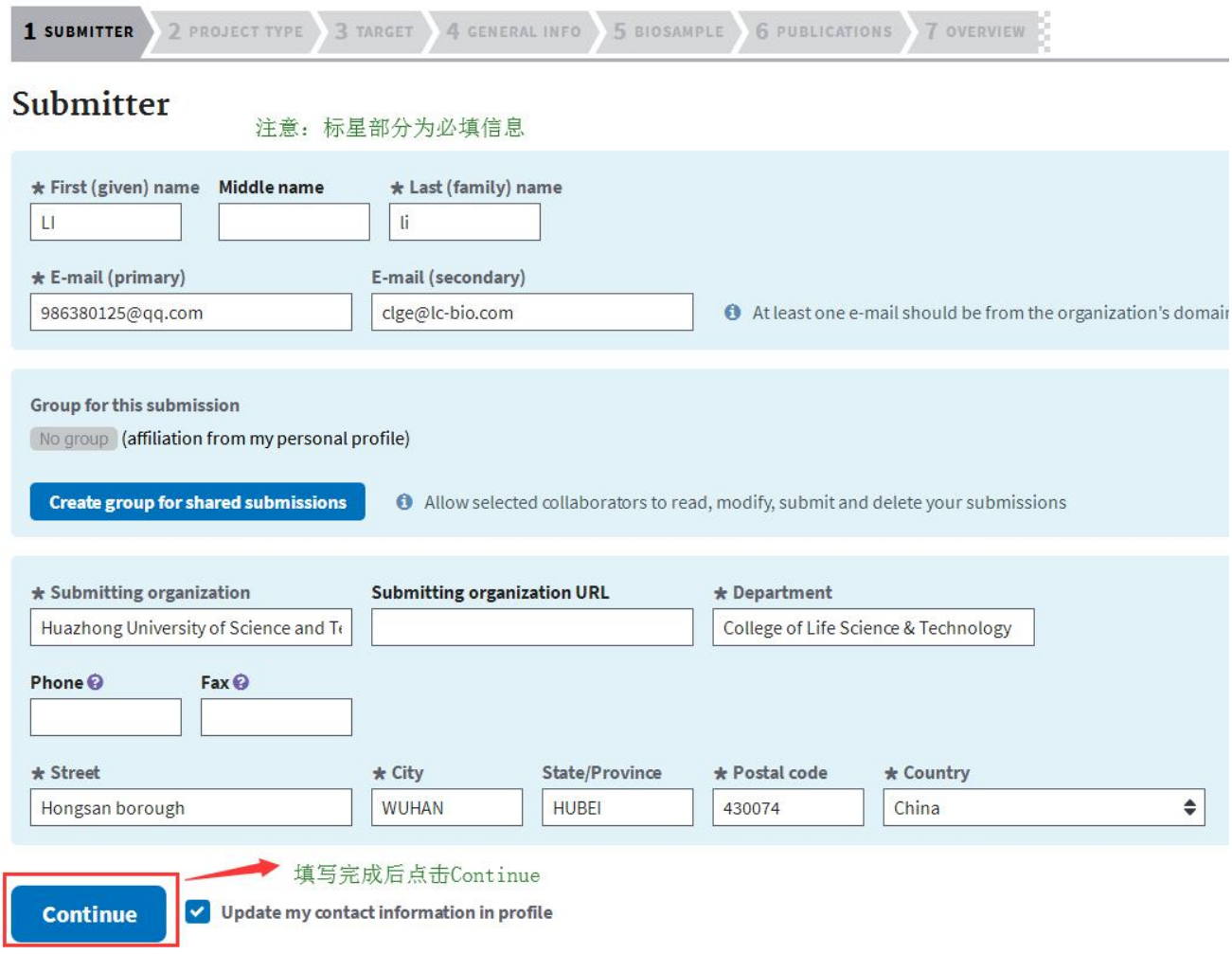



2、登录 BioProject（https://submit.ncbi.nlm.nih.gov/）,获取一个 BioProject ID

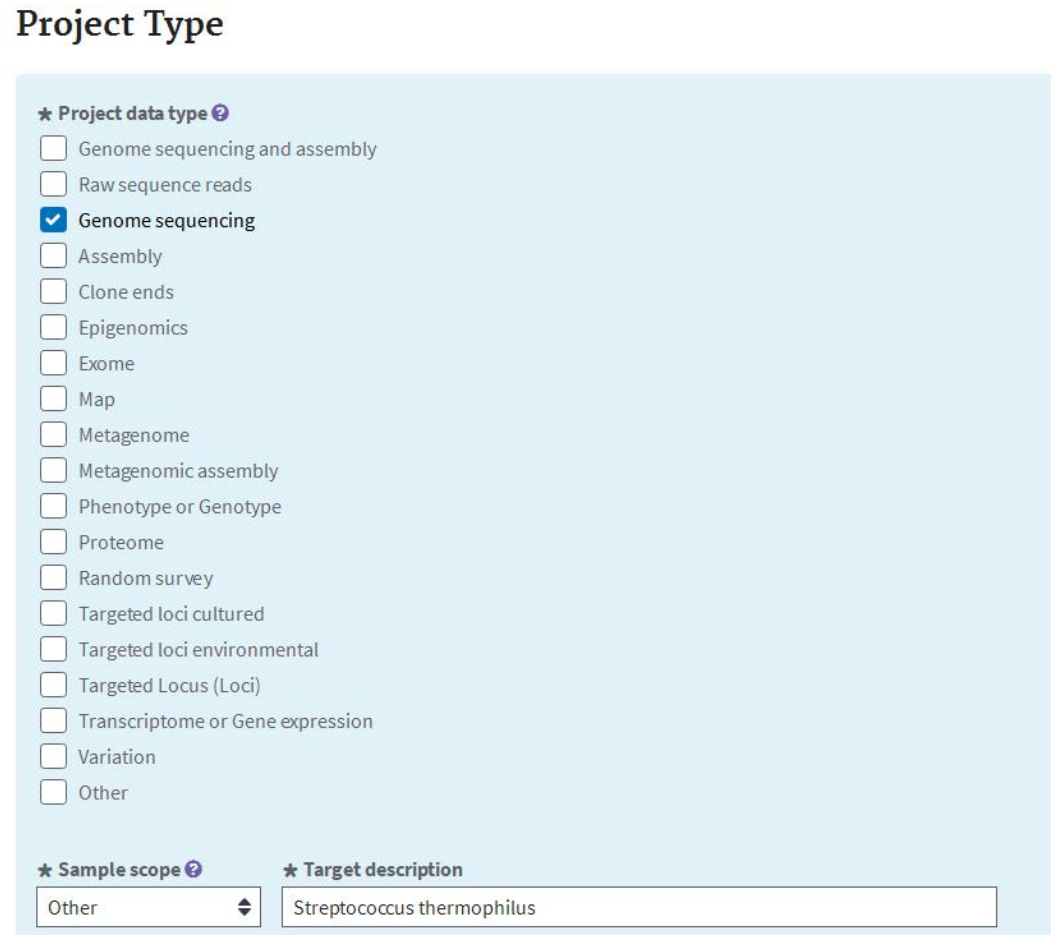


3、点击 New submission，进行提交

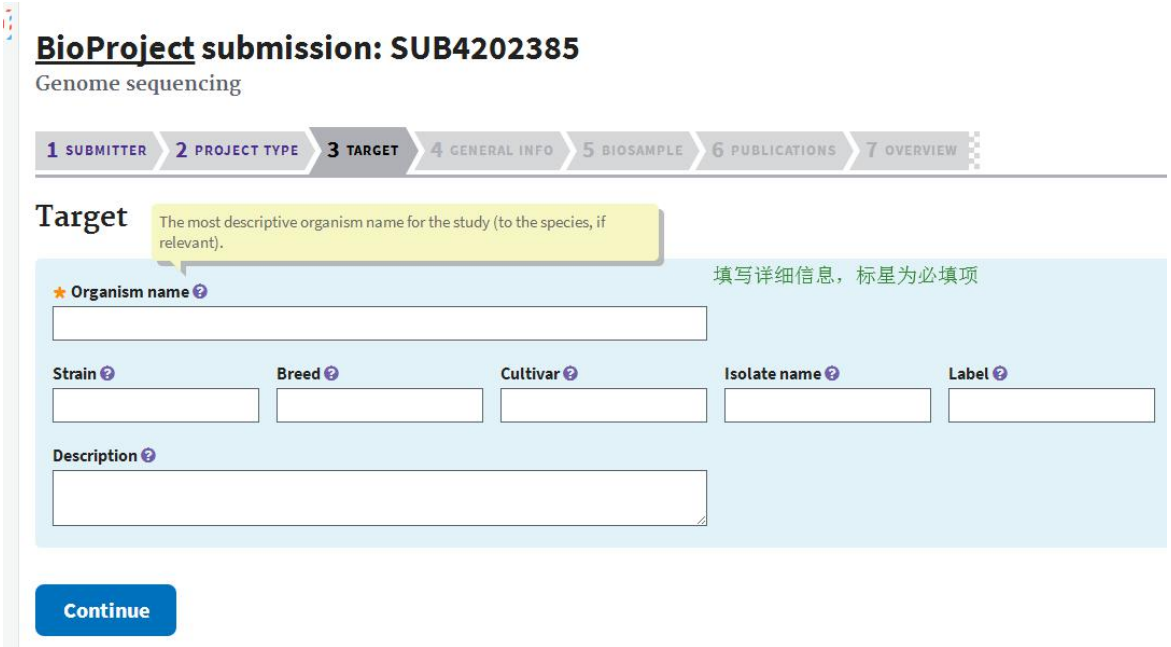

4、填写信息（该步骤需要您填写的邮箱进行验证）




5、类型填写（根据项目的实际情况进行填写）

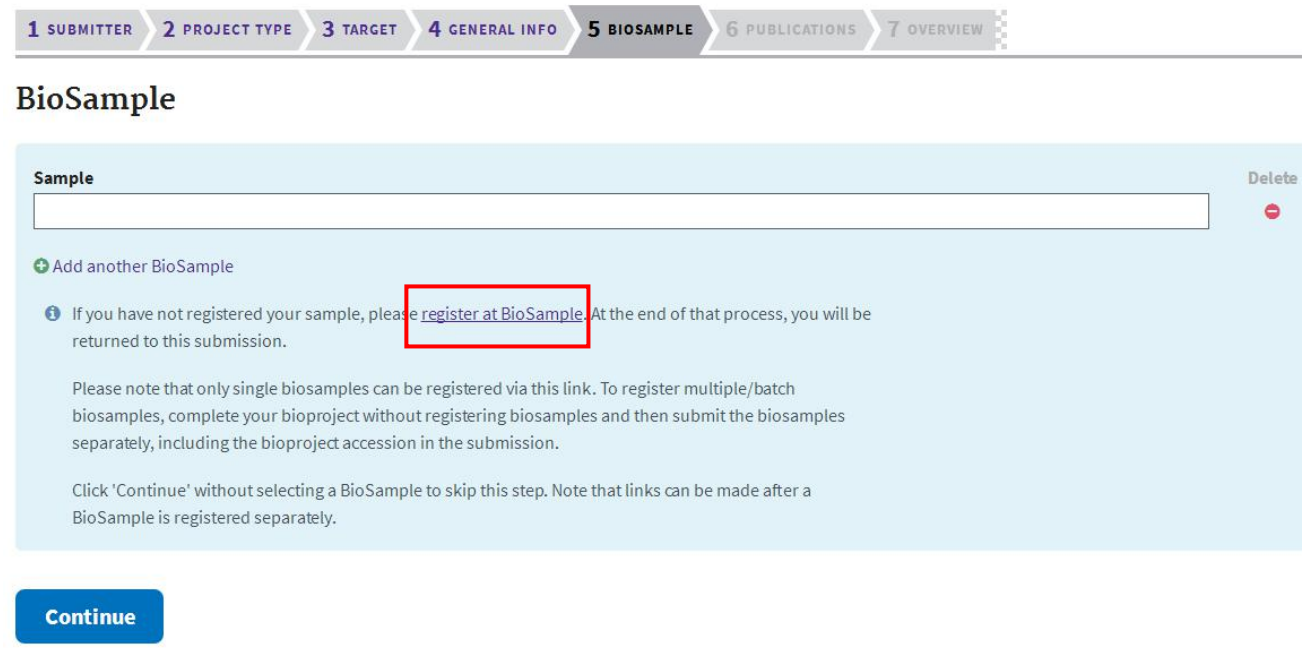



6、TARGET 填写

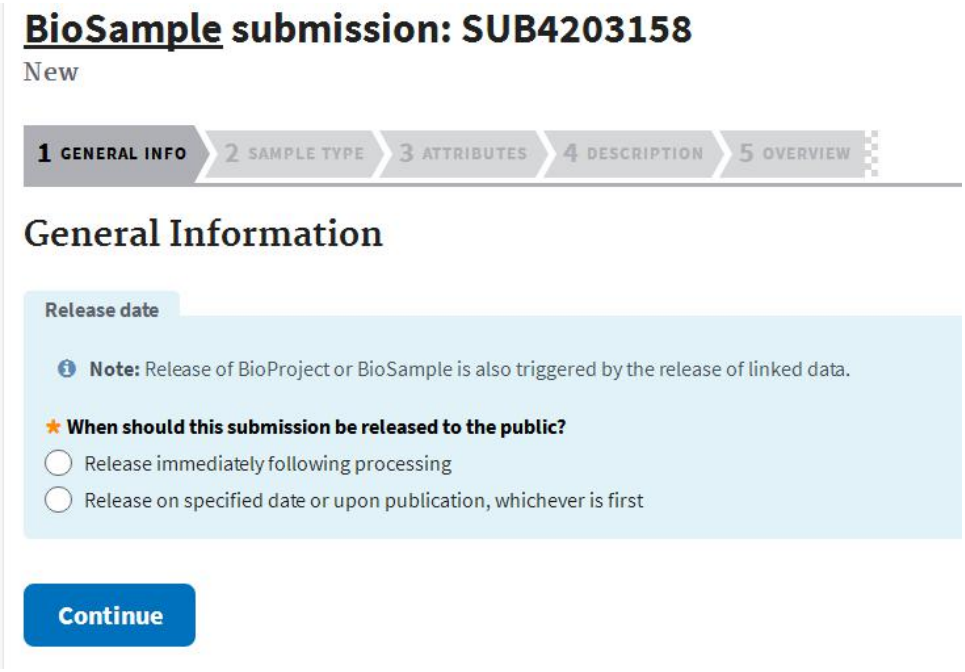



7、General Info 信息填写

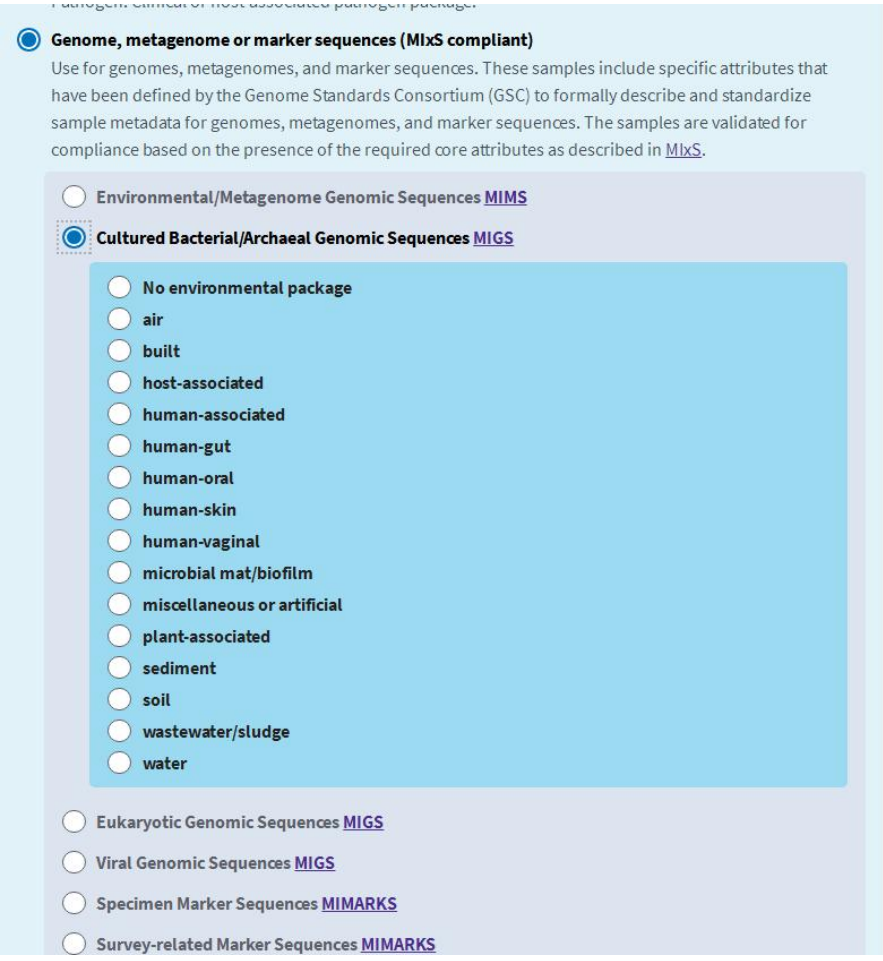

8、BIOSAMPLE 信息填写（点击 register at BioSample 进行详细信息填写）



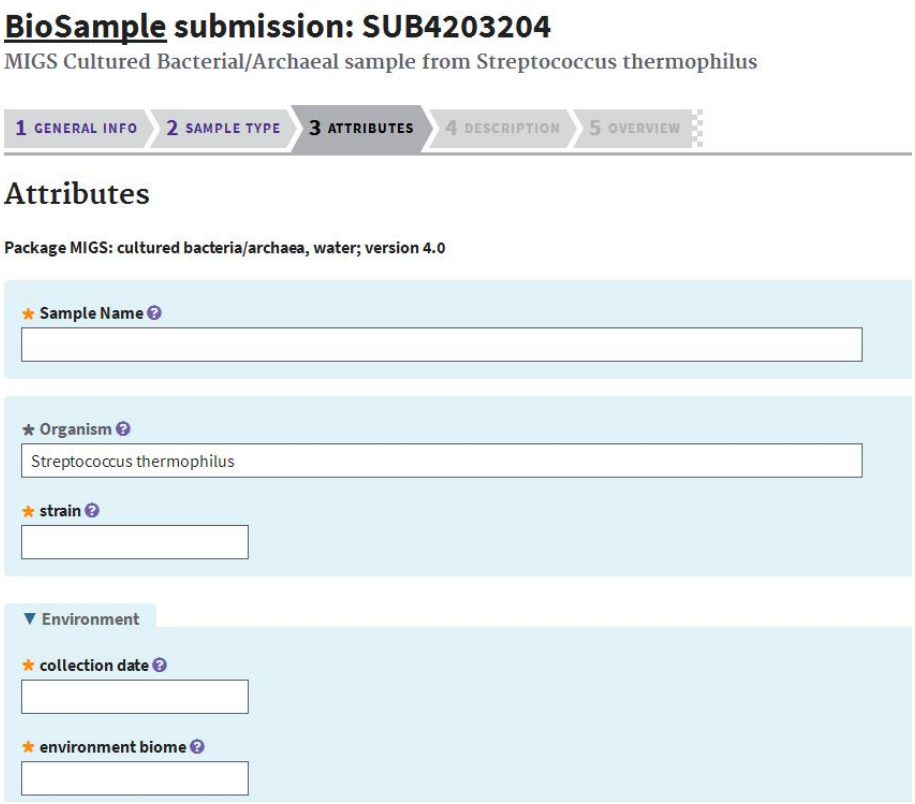

1）释放时间



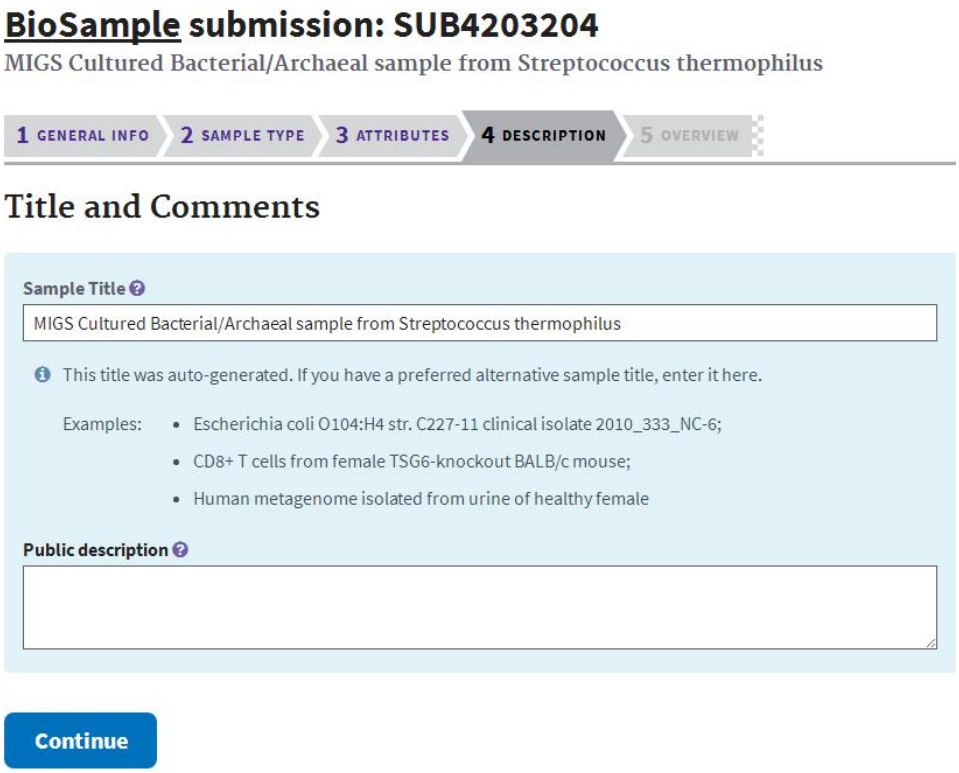

2）样本类型（根据样本来源情况进行填写）



3）属性填写（根据实际情况进行信息添加）

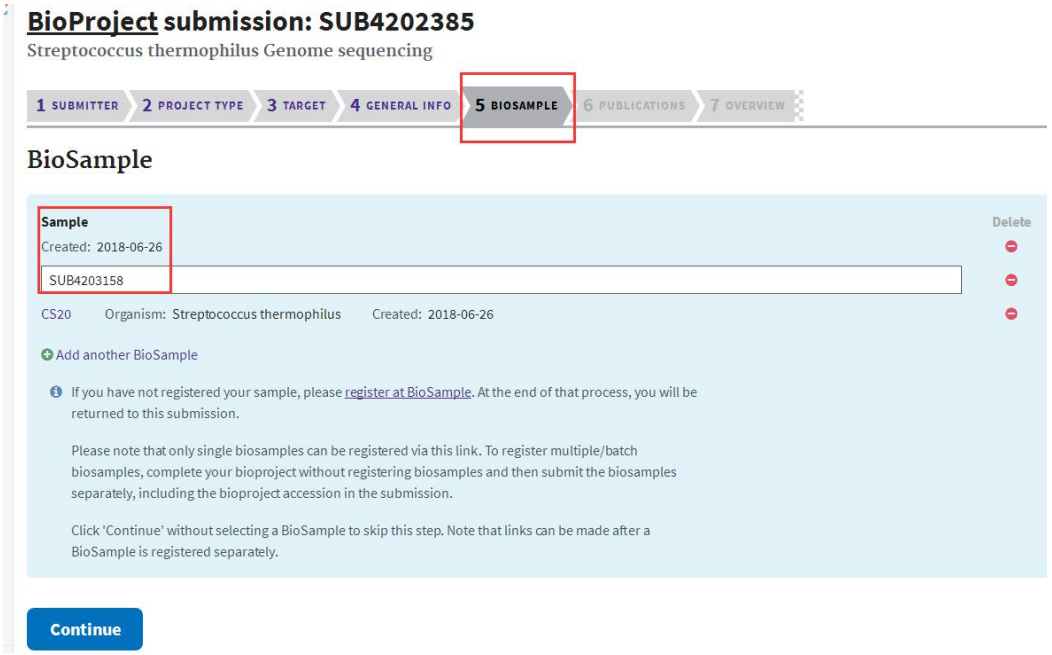



4）描述信息添加

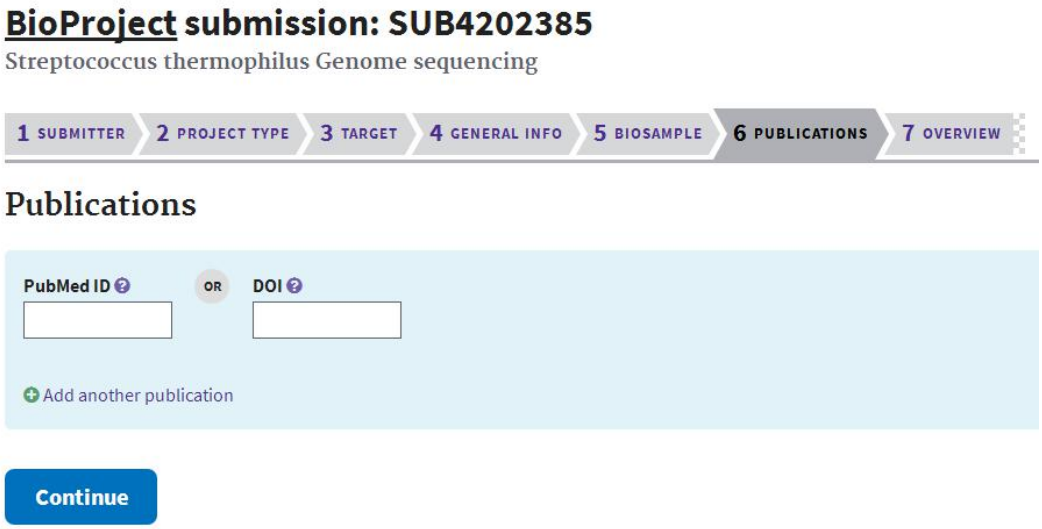


5）信息确认，确认无误后点击 submit

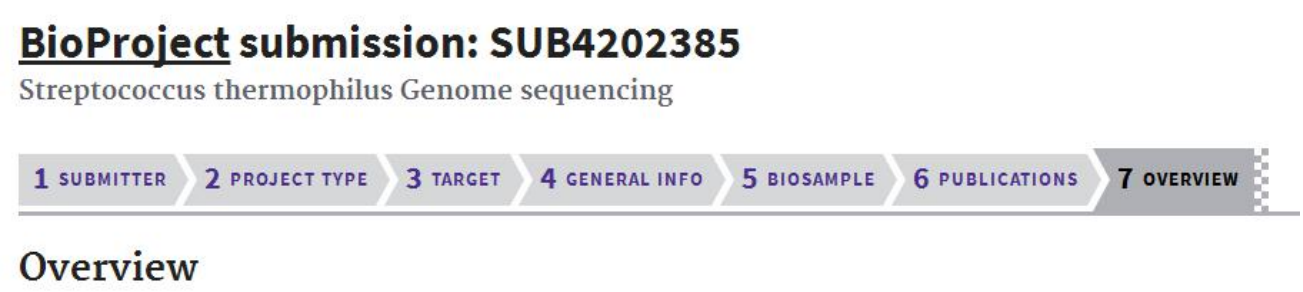

**点击 SUbmit 后会自动跳回至 BioProject**




9、发表杂志的 PubMed ID 或 DOI 信息填写，若无可不填写

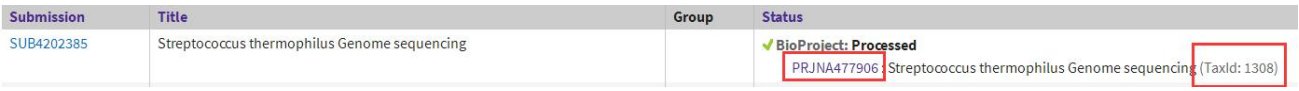


10、信息确认，确认无误后点击 submit



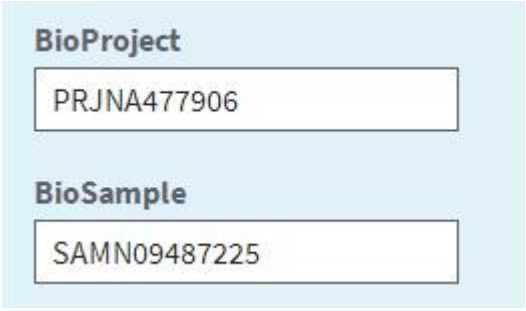

填写完成后，约几分钟的时间就收到 NCBI 审核的邮件，刷新上传的界面，状态从 Processing 变成 Processed

11、获得的 BioProjcet ID，以 PRJNA 字符为前缀，并且获得通过自动分配方式获得项目唯一的 Locus Tag Prefix 值，比如下面例子的 1306，该值用于注释结果 locus_tag 这一项的前缀（**注意：注释文件必须含有该值！**）如下图红框所示：




**组装结果的准备（该部分步骤先跳过，先进行序列的提交部分的上传步骤，如后续上传有问题再进行该步骤）**

1、生成 template（*.sbt）文件，填写 submission template form（http://www.ncbi.nlm.nih.gov/WebSub/template.cgi）。 template form 如下图所示（需要填写提交人的各项信息，姓名，地址，单位，联系方式等等，还有文章的题目及签名获得 BioProject ID）：




填写完成后，点击最下方的 Create Template，生成 sbt 文件，备用；

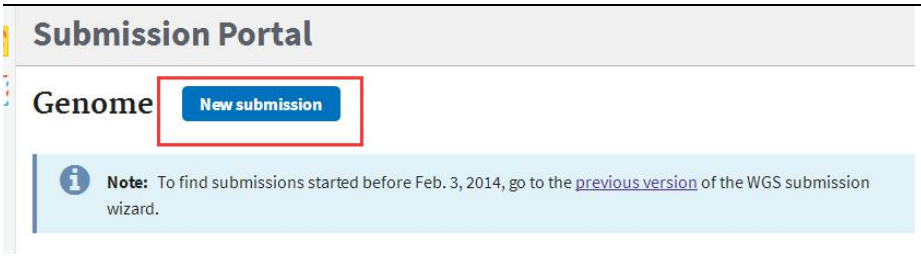

2、准备基因组文件，要求基因组序列文件，无 gap，即序列中不含 N，每个文件不超过 10，000 条序列。

格式：该文件为标准的 fasta 格式，第一行是描述信息，以“>”开头；第二行起始序列信息，每行长度不超过 80 个字符。如下图所示：

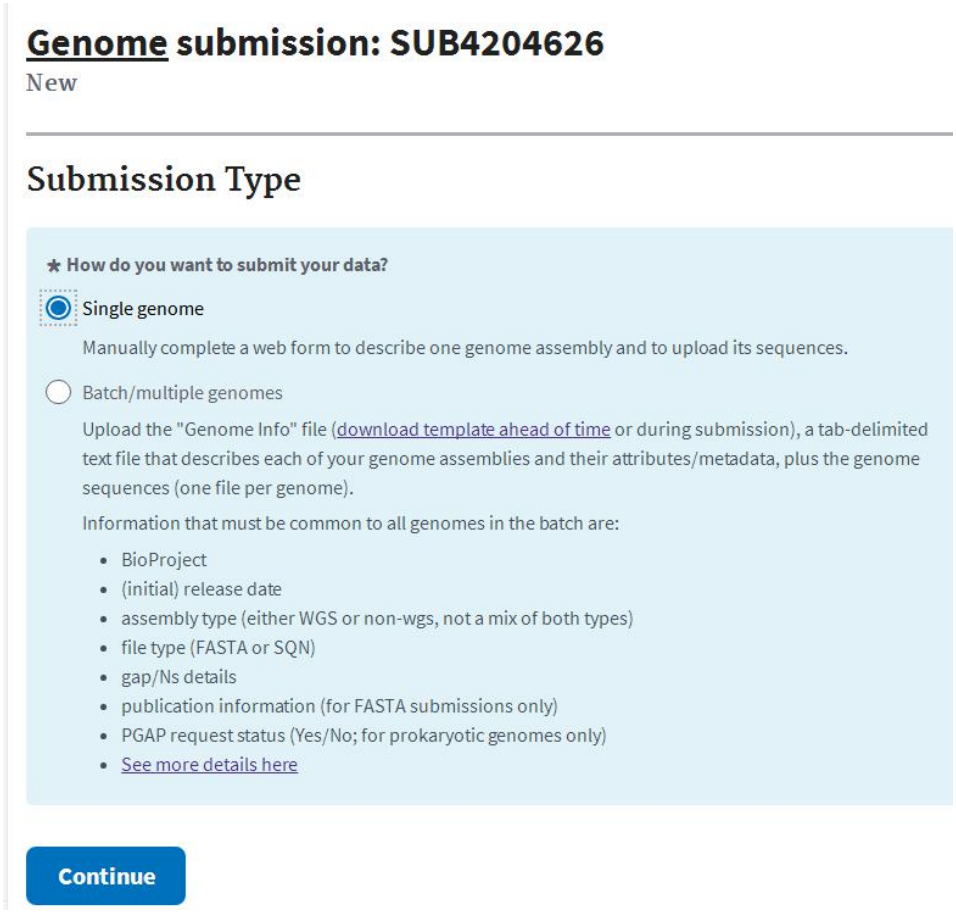

## 序列的提交

1.上传生成得到的*.sqn 文件( 具体生成步骤详见最后一部分**利用 tbl2asn 软件生成符合 NCBI 上传规则的 *.sqn 文件**）或者 FASTA 文件，网址如下：https://submit.ncbi.nlm.nih.gov/subs/wgs/

2.按照上传引导填写相应信息，上传文件，整个过程需要使用一致的物种名称。

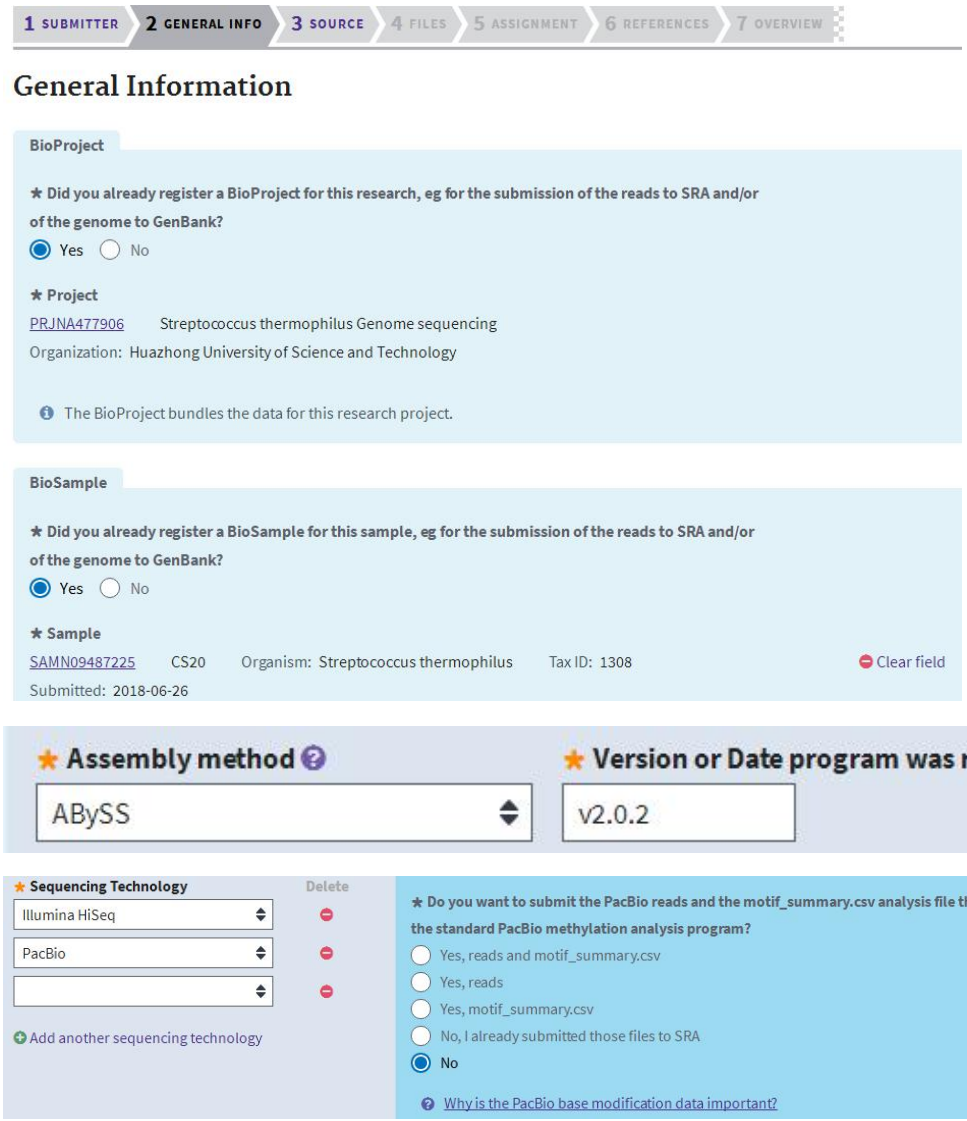



基于序列类型进行选择上传




剩下的步骤按照跳转提示一步一步进行填写即可（**这里部分重复的步骤省略**）；

1）其中信息部分，前面已经生成过 BioProject 和 BioSample 填写相应的 ID 即可，其中 BioSample 为 SAMN 开头的信息；

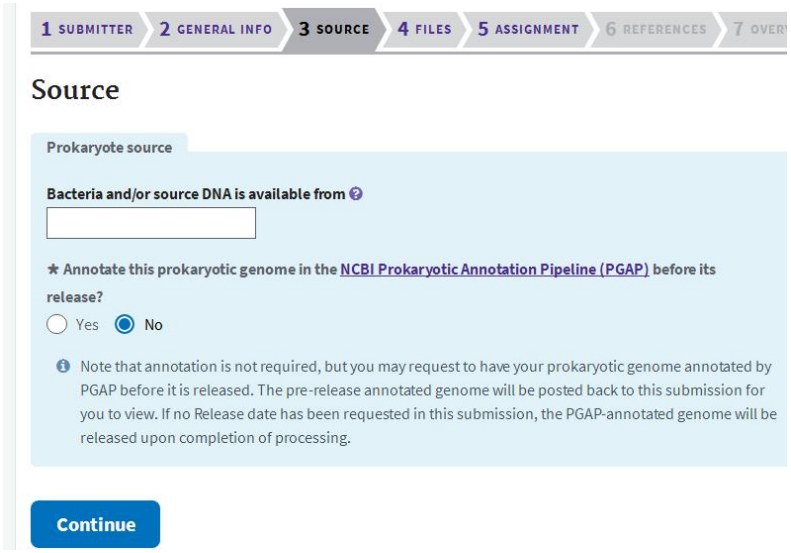




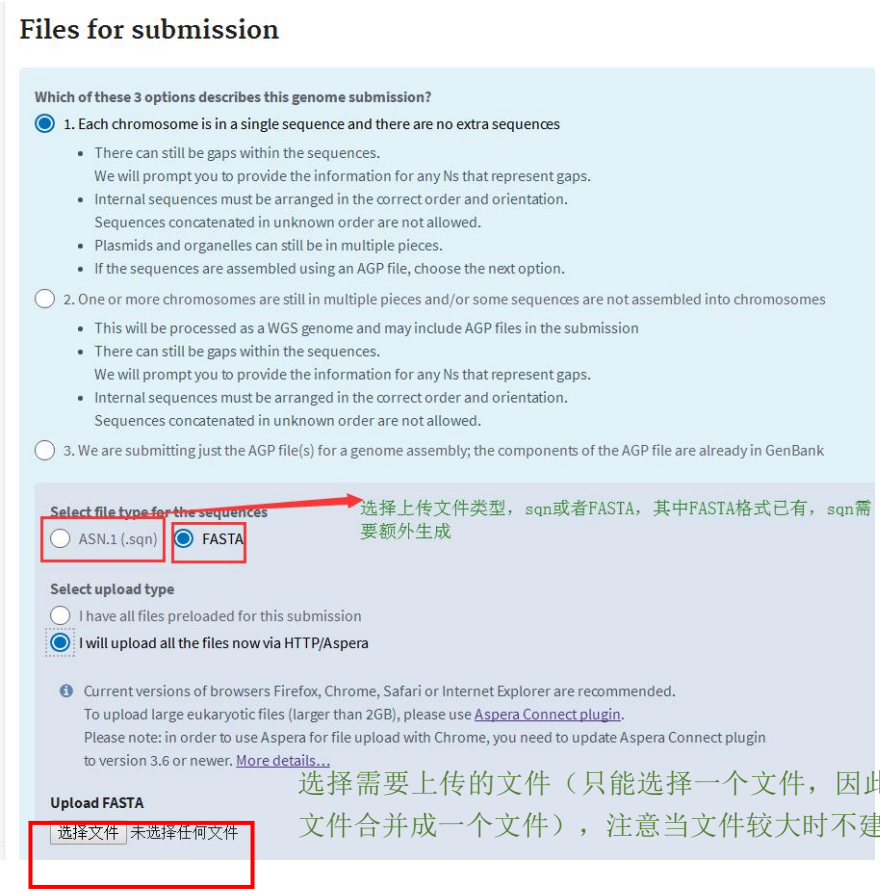




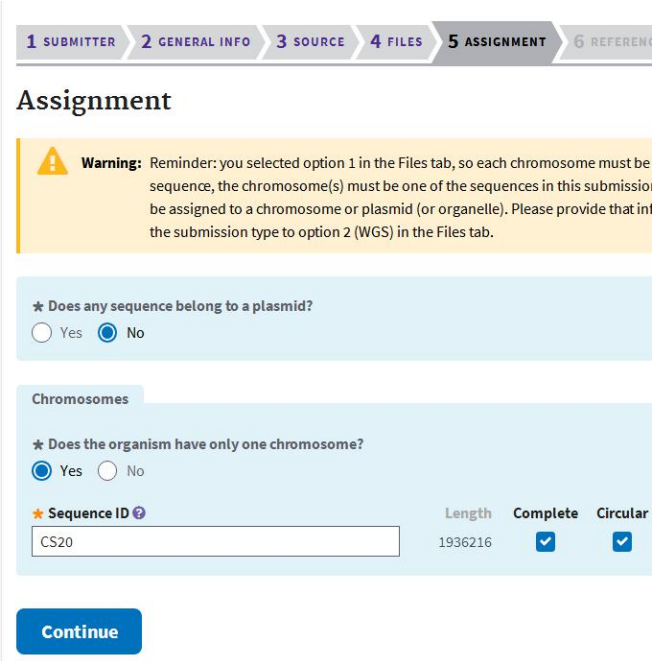

2）Source 部分选择 No



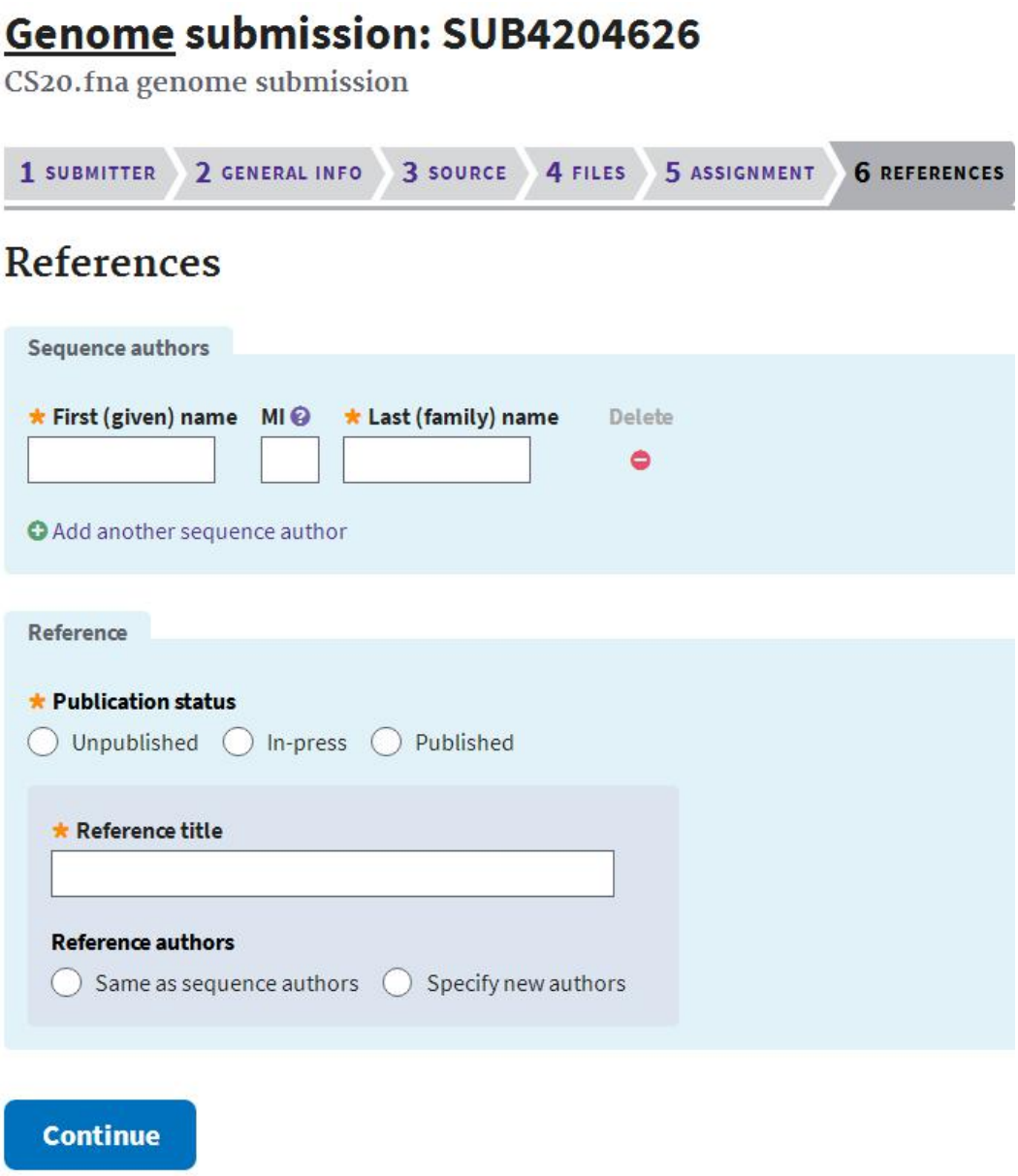
3）数据上传类型选择（如果选择.sqn 格式的文件需要按照**利用 tbl2asn 软件生成符合 NCBI 上传规则的 *.sqn 文件**步骤生成该格式文件，选择 fasta 格式直接上传即可）




4）该部分根据样本实际情况进行填写（是否含有质粒，是否完整基因组是否为环状等等）



5）文献情况，请根据实际情况填写




6）信息确认，确认无误后点击 submit

上传完成后，需要等 NCBI 的审核，审核完成后会邮件通知上传者。

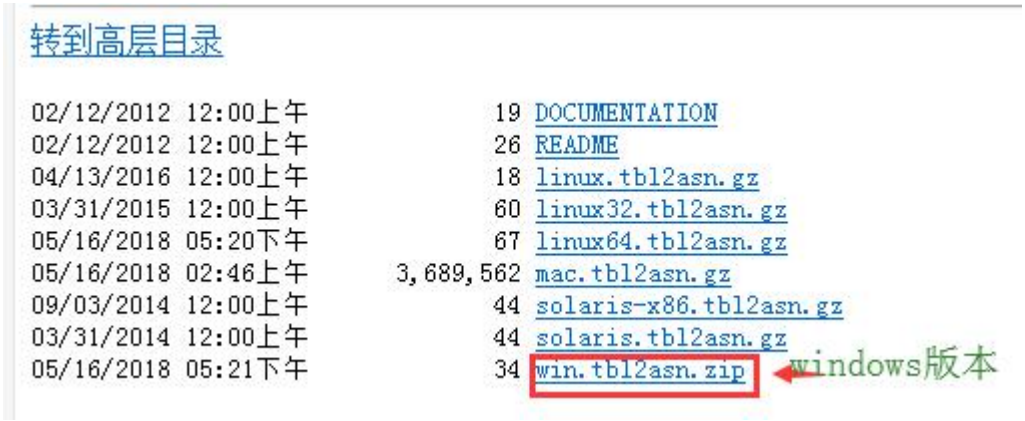
**最后简单介绍下*.sqn 文件生成的步骤，如有需要请使用：**

**利用 tbl2asn 软件生成符合 NCBI 上传规则的*.sqn 文件**

1.准备生成*.sqn 文件。该文件需要两个文件：1）前面生成的*.sbt；2）基因组序列文件--即 03.Assmebly 文件夹下的 fna 文件；

2.下载 tbl2asn 软件，下载地址如下：

ftp://ftp.ncbi.nih.gov/toolbox/ncbi_tools/converters/by_program/tbl2asn




该软件说明如下：

http://www.ncbi.nlm.nih.gov/genbank/tbl2asn2.html

3.将以上两个文件置于 tbl2asn 软件目录下，进行命令提示行命令（开始-cmd），进入 tbl2asn 软件目录，输入如下命令后回车运行（注意空格和“”号）：

tbl2asn.exe -i *.fna -t *.sbt -a s -V v -Z log -j “[organism=*][strain=*]”

[organism=*][strain=*]:*部分内容需要自行添加；顺利运行后，tbl2asn 将出现由三个后缀名的文件 *.sqn,*.val,log。*.sqn 文件用于最后的提交作业；一般来说，*.val 文件大小为 0k 则整个转换过程无问题。

4.检查输出的*.val 文件和报告文件。查看*.val 文件内是否提示错误信息，如果有，找出并解决，以减少上传审核时间。