



数据完整性检验

为了确保客户下载的文件与我们释放的一致，在数据下载到本地之后，需要做一个完整性检验，确保下载到本地的文件与我们提供的是完全一致的。

我们这里提供一种 MD5 文件校验的方法，因文件经过网络传输、拷贝或其他操作后，可以通过文件的 md5 值判断文件内容是否发生了改变，文件名改变、但文件内容不变时，文件的 md5 值会保持不变，从而判断出数据是否在传输过程中缺失。md5 是校验文件内容，与文件名是否相同无关；md5 是逐位校验，所以文件越大，校验时间越长。

注意：下载数据之后不要挪动这两个文件的位置，保持与其它文件的相对位置不变。

操作系统不同使用方法也不同：

数据校验操作视频教程：<https://edu.majorbio.com/lesson/4723>

Windows 用户

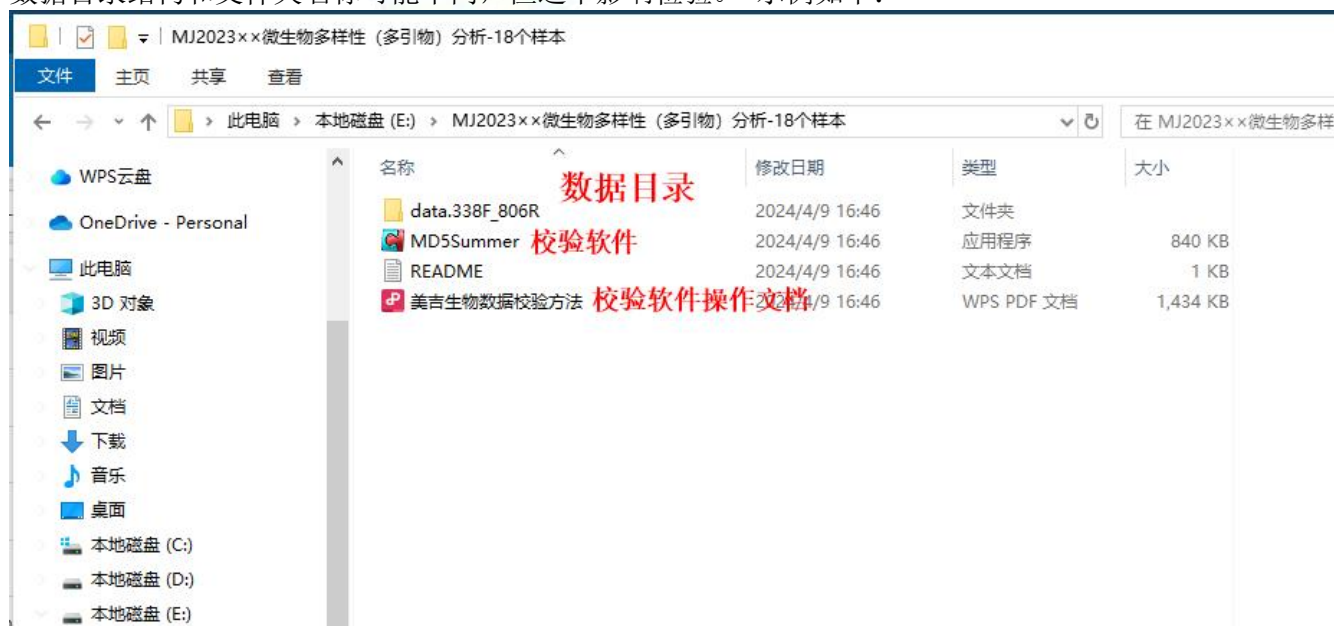
方法一：

这里我们提供了需要校验的数据文件以及对应的 MD5 校验值文件（md5.*.txt 文件）以及校验小程序（md5summer.exe）。

检验步骤如下：

第一步：打开下载数据目录

打开下载数据的文件夹，其中有引物对应数据的文件夹和校验小程序（md5summer.exe），不同产品的数据目录结构和文件夹名称可能不同，但这不影响检验。示例如下：



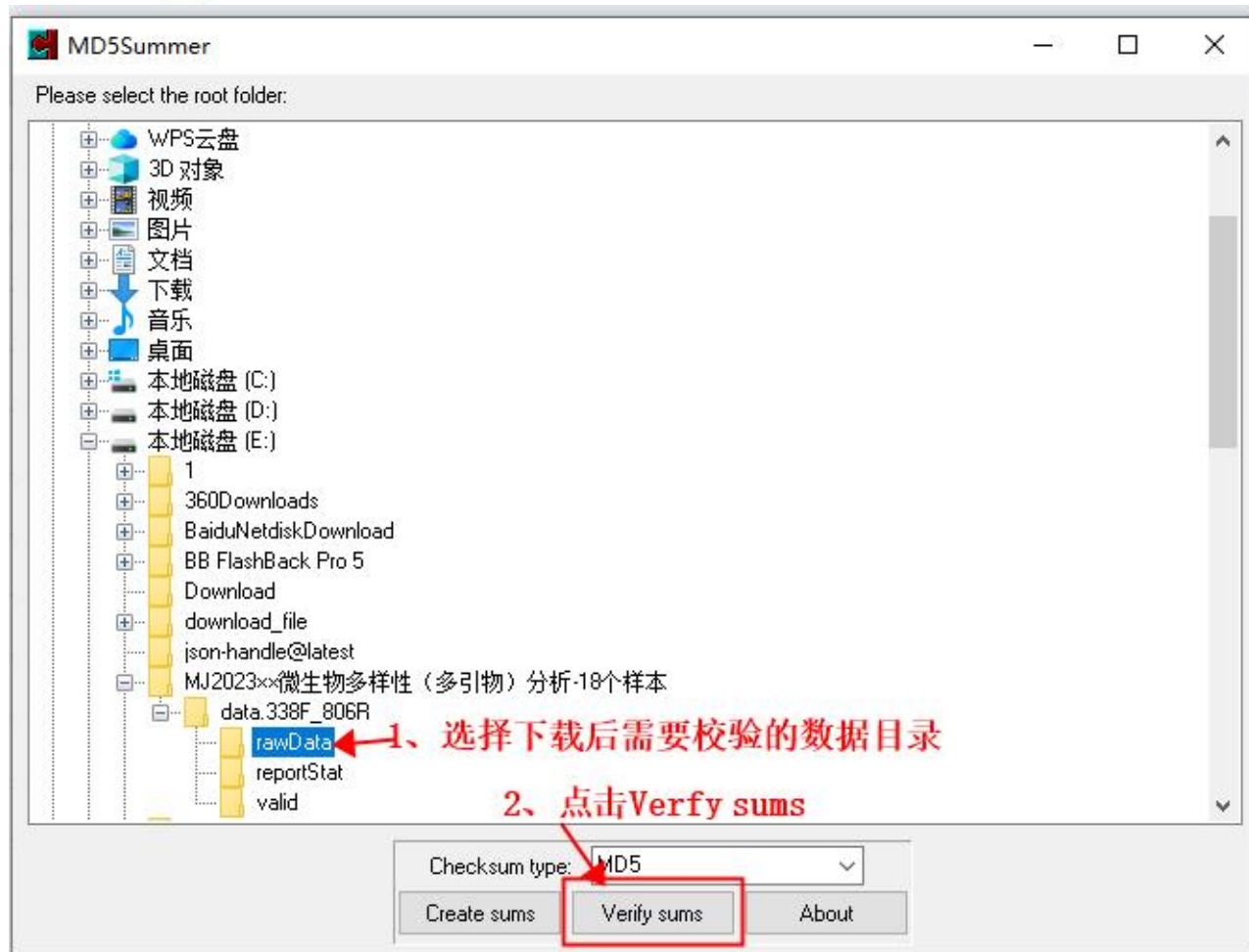
第二步：打开“md5summer.exe”

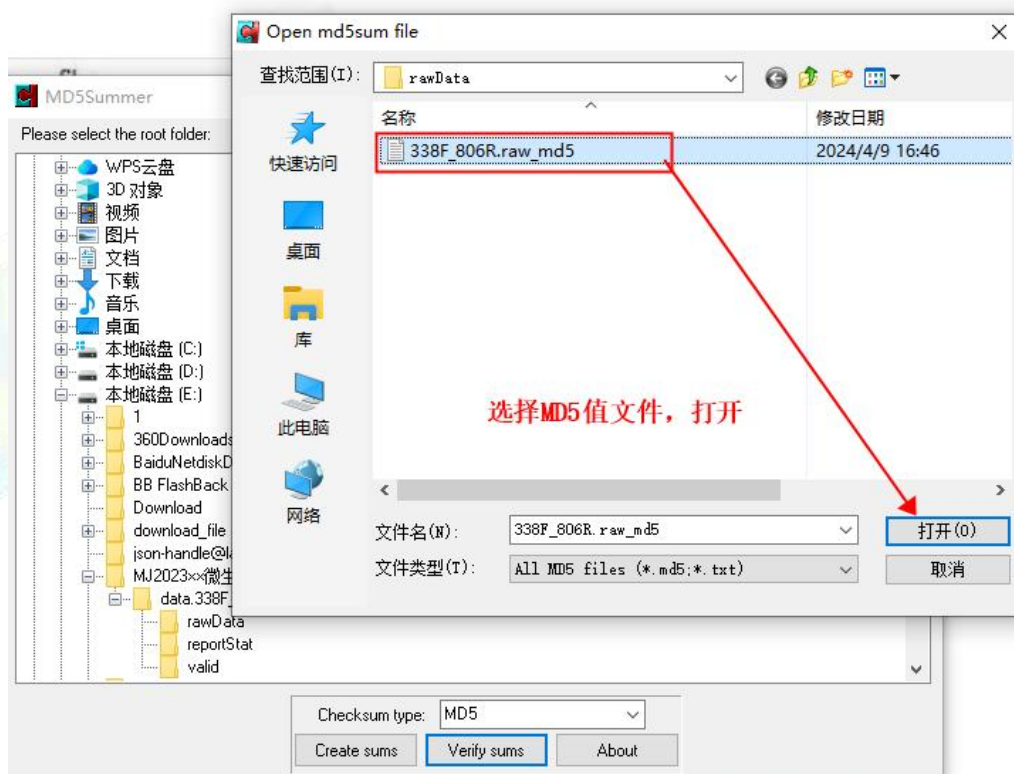
如果出现如下对话框，点击ok 即可，不影响使用。



第三步：左键点击 Verify sums

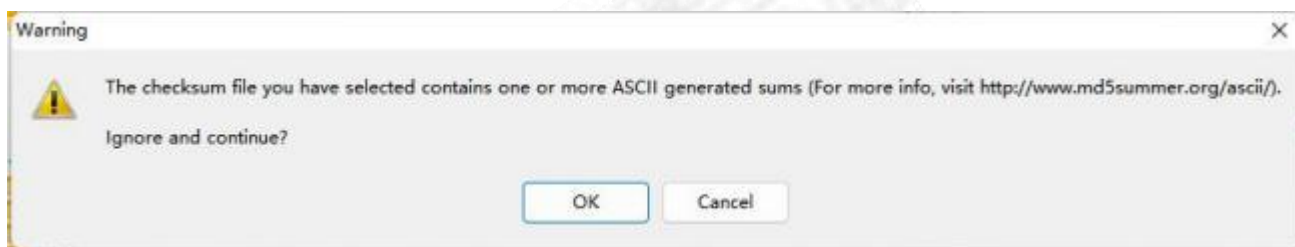
选择需要校验的数据目录，点击 Verify sums, 在弹出的对话框中选择对应的 md5. txt 文件打开



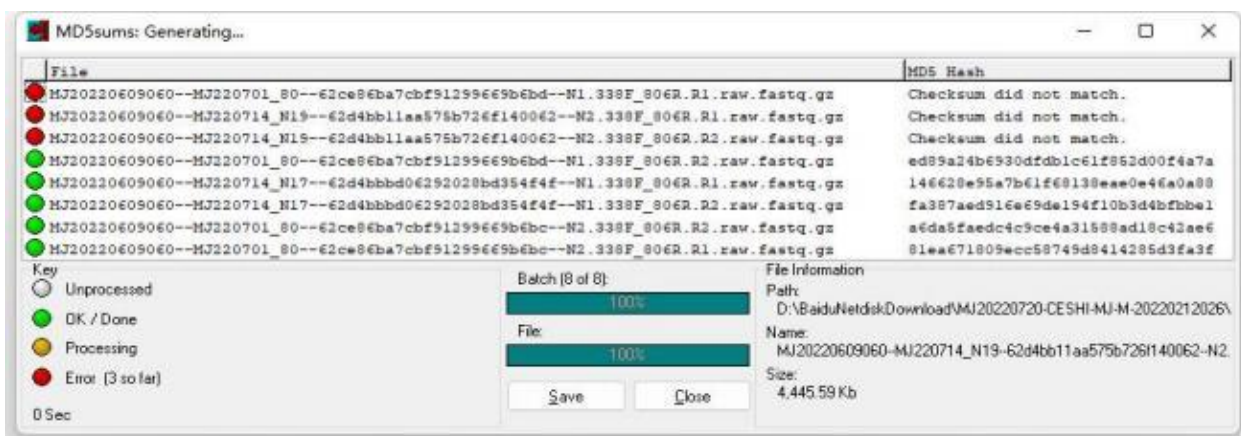


第四步：执行校验操作

如果点击打开出现此对话框，点击 ok 即可：



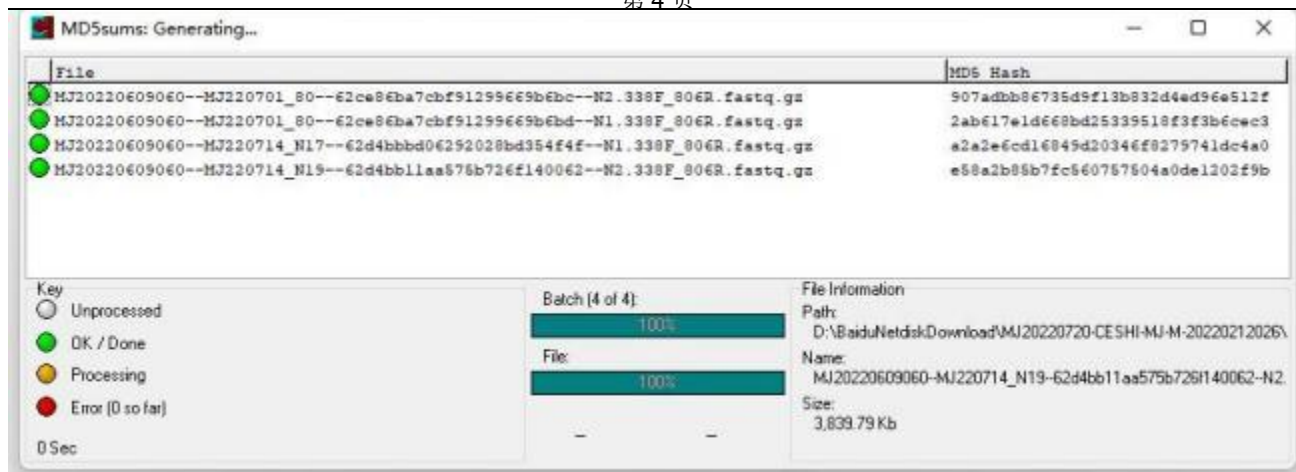
这时候需要耐心等待，所需要的时间和文件大小、所使用的计算机性能等有关系。如下图所示，文件名前边的 Key 的颜色表示所处状态，灰色表示等待处理，棕色表示处理中，绿色表示校验通过，而红色则表示没有通过校验。



第五步：校验结果检查

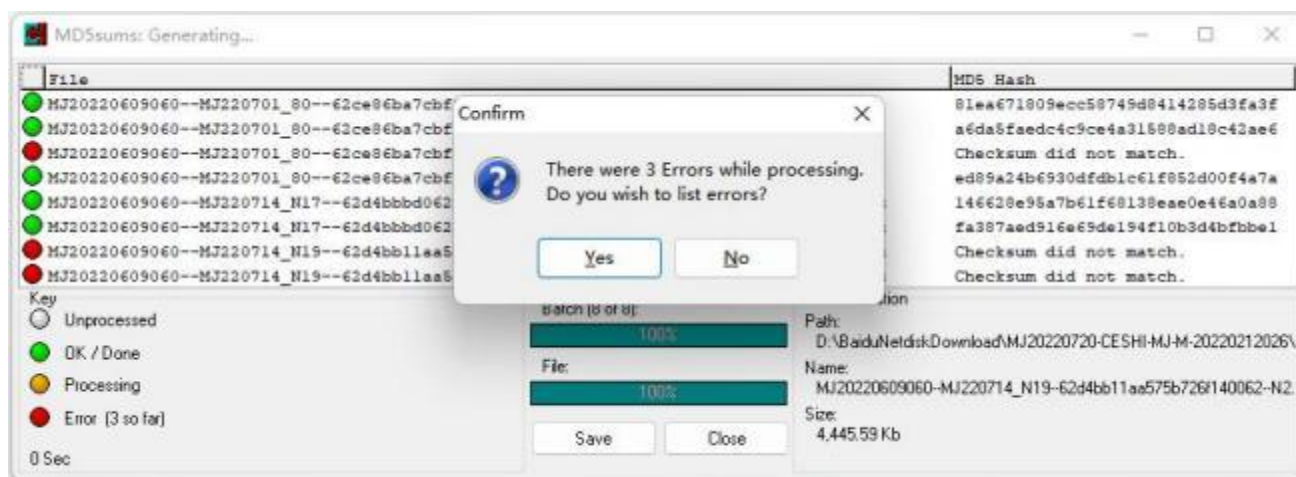
1、匹配结果示例

绿色表示文件 MD5 校验匹配，error 为 0，下载到本地的文件与我们提供的完全一致，下载过程未出现错误。



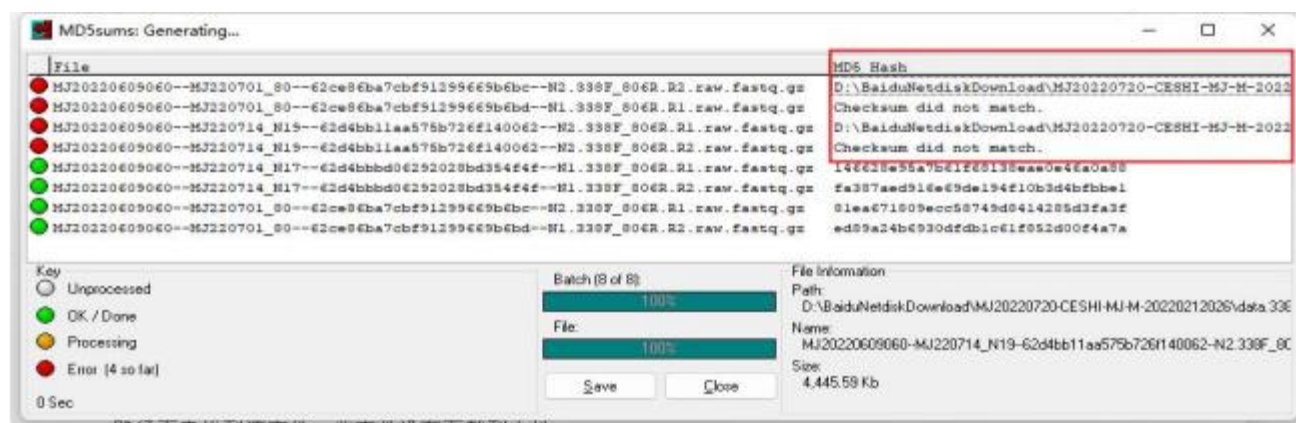
2、不匹配结果示例

红色表示文件 MD5 校验出错。在校验结束后有提示是否将错误文件列出，若点击 Yes 则所有的不匹配文件会置顶。



MD5 Hash 列错误信息若为“Checksum did not match.”则表示此文件与我们提供的不完全一致，可能是传输过程出错导致的；

若错误信息为文件路径，例如下图中“D:\BaiduNetdiskDownload\MJ20220720-CESHI-MJ-M-20220212026\data.338F_806R\rawData”则表示在径下未找到该文件，此文件没有下载到本地



如果出现上面的情况，则这些出错的文件需要重新下载，下载之后再进行验证。这次只需要验证重新下载的文件，故可以将 md5sum.txt 文件中的未通过验证的文件拷贝到新建文件中，然后用此新文件重新进行前面的步骤，也可以将 md5sum.txt 文件中的已通过验证的文件删掉，然后重复操作前面步骤。

如果再三下载之后都不能通过校验，请及时与我司项目运营人员联系。

Linux 用户

第一步：进入下载的数据目录

cd 进入需要校验数据的目录

```
cd MJ20220720-CESHI-MJ-M-20220212026/data.338F_806R/rawData/
```

第二部：md5sum -c md5*.txt 命令验证

md5sum -c md5*.txt

```
$ md5sum -c md5.rawData.txt
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bc--N2.338F_806R.R1.raw.fastq.gz: 失败
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bd--N1.338F_806R.R1.raw.fastq.gz: 失败
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bd--N1.338F_806R.R2.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N17--62d4bbbd06292028bd354f4f--N1.338F_806R.R1.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N17--62d4bbbd06292028bd354f4f--N1.338F_806R.R2.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N19--62d4bb11aa575b726f140062--N2.338F_806R.R2.raw.fastq.gz: 成功
md5sum: 警告: 2 个校验和不匹配
```

第三步：校验结果检查

校验结果如下图所示：

```
$ md5sum -c md5.rawData.txt
md5sum: MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bc--N2.338F_806R.R1.raw.fastq.gz: 没有那个文件或目录
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bd--N2.338F_806R.R1.raw.fastq.gz: 打开或读取失败
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bd--N1.338F_806R.R1.raw.fastq.gz: 失败
MJ20220609060--MJ220701_80--62ce86ba7cbf91299669b6bd--N1.338F_806R.R2.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N17--62d4bbbd06292028bd354f4f--N1.338F_806R.R1.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N17--62d4bbbd06292028bd354f4f--N1.338F_806R.R2.raw.fastq.gz: 成功
MJ20220609060--MJ220714_N19--62d4bb11aa575b726f140062--N2.338F_806R.R2.raw.fastq.gz: 成功
md5sum: 警告: 1 个列出的文件无法读取
md5sum: 警告: 1 个校验和不匹配
```

最后会提示校验结果，成功、失败、打开读取失败，分别对应校验成功，失败和文件没有下载到本地或不在此路径下，如果最下面出现 WARNING 信息则表示有文件没通过验证，则这些出错的文件需要重新下载，下载之后再进行验证。这次只需要验证重新下载的文件，故可以将 md5sum.txt 文件中的未通过验证的文件拷贝到新建文件中，然后用此新文件重新进行前面的步骤，也可以将 md5sum.txt 文件中的已通过验证的文件删掉，然后重复操作前面步骤。

如果再三下载之后都不能通过校验，请及时与我司项目运营人员联系。

注：截图系统为 Red Hat 8.3.1-5

Windows 用户

方法二：

这里我们提供了需要校验的数据文件以及对应的 MD5 校验值文件（md5*.txt 文件）以及校验小程序（md5summer.exe）。

检验步骤如下：

第一步：打开下载数据目录

打开下载数据的文件夹，其中有引物对应数据的文件夹和校验小程序（md5summer.exe），不同产品的数据目录结构和文件夹名称可能不同，但这不影响检验。示例如下：



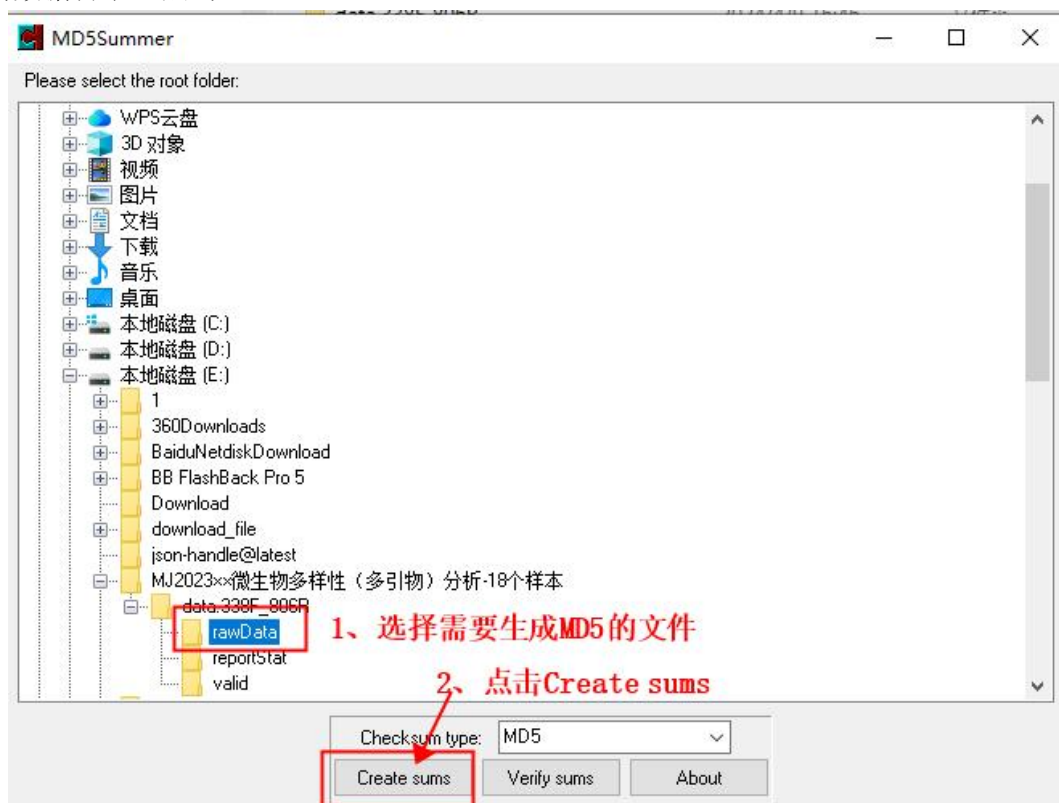
第二步：打开“md5summer.exe”

如果出现如下对话框，点击OK 即可，不影响使用。

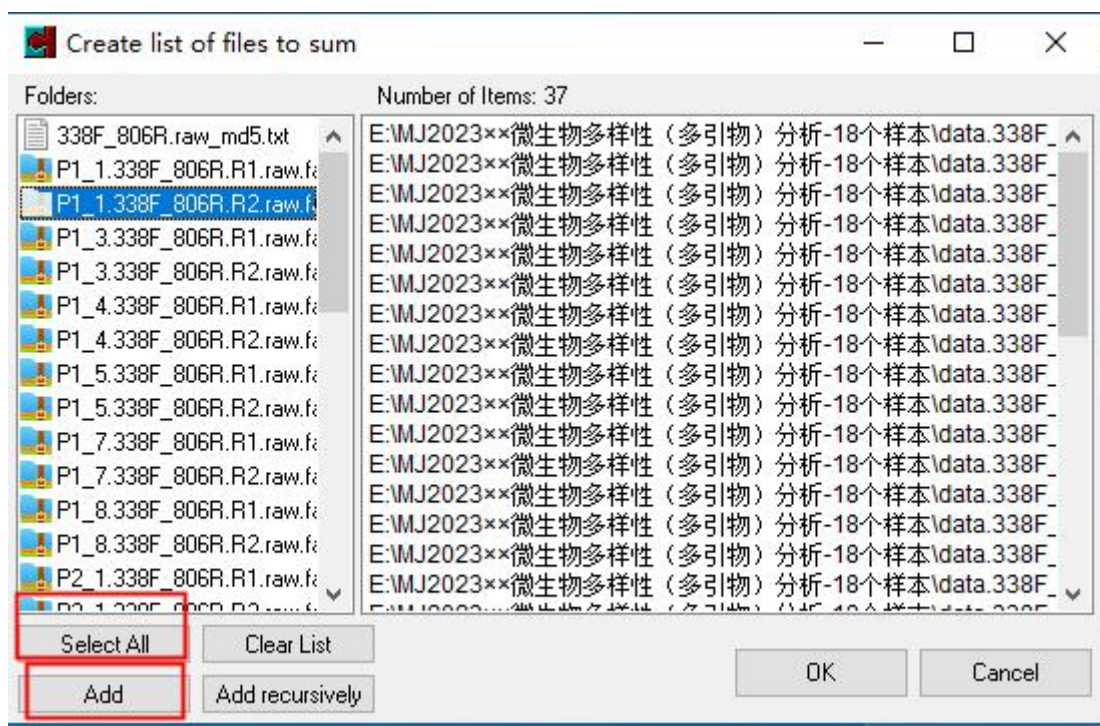


第三步：左键点击 Create sums

选择需要校验的数据目录，点击 **Create sums**

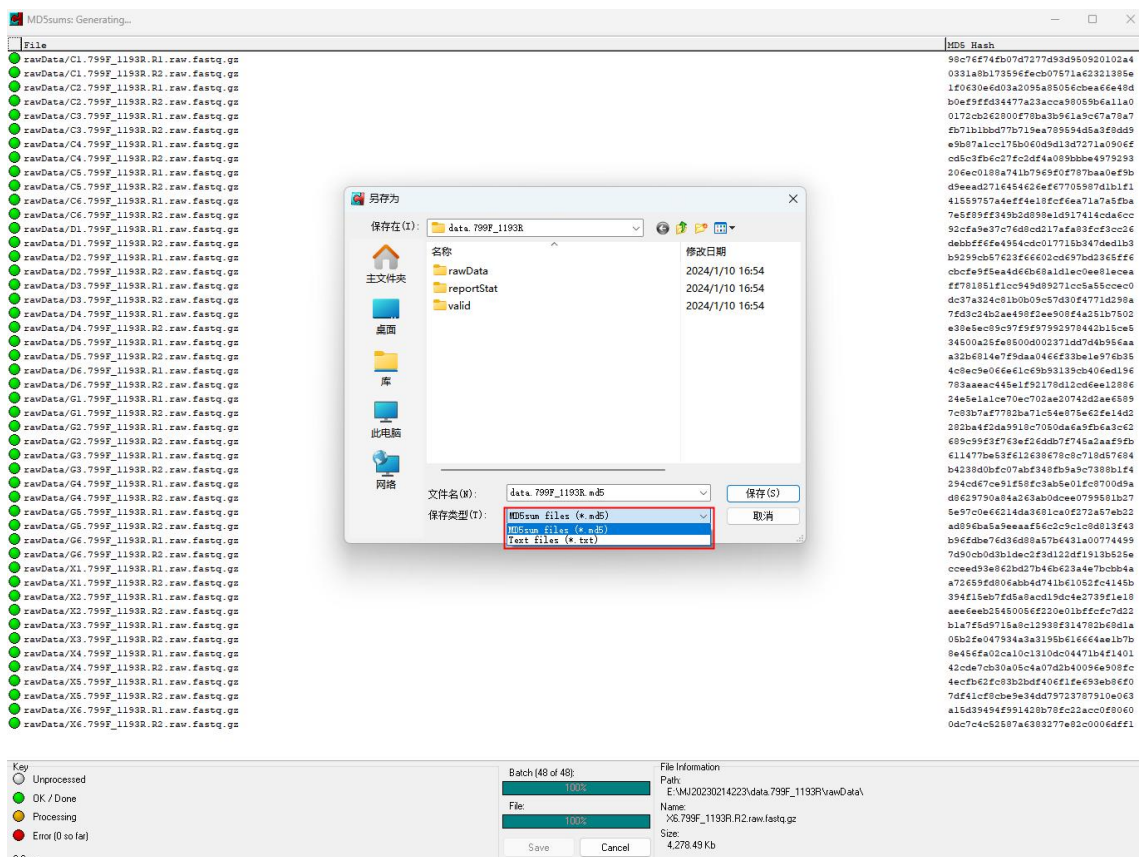


第三步: 选择需要校对的数据目录进行添加 (Add) 或者全部添加 (Select All) 点击OK进行生成MD5值



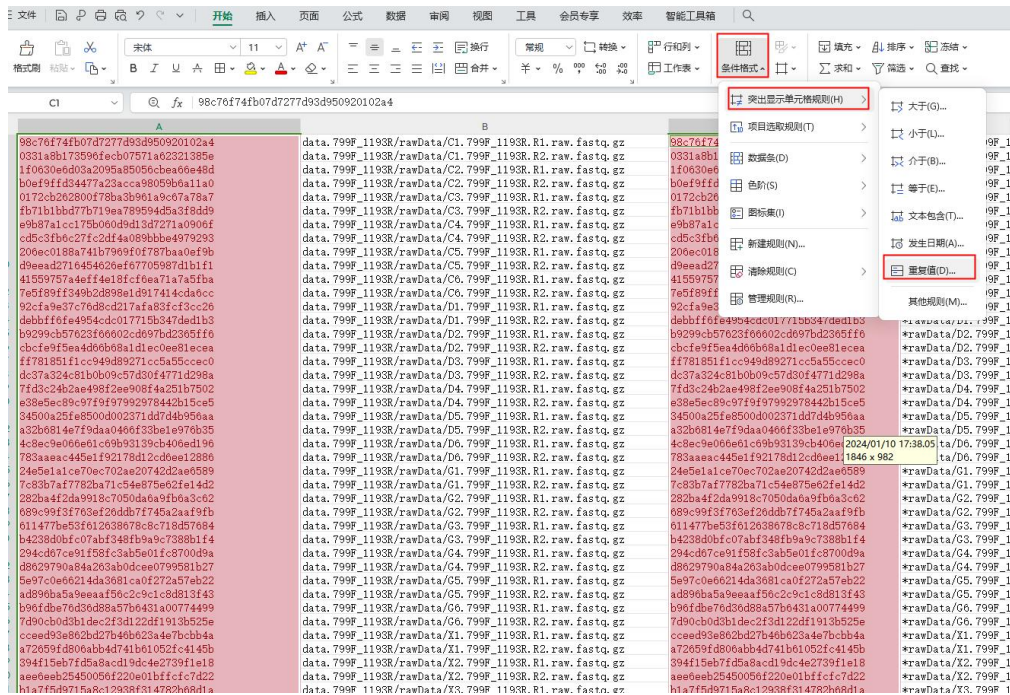
四、保存生成的MD5值

将生成好的MD5值保存成txt, 用WPS表格或者Excel打开



六、进行MD5值查重，MD5值一致说明数据完整，不一致则数据有缺失。

将数据黏贴到我们目录中md5.txt 文件中，对两列数据进行按空格分列，然后进行两列MD5值的查重



注：如果查重有不重复，说明数据缺失，建议查看数据下载是否有问题，可以尝试重新下载，校正。

如果校正还是出现不重复，请及时与我们运营进行联系，我们会尽快给您提供完整的数据链接。

MAC 用户

第一步：打开终端

首先打开 Launchpad（像火箭一样的图标），然后找到“实用工具”，最后找到终端并且打开。

第二步：校验文件 MD5 值

首先在终端中输入md5 然后再输入一个空格，最后将要验证的文件直接用鼠标拖到终端里，然后按回车键，等待（视文件大小和电脑性能而定），在下面会出现计算得到的md5 值，将这个值与我们提供的md5 值对照。若全部相同，则验证通过；若有不一样的，表示该下载文件与我们提供的不一致，请重新下载错误文件，并再次进行验证。若多次验证均不能通过，请及时与我司项目运营人员联系。