

In [6]:

```
#Скрипт для чтения текстовых файлов и определения генотипов

import glob, os
import pandas as pd

txtFileDir = r'O:\SRA\GT\' #Директория файла с отчетами в виде текстовых файлов
align_list = pd.read_excel(r'O:\SRA\GT.xlsx', header = None) #Тут читается файл с последовательностями

Allele1 = 'G' #Тут пишется первый аллель
Allele2 = 'T' #Тут пишется второй аллель

os.chdir(txtFileDir) #Тут путь, по которому происходит формирование списка файлов txt

txt_file_list = []
for txt_file in glob.glob("*.txt"): #Тут указать расширение файла
    txt_file_list.append(txt_file)
txt_file_list = sorted(txt_file_list)

SnpList = []
for i in align_list[0]:
    SnpList.append(i)

FileList = []
for i in txt_file_list:
    i = i.split('$')
    FileList.append(i[1])
FileList = sorted(set(FileList))

df_gen = pd.DataFrame(index = SnpList, columns = FileList)
df_cover = pd.DataFrame(index = SnpList, columns = FileList)

SNP_count = 0
for i in range(len(SnpList)): #SnpList - список SNP
    for j in txt_file_list:
        if SnpList[i] in j:
            l1 = align_list[1][SNP_count] + Allele1
            l2 = align_list[1][SNP_count] + Allele2
            r1 = Allele1+align_list[2][SNP_count]
            r2 = Allele2+align_list[2][SNP_count]
            TextFile = open(txtFileDir + j, 'r')
            text = TextFile.read()
            TextFile.close()
```

```
countl1 = (text.count(l1))
countl2 = (text.count(l2))
countr1 = (text.count(r1))
countr2 = (text.count(r2))
CoverAllele1 = max(countl1, countr1) #Выбор максимального значения
CoverAllele2 = max(countl2, countr2)
CommonCover = CoverAllele1 + CoverAllele2
genotype = ''
if CommonCover > 100:
    genotype = 'NA'
elif CoverAllele1 >= 2 and CoverAllele2 == 0:
    genotype = Allele1 + '/' + Allele1
elif (CoverAllele2 >= 2) and (CoverAllele1 == 0):
    genotype = Allele2 + '/' + Allele2
elif (CoverAllele1 >= 2) and (CoverAllele2 >= 2):
    genotype = Allele1 + '/' + Allele2
else:
    genotype = 'NA'

df_gen.loc[SnpList[i], j.split('$')[1]] = genotype
df_cover.loc[SnpList[i], j.split('$')[1]] = CommonCover
SNP_count += 1

df_gen.to_excel(excel_writer = r'O:\SRA\' + Allele1 + Allele2 + '_Report_genotype' + '.xlsx')
df_cover.to_excel(excel_writer = r'O:\SRA\' + Allele1 + Allele2 + '_Report_cover' + '.xlsx')
```

In []:

In []: