Instruction Manual for Windows User

Qian Feng

April 22, 2021

Download

1. Copy the downloaded .zip file into a folder to install.

2. Right-click on the .zip file and select Extract All. Please put in a folder of `C:\`.

You can find two folders in it, **Windows_scripts_for_Darren** and **Test_files**. The former one has all .exe files, you will see how these .exe work in Usage section. The latter stores source code and all middle files of implementing .exe.
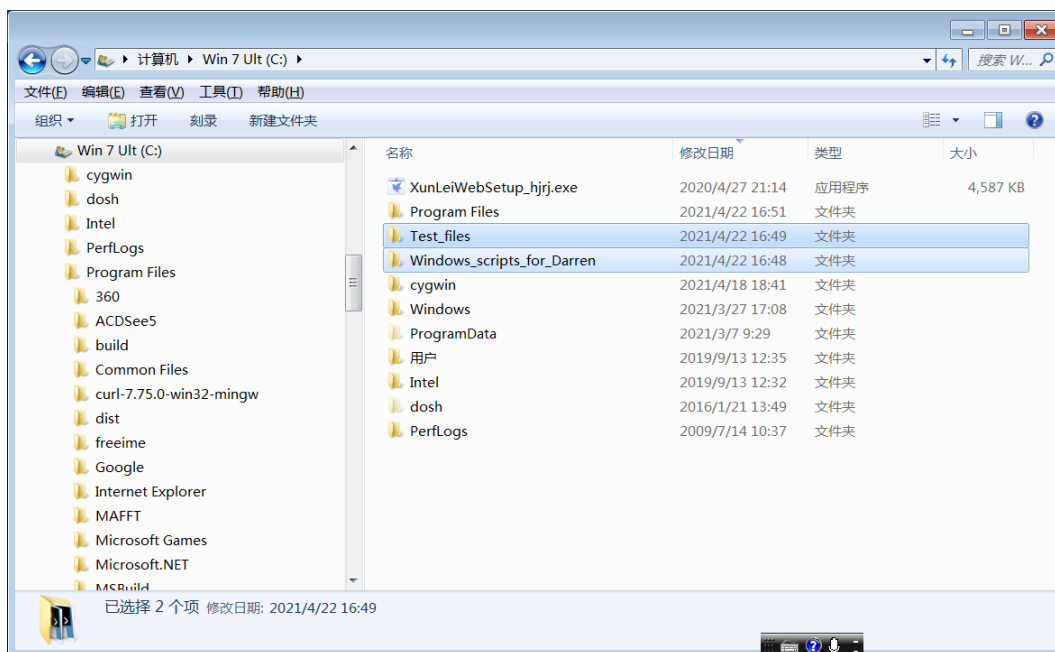
Usage

In general, we run with the following order:
mosaic → **det_rec1.exe** → **det_rec2.exe** → **det_rec3.exe**. Usage of mosaic is the most complex.

mosaic

This mosaic program (written by C language) was kindly provided by Martine Zilversmit. In order to run it, we need the input file. Therefore, just remember to copy **input.fasta** from **Test_files** to our working dir `C:\Windows_scripts_for_Darren\MZmosaic`.

The usage of running mosaic is as below.

```
cd Windows_scripts_for_Darren\MZmosaic

./mosaic -estimate -seq input.fasta  -rec 0 -aa -tag middle_file

delta=$(grep -o 'Gap initiation: .*$' middle_file_align.txt | cut -c17-)

epsl=$(grep -o 'Gap extension:  .*$' middle_file_align.txt | cut -c17-)

./mosaic -seq input.fasta -del $delta -eps $epsl -aa -tag output -grid 0.001 0.010 10
```

Once you finish running this program, you would obtain **output_align.txt**. You can also access this *so-called* middle file **output_align.txt** in your **Test_files** folder. Please copy it along with **input.fasta** to dir `C:\Windows_scripts_for_Darren` for our next steps.
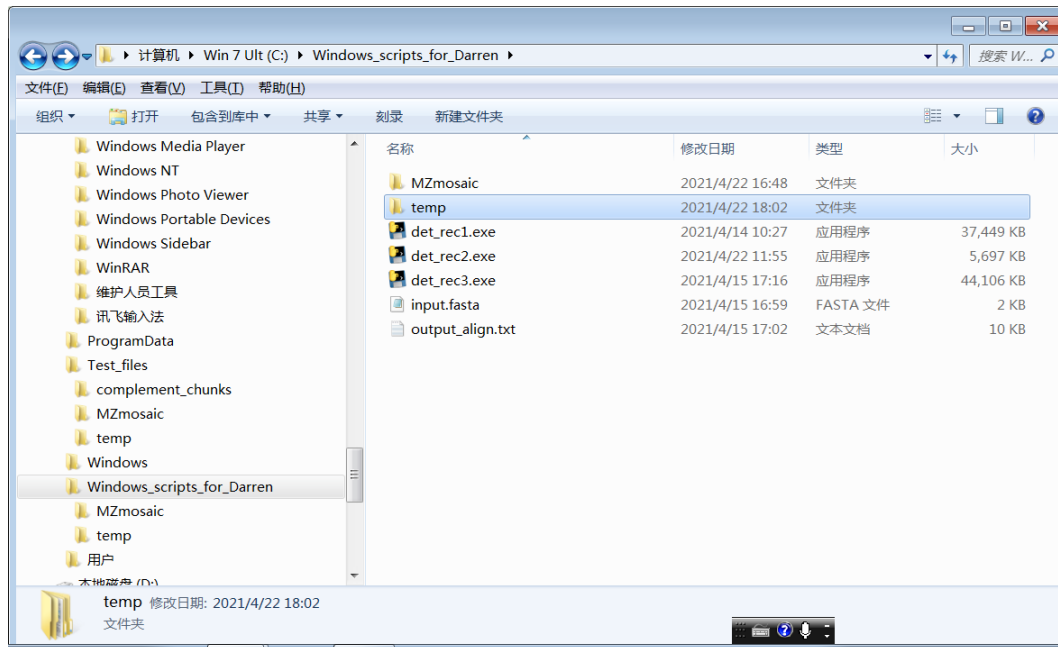
det_rec1.exe

Just simply type

```
cd Windows_scripts_for_Darren

det_rec1.exe output_align.txt input.fasta output.csv
```

After this step, you should have a temp folder.

det_rec2.exe

Our algorithm requires a third-party software, MAFFT. So please install it in your computer before running this second .exe. Here is the installation link: https://mafft.cbrc.jp/alignment/software/windows_without_cygwin.html.

Note that the installation package in this official website is only suitable for 64 bit Windows. Kazutaka Katoh (main author of MAFFT) has kindly provided us a 32 bit version mafft all-in-one package (32-bit).

Mafft is installed in dir C:\Program Files\MAFFT\mafft-7.428-win32-signed\mafft-win in my own computer, for instance. Since it's difficult to run mafft outside its installation dir, we need to run in this mafft-win dir. Please copy **det_rec2.exe**, **temp**, and **output_align.txt** in this mafft installation dir.

Just simply type

```
cd C:\Program Files\MAFFT\mafft-7.428-win32-signed\mafft-win

det_rec2.exe output_align.txt

cd temp

for /r %F in (*.fasta) do @if %~zF==0 del "%F"
```

After this step, you will get an updated **temp** folder, replace the old **temp** in usual working dir C:\Windows_scripts_for_Darren with this new **temp** folder.

det_rec3.exe

Just simply type

```
cd Windows_scripts_for_Darren

det_rec3.exe output_align.txt input.fasta output.csv
```

Our output is given in **output.csv**. The first three columns in **output.csv** show the sequence names of each identified triple, the fourth and last column provides identified recombinant sequences from triples and related support values.

triple seq1	triple seq2	triple seq3	recombinant sequence	support value
seq1	seq8	seq6	seq8	0.59
seq1	seq4	seq8	seq1	0.78
seq2	seq7	seq5	seq2	0.88
…	…	…	…	…