

FastQ.gz files Management

Pierre Lecocq

Service de Génétique CHU Liège

March 4, 2016

Overview

Perl Code (location on computer does not matter).

e.g.

Current version: FetchFastQgz_v1i1.pl (version 1, issue 1)

Windows Platform version. Tested on Win7 64b and Win10 64b

1. Perl installed (tested on version 5.22 for Windows 64)
2. 7-Zip installed (command line version)
2. Some Perl Modules installed (command line)
3. Run (command line)

Page 3

Page 4

Pages 5-6

Pages 7-8

Perl Installation

The screenshot shows the ActivePerl Downloads page. A red box highlights the 'Support' link in the top navigation bar. Another red box highlights the 'Download Perl: Other Platforms and Versions' section, which contains a table of download links. A third red box highlights the 'Request a Quote' button. A fourth red box highlights the 'Administrator' icon in the Windows taskbar. A red text overlay on the left side of the page provides instructions: 'Select the platform specific version (e.g. Windows 64 bit MSI)', 'Execute or Save then Execute (no need to register)', 'Follow Instructions. Use Default Settings', 'May take 3-5 minutes', and 'Requires Administrator Rights'. An Avira Guard notification window is visible in the bottom right corner.

<http://www.activestate.com/activeperl/downloads>

Select the platform specific version (e.g. Windows 64 bit MSI)
Execute or Save then Execute (no need to register)
Follow Instructions. Use Default Settings
May take 3-5 minutes
Requires Administrator Rights

Version	Windows (x86)	Windows (64-bit, x64)	Mac OS X (Universal)	Linux (x86_64)
5.22.1.2201	Windows Installer (MSI)	Windows Installer (MSI)	Mac Disk Image (DMG)	AS Package
5.20.3.2003	Windows Installer (MSI)	Windows Installer (MSI)	Mac Disk Image (DMG)	AS Package

7-Zip Installation



7ZIP

Home
7z Format
LZMA SDK
Download
FAQ
Support
Links

English
Chinese Simpl.
Chinese Trad.
Esperanto
French
German
Japanese
Polish
Portuguese Brazil
Spanish
Thai
Vietnamese

Download

Download 7-Zip 15.14 (2015-12-31) for Windows:

Link	Type	Windows	Description
Download	.exe	32-bit x86	7-Zip for 32-bit Windows
Download	.exe	64-bit x64	7-Zip for 64-bit Windows x64 (Intel 64 or AMD64)
Download	.7z	x86 / x64	7-Zip Extra: standalone console version, 7z DLL, Plugin for Far Manager
Download	.7z	Any	7-Zip Source code
Download	.7z	Any / x86 / x64	LZMA SDK: (C, C++, C#, Java)
Download	.msi	32-bit x86	(alternative MSI installer) 7-Zip for 32-bit Windows
Download	.msi	64-bit x64	(alternative MSI installer) 7-Zip for 64-bit Windows x64 (Intel 64 or AMD64)

Download 7-Zip 9.20 (2010-11-18) for Windows:

Link	Type	Windows	Description
Download	.exe	32-bit x86	7-Zip for 32-bit Windows
Download	.msi	64-bit x64	7-Zip for 64-bit Windows x64 (Intel 64 or AMD64)
Download	.msi	IA-64	7-Zip for Windows IA-64 (Itanium)
Download	.exe	ARM-WinCE	7-Zip for Windows Mobile / Windows CE (ARM)
Download	.zip	32-bit	7-Zip Command Line Version
Download	.tar.bz2	Any	7-Zip Source code
Download	.7z	32-bit	7-Zip Extra: 7z Library, SFXs for installers, Plugin for Fare Manager
Download	.tar.bz2	Any	LZMA SDK (C, C++, C#, Java)

You can download any versions of 7-Zip (including latest beta versions) from SourceForge:
[7-Zip files at SourceForge](#)
[7-Zip at SourceForge](#)

Download p7zip for Linux (Posix) (x86 binaries and source code):
[Download p7zip](#)

Select the platform specific version (e.g. Windows 64 bit)

Execute or Save then Execute (no need to register)

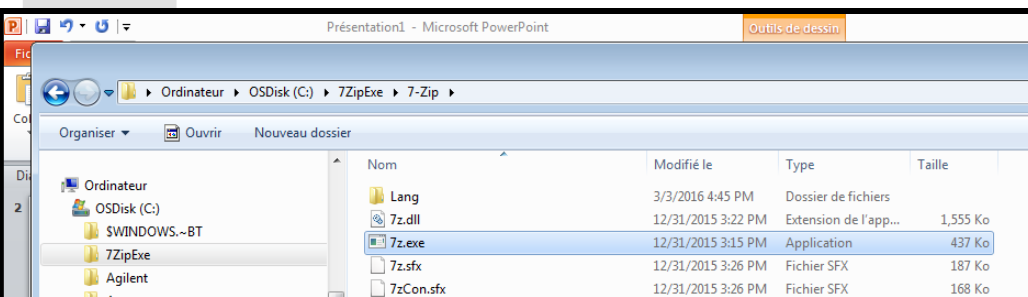
IMPORTANT: Create a new installation folder that is not under e.g. Program files
So e.g. C:\7Zipexe should make it.

Follow Instructions. May take 1-2 minutes

Requires Administrator Rights

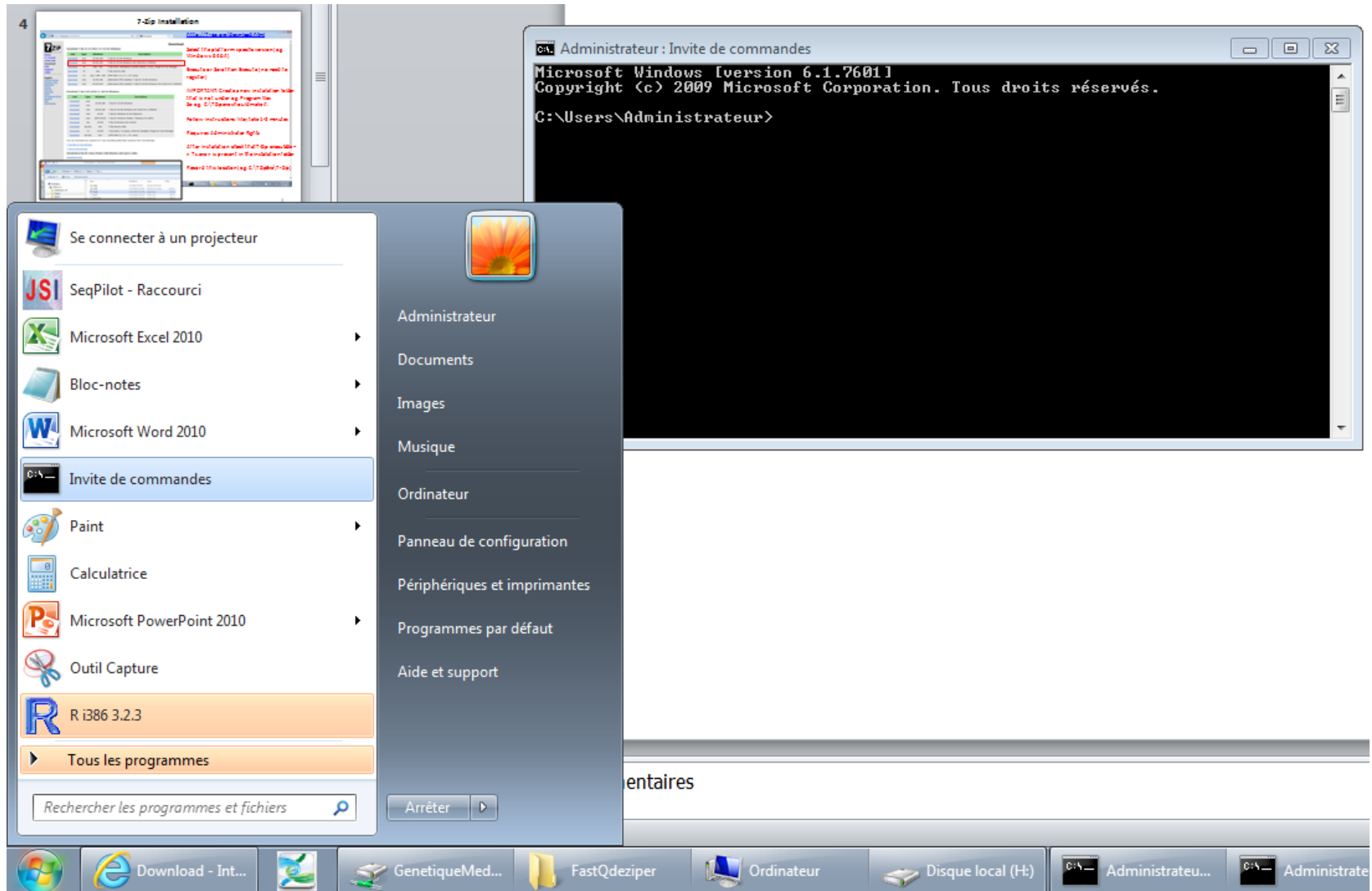
After Installation check that 7-Zip executable = « 7z.exe » is present in the installation folder

Record this location (e.g. C:\7ZipExe\7-Zip)



Perl Modules Installation

From Command line.



Perl Modules Installation

```
Administrateur : Invite de commandes
57>Data\Intensities\BaseCalls\S1-151130-0093_S1_L001_R2_001.fastq.gz

Job completed by 20160303_160732 (started by 20160303_160723).
2 FastQ files are in output folder c:\FastQdeziper\Testrun01 as well as a REPORT
and a DATA file summarizing process.

See you next time.
FetchFastQgz_v1i1.pl

C:\FastQdeziper>cpan install Win32::DriveInfo
Loading internal null logger. Install Log::Log4perl for logging messages
CPAN: Storable loaded ok (v2.53_01)
Reading 'C:\Perl64\cpan\Metadata'
Database was generated on Thu, 03 Mar 2016 05:53:38 GMT
Running install for module 'Win32::DriveInfo'
CPAN: LWP::UserAgent loaded ok (v6.13)
CPAN: Time::HiRes loaded ok (v1.9726)
Fetching with LWP:
http://ppm.activestate.com/CPAN/authors/id/M/MB/MBLAZ/Win32-DriveInfo-0.06.tar.g
z
CPAN: YAML::XS loaded ok (v0.59)
CPAN: Digest::SHA loaded ok (v5.95)
Fetching with LWP:
http://ppm.activestate.com/CPAN/authors/id/M/MB/MBLAZ/CHECKSUMS
CPAN: Compress::Zlib loaded ok (v2.069)
Checksum for C:\Perl64\cpan\sources\authors\id\M\MB\MBLAZ\Win32-DriveInfo-0.06.t
ar.gz ok
CPAN: Archive::Tar loaded ok (v2.04)
Win32-DriveInfo-0.06/
Win32-DriveInfo-0.06/Makefile.PL
Win32-DriveInfo-0.06/DriveInfo.pm
Win32-DriveInfo-0.06/Changes
Win32-DriveInfo-0.06/test.pl
Win32-DriveInfo-0.06/README
Win32-DriveInfo-0.06/MANIFEST
CPAN: File::Temp loaded ok (v0.2304)
CPAN: CPAN::Meta::Requirements loaded ok (v2.133)
CPAN: CPAN::Meta loaded ok (v2.150005)
Configuring M\MB\MBLAZ\Win32-DriveInfo-0.06.tar.gz with Makefile.PL
Checking if your kit is complete...
Looks good
Generating a dmake-style Makefile
Writing Makefile for Win32::DriveInfo
Writing MYMETA.yml and MYMETA.json
MBLAZ\Win32-DriveInfo-0.06.tar.gz
C:\Perl64\bin\perl.exe Makefile.PL INSTALLDIRS=site -- OK
Running make for M\MB\MBLAZ\Win32-DriveInfo-0.06.tar.gz
CPAN: Module::CoreList loaded ok (v5.20151213)
cp DriveInfo.pm blib\lib\Win32\DriveInfo.pm
MBLAZ\Win32-DriveInfo-0.06.tar.gz
C:\Perl64\site\bin\dmake.exe -- OK
Running make test
"C:\Perl64\bin\perl.exe" "-Iblib\lib" "-Iblib\arch" test.pl
1..11
ok 1
ok 2
ok 3
ok 4
ok 5
ok 6
ok 7
ok 8
ok 9
ok 10
Drive C: 'dir' does not return the exact free space
not ok 11
MBLAZ\Win32-DriveInfo-0.06.tar.gz
C:\Perl64\site\bin\dmake.exe test -- OK
Running make install
Prepending C:\Perl64\cpan\build\Win32-DriveInfo-0.06-JPbxlo\blib\arch C:\Perl64\
cpan\build\Win32-DriveInfo-0.06-JPbxlo\blib\lib to PERL5LIB for 'install'
Installing C:\Perl64\html\site\lib\Win32\DriveInfo.html
Installing C:\Perl64\site\lib\Win32\DriveInfo.pm
Appending installation info to C:\Perl64\lib\perllocal.pod
MBLAZ\Win32-DriveInfo-0.06.tar.gz
C:\Perl64\site\bin\dmake.exe install -- OK

C:\FastQdeziper>perl FetchFastQgz_v1i1.pl --inputfolder I:\runtest outputfolder
c:\FastQdeziper\Testrun01 -sevenzip C:\Programmes\7-Zip
```

From any location on drive just type in:

« cpan install <modulename>

e.g. cpan install Win32::DriveInfo

Wait until the installation completes

Remark: from times to times you may be asked to answer a question.

Some Modules may take several minutes.

Modules used by FetchFastQgz_v1i1.pl:

- Modern::Perl
- Smart::Comments
- Getopt::Euclid
- File::Basename
- File::Copy
- File::Compare
- File::Find
- Win32::DriveInfo

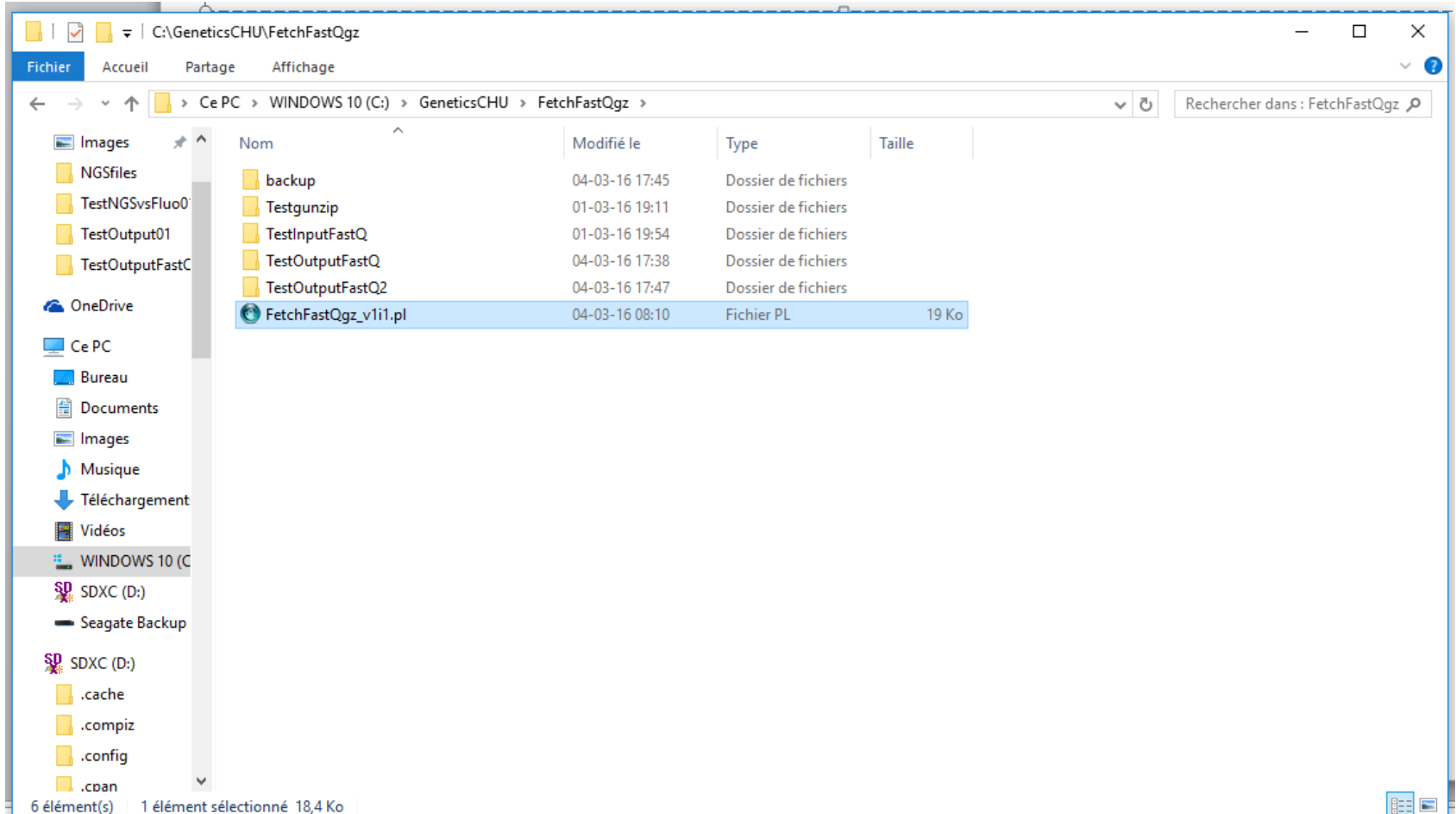
Running FetchFastQgz_v1i1.pl from command line (1/4)

First of all: Create a Folder on the computer and place the Perl code in it.

e.g. C:\GeneticsCHU\FetchFastQgz

Remark: The code will work whatever the location but it's best practice to have in a dedicated environment. ALSO if you do not have the code anymore download it from:

<http://ngyx.eu/support/various%20support%20files.html>



Running FetchFastQgz_v1i1.pl from command line (2/4)

The Perl code FetchFastQgz_v1i1.pl runs from command line. Here is an example:

```
For Info. Archiving FastQ.gz files will leave 80.2% free space
means 3208.7 Gb on target drive F.
Died at FetchFastQgz_v1i1.pl line 366.

C:\GeneticsCHU\FetchFastQgz\backup>perl FetchFastQgz_v1i1.pl --inputfolder C:\GeneticsCHU\FetchFastQgz\TestInputFastQ\*
--outputfolder C:\GeneticsCHU\FetchFastQgz\TestOutputFastQ2 --sevenzip "c:\Program Files\7-Zip" --overwrite Yes --gzarch
ivingfolder F:\Testfastgz

Welcome to FetchFastQgz_v1i1.pl.

### 7-zip executable available from: 'C:\\PROGRAM FILES\\7-ZIP'
### Overwriting set to (can be modified later on): 'YES'
### Output folder identified as: 'C:\\GeneticsCHU\\FetchFastQgz\\TestOutputFastQ2'

Starting from now the run follow-up is recorded in the REPORT file:
C:\GeneticsCHU\FetchFastQgz\TestOutputFastQ2\FASTQREPORT.TXT

Identified 5 FastQ.gz file(s) (duplicate(s) removed = 0)

For Info. Decompressing FastQ.gz files will leave 84.2% free space
means 421.5 Gb on target drive C.

For Info. Archiving FastQ.gz files will leave 80.2% free space
means 3208.7 Gb on target drive F.

7-Zip [64] 15.14 : Copyright (c) 1999-2015 Igor Pavlov : 2015-12-31
```

First this example suppose that you are in the folder where the code is located. To reach this folder from command line use:

- c: or d: or ... to select the drive

- cd subfoldername to move down one level

- cd .. To move up one level

Running FetchFastQgz_v1i1.pl from command line (3/4)

Let's review this command line run.

1. C:\GeneticsCHU\FetchFastQgz\backup>
2. perl FetchFastQgz_v1i1.pl
3. --inputfolder C:\GeneticsCHU\FetchFastQgz\TestInputFastQ*
4. --outputfolder C:\GeneticsCHU\FetchFastQgz\TestOutputFastQ2
5. --sevenzip "c:\Program Files\7-Zip"
6. --overwrite Yes
7. --gzarchivingfolder F:\Testfastgz

Explanations:

1. Location (where you are) on the drive
2. Means run code using perl code
3. Location to start searching recursively for all Fastq.gz files (you can use wildcard *)
4. Location where to record decompressed FastQ files
5. Location of the 7-Zip executable (see previous slides)
6. Option to overwrite(yes) or not (no) if found FastQ already decompressed at location (default is: Do not overwrite)
7. Optional Location where to archive source FastQ.gz files (default is: Do not archive)

Running FetchFastQgz_v1i1.pl from command line (4/4)

Important.

All the output folders should exist prior the run.

At the beginning of the run you will probably be asked to answer some questions. E.g. confirm overwriting or continuing process once drives may not have sufficient space to store files.

Also once you see that extraction start for a first file, you can leave it working alone. Nothing should happen, until the run complete.

For Info on a Icore7 2700/3500 16Gb DDR4 RAM 7200 RPM Internal and External drive a 200 Mb FastQ.Gz is fully processed (decompressed and archived) in about 20 seconds.

Finally: In the FastQ file output folder you will find REPORT and DATA TxT tab delimited files. The first one collect behavior of runs (successive if so) and the second gives an update list of FastQ files.

Last remark: We have not implemented a function to erase at end of process the source FastQ.gz files. It is up to user to check that everything went correctly and then to decide to erase these files to free space on the source drive.