illumina®

DRAGEN ORA Decompression v2.6.1

Software Guide

ILLUMINA PROPRIETARY
Document #1000000138036 v03
March 2023

For Research Use Only. Not for use in diagnostic procedures.

Revision History

Document	Date	Description of Change
1000000138036 v03	March 2023	Added • Section for the installation of DRAGEN ORA decompression v2.6.1 for Mac and Windows OS. • Recommendation for a fully transparent usage of FASTQ.ORA. Removed instructs for mkfifo command from analysis software section.
1000000138036 v02	August 2022	Updates for v2.6.1 Added the following command options: • --check-ora-reference-path • --ora-reference <refbin file DIR> • --quiet • --empty-third-line • - Removed the following command options: • --gz • --debug • -V, --verbose
1000000138036 v01	August 2021	• Changed DRAGEN Decompression to DRAGEN ORA Decompression throughout the document. • Added the Verify Lossless Compression section.
1000000138036 v00	October 2020	Initial release

Table of Contents

Overview & Installation

DRAGEN ORA Decompression Software decompresses `fastq.ora` files into `fastq.gz` files. `Fastq.ora` files are generated using lossless compression technology as part of DRAGEN. Orad is the executable file that runs the DRAGEN ORA Decompression Software, which is a standalone piece of software.

The DRAGEN ORA Decompression Software is available for the following operating System:

- Linux
- Mac
- Windows

Decompression of `FASTQ.ORA` stored on local storage is supported on Linux, Mac, and Windows.

Decompression of `FASTQ.ORA` stored on AWS S3 is only supported on Linux.

Decompression of `FASTQ.ORA` stored on Azure Blob storage is only supported on Linux.

When `FASTQ.ORA` is located on AWS S3 or Azure Blob storage, the decompression occurs on a streaming mode: the `FASTQ.ORA` file does not need to be fully transferred before decompression can start.

Do not use this documentation for decompression on the DRAGEN Server. The DRAGEN Server has its own integrated DRAGEN ORA Decompression tool.

Installation Requirements

The following are the minimum requirements for the DRAGEN ORA Decompression Software:

Component	Minimum Requirements
System memory	8 GB RAM
Free disk space	2 GB
Compatible Linux distributions	• CentOS 7 • Ubuntu 14 • Debian 8 • Fedora 26
Compatible Mac distributions	Mac 10.15 (Apple silicon and Intel CPU processors)
Compatible Windows distributions	Windows 10

Install the DRAGEN ORA Decompression Software

Use the following steps to install the DRAGEN ORA Decompression Software once DRAGEN ORA has been downloaded from the support site.

Linux or Mac

1. Extract the archive files using the following command:

   ```
   tar -xzvf orad.2.6.1.linux.tar.gz (Linux)
   tar -xzvf orad.2.6.1.mac.tar.gz (Mac)
   ```

2. Navigate to the Orad directory as follows:

   ```
   cd orad_2_6_1
   ```

3. Move the executable to your preferred location as follows:

   ```
   mv orad your_preferred_location/
   ```

4. Add `Orad` to your path as follows:

   ```
   echo 'PATH=$PATH: your_preferred_location/'» ~/.bashrc
   source ~/.bashrc
   ```

5. Move the oradata folder content into the home repository as follows:

   ```
   mv oradata ~
   ```

   To store the folder in a different location, use the `ORA_REF_PATH` environment variable as follows.

   ```
   mv oradata ~/otherlocation/
   export ORA_REF_PATH=~/otherlocation/oradata/
   ```

Windows

1. Extract the downloaded archive with a software that can handle gziped tarballs, such as 7-Zip. Right-click on the archive and select **extract with**. The following two files are extracted:
   - `orad.exe`
   - `refbin`

   The following steps use `C:\Users\user1` as an example location. Change `C:\Users\user1` to the location where you extracted the archive.
2. Open the Command Prompt application.

3. Set the environment variables to use the `orad.exe` and the `refbin` file with the `set` command or the `setx` command. The `set` command configures the variables temporarily (for the current console window) while the `setx` command configures the variables permanently.
4. Set the path to the `orad.exe` to the `PATH` environment variable as follows:

```
set PATH=%PATH%; C:\Users\user1
```

or

```
setx PATH=%PATH%; C:\Users\user1
```

5. Set the path to the `refbin` file to an ORA_REF_PATH environment variable as follows:

```
set ORA_REF_PATH= C:\Users\user1
```

or

```
setx ORA_REF_PATH= C:\Users\user1
```

Using the DRAGEN ORA Decompression

Use the following commands to decompress the files.

On Windows, replace `orad` with `orad.exe`. Example is `orad.exe FILE [args]`.

```
orad FILE [args]
```

or

```
orad [args] FILE
```

Available Command Line Options

Command	Description
-C --check	Checks the integrity of the specified ORA file. This option decompresses the file in memory and verifies that the checksum of the decompressed data and the checksum of the original data are identical. The decompressed file is not saved.
--raw	Decompresses the ORA file into an uncompressed FASTQ file. By default, the DRAGEN ORA Decompression Software decompresses to `gzip` format.

Command	Description
--rm	Deletes the input file after successful execution. By default, the input file is not deleted. This option is not supported for files in AWS S3 or Azure Blob Storage.
-t <INT> --threads <INT>	Sets the maximum number of threads allowed by the system. The default value is 8.
-f --force	Overwrites the output file without prompting. By default, if the output file exists, the software exits without overwriting.
-h -help --help	Prints help and exits.
-v -version --version	Prints software version.
-i --info	Prints information about the compressed ORA file. The following information is included: • Software version used to compress the file. • Total number of sequences in the file. • Total number of bases in the file. • If the file contains interleaved data. • The original name of the file if it was saved in the `fastq.ora` file. This option is not supported for files in AWS S3 or Azure Blob Storage. Although the ORA file format supports concatenation of `fastq.ora` files, using this command on a concatenated `fastq.ora` file prints erroneous information.
-c --stdout	Prints the decompressed file to the default standard output `stdout`. This is useful to share the results with another application without writing the decompressed file to disk.
-	Reads an input `fastq.ora` file from the default standard input `stdin`. This option is not supported for Windows OS.
-P <PATH> --path <PATH>	Sets the path location of the output file. The default file name is used. If a path is not specified, the file is created in the same location as the input file. This option overwrites the path if it is used with the -o option.

Command	Description
-o <FILENAME> --out <FILENAME>	Sets the name of the output file and the path when used with `-P`. The default is the name of the input `fastq.ora` file.
-N --name	Restores the original name saved in the `fastq.ora` file, at the time it was compressed to a `fastq.ora` file.
-I --interleave	Decompresses the output file into a single interleaved file. By default, when the input is a single interleaved `fastq.ora` file, the decompression automatically decompresses into two separate paired read files. If the interleaved `fastq.ora` file was generated with DRAGEN ORA v.4.0 or later, `-interleaved` is included in the file name.
--ora-reference <refbin file DIR>	Changes the directory of the ORA reference file `refbin`. By default the software looks for the reference file in the following locations: • `./refbin` • `$HOME/oradata/refbin` • `./oradata/refbin` If you specified a location in the environment variable, the software also looks in the location `ORA_REF_PATH`. For example, set with export `ORA_REF_PATH=/some/path/`.
--check-ora-reference-path	Verifies if the ORA reference file `refbin` is accessible and prints the `refbin` path. Decompression does not occur when this option is added.
--quiet	Sets decompression to quiet mode. In quiet mode, nothing is written to the standard output and standard error. This mode is ignored when used with `-c (--stdout) or -C (--check)`.
--empty-third-line	Outputs the third line in the FASTQ format, (which is, the line that starts with +) as an empty line. By default this line is preserved.
-r --repeat-header	Adds the read header to the third line in the FASTQ format, (which is, the line that starts with +).

Command Examples

Using Windows, replace `orad` with `orad.exe`. Example is `orad.exe myfile.fastq.ora --check`.

Command	Description
orad myfile.fastq.ora --check	Checks the integrity of an ora file.
orad myfile.ora	Decompression command.
orad myfile.fastq.ora --info	Prints information summary of an ora file.
orad -c --raw myfile.fastq.ora | head	Prints the first lines of the corresponding `.FASTQ` file in the terminal.
orad --check-ora-reference-path	Verifies the accessibility of the ora reference file and print its path.

DRAGEN ORA Decompression Combined With Analysis Software

For a fully transparent usage of `fastq.ora` files (no changes in the command, no overhead, no additional footprint) with third-party bioinformatics tools, DRAGEN ORA Helper Suite Software is recommended and available for download on the DRAGEN ORA support site page. This software is only supported on Linux.

For a semi transparent usage of `FASTQ.ORA` files with third-party bioinformatics tools, use DRAGEN ORA Decompression with the pipe function or process substitution. This method improves system performance by reducing reads and writing to the disk versus a full decompression step.

- If the analysis tool can read from the standard input, such as BWA, use the following command:

  ```
  orad file.fastq.ora -c --raw | bwa mem humanref.fasta - > resu.sam
  ```

  The `-c` option decompresses to standard output. The result is sent `|` to BWA, which uses the dash option `-` to read from standard input. This also works for paired reads, which uses the `-p` option of BWA to specify that the input contains interleaved paired reads.

- If the analysis tool cannot read from the standard input, you can use process substitution.

  ```
  bwa mem humanref.fasta <(orad file.fastq.ora -c --raw) > resu.sam
  ```

  For the file name, use the `<( )` syntax containing the command that generates the file to standard output. In this case, `orad` with the `-c` option as in the command above. This method does not work when the third-party tool checks the input file name or when the third-party tool does not read the file sequentially.

On the Windows, replace `orad` with `orad.exe`.

Verify Lossless Compression of FASTQ.ORA Files

To make sure that data was not lost during the compression of the `FASTQ.ORA` file, compare the MD5 checksum of the decompressed `FASTQ.ORA` file and the MD5 checksum of the decompressed `FASTQ.GZ` file.

On the Windows OS, replace `orad` with `orad.exe`.

1. Compute the md5 checksum of the uncompressed `FASTQ.ORA` content as follows.

   ```
   md5sum <(orad myfile.fastq.ora --raw -c )
   ```

2. Compute the md5 checksum of the uncompressed `FASTQ.GZ` content as follows.

   ```
   md5sum <(gzip -d -c myfile.fastq.gz)
   ```

Resources & References

The DRAGEN ORA support page on the Illumina support site provides additional resources. These resources include training, compatible products, and other considerations. Always check support pages for the latest versions.

Illumina, Inc.
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

illumina®