

CUSHAW3-UPC

Reference Manual



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Authors:

Jorge González-Domínguez

Yongchao Liu

Bertil Schmidt

Institution:

Parallel and Distributed Architectures Group

Institute of Computer Science

Johannes Gutenberg University Mainz, Germany

Date:

August 6, 2015

Contents

1	Introduction	2
2	Unified Parallel C++ (UPC++)	3
3	Installation	4
4	Execution	5

1 Introduction

CUSHAW3-UPC is an open-source parallelized, sensitive and accurate short-read aligner for multicore clusters. It obtains the same alignments as the original *CUSHAW3* tool, which consistently outperforms *CUSHAW2*, *BWA-MEM*, *Bowtie2* and *GEM* in terms of single-end and paired-end alignment, as well as *Novoalign* for short-reads with high error rates. *CUSHAW3-UPC* also works with color-space reads and, similarly to *CUSHAW3*, it is consistently one of the best aligners compared to *SHRiMP2* and *BFAST*. The main advantage of *CUSHAW3-UPC* is that it is able to exploit the compute capabilities of multicore clusters in order to accelerate the alignments.

The corresponding licence file is shipped with the software but can also be accessed via:

`http://www.gnu.org/licenses/gpl-3.0.en.html`

If you want to reuse the code, please ensure compliance to the aforementioned licence and a proper attribution/citation of the original work/authors.

2 Unified Parallel C++ (UPC++)

The parallelism in *CUSHAW3-UPC* is controlled thanks to a UPC++ code, a parallel language that follows the Partitioned Global Address Space (PGAS) paradigm. Similar to the popular Message Passing Interface it allows for the distributed execution of parallel programs on a compute cluster. However, all nodes can access the whole address space across different nodes without the need for explicit send and receive statements in contrast to MPI. Instructions about how to download, install and execute UPC++ can be found at the corresponding web page <https://bitbucket.org/upcxx/upcxx/wiki/Home>.

3 Installation

To complete the installation of *CUSHAW3-UPC* follow these steps:

1. Untar the archive and move into the *CUSHAW3-UPC* directory.
2. Update the file *Makefile* of the root directory in order to indicate the correct path and libraries for the UPC++ compiler installed in your system. More information can be found at <https://bitbucket.org/upcxx/upcxx/wiki/Home>.
3. Type `make all` to build *CUSHAW3-UPC*.

4 Execution

The following steps must be completed in order to execute parallel short read alignments on multicore clusters with *CUSHAW3-UPC*.

1. Build the BWT and FM-index using the command `index` of *CUSHAW3*. The command `index` is not included in *CUSHAW3-UPC* as the creation of the structures to represent the genome was not parallelized.
2. Use the commands `align` or `calign` of *CUSHAW3-UPC* for base- or color-space alignment, respectively. The parameters of these commands are exactly the same as for the original *CUSHAW3* tool (see <http://cushaw3.sourceforge.net/homepage.htm>)

Note: The current version of the tool is able to work with FASTA or FASTQ files as input. However, it cannot work with binary SAM files.

Note 2: The cluster must be able to access the BWT and the FM-index files from different nodes