FLASH (Fast Length Adjustment of SHort reads) is a very accurate and fast tool
to merge paired-end reads that were generated from DNA fragments whose lengths
are shorter than twice the length of reads. Merged read pairs result in unpaired
longer reads. Longer reads are generally more desired in genome assembly and
genome analysis processes.

INSTALLATION
Download the package and run

tar -xvf FLASH.tar (to untar the file)
./buildBinaries.sh (to build the executable file)

A simple installation script for Linux x86_64 system using gcc compiler is provided.
Running ./buildBinaries.sh on the command prompt will create executable file "flash".
FLASH does not depend on any particular architecture, so if you are using a different
compiler, you should be able to write your own script to install FLASH.

USAGE
flash <mates1.fastq> <mates2.fastq> [-m minOverlap] [-M maxOverlap] [-x mismatchRatio]
[-p phredOffset] [-o prefixOfOutputFiles] [-d pathToDirectoryForOutputFiles]
[-f averageFragment Length] [-s standardDeviationOfFragments] [-r averageReadLength]
[-h displayHelp]

mates1.fastq and mates2.fastq are fastq files of paired-end reads from the short
fragment library (with insert size less than twice the length of reads). The
corresponding mates should be in the same order in both files.

Options:

-m: minOverlap is the minimum required overlap length between two reads to provide
a confident overlap. Default: 10bp.

-M: maxOverlap is the maximum overlap length expected in approximately 90% of read
pairs. It is by default set to 70bp, which works well for 100bp reads generated from
180bp library (normal distribution of fragment lengths is assumed). Overlaps longer
than maxOverlap are still considered as good overlaps, but the mismatch ratio
(explained below) is calculated over the maxOverlap rather than the true overlap
length. If you enter a value for maxOverlap, then the read length, fragmetn length
and standard deviaiton of fragment lengths that you enter will be ignored for
calculation of maxOverlap parameter. Default: 70bp.

-x: mismatchRatio is the maximum allowed ratio of the number of mismatches and the
overlap length. An overlap with mismatch ratio higher than the set value is considered
incorrect overlap and mates will not be merged. Any occurence of an "N" in any read is
ignored and not counted towards the mismatches or overlap length. Our experimental
results suggest that higher values of mismatchRatio yield larger number of correctly
merged read pairs but at the expense of higher number of incorrectly merged read
pairs. Default: 0.25.

-p: phredOffset is the smallest ASCII value of the characters used to represent
quality values of bases in fastq files. It should be set to either 33, which corresponds
to the later Illumina platforms and Sanger platforms, or 64, which corresponds to
the earlier Illumina platforms. Default: 33.

-o: prefix of output files. Default: out.

-d: path to directory for the output files. Default: current working directory.

-r: average read length. Default: 100.

-f: average fragment length. Default: 180.

-s: standard deviaiton of fragment lengths. If you do not know standard deviation of the
fragment library, you can probably assume that the standard deviation is 10% of the average
fragment length. Default: 20.

```
-h: display help information.
```

PERFORMANCE
FLASH is extremely fast tool for merging mates from short paired-end library. It processes one million 100-bp read pairs in 120 seconds on a server with 256 GB of RAM and a six-core 2.4 GHz AMD Opteron, or in 129 seconds on a desktop with 2 GB of RAM and a dual-core Intel Xeon 3.00 GHz CPU. FLASH currently runs only in single threaded mode.

ACCURACY
With reads' error rate of 1% or less, FLASH processes over 99% of read pairs correctly. With error rate of 2%, FLASH processes over 98% of read pairs correctly when default parameters are used. With more aggressive parameters (i.e., -x 0.35), FLASH processes over 90% of read pairs correctly even when the error rate is 5%.

COMMENTS/QUESTIONS/REQUESTS
Send an e-mail to flash.comment@gmail.com