Blue Distribution 1.0.1   13th May 2013

Changes from Blue 1.0.0 to 1.0.1

- Added –o <output directory> option to Blue.

  Allows the user to specify where the corrected reads files will be written. The default is that they are written to the directory where the reads files were found.

- Maintain the qual line header convention for Fastq files

  Use full headers on the corrected reads if the original reads had full headers, and use just '+' if the original file used '+'

- Stopped rewriting reads with trailing Ns

  The test used to detect when Blue was rewriting (rather than repairing) a read did not catch the case of reads with many trailing N bases. This has now been fixed and such reads are left alone (or dropped with the 'good' option).

- Improved behavior on 'unbalanced' k-mers

  Illumina data sometimes has asymmetric reads, where the depth of coverage on one strand is very much greater than on the other strand. The correction of such reads has been improved.

- Improved behavior on error k-mers coming off high repetition k-mers

  High repetition k-mers (coming from repeated regions such as the ribosome) can have error variants that look as if they might be correct. Low-repetition variants of high repetition k-mers are now treated as potential errors during the detection pass, and as non-preferred replacements during the correction phase. This change has resulted in a small improvement in the accuracy of corrections.

- The 'OK' and 'poor' depth for a read are now recalculated at the end of the correction pass and the new values are used to determine if the read still needs correction (and if it should be dropped if the 'good' option is set)

Changes from Tessel 1.0.0 to 1.0.1

- Added the '-min' option to specify a cutoff for k-mers written to the .cbt file

  By default, Tessel writes all k-mers+counts to the .cbt file, even singleton k-mers. This accurately captures the counts for all k-mers, at the cost of larger than needed .cbt files. Subsequent programs, such as Blue, discard low-repetition k-mers when they load the .cbt file into memory, so saving such k-mers is often unnecessary. This option lets the user reduce the size of the .cbt file by dropping low-repetition k-mers during the writing phase.

Changes from GenerateMerPairs 1.0.0 to 1.0.1

- GenerateMerPairs now takes the same form of parameters as Blue and Tessel

usage: GenerateMerPairs [-m min] [-t threads] cbtFN readsPattern or file names

- Added the '-min' option to specify a lower cut-off for the depth of k-mers loaded into memory

  This option reduces the memory requirements of GenerateMerPairs by not loading low-repetition k-mers into memory from the .cbt file. The default value for this parameter is 3.

- Added the '-t' option to specify how many parallel threads should be used.

  The default number of threads is 1.

- No longer generate a 'pair' where one of the k-mers has a high depth count and the other is very low. These come from low-repetition errors occurring in high depth reads and their generation was a bug that has been fixed. This change also reduces the size of the .prs file (and subsequent 'pairs' table).