

deMultiplexing

Nate Olowo

October 16, 2018

Demultiplexing Assignment

Code developed for paired end reads in which reads and indexes are in separate files. The following scripts will output write statistics and a graph of relative sample abundances for the demultiplexing. Talapas queueing bash script **Usage** DeMultiAlgoArg.py [options]

Input flags

Required

-r1 [readfile1]

-i1 [indexfile1]

-r2 [readfile2]

-i2 [indexFile2]

Optional

-t [mean quality score threshold] *Default: 25*

-b [Barcodes mapping file] *Default: indexes.txt*

Statistics defined

Records Read: Total number of reads processed in each file. All four files should have the same amount of lines, if not, the script will only read until your shortest file.

bad Reads: N found in barcode of either of the 2 index reads

Below Threshold: Average quality scores of each index read does not surpass threshold. No N's in barcode though

unmatched indices: Indexes don't match. Passes threshold and N tests stated above.

___Not found in indexes file:___ Index of forward read not found in indexes.txt input file. Passes all tests above.

Matched Reads: Passes all above test. Read written to respective sample with forward and reverse written separately. Forward index place inline with read headers for forward and reverse reads.

Expected output from Test files

Write statisitics

Records Read: 9

Matched Reads: 2

Below threshold: 1

unmatched indices: 2

bad reads(Ns): 3

Matched barcodes above threshold not found in indexes file: 1

Output

This script will output $2n+2$, where $n = \#$ of Barcode IDs found in indexes.txt file.

All bad reads are written to unmatched*.fastq file pairs.

Output of 1294 read files.

Relative abundancies

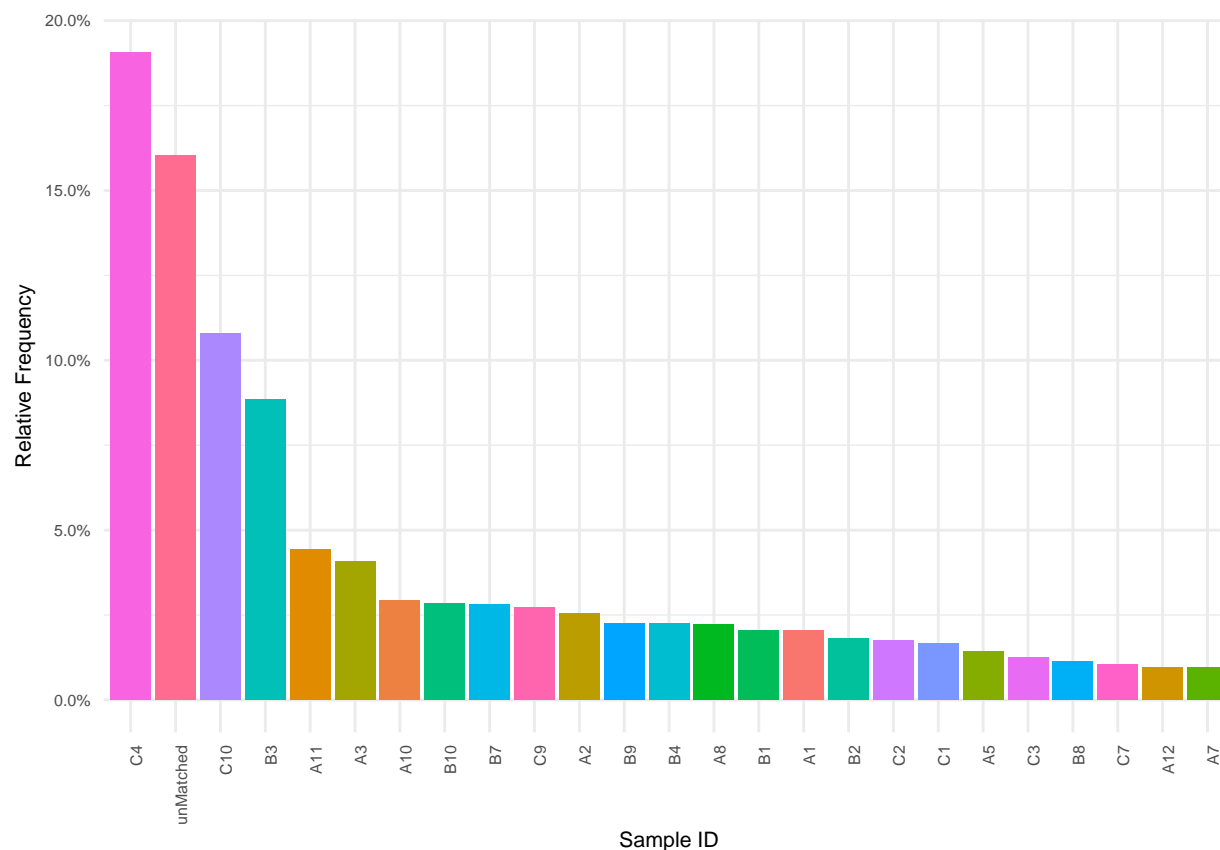


Figure 1: Relative abundance of forward and reverse reads of each sample deMultiplexed. Forward and reverse files contain same number of reads.

Output Statistics:

Write statistics

Records Read: 363,246,735

Matched reads: 304,980,270

Below threshold: 44,557,157

unmatched indices: 9,499,159

bad reads(Ns): 4,205,183

Matched barcodes above threshold not found in indexes file: 4,966

Estimated amount of index-hopping

$$(44557157 + 9499159 + 4205183 + 4966) / 363246735 * 100 = 16\% \text{ indexes hopped}$$

84% matched indexes

Output Snippets

[illegible]