

Sep 29, 2013

Dear Editors,

We would like the editors to consider our research article entitled “These are not the k-mers you are looking for: efficient online k-mer counting using a probabilistic data structure” for publication in PLoS One. In this manuscript, we present the khmer software package - a CountMin Sketch implementation of a memory-efficient k-mer counting system.

Efficient k-mer counting plays an important role in many bioinformatics approaches including data preprocessing for de novo assembly, repeat detection, and sequencing coverage estimation. How to count the large number of k-mers in next-gen sequencing data sets efficiently and effectively has become an unavoidable and challenging task. To tackle this problem we present an analysis, comparison, and discussion of software based on a CountMin Sketch data structure, which can be particularly memory efficient for certain problems. There has been increased activity in the area of k-mer counting recently, including:

Kurtz S, Narechania A, Stein JC, Ware D (2008) *A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes*. BMC Genomics 9: 517.

Marcais G, Kingsford C (2011) *A fast, lock-free approach for efficient parallel counting of occurrences of k-mers*. Bioinformatics 27: 764–770.

Melsted P, Pritchard JK (2011) *Efficient counting of k-mers in DNA sequences using a bloom filter*. BMC bioinformatics 12: 333.

Rizk G, Lavenier D, Chikhi R (2013) *Dsk: k-mer counting with very low memory usage*. Bioinformatics 29: 652-3.

Deorowicz S, Debudaj-Grabysz A, Grabowski S (2013) *Disk-based k-mer counting on a pc*. BMC Bioinformatics 14: 160.

Roy RS, Bhattacharya D, Schliep A (2013) *Turtle: Identifying frequent k-mers with cache-efficient algorithms*. Arxiv preprint .

For Academic Editors, we suggest:

Jan Aerts

Dongxiao Zhu

Haixu Tang



**MICROBIOLOGY &
MOLECULAR GENETICS**

Michigan State University
2215 Biomedical &
Physical Sciences
East Lansing, MI
48824-4320
517-355-6463
Fax: 517-353-8957
Email: mmg@msu.edu
Web: mmg.msu.edu

MICHIGAN STATE
UNIVERSITY

For reviewers, we suggest:

Rayan Chikhi (author of DSK)

Jared Simpson

Pall Melsted (author of BFCOUNTER)

We request that you do not ask the authors of Turtle (Rajat Shuvro Roy, Debashish Bhattacharya, or Alexander Schliep to review the paper as I am (openly) reviewing their paper for another journal at the moment.

Sincerely,

C. Titus Brown (corresponding author)
Assistant Professor
Computer Science and Engineering /
Microbiology and Molecular Genetics
Michigan State University



**MICROBIOLOGY &
MOLECULAR GENETICS**

Michigan State University
2215 Biomedical &
Physical Sciences
East Lansing, MI
48824-4320

517-355-6463
Fax: 517-353-8957
Email: mmg@msu.edu
Web: mmg.msu.edu