

UC Davis

UC Davis Previously Published Works

Title

Correction to: IonCRAM: a reference-based compression tool for ion torrent sequence files

Permalink

<https://escholarship.org/uc/item/09q469t1>

Journal

BMC Bioinformatics, 21(1)

ISSN

1471-2105

Authors

Shokrof, Moustafa
Abouelhoda, Mohamed

Publication Date

2020-12-01

DOI

10.1186/s12859-020-03766-1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-ShareAlike License, available at <https://creativecommons.org/licenses/by-sa/4.0/>

Peer reviewed

CORRECTION

Open Access



Correction to: IonCRAM: a reference-based compression tool for ion torrent sequence files

Moustafa Shokrof¹ and Mohamed Abouelhoda^{2,3,4*} 

The original article can be found online at <https://doi.org/10.1186/s12859-020-03726-9>.

*Correspondence: mabouelhoda@yahoo.com
⁴ Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, University Square, Giza, Egypt
Full list of author information is available at the end of the article

Correction to: BMC Bioinformatics (2020) 21:397

<https://doi.org/10.1186/s12859-020-03726-9>

Following publication of the original article [1], the authors identified a missing section in the published article. The missing section is given below.

Algorithm IonCRAM-CompressBAM

- 1 Sort the BAM file (if not sorted) by genomic coordinates. Then sort the reads starting at the same locus lexicographically via sorting their CIGAR string.
- 2 Separate the signals of the forward reads from those of the reverse ones and process each group independently (in parallel) using Steps 3 and 4.
- 3 Remove the flow signals from the BAM file and store them separately. Compress the remaining fields of the BAM file (sequence, quality, and other fields) using a reference based method (We use the program Scramble [20] for this step.)
- 4 Define blocks of flow signals, such that the reads in each block are mapped to the same locus. Each block B can be processed in parallel using the steps 4.1 to 4.3:
 - 4.1 Let F_r denote the r th flow-signal vector in B ($r \in [1..m]$), and let $F_r[i] \in \mathbb{Z}$ denote the i th component of it, $1 \leq i \leq n$. Take F_1 as a reference vector and compute the difference vector D_j , where $D_j[i] = F_j[i] - F_1[i]$, $1 \leq i \leq n$, $1 \leq j < m$.
 - 5.1 For each D_j , allocate a vector V_{j1} of n bytes to store its values. If $D_j[x] > 255$ for any x , then we write 255 in $V_{j1}[x]$ and write the values $(D_j[x]-255)$ in a separate list V_{j2} . (The length of V_{j2} list equals the number of values larger than 255 in D_j and they are very rare in practice.)
 - 6.1 Concatenate F_1 and the V vectors and compress them. (We use the XZ algorithm as default method for that purpose.) (F_1 is a reference flow signal vector that will be used in decompression).
- 5 Wait until all (parallel) processes finish. Use the Linux tar package to create a compressed folder including the compressed B blocks files and the other compressed CRAM files computed in Step 1.



The author group has been updated above and the original article [1] has been corrected.

Author details

¹ Faculty of Computer Science, University of California at Davis, Davis, CA, USA. ² King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia. ³ Saudi Human Genome Program, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia. ⁴ Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, University Square, Giza, Egypt.

Published online: 06 October 2020

Reference

1. Shokrof M, Abouelhoda M. IonCRAM: a reference-based compression tool for ion torrent sequence files. *BMC Bioinformatics*. 2020;21:397. <https://doi.org/10.1186/s12859-020-03726-9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.