Erratum to: A representation of a compressed de Bruijn graph for pan-genome analysis that enables search

Timo Beller and Enno Ohlebusch*

Erratum to: Algorithms Mol Biol (2016) 11:20
DOI 10.1186/s13015-016-0083-7

After publication of the original article [1], the authors noticed errors in Algorithm 2 and the caption of Table 4. In Algorithm 2, the term “rank_1(B_{i}, i−1) + 1” should be included on line 28 and not line 29. In addition, in the caption of Table 4, the word “BV_l” should be replaced by “B_l” and the word “BV_r” should be replaced by “B_r.” The correct versions of Algorithm 2 and Table 4 are included in this erratum.

Table 4 Breakdown of the space usage of the variants of Algorithm A4

| Algorithm | Part     | 62 E.coli | 7 x Chr1 | 7 x HG  |
|-----------|----------|-----------|----------|---------|
| A4        | wt-bwt   | 0.42 (23.83%) | 0.44 (36.23%) | 0.43 (22.68%) |
| A4        | Nodes    | 0.10 (5.94%)  | 0.03 (2.61%)  | 0.04 (2.02%)  |
| A4        | B_r      | 0.16 (8.93%)  | 0.16 (12.86%) | 0.16 (8.25%)  |
| A4        | B_l      | 0.14 (8.04%)  | 0.14 (11.57%) | 0.14 (7.42%)  |
| A4        | wt-doc   | 0.93 (53.26%) | 0.45 (36.73%) | 1.13 (59.63%) |
| A4compr1  | wt-bwt   | 0.42 (28.57%) | 0.44 (47.83%) | 0.43 (26.85%) |
| A4compr1  | Nodes    | 0.10 (7.12%)  | 0.03 (3.44%)  | 0.04 (2.39%)  |
| A4compr1  | B_r      | 0.00 (0.23%)  | 0.00 (0.12%)  | 0.00 (0.09%)  |
| A4compr1  | B_l      | 0.00 (0.23%)  | 0.00 (0.12%)  | 0.00 (0.08%)  |
| A4compr1  | wt-doc   | 0.93 (63.85%) | 0.45 (48.49%) | 1.13 (70.59%) |
| A4compr2  | wt-bwt   | 0.16 (13.03%) | 0.22 (31.01%) | 0.22 (15.62%) |
| A4compr2  | Nodes    | 0.10 (8.67%)  | 0.03 (4.55%)  | 0.04 (2.76%)  |
| A4compr2  | B_r      | 0.00 (0.28%)  | 0.00 (0.16%)  | 0.00 (0.10%)  |
| A4compr2  | B_l      | 0.00 (0.28%)  | 0.00 (0.16%)  | 0.00 (0.10%)  |
| A4compr2  | wt-doc   | 0.93 (77.74%) | 0.45 (64.11%) | 1.13 (81.42%) |

The first column shows the algorithm used in the experiment (the k-mer size is 50). The second column specifies the different data structures used: wt-bwt stands for the wavelet tree of the BWT (including rank and select support), nodes stands for the array of nodes (the implicit graph representation), B_r and B_l are the bit vectors described in “Computation of right-maximal k-mers and node identifiers” section (including rank support), and wt-doc stands for the wavelet tree of the document array. The remaining columns show the memory usage in bytes per base pair and, in parentheses, their percentage.

*Correspondence: Enno.ohlebusch@uni-ulm.de
Institute of Theoretical Computer Science, Ulm University, James-Franck-Ring O27/537, 89069 Ulm, Germany

© The Author(s) 2016. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Algorithm 2 Construction of the implicit compressed de Bruijn graph.

1: function CREATE-COMPRRESSED-GRAPH(k, BWT)
2: create an empty graph G
3: create an empty queue Q
4: \((B_r, B_l) \leftarrow \text{CREATE-BIT-VECTORS}(k, \text{BWT, } G, Q)\)
5: \(\text{rightMax} \leftarrow \text{rank}_1(B_r, n)/2\)
6: \(\text{leftMax} \leftarrow \text{rank}_3(B_l, n)\)
7: for \(s \leftarrow 1\) to \(d\) do \(\triangleright\) add the stop nodes for the \(d\) sequences
8: \(id \leftarrow \text{rightMax} + \text{leftMax} + s\)
9: \(G[id] \leftarrow (1, s, 1, s)\)
10: enqueue(Q, id)
11: \(B[s] \leftarrow 0\)
12: while Q is not empty do
13: \(id \leftarrow \text{dequeue}(Q)\)
14: repeat
15: \(\text{extendable} \leftarrow \text{false}\)
16: \(lb \leftarrow G[id].lb\)
17: \(rb \leftarrow lb + G[id].size - 1\)
18: \(\text{list} \leftarrow \text{getIntervals}(\text{lb}, \text{rb})\)
19: for each \((c, [i..j])\) in list do
20: \(\text{ones} \leftarrow \text{rank}_1(B_r, i)\)
21: if \(\text{ones} \text{ is even and } B_r[i] = 0\) then
22: if \(c \notin \{\#, \$\}\) then \(\triangleright\) Case 1
23: if list contains just one element then
24: \(\text{extendable} \leftarrow \text{true}\)
25: \(G[id].len \leftarrow G[id].len + 1\)
26: \(G[id].lb \leftarrow i\)
27: else \(\triangleright\) Case 2
28: \(\text{newId} \leftarrow \text{rightMax} + \text{rank}_1(B_l, i - 1) + 1\)
29: \(G[\text{newId}] \leftarrow (k, i, j - i + 1, i)\)
30: enqueue(Q, newId)
31: until not extendable

The online version of the original article can be found under doi:10.1186/s13015-016-0083-7.

Received: 9 November 2016   Accepted: 9 November 2016
Published online: 28 November 2016

Reference
1. Beller T, Ohlebusch E. A representation of a compressed de Bruijn graph for pan-genome analysis that enables search. Algorithms Mol Biol. 2016;11:20. doi:10.1186/s13015-016-0083-7.

Submit your next manuscript to BioMed Central and we will help you at every step:
• We accept pre-submission inquiries
• Our selector tool helps you to find the most relevant journal
• We provide round the clock customer support
• Convenient online submission
• Thorough peer review
• Inclusion in PubMed and all major indexing services
• Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit