New high-throughput sequencing technologies have promoted the production of short reads with dramatically low unit cost. The explosive growth of short read data poses a challenge to the mapping of short reads to reference genomes, such as the human genome, in terms of alignment quality and execution speed.

We present CUSHAW, a parallelized short read aligner based on the compute unified device architecture (CUDA) parallel programming model. We exploit CUDA-compatible graphics hardware as accelerators to achieve fast speed. Our algorithm employs a quality-aware bounded search approach based on the Burrows-Wheeler transform (BWT) and the Ferragina Manzini (FM)-index to reduce the search space and achieve high alignment quality. Performance evaluation, using simulated as well as real short read datasets, reveals that our algorithm running on one or two graphics processing units (GPUs) achieves significant speedups in terms of execution time, while yielding comparable or even better alignment quality for paired-end alignments compared to three popular BWT-based aligners: Bowtie, BWA and SOAP2.

CUSHAW, a parallelized short read aligner based on the CUDA programming model, exploits CUDA-compatible graphics hardware as accelerators to achieve fast speed. Our algorithm employs a quality-aware bounded search approach based on the Burrows-Wheeler transform (BWT) and the Ferragina Manzini (FM)-index to reduce the search space and achieve high alignment quality. Performance evaluation, using simulated as well as real short read datasets, reveals that our algorithm running on one or two graphics processing units (GPUs) achieves significant speedups in terms of execution time, while yielding comparable or even better alignment quality for paired-end alignments compared to three popular BWT-based aligners: Bowtie, BWA and SOAP2.

A Fast CUDA Compatible Short Read Aligner to Large Genomes

Yongchao Liu, Bertil Schmidt and Douglas L. Maskell

References


Performance Evaluation

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>CUSHAW</th>
<th>Bowtie</th>
<th>BWA</th>
<th>SOAP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR0002273</td>
<td>SE 92.58</td>
<td>94.69</td>
<td>90.26</td>
<td>94.85</td>
<td></td>
</tr>
<tr>
<td>SR001355</td>
<td>SE 97.72</td>
<td>92.42</td>
<td>94.80</td>
<td>91.09</td>
<td></td>
</tr>
<tr>
<td>ER000589</td>
<td>SE 94.76</td>
<td>96.51</td>
<td>94.06</td>
<td>96.17</td>
<td></td>
</tr>
<tr>
<td>ER0034966</td>
<td>SE 89.99</td>
<td>90.55</td>
<td>95.51</td>
<td>73.11</td>
<td></td>
</tr>
<tr>
<td>ER0024139</td>
<td>SE 89.99</td>
<td>90.55</td>
<td>95.51</td>
<td>73.11</td>
<td></td>
</tr>
</tbody>
</table>

A highest speedup of 3.1 (3.18) is achieved by CUSHAW on a single Tesla C2050 GPU over Bowtie on a single AMD 1.4 GHz CPU core for single-end (paired-end) alignments.

A highest speedup of 12 (14.5) is achieved by CUSHAW on a single Tesla C2050 GPU over BWA on a single AMD 1.4 GHz CPU core for single-end (paired-end) alignments.

A highest speedup of 7.9 (24.3) is achieved by CUSHAW on a single Tesla C2050 GPU over SOAP2 on a single AMD 1.4 GHz CPU core for single-end (paired-end) alignments.

Speedups over Bowtie

- Single-end alignment
- Paired-end alignment

Speedups over BWA

- Single-end alignment
- Paired-end alignment

Speedups over SOAP2

- Single-end alignment
- Paired-end alignment

Contact

Yongchao Liu: liyu@uni-mainz.de
Bertil Schmidt: bertil.schmidt@uni-mainz.de
Douglas L. Maskell: as douglas@ntu.edu.sg