

GateKeeper-GPU

Fast and Accurate Pre-Alignment Filtering in Short Read Mapping

Zülal Bingöl^{1,3}

zulal.bingol@bilkent.edu.tr

Mohammed Alser^{2,4}, Onur Mutlu^{2,4}, Ozcan Ozturk¹, Can Alkan^{1,3}



Bilkent University

2

ETH zürich

3



Alkanlab

4

SAFARI

Quantitative Biology > Genomics

[Submitted on 27 Mar 2021 (v1), last revised 31 Mar 2021 (this version, v2)]

GateKeeper-GPU: Fast and Accurate Pre-Alignment Filtering in Short Read Mapping

Zülal Bingöl, Mohammed Alser, Onur Mutlu, Ozcan Ozturk, Can Alkan

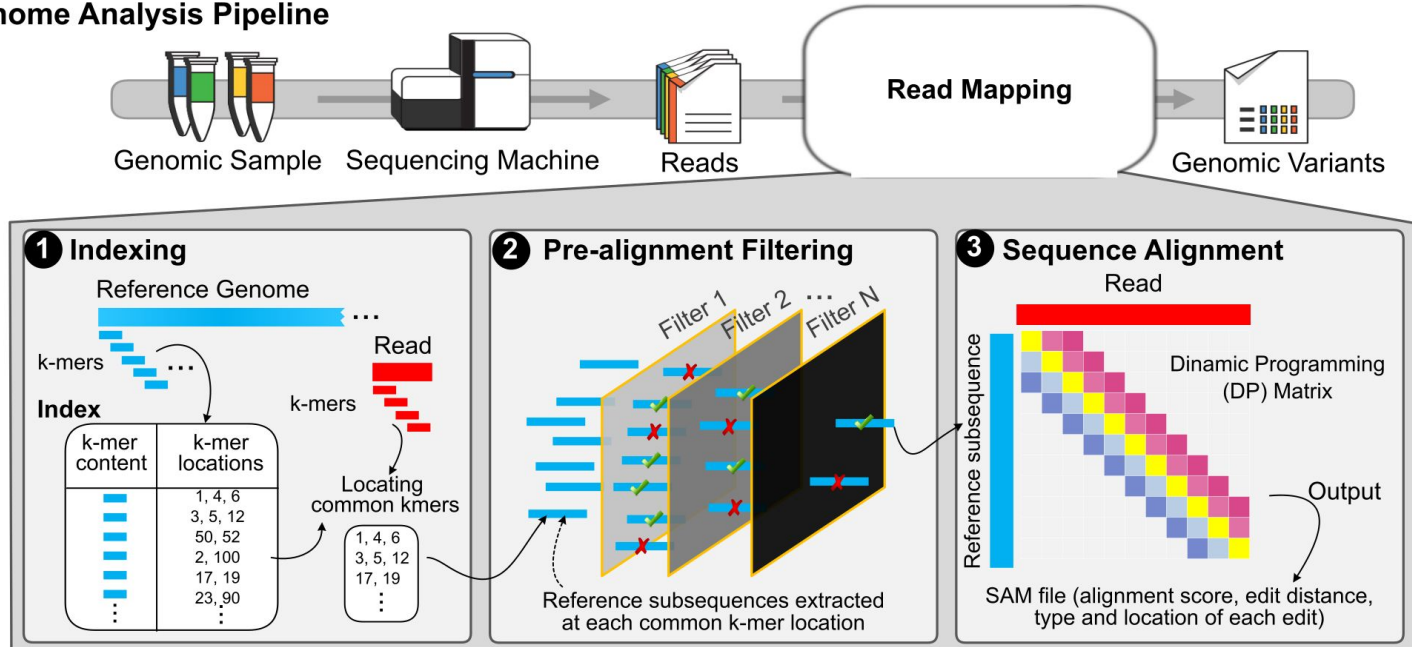
At the last step of short read mapping, the candidate locations of the reads on the reference genome are verified to compute their differences from the corresponding reference segments using sequence alignment algorithms. Calculating the similarities and differences between two sequences is still computationally expensive since approximate string matching techniques traditionally inherit dynamic programming algorithms with quadratic time and space complexity. We introduce GateKeeper-GPU, a fast and accurate pre-alignment filter that efficiently reduces the need for expensive sequence alignment. GateKeeper-GPU provides two main contributions: first, improving the filtering accuracy of GateKeeper(state-of-the-art lightweight pre-alignment filter), second, exploiting the massive parallelism provided by the large number of GPU threads of modern GPUs to examine numerous sequence pairs rapidly and concurrently. GateKeeper-GPU accelerates the sequence alignment by up to 2.9x and provides up to 1.4x speedup to the end-to-end execution time of a comprehensive read mapper (mrFAST). GateKeeper-GPU is available at [this https URL](#).

Preprint at
[arXiv:2103.14978](https://arxiv.org/abs/2103.14978)



Problem Statement

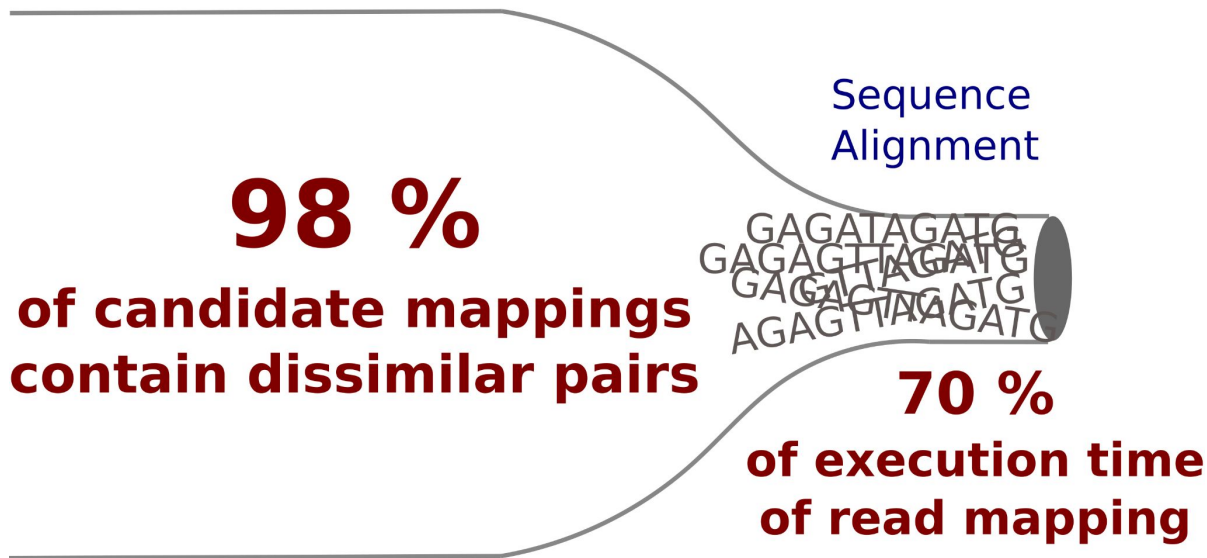
Genome Analysis Pipeline



(Alser et al., 2020)

Sequence alignment is a bottleneck

Read Mapping Pipeline



Time- and space- efficient sequence alignment is crucial

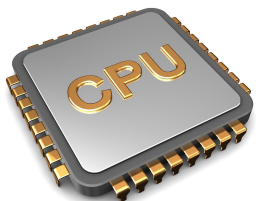
How can we resolve this bottleneck?

Reducing the need for sequence alignment

Our Aim:

Eliminate the candidate sequence pairs
that contain more edits than a predefined error threshold
with **a fast and accurate filter**

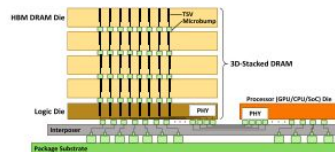
Hardware Accelerated Filter Design



CPU/SIMD: SHD
(Xin et al., 2015)



FPGA: GateKeeper
(Alser et al., 2017)



PIM: GRIM-Filter
(Kim et al., 2018)

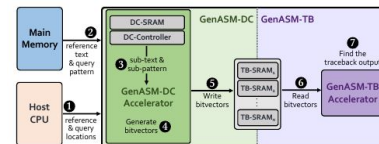


Figure 4. Overview of GenASM.

Specialized Hardware: GenASM
(Senol Cali et al., 2020)

- Filtering algorithm: **GateKeeper**
 - **Accurate**
 - Lightweight bitwise operations: **Suitable for hardware acceleration**
- Target Platform: **Graphics Processing Unit (GPU)**
 - **Ideal** for throughput-critical work
 - Economically **more viable**
 - GPU codebase is more **flexible**

GateKeeper-GPU

- Improved GateKeeper algorithm with **better accuracy**
- Adapted it to **GPU with CUDA framework: High level of data parallelism**
 - Intra-sequence processing
 - Each CUDA thread retrieves one candidate reference segment and read pair
 - Single GPU thread → One GateKeeper Operation
 - Inter-sequence parallelism
 - Many CUDA blocks are executed at once
 - Multi-GPU support
- Integrated **with full short read mapper** - mrFAST (*Alkan et al., 2009*)
- Provided **comprehensive analyses** using two different GPU architectures

Key Results: Filtering Accuracy

mrFAST Candidate Mappings



Sequence pairs with read length 100bp.

GateKeeper-GPU can reject **more than 90%** of false candidate mappings
with **less than 10%** false accept ratio

Key Results: Filtering Accuracy (Cont'd.)

minimap2 Candidate Mappings

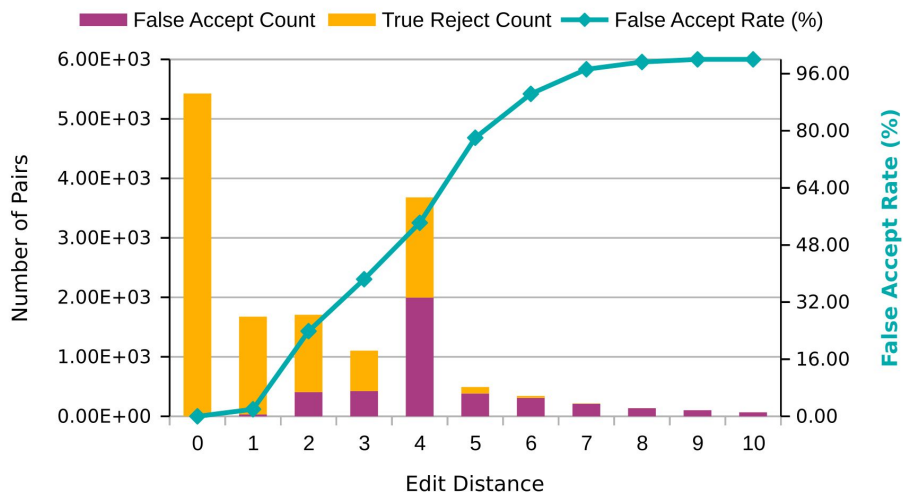


Sequence pairs with read length 100bp.

GateKeeper-GPU can eliminate
all dissimilar candidate mappings in **exact matching**

Key Results: Filtering Accuracy (Cont'd.)

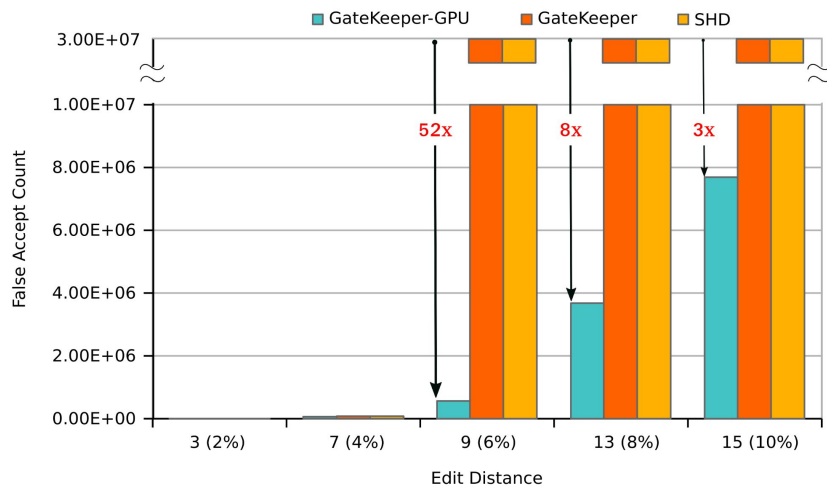
BWA-MEM Candidate Mappings



Sequence pairs with read length 100bp.

GateKeeper-GPU can eliminate
up to 98% of false candidate mappings

Key Results: Filtering Accuracy (Cont'd.)

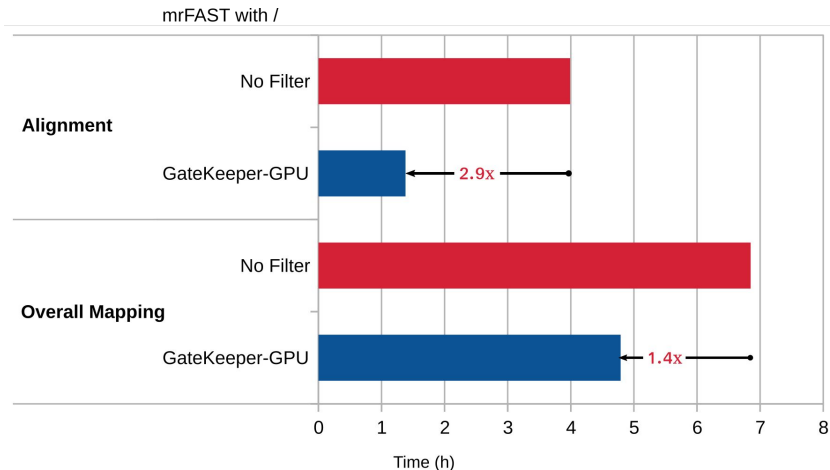


Sequence pairs with read length 150bp.

GateKeeper-GPU produces up to **52x less** false accepted pairs compared to GateKeeper

Key Results: Performance

Integrated into **mrFAST** (Alkan et al., 2009) with **single GPU**



Time taken for mrFAST procedures for read length 100bp.

GateKeeper-GPU can accelerate **alignment** step up to **2.9x**

overall mapping procedure up to **1.4x**

Executive Summary

- **Problem:** Comparing sequence pairs in read mapping is a **computationally-costly** procedure.
- **Observation:** Applying dynamic programming algorithms for the alignment step creates a **bottleneck for read mapping pipeline**.
- **Goal:** Reducing the workload on sequence alignment with **a fast and accurate pre-alignment filter**
- **Key Results:**
 - **Significantly reducing the number of sequence pairs** to be compared with dynamic programming algorithms
 - **Correctly eliminating more** sequence pairs than GateKeeper.
 - Providing **2.9x speedup** to alignment stage, and **1.4x speedup** to overall mapping procedure.

GateKeeper-GPU

Fast and Accurate Pre-Alignment Filtering in Short Read Mapping

Zülal Bingöl^{1,3}

zulal.bingol@bilkent.edu.tr

Mohammed Alser^{2,4}, Onur Mutlu^{2,4}, Ozcan Ozturk¹, Can Alkan^{1,3}



Bilkent University

2

ETH zürich

3



Alkanlab

4

SAFARI



Preprint at
[arXiv:2103.14978](https://arxiv.org/abs/2103.14978)



Source Code at Github:
[BilkentCompGen/GateKeeper-GPU](https://github.com/BilkentCompGen/GateKeeper-GPU)