Poster 118

FastHASH:
A New Algorithm for Fast and Comprehensive Next-Generation Sequence Mapping

Hongyi Xin[†] Donghyuk Lee[†]
Farhad Hormozdiari[‡] Can Alkan[§] Onur Mutlu[†]

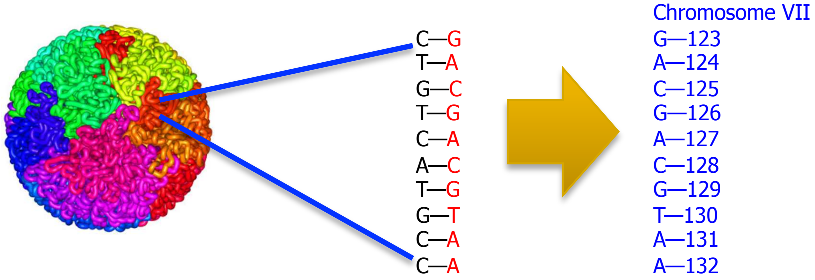
[†] **Carnegie Mellon University** [§] **University of Washington**
[‡] **University of California Los Angeles**

SAFARI Carnegie Mellon

Next-Generation DNA Sequencing (NGS)

- DNA sequencing is important

- Basic Approach: Read and map



C—G
T—A
G—C
T—G
C—A
A—C
T—G
G—T
C—A
C—A

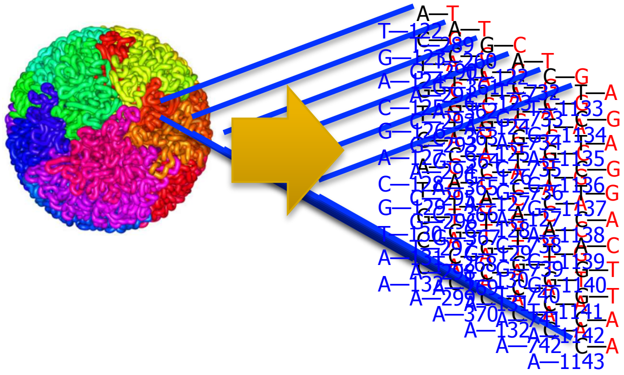
Chromosome VII
G—123
A—124
C—125
G—126
A—127
C—128
G—129
T—130
A—131
A—132

- Next-generation sequencing technologies (NGS) produce many short DNA fragments
    - More computationally intensive
    - Harder to map a fragment to entire genome
    - Especially when allowing **polymorphisms**

- Goal: Design fast and comprehensive algorithms to analyze enormous amounts of NGS data

Challenge of Existing NGS Mapping Tools

- We want a tool that is both Fast and Comprehensive



→ need a fast tool

→ need a comprehensive tool

Want to find all possible locations with some mutations allowed

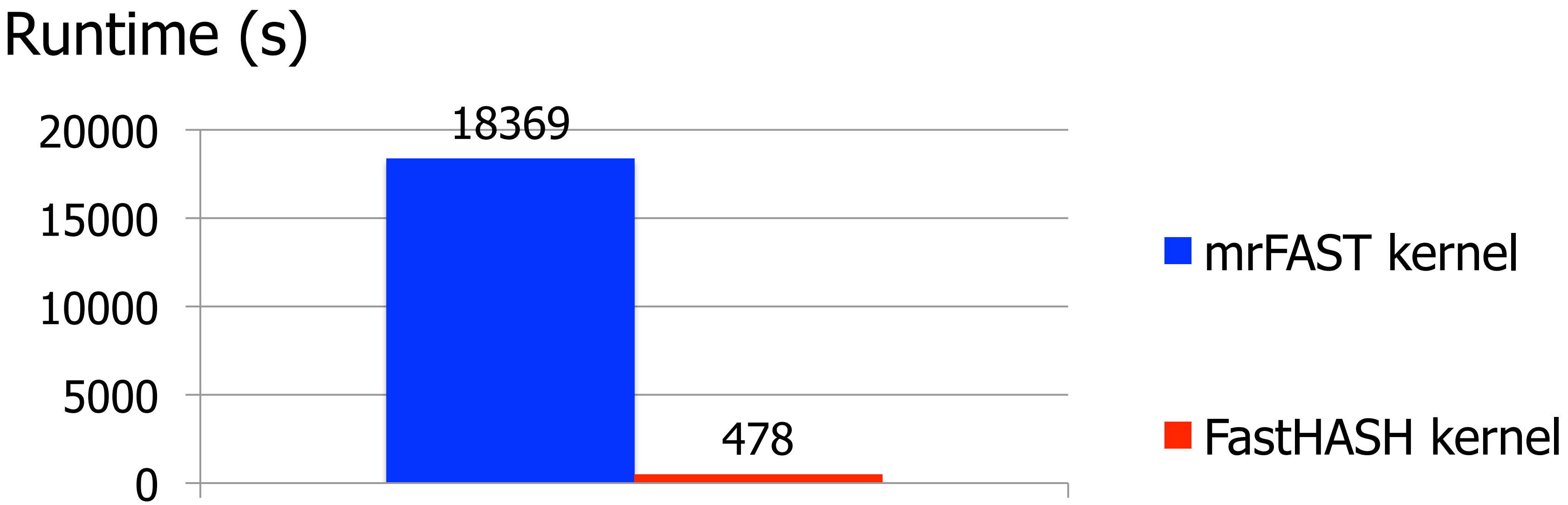
- Some tools are **only fast**
  - BWA, SOAPv2, Bowtie
  - Lose mapping

- Some tools are **only comprehensive**
  - mrFAST, mrsFAST
  - Slow

FastHASH NGS Mapping Kernel

- **Goal**
  - Best of both worlds: Fast and Comprehensive
- **Observation**
  - Some tools do unnecessary work to guarantee comprehensiveness → main cause of slowness
- **Main Idea of FastHASH**
  - Cut down unnecessary work by exploiting knowledge of reference genome

Evaluation



- 38x speedup compared to state-of-the-art comprehensive tool
  - Without sacrificing comprehensiveness

- Also implemented on GPU for further acceleration

Poster 118

FastHASH:
A New Algorithm for Fast and Comprehensive Next-Generation Sequence Mapping

Hongyi Xin[†] Donghyuk Lee[†]
Farhad Hormozdiari[‡] Can Alkan[§] Onur Mutlu[†]

[†] **Carnegie Mellon University** [§] **University of Washington**
[‡] **University of California Los Angeles**

SAFARI Carnegie Mellon