ALI GHAFFAARI

0171 9499741 ali@ghaffaari.me cartoonist alighaffaari ali.ghaffaari.me

PROFILE

Innovative computer scientist with an extensive background in computer engineering, expertise in algorithm design and engineering, skilled in developing scalable, high-performance solutions for complex computational challenges on parallel and distributed systems.

SELECTED PROJECTS

PAIRED-END SHORT READ MAPPING TO SEQUENCE GRAPHS | `GraphAligner` ON GOING

- Tackling the unreliability of current methods in complex and repetitive genomic regions
- Higher accuracy and speed in preliminary results compared to the state-of-the-art

FAST AND SCALABLE DISTANCE VALIDATION IN SEQUENCE GRAPHS | `diverg` Sep 2024

- Scalable index construction on **CUDA** and **OpenMP** capable of handling very large graphs
- **Up to 340x** smaller index size using a novel sparse matrix representation
- Faster construction time, and **2.5–4x speed-up** in query time

FULLY-SENSITIVE SEED FINDING INDEX FOR SEQUENCE GRAPHS | `psi` Jan 2019

- The **first fully sensitive** method for fixed-length query searches in sequence graphs
- **Increased sensitivity from 6.24% to 100%** while maintaining the same index size and query time compared to state-of-the-art methods

EXPERIENCE

BIELEFELD UNIVERSITY | *Postdoctoral researcher* Bielefeld

Genome Data Science Group Apr 2021 – Dec 2024

- **Project Manager** | Algorithms for Pangenome Computational Analysis (ALPACA)
- Devised and implemented a scalable method for distance validation of paired-end alignments in sequence graphs on **CUDA** and **OpenMP** backends

HEINRICH HEINE UNIVERSITY DÜSSELDORF | *Research Fellow* Düsseldorf

Institute for Medical Biometry and Bioinformatics Apr 2020 – Mar 2021

- Led the development of a fast and novel approach for paired-end short-read mapping to sequence graphs (implemented in C++)

MAX PLANCK INSTITUTE FOR INFORMATICS | *Research Fellow* Saarbrücken

Department of Algorithms for Computational Genomics Apr 2015 – Mar 2020

- Devised a **full sensitive** approach for seed finding in sequence graphs using a **cache-oblivious** hybrid index (implemented in C++)

INSTITUTE FOR RESEARCH IN FUNDAMENTAL SCIENCES (IPM) | *Student Researcher* Tehran

Department of Computer Science, HPC Laboratory Sep 2009 – Apr 2013

- Research on HPC technologies such as **GPGPU, IBM Cell/B.E., OpenMPI**, as well as cluster and grid computing to develop high-performance scientific applications
- Led and contributed in diverse projects from computational biology to machine learning

Saarland University | *Part-time HPC Cluster Administrator* Saarbrücken
Department of Spoken Language Systems 2016 – 2017

- Configured and maintained the HPC cluster system (SGE) for better resource management
- Linux system administration

GenapSys Inc. | *Remote Working Student* Redwood City, CA
May 2013 – Aug 2013

- Developed a full-stack solution on Google Cloud for DNA sequencing and analysis
- Implementation of algorithms for DNA base calling with **real-time constraints**
- System-level implementation of **high-throughput data communication** over network

University of Tehran | *Head Teaching Assistant* Tehran
Computer Network Laboratory WS 2010/2011
Operating Systems Laboratory SS 2010
Operating Systems SS 2010

EDUCATION

HHU Düsseldorf | *PhD in Computer Science* Apr 2015 – Mar 2021

- *expected to defend in 2025*
- **Thesis** | Indexing Schemes for Short-Read Mapping to Pangenome Graphs
- **Focus** | Algorithm design to address computational challenges in *pangenomic* analyses

University of Tehran | *M.Sc. in Computer Engineering* Sep 2011 – Sep 2014

- Ranked in the top 5 for GPA (≡ 4/4)

University of Tehran | *B.Sc. in Computer Engineering* Sep 2006 – Sep 2011

SKILLS

Programming *Proficient in C, C++, Python* | Rust, Elisp, R, Ruby, Java, JavaScript, x86 Assembly

Technical GNU/Linux development tools including Git, CMake, GNU Autotools | mySQL, Redis

Libraries CUDA, Kokkos, OpenMPI, SeqAn, Boost, Qt

DevOps Linux administration, CI/CD (Travis CI, GitHub), Docker

Workflow Snakemake, Nextflow

Laboratory Wet lab experience sequencing strains of L. elongisporus using ONT MinION

Soft Skills Problem solving, creativity, teamwork, presentation, project management

Language Fluent in English, Basic German, Native in Persian

Other System-level and hardware-aware programming, character and digital graphic design

HONOURS AND AWARDS

Top 0.1% (525 out of ca. 500,000 participants) The Nationwide University Entrance Exam 2006

Outstanding Student Award Tehran Scientific Fair, Tehran Central Dept. of Education 2005

Ranked 1st and 8th Tehran RoboCup Programming Contest, in Round 1 and 2 Sep, Jul 2004

PUBLICATIONS

[1] **A. Ghaffaari**, T. Marschall, and A. Schoenhuth. “DiVerG: Scalable Distance Index for Validation of Paired-End Alignments in Sequence Graphs”. Genome Biology (under review). 2024.

[2] J. M. Eizenga, A. M. Novak, J. A. Sibbesen, S. Heumos, **A. Ghaffaari**, et al. “Pangenome Graphs”. In: *Annual Review of Genomics and Human Genetics* 21.1 (May 2020). DOI: `10.1146/annurev-genom-120219-080406`.

[3] **A. Ghaffaari** and T. Marschall. “Fully-sensitive seed finding in sequence graphs using a hybrid index”. In: *Bioinformatics (Proceedings of ISMB)* 35.14 (July 2019), pp. i81–i89. DOI: `10.1093/bioinformatics/btz341`.

[4] T. Marschall, M. Marz, T. Abeel, L. Dijkstra, B. E. Dutilh, **A. Ghaffaari**, et al. “Computational Pan-Genomics: Status, Promises and Challenges”. In: *Briefings in Bioinformatics* (2016). DOI: `10.1101/043430`.