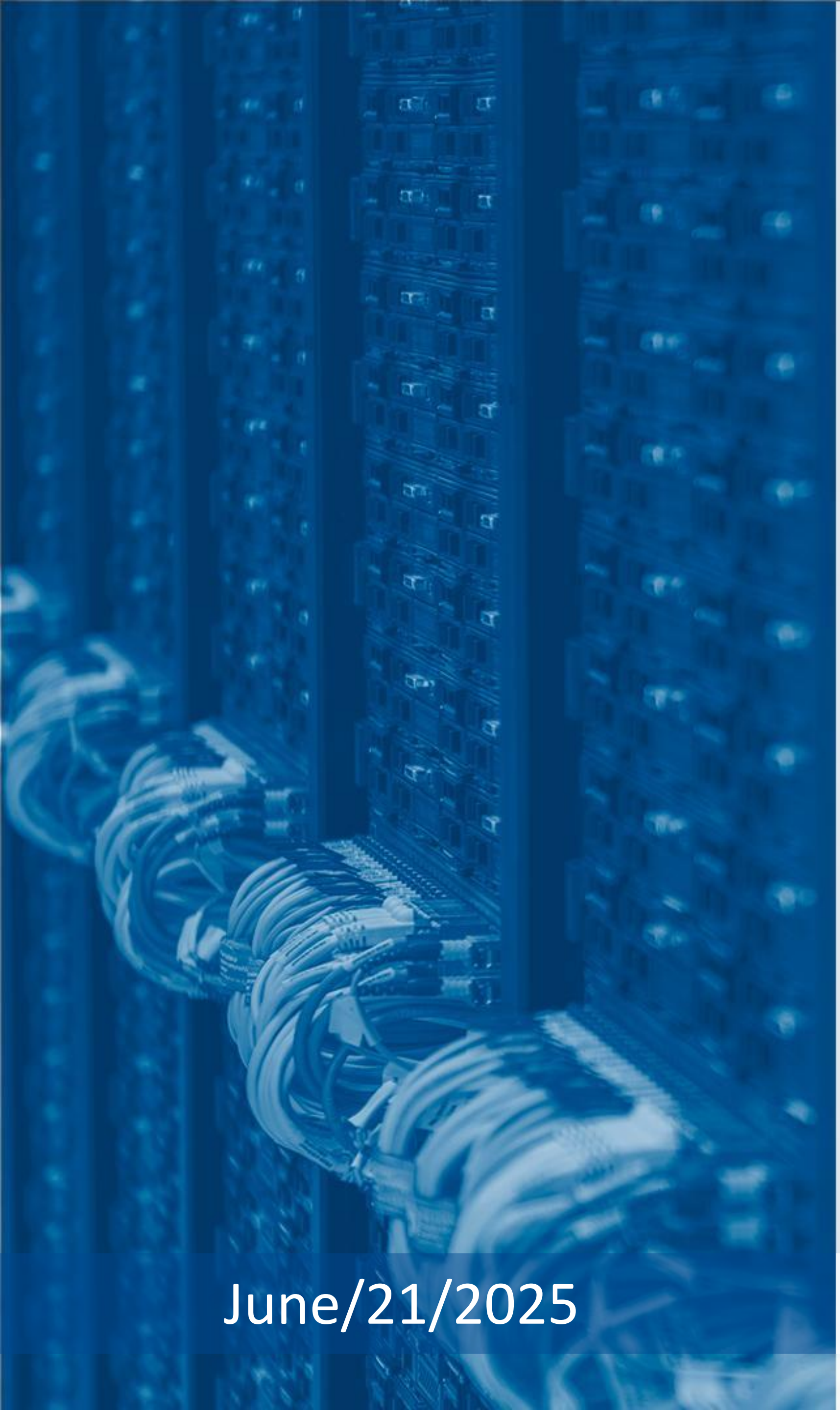


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

On the Use of Fujitsu A64FX Processor for Genome Sequence Analysis Workloads

Julian Pavon, Ivan Vargas Valdivieso Osman Unsal and Adrian Cristal

June/21/2025

Barcelona Supercomputing Center

Genome sequence analysis

The declining cost and rising throughput of genome sequencing technologies, combined with the availability of complete reference genomes, are driving breakthroughs in:

- Personalized medicine
- Evolutionary biology
- Forensic science

Genome sequence analysis

The declining cost and rising throughput of genome sequencing technologies, combined with the availability of complete reference genomes, are driving breakthroughs in:

- Personalized medicine
- Evolutionary biology
- Forensic science

However, while sequencing produces longer and more accurate reads, the computational cost of processing these sequences is growing significantly.

Opportunities for ARM HPC systems

To meet the rising performance and energy-efficiency demands, modern genomic platforms increasingly rely on hardware–software co-design and advanced instruction set architectures.

Opportunities for ARM HPC systems

To meet the rising performance and energy-efficiency demands, modern genomic platforms increasingly rely on hardware–software co-design and advanced instruction set architectures.

With the shift toward longer sequencing reads, memory subsystems must evolve to handle growing data volume and access complexity.

Opportunities for ARM HPC systems

To meet the rising performance and energy-efficiency demands, modern genomic platforms increasingly rely on hardware–software co-design and advanced instruction set architectures.

With the shift toward longer sequencing reads, memory subsystems must evolve to handle growing data volume and access complexity.

A notable example is Fujitsu's A64FX processor, the first Arm-based CPU to power a top-ranked supercomputer. Featuring:

- High memory bandwidth
- Scalable Vector Extension (SVE)

These capabilities make A64FX a powerful platform for genomics workloads with high data-level parallelism.

Opportunities for ARM HPC systems

In this work, we explore the performance of the Arm ecosystem for genome sequence analysis by evaluating widely-used sequence alignment and edit distance algorithms on the A64FX processor.

Opportunities for ARM HPC systems

In this work, we explore the performance of the Arm ecosystem for genome sequence analysis by evaluating widely-used sequence alignment and edit distance algorithms on the A64FX processor.

To maintain generality, we used the GNU Compiler Collection (GCC), we investigate how effectively current toolchains support parallelism and vectorization.

Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Opportunities for ARM HPC systems

In this work, we explore the performance of the Arm ecosystem for genome sequence analysis by evaluating widely-used sequence alignment and edit distance algorithms on the A64FX processor.

To maintain generality, we used the GNU Compiler Collection (GCC), we investigate how effectively current toolchains support parallelism and vectorization.

We compared A64FX's performance and efficiency to a baseline server class x86 system.

Evaluation methodology

CPU platforms:

1. Fujitsu A64FX processor with 48 cores and 32GB HBM2 (referred as a64fx).
2. A server class Intel Xeon Gold 5318N with 24 (48) cores (threads) processor (Cascade Lake architecture) with 192GB DDR4-2667 (referred as Xeon).

Evaluation methodology

CPU platforms:

1. Fujitsu A64FX processor with 48 cores and 32GB HBM2 (referred as a64fx).
2. A server class Intel Xeon Gold 5318N with 24 (48) cores (threads) processor (Cascade Lake architecture) with 192GB DDR4-2667 (referred as Xeon).

Benchmarks:

Modern aligners - Wavefront Align [1] and BiWavefront Align [2]

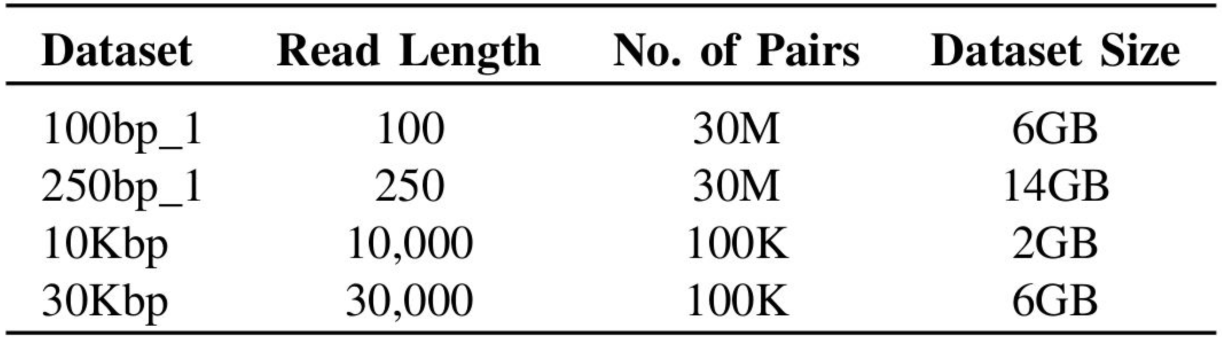
Edit distance approximation - SneakySnake [3]

End-to-end framework - Minimap2 [4]

[1] Marco-sola et al., Fast gap-affine pairwise alignment using the wavefront algorithm, Bioinformatics, 2020.
[2] Marco-sola et al. Optimal gap-affine alignment in O (s) space. Bioinformatics, 2023, vol. 39, no 2, p. btad074.
[3] Alser et al., SneakySnake: a fast and accurate universal genome pre-alignment filter for CPUs, GPUs and FPGAs. Bioinformatics, 2020, vol. 36, no 22-23, p. 5282-5290.
[4] Li, Heng. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 2018, vol. 34, no 18, p. 3094-3100.

Evaluation methodology

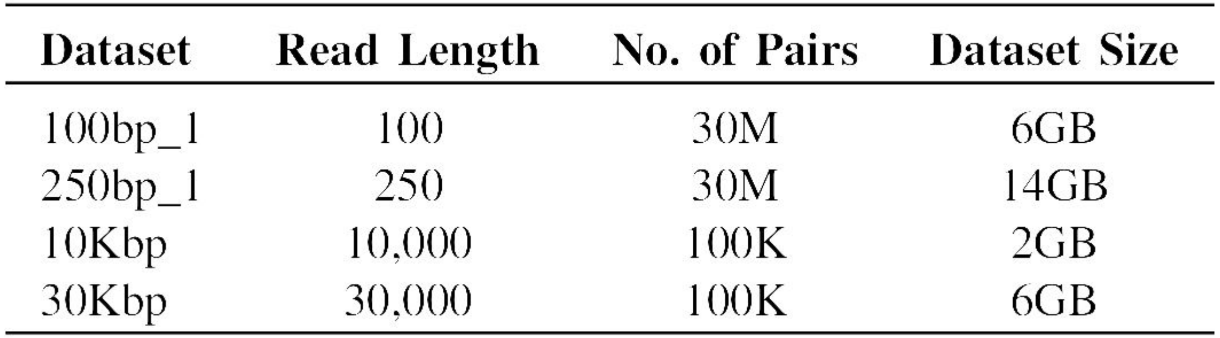
Input datasets:



Dataset	Read Length	No. of Pairs	Dataset Size
100bp_1	100	30M	6GB
250bp_1	250	30M	14GB
10Kbp	10,000	100K	2GB
30Kbp	30,000	100K	6GB

Evaluation methodology

Input datasets:



Dataset	Read Length	No. of Pairs	Dataset Size
100bp_1	100	30M	6GB
250bp_1	250	30M	14GB
10Kbp	10,000	100K	2GB
30Kbp	30,000	100K	6GB

For the end-to-end framework evaluation we used the reference and long input sequences provided by the minimap cookbook repository.

We evaluate the end-to-end execution of minimap2 going from index creation, through seeding and sequence alignment.

Evaluation methodology

Experiments:

Scalability

In each experiment, we utilize up to 24 threads, pinning each thread to a dedicated core.

To evaluate scalability, we begin with a single thread and increment the thread count in steps of four, up to the maximum of 24 threads.

Evaluation methodology

Experiments:

Scalability

In each experiment, we utilize up to 24 threads, pinning each thread to a dedicated core.

To evaluate scalability, we begin with a single thread and increment the thread count in steps of four, up to the maximum of 24 threads.

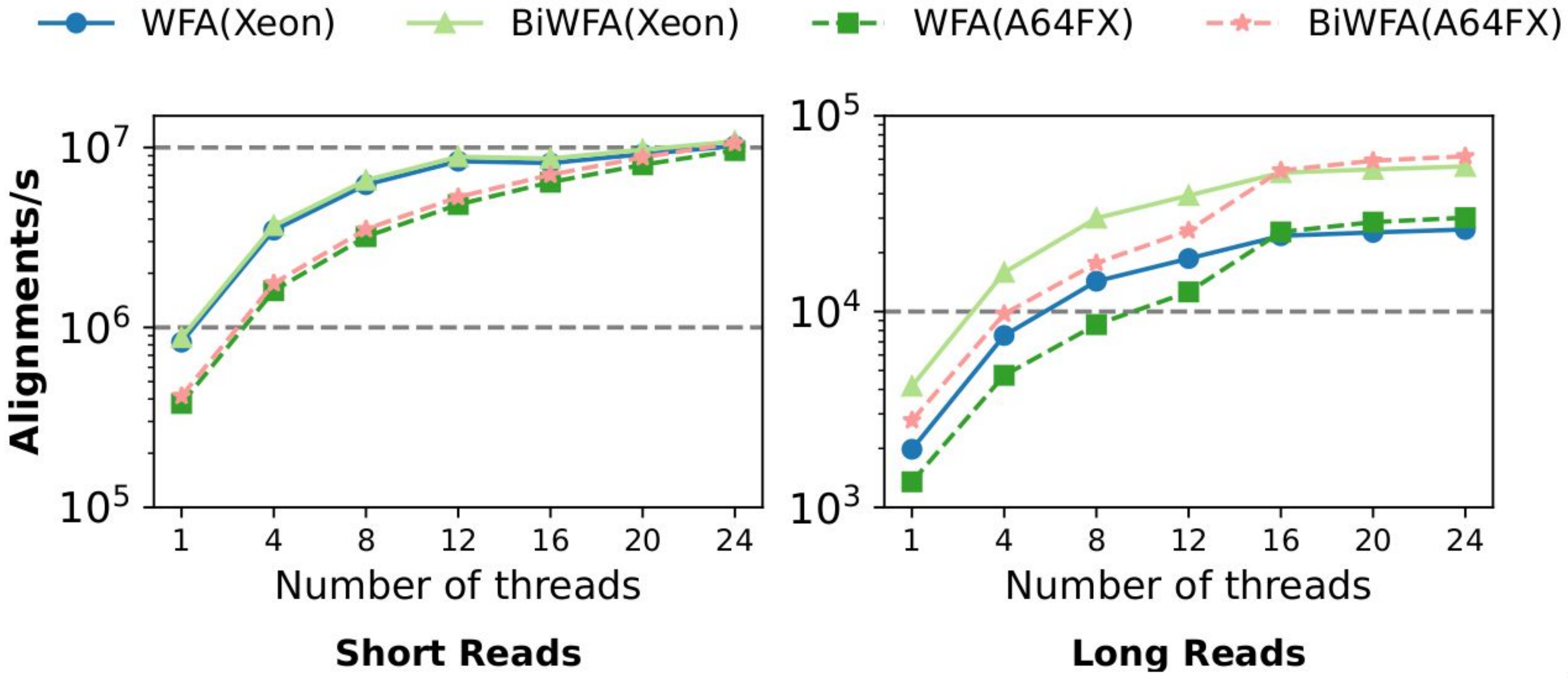
Vectorization

All evaluated benchmarks include optimized implementations that leverage the SIMD capabilities of both x86 (AVX512) and Arm (NEON) CPUs.

However, the A64FX processor supports the SVE ISA. Thus, we evaluate the performance impact of incorporating SVE optimizations across all benchmarks.

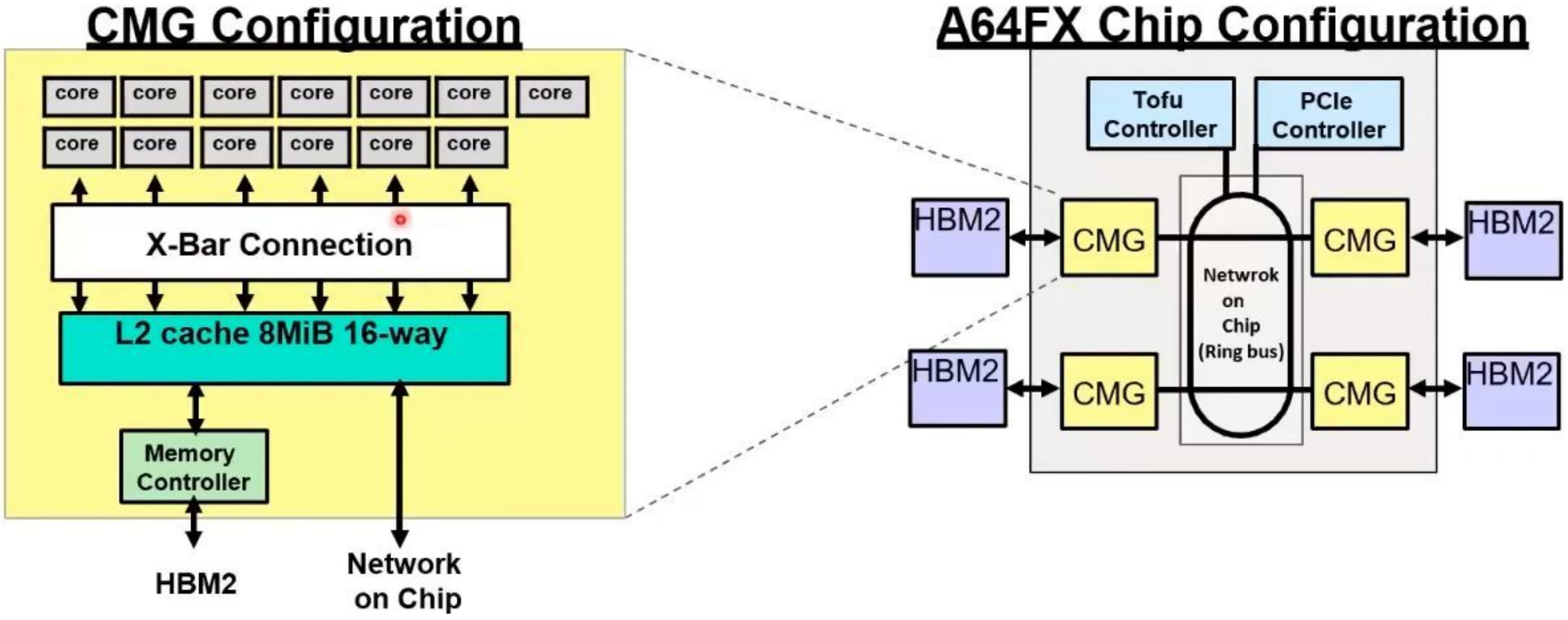
Scalability results

Modern aligners



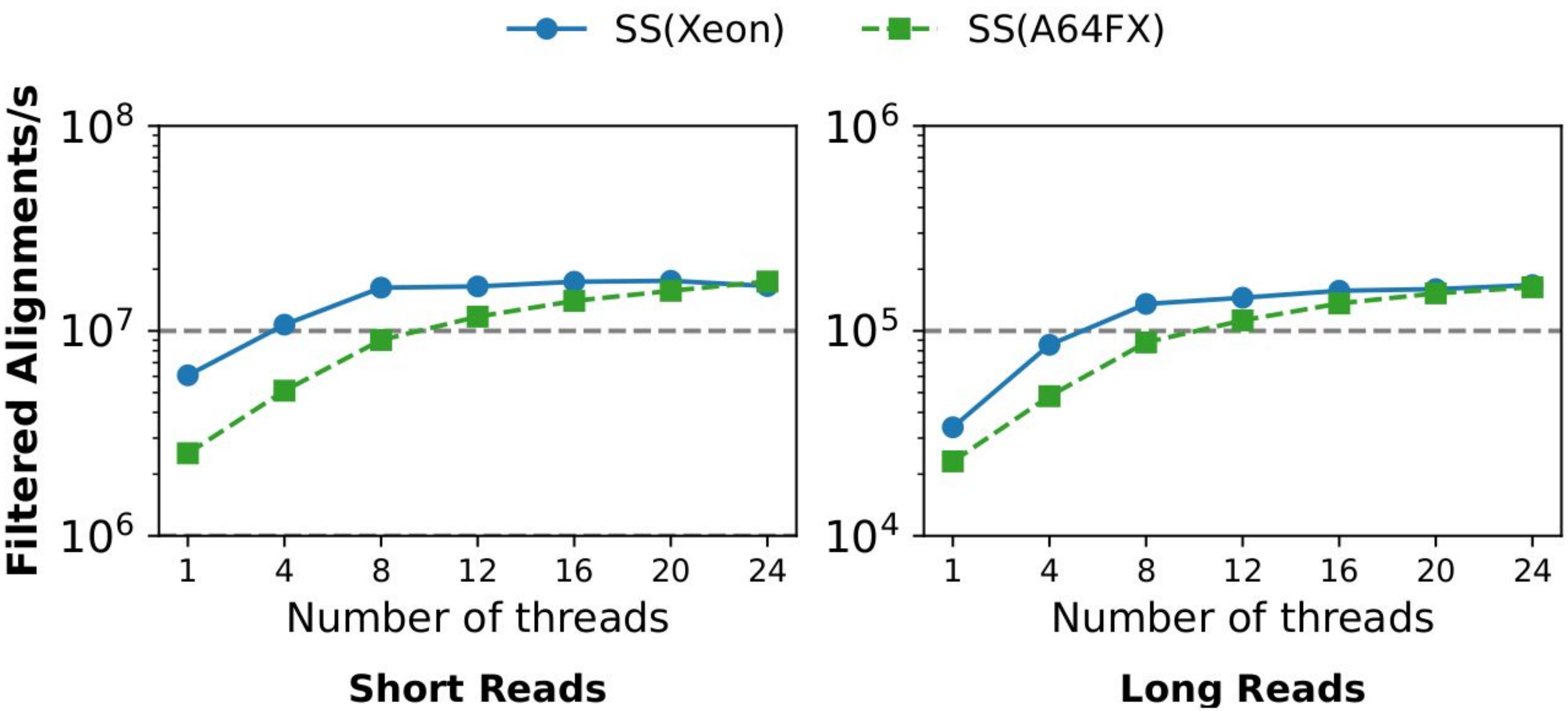
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Scalability results



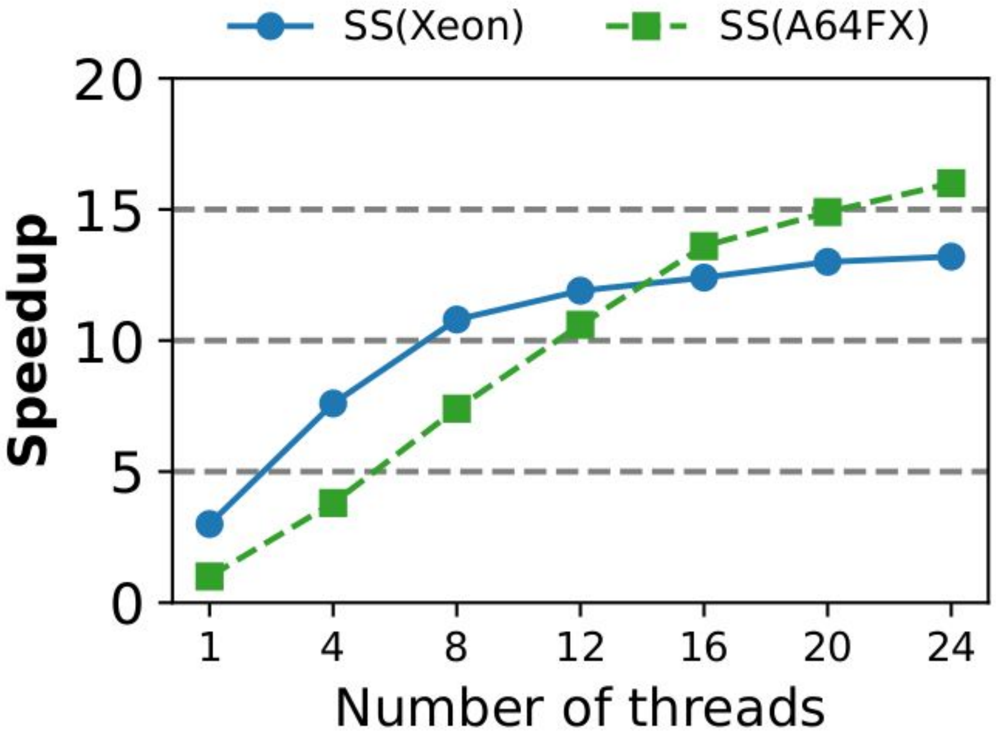
Scalability results

Edit distance



Scalability results

End-to-end framework



Minimap2

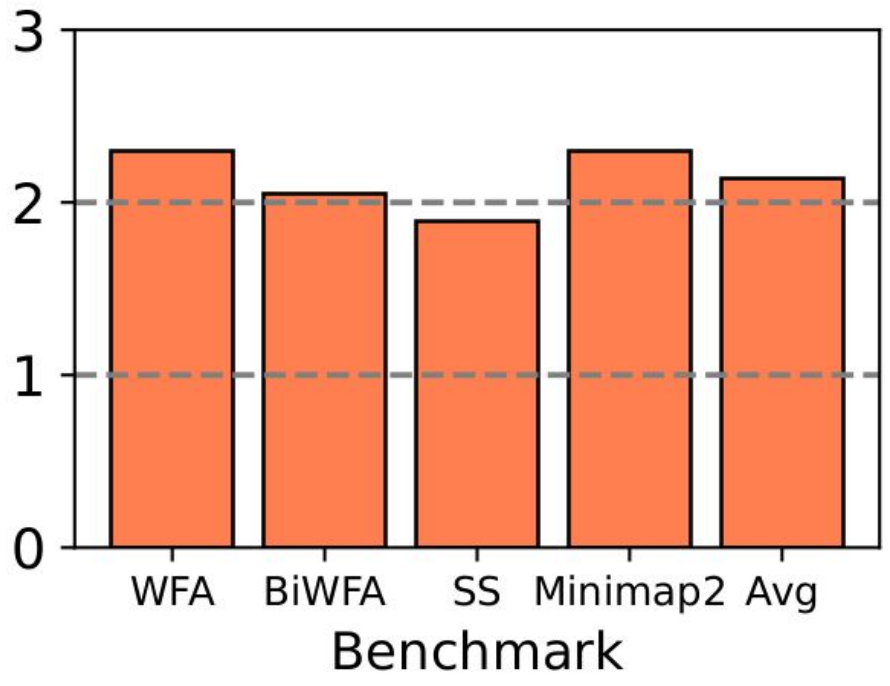
Impact of vectorization

In this experiment we compared ARM NEON vs ARM SVE.

All experiments used 24 threads.

On average, SVE outperforms by 2.1x to NEON.

This results come from the larger vector length and more sophisticated instructions, e.g., predication.



Vectorization

Conclusions and future work

A64FX outperforms Xeon in multi-threaded bioinformatics workloads, thanks to:

- High-bandwidth HBM2
- Efficient cache hierarchy
- Superior scalability

Xeon leads in single-threaded tasks, but suffers from memory bandwidth bottlenecks under parallel workloads.

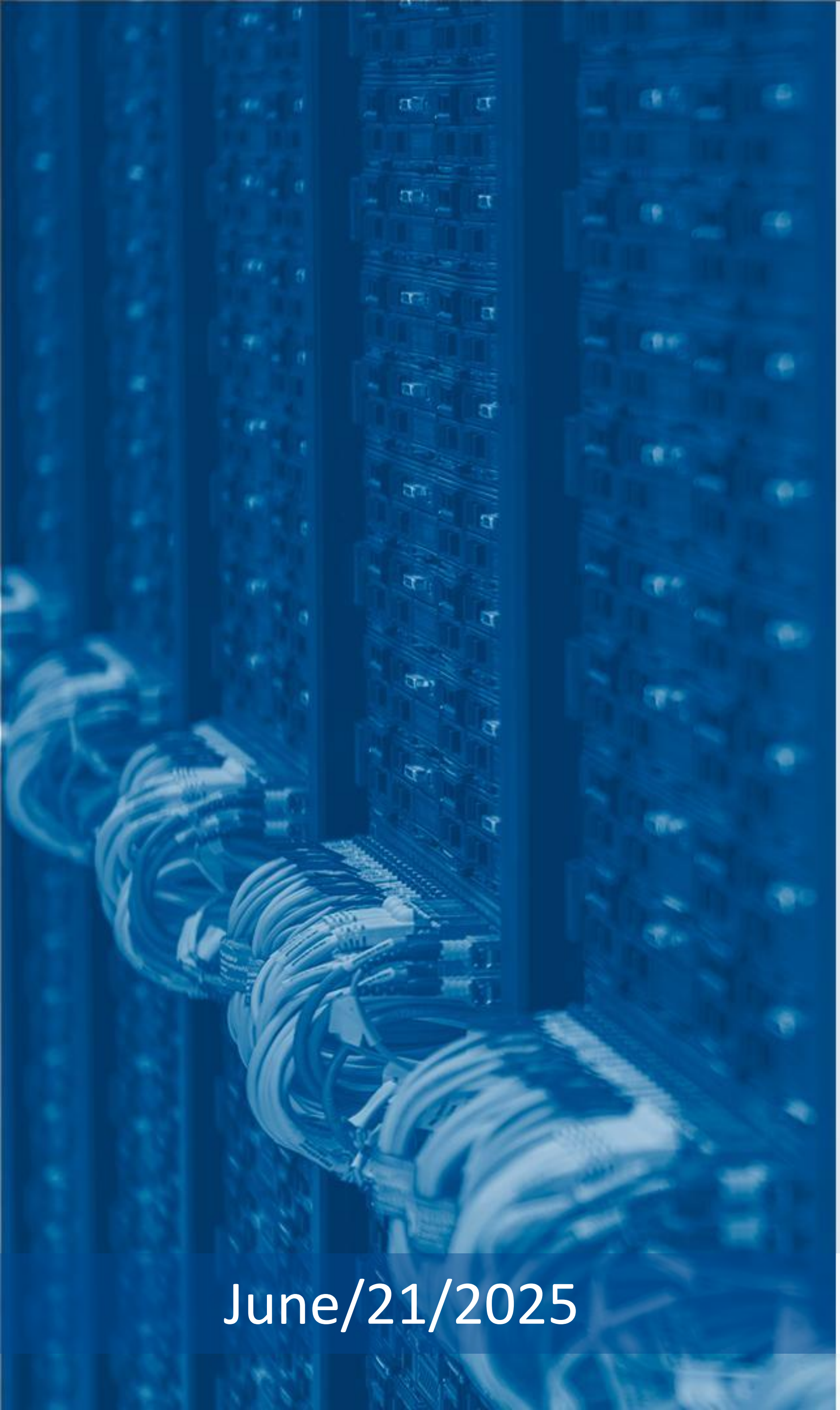
SVE vectorization delivers an average 2.1× performance boost, highlighting the importance of ISA-aware optimizations for ARM platforms.

Conclusions and future work

Further bioinformatics pipeline tuning for SVE.

Exploration of Arm Neoverse platforms for scalable genomics.

Evaluation of NVIDIA Grace Hopper Architectures.



On the Use of Fujitsu A64FX Processor for Genome Sequence Analysis Workloads

Julian Pavon, Ivan Vargas Valdivieso Osman Unsal and Adrian Cristal

June/21/2025

Barcelona Supercomputing Center