

April 17, 2020 | By Jan Zverina

UC San Diego Researchers Optimize Microbiome Tool for Computer GPUs

Microbiome analyses including COVID-19 research made significantly faster

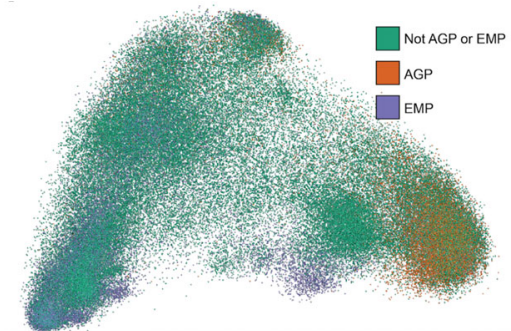
Researchers at the University of California San Diego have been applying their high-performance computing expertise by porting the popular UniFrac microbiome tool to graphic processing units (GPUs) in a bid to increase the acceleration and accuracy of scientific discovery, including urgently needed COVID-19 research.

“Our initial results exceeded our most optimistic expectations,” said Igor Sfiligoi, lead scientific software developer for high-throughput computing at the San Diego Supercomputer Center (SDSC) at UC San Diego.

“As a test we selected a computational challenge that we previously measured as requiring some 900 hours of time using server class CPUs, or about 13,000 CPU core hours. We found that it could be finished in just 8 hours on a single NVIDIA Tesla V100 GPU, or about 30 minutes if using 16 GPUs, which could reduce analysis runtimes by several orders of magnitude. A workstation-class NVIDIA RTX 2080TI would finish it in about 12 hours.”

“The new executable will also be of tremendous value for exploratory work, as the moderate-sized EMP dataset that used to require 13 hours on a server class CPU can now be run in just over one hour on a laptop containing a mobile NVIDIA GTX 1050 GPU,” added Sfiligoi.

Sfiligoi has been collaborating with Rob Knight, founding director of the Center for Microbiome Innovation, and a professor of Pediatrics, Bioengineering, and Computer Science & Engineering at the university, and Daniel McDonald, scientific director of the American Gut



An ordination plot of unweighted UniFrac distances over 113,721 samples. AGP: American Gut Project. EMP: Earth Microbiome Project. Credit: Rob Knight, Daniel McDonald, UC San Diego

Project. Microbiomes are the combined genetic material of the microorganisms in a particular environment, including the human body.

“This work did not initially begin as part of the COVID-19 response,” said Sfiligoi. “We started the discussion about such a speed-up well before, but UniFrac is an essential part of the COVID-19 research pipeline.”

UniFrac compares microbiomes to one another using an evolutionary tree that relates the DNA sequences to each other. “UniFrac played a key role in the Human Microbiome Project, allowing us to understand how microbes are related across our bodies, and in the Earth Microbiome Project, allowing us to understand how microbes are related across our planet,” said Knight. “We are using it to understand how a person’s microbiome might make them more or less susceptible to COVID-19, and what microbes in environments ranging from health care facilities to sewage to ocean spray make the environment more or less hospitable to SARS-CoV-2, the coronavirus that causes COVID-19.”

Knight noted that Sfiligoi had sped up the latest version of the algorithm, published less than two years ago in *Nature Methods*, which itself already represented a dramatic speed improvement over previous implementations.

“As microbial sequence data increase exponentially, from dozens of sequences to billions, we have to re-implement all the algorithms,” he said. “This latest step really shows how optimizing the research infrastructure can dramatically reduce time-to-result while preserving the accuracy of the findings and enabling completely new scales of questions to be asked.”

Specifically, Sfiligoi used [OpenACC](#), a user-driven, directive-based parallel programming model to port the existing Striped UniFrac implementation to GPUs because this allows a single codebase for both CPU and GPU code. Additional speedup was obtained by carefully exploiting cache locality. Also explored was the use of lower-precision floating point math to effectively exploit consumer-grade GPUs typically found in desktop and laptop computers.

UniFrac was originally designed and always implemented using higher precision floating point math, often called fp64 code path. The higher-precision floating point math was used to maximize reliability of the results. After implementing the lower-precision floating point math, usually called fp32 code path, researchers observed nearly identical results, but with significantly shorter compute times.

“We saw a 3x speed-up in the fp32 code path for gaming GPUs such as the 2080 Ti and the mobile 1050, and we believe that precision should be adequate for the vast majority of studies,” explained Sfiligoi.

Moreover, the code changes introduced to speed up GPU computation also significantly sped up the execution on CPU resources. The computational challenge mentioned above can now be completed in about 200 hours on the same server-class CPU, a 4x speedup, according to the researchers.

“Making computation available on GPU-enabled personal devices, even laptops, eliminates a large barrier within the resource infrastructure for many scientists,” said Sfiligoi.

More about Sfiligoi’s recent groundbreaking research on GPU cloudbursting [can be found here](#).

This work was sponsored by the U.S. National Science Foundation (NSF) under grants OAC-1826967, OAC-1541349, and CNS-1730158; and by the U.S. National Institutes of Health (NIH) under grant DP1-AT010885. Compute resources from both the Knight Lab and the Pacific Research Platform (PRP) were used in assessing the executable speeds.

MEDIA CONTACT

Jan Zverina, 858-534-5111, jzverina@sdsc.edu

Igor Sfiligoi, isfiligoi@sdsc.edu

Daniel McDonald, danielmcdonald@ucsd.edu

Rob Knight, robknight@eng.ucsd.edu

UC San Diego’s [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.