New Convey GraphConstructor Leverages Hybrid-Core Architecture to Speed De Novo Genome Assembly

From Cows to Grapes, Efficient Use of Memory Fuels More Science

Richardson, Texas, May 17, 2011 – Convey Computer today introduced the Convey GraphConstructor™ (CGC), a new software and hardware solution that speeds up some of the world's most popular bioinformatics algorithms and helps scientists better manage and analyze escalating amounts of research data.

The Convey GraphConstructor accelerates construction and manipulation of de Bruijn graphs commonly used in short-read genome assembly applications such as Velvet [1] and Abyss. [2] The Convey GraphConstructor is the newest component in the company's bioinformatics suite.

"Convey's hybrid-core architecture provides powerful advantages to scientists dealing with large datasets across many disciplines. Nowhere is this more important than in bioinformatics—where customers are achieving performance speed-ups from 2.2 to 8.4 times," [3] says Convey CEO and co-founder, Bruce Toal. "The Convey GraphConstructor helps researchers explore and manage the data deluge spilling from next-generation sequencing technologies faster and with significantly lower computing costs than in the past."

[1] http://www.ebi.ac.uk/~zerbino/velvet/; Velvet is the most widely used program for de novo assembly of short-read sequences.
[2] http://www.ncbi.nlm.nih.gov/pubmed/19251739
[3] Performance varies considerably depending on problem size and specified kmer lengths.

Key to improving performance and capability is Convey's novel hybrid-core computing architecture. Software-only applications are limited by the performance of commodity servers executing a stream of general-purpose instructions. Convey's architecture pairs classic Intel® x86 microprocessors with an FPGA-based coprocessor. This architecture allows key segments of an application—DNA sequence alignment, for instance—to run directly in hardware.

While raw processor performance increases are important, improved memory management is often just as important to increasing research throughput. Bioinformatics applications that depend upon random access patterns to large memory spaces, such as graph-based algorithms, experience severe memory performance limitations on cache-based x86 servers. Convey's highly parallel memory subsystem allows application-specific logic to concurrently access 8,192 individual words in memory, significantly increasing effective memory bandwidth over cache-based memory systems.

Many algorithms, such as Velvet and other de Bruijn graph based, short-read, de novo assemblers, greatly benefit from this type of memory architecture. Velvet author Dr. Daniel Zerbino says, "There are a number of engineering issues we didn't fully address in 2006 when we were developing Velvet and one of those issues was the hardware footprint. Memory size is the biggest difficulty. If your machine doesn't have enough memory, you must break down the problem and that can be quite a constraint for users. Convey's GraphConstructor offers a new approach to help researchers who want to test more parameters to achieve better assemblies or look at bigger jobs such as metagenomic or mammalian genome samples."

In fact, researchers at the U.S. Department of Energy/Joint Genome Institute (JGI) and the University of Mainz are currently using advanced computer architectures, such as hybrid-core computing and Convey's GraphConstructor, to tackle problems previously deemed impractical:

- **Cow Rumen Metagenome:** As part of their work researching biofuels, JGI researchers want to discover how cows convert grass to gas so effectively. In January, JGI reported (Science, January 28, 2011) it had "sequenced and analyzed 268 gigabases of metagenomic DNA from microbes adherent to plant fiber incubated in cow rumen." The result so far is discovery of nearly 30,000 new enzymes for biofuel production improvements. Using a Convey hybrid-core computer and Convey's GraphConstructor helped JGI researchers speed up the discovery process by as much as 2.8 times and reduce the required memory footprint by up to 82 percent.[4]

- **Riesling Genome Assembly.** In doing the first sequencing of the Riesling grape, University of Mainz researchers produced a dataset of 300 M reads, or 30 B nucleotides in size. Achieving an accurate assembly required using a fairly short kmer length, which drives up required memory and runtime. The existing computer system didn't have enough memory to complete the assembly with Velvet, but a Convey system running the Convey GraphConstructor was able to do so efficiently and without difficulty.

"Convey is solving a big problem here – de novo assembly has been very difficult," says Dr. John Castle, head of Bioinformatics/Genomics at the Institute for Translational Oncology and Immunology (TrOn), University of Mainz. "At TrOn, we tried to assemble the Riesling genome with SOAPdenovo and with Velvet. Both failed because the computer ran out of memory. Other groups are trying to assemble grape genome as well but with mixed results, frequently due to hardware limitations. Convey, just by increasing the efficiency of Velvet, has made a significant accomplishment!"

[4] Poster, "Efficient Graph Based Assembly of Short-Read Sequences on a Hybrid-Core Architecture," DOE JGI User Meeting, Genomics of Energy and Environment, March 22-24, 2011, Walnut Creek, California.

Convey's use of reconfigurable technology and supercomputer-inspired memory management systems permit Convey hybrid-core systems to accelerate applications, drive next-generation solutions, and lower ownership costs. Convey's hybrid-core platforms include the HC-1 and the HC-1^{ex}. For more information about Convey or the Convey GraphConstructor visit http://www.conveycomputer.com/

About Convey Computer Corporation

Based in Richardson, Texas, Convey Computer breaks power, performance and programmability barriers with the world's first hybrid-core computer—a system that marries the low cost and simple programming model of a commodity system with the performance of a customized hardware architecture. Convey brings decades of experience and intellectual assets to performance problem solving. Its executive and design teams all come from successful backgrounds of building computer companies, most notably Convex Computer Corporation and Hewlett-Packard. Convey Computer investors include Braemar Energy Ventures, CenterPoint Ventures, Intel Capital, InterWest Partners, Rho Ventures, and Xilinx. More information can be found at: www.conveycomputer.com.

For more information, contact Bob Masson (bmasson@conveycomputer.com) or Mary Dudley (mdudley@conveycomputer.com).