Biogemma Uses Improved Memory Sharing to Minimize Research Time

SGI UV 1000 Enables Rapid Results

Key Facts

Organization:
Biogemma

Location:
Chappes, France

Application:
Life Sciences

Biogemma is, today, the only European plant biotechnology research firm founded and financed by the agriculture industry. The company's shareholders are specialists in plant improvement and representatives of the major plant production chains in France: the cooperative groups Limagrain and Euralis, the firm RAGT, and the financial institutions Sofiprotéol and Unigrains. These partners all share strong convictions about the challenges facing agriculture today and the ways to go about tackling them.

Biogemma has more than 60 researchers and technicians specialized in genetic engineering, cellular and molecular biology, genome analysis, molecular marking, bioinformatics, plant physiology and pathology.

Researcher Successes

Biogemma works on major crop species such as maize, wheat, oilseed rape and sunflower. Except for maize, these species are not fully sequenced and are highly repetitive, poorly polymorphic and polyploid (wheat, rapeseed). As a result, it is a great challenge to develop markers that are necessary for the identification of trait-related candidate genes. Most of Biogemma's marker development projects rely on DNA sequencing and *de novo* sequence assembly.

Case Study 1: Whole Genome SNP Discovery on Non-Sequenced Species

Biogemma developed for the scientific community a comprehensive sequence and SNP database for the species *Pisum sativum*. This was achieved by whole exome sequencing using the MIRA *de novo* assembler, an analysis which was impossible without the SGI UV 1000 high performance computing platform. The study resulted in more than 30,000 markers for a species on which only 384 SNPs were publicly available.

Case Study 2: Trait-Associated Genes Targeted Sequencing on Non-Sequenced Species

Biogemma developed targeted sequencing based on sequence capture for all its species of interest. The advantage of this strategy is that it can help reconstruct complete genes of species for which only ESTs are available, allowing access to intron sequence polymorphism. A bioinformatic pipeline was developed for *de novo* assembly and SNP discovery dedicated to polyploid species, which was routinely run for thousands of genes in parallel. The assembly was made individually for each targeted gene, and the computing time was significantly reduced using parallelization with the SGI UV 1000 combined with PBS Professional® (PBSpro) job scheduling software.

Configuration

Biogemma purchased the SGI UV 1000 in 2011 to support its research mission of improving plants in order to contribute to the progress of agriculture. This system is configured as follows:

- SGI UV 1000 with 22 compute blades
- 264 cores of Intel® Xeon® processor X7542 running at 2.66GHz
- 2.56TB of coherent shared memory

Biogemma researchers use many bioinformatics tools, mainly in the DNA sequence alignment and assembly areas. The purchase of the UV 1000 was driven by the company's need for coherent shared memory, because their research projects now require handling multiple GBs to TBs of data. In some cases, like sequence assembly, an entire dataset needs to be accessible in memory at one time — it cannot be split into chunks.

"We recently optimized one of our analysis pipelines for the UV 1000 with PBSpro, and we managed to make it run in 50 hours instead of the previous 130 hours, by parallelizing PBS job submissions," according to Franck Pignard, IT Manager at Biogemma.

The UV 1000 replaced a Red Hat® cluster with 48 processor cores. Pignard states, "With this new installation, we are able to run some jobs which could not fit in memory at all on our previous cluster. I am very happy with this solution." The great advantage of the UV 1000 solution is that it allows both high performance computing for jobs like whole genome data assembly, as well as highly parallelized analyses and support for multiple users within an efficient environment.

SGI Professional Services Led the Implementation

Françoise Lecomte, IT engineer at Biogemma, explains, "A lot of custom work was done to enable this solution to work in our environment, especially in the area of training us on PBSpro, the software solution used for managing HPC workloads. We also needed to integrate the system into our web portal to enable researchers to submit processes to the system."

"With the UV 1000," Lecomte continues, "we are now able to get results in a few hours instead of multiple days. This is really important for us since we always have new research projects to do. We need to be able to tune the system parameters to get the best results, which is something we can do easily with the UV. Without the UV, this would not be possible with jobs taking more than a few hours."

About Biogemma

From maize to wheat, from oilseed rape to sunflower, Biogemma researchers describe, explore and enhance genetic diversity in the main farming crops in Europe. Their work opens up new perspectives in plant biotechnology and plant breeding which meet new demands in agriculture: savings in water resources, reduction of chemical treatments, food quality and safety. Please visit biogemma.com.

Global Sales and Support: sgi.com/global