



Science Highlight

Using the actor model of computation on Beagle ?

Sébastien Boisvert¹, Fangfang Xia^{1,2}, Rick Stevens^{1,2,3,4}

¹ Computing, Environment, and Life Sciences, Argonne National Laboratory, Lemont, Illinois ²

Computation Institute, The University of Chicago, Chicago, Illinois ³ Institute for Genomics and

Systems Biology, The University of Chicago, Chicago, Illinois ⁴ Department of Computer

Science, The University of Chicago, Chicago, Illinois

Massively parallel applications are required to perform alignment-free analyses of environmental soil microbiomes. Data for such soil microbiomes are typically gathered by using shotgun DNA sequencing of all genomes found in such an environment (called a metagenome). Massively parallel tools are now required to tackle the big data in environmental research. Pipelines exist for loosely-coupled tasks (such as short read alignment), but crafting tightly-coupled genomic tools (such as assemblers, or alignment-free approaches in general) that can tackle huge datasets remains challenging. Here, we describe an ongoing project for building the next generation of tightly-coupled high-performance applications (this is not targeting loosely-coupled tasks) in genomics.

Our work is presently focused on the Spate application, a metagenome assembler that is implemented strictly with actors. Spate will allow us to tackle one of the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) grand challenges, namely the one entitled Great Prairie Soil Metagenome Grand Challenge. This grand challenge contains 6 of soil metagenome samples from Kansas, Iowa, Wisconsin (1 cultivated corn soil sample and 1 native prairie soil sample for every of these 3 states). All these samples have least roughly 2 billion DNA sequencing reads and the one with the higher number of reads has more than 5 billion reads.

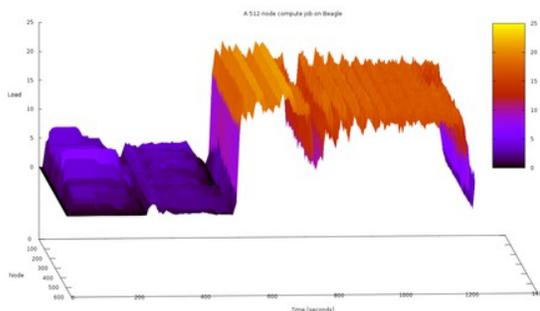


Figure 1. A 512-node metagenome assembly job on Beagle. The computation lasted 20 minutes, 21 seconds. The purple part at the beginning is input/output. This picture was generated with Spate (a metagenome assembler in development) and periodic patterns are visible.

We are currently testing the development code on Beagle using the DOE Great Prairie Soil Metagenome Grand Challenge (6 samples, see ref. [2]). One of the sample is Iowa Continuous Corn Soil which contains 2,228,341,042 Illumina DNA sequences. Using Beagle, our goal is to obtain a metagenome assembly in less than 1 hour. Right now, it takes around 20 minutes (256 Beagle nodes) to load data from storage to distributed memory and to build the assembly graph which contains 141,189,180,698 DNA sequences (called k-mers) with 140,183,562,558 overlaps.

Beagle is a Cray XE6 product and has 726 nodes connected with a Cray Gemini interconnect. This interconnect offers high throughput and low latency. Each Beagle node has 2 AMD Opteron 6100 series processors each with 12 x86-64 cores (total: 24 cores) and 32 GiB (ECC DDR3 SDRAM) of physical memory. This node architecture is suitable for our work because Thorium (the underlying computation engine) is a multi-node multi-core runtime environment and we typically use 1 single MPI rank and 22 or 23 worker threads per Beagle node. The multi-core processor (AMD Opteron "Magny-Cours") supports the x86 mfence instruction to allow us to insert memory barrier in the code to make sure that visibility of changes to memory is timely.

Resources:

Beagle Wiki

Get detailed usage information from the Beagle team

Beagle Support

Contact the Beagle experts for help

Globus Online for file transfer

Get started moving files to/from Beagle using this fast service

Other CI resources

Learn about other computing resources available at the Computation Institute



Beagle Upgrade:

We are proud to announce the substantial hardware upgrades provided by our successful Beagle 2 grant application. The upgrade would consist of:

- Addition of 2.24 PB of raw disk space in two new DDN cabinets for a total of 2.84 PB of raw storage (~2.1 PB total usable).
- All compute blades upgraded from 6-core Magny Cours processors to 8-core Abu Dhabi processors, increasing core count per node from 24 to 32.
- All compute blades upgraded from 32GB to 64GB per node.
- Login nodes network cards upgraded from 1Gbps NICs to 10Gbps NICs.
- Adding 4 compute nodes with nVidia GPU processors.

Work on those hardware upgrades are scheduled to begin on November 11. At that time, Beagle will be offline for three weeks while that work is completed. **Please note that Lustre data will be kept intact as part of the upgrade. While Beagle is offline, there is no way to access any data on Lustre.**

At this time we expect the work to happen accordingly, although these exact dates may still change in the future:

Week of Nov 10th - Beagle is down for pre-upgrade prep
Week of Nov 17th - Beagle hardware upgrade is started
Week of Nov 24th - Beagle hardware upgrade is completed
Week of Dec 1st - Beagle acceptance test is completed, users are given access to Beagle2

And that is not all ! We are happy to have **Lorenzo Pesce** back from his sabbatical where we heard that he was herding horses on Mongolian steppe.

Beagle Related Publications

Tietjen, G. T. Gong, Z. Chen, C. H. Vargas, E...Molecular mechanism for differential recognition of membrane phosphatidylserine by the immune regulatory receptor Tim4. Proc Natl Acad Sci U S A (2014)

Gumbart, J. C. Beeby, M. Jensen, G. J. Roux, B. Escherichia coli peptidoglycan structure and mechanics as predicted by atomic-scale simulations. PLoS Comput Biol

Urgent and important request for RECENT publications made using Beagle

Now that we are on the new grant we are ready to reconsider allocations and we need to weight things differently. In order to do that we would need to get your **most recent publications** made possible by Beagle. Feel free to send papers that you are not sure you might have sent already, we can deal with duplications.

Searle Chemistry Laboratory
5735 South Ellis Avenue
Chicago, IL 60637

 THE UNIVERSITY OF
CHICAGO
MEDICINE &
BIOLOGICAL
SCIENCES



Argonne
NATIONAL LABORATORY



THE UNIVERSITY OF
CHICAGO

Help E-Mail:
beagle-support@ci.uchicago.edu
Beagle CI Website:
beagle.ci.uchicago.edu