

CoMet: ExaOp Comparative Genomics for CBI on Summit

Background

- The Summit or OLCF-4 is a supercomputer developed by IBM for use at ORNL, which as of June 8, 2018 is the fastest supercomputer in the world with a speed of 200 petaflops for 64bit precision.
- *Populus* is a fast-growing perennial tree that shows promise as a low-cost, renewable feedstock for bioenergy. A large team of researchers collected data on 28 million genetic variations in *Populus* as part of the BioEnergy Science Center and the new Center for Bioenergy Innovation at ORNL. This dataset was well suited for this massively parallel comparative genomics approach on Summit.

Approach

- Modified CCC algorithm uses NVIDIA Volta Tensor Cores and cuBLAS with 16 bit mixed precision
- 25X faster than previous highly optimized bitwise algorithm on Titan, 10,000X faster than competing code.
- Near-perfect scaling to 4000 Summit nodes – 1.88 Exaops mixed precision was achieved

Outcomes

- Epistatic interactions were determined across the *Populus* genome were determined
- 95 TF half precision now achieved per GPU
 - We believe that over 2 EF mixed precision is achievable at 4500 nodes on Summit

Significance

- Millions of co-evolutionary relationships across a population of genomes were discovered at an unprecedented scale which support our ability to handle big-data in DOE biology
- This is the fastest scientific calculation ever done anywhere in the world

