

May 9, 2005

Application Name	Multiple Genome Alignment
Application Area	Bioinformatics, parallel computing
Keywords	Genome alignment, parallel application
Project/Dept. Affiliation	PhD Student, Computer Science
Value of grids to this application	Grid is the application platform for available resources
Originating institution	Georgia State University
Contact (s)	Bill Xie, PhD Student, Computer Science, (istcxx@langate.gsu.edu) Art Vandenberg, Advanced Campus Services, IS&T (avandenberg@gsu.edu) Nova Ahmed, PhD Student, Computer Science, currently Georgia Institute of Technology, (nova@cc.gatech.edu) Dr. Yi Pan, Computer Science (yipan@gsu.edu)
Participating sites	UAB, UVA, USC
General description	<p>The application takes number of genome sequences as input and gives an aligned sequence based on their structure.</p> <p>A multiple sequence alignment algorithm aligns multiple genome sequences using a pairwise alignment algorithm that implements a memory efficient method for computation and is also parallelized efficiently so that the workload is well distributed. Having compared the grid environment and a cluster environment, it can be stated that for carefully designed algorithms the grid environment is comparable for its performance to the cluster and it has the advantage of being flexible and scalable.</p> <div style="display: flex; justify-content: space-around;"> <div data-bbox="500 1144 852 1501"> <p>(a)</p> </div> <div data-bbox="860 1144 1412 1501"> <p>(b)</p> </div> </div> <p>Grid-enabled genome alignment: (a) Parallel load distribution among processors; (b) Computation time reduced with increased processors for different grid environments</p>
Anticipated system requirements for participation	Any system having MPICH to run parallel applications and minimum 1MB of memory
Anticipated non-system requirements for participation	An account for the user that contains minimum 1MB of storage space to keep the data files containing genome sequences

SURAGrid Application Description

Grid focus (data sharing, computation, access to unique resources, collaboration)	Computation, compute resource sharing with transparency via Globus
Network dependencies (bandwidth, latency, multicast, other)	The parallel program runs independently except a broadcast to the head node after some computation is done
Expected frequency of application run (one-time, occasional, monthly, weekly, daily...)	As needed, user initiated.
Estimated start date for application run	Has run several times as demonstration of capability.
Describe expected application invocation mechanism (by user submitting job, programmatically by some event or timing...)	User submits MPI job.
Is this application open to others to use with their own data or revisions?	Yes.
Additional comments	

Submitted MAY 9, 2005; revised DECEMBER 21, 2005