OSCD/PIRE research proposals

The Beijing Institute of Genomics (BIG), Chinese Academy of Sciences is one of the international members of the PIRE project. We have been trying to use cloud computing to manage our large-scale genomics data generating by high-throughput parallel DNA sequencing.

More information about BIG can be found at
http://english.big.cas.cn/

Here follows a research proposal to be performed at BIG during a period of 6 weeks in the early summer of 2013. The work is suitable for two students working on different aspects. The suitable candidates are graduated students with combined Computer Science and Biology background.

A pipeline for genome variations analyzing

The goal of this research is to develop a pipeline on cloud computing platform to analyze DNA mutations, gene copy number variations, DNA sequence insertions and deletions, which are important genome variations detected by next-generation sequencing (NGS).

The main functionalities of this pipeline will be:
- Retrieve the raw DNA sequencing data from high-throughput sequencing instrument;
- Trim the raw data and map the data to the reference genome of the species (i.e. human, mouse and etc.);
- Call DNA mutations, gene copy number variations, DNA sequence insertions and deletions from the data and return the results to the users.

To achieve the above the students will need to familiarize with the basic principle of high-throughput parallel DNA sequencing technology. They will also need to know current software which is used to analyze DNA mutations, copy number variations and DNA insertions and deletions, such as Bowtie, GATK, Varscan and Samtools.

The performance of this pipeline will be evaluated and compared with cluster-based server. The pipeline can be provided to biological users in OSDC system to help them with genome variations analyzing.

Contact: mishl@big.ac.cn