Project Title:

Whole-genome alignment of higher organisms

Name: Michiel de Hoon
Laboratory: LSA Bioinformatics Core Facility, RIKEN Omics Science Center, RIKEN Yokohama Institute

Our laboratory is interested in transcriptomics and transcriptional regulation in higher organisms. To analyze gene regulatory networks, we make use of transcription factor binding site predictions based on known binding motifs. The accuracy of transcription factor binding site prediction can be improved significantly by considering the conservation of transcription factor binding sites between organisms. For this purpose, we align the genome sequences of higher organisms against each other. Since the genome sequence of higher organisms is very long (about 3 GB for human), these whole-genome alignments take a considerable amount of time, in particular because the genome each organism needs to be aligned to the genomes of each of the other organisms. Whole-genome alignments are perfectly suited for parallelization, since the alignment of each pair of chromosomes can be run independently. We also predicted transcription factor binding sites for tens of thousands of position-weight matrices, which would be very difficult to accomplish in a reasonable amount of time without the resources of the RICC. Current usage statistics are shown in the Table.

	Limit (h)	Used (h)	Used (%)
Total	832200.0	144032.0	17.3%
+- mpc		128898.8	
+- upc		11592.7	
+- ssc		3540.5	

This project is primarily focused on supporting the FANTOM5 project, which is currently ongoing. The calculation results obtained have been analyzed and summarized in a FANTOM5 manuscript, which has been submitted for publication. I expect that more calculations will be needed depending on the reviewers' comments on the FANTOM5 manuscript, which we hope to receive within the next two months. While I expect that this project can finish within fiscal year 2013, there is a chance that the project will take longer.