Project Title:

Whole-genome alignment of higher organisms

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

Description of the project

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. For this purpose, we are using the resources of the RICC.

Current usage statistics are shown in the Table.

	Limit (h)	Used (h)	Used (%)
Total	694563.0	162822.1	23.4%
+- mpc		137403.0	
+- upc		25419.1	

The whole-genome multiple alignment, which is the final goal of our project, relies on the complete set of pairwise whole-genome alignments. As these pairwise whole-genome alignments are still ongoing, we have not yet been able to complete the multiple whole-genome alignment. Currently we have completed approximately 40% of the pairwise alignments needed for the whole-genome multiple alignment. We are hampered by the low throughput of our jobs, as we are limited to 64 running jobs under the quick usage account. We are therefore now applying for a General Account to speed up the whole-genome alignments by running more jobs in parallel.