Division of Biosciences

Department of Computational Biology and Medical Sciences

Laboratory	Faculty	Introduction of research activities and laboratory	Key words	Projects or activities summer program students can participate
Genome Informatics Laboratory (Asai Lab)	Prof. Kiyoshi ASAI	The target of our research is to clarify the “meaning” of the Genome. Genome sequences include various types of information, not only the amino-acid sequences of proteins, but their regulations, non-coding RNAs, and every history of evolution in our ancestors. We analyze DNA/RNA/protein sequences, their regulations, their functions, their structures, their interactions and their evolution using various types of mathematical models, which usually include statistics. Our favorites are stochastic models, machine learning theory, and dynamic programming. We are not the people who simply analyze biological data using software tools which someone else developed. We don’t want to use software black boxes for the analyses. We have been developing software for the analysis of DNA/RNA sequences, including gene finding, structural alignment of RNA sequences, and secondary structure prediction of RNA sequences, next-gen sequence analysis. Currently majority of the lab members are working on analyses of RNA sequences and their structures, such as RNA-seq data analysis, RNA secondary structure analysis including modified bases.	Biological sequences; RNA modification	(1) Geometry of biological sequences based and their function. Biological sequences (DNA, RNA and protein) are not simple sequences of the symbols, but are the results of evolution carrying hierarchical information. Alignment scores can be used distance measure of the sequences, but more complicated geometry should be considered to predict their function and to design the optimal sequences. (2) Genome-wide analysis of modified RNAs. It is known that most of the transcribed RNAs in the living cells are modified. Those post-transcriptional modifications change the structures of RNAs and their interacting targets. Predicting the effect of the modification is one of the major targets of epi-transcriptome research. We have energy parameters of modified bases and the tools which analyze the changes of structures/interactions of RNAs.

Frith Laboratory	Prof. Martin FRITH	Our ultimate aim is to decipher the functional and historical information in genome sequences. We do this using statistical models (such as hidden Markov models) and computational methods (such as enhanced suffix arrays and dynamic programming). A major approach is to compare and align related sequences to each other, to see how they have evolved. One recent focus is characterization of genome rearrangements in evolution and disease. Another long-term interest is promoter sequences and DNA motifs that regulate gene expression. Further interests are everything "weird": malaria genomes (80% A+T), frameshifts (especially in microbial metagenomes), unexplained evolutionary conservation, trans-splicing, etc. Our physical location is partly in Kashiwa, and partly in Odaiba in central Tokyo.	Genome; evolution; orthology; probability-based	Students are encouraged to pursue their own ideas on analyzing genetic sequences. There are broadly two types of project: biological investigation, and method development. Examples of biological investigation: survey the evolution of gene structure by gain or loss of splice sites, frameshifting, gene fusion or fission, etc; compare the evolution of mitochondrial versus plastid genomes; compare genome evolution to major body-form evolution (e.g. snakes, whales). Examples of method development: make a sensitive probabilistic model for finding distantly-related DNA sequences; devise a beautiful way to visualize complex sequence rearrangements; develop a way to extract specific rearrangement events from pair-wise alignments of long sequences (e.g. long DNA reads or whole genomes).
Morishita Laboratory	Prof. Shinichi MORISHITA	We have been attempting to develop efficient and accurate algorithms for uncovering "dark matters" in genomes that are hard to observe using traditional second generation DNA sequencers such as Illumina HiSeq and Ion torrent. Typical examples of dark matters are genomic sequences of long repetitive elements (LINE and LTR), centromeres, telomeres, homologous chromosomes, and microbiome. Towards this end, we exploit full potential of our third generation sequencers (PacBio Sequel, Oxford nanopore, 10X Chromium) that realize single-molecule sequencing and are capable of sequencing very long DNA fragments of >10,000 base pairs. Extending single-molecule sequencing, we also have been devising efficient algorithms for observing DNA methylation states of dark matters, for example, CpG methylation of centromeric repeats and highly repetitive transposons, and 6mA in gut microbiome so as to understand their biological functions.	single-molecule sequencing; single-cell sequencing; centromeres; microbiome; DNA methylation	We introduce basic ideas and algorithms for handling single-molecule sequencing data as well as how to operate third generation sequencers. We also provide a couple of open problems in this research fields. Afterwards, summer students are expected to propose and develop new ideas, applications, or algorithms through brainstorming with our graduate and undergraduate students. They can use highly parallel computers with thousands of CPU cores if they are interested.