SIII-2-1

Bioinformatics and Genome Project

Seyeon Weon
sywon@bioneer.com
Department of Bioinformatics, DNA Research Institute, Bioneer Corporation

Bioinformatics is booming in advanced countries due to the rapid and successful progress of various genome projects. High throughput sequencing facilities produce millions of bases each month and computer techniques for producing and processing these data are the requirement to make this effort become meaningful. Especially for microbiology field, it becomes almost routine to sequence the whole genome by virtue of the much smaller sizes of the genomes than those of higher organisms. We can say that microbiologists are the most vulnerable people among biologists to the new wave of biology, in which computer takes important part. In this talk, the current trend in the genome projects, computer techniques available, and how to prepare oneself for this new methodology of biological research will be discussed. Due to the interdisciplinary nature of bioinformatics, extra effort is required to produce the specialists for the field. We also have to find the way to educate the new generation of biologists toward this direction.

SIII-2-2

Inferring Relatedness of a Macormolecule to a Sequence Database Without Sequencing

Jin Kim*
Department of Computer Science, Kon-Kuk University

Derivation of biological information of a macromolecule isolate based on sequence similarity is playing a significant role in numerous areas of biological research. However, it is often the case that a researcher obtains more macromolecule isolates than can be sequenced practically, due either to the high cost of sequencing or lack of specialized equipment and personnel. To overcome this difficulty, we study the problem of obtaining biological information about a macromolecule isolate using (i) only the fragmentation pattern of that isolate obtained from digestion with enzymes and (ii) a database D of sequences. We investigate a three phase approach to solving this problem. In the first phase, we obtain a restriction pattern of the isolate while analytically deriving the corresponding restriction maps of the sequences in the database. In the second phase, we identify a set $S \subseteq D$ of sequences which have restriction maps that are most similar to the unknown isolate's restriction pattern. This task is complicated by the fact that we have only approximate fragment lengths for the unknown isolate and that we do not know the actual ordering of the unknown isolate' fragments. Despite these difficulties, we derive experimental results which indicate maximum matching techniques are effective in identifying the correct set most of the time. In the third phase, we use the set S to infer biological information about the unknown isolate. We demonstrate experimentally that the closeness of the sequences in the set S to each other can be used to infer the relatedness of the unknown isolate to the sequences of the set S. Furthermore, the confidence of this inferred information is strongly correlated to the minimum pairwise relatedness of any two elements in S.