What the next generation of sequencers may bring

Kentaro Tomii
Senior Research Scientist
Sequence Analysis Team

It seems indisputable that faster, bigger, cheaper are all watchwords of the modern age. Genome sequencing is no exception, and recent developments in sequencing technology have brought about remarkable advances1. I don't think it is too much to say that we now have DNA sequencing technologies of unprecedented power.

It is impossible to estimate the effect that these new technologies will have on the life sciences. Multiple genomes have already been sequenced, but to date we have been limited to organisms for which cell culture is possible. By using the newest technologies, however, it has become possible to sequence even genomes for organisms whose cells cannot be cultured at present, an increasingly active field known as metagenomics (or environmental sequencing). This has enabled us to obtain a great amount of information on novel genes. The scope of application of the new technologies does not stop there - it is also now possible to repeat the sequencing of genomes that have already been sequenced for validation and the determination of inter-individual variance, the analysis of gene expression, and even the sequencing of DNA from Neanderthal man and the wooly mammoth.

Although as seen from the examples above these new technologies are exceptionally promising, a number of issues do remain. For instance, the short length of the reads, which represent the sequencer output, and their precision (which varies by sample), remain problematic. For experimental subjects for which it is possible to collect highly redundant sequence information, this may not be a major issue, but in metagenomics projects involving diverse organisms, it can have a serious effect on gene identification. In order to resolve these problems, I am working on a collaborative research project at the University of California, Berkeley under the auspices of the CBRC external researcher program on algorithms to enable the efficient searching of homologous regions mutual to sequences, with consideration for the error-containing DNA sequence information using precision data at the individual nucleotide base level. By developing this method, it may become possible to take advantage that has been unavailable due to problems of precision, possibly leading to the identification of novel genes.

Regardless of these problems, the amount of information on "live" DNA that must be handled by next-generation systems is far greater than for past systems, and computers are sure to play a much greater part than in the past, particularly in the assembly and mapping. New sequencing technologies will continue developing hand in hand with new applications for sequence analysis, and it seems certain that the idea of genomes as personal information will become a much more familiar concept through its links to medicine in the future. Many issues still need to be addressed as that eventuality approaches, and I hope to continue to drive my research ahead in step with these changing times.

Reference

1. Shendure, J., Ji, H. , "Next-generation DNA sequencing", *Nat Biotechnol.*, **26**(10), pp.1135-45(2008)

View of San Francisco Bay seen from Berkeley