ABSTRACT

"More than just sequencing –
Bioinformatics tools developed in BGI"

Ruibang LUO, B.Sc.
Bioinformatics project manager
BGI (Beijing Genomics Institute)-Shenzhen, China

"The day is not far off when more biology will be done at the computer than at the bench!" Roos *et al.* said in 2002.

Nowadays, with the rapid development of information technology, not only the computational resources but also talents of this field could be acquired much more easily. It has also demonstrated its powerful ability in assistance of solving complex biological, especially genomic problems. BGI, a flagship in genomics in China for the decade, has created our own series of tools and pipelines for the analysis including the genome, SNP, structural variation, methylation, RNA, assembly and etc.

Next-generation massively parallel DNA sequencing technologies provide ultra-high throughput at a substantially lower unit data cost; however, the data is very short read length sequences, making *de novo* assembly extremely challenging. We've just described a novel method for *de novo* assembly of large genomes from short-read sequences. We successfully assembled both the Asian and African human genome sequences, achieving an N50 contig size of 7.4 and 5.9 Kb and scaffold of 446.3 and 61.9 Kb, respectively. Comparison of these assemblies to the NCBI reference genome demonstrated the capability of accurate identification of structural variations, especially small deletions and insertions that are difficult to identify using a resequencing method. The development of this *de novo* short-read assembly method creates new opportunities for building reference sequences and carrying out accurate analyses of unexplored genomes in a cost effective way.