

## **Title: Computing Opportunities in the Era of Abundant Biological Data**

**Presenter: Dr. Gane Ka-Shu Wong, Associate Director of Beijing Institute of Genomics**

### **Abstract:**

The completion of the sequence of the human genome was announced with great fanfare in the summer of 2000, with representatives from the public sector (Francis Collins of the NIH) and the private sector (Craig Venter of Celera) meeting at the White House to join hands in their brief moment of glory. At the time, it was asserted that this sequence would revolutionize the practice of medicine forever. Since those heady days, the stock market bubble has burst, and Venter has resigned from Celera. Was it all an illusion? One should hope not, because corporations like IBM, Oracle, and GE have all launched major new initiatives in life sciences. Genome sequence continues to accumulate in the databases at growth rates surpassing Moore's Law. More importantly, genome sequence is only the beginning. The output of the genome is a set of biomolecules (RNA and protein) that turn on and off under different conditions, that interact with each other, that interact with the environment, and ultimately produce the miracle known as life. A bewildering assortment of biological information is being generated, to record this molecular symphony, under all possible environmental and physiological conditions, leading to all possible outcomes. Their ultimate goal is to reverse engineer this system, by building mathematical models with as much detail as necessary. As a result, there is a monumental culture change going on in biology, which is becoming an increasingly quantitative science, driving the need for more computing power. I will show examples of where high power computing is used in biology, past present and future. Then, to put it all in perspective, I will discuss how genomes evolve. The irony of the situation is that, despite all this computing power being brought to bear on the problem, the best model for genome evolution is the haphazard development of the much-maligned Microsoft OS. What they share is that both processes were driven by immediate Darwinian needs, with little foresight or planning, and we are all left to deal with the consequences.

### **Biography:**

Gane Ka-Shu Wong is an Associate Director for the Beijing Institute of Genomics (BIG), one of the largest biological research facilities in Asia, best known for their work on the sequencing of the genomes for human, rice (*indica* and *japonica*), silkworm, and chicken. He is responsible for all of the major computational analyses performed at BIG, as well as the overall scientific direction and publications in high profile journals like *Nature* and *Science*. Eleven years ago, he was recruited into the Human Genome Project (HGP) by Maynard Olson, one of the founders of the HGP. In concert with Phil Green and Jun Yu, these four scientists launched the University of Washington Genome Center, which set new standards for data quality at the time, and developed many of the key software tools used by the HGP. In his previous life, before he moved to UW in Seattle, and made the switch into biology, he was an experimental low temperature physicist with a Ph.D. from

Cornell University. But, even before that, he was an engineer with a B.A.Sc. from the University of British Columbia, with quadruple honors in physics, electrical engineering, mathematics, and computer science.