Abstract:

To understand biodiversity, reconstruct metabolic metabolic pathways, and to harness their potential for biotechnological applications, it is desirable to obtain complete information about the genetic blueprint and epigenetic constitution of the organisms present in the samples under study. Traditional Sanger and next-generation short-read sequencing technologies have shortcomings with respect to read lengths and DNA sequence context bias, leading to fragmented and incomplete genome information. The development of long-read, single molecule, real-time (SMRT) DNA sequencing, with >10,000 bp average read lengths and a lack of sequence context bias, now allows for the generation of complete genomes in a fully automated workflow. In addition to the genome sequence, DNA methylation is characterized in the process of sequencing. I will highlight several examples where these capabilities have been leveraged in the areas of metagenomics and biodiversity studies of higher organisms, including reference-quality de novo assemblies of non-model organisms, full-length 16S sequencing, functional metagenomics, microsatellite marker discovery and surveys, elucidation of complex gene families, new natural product and antibiotic discovery, and livestock/plant microbiome interactions.