## NATIONAL UNIVERSITY OF SINGAPORE

## School of Computing

## PH.D DEFENCE - PUBLIC SEMINAR

Title:	Investigating lipid and secondary metabolisms in plants by next- generation sequencing
Speaker:	Ms Jin Jingjing
Date/Time:	15 October 2014, Wednesday, 10:30 AM to 12:00 PM
Venue:	Executive Classroom, COM2-04-02
Supervisor :	Dr Wong Lim Soon, Professor, School of Computing

## Abstract:

Plant metabolites are compounds synthesized by plants for essential functions, such as growth and development (primary metabolites, such as lipid), and specific functions, such as pollinator attraction and defense against herbivores (secondary metabolites). Many of them are still used directly, or as derivatives, to treat a wide range of diseases for humans. There is a demand to explore the biosynthesis of different plant metabolites and improve their yield.

Next-generation sequencing (NGS) techniques have been proved valuable in the investigation of different plant metabolisms. However, genome resources for primary metabolites, especially lipids, are very scarce. Similarly, using NGS, most current studies of secondary metabolites just focus on known function/metabolic pathways. Hence, in this dissertation, we systemically investigate plant lipid metabolisms and secondary metabolisms by several different studies.

We first develop a reference-based genome assembly pipeline, including misassembled scaffold and repeat scaffold identification components. From the evaluation on a gold-standard dataset, we find that these major components in our pipeline have relatively high accuracy.

Next, we use our proposed reference-based genome assembly pipeline to construct a draft genome for Dura oil palm. Then, annotations---including protein coding genes, small noncoding RNAs and long noncoding RNAs---are done for the draft genome. In addition, by resequencing 12 different oil palm strains, around 21 million high-quality single-nucleotide polymorphisms (SNPs) are found. Using these population SNP data, lots of sites with a high level of sequence diversity among different oil palms are identified. Some of these variants are associated with important biological functions, which can guide future breeding efforts for oil palm.

At the same time, a GBrowse-based database with a blast tool is developed to visualize different genome information of oil palm. It provides location information, expression information and structure information for different elements, such as protein-coding genes and noncoding RNAs.

In order to predict new functions/metabolisms for plants, a weighted pathway approach is proposed, which tries to consider dependencies between different pathways. From the validation results on two different models, we find that the weighted pathway approach is much more reasonable than traditional pathway analysis methods which do not take into consideration dependencies across pathways.

After applying this weighted pathway approach to an RNA-seq dataset from spearmint, several new functions and metabolisms are uncovered, such as energy related functions, sesquiterpene and diterpene synthesis. The presence of most of these new metabolites is consistent with GC-MS results, and mRNAs encoding related enzymes have also been verified by q-PCR experiment.