# Whole Genome Epidemiological Typing of *Salmonella*

Pimlapas Leekitcharoenphon (Shinny)
Summary of PhD thesis
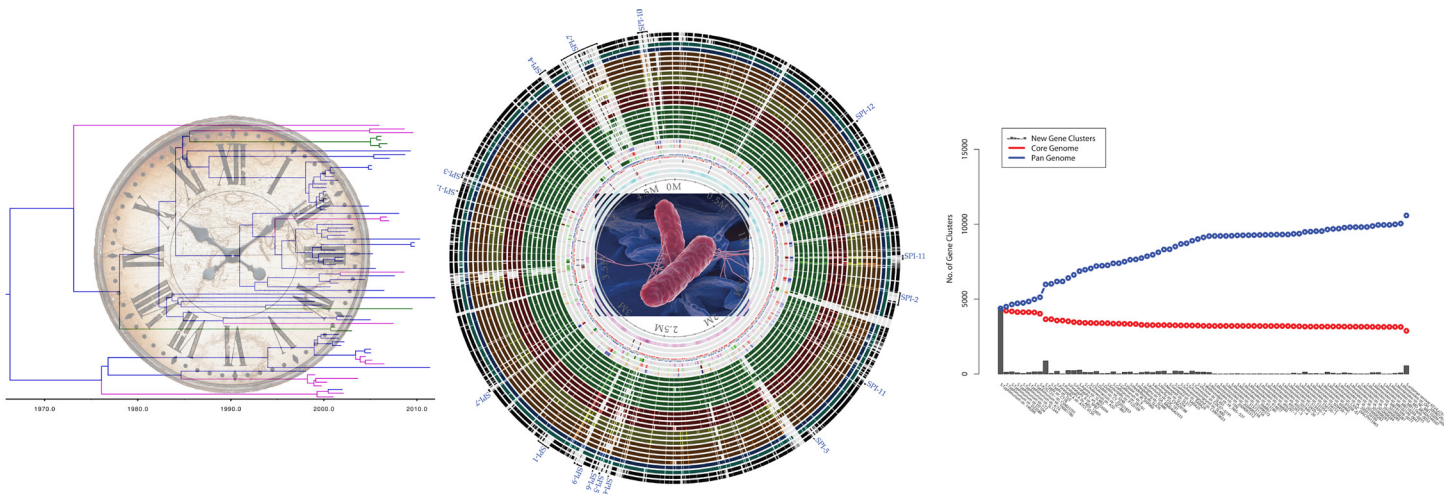
# SUMMARY

*Salmonella* is one of the most common foodborne pathogens worldwide. In the US alone, salmonellosis was estimated to cause 1.4 million cases effecting 17,000 hospitalization and almost 600 deaths each year. Particularly, *Salmonella enterica* is a common cause of minor and large food borne outbreaks. Technological advances and effective price in high throughput genome sequencing are making whole genome sequencing (WGS) available as a routine tool for bacterial typing.

Typing of *Salmonella*, especially sub-typing within the same serotype or even the same clone, the genetic variation of the target genes being used for typing is crucial for successful discrimination. The core genes or the genes that are conserved in all members of a genus or species are potentially good candidates for investigating genomic variation in phylogeny and epidemiology. A total of 2,882 core genes have been observed among 73 available *Salmonella enterica* genomes (accessed in April 2011). A consensus tree based on variation of the core genes gives better resolution than 16S rRNA and MLST that rarely provide separation between closely related strains. The performance of the pan-genome tree which is based on the presence/absence of all genes across genomes, is similar to the consensus tree but with higher branching confidence value. The core genes can be divided into two categories: a few highly variable genes and a larger set of conserved core genes, with low variance. These core genes are useful for investigating molecular evolution and remain useful as candidate genes for bacterial genome typing-even if they cannot be expected to differentiate highly clonal isolates e.g. outbreak cases of *Salmonella* [**I**].

To achieve successful 'real-time' monitoring and identification of outbreaks, rapid and reliable sub-typing is essential. A collection of thirty-four human *S*. Typhimurium strains from six different outbreaks together with background strains plus eight *S*. Enteritidis isolates from two outbreaks and five *S*. Derby isolates from a single outbreak were used to evaluate the strengths and drawbacks of different WGS approaches compared to the traditional typing, PFGE, for retrospectively outbreak typing of *Salmonella*. The resulting outcome showed that SNP analysis and nucleotide difference approach seem to be the superior methods for outbreak detection compared to other phylogenetic analytic approaches of WGS. Furthermore, WGS approaches were also superior to the more classical typing method, PFGE. Meanwhile, k-mer method constructs a tree in high speed and giving high accuracy in clade level [**II**].

SNP analysis has successfully applied in recent epidemiological studies of *Salmonella*. Currently, there are different tools and methods to identify SNPs including various cut-off values. In addition, all the tools require bioinformatics skill. In order to apply WGS in routine typing, an automatic and user-friendly tool is needed. Therefor, snpTree has been developed as a server for online-automatic SNP analysis. snpTree can identify SNPs and construct phylogenetic tree from WGS raw reads as well as from assembled genomes or contigs. The tool is freely accessible at http://cge.cbs.dtu.dk/services/snpTree/ [**III**].

Globally, *Salmonella enterica* serovar Typhimurium is the most commonly isolated serovar. *S.* Typhimurium consists of a number of subtypes that conventionally have been divided by phagetyping. During the last three decades, *S.* Typhimurium phage type DT104 emerged as the most prevalent phage type and one of the best-studied because of its rapid global dissemination. Nonetheless, the origin and transmission route of this particular phage type have not been revealed. To bridge the gaps in epidemiology of DT104, WGS and temporally structured sequence analysis within Bayesian framework have been incorporated for reconstructing temporal and spatial phylogenies, estimating rate of mutation and divergence time of global and local *S.* Typhimurium DT104 isolates sampled from 1969 to 2012 from twenty-one countries in six continents. The DT104 was estimated to initially emerge as antimicrobial-susceptible strains in ~1946 (1931-1959) and further became multidrug-resistant (MDR) DT104 in ~1974 (1966-1981) through horizontal transfer of 13-kb SGI1 MDR region into SGI1-contained susceptible strains. Changes in population size over time supported global occurrences of MDR DT104. Besides, using WGS is capable to confirm local epidemiology especially the transmission between animal herds of DT104 isolates from Denmark. Interestingly, the demographic history of Danish MDR DT104 provided an evidence for the accomplishment of an eradicating program across pig herds in Denmark during 1996 to 2000 [**IV**].

Overall, this Ph.D. thesis has assessed the usefulness of WGS epidemiological typing in *Salmonella* as well as evaluated the different WGS approaches for outbreak investigation compared to the traditional typing, PFGE. An online tool to construct phylogenetic tree based on SNPs has also been developed. Furthermore, it has revealed the application of WGS in epidemiological study of global and local occurrences of *S.* Typhimurium DT104.