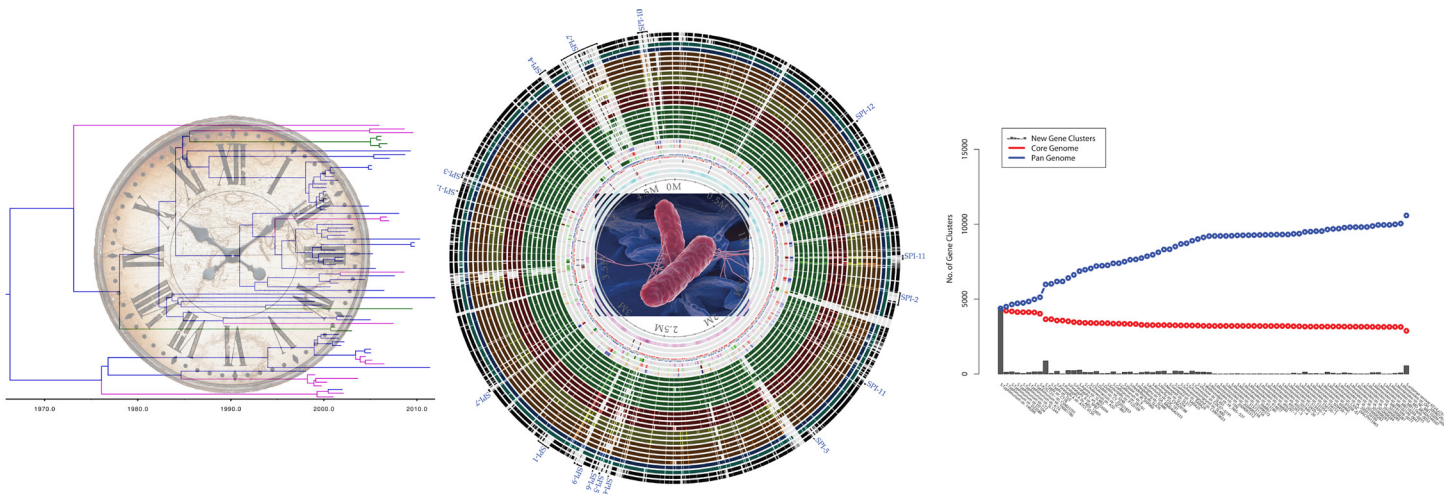


Pimlapas Leekitcharoenphon (Shinny)  
Sammenfatning af Ph.D.-afhandling

# Whole Genome Epidemiological Typing of *Salmonella*



## RESUMÉ

*Salmonella* er en af de mest almindeligt forekommende pathogene bakterier i fødevarer og fødevarereproduktion på verdensplan. Alene i USA er der ca. 1,4 millioner tilfælde af salmonellosis om året, hvilket resulterer i 17.000 hospitalsindlæggelser og næsten 600 dødsfald. Særligt er *Salmonella enterica* (*S. Enterica*) en hyppig årsag til både små og store udbrud af fødevarerforgiftning. Teknologiske fremskridt, store prisfald og stigning i antallet af prøver der kan sekventeres samtidigt, gør nu helgenom DNAsekventering (HGS) tilgængeligt som et værktøj til rutinemæssig typning af bakterier.

I *Salmonella* typning – specielt sub-typning indenfor den same serotype eller endda samme klon, er den genetiske variation i de target gener, der bruges til typningen, afgørende for, om det lader sig gøre at adskille individuelle isolater. ”Core” generne – eller de gener, der er konserverede i alle medlemmer af en genus eller art, er potentielle kandidater til at undersøge genomisk variation, udlede phylogenen og studere epidemiologien. I *S. enterica* er der fundet 2.882 fælles core gener i 73 offentligt tilgængelige *S. enterica* genomer (april 2011). Et konsensus træ baseret på variationen i core generne giver bedre differentiering end 16S RNA og Multi Locus Sekvens Typning (MLST), som sjældent muliggør separation af tæt beslægtede isolater. Et pan-genom træ, baseret på tilstedevær/fravær af alle gener i genomerne, gav et resultat, sammenligneligt med core konsensus, dog med højere konfidens på træets forgreninger. Core generne kan med fordel inddeles i to kategorier: En mindre gruppe gener med høj variation og en større, betående af konserverede core gener med lav varians. Disse core gener er anvendelige til studier af molekylær evolution og forbliver nyttige kandidater til bakteriel genotypning, selvom de ikke forventes at differentiere klonale isolater, eksempelvis fra et *Salmonella* udbrud [1].

For at opnå succesfuld ”real-time” monitorering og identifikation af udbrud er det essentielt at være i stand til at udføre hurtig og pålidelig sub-typning. En samling af 34 humane *S. Typhimurium* isolater fra seks forskellige udbrud samt baggrunds isolater, plus otte *S. enteritidis* isolater fra to udbrud og fem *S. derby* isolater fra et enkelt udbrud, blev anvendt til at evaluere styrker og svagheder i forskellige HGS baserede typninger sammenlignet med konventionel retrospekt PFGE typning af *Salmonella* udbrud. Resultatet af analyserne udpeger SNP analyse og sammenligning af parvis nukleotid forskel som superiøre metoder til detektion af udbrud set i forhold til andre, ligeledes HGS baserede, phylogenetiske metoder. Overordnet set, var samtlige HGS baserede metoder superiøre i forhold til den mere klassiske PFGE typning. K-mer metoden

var imidlertid den hurtigste og det resulterende klassifikations træ havde høj nøjagtighed på clade niveau [II].

SNP analyse har med succes været anvendt i nylige epidemiologiske studier af *Salmonella*. Aktuelt set eksisterer der forskellige værktøjer og metoder til at identificere SNPs, inklusiv adskillige, varierende cut-off værdier. Samtlige værktøjer afkræver brugeren et vist niveau af bioinformatisk kunnen. For at muliggøre anvendelsen af HGS i rutinemæssig typning er det nødvendigt at udvikle et automatisk og brugervenligt dataanalyse-værktøj. Derfor har vi udviklet snpTree – en server til automatisk, online SNP analyse. snpTree kan identificere SNPs og konstruere et phylogenetisk træ fra HGS og raw reads såvel som fra samlede genomer eller contigs. Værktøjet er frit tilgængeligt på <http://cge.cbs.dtu.dk/snpTree/> [III].

Globalt er isolater fra *Salmonella enterica* serovar Typhimurium de hyppigst forekommende. *S. Typhimurium* består af et antal subtyper, der konventionelt er blevet opdelt via phagtypning. Gennem de forrige tre årtier, er *S. Typhimurium* phagtype DT104 vundet frem som den mest udbredte phagtype og også én af de mest studerede netop på grund af dens meget hurtige, globale spredning.

Dette til trods, har man endnu ikke været i stand til at udlede denne særlige phagtypes udspring og transmissionsrute. I et forsøg på at brolægge hullerne i epidemiologien af DT104 har vi inkorporeret HGS og temporalt struktureret Bayesian baseret sekvensanalyse til at rekonstruere temporale og geografiske phylogener og estimerer dermed mutationsrate og divergenstidspunkt af en række globale og lokale *S. Typhimurium* DT104 isolater indsamlet over årene 1969 til 2012 fra 21 lande fordelt på seks kontinenter. Den globale spredning af DT104 er estimeret til at have oprindelse i en opblomstring af en antimikrobiel sensitiv i klon i ~1946 (1931-1959) , der via horisontal genoverførsel af en 13-kb SGI1 MR region til SGI1-positive sensitive isolater, videreudviklede sig til antimikrobiel multi-resisten (MR) DT104 i ~1974 (1966-1981). Ændringer i populations størrelsen over tid understøttede ligeledes den globale forekomst af MR DT104. Udover emergens studier, er HGS også anvendeligt til bekræftelse af lokal epidemiologi – specielt transmission af DT104 mellem danske dyre besætninger. Endvidere var det meget interessant at den demografiske historie af de danske MR DT104 gav evidens for at det danske bekæmpelsesinitiativ i svinebesætningerne i årene 1996 til 2000 har været en succes [IV].

Dette Ph.D. studie har vurderet anvendeligheden af HGS baseret epidemiologisk typning af *Salmonella* samt evalueret de forskellige HGS data analyse metoder med henblik på udbruds

studier i sammenligning med traditionel PFGE typning. Et online-værktøj til at konstruere fylogenetisk træ baseret på SNPs er også blevet udviklet. Endvidere er HGS anvendt i et epidemiologisk studie der der kortlægger den lokale og globale forekomst af *S. Typhimurium* DT104.