Analysis of IncHI1 plasmids in *Salmonella enterica* serovar Typhi

by

Minh-Duy Phan

This dissertation is submitted for the degree of Doctor of Philosophy

December 2009

Darwin College

UNIVERSITY OF CAMBRIDGE

// Abstract

Analysis of IncHI1 plasmids in *Salmonella enterica* serovar Typhi

Minh-Duy Phan

Plasmids play an important role in bacterial adaptation and evolution. Plasmids of incompatibility group HI1 (IncHI1) are the major vectors for the global dissemination of multiple antibiotic resistance genes among *Salmonella enterica* serovar Typhi (*S.* Typhi). *S.* Typhi is a human adapted serovar which cause the major human infection: typhoid fever. The majority of cases are in developing countries where sanitation and safe drinking water are inadequate and the true burden of disease is unknown.

This project investigated the genetic factors encoded on plasmid as well as host chromosome that are responsible for the stable maintenance of IncHI1 plasmid in *S.* Typhi. Transposon Directed Insertion-site Sequencing (TraDIS), a novel method that enables the simultaneous assay of every gene in the genome using Illumina next generation sequencing technology, was used to identify a gene set involved in plasmid stability in the bacterial host. The method successfully identified the known stability factor *sfh* among other hypothetical CDSs.

The evolution and population dynamics of IncHI1 plasmids were also studied by adapting Multi-Locus Sequence Typing for IncHI1 plasmids (PMLST). The method defined eight different plasmid sequence-types (PST), clustering into 2 groups. Group 1 was found to consist of plasmids isolated before 1993, whilst group 2 consisted of plasmids isolated after 1993. To obtain greater typing resolution on a larger strain collection, the Illumina GoldenGate SNP-typing platform was used to type both chromosomal and plasmid SNPs for 473 *S.* Typhi strains collected from 45 countries between 1916 and 2007. There is an absolute association of PST6, the predominant plasmid since 1993, with a widespread chromosomal background, H58. This suggests a competitive advantage of the ST6-plasmid/H58-haplotype combination.

In conclusion, this project demonstrates the important impact that resistance plasmids can have on the biology of a major human pathogen.

Acknowledgements

I would like to thank my supervisor Dr. John Wain for giving me the opportunity to complete this PhD. Thank you for all the guidance, discussion, supporting and especially for patience during my writing stage.

I would also like to thank Professor Gordon Dougan for supervising me after John left the Sanger Institute and for the help with my writing up. I would like to thank Dr. David Summers and my thesis committee, Dr. Julian Parkhill and Dr. Keith Turner, for their guidance.

A special thanks to Kathryn Holt, Gemma Langridge, and Keith Turner for their help in the lab and their collaboration in several projects which made up parts of my PhD.

Team 100, especially Satheesh Nair, and team 15 have been very supportive and helped me with many important little things. Thanks for taking great care of me.

I would like to extend my gratitude to all my collaborators, especially Trần Thụy Châu and Christiane Dolecek at the Oxford University Clinical Research Unit in Vietnam, François-Xavier Weill at Institut Pasteur Paris, France, among others including John Albert, Getenet Bevene, Fiona Cooke, Shanta Dutta, Rajni Gaind, Sam Kariuki and Bianca Paglietti, who helped me collect and extract DNA from clinical samples.

It is impossible to list here all the people at the Sanger Institute and elsewhere in the world who have consulted and helped me during my research. I am very grateful for their help.

My friends, Lý, Hiếu, Huy, Nguyên, Long, Quang, Hằng and many others, have been very encouraging and patiently listened to my ranting. Thanks for keeping me sane!

Last but not least, I would like to dedicate this work to my family and my partner who always believe in me and give me lots of encouragement although we have to live so far apart.

Declaration

I hereby declare that this thesis consists of work derived entirely of my own work. Due to the nature of this work, and interdisciplinary nature of biological sciences, it was not possible and impractical to perform all of these techniques, however, it was possible to design all experiments associated with this thesis. Work that was done by other persons is clearly stated in the Materials and Methods section.

This thesis is no longer than 300 pages as required by the School of the Biological Sciences.

Minh-Duy Phan

December 2009

Table of Content

Abstract 2
Acknowledgements 3
Declaration 4
Table of Content 5
List of Tables 8
List of Figures 9
Abbreviations 11
1 Introduction 12
1.1 Enteric fever 12
1.1.1 Definition 12
1.1.2 The pathogens 13
1.1.3 Treatments of enteric fever 15
1.1.4 Antibiotic resistance 16
1.1.5 Epidemiology – disease burden 20
1.1.6 Salmonella typing methods 21
1.2 Plasmid biology 28
1.2.1 Replication mechanisms 28
1.2.2 Maintaining mechanisms 34
1.2.3 Fitness cost 41
1.2.4 Toward the phylogenetic analysis of plasmids isolated from clinical strains 41
1.3 IncHI1 plasmids 43
1.3.1 Plasmids of incompatibility group H 43
1.3.2 R27 – progenitor of IncHI1 plasmids 44
1.3.3 Temperature dependent conjugation and H-NS 46
1.3.4 The development of IncHI1 plasmids in *S.* Typhi 47
1.3.5 IncHI1 plasmids in *S.* Paratyphi A 48
1.3.6 The cost of IncHI1 plasmid in *S.* Typhi 48
1.4 Mobile genetic elements associated with antibiotic resistant genes 49
1.4.1 Transposons 50
1.4.2 Class one integrons 54

1.5 Aims of the Study 57
2 Materials and Methods 59
2.1 Materials 59
2.1.1 Chemicals and reagents 59
2.1.2 Plasmid collection 59
2.1.3 Bacterial strains and isolates 59
2.1.4 Bacterial culture media 61
2.1.5 Enzymes 62
2.1.6 Oligonucleotides 62
2.2 Microbiological methods 62
2.2.1 Bacterial culture 62
2.2.2 Identification of bacteria 63
2.2.3 Antimicrobial susceptibility testing using disc diffusion 63
2.2.4 Growth curves 64
2.3 Molecular methods 66
2.3.1 Genomic and plasmid DNA extraction 66
2.3.2 Primer design and PCR 68
2.3.3 Plasmid replicon-typing by PCR 68
2.3.4 Long-range PCR 70
2.3.5 Sequencing and analysis 73
2.3.6 Plasmid multi-locus sequence typing (PMLST) 73
2.3.7 Generation of sfh knock-out 74
2.3.8 TraDIS 77
2.3.9 Genotype array (SNP typing) 82
2.3.10 PCR to detect mobile elements 87
2.3.11 Plasmid competition 88
3 Fundamental factors for plasmid stability 89
3.1 Introduction 89
3.2 Results 91
3.2.1 The generation of a one million mutant library 91
3.2.2 Experimental design 92
3.2.3 Identification of insertion sites from the library by Illumina sequencing 93
3.2.4 Essential genes and genes require for long-term survival 97

3.2.5 Chloramphenicol resistant and plasmid stability genes on Ty2 chromosome 101
3.2.6 Plasmid mediated cell death and plasmid stability genelist 105
3.2.7 Growth curves of *sfh* knock-out 111
3.3 Discussion 112
4 Plasmid Multi-Locus Sequence Typing for IncHI1 plasmids in *S.* Typhi 119
4.1.1 Introduction 119
4.1.2 Results 123
4.1.3 Discussion 135
4.1.4 Conclusions 139
5 The spread of IncHI1 plasmids in *S.* Typhi 141
5.1 Introduction 141
5.2 Results 142
5.2.1 IncHI1 plasmid SNP typing 147
5.2.2 Mobile elements on IncHI1 plasmids 151
5.2.3 The distribution of plasmids in relation to *S.* Typhi haplogroups worldwide 154
5.2.4 PST6 is linked with H58 haplotype 157
5.2.5 Plasmid competition 158
5.3 Discussion 160
6 Final Discussion 165
7 References 173
8 Appendices 192
8.1 Essential genes from day 0 192
8.2 List of chromosomal genes comparing the differences between non-CmP6 and CmP6 passages 200
8.3 List of plasmid stability genes 200
8.4 TraDIS analysis scripts 200
8.5 Isolates used for SNP typing 200
8.6 SNP data in phylip format 212
8.7 List of SNP used in the GoldenGate assay 212
8.8 Mobile elements and resistant genes detected by GoldenGate assay 212

List of Tables

Table 1-1 Common typing methods 21
Table 1-2 Bacterial plasmid partition systems........ 37
Table 1-3 Types of transposable elements covered by Robert's definition........ 50
Table 2-1 Bacterial isolates used in this study........ 59
Table 2-2 Supplements used in culture media........ 61
Table 2-3 Primers for PCR replicon-typing in three multiplex panels........ 69
Table 2-4 PCR Primers used for 5kb and 10kb PCR around IncHI plasmids........ 70
Table 2-5 Primer combinations for long-range PCRs 72
Table 2-6 Primers for PMLST........ 74
Table 2-7 Primers used to make the *sfh* knock-out........ 76
Table 2-8 Primers to detect mobile elements........ 87
Table 3-1 Top genes on the chromosome contributing to the survival in chloramphenicol passages........ 104
Table 3-2 Top plasmid gene candidates for plasmid stability recovered from control and Cm passages after 6 days........ 105
Table 4-1 Plasmids used in this chapter........ 123
Table 4-2 List of candidate PMLST genes and their annotations........ 126
Table 4-3 List of plasmids for PMLST and their sequence types 129
Table 5-1 Summary of strain collection categorised by haplogroups 143
Table 5-2 Number of plasmids isolated before and after 1993 149
Table 5-3 The presence and structure of mobile elements on various plasmid types 152
Table 5-4 Changes in the ratios of the two plasmids in competition assays 160

List of Figures

Figure 1-1 Diagram showing the classification of the *Salmonella* genus and its diseases. 14
Figure 1-2 Development of antibiotic resistance in *S.* Typhi. 16
Figure 1-3 Geographical distribution of typhoid fever. 21
Figure 1-4 RepA-*oriR* complexes in initiation of R1 plasmid replication. 30
Figure 1-5 Model for rolling-circle replication. 32
Figure 1-6 Replication of plasmid RSF1010 by the strand displacement mechanism. 33
Figure 1-7 A two-step model for conjugal DNA transport. 35
Figure 1-8 Molecular model of plasmid segregation by the R1 par operon (type II partition system). 39
Figure 1-9 Organisation of the tra1 region and the tra2 region. 46
Figure 1-10 The integron-gene cassette site-specific recombination system. 55
Figure 1-11 The mosaic structure of a resistant region in plasmid pHCM1 demonstrating several acquisition events of mobile elements into the plasmid. 56
Figure 3-1 Schematic illustration of TraDIS steps 92
Figure 3-2 Selection assay using the mutant library to investigate plasmid stability 93
Figure 3-3 Saturation of the transposon insertion sites 95
Figure 3-4 Frequency and distribution of transposon directed insert-site sequence reads across the genome of Ty2 and pHCM1 plasmid over time. 96
Figure 3-5 Identification of essential genes 98
Figure 3-6 The number of essential and probably essential genes for survival in rich, non-selective media at different time points during 6 day passages 100
Figure 3-7 Genes on chromosome potentially contributing to survival in chloramphenicol 102
Figure 3-8 Changes in insertion index of candidate genes across passages and in the chloramphenicol selection passage at day 6 (CmP6) 107
Figure 3-9 Changes in number of insertion in top candidates for plasmid stability genes 109
Figure 3-10 Pair-wise comparisons of Sfh protein on pHCM1 with its homologues 110
Figure 3-11 Growth curves of Δ*sfh* mutants (WT282, WT283 and WT284) and wildtype strains 112

Figure 3-12 Comparing essential genes of *S.* Typhi, *S.* Typhimurium and *E. coli* 117
Figure 4-1 Genetic variation in IncHI1 plasmids from *S.* Typhi as determined by microarray analysis. 122
Figure 4-2 Syntenic regions present on each IncHI1 plasmids. 125
Figure 4-3 Modified eBURST diagram for plasmid multi-locus sequence typing....... 128
Figure 4-4 Schematic structure of Tn*6062* 131
Figure 4-5 IncHI1 plasmids phylogenetic tree and observed insertion/deletion events133
Figure 4-6 Screenshots of PMLST website. 135
Figure 5-1 Relationship of plasmid sequence types 147
Figure 5-2 Spatiotemporal distribution of plasmid STs. 150
Figure 5-3 The association of plasmid type and strain background. 155
Figure 5-4 Spatiotemporal distribution of haplogroups of *S*. Typhi........ 156
Figure 5-5 The distribution of individual haplogroups, and their plasmids, across the world. 158
Figure 5-6 Comparing the growth of *S.* Typhi harbouring PST1 and PST6 plasmids. 159

Abbreviations

ALFP	Amplified fragment length polymorphism
CDC	Center for Disease Control and Prevention
CDS	Coding Sequence
Cm	Chloramphenicol
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
HIV	Human Immunodeficiency Virus
Inc	Incompatibility
Km	Kanamycin
MDR	Multi-drug resistance
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MLVA	Multiple loci VNTR analysis
Mpf	Mating pare formation
NCCLS	National Committee for Clinical Laboratory Standards
PCR	Polymerase chain reaction
PFGE	Pulse-field gel electrophoresis
PMLST	Plasmid multilocus sequence typing
PSK	Post-segregational killing
PST	Plasmid sequence type
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
S. Paratyphi A	*Salmonella enterica* serovar Paratyphi A
S. Typhi	*Salmonella enterica* serovar Typhi
SNP	Single nucleotide polymorphism
Tet	Tetracycline
VNTR	Variable number of tandem repeats