Molecular typing and evolution of *Salmonella enterica* serovar Typhimurium

Honghua Hu

School of Molecular and Microbial Biosciences

The University of Sydney

Australia

A thesis submitted for the degree of

Doctor of Philosophy

March 2005

STATEMENT OF ORIGINAL

This thesis contains no material that has been accepted for the award of any other degree or qualification at any other university. To the best of my knowledge and belief, this thesis is original and contains no material previously published or written by another person, except where due references or acknowledgements are given in the text and publications indicated below from this material.

Honghua Hu, Ruiting Lan, and Peter R. Reeves (2002). Fluorescent amplified-fragment length polymorphism analysis of *Salmonella enterica* serovar Typhimurium reveals phage-type specific markers and potential for microarray typing. *Journal of Clinical Microbiology* 40:3406-3415.

Honghua Hu, Ruiting Lan, and Peter R. Reeves (2005). Adaptation of multilocus sequencing to study variation within a major clone: evolutionary relationships of phage types in salmonella enterica serovar Typhimurium. *Genetics* Accepted.

Honghua Hu

March 2005

ACKNOWLEDGEMENTS

First, I would like to offer my appreciation to my supervisors, Professor Peter Reeves and Dr. Ruiting Lan for their excellent guidance throughout my PhD study. Their invaluable knowledge in bacterial genetics and bacterial evolution played a crucial role in the success of my studies. I thank them sincerely for their valuable advices, extensive discussions, critical comments and encouragement throughout the publication and presentation of my research work and especially thanks to Dr. Ruiting Lan for the guide in computer data analysis.

Thanks to all the people in the Discipline of Microbiology, especially the Reeves' lab for providing a warm environment and help whenever I needed it. Especially thanks to Gordon and Kathy for their friendly help in experimental technique skills and computer analysis. Thanks to Jocelyne and Monica for their enthusiastic and friendly support.

I also wish to thank Dr. Chris Murray in the Australian Salmonella Reference Centre located in the Institute of Medical and Veterinary Science, Adelaide, Australia for kindly providing strains. I also would like to thank Dr. Bing Yu in the Dept. of Molecular and Clinical Genetics, Royal Prince Alfred Hospital for his advice in using the SNaPshot method. Thanks also to the Faculty of Agriculture for providing me an A H Thurburn Postgraduate Scholarship.

I would like to dedicate this thesis to my family, especially to my husband Fengsheng Wang for his love and great support throughout and my parents Decan Hu and Xiulin Wang for their selfless love and high expectation and encouragement. Thanks also to my daughter Jenny Wang and my son Matthew Wang for bringing me the happiness and their bearing with me through the hard time.

Table of Content

List of Tables

List of Figures

List of Abbreviations

A	adenosine
AFLP	amplified fragment length polymorphism
APS	Ammonium persulfate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
C	cytidine
CDC	Centers for Disease Control and Prevention
Ci	Curie
CSPD	Disodium 3-[4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricycle decan}-4-yl]phenyl_phosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double stranded
DT	definitive phage type
dUTP	deoxyuridine triphosphate
EDTA	ethylenediaminetetra acetic acid
g	acceleration due to gravity
G	guanosine
hr	hour
Indels	insertions and deletions
IPTG	isopropyl thiogalactoside
IS	insertion sequences
kb	kilobases
LPS	lipopolysaccharide
M	moler
min	minute
ML	maximum likelihood
MLEE	multilocus enzyme electrophoresis
MLST	multilocus sequence typing
MP	maximum parsimony

MQ	milli-Q pure water
NA	nutrient agar
NB	nutrient broth
NJ	neighbour-joining
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAUP	phylogenetic analysis using parsimony
PCR	polymerase chain reaction
PFGE	pulse field gel electrophoresis
PNK	polynucleotide kinase
RAPD	random amplified polymorphic DNA
RDNC	reaction does not confirm
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
sec	second
SNP	single nucleotide polymorphism
SSC	saline sodium citrate
T	thymidine
TAE	Tris-acetate EDTA
TB	terrific broth
TBE	Tris-borate EDTA
TE	Tris EDTA
TEMED	N,N,N,N'-tetramethylethylenediamine
Tn	transposons
U	unit
UPGMA	unweighted pair group method with arithmetic means
UV	ultraviolet light
v/v	volume per volume
w/v	weight per volumn
X-gal (BCIG)	4-bromo-3-chloro-2-isopropyl galactose

Abstract

Salmonella enterica serovar Typhimurium is a common cause of salmonellosis among humans and animals worldwide. In Australia, Typhimurium is responsible for over half of the salmonellosis cases. The Anderson phage-typing scheme is the primary means of long-term surveillance of Typhimurium outbreak isolates, and has played an important role in epidemiology. However, there exist quite a number of strains of Typhimurium that cannot be defined by the phage-typing scheme. Furthermore, the knowledge of evolutionary relationships among isolates of different phage types is still very limited and the genetic basis of phage type variation remains largely unknown. To address these issues, this study focused on molecular typing and evolution of Typhimurium.

Fluorescent amplified-fragment length polymorphism (AFLP) was applied to 46 Typhimurium isolates comprising nine phage types in Australia using the restriction enzymes *Mse*I and *Eco*RI and *Mse*I +1 / *Eco*RI +1 primer pair combinations. The selected phage types, DT9, DT135, DT64, DT44, DT126, DT12a, DT1, DT141 and DT108, have been dominant or frequent phage types in animal and human infections in Australia in recent years. AFLP in the present study showed a very good discrimination power with Simpson index of diversity of 0.98, 35 different AFLP patterns were observed in the 46 isolates studied. The tree based on AFLP patterns showed good correlation with phage type, grouped most Typhimurium isolates by phage type, and differentiated all nine phage types. Furthermore, 84 phage-type specific polymorphic AFLP fragments, for which presence or absence correlated with phage type (including 25 with one exception to phage-type specificity) were observed in the 46 strains studied.

Eighteen phage-type specific AFLP fragments were cloned and sequenced. Sixteen are of known genes or have a homologue in the databases. It was found a predominance of phage and plasmid genes rather than mutational changes in the AFLP fragments studied. Of the 18 cloned and sequenced AFLP fragments, only four relate to mutational changes in the *S. enterica* chromosome, the other 14 comprise DNA of mobile elements: nine are phage related, three are plasmid related and two are gain of DNA from unknown origin.

Twelve of the 18 sequenced phage-type specific AFLP markers are polymorphic because the DNA is present or absent as indicated by Southern hybridization. Two of these markers were successfully used in preliminary PCR-based typing of 30 DT9 and 29 DT135 isolates from worldwide collections. 27 of the 30 DT9 isolates and all DT135 isolates tested were correctly categorized. The results implied a good potential to use the sequence of these fragments as the basis for a multiplex PCR or a microarray based molecular "phage" typing method for Typhimurium.

This thesis also studied the molecular evolutionary relationships among the same set of 46 Typhimurium isolates using mutational changes detected by AFLP, or analysis of intergenic regions and their flanking genes in genome sequences. The complete genome sequence of Typhimurium LT2 was analysed by computer modelled AFLP. The polymorphic AFLP fragments, which matched with the modelled LT2 AFLP fragments, were amplified and sequenced by LT2 genome based primers to determine the changes. Forty-nine intergenic regions with higher pairwise differences between LT2 and Typhi CT18 were amplified and sequenced using LT2 genome based primers for one isolate of each phage type. 51 polymorphic sites were detected consisting of 18 in AFLP fragments and 33 in intergenic regions or their flanking genes. PCR-RFLP (restriction fragment length polymorphism) and SNaPshot were used to further investigate the distribution of the single nucleotide polymorphisms (SNPs) detected in intergenic regions in all isolates studied.

Of the 18 mutational changes detected in AFLP fragments, eight were indels (insertions / deletions) and ten single base substitutions. Of the eight indels, four were in genes, three in intergenic regions, and one covered adjacent intergenic and coding regions. The four indels in genes all caused frameshift mutations, including three single base indels and one 19 bp deletion. Of the ten substitutions, one was in an intergenic region and nine in genes comprising three synonymous and six non-synonymous substitutions.

Of the 33 polymorphic sites detected from sequences of 23 intergenic regions and their flanking genes, one was IS*200* insertion and 32 single nucleotide polymorphisms (SNPs), of which 30 were single base substitutions and two were single base indels. Nine of the 33 variations were found in the flanking genes, which were all single base substitutions comprising four synonymous, four non-synonymous substitutions and one non-sense mutation.

More non-synonymous than synonymous substitutions were found for those in coding regions within Typhimurium, indicating that slightly deleterious intraspecies mutations can be fixed within clones, such as various lineages of Typhimurium.

The 51 polymorphic sites, which were inferred from sequences of both mutation related AFLP fragments, and intergenic regions and their flanking genes, gave a single phylogenetic tree of the 46 Typhimurium isolates studied. All sequences involved were compared with the homologous sequences in the available *S. enterica* genome sequences for serovars Typhi, Paratyphi A, Gallinarum, Enteritidis and Pullorum and this enabled the determination of the direction of the mutational changes in the isolates studied and the root of the phylogenetic tree. There were only two events inferred to have occurred twice, the remaining 49 polymorphisms can be explained by a single event. The data indicated that Typhimurium has a very strong clonal structure with a very low level of recombination over the time for diversification of Typhimurium as majority of clonal variations are from point mutations rather than recombination.

The phylogenetic tree based on mutational changes showed that most Typhimurium isolates of a given phage type are in the same evolutionary group, but that some phage types appear to have arisen more than once. Comparison of the phylogenetic tree with AFLP data gave examples of unrelated isolates of a given phage type having common AFLP fragments comprising plasmid or phage genes, supporting the view that phage type can be determined by presence of specific phages or plasmids. The mutation-based tree showed that six of the nine phage types studied appeared to have a single origin, at least for the isolates studied. It also found that DT1 and DT44 had two independent origins even for the limited set of strains used. The distribution of DT12a isolates into two groups could be explained that the group of three DT12a isolates were derived from the other group of four DT12a isolates, where the root of the tree might be. The data also confirmed that DT64 arose from DT9.

The phylogenetic tree that was generated based on essentially mutational changes provides clear relationships of the closely related Typhimurium isolates with high level of consistency and reasonable confidence. This study provided one of the few analyses of relationships of isolates within a clone. Matching actual AFLP with computer modeled AFLP and sequencing intergenic regions provide very good new strategies to

identify mutational polymorphisms and to study the molecular evolutionary relationships in the closely related isolates.