Infection, Genetics and Evolution 9 (2009) 996–1005

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Discussion

Population structure, origins and evolution of major *Salmonella enterica* clones

Ruiting Lan [a,*], Peter R. Reeves [b], Sophie Octavia [a]

[a] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia
[b] School of Molecular and Microbial Biosciences, University of Sydney, Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 9 April 2009
Accepted 10 April 2009
Available online 23 April 2009

Keywords:
Multilocus sequence typing
Single nucleotide polymorphism
Evolution
Salmonella
Typhi
Typhimurium
Population structure

ABSTRACT

The genus *Salmonella* consists of two species *S. enterica* and *S. bongori*. *S. enterica* has a well defined subspecies structure with seven subspecies consistently delineated by sequence variation. Frequency of recombination between subspecies and within a subspecies is markedly different. Subspecies I undergoes frequent recombination as demonstrated recently, demystifying the long-held belief that *Salmonella* is a highly clonal organism. The majority of disease causing serovars are from subspecies I with the most important serovars in human health being Typhimurium and Typhi. Typhimurium has developed considerable diversity and may be a very old serovar. The majority of the isolates belong to a single clonal complex by multilocus sequence typing. Typhimurium isolates are divided into phage types and some of the phage types do not have a single origin as determined using mutational changes. Phage type DT104 is heterogeneous and represented in multiple sequence types, with its multidrug-resistant variant most successful causing epidemics in many parts of the world. Typhi, a human restricted serovar, is relatively young compared to Typhimurium, and has a low level of sequence variation. Single nucleotide polymorphisms (SNPs) have been shown to be very useful for typing and resolving relationships within Typhi. Genome sequences of 19 isolates revealed more than 1700 SNPs. The fully resolved phylogenetic tree allows one to trace the mutational changes occurred during clonal diversification. Genome wide SNPs have greatly enhanced our understanding of the evolution of *Salmonella* clones.

1. Introduction

The genus *Salmonella* consists of two species *S. enterica* and *S. bongori*. Both species have the capacity to invade host cells due to the presence of the *Salmonella* Pathogenicity Island 1 (SPI1), but only *S. enterica* has the capacity for systemic spread due to the presence of SPI2 (Groisman and Ochman, 1997; Hensel, 2000). However, *Salmonella* is quite variable and only some forms are characteristically pathogenic. *Salmonella* is best known over much of the world, for its capacity to cause human food poisoning. When *Salmonella* passes through the food chain from one of the domestic animals, humans can become an accidental host, causing a usually short term enterocolitis, named salmonellosis. Salmonellosis is one of the most common foodborne bacterial infections worldwide, with the US alone having an estimated 1.4 million cases, resulting in 17,000 hospitalizations and almost 600 deaths each year (Voetsch et al., 2004). Serovar Typhimurium was for many years the most commonly isolated serovar globally, and the spread of the multidrug-resistant DT104 clone is a major public health concern (Threlfall et al., 1994; Helms et al., 2005). However in recent years, serovar Enteritidis has outnumbered Typhimurium in the European Union and many other areas (de Jong and Ekdahl, 2006; Galanis et al., 2006).

Some serovars are host specific or confined to a small number of hosts. Serovars Typhi and Paratyphi A cause enteric fever exclusively in humans. Typhoid fever remains a devastating disease in several regions in Asia, Africa and South America, while the disease is rare in developed countries. The global burden was estimated to be more than 21 million cases and 200,000 deaths in 2000 (Crump et al., 2004). The 2004–2005 outbreak of typhoid fever in Kinshasa, Congo involved 42,564 cases and 214 deaths (WHO, 2005). Although much less common than typhoid fever, paratyphoid fever which is caused by serovar Paratyphi A, is also a significant disease. There were more than 5 million cases in 2000 (Crump et al., 2004), and paratyphoid fever is becoming a problem in Asian countries (Ochiai et al., 2005). Recent studies on the evolution of Typhi have made significant advances (Roumagnac et al., 2006; Octavia and Lan, 2007; Holt et al., 2008) but there is little information on Paratyphi A.

Salmonella serovars are distinguished by antisera to two highly variable surface antigens, the O antigen reflecting variation in the exposed part of the lipopolysaccharide, and the H antigen,

* Corresponding author. Tel.: +61 2 9385 2095; fax: +61 2 9385 1591.
E-mail address: r.lan@unsw.edu.au (R. Lan).

1567-1348/$ – see front matter. Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2009.04.011

reflecting variation in flagellin, the major protein of the flagellum, with about 46 and 114 recognised antigenic forms, respectively (Grimont and Weill, 2007; McQuiston et al., 2004). Most *Salmonella* carry two flagellin genes that code for H1 and H2 variants of the H antigen. Isolates are identified by serotyping, with more than 2500 serovars (each a unique combination of O, H1 and H2 antigens) reported (Grimont and Weill, 2007). Most of the serovars were initially given Latin binomial species names, and these have been retained as serovar names (Le Minor and Popoff, 1987; Brenner et al., 2000). For example, the original *S. typhimurium* is now referred to as *S. enterica* serovar Typhimurium or simply Typhimurium (the latter convention is used in this review). The *Salmonella* serovars were classified into seven subspecies (I, II, IIIa, IIIb, IV, V and VI), based on DNA hybridization and biotyping studies (Crosa et al., 1973; Le Minor et al., 1986). An eighth group, designated as subspecies VII, was identified by multilocus enzyme electrophoresis (MLEE) analysis, but currently contains only four isolates of two serovars initially allocated to subspecies IV on the basis of biochemical characteristics (Boyd et al., 1996). This probably an underestimation as these subspecies were not distinguished by the traits used in the earlier studies, and only 23 of the 71 subspecies IV serovars were examined by MLEE. The relationships of the *S. enterica* subspecies appear to have been finally resolved by using four housekeeping genes and 67 isolates (McQuiston et al., 2008), with only one minor topological difference from an earlier sequence based tree (Boyd et al., 1996).

There has been considerable confusion on *Salmonella* nomenclature but the situation has been clarified by the JCICSP (2005) and subsequent commentaries on taxonomy (Heyndrickx et al., 2005; Tindall et al., 2005). It is now quite clear that the type species of the genus is *S. enterica*, with six named subspecies: *enterica* (subspecies I), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), *indica* (VI), and *salamae* (II), with the original subspecies V becoming *S. bongori*. Subspecies VII was not recognised. However at the sequence level it is as distinct as some other subspecies, but there is no phenotypic description or taxonomic name yet. These recommendations are consistent with earlier recommendations (Brenner et al., 2000).

This review discusses the diversity and the population structure of *S. enterica*, particularly subspecies I, and the evolution and epidemiology of the two major pathogenic serovars Typhimurium and Typhi.

2. Diversity of *S. enterica*

S. enterica is found in reptiles and warm-blooded vertebrates. Most subspecies are only commonly isolated from reptiles, and are often not disease associated, but subspecies I has far more serovars than the others and these are generally isolated from mammals or birds. Ninety-nine percent of salmonellosis is due to subspecies I serovars, with 70% caused by only 12 serovars (Anjum et al., 2005). Therefore, only a small fraction of serovars within subspecies I is pathogenic in this sense. Most serovars are not pathogenic in their natural hosts although a range of serovars can cause disease in domestic animals, and some specific to a particular host (Wray and Wray, 2000). There is a considerable overlap between human and farm animal serovars (Alcaine et al., 2006; Foley et al., 2008), but the frequencies are different. The top three in human infections in 2005 in the US are Typhimurium, Enteritidis and Newport, accounting for nearly half of the human infections. In non-human isolations, the top three are Heidelberg, Typhimurium and Kentucky, accounting for 40% of the total non-clinical isolations (CDC, 2007). The frequencies may be different in different regions. In Australia, the top three human serovars are Typhimurium, Virchow and Saintpaul while the farm animal serovars are Typhimurium, Dublin and Bovismorbificans (Powling, 2005).

Genetic diversity within *S. enterica* natural populations, in particular subspecies I serovars, was extensively analysed by MLEE in the late 1980s (Beltran et al., 1988, 1991; Reeves et al., 1989; Selander et al., 1991). Many serovars were found to be confined within a single cluster of closely related electrophoretic types (ETs), where each has a predominant widely distributed ET (Beltran et al., 1988, 1991; Selander et al., 1990b; Selander and Smith, 1990). However some serovars are genotypically heterogeneous, for example Derby and Newport include divergent isolates with ETs clustered distantly in MLEE trees. Clearly, these serovars comprise two or more clones that probably gained the same antigenic properties independently. However, the focus was on the frequently isolated serovars, with about 30 of the 1500 subspecies I serovars looked at in population genetics terms. Three widely used *Salmonella* strain collections, SARA (Beltran et al., 1991), SARB (Boyd et al., 1993) and SARC (Boyd et al., 1996), were based on the MLEE data, and include representative strains of Typhimurium, subspecies I, and the eight subspecies, respectively. Multilocus sequence typing (MLST) analysis showed sequence level diversity in some serovars (Torpdahl et al., 2005; Harbottle et al., 2006), with strong support for the division into subspecies (Falush et al., 2006; McQuiston et al., 2008).

There are 368 isolates belonging to 111 sequence types (STs) for subspecies I from the publicly available MLST data (note that under the policy of the *Salmonella* MLST database (http://mlst.ucc.ie/mlst/dbs/Senterica), only the published data is freely available for publication and the following analysis relates to that component), and using eBURST (Feil et al., 2004), with six out of seven genes identical as the definition of a clonal complex, we grouped 60 STs into 20 clonal complexes with an average of three STs per clonal complex (ranging from two to nine STs, Table 1), leaving 51 singletons. Three clonal complexes, ST27, 39, and 42 clonal complexes, contain more than one serovar. We constructed a neighbour-joining tree, using a representative ST for each of the 20 clonal complexes, reducing the number to 71 STs. The tree is shown in Fig. 1. It should be noted that, however, the relationships are not robust as indicated by bootstrap values which gave support

Table 1
Clonal complexes identified.

Clonal complex[a]	Total no. of STs	Member STs[b]	Serovars[c]
2	4	1, **2**, 3, 8	Typhi
10	2	10, 73	Dublin
78	2	78, 331	Gallinarum
12	2	12, 33	Hadar
13	2	13, 37	Agona
16	2	16, 38	Virchow
18	2	18, 44	Manhattan
19	9	**19**, 34, 35, 98, 99, 153, 204, 205, 209	Typhimurium
20	2	20, 65	Brandenburg and unknown
21	2	21, 22	Braenderup
23	2	23, 47	Oranienburg
25	2	25, 26	Thompson
27	3	**27**, 49, 50	Saintpaul and Haifa
29	2	29, 51	Stanley
32	2	32, 41	Infantis
39	2	39, 40	Derby, Typhimurium and unknown
42	2	42, 74	Paratyphi B and Dublin
43	3	**43**, 86, 110	Paratyphi B
45	5	**45**, 46, 116, 121, 125	Newport
85	3	**85**, 129, 130	Paratyphi A
118	7	5, 115, 117, **118**, 119, 120, 122	Newport

[a] Clonal complexes are named after the founder if known, or named after the ST with smallest number in that clonal complex if no founder or multiple founders.
[b] Founders are marked bold.
[c] Unknown for STs with no serovar information.