Infection, Genetics and Evolution 11 (2011) 92–102

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

# Evaluation of the taxonomic utility of six-enzyme pulsed-field gel electrophoresis in reconstructing *Salmonella* subspecies phylogeny

Socrates Trujillo *, Christine E. Keys, Eric W. Brown

*Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, MD 20740, USA*

ARTICLE INFO

*Article history:*
Received 30 April 2010
Received in revised form 4 October 2010
Accepted 6 October 2010
Available online 17 October 2010

*Keywords:*
*Salmonella enterica*
PFGE
Taxonomy
Phylogenetic analysis
Serovar
Subspecies
Cladistics
Congruence

ABSTRACT

Pulsed-field gel electrophoresis (PFGE) remains an important tool in the molecular epidemiological evaluation of strains emerging from disease outbreak clusters. Recent studies of *Escherichia coli* O157:H7 and *Salmonella* Enteritidis have noted marked improvements in the discriminatory power of PFGE when combining band profiles from up to six restriction enzyme datasets into a single concatenated analysis. This approach has provided more accurate assignments of genetic relationships among closely related strains and allowed effective phylogenetic inference of host and geographical reservoirs. Although this approach enhances epidemiological congruence among closely related strains, it remains unclear to what extent six-enzyme PFGE pattern similarity reiterates evolutionary relatedness among more distantly related *Salmonella* strains (*i.e.*, serovar or subspecies levels). Here, taxonomic accuracy of six-enzyme PFGE is tested phylogenetically across two distinct *Salmonella enterica* populations–Salmonella reference collection B (SARB), representing the breadth of taxonomic diversity of S. enterica subspecies I only, and Salmonella reference collection C (SARC), comprising the seven disparate subspecies of *S. enterica* plus *S. bongori*. Cladistic analysis of SAR strains revealed substantial polyphyly between the two strain collections such that numerous SARB strains clustered more closely with diverged SARC subspecies rather than with other members of subspecies I. Also, in several cases, SARC sibling strains from the same subspecies were evolutionary obscured—broken into distant locales on the most parsimonious six-enzyme trees. Genetic diversity among SARB and SARC strains was comparable at 45% and 47%, respectively, while congruence testing revealed discordance among individual enzyme datasets. While six-enzyme PFGE is effective in ascertaining accurate genetic relationships for more closely related strains (*e.g.*, strains within the same serovar), reconstitution of evolutionarily meaningful strain groupings may be elusive for *Salmonella* at the serovar level and above. Thus, caution is warranted when applying PFGE with a limited number of enzymes as the primary phylogenetic marker in these instances.

Published by Elsevier B.V.

## 1. Introduction

The re-emergence of foodborne salmonellosis in recent outbreaks associated with fresh-cut produce (Doyle and Erickson, 2008; MMWR, 2008) along with the continued rise of multi-drug resistant strains of *Salmonella enterica*, has catapulted this microorganism once more to the frontlines of public health science. The species *S. enterica* comprises more than 2500 serovars (WHO, 2005) which are partitioned phylogenetically into seven diverged subspecies (I–IIIa, IIIb, IV, VI, and VII) and *S. bongori*, previously identified as subspecies V and now widely accepted as a separate *Salmonella* species. Serovars of *Salmonella* subspecies I are some of the most common food-borne pathogens in the US, causing an estimated 1.4 million human salmonellosis cases with nearly 600 deaths annually and an estimated cost of $2.5 billion dollars in treatment (USDA, 2010).

Evolutionary relationships of *S. enterica* subspecies and strains have been described previously using multi-locus enzyme electrophoresis (MLEE) and DNA sequence analysis of several conserved housekeeping gene sequences known largely to reiterate *Salmonella* strain evolution (Beltran et al., 1991; Boyd et al., 1993, 1996; Brown et al., 2002; McQuiston et al., 2008). Additionally, concatenated gene sequence analyses or multi-locus sequence typing (*i.e.*, MLST) approaches have provided remarkable insight into the phylogenetic and evolutionary relatedness of the salmonellae (Maiden et al., 1998; Kotetishvili et al., 2002). Several reference collections reflecting distinct taxonomic levels have

* Corresponding author at: Division of Microbiology, FDA-CFSAN, 5100 Paint Branch Parkway, Mailstop HFS-712, College Park, MD, USA.
Tel.: +1 301 436 2037; fax: +1 301 436 2644.
*E-mail addresses:* socrates.trujillo@gmail.com, socrates.trujillo@fda.hhs.gov (S. Trujillo).

1567-1348/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.meegid.2010.10.004

precipitated from studies on *Salmonella* diversity. Notably, *Salmonella* reference collection C (SARC), comprising the seven diverged subspecies of *S. enterica* and *S. bongori* (Boyd et al., 1996), *Salmonella* reference collection B (SARB), restricted to subspecies I warm-blooded pathogens only (Boyd et al., 1993), and *Salmonella* reference collection A (SARA), also known as the Typhimurium strain complex—the most homogeneous *S. enterica* reference collection, consisting solely of five closely related subspecies I serovars (Typhimurium, Paratyphi B, Muenchen, Saintpaul, and Heidelberg (Beltran et al., 1991).

The molecular subtyping of bacterial strains has become an essential component of outbreak investigations augmenting the identification and traceback of clusters suspected to originate from foods, the environment, or nosocomial sources (Lukinmaa et al., 2004; Gerner-Smidt et al., 2006). Pulsed-field gel electrophoresis (PFGE) has greatly improved the accuracy of pinpointing bacterial sources of foodborne outbreaks while significantly reducing the amount of time required to complete investigations. The primary epidemiological application of PFGE has been to provide outbreak investigators evidence of the molecular genetic relatedness of two or more strains (Tenover et al., 1995). As such, PFGE has entrenched itself as a powerful tool in the molecular epidemiologic linking of strains during traceback and now serves as a cornerstone in the PulseNet national subtyping network for foodborne diseases (Swaminathan et al., 2001). PFGE is an extremely sensitive technique and is used regularly to discriminate pathogenic strains in numerous serovars of *S. enterica* including *S. enterica* serovar Typhimurium (Best et al., 2009), *S. enterica* Typhi (Kim et al., 2009), and *S. enterica* Newport (Gerner-Smidt et al., 2006) to name a few.

The genetic homogeneity of certain *Salmonella*e, however, confounds many subtyping approaches, including conventional one- and two-enzyme PFGE protocols (Lukinmaa et al., 2004; Zheng et al., 2007). Recently, more discriminatory schemes that incorporate combinations of potentially more informative enzymes have been reported (Zheng et al., 2007; Xi et al., 2008). For example, the discriminatory power created by the simultaneous analysis of six concatenated enzyme data sets allowed for a substantially more informative PFGE-based subtyping scheme for *S.* Enteritidis, a *Salmonella* serovar often associated with undercooked poultry (Zheng et al., 2007; Huehn et al., 2008; Kanki et al., 2009) as well as *S.* Heidelberg, *S.* Saintpaul, *S.* Kentucky, and *S.* Hadar (Xi et al., 2008). While such methods greatly enhance our ability to differentiate *Salmonella* strains and serovars, it remains unclear to what extent PFGE pattern similarity is indicative of actual strain phylogeny and accurate genetic relatedness above the serovar or subspecies levels in *Salmonella*. That is, are PFGE pattern clusters capable of discerning disparate evolutionary lineages of *S. enterica*? The answer to this question is essential to any PFGE approach applied to determining the epidemiological and phylogenetic relationships of feral strains of *Salmonella* isolated in conjunction with surveillance or traceback activities.

Previous studies have noted confounding results when examining isolate pairs with different restriction enzymes (Harsono et al., 1993). In many cases, isolates that are found to be similar on PFGE with one particular enzyme yield widely dissimilar banding patterns with another enzyme. Davis et al. (2003a) noted two potentially aggravating factors of using single-enzyme PFGE: (i) bands that appear equivalent in size may not be homologous in structure; and (ii) loci of identical size are difficult to resolve in a pulsed-field platform. In a subsequent report, these problems were ameliorated by analyzing a concatenated six-enzyme PFGE band matrix to more accurately resolve phylogenetic relationships among geographically disparate strains of *Escherichia coli* O157:H7 (Davis et al., 2003b). Here, we examine empirically the extent to which the simultaneous analysis of concatenated PFGE data sets can accurately resolve evolutionary relationships within two well-characterized and taxonomically distinct *Salmonella* reference collections (*i.e.*, SARB and SARC), taken to represent the breadth of *Salmonella* diversity at the serovar and subspecies levels, respectively. Moreover, we present compelling evidence that concatenated PFGE banding patterns, even among phylogenetic siblings of *Salmonella*e, lack any notable evolutionary congruence and may do little to capture an accurate phylogenetic signal among *Salmonella* strains at serovar or subspecies levels.

## 2. Materials and methods

### 2.1. Bacterial strains

The complete SARB ($n$ = 72 strains) and SARC ($n$ = 16 strains) *S. enterica* reference collections were included in this study (Boyd et al., 1993; 1996) and were acquired from the *Salmonella* Genetic Stock Centre (Calgary University, Alberta, Canada). These strains are listed in Table 1 and, taken together, largely represent the

**Table 1**
Reference strains and collections of *Salmonella enterica* analyzed in this study.

| SAR strain no.[a] | Species name | RKS no. | Subspecies name | Group no.[b] | Serovar name | MLEE cluster[c] |
|---|---|---|---|---|---|---|
| B1 | *enterica* | 1701 | enterica | I | Agona | D |
| B2 | " | 2403 | " | " | Anatum | A |
| B3 | " | 4231 | " | " | Brandenberg | F |
| B4 | " | 1280 | " | " | Choleraesuis | B |
| B5 | " | 1239 | " | " | " | C |
| B6 | " | 3169 | " | " | " | B |
| B7 | " | 4640 | " | " | " | A |
| B8 | " | 4647 | " | " | Choleraesuis | G |
| B9 | " | 246 | " | " | Derby | E |
| B10 | " | 241 | " | " | " | A |
| B11 | " | 243 | " | " | " | B |
| B12 | " | 1518 | " | " | Dublin | A |
| B13 | " | 4717 | " | " | " | " |
| B14 | " | 1550 | " | " | " | " |
| B15 | " | 4239 | " | " | Duisberg | " |
| B16 | " | 53 | " | " | Enteritidis | " |
| B17 | " | 761 | " | " | " | F |
| B18 | " | 69 | " | " | " | A |
| B19 | " | 1208 | " | " | " | " |
| B20 | " | 1216 | " | " | Emek | " |
| B21 | " | 2962 | " | " | Gallinarum | " |
| B22 | " | 4241 | " | " | Haifa | " |
| B23 | " | 539 | " | " | Heidelberg | " |
| B24 | " | 1391 | " | " | " | " |
| B25 | " | 4250 | " | " | Indiana | " |
| B26 | " | 1490 | " | " | Infantis | C |
| B27 | " | 1452 | " | " | " | F |
| B28 | " | 2833 | " | " | Miami | " |
| B29 | " | 4381 | " | " | " | " |
| B30 | " | 1762 | " | " | Montevideo | " |
| B31 | " | 1740 | " | " | " | " |
| B32 | " | 3121 | " | " | Muenchen | A |
| B33 | " | 4288 | " | " | " | G |
| B34 | " | 4300 | " | " | " | A |
| B35 | " | 4272 | " | " | " | " |
| B36 | " | 2016 | " | " | Newport | D |
| B37 | " | 1915 | " | " | " | A |
| B38 | " | 1956 | " | " | " | " |
| B39 | " | 1793 | " | " | Panama | F |
| B40 | " | 1796 | " | " | " | " |
| B41 | " | 1779 | " | " | " | " |
| B42 | " | 4993 | " | " | Paratyphi A | A |
| B43 | " | 3222 | " | " | Paratyphi B | " |
| B44 | " | 3202 | " | " | " | " |
| B45 | " | 3201 | " | " | " | " |
| B46 | " | 3274 | " | " | " | " |
| B47 | " | 3215 | " | " | " | Pb7[d] |
| B48 | " | 4587 | " | " | Paratyphi C | B |
| B49 | " | 4594 | " | " | " | " |
| B50 | " | 4620 | " | " | " | F |
| B51 | " | 2266 | " | " | Pullorum | A |
| B52 | " | 2246 | " | " | " | " |