Infection, Genetics and Evolution 28 (2014) 251–260

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Tracing the sources of human salmonellosis: A multi-model comparison of phenotyping and genotyping methods

Lapo Mughini-Gras [a,b,*], Joost Smid [c], Remko Enserink [a], Eelco Franz [a], Leo Schouls [a], Max Heck [a], Wilfrid van Pelt [a]

[a] National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control (CIb), Bilthoven, The Netherlands
[b] Utrecht University, Faculty of Veterinary Medicine, Department of Infectious Diseases and Immunology, Utrecht, The Netherlands
[c] Utrecht University, Institute for Risk Assessment Sciences, Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 29 September 2014
Accepted 5 October 2014
Available online 12 October 2014

Keywords:
Salmonella
Source attribution
MLVA
Serotyping
Phage typing
Food safety

ABSTRACT

Salmonella source attribution is usually performed using frequency-matched models, such as the (modified) Dutch and Hald models, based on phenotyping data, i.e. serotyping, phage typing, and antimicrobial resistance profiling. However, for practical and economic reasons, genotyping methods such as Multilocus Variable Number of Tandem Repeats Analysis (MLVA) are gradually replacing traditional phenotyping of salmonellas beyond the serovar level. As MLVA-based source attribution of human salmonellosis using frequency-matched models is problematic due to the high variability of the genetic targets investigated, other models need to be explored. Using a comprehensive data set from the Netherlands in 2005–2013, this study aimed at attributing sporadic and domestic cases of *Salmonella* Typhimurium/4,[5],12:i:- and *Salmonella* Enteritidis to four putative food-producing animal sources (pigs, cattle, broilers, and layers/eggs) using the modified Dutch and Hald models (based on sero/phage typing data) in comparison with a widely applied population genetics model – the asymmetric island model (AIM) – supplied with MLVA data. This allowed us to compare model outcomes and to corroborate whether MLVA-based *Salmonella* source attribution using the AIM is able to provide sound, comparable results. All three models provided very similar results, confirming once more that most *S.* Typhimurium/4,[5],12:i:- and *S.* Enteritidis cases are attributable to pigs and layers/eggs, respectively. We concluded that MLVA-based source attribution using the AIM is a feasible option, at least for *S.* Typhimurium/4,[5],12:i:- and *S.* Enteritidis. Enough information seems to be contained in the MLVA profiles to trace the sources of human salmonellosis even in presence of imperfect temporal overlap between human and source isolates. Besides *Salmonella*, the AIM might also be applicable to other pathogens that do not always comply to clonal models. This would add further value to current surveillance activities by performing source attribution using genotyping data that are being collected in a standardized fashion internationally.

1. Introduction

An estimated 93.8 million human cases of *Salmonella* gastroenteritis occur annually worldwide, resulting in approximately 155,000 deaths (Majowicz et al., 2010). In the 27 European Union (EU) member states (~502 million population), over 6.2 million cases are estimated to occur every year, ranking salmonellosis second only to campylobacteriosis among the causes of bacterial gastroenteritis (Havelaar et al., 2013). More than 2500 *Salmonella enterica* subsp. *enterica* serovars have been identified. Those most commonly implicated in human infections are *Salmonella* Typhimurium (ST), including its monophasic variant *S.* 4,[5],12:i:- (ST_{mv}), and *Salmonella* Enteritidis (SE), which account together for ~74% of all *Salmonella* isolates from human cases in the EU (EFSA and EDC, 2013).

To identify the main sources of human salmonellosis and to assess the impact of food safety interventions, source attribution of sporadic cases using the so-called "microbial subtyping approach" is being performed in several countries (David et al., 2013a, 2013b; Guo et al., 2011; Hald et al., 2007, 2004; Mughini-Gras et al., 2014a; Pires and Hald, 2010; Pires et al., 2011; van Pelt et al., 1999; Wahlström et al., 2011). This approach is based on the

* Corresponding author at: National Institute of Public Health and the Environment (RIVM), Centre for Infectious Disease Control (CIb), Epidemiology and Surveillance Unit, PO Box 1 – 3720 BA, Bilthoven, The Netherlands. Tel.: +31 (0)30 274 4350; fax: +31 (0)30 274 4409.
E-mail addresses: lapo.mughini.gras@rivm.nl, lapomugras@libero.it (L. Mughini-Gras).

http://dx.doi.org/10.1016/j.meegid.2014.10.003
1567-1348/

phenotypic and/or genotypic characterization of pathogen isolates from human cases and from a range of putative sources of infection. Given the source specificity of certain subtypes and assuming an unidirectional transmission pathway, from sources to humans (with humans representing the endpoint), the relative contribution of each source to human cases can be inferred probabilistically by comparing the human and source subtype distributions (Barco et al., 2013; Pires et al., 2009). Two frequency-matched models based on microbial subtyping, the Dutch model and Hald (or Danish) model, are being used widely for *Salmonella* source attribution. The Dutch model relies on a relatively straightforward frequentist approach which does not account for differences in the ability of subtypes and sources to infect humans (van Pelt et al., 1999). Conversely, the Hald model is based on Bayesian inference and incorporates prevalence parameters and food consumption weights, accounting explicitly for differences in subtypes and sources to cause human infections (Hald et al., 2004). To further improve identifiability and to handle uncertainty in data of poorer quality, a modified Hald model (mHM) has been developed (Mullner et al., 2009a). More recently, a modified Dutch model (mDM) has been developed and its attributions do not seem to differ significantly from those of the mHM (Mughini-Gras et al., 2014a).

Serotyping is the most popular phenotyping method used for *Salmonella* source attribution (Barco et al., 2013). Phage typing and antimicrobial susceptibility profiling are the traditional methods for *Salmonella* phenotyping beyond the serovar level, and they possess a reasonable discriminatory power for source attribution of ST and SE (David et al., 2013a, 2013b; Hald et al., 2007, 2004; Pires and Hald, 2010; Wahlström et al., 2011). However, phage typing is not always fully reproducible among laboratories (Ross and Heuzenroeder, 2005), and mechanisms such as plasmid exchange, gene mutation, and expression of temperate phages may cause phage conversion in *Salmonella* (Olsen et al., 1993). Although some antimicrobial resistance factors are very stable, others carried on plasmids, integrons and genomic islands can be transferred horizontally between strains, thereby limiting the value of antimicrobial susceptibility for the purpose of source attribution (Barco et al., 2013).

Multi-locus Variable Number of Tandem Repeats Analysis (MLVA) is one of the most common genotyping methods used for public health surveillance of *Salmonella* (Best et al., 2007; Hopkins et al., 2011; Lindstedt et al., 2013; Sintchenko et al., 2012; Torpdahl et al., 2007; Wattiau et al., 2011). The principle behind MLVA is a concurrent analysis of loci with tandem repeated DNA sequences (Variable Number of Tandem Repeats, VNTRs). There are several MLVA schemes for ST and SE, all using different VNTR loci, primers, sequencers, dye chemistries, fragment size standards and ways of assigning allele numbers. For practical and economic reasons, MLVA is replacing phage typing in routine surveillance activities for *Salmonella* in several countries (Lindstedt et al., 2013). In Europe, two standardized protocols are being used for ST (Larsson et al., 2009) and SE (Hopkins et al., 2011), and both are based on a scheme with 5 VNTR loci. MLVA is particularly helpful in supporting outbreak investigations, but the tracing of sporadic cases through the food chain using frequency-matched models is problematic, primarily due to the high variability of MLVA patterns and the uncertainty on the stability over space and time of the genetic targets investigated (Hopkins et al., 2011; Lindstedt, 2005). For this reason, MLVA data are not currently integrated in the Dutch and Hald models, in neither their original nor modified forms, as their computational methods rely on the one-to-one matching of identical strains in humans and sources. Given the multitude of possible MLVA patterns, such one-to-one matching is hardly attainable, if at all. Population genetics approaches are usually best suited to deal with such a diversity of patterns, as they can infer evolutionary and family relationships among different strains, including the occurrence of novel (combinations of) alleles in strains from humans that are unobserved in source populations (Barco et al., 2013).

Wilson et al. (2008) have proposed a (population genetics) model for *Campylobacter* source attribution: the asymmetric island model (AIM). Despite the potentiality of obtainable outcomes (Boysen et al., 2013; Mughini-Gras et al., 2014c, 2013, 2012; Mullner et al., 2009b; Smid et al., 2013), the AIM has never been extended to pathogens other than *Campylobacter* or to data other than Multi-locus Sequence Typing (MLST).

Using a comprehensive data set from the Netherlands in 2005–2013, the aim of this study was to attribute sporadic and domestic human ST/ST_{mv} and SE cases to four putative food-producing animal sources using the AIM supplied with MLVA data in parallel with the mDM and mHM based on sero/phage typing data. This allowed us to compare model outcomes and, specifically, to corroborate whether MLVA-based *Salmonella* source attribution using the AIM is able to provide sound, comparable results.

2. Materials and methods

2.1. Serotyping and phage tying data

A data set consisting of 14,061 sero/phage typed human salmonellosis cases that occurred in the Netherlands between January 2005 and December 2013 was obtained from the EU and national reference laboratory for *Salmonella* at the Dutch National Institute for Public Health and the Environment (RIVM). Cases were identified through passive surveillance in the Dutch general population. Concurrent (2005–2013) sero/phage typed *Salmonella* isolates from four food-producing animal sources, i.e. pigs ($n = 2871$), cattle ($n = 855$), broilers ($n = 2329$) and layers/eggs ($n = 942$) were provided by the Dutch Veterinary Services as part of diagnostic activities and a diversity of surveillance programmes on farms, slaughterhouses and at retail. Both human and animal isolates were sent to the RIVM for serotyping and further phage typing of ST and SE as described in van Duijkeren et al. (2002). Phage typing was discontinued from 2010 (for ST) and from 2013 (for SE) onwards.

Of the 14,061 human cases, 1733 (12.3%) and 1332 (9.5%) were discarded from the source attribution analysis because they were travel- and outbreak-related, respectively. Another 749 (5.3%) cases, mainly of non-*S. enterica* serovars, were discarded because there was no one-to-one matching with any of the considered animal sources, and therefore they could not be attributed using the mDM and mHM. These cases were therefore assigned to an "unknown" source. The final data set comprised 172 different *Salmonella* sero/phage types found in humans and in at least one of the considered animal sources; a total of 10,247 sporadic and domestic human cases were then attributed to the four animal sources (Table 1).

2.2. MLVA data

A large subset of sporadic and domestic human cases of ST/ST_{mv} ($n = 4214$) and SE ($n = 995$) that occurred in the Netherlands between January 2005 and December 2013 (ST/ST_{mv}) and between January 2011 and December 2013 (SE) were typed with MLVA by the RIVM using EU standardized protocols (Hopkins et al., 2011; Larsson et al., 2009). Contemporaneous ST/ST_{mv} ($n = 1294$) and SE ($n = 159$) isolates obtained from the four animal sources were also typed with MLVA (Table 2). There were 1562 different MLVA patterns among all (human and animal) isolates used in the source attribution analysis; 1450 among ST/ST_{mv} isolates and 112 among SE isolates, corresponding to one new MLVA pattern every ~4 ST/ST_{mv} and ~10 SE isolates (Table 2).