Food Microbiology 71 (2018) 60–67

Contents lists available at ScienceDirect

Food Microbiology

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

New paradigms for *Salmonella* source attribution based on microbial subtyping

Lapo Mughini-Gras [a, b, *], Eelco Franz [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, Department of Infectious Diseases and Immunology, Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form
24 February 2017
Accepted 3 March 2017
Available online 19 March 2017

Keywords:
Salmonellosis
Source attribution
MLVA
Case-control study
Reptiles

ABSTRACT

Microbial subtyping is the most common approach for *Salmonella* source attribution. Typically, attributions are computed using frequency-matching models like the Dutch and Danish models based on phenotyping data (serotyping, phage-typing, and antimicrobial resistance profiling). Herewith, we critically review three major paradigms facing *Salmonella* source attribution today: (i) the use of genotyping data, particularly Multi-Locus Variable Number of Tandem Repeats Analysis (MLVA), which is replacing traditional *Salmonella* phenotyping beyond serotyping; (ii) the integration of case-control data into source attribution to improve risk factor identification/characterization; (iii) the investigation of non-food sources, as attributions tend to focus on foods of animal origin only. Population genetics models or simplified MLVA schemes may provide feasible options for source attribution, although there is a strong need to explore novel modelling options as we move towards whole-genome sequencing as the standard. Classical case-control studies are enhanced by incorporating source attribution results, as individuals acquiring salmonellosis from different sources have different associated risk factors. Thus, the more such analyses are performed the better *Salmonella* epidemiology will be understood. Reparametrizing current models allows for inclusion of sources like reptiles, the study of which improves our understanding of *Salmonella* epidemiology beyond food to tackle the pathogen in a more holistic way.

Contents

1. Introduction

1.1. Scopes of source attribution

Source attribution of *Salmonella*, as well as of other zoonotic pathogens of public health significance, is a subject area of epidemiological research that is gaining momentum by incorporating a

* Corresponding author. National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control (CIb), PO Box 1 - 3720 BA, Bilthoven, The Netherlands.
E-mail address: lapo.mughini.gras@rivm.nl (L. Mughini-Gras).

http://dx.doi.org/10.1016/j.fm.2017.03.002
0740-0020/

growing number of methodological approaches and data types. Source attribution is defined as the process of inferring the relative contributions of different animal, food and/or environmental sources of infection to the human disease burden of a given pathogen. Therefore, the main goal of a *Salmonella* source attribution analysis is typically the one of partitioning human salmonellosis cases over a number of putative sources of infection representative of the epidemiological context in question. Quantitative estimates of the relative contributions of different sources to human salmonellosis morbidity is crucial for risk managers to set priorities for public health interventions, as well as to enable the measurement of the impact of such interventions. A detailed overview of definitions, terminology, and methodologies for source attribution has been presented elsewhere (Pires, 2013; Pires et al., 2009). While animals are usually defined as reservoirs or amplifying hosts, the environment, food and direct contact with animals are examples of transmission routes; meat, dairy, eggs, etc. are examples of exposures (vehicles), whilst consumption of undercooked pork, unpasteurized dairy products, etc. are examples of risk factors. In practice, however, the term "source" is used generically to refer to any point across the transmission chain, from the original reservoir up to the risk factor ultimately allowing for transmission of the pathogen to humans.

There are several methodological approaches for source attribution (Pires et al., 2009), and the best approach always depends on the data availability and research question to be addressed (Pires, 2013). These approaches are listed as follows:

- Microbial subtyping (top-down approach), e.g. Barco et al. (2013).
- Comparative exposure assessment (bottom-up approach), e.g. Pintar et al. (2016).
- Case-control/cohort studies (traditional epidemiological approach), e.g. Doorduyn et al. (2006).
- Analysis of outbreak investigations (meta-analytical approach), e.g. Domingues et al. (2012); King et al. (2011).
- Intervention studies ("learning through experience" approach), e.g. Sears et al. (2011); Tustin et al. (2011); van Pelt et al. (2004); Vellinga and Van Loock (2002).
- Expert elicitations ("last resort", when no empirical data is available), e.g. Butler et al. (2015); Havelaar et al. (2008).

For *Salmonella*, source attribution based on microbial subtyping is the approach that is mostly applied (Barco et al., 2013) and will therefore be the focus of this paper. This approach requires subtyping of a representative collection of *Salmonella* isolates from human cases and their putative sources of infection. Historically, *Salmonella* subtyping for the purposes of source attribution has mainly been based on phenotyping, i.e. serotyping, phage typing of the common serotypes (e.g. *S.* Typhimurium and *S.* Enteriditis), and antimicrobial resistance profiling. Yet, in recent years, genotyping and particularly Multiple-Locus Variable Number of Tandem Repeats Analysis (MLVA) has become the fashion for *Salmonella* subtyping beyond serotyping (de Knegt et al., 2016; Mughini-Gras et al., 2014c). Source attribution thus relies on the comparison of distributions of pathogen subtypes between human and source isolates through mathematical models that can infer probabilistically the likely sources of the human subtypes based on the distribution of these subtypes in the sources, assuming a unidirectional transmission pathway from sources to humans. Source attribution models can also incorporate epidemiological (meta-)data, as well as data on prevalence, food consumption, import-export flows, etc. as to better inform the attributions. Moreover, source attribution models do not usually address the contribution of anthroponotic sources (e.g. a-, pre- or post-symptomatic carriers), thereby focusing on zoonotic transmission only. This is because humans are usually "the endpoint" in the transmission of the pathogens attributed, i.e. humans do not represent reservoirs (amplifying hosts) for these pathogens, so a negligible contribution of human-to-human transmission can be assumed *a priori*. In any case, the contribution of anthroponotic sources to human salmonellosis morbidity (i.e. secondary transmission) can be best assessed in case-control studies (see Section 2.2.) by looking at the differences in exposure to gastroenteritis patients between cases and controls. For instance, the population attributable fraction (PAF) of "contacting people with gastroenteritis outside the household" has been estimated at 7% (95% confidence interval 3–9%) for human salmonellosis (Mughini-Gras et al., 2014b).

1.2. Brief overview of source attribution models

Looking at the "publication rate" of PubMed-indexed papers on *Salmonella* source attribution (as given by the number of PubMed-indexed papers on source attribution over the total number of PubMed-indexed papers on *Salmonella*, Fig. 1), it is revealed that this started increasing only by the second half of the nighties. In 1999, the so-called Dutch model was published (van Pelt et al., 1999). The Dutch model and its adaptations have been described in detail elsewhere (Barco et al., 2013; Mughini-Gras et al., 2014a, 2014b, 2014c; Mughini-Gras and van Pelt, 2014; Mullner et al., 2009b; Vieira et al., 2016). Briefly, in its simplest form, this model estimates the expected number of human cases caused by subtype i originating from source j, denoted as λ_{ij}, as follows:

$$\lambda_{ij} = \frac{r_{ij}}{\sum_j r_{ij}} \times e_i$$

where r_{ij} is the relative frequency of subtype i in source j and e_i is the observed frequency of human cases of subtype i. A sum over subtypes gives the total number of cases attributable to reservoir j; 95% confidence intervals can be estimated by bootstrapping. While the Dutch model always provides attributions in a rather proportional and straightforward way, it makes the arguable assumption of equal impact of the different subtypes and sources to the human population. However, this assumption has been mitigated in several modified versions of the Dutch model by incorporating modelled subtype-specific prevalences (instead of just relative frequencies of subtypes) and food consumption weights, including both the overall amount of food consumed and its likelihood to be consumed raw/undercooked by the population (Mughini-Gras et al., 2014a, 2014b, 2014c; Mughini-Gras and van Pelt, 2014).

As showed in Fig. 1, one of the main milestones in the timeline of source attribution research is certainly the development of the Danish or Hald model, which was published in 2004 (Hald et al., 2004). This model put a spotlight on the microbial subtyping approach and it relies on the same type of data as the Dutch model, but it uses a Bayesian approach to attribute stochastically human cases to putative sources of infection, while accounting for differences in subtypes and sources to cause human infection. The Danish model assumes that the expected number of human cases of subtype i, denoted as λ_i, is given by:

$$\lambda_i \sim Poisson\left(\sum_j \lambda_{ij}\right)$$

$$\lambda_{ij} = p_{ij} \times M_j \times q_i \times a_j$$

where p_{ij} is the prevalence of subtype i in source j, M_j is the amount

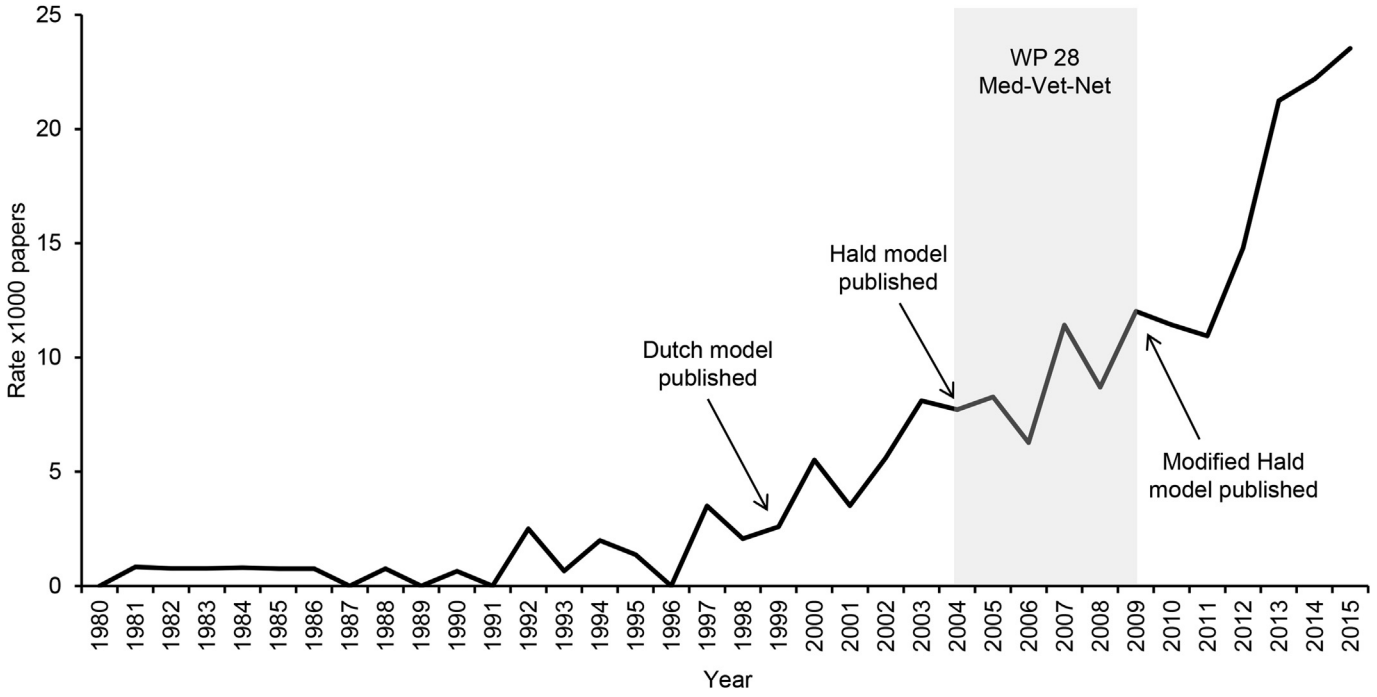

Fig. 1. Number of PubMed-indexed papers on source attribution over the number of PubMed-indexed papers on *Salmonella* per 1000 papers.

of food consumed from source j, q_i is the serotype-dependent factor, which putatively accounts for differences in the success of subtype i to infect humans (e.g. survivability, virulence and pathogenicity or population partial immunity), and a_j is the source-dependent factor, which putatively accounts for the ability of sources j to act as a vehicle for *Salmonella* (e.g. differences in pathogen load, magnitude of exposure, source characteristics influencing pathogen growth, preparation/handling procedures, differences in sensitivity of surveillance programs and randomness of sampling schemes). Posterior distributions for the parameters of interest are obtained by a Markov Chain Monte Carlo (MCMC) simulation. There have also been several adaptations of the Danish model depending on the data and epidemiological context in question. The most successful and portable adaptation is perhaps the one proposed in 2009 by Mullner et al. (2009a), the so-called "modified Hald model", which greatly improved model identifiability and helped handling uncertainty in data of poorer quality, as well as data for pathogens other than *Salmonella*. Although the Danish model and its modified versions may sometimes be problematic in terms of computational requirements, they have been widely applied to *Salmonella* source attribution in many parts of the world, including several European countries (David et al., 2013a, 2013b; de Knegt et al., 2016; Hald et al., 2004, 2007; de Knegt et al., 2015; Mughini-Gras et al., 2014a, 2014b; Pires and Hald, 2010; Ranta et al., 2011; Wahlstrom et al., 2011), the United States of America (Guo et al., 2011), Australia (Glass et al., 2016) and New Zealand (Mullner et al., 2009a).

A common trait between the Dutch and Danish models is that they are both frequency-matching models, meaning that they rely on the one-to-one matching of the same subtypes in humans and sources. While this means that those human cases infected with subtypes found only in humans cannot be attributed to sources, the resulting non-attributable fraction may give some clues about how many cases may be attributed to sources other than those included in the models. However, the extent of such non-attributable fraction strongly relies on the discriminatory power of the subtyping method used (de Knegt et al., 2016), making it more likely to reflect the characteristics inherent in the subtyping method itself rather than a directly interpretable piece of outcome.

A much different source attribution modelling framework is the one provided by population genetics models, such as the asymmetric island model (Wilson et al., 2008). The asymmetric island model is a coalescent-based model derived from a generalization of the Wright's island model that incorporates a Bayesian approach for reconstructing the genealogical history of the isolates using their allelic profiles. By accounting for mutation, recombination and migration events, this model can cope with the occurrence of (combinations of) novel alleles in isolates from humans that are not found in sources (Wilson et al., 2008).

As shown in Fig. 1, from 2011 onwards, there has been an explosion of published articles on *Salmonella* source attribution. This happened after a period of fervent activity on source attribution (2004–2009), also thanks to the work-package 28 of the Med-Vet-Net (European Network of Excellence for Zoonoses Research) project, as several PhDs and researchers were appointed in different countries to work on this topic. It was during this period that most source attribution tools were developed, calibrated and/or perfectioned (see, for example, David et al., 2013a, 2013b; Hald et al., 2007; Mullner et al., 2009a, 2009b; Pires and Hald, 2010; Pires et al., 2009). While all these advancements have definitely paid off in terms of scientific output, novel paradigms are being offered to source attributors.

2. Three new paradigms for *Salmonella* source attribution

There are at least three major challenges facing *Salmonella* source attribution today, which can be summarized as follows:

1. The use of genotyping data for *Salmonella* source attribution, particularly the use of MLVA data, a typing scheme that is increasingly being used in *Salmonella* surveillance and outbreak investigation.
2. The integration of epidemiological data into source attribution to improve the identification and characterization of risk factors and transmission routes for human *Salmonella* infection.

3. The investigation of the role of non-food sources of human salmonellosis, as typically *Salmonella* source attribution focuses only on foods of animal origin.

This paper will critically review the current state-of-knowledge for each of the aforementioned three points, trying also to delineate the bottlenecks and ways forward for the years to come.

2.1. *Salmonella* source attribution using MLVA

MLVA is amongst the most popular genotyping methods used in public health surveillance and outbreak investigation of *Salmonella* (Hopkins et al., 2011; Lindstedt et al., 2013; Torpdahl et al., 2007). The principle behind MLVA is a concurrent analysis of loci with tandem repeated DNA sequences (Variable Number of Tandem Repeats, VNTRs). An MLVA profile is usually expressed as a series of numbers of length *n* representing the number of copies of repeated sequences at each of a set of *n* loci under analysis. In Europe, two standardized protocols are being used for MLVA of *S*. Typhimurium (scheme STTR9/STTR5/STTR6/STTR10/STTR3) (Larsson et al., 2009) and *S*. Enteritidis (scheme SE9/SE5/SE2/SE1/SE3) (Hopkins et al., 2011), both based on a five (VNTR) loci scheme. For practical and economic reasons, MLVA is replacing traditional phage typing in routine surveillance activities for *Salmonella* in several countries (Lindstedt et al., 2013). Tracing the sources of (sporadic) human salmonellosis using MLVA data in the frequency-matching models (i.e. the Dutch and Danish models) is problematic due to the high variability of the MLVA profiles and the instability over space and time of the genetic targets investigated (Hopkins et al., 2011; Lindstedt et al., 2013). This makes the required one-to-one matching of identical subtypes in humans and sources quite hard to attain. Therefore, other modelling options need to be explored, and at least two research teams in the Netherlands and Denmark have performed this.

Using surveillance data from the Netherlands in 2005–2013, Mughini-Gras et al. (2014c) attributed human *S*. Typhimurium/4,[5],12:i:- and *S*. Enteritidis infections to four putative food-producing animal sources (pigs, cattle, broilers, and layers/eggs) using modified versions of the Dutch and Hald models based on serotyping and phage typing data in comparison with the asymmetric island model supplied with MLVA data. This allowed the authors to corroborate whether MLVA-based *Salmonella* source attribution using the asymmetric island model is able to provide results comparable with those of the frequency-matching models based on phenotyping information. As all the models provided very similar results and no particular warning was prompted by their empirical cross-validation, it was concluded that MLVA-based source attribution using the asymmetric island model is a feasible option, at least for *S*. Typhimurium/4,[5],12:i:- and *S*. Enteritidis, and that enough information is contained in the five VNTR loci to let this model make sound inferences. In general, the main advantage of using such a population genetics approach is that it is best suited to deal with genetic targets with a high discriminatory level, without having to make any (somewhat questionable) adjustment to such level to meet the requirements of the frequency-matching models. This means that through the modelling of evolutionary relations among isolates, it is possible to infer the sources of all these isolates, including those from humans whose subtypes were not observed in the sources. However, the use of the asymmetric island model with fast-changing and highly variable genetic markers (i.e. MLVA) is also somewhat questionable, as this is a modelling framework originally designed for use with more conserved and slowly evolving genes in the core-genome, as reflected by Multi-Locus Sequence Typing (MLST). MLST has emerged in recent years as the standard for phylogenetic studies of several bacterial species by characterizing isolates using DNA sequence analysis of internal fragments of (usually) five to seven housekeeping genes (Feil, 2004). For each housekeeping gene, the sequences are given as alleles, and the allele numbers at the different loci constitute the allelic profiles of the strains; sequences differing at even a single nucleotide are assigned to different alleles. While the large number of alleles that could potentially occur at each locus has the potential to identify many different allelic profiles, the discriminatory power of MLST is naturally limited by the slowness of accumulating nucleotide changes in the housekeeping genes. This is why MLST is well suited to reconstruct the evolutionary history and family relationships of relatively distant strains, but it is less suited to discriminate among closely related strains in, e.g., outbreak situations. In contrast, MLVA profiles change rapidly enough to identify bacterial clones, e.g. to distinguish epidemic strains from the endemic ones, offering a much higher discriminatory alternative by indexing the variation in the number of tandem repeats found mostly in noncoding genome segments. There are therefore three fundamental differences between MLVA and MLST: (1) the genetic information expressed in their profiles (allele numbers for MLST *vs*. repeat copy numbers for MLVA), (2) the magnitude and speed at which these profiles are expected to vary (limited and slow for MLST *vs*. large and fast for MLVA), and (3) the possible interpretation of such variations (evolutionary changes for MLST *vs*. transitory clonal expansions for MLVA). Because of these differences, it is apparent that MLVA data are not ideal for use with the asymmetric island model.

Since a single mutation or recombination event may delete or add either a single or multiple repeats at once, but the nature of the underlying mechanisms is unknown, modelling MLVA profiles in terms of number of different loci, rather than different repeats, is most advisable when using the asymmetric island model (Mughini-Gras et al., 2014c). Moreover, mutation and recombination events can generate either single- or multiple-locus variants, with recombination potentially masking mutation and *vice versa*. To discern recombination from mutation, the asymmetric island model looks for the allele in question elsewhere in the data set, assuming that such recombination event will result in an allele that is already present in the pool. However, the frequency of recombination is relatively low in *Salmonella*, and VNTR variation is more likely to arise from mutation events such as slipped strand mispairing. It follows, therefore, that the asymmetric island model's assumption that an allele observed previously at a given locus in a different allelic profile is due to recombination is the specific assumption that does not apply to MLVA in the same way as it does to MLST.

Moreover, currently the model cannot deal with serotypes other than *S*. Typhimurium/4,[5],12:i:- and *S*. Enteritidis, and since different loci are used for *S*. Typhimurium/4,[5],12:i:- and *S*. Enteritidis, a model for each serotype is to be run. Finally, it appears that sample size can really be a limiting factor in the successful application of the asymmetric island model using MLVA. Indeed, an Italian study on the sources of *S*. Typhimurium and its monophasic variant (Barco et al., 2015) used a much smaller data set (268 human and 325 animal isolates) than the one used in the Dutch study (4214 human and 1294 animal isolates) (Mughini-Gras et al., 2014c) and noticed considerable uncertainty around the attributions for sources other than the largest one (i.e. pigs), and only when expanding the dataset with bootstrapping, this uncertainty was reduced to reasonable levels. An alternative to the asymmetric island model could be the use of the STRUCTURE algorithm (Pritchard et al., 2000), one of the first model-based clustering methods for using multilocus genotype data to infer population structure and to assign individuals to populations. The model assumes the existence of *k* populations, each of which is represented

by the allele frequencies at each locus; isolates are then probabilistically assigned to these populations, or jointly to two or more populations if their genotypes are admixed. The application of this model for source attribution represents a special case in which a model without admixture is usually specified, meaning that the source isolates are assumed to belong to only one of the *k* populations (i.e. sources in this case). The allelic frequencies at each locus are then determined for each of the *k* sources, allowing for the estimation of relative probabilities for the attributable isolates (e.g. human cases) to originate from each of the *k* populations based on the allelic profiles in question and the specific allelic frequencies at each locus. The model does not assume a particular mutation process, and it can be applied to most genetic markers. Yet, its specific application on *Salmonella* MLVA data remained to be explored. In Denmark, de Knegt et al. (2016) developed a modelling framework for *Salmonella* source attribution based on the Hald model using a combination of serotyping, MLVA, and antibiotic resistance profiling data. Full and simplified MLVA schemes were assessed in relation to model fit and robustness of results. Simplified MLVA schemes were composed by five (*S.* Enteritidis and *S.* Typhimurium), four, and three VTNR loci (*S.* Typhimurium) to create MLVA profiles with varying discriminatory power, thereby allowing them to group under broader subtypes, as the standard five VTNR loci scheme seemed to be too discriminatory for the purposes of source attribution. The simplified MLVA schemes were based on the assumption that because an unstable locus changes more frequently than a stable one, a larger variety in VNTRs can be expected therein, meaning that this would be more discriminatory as it produces a larger number of different subtypes. Grouping those subtypes that vary greatly at an unstable locus (e.g. STTR6 or STTR5), while keeping the stable ones, reduces the number of subtypes, i.e. decreases the discriminatory level, as observed in the scheme STTR9/STTR10/STTR3. Conversely, grouping subtypes based on a stable locus, while allowing more unstable loci to vary, results in the formation of more subtypes, thus retaining the discriminatory potential, as for the scheme STTR5/STTR10/STTR3. After performing several simulations, the Danish authors concluded that the MLVA schemes STTR5/STTR10/STTR3 for *S.* Typhimurium and SE9/SE5/SE2/SE1/SE3 for *S.* Enteritidis could be best used in a Hald model with an "adjusted discriminatory level" for use with MLVA data (de Knegt et al., 2016). This approach has the advantage that relies on a well-known (and largely empirically supported) model. Moreover, the results do not seem to deviate from what can be obtained by phage typing, and in the situations where they do, model fit indicates that this is related to a higher specificity of MLVA (de Knegt et al., 2016). Conversely, the main disadvantages are that the simplification of the MLVA schemes entails some degree of loss of information, and that a number of cases cannot be attributed to sources, with such number being expected to be larger in scenarios with a higher discriminatory level, as perfect matches between subtypes in humans and sources are generally difficult to find. This also means that this approach is far less scalable to even more discriminatory typing schemes, such as extended MLST or Whole-Genome Sequencing (WGS), which are increasingly being applied and are highly likely to become the standard for pathogen subtyping in the near future.

WGS reveals the complete genetic make-up of an organism. The biggest advantage of WGS for a pathogen like *Salmonella* is that its typing can be performed at a much higher resolution than with traditional molecular typing methods, including MLVA. The application of WGS in *Salmonella* source attribution is still largely unexplored. The reasons for this are debatable, but include the generally scarce availability of modelling tools suitable for high-throughput data that are produced at a rate much faster than that with which source attribution modelling is able to cope, as well as the issue of defining the optimal discrimination level for WGS-based source attribution as to reflect the level of clonality and degree of host association of *Salmonella*. This requires a way to define host associations per locus, Single Nucleotide Polymorphism (SNP), or groups of loci (i.e. genome-wide MLST) or SNPs. Innovative approaches, like machine-learning methods, need to be developed in order to make full use of WGS in source attribution studies.

2.2. Combining case-control and source attribution data

Typically, case-control studies can only trace back the sources of human *Salmonella* infections up to the level of risk factor (e.g. consumption of specific food items, contact with animals, etc.), which, however, may not point to the original reservoirs because of, e.g., cross-contamination and alternative transmission pathways. On the other hand, source attribution based on microbial subtyping allows us to determine the relative contributions of different reservoirs to the human disease burden, i.e. to attribute the human cases up to the beginning of the transmission chain. Combining source attribution and case-control data would allow us to reconstruct the underlying transmission pathway, from a given reservoir up to the point of risk factor, providing a more complete epidemiological picture than when performing separate analyses. This type of combined analysis is called "source-assigned case control study" and has been explored for both *Campylobacter* (Mossong et al., 2016; Mughini Gras et al., 2012) and *Salmonella* (Mughini-Gras et al., 2014b), showing that the outcome of classical case-control studies can be greatly enhanced by incorporating source attribution data. The principle is to first attribute human cases included in a case-control study to sources using the microbial subtyping approach in order to determine their likely sources, and then to compare the exposures of the attributed cases with those of the controls to identify source-specific risk factors for infection, as well as to infer the underlying transmission pathways.

In 2002–2003, a large case-control study was performed in the Netherlands to identify risk factors for human salmonellosis (Doorduyn et al., 2006). About ten years later, the same data were re-analysed in combination with source attribution (Mughini-Gras et al., 2014b). In total, 414 sero- and phage-typed human *Salmonella* isolates included in the case-control study were attributed to pigs, cattle, broiler, or layers using a modified version of the Dutch model. The posterior source probabilities obtained from the source attribution analysis were then used as weights in a logistic regression analysis to study risk factors for salmonellosis of pig, cattle, broiler or layer origin, using a total of 3165 controls as common comparison group (Mughini-Gras et al., 2014b). In this way, several (source-specific) risk factors were identified (Fig. 2). For instance, eating habits like consuming raw or undercooked meat, as well as substandard kitchen hygiene practices like changing kitchen rags less often than once a week or not using a chopping board for raw meat only, were specific risk factors for infection with salmonellas originating from the meat-producing animals under study, that is, pigs, cattle and broilers. Moreover, consuming raw or undercooked eggs and products thereof were specific risk factors for layer/egg-associated salmonellosis. Non-food related risk factors, such as playing in sandpits, contact with people with gastroenteritis outside the household, and contact with canine puppies were significant risk factors for infection with *Salmonella* strains originating from reservoirs other than layers. Moreover, occupational exposure to farm animals appeared to be a specific risk factor for cattle-associated salmonellosis. Altogether, these findings suggest that the *Salmonella* strains originating from the meat-producing animals are somewhat more likely to be transmitted to humans via non-foodborne routes as well. Finally, while using antibiotics was a risk factor for pig- and cattle-

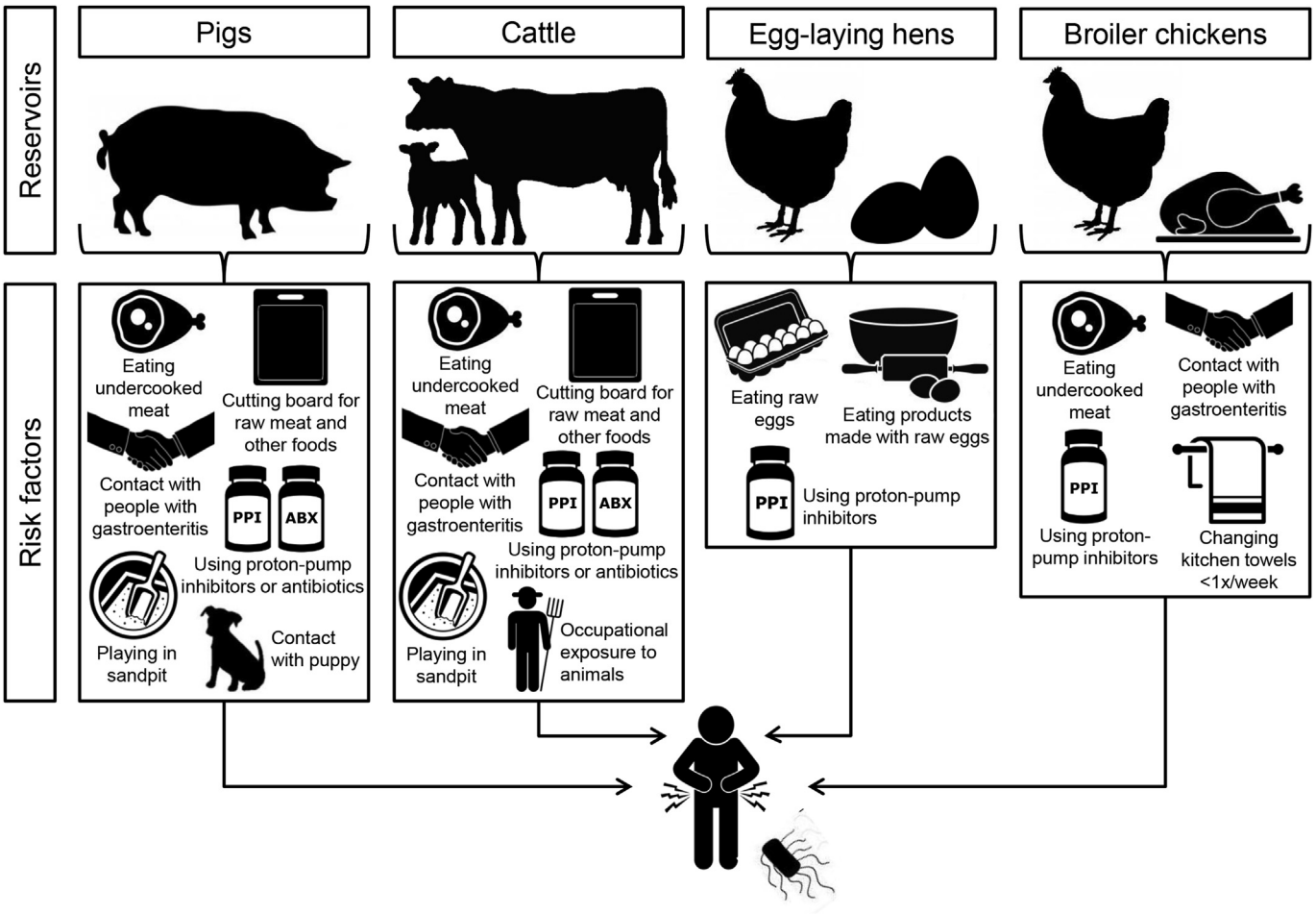

Fig. 2. Schematic exemplification of the source-specific risk factors for human salmonellosis, adapted from Mughini-Gras and van Pelt (2014).

associated salmonellosis, using proton-pump inhibitors increased the risk of acquiring salmonellosis attributable to any reservoir. The authors therefore concluded that individuals acquiring salmonellosis from different reservoirs have different associated risk factors, suggesting that salmonellas may infect humans through various transmission pathways depending on their original reservoirs, with a clear distinction between a predominantly foodborne pathway for salmonellas originating from layers and a substantial non-foodborne component in the epidemiology of salmonellas originating from meat-producing animals (Mughini-Gras et al., 2014b).

It is conceivable that the more such combined studies will be conducted, the more the epidemiology of salmonellosis will be elucidated (as risk factors can be identified in a more nuanced way), although regrettably there is still the tendency to perform case-control studies for human salmonellosis independently of source attribution and *vice versa*.

2.3. Investigating non-foodborne sources of human salmonellosis

While the contribution of the main food-related sources to human salmonellosis is relatively well documented, limited knowledge exists on the contribution of non-food related animal sources like pets, including exotic pets. This is mainly due to the lack of representative and routinely collected data for animals that do not produce food, as well as for animals that are not considered in the current legislation for *Salmonella* control in livestock. As a consequence, a part of the complex *Salmonella* epidemiology is left unexplored in most source attribution studies.

A recent study (Mughini-Gras et al., 2016) quantified and examined trends in reptile-associated salmonellosis in the Netherlands during a 30-year period (1985–2014). The authors adapted the Dutch model to accommodate for non-food sources, similar to the adaptations of the Hald model to allow for the inclusion of the environment as a potential source for human *Campylobacter* infections (Mullner et al., 2009a,b). It was estimated that 2% of all sporadic/domestic human salmonellosis cases reported in the Netherlands during the study period (n = 63718) originated from reptiles, and that there was a significantly increasing trend (+19% each year) in reptile-associated salmonellosis cases. Besides this increase, a shift toward adulthood in the age groups at highest risk was observed, with the proportion of reptile-associated salmonellosis cases among the ≤4 year olds decreasing by 4% annually and the proportion of cases aged 45–74 years increasing by 20% each year. The authors hypothesized that these findings may be the effect of the increased number and variety of reptiles that are nowadays kept as pets, calling for further attention to the issue of safe reptile-human interaction as to target and reinforce current standing recommendations (Mughini-Gras et al., 2016). This provides an example of the benefit of considering also non-food sources in *Salmonella* source attribution analyses, as the significant increase and changing epidemiology of reptile-associated salmonellosis seen in the Netherlands would have otherwise probably passed undetected or, even worse, become apparent when the situation would have already been out of

control. In situations where no typing data is available, which is often the case for sources like reptiles, quantitative microbial risk assessment (QMRA) may provide an alternative to perform source attribution using a bottom-up approach. QMRA models incorporating prevalence for the different subtypes in the different sources are difficult to build because of the data required as input and because the number of hazards to be considered increases exponentially with the number of types included, so the use of high discriminatory (geno)typing methods in QMRA requires a paradigm shift with respect to the current risk assessment framework which is essentially based on phenotyping information.

3. Concluding remarks

While current population genetics models and *ad hoc* adjustments of the discriminatory power of common genotyping data for *Salmonella* like MLVA seem to provide feasible options to perform source attribution, there is still a strong need to explore novel modelling options for (molecular-based) *Salmonella* source attribution. This is particularly relevant given the current "WGS revolution" that will increase the acquisition of high-throughput data in the years to come. Obviously, the use of new models and data types requires a careful evaluation of the changes observed in the trends over the years as to assess if they represent actual changes in the epidemiology of *Salmonella* or mere artefacts due to the different methods used.

It is clear that the outcome of classical case-control studies can be enhanced by incorporating source attribution data, as individuals acquiring salmonellosis from different reservoirs have different associated risk factors, which in turn indicates that salmonellas may infect humans through various transmission pathways depending on their original reservoirs. It is believed that the more such studies will be performed the better the epidemiology of *Salmonella* will be understood. In this regard, the proposed approach provides a relatively simple way to analyse and re-analyse case-control study data in light of the results of source attribution for the benefits of the outcomes of both types of analyses.

It has been shown that reparametrizing current source attribution models for *Salmonella* allows for the inclusion of non-food sources, such as reptile pets, but in principle also dogs and cats, horses, etc., and even non-animal sources like vegetables and the environment itself, provided that these data are available. This has the ability to broaden our understanding of the non-foodborne side of human salmonellosis as to tackle the disease in a more holistic way.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Barco, L., Barrucci, F., Cortini, E., Ramon, E., Olsen, J.E., Luzzi, I., Lettini, A.A., Ricci, A., 2015. Ascertaining the relationship between Salmonella Typhimurium and Salmonella 4,[5],12:i:- by MLVA and inferring the sources of human salmonellosis due to the two serovars in Italy. Front. Microbiol. 6, 301.

Barco, L., Barrucci, F., Olsen, J.E., Ricci, A., 2013. Salmonella source attribution based on microbial subtyping. Int. J. Food Microbiol. 163, 193–203.

Butler, A.J., Thomas, M.K., Pintar, K.D., 2015. Expert elicitation as a means to attribute 28 enteric pathogens to foodborne, waterborne, animal contact, and person-to-person transmission routes in Canada. Foodborne Pathog. Dis. 12, 335–344.

David, J.M., Sanders, P., Bemrah, N., Granier, S.A., Denis, M., Weill, F.X., Guillemot, D., Watier, L., 2013a. Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data. Prev. Vet. Med. 110, 12–27.

David, J.M., Guillemot, D., Bemrah, N., Thébault, A., Brisabois, A., Chemaly, M., Weill, F.X., Sanders, P., Watier, L., 2013b. The Bayesian microbial subtyping attribution model: robustness to prior information and a proposition. Risk Anal. 33, 397–408.

de Knegt, L.V., Pires, S.M., Lofstrom, C., Sorensen, G., Pedersen, K., Torpdahl, M., Nielsen, E.M., Hald, T., 2016. Application of molecular typing results in source attribution models: the case of multiple locus variable number tandem repeat analysis (MLVA) of Salmonella isolates obtained from integrated surveillance in Denmark. Risk Anal. 36, 571–588.

de Knegt, L.V., Pires, S.M., Hald, T., 2015. Attributing foodborne salmonellosis in humans to animal reservoirs in the European Union using a multi-country stochastic model. Epidemiol. Infect. 143, 1175–1186.

Domingues, A.R., Pires, S.M., Halasa, T., Hald, T., 2012. Source attribution of human salmonellosis using a meta-analysis of case-control studies of sporadic infections. Epidemiol. Infect. 140, 959–969.

Doorduyn, Y., Van Den Brandhof, W.E., Van Duynhoven, Y.T., Wannet, W.J., Van Pelt, W., 2006. Risk factors for Salmonella Enteritidis and Typhimurium (DT104 and non-DT104) infections in The Netherlands: predominant roles for raw eggs in Enteritidis and sandboxes in Typhimurium infections. Epidemiol. Infect. 134, 617–626.

Feil, E.J., 2004. Small change: keeping pace with microevolution. Nat. Rev. Microbiol. 2, 483–495.

Glass, K., Fearnley, E., Hocking, H., Raupach, J., Veitch, M., Ford, L., Kirk, M.D., 2016. Bayesian source attribution of salmonellosis in south Australia. Risk Anal. 36, 561–570.

Guo, C., Hoekstra, R.M., Schroeder, C.M., Pires, S.M., Ong, K.L., Hartnett, E., Naugle, A., Harman, J., Bennett, P., Cieslak, P., Scallan, E., Rose, B., Holt, K.G., Kissler, B., Mbandi, E., Roodsari, R., Angulo, F.J., Cole, D., 2011. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: adaptation of a Danish model. Foodborne Pathog. Dis. 8, 509–516.

Hald, T., Lo Fo Wong, D.M., Aarestrup, F.M., 2007. The attribution of human infections with antimicrobial resistant Salmonella bacteria in Denmark to sources of animal origin. Foodborne Pathog. Dis. 4, 313–326.

Hald, T., Vose, D., Wegener, H.C., Koupeev, T., 2004. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal. 24, 255–269.

Havelaar, A.H., Galindo, A.V., Kurowicka, D., Cooke, R.M., 2008. Attribution of foodborne pathogens using structured expert elicitation. Foodborne Pathog. Dis. 5, 649–659.

Hopkins, K.L., Peters, T.M., de Pinna, E., Wain, J., 2011. Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of Salmonella enterica serovar enteritidis. Euro Surveill. 16.

King, N., Lake, R., Campbell, D., 2011. Source attribution of nontyphoid salmonellosis in New Zealand using outbreak surveillance data. J. Food Prot. 74, 438–445.

Larsson, J.T., Torpdahl, M., Petersen, R.F., Sorensen, G., Lindstedt, B.A., Nielsen, E.M., 2009. Development of a new nomenclature for Salmonella typhimurium multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill. 14.

Lindstedt, B.A., Torpdahl, M., Vergnaud, G., Le Hello, S., Weill, F.X., Tietze, E., Malorny, B., Prendergast, D.M., Ni Ghallchoir, E., Lista, R.F., Schouls, L.M., Soderlund, R., Borjesson, S., Akerstrom, S., 2013. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill. 18, 20385.

Mossong, J., Mughini-Gras, L., Penny, C., Devaux, A., Olinger, C., Losch, S., Cauchie, H.M., van Pelt, W., Ragimbeau, C., 2016. Human campylobacteriosis in Luxembourg, 2010-2013: a case-control study combined with multilocus sequence typing for source attribution and risk factor analysis. Sci. Rep. 6, 20939.

Mughini-Gras, L., Barrucci, F., Smid, J.H., Graziani, C., Luzzi, I., Ricci, A., Barco, L., Rosmini, R., Havelaar, A.H., Van Pelt, W., Busani, L., 2014a. Attribution of human Salmonella infections to animal and food sources in Italy (2002-2010): adaptations of the Dutch and modified Hald source attribution models. Epidemiol. Infect. 142, 1070–1082.

Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., van Duynhoven, Y., van Pelt, W., 2014b. Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: a combined case-control and source attribution analysis. PLoS One 9, e87933.

Mughini-Gras, L., Heck, M., van Pelt, W., 2016. Increase in reptile-associated human salmonellosis and shift toward adulthood in the age groups at risk, The Netherlands, 1985 to 2014. Euro Surveill. 21.

Mughini-Gras, L., Smid, J., Enserink, R., Franz, E., Schouls, L., Heck, M., van Pelt, W., 2014c. Tracing the sources of human salmonellosis: a multi-model comparison of phenotyping and genotyping methods. Infect. Genet. Evol. 28, 251–260.

Mughini-Gras, L., van Pelt, W., 2014. Salmonella source attribution based on microbial subtyping: does including data on food consumption matter? Int. J. Food Microbiol. 191, 109–115.

Mughini Gras, L., Smid, J.H., Wagenaar, J.A., de Boer, A.G., Havelaar, A.H., Friesema, I.H., French, N.P., Busani, L., van Pelt, W., 2012. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One 7, e42599.

Mullner, P., Jones, G., Noble, A., Spencer, S.E., Hathaway, S., French, N.P., 2009a. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Anal. 29, 970–984.

Mullner, P., Spencer, S.E., Wilson, D.J., Jones, G., Noble, A.D., Midwinter, A.C., Collins-Emerson, J.M., Carter, P., Hathaway, S., French, N.P., 2009b. Assigning the source

of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. Infect. Genet. Evol. 9, 1311–1319.
Pintar, K.D., Thomas, K.M., Christidis, T., Otten, A., Nesbitt, A., Marshall, B., Pollari, F., Hurst, M., Ravel, A., 2016. A comparative exposure assessment of Campylobacter in Ontario, Canada. Risk Anal. http://dx.doi.org/10.1111/risa.12653.
Pires, S.M., 2013. Assessing the applicability of currently available methods for attributing foodborne disease to sources, including food and food commodities. Foodborne Pathog. Dis. 10, 206–213.
Pires, S.M., Evers, E.G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F.J., Havelaar, A., Hald, T., 2009. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathog. Dis. 6, 417–424.
Pires, S.M., Hald, T., 2010. Assessing the differences in public health impact of salmonella subtypes using a bayesian microbial subtyping approach for source attribution. Foodborne Pathog. Dis. 7, 143–151.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Ranta, J., Matjushin, D., Virtanen, T., Kuusi, M., Viljugrein, H., Hofshagen, M., Hakkinen, M., 2011. Bayesian temporal source attribution of foodborne zoonoses: Campylobacter in Finland and Norway. Risk Anal. 31, 1156–1171.
Sears, A., Baker, M.G., Wilson, N., Marshall, J., Muellner, P., Campbell, D.M., Lake, R.J., French, N.P., 2011. Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. Emerg. Infect. Dis. 17, 1007–1015.
Torpdahl, M., Sorensen, G., Lindstedt, B.A., Nielsen, E.M., 2007. Tandem repeat analysis for surveillance of human Salmonella Typhimurium infections. Emerg. Infect. Dis. 13, 388–395.
Tustin, J., Laberge, K., Michel, P., Reiersen, J., Dadadottir, S., Briem, H., Hardardottir, H., Kristinsson, K., Gunnarsson, E., Fridriksdottir, V., Georgsson, F., 2011. A national epidemic of campylobacteriosis in Iceland, lessons learned. Zoonoses Public Health 58, 440–447.
van Pelt, W., Mevius, D., Stoelhorst, H.G., Kovats, S., van de Giessen, A.W., Wannet, W., Duynhoven, Y.T., 2004. A large increase of Salmonella infections in 2003 in The Netherlands: hot summer or side effect of the avian influenza outbreak? Euro Surveill. 9, 17–19.
van Pelt, W., van de Giessen, A., van Leeuwen, W., Wannet, W., Henken, A., Evers, E.G., de Wit, M.A.S., van Duynhoven, Y.T.H.P., 1999. Oorsprong, omvang en kosten van humane salmonellose. Deel 1. Oorsprong van humane salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. Infect. Bull. 10, 240–243.
Vellinga, A., Van Loock, F., 2002. The dioxin crisis as experiment to determine poultry-related campylobacter enteritis. Emerg. Infect. Dis. 8, 19–22.
Vieira, A.R., Grass, J., Fedorka-Cray, P.J., Plumblee, J.R., Tate, H., Cole, D.J., 2016. Attribution of Salmonella enterica serotype Hadar infections using antimicrobial resistance data from two points in the food supply system. Epidemiol. Infect. 144, 1983–1990.
Wahlstrom, H., Andersson, Y., Plym-Forshell, L., Pires, S.M., 2011. Source attribution of human Salmonella cases in Sweden. Epidemiol. Infect. 139, 1246–1253.
Wilson, D.J., Gabriel, E., Leatherbarrow, A.J., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Fearnhead, P., Hart, C.A., Diggle, P.J., 2008. Tracing the source of campylobacteriosis. PLoS Genet. 4, e1000203.